International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 10 Issue 10, October-2021, Impact Factor: 7.957

Swarm Intelligence Inspired Lung Cancer Segmentation and Classification

Arshak Shan Shajahan

Computer Science Undergraduate, Anna University, Chennai, Tamil Nadu

Author bio - B.E. Computer Science undergraduate student at College of Engineering Guindy, Anna University, Chennai. An Enthusiastic Undergraduate Student eager to contribute to the society through hard work and attention to detail. Motivated to learn, grow and excel in Data Science and Machine Learning. His recent project on Covid Case Prediction using machine learning triggered his interest to contribute more to this field.

ABSTRACT

Cancer is the major cause to increase in the death rate of people in the world. In the medical field, early detection holds a significant role in precise planning of the treatment strategies against any disease to improve the overall survival rates. In the paper, the authors had integrated and evaluated various Swarm Intelligence (SI) approaches to improve lung cancer detection. In this context, the strength of existing SI techniques namely, Artificial Bee Colony (ABC), Firefly Algorithm (FFA), Particle Swarm Intelligence (PSO), and Cuckoo Search Algorithm (CSA) are evaluated at the segmentation stage. Following this, the better-performing CSA segmentation outcomes are fed to CNN for training and classification or used to extract features from the segmented region. The performance analysis of the proposed work is done at two stages, firstly at the segmentation stage to identify the best optimization approach and secondly at the classification stage, to analyse the performance of the proposed segmentation and classification work. The overall analysis against the ELCAP dataset exhibited a classification accuracy of 97.25% with sensitivity of 0.957. The comparative analysis demonstrates that the proposed SI-based lung cancer segmentation and classification work outperformed the existing works.

Keywords: Lung Cancer, Segmentation, Swarm Intelligence, K-means.

INTRODUCTION

Cancer is a very dangerous disease that quickly effected to the human body. According to the WHO, report cancer has become the topmost second disease that increases day by day worldwide death rate. Several types of cancer like; lung, breast, skin, etc. But lung cancer is mostly seen in maximum cancer patients. According to the survey of WHO against cancer, it is concluded that in the year 2020 new 2.21 million cases of lung cancer were predicted where 1.80 million were died due to lung cancer [1]. The reason behind the increased death rate due to lung cancer is that cancer is detected in the advanced stage so the treatment is also started late that did not save the life of patients. This growth might lead to metastasis, known as the adjacent tissue invasion and infiltration ahead of the lungs [2]. The prognosis and treatment are dependent on the histological cancer type, the degree of spread (the stage), and also check patients' performance. Now a days several treatments are available to fight against cancer like surgery, radiotherapy, and chemotherapy [3]. The treatment of cancer depends upon the overall health of the patient who survives against it. It seems that overall only 14% of lung cancer patients were diagnosed and they survived five years after the diagnosis [4]. The common symptoms for lung cancer detection are Hemoptysis (coughing up blood) and Dyspnea (shortness of breath with activity).

Lung cancer is also named 'Bronchogenic Carcinomas' that is further divided into two sub-types that are; SCLC (Small cell lung cancers) and NSCLC (Non-small cell lung cancers) [5]. Figure 1 describes the sub-categories that come under lung cancer. The sub-categories of lung cancer are based upon the tumor cells that are detected through microscopic visualization. These cancer types have peculiar growth, differentiation, and spreading mechanisms and hence required different medical treatment. Therefore, a clear distinction is very significant in planning treatment strategies [6].

SCLC comprises 20% of the total lung cancer types and is known to be quickly spread in the human body. They are strongly linked with the wrong lifestyle such as cigarette smoking while it affects only 1% of non-smokers. SCLC metastasizes very fast and at numerous sites in the body all together before it could be discovered or diagnosed. The specific cell appearances could only be detected with the microscopic examination. NSCLC is considered a common type of lung cancer that covers 80% of all lung cancer types [7].

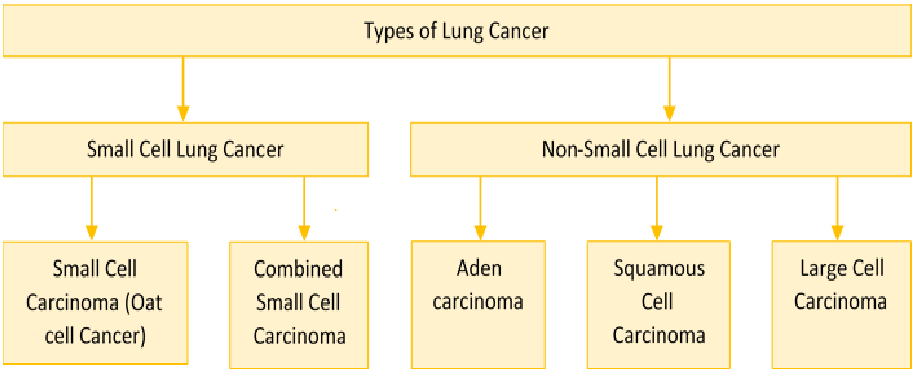


Fig. 1. Various types of Lung Cancer

Visual inspection holds very little significance in this type of cancer and testing is required for precise detection. However, there are risk factors that are related to a person's characteristics, traits, or experience that may increase the chances of developing a particular disease. Many factors such as gender, family history, age, and ethnic background not be modified but are certainly very useful in the prediction of developing the disease shortly [8-9].

Significance of Early Detection

Early detection relates to interventions that can be taken to diagnose cancer as soon as possible when it is easier to treat the disease, as researchers discover what cancer looks like in the earlier stages-under the microscope, in scans, or even on the human body [10]. Early detection in any type of cancer significantly reduces the death rate and saves many lives. For cancer diagnosis firstly check the medical history of the patient and then examine the blood test, biopsy, CT scan, and X-Ray that clarifies the cancer patient stage. To find the lung nodule from the chest is a very difficult task that is performed by the radiologist using chest based radiograph [11]. The lung nodules are the small tissue masses that are observed as small white shadows in both X-ray and CT scans. The high dimensionality of the data also challenges the Statistical analysis. Many research studies given by several authors based on radiologists prove that it is not possible to predict small lung nodules from the chest because images are based on the low trait.

