ENHANCED APPROACH TO DETECT LUNG CANCER USING GENETIC WATERSHED ALGORITHM

Seema Devi[1],Mrs.Gurjeet Kaur[2]

M.tech Research schollar[1], Assistant professor Department of CSE-SSCET Badhani[2]

Abstract: In today's world cancer is most common disease that causes death and lung cancer is one of them. lung cancer detection at early stages will increase the survival rate of the patient and also provide more option for treatment. in recent years if detection of cancer at early stages is done than the survival rate of patient is increased from 10 to 40%. the lung cancer is most common cause of death and dangerous disease, so faster method for detection required. the abnormal growth of cells present within the lungs can be detected using image processing technique. image processing technique classifies the mri images to distinguish the affected areas. in this paper a hybrid algorithm that utilizes watershed and genetic algorithm is utilized to detect lung cancer. for classification the decision tree is used that will enhance the detection rate. it will also increase the linear accuracy and recognition rate.

keyword: lung cancer, watershed algorithm, genetic algorithm.

INTRODUCTION

The main cause of cancer death is lung cancer and detection of cancer at early stage is difficult that cause high mortality rate. lung cancer is anomalous and unrestrained explosions of cells. the lung made of neurons cells; these cells are liable for lung usual activities. the lung generally creates new cells only when they are necessary to interchange old or smashed ones. most cells restore themselves by distributing to make more cells. usually, this turnover takes place in an organized and precise manner. if, for some reason, the development becomes out of control, the cells will endure to divide, evolving into an inflammation, which is called a tumour. lung cancer can be defined as an unexpected evolution of cells inside the lung or the skull, which can either be cancerous or non-cancerous. lung cancer can grow at any age, but are most collective in teenagers between the ages of 3-12, and in adults aged 55-65. the main goal of the medical researcher in today's world is curing cancer but this take time as well as need money. the researcher finds the causes of various cancer and also develop a technique to that cure them. the lung cancer is one of them that are main centre of attraction for research. it have chances to grow and spread all around the world. in world the number of people died due to lung cancer is high as compare to any other cancer like breast, skin and colon. so the early detection of lung cancer will decrease rate of deaths. recently according to who around 6.8 million deaths caused due to lung cancer worldwide and also it continue to rise.[1] increase of abnormal cells of lung is caused lung cancer. the tumour will increase the cells and form new one, also abnormal one. this will be detect at early stage so that proper treatment will provided on time, but many will notice this disease when it is too late and it is impossible to surgery. the detection of lung cancer is important at early stage is important for better treatment. [2]the mri images are used for diagnosis and the tumour starts at lungs part then it is called primary lung cancer. there are two types of lung cancer

- small cell cancer
- non-small cell cancer

lung cancer medical procedure is performed by an extraordinarily arranged thoracic expert. in the wake of emptying the tumor and the including edge of tissue, the edge is moreover considered to check whether cancer cells are accessible. if no cancer is found in the tissue incorporating the tumor, it is seen as a "negative edge." a "positive edge" may require the authority to oust a more prominent measure of the lung tissue.

lung cancer medical procedure can be recuperating or palliative. recuperating medical procedure intends to fix a patient with beginning period lung cancer by ousting most of the cancerous tissue. palliative medical procedure intends to empty an obstacle or open an avionics course, making the patient more pleasing yet not by any means removing the cancer. [3] in our work we focused on detecting tumour and its stages using hybrid algorithm that contains watershed and genetic algorithm.

LITERATURE SURVEY

In **m. e. s. processing et al.** [4]proposed a region growing algorithm that segment the ct scan images of lung. it starts with seed pixel and then continues on checking other pixels of its neighbourhood. **i. histogram, e. technique, f. handling, and p. nois et al.**[5] it utilizes a criterion for similarity index and it is similar then it include region. then this region is examined further.
In **f. taher, n. werghi, h. al-ahmad, and r. sammouda et al.** [6]proposed algorithm to detect cancer cells from lung mri image and it minimize the error rate in detection. it utilizes sobel edge for detection and matrix for label the edges. it uses gradient to find the sobel edges. the change in intensity of image is given by image gradient.

