



Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks



P. Mohamed Shakeel ^{a,*}, M.A. Burhanuddin ^b, Mohamad Ishak Desa ^a

^a Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia

^b Advanced Manufacturing Centre, Universiti Teknikal Malaysia Melaka, Malaysia

ARTICLE INFO

Article history:

Received 26 February 2019

Received in revised form 29 April 2019

Accepted 8 May 2019

Available online 15 May 2019

Keywords:

Image denoising

Clustering

Improved profuse clustering technique (IPCT)

ABSTRACT

Automatic lung disease detection is a critical challenging task for researchers because of the noise signals getting included into creative signals amid the image capturing process which may corrupt the cancer image quality thusly bringing about the debased performance. So as to evade this, Lung cancer preprocessing has turned into an imperative stage with the key parts as edge detection, lung image resampling, lung image upgrade and image denoising for improving the nature of input image. Image Denoising is a critical pre-processing task preceding further preparing of the image like feature extraction, segmentation, surface examination, and so forth which eliminates the noise whereas retaining the edges and additional complete features to the extent possible. This paper deals with improvement of the quality of lung image and diagnosis of lung cancer by reducing misclassification. The lung CT images are collected from Cancer imaging Archive (CIA) dataset, noise present in the images are eliminated by applying weighted mean histogram equalization approach which successfully removes noise from image, also enhancing the quality of the image, using improved profuse clustering technique (IPCT) for segmenting the affected region. Various spectral features are derived from the affected region. These are examined by applying deep learning instantaneously trained neural network for predicting lung cancer. Eventually, the system is examined by the efficiency of the system using MATLAB based simulation results. The system ensures that 98.42% of accuracy with minimum classification error 0.038.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Continuous development of the technologies has caused changes in people life style due to various scientific factors. Changes in life style lead to changes in human genetic cell (Deoxyribonucleic acid (DNA)) structure function [1]. The structure changed DNA has been spilt into two new cells that leads to the generation of duplicate DNA which is used for the replacement of the old DNA while dying DNA that is generally referred to as mutation [2]. The wrong mutation of DNA cell generally leads to the creation of cancer because the mutation process is affected by several factors such as radon gas exposure, smoking, asbestos fiber breathing and drinking behavior [3]. Depending on the North American Association of Central Cancer Registries (NNACCR) report clearly indicate 234,030 peoples are influenced by lung cancer in USA in the year 2018. Not only men (14%), 13% of women in USA also were influenced by lung cancer [4]. Furthermore, worldwide analysis 154,050 out of 234,030 analyzed as fatal. Based on the

NNACCR report, the conclusion is that lung cancer is one of the most serious diseases which have several symptoms [5] such as breath shortage, cough, chest pain, voice change, sputum color change and blood coughing. In addition, lung cancer is recognized by self-using fatigue, joint pains, weight loss, bone fracture, memory loss, cachexia, headache, bleeding, neurological problem, blood clots and swelling in facial [6]. These symptoms [7] have been analyzed by clinical doctors using spirometer that measures the quantity of breath air that helps diagnosis of lung cancer through application of different screening procedures [8] such as reflex testing, bronch scopy, biopsy, genetic testing and liquid biopsy. In addition to this, blood test is conducted to predict chest infection. National Institute for Health and Care Excellence provides guidelines and symptoms [9] for speedy prediction of the lung cancer. Among the various screening processes, biopsy and bronch scopy examine the cell extracted from the lung to predict cancer related cells. Biopsy predicts the lung cancer but maintain the accuracy and precision is rather difficult. So, computerized tomography (CT) [10] scan is carried out by passing X-rays for examining the changes that occur in the body. During this process, dye or liquid is applied to the chest and images are taken by scanning the

* Corresponding author.

E-mail address: shakeelji@ieee.org (P.M. Shakeel).

body for 10–30 min. Medical imaging shows, CT as an effective process during analysis due to the successive examination of soft organs, tissues in lung, providing useful information about the affected part compared to other imaging such as MRI and PET [11]. Screening methodology has helped prediction of lung cancer but an earlier detection of cancer and accuracy of cancer detection is difficult to maintain. So, computer aided automatic detection (CAD) [12] process has to be applied to the clinical center for developing an effective cancer prediction [13] system using an optimized and intelligent technique. The optimized lung image processing methods are utilized for the examination of inner details of body, restore the details, extract valuable information, and also create a knowledgeable system for the diagnosis of lung cancer. Processing techniques [14] include several stages such as lung image preprocessing, affected part segmentation, feature extraction and lung cancer prediction. Among the different stage, segmentation placed a crucial role as analyzes every pixel present in the lung image, and segment the affected region related cells in separate that helps to making the decision about cancer and non-cancer. Various methods such as fuzzy c-means clustering, k-means clustering, Hopfield neural network, self-organizing map, agglomerative clustering, distributed clustering, sobel are used for making region segmentation [15]. Canny edge detection methods are utilized for the prediction the affected region. In addition to this, several optimization methods for example particle swarm optimization method, genetic algorithm, ant colony, firefly's algorithm are used for the improvement of the clustering process. Various features [16] such as local binary patterns, spectral, statistical, robust features are extracted from the segmented lung region and processed by effective machine learning methods called support vector machine and also a neural network that utilizes the data for training the CAD system considering the need for the incoming new testing to be recognized without any failure. Despite the CAD system effectively utilizing the optimized techniques, sometimes the large set of lung image dataset creates complexity reducing the cancer prediction accuracy [15]. The measurement of precision has been analyzed using various image coordinates, In general entropy is detected based on two ways where the uniform regions of the images are analyzed using low frequency values. In turn image edges are analyzed using high frequency values to detect the edges. In general high frequency signals are used to detect the edges rendering comparison about the dark back ground. Here in imaging of signalling entropy is used to check the available data sets [16]. The entropy provides robust results which helps to measure the indicators for lung cancer. Further characteristic features of Hough transform model of various lung images has been analyzed to check the complex shapes due to potential undetermined significant analysis. This hough transform is used to analyse the noise component of the continuous transform [17,18]. In addition to this, traditional systems fail to process any low-quality lung image resulting in the false features extracted and increasing the misclassification error rate. Considering these issues, this paper utilizes the improved profuse clustering technique (IPCT) for segmenting the affected region considering the weighted mean enhancement technique effectively removes the noise image and the IPCT method extracting the affected region without avoid any original and normal pixels. Various spectral features are derived from the affected region and are examined by influencing deep learning instantaneously accomplished neural network (DITNN) for predicting lung cancer. Finally the proficiency of the framework is inspected by utilizing MATLAB based simulation results. On the basis of discussions, this paper analyses various lung cancer prediction processes in Section 2. Optimal lung cancer prediction technique called IPCT and DITNN method is explained in Section 3. Efficiency of the cancer image prediction process is conferred in Section 4 and Section 5 provides conclusion of the work.

