

GENETIC BEE OPTIMIZED RECURRENT NETWORK BASED LUNG CANCER DETECTION

P. Sankar Ganesh¹, Teja², Marish³ and Dr. K. Devi^{4*}

^{1,2,3}Dr. MGR Educational and Research Institute.

⁴SRM Valliammai Engineering College.

Article Received on 04/04/2022

Article Revised on 25/04/2022

Article Accepted on 15/05/2022

*Corresponding Author

Dr. K. Devi

SRM Valliammai

Engineering College.

ABSTRACT

Lung cancer is one of the serious issue in healthcare application which occurs in smoking people. The lung cancer occurred in types such as small cell and non-small cell lung cancer. It also created by exposure of toxin gases, family history, second-hand smoke etc. The disease requires immediate treatment to avoid the serious issue. Therefore, several researcher's focus on the automatic cancer detection system; however the existing systems fails to concentrate the exact disease affected region prediction. This causes to reduce the efficiency of cancer prediction accuracy. To overcome this issues, in this, genetic bee optimized recurrent network is applied to improve the overall prediction accuracy. Initially, the images are collected from Chest CT-Scan images Dataset which is processed by median filter to removing noise. Then, dual clustering approach is applied to segment the affected region which helps to identify the cancer related features. The extracted features are classified using recurrent approach that uses the optimization algorithm to regularize the network performance. The effectiveness of the system is evaluated using MATLAB based results.

KEYWORDS: Lung cancer, recurrent network, Chest CT-Scan images Dataset, image segmentation.

1. INTRODUCTION

Presently most of the people affected by various diseases due to the life style changes and scientific factors. These changes creates the changes in the genetic cell which affects the cell

mutations.^[1] Then the old cell is replaced by the wrong mutated cell^[2] which is converted into the cancer. According to the report of cancer registries (NNACCR)^[3] around 234030 peoples are suffered by lung cancer. It has several symptoms^[4] such as cough, shortage in breath, voice change, chest pain, blood coughing, sputum color changes etc. In addition to this, people feels weight loss, neurological problems, memory loss, bone fracture, joint loss, bleeding, cachexia and headaches. Once the people suffered by these symptoms, they are requested to keep checking their health status with the help of the screening procedures.^[5] There are several screening techniques such as liquid biopsy, genetic testing, biopsy, bronchoscopy and reflex testing etc. The recent research suggest that radio imaging analysis^[6] are more useful to predict the lung cancer with fastest manner. The radio-imaging is helps to improve the recognition accuracy with minimum computation complexity.

The lung cancer is recognized with the help of computerized tomography (CT), MRI and PET imaging.^[7] Among the various imaging, CT gives the more useful information because it gives the inner details about lung tissues and soft organs. The screening process applied in the automatic disease detection system^[8] recognize the lung cancer effectively. The capture CT images are investigated by applying the image processing and machine learning algorithms. The image processing techniques^[9] uses the filtering approaches, image enhancement technique and segmentation procedures to derive the valuable information. Comparing to other technique, segmentation process^[10] gives the detailed representation of each pixels and changes involved in the pixels effectively. The affected regions are identified using various methods such as Hopfield neural networks, self-organizing maps, fuzzy-c-means clustering, distributed clustering, canny, Sobel, and other region segmentation techniques. The segmentation process investigate the each pixel and similar pixels are grouped to identify the disease affected region. This process is further improved by using the intelligent techniques^[11] such as fireflies, genetic algorithm, artificial bee colony, butterfly optimization, particle swarm optimization etc. The optimization algorithms reduces the complexity while selecting the similar pixel and clustering center. After that, various features such as statistical, spectral, robust and binary patterns are derived to identify the exact cancer region. The derived features are classified by applying various classifiers that classifies the lung cancer by matching with the training pattern. There are several methods^[12] are applied in this field to improve the classification accuracy; however, the existing systems are fails to predict the exact cancer region. The segmentation approach requires the improvement to identify the edges relevant features accurately. If the inner edge details eliminated during the

segmentation it will affect the lung cancer prediction rate. Moreover, the false segmentation process, influence the lung features which causes to the mis-classification error rate.^[13] To resolve this issues, in this work effective and optimized clustering process is introduced. This work uses the chest CT-scan image dataset information is utilized to analyze the effectiveness of the system. The collected images are investigated using the dual-clustering approach that segments the affected region with maximum segmentation accuracy. Finally the optimized neural network is applied on the region related features that classifies the cancer with minimum error rate and high recognition accuracy.

