RESEARCH ARTICLE

Heuristic Classifier for Observe Accuracy of Cancer Polyp Using Video Capsule Endoscopy

Geetha K[1]*, Rajan C[2]

Abstract

Methods: Colonoscopy is a technique for examine colon cancer, polyps. In endoscopy, video capsule is universally used mechanism for finding gastrointestinal stages. But both the mechanisms are used to find the colon cancer or colorectal polyp. The Automatic Polyp Detection sub-challenge conducted as part of the Endoscopic Vision Challenge (http://endovis.grand-challenge.org). **Method:** Colonoscopy may be primary way of improve the ability of colon cancer detection especially flat lesions. Which otherwise may be difficult to detect. Recently, automatic polyp detection algorithms have been proposed with various degrees of success. Though polyp detection in colonoscopy and other traditional endoscopy procedure based images is becoming a mature field, due to its unique imaging characteristics, detecting polyps automatically in colonoscopy is a hard problem. So the proposed video capsule cam supports to diagnose the polyps accurate and easy to identify its pattern. Existing methodology mainly concentrated on high accuracy and less time consumption and it uses many different types of data mining techniques. To analyse these high resolution video scale image we have to take segmentation of image in pixel level binary pattern with the help of a mid-pass filter and relative gray level of neighbours. This work consists of three major steps to improve the accuracy of video capsule endoscopy such as missing data imputation, high dimensionality reduction or feature selection and classification. The above steps are performed using a dataset called endoscopy polyp disease dataset with 500 patients. Our binary classification algorithm relieves human analyses using the video frames. SVM has given major contribution to process the dataset. **Results:** In this paper the key aspect of proposed results provide segmentation, binary pattern approach with Genetic Fuzzy based Improved Kernel Support Vector machine (GF-IKSVM) classifier. The segmented images all are mostly round shape. The result is refined via smooth filtering, computer vision methods and thresholding steps. **Conclusion:** Our experimental result produces 94.4% accuracy in that the proposed fuzzy system and genetic Fuzzy, which is higher than the methods, used in the literature. The GF-IKSVM classifier is well-organized and provides good accuracy results for patched VCE polyp disease diagnosis.

Keywords: Polyps- Colon cancer- Video Capsule Endoscopy-Colonoscopy- Segmentation- binary pattern

Asian Pac J Cancer Prev, **18 (6)**, 1681-1688

Introduction

The advances in medicine and technology, but we are facing with different challenges to improve both lifespan and quality of life. One of the main health problems of modern society is cancer, with colonic cancer being one of the leading causes of death by cancer (Geetha and Rajan, 2016) (third most commonly diagnosed cancer in the world). Colorectal polyps are abnormal growths in the mucosa of the colon or the rectum. Several methods are available to perform the imaging of the colorectal area and available to gastroenterology, such as conventional endoscopy and virtual colonoscopy. Wireless capsule endoscopy is a fairly recent imaging method with little invasion, consisting on a miniature camera with wireless data transfer capabilities (Adler and Gostout, 2003).

Video capsule endoscopy (VCE) is an novel strategic diagnostic imaging modality in gastroenterology, which acquires digital photographs of the gastrointestinal (GI) tract using a swallow able miniature camera device with LED flash lights (Li et al., 2014). The capsule sends images of the gastrointestinal tract to a portable recording device. The captured images are then analyzed by gastroenterologists, who locate and detect abnormal features such as polyps, lesions, bleeding, etc. and carry out diagnostic assessments. A typical capsule exam consists of more than 60,000 images during its operation time, which spans a duration of 8 to 10 h. Hence, examining each image sequence produced by VCE is an extremely time-consuming process.

Clearly, an efficient and accurate automatic detection procedure would relieve the diagnosticians of the burden of analyzing a large number of images for each patient. After the whole video sequence is recorded, it has to be analyzed for the presence of polyps. In that analysis a single patient may contain thousands of frames, which

*[1]Department of Information Technology, Excel Engineering College, [2]Department of Information Technology, K S Rangasamy College of Technology, India. *For Correspondence: geetharajsri@gmail.com*

makes manual analysis of all frames a burdensome task. Using an automated procedure for detecting the presence of polyps in the frames can greatly reduce such burden. Thus, an efficient algorithm should not only be able to detect the polyps accurately (high sensitivity), but should also have a reasonably low rate of false positive detections (high specificity) to minimize the number of frames that have to be analyzed manually.

