Detection of small bowel tumor in wireless capsule endoscopy images using an adaptive neuro-fuzzy inference system

Mahdi Alizadeh, Omid Haji Maghsoudi, Kaveh Sharzehi, Hamid Reza Hemati, Alireza Kamali Asl, Alireza Talebpour

1 Department of Bioengineering, Temple University, Philadelphia, PA 19121, USA; 2 Department of Medicine, Section of Gastroenterology, School of Medicine, Temple University, Philadelphia, PA 19140, USA; 3 Department of Radiation Medicine Engineering, and 4 Department of Electrical and Computer Engineering, Shahid Beheshti University, Tehran 1983963113, Iran.

Abstract

Automatic diagnosis tool helps physicians to evaluate capsule endoscopic examinations faster and more accurate. The purpose of this study was to evaluate the validity and reliability of an automatic post-processing method for identifying and classifying wireless capsule endoscopic images, and investigate statistical measures to differentiate normal and abnormal images. The proposed technique consists of two main stages, namely, feature extraction and classification. Primarily, 32 features incorporating four statistical measures (contrast, correlation, homogeneity and energy) calculated from co-occurrence metrics were computed. Then, mutual information was used to select features with maximal dependence on the target class and with minimal redundancy between features. Finally, a trained classifier, adaptive neuro-fuzzy interface system was implemented to classify endoscopic images into tumor, healthy and unhealthy classes. Classification accuracy of 94.2% was obtained using the proposed pipeline. Such techniques are valuable for accurate detection characterization and interpretation of endoscopic images.

Keywords: adaptive neuro-fuzzy inference system, co-occurrence matrix, wireless capsule endoscopy, texture feature

Introduction

Until the last decade, visualization of the entire small bowel was a major challenge for gastroenterologists. The advent of Wireless capsule endoscopy (WCE) has enabled gastroenterologists to examine the entire small bowel tract[1-3]. WCE is a pill shape device 26 mm × 11 mm in size and weighing 4 g. It is equipped with a camera, light source (usually four white LEDs), radio-transmitter and batteries (Fig. 1)[4].

Once the WCE is swallowed by the patient, it starts sending images to the recorder which is attached to the patients' waist. Imaging will continue until battery dies out (6-8 hours). In the interim, the capsule sends approximately 50,000 images including 100 images from the mouth and esophagus, 4,000 from the stomach, and 30,000 from the small bowel and 20,000 from the large intestine[5-6]. Due to the large amount of images, diagnosis, interpretation and analysis of images is very time consuming; gastroenterologist might spend at least

The authors reported no conflict of interests. This is an open access article under the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited.

© 2017 by the Journal of Biomedical Research. All rights reserved doi:10.7555/JBR.31.20160008
2-3 hours reading and analyzing an examination, depending on the clinician's experience and complexity and type of gastric diseases\cite{6}.

Recently, several computer aided diagnosis systems have been developed. In general, a computer based diagnostic system consist of several units, namely, data acquisition system (WCE), feature extraction unit, decision system or classification unit and user interaction unit\cite{7}. Feature extraction unit is used to prepare data in a format that is easy for a decision support system or a classification unit to use\cite{7}. Different feature extraction and classification methods have been used for detection of abnormalities including: statistical texture features obtained by the texture unit numbers over the histogram spectrum\cite{8-9}, co-occurrence matrix\cite{10}, curvelet transformation\cite{11}, Gaussian modeling of color curvelet covariance coefficients\cite{12}, local binary pattern and wavelet transform\cite{8} and nonlinear analysis in RGB color space\cite{13} have been used as a part of detection approaches. The decision system or classification unit makes decisions using data achieved by the feature extraction unit. This system provides a list of cause-and-effect reasoning from the features and/or their possible corresponding treatment\cite{7}. There are a variety of texture classification systems such as Neuro-Fuzzy, Extended Normalized Radial Basis Function (ENRBF), Radial Basis Function (RBF), Adaptive Neuro-Fuzzy, Support Vector Machine (SVM), Gaussian Mixture Model (GMM), k-nearest-neighbor (kNN) and Multi-Layer Perception (MLP)\cite{11-15}.

Computer based diagnostic system is able to classify an image as normal or abnormal and acts as a second "eye" by processing the endoscopic video, which facilitates abnormality detection by multiple techniques. However, literature indicates that post processing of endoscopic images has been poorly explored and is still undergoing testing\cite{7-14}. Given that only small bowel tumor images exhibit great variations in color, size and shape, it will be even more complicated if we consider other abnormalities to be differentiated in the computer aided system. Therefore, a robust and well developed classification system is needed to make a right decision about the status of these complex images.

