Multimedia Tools and Applications (2020) 79:28825–28840
https://doi.org/10.1007/s11042-020-09468-3

Gastrointestinal tract classification using improved LSTM based CNN

Şaban Öztürk[1] · Umut Özkaya[2]

Received: 6 August 2019 / Revised: 28 June 2020 / Accepted: 28 July 2020 /
Published online: 6 August 2020

Abstract
Automated medical image analysis is a challenging field of research that has become quite widespread recently. This process, which is advantageous in terms of both cost and time, is problematic in terms of obtaining annotated data and lack of uniformity. Artificial intelligence is beneficial in the automatic detection of many diseases where early diagnosis is vital for human life. In this study, an effective classification method is presented for a gastrointestinal tract classification task that contains a small number of labeled data and has a sample number of imbalance between classes. According to our approach, using an effective classifier at the end of the convolutional neural network (CNN) structure produces the desired performance even if the CNN structure is not strongly trained. For this purpose, a highly efficient Long Short-Term Memory (LSTM) structure is designed and added to the output of the CNN. Experiments are conducted using AlexNet, GoogLeNet, and ResNet architectures to test the contribution of the proposed approach to the classification performance. Besides, three different experiments are carried out for each architecture where the sample numbers are kept constant as 2500, 5000, and 7500. All experiments are repeated with CNN + ANN and CNN + SVM architectures to compare the performance of our framework. The proposed method has a more successful classification performance than other state-of-the-art methods with 97.90% accuracy.

Keywords Colorectal cancer · Endoscopic images · Deep learning · LSTM · Transfer learning

1 Introduction

Colorectal cancer (CRC) is caused by the occurrence of malignant polyps in the colon. According to obtained statistical information, colon cancer is one of the most common types

✉ Şaban Öztürk
saban.ozturk@amasya.edu.tr

[1] Technology Faculty, Electrical and Electronics Engineering, Amasya University, Amasya, Turkey

[2] Engineering and Natural Science Faculty, Electrical and Electronics Engineering, Konya Technical University, Konya, Turkey

of cancer in the United States [33]. This type of cancer is the third most common type of cancer worldwide and is the second most common cause of mortality [7]. Since early diagnosis is very important for human life, it is highly burdened by expert doctors. Some polyps are likely to be missed if only expert knowledge is used to detect these polyps. Therefore, it is advantageous to use the computer-aided diagnosis (CAD) systems and machine learning algorithms.

Machine learning (ML) systems have been used effectively in the analysis of images and videos for many years [23]. When we look at the recent trend of ML, it is generally seen that there is a concentration in the medical field [34]. Firstly, medical imaging systems have become widespread, especially with the increasing hardware force in recent years. As a result of this spread, many medical images and videos emerge. It takes a long time and tiring to process and analyze them by expert doctors. Secondly, they can be used as a counseling system for specialist doctors and can easily catch missed situations. Last but not least, the use of these systems to support human health is very virtuous.

In the determination of gastrointestinal (GI) tract disease, machine learning methods are essential. Because the polyp marking process applied by specialist doctors has some critical problems in itself, these are folds, recesses, and stool deposits that make the detection of polyps difficult. Also, it is not possible to clean all the polyps in the colon. Therefore, the patient should be investigated gradually. There is the possibility that unspecified polyps may turn into cancer [38].

1.1 Related works

When GI tract detection studies in the literature are examined, most of the studies are limited to polyp detection in the colon. The reasons for this can be the insufficient number of medical data or simplifying the problem [12]. When the first studies in this area are examined, it is understood that the analysis is made according to the shape and texture information. Wang et al. [40] use morphological, geometrical, and texture features for the detection of colonic polyps. According to their method, they try to detect the growing region of suspected polyp using local geometry. Chowdhury et al. [8] propose the colonic polyp detection method using statistical features derived from the local surface. Their technique shows 87.5% sensitivity for polyps greater than 5 mm. Hwang et al. [16] introduce an elliptical shape-based polyp detection method to detect small colon polyps. Zhao et al. [47] use lines of curvatures to characterize the polyp surface, and they use these lines for polyp detection. These hand-crafted methods have two drawbacks when compared with current automated methods. First, these methods are generally recommended for the solution of a single problem and require considerable experience. The second is that it cannot work with high performance on the raw samples without preprocessing.

The development of machine learning methods has led to the quick entry of automated methods into almost every field. Satisfactory achievements have been obtained in many areas, including GI tract detection. Tajbakhsh et al. [36] propose a novel method integrating the global geometric features with local intensity variations of polyp boundaries. In this way, they can discriminate polyp edges and non-polyp edges. Bernal et al. [5] analyzed colonoscopy videos to detect automatic polyps. The proposed method consists of three stages: these are regional segmentation, region definition, and region classification. The output of the proposed algorithm also defines which areas can be considered non-informative. Yuan et al. [43] use the bag of feature (BoF) method to classify polyps in wireless-capsule endoscopy. They extract many textural features from neighborhoods of the key points instead of the scale-invariant feature transform (SIFT). David et al. [11] present an automatic detection method for colon

polyps using color, geometrical features, and histogram of gradients. Then, they classify these features using multilayer perceptron neural networks for the classification of these features. Although these studies overcame many problems, such as focusing on single events and additional preprocessing steps, which were common problems for traditional algorithms, it still required a lot of experience. Still, it cannot works well in real-time.

Recently, CNN has been used for image processing to overcome all the mentioned problems. Thanks to the high success achieved by Alexnet architecture [20], CNN architecture spreading to almost all fields did not take a long time to enter the field of medical image analysis. In addition to Alexnet, GoogLeNet [35], Resnet [37], etc. which is deeper and more stable models, have been proposed in the following years for classification tasks. Many researchers working in the field of GI tract detection have benefited from the power of CNN architecture. Karnes et al. [18] proposed the CNN model that can accurately identify colonic polyps in colonoscopic images with high AUROC. Cogan et al. [9] propose the MAPGI framework to increase the classification accuracy of small datasets like the Kvasir dataset. They use edge removal, contrast enhancement, image filtering, color mapping, and scaling for each image with CNN architecture. Zhang et al. [45] proposed a new convolutional neural network (CNN) structure based on regression for the detection of polyps during a colonoscopy. The proposed method consists of two stages. In the first stage, transfer learning was used to train the ResYOLO algorithm. And then, Efficient Convolution Operator (ECO) improved trained performance. The proposed method was able to detect polyphonic frames with a sensitivity of 88.6%. Zeng et al. [44] proposed LSTM based new method on polyp region detection. They used the LSTM algorithm to decode feature vectors. Experimental results show that their method can precisely determine the position of the focus area, while at the same time, BLEU-1 increases the BLEU-2 scores by 1%, with fewer parameters. Yu et al. [42] present a novel online and offline 3D deep learning approach for automatic polyp detection. According to their approach, it can produce a polyp probability map for online colonoscopy videos. Kang and Gwak [17] suggested a robust object detection architecture to detect polyps in colonoscopy images. They also proposed a hybrid method for combining two Mask R-CNN models with different structures (ResNet50 and ResNet101) to improve performance.

