

Decomposition of color wavelet with higher order statistical texture and convolutional neural network features set based classification of colorectal polyps from video endoscopy

A. S. M. Shafi, Mohammad Motiur Rahman

Department of Computer Science and Engineering, Mawlana Bhashani Science and Technology University, Bangladesh

Article Info

Article history:

Received Apr 21, 2019

Revised Nov 25, 2019

Accepted Dec 1, 2019

Keywords:

Colorectal cancer
Convolutional neural network
Gastrointestinal lesions
Run length matrix
Support vector machine
Video endoscopy
Wavelet

ABSTRACT

Gastrointestinal cancer is one of the leading causes of death across the world. The gastrointestinal polyps are considered as the precursors of developing this malignant cancer. In order to condense the probability of cancer, early detection and removal of colorectal polyps can be cogitated. The most used diagnostic modality for colorectal polyps is video endoscopy, but the accuracy of diagnosis mostly depends on doctors' experience that is crucial to detect polyps in many cases. Computer-aided polyp detection is promising to reduce the miss detection rate of the polyp and thus improve the accuracy of diagnosis results. The proposed method first detects polyp and non-polyp then illustrates an automatic polyp classification technique from endoscopic video through color wavelet with higher-order statistical texture feature and Convolutional Neural Network (CNN). Gray Level Run Length Matrix (GLRLM) is used for higher-order statistical texture features of different directions ($\Theta=0^0, 45^0, 90^0, 135^0$). The features are fed into a linear support vector machine (SVM) to train the classifier. The experimental result demonstrates that the proposed approach is auspicious and operative with residual network architecture, which triumphs the best performance of accuracy, sensitivity, and specificity of 98.83%, 97.87%, and 99.13% respectively for classification of colorectal polyps on standard public endoscopic video databases.

Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

A. S. M. Shafi,
Department of Computer Science and Engineering,
Mawlana Bhashani Science and Technology University,
Santosh, Tangail-1902, Bangladesh.
Email: shafi.cse.mbstu11@gmail.com

1. INTRODUCTION

Colorectal cancer or bowel is the second most prominent cause of cancer in women and the third most leading cause in men [1]. Although colorectal polyps are forerunners to colorectal cancer, for these polyps it takes several years to potentially transform into cancer [2], which may develop if these polyps are left untreated. Major types of colorectal polyps are adenomas, hyperplastic and serated. If early detection and classification of these polyps is possible, they can be removed before this transmission arises. Several tests are recommended in all colon cancer screening guidelines, particularly video endoscopy and fecal occult blood testing. Within the United States, video endoscopy is the most commonly utilized test and should be performed every 10 years in average-risk persons. During an endoscopy, a long, flexible tube (colonoscope) is inserted into the body. A tiny video camera at the tip of the tube allows the doctor to view the inside of the entire colon to detect and remove a polyp. Distinguishing from low-risk polyps with high-risk colorectal polyps is an important part of colorectal cancer screening through the detection and histopathological characterization of colorectal polyps. The generation of such a typical endoscopy video proceeds for a long period. For back-to-back endoscopy, it is so tough for an endoscopist to examine it with sufficient

attentiveness during such a long period, as it is an operator-dependent procedure. The accuracy of such a challenging diagnosis depends on a qualified physician. There is a large degree of variability for how the physician characterizes and diagnose these polyps, as not all polyps have same malignant influence. As an example, serrated polyps can potentially develop more aggressively into colorectal cancer as compared to other colorectal polyps, because of the serrated pathway in tumorigenesis [3]. There are only consistent prevailing methods for diagnosing serrated polyps is histopathological characterization because other screening methods designed to detect premalignant lesions (such as fecal blood, fecal DNA, or virtual colonoscopy) and are not well matched for differentiating serrated polyps from other polyps [4]. The challenging task for a doctor is to differentiate between a serrated polyps and hyperplastic polyps. This is because hyperplastic polyps often lack the dysplastic nuclear changes that characterize conventional adenomas polyp, and their histopathological diagnosis of hyperplastic polyps is entirely based on morphological features, such as serration, dilatation, and branching and often lack the dysplastic nuclear changes that depict conventional adenomas polyps [5].

In the field of artificial intelligence, convolutional neural network models have been proposed and applied for computer-aided polyp detection and classification system. Color wavelet covariance (CWC) of different color bands based on the covariance of second-order textual measures have been used to the detection of tumors in colonoscopic video with a specificity and sensitivity of 97% and 90% respectively [6]. Intelligent processing techniques of SVMs and color-texture analysis methodologies have been proposed in their consecutive work for automatic detection of gastrointestinal adenomas in video endoscopy having the accuracy of 94% [7]. 94.20% accuracy was obtained with the combination of color and shape features to recognize intestinal polyp from capsule endoscopy as a classifier of multilayer perceptron (MLP) [8]. A deep convolutional neural network-based classification for digestive organs in wireless capsule endoscopy images was considered in Y. Zou *et al.* [9]. A trainable feature extractor based on a convolutional neural network is utilized for lesion detection from endoscopy images in R. Zhu *et al.* [10]. In paper [11] CNN features have been proposed for the automated classification of colonic mucosa for colon polyp staging in the context of colon cancer screening with a sensitivity of 95.16% and specificity of 74.19%. The CNN-derived features show greater invariance to viewing angles and image quality factors when compared to the eigenimage model [12]. Color wavelet (CW) features and convolutional neural network features of video frames are extracted and combined which are used to train a linear support vector machine, gaining accuracy of 98.65%, sensitivity of 98.79% and specificity of 98.52% [13]. Authors [14] enhanced automatic polyp detection accuracy by utilizing feature fusion (wavelet, local binary pattern, and Gabor features) and multiple classifier techniques. They achieved 80% true positive rate by incorporating local binary pattern and wavelet features. Authors [15] relies on a faster region-based convolutional neural network (Faster R-CNN) model for polyps' detection in endoscopic videos. Tajbakhsh *et al.* [16] focused on a novel vote accumulation scheme for detecting colonic polyps that enable polyp detection from partially identified boundaries of polyps. However, several novel technologies are emerging within the field of endoluminal imaging and confocal laser endomicroscopy to perform the imaging of the colorectal area, the number of computer-aided decision support system (CADSS) related to colorectal polyps' classification is still limited. Hence, the main contribution of this paper is to effectively use machine learning to successfully detect and classify colorectal polyps presented in endoscopy video. However, the other intentions of this paper include the followings:

- A novel intelligent approach utilizing higher-order statistical texture feature on the color wavelet and convolutional neural network learned from video endoscopy dataset.
- The proposed approach is favorable and effective, which achieves the best performance of accuracy, compare to only CNN feature-based method.
- Drastically reducing the necessary time for a qualified physician to examine the entire endoscopy video, by indicating colorectal polyps and/or classify them.

The rest of the paper is organized as follows: section 2 describes the architecture and methods of the proposed system. Section 3 deals with the analyzing of experimental results. Finally, discussion of the results as well as the conclusions of this study is summarized in sections 4 and 5 respectively.

2. SYSTEM ARCHITECTURE

The implementation of the proposed system is based on MATLAB 2017b that can accept standard video files of different formats such as AVI, MP4, and WVM as input and produce outputs with a classified polyp in the video frame sequence. When endoscopy video is fed into the proposed system, it utilizes color wavelet with higher-order statistical texture image features and convolutional neural network features. The fusion of all the features is incorporated into SVMs to achieve improved detection and classification

accuracy. There are five stages of this section namely the acquisition of video endoscopy with normal lesions and different types of colorectal polyps, pre-processing, feature extraction, classification, and post-processing module as shown in Figure 1.

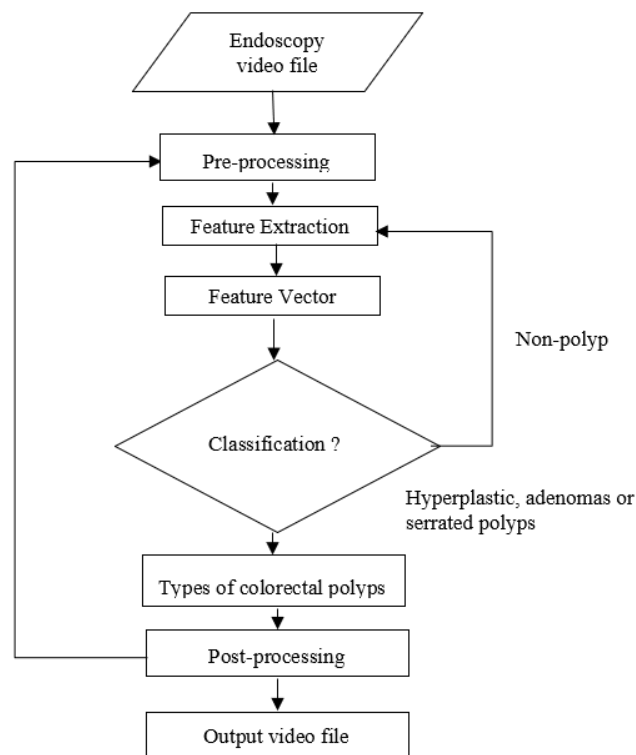


Figure 1. A framework of the proposed approach

2.1. Acquisition of video endoscopy

The availability of endoscopy datasets is an important issue for this purpose. However, the performance of any Computer-aided design (CAD) system depends on the training dataset, this proposed system utilizes more than 86 standard endoscopy videos from different sources. Sample dataset of different polyp classes and normal lesion are shown in Figure 2. Most of the endoscopy video embrace colorectal polyps and only small part of the endoscopy video are associated with the normal colon. Important sources of the database include Department of Electronics, University of Alcala (http://www.depeca.uah.es/colonoscopy_dataset/) [17], Endoscopic Vision Challenge (<https://polyp.grand-challenge.org/databases/>) [18]. Table 1 shows the dataset collected from video endoscopy. For our experiment, we have extracted 4,216 frames from endoscopy video, which consists of 3,162 colorectal polyps, and 1,054 normal frames.

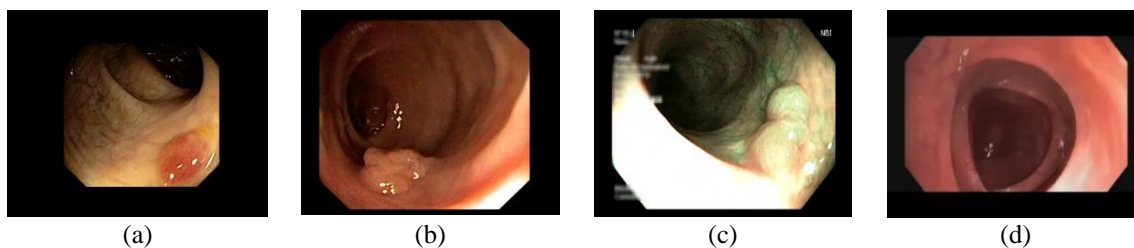


Figure 2. Sample dataset: (a) hyperplastic polyps, (b) adenomas polyps, (c) serrated polyps, and (d) normal lesion

Table 1. Our dataset: the distribution of colorectal polyp types in the pre-processed image frame used in our proposed system

Colorectal polyps' type	Short-form	Number of endoscopy video	Size of video endoscopy (MB)
Hyperplastic polyps	Hp	21	40.8
Adenomas polyps	Ad	40	126
Serrated polyps	Sr	15	36.4
Normal lesion	-	10	32.6
Total	-	86	235.8

80% (3,372 video frame) of this dataset was used for training, while the remaining 20% (844 video frames, 211 from each class) was used for validation.

