NFF: A Novel Nested Feature Fusion Method for Efficient and Early Detection of Colorectal Carcinoma

Amitesh Kumar Dwivedi, Gaurav Srivastava, and Nitesh Pradhan

Abstract Colorectal cancer is one of the most common cancer types and causes of death due to cancer in the world. Wireless curated endoscopy is used to diagnose and classify colorectal carcinoma. However, the major drawback of wireless curated endoscopy is that it presents many images to be analyzed by the medical practitioner. Therefore, many studies have been performed to automate the detection and classification of colorectal carcinoma using machine learning and deep learning models. Studies vary from traditional image classification techniques to image processing algorithms combined with data augmentation combined with pre-trained neural networks for early detection and type classification of colorectal carcinoma. In this manuscript, we proposed a novel nested feature fusion method to fuse the deep features extracted by the pre-trained EfficientNet family to devise an approach for early detection and classification of colorectal carcinoma. We have used the WCE curated colon disease dataset, which consists of 4 classes: normal, ulcerative colitis, polyps, and esophagitis. Our proposed method and experimental results outperformed compared to the state of the art with the fused model showing an accuracy of 94.11%. Medical centers can use the proposed method to detect colorectal cancer efficiently in real life.

Keywords Colorectal carcinoma · Deep learning · Feature extraction · EfficientNet · Nested feature fusion

1 Introduction

Colorectal carcinoma (CRC) is ubiquitous and is the underlying cause of death due to cancer worldwide [1, 2]. Unfortunately, colorectal carcinoma is mainly discovered in very late stages in patients for its effective treatment [3]. Mainly, colonoscopy is used to detect the various types of CRCs. However, such methods also impose risks to the patient, such as bleeding, negative consequences of sedation, colonic perforation, and

A. K. Dwivedi · G. Srivastava · N. Pradhan (✉)
Department of Computer Science and Engineering, Manipal University Jaipur, Rajasthan, India
e-mail: nitesh.pradhan@jaipur.manipal.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

K. A. Reddy et al. (eds.), *Proceedings of Fourth International Conference on Computer and Communication Technologies*, Lecture Notes in Networks and Systems 606,
https://doi.org/10.1007/978-981-19-8563-8_28

other clinical risks [4, 5]. Furthermore, due to wide-ranging variation in data from one patient to another, traditional learning methods of diagnosis are not extremely reliable [6].

Biomedical image processing is the mainstay of scientific research and an essential part of medical care, which is being highly sought after in the field of deep learning [7]. Although clinical detection of diseases based on traditional medical imaging methods has provided factual accuracy, developments in machine learning have pushed deep learning research developments in biomedical medical imaging [6].

To augment the process of colorectal carcinoma detection, a tremendous amount of research is focused on detecting CRCs through medical image processing and computer-aided diagnosis.

Machine learning methods have provided accurate classification and prediction abilities and have been deployed to be used for the diagnosis and prognosis of various medical ailments and health conditions due to their data-backed method of analysis, which unifies diverse risk factors into a classification/prediction algorithm [8–10]. However, deep learning methods are more effective than conventional machine learning methods due to their ability to process a high number of available samples during the training stage [11], their ability to execute feature engineering on its own, and their need for less human intervention while training which is highly suitable for datasets with a large number of samples. Furthermore, deep neural network models and frameworks can be retrained using a custom dataset compared to traditional computer vision algorithms, which are highly domain-specific. This provides much flexibility in deep learning compared to traditional machine learning algorithms [12].

With deep learning, an image dataset with object classes annotated to each image is presented to the machine to facilitate end-to-end learning [13], which is, in comparison with traditional computer vision techniques where parameters have to be fine-tuned by the CV engineer, is much easier.

The remaining contents of the proposed experimentation can be summarized as follows: Sect. 2 briefs about the previous academic works of various scholars in detecting colorectal carcinoma. Section 3 explores EfficientNet models, other deep learning strategies, and the materials and methods used. Section 4 describes the deep feature extraction and model training. Finally, Sect. 5 presents the experimentation and results from the mentioned experimentations.

2 Related Works

A variety of research has been performed on the automated detection and classification of colorectal cancer using machine learning and computer vision algorithms. Recently, deep learning has become the state-of-the-art approach for performing the classification of colorectal cancer due to its current popularity in biomedical image classification experimentations.

