DOI: 10.24996/ijs.2026.67.6.36

ISSN: 0067-2904

An Efficient Optimized Deep Model for Improving Gastrointestinal Diseases Classification

Ameer K. AL-Mashanji[1], Tahseen A. Wotaifie[1]*,Hasanain Ali Al Essa[1] Ali K. Abdul Raheem[2]

[1]*Software Department Presidency of the University of Babylon, University of Babylon, Babil, Iraq*

[2]*University of Warith Al-Anbiyaa, Kerbala, Iraq*

Received: 8/ 11/2024 Accepted: 24/ 6/2025 Published: 30/6/2026

Abstract

Gastrointestinal (GI) diseases are increasing day by day due to the continuous change in people's dietary habits, as these changes play a major role in several intestinal problems. Endoscopy is a medical imaging device used to detect and diagnose gastrointestinal diseases such as esophagitis and benign tumors. Manual diagnosis consumes a long time, so there is an urgent need to use computer techniques for GI disease diagnosis accurate and fast. In this study, a Bayesian optimizer-based pre-trained model architecture (BOPMA) is proposed to improve GI disease detection. The BOPMA is concerned with adapting the Bayesian optimizer in two important aspects: improving the EfficientNetV2s architecture and fine-tuning the hyperparameters of AdamW, which is the latest version currently available in the concept of backpropagation to improve the training process. In the proposed model, the Kvasir dataset of 8000 images, which includes 1,000 images for each of the eight GI disease classes, is used for training. Several transfer learning models, such as ResNet50, InceptionV3, and Xception, are adopted for comparison. To measure the quality of the model, the accuracy measure is used, whereby the accuracy of the BOPMA reached (0.9479), while the accuracy of ResNet50, InceptionV3, and Xception is (0.7995), (0.8568), and (0.9141), respectively. Additionally, the results demonstrated that the proposed model achieved better classification accuracy compared to previous studies that adopted the same dataset (Kvasir dataset). Finally, our study adopts the issuance of a prediction model that works as a doctor’s assistant to improve diagnostic accuracy within the complex domain of gastrointestinal health, opening new avenues for clinical application and contributing to better patient outcomes.

Keywords: Image Classification, Gastrointestinal Diseases, Hyperparameters, Transfer Learning, and Optimization.

الضبط التلقائي للمعلمات الفائقة باستخدام التحسين البايزي لتحسين تصنيف أمراض الجهاز الهضمي

امير علي كاظم[1]، تحسين علي حسن*[1]، حسنين علي عبيس [1] ، علي كريم عبد الرحيم[2]

[1]رئاسة جامعة بابل، جامعة بابل، بابل، العراق

[2]جامعة وارث الانبياء ، كربلاء ، العراق

---

* Email: tahseen.ali@uobabylon.edu.iq

الخلاصة

تتزايد أمراض الجهاز الهضمي يومًا بعد يوم بسبب التغيير المستمر في العادات الغذائية للأشخاص، حيث تلعب هذه التغييرات دورًا رئيسيًا في العديد من مشاكل الأمعاء. التنظير الداخلي هو جهاز تصوير طبي يستخدم للكشف عن أمراض الجهاز الهضمي وتشخيصها مثل التهاب المريء والأورام الحميدة. يستغرق التشخيص اليدوي وقتًا طويلاً، لذلك هناك حاجة ملحة لاستخدام تقنيات الكمبيوتر لتشخيص أمراض الجهاز الهضمي بدقة وسرعة. في هذه الدراسة، تم اقتراح بنية نموذج مدرب مسبقًا يعتمد على مُحسِّن بايز (BOPMA) لتحسين اكتشاف أمراض الجهاز الهضمي. يهتم BOPMA بتكييف المُحسِّن البايزي في جانبين مهمين: تحسين بنية EfficientNetV2s وضبط المعلمات الفائقة لـ AdamW، وهو أحدث إصدار متاح حاليًا في مفهوم الانتشار الخلفي لتحسين عملية التدريب. في النموذج المقترح، يتم استخدام مجموعة بيانات Kvasir المكونة من 8000 صورة، والتي تتضمن 1000 صورة لكل فئة من فئات أمراض الجهاز الهضمي الثمانية للتدريب. تم اعتماد العديد من نماذج التعلم الانتقالي مثل ResNet50 و InceptionV3 و Xception للمقارنة. لقياس جودة النموذج، تم استخدام مقياس الدقة، حيث بلغت دقة BOPMA (0.9479)، بينما بلغت دقة ResNet50 و InceptionV3 و Xception (0.7995) و(0.8568) و (0.9141) على التوالي. بالإضافة إلى ذلك، أظهرت النتائج أن النموذج المقترح حقق دقة تصنيف أفضل مقارنة بالدراسات السابقة التي اعتمدت نفس مجموعة البيانات. أخيرًا ، تتبنى دراستنا إصدار نموذج تنبؤ يعمل كمساعد للطبيب لتحسين دقة التشخيص في المجال المعقد لصحة الجهاز الهضمي ، وفتح طرق جديدة للتطبيق السريري والمساهمة في نتائج المرضى بشكل أفضل.

1. Introduction

Gastrointestinal (GI) is a long, somewhat tube-like path through which the food that a person eats passes directly after the mastication process from the mouth to the anus. In anatomy, the GI tract is usually divided into two main parts [1]. The extension from the mouth to the duodenum is referred to as the "Upper" GI tract, While the part extending from the opening of the small intestine to the anus is referred to as the "Lower" GI tract. The "Upper" GI tract in living organisms generally functions to ingest and digest food [2]. As for the enzymes and acids present in this area, they work to digest the food before it finally travels to the "Lower" GI tract. It promotes digestion in addition to the effective and primary role of the small intestine, where electrolytes, water, and nutrients are drawn up and absorbed by the large intestine tissues. The remaining material is retained in the rectum as solid waste, and the body then tries to expel it through the anus [3].

The GI tract can be exposed to various pathological conditions, which require direct intervention by specialized doctors. These include tissue infections, gastrointestinal diseases, and abnormal growth. Periodically, specialist doctors request visual examinations of the internal organs located in the cavity of the GI tract. Specialized doctors use a device called an endoscope to examine the internal parts of the GI tract [4]. The endoscope is a relatively long tube that is characterized by high flexibility in movement and is equipped with a fiber-optic camera in front. This device is inserted through the mouth or any other opening in the body. With the help of this device, the doctor can visually access the internal tissues of the GI tract and identify abnormalities to which the GI tract has been exposed [5]. Classification of medical images is a very important field for specialist doctors, as it helps them diagnose the disease accurately. Recently, deep learning (DL) models have become a powerful and effective tool for endoscopic image classification. Features are extracted directly from images using convolutional neural networks (CNNs) and have provided more accurate medical imaging classification [7].

