Efficient Gastric Cancer Classification from Endoscope Images

Thi Thi Tun
Faculty of Computer Science,
University of Computer Studies (Myitkyina)
thithitun.cumkn@gmail.com

Thwe
Faculty of Computer Science,
University of Computer Studies (Taunggyi)
thwe.ucstaunggyi@gmail.com

Abstract

Nowadays, computer assisted approaches for analyzing the images have increased because the manual interpretation of the image is a time consuming process and suitable to human errors. So, this system is implemented as the computer assistance classification system for analyzing endoscope image about gastric cancer. To classify the gastric cancer stage, this system is based on image processing technique. Before classification, this system extracts geometrical features and first-order (FO) statistic features from the endoscope image. By using these features, this system classifies the gastric cancer stage according to the support vector machine (SVM) classifier. So, this system helps the physician by identifying the gastric cancer stage.

Keywords: Gastric Cancer, Classification, SVM.

1. Introduction

In the worldwide, cancer is the dangerous disease in any part of the human body. Among many cancer disease, gastric cancer is a highest ratio of deaths. Doctors examines the biological tissues of a patient to detect cancer. But, this process consumes more time and manpower. In this situation, image processing techniques are useful in several medical areas. So, the proposed system uses the image processing techniques to classify the gastric cancer stage. This system classifies the gastric cancer stage by using training endoscope gastric cancer image.

For classification, this system extracts the geometrical features and first-order statistic features from all training endoscope images. To extract these features from the endoscope image, this system calculates area, diameter, perimeter, irregularity index, standard deviation, smoothness and entropy. Using these features, this system classifies the gastric cancer stage according to the support vector machine (SVM) classifier. For medical domain, this system can help in the detection of gastric cancer. Computer-aided classification systems have reduced the misclassification rate in cancer diagnosis and improve the time efficiency in analysis.

2. Related Work

In 2017, G. Chen, Y. Zhang and B. Chen [1] solved the electromagnetic inverse scattering problem. Using support vector machine (SVM), this technique converted this problem into a regression problem. SVM deals with nonlinearity and ill-posedness inherent in this problem. SVM method effectively located the tumor of stomach regardless of the presence of noise.

In 2018, A. Chanchlani, A. Bhatt and P. Awaghade [2] described a framework for automatic detection and classification of gastric cancer based on deep learning. A deep neural network has been built for detecting and classifying abnormal and normal CT images from obtained datasets. This detection and classification of gastric cancer had a greater value as it helped doctors and medical field.

In 2020, M. A. Aslam, C. Xue and K. Wang [3] presented a diagnosis of gastric cancer classification system. They used the support vector machine (SVM) for binary classification. The processed Raman dataset was used to train and test the developed model. SVM based neural networks were established using different kernels, which produced different results.

3. Pre-processing

Image preprocessing means standardization of images before feature extraction. In this system, pre-processing includes:

- RGB to Grayscale Conversion: The difference color may be difficult to train a system based on recognition. So, this system converts the RGB image to grayscale image.
- Binarization: Otsu's thresholding technique is used to isolate objects of interest having values different form the background. Images were converted to black and white images. White is represented by 1 and black by 0. The stomach part of the images was represented by 1 and black part of the image by 0. Otsu's thresholding technique is as follows [4]:

$$g(x, y) = \begin{matrix} 1\{f(x, y) > T\} \\ 0\{f(x, y) <= T\} \end{matrix} \quad (1)$$

where, g(x, y) is the output image, f(x, y) is the input image and T is the threshold value.

- Noise Filtering: Adaptive filter is performed on the degraded image that contains original image and noise.

$$\hat{f}(x,y) = g(x,y) - (\delta_n^2 / \delta_l^2)[g(x,y) - m_l] \quad (2)$$

where δ_l^2 is the local variance of the local region, m_l is the local mean, δ_n^2 is the variance of overall noise and g(x, y) is the pixel value at the position (x, y) [5].

4. Feature Extraction

In the feature extraction step, this system can extract different geometrical and first-order (FO) statistic features from the gastric endoscope image [6, 7]. The area (A) in the object is the just the count of the ones in the image array. For computing area, binary image is used.

$$A = n[1] \quad (3)$$

where, n[ ] represents the count of number of the patterns within the parenthesis.