2.2. Pre-processing module

Endoscopy video is loaded in the computer to find possible categories of colorectal polyps that can produce the running sequence of still images called frame using the proposed intelligent system. Original video frame as shown in Figure 3(a) with superfluous regions are discarded resulting in a pre-processed frame as shown in Figure 3(b) in order to exclusion of irrelevant textual information, such as patient’s name, date of birth and time of the examination.



Figure 3. Pre-processing: (a) original video frame and (b) pre-processed frame

2.3. Feature extraction module

A sliding window of user-defined size and sliding step is slid over the pre-processed video frames to generate small images called window as shown in Figure 4. Depending on the size, dimensions, and sliding step, each window produces a number of feature vectors.

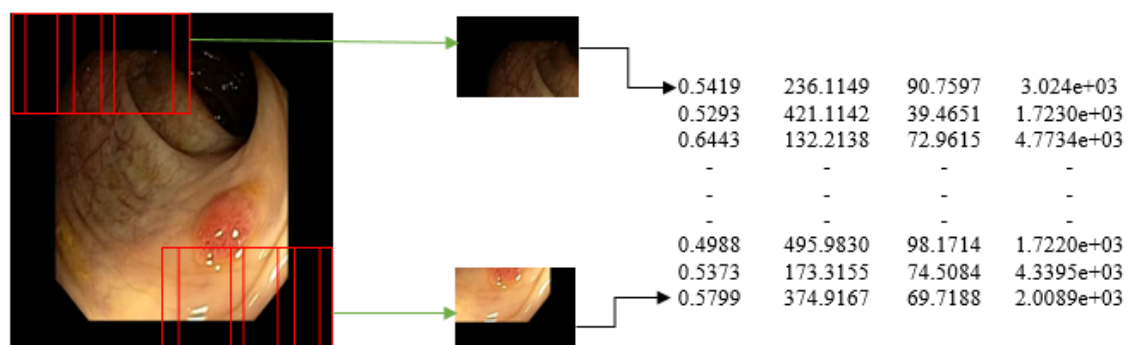


Figure 4. Feature extraction technique

2.3.1. Higher order statistics on the wavelet domain as gray scale texture features

Multiresolution analysis of an image is achieved by using a discrete wavelet transform for our proposed framework as the size of the polyp varies. The most relevant texture information often appears in middle-frequency channels [19]. Texture provides information about the spatial arrangement of colors or intensities in an image that helps image segmentation and classification. Wavelets perform well for texture analysis. The decomposition of the wavelet transform provides spatial/frequency representation from original

images where every sub-image preserves both local and global information of a specific scale and orientation. When decomposition level decreases in the spatial domain, it increases in the frequency domain providing zooming capabilities and local characterization of the image [6]. Since major texture information produced by this transformation does not contain in the low-frequency image and most of the substantial texture information often looks as if in the middle frequency channel, so our proposed approach uses a discrete wavelet transform (DWT) for the decomposition of the frequency domain of the image.

A two-dimensional (2-D) discrete wavelet transformation is obtained by applying the filtering consecutively along with vertical and horizontal directions (separable filter bank). This filtering procedure is grasped by convolving the image with a high pass filter H and a low pass filter L which produces a low-resolution image $B_0(k)$ at scale k and detail images $D_j(k)$, $j = 1, 2, 3$, at scale k as described by the following recursive [19]:

$$\begin{aligned} B_0(k) &= \{ L_x * [L_y * B_0(k-1)]_{\downarrow 2,1} \}_{\downarrow 1,2}. \\ D_1(k) &= \{ H_x * [L_y * B_0(k-1)]_{\downarrow 2,1} \}_{\downarrow 1,2}. \\ D_2(k) &= \{ H_x * [H_y * B_0(k-1)]_{\downarrow 2,1} \}_{\downarrow 1,2}. \\ D_3(k) &= \{ L_x * [H_y * B_0(k-1)]_{\downarrow 2,1} \}_{\downarrow 1,2} \end{aligned} \quad (1)$$

where arrow (\downarrow) denotes the sub-sampling procedure, the asterisk (*) is the convolution operator and H and L are the two filters for all $k = 1, 2, 3, \dots, n$.

In this paper, the gray level run length matrix has been considered for the description of higher-order statistical texture features implemented within the decomposed sub-images. A run-length matrix P (i, j) for a given image is defined by the specifying direction of $0^\circ, 45^\circ, 90^\circ, 135^\circ$ and then count the occurrence of a run for each gray levels i and run-length j in this direction. We consider only four run-length matrix features namely Short Run Low Gray-Level Emphasis (SRLGE), Short Run High Gray-Level Emphasis (SRHGE), Long Run Low Gray-Level Emphasis (LRLGE), Long Run High Gray-Level Emphasis (LRHGE) [20].

$$SRLGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P(i,j)}{i^2 j^2} \quad (2)$$

$$SRHGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P(i,j) \cdot i^2}{j^2} \quad (3)$$

$$LRLGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P(i,j) \cdot j^2}{i^2} \quad (4)$$

$$LRHGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N P(i,j) \cdot i^2 \cdot j^2 \quad (5)$$

where P(i,j) is the run-length matrix and n_r is the total number of runs.