In this work, an automatic deep learning model exploiting a Bayesian optimizer is proposed to tune neural network hyperparameters for improving the accuracy of

gastrointestinal disease image classification. Extensive analyses have been implemented on the Kvasir dataset to test the proposed model's effectiveness. It contains high-resolution endoscopic images of different GI diseases. However, current work adopts different data augmentation methods and strategies for improving the deep learning model's performance and enhancing its generalization ability. The main contributions of this study are as follows:
1- A new model, BOPMA, is proposed by integrating Bayesian optimization with EfficientNetV2s, allowing for automatic adaptation of model architecture.
2- Fine-tuning of the AdamW optimizer's hyperparameters using Bayesian optimization is implemented. This enhances model convergence and classification accuracy compared to traditional techniques.
3- Comprehensive experimental evaluation is conducted on the Kvasir dataset, demonstrating the superiority of the proposed model over widely used transfer learning architectures (ResNet50, InceptionV3, and Xception), and establishing its potential for clinical decision support in gastrointestinal diagnostics.
3- A comprehensive experimental evaluation is performed on the KVASIR dataset, demonstrating the superiority of the proposed model over widely used transfer learning architectures and ensuring its potential for clinical decision support in GI disease diagnosis.

The rest of the outline of this paper is organized as follows. Section 2 gives an overview of related work. Section 3 details the materials and methods used. Section 4 discusses the experimental results and comparative analysis. Finally, in Section 5, concluding remarks are listed.

2. Related work

Demirbaş et al. [8] developed a new architecture, called Spatial-Attention-ConvMixer (SAC), for endoscopic image classification. In addition, they refined a method called DropBlock to organize and address problems related to dealing with noise and overfitting. To evaluate the accuracy of the classification process, they used the Kvasir dataset for gastrointestinal diseases. The analytical results revealed that the proposed method achieved accuracy results of up to 93.37% and outperformed the other architectures used in the study. The other architectures achieved classification accuracy, 74.52% for Swin Transformer, 79.52% for Vanilla ViT,63.04% for MLPMixer, 92.48% for ConvMixer, 85.59% for SqueezeNet, and 87.44% for ResNet50.
Yoshiuk et al. [9] analyzed the four different transfer learning CNN architectures' performance (GoogleNet, MobileNetV2, ResNet-50, and MobileNetV3) to diagnose and detect GI infections based on the Kvasir dataset. The results demonstrated that the values of accuracy acquired for the models were 83% for MobileNet V2, 83.3% for ResNet-50, 84.2% for MobileNet V3, and 84.6% for GoogleNet.

Varalaxmi et al. [10] concluded that traditional techniques were time-consuming in the process of GI disease classification. They utilized the well-known ResNet50 architecture, a network with very accurate classification capability that saves time required for the training process. The proposed network achieved an accuracy of 88.05%.
Amin et al. [11] developed a new approach known as deep semantic segmentation to detect different types of GL infections. The proposed method adopted Deeplabv3, which represents the ResNet-50 network cornerstone and classification of the pixel regions correctly. The results reached predictive values of up to 90%, which indicates the proposed method's efficiency in accurately classifying the infections of GI.
Ahmed et al. [12] developed an architecture that uses denoising CNNs (DnCNNs) for the classification of medical images. They used DnCNN as a pre-processing method and AlexNet

(a pre-trainable network) as a model for the process of classification on the Kvasir dataset. The classification results demonstrated that the proposed DnCNN network achieved an accuracy of 90.17% and exceeded several comparable methods.

Lonseko et al. [5] proposed a deep-learning model to classify GI diseases using images collected endoscopically. They compared their model with the modern models used, and the model proved to outperform those models with an accuracy rate of 93.19%. Finally, Agrawal et al. [13] presented an architecture consisting of InceptionV3 and VGG to classify abnormalities of the GI diseases on the Kavasir dataset. They concluded that the experimental analyses of their proposed method obtained an F1_Score value of 84.7%.

3. Materials and Methods

3.1. About the Kvasir Dataset

The Kvasir dataset is a set of high-resolution images accumulated from endoscopic equipment utilized in the hospitals of Norway. The dataset contains 8,000 images of many different GI tract diseases, including infections, normal tissue, benign tumors, and ulcers [14]. In the field of GI endoscopy, the Kvasir dataset is widely recognized as a resource for developing many projects. One of the strengths of the Kvasir dataset is that it contains high-quality images. Specialized doctors in hospitals used endoscopes to obtain high-definition image data with fine details [15]. Since the Kvasir data set is available to the public, it has attracted the attention of many researchers and has been used in numerous studies and benchmarks. It comprises eight classes, involving ‘Dyed-lifted-polyps’, Ulcerative_colitis, Normal (‘pylorus’, ‘cecum’, and ‘z-line’), ‘Polyps’, ‘Esophagitis’, and ‘Dyedresection-margins’. Table 1 shows the number of images for each class [16]. Figure 1 depicts a sample of images for each class in the Kvasir dataset.

Table 1: Description of Kvasir Dataset

Label	Class	No. of images
C1	Dyedresection_margins	1000
C2	Dyed_lifted_polyps	1000
C3	Esophagitis	1000
C4	Ulcerative colitis	1000
C5	Polyps	1000
C6	Normal-Z-line	1000
C7	Normal-pylorus	1000
C8	Normal-cecum	1000

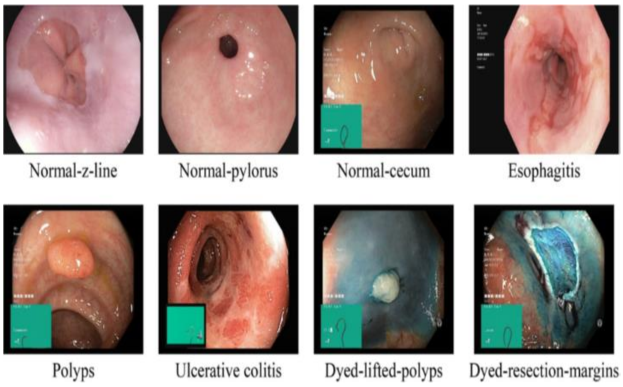

Figure 1: Image samples of the various classes

3.2 Data Augmentation

Data augmentation is a commonly employed strategy in the application of computer vision that includes several different transformations to the original image set to create new images. These transformations comprise rotation, scaling, translation, reflection, flipping, zooming, and other operations [17]. Usually, to train deep learning models effectively, this procedure requires more data to increase generalization. In this work, data augmentation is performed on the Kvasir dataset to increase the number of images for the classes and thus enhance the proposed model's generalizability [18].

3.3. Fine-Tuning Hyperparameters

Hyperparameters are a set of values that are used during the training and testing process. In general, hyperparameters include batch size, learning rate, hidden layers, number of iterations, and activation functions. The basic architecture of deep learning models consist of different layers, and thus these layers must be configured by the user before fitting them into the model [19]. The selection of hyperparameters plays a key role in improving prediction performance. Selecting hyperparameters that lead to a correct learning process for the model can achieve better classification results. Hyperparameter optimization aims to choose the values of the hyperparameter that exhibit the best results in the model testing process [20].

3.4. Bayesian Optimizer

Bayesian optimization is very useful for areas where human expertise cannot contribute to improving accuracy. The principle of Bayesian optimization is based on using prior information and updating posterior information, which helps to minimize the loss and increase the model accuracy to the maximum extent. Manual tuning is a very difficult procedure because it depends on the trial and error principle [21]. Random search and grid search are not scalable to higher dimensions and need a considerable number of training cycles. Bayesian optimization overcomes such limitations by revealing the global optimum of the function of a neural network. In addition, it scales with maximum resource usage, handles noisy data well, and exploits discontinuous spaces [21]. The following equation states that the optimization procedure is dependent on Baye’s Theorem.

$$\mathrm{P}(\mathrm{T}/\mathrm{t}) = (\mathrm{P}(\mathrm{t}/\mathrm{T})\mathrm{P}(\mathrm{T}))/\mathrm{P}(\mathrm{t}) \qquad (1)$$

Where P(t/T): refers to the ‘posterior probability’ of t given T, P(T): refers to the ‘prior probability’ of T, and P(t): refers to the ‘marginalization probability’ of t.