b. abdillah, a. bustamam, and d. sarwinda et al. [8]suggests an approach the use cad (computer aided design) to detect the lung cancer by using edges of mri images. **i. engineering et al.**[9]proposes threshold algorithm that detect sputum cell from the image. in **c. engineering et al**.[10] proposes water shed transformation algorithm that segment the images. in this to process the image morphological operations are used and eliminate the over segmented area. after that gradient is reconstruct and the shape of image gradient is maintained. **p. rao, n. a. pereira, and r. srinivasan, et al.** [11] in this paper we mainly focused on detecting tumour affected regions in image and also evaluate accuracy of the system. **azizi et al.** [12] proposed a mechanism to determine breast cancer by feature extraction iterative approach is followed along with support vector machine. supervised learning mechanism produce best possible result. **r. rani and s. gupta et al**.[13] in case of complex images, time consumption in generating the result is high. also nominal values cannot be tackled through the existing literature. gaussian kernel method is associated with the existing literature. the modification to this kernel function yield best possible result. improvement in terms of classification accuracy and recognition rate is still required.

lie et al. [15] proposed a method in which the l-fold down sampling is used for reducing data rates and coefficient of filters that conduct compression with the complex data. so it decreases the overall computational complexity. the simulation result will show that the proposed method produced better axial resolution as compared to the existing methodology. **a. verma and v. singh et al.**[16] the ser is >36 and the performance of the overall system is better.

wu et al. [17]demonstrated that a high-arrange bunched differential heartbeat code regulation technique with expulsion of nearby ghostly anomalies (c-dpcm-rlso) is proposed for the lossless pressure of hyperspectral pictures. by adaptively expelling the nearby phantom anomalies, the c-dpcm-rlso strategy enhances the forecast exactness of the high-arrange relapse indicator and decreases the residuals between the anticipated and the first pictures. **r. jhawar, v. piuri, and m. santambrogio et al.**[18] the analysis on an arrangement of the nasa airborne visible infrared imaging spectrometer (aviris) test pictures demonstrate that the c-dpcm-rlso strategy has a practically identical normal pressure pick up however a much lessened execution time as contrasted and the past lossless techniques.

islam et al. [20] found a prosperous strategy to recognize vehicle number plates. the proposed method is based on morphological activities in view of various organizing components with a specific end goal to maximally bar non-intrigued locale and enhance protest zone. **c. science and t. rourkela et al.** [21]this framework has been experienced utilizing a database of number plates and recreated comes about show significant changes when contrasted with other customary frameworks. the achievement rate of the proposed technique is around 92% with differing light conditions.

PROPOSED METHODOLOGY

The image set is required to be operated upon by the modified watershed algorithm with classification. feature extraction is performed by the use of genetic algorithm. rules are implemented using decision tree for classification purpose. the detailed description of proposed system is listed as under

Dataset Description

The dataset which is used is derived from internet. lung cancer dataset is used for extraction. lung cancer dataset contains 10 images of 40 to 44 kb in size. the images are grey scale in nature. dimension of images is 1024×1024.

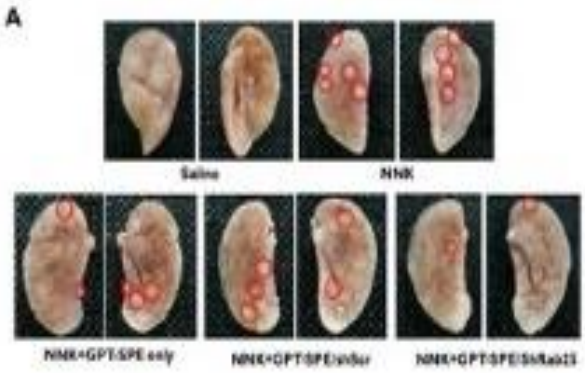

figure 1:dataset

lung cancer dataset contains group of images with and without cancerous cell. through the proposed system classification is generated to detect the disease and compared against the original image set.

Applying watershed algorithm for segmentation

Watershed is a transformation commonly applied on grey scale images. during the flooding operation adjacent catchment are constructed. the flooding process is performed on gradient images. the basins are constructed near the edges. it leads to over segmentation of images especially on noisy images. the algorithm is as follows

- set of markers from where flooding should begin are selected.
- priority is assigned to the neighbourhood of each marked pixel on the basis of gradient magnitude value selected. label is assigned to each pixel inserted into priority queue.
- the pixel with least priority is extracted from priority queue and labels are checked.
- if labels are assigned and are same then same label pixels accept one are removed.
- the other pixels which are unlabelled are pushed into priority queue and steps are performed again.

after performing watershed algorithm, result is as follows

figure2: after performing watershed algorithm

The watershed algorithm is followed by genetic algorithm for feature extraction.