The study presented by Jesmar et al. proposed a model that integrates EfficientNet, MobileNetV2, and ResNetV2 into a single feature extraction pipeline called multi-fused residual convolutional neural network (MFuRe-CNN) with Auxiliary Fusing Layers (AuxFL) and a Fusion Residual Block (FuRB). The fusion model along with the Alpha Dropouts diagnosed a diverse set of endoscopic images of gastrointestinal ailments and handled conditions such as ulcerative colitis, esophagitis, polyps, and a healthy colon. The datasets used in the experimentation consisted of KBASIR and ETIS-Laris PolyDB. The fusion model showed an accuracy of 97.25% with only 4.8 million parameters. Furthermore, the FuRB and Alpha Dropouts substantially contributed to reducing overfitting and performance saturation [14]. However, due to lack of testing on other datasets, the proposed model does not immediately guarantee similar results for other medical images. Additionally, FuRB and Alpha Dropouts tend to slow down interference. In another study conducted by Khan et al. [15], an automated system is used to distinguish gastrointestinal infections based on WCE imaging. Automated functions within the research experiment included preprocessing ulcer frames with a dark channel, decorrelation, optimization of saliency-based segmentation to improve ulcer visibility, feature extraction using deep learning, selection of best frames, and classification of the selected features. A multi-class cubic SVM was used to classify the selected features, which attained an accuracy of 98.40%. However, in this study, if the segmentation of ulcers is incorrect, then the deep learning model can be mistrained.

Furthermore, in a study by Poudel et al., a neural network for endoscopic image classification is provided using an adequate dilation in convolutional neural networks (CNNs). To deal with overfitting and extraneous noise and miscellaneous features, DropBlock, a regularization technique has been used. The proposed study compares its proposed model's efficiency with that of other state-of-the-art models such as VGG16, InceptionResnetV2, Xception, ResNet, DenseNet, and NasNet. Using the proposed model, 95.7% accuracy is achieved, and an F1 score of 0.93 is achieved with the colorectal dataset, and an F1 score of 0.88 is obtained with the KVASIR dataset. The achieved accuracy gives better results than the traditional methods. However, the model has not been tested on other medical image datasets for classification purposes [16]. In the study presented by Silva et al. [17] likely polyps within image samples were withdrawn using geometric shape features. Further, the regions containing polyps were boosted using textural features. Evaluation of this method was conducted on datasets that contained 300 images of polyps and 1,200 images without polyps. According to the proposed method, 91.2% sensitivity, 95.2% specificity, and a deceit detection rate of 4.82% were achieved, which are comparable to the analysis systems developed for online video colonoscopy images. In the study presented by Fan et al., AlexNet convolutional neural network was used and trained to a database containing more than ten thousand images of wireless capsule endoscopy images to detect ulcers and erosion. Based on the proposed model, the accuracy was 95.16% and 95.34%, the sensitivity was 96.80% and 93.67%, and the specificity was 94.79% and 95.98%, correspondingly. Despite the fact that the method used in this experiment had great results in detecting ulcers and erosions, and it was unable to identify some ulcers, erosions, and other WCE images. After the experimentation, approximately 5% of images were not incorrectly [18].

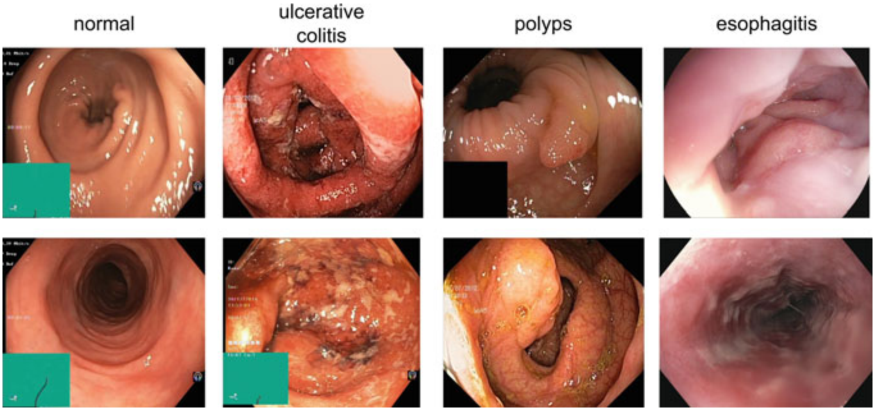

Fig. 1 Dataset samples of normal, ulcerative colitis, polyps, and esophagitis

We observed that most experiments focused on metrics such as accuracy, sensitivity, and F1 score and observed negligence in the area of efficiency. Thus, we decided to propose a novel and efficient neural network to tackle the early detection of colorectal carcinoma (Fig. 1).