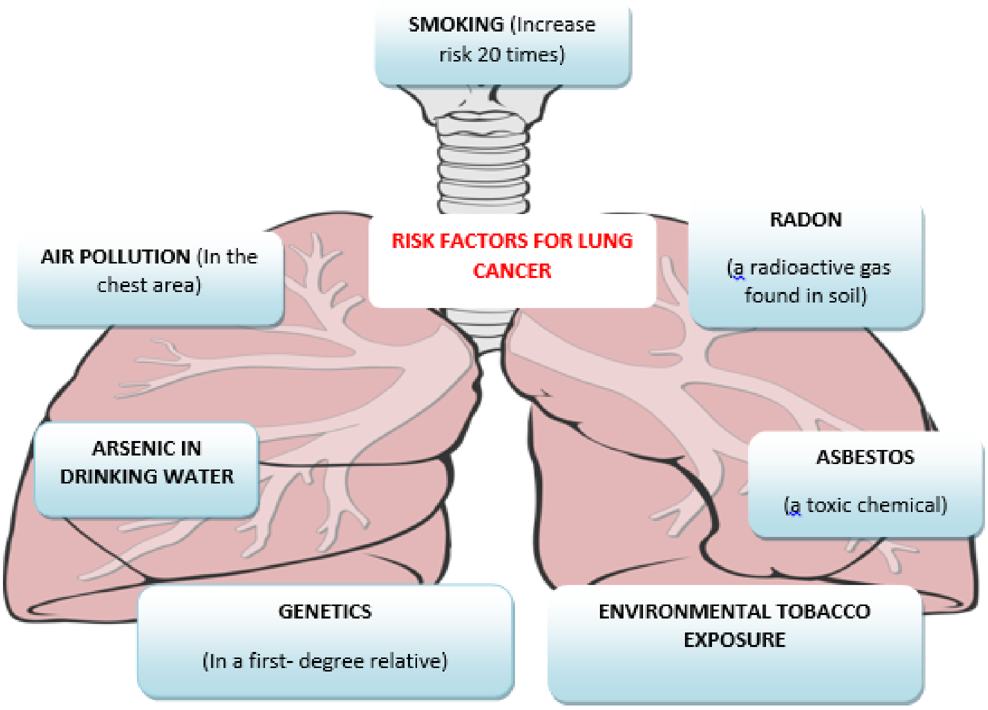


Fig. 2. Risk Factors of Lung Cancer

LITERATURE REVIEW

Cancer is termed a fatal disease because of its rapid progression and because of this, it could only be detected at a very late stage. To reduce the death rate, early detection is essential that can be attained with the integration of machine learning techniques to help automated detection. In 2015, Kumar et al. had implemented a Computer Aided Designing (CAD) using deep encoding features to recognize lung nodules in lung scans. The overall accuracy of the designed technique exhibited a 6% improvement in comparison to the existing work. The work was evaluated against 4323 images from NCI, LIDC datasets that computed an overall accuracy of 75.01%, the sensitivity of 83.35% with a false positive of 0.39 according to the patient [12]. Kurkure and Thakare (2016) also emphasised on developing and designing the initial diagnosis and detection method for CT, PET, and X-ray lung cancer. They implemented a Genetic Algorithm (GA) as natural computing architecture has improved the results, allowing the detection of the lungs' nodules at the early stages of lung cancer. The combination of GA and Naïve Bayes was used for the selection of best features to precisely identify the cancer stage. The simulation analysis had exhibited an 80% reliable classification of lung cancer images based on nodule analysis. [13].

The wavelet feature descriptors were used by Arulmurugan and Anandakumar (2018) for feature extraction of the lung cancer images. Wavelet feature descriptor has been applied for feature extraction by using NN as a classification algorithm. From the total images, 70% of images are used for the training and 30% images are used for the testing. The experimental results of the classifier computed accuracy of 92.6%. [14]. Razmjooy et al. (2018) have introduced a novel methodology that is used to detect malignancy in melanoma by using multiple images. Firstly apply a smoothing and edge detection technique that removed the unnecessary scales from the images. After that perform feature extraction and segmentation on the cancer images. At the end, information of objects that are not required in this task has to be filtered by applying morphological operations and utilized to focus on the area where melanoma boundaries are present. To perform this operation, an algorithm of World Cup Optimization (WCO) has been utilized to optimize an MLP (Multi-Layer Perceptron) Artificial Neural Network (ANN). This proposed scheme WCO is a derivative-free, Meta-Heuristics and global search algorithm however gradient-based back-propagation method is local search. The MLP employs the problems constraint and WCO algorithms have utilized to reduce the root mean square error [15].

Makaju et al. (2018) demonstrated a practical method for detecting and diagnosing lung cancer-related images of CT scan that differentiate into malignant and benign types. The author applied image processing techniques that processed the selected images and after that apply a supervised ML algorithm for the classification of images. Seven different classifiers SVM, K-Nearest, multinomial NB classifier, DT, RF, Stochastic GD, and MLP are used. The dataset is collected from the 15750 clinical images that contain 8840 images of malignant and 6910 images of benign. The experiment shows that the accuracy for the classifier of MLP is more significant with the value of 88.55% in contrast with the other classifiers [16]. Da et al. (2018) presented an FP elimination approach through a deep learning system in tandem with such an evolutionary scheme by using Particle Swarm Optimization (PSO) algorithm that provides the optimization based on hyperparameters of the network into CNN to speed up the system performance that overcomes the use of manual search. The results are computed by using CT scan images datasets from LIDC- IDRI with accuracy (97.62%), sensitivity (92.20%), specificity (98.64%), and the zone underneath the receiver operating characteristics (ROC) curve is 0.9555. Finally, the overall evaluation describes that PSO algorithm provides better performance with CNN parameters based on the classification of nodules and non-nodules [17]. Senth et al. (2019) have evaluated the performance of five different algorithms based on optimization of tumor extraction by using images of lung cancer. The authors targeted the algorithms are; PSO, guaranteed convergence particle swarm optimization (GCPSO), clustering of k-median, clustering of k-means, and inertia-weighted particle swarm optimization. For preprocessing the performance of median, average filters, and adaptive median in the stage has been equated, and it was verified that the filter of the adaptive median is more appropriate for images of medical CT. Moreover, the contrast of the image has been boosted through the equalization of the adaptive histogram.

The pre-processed image with better quality is a matter of four algorithms. The experimental consequences are proved for pictures of 20 samples of the lung via MATLAB, and it was shown that the GCPSO has the maximum accuracy of 95.89% [18]. Prabukumar et al. (2019) implemented an optimized diagnostic method for the prediction of lung cancer nodules early and improving the Fog computing environment's effectiveness. For attaining the high security, low latency, and flexibility aid, the cloud environment is used to store high-volume CT scanned images. Essential aspects of the interest nodule including geometric characteristics, texture, and numerical or intensity characteristics are extracted. And to utilize the Cuckoo search enhancement technique, the best characteristics utilized to identify lung cancer are defined according to the above-selected characteristics. In the end, the authors have trained the SVM classifier through given optimal features, which in turn allows the classification of lung cancer as malignant or benign. The accuracy of this work is evaluated with help of public CT lung images extracted from the Early Lung Cancer Action Program (ELCAP) [19]. A comprehensive survey of various types of systems employed for the early detection various types of cancers including lung cancer was conducted by Patel et al. (2020). They concluded that integration of AI techniques leads to highly accuracy and efficient cancer detection. This significantly helped in early diagnosis and planning effective treatment and therapy. However, lots of challenges and hurdles were found in the existing techniques such as image preprocessing and management of large volume cancer data. Hence, traditional AI techniques need to be enhanced so as to incorporate in automated cancer detection design [20].