2. Related works

The lung cancer identification method, procedures, ideas, thoughts and processing steps of lung images in the views of different authors are discussed in this section. In [19] examination of the lung tumor related affected region from positron emission tomography (PET) and computerized tomography (CT) image using fuzzy markov random field segmentation approach have been discussed. The method examines the distribution of pixels present in the image by computing the Gaussian distribution and probability distribution function. After computing the pixel distribution, various features are examined in terms of calculating the pixel similarity. Similar pixels are the formed as cluster and affected region is detected from the formed cluster. Then the efficiency of the system is examined using PET and CT image. The author introduced a system that ensures a high dice similarity coefficient value such as 0.85 which indicates that fuzzy random markov model successfully examines the affected lung tumor region. In [20] creating the computer aided diagnosis (CAD) system for predicting lung cancer from CT and PET image. The author examines different challenges and methodologies such as segmentation of image and nodule detection for maximizing the lung cancer prediction process. During the analysis, the captured images are split into two halves, namely, training and testing image for evaluating the efficiency of the CAD system introduced. Along with this, the author has discussed drawbacks seen in the traditional cancer detection system because the introduced CAD system helps to resolve those prediction issues with effective manner.

In [21] considers the implementation of the lung cancer prediction system by utilizing the convolution neural network that overcomes the issues relating to manual cancer prediction. During this process CT scan images are collected and processed using the layer of neural network that makes automatic extraction of the image features, which are processed using the deep learning process for prediction of the cancer related features making use of the utilization of huge volume of images. The authors have created a system that helps making the decision while analyzing the patient CT scan report. In [22] have predicted lung nodules from CT scan images using the convolution neural network method. During this process LIDC IDRI database images are collected and fed into the stack encoder (SAE), convolution neural network (CNN) and deep neural network (DNN) for effective classification of the lung cancer related feature as benign and malignant. The author has introduced a system that ensures upto 84.32% accuracy. According to the various authors' discussions, CT images are examined by segmentation and optimized machine learning technique for the prediction of cancer with accuracy. In this paper lung CT images are examined by applying an improved profuse clustering technique (IPCT) and deep learning based instantaneously trained neural network (DITNN) for classifying lung cancer which is explained in the following section.

3. Improved profuse clustering and deep learning instantaneously trained neural networks based dental disease recognition

3.1. Materials and methods

The lung cancer prediction process is performed with the help of improved profuse clustering and deep learning instantaneously trained neural network. At the time of lung cancer prediction process CT lung image was collected from Cancer imaging Archive (CIA) dataset [23] which contains the images that was gathered from national cancer institute tumor analysis consortium lung cohort. The gathered patient details were correlated with the

genomic and proteomic along with clinical data. The collected images were stored in the DICOM format with particular tags such as imaging, birth date, study dates and so on. The description of the collected image modalities are CT, PT, DX and CR. Totally, 5043 images were utilized for the analysis of the lung cancer from different patients with 48 series. Among the 5043 images, 3000 was used for training and 2043 images as testing images for the examination of lung cancer related modalities. Sample lung cancer images based on the dataset description are shown in Fig. 1.

The captured radiographic CT lung images have been analyzed by medical image processing [24] techniques for the prediction of the changes present in the lung cell. The processing structure of lung cancer is shown in Fig. 2.

Depending on the processing structure of the lung image, lung cancer has been successfully predicted by applying the medical image processing procedure. A detailed explanation of every processing step is provided in the following section.

3.2. Lung cancer detection process

The collected Cancer imaging Archive (CIA) dataset based lung CT images have been processed by pre-processing; lung image segmentation and classification process are discussed in this section.

3.2.1. Lung CT image preprocessing

The captured images are examined in terms of predicting pixel noise, contrast details for improving the quality [25] of the CT lung image as the captured images consists of several inconsistent details, low quality of pixels which reduces the accuracy of Predicted lung cancer. The quality of CT lung image is improved with the help of the pixel intensity examination process that effectively changes the perception of the image pixel. The continuous change of the pixel eliminates the inconsistent pixel, noise pixel with effective manner. Image histogram techniques are utilized for improvement to the image quality because it works on different images with excellence and simplicity. This paper utilizes the weighted mean histogram equalization approach for examining

the captured CT lung image as the introduced method examines the images by dividing them into sub-images. The method also inspects the images in terms of qualitative and number. Before enhancing the quality of the image [13], every pixel is registered against the limit esteem and the clamor present in the pixel is supplanted by the middle esteem. This procedure is rehased for wiping out the clamor present in the image. Then the noise removed CT lung image is analyzed in terms of separating image histogram according to weighted mean function and qualizing sub histogram [26] for enhancing image contrast. S is considered, the sample CT lung image which has G discrete gray level pixels represented as P_0, P_1, \dots, P_{G-1} . After initializing, the image pixel value is determined according to the pixel density and cumulative distribution of the pixel is computed. The pixel density used for inspecting how the pixels are related to a particular range of pixel. Based on the density value, the quality of the pixel is computed, then the density function is calculated using Eq. (1).

$$PDF(P_n) = \frac{n^h}{n} \quad (1)$$

In Eq. (1), $h = 0, 1, 2, 3, \dots, G-1$

n^h is represented as pixel value that is related to the value of h . n represents the total pixels present in the image S . After computing the pixel distribution value, weighted mean value of each pixel is computed as follows.

$$X_t = \frac{\sum_{l=a}^b l * CDF(l)}{\sum_{l=a}^b CDF(l)} \quad (2)$$

In Eq. (2), (a, b) denoted as sub interval value of histogram which is initialized upto $(0, 255)$.