2.3.2. Higher order color wavelet covariance features

In our proposed approach color texture features extracted by using DWT for the decomposition of the frequency domain of the color image are estimated over the GLRLM. The input image is decomposed into three color channels:

$$I^{C_i}, i = 1, 2, 3. \quad (6)$$

In this work, the frames are decomposed at level 3 using Daubechies 2 (db2) wavelet family. A three-level two dimensional discrete wavelet transformation is consequently applied on each color channel (I^{C_i}), producing a low resolution image $B_i(k)$, at scale k and the detail images $D_i(k)$, where $i = 1, 2, 3$ and $k = 1, 2, \dots, 9$ according to the wavelet decomposition of (1).

Therefore, we have:

$$I^{C_i} = \{ B_i(k), D_i(k) \}, i = 1, 2, 3 \text{ and } k = 1, 2, 3, \dots, 9. \quad (7)$$

where k is the decomposition level.

As it has already been noted, the most significant textual information is presented in the middle wavelet detailed channels. So, we consider only the detail images for $k = 4, 5, 6$. So, the nine different sub-images produced from (7) for the values $k = 4, 5, 6$.

$$D_i(k), \quad i = 1, 2, 3 \text{ and } k = 4, 5, 6. \tag{8}$$

For extracting the higher-order statistical textual information, we consider run-length matrices calculated over the above nine different sub-images. These matrices reflect the spatial relationship between more than two pixels in a definite direction. Run-length matrices generate 36 matrices calculated from four different directions of intensities relation $0^\circ, 45^\circ, 90^\circ, \text{ and } 135^\circ$.

$$R_\alpha\{D_i(k)\}, \quad i = 1, 2, 3, \quad k = 4, 5, 6, \text{ and } \alpha = 0^\circ, 45^\circ, 90^\circ, 135^\circ \tag{9}$$

Finally, the four run-length matrix features namely Short Run Low Gray-Level Emphasis, Short Run High Gray-Level Emphasis, Long Run Low Gray-Level Emphasis, and Long Run High Gray-Level Emphasis are calculated for each 36 matrices resulting in 144 wavelet features.

$$F_m[R_\alpha\{D_i(k)\}], \quad i = 1, 2, 3, \quad k = 4, 5, 6, \quad \alpha = 0^\circ, 45^\circ, 90^\circ, 135^\circ, \text{ and } m = 1, 2, 3, 4. \tag{10}$$

where m is the respective run-length matrix feature.

2.3.3. Convolutional neural network features

Convolutional neural network features are extracted from each window of size $227 * 227$. A convolutional neural network, a hierarchical neural network is the branch of deep learning, can be composed of convolution layers, pooling layers, rectified linear unit (ReLU) layers, fully connected layers, and loss layers. In a simple CNN architecture, a Rectified Linear Unit layer follows each convolution layer. After each convolution layer, there is a max-pooling layer. Finally, one or more fully connected layers, which can be attained after one or more convolution layers. An important characteristic that distinguishes CNN to traditional multilayer perceptron (MLP) models is taking into account the structure of the images while processing them. Due to full connectivity between the layers, MLP models suffer from the curse of dimensionality thus do not scale well to high-resolution images and less sensitive to positional changes. [21] Inspire the structure of CNN used in this paper contains the following representation as shown in Figure 5. Table 2 can describe about the neural network architecture implementation.

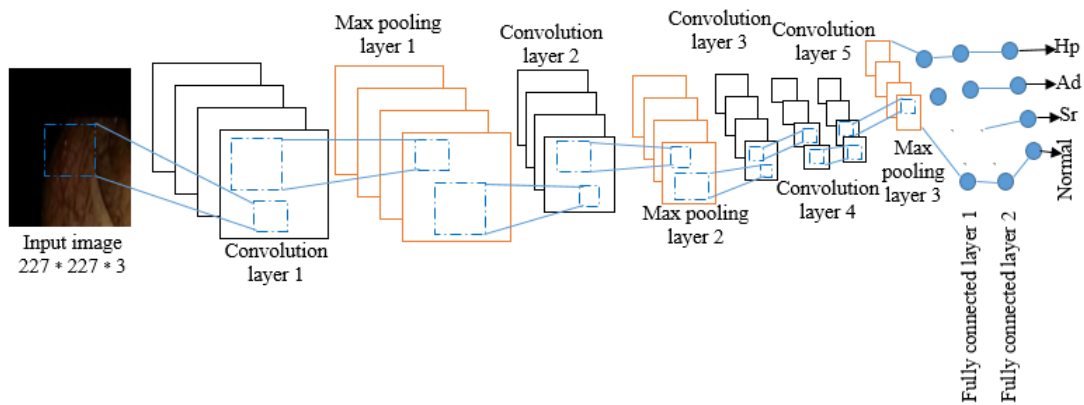


Figure 5. An illustration of the proposed CNN feature extraction module

Table 2. The structure of the neural network used in this paper

Layer Type	Dimensions
Convolution	11 x 11 - Padding: 0 - Stride - 4
Max-pooling	3 x 3 - Stride: 2
Convolution	5 x 5 - Padding: 2 - Stride: 2
Max-pooling	3 x 3 - Stride: 2
Convolution	3 x 3
Convolution	3 x 3
Convolution	3 x 3
Max-pooling	3 x 3 - Stride: 2
Fully Connected	4096
Fully Connected	4096

