RISE:2016
RESEARCH, INNOVATION AND SCHOLARSHIP EXPO
Category: Computer and Information Sciences
Degree Level: Graduate
Abstract ID# 1385

Northeastern University

Introduction

- Gastric cancer is the fourth most common cancer and the second major cause of cancer causing deaths worldwide.
- Automated diagnostic system for gastric cancer detection can assist physicians in faster and efficient detection of cancer.

The goal is to classify images as cancerous and non-cancerous in endoscopic images. Feature extraction, feature selection and ranking technique is perform to identify the pertinent set of features.These features are used to build and train Binary Classfier.

An accuracy of 87% on training dataset and 89% on testing dataset is achieved used Random Forest Classifier. It improved the overall system efficiency by 11% and reduced the manpower by 40%

Method of Image Analysis

At the most essesntial level, a digital image is represented by a rectangular array of numbers, partitioned into small regions: Pixels. The intensity value at each cell reflects the brightness of image at the corresponding point.

The digital image processing can be divided into fundamental steps. Image preprocessing, segmentation, feature extraction, feature selection and classification are the five stages used for gastric cancer detection.

Image Acquisition & Preprocessing

A total of 100 gastric endoscopic images were selected for the purpose of carrying out this study. These encorporated 70 instances of cancerous and 30 instances of non-cencerous conditions. The selected images contained endoscopic images of both male and female subjects. During the acquisition, the administrator inspected the entire stomach region from different orientation and saved a single frame.

The image pre-processing techniques that are performed on images can be named as following: cropping, conversion of RBG image to gray-scale, edge detection, region-based segmentation and background subtraction

A. Cropping

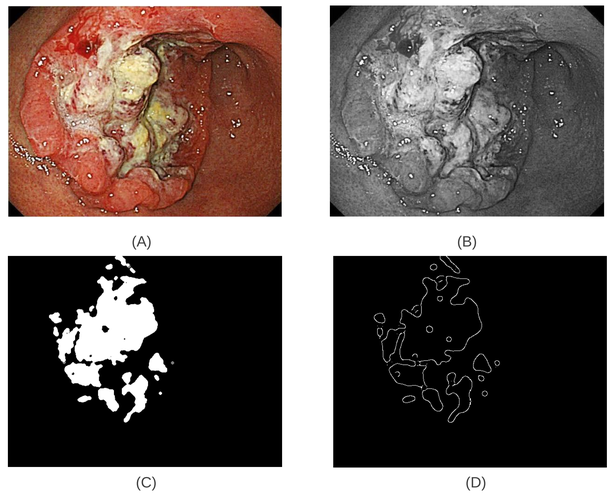
Cropping is an operation performed on all the acquired images to give more prominence to the region of interest (i.e., the ulcerous region) and to uproot all the undesirable artefacts such as written labels and noise in the background from these images in order to improve the framing of the image.

B. Edge-Detection

The shape as well as structure of the stomach is pyramidal; hence edge detection operation is necessary to define the edges of the stomach.

C. Region-based Segmentation

Region Growing is a method that gathers pixels or sub regions into bigger areas. It is otherwise called pixel-based image segmentation method since it includes determination of introductory seed points. The areas are then developed from these seed point to contiguous point relying upon a region membership criterion.



(A) Original Image. (B) Image on grey scale. (C) Region-based Segmentation (D) Edge-Detection of Region of Interest.

D. Background Subtraction

The process of getting rid of the intensity values present outside the region of interest.

Image Filtering

The Median filter is a spatial filter which replaces the value of the central pixel with the median of the intensity values in the neighbourhood of that pixel. Median filters are generally used to remove ”salt and pepper” noise.

The Laplacian of an image is often used for edge detection as it highlights the region of rapid change in intensity. The Laplacian applied to an image that is first smoothed with a Guassian smoothening filter so as to reduce its sensitivity to noise.

Image Processing Techniques

Texture is used to inspect important characteristics of an image for surface and object identification. The spatial distribution of gray levels in a neighbourhood in an image characterizes the texture of an image. The resolution of the image determines the perception of texture. Texture can be defined as repeating patterns of local variations of intensity of image which are too fine to be distinguished as separate objects for this particular resolution.

In image processing technique, we first identify the sets of essential features and then extract those features from the endoscopic image for further processing. “Feature extraction” and “Feature selection” are two most vital steps of image processing.

A. Feature Extraction

Feature extraction makes use of several algorithms and techniques to detect and isolate prominent features from an image based on the pixel intensity to find out final result and determine if it is normal or abnormal. A set of four features (i.e., statistical texture features), namely, intensity histogram, gray-level co-occurrence matrix (GLCM), graylevel run-length matrix (GLRLM), and invariant moments, were extracted from each of the total 100 images in MATLAB. Each feature set comprises individual image parameters.

B. Feature Selection

Feature selection is used to extract those features from the selected feature set which best describes the characteristics of the image. A total of 46 features were extracted in the feature extraction process from each image. But since the number of features is high, we cannot supply all of these features to the classification algorithm. Therefore, only those features, which are very significant in classifying and identification of the disease conditions, were selected.

Feature Category	Number of features before selection	Number of features after selection
Intensity Histogram	6	4/6
GLCM	22	7/22
GLRLM	11	4/11
Invariant Moments	7	0/7
Mixed Features	46	15/46

Classification

Classification is considered an instance of supervised learning, i.e. learning where a training set of correctly identified observations is available. In this project, we employed four classification techniques, namely, Nave Bayes’ classification, Support Vector Machine, Bagging and Random Forest, to investigate the most suited and accurate classification algorithm for our dataset for the diagnosis of gastric cancer.

Conclusion

In this study, five feature classifiers have been investigated for diagnosing the Cancer conditions. The accuracy of the classifier was based upon the feature set used, selected training 10 samples, and the classifiers ability to learn from the training samples. From the above results, we have achieved our objective in finding the best classifier for cancer diagnosis. Five sets of features such as GLCM, intensity histogram, GLRLM, invariant moments, and mixed features were extracted. These features were then selected and trained in Weka to determine the best set of features, which can determine the presence of ulcer conditions in the stomach. A comparative approach revealed that both GLRLM and mixed feature set showed excellent accuracy in training as well as testing when classified using the Random Forest classifier.