L is denoted as gray pixel value in image.

T is recursion level having 1 as the value which is set as 6 according to the experiment. In addition, t value interval is set as $(0 \leq r \leq t-1)$. r is the range of the sub histogram (X_r, X_{r+1}) . Then the weighted value of the pixel, and the number of sub images are determined according to the histogram value that is estimated as follows

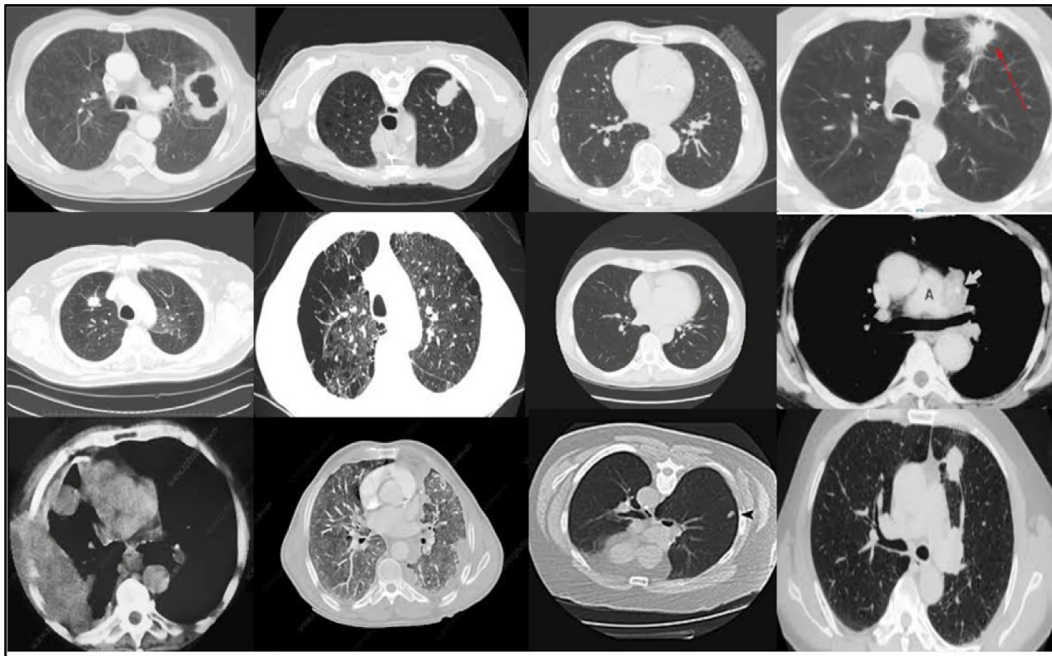


Fig. 1. Sample lung cancer CT images.

