

Rule Generation for Gallbladder Cancer Prediction Using Decision Tree Classifier

S. Sharmila and D. Dharunya Santhosh

Abstract--- *Cancer is a disease, where cells grow out of control, divide and attack other tissues of the patient. There are many forms of cancer, among which the “Gallbladder Cancer” is relatively uncommon. This cancer can be permanently cured only by removing the Gallbladder. Else, the patient must be treated with high dosage of medicine for the survival, which in turn cures the disease but reduces the lifetime of the patient. Instead, if the gall bladder cancer is predicted at the earlier stage, it can be cured by treating with less dosage of medicines. In the proposed methodology, earlier detection of gall bladder cancer is done through imaging, which contains three major steps as, Pre-processing, Feature Extraction and Feature Classification. The noise in the image can be removed using Pre-processing step. The essential features for the cancer prediction can be extracted from the segmented image. Further, using decision tree algorithm, gallbladder cancer prediction is done.*

Keywords--- *Pre-processing, Segmentation, Feature Extraction, Classification, Decision Tree.*

I. INTRODUCTION

CANCER is a group of disease including abnormal cell development with the possibility to attack or spread to different parts of the body. These appear differently in relation to benign tumors, which don't spread to different parts of the body. Possible signs and effects include a lump, irregular bleeding, prolonged cough, unexplained weight reduction. While these symptoms may demonstrate cancer, they may have different causes. More than 100 sorts of cancer affect humans.

Tobacco utilize is the reason for around 22% of cancer deaths. Another 10% are because of stoutness, poor diet, absence of physical activity, and excessive consumption of liquor. Other elements include certain infections, exposure to ionizing radiation and environmental pollutants. Ordinarily numerous hereditary changes are required before cancer develops. Around 5– 10% of cancer are because of acquired hereditary imperfections from a person's parents. Cancer can be recognized by specific signs and symptoms or screening tests.

Gallbladder cancer is a generally uncommon. It has peculiar geographical distribution being common in central and South America, central and eastern Europe, Japan and

northern India; it is also common in certain ethnic groups e.g. Native American Indians and Hispanics. If it is diagnosed early enough, it can be cured by removing the gallbladder, some portion of the liver and associated lymph nodes. Regularly it is found after side effects, for example, abdominal pain, jaundice and vomiting occur, and it has spread to other organs, for example, the liver.

It is uncommon type of cancer that is believed to be identified with constructive development of gallstones, which additionally can lead to calcification of the gallbladder, a condition known as porcelain gallbladder. Porcelain gallbladder is also uncommon. A few studies show that people with porcelain gallbladder have a high danger of developing gallbladder cancer, however different studies question this. The outlook is poor for recovery if the cancer is found after symptoms have begun to happen, with a 5-year survival rate near 3%. The most widely recognized and best treatment is surgical removal of the gallbladder (cholecystectomy) with part of liver and lymph node dissection. However, with gallbladder cancer's extremely poor prognosis, most patients will die within a year of surgery.

If surgery is not possible, endoscopic stenting of the biliary tree can reduce jaundice and a stent in stomach may relieve vomiting. Chemotherapy and radiation may also be used with surgery. If gall bladder cancer is diagnosed after cholecystectomy for stone disease (incidental cancer), reoperation to remove part of liver and lymph nodes is required in most cases. When it is done as early as possible, patients have the best chance of long-term survival and even complete cure.

The remainder of the paper is organized as follows. In section II, related works will be discussed. The proposed work is viewed in section III, experimental results are represented in section IV, conclusion and future work is mentioned in section V. Finally, ended with references in section VI.

II. RELATED WORK

In [1] MRI images were taken into account to classify the presence of tumour cells using the image processing techniques. Initially the test brain images were pre-processed and the external disturbances say noise were removed. The pre-processed brain images were segmented, such that the affected and unaffected part were segregated. The important features say energy, local variation and correlation were extracted. These processes were considered under training phase. Using the extracted features and the concept of decision tree classifier the segmented images were classified into tumorous and non-tumorous.