METHODOLOGY

In the present article, authors had proposed an automatic lung cancer segmentation approach using lung CT scan images. The paper evaluates the effectiveness of various existing Swarm Intelligence (SI) approaches, namely, Artificial

Bee Colony (ABC), Particle Swarm Optimization (PSO), FireFly Algorithm (FFA) and Cuckoo Search Algorithm (CSA) to improve the k-means based segmentation. The features of the k-mean optimized segmentation region are then fed to Convolutional Neural Network (CNN) for training and classification. The overall all methodology steps are summarized in Figure. The performance of the segmentation and the classification architecture is evaluated in terms of performance parameters, namely, precision, sensitivity and accuracy.

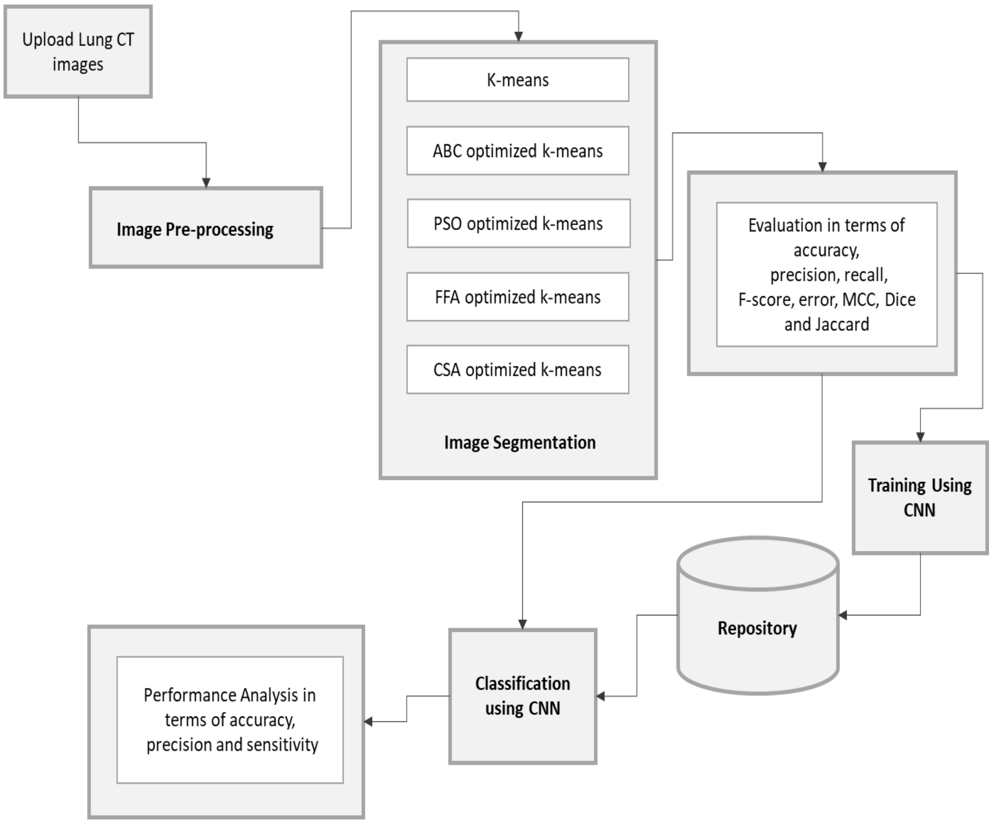


Fig. 3. Outlined steps of the proposed methodology

Dataset

The dataset used for the evaluation of the proposed segmentation approaches is obtained from ELCAP. The dataset is a collaboration between two research groups, namely, ELCAP and VIA. This online repository comprises of image set of 50 low dose images documented with the whole lung CT scan images for lung cancer detection [21-22]. The images have a slice of 1.25 mm thickness and are annotated by the radiologists for the location of lung nodules within the CT scan image. In the present paper, the labelled lung nodule images are used for the evaluation of the proposed SI-inspired lung cancer classification.

Pre-processing of Dataset

The quality of CT scan images is improved by reducing the various type of noise that incorporate during scanning process. The morphological operations applied on the lung cancer image in pre-processing stage is color conversion. It converts the RGB image into gray scales image using following relationship.

$$L_{Img}\,(Gray) = 0.299 * \, L_{img}\,(R) + 0.587 \, * \, L_{img}\,(G) + 0.114 \, * \, L_{img}\,(B)$$

Where, $L_{Img}\,(Gray)$ represents the gray scale image, $L_{img}\,(R)$, $L_{img}\,(G)$ and $L_{img}\,(B)$ represents the red, green, and blue components of the image are respectively. It is observed that when image enhancement is applied on the image it is equally applied on all the pixel values of the image. Therefore, to customize the regions that need to be enhanced the authors had implemented a limited contrast-based image enhancement strategy.

Image Enhancement using Limited Contrast

The image enhancement is used to improve the better quality of the input image to make certain features to be better visualized. Intensity enhancement is a type of image adjustment in which the intensity of the image is mapped into new range to construct a better quality image. In the proposed image enhancement technique, the intensity of the lung cancer image is selectively improved. To achieve this, first of all the upper and lower pixel values of the image are determined. In a gray scale image, the lower and upper limits might be 0 and 255 that are represented as L and H, respectively. The simplest way to sort the image pixel is to find the lowest and highest pixel values present in the image under study and which needs enhancement. These pixels values here are represented as L_N and H_N. Then each pixel $L_{Img}\,(Gray)$ of lung image is enhanced using the following equation:

$$L_{enh} = \left(L_{Img} - L_N\right)\left(\frac{H - L}{H_n - L_N}\right) + L$$

Where P_{enh} is denoted as an enhanced pixels P_{image} is used to define the original image of lung cancer. For detailed description following figure describes both low-intensity images and enhance images with their histogram by applying the intensity enhancement technique.