Perimeter is calculated by counting the pixels contained in the boundary. The irregularities are computed by an index:

$$I = \frac{4\pi \times \text{Area}}{(\text{Perimeter})^2} \quad (4)$$

Diameter is computed as follows:

$$\text{Diameter} = \sqrt{\frac{4 \times \text{area}}{\pi}} \quad (5)$$

Variance feature is given by

$$\sigma^2 = \frac{1}{N \times M} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (P(i,j) - \mu)^2 \quad (6)$$

where "μ" is the average gray level, "i" is x coordinate and "j" is y coordinate of the pixel and p(i, j) is a pixel value at that location. Standard deviation feature is given by

$$\text{Std} = \sqrt{\sigma^2} \quad (7)$$

Smoothness feature is given by

$$R = 1 - \frac{1}{1 + (\text{Std})^2} \quad (8)$$

Entropy feature is given by

$$E = \sum_{i}^{M} \sum_{j}^{N} p(i,j) \log p(i,j) \quad (9)$$

5. Support Vector Machine (SVM)

Support vector machine (SVM) is a supervised learning method that is be divided into two categories: linear type and non-linear categories. Among them, this system uses the linear support vector machine.

For the linear problem, "n" is the number of samples, the sample can is x_i and its subordinate classes is y_i. In two dimensional space, their training sets is as: $\{x_i, y_i\}$, i=1,…, n, y_i € {-1, 1}, x_i € R and the discrimination function is as: g(X) = w. x + b. In each formula, "w" is weight vector and "b" is classification threshold value. For w. x + b = 0, if w, b can be amplified or narrow several times at the same time, the classifying face determined by this equation remains constant.

The training samples in the hyperplanes H1 and H2 are the samples that satisfy y_i (w. x + b)-1≥0, i=1,…,n, which are support vector. To improve the classification accuracy, relaxation factor $\xi_i \geq 0$ (i = 1, 2, ..,n) is introduced, at this time constraints condition becomes

$$y_i(w.\ x + b) - 1 + \xi_i \geq 0 \ (i = 1, 2, .., n) \quad (10)$$

Optimal objective function expression becomes

$$\min = 1/2\ (w\ .\ w) + C\ .\ \sum \xi_i \quad (11)$$

In the above expression, "C" is the punishment coefficient. The greater "C" is, the greater is the punishment level of wrong point's samples. Using Lagrange optimization method can put the above optimization problem into its dual problem, the maximum value of the formula is:

$$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \bullet y_j) \quad (12)$$

If "α" is the optimal solution, $w = \sum \alpha_i y_i x_i$. In the above formula, if $\alpha_i = 0$, the sample x_i should be called non-support-vector; if $\alpha_i > 0$, the sample x_i can be called support vector. Solving the above problems later got the discriminate function for:

$$f(x) = \text{sgn}\ \{\sum_{i=1}^{n} a_i y_i (x_i \bullet x) + b\} \quad (13)$$

In the above formula, the summation operates for support vectors. The "sgn()" is the sign function "s" is classification threshold and "α" is the optimal solution [8].

6. Proposed System Design

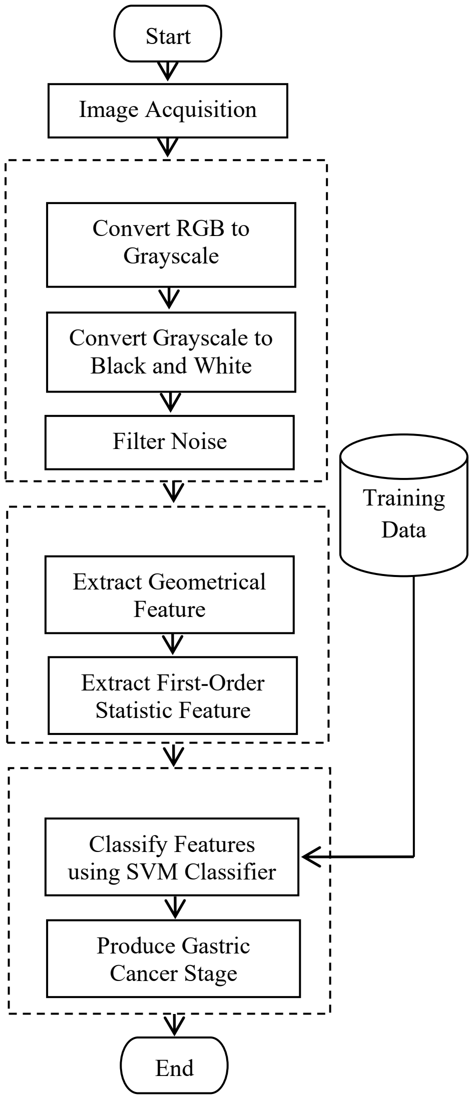


Figure 1. Proposed System Design

Proposed system design is shown in Figure 1. In the proposed gastric cancer classification system, there are three main processes that are pre-processing, feature extraction, and feature classification. Firstly, the input endoscope image is acquired from the patient or doctor. The input image is pre-processed by binarization and noise filtering. This system uses the Otsu's method for binarization. In this system, "1" is for the stomach part of the image and "0" is for the black part of the image.

