NATURENGS
E-ISSN 2717-8013
https://dergipark.org.tr/en/pub/naturengs

Automatic Detection of Gastrointestinal Diseases Using Wireless Capsule Endoscopy Images

Harun Bingol [a,*], Muhammed Yildirim [b]

[a] Malatya Turgut Ozal University, Department of Software Engineering, Malatya Turkey – 44210
[b] Malatya Turgut Ozal University, Department of Computer Engineering, Malatya Turkey - 44210

*Corresponding author

ARTICLE INFO

Received 15.10.2024
Accepted 26.11.2024

Doi: 10.46572/naturengs.1567342

ABSTRACT

Gastrointestinal (GI) diseases are various disorders related to the digestive system. This system includes the esophagus, stomach, small and large intestines, liver, gallbladder and pancreas, starting from the mouth. Early diagnosis is very important in the treatment of the disease. The earlier the disease is diagnosed, the higher the chance of the patient being treated. In recent years, it is known that artificial intelligence techniques have been widely used in disease diagnosis and classification. Among the artificial intelligence techniques, deep learning methods that produce very successful results in image classification are frequently used. This success of deep learning methods has been tried to be used in the classification of GI diseases. Within the scope of this study, it was tried to detect bleeding GI or lesions from publicly available wireless capsule endoscopy (WCE) images. As a result of the experiments, 5 different deep learning architectures were used. Features were extracted from the two architectures that showed the highest accuracy and were combined. Neighborhood Component Analysis (NCA) dimension reduction method was applied to the obtained feature map and a hybrid model was obtained. It was seen that the proposed hybrid model achieved an accuracy value of 86.3%.

Keywords: Gastrointestinal bleeding, Artificial intelligence, Deep learning, Classification

1. Introduction

The gastrointestinal (GI) system can be called the route that nutrients taken into the body pass through until they are excreted. This route consists of the mouth, esophagus, stomach, small and large intestines, and finally the anus. The fact that there are so many different organs on the route greatly increases the importance of the GI system. GI bleeding can originate from the upper GI system as well as the lower GI system. In some cases, patients may not be aware of GI bleeding. In some cases, the exact location of the bleeding may not be determined by experts [1]. For these reasons, if a disease or bleeding in the GI system is not treated early, its consequences can be hazardous [2].

As it is known, early diagnosis of diseases can allow the patient to start treatment early. Early treatment increases the likelihood of success of the treatment. For this reason, computer-aided systems that will make it easier for doctors to make correct diagnoses under their intense workload have become quite popular in recent years.

Computer-aided systems are very important in terms of both reducing the workload of the doctor and preventing human errors. These systems can help diagnose. They can be very useful, especially in rural areas where there are no physicians or in places where the workload is extremely high and there are not enough physicians. Convolutional neural networks (CNN), which have proven their success in image classification among computer-aided systems and are widely used in biomedical images, were used in this study to classify GI bleeding and lesion images [3]. In this study, GI bleeding and lesion images were classified with 5 different CNN architectures. Very competitive results were obtained.

The contribution and innovation of the study to the literature;

1. In this study, a deep learning-based hybrid method was developed to classify GI bleeding and lesion images.

2. The proposed hybrid model was compared with 5 different deep architectures accepted in the literature and it was seen that it achieved highly competitive results.

3. In the proposed model, GoogleNet and ResNet50 architectures, which gave the highest accuracy value among the pre-trained architectures, were used for feature extraction.

4. For different features of the same image to be used together, the features extracted from GoogleNet and ResNet50 architectures were combined.

5. For the proposed model to achieve competitive results, feature selection was made with the NCA

* Corresponding author. e-mail address: harun.bingol@ozal.edu.tr
ORCID : 0000-0002-6300-9016

method from the combined feature map. The number of features, which was 2000 for each image after the merging process, was reduced to 52 after the NCA method [4]. This enabled the proposed model to work faster and more effectively.

6. Finally, the optimized feature map was classified in the SVM classifier and a competitive result of 86.3% was obtained.

The flow of this study is as follows. In the second section, the data set used in the study and the proposed method were examined. In the third section, the Experimental result section was included, and in the fourth section, the study was completed with the Conclusion section.

2. Material and Methods

2.1. Dataset

Capsule endoscopy (CE) is a technology that helps specialists diagnose bleeding and polyps in the system by imaging the entire digestive system [5]. This project was obtained from the Mendeley dataset, which includes publicly available wireless capsule endoscopy (WCE) images of GI bleeding and normal patients [6,7]. The dataset contains a total of 226 WCE images, 113 normal, and 113 bleeding or lesion images. The images were collected in a hospital in Pakistan.

