Biomedical Signal Processing and Control 40 (2018) 415–424

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Performance assessment of a bleeding detection algorithm for endoscopic video based on classifier fusion method and exhaustive feature selection

Farah Deeba*, Monzurul Islam, Francis M. Bui, Khan A. Wahid

Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, Saskatchewan, S7N5A9, Canada

ARTICLE INFO

Article history:
Received 30 January 2017
Received in revised form 15 July 2017
Accepted 15 October 2017
Available online 5 November 2017

Keywords:
Capsule endoscopy
Color features
Automated bleeding detection
Classifier fusion
Nested cross validation
SVM score

ABSTRACT

Capsule Endoscopy (CE) is a non-invasive clinical procedure that allows examination of the entire gastrointestinal tract including parts of small intestine beyond the scope of conventional endoscope. It requires computer-aided approach for the assessment of video frames to reduce diagnosis time. This paper presents a computer-assisted method based on a classifier fusion algorithm which combines two optimized Support Vector Machine (SVM) classifiers to automatically detect bleeding regions present in CE frames. The classifiers are based on RGB and HSV color spaces; the image regions are characterized on the basis of statistical features derived from the first-order histogram probability of respective color channels. A nested cross validation strategy has been adopted for the parameter tuning and feature selection to optimize the classifiers. The optimum feature sets for the best performance are evaluated after exhaustive analysis. The proposed fusion approach achieves an average accuracy of 95%, sensitivity of 94% and specificity of 95.3% for a dataset of 8872 CE frames, which is higher than that obtained from a single classifier. Comparison with the state-of-the-art algorithms exhibits that the proposed method yields superior performance for diverse dataset.

1. Introduction

The invention of fiberoptic endoscopy made it possible to visualize the gastrointestinal (GI) tract from the outside world, but was associated with pain and discomfort of patients. Wireless Capsule Endoscopy (WCE) is thus an important innovation in diagnostic endoscopy to examine human GI tract without any pain, sedation or air insufflation. It is especially useful in finding the origin of obscure gastrointestinal bleeding (OGB), which is defined as chronic bleeding from a source not found from traditional endoscopy. Although OGB is responsible for approximately 5% of all GI bleeding [1], it has crucial economic impact on health budget, with an estimated cost per patient of $35,000 [2]. For the ability to directly image the entire small bowel, CE is considered as a breakthrough in diagnosis of OGB.

A major drawback of CE procedure that affects correct diagnosis is the time consumed in image screening process, demanding an undivided attention of the clinician to review over 14,400–72,000 captured frames among which only 1% may be of clinical interest [3]. The reported time for the complete review of a single VCE case ranges from 45 min to two hours [4]. Some image processing techniques have been incorporated by the manufacturers to speed up the process, such as, Given Imaging's Rapid Reader Software that includes a blood detection tool called Suspected Bleeding Indicator (SBI). However, studies have shown that accuracy profile of SBI is suboptimal and highly variable for different cases [3,4]. This has motivated a lot of studies in the development of computer assisted diagnosis tools to automatically detect bleeding areas in capsule endoscopy images. These studies have the potential to not only reduce the time of reading but also improve the diagnosis, by drawing attention to clinically significant images, which could have otherwise been missed among many thousands normal images.

In current literature, several supervised and unsupervised learning algorithms have been employed for bleeding detection. The unsupervised methods [37] primarily rely on salient features characterizing the particular abnormality. However, these methods have not be tested on large dataset. Supervised bleeding detection methods are more commonly found in literature. Among the supervised methods, pixel-wise classification [5–8] is a widely used approach which is suitable for obtaining pixel-perfect delineation of bleeding region. Due to high computational complexity, this approach is not suitable for processing large number of video

* Corresponding author.
E-mail address: farah.deeba@usask.ca (F. Deeba).

https://doi.org/10.1016/j.bspc.2017.10.011
1746-8094/

frames, especially in case of high resolution images. The limitations have inspired feature extraction from image patches [9,10] or whole images [11,12]. These methods are likely to have reduced sensitivity for the detection of small bleeding regions [13]. To generate features precisely characterizing bleeding or non-bleeding regions, regions of arbitrary shape are considered in recent papers. These regions are obtained by image segmentation techniques, for example, region growing [14], super pixel [15], growcut [36] or mean-shift [16] segmentation algorithms. In this paper, an improved region growing algorithm has been proposed for selecting regions of arbitrary shape requiring reduced human interaction. Studies show that color features are more discriminative for detection of bleeding than texture features [17]. So we limit the search space for feature selection into first order histogram moment of selected color channels. RGB and HSV color space are chosen for feature extraction based on the assumption that (1) bleeding exhibits itself as specific shades of red and (2) HSV color space is less sensitive to illumination change [18], which is suitable for CE images.

Recently, deep learning is bringing a new paradigm shift in the field of bleeding detection [38]. However, deep learning framework requires intensive and time-consuming training involving very large training dataset and additional hardware. We can argue that traditional supervised algorithms are sufficient to achieve superior bleeding detection performance. The objective of this paper is to propose an efficient cascade system for bleeding detection that is suitable for practical application and with minimal human interaction and superior performance index. We have employed a two-step preprocessing step that utilizes uniform black background of the frames as reference that ensures consistent brightness for all video frames taken from different parts of GI tract, different patients and also different sources. This will enable the applicability of the proposed system to a wide variety of situations prevalent in clinical setting. Again, we have adopted region selection methods for "bleeding" frame (frame containing bleeding region) which require minimum human interaction by taking advantage of continuous bleeding frames. And the region selection from "normal" frame (frame containing no bleeding) is an automated segmentation process. For the classifier optimization, we have reviewed the problem of biasing and overfitting [19,20] associated with cross-validation and other standard methods used in current literature very carefully. Though SVM has been widely used in field of CE video classification [11,12,21,22] little attention has been given to prevent possible biasing and overfitting while performing SVM optimization. As a result, a classifier optimization scheme for ensuring unbiased optimization is presented in this work to address the issue. Finally, a classifier fusion algorithm based on SVM score has been proposed to combine the results of two classifiers: one based on RGB color space and another on HSV color space.

Incorporating all the above mentioned features, we are able to build a classification system which is applicable for a general and diverse data set. This fact is validated by applying the system on a test dataset taken from different sources to ensure high variability and minimum training. The results show high performance indices in terms of accuracy, sensitivity and specificity which shows promise to be used in clinical setting. Also the low computation complexity (0.215 s/frame) makes it possible to perform real time evaluation of CE frames where CE frames are generated at a rate of 2 frames/s.

The proposed system can be divided into three stages: training, classifier optimization and application stage as illustrated in Fig. 1. In the following sections, we present the detailed description of each processing stage.

2. Proposed system: training stage

2.1. Pre-processing

2.1.1. Eliminating unwanted information in background

Generally, CE frames are circular in shape with black background containing texts in white. The texts can interfere in the subsequent segmentation stage. So in the pre-processing stage, we eliminate the text by: (1) extracting an elliptical ROI from the image, (2) creating a mask using this ROI and (3) multiplying the original image with the mask as shown in Fig. 2. For the automation of the process, we select a fixed size ROI, which is applied to all the video frames in training and testing dataset.