S. Sharmila, Sri Krishna College of Engineering and Technology, Coimbatore, India.

D. Dharunya Santhosh, Sri Krishna College of Engineering and Technology, Coimbatore, India.

DOI:10.9756/BIJDM.8354

In [2], a bio-metric system using iris recognition was considered in-order to increase its efficiency in terms of security. The main methodology behind iris recognition for ensuring security is that, the sample iris images were taken and processed to get segmented image. From the segmented images knowledge based features were extracted which in turn helped them to classify iris using decision tree classifier accordingly in 1.8 seconds. They provided a technological solution for enabling high security everywhere.

The ensemble learning is the main concern in [3], where the noisy data were classified with new technique which inculcated boosting and decision tree classifier together. AdaBoost algorithm is used for boosting here. For every instances separate decision tree algorithm is applied and those results were recorded. Finally, the obtained instances were classified using voting method. Thus noisy data were classified by combining both boosting from ensemble learning and decision tree classifier to obtain accurate results.

According to [4], the prediction of precipitation is done using decision tree classifier. The prediction of precipitation is necessary tool for meteorology. Though many features were involved in predicting precipitation, they used decision tree classifier for classification. Approaches such as back propagation of neural network, Bayesian network, Support vector machines, linear regression and fuzzy logic can also be applied for predicting precipitation. But, their rate of prediction is low which may lead to unpredictable errors.

Many forms of gas classification is made possible with the help of other data mining algorithms. As their results were not that much efficient in[5] they used binary decision tree algorithm for effective classification. The gas concentration is recorded using array of seven metal oxide gas sensor for five different gases. Their proposed system accurately classified gases with respect to different concentrations.

III. PROPOSED SYSTEM

The proposed methodology is used to identify or predict Gallbladder cancer in its initial stage thus reducing complexity of life span. It contains three main stages say 1) Preprocessing-The process of removing external noise from diagnostic images. 2) Segmentation-The process of segmenting the affected and unaffected part from an image. 3) Feature Extraction- The process of extracting necessary features in order to classify whether a person is having cancer or not. 4) Feature Classification- The process of Classifying using Decision Tree Classifier to predict the discrete output either having cancer or not.

A. Pre-Processing

Whenever an electronic device processes some data, it will be always associated with some external disturbance in terms of noise. In proposed methodology, the CT scan images were taken as input test data which will also contain noise in gaussian distribution over the image. If gaussian noise is removed using gaussian filter, the edges of parts will get blurred. In-order to avoid this, Weiner filter is used for removal of gaussian noise.

B. Segmentation

The main goal of applying segmentation over any image is to segregate meaningful parts from it. The tomography images when segmented, resulted in visually classifiable images which has shown affected and unaffected parts separately. After segmenting main features were extracted using feature extraction process.

C. Feature extraction

At this stage various statistical features were available for feature classification. The main challenge is to select the features which contribute more for the classification of tumorous images. In proposed methodology, the main features extracted were Energy, Correlation, Local variations, Entropy and Homogeneity.

Energy (k1) which expresses the repetition of pixel pairs of an image,

$$k1 = \sum_{i=0}^{N-1} \sum_{j=0}^{k-1} p_{\mu}^2(i, j) \quad (1)$$

Local Variation (k2) which present in the image is calculated by Contrast. If the contrast value is high means the image has large variations.

$$k2 = \sum_{i=0}^{N-1} t^2 \left\{ \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{\mu}(i, j) \right\} \quad (2)$$

Correlation (k3) is a measure linear dependency of gray level values in co-occurrence matrices. It is a two dimensional frequency histogram in which individual pixel pairs are assigned to each other on the basis of a specific, predefined displacement vector.

$$k3 = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \frac{(i, j)p(i, j) - \mu_1\mu_2}{\sigma_1^2\sigma_2^2} \quad (3)$$

Entropy (k4) is a measure of non-uniformity in the image based on the probability of Co-occurrence values, it also indicates the complexity of the image.