Then the rest of the paper is organized as follows. Section 2 discussing the various researcher's work on the lung cancer prediction process. Section 3 analyze the working process of the optimized neural network based lung cancer prediction and efficiency of the system is evaluated in section 4. Conclusion described in section 5.

2. Related Works

Coudray, et al., 2018^[14] applying deep learning approach to predict the lung cancer from the histopathology images. The collected images are processed by convolute model that predicts the changes in every tissues. The tissues related features are trained by using the inception v3 layer which helps to classify the lung cancer with 85.6% of accuracy.

Zhang, et al., 2019^[15] detecting lung cancer from CT images using the deep convolution neural networks. This work designed to identifying the pulmonary nodules in the lung region. The nodules are more useful to recognize the predicted region is benign or malignant. The effective utilization of the fully convoluted network recognize the lung cancer with 84.4% of accuracy compared to the other methods. The effectiveness of the system is assessed with the manual doctors to predict the efficiency of the system.

Shakeel, P. Mohamed, et al., 2020^[16] predicting lung cancer from CT images by applying the deep ensemble classifier. The main intention of the work is to improving the cancer recognition accuracy and reducing the mis-classification error rate. The collected CT images are investigated according to the multilevel brightness approach that removes the noise from image also enhance the image quality. Then affected regions are segmented according to the deep learning concept and the feature are extracted from region. Finally, ensemble classifiers are applied to recognize the lung cancer by utilizing the selected features.

ALzubi *et al.*, 2019^[17] introducing the ensemble boosted neural network approach to classify the lung cancer. The collected lung images are processed by using the newton-Raphson's maximum likelihood approach that removes the noise image. Then different features are extracted which are classified using ensemble classifier that recognize the cancer with minimum delay and maximum recognition accuracy.

Prabukumar, *et al.*, 2019^[18] introducing the cuckoo optimized support vector machine for recognizing the lung cancer. The intention of this work is to minimize the latency, increases the privacy while classifying the cancer. The collected images are investigated according to the fuzzy c-means clustering approach that recognize the affected region. Then different features such as intensity, statistical, texture and geometric are extracted. The derived features are investigated based on the cuckoo search algorithm. Finally, support vector machine is applied to classify the benign and malignant tumor. The effectiveness of the system is evaluated using public database and cancer action program (ELCAP) in which system classifies the cancer with 98.79% and 98.13% of accuracy. According to the various researcher's opinions, lung cancer is recognized by applying different machine learning techniques. However, the existing systems are fails to minimize the false classification rate while analyzing the lung cancer features. To overcome the above issues, in this work, optimized deep recurrent network is applied to recognize the lung cancer with maximum recognition rate.

3. Genetic bee optimized recurrent network based lung cancer detection

This section discussing the genetic bee optimized recurrent network based lung cancer detection process. The CT images are collected from the cancer imaging archives (CIA) dataset^[19] which has 5043 images. The sample CT images are illustrated in figure 1, in which 3000 images are treated as testing and 2043 images are used as testing images.

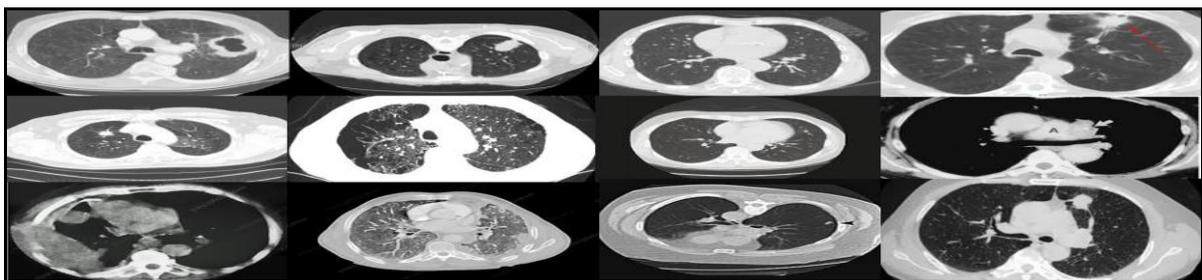


Figure 1: Sample CT images.