3 Materials and Methods

3.1 Data Collection

WCE curated colon disease dataset deep learning is an image dataset for gastrointestinal tract or simply, a colon disease image dataset [19, 20]. These are images of the gastrointestinal tract captured during the procedure of wireless capsule endoscopy, which in the scope of current experimentation, will be used to devise a deep learning model for the early detection of colorectal carcinoma. The dataset contains 6000 colored images and the dataset contains four classes: Normal, ulcerative colitis, polyps, and esophagitis as given in Table 1.

Table 1 Dataset description

Classes	Normal	Ulcerative colitis	Polyps	Esophagitis
No. of samples	1500	1500	1500	1500

3.2 Data Preprocessing

Data preprocessing is an essential step for deep learning model training. It outlines the processes required to alter or encode data so the model can parse it effectively. In neural networks, the model expects the input image to be the same size. However, the images gathered are not the same size or form. The images in our dataset originally ranged in size from 400 × 300 to 936 × 768 pixels. We converted all the images into a common size of 128 × 128 pixels as a preprocessing step before training because the dataset's images were not homogeneous and came in varied sizes. After applying RGB reordering to all images, the model's final input was delivered as a 128 × 128 × 3 matrix.

While downscaling the images, we can sometimes lose some vital information, so this has to be done carefully by observing the dataset. For example, suppose we have a dataset of MRI scans for brain tumor classification. In that case, if we downscale the images to a minimal size, the tumor will almost disappear from MRI scans, which can impact training accuracy. Also, resizing the image to a very large size like 512 × 512 can exceed the GPU memory. Therefore, to make it both memory efficient and not lose any critical information from the image, we have to choose the best image size based on the experiments.

We scaled and ran our trials on all 128 × 128, 196 × 196, and 256 × 256 image sizes in this study, and we found that the accuracy is similar for all three image sizes. However, training time is considerably shorter on 128 × 128, saving significant computational efforts.

3.3 Dataset Division

A deep learning model may obtain a 99% accuracy rate, but it fails when evaluated on real-world images. In order to prevent model selection bias and overfitting, it is ethical to divide the dataset into training, validation, and testing sets. Furthermore, our parameter estimations are more variable when we have a scant amount of data. Similarly, our performance measure will be more variable if we have fewer testing data. As a result, we should split the data so that no variances are excessive.

Adding more data to the final testing set ensures the method's resilience and minimizes the chance of failure in real-world tests. As a result, as given in Table 2, we partitioned the entire dataset into three sections: 70% training, 10% validation, and 20% testing.

Table 2 Dataset division

Classes	Normal	Ulcerative colitis	Polyps	Esophagitis
Training set	1050	1050	1050	1050
Validation set	150	150	150	150
Testing set	300	300	300	300

3.4 Transfer Learning

Transfer learning was initially discussed in the NeurIPS (Conference on Neural Information Processing System), which talked about using previously learned knowledge to augment further future learning. Deep transfer learning (DTL) combines deep learning architecture with transfer learning. Deep neural networks (DNNs) provide a powerful way to learn features, making them useful in feature-based transfer learning. Methods based on latent feature spaces utilize DNNs to discover a common latent feature space where both source and target data can exhibit the same probability properties. Consequently, the source data can be used as a training set for target data in the latent feature space, which improves the model's performance with target data [21].

3.5 EfficientNet

EfficientNet is a simple convolutional neural network that is known for its profound effective compound scaling method that helps researchers to scale up a convolutional neural network to any target resource constraints in a highly fundamental way, quickly. Unlike other architectures, EfficientNet uniformly scales network resolution, depth, and width. EfficientNets are also highly used in transfer learning which is why they are being used in the scope of this experiment [22].