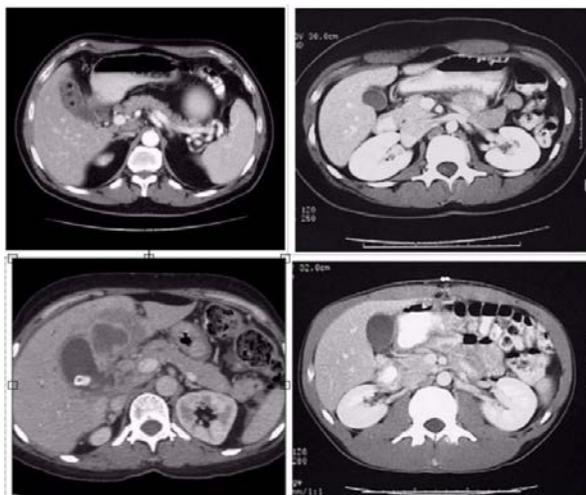
$$k4 = - \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} p_{\mu}(i, j) \log(p_{\mu}(i, j)) \quad (4)$$

Homogeneity (k5) is inversely proportional to contrast at constant energy whereas it is inversely proportional to energy.

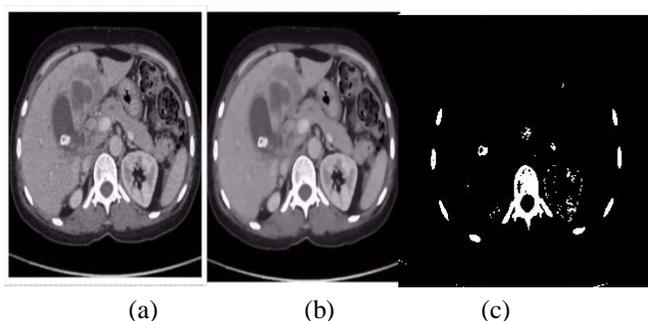
$$k5 = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \frac{P_{\mu}(i, j)}{1 + (i - j)^2}, i \neq j \quad (5)$$

IV. EXPERIMENTAL RESULTS

The proposed methodology was implemented in MATLAB platform. The CT scan images were taken from internet sources. The data source which includes both tumorous and non-tumorous computerised tomography images are as follows:



The comparison between image before denoising, after denoising and segmented image is as follows:



Where

- (a) - normal image before pre-processing.
- (b) - image after pre-processing.
- (c) - final image after segmentation.

V. CONCLUSION AND FUTURE WORK

Thus an effective method was proposed which in-turn is used to predict Gallbladder cancer in its starting stage using Decision tree classifier. As the proper imaging will be done and diagnosis will be done in its starting stage there is no need for high dosage of medicine, also it reduces the risk factors of patient which will be experienced by them after removal of Gallbladder. In future, the same proposed methodology is planned to get implemented using different data mining algorithms for more accurate results.

REFERENCES

- [1] G.K. Sundararaj and V. Balamurugan, "An expert system based on texture features and decision tree classifier for diagnosis of tumor in brain MR images", IEEE International Conference on Contemporary Computing and Informatics, Pp. 1340-1344, 2014.
- [2] V.S. Tallapragada and E.G. Rajan, "Morphology based non ideal iris recognition using decision tree classifier", IEEE International Conference on Pervasive Computing, Pp. 1-4, 2015.
- [3] D.M. Farid, G.M. Maruf and C.M. Rahman, "A new approach of Boosting using decision tree classifier for classifying noisy data", IEEE International Conference on Informatics, Electronics & Vision, Pp. 1-4, 2013.
- [4] N.P. Lakkakula, M.M. Naidu and K.K. Reddy, "An Entropy Based Elegant Decision Tree Classifier to Predict Precipitation", IEEE Symposium on European Modelling, Pp. 11-19, 2014.
- [5] Hassan, M and Bermak, A, "Gas classification using binary decision tree classifier", IEEE International Symposium on Circuits and Systems, Pp. 2579-2582, 2014.
- [6] S.R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology", IEEE transactions on systems, man, and cybernetics, Vol. 21, No. 3, Pp. 660-674, 1991.
- [7] K. Polat and S. Güneş, "Classification of epileptic form EEG using a hybrid system based on decision tree classifier and fast Fourier transform", Applied Mathematics and Computation, Vol. 187, No. 2, Pp. 1017-1026, 2007.
- [8] G. Stein, B. Chen, A.S. Wu and K.A. Hua, "Decision tree classifier for network intrusion detection with GA-based feature selection", 43rd Annual Conference on Proceedings of the Southeast regional, Pp. 136-141, 2005.