When the studies are examined, it is seen that CNN architecture produced very successful results and overcame all the mentioned problems. However, the success of the CNN structure is highly correlated with the number of samples used for training [19]. Considering that one of the biggest problems with medical images is the limited number of labeled data, the magnitude of the problem is understood. To achieve this problem, transfer learning has been actively used to detect polyps [28]. Ribeiro et al. [27] use an 8-HD endoscopic database to train CNN networks using scratch and pre-trained models with transfer learning. Comparing the results of classical methods and proposed CNN algorithms, it is understood that off-the-shelf CNN features should be improved. Shin et al. [32] work on a region-based evolutionary neural network (RCNN) approach that is applied for automatic detection of images from colonoscopy examinations. Experimental results by using large-scale colonoscopy databases have been analyzed that suggested detection systems perform better than non-pretrained systems in the literature. Urban et al. [39] obtained 96.4% cross-validation accuracy with the transfer learning CNN models proposed for the detection of polyps in the 8641 colonoscopy images. The proposed method is tested in real-time with the help of GPU.

1.2 Contribution of this study

In this study, a GI tract classification method is presented by utilizing the power of CNN architecture. Training and test processes of the proposed method are performed using the

Kvasir dataset [24]. Recently, several methods have been presented in the literature to classify images of the Kvasir dataset, and all methods are CNN based [1, 2, 6, 10, 12, 13, 19, 21, 25]. These studies are usually based on adding various layers to classical CNN architectures, updating layer parameters according to different approaches, or optimizing the parameters in the layers. In some methods, network input or network output is supported by hand-crafted methods to increase classification accuracy. The main problem in such studies is that there is not enough data to train the CNN model strongly. The layers in the CNN architecture have always been limited to represent endoscopic features. To overcome this problem, it is necessary to use a classifier with strong representation capability with little data at the output of the CNN model. In the literature, it is seen that the results of the studies aimed at increasing the feature representation power are more successful [10, 12]. But there is no study focusing on the classification of properties section for the Kvasir dataset. Accordingly, two LSTM layers are combined with the dropout layer. These LSTM layers contain 200 neurons and 100 neurons, respectively. Since the learning power of the LSTM structure is quite high, it is supported by dropout layers against memorization. The Stacked LSTM model is one of the most effective univariate LSTM models used for time series forecasting problems. Compared to other univariate LSTM models, the stacked LSTM architecture is more suitable for sequential and independent prediction methods. For this reason, stacked LSTM is preferred in this study. The proposed method has been tested with AlexNet, GoogLeNet, and ResNet-50. Experiments are repeated by adding ANN and SVM to the classifier layers of the same CNN architectures, respectively, to reveal the contribution of the method. The most important contribution of this study is that it includes a highly representative classifier approach for datasets containing limited samples. In contrast to the trend of producing more robust features in the literature, the proposed study can classify features with limited representation power with high performance.

The structure of this article is the following; in Section 2, the fundamentals of CNN and proposed method details are presented. The dataset and experimental results are reported in Section 3. Finally, the conclusion is giving in Section 4.

2 Methods

2.1 Fundamentals of the proposed method

In 2006, Geoffrey Hinton et al. showed that deep neural networks could be effectively trained by the greedy-layered pre-training method [15]. Other research groups have used the same strategy to train many other deep networks. The rise of deep learning began when CNN architecture won the ImageNet challenge [20]. This architecture, called AlexNet, is more successful than other methods because it is very suitable for image processing problems. In later years, CNN becomes widespread, and many successful CNN architecture is recommended. Special CNN structures have been proposed for many image processing problems such as classification, semantic segmentation, instance segmentation, object detection, retrieval, etc. The reason behind its spread is that it represents the human vision system quite successfully. CNN architecture includes many layers such as convolution, pooling, ReLU, dropout, fully connected network, deconvolution, unpooling, softmax, concatenate, etc. The arrangement of these layers can create a new architecture. In addition, the state of the inter-layer connections is sufficient to form a new deep network. In a CNN architecture, the most basic layer is the

convolution layer. This layer is where the properties of the image are learned and stored. If we call the height of this filter *W*, the filter size will be *WxW* (width and height). This filter works according to the 2D conventional convolution process by sliding over the image as in Eq. 1.

$$y_i^l = f\left(\sum_j x_j^{l-1} \otimes w_{ij}^l + b_i^l\right) \tag{1}$$

in which *x* is input, *y* is output, *w* is convolution filter, and *b* represents bias. According to Eq. 1, the output of linear CNN is $y = F(x)$. As the number of filters in convolution layers increases, the number of parameters increases. Sliding of convolution filters over the images is called feature sharing and significantly reduces the number of parameters. Besides, the pooling layer is a highly effective layer used to reduce parameters. This layer, which is used for transferring the most important properties in the property matrix to other layers, has various types, including max-pooling, sum-pooling, and average pooling. The max-pooling used in this study is calculated as in Eq. 2.

$$P_{jm} = \max_{k=1}^{r} \left(x_{j(m-1)n+k}\right) \tag{2}$$

where *x* is the input matrix. Deep networks often suffer from the memorization of training data. In CNN architecture, the ReLU layer is very effective for overcoming this problem. $y = max(0,x)$ is used for ReLU layer in our method.

The classification of the features obtained automatically by the CNN structure is the main aim of this study. Considering that it is quite challenging to find labeled medical data, the importance of creating a robust classifier with less data is understood. In the literature, the transfer learning approach is used to learn features from such datasets [31]. The weights used in the structure of pre-trained models for a different task are referred to as transfer learning. In our study, pre-trained AlexNet, GoogLeNet, and ResNet-50 are used. These CNN architectures are trained with ImageNet challenge data. Afterward, various modifications and parameter changes detailed in the proposed section have been performed, and the training has been repeated with the Kvasir dataset.