2.4. Classification module

This module handles the classification of the feature vectors into one of four classes: hyperplastic, adenomas, serrated and normal. For computer-aided histopathological classification systems, many classifiers have been developed such as linear discriminant analysis (LDA) [6], neural networks [8], adaptive neuro-fuzzy inference system [22], binary classifier [23], and support vector machine (SVM) [13, 24, 25]. In this proposed system, SVM has been used for better performance in the case of a large number of features, training, and sparse data. Different types of colorectal polyps' frames are extracted from video endoscopy that operates in two modes: training and testing modes. In training mode, decomposition of color wavelet with GLRLM and CNN features are extracted from a hyperplastic, adenomas, serrated and normal video frames. The input feature vector consists of 144 color wavelet features and 4096 CNN features, which are combined for SVM training. In the testing mode of operation, the classification of new samples is extracted from unknown video frames based on the utilization of knowledge gained from the training samples. If the unknown sample is classified as hyperplastic, adenomas or serrated polyps it goes to the post-processing module otherwise a new subsequent unidentified video frame comes under the feature extraction module.

2.5. Post-processing module

The outcome of the classification module is used to produce new video frame on which possible types of colorectal polyps are appropriately labeled. An illustration of this technique shown in Figure 6.

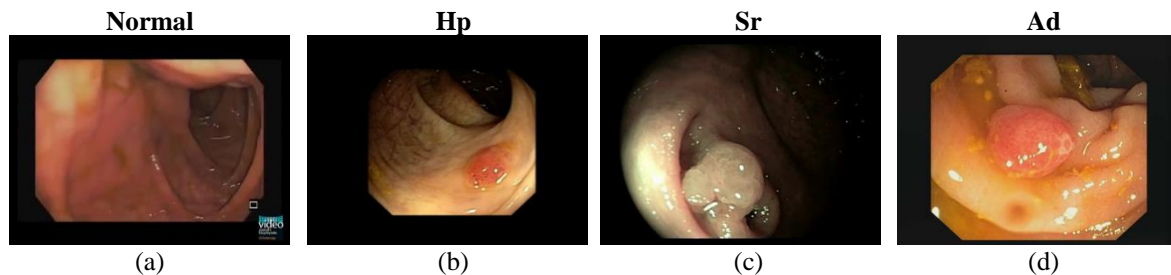


Figure 6. Post-processing module: System's output after post-processing: (a) normal lesion, (b) hyperplastic polyps, (c) serrated polyps, and (d) adenomas polyps

3. RESULTS

To evaluate the performance of the proposed method, we examined three metrics per class: sensitivity (Sen), specificity (Spe), and accuracy (Acc). Sensitivity also called the true positive rate refers to the ability to measure the proportion of actual positives that are correctly identified as such. Specificity refers to the true negative rate, measuring the proportion of actual negatives that are correctly identified as such. Accuracy is derived from sensitivity and specificity and is defined as the sum of true positives and false positives divided by the total number of evaluated cases (true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN)). Twelve different performance metrics were used for classification purposes (Sen, Spe, and Acc for each one of the 4 classes understudy). The final confusion matrices can be seen in Table 3 and the performance analysis with respect to Sen, Spe, and Acc scores across the four different classes is shown in Table 4. In comparison, as is manifested in Table 4, the outcomes indicate that the proposed method (Color wavelet + CNN + GLRLM + SVM) achieves competitive performance not only for detecting polyp and non-polyp but also for classifying colorectal polyps compare to CNN feature-based method.

Table 3. Confusion matrix of our proposed system for different kind of polyp classification

Actual Class	Predicted Class			
	Ad	Hp	Normal Lesion	Sr
Ad	197	6	0	8
Hp	2	209	0	0
Normal Lesion	0	0	211	0
Sr	2	0	0	209

Table 4. Performance analysis of our proposed method in the 4-class problem

Polyp Type	Methodology	Sensitivity	Specificity	Accuracy
Ad	CNN	89.57	98.73	96.44
	Proposed Method	93.36	99.37	97.87
Hp	CNN	98.10	98.89	98.69
	Proposed Method	99.05	99.05	99.05
Normal Lesion	CNN	99.53	99.52	99.52
	Proposed Method	100	100	100
Sr	CNN	97.63	97.78	97.74
	Proposed Method	99.05	98.10	98.41

3.1. Comparison with existing methods

However, the comparison with existing approaches has been extremely difficult for several reasons. First, the widely acceptable research papers of related topics attempt to explain different problems: polyp detection [10, 12, 13] commonly using convolution neural network features. Second, even if the same problem was solved by those methods, there is an entire lack of openly accessible codes. We compare the methods focused on the classification of gastrointestinal lesions in video endoscopy as shown Table 5. In addition, the proposed system is appraised alongside with standard dataset.