Genetic algorithm

Genetic algorithm is multiheuristic algorithm having multiple objectives associated with it. genetic algorithm is associated with different phases. the pixels correspond to chromosomes. to perform feature extraction, selection operation takes place. the proposed system uses random selection operation. the fitness function evaluation is used to generate next population for feature extraction. the extracted features are compared against the threshold value. the threshold value is assumed to be base value above which optimality is achieved.

$$Threshold = Optimal_{base-value}$$

the fitness function evaluation takes place in order to obtain optimal results for classification.

$$F(t) = Pixel_i(Features_j > Threshold)$$

the mutation and crossover is performed only if threshold value is invalidated. mutation and crossover is accomplished by identifying pixels having intensity values lower than desired levels. after which selection operation is performed again.

Decision tree implementation

Decision tree is used for classification purpose. classification of results required certain rules to be created. the training rules correspond to features which are identified through genetic algorithm. the disease is detected if rules are violated. decision tree rules are listed as under

if gm>0.5

if std>0.5

if kurtosis>0.5

if moment>0.5

if mean>0.5

disease detected

else

disease not detected

Membership rules.

Maximum value 1 and minimum value 0

a feature extracted from the segmented image is compared against these rules. membership is decided only if feature lies within the range of 0 and 1. in order to detect the disease membership should be greater than 0.5.

Performance analysis and results

The performance of proposed system is analysed in terms of recognition rate and sigmoid function. sigmoid function predicts the activation of training images. training images matches against the testing images. threshold must be satisfied in order to determine whether given image lie within the desired levels of features. if they do then they are classified as having given disease.

Recognition rate

Recognition rate indicates the success of simulation in the detection of problems within the given image set. for the success of proposed system recognition rate must be high.

Sigmoid function

A sigmoid function is a mathematical function having a characteristic "s"-shaped curve or **sigmoid curve**. often, *sigmoid function* refers to the special case of the logistic function shown in the figure and defined by the formula

$$S = \frac{1}{1+e^{-x}}$$

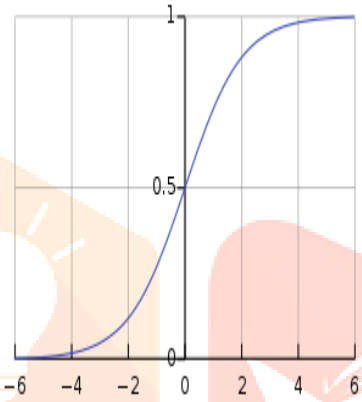

figure3: showing sigmoid function

other examples of similar shapes include the gompertz curve (used in modelling systems that saturate at large values of x) and the ogee curve (used in the spillway of some dams). sigmoid functions have domain of all real numbers, with return value monotonically increasing most often from 0 to 1 or alternatively from −1 to 1, depending on convention.

A wide variety of sigmoid functions have been used as the activation function of artificial neurons, including the logistic and hyperbolic tangent functions. sigmoid curves are also common in statistics as cumulative distribution functions (which go from 0 to 1), such as the integrals of the logistic distribution, the normal distribution, and student's *t* probability density functions.

Snapshots

Snapshots corresponding to proposed system are as under

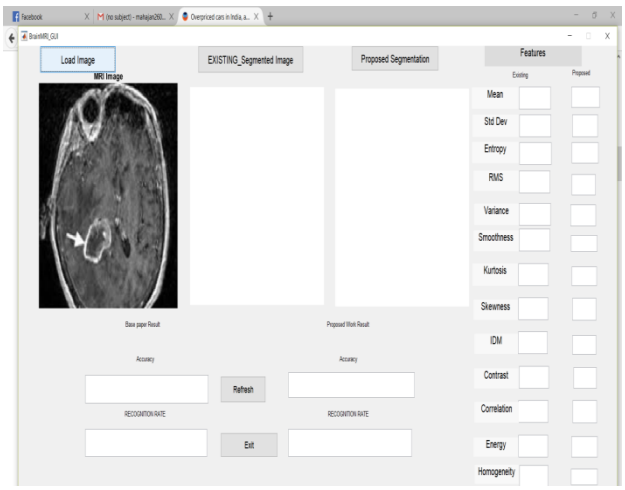

figure4: snapshot corresponding to load image

This screen demonstrates the selection of image when user clicks on the load button. the next screen demonstrates the existing system segmentation.

