How to Cite:
Jayaraj, D., Kumar, C. S., Nagarajan, N., Kuma, B. S., Govindasamy, S., Vennila, L., & Panbilnathan, A. (2022). Lung cancer prediction and retrieval using multistage hybrid prediction approach. *International Journal of Health Sciences*, *6*(S2), 13161–13179. https://doi.org/10.53730/ijhs.v6nS2.8484

Lung cancer prediction and retrieval using multistage hybrid prediction approach

D. Jayaraj
Assistant Professor/Programmer, Department of Computer Science and Engineering, FEAT, Annamalai University, Annamalai Nagar-608002, Tamil Nadu, India
Corresponding author email: jayarajvnr@gmail.com

C. Senthil Kumar
Assistant Professor/Programmer, Department of Computer and Information Science, Annamalai University, Annamalai Nagar-608002, Tamil Nadu, India
Email: csenthilkumar77@gmail.com

N. Nagarajan
Assistant Professor/Programmer, Department of Computer Science and Engineering, FEAT, Annamalai University, Annamalai Nagar-608002, Tamil Nadu, India
Email: nnr764@gmail.com

B. Suresh Kuma
Assistant Professor/Programmer, Department of Computer and Information Science, Annamalai University, Annamalai Nagar-608002, Tamil Nadu, India
Email: sureshaucis@gmail.com

S. Govindasamy
Assistant Professor/Programmer
Department of Computer and Information Science, Annamalai University, Annamalai Nagar-608002, Tamil Nadu, India
Email: click2govind@gmail.com

L.Vennila
Assistant Professor, Department of Biochemistry & Biotechnoloigy, Annamalai University Annamalai Nagar-608002, Tamil Nadu, India
Email: vennilajnr@gmail.com

Aranga Panbilnathan
Assistant Professor, Department of Physical Education, Annamalai University Annamalai Nagar-608002, Tamil Nadu, India
Email: apanbil@gmail.com

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
Manuscript submitted: 09 March 2022, Manuscript revised: 18 April 2022, Accepted for publication: 27 May 2022

***Abstract*---**In cancer diagnosis computer-aided Prediction is considered as significant research domain. The generation and processing of lung cancer identification and Prediction is expensive. During past decades, content-based image retrieval (CBIR) technique has been applied in various medical applications. For effective diagnosis of lung cancer radiologist's requires effective approach for cancer prediction and diagnosis. In this research, for prediction of CT lung images is achieved using hybrid Prediction approach is adopted with integration of logistic regression and Adaboost classifier. Feature extraction and retrieval of lung cancer image applied through optimization technique known as firefly approach for processing several vector features of lung images to derive salient characteristics for achieving best features. Results illustrated that proposed prediction and retrieval approach exhibits significant performance rather than conventional technique. The proposed approach provides classification accuracy of 97% which is significantly higher.

***Keywords*---**lung nodule, retrieval, prediction, feature extraction, classifier.

Introduction

In recent years, throughout the world Cancer is considered as severe health problem. According to the survey conducted in 2016, throughout the world cancer is considered as common cause of death [1]. As per NNACCR report, lung cancer exhibits several symptoms such as change in voice, pain in chest, shortage of breath, cough and blood coughing [5]. But this, methods are difficult for accurate detection of cancer even cancer diagnosis, earlier detection of tumor improves the survival rate of patient [2]. Even though experts are subjected to certain difficulties which affect the decision-making process related to cancer. To facilitate the medical professional for cancer diagnosis Computer Aided Diagnosis (CAD) has been useful for professional in process of decision making [3,4]. In the field of medial application CAD system exhibits a promising performance [5,6].Medical imaging technique through CAD system involves the significant examination of lung tissues, soft organs, offering information related to affected part through MRI [7]. Even though screening technique provides prediction of lung cancer but accuracy is highly difficult for processing.

For diagnosis of lung cancer CAD approaches are utilized for inner body details examination, restore details and extraction of valuable information related to lung cancer image diagnosis. In the field of medical, Content based image retrieval offers radiologist for retrieval of images with similar characteristics features which involved in effective diagnosis of applied input images. As mentioned earlier, lung nodule diagnosis is considered as progressive technique for diagnosis of lung cancer since it involves millions of death per year. This causes the requirement of earlier diagnosis of nodule detection for CT images. Few researcher utilizes Content based medical image retrieval technique for diagnosis [8]. In recent years, diagnosis of lung cancer at earlier stage is considered as effective factor through

Computer Aided Diagnosis (CAD) [9, 10] for diagnosis. By application of this CAD system performance of preprocessing improved for lung nodules.

Lung nodule or cancer prediction is performed by using suppression of lung background structure such as ribs, bronchi and blood vessels [9]. The image which capture excellent structure of chest provides nodule region and classification is based on the consideration of shape, size and contrast [12]. In existing, simple rule based prediction is applied but it exhibits higher false positive values to overcome this drawback CAD has been applied widely for lung nodule. At present, various machine learning technique has been evolved for CT image classification such as SVM (support vector machines), KNN (K-nearest neighbor) algorithm, NN (Neural Network), Ensemble and regression models. Even though this technique provides significant classification for lung nodule prediction results are not desire. To overcome this limitation desire results are achieved through certain factors for achieving prediction results by use of multistage and ensemble classifier. Based on this merits and limitation of individual classifier are combined together for improving accuracy [13].