2.1.2. Vignetting correction

The CE hardware is equipped with four to six white light emitting diodes (LED) as illumination source [23]. This illumination system causes a non-uniform spatial light distribution resulting in poor image quality [24]. Especially, the directional nature of illumination source often causes an artifact in the CE frames, known as vignetting, which refers to the brightness attenuation away from the image center. To minimize this artifact, we apply a vignetting correction algorithm [25] on the CE frames.

CE frames from different videos largely vary in brightness due to different beam angle of LED sources [26] as can be seen from Fig. 3. The high variation in brightness can cause overlapping between bleeding and non-bleeding features obtained from different video frames, thus leading to less separability between different classes and inferior classification performance. Vignetting correction of frames minimizes this problem. From Fig. 3, it is evident that the estimated vignetting provides an estimation of the light distribution of the frame. For example, frame (iv) is more contrasting than frame (i), if we take the black background as reference. The estimated vignetting in image (ii) and (v) reflect the difference in contrast in image (i) and image (iv). The estimation of non-uniform brightness allows normalization of illumination of different frames in the vignetting correction step, which can be seen from the vignetting corrected images in Fig. 3.

2.2. Region selection for training phase

After the pre-processing step, we create a training dataset containing CE videos. Then we manually sort the videos in two classes: 1) videos containing occasional bleeding region and 2) videos containing no bleeding. For region selection from "bleeding" frames and "normal" frames, we apply two different methods.

2.2.1. Modified semi-automatic region-growing algorithm for "Bleeding" region selection

Acute bleeding typically occurs in several consecutive frames [18], as can be seen in Fig. 4. This prior knowledge can be utilized for devising a semi-automatic bleeding-region selection algorithm. For taking advantage of the occurrence of bleeding in consecutive frames, we propose a semi-automatic region growing method for bleeding region selection, which is an improved version of the semi-automatic ground truth annotation scheme proposed in [14]. According to that work, the seed for the first frame of a consecutive frame sequence containing bleeding, is manually provided by the user which is used for region growing. Then the centroid of the current frame is calculated and the coordinate of the centroid is used as the seed point for the next frame. One limitation of that method was that it does not take into account the absence of bleeding in one or few intermediate frames which belongs to a frame sequence containing bleeding in all other frames. For the forward-backward motion and tilting of the CE camera, there could be sudden absence or major displacement of bleeding region (Fig. 4) which can fail the

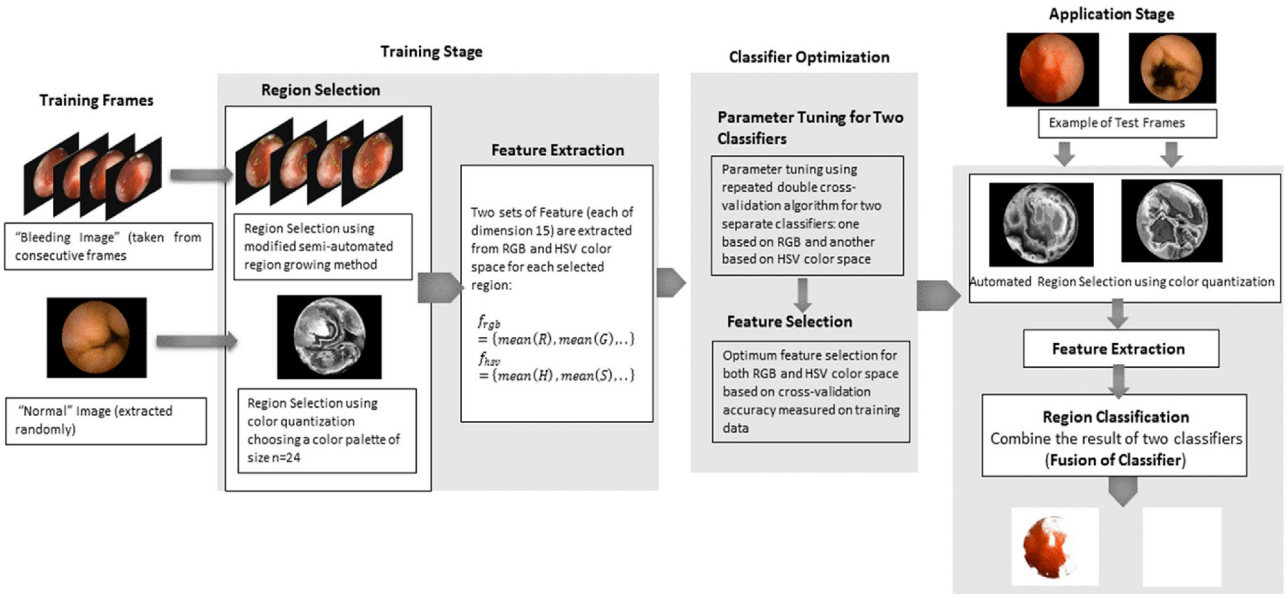

Fig. 1. Overview of the proposed system showing the three major parts of the system: training stage, classifier optimization stage and application stage.

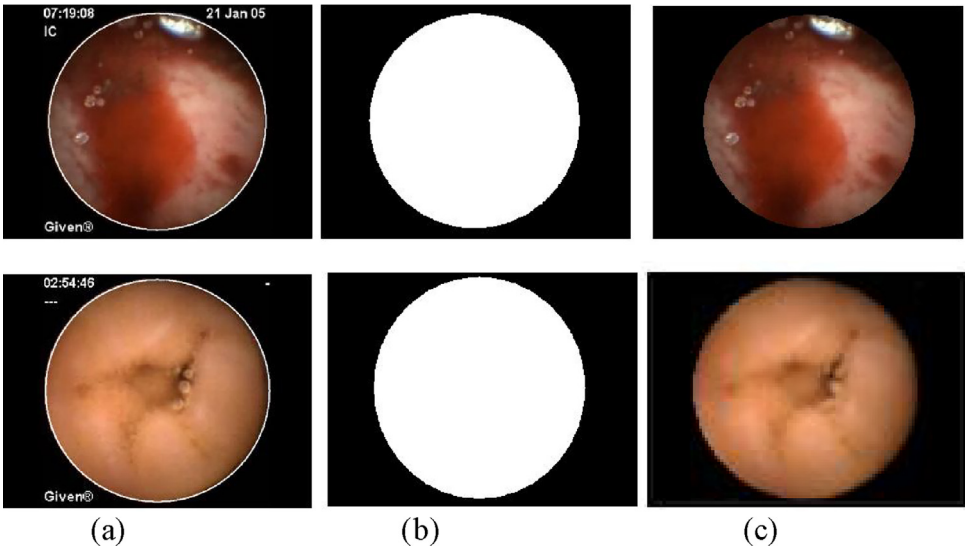

Fig. 2. Elimination of unwanted information in the background of CE frame: (a) Original Image with the delineated ROI, (b) Binary mask made from ROI, (c) Final image after replacing the background with a plain black one.

condition of automatic seed search. The previous algorithm [14] will terminate in this case and has to be started again with providing new seed for the next frames of bleeding sequence. In our proposed work, we improve the scheme to minimize the human intervention by keeping a provision of skipping one or few frames based on some condition, which can be seen in Fig. 6. This allows automatic seed selection for more frames compared to the original method.

