IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-10, Oct 2018 E-ISSN: 2347-2693

Lung Cancer Detection Using Semantic Based ANN Approach

Chandni Kumari[1*], Kamlesh Chandravanshi[2], Gaurav Soni[3]

[1,2]CS, Technocrats Institute of Technology College, Bhopal, India

[*]*Corresponding Author: singhxchandni@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 17/Oct/2018, Published: 31/Oct/2018

***Abstract*—**Medical science is an important part segment of our day to day life. It generates medical report and other statistics in the form of multimedia format. The analysis and prediction of related disease can be done using the data analysis. Every prediction and analysis required image processing and its feature extraction. Processing a data with multiple feature aspect is still a challenging issue. Lung is the parts which carry multiple diseases as it's a major body cycle process. Further which a proper classification and prediction is problem formulation with several research algorithm. Network based processing of data make use of its phases and help out to find more feature analysis and further detailed classification. In this paper an advance algorithm with semantic and ANN model is proposed for the lung cancer disease prediction from its image format. The analysis is compared with traditional SVM approach and proposed Semantic ANN approach. Implementation is performed using the images collected from web medical resources using MATLAB platform. The computed result shows the efficiency of proposed network layer based semantic solution for the lung cancer prediction and its feature analysis.

***Keywords*—**Machine Learning Algorithm; Artificial Neural networks; MATLAB; Data Sets; Genetic Algorithm.

I. INTRODUCTION

"Lung cancer, also known as lung carcinoma, is a malignant lung tumor characterized by uncontrolled cell growth in tissues of the lung. This growth can spread beyond the lung by the process of metastasis into nearby tissue or other parts of the body. Most cancers that start in the lung, known as primary lung cancers, are carcinomas. The two main types are small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma (NSCLC). The most common symptoms are coughing (including coughing up blood), weight loss, shortness of breath, and chest pains."

Some Facts about Lung Cancer

1. Lung cancer is the number one cause of cancer deaths in both men and women in the U.S. and worldwide.
2. Cigarette smoking is the principal risk factor for development of lung cancer.
3. Passive exposure to tobacco smoke (passive smoking) also can cause lung cancer.
4. The two types of lung cancer, which grow and spread differently, are small cell lung cancers (SCLC) and non-small cell lung cancers (NSCLC).
5. The stage of lung cancer refers to the extent to which the cancer has spread in the body.
6. Treatment of lung cancer can involve a combination of surgery, chemotherapy, targeted therapy, immunotherapy, and radiation therapy as well as newer experimental methods.
7. The general prognosis of lung cancer is poor because doctors tend not to find the disease until it is at an advanced stage. Five-year survival is around 54% for early stage lung cancer that is localized to the lungs, but only around 4% in advanced, inoperable lung cancer.
8. Smoking cessation is the most important measure that can prevent the development of lung cancer.

We have explained about lung cancer along its types and the reasons why it happens. So, we will somehow work with the machine learning mining algorithms that will be helpful in analyzing the from the medical reports and the work will focus on the Semantic Based ANN Approach with its efficiency over traditional algorithms to provide better results.

II. RELATED WORK

Abdullah et al. [2012] [1] stated that the segmentation of the lung region due to the limitation regarding on the similarities of the intensity in the X-ray image. As for lung cancer nodule detection process, it does not seem to be the problem because of the absent of the similar intensity due to the lung segmentation done. It can be used in the lung cancer application, the system can also be used in the application

such as the detection and classification of breast tumour in mammography images regarding on the higher variation of intensity present.

Zare et al. [2011] [2] declared that the approaches of content-based image retrieval (CBIR) using low level features such as shape and texture are investigated in order to create a framework that classify medical X-ray image automatically. grey level Co- occurrence Matrix, Canny Edge Operator, Local Binary Pattern and pixel level information of the images act as image based feature representations. The performance of image classification offered by combining the promising features stated above is investigated. Experimental results using 116 different classes of 11,000 X-ray images showed 90.7% classification accuracy.

Gomathi et al. [2010] [3] expressed that Computer Aided diagnosing system which uses FPMC algorithm for segmentation to improve the accuracy. Rule based technique is applied to classify the cancer nodules after segmentation. For its better classification, the learning is performed with the help of Extreme Learning Machine. Patil et al. [2009] conveyed that image segmentation is important for medical image analysis. It helps to detect the absence or presence of disease in an image. The Grey Level Co-occurrence Matrix (GLCM) technique is used to estimate texture features. It is applied on two main types of lung cancer images, like Small-cell, Non-Small Cell type and as well as on TB database.