Image retrieval is is achieved effectively got prediction for increasing the retrieval performance and timeliness of larger data set. Generally, image retrieval consists of two basic steps such as feature generation based on the query images and comparison of generation features for the stored datasets. For this images content of image does not offers significant performance for processing and operation. Further this requires appropriate distance distribution between the features of images. To withstand this requirement for learning and prediction of image features Support Vector based machine learning (SVM) has been widely adopted for medical image processing. Based on this several machine learning algorithm are developed such as decision tree, Adaboost and so on are evolved. In case of Medical Image database datasets consists of numerous size, also data sets consists of hundreds and thousands of images in it. For medical image retrieval system traditional system uses categorizer of human brain which means image is assigned with descriptive keyword for retrieval and selection of features in image [14,15]. For retrieval and feature selection of medical images Content Based Medical Images(CBMIR) has been widely adopted but it subjected to certain limitations. The drawbacks associated with CBMIR system are it is computationally expensive and minimal accuracy rather than image insexing technique. This resulted in tradeoff lies between cost of computation and accuracy, this tradeoff can be limited through utilization of efficient algorithm for increasing computation power. In other hand, recent years, optimization algorithm exhibits significant performance for vast number of application such as power electronics, power system, image processing, signal processing and so on [16]. This optimization performs based on the consideration of behaviour of individual for achieving optimal results. In the field of image processing optimization algorithm perform based on consideration of image pixels which will be expected as helpful tool for identification of features and retrieval of medical images.

In this research proposed a novel adaboost integrated with regression classifier for lung nodule cancer prediction. The conventional adaboost classifier involved in identification of strong and week learner here it involved in identification of image pixel those are all unwanted and irregular. Regression classifier involved in conversion of those irregular and unwanted pixels in an CT images of lung nodule. Lung nodule retrieval is performed through optimization approach firefly with identification and feature extracted. Simulation results demonstrated that proposed approach exhibits significant performance rather than conventional technique. Adaboost based regression classifier offers classification accuracy of 97% which is significantly higher than other approaches.

Related Works

Image retrieval in medical applications (IRMA) [22] examined the CBIR approach for medical image processing with consideration of various modalities. IRMA integrates the central database and architecture of distributed system for supporting prototype and integration of efficient analysis of medical images. The algorithm included the different factors for retrieval of image through consideration of region of interest by means of localization, overlapping and irregular factors [20]. For image retrieval several approaches are developed specifically for high-resolution CT (HRCT) images of lungs.

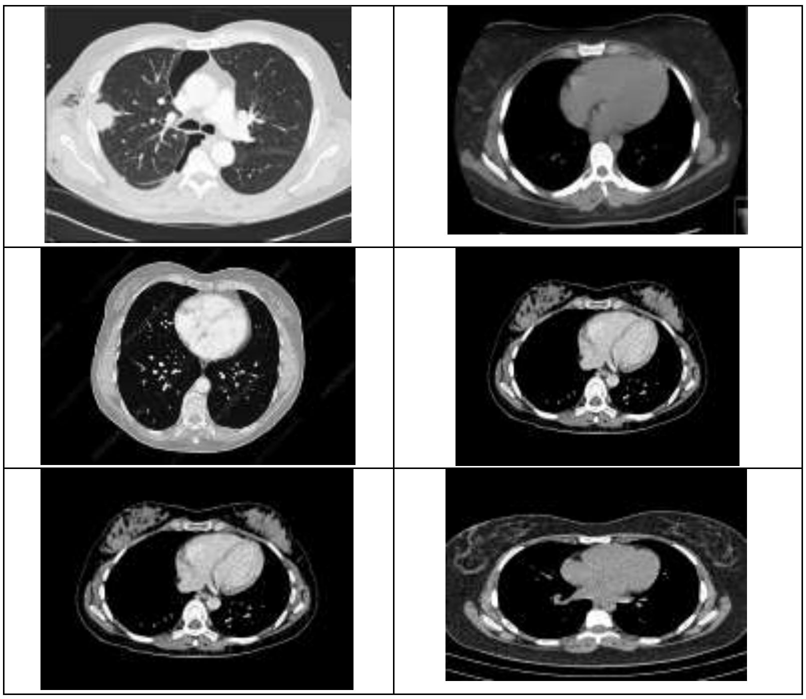
Lung image retrieval system has been examined in several researches and observed techniques are ASSERT [12], medGIFT [13] and comparison is performed using Digital Image Databases (CANDID) system [14]. Through global signature CANDID system evaluate the database of stored images which represents several features such as color, shape and texture. Image visual features such as edges, texture, shape and other pathology bearing characteristics are evaluated using feature-based representation. The retrieval system integrated with open-image identification tool with PACS has been utilized in medDIFT [15]. This system combines the characteristics of labels texture and features exhibited in visual representation of retrieval for various images like MRI, CT and other color images.

Researcher [23] and [24] utilized CBIR approach for retrieval of pulmonary nodules of image. In this [23] integrated system is defined as 'BRISC' for database of CT images obtained through integrated image slices of pulmonary nodules of lung. For every CT images slices of lung nodules various texture features like Gabor filter, Haralick and Markov are extracted for nodule representation. This approach has been applied to 2424 images databases with 141 nodules and provides precision rate of 88% for image retrieval. Researcher [24] designed a system for evaluation of nodules using CBIR approach with rating of malignancy rating from 1 or 2 as benign, 4 or 5 as malignant and 3 as unknown region. In this image 64 features of nodules are extracted with nodule slice representation. For the developed approach similarity measures are performed using Euclidean distance. The measured precision are at the rate of 82.77, 82.57, 81.94, 80.90, and 79.23 % for retrieval of image. Even though this research provides significant retrieval performance it requires manual segmentation of nodule. However, this manual segmentation approach consumes lot of time subjected to higher error rate.