3.6 Proposed Nested Feature Fusion Method

In order to construct a CNN, you need to extract features and classify them. The model's first layers may be considered as descriptors of image features, whereas the latter layers are associated with specific categories. In feature extraction, many convolution layers are utilized, followed by max-pooling and an activation function. A fully connected layer and a softmax activation function are standard components of a classifier. Since the number of classes in a dataset is directly proportional to the number of features in a model to learn, to learn complex features, the feature extraction component of the convoluted neural network should be more complex and deeper.

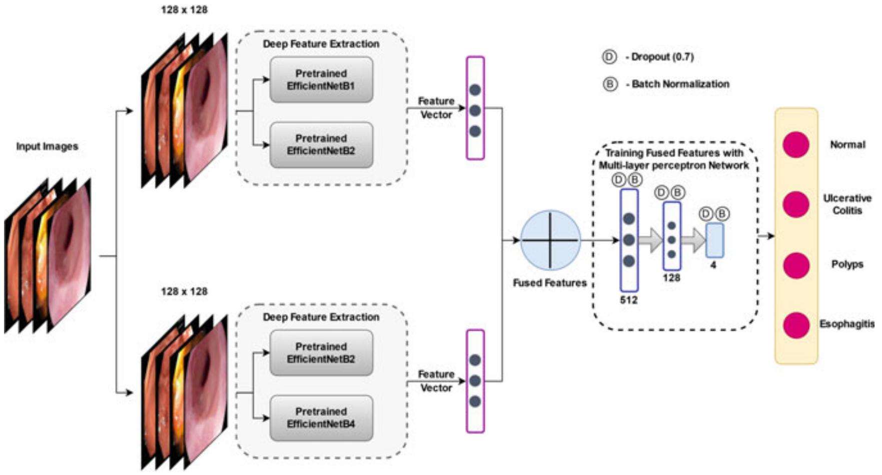

Fig. 2 Graphical abstract of the proposed nested feature fusion method

A feature in an image is a component or pattern of an object that helps with identification. In computer vision and image processing, it is a piece of information regarding the content of an image, usually pertaining to whether a particular section of the image contains specified properties. Different structures in an image, such as points, edges, or objects, are examples of features. Each CNN produces a feature vector with a distinct set of features extracted. They can overlap, but they are not always the same, which is why the accuracy varies from time to time. The key idea behind this proposed nested fusion model is that each CNN will produce a feature vector. By integrating those feature vectors, we will not miss any features the model ignores, resulting in a significant set of features being omitted.

So, in the proposed method, we first fused EfficientNetB1 and EfficientNetB2 and EfficientNetB2 and EfficientNetB4 individually. Both provide two output feature vectors, which we fuse further to create our final model. After the feature extraction, we use a multi-layer perceptron network with a softmax activation function to classify the input image into their respective categories. The proposed methodology is depicted in Fig. 2.

4 Deep Feature Extraction and Model Training

4.1 Loss Function: Categorical Cross-Entropy

The loss function is used to measure the deviation of the estimated value from the true value. It is a computational procedure to assess how the algorithm used models the data. In this experiment, cross-entropy loss function is used because of its ability to increase in magnitude when predicted probability skews from the actual results. The following mathematical Eq. 1 explains the computation of the cross-entropy loss function:

$$L_{\mathrm{CE}} = -\sum_{i=1}^{n} t_i \log(p_i)\,,\ \text{for n classes,} \tag{1}$$

where t_i is the truth label and p_i is the Softmax probability for the ith class.

4.2 Classifier: Softmax

The softmax classifier is an output function that outputs the probabilities for each class label in the form of a vector. It is usually used for multi-class classification purposes. Softmax function is defined in Eq. 2.

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{2}$$

where σ = softmax, $\mathbf{z}$ = input vector, e^{z_i} = standard exponential function for input, K = number of classes in the multi-class, and e^{z_j} = standard exponential function for output.

4.3 Learning Rate Decay

Learning rate decay is an actual practical technique that is used to instruct modern neural networks. It initializes with an enormous learning rate and then declines multiple times: Decomposition of learning rate—decay. It is used to enhance optimization and generalization in the experimentation process. Learning rate decay can be time-based, step-based, and exponential.