After finishing binarization process, this system uses the adaptive filter for noise filtering. This filter can reduce the impulse high density noise. In the next step, this system extracts features from the input image. In this feature extraction process, the geometrical and first-order statistic features are extracted. These features include the area, perimeter, irregularities, diameter, variance, standard deviation, smoothness and entropy of the endoscope image.

And then, this system performs the classification process by using support vector machine (SVM) classifier. According to the SVM classifier, this system classifies the geometrical and first-order statistic features of the training data with the testing data from the patient or doctor. After classifying, this system produces the gastric cancer stage that the patient suffers.

6.1. Image Acquisition

Firstly, the patient or doctor must input the stomach endoscope image into the system.

Figure 2. Input Endoscope Image

Input endoscope image is shown in Figure 2. Using this inputted image, this system performs feature extraction and classification.

6.2. Color to Gray Scale Image

The difference color may be difficult to train a system based on recognition. The grayscale image is 0 (black) to 255 (white) each pixel. Grayscale image is shown in Figure 3.

Figure 3. Grayscale Image

6.3. Gray Scale to Black and White Image

All grayscale images were converted to black and white images which is represent by 1 and black by 0. The main part of the image was represented by 1 and white parts of the image by 0. Black and white image is shown in Figure 4.

Figure 4. Black and White Image

6.4. Noise Reduction

The noise may come from the camera or the optical device captures the image. Most scanned image contains noise in form of darker dots and disturbances caused by the scanning process. Adaptive filter method can remove from an image without significantly reducing the sharpness of the image. Noise filtered image is shown in Figure 5.

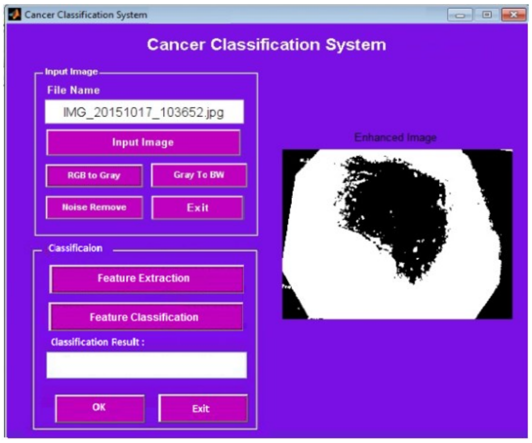
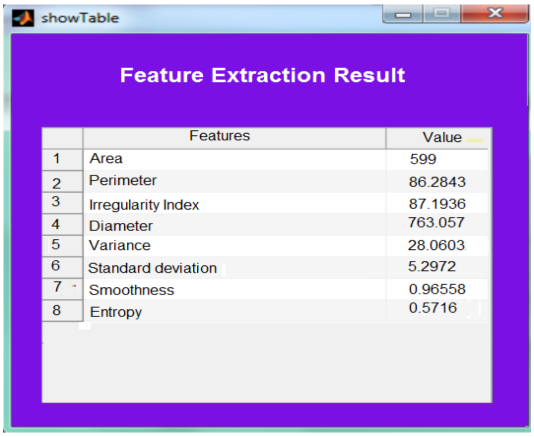


Figure 5. Noise Filtered Image

6.5. Feature Extraction

After finishing pre-processing step, this system extracts features from the stomach endoscope image. Feature extraction results are shown in Figure 6.



	Features	Value
1	Area	599
2	Perimeter	86.2843
3	Irregularity Index	87.1936
4	Diameter	763.057
5	Variance	28.0603
6	Standard deviation	5.2972
7	Smoothness	0.96558
8	Entropy	0.5716

Figure 6. Feature Extraction Results

This system extracts geometrical and first-order (FO) statistic features. Geometrical features include the area, perimeter, diameter and irregularity index. The FO statistic features includes the variance, standard deviation, smoothness and entropy. So, this system extracts eight features from the inputted image.

6.6. Gastric Cancer Classification

By using testing and each training gastric (stomach) features, this system classifies the patient who suffers gastric cancer or not. For classification, this system uses the support vector machine (SVM) classifier. Based on classification result, if the patient suffers gastric cancer, this system classifies the cancer stage. Gastric cancer classification result is shown in Figure 7.