Although transfer learning capabilities provide powerful learned features, sufficient classification level cannot be achieved for medical databases. Two of the most important reasons for this are the lack of labeled data and the unbalanced class distribution of medical datasets. In the literature, the artificial neural network (ANN) approach is generally used as a classifier as part of CNN architecture [14]. This layer is called the fully connected layer. Although the ANN structure produces highly successful results for classification problems, it is dependent on the number of training samples. In particular, as the number of ANN parameters increases, the number of samples should be high enough to avoid the problem of memorization [3]. On the other hand, the support vector machine (SVM) is more effective in classifying multiple classes. It can produce high performance, especially in classification problems with low data for many classes [41]. However, if the target classes are very close to each other, SVM performance may not reach the desired level. ANN structure is used in this study for comparison purposes because it is found in almost every structure proposed in the literature and is a primary classifier. SVM structure has been used for comparison in this study thanks to the solution of many multi-class problems that the ANN model cannot solve. However, these two classifiers have significant deficiencies, and these deficiencies prevent successful classification, no matter how well represented the features.

The LSTM structure is preferred for our CNN architecture. It can predict future situations. With LSTM, time-series data can be analyzed to predict the future situation of the data. Generally, it may be trained on different portions of the sequence rather than fixed-size inputs such as feedforward neural networks. It is very similar to a feedforward neural network except that they have a feedback loop that manifests itself in a series of timelines. This structure gives the LSTM a second chance to correct itself during training [29]. The overall process of this study is shown in Fig. 1.

2.2 Proposed method overview

Medical images often contain unwanted artifacts, annotation information, or various markings [24]. These disruptive factors in the medical images adversely affect the automatic image analysis algorithms. Elimination of these disruptive factors will improve the performance of classification and segmentation procedures. Therefore, studies have been made in the literature to eliminate these unwanted factors, and their positive contributions have been shown in [9, 19]. These methods increase the success of medical image analysis but are time-consuming. Therefore, in this study, a method that can produce successful results without the need for preprocessing is proposed. The proposed method can be trained end-to-end, but theoretically can be divided into two parts, feature extraction, and classification.

The feature extraction section is implemented using CNN, the most powerful algorithm of today. As mentioned before, the extracted features are not strong enough due to the drawbacks of the medical image datasets. In this study, unlike other studies, the classifier is strengthened rather than the feature extraction algorithm. Because it is seen that the contribution of feature reinforcement to classification accuracy is not sufficient.

As seen in Eq. 1, there are more parameters in a larger architecture. The number of samples required for the training of these weights should be higher. Besides, the activation function appropriate to the problem must be used. All this requires both experiences, and the behavior of the network cannot be predicted. In multiclass cases, the probability distribution of ANN

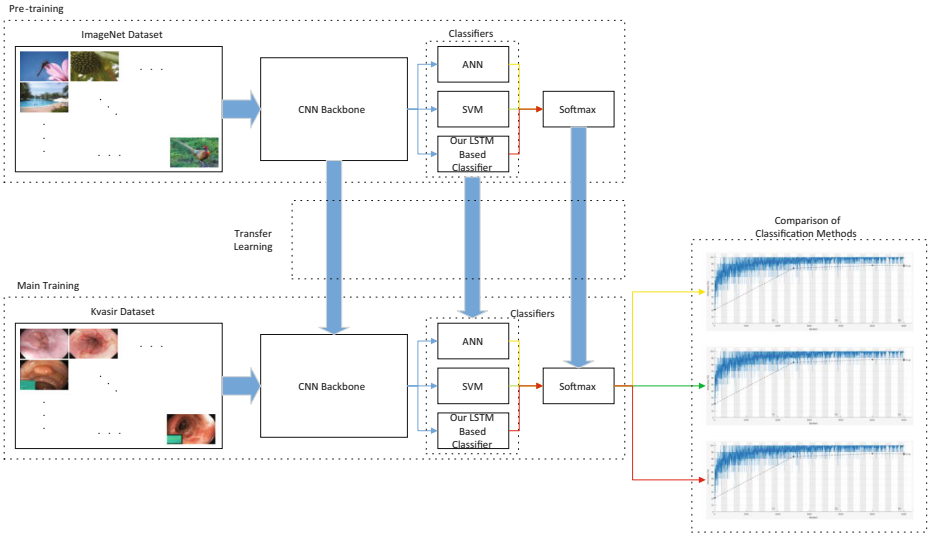


Fig. 1 The overall process of our approach

output is calculated by the softmax function. $1/\Sigma_j e^{fj}$ term normalizes the distribution. That is, the sum of the values equals 1. Therefore, it calculates the probability of the class to which the class belongs. When a test input is given x, the activation function in $j = 1,...,k$ is asked to predict the probability of $p\ (y = j \mid x)$ for each value.

Classification using the CNN + SVM approach has recently become popular. For many problems, more successful results have been achieved [26]. It works compatible with softmax function, especially in classification problems involving multiple classes. Let assigns w for weights, a for input features, and y for output. SVM method tries to reduce the number of nonzero weights. When Eq. 3 is examined, it is possible to see softmax compatibility with SVM.

$$y_i(x_i w + b) - 1 = 0 \tag{3}$$

Although the CNN + SVM supported by softmax function seems to be very useful theoretically, there is a significant problem in the experimental part. If the number of samples in each class in the dataset is unbalanced, the performance of this combination decreases. In particular, medical datasets sometimes contain very few examples for one or two classes. In this case, less class information disrupts the support vector's stability. To gain a better understanding of the dataset, previous experiences are of significant contribution. Recurrent neural networks (RNN) can estimate the current status using historical information [22]. However, RNN performance decreases as the gap increases between the samples [4]. To solve this gap problem, LSTM, which is an RNN structure, is used. LSTM is specially designed to solve the problem of long-term dependency. The output values generated by an LSTM cell are expressed in Eq. 4.

$$y(t) = \phi\left(x_{(t)}.w_x + y_{(t-1)}.w_y + b\right) \tag{4}$$

When we examine Eq. 4, we can see that classification can be performed without the need for softmax function. Besides, thanks to the short-term contribution, there is no distortion of the classifier curve of the classes containing fewer samples.