3.5. Transfer Learning

Transfer learning is a technique in which a model that has been pre-trained on a suitable amount of data is used to execute a similar operation or task. The process of training a new model from scratch is time-consuming and requires several procedures, such as a large amount of data, computational power, and many iterations before the model is ready for production [22]. CNNs are a neural network set intended to comprehend patterns and maps within image data. It has been used as a significant tool in several life fields, as in the field of health care. CNN played an important role in detecting and diagnosing medical images of people in the early stages of the disease [23].

3.5.1 Inception v3

Inception V3 is a developed version of the well-known network of GoogleNet, which has acceptable classification outcomes for medical images and other biomedical applications. The Inception V3 network contains a model that concatenates convolutional filters of multiple sizes into a single new filter. Inception V3 involves a block of parallel convolutional layers with different sizes of filters [24].

3.5.2 Xception

Inception Xception is an open-source CNN developed by Google. It is an abbreviation for the extended Inception network version. Xception consists of a convolutional layer that combines pointwise convolutions and deep convolutions [25]. It has been proven that this combination provides effective and high accuracy in image classification. The basic structure of the Xception network contains 36 convolutional layers, which represent the cornerstone for the extraction of features. These convolutional layers are distributed into 14 modules. The built modules include linear connections wrapped, except for the initial and last modules [25].

3.5.3 ResNet-50

ResNet-50 is the winning model of the 2015 Image Classification Challenge and serves as the backbone for several tasks in the computer vision field. The basic structure of the ResNet-50 model comprises 16 blocks, including 177 layers distributed over the input layer. The input layer is responsible for inputting images of size 224 × 224 pixels, in addition to 49 convolutional layers that utilize filters of various sizes [26]. Deep features extracted from the input image data by the convolutional layer and the maps are stored in a vector, with one layer of pooling for both maximum and average. The pooling layers reduce the feature map dimensions space. Convolutional layers follow a rectified linear activation (ReLU) function that directly passes positive output values and converts negative values to zero [26].

3.5.4 EfficientNet v2

The basic architecture of EfficientNet v2 demonstrates an improvement in performance and efficiency compared to the architecture of other traditional convolutional methods. EfficientNet version 2 represents an improvement over the base framework. It introduces variables designed for balancing performance and efficient processing of high-resolution images [27]. The EfficientNet v2 architecture includes key components: fused Mobile Vision Convolutional Network (MBConv) layer, scaling, and kernel sizing. Operational efficiency improves through Fused-MBConv layers, which combine a depth-wise convolution layer (DWConv) and a convolutional layer in the initial stages of image input. The EfficientNet v2 network is used to maintain computational efficiency and to expand the size of the model as a scaling strategy [27].

The small model (EfficientNet v2-S) general scheme is depicted in Figure 2 EfficientNet v2-s. In this figure, (a) the network architecture represents the stage of image input to the stage of logit output. The numbers on the right side of the squares indicate the layer numbers, while the numbers below the squares indicate the number of strides and corresponding channels. The blocks from b to e represent the schematic diagrams of each algorithm defined in (a); 'Fused-MBConv1(k3x3)', 'Fused-MBConv4(k3x3)', 'MBConv4(k3x3)', and 'MBConv6(k3x3)', respectively. The abbreviations that appear in the schematic are expanded as follows: MBConv as 'mobile vision convolutional network', DW as 'Depth Wise', FC as 'Full Connect' BN as 'Batch Normalization', SE as 'Squeeze and Excitation', H as 'height', W as 'width', and F as 'number of channels' [28].

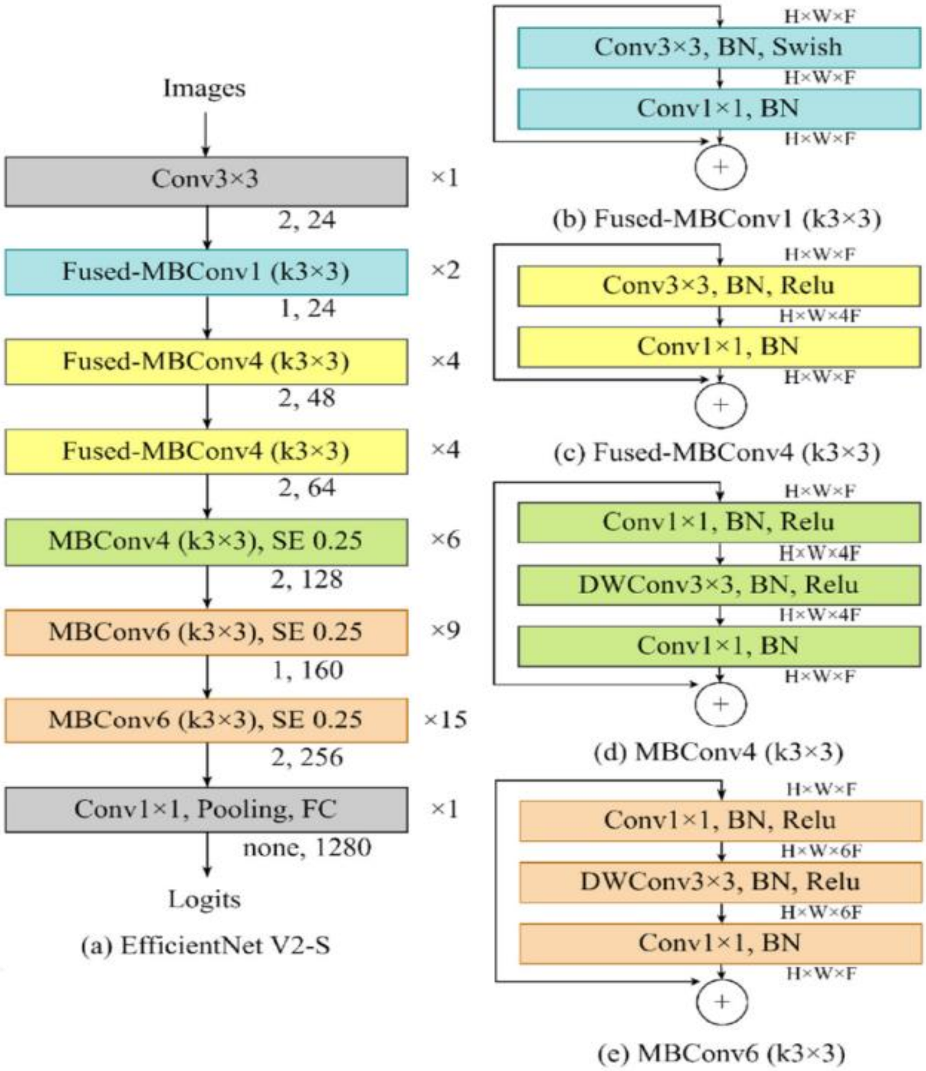

Figure 2: The Efficientnet v2-s architecture model

3.6. Evaluation metrics

Several performance metrics are employed to evaluate the performance of the classification process of the proposed approach [29].