Intensity Enhancement Algorithm

1. $Input$: L_{Img} // enhanced lung cancer image
2. $Calculate\ the\ number\ of\ dimension\ (Dia)$
3. $If\ D == 3$ // for three frames of the image
4. $L_R = Red\ Part\ of\ L_{Img}$
5. $L_G = \ Green\ Part\ of\ L_{Img}$
6. $L_B = \ Blue\ Part\ of\ L_{Img}$
7. $Using\ enhancement\ equation\ L_{enh} = \left(L_{Img} - L_N\right)\left(\frac{H-L}{H_n-L_N}\right) + L$
8. $Red\ Enhanced = L_{enh}\ (L_R)$
9. $Green\ Enhanced\ = L_{enh}\ (L_G)$
10. $Blue\ Enhanced\ = L_{enh}\ (L_B)$
11. $Enhanced\ Image = cat\ (3, Red\ Enhanced, Green\ Enhanced, Blue\ Enhanced)$
12. $Else$
13. $Img_{lung}\ = \ L_{enh}\ (L_{Img})$ // image enhancement of the complete image
14. End_{if}
15. $Return$: Img_{lung} //Enhanced Lung cancer Image

The intensity enhancement approach is used to increase the intensity of the image by converting each grayscale value into a new one. It is used a cumulative distribution function that is derived from the histogram.

Segmentation

The segmentation is performed using k-means to divide the lung nodule image into foreground and background image. To improve the segmentation using k-means SI approaches are integrated at this stage. The segmentation is performed using k-means, k-means optimized with ABC, k-means optimized with PSO, k-means optimized using FFA and k-means optimized using CSA.

K-means based segmentation

The unsupervised clustering is applied using k-means to distinguish foreground and the background of the image. This step distinguishes the Region-of-Interest (RoI) from the irrelevant part of the image. In the process, k-number of clusters are made from the input image of size x by y. The steps followed in the clustering process are summarized as follows:

K-means for segmentation

1. Input: Img_{lung} // input the lung CT scan image from the dataset
2. Determine the size of the Input Image
3. $[R, C] = size\ (Img_{lung})$ // size of the input lung image is expressed in the form of number of rows (R) and columns (C)
4. Initiate loop for iteration purpose to analyse all pixel values
5. $For_{each}\ i\ in\ R$ // read all values present in the row and iteratively works till the last value present in the row
   $For_{each}\ j\ in\ C$ // reads all values present in the column and iteratively works till the last value is present in the column
      $if\ Pix.dist_{i,j}\ = \ Forground_{(i,j)}$ // Implement the condition statement to distinguish the background and the foreground pixels
      $RoI_1\ = \ Forground_{(i,j)}$ // assign the pixel to the foreground cluster
      Else: $RoI_2\ = \ Background_{(i,j)}$ // assign the pixel to the background cluster
      End_{if} // close the condition statement
   End_{for}
6. End_{for} /// close the loop when all the pixels have been analysed and clustered accordingly
7. $Img_{RoI}\ = \ \min(\ RoI_1 RoI_2)$ // identify the pixels that forms the segmented region
8. Returns: Img_{RoI} // The image represents the foreground cluster representing the segmented region using k-means based clustering

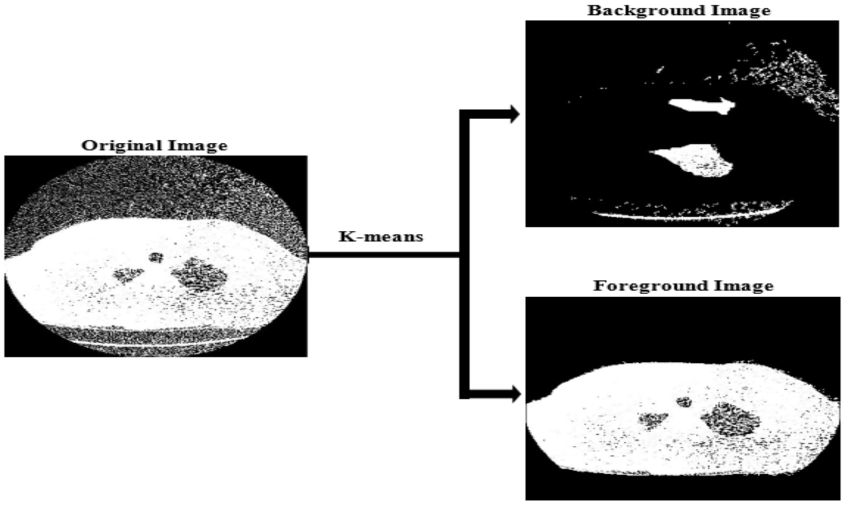


Fig.4. Segmentation using K-means

The above algorithm clusters the image by analysed the pixels. K-means analyses the each and every pixel of the image and divides it into two clusters where RoI_1 represents the foreground cluster and RoI_2 representing background cluster. The visualization of the segmented results using k-means is shown in Figure 4.

It is observed that the foreground and the background data shows mixed-up pixels and irregular patterns for the segmented cancer nodules are obtained. Using such data for classification will not result in high classification accuracy, Therefore, SI approaches are used to resolve the issue by optimizing the k-means segmentation results.

K-means optimized using CSA

The CSA is one of the SI approach used to optimize k-means based segmentation of the lung cancer image. The optimised k-means segmentation results had shown that CSA optimized k-means outperformed various SI algorithms namely, ABC, PSO, FFA and CSA is illustrated by the visual differences in the lung cancer image shown in Table 1. The cuckoo follows hatchery production and proved to be very efficient in searching for the best fit pixels. The algorithmic steps used for the CSA based optimization are as follows:

CSA optimization on k-means segmented

1. Input: Img_{RoI} // segmented region using k-means based clustering
2. Initialize CSA parameters
3. E_{size} // it gives the egg size of the cuckoo
4. E_{other} // it gives the egg size of other than cuckoo's
5. $Intitialize\ OImg_{lung}\ \ an\ emply\ matrix\ for\ storing\ the\ optimized\ segmented\ image\ data$
6. $For_{each}\ \ i\ pixel\ in\ the\ segmented\ image\ Img_{RoI}$

   $E_{curr}\ =\ Img(i)$ // properties of the ith image pixel

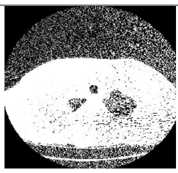
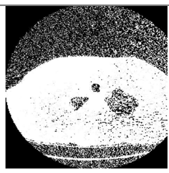
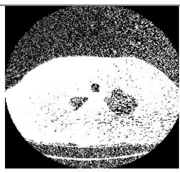
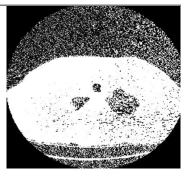
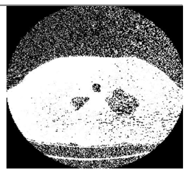
   $E_{th}\ =\ Img(i)$ // threshold properties of the ith image pixel