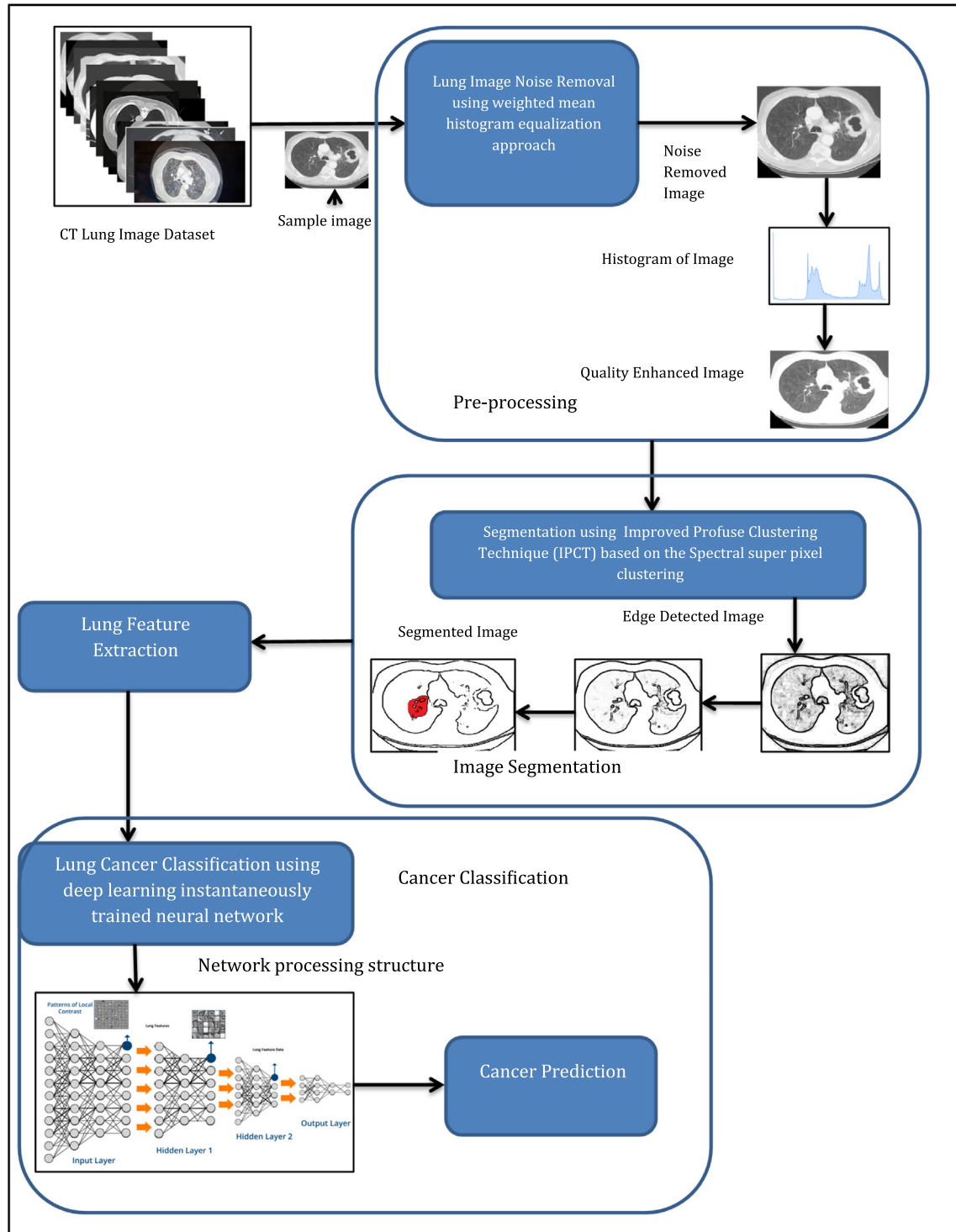


Fig. 2. Lung cancer classification structure.

$$S_k = \{S(x, y) | X_k \leq S(x, y) X_{k+1}, \forall S(x, y) \in S\} \quad (3)$$

On the basis of Eq. (3), S_k is denoted as sub image of captured image S .

$k = 0, 1, 2, \dots, t-1$. Then the relationship of image gray level and sub image is determined based on the probability distribution is defined as follows.

$$PDF_k(P_h) = \frac{n_k^k}{n_k} \quad (4)$$

In Eq. (4) $k = 0, 1, 2, \dots, t-1$.

$$h = X_k + 1, X_k + 2, \dots, X_{k+1}$$

$PDF_k(P_h)$ is related to the k -sub image histogram and gray level image, n_k represents the total number of pixels. Then the cumulative distribution function is defined as follows.

$$CDF(G_j) = \sum_{j=X_0+1}^n PDF_k(G_j) \quad (5)$$


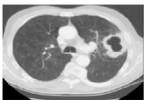


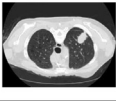








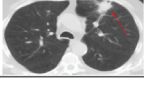
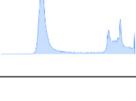
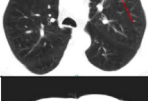
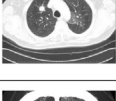
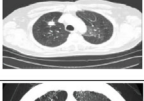

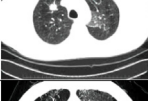
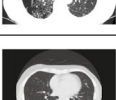
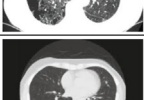
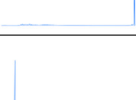
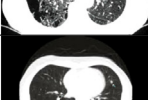
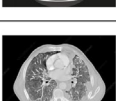
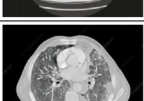






In Eq. (5), $k = 0, 1, 2, \dots, t-1$ and $h = X_k + 1, X_k + 2, \dots, X_{k+1}$. Based on the successful examination of above weighted mean function and piecewise histogram the equalized pixel is mapped. The effective examination process of each pixel is done and the process is split into sub image based histogram computation process improves the quality of image also eliminates the noise pixel present in the CT lung image. According to the above process noise removed, histogram and quality improved CT lung image is shown in Table 1.

Table 1, depicts the output of the pre-processing stage which consists of sample input lung CT image, noise removed lung image, histogram of the lung image and enhanced lung image. After enhancing the quality of the image, the affected region is predicted and segmented by applying the improved profuse clustering technique (IPCT) based on the Spectral super pixel clustering technique.

3.2.2. Lung CT image segmentation using improved profuse clustering technique (IPCT)

The next important step is segmentation of the [27] the cancer affected region from enhanced lung CT image which is done by applying the improved profuse clustering technique (IPCT). The introduced segmentation method inspects the lung CT images according to the pixel similarity and segments the images into several sub images for predicting the affected region. The improved profuse technique works according to two procedures, namely,

Table 1
Sample pre-processed CT lung image.

CT Lung Image	Noise Removed Image	Histogram Image	Enhanced Image
			
			
			
			
			
			
			
			

analyzing the image pixels present in the image and grouping the similar super pixels into the same group for detecting the abrupt image pixels. During the image segmentation process [28], each pixel or data is examined continuously for predicting the similarity of data utilizing the spectrum or pixel eigenvalue. The pixel similarity value includes the image quantitative assessment that used to form the cluster with effective manner. The quality enhanced lung CT image pixels are examined and each pixel is treated as a point in the feature space. An undirected graph is constructed for segmenting the affected region. The lung image related graph is represented as $G(V, E)$ in which V is denoted as the vector or data point and E is the edge between the data points. The connection between the edges has a particular weight value that is represented as $w_{ij}(i, j) \in E$. E is the similarity function between data points i and j . After initializing the lung image related undirected graph, the segmentation process is carried out in which the graph partitioning problem is resolved by splitting the vertex into a high similarity vertex and a low similarity set vertex. Then A and B are considered as two subset vertices and the related normalized cut of the image is represented as follows.

$$w(A, B) = \sum_{i \in A, j \in B} w_{ij} \quad (6)$$

After defining the weighted value of vertices, the n -cut process should be performed as follows.

$$ncut(A, B) = \frac{w(A, B)}{w(A, V)} + \frac{w(A, B)}{w(B, V)} \quad (7)$$

According to the $ncut$ process, the similarity between the data points or pixels is examined in different parts of the image. After that total similarity between the pixels are computed using Eq. (8).

$$nassoc(A, B) = \frac{w(A, A)}{w(A, V)} + \frac{w(B, B)}{w(B, V)} \quad (8)$$

During the similarity computation process, NP-hard problem is resolved by estimating the minimum cut value that is computed as follows,

$$d(i) = \sum_j w_{ij} \quad (9)$$

D is considered as the diagonal matrix of with d diagonal that is represented as $n \times n$. Then symmetric matrix is denoted as W with dimension $n \times n$. By using this representation, minimum cut is computed as follows

$$\min ncut(A, B) = \min_y \frac{y^T(D - W)y}{y^T D y} \quad (10)$$

In above Eq. (10), $y_i \in \{1, -b\}$ where $-b$ is constant and $y^T D 1 = 0$.