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}} \quad (3)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}} \quad (4)$$

$$\text{F1 score} = \frac{2\text{TP}}{\text{FP} + \text{FN} + 2\text{TP}} \quad (5)$$

4. The proposed methodology

This study introduces a deep model whose structural configuration is optimized using meta-heuristic algorithms. The proposed system focuses on modifying the EfficientNetV2s architecture and adjusting its hyperparameters to improve the prediction of patients with GL disease. Our proposed model is trained based on the KVASIR dataset that contains images classified into eight categories. As shown in Figure 3, the process of classifying GI diseases according to our proposed model goes through three main stages: data preprocessing, prediction model construction, and evaluation.

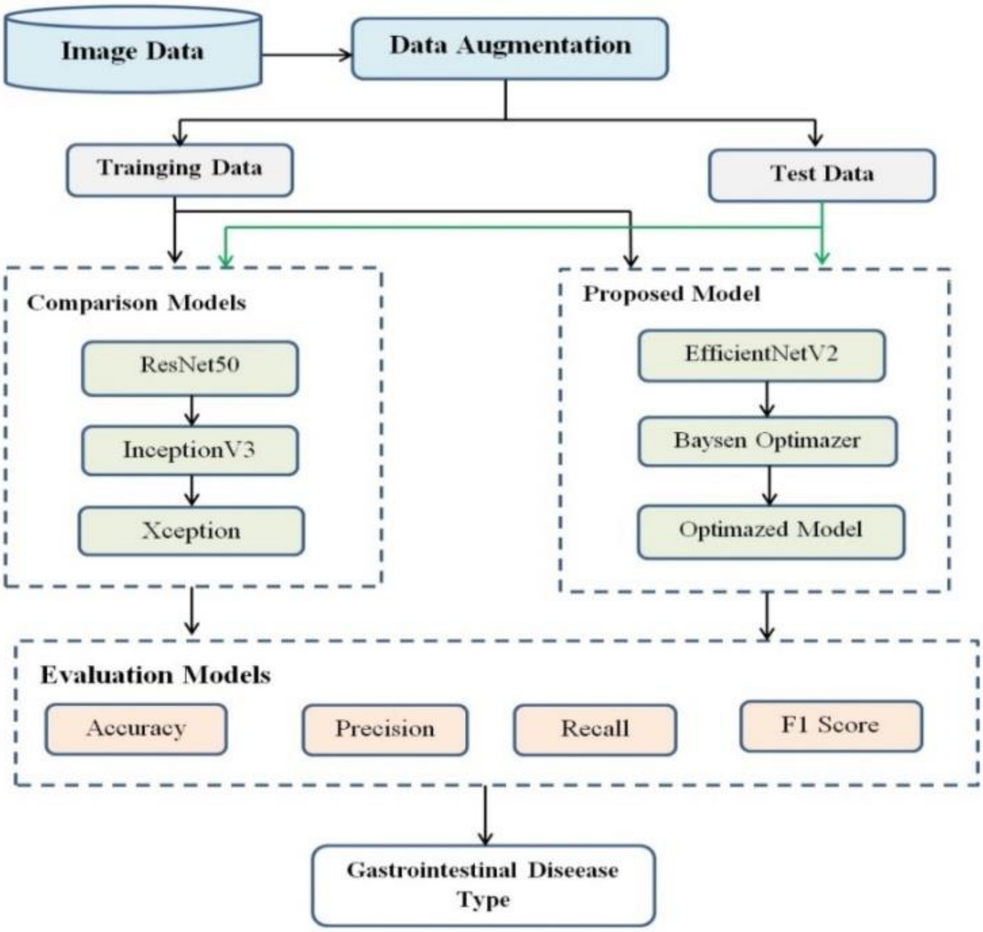

Figure 3: General steps of the proposed methodology

Firstly, data processing techniques were applied to the KVASIR dataset. One of the necessary processes is to resize images to suit pre-trained models. As such, all images in the dataset are resized to 224*224. Because deep learning models are effective and efficient whenever they are trained on a sufficient amount of data, data augmentation techniques were used in this study. Three techniques of data augmentation are applied: rotation, zooming, and cropping.

The most important part of this study is the prediction model construction. Classification of patients with GL disease is achieved through the use of one of the most important pre-trained models, which is the EfficientNetV2s. The proposed model is interested in improving the performance of the EfficientNetV2s through two important aspects: 1) Building an appropriate architecture that can be applied to the features extracted from the EfficientNetV2s in the concept of forward-propagation; 2) Adjusting the parameters of the AdamW optimizer and determining the best one within the concept of back-propagation.
Bayesian optimization, which is one of the most important meta-heuristic techniques, has been used to achieve the purpose of our proposed model. A Bayesian optimizer is efficient for fine-tuning hyperparameters in machine learning models. The basic idea is to build a probabilistic model of the objective function (e.g., accuracy) and use it to select the promising points (hyperparameter) to evaluate next. Figure 4 illustrates the process of determining optimal parameters using the Bayesian optimizer.

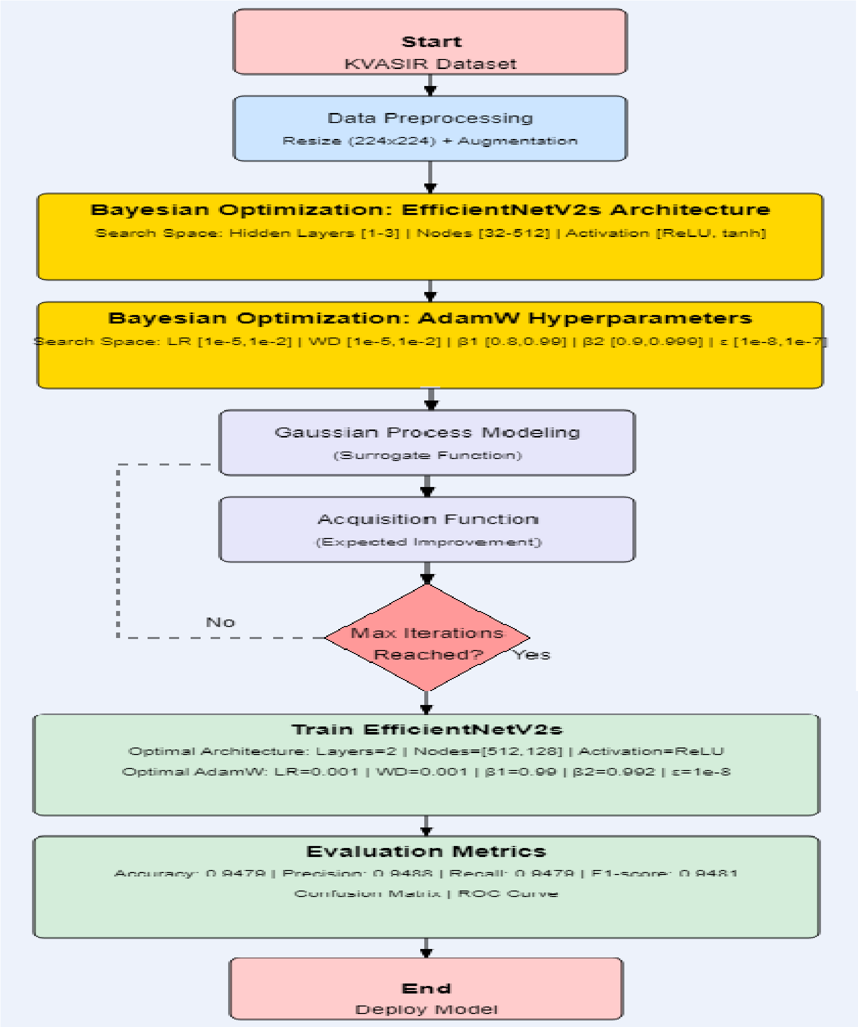

Figure 4: Steps of hyperparameter tuning

Regarding back-propagation, AdamW is the latest version currently available and has achieved promising results compared to other optimization techniques. This optimizer relies on five basic hyperparameters in the process of updating weights and bias: learning rate, weight decay, epsilon, beta_1, and beta_2. Instead of specifying these hyperparameters by the user, this study has adopted a Bayesian optimizer to determine the best values. By identifying the best values that fit the dataset, our methodology achieved the desired goal, which is speed of training or speed of convergence of the proposed model and not falling into the local minima. In forward propagation, the study focused mainly on modifying the EfficientNetV2s architecture through the same optimizer (which is the AdamW optimizer).