2.2. Proposed Model

Within the scope of this project, ESA architectures Alexnet [8,9], Googlenet [10], Inceptionv3 [11], Mobilenetv2 [12] and Resnet50 [13] were used for feature extraction in the detection of GI bleeding images. Alexnet architecture consists of 25 layers. It receives the incoming image in 227x227 size. It uses softmax as the classification function. Googlenet architecture has a depth of 22 layers. It accepts the incoming image in 224x224 size. Resnet50 architecture uses residual blocks in order to eliminate the gradient problem. It accepts the incoming image in 224x224 size just like Googlenet architecture. Inceptionv3 architecture is the addition of cluster normalization and fully connected layers as auxiliary classifiers together with convolution layers to Inceptionv2 architecture. Inceptionv3 architecture accepts the incoming image in 299x299 size. Mobilenetv2 architecture is an ESA architecture that produces quite good results even on devices with small hardware specifications. It accepts the incoming image as 224x224. Since the highest performance rates among the five models are achieved in GoogleNet and ResNet50 architectures, these two models are preferred for feature extraction in the proposed model. The diagram of the proposed model is given in Figure 1

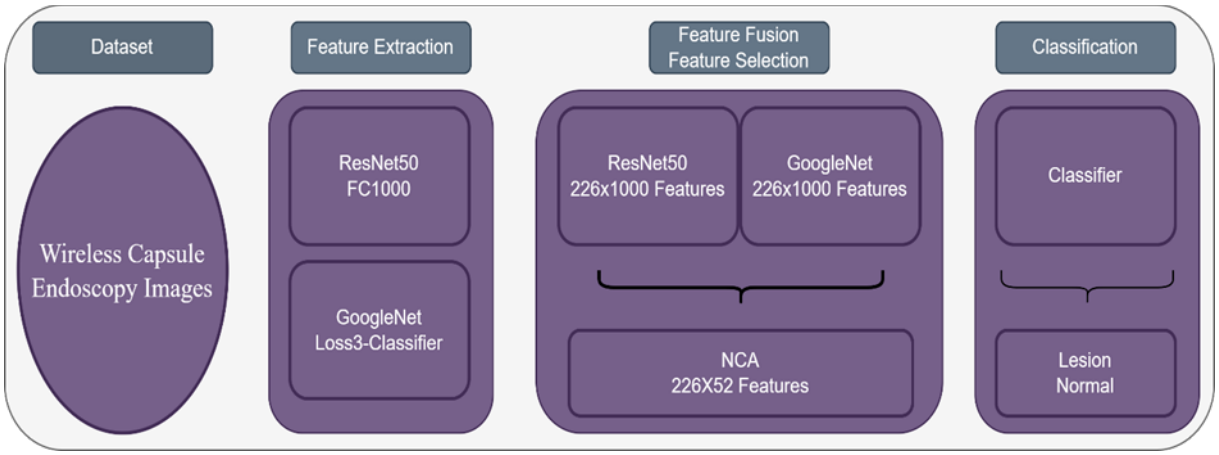

Figure 1. Flowchart of the proposed model

The size of the feature maps obtained with GoogleNet and ResNet50 architectures is 226x1000 in each architecture. The size of the feature map obtained after the merging process to use different features of the same image together is 226x2000. The features were obtained from the “loss3-classifier” layer in the GoogleNet architecture and the “FC1000” layer in the ResNet50 architecture. Then, feature selection was performed with the NCA method from the merged feature map and as a result, an optimized feature map of 226x52 size was obtained. Finally, the optimized feature map was classified in the SVM classifier and competitive results were obtained.

3. Experimental Results and Discussion

The dataset used in the experiments contains a total of 2 classes and 226 images. All architectures were run under equal conditions in the experiments. Each deep architecture was run for 20 Epochs. The MiniBatchSize value was determined as 8, the InitialLearnRate value as 1e-4, and the ValidationFrequency value as 18. The experiments were performed on a laptop computer with an i5 processor and 16 GB RAM. During the experiments, the dataset was divided into 30% test and 70% training data. To measure the performance of the models in the study, Sensitivity, Specificity, Negative Predictive Value (NPV), False Positive Rate (FPR), False Discovery Rate (FDR), False Negative Rate

(FNR), and Matthews Correlation Coefficient (MCC) metrics were used. The confusion matrix of the Alexnet architecture is in Figure 2.