Another limitation of the previous work [14] is that it extracts a single bleeding region from the successive video frames containing bleeding. This feature has a potential limitation which can lead to the imbalance between bleeding class and non-bleeding class of training data. Generally, bleeding frames are less probable than the non-bleeding frames. This unequal distribution is intensified by extracting single region from multiple frames, instead of a single frame. To address this problem, we extract bleeding region from single frame, rather from the entire video sequence containing bleeding. The improved semi-automatic bleeding region selection algorithm is presented in Fig. 5.

We have also experimented with different threshold values for optimum region selection result. Empirically it can be found that, a threshold between 1.5–3.0 yields optimum results for region growing from the centroid in a single frame. And a threshold value between 2.0–3.0 is used for the condition of centroid transfer from one frame to next frame. HSV color space is used for calculating the color distance in both cases.

2.2.2. *Region selection from "non-bleeding" image*

We employ Minimum Variance Quantization technique [14] for non-bleeding region selection. Minimum Variance Quantization algorithm cuts the RGB color cube into maximum n number of smaller boxes of different sizes depending on the color distribution of the image where each box represents a single color in the output image.

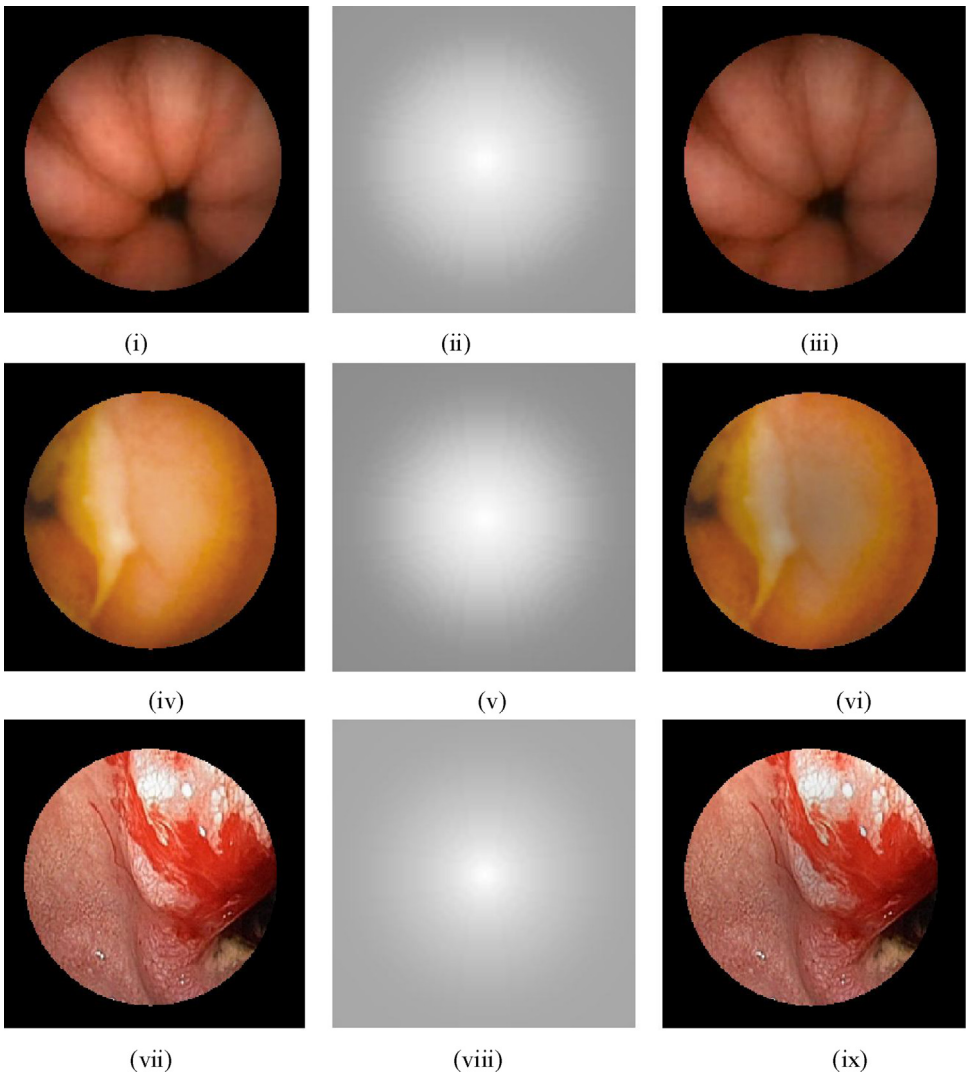

Fig. 3. Vignetting Correction: left: Original image, middle: Estimated vignetting, right: Final image after correction.

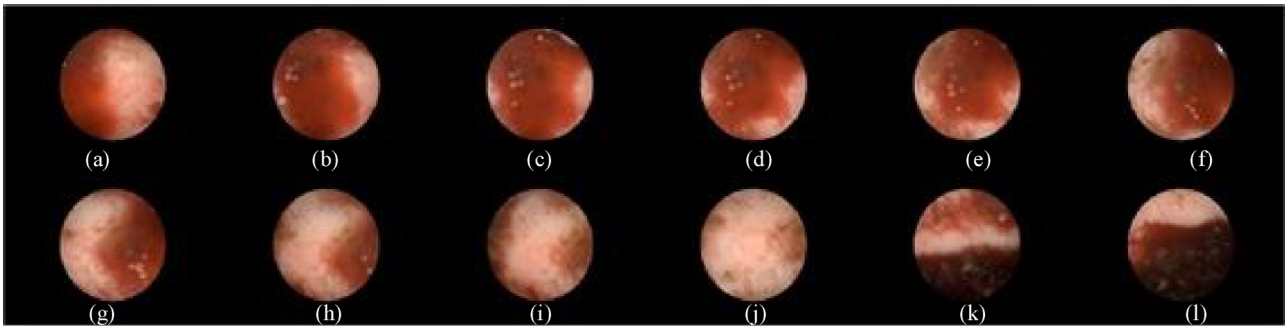

Fig. 4. Example of bleeding region in continuous frames with intermediate frames containing no blood (frame (j)) or major displacement of bleeding region (frame (i)).

The quantization of the image can be considered as a rough segmentation, where all the pixels with same color values are grouped as the same region. We have adopted $n = 24$ as the number of reduced color palette in output image due to the satisfactory performance exhibited in the other work [14]. Fig. 7 gives an example of segmentation using color quantization. For the training purpose, we eliminate the region corresponding to the background as it does not belong to either "bleeding" or "normal" class.