The EfficientNetV2s mainly aims to extract important features in the image. This pre-trained model is one of the best models currently available, so it was exploited in this study to extract features. Classification is achieved through the EfficientNetV2s by adding a fully connected layer with a softmax activation function. Instead of manually designing the architecture, a Bayesian optimizer was employed to build a multi-layer perceptron on top of the EfficientNetV2s pre-trained model. The optimizer was used to automatically identify the most suitable activation functions (tanh or ReLU) and determine the optimal number of nodes in each hidden layer. The architecture of the proposed model, based primarily on the Bayesian optimizer, is shown in Figure 5.

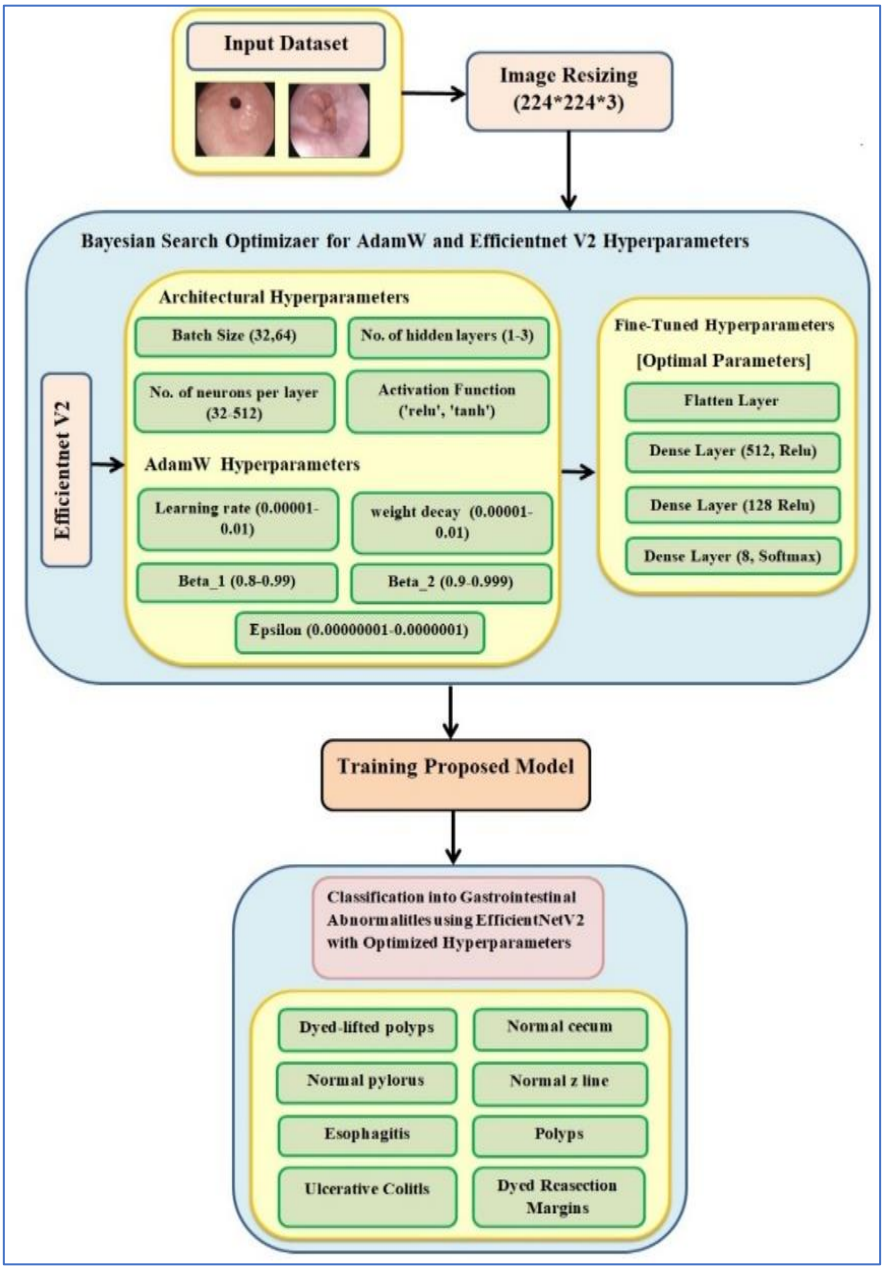

Figure 5: Proposed model architecture

As shown in Figure 5 above, our study focused on employing the Bayesian optimizer to improve the architecture of the EfficientNetV2s pre-trained model on the one hand and adjusting the optimal parameters of the AdamW optimizer on the other hand. Regarding the model architecture, the focus was placed on key structural elements, including the number of layers, the number of nodes per layer, the activation function, and the batch size. In contrast, the AdamW optimizer, which focuses on network training and weight updates, was configured using optimal values for its five hyperparameters. Finally, the efficiency of the proposed model is verified through the evaluation and comparison stages. The performance of the proposed model is evaluated through the essential metrics, namely precision, accuracy, f-measure, and recall. Both ResNet50, InceptionV3, and Xception, which are pre-trained models, were used for comparison, and our proposed model achieved better results in terms of performance metrics.

5. Experimental results and discussion

Throughout this study, the KVASIR dataset was utilized, which contains endoscopic images of the human GI. The KVASIR dataset consists of 8000 images classified into eight categories: normal-pylorus, normal-cecum, normal-z-line, dyed-lifted, polyps, ulcerative colitis, esophagitis, polyps, and dyed-resection. Three of these categories are normal (healthy), while five are abnormal (infected).

The dataset was randomly divided into 80% for training and 20% for testing. The images contained in the KVASIR dataset are of different sizes, so it is necessary to standardize these sizes. Taking into account the concept of the EfficientNetV2s pre-trained model, all images are resized to 224*224. Finally, rotation, zooming, and cropping transforms have been used to train the model on the largest number of images. This study is mainly interested in employing the Bayesian optimizer with transfer learning techniques. This optimizer is generally utilized to optimize machine learning models by determining the best hyperparameters to obtain accurate results. As such, this optimizer was used to modify the EfficientNetV2s architecture on the one hand and, on the other hand, to adjust the hyperparameters of the AdamW optimizer for achieving faster training and better accuracy.