The gathered CT images investigated using several steps such as image preprocessing, segmentation, feature extraction and cancer detection. Each step has specific functions and procedure that used to analyze the CT image pixels with minimum computation complexity. According to the discussion, the working process of lung cancer detection architecture illustrated in figure 2.

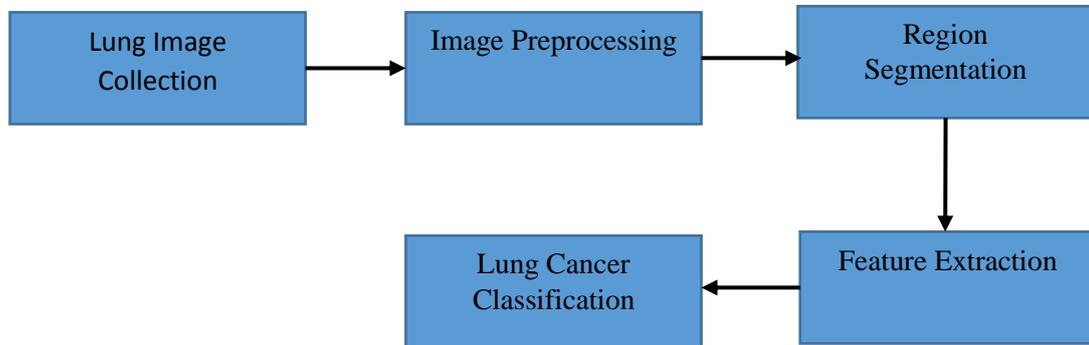


Figure 2: Working flow of lung cancer classification.

Image preprocessing

The first step of this work is image collection and noise removal because the gathered information consists of several inconsistent and noisy pixel details. Once the image is collected, noise pixel is identified and replacement should be performed to improve the image quality. In this work, median filter is applied to eliminate the noisy pixel. Each pixel is computed against the threshold value, if the pixel is corrupted by any noisy details, it has been replaced by median value. The median value is estimated with the help of neighboring pixel. The neighboring pixels are arranged in the sorting order and the median value is computed which utilized to replace the noise pixel. This process repeated until to reach the threshold value. The effective computation of each pixel and neighboring details helps to improve the image quality. According to the above discussion, the noise removed images are illustrated in table 1.

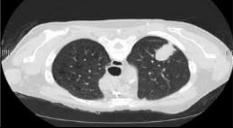
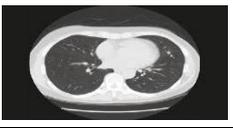
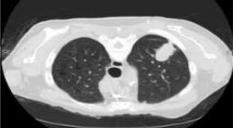
Input Images				
Preprocessed Images				

Image Segmentation

The second part of this work is to segmenting the cancer affected region which is done by applying the dual clustering approach. The segmentation is done according to the images characteristics such as histogram and border gradients. Image are investigated according to the histograms and the similar histogram pixels are grouped. The clustering process is performed along with the border gradient. Therefore, the regions are investigated according to the image brightness (B) in one space and second space. In one space, histogram brightness is evaluated $H = H(B)$ and the second space, dual-three-dimensional image brightness value is computed $B = B(x, y)$. In addition to this, the distributed of brightness pixel should estimate in one-space to compute the minimal clustering deviation. Then threshold value of the brightness should be estimated to reduces the distance. The clustering is formed by estimating the image black and white pixels. Then the threshold value of the image brightness is computed using eqn (1)

$$\text{image} - \text{bitmap} (b) = \varphi(x, y) \quad (1)$$

The computed image pixels are investigated with binary image threshold value (T) and the clustering is formed according to the below condition.