To identify whether the nodule of lung is cancerous or not prediction perform the significant role. For lung cancer Artificial Neural Network offers significant results with accuracy rate of 90% for lung cancer classification [17]. Generally, in many research work SVM model has been utilized for mapping and space for broader range of applications [12]. SVM can be classified in to two classes such as class - 1 and class - 2 in which class - 1 is normal and class - 2 is cancerous. Based on the performance of feature extracted prediction is performed. Some technique utilized for prediction of cancer is feed forward back propagation neural network [25]. Author [26] developed a neural network approach for hybrid classification with integration of fuzzy logic and Multilayer Perceptron (MLP) [11]. Another, researcher [27], presented a approach for segmentation of images using supervised learning method for prediction of cancer in lung [14]. Research conducted by [28] and [29] utilized rule-based classification integrated with SVM and an intelligent and dynamic Prediction method, Intelligent Fuzzy C-means to detect the tumor respectively. Further, [30] also utilized SVM approach for lung cancer prediction. Through analysis it is observed that many prediction approaches uses SVM classifier but this take higher processing time and difficult to compute for larger medical databases.

Data and Methodology

This research utilizes Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI)[6]. This database contains 1018 CT images which is associated with annotation file. Each and every file in this system are differ in its characteristics and quality of image also differs. Through radiologist certain annotation files are created for processing. In below figure 1 sample images for lung cancer has been presented.



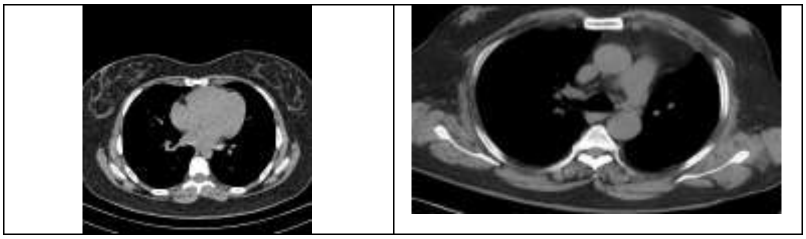


Figure 1. Lung Nodule CT datasets

Pre-processing

The collected data may contain fewer unwanted data which involves the combination of missing values and redundant information which affect the performance of results. The data quality need to be improved before performing lung cancer prediction hence filtering approach is applied in this research to enhance the performance of image retrieval and prediction process [17]. The performance of preprocessing involved in compensation, correction, normalizing and smoothing.

Image Retrieval

Retrieval of image is obtained through the spatial relationship exists between image objects or regions [19 - 20]. Rather it offers interactive system for extracting regions in an image. Generally, image retrieval involved in decomposition of image regions or objects through segmentation approach. The performance of image retrieval has been applied in examination of digital medical images and provides effective storage and retrieval of medical images with consideration of attributes.

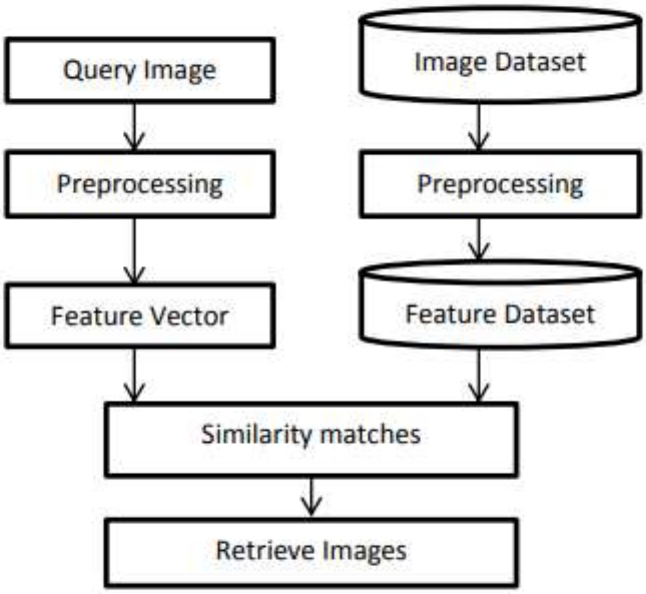


Figure 2. Block of Image Retrieval

Proposed Hybrid Adaboost Regression Classifier

In this research for prediction of lung cancer proposed a classifier technique for accurate prediction of CT lung cancer. The proposed prediction technique involves combination of ada-boost and logistics regression classifier. In this ada-boost involved in attack identification for the selected data set for classification of cancer and non-cancer images of obtained CT images. The other classifier approach logistics regression involved in conversion of 0's and 1's where images may contains fewer values which affect the accuracy of images. The steps involved in adaboost based image processing are presented as follows:

Algorithm: Proposed AFRC
Step 1: Pre- processing of collected CT image dataset. *Step 2: Training subsets are trained for the cancer prediction.* *Step 3: For the trained dataset initial datasets are evaluated.* Step 4: Redundant features are evaluated and removes noisy data. *Step 5: Through the application of rule set cancer and non-cancer images are identified and classified.* *Step 6:The data are tested for identification of cancer.* *Step 7: Evaluation of cancer and non-cancer images in CT imagess* *Step 8: Calculation of* FP, FN, TP and TN parameters.