2.3. Feature extraction

Color is the most important aspect that differentiates bleeding and non-bleeding region. In this paper, first order histogram features from each channel of RGB plane and HSV plane are extracted from the segmented regions obtained from previous section. The features used in this paper are: Mean, Standard deviation, Entropy, Skew and Energy [27]. These features are calculated from the histogram probability of the extracted regions from CE frames. So each region will be represented by 15 (=3 (channels) $\times$5 (features))

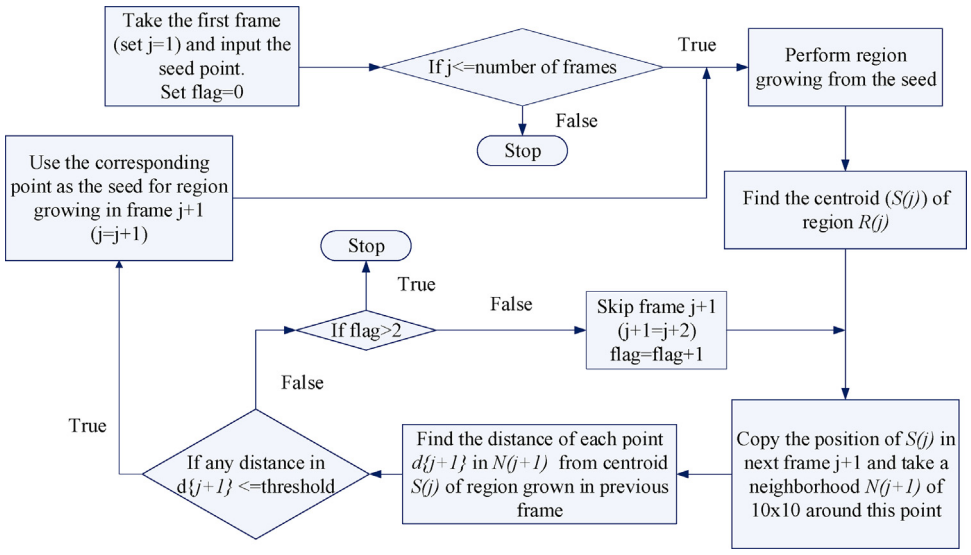

Fig. 5. Outline for the improved semi-automatic region-growing algorithm for bleeding region selection.

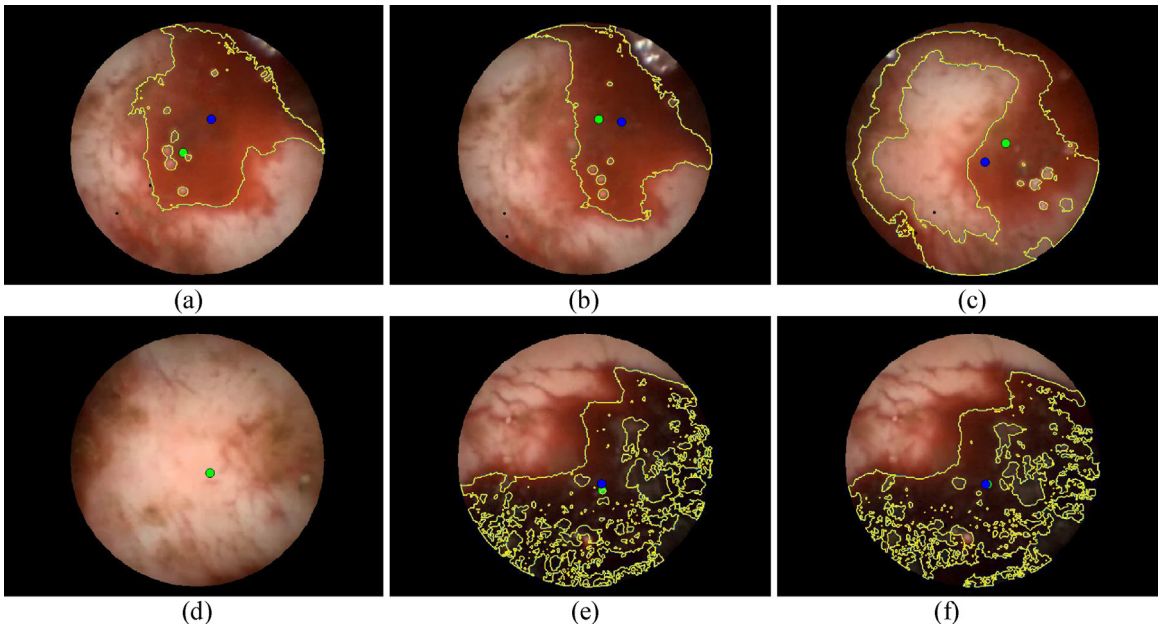

Fig. 6. Example of Bleeding Region Selection using Modified Ground Truth Annotation Algorithm. The green dot indicates the seed for the current frame and the blue dot indicates the centroid calculated from the grown region in current frame, which will be transferred to the next frame and will act as the seed for that frame. Seed for the first frame (image (a)) will be provided by the user. In frame (b) and (c), condition for transferring centroid position from previous frame is satisfied and the corresponding centroid is transferred and used for region growing. In frame (d), the condition is not met, so this frame is skipped and centroid position in frame (c) is copied in frame (e), which again meet the condition and continuous region growing is performed for subsequent frames. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

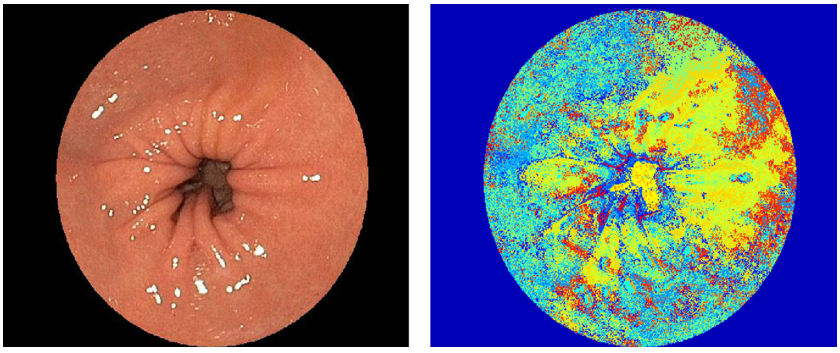

Fig. 7. Region selection from "Normal" image using color quantization method. Here, 24 colors have been used to approximate the original image. So the image is segmented into 24 normal regions.

dimensional feature vector in RGB space and 15 (=3 (channels) ×5 (features)) dimensional feature vector in HSV space. These two feature sets will be used to train two SVM classifiers for bleeding detection.

3. Proposed system: classifier optimization stage

3.1. Support vector machine

Support vector machine (SVM) [28] is a powerful supervised classifier algorithm which theoretically finds a global optimal solution by finding a hyperplane which maximizes the margin between data of different classes in a certain feature space. For non-linearly separable data, SVM maps the original feature space to a high dimensional space where the data are linearly separable using "kernel trick". In our paper, radial basis function (rbf) kernel is chosen, which is given by:

$$K(x, x_i) = \exp\left(-\frac{|x - x_i|^2}{2\sigma^2}\right), \gamma = \frac{1}{2\sigma^2} \quad (1)$$

This particular selection of classifier and kernel function leave us to the tuning of two parameters for the optimization of the classifier: the regularization parameter C and kernel parameter (specific to rbf kernel)γ.

3.2. Feature selection and parameter tuning: using a nested CV approach

Exhaustive feature selection was performed to find optimal feature subset from the two feature sets (one in RGB color space and another in HSV color space) separately. For classifier optimization (by means of parameter selection and feature selection), cross-validation is an effective method. However cross-validation error could be a very biased estimate of the true error and could pose risk of overfitting if the parameter and features are selected based on cross-validation error [19,20]. To ensure that the optimization process is unbiased, we adopt a nested cross-validation approach [19], where 20% of the training data (validation set) are left out and not used in parameter tuning and feature selection. The classifier is trained on the 80% data and tested on the left out 20% data. So we are effectively implementing two nested cross-validation loops where the inner loop is used for parameter tuning and the outer loop computes the estimate of the true error. Finally, the optimal parameters and feature sets are selected based on the minimum cross-validation error. This approach results in an optimization of classifier which is unbiased of the training data with good generalization capability. The detail of the optimization is described in Section 5–5.3.