Jia Tong et al. [2007] [4] acknowledged that several steps are followed to detect the cancer like segmentation of lung parenchyma, the detection of suspicious nodule candidates, the feature extraction and classification. Here, the author used adaptive threshold segmentation, math morphologic, Gaussian filter, Hessian matrix algorithms.

The ANN (Artificial Neural Networks) plays an active as well as important role to solve various problems related to the health. For example acute diseases (those diseases which happens suddenly in onset) and mild diseases (which are cold in nature). In this paper author used HNN (Hopfield Neural Networks) and FCM (Fuzzy C-Means) for the separation of the colorful images and to compare their classification accordingly is shown in their study. This study shows that HNN provides better results as compare to the FCM and allows extracting the cytoplasm as well as nuclei regions [5]. Jia Tong, has worked on the HRCT images to find out the nodule (it is the swelling or the aggregation cells in the body) patients in lung field [6]. A unique lung segmentation technique was introduced by the author Lin-yu-Tseng to improve the accuracy and to destroy the trachea from lungs completely [7]. Anita Choudhary, in this paper author has worked on the Digital Image Processing Technique that will provide more accurate quality of the images [8]. Azian Azamini Abdullah Described the algorithm that provides the detection of the symptoms of lung cancer in the X-Ray films by CNN (Cellular Neural Networks) simulation [9].

The affectability is additionally imperative in this perspective thus by utilization of the Massive Training ANN (MTANN) channel, the specificity and affectability of CAD framework was improved altogether with a difference image strategy accomplished 96% affectability detailed by Suzuki K et al. In the result of Shi Z et al found a neural system gathering (NNE) in chest radiograph for lessening false positive for knobs location in PC aided diagnosis [10] [11].

In this paper author has Parveen S S, described the new CAD technique for the detection of the crucial regions images in order to divide the areas of the lungs by using the automated region growing methods to enhance the accuracy rate of the diagnosis [12]. Another method is described by the author Yang Song that is PET-CT thoracic images which are based on low-level intensity and the features of the neighbor also the high-level contrast type features, along with the 2-level classification of the SVM technique. It is also very helpful in handling the large number of unsynchronized patterns that showed the high detection performance along with the capability [13].

In this paper author Ye X proposed the work for the detection of the nodules for both the solid and the ground glass opacity (GGO) by calculating the values of the 3-D geometric and the features of the statistical intensity [14]. Another author Raman [15] proposed review of the enhancement techniques of the images. In which the point processing method is much in use, still some of the image processing images are used for the contrast enhancement.

III. METHODOLOGY

In order to perform to perform the better classification with low computation and high accuracy, A proposed algorithm scenario is presented which use the ANN Layer model with enhance feature extraction approach.

The proposed algorithm SYNANN is presented which is an algorithm using region of interest, applying prop filtering approach for pre-processing and feature extraction. Further the data is processed with ANN model. The obtained features are trained using the NPR tool and find the advantage accuracy over SVM classification. The algorithm SYNANN uses multiple layer input and output values which are able to process number of features. Thus finding a proper relevant classification is performed. The proposed methodology and its steps which are followed are presented below:

1. Finding the dataset from different resources which is extracted from the Lung units sample dataset availability.

2. Linearizing and Binaries the image. Here a binary conversion of image is performing which strength the complete Lung Image.
3. Thresholding and Smoothening the Image. Here a thinning is applied which help in finding the binary thinning to the image.
4. Finding the segmented area and feature extraction. The prop region approach is provided to extract the image features.
5. Prop Region help in extraction of feature, here 23 features are obtained. Also a ROI region of interest is extraction using the particular area selection. It helps in minimizing the processing.
6. Apply the ANN approach using the NPRtool. The Neural network approach help in processing a finding the accuracy over detection.
7. Finally the computation parameters were computed and compared with the current approach to show the efficiency of proposed algorithm.

The discussed step shows the proposed mechanism followed for better segmentation and classification.

