

Development of algorithm for identification of malignant growth in cancer using artificial neural network

R. Pandian¹, D. N. S. Ravi Kumar², R. Raja Kumar³

^{1,2}School of Electrical and Electronics Engineering, Sathyabama Institute of Science and Technology, India

³School of Science and Humanities Sathyabama Institute of Science and Technology, India

Article Info

Article history:

Received Mar 23, 2020

Revised May 6, 2020

Accepted May 15, 2020

Keywords:

Classification

CT image

GLCM

Texture

ABSTRACT

The precise identification and characterization of small pulmonary nodules at low-dose CT is a necessary requirement for the completion of valuable lung cancer screening. It is compulsory to develop some automated tool, in order to detect pulmonary nodules at low dose ct at the beginning stage itself. The various algorithms had been proposed earlier by many researchers within the past, but the accuracy of prediction is usually a challenging task. During this work, a man-made neural network based methodology is proposed to seek out the irregular growth of lung tissues. Higher probability of detection is taken as a goal to urge an automatic tool, with great accuracy. The best feature sets derived from Haralick Gray level co occurrence Matrix and used because the dimension reduction way for feeding neural network. During this work, a binary Binary classifier neural network has been proposed to spot the traditional images out of all the images. The potential of the proposed neural network has been quantitatively computed using confusion matrix and located in terms of accuracy.

Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Pandian R,
School of Electrical and Electronics Engineering,
Sathyabama Institute of Science and Technology,
Chennai, Tamil Nadu, India.
Email: rpandianme@rediffmail.com

1. INTRODUCTION

It is the way toward procuring more elevated level data of a picture, for example, shading, shape and Texture. It is one of the significant qualities utilized in recognizing items or Region of enthusiasm for a image. Surface, the example of data or exhibit of the structure set up in a image, is a huge component of many images types [1]. From a general perspective, surface alludes to surface attributes and presence of Texture highlights can be removed in a few strategies, utilizing factual, auxiliary, and model-based and change data, in which the most well-known way is utilizing the gray level co event matrix (GLCM). GLCM contains the second-request factual data of spatial relationship of pixels of a image [2].

Generally, the exactness of any arrangement framework mostly relies upon the best possible decision of the highlights. Thus it is fundamental to locate a decent arrangement of highlights. Essential period of structure in any grouping framework is the choice of a decent arrangement of highlights which have the capacity of sign detachment in the component space. An order calculation will consistently give an outcome, yet a poor element portrayal will prompt an outcome that doesn't mirror the genuine idea of the hidden information. The information ought to be streamlined without loss of data. Finding the best highlights is a troublesome assignment, and it frequently must be practiced through an experimentation procedure. In this work, a dim level co-event grid (GLCM) is utilized, which is a measurable technique that uses the spatial relationship of pixels.

Davis et al., [3] started the use of dim level co-event grid (GLCM) so as to discover the highlights that are to be produced dependent on a pixel's neighborhood. Davis et al., proceeded with crafted by in such a method of the directional dispersion of GLCM includes and proposed a lot of polar gram measurements which are rotationally invariant. Haralick et al., [4] recommended that revolution invariant highlights could be acquired from co-event networks by taking the normal and scope of each element type over the four points that is utilized. The dark level distinction insights is another surface depiction technique, which is firmly identified with GLCM, Weszka et al., [5]. A co-event grid, additionally alluded to as co-event appropriation, is characterized over a picture to be the dissemination of co-happening esteems at a given balance. It speaks to the separation and rakish spatial relationship over a picture sub-locale of explicit size. The GLCM is made from a dim scale picture. The GLCM is determined how regularly a pixel with dim level (grayscale force or Tone) esteem I happens either on a level plane, vertically, or corner to corner to adjoining pixels with the worth j . A notable factual apparatus for removing second-request surface data from pictures is the dark level co-event. The GLCM lattice is one of the most well known and viable wellsprings of highlights in surface examination. For a district, characterized by a client determined window, GLCM is the lattice of those estimations over all dim level sets. In this technique, highlights are determined dependent on the total contrasts between sets of dim levels or normal dim levels rather than unique dark level pixel values [6]. This methodology makes the measurements somewhat more powerful to brightening varieties than on account of GLCM. The paper is organized as follows. The image information base is portrayed in segment 2. The component extraction strategies are portrayed in segment 3. The arrangement strategy is listed in area 4. The examination work is finished up in segment 5.