2.3 Parameter setting

The proposed method explores the effect of a robust classifier on CNN architecture. For this purpose, powerful CNN architectures such as AlexNet, GoogLeNet, ResNet are used as a backbone. The essential features of each architecture are learned with the help of the transfer learning approach, and then they are shallowly trained with the Kvasir dataset. In the feature extraction process, the original architectures have not been changed. The batch size is 16, the learning rate is 0.0001, and dropout is selected as 0.5 for training. At higher learning rates, CNN models cannot converge a solution for the problem. Lower learning rates require more training time. We use stochastic gradient descent (SGD) to update network parameters.

An LSTM structure is designed to replace the ANN layers at the output of the CNN architecture. A two-layer LSTM block is added to reduce response time and computational complexity. The first LSTM block consists of 200, and the second LSTM block consists of 100 cells. Dropout layers are added so that the LSTM structure, which can learn the data strongly, does not suffer from the problem of memorization. The dropout ratios of the dropout layers used in the LSTM block are selected as 0.5. The proposed method is shown in Fig. 2.

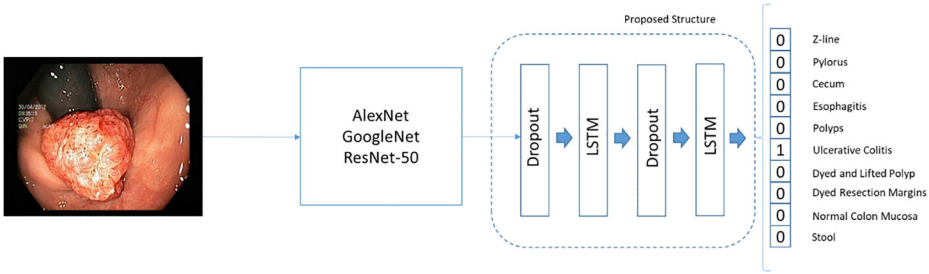


Fig. 2 Proposed Stacked LSTM structure with two LSTM layers and dropout layers

3 Results

3.1 Data

The Kvasir dataset consists of endoscopy images with pathological or regular findings labeled by specialist physicians. There are 8000 images in the Kvasir dataset, which has a total of 8 classes. Pathological findings in the dataset are three classes as esophagitis, polyps, and ulcerative colitis. It contains Z-line, pylorus, and cecum data as anatomical images. There are dyed and lifted polyp and also dyed resection margins labels indicating surgical intervention. Finally, there are healthy Colon Mucosa and Stool structures. Figure 3 includes all categories about the Kvasir dataset.

The Z-line, one of the anatomical regions, forms the transition line between the esophagus and the stomach. Detection of the Z line in this region provides essential information about the presence of the disease. The area defined up to the first parts of the stomach, and the small intestine is called the pylorus. There are muscle structures that facilitate digestion in this region. The cecum is the closest part of the large intestine in human anatomy. Colonoscopy starts with the structure of the cecum. There are three pathological findings in endoscopic procedures. Esophagitis, one of these species, is a type of inflammation seen in the Z-line. Such findings occur in cases of reflux. Treatment should be initiated to stop the progression of the disease with clinical findings. Polyps are mostly lesions in the intestine.

Polyps can be distinguished from normal mucosa by their color and pattern. Although most polyps are benign, there is a possibility that they will develop into cancer. Ulcerative Colitis is a chronic disease that has a direct effect on the large intestine. It may have a direct impact on the patient's standard of living. In addition to the pathological findings, normal findings can be seen as Normal Colon Mucosa and Stool.

3.2 Experimental results

The proposed method is trained on a computer with Intel Core i7-7700K CPU (4.2 GHz), 32 GB DDR4 RAM and NVIDIA GeForce GTX 1080 graphic card.

Three different CNN architectures are used to prove the effectiveness of the proposed model. In the training process, 2500, 5000, and 7500 samples are used. The reason for this process is to reveal the performance of the proposed classification approach, even with fewer examples. Also, training-validation graphs are presented in detail for a better understanding of learning speeds: Figs. 4, 5 and 6 show training and loss graphs. The blue lines in the training curves represent the accuracy of the training data. To detect the overfitting problem in CNN

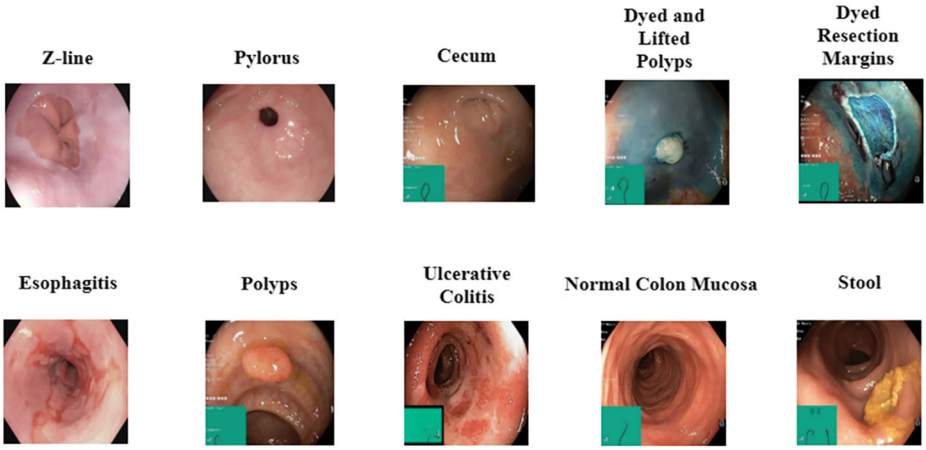


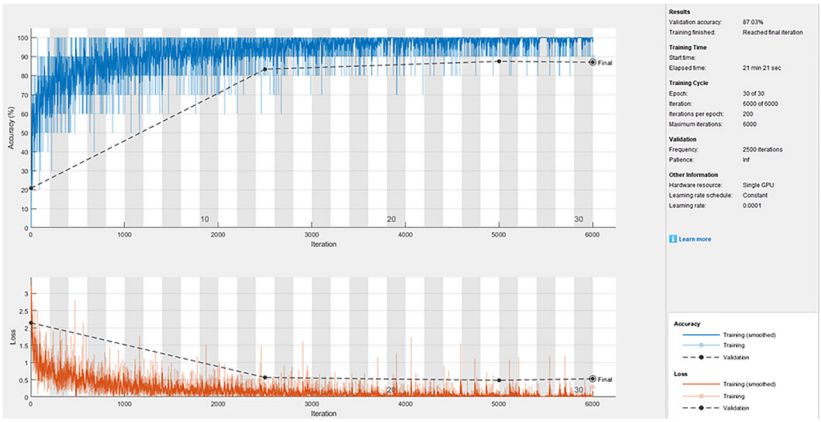
Fig. 3 Sample images for each class from the Kvasir dataset

networks, they are tested at specific iterations. The black lines in the curves show the accuracy of the test data. In the loss graph, the red lines indicate the loss of the training data, and the black lines indicate the loss of the validation data.