5 Experimental Results and Discussion

5.1 Experimental Setup

All the models mentioned in the proposed research were implemented with TensorFlow in Python. Further, Kaggle was used to train the models mentioned, with the following specs - GPU Tesla P100-PCIE-16GB compute capability: 6.0 and 16 GB GPU RAM.

Table 3 Experimental results of efficientnet family

Model	Training accuracy	Validation accuracy	Testing accuracy
EfficientNetB0	95.79	91.71	91.25
EfficientNetB1	95.95	91.71	92.42
EfficientNetB2	95.43	92.45	92.93
EfficientNetB3	96.10	91.87	92.09
EfficientNetB4	94.55	92.04	93.09
EfficientNetB5	94.90	92.87	90.40
EfficientNetB6	93.79	91.04	91.92
EfficientNetB7	95.76	93.20	91.08

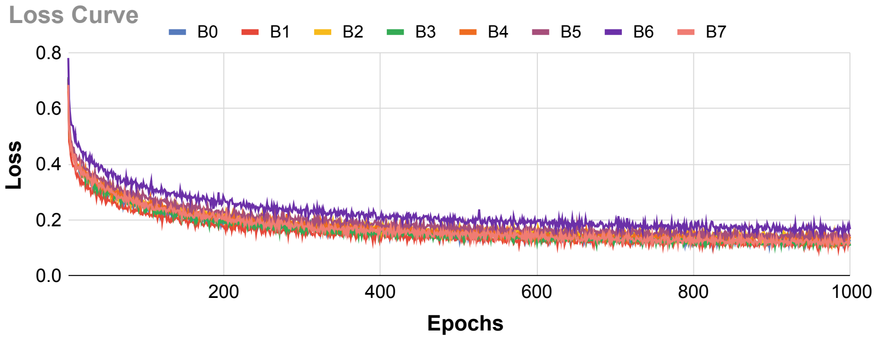

Fig. 3 Loss curve

5.2 *Classifier Performance*

The first and most crucial step in constructing a deep learning model is to define the network architecture. We prefer to use pre-trained networks to extract deep features as they have been initially trained on a large-scale ImageNet dataset. Therefore, we save a lot of computational power when adjusting weights to match our WCE dataset. In this study, we have used pre-trained networks of the EfficientNet family for feature extraction. The extracted deep features were then trained with a multi-layer perceptron network with a softmax activation function. The accuracy achieved on each of the networks is reported in Table 3. The loss and accuracy curve of the training of EfficientNet family are shown in Figs. 3 and 4, respectively.

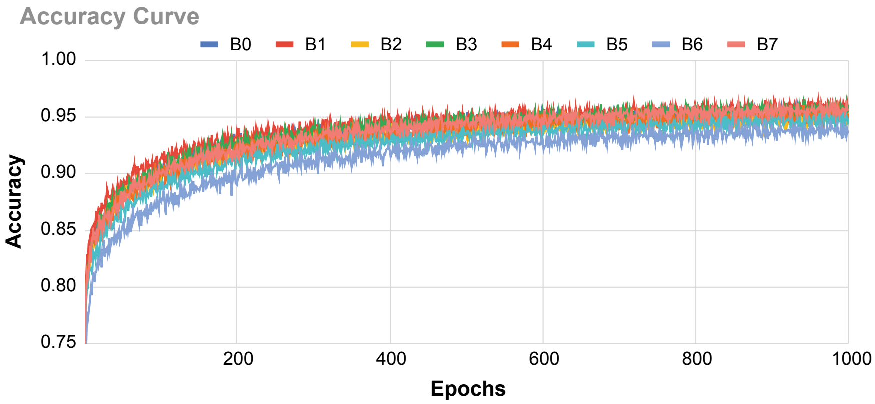

Fig. 4 Accuracy curve

5.3 Nested Fusion Model

Three classifiers are required to generate the fusion model. After working with the whole EfficientNet family, it was discovered that EfficientNetB1, EfficientNetB2, and EfficientNetB4 provided the best testing accuracy. As a result, Fused Model 1 was created by combining EfficientNetB1 and EfficientNetB2, while Fused Model 2 was created by combining EfficientNetB2 and EfficientNetB4. Furthermore, we have fused models 1 and 2 together to generate our final nested fusion model.