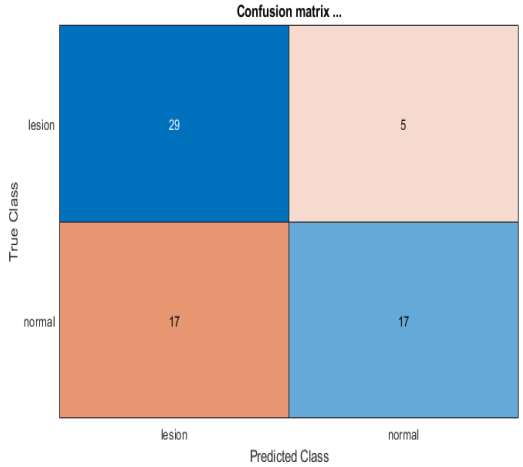

Figure 2. Alexnet mimarisine ait karmaşıklık matrisi

Alexnet predicted 29 of 34 Lession images correctly and 5 incorrectly. Similarly, in the Normal class, it predicted 34 test images halfway through the arc. As a result, 46 of 68 test images were obtained correctly. The accuracy value of Alexnet architecture was 67.65%. The performance metrics of Alexnet architecture are given in Table 1.

Table 1. Alexnet mimarisinin başarım metrikleri (%)

Sensitivity	63.04
Specificity	77.27
NPV	50.00
FPR	22.73
FDR	14.71
FNR	36.96
MCC	37.72

Another architecture used in the study is the GoogleNet architecture. The confusion matrix obtained in this architecture is presented in Figure 3.

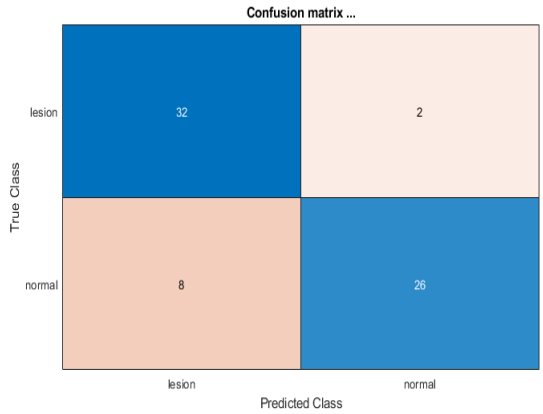

Figure 3. Googlenet mimarisine ait karmaşıklık matrisi

Googlenet architecture predicted 32 out of 34 Lession images correctly and 2 incorrectly. Similarly, in the Normal class, it predicted 26 out of 34 test images correctly. As a result, 58 out of 68 test images were obtained correctly. The accuracy value of Googlenet architecture was 85.29%. The performance metrics of Googlenet architecture are given in Table 2.

Table 2. Googlenet mimarisinin başarım metrikleri (%)

Sensitivity	80
Specificity	92.86
NPV	76.47
FPR	7.14
FDR	5.88
FNR	20
MCC	71.71

The confusion matrix obtained from the Inceptionv3 architecture used in the study is also presented in Figure 4.

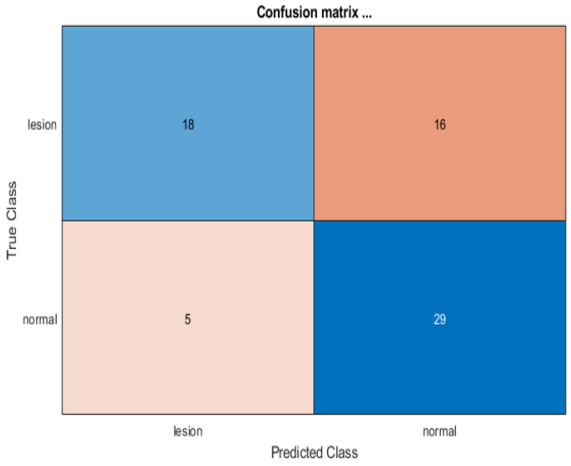

Figure 4. Confusion matrix of Inceptionv3 architecture

Inceptionv3 architecture predicted 18 of 34 Lession images correctly and 16 incorrectly. Similarly, in the Normal class, it predicted 29 of 34 test images correctly and classified 5 incorrectly. As a result, 47 of 68 test images were obtained correctly. The accuracy value of Inceptionv3 architecture was 69.12%. The performance metrics of Inceptionv3 architecture are given in Table 3.