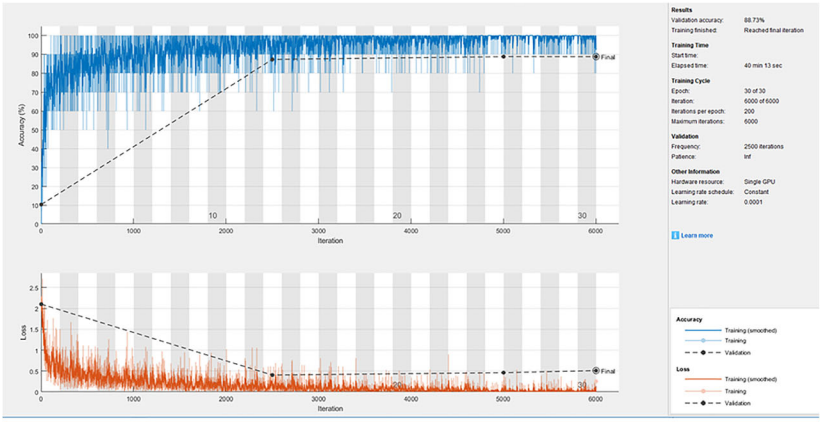
As can be seen from the training curves, the low number of training samples reduces the success. However, as we understand from the validation curves, the problem of memorization does not occur with the proposed LSTM framework. Notably, the error curve in Fig. 4, where the number of data is low, reaches saturation very quickly. The main reason for this is that they do not have enough data to train the CNN architecture strongly. It can be seen from Fig. 4 that the LSTM structure is quite successful without the problem of memorization. But the learning process has stopped. The increase in the number of samples allowed the learning process to continue even in the last epochs, as shown in Figs. 5 and 6. From the changes in the error curve, this process can be seen very clearly. Besides, the increase in the number of samples increases the performance of classification. The comparative results of the proposed method are shown in Table 1.

When Table 1 is examined, it is seen that the SVM classifier is generally more successful than Softmax. The lowest accuracy rates are obtained in the softmax classifier by using 2500 training data. These rates are 87.03% for AlexNet+softmax, 88.79% for GoogleNet+softmax and 90.38% for ResNet-50 + softmax. The highest success rate for the Softmax classifier is obtained by using 7500 training data at 91.55% in the ResNet-50 + softmax model. The lowest accuracy for the SVM classifier is obtained when 2500 training data are used. These accuracies are 91.07%, 91.53%, and 92.93%, respectively. When Table 1 is evaluated, the highest classification accuracy is 97.90% by ResNet-50 + proposed structure with 7500 images.

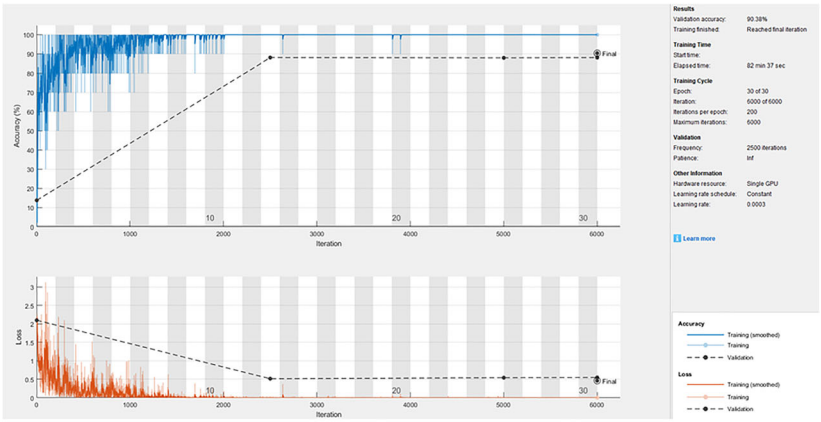
In terms of training time, our proposed method has a shorter training time than other pre-trained CNNs. ResNet-50 + Softmax structure has the longest training period among the pre-trained networks. The training time of this model is 241 min 46 s with 7500 training data and 30 epochs. The shortest training time belongs to AlexNet + Softmax architecture with 30 epochs and 2500 training data. It measured as 21 min 21 s. The proposed method with the ResNet-50 model showed the highest performance with a classification accuracy of 97.90%, as in Fig. 6 (c). Also, its training was completed in 14 min 53 s with 50 epochs and 7500 training data.



a)

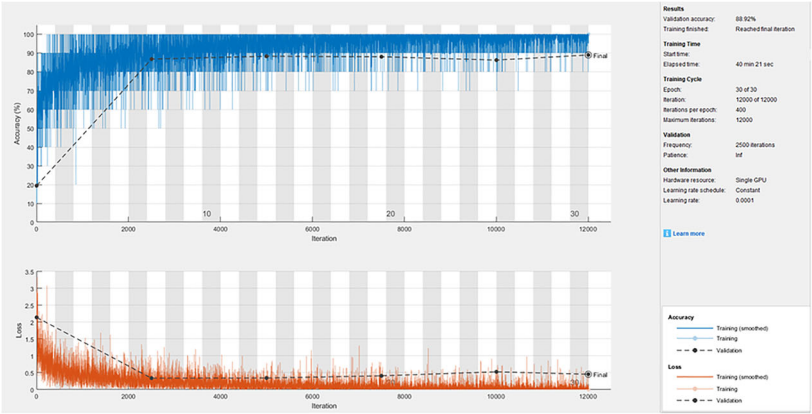


b)

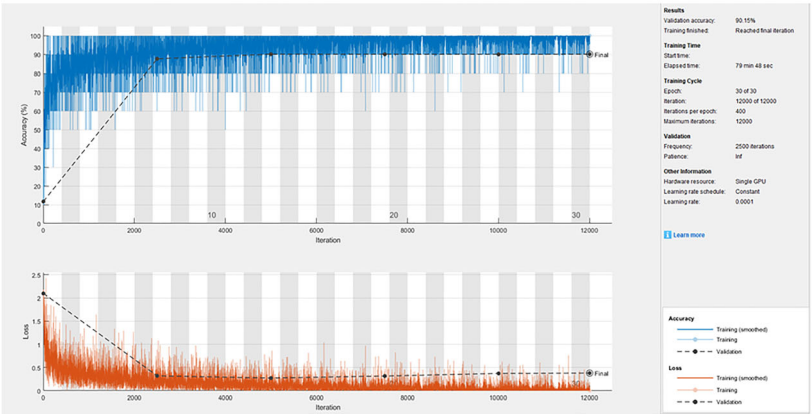


c)