There are several research work have been concentrated to detecting polyps on classical imaging techniques of colonoscopies (not capsule endoscopy images) (Park et al., 2012; Iakovidis et al., 1999). Polyp detection schemes applicable to colonoscopy and Computed Tomography (CT) colonography use mainly geometry based techniques However, due to the different imaging modality in VCE, images have different characteristics and hence require unique methods for efficient polyp detection across various frames. Several shapes based schemes that were proposed to find polyps in virtual colonoscopy or computed tomography. Colonography have been addressed (Ruano et al., 2013), Most of these methods take the already restructure surface which describes the colon's interior or rely on some specific imaging techniques for reviews. In contrast, VCE comes with an un-aided, uncontrolled photographic device, which moves automatically and is highly susceptible to illumination saturation due to near-field lighting (El Khatib et al., 2015).

Colonoscopy may be primary way of improve the ability of colon cancer detection especially flat lesions. Which otherwise may be difficult to detect. A defect we found the VCE diagnoses, as the images from VCE differ significantly from images obtained with the traditional colonoscopy. Due to the unaided movement of the capsule camera, blurring effects make the image looks less sharper. Moreover, the color of mucosal tissue under VCE has some peculiar characteristics (Prasath , 2015). Nevertheless, a recent meta-analysis showed that capsule endoscopy is effective in detecting colorectal polyps (at-least in the colon capsules, though the jury is still out on the small-bowel and oesophagus). Newer advances in sensors, camera system results in second generation capsule endoscopes and sensitivity and specificity for detecting colorectal polyps was improved. However, the increased imaging complexity and higher frame rates, though they provide more information, inevitably put more burdens on the gastroenterologists. Thus, having efficient, robust automatic computer aided detection and segmentation of colorectal polyps is of great importance and needed medical services now.

One of the approaches of polyp detection and classification, both in wireless capsule endoscopy imaging and in virtual colonoscopy imaging is a curvature-based approach. This approach is present where multiple curvature descriptors are used (Gaussian curvature, mean curvature, principal curvatures) to detect and segment polyps.

In order to attain the proposed segmentation and detection of human colorectal polyps the following sequential approach was performed:

1. Segmentation and detection. For each of the video sequences, a preprocessing step is performed, aiming to reduce the degradation of the images due to illumination and high reactance factors. The segmentation step of the preprocessed images consists in the application of a LPA filter and use of principal curvature and morphological operators in multiple color subspaces to obtain a rough segmentation of the images. The detection stage consists on the application of information from the LPA filtering, the Laplacian of the images and the principal curvatures to select, in an heuristic fashion, the polyp candidates from the segmentation data, aiming to remove non-polyp segmented regions, making the lowest compromise possible regarding the non consideration of a polyp segmented region as a polyp candidate.

2. Classification. First examine the set of features for polyp detection. A stepwise forward wrapper subset selection algorithm is applied to a large set of features (curvature based descriptors calculated inside gray-scaled candidate regions of the image and contour based descriptors calculated from the contour of the candidate regions). The wrapper subset selection algorithm consists on the application of a greedy forward feature selection algorithm to all available data using Support Vector Machines (SVM) with (N - 1) fold cross validation. The effect of the use of non-linear kernels in the SVM and the effect of artificially balancing the training set is also tested.

3. Assessment. The set of features determined previously. Multiple 10-fold cross validations with random partitioning are performed, using the determined set of features, to each video sequence in order to assess the validity of the detection and classification using as training data the same video sequence. Furthermore, we apply the detection and classification methods to a video sequence using as training data a different video sequence; and perform analysis of a separate image acquired by normal endoscopy imaging.

In this comprehensive survey, we have given an overview on distinct automatic image/video based polyp detection (localization) and segmentation methods proposed in the literature so far (up to February 2017, and we refer the reader to the project website that is updated continuously with links to all the papers presented here to obtain more details about this research area (Surya Prasath, 2017) A new feature selection method is used to discriminative features to the patients' records which have a significant impact on prediction ability of the algorithms. The rest of this paper is organized as follows: The Survey of video capsule endoscopy details are introduced in Section 2. Discuss about the methodology segmentation and binary pattern using data mining in Section 3. The proposed heuristic methods has applied (Fuzzy and Genetic for detect exact image quality and accuracy using improved kernel SVM in section 4. The experimental results are discussed in Section 5, and finally, Section 6, we discuss about the overall accuracy of proposed and existing methods and conclude our expected result.

Polyp Detection in Capsule Endoscopy Videos

In Figure 1, we show a example polyps (selected from the VCE frames) indicating the amount of protrusion out of the mucosa surface

Figure 2 shows a pedunculated/Stalked polyp that appears in consecutive frames from a VCE exam of a

patient. The stalk of the polyp is not visible, thus appearing like a sessile polyp attached to the mucosal fold at the bottom.