4. Proposed system: application stage

4.1. Preparing test data

After the completion of region selection, feature extraction, feature selection, classifier training and optimization, the algorithm is ready to perform the intended task, i.e., classification of test data. Here, at first each video frame will undergo the automatic preprocessing steps (described in 2-2.1). Then region segmentation will be performed according to the method described in 2-2.1 (*2*). After region selection, we extract 15 (=3 (channels) ×5 (features)) dimensional feature vector in RGB space and 15 (=3 (channels) ×5 (features)) dimensional feature vector in HSV space. These feature sets constitute the test data. In the classifier optimization stage (as described in Section 3), the optimal parameters for SVM and feature sets have already been obtained. The Classifier Optimization Stage results in two trained SVM classifiers, one trained on RGB feature set and the other on HSV feature set. These two classifiers separately classify the test data either as bleeding or non-bleeding.

4.2. Fusion of classifier

According to [29], a higher accuracy can be achieved by combining the results of multiple classifiers than that of the best classifier. Based on this hypothesis, we propose a new method for classification of CE frames based on SVM score. SVM score is the distance from the decision hyperplane to the test vector and can be given by the following equation [28]:

$$d(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b \quad (2)$$

Here,x_i $(i = 1, 2, \cdots, n)$ are the support vectors which define the margin of the SVM, $y_i \in \{-1, 1\}$ are the associated class labels, K is the kernel function, α_i is the Lagrange Multiplier associated with support vector x_i and b is the bias term. If $d(x)$ is positive, predicted label for the test vector x is +1("bleeding" in our case) or −1 ("non-bleeding") otherwise. And the numeric value of $d(x)$ gives the distance of the test vector from the decision hyperplane which is called the SVM score. A higher SVM score implies greater probability of the test vector to belong to the corresponding class. Though research has been going on for extracting probability from the SVM score, it has been found that for equal prior probability, raw SVM score using zero threshold gives better result than that obtained from posterior probability calculated by incorporating sigmoid function with SVM [30]. As we have trained our classifiers on a training set with equal number of data from both classes (which leads to equal prior probability of two classes), adopting uncalibrated SVM score is more appropriate for calculating class confidence.

We will now describe the fusion method for combining the performance of two classifiers. In the classifier optimization stage, we have separately optimized two classifiers and two sets of features: one for RGB and another for HSV color space. Now using these two optimized classifiers, each test region of every CE frame will be classified separately based on the corresponding best feature subset. The regions, for which both classifiers agree with each other, will be classified accordingly. The remaining regions, for which two classifiers give different results, SVM score will be the metric for tie-breaking. So the SVM score, i.e., the distance of the feature point (which represents the test region) from the decision boundary of SVM of both classifiers will be calculated using Eq. (2). The classifier which gives higher value of SVM score is taken as the deciding classifier, as it implies higher class confidence. An example for the classifier fusion method is given in Fig. 8.

5. Experimental results

5.1. Dataset

For optimum training of classifier, a representative data set is required which accurately reflects the entire population. Again for validation of the trained classifier, a diverse and large data set is required. For this purpose, we create two database: database 1 (taken from [31]) and database 2 (taken from [32]). Database 1 includes videos taken by Pillcam SB of Given Imaging and Database 2 includes videos taken by Pilcam SB1 and Pillcam SB2 (Given Imaging) and EndoCapsule1 (Olympus). In a real world scenario, test set could be very different from the training set. To simulate this condition, we restrict selection of training set from database 1 only and testing data are extracted from both database 1 and database 2. This will ensure the diversity of test data set. A robust and generalized

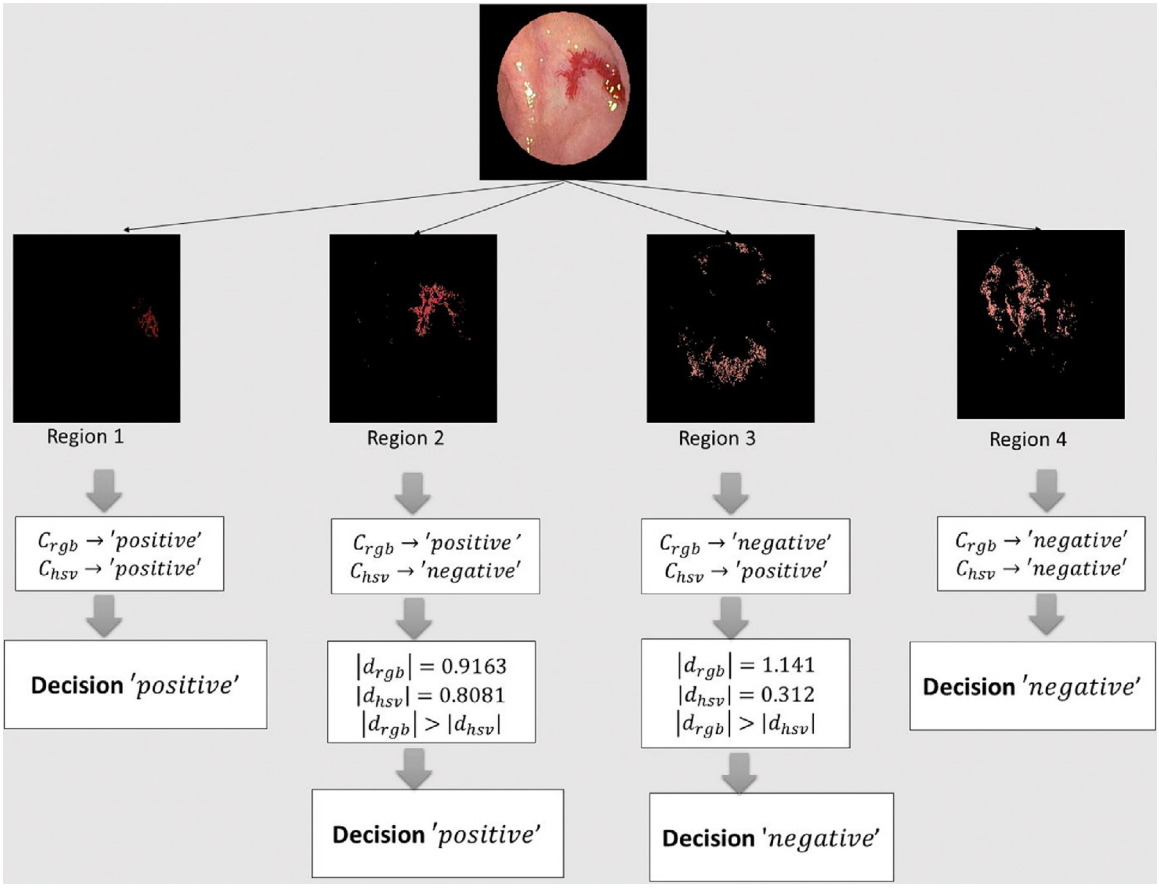

Fig. 8. Classifier Fusion Scheme for four arbitrary regions extracted from a bleeding frame. For region 1 and region 4, two classifiers give the same result, which is taken as the final result. For region 2 and region 3, two classifiers give different results. So the distance from the test data representing the region to the respective SVM hyperplane is calculated and the classifier which gives the greater distances is taken as the deciding classifier.