Regarding building the appropriate architecture on the features extracted from the EfficientNetV2s pre-trained model, the study focused on four basic hyperparameters. These hyperparameters are the number of layers, the number of nodes in each layer, the batch size, and the activation function. On the other hand, which is the most important part, updating the weights and bias in backpropagation depends on the type of optimizer used. AdamW optimizer, which is based on five hyperparameters, is used as it is the latest version currently available. The hyperparameters of this optimizer are learning rate, weight decay, epsilon, beta_1, and beta_2. Accelerating deep model training mainly depends on choosing appropriate values for these hyperparameters. Therefore, this study used the Bayesian method and adapted it to the EfficientNetV2s pre-trained model to determine and adjust the best values of AdamW hyperparameters. Table 2 shows the target hyperparameters and the possible values of these parameters, which AdamW depends on, as well as the best values.

Table 2: Hyperparameter tuning for the proposed model based on the Bayesian optimization method

Hyperparameter	Search space	Optimal hyperparameter values
Number of hidden layers	[1-3]	2
Number of nodes/neurons per layer	[32-512]	[512,128]
Activation function	['relu', 'tanh']	'relu'
Batch size	[32,64]	64
Learning rate	[0.00001- 0.01]	0.001
Weight decay	[0.00001- 0.01]	0.001
Beta_1	[0.8-0.99]	0.99
Beta_2	[0.9-0.999]	0.992
Epsilon	[0.00000001-0.0000001]	0.00000001
Number of hidden layers	[1-3]	2
Number of nodes/neurons per layer	[32-512]	[512,128]

Based on the best hyperparameters shown in Table 2 above, the complementary architecture of the EfficientNetV2s pre-trained model was built. Then, adjust the AdamW optimizer parameters to train the model and update the weights and bias. Training the model depends on a hyperparameter called an epoch. The challenge of determining the optimal value for this hyperparameter is addressed by using the concept of early stopping.

On the other hand, the early stopping technique helps the model not fall into the problem of overfitting. Therefore, early stopping with 5 patience was used in the proposed model. This procedure led to setting the epoch parameter to 10. Three pre-trained models have been trained on the KVASIR dataset to compare with the results of our proposed model. These models are ResNet50, InceptionV3, and Xception. As shown in Table 3, the performance of our proposed model outperformed all of these models in terms of all evaluation metrics.

Table 3: Hyperparameter tuning for the proposed model based on the Bayesian optimization method

Model	Accuracy	Precision	Recall	F1-score
ResNet50	0.7995	0.8251	0.7995	0.7868
InceptionV3	0.8568	0.8698	0.8568	0.8520
Xception	0.9141	0.9191	0.9141	0.9141
Proposed Model (BOPMA)	**0.9479**	**0.9488**	**0.9479**	**0.9481**

By looking at Table 4, it was found that the proposed model outperformed the compared models and the previous study that used the KVASIR dataset. Moreover, the improvement ratio in accuracy achieved through the proposed models is 94.79 %, 91.41%, 85.68%, and 79.95% in comparison with the Xception, InceptionV3, and ResNet50, respectively. To display the results of the proposed model visually, Figure 6 and Figure 7, explain the confusion matrix and receiver operating characteristic (ROC) curve, respectively.

The superior performance of the BOPMA model, demonstrated by its highest accuracy score of 94.79%, indicates a substantial improvement in the automated detection of gastrointestinal diseases. This high level of precision can reduce diagnostic errors and support clinicians in making accurate decisions. By minimizing reliance on time-consuming manual interpretation, the proposed model enhances the efficiency of endoscopic examinations and contributes to more timely treatment interventions, ultimately improving patient care and outcomes in clinical practice.

Table 4: Hyperparameter tuning for the proposed model based on the Bayesian optimization method

Author	Methods	Acc (%)
Abhishek. et al, [30]	FocalConvNet	63.73
Roger. et al, [31]	Multi Model Classification(MMC)	90.20
Qilong. et al, [32]	Efficient Channel Attention (ECA)	92.81
Xiaoqi. et al, [33]	Transfer Learning Framework(TLF)	93.00
Zenebe. et al, [5]	Deep CNN Layers based SAM model	93.19
Ayyüce. et al,[8]	Spatial- Attention ConvMixer (SAC)	93.37
Ours	**BOPMA**	**94.79**

The results presented in Table 4 demonstrate the superior performance of our model, BOPMA, which achieved an accuracy of 94.79%, significantly outperforming all previous studies. The closest competitor, the Spatial-Attention ConvMixer (SAC) model by Ayyüce et al., achieved an accuracy of 93.37%, resulting in an improvement ratio of 1.63. Compared to the least performing model, FocalConvNet by Abhishek et al. (63.73%), the improvement ratio is 48.72%. These results ensure the effectiveness of integrating the Bayesian optimizer with both architectural optimization and hyperparameter tuning, allowing the model to improve automated classification of gastrointestinal diseases.