Table 3. Inceptionv3 mimarisinin başarım metrikleri (%)

Sensitivity	78.26
Specificity	64.44
NPV	85.29
FPR	35.56
FDR	47.06
FNR	21.74
MCC	40.41

Another architecture used in the study is the Mobilenetv2 architecture. The confusion matrix obtained in this architecture is presented in Figure 5.

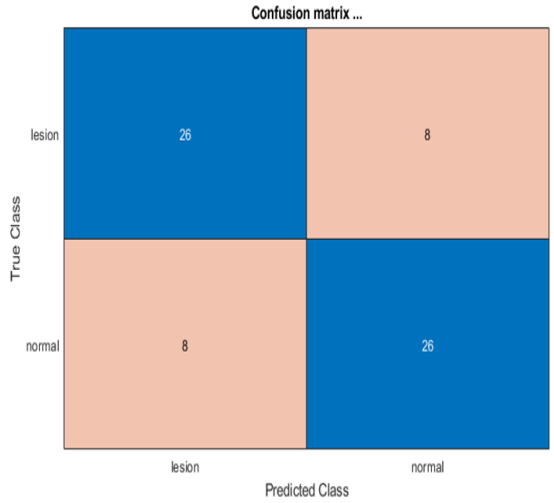

Figure 5. Confusion matrix of Mobilenetv2 architecture

Mobilenetv2 architecture predicted 26 of 34 images correctly and 8 incorrectly in both normal and lesion classes. As a result, 52 of 68 test images were correctly predicted and 16 incorrectly. The accuracy value of Mobilenetv2 architecture was 76.47%. The performance metrics of Mobilenetv2 architecture are given in Table 4.

Table 4. Performance metrics of Mobilenetv2 architecture (%)

Sensitivity	76.47
Specificity	76.47
NPV	76.47
FPR	23.53
FDR	23.53
FNR	23.53
MCC	52.94

Another architecture used in the study is the Resnet50 architecture. The confusion matrix obtained in this architecture is presented in Figure 6.

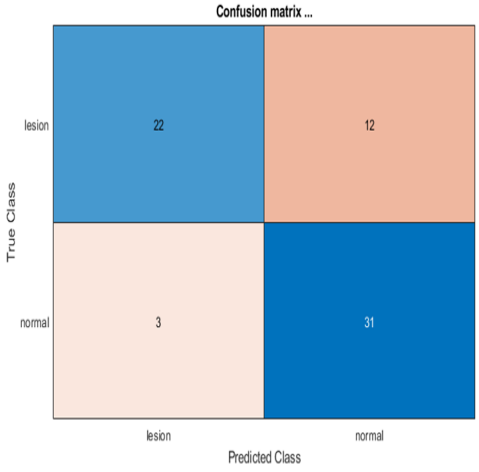

Figure 6. Confusion matrix of Resnet50 architecture

Resnet50 architecture correctly predicted 22 and incorrectly predicted 12 images out of 34 in the lesion class. In the normal class, it correctly classified 31 images. As a result, 53 of the 68 test images were correctly predicted and 15 were incorrect. The accuracy value of Resnet50 architecture was 77.94%. The performance metrics of Resnet50 architecture are given in Table 5.

Table 5. Performance metrics of Resnet50 architecture (%)

Sensitivity	88
Specificity	72.09
NPV	91.18
FPR	27.91
FDR	35.29
FNR	12
MCC	57.95

As can be seen, Alexnet, Googlenet, Inceptionv3, Mobilenetv2 and Resnet50 correctly classified 46, 58, 47, 52 and 53 images in 68 test images, respectively. Thus, it was seen that Googlenet architecture achieved the highest performance with an accuracy value of 85.29%. The second highest performance belongs to Resnet50 architecture. The confusion matrix of the hybrid model proposed within the scope of this study is given in Figure 7. In Figure 7, lesion and normal classes are expressed as 1 and 2, respectively.