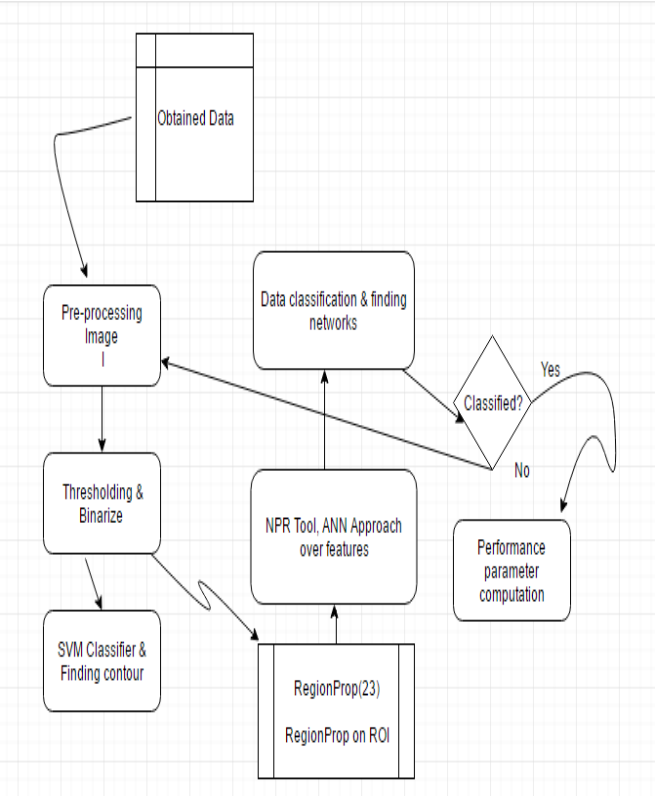

Figure 1: Overall Architecture of Complete Framework.

The figure 1 above, it shows the overall proposed architecture framework which uses for the execution. Proposed scenario shows the outcome performed are processed in systematic way and obtained high performance.

The different provided component works with the flow and providing different steps to process the algorithm data.

- Pre-Processing:
  Image pre-processing is performed which help in smoothing and binarizing the image data. It helps in getting sharpness among the image.
- Thresholding:
  This is the process of thinning the image data which again help in finding the edges, corners and capturing all the corners of an image. It also helps in removing the low intensity level of data available in image.
- Feature Extraction:
  Feature extraction is the further performed process using which the number of feature from the selected area is extracted. It helps in analyzing the image data using the features available in the multimedia.
- Classification:
  This is the final process of prediction where the algorithm of processing the features and finally predicting the classification results such as accuracy and confusion matrix. Finally a result on computation is also getting performed.
  Thus the given execution application steps help in completing the execution. This is the main advantage of processing data in multiple steps.
- Pseudo Code Of SYMANN Algorithm
  As per the discussion, proposed algorithm SYNANN is presented and it works with the proper brain tumour data classification. The following is the algorithm Pseudo code which helps in executing the proposed approach.

Upon applying the synaptic approach the proposed algorithm take advantage of ANN layered based approach for finding the efficient classification over disease prediction. In order to find a proper accuracy and computation SYMANN algorithm is proposed which is a hybrid model of finding prediction over short text and lung cancer dataset analysis.

Input: Dataset Attributes, Lung Cancer Di-N, Ci Word Concept.

Output: Semantic Classification, Lung Cancer Prediction, Computations.

Steps: Begin [

Data: For (Di-N)

{

Similarity= ∑i-n Attribute Matching;

Compute p(n) = P r[N];

Entropy= X N i=1 f(ci)E(ci);

Return E(i);

}

Apply ANN (Ei-n)

{

Back propagation;

Compute sigmoid ();

Prediction classification ();

}

Compute utilization parameters;
Exit;
] End;

The algorithm above is the hybrid approach which is presented in proposed work for text similarity disease prediction. The algorithm takes input as text phrase and disease dataset which is required to process. Further a finding of disease prediction by utilizing the textual information is performed. A large textual data is processed and final prediction is made using the ANN approach. The entropy calculation and depth computation over the relational data associated with textual data is analyzed. Thus a final text classification improves the prediction.

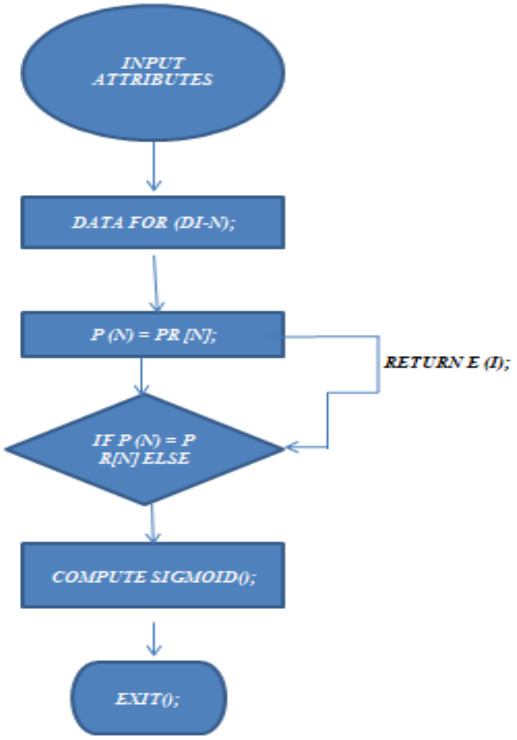

Figure 2: Flow Chart of the Proposed Algorithm.