Generally, the incorporated ada-boost classifier relies on the category of ensemble classifier for identification and prediction of strong and weak classifier. By using binary classification approach training data is used for identification of error and correct values. This approach is generally belongs to the class of machine learning approach which effectively identifies weak learner and improves the overall classification accuracy of the system [21]. The general equation considered in this research for adaboost classifies is presented below:

$$H = sign\left(\sum_t \alpha_t h_t\left(x_t\right)\right)$$

To improve the overall prediction accuracy for cancer and non-cancer classification ada-boost algorithm is combined with logistics regression. Based on the logistics function data were fitted for the CT images where logistics values are ranges from 0 and 1, in this pixel intensity is below or equal to 0.5 than it automatically considered as 0. The general logistics equation is presented below [16]:

$$h_\theta(x) = g\left(\frac{1}{1+e^{-\theta T_x}}\right)$$

In order to enhance the prediction accuracy logistics equation is modified for deriving sigmoidal function with reduction of computational time. This can be denoted as:

$$y = \omega^T x$$

Based on the linear equation model linear function is evaluated for deriving sigmoidal function which has limit of $(-\infty, \infty)$ is

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

Taking probability values for regression

$$P = a_0 + a_1 x_1 + a_2 x_2 + \ldots\ldots + a_k x_k$$

$$\left[\frac{P}{(1-P)}\right] = b_0 + b_1 x_1 + b_2 x_2 + \ldots\ldots + b_k x_k$$

Taking probability on both sides,

$$\log\left(\frac{P}{1-P}\right) = \log\left(\omega^{T} x\right)$$

After applying natural exponential property,

$$\log\left(\frac{P}{1-P}\right) = \sum b_j x_j$$

Where $P = \sum b_j x_j$ hence for logistics regression equation

$$P = \frac{\exp\left(b_j x_j\right)}{\left[1+\exp\left(b_j x_j\right)\right]}$$

For enhancing prediction performance of hybrid approach property of chain rule and maximum likelihood property is applied as below:

$$F'(x) = F' g(x) g'(x)$$

After applying above property and simplification we obtained equation as,

$$P = P(k)\left(1-P(k)\right)$$

Maximum likelihood estimation for P is,

$$\hat{l}(\theta; x) = \frac{1}{n}\sum_{i=1}^{n} \ln f\left(x_i | \theta\right)$$

$$P = \sum \log P(k_i) + \sum \log\left(1-P_i(k_i)\right)$$

Removal of negative term offers,

$$\sum P = \sum P_i$$

Now, $P = \left(a_0 + a_1 x_1 + a_2 x_2 + \ldots\ldots + a_k x_k\right) \sum P_i$

Hence the final equation for proposed AFRC algorithm is,

$$H = Sigmoid\left(\sum_{j=1}^{N} P_i \alpha_t h_t\left(x\right)\right)$$

The above equation presents a final equation derived for prediction of cancer and non-cancer on the CT images with improved accuracy and minimal computational time.

Algorithm 1: Pseudo-code for prediction algorithm.

1. Collection of CT lung image datasets
2. Applying regression value for the images for optimizing images
3. Apply AdaBoost approach for optimized CT lung image datasets
4. Evaluation of cancer in the present datasets
5. Identification of cancer in the framed datasets.
6. Calculation of distance between the every pixel intensity in the images.
7. In case value of present value is higher than previous value {
Consider as CT image with cancer
}
Else {
Consider CT image as non-cancer
8.In case classifier not able to locate the cancer prediction probability
{
Consider random position
}
Else {
Go step 9
}
9. Cancer and non-cancer is evaluated using adabost regression algorithm.
End