On the test dataset, combining the EfficientNetB1 and EfficientNetB2 generated an accuracy of 93.43%, while combining the EfficientNetB2 and EfficientNetB4 gave an accuracy of 93.63%. Finally, when the previous two fused models were combined, an accuracy of 94.11% was achieved on the test dataset as given in Table 4. The loss and accuracy curve of the training of fusion models are shown in Fig. 5. The confusion matrix and AUC-ROC plots of each fusion model are shown in Fig. 6.

Table 4 Experimental results of nested feature fusion model

Fused model	Training accuracy	Validation accuracy	Testing accuracy
EfficientNetB1 and EfficientNetB2	98.71	93.28	93.43
EfficientNetB2 and EfficientNetB4	98.76	93.45	93.63
Final fused model	99.50	93.95	94.11

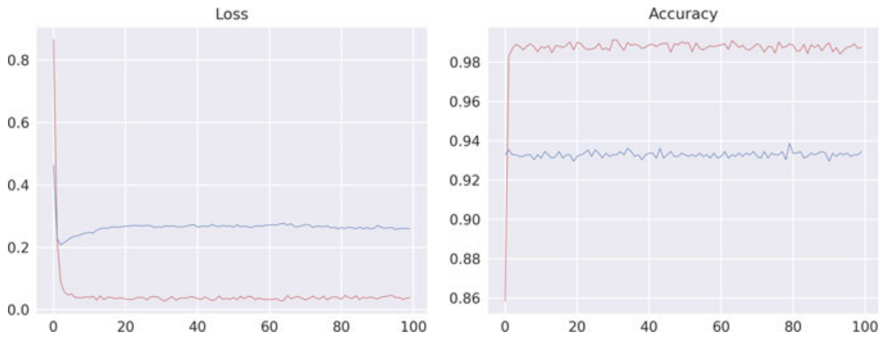

Fig. 5 Loss and accuracy curve of the training of final fused model

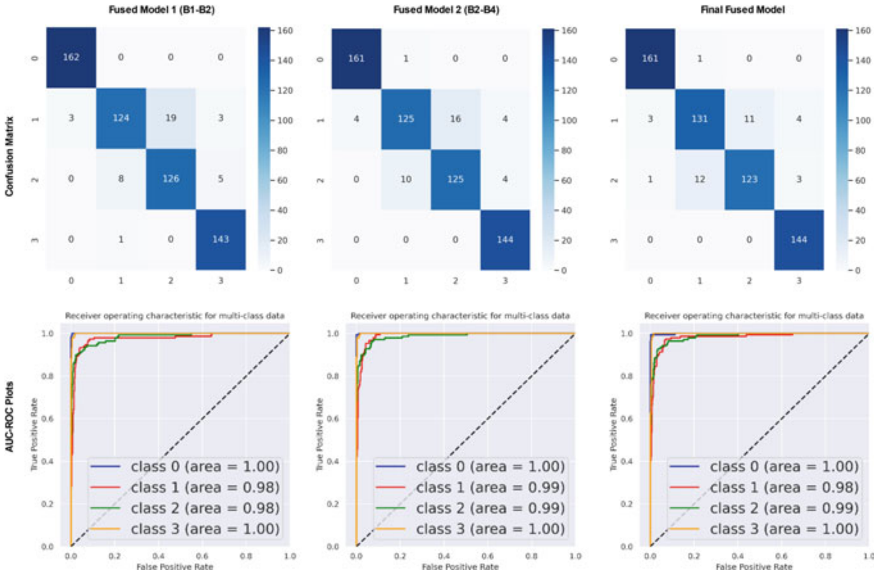

Fig. 6 Confusion matrix and AUC-ROC plots of each fused model

6 Conclusion and Future Directions

Early stage detection of colorectal carcinoma is essential for correctly diagnosing and curing the disease. Our research experimentations establish that a nested fused model can be used to predict colorectal carcinoma in its early stages accurately and can also perform classification upon the type of colorectal carcinoma in its early stage. First, we use pre-trained networks of the EfficientNet family for feature extraction. Later, the deep features are trained in a multi-layer perceptron network with a softmax activation function. We experimented with pre-trained networks of the EfficientNet family. Afterward, we fused EfficientNetB1 and EfficientNetB2 and

EfficientNetB2 and EfficientNetB4 and developed a model and a novel approach for early detection and classification of colorectal carcinoma. Our proposed model gives a testing accuracy of 94.11%. This is a novel approach to early stage detection and classification of colorectal carcinoma. Furthermore, this method can also be used in other biomedical classification tasks for fast and automated detection and classification of diseases.