Based on the above process, the similarity between the data point or pixel is computed by minimizing the NP-hard problem and $ncut$ process. Then the clustering process is performed on the basis of the number clusters. During the clustering process, the k -means problem is overcome by applying the kernel k -means [29] process which works in a large number of data points in the feature and mapping the non-linear pixel in the feature space with effective manner. Then the kernel k -means function is defined as follows

$$k(x_i, x_j) = \varphi^T(x_i) \varphi(x_j) \quad (11)$$

In Eq. (11), $k(x_i, x_j)$ is kernel function of a particular pixel cluster in the image. The kernel problem is further extended by computing reciprocal of cluster elements. After resolving the cluster problem, similar pixels are grouped in one cluster and dissimilar pixels are grouped into another, this helps prediction of the affected part.

In addition to this, IPCT algorithm works according to the undirected graph which relates each pixel in the image. This helps effective prediction of the edges of the lung image. The sample segmented lung CT images are shown in Table 2.

Based on the above process the affected regions are segmented from the quality enhanced lung CT image. Several statistical and spectral features are extracted from the region for recognizing the lung cancer with effective manner.

3.2.3. Spectral feature extraction

The segmented region is passed on to the feature extraction [30] stage that derives the various spectral features such as mean, third moment skewness, standard deviation, and fourth moment kurtosis because it effectively detects the lung cancer related features. The estimated spectral features are shown in Table 3.

The extracted lung spectral features are processed by deep learning instantaneously trained neural network for recognizing lung cancer which is described in sub Section 3.2.4.

3.2.4. Lung cancer classification

The final stage of this research work is the recognition of the lung cancer with the help of deep learning instantaneously trained neural network (DITNN). Before making the classification [31] process, image training should be performed using deep learning neural network because it does not require manual extracted features. Instead of this process, the deep learning process utilizes the captured lung CT image or segmented image which uses a large number of hidden layers for recognizing the edges present in the image

Table 2
Sample lung cancer segmented image.

Enhanced Image	Edge Detected Image	Segmented Image	

Table 3
Spectral features.

Features	Formula
Mean	$\mu = \frac{1}{N} \sum_{i=1}^N S_i$ N is total amount of pixel present in the segmented region
Standard Deviation	$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (S_i - \mu)^2}$
Third moment skewness	$sk = \left(\frac{1}{N\sigma^3} * \sum_{i=1}^N (S_i - \mu)^3 \right)^{1/3}$
Fourth moment kurtosis	$ku = \left(\frac{1}{N\sigma^4} * \sum_{i=1}^N (S_i - \mu)^4 \right)^{1/4}$

and relevant features that are used in the training of the network with a large quantum of data. Usually the network utilizes 150 hidden layer because of high dimensional dataset. In the below Fig. 3 shows that the deep learning based lung cancer training process structure.

The trained features are saved in the databased as templates for predicting the lung cancer related manual extracted feature. During the classification process, the manually extracted spectral features are self trained and classified with the help of instantaneously trained neural network. The network [32] utilizes the extracted features works according to the unsupervised classification process by inserting the hidden node while examining the lung features. At the time of training process the network has a specific weight value and CC4 network used for generalizing the network and minimizing the irrelevant features while classifying features in feature space. The training process uses the input nodes that is more than one of total number of training features in feature space and network bias value is 1. The weight value of the network is denoted as follows

$$\omega_{ij} = \begin{cases} -1 & \text{for } x_i = 0 \\ +1 & \text{for } x_i = 1 \\ r + s - 1 & \text{for } i = n + 1 \end{cases} \quad (12)$$

In Eq. (12), r is radius generalization and s is the hamming weight. Following this, hidden to output layer weight value is mentioned as -1 or 1 that is belongs to the extracted lung features which used for predicting extracted features as belonging to cancer or non-cancer stage. Finally the output value of hidden to output layer is 1. Then the training process output is defined as follows,

$$y = \begin{cases} 1 & \text{if } \sum x_i \geq 0 \\ 0 & \text{if } \sum x_i < 0 \end{cases} \quad (13)$$

The computed value is compared with that of the deep learning based extracted and trained features for classifying the cancer. The double time matching process improves the overall recognition accuracy also reduces the miss-classification rate with effective manner. Finally the efficiency of lung cancer prediction was evaluated in the following section.

4. Experimental analysis and results

Improved profuse clustering technique (IPCT) and deep learning instantaneously trained neural network (DINN) based lung cancer prediction help evaluation of the process efficiency as discussed in Section 4. During the process of analyzing lung CT images were collected from Cancer imaging Archive (CIA) dataset [33] which consists of 5043 images, 3000 images were used as training and 2043 images as testing images. After dividing the images, MATLAB tool was used for making pre-processing, segmentation, feature extraction and classification process which was done using the algorithm referred to above. Based on the above discussion, the accuracy of segmented lung region images were compared with several

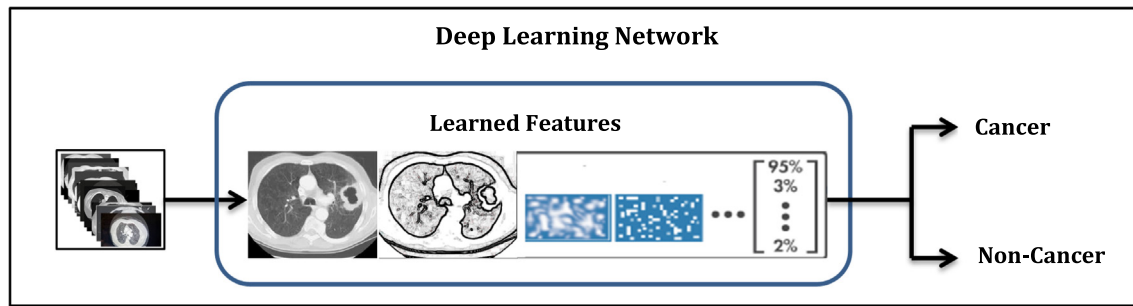


Fig. 3. Deep learning training process structure.