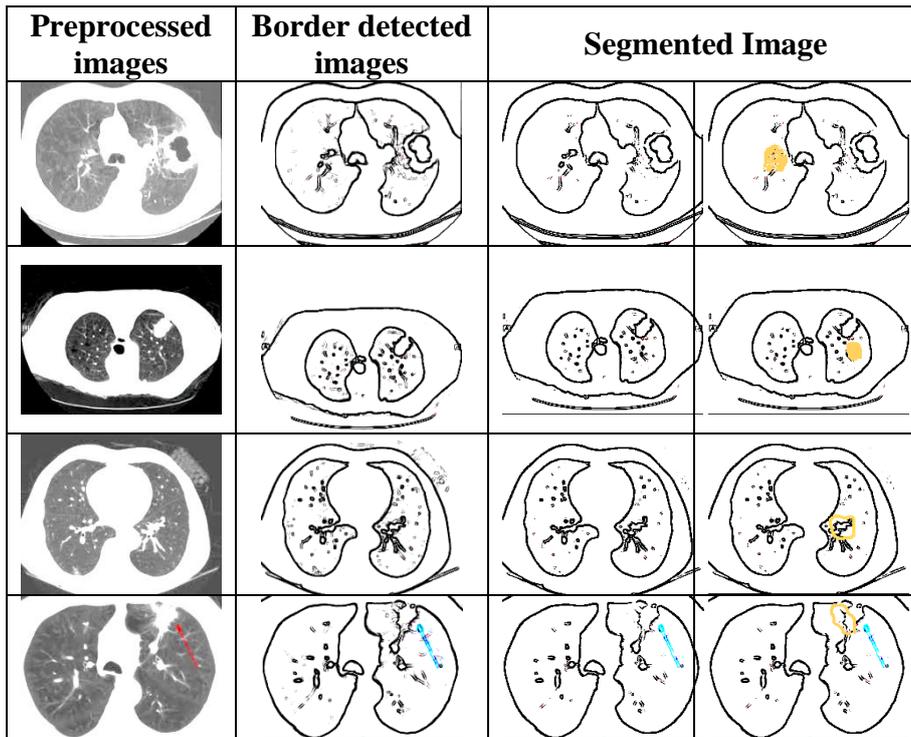
$$\begin{cases} \varphi(x, y) = 0 & \text{if } B(x, y) < T \\ \varphi(x, y) = 1 & \text{if } B(x, y) \geq T \end{cases} \quad (2)$$

In eqn (1) b is the bitmap of the objects in the dual space. The bitmap is measure to estimate the distribution of the black and white pixels in the images. According to the analysis, image border should computed using eqn (3)

$$M_{DC} = G / (k * L) \quad (3)$$

In eqn (3), image border should be identified from the border gradient value G , length of the border and the difference between the background and object brightness value (k). From the computation, the minimum clustering difference and maximum M_{DC} value should be investigated in one and second space. According to the computation, the condition related region should be segmented to improve the overall cancer recognition accuracy. Then the respective segmented region is illustrated in table 2.

Table 2: Region Segmented Images.



Feature Extraction

The third step is feature extraction which is the representation of the segmented region. In this work, various statistical features are derived as the lung cancer related features. The extracted features are illustrated in table 3.

Table 3: Feature Extraction.

Features	Formula
Mean(μ)	$\mu = \frac{1}{M} \sum_{i=1}^M A_i$ M is total amount of pixel present in the segmented region (A)
Standard Deviation (σ)	$\sigma = \sqrt{\frac{1}{M} \sum_{i=1}^M (A_i - \mu)^2}$
3 rd moment skewness (3 – msk)	$3 - msk = \left(\frac{1}{M * \sigma^3} * \sum_{i=1}^M (A_i - \mu)^3 \right)^{1/3}$
4 th moment kurtosis (4 – mku)	$4 - mku = \left(\frac{1}{M * \sigma^4} * \sum_{i=1}^M (A_i - \mu)^4 \right)^{1/4}$

The derived features are reflect the cancer region characteristics that are more useful to identify the patient disease affected rate. The detailed explanation of the cancer classification process discussed in the following section.