Karargyris and Bourbakis (2009) performed Log–Gabor filter based segmentation along with a smallest univalue segment assimilating nucleus (SUSAN) edge detector, curvature clusters, and active contour segmentation to identify polyp candidates. On a 50-frame video containing 10 polyp frames, they achieved a sensitivity of 100%.

Hwang and Celebi (2010) used watershed segmentation with initial markers selected using Gabor texture features and K-means clustering that avoided the requirement of accurate edge detection. On a set of 128 images with 64 polyp frames, this method achieved a sensitivity of 100%.

Nawarathna et al., (2014) later extended this approach with the addition of local binary patterns (LBP) feature. In addition, a bigger filter bank (Leung–Malik), which includes Gaussian filters, were advocated for capturing texture more effectively. Note that these approaches rely only on texture features and do not include any color or geometrical features. The best result of 92% accuracy was obtained for the Leung–Malik-LBP filter bank with K-NN classifier for 400 images with 25 polyps.

Zhao and Meng (2011) proposed using opponent color moments with LBP texture features computed over contour let transformed images with an SVM classifier, and reported 97% accuracy. Their work unfortunately did not mention how many total frames and polyp frames were used or how the color and texture features were fused.

Mamonov et al., (2014) proposed a system based on sphere fitting for VCE frames. The pipeline consisted of considering the gray-scale images and applying a cartoon and texture decomposition. The texture part of the given frame is enhanced with nonlinear convolution and then mid pass filtered to obtain binary segments of possible polyp regions. Then, a binary classifier uses a best-fit ball radius as an important decision parameter to decide whether there is a polyp present in the frame or not. This decision parameter is based on the assumption that the polyps are characterized as protrusions that are mostly round in shapes, hence polyps which violate this do not get detected.

Materials and Methods

Polyp Segmentation within a VCE Frame

Segmentation of polyps in a single frame of VCE is an (relatively easier) object identification problem. The general task describes under the category of image segmentation, which is a well-defined problem in various biomedical imaging domains. As we have seen before, color, texture, and shape features individually are not discriminative enough to obtain reliable polyp segmentations from VCE frames. There have been a number of efforts in polyp segmentation from VCE which we classify based on which segmentation techniques are utilized. Note that the majority of the previously discussed polyp detection algorithms employ a polyp segmentation step to identify candidate polyp frames, although accurately segmenting the polyp region is not required for subsequent polyp detection in VCE.

The discussion reveals proposed approach of local binary pattern and segmentation system. Next, the polyp detection testing is done using traditional colonoscopy images. Apart from these technological constraints, these approaches require robust embeddable computer vision systems that need to operate under a strict energy budget (battery constraints). However, a computer vision embedded VCE system can revolutionize the diagnosis procedures that have been done manually through tedious processes so far.

Binary pattern classification algorithm (BCA)

In our approach we can't fitting the image itself for this VCE, but we first suite a certain type of a mid-pass filter to it. This allows us to isolate the protrusions with certain size limits. We use the radius of the best fit ball as the decision parameter in a binary classifier. If the decision parameter is larger than the discrimination threshold, then the frame is classified as containing a polyp. Another feature that distinguishes our approach from the ones mentioned above is the use of texture information. The surface of polyps is often highly textured, so it makes sense to discard the frames with too little texture content in them. On the other hand, too much texture implies the presence of bubbles and/or trash liquids in the frame. These unwanted features may lead a geometry-based classifier to classify the frame as containing a polyp when no polyp is present, i.e. they lead to an increased number of false positives. Thus, in order to avoid both of the situations mentioned above, we apply a pre-selection procedure that discards the frames with too much or too little texture content. Combined with the binary classifier this gives the algorithm that we refer to as binary classification with pre-selection.

The algorithm single frames from a video capsule endoscope sequence are moderated. The algorithm makes a decision for every frame whether to classify it either as containing polyps ("polyp" frame) or as containing normal tissue only ("normal" frame) that requires a separate study is the choice of the color space of the frame. In this work we convert the captured color frames to grayscale before processing. This choice provides good polyp detection results, as we observe from the numerical experiments an images acquired by the endoscope are of circular shape. The area of the rectangular frame outside the circular mask is typically filled with a solid color. This creates a discontinuity along the edge of the circular mask, which may cause problems in the subsequent steps of the algorithm. To remove this discontinuity we use a simple linear extrapolation to extend the values from the interior of a circular mask to the rest of the rectangular frame.