It has been found that substantial recognition improvements may be obtained in difficult pattern recognition problems by combining or integrating outputs of multiple classifiers. To the best of our knowledge, ensemble methods of classification have received little attention in the field of CE images.

The combination of neural networks (NNs) and fuzzy logic were applied for the purpose of discriminating small bowel tumor endoscopic images from normal and other abnormal small bowel endoscopic images. Although, NNs and fuzzy logic systems have well-established strengths and weaknesses, they are both capable of modeling highly complex nonlinear relationships. In general, neuro-fuzzy systems attempt to combine the low-level numerical modeling capabilities of NNs with some of the representational transparencies of fuzzy logic, but like fuzzy logic often suffer from the so-called "curse of dimensionality"\cite{7-14}.

In this research, a new approach of obtaining statistical features/parameters from the texture spectra was proposed in both chromatic and achromatic domains of the image. Multiple-classifier scheme has been adopted, where the fusion of individual outputs was realized using fuzzy integral. Texture features with maximal dependence on the target class as defined by an independent board certified gastroenterologist and with minimal redundancy between features were selected to
build a classifier, Adaptive Neuro-Fuzzy Interface System (ANFIS) written in MATLAB, to be used to discriminate between WCE images.

Materials and methods

Subjects

A total of 435 frames selected from 35 patients (with a mean age of 45.2 years; range 21-58 years) including 10 healthy subjects, 10 subjects with small bowel tumors and 15 patients with gastric diseases (5 patients with celiac disease, 5 with Crohn’s disease and 5 with lymphangiectasia disease) were examined. Small bowel tumors included adenocarcinoma (n = 3), sarcoma (n = 4) and carcinoid tumors (n = 3).

Feature extraction

The normalized 2D co-occurrence matrices for each image using four directions with distance of D = 1 were implemented. The displacement vectors are listed in Table 1. Assume that p is the number of gray level co-occurrence matrices, the following features were used in the proposed algorithm:

- Contrast:
  \[ \sum_{i=1}^{M} \sum_{j=1}^{N} \|i-j\|^2 p(i, j) \]  
  (1)

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Displacement vector of 2D co-occurrence matrices.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direction</td>
<td>0°</td>
</tr>
<tr>
<td>Displacement Vector</td>
<td>(D,0)</td>
</tr>
</tbody>
</table>

It measures local variations in the gray level co-occurrence matrix and provides evidence of how sharp the structural variations in the image are.

- Energy (angular second moment):
  \[ \sum_{i=1}^{M} \sum_{j=1}^{N} p(i, j)^2 \]  
  (2)

It gives a strong measure of uniformity. Higher non-uniformity values provide evidence of higher structural variations.

- Correlation:
  \[ \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{(i-\mu)(j-\mu)p(i, j)}{\sigma_i \sigma_j} \]  
  (3)

The correlation feature is the measure of gray level linear dependency of the image.

- Homogeneity:
  \[ \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{p(i, j)}{1 + |i-j|} \]  
  (4)

It measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal and increases with less contrast in the window.

Dimensionality reduction is very important in statistical pattern recognition, machine learning, data mining and statistics. In theory, more features can be effective in classification but in practice causes overfitting in training data set [15-16]. In this paper, features in RGB color space were not used because in natural images the correlation between these components is not zero. Also, proposed features were not calculated from V channel in HSV color space. V shows the intensity of each pixel and highly depends on the position of the camera [39].

Mutual information

Feature selection is a routine that excerpts a subset of primitive features. Evaluation criterion acts as a measuring tool to generate optimum feature subset from the extraction. In this paper, mutual information was used as feature selection method. It has several advantages compared to the other methods. First, it measures general statistical dependence between variables which makes no assumption about the nature of the relationship between variables. Second, mutual information (MI) is invariant to monotonic transformations performed on the variables and finally dependent on the decision algorithm, thus reducing computational complexity [17-18]. Consider X and Y are random variable, the MI \(I(X;Y)\) is defined as:

\[ I(X;Y) = H(X) - H(X|Y) \]  
(5)

where \(H()\) is the entropy of a random variable, which measures uncertainty. We can estimate MI between two variables by using histogram approach as follows:

\[ I(x;y) = \sum_{x} \sum_{y} \frac{P(X,Y)}{P(X)P(Y)} \log_{2} \frac{P(X,Y)}{P(X)P(Y)} \]  
(6)

where \(P(X)\) and \(P(Y)\) are the marginal distributions and \(P(X,Y)\) is the joint probability distribution.