IV. Results and Discussion

Simulation Setup

To evaluate the performance of existing method local and global image segmentation use MATLAB software 17 with a variety of dataset used for experimental task. This research uses the various tools and functions from MATLAB 2017b. Script generation for thresholding and feature extraction is also getting performed using the MATLAB scripting tool.

Data Set Description

In order to deal with the algorithm execution and result analysis from it. A data of brain tumour images is taken; A data from fig share is taken which is the collection of images. These images contain the images having the brain tumour and thus identification and processing of such images is going to perform. The results are compared on the basis of execution accuracy, precision, recall and ROC Curve by both algorithms.

Comparative Performance Evaluation

Execution Time

The execution time is computed which is taken by algorithm to process the input image. The time difference between the initial and completed time is shown.

Syntax:-

Computation time:

TIC
{
Algorithm;
}
TOC;

Accuracy:

Accuracy = (TP+TN) / (TP+TN+FP+FN)

TPR:

TPR = TP / (TP+FN)

Statistical and Graphical Analysis

Table 1: Comparative Performance Evaluation for the Traditional and SYNANN Techniques with the Various Images.

ALGORITHM/PARAMETER	ACCURACY	ROC CURVE	DETECTION RATE
BAYES ALGORITHM	75.23	71.60	79.32
SVM	81.90	82.66	86
FEATURE BASED ANN	88.98	83.40	91.22
SYMANN	92.20	89.38	94.10

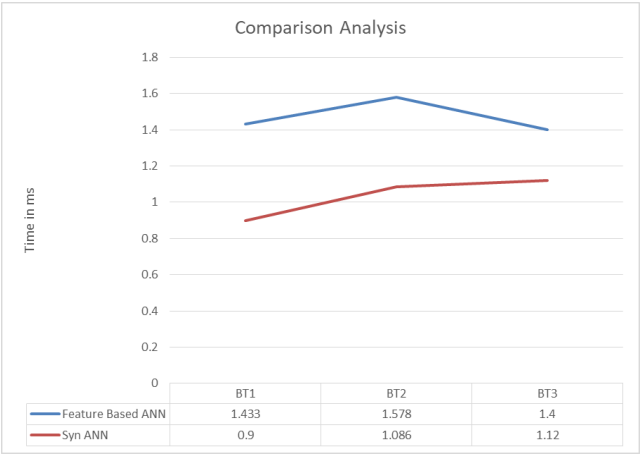

Figure 3: Comparative Result Graph for the Value of Execution Time in Seconds.

An execution with the given dataset and processing with the proposed algorithm shows the advantage in computation while comparing with traditional approach. The execution shows the low computation time.

Table 2: Comparative Performance Evaluation for the Traditional and SYNANN Techniques with the Various Images.

NAME OF IMAGE	NAME OF METHOD	EXECUTION TIME IN MILLISECONDS
BT1	TRADITIONAL	1.433
	SYNANN	0.900
BT2	TRADITIONAL	1.578
	SYNANN	1.086
BT3	TRADITIONAL	1.40
	SYNANN	1.12

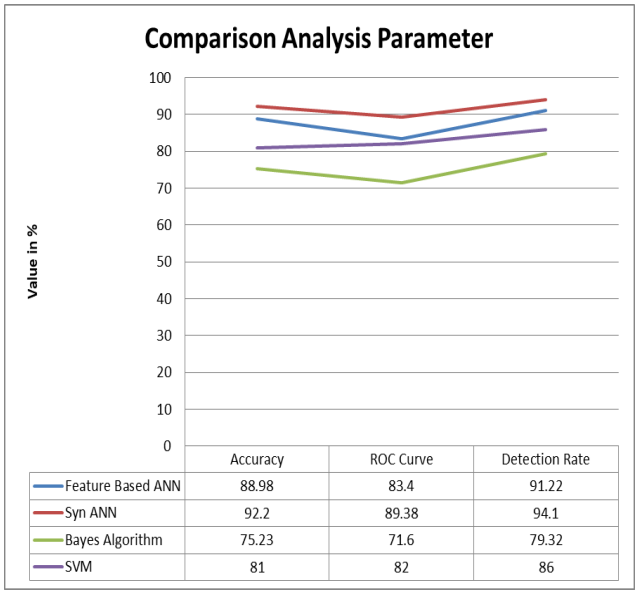

Figure 4: Comparative Result Graph Finds the Value of Accuracy in % Age.

The figure 4 above shows the graphical bar chart analysis of the observed result. The algorithm graphical comparison analysis over proposed solution and existing outcomes.