traditional segmentation methods. Details are shown in Table 4. The excellence of the improved profuse clustering technique (IPCT) method was examined in terms of accuracy, recall, precision, specificity and f-score those metrics were estimated as,

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) * 100\% \quad (14)$$

$$\text{Specificity} = TN / (TN + FN) * 100\% \quad (15)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (16)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (17)$$

$$F1 \text{ score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (18)$$

According to the metrics, the obtained value is depicted in Table 4.

Table 4 shows details of the improved profuse clustering technique (IPCT) and various segmentation method. Based on the discussion, each method attains high value in each metric but the improved profuse clustering technique (IPCT) ensures highest the accuracy for every metric due to the successful examination of pixels, interrelation, and similarity between the pixel examination processes. The fuzzy cmeans approach help prediction of cancer affected region up to 88% compared to the other methods, even though the global thresholding approach predicts the affected part up to 95% of specificity but fails to segment the predicted region. In addition to this, watershed approach gather the cancer related pixels from a collection of pixels up to 83% and the sobel method selects the segmented region pixels up to 83% but both methods fail in providing accurate segmentation of the gathered and selected pixels. But the improved profuse clustering technique (IPCT) introduced successfully locates the affected region and segments the particular region up to 95%. Fig. 4 is the graphical representation of the segmentation accuracy.

The computation of super pixel, Eigen value of pixel and similarity between the pixels are used for determination of the interrelationship between the pixels. The pixel relationship helps

grouping similar pixels into one cluster and the remaining pixels in another cluster. The reduced NP hard problem and k-means problem by kernel k means function are used for the improvement of the overall segmentation accuracy. This is shown in Fig. 4. The successful prediction of segmented region helps classification of the lung cancer due to effective learning and classification process. The classified method is known for minimum error rate for reducing the miss-classification rate. The mean square error value of lung cancer classification process obtained is shown in Table 5.

Table 5 depicts the error rate of various lung cancer classification techniques Radial Basis Neural Network (RBNN) [34] (0.63), Convolution Neural Networks (CNN) [35] (0.585), Hopfield Neural Network (HNN) [36] (0.543), Learning Vector Quantization (LVQ) [37–39] (0.437) is discussed. The cancer prediction process examined using various lung images and the Deep Learning with Instantaneously Trained Neural Networks (DITNN) reduces the miss-classification rate arising from the minimum error rate to (0.038). Fig. 5 is the graphical presentation of the error rate.

The successful utilization of deep learning reduces the error rate due to its have high dimensionality of images. The method uses a large set of images which never need manual features. Instead, the deep learning process extracts the features from the segmented region and uses it for matching with the testing lung image feature. In addition to this, the neural network itself has a training process. So, the features are matched with template that completely minimizes the miss-classification process. The minimized error rate enhances the overall lung cancer recognition rate that is measured using precision value. The obtained value is shown in Table 6.

Table 6 demonstrates precision value of various lung cancer prediction approach. The Deep Learning with Instantaneously Trained Neural Networks (DITNN) approach successfully gathers the lung cancer related features due to collection of images in the training process. The successful utilization of each layers, weight updating process improves the overall precision value.

Fig. 6 shows the precision of different lung cancer classification techniques in which the Deep Learning with Instantaneously Trained Neural Networks (DITNN) has an overall 98.43% precision, which is maximum compared to other method such as RBNN (88.8%), CNN (92.26%), HNN (93.92%) and LVQ (96.30%). Depending on the effective training and analyzing process lung cancer classi-

Table 4
Efficiency of different lung image segmentation methods.

Methods	Accuracy	Specificity	Precision	Recall	F1 score
Region Growing (RG)	0.738	0.83	0.73	0.746	0.738
Global Threshold (GT)	0.824	0.952	0.79	0.835	0.8125
Fuzzy c-means (FCM)	0.882	0.943	0.75	0.69	0.72
Canny Method (CM)	0.863	0.89	0.69	0.77	0.73
Sobel Method (SM)	0.874	0.931	0.72	0.834	0.777
Watershed Approach (WSA)	0.823	0.868	0.838	0.81	0.824
Improved profuse clustering technique (IPCT)	0.945	0.972	0.94	0.968	0.954

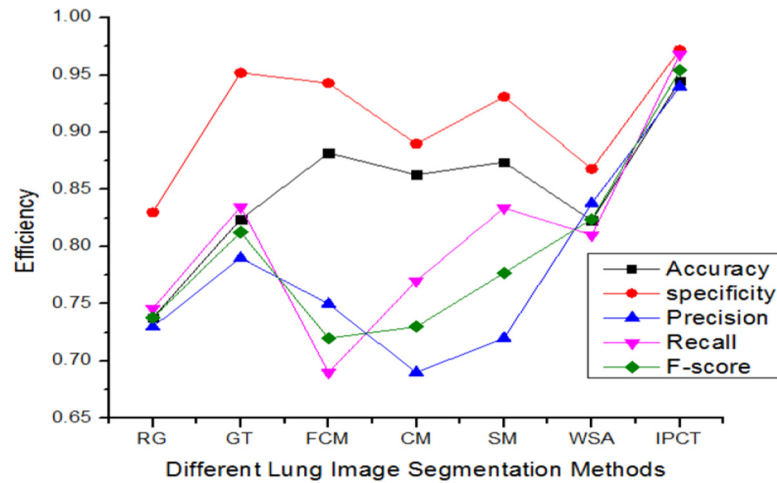


Fig. 4. Efficiency of lung image segmentation.