Table 1
Description of Test Data.

Database	Total	Bleeding	Non-Bleeding
Database 1	1224	534	690
Database 2	7648	1933	5715
Total	8872	2467	6405

system should be able to give satisfactory result for this large and diverse dataset depending on a comparatively smaller training set, which is the case in clinical practice. Again it has been made sure that, training and testing data have not been extracted from the same video to avoid biasing. Each image has been downscaled to a resolution of 256×256 for increasing computational efficiency.

5.2. Experimental setup

All experiments are performed using MATLAB R2015a. During experiments, we prepare separate training data set and test data set. The training dataset consists of 124 bleeding instants and 124 non-bleeding instants, drawn from 10 CE videos. The test data are comprised of two subsets. First subset which is taken from database 1 contains 1224 images among which 534 images are bleeding and 690 images are non-bleeding. Second subset, taken from database 2, contains 7648 images among which 1933 images are bleeding and 5715 images are non-bleeding. The description of test data set has been given in Table 1.

5.3. Feature selection and classifier optimization using two-nested CV loops

The first experiment is performed to train the classifiers by tuning classifier parameter and selecting the best feature combination. For this experiment, we have used the training data adopting nested cross validation method to avoid the biasing effect typically arising from cross-validation. We have used global optimization toolbox in MATLAB to tune the parameters of SVM, i.e., regularization parameter C and kernel parameter γ. For each possible combination of features, first the training data have been randomly divided into calibration set and test set, where calibration set constitute the 80% of training data (198 samples) and remaining (50 vectors) are the test data. SVM parameters are optimized using this calibration set. In the next step (the inner loop), a 5-fold cross validation is performed on the calibration step by splitting the set into 5 segments. Each segment is taken as the validation set while the rest 4 segments as the training set in turn. On this setting, optimum parameter (C, γ) is found using Global Optimization toolbox. For five folds, five misclassification error values are obtained, averaging which will give the misclassification error for the certain parameter(C, γ). This inner loop is repeated five times yielding five (C, γ) values. The best parameter(C, γ) is chosen based on the minimum classification error, i.e. maximum accuracy. Now in the outer loop, an SVM model is trained on the entire calibration set using the optimum parameters in previous step and the classification accuracy is evaluated on the test set. This is the best accuracy achievable for a specific feature combination. So we are effectively implementing two nested cross-validation loops where the inner loop is used for

Table 2
Optimum Feature Subsets in RGB Space for Different Dimensions.

Dim.	Feature Subset	Accuracy
1	Mean (G)	0.8
2	Mean (R) + Mean (G)	0.96
3	Mean(R) + Mean (G) + Std. Dev. (G)	0.996
4	Mean(R) + Mean (G) + Std. Dev. (G) + Skew (B)	0.996
5	Mean(R) + Mean (G) + Energy (G) + Skew (R) + Energy (R)	0.992
6	Mean (B) + Std. Dev. (G) + Entropy (R) + Entropy (G) + Skew (R) + Skew (G)	0.996
7	Mean (R) + Entropy (R, G) + Skew (R, G) + Energy (R + G)	0.992
8	Mean (R, G, B) + Std. Dev (G) + Entropy (R, G) + Energy (R, G)	0.992
9	All Features except Std. Dev. (R, G, B) + Entropy (B) + Skew (B) + Energy (B)	0.992
10	All Features except Std. Dev. (B) + Entropy (R) + Skew (G & B) + Energy (B)	0.992
11	All Features except Std. Dev. (R &B) + Skew (B) + Energy (B)	0.996
12	All Features except Std. Dev. (R) + Entropy (R & G)	0.99
13	All Features except Skew (B) and Energy (G)	0.98
14	All Features except Skew (B)	0.976
15	All Features	0.92

Dim = Dimension.

Table 3
Optimum Feature Subsets in HSV Space for Different Dimensions.

Dim.	Feature Subset	Accuracy
1	Mean (H)	0.84
2	Mean (S) + Std. Dev. (S)	0.92
3	Mean(H) + Mean (S) + Entropy (H)	0.98
4	Mean(H) + Mean (S) + Mean (V) + Std. Dev. (S)	0.996
5	Mean(H) + Mean (S) + Std. Dev. (H) + Entropy (V) + Skew (H)	0.996
6	Mean (H, S, V) + Std. Dev. (H,S) + Skew (V)	0.992
7	Mean (H, S) + Std. Dev. (H) + Entropy (H, S) + Energy (H, S)	0.98
8	Mean (H, S, V) + Std. Dev (H, S) + Entropy (S) + Energy (S) + Skew (V)	0.98
9	All Features except Std. Dev. (V) + Entropy (V) + Skew (H, S, V) + Energy (V)	0.96
10	All Features except Mean (V) + Std. Dev. (V) + Skew (S) + Energy (H, V)	0.996
11	All Features except Std. Dev. (S) + Entropy (S, V) + Energy (H)	0.996
12	All Features except Entropy (H, S) + Energy (H)	0.996
13	All Features except Std. Dev. (V) and Energy (V)	0.98
14	All Features except Energy (H)	0.96
15	All Features	0.976

Dim = Dimension.

tuning parameter and outer loop is used for evaluating the generalization error [19]. This procedure will safely discard any possibility of bias and overfitting.

After the previous step, we obtain the estimate of classification accuracy for all possible combination of features for the optimized classifier. So the feature combination which accounts for the maximum accuracy will be the selected as best subset. For RGB and HSV color space, we separately select the best feature subset for each dimension which are given in Tables 2 and 3. From Table 2, we can see that the best accuracy (99.6%) is achievable with feature subsets of dimension three in RGB space. So, for RGB color space, we select the feature triplet constituting of mean of red channel, mean of green channel and standard deviation of green channel as the optimum feature subset. The optimum (C, γ) value for this combination is (0.668, 3.004). And Table 3 shows that with a minimum feature set of dimension four, best accuracy (99.6%) can be obtained in HSV color space. So, subset consisting of mean of all three channels (hue, saturation and value) and standard deviation of saturation channel is selected as the best feature subset for HSV color space. And the optimum value of (C, γ) for this combination found from the optimization process is (36.23, 2.447).

5.4. Performance metrics

For quantitatively evaluate the performance of our proposed system, we have used the following performance criteria [33]:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Sensitivity/Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{F}_1\text{Score} = 2 * \frac{\text{Precion} * \text{Recall}}{\text{Precion} + \text{Recall}}$$

where TP is the number of "bleeding" frames classified correctly, TN is the number of "non-bleeding" frames classified correctly, FP is the number of "non-bleeding" frames classified incorrectly as "bleeding" frame and FN is the number of "non-bleeding" frames classified incorrectly as "bleeding" frame. We have also constructed confusion matrix to show the overall picture of the classification performance indicating the values of sensitivity, specificity, FNR (False Negative Rate = 1-sensitivity) and FPR (False Positive Rate = 1-specificity).

5.5. Experimental results for bleeding detection

After the classifier parameter optimization and feature selection, the trained classifier is tested on a large dataset as described in Section 5-5.2. Table 4 shows the performance of three classifiers in terms of accuracy, sensitivity, specificity, precision and F1-score. While assessing the performance of a classifier (or system), one cannot simply look at just one performance index as it may be misleading. A combination of all performance indices should be considered to make an informed judgment. This is where the proposed fusion method is useful.