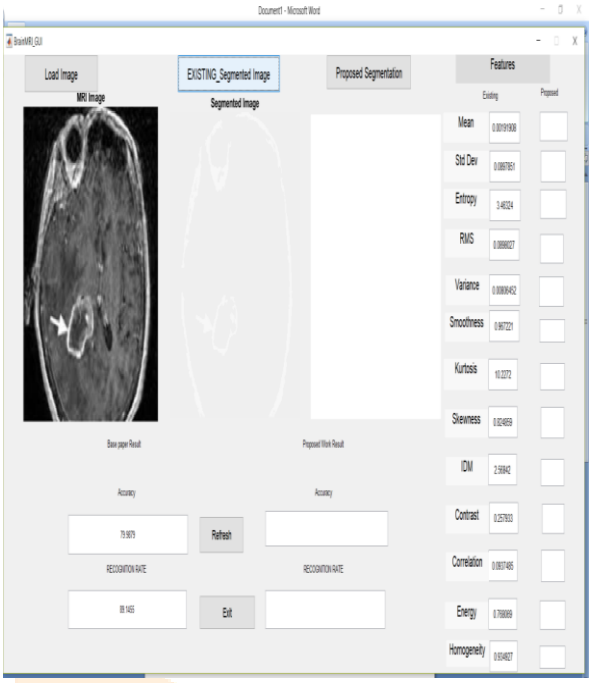

figure5: snapshot corresponding to existing segmentation image

In this case non small cell lung cancer is detected. in other words classification will be of benign cancer

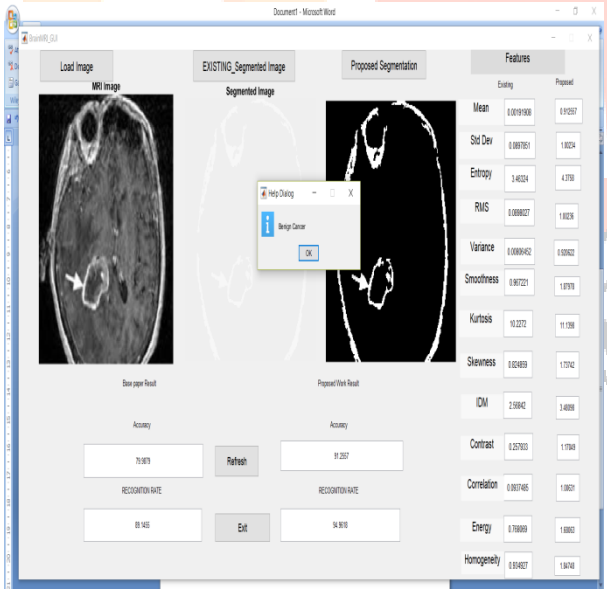

figure6: snapshot corresponding to proposed segmentation image

In this case classification result gives lung cancer. the accuracy of proposed system along with recognition rate is high as compared to existing techniques.

table 1: recognition rate

images	Result	threshold=0.4
image_name	Existing	proposed
lung1	79	91
lung2	81	92
lung3	79.879	90.23323
lung4	85	95.878
lung5	83	93.4343

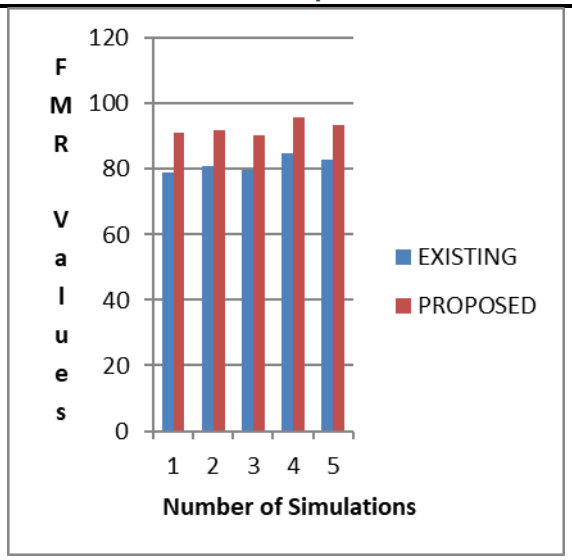

figure7: false matching rate (fmr)

table 2: sigmoid function

images	without GA	with GA and watershed
lung1	89	94
lung2	89.433	94.4334
lung3	90.2323	95.343
lung4	90.2324	95.787
lung5	91.1223	96.233434