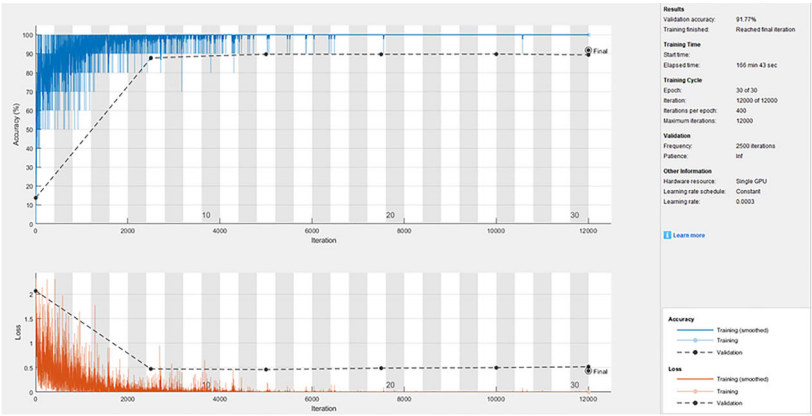
Fig. 4 Training of 2500 Images for (**a**) AlexNet (**b**) GoogleNet (**c**) ResNet-50 with Softmax



a)

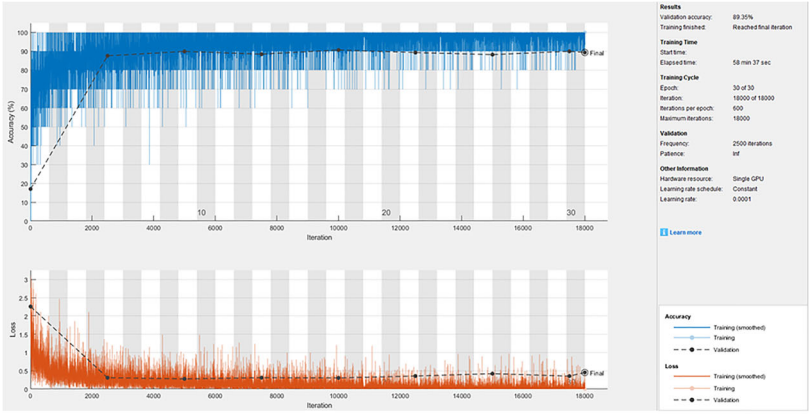


b)

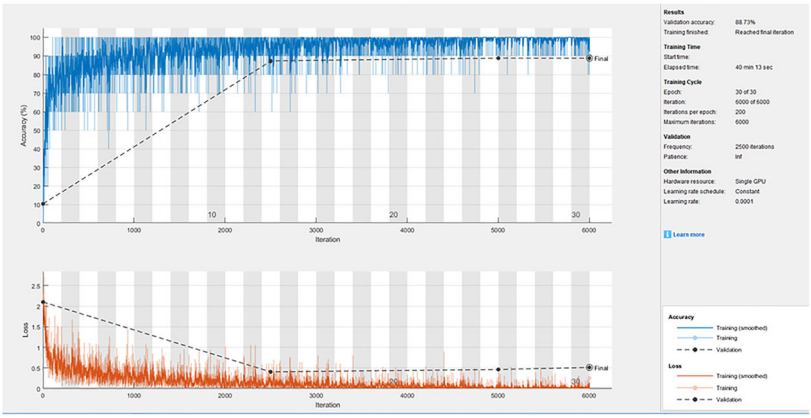


c)

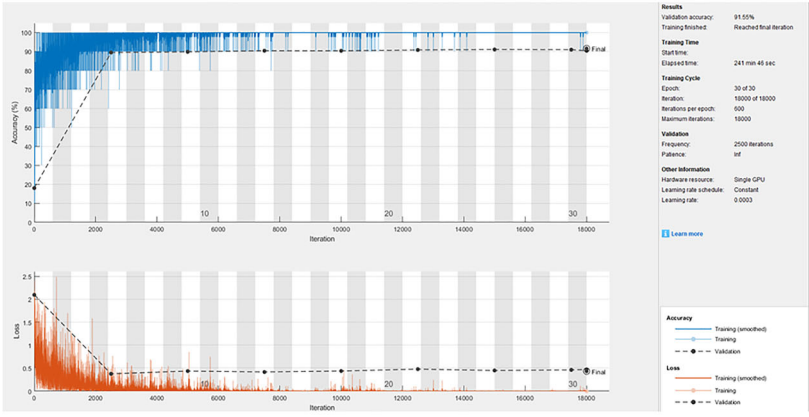
Fig. 5 Training of 5000 Images for (**a**) AlexNet (**b**) GoogleNet (**c**) ResNet-50 with Softmax



a)



b)



c)

Fig. 6 Training of 7500 Images for (**a**) AlexNet (**b**) GoogleNet (**c**) ResNet-50 with Softmax

Table 1 Comparison of classification performance and computational complexity of the proposed framework with the other classifiers

Classifier	Number of training data	Accuracy of Alexnet (%)	Accuracy of GoogleNet (%)	Accuracy of ResNet-50 (%)
Softmax	2500	87.03	88.73	90.38
	5000	88.92	90.15	91.77
	7500	89.35	88.73	91.55
SVM	2500	91.07	91.53	92.93
	5000	94.33	94.68	95.43
	7500	97.50	96.50	97.65
Proposed method	2500	90.37	90.28	93.02
	5000	94.50	94.58	95.45
	7500	96.95	97.15	97.90

Table 2 shows the performances of deep architectures in test data. It makes more sense if you evaluate this table from three different perspectives. When evaluated first in terms of classifiers, the proposed method contributes to the whole number of data and almost all performance parameters. If this contribution is expressed numerically, it is around 4% on average. The second evaluation criterion is the number of training samples. As stated earlier, the high number of training samples contributes positively to test performance. Finally, the proposed method should be evaluated based on deep architectures. This stage clearly shows