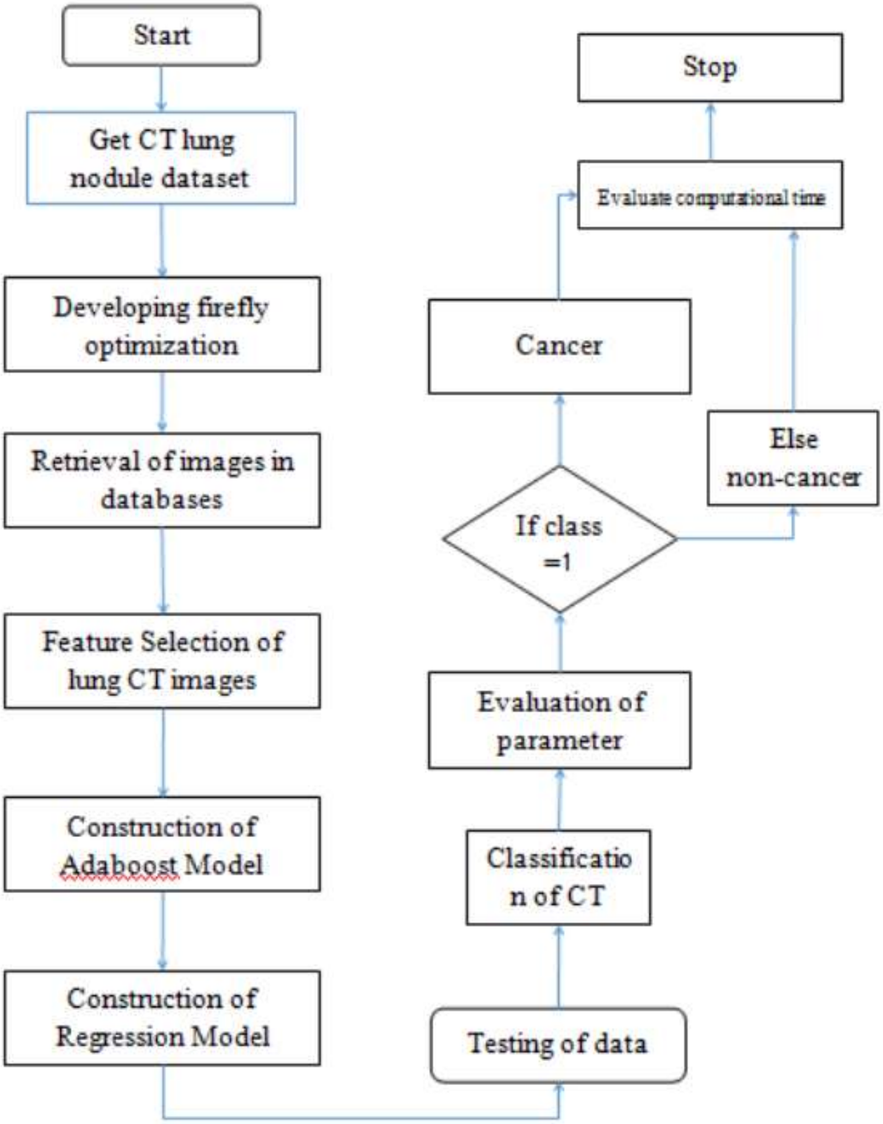


Figure 3. Flow Chart of Adaboost Regression Classifier

Experimental Results Analysis

The performance of adaboost regression classifier is evaluated using consideration of metrics which is examined in this section through using confusion matrix. Prediction or classification subjected to certain error rate even for correctly classified instances which is described as [22]:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{2}$$

Where, True positive value is described as TP; TN is True Negative; False Positive as FP and False Negative as FN. Generally, TP is also known as sensitivity. For any classification instance true positive value must be high hence TP rate is described in equation (2):

$$TP\ Rate = \frac{TruePositive}{ActualPositive} \tag{3}$$

FP denotes the number of positive value described as positive. For effective classifier FP rate should be minimal as denoted in eq (3):

$$FP\ Rate = \frac{FalsePositive}{ActualNegatives} \tag{4}$$

Another factor considered in this research is precision or positive predictive value (PPV). This is used to measure the quality and exactness of the classifier as shown in (4) :

$$\Pr ecision = \frac{True\ Positive}{(True\ Positive + False\ Positive)} \tag{5}$$

Through the evaluation of precision and recall value of the images plays significant role where minimal recall increases the value of FN:

$$\operatorname{Re} call = \frac{True\ Positive}{(True\ Positive + False\ Positive)} \tag{6}$$

To calculate the accuracy of classification Tradeoff value For classification accuracy tradeoff points for data of same classed are evaluated for evaluating class accuracy of every class through following equation:

$$F - measure = 2 \times \left(\frac{\Pr ecision \times \operatorname{Re} call}{\Pr ecision + \operatorname{Re} call} \right) \tag{7}$$

The performance of classifier is estimated using the consideration of FP and FN value of the images.

CT image feature selection and retrieval

Feature selection belongs to the class of image reduction approach which has been used in knowledge discovery and image mining approaches for elimination of redundant features and irrelevant data. Through image feature selection approach irrelevant and redundant features of CT lung images are processed and removed in order to enhance the learning system performance. The conventional technique utilizes SVM approach due to drastic development of Content Based Image Retrieval needs efficient scheme for processing medical databases.

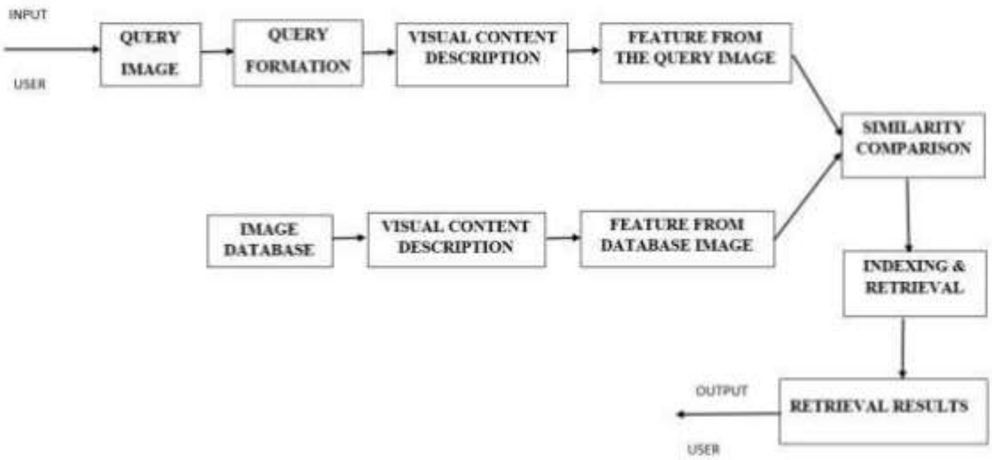


Figure 4. Flow Chart of Image retrieval

As illustrated in above figure feature selection approach is performed prior to image retrieval system. For this images are extracted with respect to query images and database images with computation of distance between database images and query images. This is processed using the conventional optimization algorithm firelfly.