Cancer Classification

The last step of this work is cancer classification which is done by applying the genetic bee optimized recurrent neural networks. The recurrent network analyze the derived features and predicts the new data with minimum computation complexity. The training samples are investigated by different layers and labels are generated according to the previous learning patterns. The trained pattern more useful to identify the new incoming features cancer status. The classification process predicts the output value; the main intention of the work is to minimize the error value. Suppose the network having maximum deviation between the output and input value, network parameter should be updated to reduce the false classification rate. The extracted input is processed in the input layer with respective weight parameters and the output is estimated in hidden layer as follows.

$$H_t = \phi(X_t W_{xh} + H_{t-1} W_{hh} + \delta_h) \quad (4)$$

In eqn (4), the input X_t is processed with respective network parameters such as weight W_{xh} and bias value δ_h to get the hidden layer output value H_t . The hidden layer uses the continuous information and the output is predicted in each layer. The entire outputs are combined and the exact output value is calculated using eqn (5).

$$O_t = H_t W_{hq} + \delta_q \quad (5)$$

The computed output value is compared with the trained pattern to estimate the deviation between the actual and predicted value. Here, the network process is regularized by applying the genetic bee optimization algorithm. The algorithm helps to resolve the objective problem by updating the network parameter using eqn (6).

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; x^{(i:i+k)}; y^{(i:i+k)}) \quad (6)$$

Here, learning rate η and gradient ∇_{θ} values are continuously utilized to update the network parameters. According to the process, the momentum optimization is performed as $v_i^{(t+1)} = \alpha v_i^{(t)} - \eta \nabla_{\theta} J(\theta)$ and $\theta_i^{(t+1)} = \theta_i^{(t)} + v_i^{(t+1)}$. For every computation, the network parameters are selected based on the objective function. Based on the objective function, the probability value of the parameter should estimate as $p_i = \frac{fit_i}{\sum_{j=1}^{PS} fit_j}$. According to the computation the searching process is initiated and parameters are updated to minimize the error value. Thus the system successfully classifies the lung cancer by solving the optimization problem.

4. RESULTS AND DISCUSSIONS

This section examines the excellence of the introduced genetic bee optimized recurrent network (GBRNN) approach based lung cancer prediction process. As said, the lung images are collected from CIA dataset which has 5043 images. Gathered images are processed by above discussed processing steps and the effectiveness of the system is evaluated using accuracy, precision, recall and error-rate metrics. The introduced system effectiveness is compared with the existing researcher's work such as deep convolution neural networks (DCNN)^[15], deep ensemble neural networks (DSNN)^[16] and ensemble boosted neural network (EBNN).^[17] The metrics are computed using following eqns.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) * 100\% \quad (7)$$

$$precision = \frac{TP}{TP+FP} \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

Based on the metrics, the dual clustering based segmentation efficiency is examined first to understand how effectively system recognize the lung cancer affected region. The segmentation comparison results are illustrated in 4.

Table 4: Segmentation Efficiency.

Methods	Accuracy	Specificity	Precision	Recall
DCNN	0.834	0.857	0.848	0.862
DSNN	0.858	0.872	0.875	0.893
EBNN	0.867	0.893	0.88	0.90
GBRNN	0.965	0.958	0.967	0.973

The table 4 illustrate the segmentation efficiency of the introduced GBRNN approach. The obtained results are compared with existing research's works such as^{[15],[16]} and.^[17] Among the computation, GBRNN approach utilizes the dual clustering process which analyze the image pixels in one and second space. The effective utilization of image borders and brightness helps to improve the overall segmentation accuracy. The obtained results are high compared to the existing methods. This directly shows that the introduced system extracted regions are more relevant to the lung cancer related information. The segmented region related features are extracted from region which is processed by the introduced classifier to recognize the cancer status. The effective utilization of the learning rate, activation function and updating process improves the minimize the deviation error. Then the obtained error rate results are illustrated in figure 3.