2. IMAGE DATABASE

The images used for testing the algorithms developed are medical images. In this work also, the normal lung images and a cancer affected lung images are taken from fifty different peoples to characterize the lung cancer. The image data for this research work is grouped as normal and abnormal images. The Medical Images are CT Lung axial view images, CT Lung sagittal view images and CT Lung coronel view images are taken from ten different people in DICOM format. A sample of normal CT lung and cancer affected CT lung images are shown in Figures 1 and 2 [7, 8].

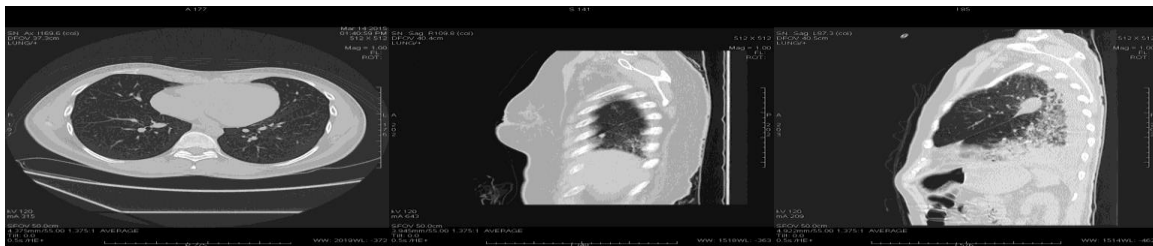


Figure 1. CT images of normal lung image in DICOM



Figure 2. CT images of lung cancer image in DICOM

3. HARALIC TEXTURE FEATURES

Haralick implemented fourteen texture features from GLCM for an image. These features are as follows:

- Entropy: It is a measure of randomness of the input image. Mathematically, it can be represented by

$$\sum_{i,j=0}^{N-1} P(i, j) * [-\ln P(i, j)] \quad (1)$$

- Contrast: It will measure the intensity or gray-level variations between the reference pixel and its neighbor.

$$\text{Contrast} = \sum_{n=0}^{N-1} n^2 \sum i - j = n P_d(i, j) \quad (2)$$

- Correlation: This feature measures how a pixel is correlated to its neighborhood. Its equation is given by

$$\frac{(i-\mu)(j-\mu) * P(i, j)}{\sigma_i \sigma_j} \quad (3)$$

- HomogeneityL: The equation is given by

$$\sum_{i,j=0}^{N-1} \{P(i, j) | 1 + (i - j)\} \quad (4)$$

- Energy : it is represented by

$$E = \sum_{i,j=0}^{N-1} P(i, j)^2 \quad (5)$$

- Autocorrelation: It is the measure of the coarseness of the image. If low value, texture is rough.
- Dissimilarity: It is a measure of dissimilarity between gray pixels of an image.
- Cluster shade: The cluster shade will give information about the skewness of the matrix.
- Variance: The equation is given by

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i - \mu)^2 P_d(i, j)^2 \quad (6)$$

- Angular second moment (ASM): The equation is given by

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_d(i, j)^2 \quad (7)$$

- Inverse difference moment (IDM): IDM measures the closeness of the distribution of the GLCM elements to the GLCM diagonal.

Here Fast Fuzzy C means algorithm used for segmentation. Threshold segmentation is called the threshold process, the easiest for of segmentation of images. This approach is based on a clip-level (or threshold value) to turn a gray-scale image into a binary one. The trick to this approach is to pick the threshold value [9]. The segmentation goal is to reduce or alter an images portrayal into one that is more concrete unit simpler to interpret. I make segmentation usually used in the CT images to identify normal and cancer. In this research work many classification algorithms have been developed for finding cancer in Lung images [10]. In this work, the identification task done by classification method from a set of groups the images [11].

4. RESULT AND DISCUSSION

This section centers around back-propagation neural network (BPN) system-based classifiers that are appropriate for deciding sickness [12]. One of the generally experienced dynamic errands of human movement is order. The order can be characterized as the distinguishing proof undertaking to which a lot of the gathering, another perception has a place, based on preparing a lot of information containing observation [13, 14]. In this examination work, the diverse CT images are the gatherings and the preparation information incorporate the highlights, which were removed from typical and strange images [15, 16].