Firefly algorithm is involved in formation of various species of animals in marine through non-random and under-dispersed factor. This firefly is formulated based on the consideration of antarctic studies with consideration of marine animal. Without any aggregation of image pixels firefly hears is orientation are evaluated for consideration of different species with the space scale of 10s to 100 meters. The conceptual model involved in proposed weighted firefly is presented as follows [24]:

Model for KH

To evaluate the multi-objective path identification optimization algorithm considered in this research is based on objective function which is defined as:

$$F(x) = \left[F_1(x), F_2(x)....., F_k(x)\right]^T$$

$$g_j(x) \leq 0; j = 1,,...m$$

In above equation m defines a inequality constraints and k illustrate the objective function for the lung image retrieval.
Through consideration of arbitrary dimension involved in optimization with consideration of search space. The weighted function is defined as:

$$U = \sum_{i=1}^{k} w_i F_i(x)$$

In this weighted function relies on network functionality. The evaluation is based on the consideration of Pareto Optimality condition. The decision is defined as:

$$\frac{dX_i}{dt} = N_i + \tau_i + D_i$$

Where, N_i individual firefly motion induced is denoted; τ_i is the foraging motion, and D_i is the physical diffusion of the ith firefly individuals.

Path formulated by the firefly

Based on the theoretical arguments, firefly with weighted function moves towards higher density pixel with consideration of mutual effects. The induced motion direction and estimated image pixel density, density of query image pixel and repulsive density of CT databases. The movement of firefly is stated as

$$N_i^{new} = N^{\max}\alpha_i + \omega_n N_i^{old}$$

Where, $\alpha_i = \alpha_i^{local} + \alpha_i^{target}$

Where, maximal computation speed is denoted as N^{max}; Direction of firefly is defined as α_i, inertia weight is defined as ω_n; last motion induced is defined as N_i^{old}; Effect of neighbor locally is stated as α_i and target effect of individual firefly is defined as α_i^{target}.

Experimental Analysis

For analysis of collected data related to lung cancer collected lung image datasets LIDC-IDRI has been examined. The simulation analysis is performed through the simulation software MATLAB 2019a with RAM memory of 4GB. The collected images are examined for the evaluation of proposed technique. The analysis of proposed classifier method is evaluated using conventional techniques. The analysis is performed through consideration of MSE and PSNR value of the different technique. For analysis 4 lung images are presented comparatively to evaluate the performance of proposed algorithm technique.

Input Images	
	

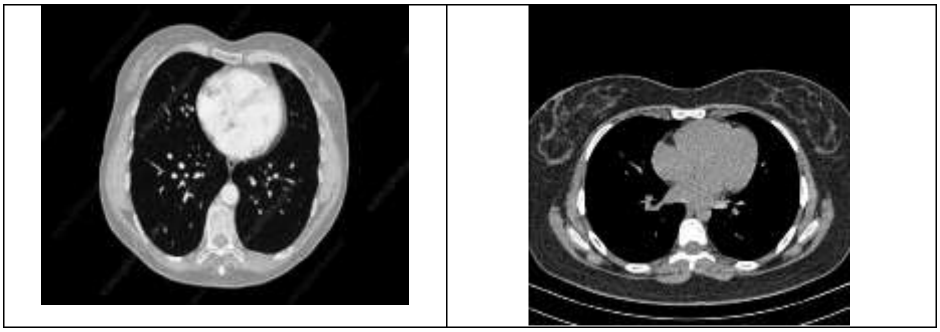


Figure 5. Image Selected as input

This research aimed to perform image retrieval and feature selection using optimization approach. For analysis firefly algorithm is utilized for feature selection of query images and retrieval of images in the medical system. To improve the performance of feature selection for analysis only segmented images are used as query images. The analysis is performed for selected 4 sample images the obtained geometric parameters are presented in below table as follows:

Table 1: Geometric Features of Sample Images

Image	Mean	Entrophy	Calcification degree	Cavitary ratio	Area	Roundness	Rectangle	ASM	CON	IDM
1	0.5864	0.3765	0.1255	0.295	346	0.365	0.458	0.0246	1.252	0.9452
2	0.5795	0.2476	0.1376	0	429	0.468	0.5186	0.0125	1.2856	0.6764
3	0.4685	0.3275	0.2867	0.1346	375	0.567	0.3792	0.1258	0.946	0.4831
4	0.5763	0.2764	0.1894	0.2767	479	0.486	0.4863	0.2556	1.7986	0.6127

In table 1 geometric features associated with selected images are presented with inclusion of geometric features such as area, rectangle, roundness. In other hand morphological characters such as calcification degree and cavitary ratio are presented. The other features such as ASM, CON and IDM belongs to the class of texture features of the lung nodules. Since this research focused on lung nodules, region of interest is obtained. The aim of this approach is prediction of cancer in lung nodule hence for retrieval region with identified pixel intensity based on firefly optimization is selected and applied as query image. The figure illustrate the selected region of interest using optimization approach.

Figure 6. Feature Extraction

Once the optimization is performed query image is applied to the retrieval system for obtaining cancer part of lung nodules. The retrieved images are shown as below in which red box indicates the input images and green boxes denotes the retrieved images.