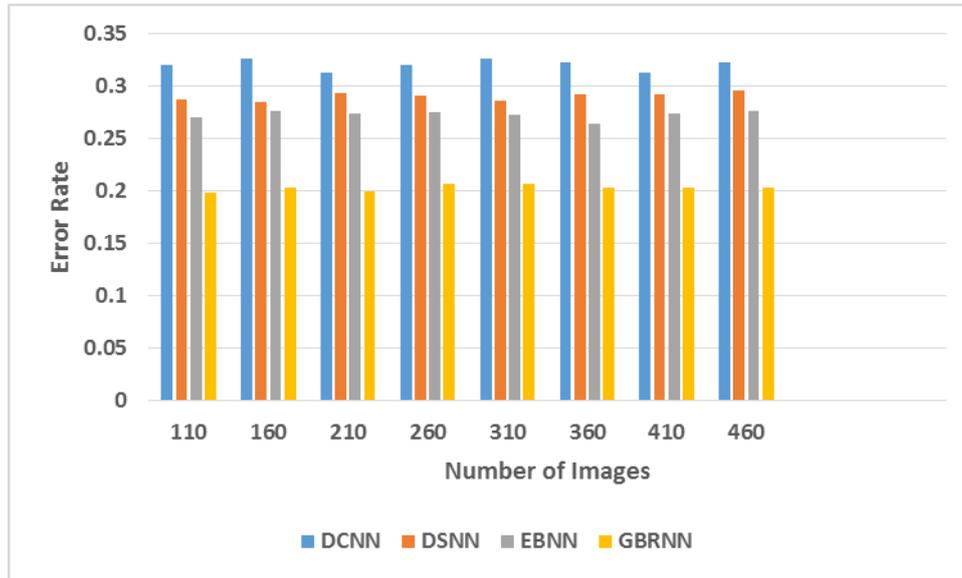


Figure 3: Error Rate.

From the figure 3 it clearly states that introduced genetic bee optimized recurrent neural network (GBRNN) method attains the minimum error rate. The low error rate directly shows that system ensures the low difference between the actual and predicted output value. The GBRNN approach uses the memory unit to save the processing inputs that helps to identify the new features output with minimum deviation. In addition to this, the network uses the effective training patterns that helps to classify the inputs with minimum error rate. Moreover, the optimization algorithm regularize the network parameters according to their characteristics that further improve the system performance. Then the obtained system efficiency is illustrated in figure 4.

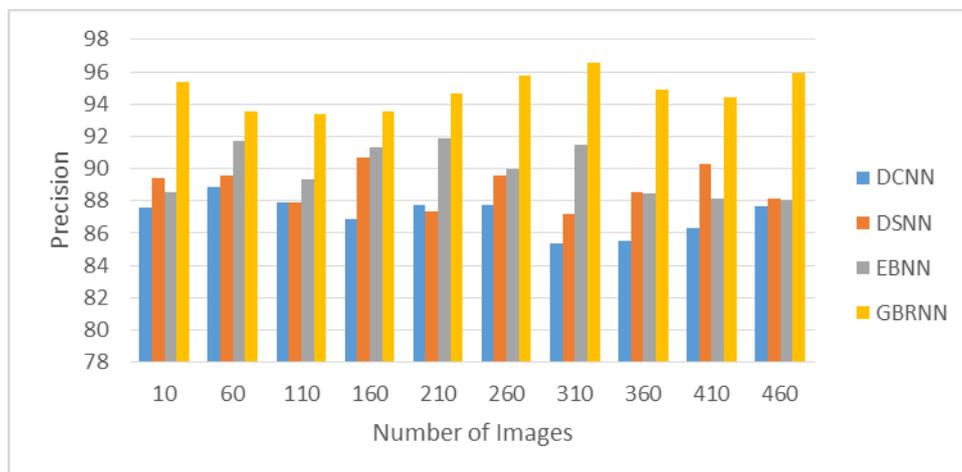
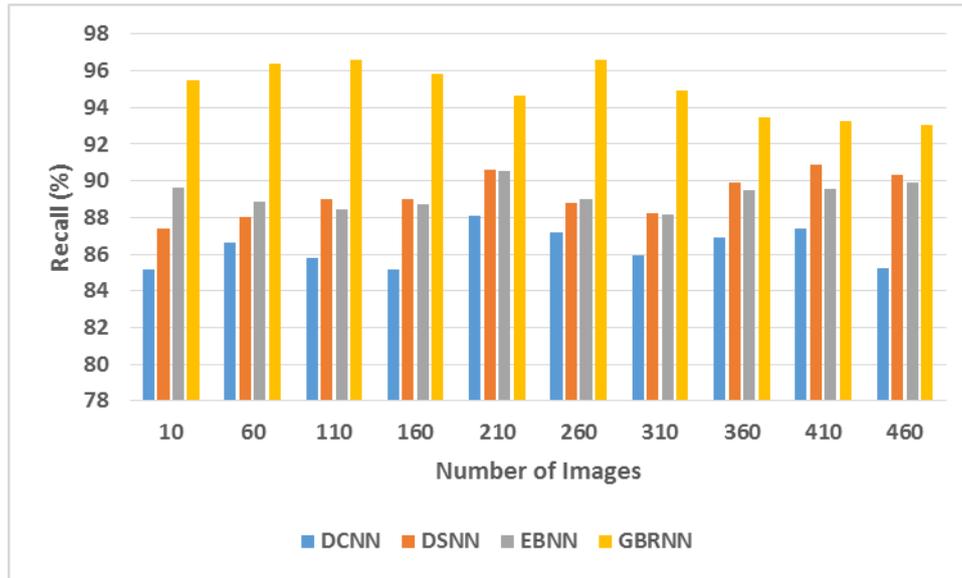