Table 5. Comparison among different methods

Author	Used Methodology	Used Dataset	Result		
			Sensitivity	Specificity	Accuracy
Mustain Billah et al. [13]	Color wavelet + CNN + SVM	100 videos	98.65%	98.79%	98.52%
Xi Mo et al. [15]	Faster R - CNN	29 videos	98.1%		
Ribeiro et al. [11]	CNN	100 images	95.16%	74.19%	90.96%
Alaa El Khatib et al. [14]	LBP + Wavelet features	13 videos	80%		
Karkanis et al. [6]	CWC + LDA	60 videos	97%	90%	
Nima Tajbakhsh et al. [16]	Vote accumulation scheme	300 images	86%		
Proposed method	Color wavelet + CNN + GLRLM + SVM	86 videos	97.87%	99.13%	98.83%

3.2. Comparison with human experts

We have considered the diagnostic efficacy of human experts to compare the performance with our proposed approach. Although the range of years of experience performing endoscopies goes from 8 to 40 for the human expert's category, several human issues lead to polyp misdetection and misclassification. The lesions wrongly classified by all the humans are shown in Figure 7. The lesions correctly classified by all humans and the best machine learning model, the ones wrongly classified by the best model and correctly by all humans, and the lesions correctly classified by the best model and wrongly classified by all humans are displayed in Figures 8, 9 and 10, respectively. Text included in the figures shows the lesion name ('Hp', 'Ad', 'Sr', 'N' refers to hyperplastic, adenomas, serrated, and normal lesion respectively).

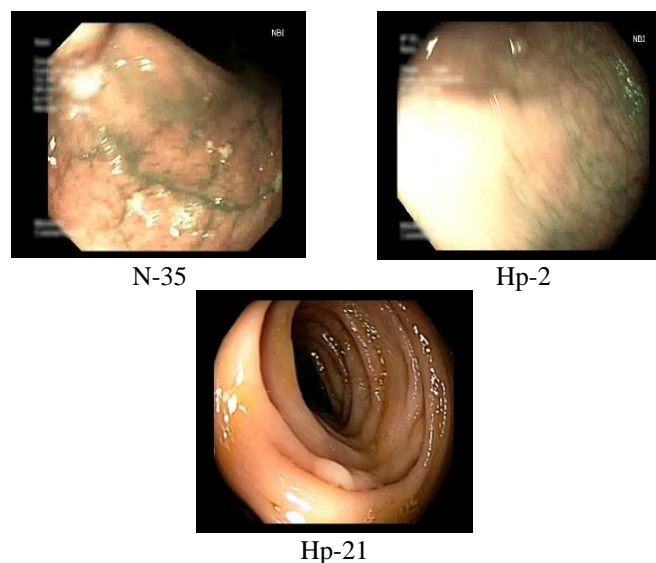


Figure 7. Lesions wrongly classified by all humans

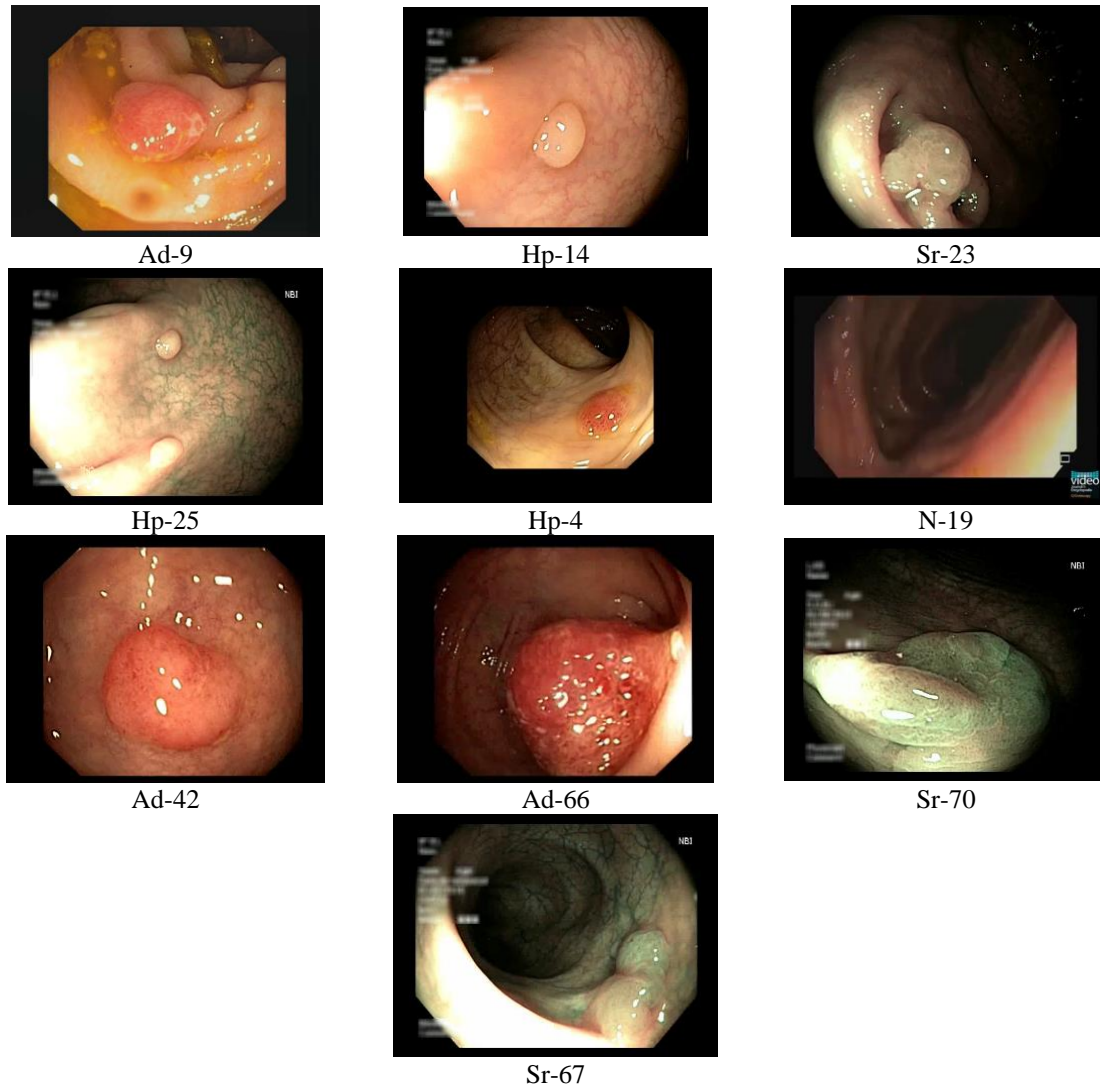


Figure 8. Lesions correctly classified by the proposed model and all humans



Figure 9. Lesions wrongly classified by the proposed model but correctly classified by all humans