In this work, the BPN organize is utilized for characterizing the CT images. The highlights are separated from CT images of ordinary lung and malignant growth influenced lung are taken into the examination. GLCM based highlights are extremely helpful in distinguishing sicknesses since the highlights are indicating the wide contrast between the two classes [17, 18]. The BPN is utilized to arrange these images [19, 20]. The exactness of the system is changed so as to improve the order of the classes dependent on the capacity of created calculation. The inferred highlights, which are tabulated in Table 1, gives a wide distinction between the typical and malignant growth pictures and the proposed pressure calculations don't influence the qualities much. The ideal calculation is picked dependent on the arrangement accuracy [21, 22]. The GLCM features are analysed in this work, so as to characterize the lung CT images. The variation between the normal image and cancer image, clearly reveal the capability of this method to classify the normal image from the cancer image [23-25].

Table 1. GLCM image feature of normal and cancer lung image

Features	Normal Lung	Cancer Lung
Image entropy	5.73	3.56
Auto correlation	21.56	7.34
Contrast	0.56	0.31
Correlation	0.93	0.95
Cluster prominence	535.46	234.78
Cluster shade	82.96	78.95
Dissimilarity	0.25	0.23
Sum of square	21.71	8.04
Sum of average	8.52	4.45
Sum of variance	61.93	19.47
Information measure of correlation	0.61	0.63
INM	0.97	0.43
Energy	0.31	0.18
Maximum probability	0.52	0.36
Homogeneity	0.91	0.67

5. CONCLUSION

In this work, the algorithm is developed for classify the bio medical images. The features are extracted from CT images of normal Lung and Cancer affected Lung is taken into the study. Even though, each disease type has unique characteristics and patterns, some similarities are also found among these categories that will lead to difficulty in designing a classifier with a correct decision boundary. Hence, the selection of features is a complex problem, which is overcome by careful trial and error process. Moreover, efficient feature selection is still a problem in medical Images and it can be addressed in future, in an effective manner to achieve better results. The classification accuracy of the Binary classifier finds the proposed algorithms suitable for identifying the cancer disease.

ACKNOWLEDGEMENTS

We would like to thank the management of Sathyabama Institute of Science and Technology Chennai, Tamil Nadu for allowing us to conduct study.

REFERENCES

- [1] A. Said and W. A. Pearlson, "An image multi resolution representation for lossless and lossy compression," *IEEE Transactions on Image Processing*, vol. 5, no. 9, pp. 1303-1310, 1996.
- [2] C. Deng, et al., "Content-based image compression for arbitrary-resolution display devices," *Proceedings of 2011 IEEE International Conference on Communications*, pp. 1-5, 2011.
- [3] L. S. Davis, et al., "Texture analysis using generalized co-occurrence matrices," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 3, pp. 251-259, 1979.
- [4] R. M. Haralick, et al., "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. SMC-3, no. 6, pp. 610-621, 1973.
- [5] J. S. Weszka, et al., "A Comparative Study of Texture Measures for Terrain Classification," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. SMC-6, no. 4, pp. 269-285, 1976.
- [6] R. Pandian, et al., "Characterization of CT cancer lung image using image compression algorithms and feature Extraction," *Journal of scientific and industrial research*, vol. 75, no. 12, pp. 747-751, 2016.
- [7] R. Pandian and T. Vigneswaran, "Adaptive wavelet packet basis selection for zerotree image coding," *International Journal of Signal and Imaging Systems Engineering*, vol. 9, no. 6, pp. 388-392, 2016.
- [8] A. El-Bazl, et al., "Automatic identification of lung abnormalities in chest spiral CT scans," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 261-264, 2003.
- [9] R. Pandian, "Evaluation of image compression algorithm," *2015 IEEE Under water Technology (UT)*, pp. 1-3, 2015.
- [10] K. Kanazawa, et al., "Computer aided diagnosis for pulmonary nodules based on helical CT image," *Computerized Medical Imaging and Graphics*, vol. 22, no. 2, pp. 157-167, 1998.
- [11] D. T. Lina and C. R. Yan, "Lung nodules identification rules extraction with neural fuzzy network," *Proceeding of the 9th International Conference on Neural Information Processing*, vol. 4, pp. 2049-2053, 2002.
- [12] G. P. Zhang, et al., "Neural networks for classification survey," *IEEE Transactions on Systems, Man, and Cybernetics Part C*, vol. 30, no. 4, pp. 451-462, 2002.
- [13] C. Deng, et al., "Content-Based Image Compression for Arbitrary-Resolution Display Devices," *IEEE Transaction on Multimedia*, vol. 14, no. 4, pp. 1127-1139, 2012.
- [14] I. Maglogiannis, et al., "Waveleth ROI coding support for mobileaccess to DICOM images over heterogeneous radio networks," *Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 458-466, 2009.