Figure (a) Precision.



(b) Recall

Figure 4: Efficiency Analysis.

According to the above figure it clearly states that introduced GBRNN approach attains the high precision and recall value compared to the other methods. As said earlier, the optimization parameter involvement chose the right parameters to regularize the network performance. The effective utilization of these parameters helps to improve the selection of lung cancer features with minimum deviation value. More ever, the method uses the dual clustering approach based segmented region. The regions are extracted without losing any edge and border details that helps to improves the overall lung cancer recognition accuracy. Then the obtained accuracy value is shown in table 5.

Table 5: Classification Efficiency.

Methods	Accuracy	Specificity	Precision	Recall
DCNN	88.25	89.9	87.35	87.50
DSNN	88.8	88.29	89.51	88.77
EBNN	90.4	90.3	90.38	90.56
GBRNN	95.8	96.35	95.7	95.4

Thus the system ensures the 95.8% of accuracy while classifying the lung cancer with minimum error rate and objective problems.

5. CONCLUSION

Thus the paper analyzing the genetic bee optimized deep recurrent network based lung cancer classification process. The work uses the CIA database images for investigating the lung cancer status. The collected images are investigated using median filter to remove the noise

from the images that helps to reduce the false classification rate. Then the cancer affected regions are predicted according to the dual clustering procedure. The method uses the one and second phase analyzes to identify the affected region. Finally, the region related features are classified using the recurrent network which improves the overall classification rate. At the time, the network performance is optimized using genetic bee algorithm characteristics that enhance the entire classification system performance. In future, the optimization algorithm is applied to improve the overall effectiveness of the system.

REFERENCE

1. Inage, Terunaga, Takahiro Nakajima, Ichiro Yoshino, and Kazuhiro Yasufuku. "Early lung cancer detection." *Clinics in chest medicine*, 2018; 39(1): 45-55.
2. Yang, Gaojian, Ziqi Xiao, Congli Tang, Yan Deng, Hao Huang, and Ziyu He. "Recent advances in biosensor for detection of lung cancer biomarkers." *Biosensors and Bioelectronics*, 2019; 141: 111416.
3. Hasyagar, Keerti Shreedhar, and B. S. Maya. "ANALYSIS ON TECHNIQUES FOR LUNG CANCER CLASSIFICATION USING CT IMAGES."
4. Shakeel, P. Mohamed, Mohd Aboobaider Burhanuddin, and Mohamad Ishak Desa. "Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks." *Measurement*, 2019; 145: 702-712.
5. Huang, Kai-Lin, Shih-Yuan Wang, Wan-Chen Lu, Ya-Hui Chang, Jian Su, and Yen-Ta Lu. "Effects of low-dose computed tomography on lung cancer screening: a systematic review, meta-analysis, and trial sequential analysis." *BMC pulmonary medicine*, 2019; 19(1): 1-11.
6. Wang, Shidan, Donghan M. Yang, Ruichen Rong, Xiaowei Zhan, Junya Fujimoto, Hongyu Liu, John Minna, Ignacio Ivan Wistuba, Yang Xie, and Guanghua Xiao. "Artificial intelligence in lung cancer pathology image analysis." *Cancers*, 2019; 11(11): 1673.
7. Shi, Liting, Yaoyao He, Zilong Yuan, Stanley Benedict, Richard Valicenti, Jianfeng Qiu, and Yi Rong. "Radiomics for response and outcome assessment for non-small cell lung cancer." *Technology in cancer research & treatment*, 2018; 17: 1533033818782788.
8. Zhang, Guobin, Shan Jiang, Zhiyong Yang, Li Gong, Xiaodong Ma, Zeyang Zhou, Chao Bao, and Qi Liu. "Automatic nodule detection for lung cancer in CT images: A review." *Computers in biology and medicine*, 2018; 103: 287-300.