In my previous research work related to polyps, I accomplished with solving linear system which corresponds to an upwind discretization of the following PDE assuming that the frame f is given on a N_y X N_x uniform Cartesian grid (1.1). Here the vector field $\vec{r}$ at the pixel (i, j) is the unit vector

$$\nabla \mathbf{f} \cdot \vec{\mathbf{r}} = 1, \qquad (1)$$

With the values of f inside the circular mask fixed, the solution outside the mask provides the desired

extrapolated values (1). To obtain linear system result, we use a standard upwind discretization scheme.

$$\vec{\mathbf{r}}_{ij} = \frac{1}{\sqrt{(i-N_y/2)^2+(j-N_x/2)^2}} \begin{bmatrix} i-N_y/2 \\ j-N_x/2 \end{bmatrix} \quad (1.1)$$

Binary pattern is great in the domain of texture classification retrieval. Traditional binary pattern are rotational invariant form. Today IR domain helps us to find the huge quantity of digital images. The proposed automated system of processing images are accessed from database is significant Jabid et al., (2010). This binary patterns are mostly helps us in common textural descriptions in IR analysis. The segmented pixels are labelled via thresholding P-Neighbour values with centre value and converting result in binary numbers through

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{p-1} s(g_p - g_c)2^p, s(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x<0. \end{cases} \quad (2)$$

Wherein g_c represents the grey value of the centre pixel x_c, y_c and g_p relates to the grey values of equally spaced pixels P on the circumference of circles with radius R. The pixel venues are not exactly estimated via bi-linear mode in the neighbour values. It signs the variant of neighbourhood realized as *P*-bit binary numbers, resulting in 2^p distinct values for binary patterns. Individual pattern values are capable of describing texture data at centre pixel g_c. Binary pattern technique uses 16 = 256 possible texture units rather than the 81 =6561 units used in the textural spectrum technique, resulting in a more effective abstraction of textures that lead to a comparable textures discrimination performance.

Fuzzy Logic for Improved-SVM

In order to avoid the crisp definition of a row in the video capsule endoscopy the polyps disease dataset belonging to one of the classes, fuzzy technique is involved. Initially, we deem to divide the N into P classes and each dataset samples is converted as fuzzy values. The SVM learning process has the task, which offers the architecture and decision function parameters which are representing the largest margin. Those parameters are represented by the vectors in the class boundary and their associated Lagrange multipliers (Winters-Hilt et al., 2006). In order to take into account nonlinearities, a higher dimension space is obtained; this is done by transforming data vectors xi ϵ R^n through a function φ (x). In this transformation, explicit calculation of φ (x) is not necessary; instead of that, just the inner product between mapped vectors is required. For this inner product, kernel functions fulfilling the Mercer condition are usually used. An example of a kernel function is given (3).

$$Q_i^{(x)} = Q_{min} + (Q_{max} - Q_{min})U(0,1) \quad (3)$$

$$\max_{\alpha} \sum_{a=1}^{l} \alpha_a - \frac{1}{2}\sum_{a,b=1}^{l} \alpha_a \alpha_b y_a y_b k(x_a, x_b) \quad (4)$$

Subject to

$$\sum_{a=1}^{l} y_a \alpha_a = 0, 0 \leq \alpha_a \leq \varsigma, a, b = 1, \ldots l \quad (5)$$

where $y_a y_b$ are labels for data vectors a, b respectively and α_i are the Lagrange multipliers with linear inequality constraints. The ζ parameter controls the misclassification level on the training data and therefore the margin. Once a solution is obtained, a endoscopy decision rule used to classify data is defined as in (6):

$$f(x) = sign[\sum_{a=1}^{nsv} \alpha_a y_a k(x_a, x) + bias] \quad (6)$$

Where NSV is the support vectors dimension, BIAS is the projection xi onto the hyper plane that separates the classes, and only non-zero αi for the decision rule. As a result, the data vectors are associated to these multipliers and those are called as support vectors. So there is a solution to separate the problem in smaller sub problems, which are easier to control and resolve. But, an important disadvantage of this approach is that data vectors are selected randomly to build sub problems that can affect the performance and it is providing an inferior learning rate. To solve this problem next section fuzzy rules are created for SVM classifier.

Genetic Fuzzy rules

The fuzzy rules would be optimized using double coding scheme, after it created. This double coding scheme is used for both rule selection CS and lateral tuning CT (Alcala-Fdez et al., 2006).

1) In the part, each chromosome is a binary vector that determines whether the rule is selected or not (alleles "1" and "0," respectively). Considering the M rules that are contained in the candidate rule set, the corresponding part, i.e., $C_s = \{c_1, \ldots, c_M\}$, represents a subset of rules composing the final RB so that IF $c_i = 1$ THEN ($R_i \epsilon$ RB) else (Ri ϵ RB), with R_i being the corresponding i_{th} rule in the candidate rule set and RB being the final RB.