The confusion matrix for the result of testing is given in Table 5. From the tables, we can see that the classifier based on RGB color space has a tendency to incorrectly label data as bleeding, which is reflected in its poor specificity (as highlighted in Table 5). So the high sensitivity index from this classifier is somewhat misleading. On the other hand, though the classifier based on HSV space generally show higher performance indices, there are cases where it suffers from lower sensitivity by incorrectly labelling bleeding frames as normal. Here we see that in both cases, the fusion method minimizes the error by correctly labelling the frames for different types of datasets which shows the robustness of the proposed algorithm.

Table 6 provides the computation cost (time required for classification of each CE frame) for three classifiers. As training classifier

Table 4
Performance Evaluation of Three Classifiers.

	Classifier	Accuracy	Sensitivity	Specificity	Precision	F_1 Score
Database 1	C_{rgb}	92.97%	97.4%	89.6%	87.84%	89.12%
	C_{hsv}	96.73%	98.13%	95.65%	94.58%	96.32%
	C_{fusion}	97.22%	98.13%	**96.52%**	**95.66%**	**96.86%**
Database 2	C_{rgb}	49.53%	100%	32.46%	33.37%	49.31%
	C_{hsv}	93.89%	91.05%	94.86%	85.69%	88.2%
	C_{fusion}	**94.5%**	**92.6%**	**95.2%**	**86.6%**	**90.0%**

Table 5
Confusion Matrix for Bleeding Detection Resulting From Three Classifiers.

			Predicted Class					
			RGB Classifier C_{rgb}		HSV Classifier C_{hsv}		Fusion of Classifier C_{fusion}	
Database 1 n = 1224	True Class	n = 1224	Non-Bleeding	Bleeding	Non-Bleeding	Bleeding	Non-Bleeding	Bleeding
		Non-Bleeding	TN = 618	FP = 72	TN = 660	FP = 30	TN = 666	FP = 24
			Spec. = 89.6%	FPR = 10.4%	Spec. = 95.65%	FPR = 4.35%	**Spec. = 96.52%**	FPR = 3.48%
		Bleeding	FN = 14	TP = 520	FN = 10	TP = 524	FN = 10	TP = 524
				FNR = 2.6%	**Sens. = 97.4%**	FNR = 1.87%	**Sens. = 98.13%**	FNR = 1.87%
Database 2 n = 7648	True Class	Non-Bleeding	TN = 1855	FP = 3860	TN = 5421	FP = 294	TN = 5438	FP = 277
			Spec. = 32.46%	FPR = 67.54%	**Spec. = 94.86%**	FPR = 5.14%	**Spec. = 95.2%**	FPR = 4.8%
		Bleeding	FN = 0	TP = 1933	FN = 173	TP = 1760	FN = 144	TP = 1789
			FNR = 0%	**Sens. = 100%**	FNR = 8.95%	**Sens. = 91.05%**	FNR = 0.6%	**Sens. = 92.6%**

Sens. = Sensitivity, Spec. = Specificity.

Table 6
Computation Cost Comparison AMONG THREE Classification Approach.

Classifier	C_{rgb}	C_{hsv}	C_{fusion}
Time (second/Frame)	0.0924	0.1584	0.2145

is done offline and can be done beforehand, time for training has not been considered. Classifier based on HSV space (C_{hsv}) requires more time than classifier based on RGB space (C_{rgb}) because of the higher dimension of optimum feature subset and conversion requirement from RGB to HSV space. The computational time for classifier based on fusion (C_{fusion}) is not much higher than a single classifier, as the extra computation of comparing SVM scores is performed only in cases where the decision of C_{rgb} and C_{hsv} does not match. Therefore, the fusion based classifier stands out to be the preferred choice of classifier.

We have compared the efficacy of our method with the state-of-the-art bleeding detection algorithms [12,14,15]. All the algorithms were implemented using MATLAB 2016a on a computer with Intel(R) Core(TM) i5-6320, 3.9 GHz CPU, 32 GB of RAM. For meaningful comparison, we used the same training and test dataset for all the experiments. Similar to the proposed method, these algorithms were tested on both database 1 and database 2. While implementing the algorithms for the performance comparison, the parameters were selected to yield the best accuracy on database 1. To emulate practical clinical setting, the selected algorithm parameters were kept unchanged while testing on database 2. For the algorithm proposed in Fu et al. [15], the chosen parameters are: the standard deviation of the Gaussian function, $\sigma = 1$, sensitivity threshold, $T = 0.1$, superpixel number, $K = 20$, and the size of structural elements B in dilation is "Square" with 10 pixels width. For the algorithm proposed in Sainju et al. [14], the number of neurons used was 15. In the algorithm proposed in Yuan et al. [12], YCbCr color space was used. The vocabulary size $K = 80$ was found to give the best performance. The performance results have been summarized in Table 7. From the comparison on Database 1, we found that all the state-of-the-art methods give competitive performance. Our method achieves slightly better sensitivity compared to other methods. However, it has only a better accuracy compared to Fu et al. [15], and comparable accuracy with the remaining methods. Interestingly, the comparison on the database 2 presents a different scenario. For database 2, all the algorithms except the proposed algorithm have an inclination to incorrectly label most of the images as bleeding. These algorithms suffer from low specificity and moderate accuracy. Therefore, it can be concluded that these algorithms are not effective in discriminating bleeding and non-bleeding images for a large and diverse test dataset. On the contrary, the proposed algorithm yields high accuracy of 94.5% and maintains high sensitivity and specificity. Therefore, the proposed method outperforms the state-of-the-art performance in terms of bleeding classification accuracy and specificity. The computation complexity is slightly higher compared to the methods Sainju et al.

Table 7
State of the Art Methods for Bleeding Detection.

Study	Database 1		Database 2		Time (Sec/frame)
	Accuracy (%)	Sens./Spec. (%)	Accuracy (%)	Sens./Spec. (%)	
Fu et al. (2014)[15](Fu, Zhang, Mandal, & Max Q.-H. Meng, 2014) (Fu, Zhang, Mandal, & Max Q.-H. Meng, 2014)	91.96	97.50/87.92	86.54	98.89/76.48	0.99
Sainju et al. (2014)CITATION SSa14 \l 1033 [14](Sainju, Bui, & Wahid, 2014) (Sainju, Bui, & Wahid, 2014)	97.52	94.83/99.48	71.44	97.40/50.32	0.11
Yuan et al. (2016) CITATION Yua15 \l 1033 [12] (Yuan, Li, & Meng, 2015)	98.05	97.15/98.70	89.34	99.35/81.20	0.04
Proposed Method	97.22	98.13/96.52	94.5	92.32/95.07	0.15

[14] and Yuan et al. [12]. Considering the current frame rate of capsule endoscopy (2fps), our method is fast enough to enable real time bleeding detection. This validation step is very important to ensure that the system will perform well in practical cases where testing data will be very different from the training data. To the best of our knowledge, no other previous work has been taken this aspect into account. Thus our method is applicable for general application including videos from different sources and different patients based on a very limited training dataset.