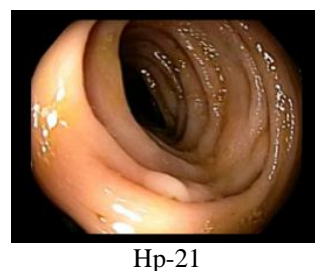


Figure 10. Lesions correctly classified by the proposed model but wrongly classified by all humans

4. DISCUSSION

In this paper, we presented and developed a set of algorithms for computer-aided automatic colorectal polyp detection and classification system from video endoscopy. Although the polyp detection and classification is a disputing task because of its legion factors such as the presence of trash and liquids and bubbles, vignetting and different types of polyp shape, we attempt to overwhelm this problem by

incorporating both higher-order statistical texture and convolutional neural network features. We have used a linear SVM for classification solution for its stable results and faster training instead of a conventional softmax loss layer. Here, the input to the fully connected layer has been used as the input to the SVM classifier. As there are several kinds of important low-level features, we consider that the texture feature suits this paper best. We have used higher-order statistical texture descriptors for feature extraction instead of second-order statistics because higher-order statistics determine the relationship between three or more pixels whereas second-order statistics examine the relationship between two pixels. Therefore, we have considered DWT and GLRLM features for texture recognition. We consider that the CNN feature is another suitable choice for feature extraction as CNN is an end-to-end learning process. On CNN, we provide the dataset and their corresponding labels, the entire process of feature engineering is done. It has both feature extraction and machine learning inside it. The proposed system for colorectal polyp's classification showed that the use of color wavelet with GLRLM and CNN features achieves almost high sensitivity (97.87%) and, equally importantly, displays high specificity (99.13%).

5. CONCLUSIONS AND FUTURE WORKS

In this paper, we tackled a complete pipeline for computer-aided automatic colorectal polyps' detection and classification system capable of supporting the decision of a medical person. Its aims to enhance the ability of an endoscopist to locate colorectal polyps more accurately which may go undetected and evolve into malignant tumors. The system exploits the higher-order statistical texture feature calculated over the strength of wavelet frame transformation of different color bands and convolutional neural network. The classification was performed using a linear SVM classifier.

The results of the extensive experimental study have to lead us to the use of higher-order statistical texture features on the wavelet domain results in higher classification accuracy in terms of specificity and sensitivity. The proposed GLRLM performs significantly better than Gray Level Co-Occurrence Matrix (GLCM) for polyp detection as the Gray Level Run Length Matrix provides information about the connected length of a particular pixel in a definite direction. Convolutional neural networks are a special architecture of artificial neural networks that have been widely used in automatic image classification systems by reducing learning complexity with sharing the weights in different layers. The use of linear SVM kernel positively affects the discrimination of normal and abnormal samples.

Finally, we may conclude that the experimental results showed the enhanced performance of the proposed detector and classifier, compared to other state-of-the-art detectors. In future, we would like to determine a more robust classification scheme based on video endoscopy frames and wireless capsule endoscopy (WCE) with a border range of images containing colorectal polyps of different qualities. We would like to add another functionality that facilitates the improvement (enhancement) of video endoscopy by applying the super-resolution technique.

ACKNOWLEDGEMENTS

We would like to acknowledge Prof. Dr. Md. Zulfikar Ali, Professor and Head, Dept. of Medicine, Khwaja Yunus Ali Medical College and Hospital, Enayetpur, Sirajgonj, Bangladesh for his valuable support, suggestions, and contribution to the evaluation of the results.

REFERENCES

- [1] L.A. Alexandre, N. Nobre, J. Casteleiro, "Polyp Detection in Endoscope Videos Using SVMs," *European Conference on Principles and Practice of Knowledge Discovery in Databases*, pp. 358-36, 2007.
- [2] Wong NA, Hunt LP, Novelli MR, Shephered, Warren BF, "Observer agreement in the diagnosis of serrated polyps of the large bowel," *Histopathology*, vol. 55, no. 1, pp. 63-6, 2009.
- [3] Leggett B, Whitehall V, "Role of the serrated pathway in colorectal cancer pathogenesis," *Gastroenterology*, vol. 138, no. 6, pp. 2088-100, 2010.
- [4] Kahi CJ, "How does the serrated polyp pathway alter CRC screening and surveillance?," *Digestive Diseases and Sciences*, vol. 60, no. 3, pp. 773-80, 2015.
- [5] Korbar B, Olofson AM, Miraflor AP, Nicka CM, Suriawinata MA, Torresani L, "Deep learning for classification of colorectal polyps on whole-slide images," *J Pathol Inform*, vol. 1, no. 30, pp. 1-17, 2017.
- [6] S. A. Karkanis, *et al.*, "Computer-aided tumor detection in endoscopic video using color wavelet features," *IEEE Transactions on Information Technology in Biomedicine*, vol. 7, no. 3, pp. 141-152, 2003.
- [7] D. K. Iakovidis, *et al.*, "An intelligent system for automatic detection of gastrointestinal adenomas in video endoscopy," *Computers in Biology and Medicine*, vol. 36, no. 10, pp. 1084-1103, 2005.