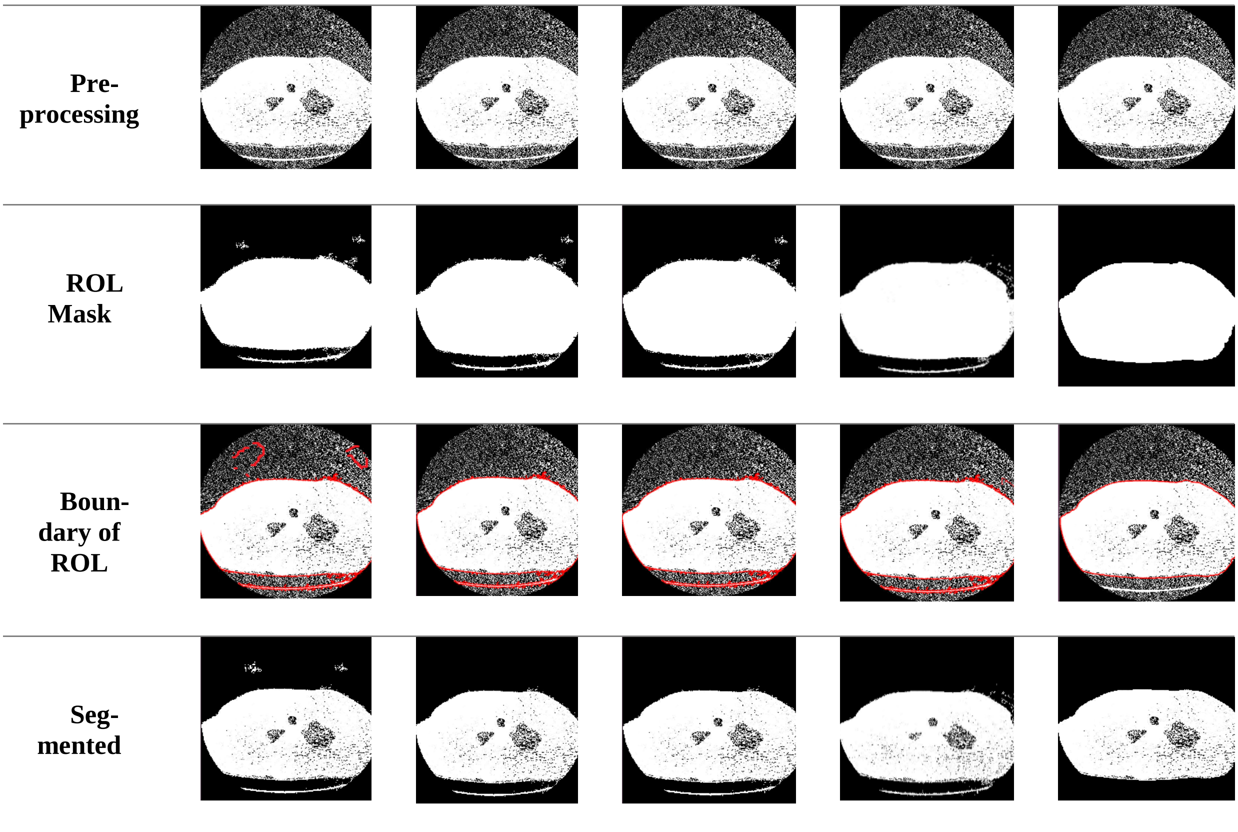
   $Call\ CSA\ fitness\ funtion$

   $$F_{fit} = \begin{cases} 1\ (True); & if\ E_{curr} < E_{th} = Other\ Threshold_{Properties} \\ 0\ (False); & Otherwise \end{cases}$$

   $OImg_{lung}\ =\ CSA\left(F_{fit}\ ,Img_{RoI}\right)$
7. End_{for}
8. $Return,\ OImg_{lung}$ // optimized segmented lung nodule image
9.

Table 1. Comparative analysis of Segmentation Techniques

Images	K-means	K-means with PSO	K-means with ABC	K-means with FFA	K-means with CSA
CT-scan					



The above algorithm initializes the CSA with the k-means segmented RoI image. The pixel values present in the foreground cluster are then used as the number of cuckoo eggs. The CSA is implemented to optimize the pixels for accurate and precise location of pixels to the region it actually belong Table 1 presents the comparative analysis of the segmentation performed using different techniques and table 2 verifies the implementation of CSA as the most effective in performed segmentation.

Classification using CNN

It is observed that the k-mean optimized using CSA resulted in precise segmentation of the lung cancer nodules in comparison of other optimization approaches. Therefore, the optimized segmented region is used for the training and classification using neural network architecture. The Convolutional Neural Network (CNN) comprises of several convolutional and pooling layers. In the present work, training data is distributed into groups comprising of normal and the cancer data. The query segmented lung cancer image is used at the input layer and probability of prediction into any of the defined classes is returned by the output layer. The steps involved in training and testing are given below.

CNN for training and classification

1. Input: $OImg_{lung}$ // optimized segmented lung nodule image
2. $Initialize\ variables$
3. Gp // number of groups for the input data
4. N // number of neurons
5. Initialize CNN parameters
6. E // number of epoches
7. N // number of neurons
8. $Performance_{parameters} \rightarrow Cross\ Entropy, Gradient, Mutation\ and\ Validation$
9. $Training_{technique} \rightarrow Scaled\ Conjugate\ Gradient$
10. $Data_{division} \rightarrow random$
11. $for_{each}\ i\ in\ OImg_{lung}$

    $if\ (OImg_{lung}(i)\ belongs\ to\ normal_{cat})$

    $Gp_1 = OImg_{lung}(i)$

    $Elseif\ (OImg_{lung}(i)\ belongs\ to\ cancer_{cat})$

    $Gp_2 = OImg_{lung}(i)$

    End_{if}

    Initialize CNN with $OImg_{lung}\ and\ Gp$ //initialize CNN based on training data and groups

    $net = patternet(N)$

    Set parameters and train system

$Net = train(net, OImg_{lung}, Gp)$
$Sim_{results} = simulate(net, OImg_{lung_{properties}})$
$if\ Sim_{results} == True$
$show\ Sim_{results}\ in\ terms\ of\ lung_{cancer}\ classes$
$Compute\ performance$ // performance in terms of sensitivity and accuracy
End_{if}

12. End_{for}
13. Return: $Sim_{results}\ and\ performance$ // performance for the classification of cancer and normal classes is computed

RESULTS AND ANALYSIS

The performance of the proposed SI inspired segmentation and classification is analysed at two stages, at the segmentation and classification stage. Table 2 summarises the performance analysis of k-means and optimized k-means based segmentation performed using PSO, ABC, FFA and CSA in terms of precision, recall, f-score, accuracy and error.