2) An actual coding is considered in CT part. This part is the combination of all α parameters of each fuzzy partition. By considering the following number of labels per variable: $(m_1, m_2, \ldots, m_n)$ with n being the number of system variables. Next, each gene is associated with the tuning value of the corresponding label: $C_T = (c_{11}, \ldots c_{1m}, \ldots c_{n1}, \ldots c_{nm}^n)$

3) At last, a chromosome C is coded in the following way: $C = C_s C_T$

All the candidate rules are included in the population as an initial solution in order to use the rules. The initial pool is obtained with the first individual having all genes with value "1" in the C_s part and all genes with value "0.0" in the C_T part. The remaining individuals are generated at random.

Chromosome Evaluation: To estimate a particular chromosome based on a large number of rules, the classification rate and the following function are computed and maximized:

$$Fitness(C) = \frac{\#Hits}{N} - \delta . \frac{NR_{initial}}{NR_{initial} - NR + 1.0} \quad (7)$$

where # Hits is the number of patterns that are correctly classified, $NR_{initial}$ is the number of candidate rules, NR is the number of selected rules, and δ is a weighting percentage given by the system expert that finds

the tradeoff between accuracy and complexity. If there is minimum one class without selected rules or if there are no covered patterns, the fitness value of a chromosome will be penalized with the number of classes without selected rules and the number of uncovered patterns.

Crossover Operator: The crossover operator will depend on the chromosome part where it is applied.

01. For the CT part, we consider the Parent Centric BLX (PCBLX) operator (Lozano et al.,2004) (an operator that is based on BLX-α). This operator is based on the concept of neighborhood, which allows the offspring genes to be around the genes of one parent or around a wide zone that is determined by both parent genes. Let us assume that $X=(x_1,...x_n)$, where $x_i,y_i \in [a_i,b_i]$ c R =1,....n, are two real-coded chromosomes that are going to be crossed. Generate the following two offspring.

a) $O_1=(x_{11},...x_{1n})$, where O_{1i} is a randomly (uniformly) chosen number from the interval $[i_i^2,u_i^2]$, with. $l_i^1 = \max\{a_i, x_i - I_i,\alpha\}, u_i^1 = \min\{b_i, x_i - I_i,\alpha\}$ and $I_i = |x_i - y_i|$

b) $O_2=(x_{21},...x_{2n})$, where O_{1i} is a randomly (uniformly) chosen number from the interval $[i_i^2,u_i^2]$, with. $l_i^2 = \max\{a_i, x_i - I_i,\alpha\}, u_i^2 = \min\{b_i, x_i - I_i,\alpha\}$ and $I_i = |x_i - y_i|$

02. In the CS part, the half-uniform crossover scheme (HUX) is employed to interchange the mid of the alleles that are altered in the parents (the genes to be crossed are randomly selected from among those that are different in the parents). This operator ensures the maximum distance of the offspring to their parents (exploration).

The two from C_T with the two from C_S are combined to create four offspring. The two best offspring will be considered as two corresponding descendents. In order to do Gray code (binary code) with a fixed number of bits per gene (BITSGENE), the hamming distance is calculated between two individuals to apply the crossover operators. Example of the selected rules

1. If [BMI > 25, Pulse rate >80 , Homocystein = High] then class C = 1 else C=0.

2. If [PulseRate < 60 or PulseRate > 100, Alcohol= true] then C = 1 else C=0.

3. If [BMI > 25, Smoking = true, Sugar > 140] or [BMI = 25, Smoking = true, PulseRate > 100] then C = 1 else C=0.

Figure 1. Example of Polyps

Table 1. Confusion Matrix

Visual Class	Actual class	Actual class
Predicted class	True positive (TP)	False positive (FP)
Predicted class	False Negative (FN)	True Negative (TN)

4. if [PulseRate > 100, Sugar > 140, Alcohol= true] or if [PulseRate > 100, Sugar > 140, Stress = high] then C = 1 else C=0.

5. If [PulseRate > 100, Sugar > 140, High fatty diet = true] or if [BMI > 25, Sugar > 140, Stress = high] then C = 1 else C=0.

6. If [CPK > 25, BMI > 25, High fatty diet = true] or if [BMI > 25, PulseRate > 100, Alcohol= true] then C = 1 else C=0 .

7. If [Cardiac Troponin I > 10, BMI > 25, Stress = high] or if [BMI > 25, Sugar > 140, Alcohol= true] then C = 1 else C=0 .

8. If [Troponin T > 0.01, PulseRate > 100, Homocystein = high] or if [PulseRate > 100, Sugar > 140, Stress = high] then C = 1 else C=0 .