- [8] B. Li, Y. Fan, M. Q.-H. Meng, and L. Qi, "Intestinal polyp recognition in capsule endoscopy images using color and shape features," *IEEE International Conference on Robotics and Biomimetics*, Institute of Electrical and Electronics Engineers, Guilin, China, pp. 1490–1494, 2009.
- [9] Y. Zou, L. Li, Y. Wang, J. Yu, Y. Li, and W. J. Deng, "Classifying digestive organs in wireless capsule endoscopy images based on deep convolutional neural network," *IEEE International Conference on Digital Signal Processing*, Institute of Electrical and Electronics Engineers, Singapore, pp. 1274–1278, 2015.
- [10] R. Zhu, R. Zhang, and D. Xue, "Lesion detection of endoscopy images based on convolutional neural network features," *8th International Congress on Image and Signal Processing*, Institute of Electrical and Electronics Engineers, Shenyang, China, pp. 372–376, 2015.
- [11] E. Ribeiro, A. Uhl, and M. Hafner, "Colonic polyp classification with convolutional neural networks," *IEEE 29th International Symposium on Computer-Based Medical Systems*, Institute of Electrical and Electronics Engineers, Dublin, Ireland, 2016.
- [12] S.Y. Park and D. Sargent, "Colonoscopic polyp detection using convolutional neural networks," *International Society for Optics and Photonics*, of SPIE Medical Imaging, San Diego, Cali, USA, vol. 9785, 2016.
- [13] Mustain Billah, Sajjad Waheed and Mohammad Motiur Rahman, "An Automatic Gastrointestinal Polyp Detection System in Video Endoscopy Using Fusion of Color Wavelet and Convolutional Neural Network Features," *International Journal of Biomedical Imaging*, vol. 2017, pp. 1-9, 2017.
- [14] Alaa El Khatib, Naoufel Wergghi, and Hussain Al-Ahmad, "Enhancing Automatic Polyp Detection Accuracy using Fusion Techniques," *IEEE 59th International Midwest Symposium on Circuits and Systems*, 2016.
- [15] Xi Mo, Ke Tao, Quan Wang, and Guanghui Wang, "An Efficient Approach for Polyps Detection in Endoscopic Videos Based on Faster R-CNN," *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018.
- [16] N. Tajbaksh, S. Gurudu, and J. Liang, "A classification-enhanced vote accumulation scheme for detecting colonic polyps," in *Abdominal Imaging. Computation and Clinical Applications, ser. Lecture Notes in Computer Science*, pp. 53–62, 2013.
- [17] P. Mesejo, D. Pizarro, A. Abergel, "Computer-aided classification of gastrointestinal lesions in regular colonoscopy," *IEEE Transactions on Medical Imaging*, vol. 35, no. 9, pp. 2051–2063, 2016.
- [18] J. Bernal, N. Tajbaksh, F. J. Sanchez, "Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge," *IEEE Transactions on Medical Imaging*, vol. 36, no. 6, pp. 1231–1249, 2017.
- [19] G. Van deWouwer, P. Scheunders, and D. van Dyck, "Statistical texture characterization from discrete wavelet representations," *IEEE Transaction on Image Processing*, vol. 8, no. 4, pp. 592–598, 1999.
- [20] Belur VD, Edwin B. Holder, "Image characterizations based on joint gray level-run length distributions," *Pattern Recognition Letters*, vol. 12, no. 8, pp. 497-502, 1991.
- [21] Donahue, Jeff, "Decaf: A deep convolutional activation feature for generic visual recognition," *Proceedings of the 31st International Conference on Machine Learning*, vol. 32, no. 1, pp. 647-655, 2014.
- [22] K. Vassilis and B. Maria, "An adaptive neurofuzzy approach for the diagnosis in wireless capsule endoscopy imaging," *International Journal of Information Technology*, vol. 13, no. 1, pp. 46–56, 2007.
- [23] Alexander V. Mamonov, et al., "Automated polyp detection in colon capsule endoscopy," *IEEE Transactions on Medical Imaging*, vol. 33, no. 7, pp. 1488–1505, 2014.
- [24] N.N. Mohd Hasri, et al., "Improved Support Vector Machine Using Multiple SVM-RFE for Cancer Classification," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, no. 4–2, pp. 1589, 2017.
- [25] L. Gao, M. Ye, and C. Wu, "Cancer Classification Based on Support Vector Machine Optimized by Particle Swarm Optimization and Artificial Bee Colony," *Molecules*, vol. 22, no. 12, pp. 2086, Nov. 2017.

BIOGRAPHIES OF AUTHORS



A. S. M. Shafi received the B.Sc. (Engg.) degree in Computer Science and Engineering (CSE) in 2014 from the University of Mawlana Bhashani Science and Technology (MBSTU). He is currently pursuing his M.Sc (Engg.) in CSE from the same university. His area of research interests are medical image processing, digital image processing, deep learning, neural networks, computer vision and data mining.



Mohammad Motiur Rahman is currently working as a professor in the Dept. of Computer Science and Engineering (CSE) at Mawlana Bhashani Science and Technology University (MBSTU). He received his Ph.D. from Jahangirnagar University in CSE. He worked in visual analysis lab, ICM, university of Warsaw, Poland as a Ph.D. exchange student. His research is focused on bio imaging and image analysis, computer vision, pattern recognition, medical diagnosis. Mr. Rahman has many international journal and conference publications.