Table 2. Performance analysis of the segmentation approaches

K-means					
Sample Size	Precision	Recall	F-score	Accuracy (%)	Error (%)
10	0.840	0.820	0.830	83.336	16.664
20	0.858	0.830	0.844	83.633	16.367
50	0.870	0.837	0.853	84.227	15.774
100	0.872	0.842	0.857	84.673	15.327
200	0.882	0.844	0.863	86.118	13.882
500	0.882	0.859	0.870	86.387	13.613
1000	0.889	0.863	0.876	87.165	12.849
Average	0.871	0.842	0.856	85.077	14.925
K-means with PSO					
10	0.876	0.861	0.868	88.747	11.253
20	0.890	0.863	0.876	88.805	11.196
50	0.893	0.866	0.879	89.456	10.544
100	0.894	0.867	0.881	90.417	9.583
200	0.913	0.868	0.890	90.579	9.421
500	0.918	0.869	0.893	90.628	9.372
1000	0.918	0.871	0.894	91.427	8.573
Average	0.900	0.867	0.883	90.008	9.992
K-means with ABC					
10	0.901	0.895	0.898	90.918	9.082
20	0.909	0.904	0.906	91.061	8.939
50	0.911	0.907	0.909	91.889	8.111
100	0.916	0.911	0.913	93.239	6.761
200	0.920	0.914	0.917	94.199	5.801
500	0.937	0.919	0.928	94.629	5.371
1000	0.949	0.926	0.937	94.890	5.110
Average	0.920	0.911	0.916	92.975	7.025
K-means with FFA					
10	0.931	0.925	0.928	94.027	5.973
20	0.937	0.933	0.935	94.502	5.498
50	0.938	0.938	0.938	94.695	5.305
100	0.944	0.939	0.941	95.696	4.304
200	0.950	0.940	0.945	95.704	4.296
500	0.954	0.942	0.948	96.136	3.864
1000	0.959	0.950	0.954	96.145	3.855

Average	0.945	0.938	0.941	95.272	4.728
K-means with CSA					
10	0.922	0.921	0.921	96.055	3.945
20	0.923	0.929	0.926	96.337	3.663
50	0.948	0.938	0.943	96.620	3.380
100	0.948	0.943	0.946	96.718	3.282
200	0.953	0.945	0.949	97.426	2.574
500	0.960	0.965	0.963	97.708	2.292
1000	0.977	0.968	0.972	97.821	2.179
Average	0.947	0.944	0.946	96.955	3.045

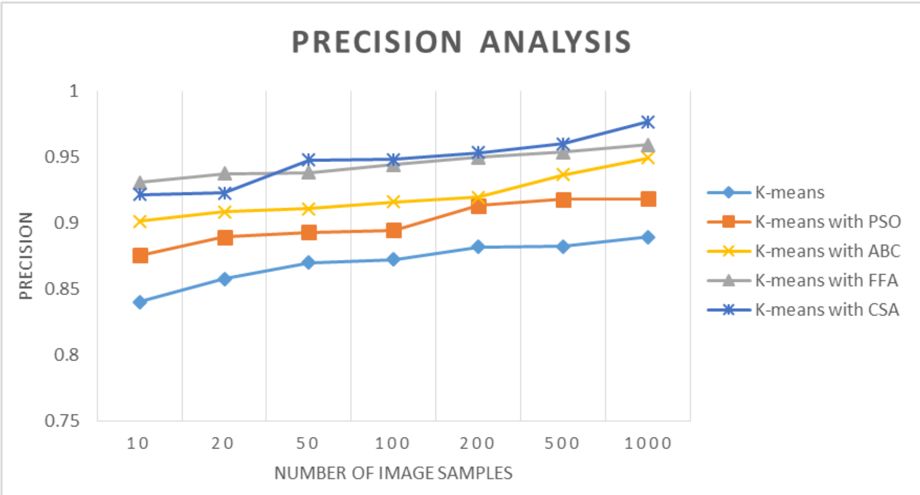


Fig. 5. Precision analysis of segmentation results

The precision analysis of the lung cancer segmentation performed using different segmentation approaches is shown in Figure 5. The number of samples used in the analysis ranges from 10 to 1000. It is observed that k-means optimized using CSA exhibited highest precision over all the image samples with an average value of 0.947 as compared to k-means, k-means with PSO, k-means with ABC and k-means FFA. Similar trend was also observed using various segmentation approaches for recall analysis as illustrated in figure 6. The highest average recall of 0.944 was observed using k-means with CSA.