6. Conclusion

In this paper, we have proposed an automated and computationally simple system for bleeding detection. The extensive experiment on test data containing a variety of CE frames establishes the superiority of fusion based classifier over other classifiers. The use of pre-processing technique and nested cross-validation improves the performance of the proposed system. Moreover, the low computational time makes it possible to assess the video frames in real-time which could be applied in intelligent decision support system. Comparison with state-of-the art algorithm reveals that the proposed method is robust against variation in test data. Therefore, the method can be expected to yield superior performance in clinical applications. The algorithm can be further improved by incorporating more classifiers and fusion techniques which is left for future exploration.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] L.B. Katz, The role of surgery in occult gastrointestinal bleeding, in: Seminars in Gastrointestinal Disease, 1999.
[2] A. Goldfarb, M. Phillips, B.S. Lewis, D.B. Nash, Economic and health outcomes of capsule endoscopy: opportunities for improved management of the diagnostic process for obscure gastrointestinal bleeding, Dis. Manage. 5 (3) (2004) 123–135.
[3] A. Koulaouzidis, E. Rondonotti, A. Karargyris, Small-bowel capsule endoscopy: a ten-Point contemporary review, World J. Gastroenterol. 19 (24) (2013) 3726–3746.
[4] S. C. Park, C., HJ, K., ES, K., B, S., YS, K., YS, J., YT, L., HS, U., SH, K. CD and R. HS, Sensitivity of the Suspected Blood Indicator: An Experimental Study, World Journal of Gastroenterology, vol. 18 (31), p. 4169–4174, 2012.
[5] G. Pan, G. Yan, X. Qiu, J. Cui, Bleeding detection in wireless capsule endoscopy based on probabilistic neural network, J. Med. Syst. 35 (2011) 1477–1484.
[6] S. Hwang, J. Oh, J. Cox, S.J. Tang, H.F. Tibbals, Blood detection in wireless capsule endoscopy using expectation maximization clustering, Proc. SPIE (2006).
[7] Y.S. Jung, Y.H. Kim, D.H. Lee, J.H. Kim, Active blood detection in a high resolution capsule endoscopy using color spectrum transformation, Int. Conf. Biomed. Eng. Inf. (2008).
[8] G. Pan, G. Yan, X. Song, X. Qiu, BP neural network classification for bleeding detection in wireless capsule endoscopy, J. Med. Eng. Technol. 33 (7) (2009).
[9] B. Li, M.Q.-H. Meng, Computer-aided detection of bleeding regions for capsule endoscopy images, IEEE Trans. Biomed. Eng. 56 (2009) 1032–1039.
[10] P. Szczypiński, A. Klepaczkoa, M. Pazurekb, P. Danielb, Texture and color based image segmentation and pathology detection in capsule endoscopy videos, Comput. Methods Programs Biomed. 113 (2014) 396–411.
[11] G. Lv, G. Yan, Z. Wang, Bleeding detection in wireless capsule endoscopy images based on color invariants and spatial pyramids using support vector machines, Conf. Proc. IEEE Eng. Med. Biol. Soc. (2011).
[12] Y. Yuan, B. Li, Q. Meng, Bleeding frame and region detection in the wireless capsule endoscopy video, Biomed. Health Inf. IEEE J. 99 (2015).
[13] D.K. Iakovidis, A. Koulaouzidis, Software for enhanced video capsule endoscopy: challenges for essential progress, Nat. Rev. Gastroenterol. Hepatol. (2015) 172–186.
[14] S. Sainju, F.M. Bui, K.A. Wahid, Automated bleeding detection in capsule endoscopy videos using statistical features and region growing, J. Med. Syst. 25 (2014) (38).
[15] Y. Fu, W. Zhang, M. Mandal, F. Max, Q.-H. Meng, Computer-aided bleeding detection in WCE video, IEEE J. Biomed. Heal. Inf. (2014) 636–642.
[16] K. Sun, Y. Wu, X. Lin, S. Cheng, Y.-M. Zhu, S. Zhang, Mean shift-based lesion detection of gastroscopic images, Proceedings of the Second Sino-foreign-interchange. Conference on Intelligent Science and Intelligent Data Engineering (2012).
[17] R. Kumar, Q. Zhao, S. Seshamani, G. Mullin, Assessment of crohn's disease lesions in wireless capsule endoscopy images, Biomed. Eng. IEEE Trans. 59 (2) (2012) 355–362.
[18] Y. Chen, J. Lee, A review of machine-vision-based analysis of wireless capsule endoscopy video, Diagn. Ther. Endoscopy (2012) 2012.
[19] S. Varma, R. Simon, Bias in error estimation when using cross-validation for model selection, BMC Bioinf. 7 (2006) 1.
[20] R.B. Rao, G. Fung, R. Rosales, On the dangers of cross-validation. an experimental evaluation, in SDM, SIAM (2017) 588–596.
[21] M.W. Mackiewicz, M. Fisher, C. Jamieson, Bleeding detection in wireless capsule endoscopy using adaptive colour histogram model and support vector classification, Proc. SPIE (2008).
[22] A. Karargyris, N. Bourbakis, Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos, IEEE Trans. Biomed. Eng. 58 (2011) 2777–2786.
[23] N. Kurniawan, M. Keuchel, Technology, in: Video Capsule Endoscopy, Springer-Verlag, Berlin Heidelberg, 2014, pp. 15–20.
[24] D. Filip, C.N. OrlyYadid-Pecht Andrews, M.P. Mintchev, Design, implementation, and testing of a miniture self-stabilizing capsule endoscope with wireless image transmission capabilities, Int. J. Inf. Technol. Knowl. 5 (1) (2011) 3–24.
[25] Y. Zheng, J. Yu, S.B. Kang, S. Lin, C. Kambhamettu, Single-image vignetting correction using radial gradient symmetry, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2008).
[26] R. Shrestha, X. Zhang, Z. Gias, K. Wahid, Adaptive illumination in wireless capsule endoscopy system, Circuits Syst. (ISCAS) 2015 IEEE Int. Symp. (2015) 2015.
[27] S. Sergyan, Color Histogram Features Based Image Classification in Content-based Image Retrieval Systems, 2008.
[28] V.N. Vapnik, V. Vapnik, Statistical Learning Theory, John Wiley and Sons, Inc., 1998.
[29] L. Kuncheva, A theoretical study on six classifier fusion strategies, in Pattern Analysis and Machine Intelligence, IEEE Trans. 24 (2) (2002) 281–286.
[30] J. Platt, Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, Adv. Large Margin Classifiers 10 (3) (1999) 61–74.
[31] D.O. Faigel, D.R. Cave, Capsule Endoscopy, Saunders Elsivier, 2008.
[32] Video Journal and Encyclopedia of GI Endoscopy, Elsevier Inc, [Online]. Available: http://www.vjgi-endoscopy.com/toc_a-c/. [accessed November 2015].
[33] D.M. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, Bioinf. Publ. (2011).
[36] F. Deeba, F.M. Bui, K.A. Wahid, Automated Growcut for segmentation of endoscopic images, Neural Networks (IJCNN), 2016 International Joint Conference on (2016) 4650–4657.
[37] F. Deeba, S.K. Mohammed, F.M. Bui, K.A. Wahid, Unsupervised abnormality detection using saliency and Retinex based color enhancement, Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the IEEE (2016) 3871–3874.
[38] X. Jia, Q.-H. Max Meng, A study on automated segmentation of blood regions in Wireless Capsule Endoscopy images using fully convolutional networks, Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on (2017) 179–182.