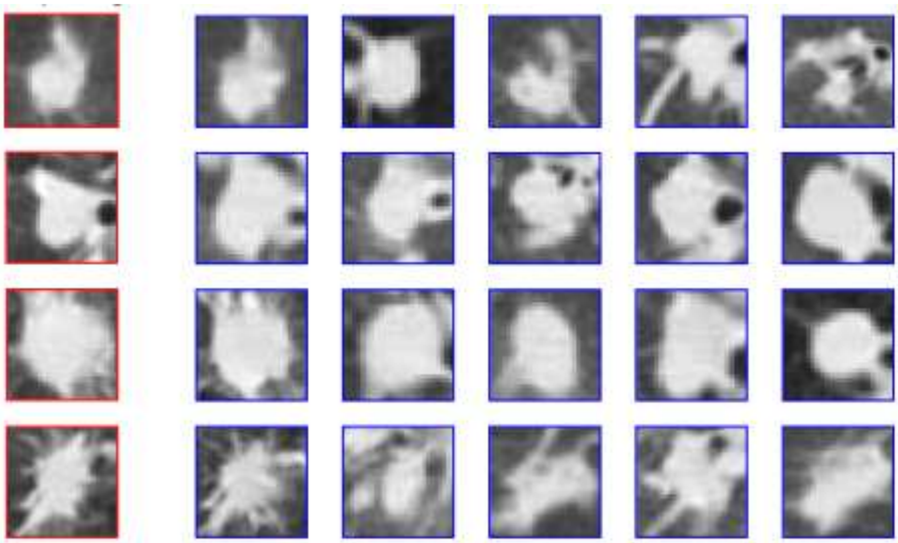


Figure 7. Image Retrieval Process

After the completion of retrieval process next stage is prediction of cancer using the proposed Adaboost based regression classifier. The images those are all retrieved are applied to the prediction system. The classification of lung nodules are performed for selected sample images.

Table 2. Prediction Rate

Images	No. Of Frames	Correct Detection Rate (%)	Error Rate (%)
1	30	97.56	2.44
2	30	96.46	3.54
3	30	98.56	1.44
4	30	96.56	3.44

The classification instances of sample images are presented in above table in which correct detection % is presented the proposed approach offers higher detection rate of 98.56% with error rate of 1.44% which implies the significant classification performance rather than conventional technique. The performance of proposed classification approach is comparatively examined using the other conventional technique such as SVM, KNN and BP.

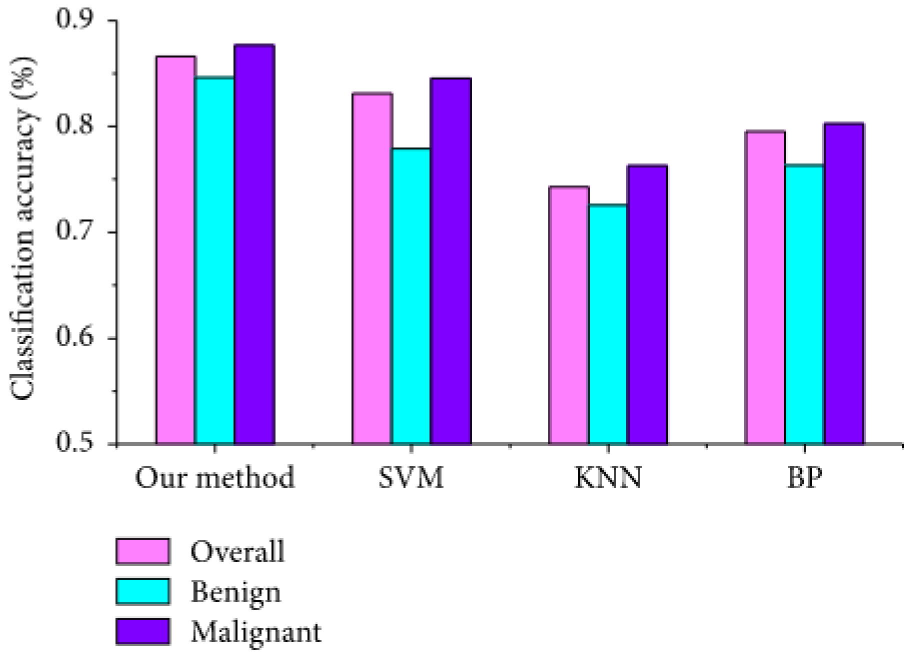


Figure 8. Comparison of Classifier

The comparative analysis of proposed adaboost based classifier with other technique is performed with evaluation of all lung nodule data benign nodule, and malignant nodule accuracy in the test dataset. Convention KNN technique and SVM employees Euclidean distance classification which is not suitable for high-dimensional features of images while BP uses complex performance. The proposed Adaboost based Regression classifier exhibits outperforms rather than other classification approaches. The classification accuracy of proposes approach is reaches around 90% for maligant, 85% for being and 97% for overall lung nodule lesions.

Conclusion

Content based image retrieval offers radiologist for retrieval of images with similar characteristics features which involved in effective diagnosis of applied input images. In this research proposed a novel adaboost integrated with regression classifier for lung nodule cancer prediction. The conventional adaboost classifier involved in identification of strong and week learner here it involved in identification of image pixel those are all unwanted and irregular. Regression classifier involved in conversion of those irregular and unwanted pixels in an CT images of lung nodule. Lung nodule retrieval is performed through optimization approach firefly with identification and feature extracted. Simulation results demonstrated that proposed approach exhibits significant performance rather than conventional technique.