Table 5
Error rate.

Methods	Lung Images									
	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7	Image 8	Image 9	Image 10
Radial Neural Networks (RBNN)	0.74	0.654	0.63	0.632	0.631	0.58	0.592	0.63	0.623	0.62
Convolution Neural Networks (CNN)	0.53	0.63	0.64	0.52	0.65	0.62	0.52	0.58	0.535	0.63
Hopfield Neural Networks (HNN)	0.52	0.546	0.52	0.53	0.584	0.56	0.546	0.563	0.523	0.54
Learning Vector Quantization (LVQ)	0.43	0.47	0.438	0.45	0.431	0.423	0.45	0.42	0.433	0.428
Deep Learning with Instantaneously Trained Neural Networks (DITNN)	0.043	0.0343	0.043	0.0456	0.043	0.0382	0.0325	0.038	0.031	0.0382

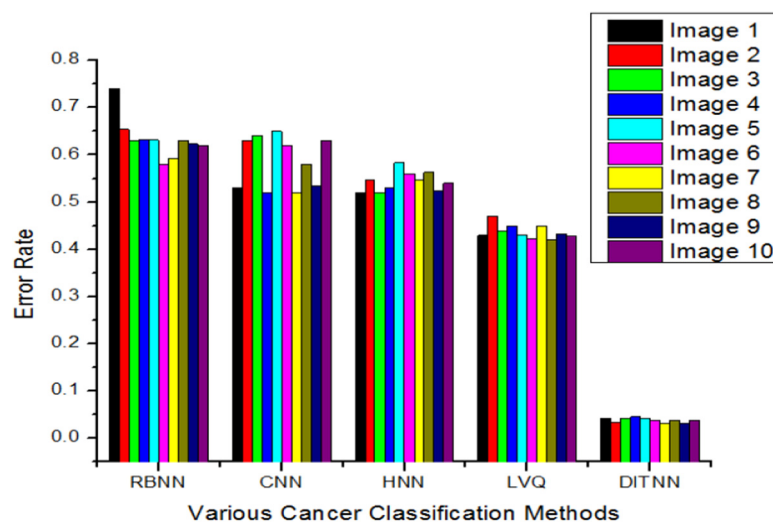


Fig. 5. Error rate.

fication efficiency is further discussed using recall metric that is shown in Table 7.

Table 7 relates to the recall value of various lung cancer prediction approach. The Deep Learning with Instantaneously Trained Neural Networks (DITNN) approach does the selection of the lung cancer related features from a collection of images in the training process. The successful utilization of each layer, weight updating

process improves the overall recall value. The graphical representation of recall is shown in Fig. 7.

Fig. 7 shows the recall of different lung cancer classification techniques in which the Deep Learning with Instantaneously Trained Neural Networks (DITNN) has 98.36% precision overall, which is maximum compared to other methods image processing techniques such as RBNN (91.45%), CNN (93.25%), HNN (97.11%)

Table 6
Precision.

Methods	Lung Images									
	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7	Image 8	Image 9	Image 10
Radial Neural Networks (RBNN)	88.6	89.2	88.7	88.2	89.2	88.4	89.21	89.4	88.2	89.7
Convolution Neural Networks (CNN)	89.8	89.5	90.90	91.21	92.9	93.8	93.65	92.8	93.5	94.2
Hopfield Neural Networks (HNN)	92.2	93.9	93.5	93.73	94.2	94.39	95.21	93.4	94	94.7
Learning Vector Quantization (LVQ)	95.3	96.3	96.2	96.4	95.24	96.9	95.9	96.87	97.3	96.6
Deep Learning with Instantaneously Trained Neural Networks (DITNN)	98.5	98.9	99.24	98.7	97.6	98.3	98.4	98.29	98.2	98.3

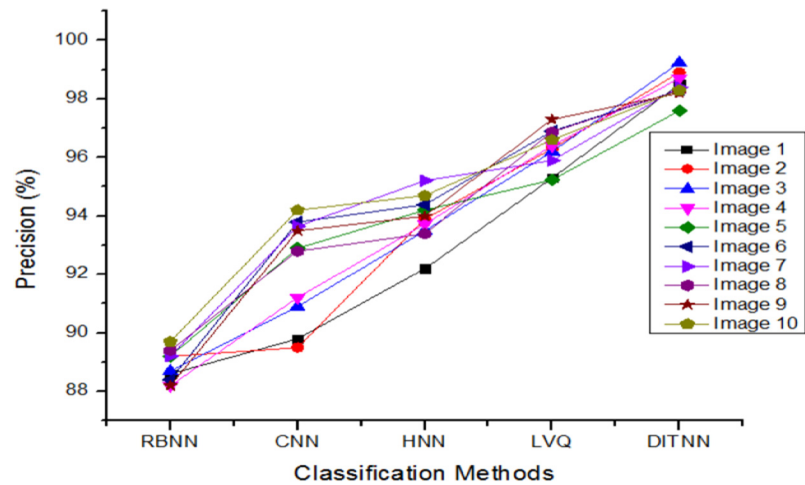


Fig. 6. Precision.

Table 7
Recall.