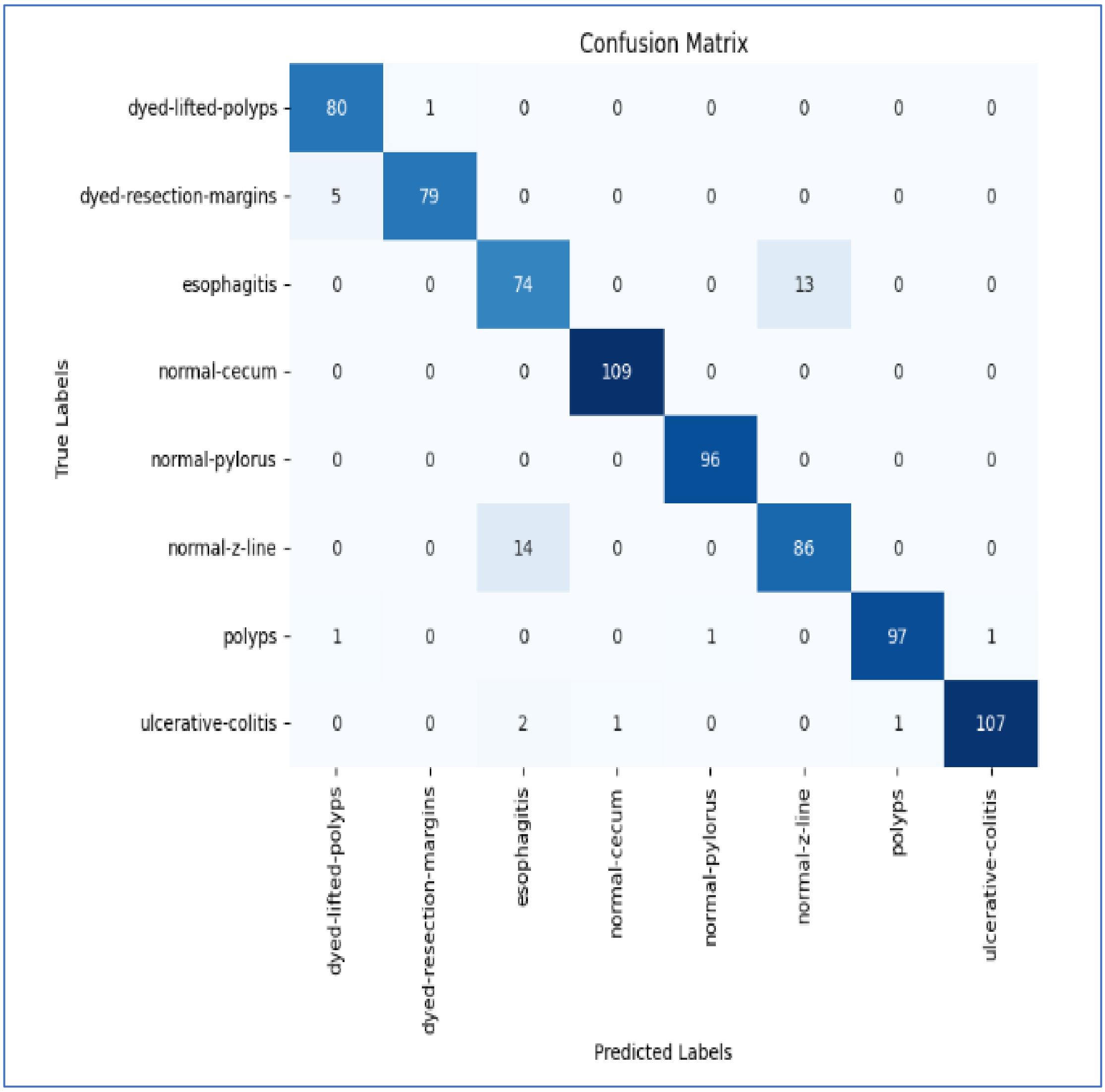

Figure 6: Confusion matrix of proposed model

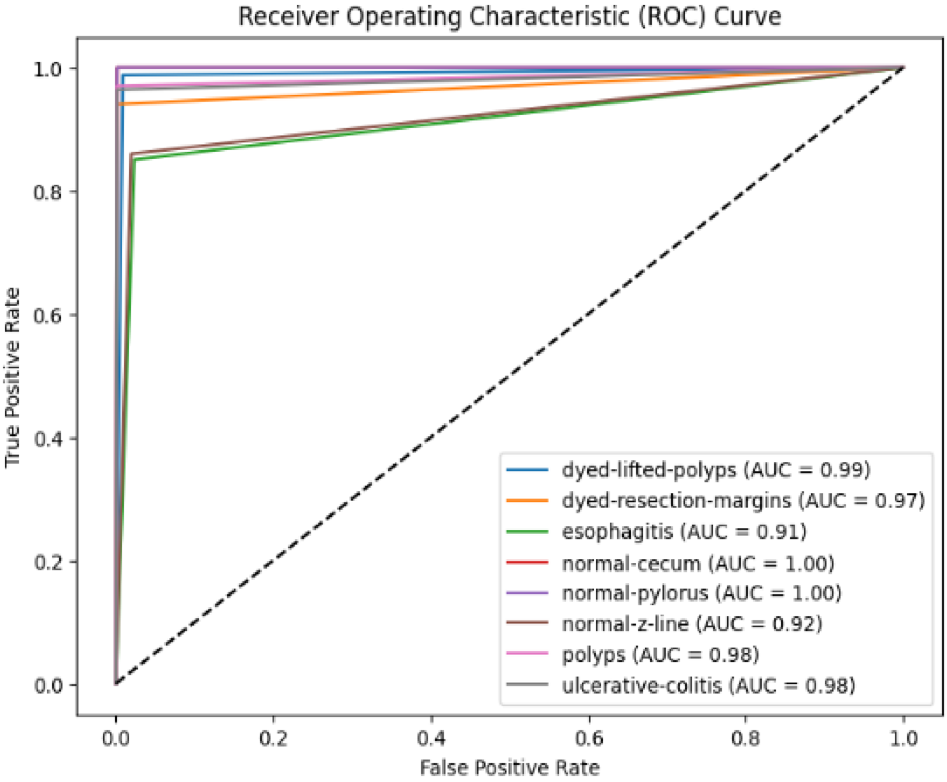

Figure 7: ROC curve of the proposed model

Despite the promising results, the BOPMA model has some limitations. The model was trained and tested only on the Kvasir dataset, which may limit its generalizability to other gastrointestinal datasets. Future work will focus on training and evaluating the model's scalability across diverse datasets and clinical environments, as well as extending its application to other medical conditions beyond gastrointestinal diseases to ensure broader clinical utility.

6. Conclusion

This paper provides a reliable framework for classifying GI disorders in the public Kvasir dataset. By aiding early detection, deep learning algorithms may reduce the likelihood of malignant diseases . As such, a new model was proposed that is capable of improving the accuracy of GI disease detection on the one hand and is fast in terms of convergence on the other hand. These deep learning models help guide the focus of doctors and highlight the most important endoscopic image parts that they may have missed. Bayesian search optimization is adopted to obtain the best hyperparameters for the proposed model architecture and thus improve the results of the detection process. Compared to previous studies and three transfer learning models, which are ResNet50, InceptionV3, and Xception, our proposed model achieved better results in terms of evaluation metrics. In future work, it is possible to build hybrid models such that the decision process of infection or non-infection depends on more than one pre-trained model. Also, using other optimizers from the family of met-heuristic algorithms and adapting them to other transfer learning models, as well as training these models on a variety of data to increase the possibility of generalization.

References

[1] M. Sharif, M. Attique Khan, M. Rashid, M.Yasmin, F. Afza , and U. J. Tanik, “Deep CNN and geometric features-based GI tract diseases detection and classification from wireless capsule endoscopy images", Journal of Experimental & Theoretical Artificial Intelligence, Vol. 33, No. 4, pp. 577-599, 2021. Doi : 10. 1080/ 09528 13X. 2019. 15726 57

[2] A. K. AL-Mashanji, and S. Z. AL-Rashid, "Predicting with the quantify intensities of transcription factor-target genes binding using random forest technique", International Journal of Nonlinear Analysis and Applications, Vol. 12, No. 2, pp. 145-161, 2021. doi: 10. 1016/ B978-0-12-822260-7. 00001-7

[3] I. Iqbal, K. Walayat, M. U. Kakar, and J. Ma, "Automated identification of human GI tract abnormalities based on deep convolutional neural network with endoscopic images", Intelligent Systems with Applications, Vol. 16, pp.200149, 2022. doi: 10.1016/j. iswa. 2022. 200149

[4] T. Cogan, M. Cogan, and L. Tamil, "MAPGI: Accurate identification of anatomical landmarks and diseased tissue in GI tract using deep learning", Computers in biology and medicine, Vol. 111, pp.103351, 2019. doi: 10. 1016/j. compb iomed. 2019.103351

[5] Z. M. Lonseko, P. E. Adjei, W. Du, C. Luo, D. Hu, L. Zhu, and N. Rao, "GI disease classification in endoscopic images using attention-guided convolutional neural networks", Applied Sciences, Vol. 11, No. 23, pp.11136, 2021. doi: 10. 3390/ app11 23111 36

[6] M. A. Khan, M. A. Khan, F. Ahmed, M. Mittal, L. M. Goyal, D. J. Hemanth, and S. C. Satapathy, "GI diseases segmentation and classification based on duo-deep architectures", Pattern Recognition Letters, Vol. 131, pp. 193-204, 2020. doi: 10.1016/j.patrec.2019.12.024

[7] H. Borgli, V. Thambawita, P. H. Smedsrud, S. Hicks, D. Jha, S. L. Eskeland, and T. de Lange, "HyperKvasir, a comprehensive multi-class image and video dataset for GI endoscopy", Scientific data, Vol. 7, No. 1, pp. 1–14, 2020. doi: 10.1038/s41597-020-00622-y