Texture features were selected from HSV color space, since it is more similar to the physiologic perception of
human eye\cite{10}, and therefore is more adequate than the standard RGB color space. Primarily, 32 features including homogeneity, correlation, energy and contrast from co-occurrence matrices of H and S channels in degrees of 0, 45, 90 and 135 were calculated. The features were ranked by MI and then the top $k$ ranked features (8 features are shown in Table 2) were selected. The number of top features ($k$) can be determined by user or adjusted experimentally.

**ANFIS**

ANFIS combines the benefits of two powerful methods (i.e. NNs and Fuzzy Inference System (FIS)) and works by applying neural learning rules to identify and tune the parameters and structure of a fuzzy inference system. The attractive features of an ANFIS include: (1) ease of implementation, (2) fast and accurate learning, (3) strong generalization abilities, (4) excellent explanation facilities through fuzzy rules and (5) easy to incorporate both linguistic and numeric knowledge for problem solving. Fig. 2 shows the first order Sugeno ANFIS architecture\cite{19}.

In this connected structure, the input (8 inputs) and output (1 output) nodes represent the training values and the predicted values, respectively, and in the hidden layers, there are nodes functioning as membership functions using Gaussian distribution function and rules (311 membership functions and 311 rules). ANFIS is based on a first-order Sugeno fuzzy model and uses two fuzzy if-then rules as defined below:

Rule 1: If ($x$ is $A_1$) and ($y$ is $B_1$)

then ($f_1 = p_1x + q_1y + r_1$) \hspace{2cm} (7)

Rule 2: If ($x$ is $A_2$) and ($y$ is $B_2$)

then ($f_2 = p_2x + q_2y + r_2$)

where $x$ and $y$ are inputs, $A_i$ and $B_i$ are the fuzzy sets, $f_i$ are the outputs, $p_i$, $q_i$ and $r_i$ are the linguistic labels associated with the node function.

The error measure to train the above-mentioned ANFIS was defined as:

$$E = \sum_{k=1}^{n} (f_k - f_k)^2$$ \hspace{2cm} (8)

Where $f_k$ and $f_k$ are the $k$th desired and estimated output, respectively and $n$ is the total number of pair (inputs-outputs) of data in the training set\cite{20-22}.

ANFIS model implemented in MATLAB was trained to classify test samples into one of those three classes defined by independent board certified gastroenterologist: set A (abnormal), set B (normal) and set C (tumor). The criteria used to define and label target classes are described in Table 3.

The performance of the classifiers can be determined by the computation of sensitivity, specificity and total accuracy by using the following statistics:

**Classification Accuracy**

$$\text{Classification Accuracy} = \frac{\text{correctly classified samples}}{\text{classified samples}}$$

**Sensitivity**

$$\text{Sensitivity} = \frac{\text{correctly classified positive samples}}{\text{true positive samples}}$$ \hspace{2cm} (9)

**Specificity**

$$\text{Specificity} = \frac{\text{correctly classified negative samples}}{\text{true negative samples}}$$

where true positive samples are the total of normal samples, and true negative samples are the total of abnormal samples. The threshold value of 0.125 was considered during performance measurements which means that the values in the range of 0.25±0.125 are considered as normal image, 0.5±0.125 (mean±stand-

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal image</td>
<td>0.25</td>
</tr>
<tr>
<td>Abnormal image with gastric tumor</td>
<td>0.5</td>
</tr>
<tr>
<td>Other types of abnormalities</td>
<td>0.75</td>
</tr>
</tbody>
</table>

**Table 2** Final selected features using the mutual information method.

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Channel</th>
<th>Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Homogeneity</td>
<td>H</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Energy</td>
<td>S</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Homogeneity</td>
<td>S</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Contrast</td>
<td>H</td>
<td>45</td>
</tr>
<tr>
<td>5</td>
<td>Energy</td>
<td>H</td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td>Contrast</td>
<td>S</td>
<td>90</td>
</tr>
<tr>
<td>7</td>
<td>Energy</td>
<td>S</td>
<td>90</td>
</tr>
<tr>
<td>8</td>
<td>Contrast</td>
<td>S</td>
<td>135</td>
</tr>
</tbody>
</table>

**Table 3** Criteria for defining target class.
standard deviation) as abnormal image with gastric tumor and $0.75\pm0.125$ as abnormal image with other gastric diseases.