V. CONCLUSION AND FUTURE SCOPE

Data mining helps in proper knowledge extraction from a large and dynamic dataset. Data mining help in many areas such as security, intrusion detection and advance health care detection. Lung cancer prediction contains a dataset and real time usage. It needs a human skillset which is time consuming and costly while analysing large data input. A short text data value analysis is required to process the dataset and finding a prediction for lung cancer disease. In this paper a hybrid ANN based approach with Synaptic analysis approach is presented which compared with traditional classification approach. The result outcome shows the efficiency of proposed hybrid structure over the traditional semantic and classification approach. A further work to deal with real time dataset and continuous value parameter can be done. The work also can lead in hardware integration of current values and real time values to monitoring its usability. The concept can incorporate with IoT, so that the remote area and extended laboratory can be used as a service on demand. Thus the further concept can be done with IOT as well as cloud based accessing system of algorithm.

ACKNOWLEDGMENT

I am very thankful to my guide who has helped me throughout.

REFERENCES

[1]. Azian Azamimi Abdullah, Syamimi Mardiah Shaharum, "Lung Cancer Cell Classification Method utilizing Artificial Neural Network", Information Engineering Letters, ISSN:21604114, Volume 2, Issue 1, March 2012.

[2]. Mohammad Reza Zare, Ahmed Mueen, Woo Chaw Seng, Mohammad Hamza Awedh, "Correspondence Systems and Networks Combined Feature Extraction on Medical X-beam Images" Third International Conference on Computational Intelligence, Bali, Indonesia, July 26-July 28, 2011.

[3]. M. Gomathi, Dr. P. Thangaraj, "A Computer Aided Diagnosis System for Detection of Lung Cancer Nodules Using Extreme Learning Machine", International Journal of Engineering Science And Technology Vol.2(10), 2010.

[4]. Jia Tong, Zhao Da-Zhe, Wei Ying, Zhu Xin-Hua, Wang Xu, "PC Aided Lung Nodule Detection Based On CT Images", IEEE/ICME International Conference on Complex Medical Engineering, 2007.

[5]. Taher F, Sammouda R, ―Lung Cancer Detection By Using Artificial Neural Network and Fuzzy Clustering Methods‖, IEEE GCC Conference and exhibition(GCC), 2011, 295-298.

[6]. Tong J, ZHAO D, Ying W, ZHU X and Xu W, ―Computer-Aided Lung Nodule Detection Based on CT Images ― IEEE-International Conference on Complex Medical Engineering, 2007, 816 – 819.

[7]. Tseng L, Huang L, ―An Adaptive Thresholding Method for Automatic Lung Segmentation in CT Images ― IEEE AFRICON, September 23-25, 2009, pp 1-5

[8]. Chaudhary A, Singh S, ―Lung Cancer Detection Using Digital Image Processing‖, IJREAS vol 2, Issue 2, 2012, 1351-1359.

[9]. Abdullah An and Mohamaddiah H, Development of Cellular Neural Network Algorithm for Detecting Lung Cancer Symptoms, IEEE EMBS Conference on Biomedical Engineering and Sciences, 2010, 138-143.

[10]. Suzuki K, Shi Z, and Zhang J, ―Supervised Enhancement of Lung Nodules by utilization of a Massive-Training Artificial Neural Network (MTANN) in Computer-supported diagnosis‖, IEEE, nineteenth International Conference on Pattern Recognition, 2008,1-4.

[11]. Shi Z, Suzuki K and He L, ―Reducing FPs in Nodule Detection utilizing Neural Networks Ensemble‖, IEEE, Second International symposium on data science and building, 331-333.
[12]. Parveen S and Kavitha C, ―Detection of Lung Cancer Nodules Using Automatic Region Growing Method‖, IEEE-31661 International Conference on Computing and Networking Technologies (ICCCNT, 2013) 1-6.
[13]. Tune Y, Cai W, Kim J and Feng D,―A Multistage Discriminative Model for Tumor and Lymph Node Detection in Thoracic Images‖, IEEE Transactions on Medical Imaging, Vol 31, No 5, 2012, 1061-1075.
[14]. Ye X, Lin X, Dehmeshki J, Slabaugh G and Beddoe G, ―Shape-Based Computer-Aided Detection of Lung Nodules in Thoracic CT Images‖, IEEE Transactions on Biomedical Engineering, , vol. 56, no. 7, 2009, 1810-1820.
[15]. Maini R and Aggarwal H ―A Comprehensive Review of Image Enhancement Techniques‖, Journal of Computing, Vol 2, Issue 3, 2010, 8-13.