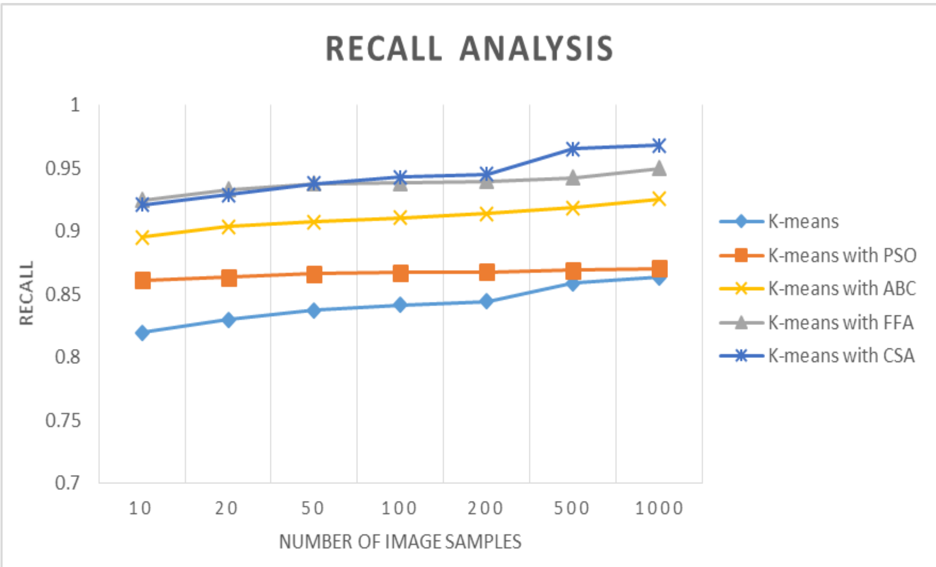


Fig. 6. Recall analysis of segmentation results

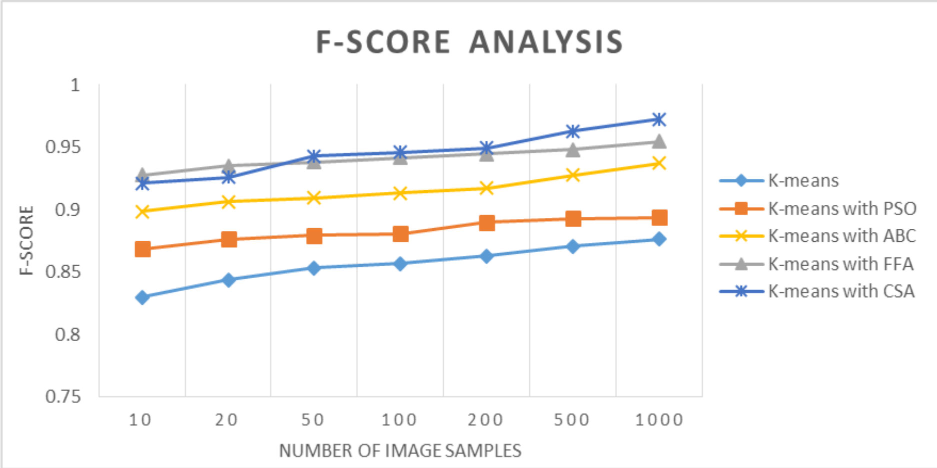


Fig. 7. F-score analysis of segmentation results

F-score is the harmonic mean of precision and recall. The computed f-score for the k-means and k-means optimize using various SI techniques is depicted in figure 7. It is observed that k-means with CSA exhibited higher f-score over all the variation in the number of image samples used for the segmentation analysis. The average f-score observed using k-means with CSA is observed to be 0.946.

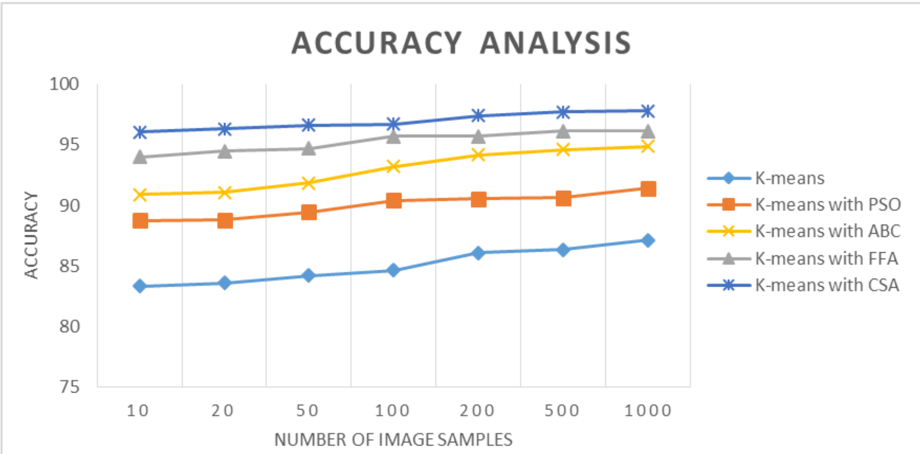


Fig. 8. Accuracy analysis of segmentation results

The segmentation accuracy of the lung cancer image is comparatively analysed in figure 8 and error percentage in figure 9. It is observed that the using k-means optimized using CSA exhibited highest classification accuracy of 96.95% with a least average segmentation error of 3.045% in comparison to other approaches.

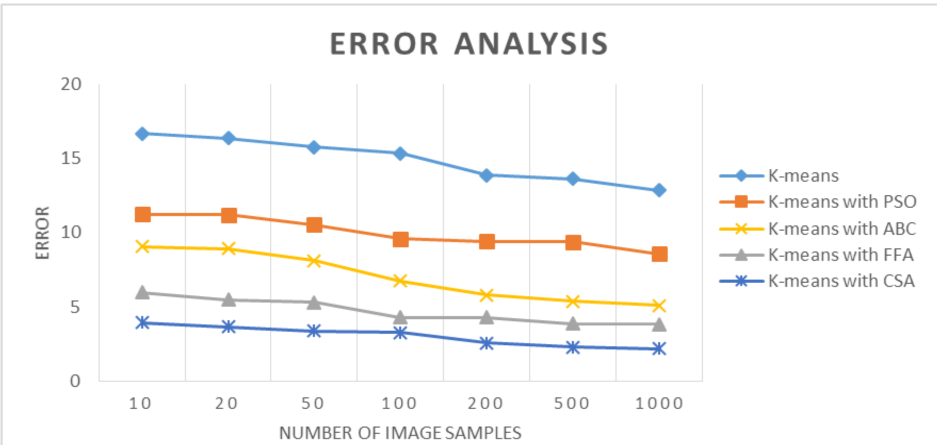


Fig. 9. Error analysis of segmentation results

The performance of the CNN based classification using k-means optimized with CSA is further evaluated to in terms of sensitivity and accuracy. The work is compared against the existing work of Nanglia et al. (2020)who had implemented Genetic Algorithm (GA) with Feed Forward and Back Propagation Neural Network (FFBP-NN) for lung cancer segmentation and classification [23]. The comparative analysis of the CNN based classification is summarized in Table 3 in terms of classification accuracy and sensitivity.

Table 2. Comparative analysis of the lung cancer classification

Sample Size	Accuracy		Sensitivity	
	Proposed	Nanglia et al., 2020	Proposed	Nanglia et al., 2020
10	96.05	94.753	0.923	0.914
20	96.34	94.984	0.934	0.924
50	96.62	95.641	0.946	0.936
100	96.72	95.872	0.953	0.944
200	97.43	96.784	0.968	0.949
500	98.48	97.083	0.984	0.973
1000	99.14	97.371	0.993	0.984
Average	97.254	96.070	0.957	0.946

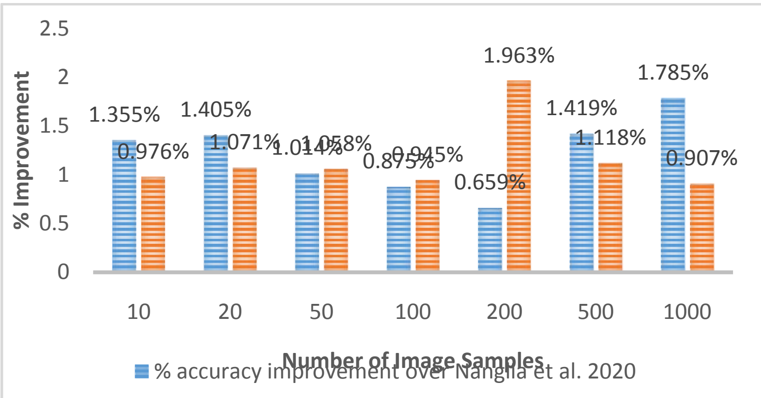


Fig. 9. % improvement of the proposed work over existing work

The % improvement of the proposed CNN based architecture is graphically illustrated in Figure 10. It is observed that the average improvement of the proposed CSA and CNN based architecture over existing work of Nanglia et al., 2020 is 1.216% and 1.148% in terms of classification accuracy and sensitivity.