The experimental training set consisted of 315 frames from different endoscopic examinations taken at Shariati Hospital in Tehran, Iran. One hundred and ten of these frames did not present any abnormality, while the rest were selected as representing tumor pathology pattern (73 frames) and other types of gastric diseases including lymphangiectasia, Celiac and Crohn’s (132 frames). The selection was made by an independent board certified gastroenterologist at Temple Hospital. The test set contains 120 frames (44 frames from healthy exams, 20 frames contain small bowel tumors and 56 frames with other gastric diseases (i.e., lymphangiectasia, celiac and Crohn’s disease) which randomly selected from different endoscopic exam and are different than the training set. Fig. 3 shows some frames belonging to the training and test set. Endoscopic images were obtained initially from the M2A microcapsule. They had spatial resolution of 258 pixels, a brightness resolution of 256 levels per color.

**Fig. 2** ANFIS architecture.

**Fig. 3** WCE images from healthy subjects (A), and subject with angiectasia (B), Celiac disease (C), Crohn’s disease (D) and small bowel tumor (E).
plane (8bits), and consisted of three color planes (red, green and blue) for a total of 24 bits per pixel.

Results

The training error converged to 0.03 after 1000 epochs in the system which subtractive clustering was used to determine the number and type of membership functions and back propagation method was used to update the parameters of the proposed Second Order Sugeno-ANFIS, while it reached to 0 only after 20 epochs in the same system but used hybrid optimization technique. The premise parameters were not fixed in the first method, because the convergence of training took more iteration. Subtractive clustering function was used to produce accurate output values by using a large number of membership functions. Table 4 and 5 represent the performance of the classification unit of the proposed computer aided systems.

Table 6 shows an overview of the overall accuracy, specificity and sensitivity values which have been reported in literature and what achieved in this paper using the proposed technique.

Discussion

A computer aided system was presented to be used as a supportive tool for automatic detection of a broad spectrum of abnormalities in WCE, considering color as a discriminative feature of importance. The diagnostic accuracy of the proposed computer-based algorithm is high and not inferior to the results obtained by experts. In this study, as well as in previous studies by other groups, overall diagnostic accuracy of experts in academic centers is in the range of 90% to 95%, which is considered to be high enough for a feasible “resect and discard” model. Such techniques are extremely valuable for accurate detection characterization and interpretation of the endoscopic images in normal and patients with gastric tumors and other gastric diseases. In addition, the standard, formal histopathology, does not reach a 100% diagnostic accuracy in general, because of confusion of samples or misdiagnosis. The superiority of this study is to detect and distinguish different abnormalities as well as tumors with higher performance using fewer features and less complicated steps which is important in supervised techniques.

Regarding texture characterization, it is well known that the most important information lies in the H and S channels of the HSV color transform depending on the image intrinsic spatial resolution. The technique works completely in the spatial domain and uses the texture information to differentiate endoscopic images. The design of algorithm has been based on the concept that different types of tissues have different textural features. This information is used to classify an image based on the pre-defined class by a clinician.

However, ANFIS was trained without filtering the data for bad signals, lost signals or other signal procurement deficiencies. Therefore, better accuracies can be expected if the signal is filtered. Regarding the clinical application of the proposed method, we aim indeed to introduce the developed method in a supervised computer-aided diagnostic system, where the clinicians acts as a final reviewer of the selected frames containing potential textural abnormalities. Therefore, the primary aim of the present work is to facilitate the analysis process of a capsule endoscopy video, reducing the time required by the physician to review the acquired data and to detect abnormalities. Additionally, it may contribute to a smaller dependency on physician expertise, allowing users with less experience to achieve better diagnosis performance.

From Table 4 and Table 5, we notice that ANFIS

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Performance measurements for the first classifier.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>Set A (abnormal)</td>
<td>87.5%</td>
</tr>
<tr>
<td>Set B (normal)</td>
<td>75%</td>
</tr>
<tr>
<td>Set C (tumor)</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Performance measurements for the second classifier.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>Set A (abnormal)</td>
<td>98.2%</td>
</tr>
<tr>
<td>Set B (normal)</td>
<td>86.36%</td>
</tr>
<tr>
<td>Set C (tumor)</td>
<td>100%</td>
</tr>
</tbody>
</table>
classifier can improve detection accuracy of the proposed textural features up to 94.2%, respectively. In our case, since the candidate features are from different color spaces, the features may be complementary. That might be one reason why ANFIS produces better detection performance.