[8] A. A. Demirbaş, H. Üzen, and H . Fırat, "Spatial-attention ConvMixer architecture for classification and detection of GI diseases using the Kvasir dataset", Health Information Science and Systems, Vol. 12, No. 1, pp.32, 2024. doi: 10.1007/s13755-024-00290-x

[9] K., Yoshiok, K., Tanioka, S., Hiwa, and T. Hiroyasu,"Deep-learning models in medical image analysis: Detection of esophagitis from the Kvasir Dataset", arxiv. org/ abs/ 2301. 02390. arXiv preprint arXiv:2301.02390

[10] G.Varalaxmi, S. R. Baddam, E. S. Yalamarthi, K. Swaraja, K. R. Madhavi, and C. Sujatha, "Diagnosis of GI diseases using modern CNN techniques," In 2023 IEEE 8th International Conference for Convergence in Technology (I2CT) ,IEEE, pp. 1-6, 2023. doi: 10.1109/I2CT57861.2023.10126259

[11] J. Amin, M. Sharif, E. Gul, and R. S. Nayak, "3D-semantic segmentation and classification of stomach infections using uncertainty aware deep neural networks", Complex & Intelligent Systems, Vol. 8, No. 4, pp. 3041-3057, 2022. doi: 10.1007/s40747-021-00328-7

[12] A. Ahmed, "Classification of GI images based on transfer learning and denoising convolutional neural networks", In Proceedings of International Conference on Data Science and Applications: ICDSA, Springer Singapore, Vol. 1 , pp. 631-639, 2022. doi: 10.1007/978-981-16-5120-5_48

[13] T. Agrawal, R. Gupta, S. Sahu, and C. Y . Espy-Wilson, " SCL-UMD at the Medico Task-MediaEval 2017: Transfer Learning based Classification of Medical Images", MediaEval, Vol. 17, pp. 13-15, 2017.

[14] M.Ramzan, M. Raza, M. Sharif, M. A. Khan, and Y. Nam. "GI tract infections classification using deep learning", Comput. Mater. Contin, Vol. 69, pp. 3239-3257, 2021. doi : 10.32604/cmc.2021.015920

[15] M. A. Khan, N. Sahar, W. Z. Khan, M. Alhaisoni, U. Tariq, M. H. Zayyan, and B. Chang, "GestroNet: a framework of saliency estimation and optimal deep learning features based GI diseases detection and classification", Diagnostics, Vol. 12, No. 11,pp. 2718, 2022. doi: 10.3390/diagnostics12112718

[16] J. V. Thomas Abraham, A. Muralidhar, K. Sathyarajasekaran, and N. Ilakiyaselvan, "A deep-learning approach for identifying and classifying digestive diseases", Symmetry, vol.15, no.2, pp. 379, 2023. doi: 10.3390/sym15020379

[17] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications", Journal of Medical Imaging and Radiation Oncology, Vol. 65, No. 5, pp. 545-563, 2021. doi:10.1111/1754-9485.13261

[18] A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan,“Deep learning approaches for data augmentation in medical imaging: a review”, Journal of Imaging, Vol. 9, No. 4, pp. 81, 2023. doi: 10.3390/jimaging9040081

[19] S. A. El-Ghany, M. Azad, and M. Elmogy, “Robustness Fine-Tuning Deep Learning Model for Cancers Diagnosis Based on Histopathology Image Analysis”, Diagnostics, Vol. 13, No. 4,pp. 699, 2023. doi: 10.3390/diagnostics13040699

[20] K.Shankar, S. Kumar, A. K. Dutta, A. Alkhayyat, A. J. A. M. Jawad, A. H. Abbas, and Y. K. Yousif, “An automated hyperparameter tuning recurrent neural network model for fruit classification”, Mathematics, vol.10, No. 13, pp. 2358, 2022. doi: 10.3390/math10132358

[21] A. H. Victoria, and G. Maragatham,“Automatic tuning of hyperparameters using Bayesian optimization”, Evolving Systems, Vol. 12, No. 1, pp. 217-223, 2021. doi: 10.1007/s12530-020-09345-2

[22] T. A. Wotaifi, and B. N. Dhannoon,“ Developed Models Based on Transfer Learning for Improving Fake News Predictions”, JUCS: Journal of Universal Computer Science,Vol. 29, No. 5, 2023. doi: 10.3897/jucs.94081

[23] T. A. Wotaifi and B. N. Dhannoon, “Attention Mechanism Based on a Pre-trained Model for Improving Arabic Fake News Predictions,” Iraqi Journal of Science, pp. 6041–6054, Nov. 2023, doi: 10.24996/ijs.2023.64.11.45.

[24] M. Mujahid, F. Rustam, R. Álvarez, J. Luis Vidal Mazón, I. D. L. T. Díez, and I ,“ Ashraf, Pneumonia classification from X-ray images with inception-V3 and convolutional neural network”, Diagnostics, Vol. 12, No. 5, pp. 1280, 2022. doi: 10.3390/diagnostics12051280

[25] J. O.Carnagie, A. R. Prabowo, E. P. Budiana, and I. K. Singgih,“ Essential oil plants image classification using xception model ”, Procedia Computer Science, Vol. 204, pp.395-402, 2022. doi: 10.1016/j.procs.2022.08.048

[26] L. V. Fulton, D. Dolezel, J. Harrop, Y. Yan, and C. P. Fulton,“ Classification of Alzheimer’s disease with and without imagery using gradient boosted machines and ResNet-50”, Brain sciences,Vol. 9,No. 9, pp. 212, 2019. doi: 10.3390/brainsci9090212

[27] L. Deng, H. Suo, and D. Li.“ Deepfake Video Detection Based on EfficientNet-V2 Network”,Computational Intelligence and Neuroscience, No. 1,pp. 3441549, 2022. doi: 10.1155/2022/3441549

[28] B. Li, B. Liu, S. Li, and H. Liu. “An improved EfficientNet for rice germ integrity classification and recognition,” Agriculture,Vol. 12,No. 6, pp. 863, 2022.

[29] T. A. Wotaifi and E. S. Al-Shamery, “Modified random forest based graduates earning of higher education mining,” International Journal of Computer Information Systems and Industrial Management Applications., vol. 12, no. 3, pp. 56–65, Mar. 2020.

[30] A. Srivastava, N. K. Tomar, U. Bagci, and D. Jha,“ Video capsule endoscopy classification using focal modulation guided convolutional neural network,” In 2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS), IEEE, pp.323-328, 2022. doi: 10.1109/CBMS55023.2022.00064

[31] R. Fonolla, F. van der Sommen, R. M. Schreuder, E. J. Schoon, and P. H. de With , “Multi-modal classification of polyp malignancy using CNN features with balanced class augmentation,” In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE, pp. 74-78, 2019. doi: 10.1109/ISBI.2019.8759320

[32] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu,“ ECA-Net: Efficient channel attention for deep convolutional neural networks”, In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition , IEEE, pp. 11534-11542, 2020.

[33] X., Liu, C. Wang, J. Bai, and G. Liao, “Fine-tuning pre-trained convolutional neural networks for gastric precancerous disease classification on magnification narrow-band imaging images”, Neurocomputing, Vol. 392, pp. 253-267, 2020. doi: 10.1016/j.neucom.2018.10.100