- [15] D. Dragan and D. Ivetic, "An approach to DICOM extension for medical image streaming," *DAAAM International Scientific Book*, pp. 25-35, 2009.
- [16] R. A. Osornio-Rios, "Identification of positioning system for industrial application using neural network," *Journal of Scientific and Industrial Research*, vol. 76, no. 3, pp. 141-144, 2017.
- [17] W. J. Choi, et al., "Automated Pulmonary Nodule Detection Based on Three-Dimensional Shape-Based Feature Descriptor," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 37-54, 2014.
- [18] D. M. Peña, et al., "Auto Diagnostics of Lung Nodules Using Minimal Characteristics Extraction Technique," *Diagnostics*, vol. 6, no. 1, pp. 13, 2016.
- [19] N. Camarling, et al., "Combination of Computer-Aided Detection Algorithms for Automatic Lung Nodule Identification," *International Journal of Computer Assisted Radiology and Surgery*, vol. 7, no. 3, pp. 455-464, 2012.
- [20] A. Teramoto and H. Fujita, "Fast Lung Nodule Detection in Chest CT Images Using Cylindrical Nodule-Enhancement Filter," *International Journal of Computer Assisted Radiology and Surgery*, vol. 8, no. 2, pp. 193-205, 2013.
- [21] S. R. Suralkar and P. M. Patil, "Correlation based Fingerprint Image Segmentation," *International Journal of Computer Applications*, vol. 68, no. 7, pp. 1-3, 2013.
- [22] M. F. Fahmy and M. A. Thabet, "A fingerprint segmentation technique based on morphological processing," in *IEEE International Symposium on Signal Processing and Information Technology*, pp. 000215-000220, 2013.
- [23] K. Shreedarshan and Selvi S. S., "Crowd Recognition System Based on Optical Flow Along with SVM classifier," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 4, pp. 2451-2459, 2019.
- [24] A. A. Abdullah, A. F. D. Giong, and N. A. H. Zahri, "Cervical cancer detection method using an improved cellular neural network (CNN) algorithm," *Indonesia Journal of Electrical Engineering and Computer Science*, vol. 14, no. 1, pp. 210-218, 2019.
- [25] M. Y. Kamil and A. M. Salih, "Mammography Images Segmentation via Fuzzy C-mean and K-mean," *International Journal of Intelligent Engineering and Systems*, vol. 12, no. 1, pp. 22-29, 2019.

BIOGRAPHIES OF AUTHORS



R. Pandian is currently working as an Associate Professor in the Department of Electronics and Instrumentation Engineering at Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India. He has graduated from Madras University, Chennai in 1999 with Bachelor's Degree in Electrical and Electronics Engineering. He has obtained his M.E Degree in Applied Electronics from Anna University, Chennai 2007. He has obtained his Ph.D. from Sathyabama Institute of Science and Technology.



D.N.S Ravi Kumar has completed his M.Sc.(Applied Electronics) in the year 2003 from Bharathiyar University Coimbatore, India and M.E ((Applied Electronics) in the year 2006 from Sathyabama Institute of Science & Technology Chennai India. He is pursuing PhD in the field of Vehicular Communication (Distributed Embedded Systems) from Sathyabama Institute of Science and Technology since 2015, Chennai India. His field of interests includes Embedded Systems, Vehicular Communication and Robotics. He is currently working on Smart car Technology (V2V). Apart from research an innovator to spark startup ideas to implementation



Dr. R. Rajakumar is a Professor in Department of Mathematics having experience of 25 years in teaching and working in our University for the past 13 years with Doctorate in the field of Chaos Theory, joined Sathyabama University as Lecturer in 2003, completed PhD in the year 2013 from Sathyabama Institute of Science and Technology.