It is shown that the second proposed classifier has better performance than the first classifier. One explanation for the superior performance of the second classifier over the first classifier is that the second classifier has a built-in regularization mechanism to overcome the possible over-fitting on the training data. Concerning the efficiency, the first classifier produces a great computational burden due to its large search range in the search algorithm. Hence, the first classifier is very time consuming. As a matter of fact, the time consumption of each run of the first classifier in our experiments is much larger than that of the second classifier.

To approach a reliable and trustable model, it is essential to have proper training and testing data set. If the data sets are not selected properly, the testing data set will not validate the model. For the proper data set, the testing error decreases with the training proceeding until a jump point. Over fitting occurs when the training passes that point. The optimization methods are used to learn about the training data. During the learning process, the parameters of the memberships are updated. In MATLAB, the two ANFIS parameter optimization methods are hybrid (the default, mixed least squares and back propagation) and back propagation. Error tolerance is used as training stopping criterion, which is related to the error size. The training will stop after the training data error remains within this tolerance\(^ {[23-24]}\).

The time required for training procedure depends on the number and the size of the training samples, and the number of iterations selected for ANFIS. Increasing number of iterations decreases the training error while increasing the computation time. Computation time required for training system in the first system was 28.18 s (mean ± SD) and in the second system was 253.4 ± 5.9 s (mean ± SD). This was achieved using Intel i5, 2.66 GHz processor with 3G of RAM. Computation time was not optimized as the whole procedure was coded in MATLAB.

To assess the performance of the proposed scheme from medical point of view, certified experienced physician further evaluated the experimental results. Most WCE images used in our experiments are representative ones that may be present in the gastrointestinal tract. For those cases that the proposed scheme fails to recognize, one typical case is that a WCE image contains bubbles, which may hamper computerized detection. In addition, a physician judged the status of all the experimental data using the other physician’s evaluation as the ground truth, and the other physician also implemented such an experiment. Excluding uninformative regions using pre-processing steps will improve detection accuracy but this procedure will be time consuming. Also, by increasing the number of training images, accuracy will be improved.

For computer-aided diagnosis for tumor in WCE images, it is worth pointing out that there is no real-time need at present since the diagnosis itself is not a real-time process and we can implement feature selection in a complete off-line mode.

| Table 6 | An overview of the common available methods reported in literature. |
|---------|-----------------|-----------------|-----------------|
|         | Sensitivity     | Specificity     | Accuracy        |
| kNN based on multi scale local binary patterns features\(^ {[22]}\) | 84.5%            | 79.33%          | 81.92%          |
| MLP based on multi scale local binary patterns features\(^ {[22]}\) | 87.16%           | 84.33%          | 85.75%          |
| SVM based on multi scale local binary patterns features\(^ {[22]}\) | 87.99%           | 86.00%          | 87.00%          |
| kNN based on color wavelet covariance features\(^ {[22]}\) | 73.83%           | 68.33%          | 63.32%          |
| MLP based on color wavelet covariance features\(^ {[22]}\) | 74.67%           | 71.17%          | 72.92%          |
| SVM based on color wavelet covariance features\(^ {[22]}\) | 67.00%           | 71.67%          | 69.33%          |
| RBF using NTU-based features\(^ {[7,23]}\) | -                | -               | 91.43%          |
| Adaptive fuzzy logic system (AFLS) using histogram based features\(^ {[7,23]}\) | -                | -               | 92.85%          |
| Fuzzy inference neural network (FINN) using histogram based features\(^ {[7,23]}\) | -                | -               | 88.57%          |
| Adaptive fuzzy logic system (AFLS) using NTU-based features\(^ {[7,23]}\) | -                | -               | 95.71%          |
| Fuzzy inference neural network (FINN) using NTU-based features\(^ {[7,23]}\) | -                | -               | 94.28%          |
| ANFIS1 based on histogram and co-occurrence matrix based texture features in this paper | 85%              | 95.00%          | 85%             |
| ANFIS2 based on histogram and co-occurrence matrix based texture features in this paper | 94.16%           | 96.27%          | 94.2%           |
In conclusion, nowadays, computer-aided diagnosis plays an important role in medicine that helps physicians in the interpretation of medical images. Using fuzzy set theory will help us to use uncertainty in classifier design and consequently will increase the accuracy of detection systems. The two ANFIS classifiers were used to classify WCE images into normal, abnormal, and tumor groups. The total classification accuracy of the proposed ANFIS models was 85\% and 94.2\%. Therefore, we concluded that the ANFIS model with subtractive FIS and hybrid optimal method can be used to classify abnormalities in WCE images.

Acknowledgments

The authors thank all the staff of the Endoscopy Unit of Internal Medicine at Shariati Hospital for their helpful contribution.

References