9. If [Troponin T > 0.01, PulseRate > 100, High sensitive C reactive protein in blood = true] or if [PulseRate > 100, Sugar > 140, Homocystein = high] then C = 1 else C=0 .

10. If [CPK > 25, BMI > 25, Troponin T > 0.01, Stress = high ,High sensitive C reactive protein in blood = true] or if [BMI = 25, Sugar > 140, High sensitive C reactive protein in blood = true] then C = 1 else C=0.

Restarting Approach: In restarting procedure, the best chromosome is maintained, and the remaining is generated randomly. If the threshold value L is below zero the restart procedure is used. That means that all the individuals synchronized in the population. Then classification task of the dataset is performed and the results are conceived and it is presented in the next section.

Results

The proposed work is implemented in MATLAB environment. MATLAB is mainly used for machine learning, data mining, text mining and business analytics. It is applied in the area of research, education, training and industrial applications. The experiments are designed with different parts of the work. These different parts include the evaluation of the features of the dataset and the feature selection. In order to achieve this, first the

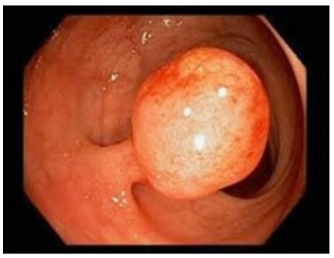

Figure 2. Pedunculated/Stalked Polyp

Table 2. Results of the VCE Cancer Dataset

Methods	F-measure (%)	Precision (%)	Recall (%)	Accuracy (%)	Error Rate (ER) (%)
SVM	91.82	93.83	90.04	86	12
ISVM	93.341	94.97	92.12	88.695	9.241
FISVM	93.463	93.42	94.06	90.25	7.4
GF-IKSVM	96.201	96.33	95.724	93.2	4.6

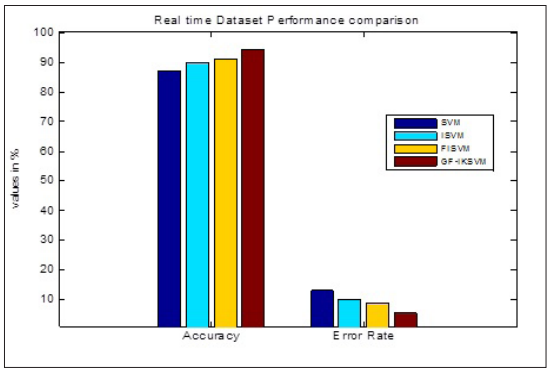

Figure. 3. Accuracy and Error Results Comparison of Methods for the Cancer Disease Dataset

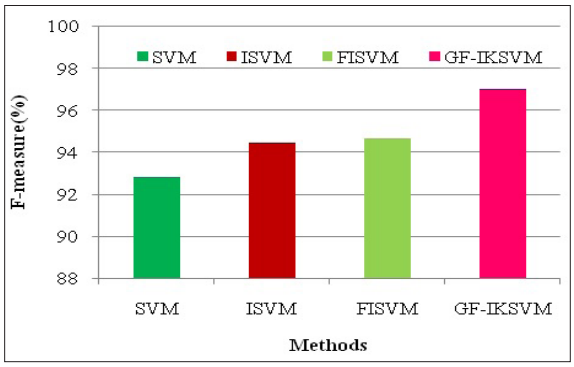

Figure 5. F-Measure Results Comparison of Methods for the Endoscopy Colon Poly Dataset

features were selected by the feature selection method and their importance are discussed. Second, all the two possible combinations of the feature selection and classification methods are tested over the cancer dataset. Finally, result section presented accuracy. This is the most important performance measurement parameter in the medical field (Lavrac, 1999), which are mostly discussed in the literature. So this measurement is used to find the performance of algorithms.

Confusion matrix

A confusion matrix is a table that allows visualization of the performance of an algorithm. In a two class problem (with classes C1 and C2), the matrix has two rows and two columns that specifies the number of False Positives (FP), False Negatives (FN), True Positives (TP), and True Negatives (TN). These measures are defined as follows: TP is the number of polyp samples of class C1 which has been correctly classified. TN is the number of samples of class C2 which has been correctly classified. FN is the number of samples of class C1 which has been falsely classified as C2. FP is the number of samples of class C2 which has been falsely classified as C1. Table 1 shows confusion matrix.