CONCLUSION

The paper had addressed the lung cancer segmentation followed by classification challenge. The segmentation is performed using unsupervised k-means clustering which is improved with the integration of various SI approaches. This comparative analysis at the segmentation shows that k-means optimised with CSA outperformed the other SI approaches in terms of precision, recall, f-measure and segmentation accuracy and error. Based on this observation, classification of lung cancer images into normal and cancerous is performed using CSA optimized segmented images. At classification stage, CNN is used for the training and classification in the proposed work to exhibit an average accuracy of 97.25%. The comparative analysis of the proposed work against the existing work using 1000 lung cancer images exhibited a % improvement of 1.21% in terms of classification accuracy and 1.145 in terms of sensitivity. In future the possibilities exists in the variation of fitness function by involving more recent SI approaches such as Whale Optimization Algorithm (WOA) to further improve the segmentation results that could further improve the classification accuracy of the proposed work.

REFERENCES

[1]. World Health Organization. Cancer fact sheet, 2010. http://www. who.int/mediacentre/factsheets/fs297/en/. Accessed May 30, (2016).

[2]. Huang, X., Shan, J., & Vaidya, V. : Lung nodule detection in CT using 3D convolutional neural networks. In 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017) pp. 379-383, IEEE, (2017).
[3]. Al-Tarawneh, M. S.: Lung cancer detection using image processing techniques. Leonardo Electronic Journal of Practices and Technologies, vol. 11, no. 21, pp. 147-58. (2012).
[4]. Chaudhary, A., & Singh, S. S.: Lung cancer detection on CT images by using image processing. In 2012 International Conference on Computing Sciences pp.142-146, IEEE, (2012).
[5]. Virmani, A. K., Fong, K. M., Kodagoda, D., McIntire, D., Hung, J., Tonk, V., & Gazdar, A. F.: Allelotyping demonstrates common and distinct patterns of chromosomal loss in human lung cancer types. Genes, Chromosomes and Cancer, vol. 21, no. 4, pp. 308-319, (1998).
[6]. Wu, A. H., Henderson, B. E., Thomas, D. C., & Mack, T. M.: Secular trends in histologic types of lung cancer. Journal of the National Cancer Institute, vol. 77, no. 1, pp. 53-56, (1986).
[7]. Tian, R. H., Zhang, Y. G., Wu, Z., Liu, X., Yang, J. W., & Ji, H. L.: Effects of metformin on survival outcomes of lung cancer patients with type 2 diabetes mellitus: a meta-analysis. Clinical and Translational Oncology, vol. 18, no. 6, pp. 641-649, (2016).
[8]. Alam, N., Park, B. J., Wilton, A., Seshan, V. E., Bains, M. S., Downey, R. J.,& Amar, D.: Incidence and risk factors for lung injury after lung cancer resection. The Annals of thoracic surgery, vol. 84, no.4, pp. 1085-1091, (2007).
[9]. Fink-Neuboeck, N., Lindenmann, J., Porubsky, C., Fediuk, M., Anegg, U., Maier, A., & Smolle-Juettner, F. M.: Hazards of recurrence, second primary, or other tumor at ten years after surgery for non–small-cell lung cancer. Clinical lung cancer, vol. 21, no. 4, pp. 333-340, (2020).
[10]. Blandin Knight, S., Crosbie, P. A., Balata, H., Chudziak, J., Hussell, T., & Dive, C. : Progress and prospects of early detection in lung cancer. Open biology, vol. 7, no. 9, 170070, (2017).
[11]. Bhattacharjee, A., & Majumder, S.: Automated computer-aided lung cancer detection system. In Advances in Communication, Devices and Networking, pp. 425-433,Springer, Singapore, (2019).
[12]. Kumar, D., Wong, A., & Clausi, D. A. : Lung nodule classification using deep features in CT images. In 2015 12th Conference on Computer and Robot Vision, pp.133-138, IEEE, (2015).
[13]. Kurkure, M., & Thakare,: A. Lung cancer detection using genetic approach. In International Conference on Computing Communication Control and automation (ICCUBEA) pp. 1-5, IEEE, (2016)
[14]. Arulmurugan, R., & Anandakumar, H. : Early detection of lung cancer using wavelet feature descriptor and feed forward back propagation neural networks classifier. In Computational Vision and Bio Inspired Computing Springer, Cham, pp. 103-110, (2018).
[15]. Razmjooy, N., Sheykhahmad, F. R., & Ghadimi, N. (2018). : A hybrid neural network–world cup optimization algorithm for melanoma detection. Open Medicine, vol. 13, no. 1, pp. 9-16, (2018).
[16]. Makaju, S., Prasad, P. W. C., Alsadoon, A., Singh, A. K., & Elchouemi, A.: Lung cancer detection using CT scan images. Procedia Computer Science, 125, pp. 107-114, (2018).
[17]. da Silva, G. L. F., Valente, T. L. A., Silva, A. C., de Paiva, A. C., & Gattass, M. : Convolutional neural network-based PSO for lung nodule false positive reduction on CT images. Computer methods and programs in biomedicine, 162, pp. 109-118, (2018)
[18]. Senthil Kumar, K., Venkatalakshmi, K., & Karthikeyan, K.: Lung Cancer Detection Using Image Segmentation by means of Various Evolutionary Algorithms. Computational and mathematical methods in medicine, (2019).
[19]. Prabukumar, M., Agilandeeswari, L., & Ganesan, K.: An intelligent lung cancer diagnosis system using cuckoo search optimization and support vector machine classifier. Journal of Ambient Intelligence and Humanized Computing, vol. 10, no. 1, pp. 267-293, (2019).
[20]. Patel, D., Shah, Y., Thakkar, N., Shah, K., & Shah, M. : Implementation of artificial intelligence techniques for cancer detection. Augmented Human Research, vol. 5, no. 1, pp. 1-10, (2020).
[21]. Accessed online from https://www.nationaljewish.org/NJH/media/img/stock/lung-infographic-full.JPG
[22]. Dataset URL: http://www.via.cornell.edu/lungdb.html accessed on 2021
[23]. Nanglia, P., Mahajan, A. N., Rathee, D. S., & Kumar, S. : Lung cancer classification using feed forward back propagation neural network for CT images. International Journal of Medical Engineering and Informatics, vol. 12, no.5, pp. 447-456, (2020).