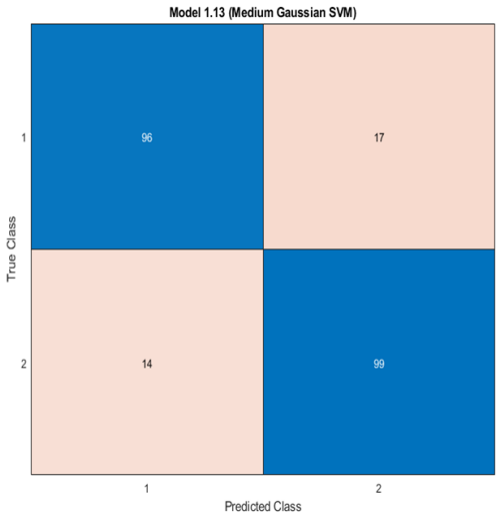

Figure 7. Confusion matrix of the proposed model

The proposed model correctly predicted 96 images out of 113 in the lesion class and incorrectly predicted 17. In the normal class, it correctly classified 99 images. As a result, it correctly predicted 195 images out of 226 and incorrectly predicted 31. The accuracy value of the proposed model was 86.30%. The performance metrics of the proposed model are given in Table 6.

Table 6. Performance metrics of the proposed model (%)

Sensitivity	87.27
Specificity	85.34
NPV	87.61
FPR	14.66
FDR	15.04
FNR	12.73
MCC	72.59

As a result, when the results obtained in the study are examined, it is seen that the proposed model achieves competitive results. One of the biggest limitations of the study is the small number of data and the collection of data from a single center. In addition, testing the proposed model on a single data set is another limitation of the study.

4. Experimental Results and Discussion

GI bleeding can occur anywhere along the digestive tract, from the mouth to the anus, and is typically classified based on the site of bleeding (upper GI or lower GI) and the severity of blood loss. It is a known fact that GI bleeding and lesions have many aspects that negatively affect human life. When the disease is not treated and progresses, it can cause other diseases. In the advanced period, the treatment of the disease becomes difficult and affects the daily life of the patient extremely. It is aimed to use ESA architectures, which have been widely used in the classification of biomedical images in recent years, within the scope of this study. WCE images were examined with ESA architecture for the purpose of GI bleeding and lesion detection. In this context, the proposed hybrid model achieved extremely competitive results. It was observed that the proposed hybrid deep architecture achieved a higher accuracy rate than the other 5 different ESA architectures. It was understood that the hybrid architecture, which achieved an accuracy value of 86.30%, can be used for GI bleeding and lesion detection and can help experts in making a diagnosis.

Among the future studies, increasing the number of ESA architectures used will be the first priority. Also, increasing the number of images used is among the targets. In addition, it is aimed to obtain patient images from different centers and create a multi-center data set.

References

[1] Rockey, D. C. (2005). Gastrointestinal bleeding. Gastroenterology Clinics, 34(4), 581-588.

[2] Kiziloluk, S., Yildirim, M., Bingol, H., & Alatas, B. (2024). Multi-feature fusion and dandelion optimizer based model for automatically diagnosing the gastrointestinal diseases. PeerJ Computer Science, 10, e1919.

[3] Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., ... & Ghayvat, H. (2021). CNN variants for computer vision: History, architecture, application, challenges and future scope. Electronics, 10(20), 2470.

[4] Bingol, H. (2022). Classification of OME with Eardrum Otoendoscopic Images Using Hybrid-Based Deep Models, NCA, and Gaussian Method. Traitement du Signal, 39(4).

[5] Pannu, H. S., Ahuja, S., Dang, N., Soni, S., & Malhi, A. K. (2020). Deep learning based image classification for intestinal hemorrhage. Multimedia Tools and Applications, 79, 21941-21966.

[6] Access Date: 05.09.2024, https://data.mendeley.com/datasets/8pbbjf274w/1

[7] Khan, Ali; Malik, Hassaan (2023), "Gastrointestinal Bleeding WCE images Dataset", Mendeley Data, V1, doi: 10.17632/8pbbjf274w.1

[8] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

[9] Kiziloluk, S., Sert, E., Hammad, M., Tadeusiewicz, R., & Pławiak, P. (2024). EO-CNN: Equilibrium Optimization-Based hyperparameter tuning for enhanced pneumonia and COVID-19 detection using AlexNet and DarkNet19. Biocybernetics and Biomedical Engineering, 44(3), 635-650.

[10] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[11] Xia, X., Xu, C., & Nan, B. (2017, June). Inception-v3 for flower classification. In 2017 2nd international conference on image, vision and computing (ICIVC) (pp. 783-787). IEEE.

[12] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[13] You, Y., Zhang, Z., Hsieh, C. J., Demmel, J., & Keutzer, K. (2018, August). Imagenet training in minutes. In Proceedings of the 47th international conference on parallel processing (pp. 1-10).