Visual Class Actual class
Actual class
Predicted class
True positive (TP) False positive (FP)
Predicted class
False Negative (FN) True Negative (TN)
Precision and Recall

According to confusion matrix, precision and recall are explained as following,

$$Precision = \frac{TP}{(TP + FP)} \quad (8)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (9)$$

Accuracy: Accuracy shows ratio of correctly classified samples to the total number of tested samples. It is defined as:

$$Accuracy = \frac{(TN + TP)}{(TP + TN + FN + FP)} \quad (10)$$

F-measure: Harmonic mean value of the precision and recall is known as F-measure is defined as follows:

$$F - measure = \frac{2 * P * R}{(P + R)} \quad (11)$$

Table 2 shows the results for the cancer disease dataset and Figure 3 shows the performance comparison results of accuracy and error between existing and proposed algorithms for the cancer dataset.

From the experimental results accuracy and error rate is explained that for the cancer dataset the proposed GF-IKSVM algorithm performs 7.6 % better than the SVM algorithm, 4.462% better than the ISVM algorithm and 3.4% better than the FISVM algorithm is illustrated in Figure 3. Proposed GF-produces 5.2 % error value which is 3.5 % lesser than the FISVM algorithm, 4.46% lesser than the ISVM algorithm and 7.3% lesser than the FISVM

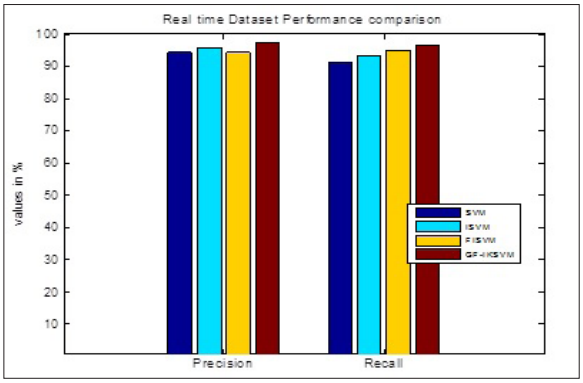

Figure 4. Precision and Recall Results Comparison of Methods for the Cancer Dataset

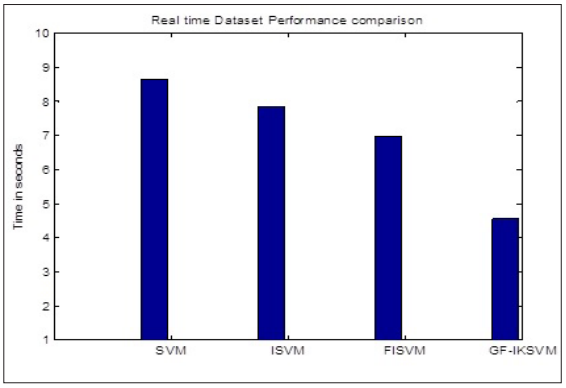

Figure 6. Time Comparison Results of Methods for the Endoscopy Colon Poly Cancer Dataset

algorithm is illustrated in Figure 3. It concludes that the proposed GF-IKSVM algorithm produces higher accuracy value and less error rate when compared to all methods.

From the experimental results it is concluded that TP rate for the cancer dataset the proposed GF-IKSVM algorithm 2.486% better than the SVM algorithm, 1.346% better than the ISVM algorithm and 2.898% better than the FISVM algorithm is illustrated in Figure 4. Similarly FP rate for the cancer dataset the proposed GF-IKSVM algorithm performs 1.675% better than the SVM algorithm, 3.607% better than the ISVM algorithm and 5.4107% better than the FISVM algorithm is illustrated in Figure 4. It is clear from the results that the proposed classifier produces less error results.

From the experimental results it is concluded that F-Measure for the cancer dataset the proposed GF-IKSVM algorithm performs 4.201% better than the SVM algorithm, 2.59% better than the ISVM algorithm and 2.378% better than the FISVM algorithm is illustrated in Figure 5.

Discussions

From the experimental results it is concluded that F-Measure for the cancer dataset the proposed GF-IKSVM algorithm takes 4.35 seconds, whereas the FISVM algorithm takes 6.79 seconds, ISVM algorithm takes 7.76 seconds and SVM algorithm takes 8.58 seconds is illustrated in Figure 6. It concludes that the proposed GF-IKSVM performs quicker when compared to other methods.

Conclusion and future work

In this work, numerous algorithms were applied on the cancer disease dataset and the results were discussed. The features used in this dataset are important of VCE and it is taken by using medical knowledge. Additionally, data mining techniques with feature selection were used to improve the accuracy. Segmentation method is proposed to overcome incomplete dataset problem. Then optimal features are selected using binary classification evolutionary algorithm. As well, the features used in this study, can be calculated by using the objective function. Genetic Fuzzy based Improved Support Vector Machine (GF-ISVM) classifier is proposed in our work for optimisation of fuzzy rules. Fuzzy logic is introduced to create rules of a dataset belonging to one of the classes. The rules are fully based on the classes. Genetic Algorithm (GA) is used to perform fuzzy association rule extraction, rule pre-screening, rule selection and lateral tuning. The accuracy value achieved in this study is higher than currently reported values in the literature. It was shown that significant improvement was gained over by using heuristic methods. The GF-IKSVM is evaluated using the performance metrics precision, recall, F-measure and classification accuracy. In future work it needs to propose cost sensitive algorithms for feature selection. At last, big datasets, more features and also broader data mining approaches, could be used to achieve better and more interesting results.