Table 2 The classification performance of deep structures for test data

CNN model	Classifier	Number of training data	Sensitivity (%)	Specificity (%)	Precision (%)	F-score (%)
AlexNet	Softmax	2500	79.95	97.13	87.89	80.73
		5000	81.38	97.41	88.51	82.1
		7500	82.66	97.76	89.30	82.94
	SVM	2500	84.56	98.09	89.93	84.97
		5000	86.94	98.40	90.92	87.53
		7500	91.10	98.92	93.71	91.60
	Proposed Method	2500	83.16	97.97	89.87	83.58
		5000	87.05	98.61	91.09	87.74
		7500	89.48	98.91	92.46	89.99
GoogleNet	Softmax	2500	81.04	97.24	88.21	87.41
		5000	82.90	97.74	89.28	83.16
		7500	81.1	97.11	88.02	87.30
	SVM	2500	84.72	98.25	90.01	85.10
		5000	87.53	98.47	91.19	88.13
		7500	89.32	98.70	92.07	89.85
	Proposed Method	2500	83.08	97.86	89.55	83.23
		5000	87.19	98.76	91.21	87.90
		7500	90.51	98.85	93.38	91.07
ResNet-50	Softmax	2500	83.19	98.02	89.96	83.65
		5000	85.15	98.17	90.16	85.64
		7500	84.94	98.21	89.97	85.07
	SVM	2500	85.75	98.25	90.40	86.29
		5000	88.13	98.55	91.47	88.72
		7500	91.70	99.00	94.07	92.11
	Proposed Method	2500	85.83	98.34	90.52	86.43
		5000	88.26	98.71	91.60	88.86
		7500	92.32	99.10	94.46	92.64

Table 3 Comparison of classification performance of proposed method with other state-of-the-art methods

Method	Accuracy (%)
Mahmood et al. [21]	95.4
Shin and Balasingham [30]	95
Zhang et al. [45]	88.6
Urban et al. [39]	96.4
Borgli et al. [6]	97.0
Zhang et al. [46]	90.4
Agrawal et al. [1]	83.8
Gamage et al. [12]	97.38
Proposed Method	**97.90**

that if the proposed method uses it in conjunction with other architectures, it will make a similar contribution. The most successful algorithm in Table 2 is ResNet-50 architecture, thanks to its depth and residual structure.

Table 3 contains the results of the classification studies performed in the Kvasir dataset. When Table 3 is examined, it is seen that the proposed method produces more successful classification results than other state-of-the-art methods. Accordingly, the proposed method for the classification of datasets containing a labeled data problem such as the medical image dataset is quite successful. Also, it is understood that the proposed method will be successful in datasets with data number imbalance between classes.

When Table 3 is examined, Agrawal et al. have the lowest performance with 83.8% accuracy [1]. There are two studies under 90% accuracy in the literature. Gamage et al. [12] used the Kvasir dataset for training as in our study. They have achieved 97.38% accuracy for test data. The proposed method showed higher performance than other studies in the literature with an accuracy of 97.90%.

4 Conclusion

In this study, a robust classification method is presented for datasets that contain a small number of labeled data or have a sample number of imbalance between classes. When the proposed method is subdivided into feature extraction and classification of extracted features, our most important contribution is to reveal the importance of the classification part for the GI tract. A proposed framework consisting of two LSTM layers is designed for this process. Also, the dropout layer is added to avoid the problem of memorization in the LSTM layer. To test the performance of the proposed approach, the common CNN architectures in the literature, AlexNet, GoogLeNet, and ResNet are used. The results show that even in a small number of data, better performance is produced then state-of-the-art studies in the literature without the problem of memorization.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Human and animal rights The paper does not contain any studies with human participants or animals performed by any of the authors.

References

1. Agrawal T, Gupta R, and Narayanan S (2019) On evaluating CNN representations for low resource medical image classification," arXiv e-prints, https://ui.adsabs.harvard.edu/abs/2019arXiv190311176A, [March 01,, 2019].
2. Ahmad J, Muhammad K, Lee M, and Baik S. W. J. J. o. M. S. (2017) Endoscopic image classification and retrieval using clustered convolutional features, vol. 41, no. 12, pp. 196, October 30, 2017.
3. Al-Bulushi NI, King PR, Blunt MJ, Kraaijveld M (2012) Artificial neural networks workflow and its application in the petroleum industry. Neural Computing and Applications 21(3):409–421 2012/04/01
4. Bengio Y, Simard P, Frasconi P (Mar, 1994) Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 5(2):157–166
5. Bernal J, Sánchez J, Vilariño F (2012) Towards automatic polyp detection with a polyp appearance model. Pattern Recognition 45(9):3166–3182 2012/09/01/
6. Borgli RJ, H. K. Stensland, M. A. Riegler, and P. Halvorsen (2019) Automatic hyperparameter optimization for transfer learning on medical image datasets using bayesian optimization." pp. 1–6.
7. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (Nov, 2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68(6):394–424
8. Chowdhury T, Ghita O, Whelan P (2005) A statistical approach for robust polyp detection in CT colonography. Conf Proc IEEE Eng Med Biol Soc 3:2523–2526
9. Cogan T, Cogan M, Tamil L (Aug, 2019) MAPGI: accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning. Comput Biol Med 111:103351
10. Cogan T, Cogan M, Tamil L (2019) MAPGI: Accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning. Computers in Biology and Medicine 111:103351 2019/08/01/
11. DavidE, R. Boia, A. Malaescu, and M. Carnu (2020) Automatic colon polyp detection in endoscopic capsule images. pp. 1–4.
12. Gamage C, I. Wijesinghe, C. Chitraranjan, and I. Perera (2019) GI-Net: Anomalies classification in gastrointestinal tract through endoscopic imagery with deep learning. pp. 66–71.
13. Ghatwary N, Ye X, Zolgharni M (2020) Esophageal abnormality detection using DenseNet based faster R-CNN with Gabor features. IEEE Access 7:84374–84385
14. Habibzadeh M, Jannesari M, Rezaei Z, Baharvand H, Totonchi M (2018) Automatic white blood cell classification using pre-trained deep learning models: ResNet and Inception, p.^pp. SPIE, MV
15. Hinton GE, Osindero C, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," Neural Computation, vol. 18, no. 7, pp. 1527–1554, 2006/07/01, 2006.
16. Hwang S, Oh J, Tavanapong W, Wong J, and d. Groen PC(2020) Polyp detection in colonoscopy video using elliptical shape feature." pp. II - 465-II - 468.
17. Kang J, Gwak J (2019) Ensemble of Instance Segmentation Models for polyp segmentation in colonoscopy images. IEEE Access 7:26440–26447
18. Karnes WE, Alkayali T, Mittal M, Patel A, Kim J, Chang KJ, Ninh AQ, Urban G, Baldi P (2017) Su1642 automated polyp detection using deep learning: leveling the field. Gastrointest Endosc 85(5):AB376–AB377
19. Kirkerød M, Borgli RJ, Thambawita V, Hicks S, Riegler MA, and Halvorsen P (2019) Unsupervised preprocessing to improve generalisation for medical image classification. pp. 1–6.
20. Krizhevsky A, Sutskever I, and Hinton GE (2012) ImageNet classification with deep convolutional neural networks," in Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, Lake Tahoe, Nevada, pp. 1097–1105.
21. Mahmood F, Yang Z, Ashley T, and Durr NJ (2018) Multimodal densenet, arXiv e-prints, https://ui.adsabs.harvard.edu/abs/2018arXiv181107407M, [November 01, 2018, 2018].
22. Mikolov T, Kombrink S, Burget L, Černocký J, and Khudanpur S (2011) Extensions of recurrent neural network language model. pp. 5528–5531.
23. Pei SC, Cheng CM (1999) Color image processing by using binary quaternion-moment-preserving thresholding technique. IEEE Trans Image Process 8(5):614–628
24. Pogorelov K, Randel KR, Griwodz C, Eskeland SL, d. Lange T, Johansen D, Spampinato C, Dang-Nguyen D-T, Lux M, Schmidt PT, Riegler M, #229, and Halvorsen I (2017) KVASIR: a multi-class image dataset for computer aided gastrointestinal disease detection, in Proceedings of the 8th ACM on Multimedia Systems Conference, Taipei, Taiwan, pp. 164–169.
25. Pogorelov K, Ostroukhova O, Jeppsson M, Espeland H, Griwodz C, d. Lange T, Johansen D, Riegler M, and Halvorsen P (2018) Deep learning and hand-crafted feature based approaches for polyp detection in medical videos. pp. 381–386.