References

1. Siegel, R., Ma, J., Zou, Z., & Jemal, A. (2014). Cancer statistics, 2014. *CA: a cancer journal for clinicians*, *64*(1), 9-29.
2. Sone, S., Li, F., Yang, Z. G., Honda, T., Maruyama, Y., Takashima, S., ... & Yamanda, T. (2001). Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner. *British journal of cancer*, *84*(1), 25.
3. MacMahon, H., Austin, J. H., Gamsu, G., Herold, C. J., Jett, J. R., Naidich, D. P., ... & Swensen, S. J. (2005). Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology*, *237*(2), 395-400.
4. Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, *31*(4-5), 198-211.
5. Kostis, W. J., Reeves, A. P., Yankelevitz, D. F., & Henschke, C. I. (2003). Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. *IEEE transactions on medical imaging*, *22*(10), 1259-1274.
6. Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., ... & Kazerooni, E. A. (2011). The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Medical physics*, *38*(2), 915-931.
7. Han, F., Wang, H., Zhang, G., Han, H., Song, B., Li, L., ... & Liang, Z. (2015). Texture feature analysis for computer-aided diagnosis on pulmonary nodules. *Journal of digital imaging*, *28*(1), 99-115.
8. Armato III, S. G., Giger, M. L., & MacMahon, H. (2001). Automated detection of lung nodules in CT scans: preliminary results. *Medical physics*, *28*(8), 1552-1561.
9. Gomathi, M., & Thangaraj, P. (2010). A computer aided diagnosis system for lung cancer detection using support vector machine. *American Journal of Applied Sciences*, *7*(12), 1532.
10. Wiemker, R., Rogalla, P., Zwartkruis, A., & Blaffert, T. (2002, May). Computer-aided lung nodule detection on high-resolution CT data. In *Medical*

Imaging 2002: Image Processing (Vol. 4684, pp. 677-688). International Society for Optics and Photonics.

11. Zhao, J., Ji, G., Han, X., Qiang, Y., & Liao, X. (2016). An automated pulmonary parenchyma segmentation method based on an improved region growing algorithmin PET-CT imaging. *Frontiers of Computer Science*, *10*(1), 189-200.
12. Ng, G., Song, Y., Cai, W., Zhou, Y., Liu, S., & Feng, D. D. (2014, August). Hierarchical and binary spatial descriptors for lung nodule image retrieval. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 6463-6466). IEEE.
13. Yu, X., Zhang, S., Liu, B., Zhong, L., & Metaxas, D. (2013). Large scale medical image search via unsupervised PCA hashing. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 393-398).
14. Miah, M. B. A., & Yousuf, M. A. (2015, May). Detection of lung cancer from CT image using image processing and neural network. In *2015 International conference on electrical engineering and information communication technology (ICEEICT)* (pp. 1-6). ieee.
15. Vijaya, G., Suhasini, A., & Priya, R. (2014). Automatic Detection of Lung Cancer in CT Images. *IJRET: International Journal of Research in Engineering and Technology*, *3*(7), 182-186.
16. Sharma, N. (2014). Size Estimation of Lung Cancer Using Image Segmentation and Back Propagation. *International Journal for Research in Technological Studies*, *1*(9), 14-17
17. Byra, M., Dobruch-Sobczak, K., Piotrzkowska-Wróblewska, H., & Nowicki, A. (2017). Added value of morphological features to breast lesion diagnosis in ultrasound. *arXiv preprint arXiv:1706.01855*.
18. Spackman, K. A. (1989). Signal detection theory: Valuable tools for evaluating inductive learning. In *Proceedings of the sixth international workshop on Machine learning* (pp. 160-163).
19. Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, *30*(7), 1145-1159.
20. Sakamoto, M., & Nakano, H. (2016). Cascaded Neural Networks with Selective Classifiers and its evaluation using Lung X-ray CT Images. *arXiv preprint arXiv:1611.07136*.
21. Huang, G. B. (2014). An insight into extreme learning machines: random neurons, random features and kernels. *Cognitive Computation*, *6*(3), 376-390.
22. Miah, M. B. A., & Yousuf, M. A. (2015, May). Detection of lung cancer from CT image using image processing and neural network. In *2015 International conference on electrical engineering and information communication technology (ICEEICT)* (pp. 1-6). ieee.
23. McWilliams, A., Tammemagi, M. C., Mayo, J. R., Roberts, H., Liu, G., Soghrati, K., ... & Atkar-Khattra, S. (2013). Probability of cancer in pulmonary nodules detected on first screening CT. *New England Journal of Medicine*, *369*(10), 910-919.
24. Seitz Jr, K. A., Giuca, A. M., Furst, J., & Raicu, D. (2012, February). Learning lung nodule similarity using a genetic algorithm. In *Medical Imaging 2012: Computer-Aided Diagnosis* (Vol. 8315, p. 831537). International Society for Optics and Photonics.

25. Ko, J. P., Berman, E. J., Kaur, M., Babb, J. S., Bomsztyk, E., Greenberg, A. K., ... & Rusinek, H. (2012). Pulmonary nodules: growth rate assessment in patients by using serial CT and three-dimensional volumetry. *Radiology*, *262*(2), 662-671.
26. Tariq, A., Akram, M. U., & Javed, M. Y. (2013, April). Lung nodule detection in CT images using neuro fuzzy classifier. In *2013 Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI)* (pp. 49-53). IEEE.
27. Zhang, Y., Qiang, J. W., Ye, J. D., Ye, X. D., & Zhang, J. (2014). High resolution CT in differentiating minimally invasive component in early lung adenocarcinoma. *Lung Cancer*, *84*(3), 236-241.
28. Valente, I. R. S., Cortez, P. C., Neto, E. C., Soares, J. M., de Albuquerque, V. H. C., & Tavares, J. M. R. (2016). Automatic 3D pulmonary nodule detection in CT images: a survey. *Computer methods and programs in biomedicine*, *124*, 91-107.
29. Thabsheera, A. A., Thasleema, T. M., & Rajesh, R. (2019). Lung cancer detection using CT scan images: A review on various image processing techniques. In *Data Analytics and Learning* (pp. 413-419). Springer, Singapore.
30. Makaju, S., Prasad, P. W. C., Alsadoon, A., Singh, A. K., & Elchouemi, A. (2018). Lung cancer detection using CT scan images. *Procedia Computer Science*, *125*, 107-114.