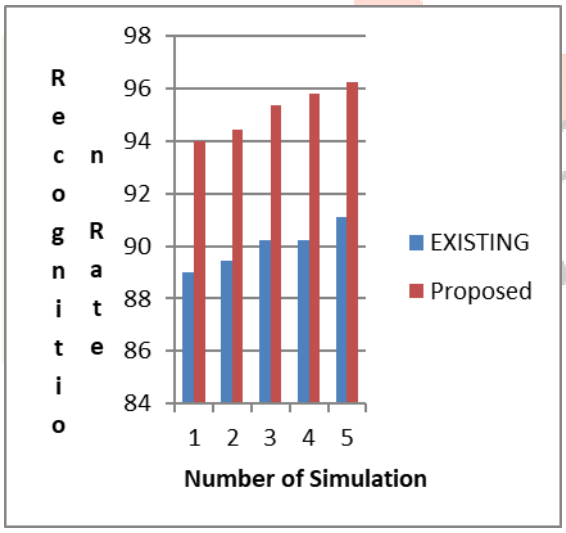

figure8: sigmoid function

Where 1,2,3,4,5 are the images of lung cancer.

CONCLUSION AND FUTURE SCOPE

The proposed scheme of thing produces efficient mechanism in order to determine the accuracy of disease detection using hybrid approach of watershed and fsvm. analysis process detects disease efficiently and handles the complex image with least complexity.

In future, detection process can be accomplished using lbp with genetic algorithm.

REFERENCES

[1] "selected automobile companies and its," no. 10, 2014.

[2] n. kureshi, s. s. r. abidi, and c. blouin, "a predictive model for personalized therapeutic interventions in non-small cell lung cancer," *ieee j. biomed. heal. informatics*, vol. 20, no. 1, pp. 424–431, 2016.

[3] i. d. t, b. goossens, and w. philips, "mri segmentation of the human brain : challenges , methods , and applications," vol. 2015, 2015.

[4] m. e. s. processing, "detection of lung cancer using marker-controlled watershed transform," vol. 0, no. c, 2015.

[5] i. histogram, e. technique, f. handling, and p. noise, "and pepper noise," vol. 6, no. may, pp. 17–22, 2016.

[6] f. taher, n. werghi, h. al-ahmad, and r. sammouda, "lung cancer detection by using artificial neural network and fuzzy clustering methods," vol. 2, no. 3, pp. 136–142, 2012.

[7] a. analysis, o. f. dengue, and u. fuzzy, "i nternational j ournalof e ngineering s ciences & m anagement," vol. 6, no. june, pp. 41–47, 2016.

[8] b. abdillah, a. bustamam, and d. sarwinda, "image processing based detection of lung cancer on ct scan images image processing based detection of lung cancer on ct scan images," 2017.

[9] i. engineering, "lung cancer recognition in ct image using," pp. 3405–3409, 2016.

[10] c. engineering, "region of interest based medical image compression," pp. 1–12, 2014.

[11] p. rao, n. a. pereira, and r. srinivasan, "convolutional neural networks for lung cancer screening in computed tomography (ct) scans," pp. 489–493, 2016.

[12] w. ghwhfwlrq *et al.*, "i pdjh s hjphqwdwlrq iru â€tm hwhfwlrq ri b hqljq dqg 0doljqdqw 7xpruv," no. 1, pp. 51–54, 2016.

[13] r. rani and s. gupta, "lung tumour detection by applying watershed method," vol. 13, no. 5, pp. 955–964, 2017.

[14] p. yuvarani, "image denoising and enhancement for lung cancer detection using soft computing technique," pp. 27–30, 2012.

[15] d. lie, k. chae, and s. mukhopadhyay, "analysis of the performance, power, and noise characteristics of a cmos image sensor with 3-d integrated image compression unit," *components, packag. manuf. technol. ieee trans.*, vol. 4, no. 2, pp. 198–208, 2014.

[16] a. verma and v. singh, "lung tumor detection using watershed algorithm," pp. 406–417.

[17] j. wu, w. kong, j. mielikainen, and b. huang, "lossless compression of hyperspectral imagery via clustered differential pulse code modulation with removal of local spectral outliers," vol. 9908, no. c, pp. 1–10, 2015.

[18] r. jhawar, v. piuri, and m. santambrogio, "a comprehensive conceptual system-level approach to fault tolerance in cloud computing," pp. 0–4, 2012.

[19] v. t. u. pg-center, "an extensive research survey on brain mri enhancement , segmentation and classification," 2017.

[20] r. islam, k. f. sharif, and s. biswas, "automatic vehicle number plate recognition using structured elements," no. december, pp. 18–20, 2015.

[21] c. science and t. rourkela, "pattern clustering using soft-computing approaches mohit agarwal pattern clustering using soft-computing approaches."

[22] m. zortea, s. o. skrøvseth, t. r. schopf, h. m. kirchesch, and f. godtliebsen, "automatic segmentation of dermoscopic images by iterative classification," vol. 2011, 2011.