9. Masood, Anum, Bin Sheng, Po Yang, Ping Li, Huating Li, Jinman Kim, and David Dagan Feng. "Automated decision support system for lung cancer detection and classification via enhanced RFCN with multilayer fusion RPN." *IEEE Transactions on Industrial Informatics*, 2020; 16(12): 7791-7801.
10. Singadkar, Ganesh, Abhishek Mahajan, Meenakshi Thakur, and Sanjay Talbar. "Deep deconvolutional residual network based automatic lung nodule segmentation." *Journal of digital imaging*, 2020; 33(3): 678-684.
11. Wu, Jianrong, and Tianyi Qian. "A survey of pulmonary nodule detection, segmentation and classification in computed tomography with deep learning techniques." *J. Med. Artif. Intell*, 2019; 2(8): 1-12.
12. Lakshmanprabu, S. K., Sachi Nandan Mohanty, K. Shankar, N. Arunkumar, and Gustavo Ramirez. "Optimal deep learning model for classification of lung cancer on CT images." *Future Generation Computer Systems*, 2019; 92: 374-382.
13. ALzubi, Jafar A., Balasubramaniyan Bharathikannan, Sudeep Tanwar, Ramachandran Manikandan, Ashish Khanna, and Chandrasekar Thaventhiran. "Boosted neural network ensemble classification for lung cancer disease diagnosis." *Applied Soft Computing*, 2019; 80: 579-591.
14. Coudray, Nicolas, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Fenyö, Andre L. Moreira, Narges Razavian, and Aristotelis Tsirigos. "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning." *Nature medicine*, 2018; 24(10): 1559-1567.
15. Zhang, C., Sun, X., Dang, K., Li, K., Guo, X.W., Chang, J., Yu, Z.Q., Huang, F.Y., Wu, Y.S., Liang, Z. and Liu, Z.Y., 2019. Toward an expert level of lung cancer detection and classification using a deep convolutional neural network. *The oncologist*, 24(9): 1159-1165.
16. Shakeel, P. Mohamed, M. A. Burhanuddin, and Mohammad Ishak Desa. "Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier." *Neural Computing and Applications*, 2020; 1-14.
17. ALzubi, Jafar A., Balasubramaniyan Bharathikannan, Sudeep Tanwar, Ramachandran Manikandan, Ashish Khanna, and Chandrasekar Thaventhiran. "Boosted neural network ensemble classification for lung cancer disease diagnosis." *Applied Soft Computing*, 2019; 80: 579-591.
18. Prabukumar, M., Loganathan Agilandeeswari, and K. Ganesan. "An intelligent lung cancer diagnosis system using cuckoo search optimization and support vector machine

classifier." *Journal of ambient intelligence and humanized computing*, 2019; 10(1): 267-293.

19. National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). (2018). Radiology Data from the Clinical Proteomic Tumor Analysis Consortium Lung Squamous Cell Carcinoma [CPTAC-LSCC] Collection [Data set]. The Cancer Imaging Archive. <https://doi.org/10.7937/k9/tcia.2018.6emub512>