References

1. Ponzio F, Macii E, Ficarra E, Cataldo SD (2018) Colorectal cancer classification using deep convolutional networks. In: Proceedings of the 11th international joint conference on biomedical engineering systems and technologies, vol 2, pp 58–66
2. Matthew F, Sreelakshmi R, Tatishchev Sergei F, Wang Hanlin L (2012) Colorectal carcinoma: pathologic aspects. J Gastrointest Oncol 3(3):153
3. Wan N, Weinberg D, Liu T-Y, Niehaus K, Ariazi EA, Delubac D, Kannan A et al (2019) Machine learning enables detection of early-stage colorectal cancer by whole-genome sequencing of plasma cell-free DNA. BMC Cancer 19(1):1–10
4. Young Patrick E, Womeldorph Craig M (2013) Colonoscopy for colorectal cancer screening. J Cancer 4(3):217
5. Su H, Lin B, Huang X, Li J, Jiang K, Duan X (2021) FFNet: multi-branch feature fusion network for colonoscopy. Front Bioeng Biotechnol 515
6. Razzak MI, Naz S, Zaib A (2018) Deep learning for medical image processing: overview, challenges and the future. Classification BioApps 323–350
7. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier KH (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18(2):203–211
8. Liyan P, Guangjian L, Fangqin L, Shuling Z, Huimin X, Xin S, Huiying L (2017) Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. Sci Rep 7(1):1–9
9. Konstantina K, Exarchos Themis P, Exarchos Konstantinos P, Karamouzis Michalis V, Fotiadis Dimitrios I (2015) Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 13:8–17
10. Passos IC, Mwangi B, Kapczinski F (2016) Big data analytics and machine learning: 2015 and beyond. Lancet Psychiatry 3(1):13–15
11. Dinggang S, Guorong W, Heung-Il S (2017) Deep learning in medical image analysis. Annual Rev Biomed Eng 19:221
12. O'Mahony N, Campbell S, Carvalho A, Harapanahalli S, Hernandez GV, Krpalkova L, Riordan D, Walsh J (2019) Deep learning vs. traditional computer vision. In: Science and information conference. Springer, Cham, pp 128–144
13. Montalbo Francis Jesmar P (2022) Diagnosing gastrointestinal diseases from endoscopy images through a multi-fused CNN with auxiliary layers, alpha dropouts, and a fusion residual block. Biomed Signal Process Control 76:103683
14. Poudel S, Kim YJ, Vo DM, Lee S-W (2020) Colorectal disease classification using efficiently scaled dilation in convolutional neural network. IEEE Access 8:99227–99238
15. Khan MA, Kadry S, Alhaisoni M, Nam Y, Zhang Y, Rajinikanth V, Sarfraz MZ Computer-aided gastrointestinal diseases analysis from wireless capsule endoscopy: a framework of best features selection. IEEE Access 8:132850–132859
16. Juan S, Aymeric H, Olivier R, Xavier D, Bertrand G (2014) Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. Int J Comput Radiol Surgery 9(2):283–293

17. Fan S, Lanmeng X, Fan Y, Wei K, Li L (2018) Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. Phys Med Biol 63(16):165001
18. Chenjing C, Shiwei W, Youjun X, Weilin Z, Ke T, Qi O, Luhua L, Jianfeng P (2020) Transfer learning for drug discovery. J Med Chem 63(16):8683–8694
19. Pogorelov K, Randel KR, Griwodz C, Eskeland SL, de Lange T, Johansen D, Spampinato C et al (2017) Kvasir: a multi-class image dataset for computer aided gastrointestinal disease detection. In: Proceedings of the 8th ACM on multimedia systems conference, pp 164–169
20. Juan S, Aymeric H, Olivier R, Xavier D, Bertrand G (2014) Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. Int J Comput Radiol Surgery 9(2):283–293
21. Pan SJ, Yang Q (2009) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359
22. Tan M, Le Q (2019) Efficientnet: rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR, pp 6105–6114