26. Razavian AS, Azizpour H, Sullivan J, and Carlsson S (2014) CNN Features Off-the-Shelf: An Astounding Baseline for Recognition." pp. 512–519.
27. Ribeiro E, Häfner M, Wimmer G, Tamaki T, Tischendorf JJW, Yoshida S, Tanaka S, and Uhl A (2017) Exploring texture transfer learning for colonic polyp classification via convolutional neural networks. pp. 1044–1048.
28. Shao L, Zhu F, Li X (May, 2015) Transfer learning for visual categorization: a survey. IEEE Trans Neural Netw Learn Syst 26(5):1019–1034
29. Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-k, and Woo W-c (2015) Convolutional LSTM Network: a machine learning approach for precipitation nowcasting," in Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, Montreal, Canada, pp. 802–810.
30. Shin Y, Balasingham I (2018) Automatic polyp frame screening using patch based combined feature and dictionary learning. Computerized Medical Imaging and Graphics 69:33–42 2018/11/01/
31. Shin H, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 35(5):1285–1298
32. Shin Y, Qadir HA, Aabakken L, Bergsland J, Balasingham I (2018) Automatic Colon polyp detection using region based deep CNN and Post learning approaches. IEEE Access 6:40950–40962
33. Siegel RL, Miller KD, Jemal A (Jan, 2019) Cancer statistics, 2019. CA Cancer J Clin 69(1):7–34
34. Sudharshan PJ, Petitjean C, Spanhol F, Oliveira LE, Heutte L, Honeine P (2019) Multiple instance learning for histopathological breast cancer image classification. Expert Systems with Applications 117:103–111, 2019/03/01/
35. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, and Rabinovich A (2014) Going deeper with convolutions," arXiv e-prints, https://ui.adsabs.harvard.edu/abs/2014arXiv1409.4842S, [September 01, 2014, 2014].
36. Tajbakhsh N, Gurudu SR, and Liang J (2014) Automatic polyp detection using global geometric constraints and local intensity variation patterns. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014. pp. 179–187.
37. Targ S, Almeida D, and Lyman K (2016) Resnet in resnet: generalizing residual architectures. arXiv e-prints, https://ui.adsabs.harvard.edu/abs/2016arXiv160308029T, [March 01, 2016, 2016].
38. Tulum G, Bolat B, Osman O (Apr, 2017) A CAD of fully automated colonic polyp detection for contrasted and non-contrasted CT scans. Int J Comput Assist Radiol Surg 12(4):627–644
39. Urban G, Tripathi P, Alkayali T, Mittal M, Jalali F, Karnes W, and Baldi P, "Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy," Gastroenterology, vol. 155, no. 4, pp. 1069–1078.e8, 2018/10/01/, 2018.
40. Wang Z, Li L, Anderson J, Harrington DP, and Liang Z (2004) Computer-aided detection and diagnosis of colon polyps with morphological and texture features, p.^pp. MI: SPIE.
41. Yi-Min H, and Shu-Xin D (2005) Weighted support vector machine for classification with uneven training class sizes. pp. 4365–4369 Vol. 7.
42. Yu L, Chen H, Dou Q, Qin J, Heng PA (2017) Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos. IEEE Journal of Biomedical and Health Informatics 21(1):65–75
43. Yuan Y, Li B, Meng MQ (2016) Improved bag of feature for automatic polyp detection in wireless capsule endoscopy images. IEEE Trans Autom Sci Eng 13(2):529–535
44. Zeng X, Wen L, Liu B, and Qi X (2019) Deep learning for ultrasound image caption generation based on object detection Neurocomputing, 2019/04/27/.
45. Zhang R, Zheng Y, Poon CCY, Shen D, Lau JYW (2018) Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker. Pattern Recognition 83:209–219, 2018/11/01/
46. Zhang X, Chen F, Yu T, An J, Huang Z, Liu J, Hu W, Wang L, Duan H, Si J (2019) Real-time gastric polyp detection using convolutional neural networks. PLoS One 14(3):e0214133–e0214133
47. Zhao L, Botha CP, Bescos JO, Truyen R, Vos FM, Post FH (Sep-Oct, 2006) Lines of curvature for polyp detection in virtual colonoscopy. IEEE Trans Vis Comput Graph 12(5):885–892