References

Alcala-Fdez J, Alcala R, Herrera F (2011). A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. *IEEE Trans Fuzzy Syst*, **19**, 857-72.

El Khatib A, Werghi N, Al-Ahmad H (2015). Automatic polyp detection: A comparative study. In proceedings of the 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC), Milano, Italy, 25–29 August 2015, pp 2669-72.

Geetha k, Rajan C (2016). Automatic colorectal polyp detection in colonoscopy video frames. *Asian Pac J Cancer Prev,* **17** 4869-73.

Gostout C Adler D (2003). Wireless capsule endoscopy. *Hosp Physician*, **1**, 16-22.

Hwang S, Celebi ME (2010). Polyp detection in wireless capsule endoscopy videos based on image segmentation and geometric feature. In proceedings of the 2010 IEEE international conference on Acoustics, Speech and signal processing, Dallas, TX, USA, 14–19 March 2010, pp 678–81.

Iakovidis DK, Maroulis DE, Karkanis SA, Brokos A (2005). A comparative study of texture features for the discrimination of gastric polyps in endoscopic video. In Proceedings of the 18th IEEE symposium on computer-based medical systems, Washington, DC, USA. June 2005, pp 575–80.

Iddan G, Meron G, Glukhovsky A, Swain F (2000). Wireless capsule endoscopy. *Nature,* **405**, 417.

Jabid T, Kabir MH, Chae O (2010). Robust facial expression recognition based on local directional pattern. *ETRI J Journal*, **32**, 784-94.

Karargyris A, Bourbakis N (2009). Identification of polyps in wireless capsule endoscopy videos using log Gabor filters. in proceedings of the 2009 IEEE/NIH life science systems and applications workshop, Arlington, TX, USA, 9–10 April 2009, pp 143–7.

Lavrac N (1999). Selected techniques for data mining in medicine. *Artif Intell Med*, **16**, 3-23.

Lin Z, Liao Z, McAlindon M (2014). Handbook of capsule endoscopy; Springer: Dordrecht, the Netherlands.

Mamonov AV, Figueiredo IN, Figueiredo PN, Tsai YHR (2014), Automated polyp detection in colon capsule endoscopy. *IEEE Trans Med Imaging*, **33**, 1488–1502.

Nawarathna R, Oh J, Muthukudage J, et al (2014). Abnormal image detection in endoscopy videos using a filter bank and local binary patterns. *Neurocomputing*, **144**, 70–91.

Park SY, Sargent D, Spofford I, Vosburgh KG, A-Rahim Y (2012). A colon video analysis framework for polyp detection. *IEEE Trans Biomed Eng*, **59**, 1408–18.

Potter J, Slattery M, Bostick R, Gapstur S (1993). Colon cancer: a review of the epidemiology. *Epidemiol Rev*, **15**, 499-545.

Prasath VBS (2015). On fuzzification of color spaces for medical decision support in video capsule endoscopy. In proceedings of the 26th modern artificial intelligence and cognitive science conference, Greensboro, NC, USA, 25–26 April 2015, pp 1–5.

Ruano J Martinez F, Gomez M, Romero E (2013). Shape estimation of gastrointestinal polyps using motion information. In proceedings of the IX international seminar on medical information processing and analysis, Mexico City, Mexico, 11–14 November 2013.

Surya Prasath VB (2017). Polyp detection and segmentation in video capsule endoscopy: A review. *J Imaging*, **3**, 1-15.

van Wijk C, Van Ravesteijn VF, Vos FM, Van Vliet LJ (2010). Detection and segmentation of colonic polyps on implicit iso surfaces by second principal curvature flow. *IEEE Trans*

Med Imaging, **29**, 688–98.
Winters-Hilt S, Yelundur A, McChesney C, Landry M (2006). Support vector machine implementations for classification and clustering. *BMC Bioinformatics*, **7**, 1.
Zhao Q, Meng MQH (2011). Polyp detection in wireless capsule endoscopy images using novel color texture features. In proceedings of the 9th world congress on intelligent control and automation, Taipei, Taiwan, 21–25 June 2011, pp 948–52.