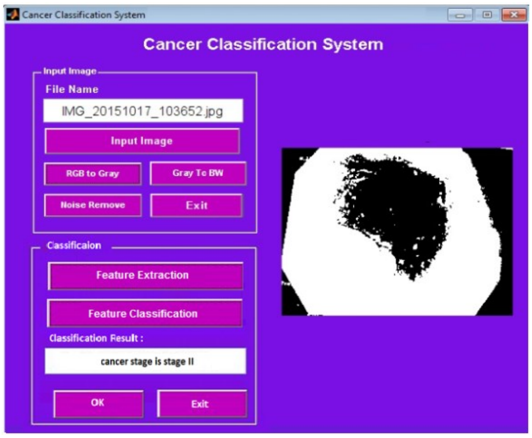


Figure 7. Gastric Cancer Classification Result

7. Experimental Result of the System

This system is implemented by using MATLAB programming language. To show the performance of gastric cancer classification, this system uses total number of 150 gastric cancer endoscope images as training data that are stored in the system database. These training data format is JPEG (Joint Photographic Experts Group) type. This system uses the 2048 pixels in width and 1536 pixels in height.

Precision is the ratio of number of images that are correctly classified images to the number of tested images.

$$\text{Precision} = (\text{Number of Correctly Classified Images}/\ \text{Number of Trained Images}) * 100\% \quad (14)$$

Table 1 and Figure 8 shows the precision for gastric cancer classification system.

Table 1. Precision Results of the System

Number of Trained Endoscope Images	Precision	Error Rate
150	95%	5%
120	92%	8%
100	94%	6%
90	89%	11%

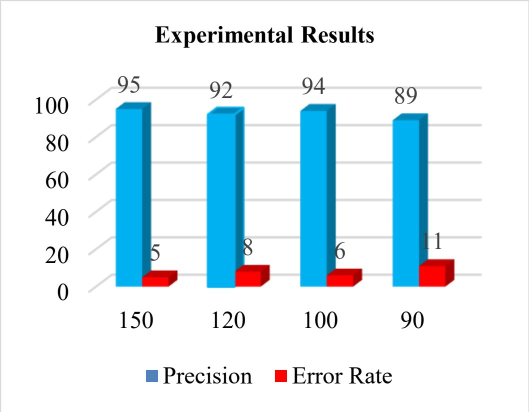


Figure 8. Experimental Result of the System

8. Conclusion

In conclusion, this system provides any physician and any patients by classifying the gastric cancer stage that the patient suffers. According to the technology view, the user can know support vector machine (SVM) classifier is more effective for the gastric cancer classification. So, this system is useful and important for both medical domain and technology domain. Moreover, this system can eliminate time consuming and manpower about cancer classification.

References

[1] G. Chen, Y. Zhang and B. Chen, "Stomach Tumor Localization Method of a Support Vector Machine Based on Capsule Endoscopy", *Progress in Electromagnetics Research B*, vol-78, pp. 125-142, pp. 125-142, 2017.

[2] A. Chanchlani, A. Bhatt and P. Awaghade, "A Survey on Automatic Detection and Classification of Gastric Cancer using Deep Learning", *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 6, no. 11, pp. 8794-8799, 2018.

[3] M. A. Aslam, C. Xue and K. Wang, "SVM Based Classification and Prediction System for Gastric Cancer using Dominant Features of Saliva", *Nano Biomed Eng*, vol. 12, no. 1, pp. 1-13, 2020.

[4] T. Kumar, K. Verma, "A Theory Based on Conversion of RGB image to Gray Image", Computer Science and Engineering Department, Vidya College of Engineering, Meerut, September 2010.

[5] Anonymous: http://angeljohnsy.blogspot.com/2011/12/adaptive-filtering-local-noise-filter.html

[6] N. S. Lingayat and M. R. Tarambale, "A Computer Based Feature Extraction of Lung Nodule in Chest X-Ray Image", *International Journal of Bioscience, Biochemistry and Bioinformatics*, vol. 3, no. 6, pp. 624-629, November, 2013.

[7] S. A. Patil and V. R. Udupi, "Chest X-ray Features Extraction for Lung Cancer Classification", *Journal of Scientific & Industrial Research*, vol. 69, pp. 271-277, April, 2010.

[8] J. Xiuming and W. Jing, "Research of Remote Sensing Classification About Land Survey Based on SVM", *IEEE*, pp. 3230-3233, 2011.