Methods	Images									
	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7	Image 8	Image 9	Image 10
Radial Neural Networks (RBNN)	88.6	89.39	90.4	91.35	92.9	91.7	90.91	92.8	93.78	92.7
Convolution Neural Networks (CNN)	92.8	93.6	92.54	92.31	93.4	92.97	92.5	93.78	94.2	94.4
Hopfield Neural Networks (HNN)	96.5	97	97.4	96.17	97.6	96.2	97.4	97.4	97.62	97.83
Learning Vector Quantization (LVQ)	97.24	97.56	97.14	97.89	97.89	97.45	97.98	98.01	97.98	98.21
Deep Learning with Instantaneously Trained Neural Networks (DITNN)	98.09	98.46	98.23	98.45	98.65	98.13	98.53	98.23	98.4	98.45

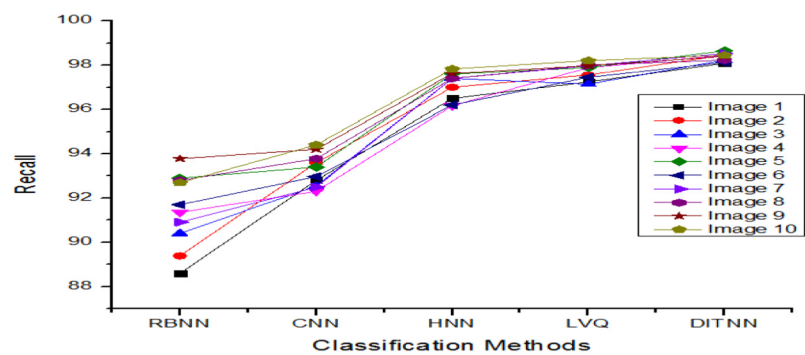


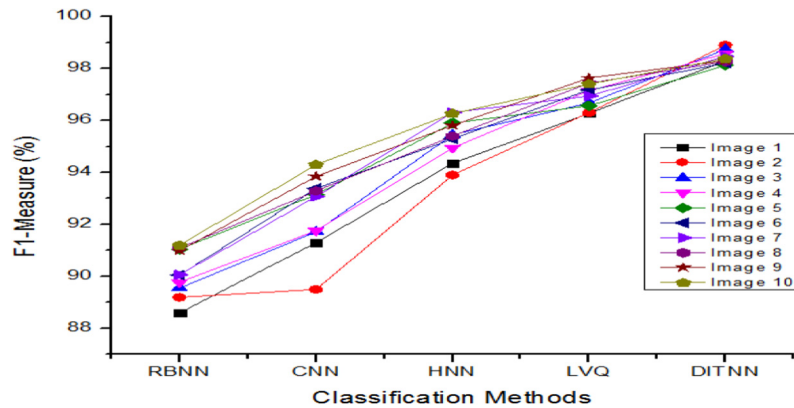
Fig. 7. Recall.

and LVQ (97.73%). Depending on the training feature analysis, selection process lung cancer classification efficiency is further discussed using F1 measure metric that is shown in Table 8.

Table 8 demonstrate the F1 measure value of various lung cancer classification approach. Effective prediction and selection of lung features are used for the prediction of the affected lung region

Table 8
F1 measure.

Methods	Images									
	1	2	3	4	5	6	7	8	9	10
Radial Neural Networks (RBNN)	88.6	89.2	89.55	89.775	91.05	90.05	90.06	91.1	90.99	91.2
Convolution Neural Networks (CNN)	91.3	89.5	91.72	91.76	93.15	93.385	93.075	93.29	93.85	94.3
Hopfield Neural Networks (HNN)	94.35	93.9	95.45	94.95	95.9	95.295	96.305	95.4	95.81	96.265
Learning Vector Quantization (LVQ)	96.27	96.3	96.67	97.145	96.565	97.175	96.94	97.44	97.64	97.405
Deep Learning with Instantaneously Trained Neural Networks (DITNN)	98.295	98.9	98.735	98.575	98.125	98.215	98.465	98.26	98.3	98.375

**Fig. 8.** F1 measure.

by utilizing the successful feature training process. Fig. 8 is the graphical presentation of the F1 measure value.

Thus the Deep Learning with Instantaneously Trained Neural Networks (DITNN) recognizes the lung cancer with 98.42% of accuracy which is the maximum compared to other classification methods.

5. Conclusion

This paper has done evaluation of the lung CT images for predicting lung cancer through use of the improved profuse clustering technique (IPCT) and Deep Learning with Instantaneously Trained Neural Networks (DITNN) approach. Initially, the lung CT images were collected from Cancer imaging Archive (CIA) dataset which consists of 5043 DICOM format images that was divided into 3000 training images and 2043 testing images. Then the quality of the images was enhanced by computing the weighted mean function that replaced the pixel using probability distribution and cumulative distribution process. After enhancing the representation of the image, the affected part was segmented by computing the pixel similarity value. Based on the similarity measure clusters were formed for the extraction of the spectral related features. The features were trained and classified by classifier methods which successfully predict the cancer upto 98.42% of accuracy with minimum classification error of 0.038.

Acknowledgement

The authors would like to thanks Advance Manufacturing Centre and Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka for providing all facilities to conduct this study.

References

- [1] Prevalence of beliefs about actual and mythical causes of cancer and their association with socio-demographic and health-related characteristics: Findings from a cross-sectional survey in England.
- [2] Tamara Baker, Early cancer detection behaviors among black males, *J. Mens Health* 14 (3) (2018), ISSN: 1875-6859.
- [3] Z. Liu, J. Wang, Z. Yuan, B. Zhang, L. Gong, L. Zhao, P. Wang, Preliminary results about application of intensity-modulated radiotherapy to reduce prophylactic radiation dose in limited-stage small cell lung cancer, *J. Cancer* 9 (15) (2018) 2625–2630, <https://doi.org/10.7150/jca.24976>.
- [4] Catharina Balmelli, Nikola Railic, Marco Siano, Kristin Feuerlein, Richard Cathomas, Valerie Cristina, Christiane Güthner, Stefan Zimmermann, Sabine Weidner, Miklos Pless, Frank Stenner, Sacha I. Rothschild, Lenvatinib in advanced radioiodine-refractory thyroid cancer - a retrospective analysis of the Swiss Lenvatinib named patient program, *J. Cancer* 9 (2) (2018) 250–255, <https://doi.org/10.7150/jca.22318>.
- [5] R. Manser, A. Lethaby, L.B. Irving, C. Stone, G. Byrnes, M.J. Abramson, D. Campbell, Screening for lung cancer, *Cochrane Database Syst. Rev.* 6 (6) (June 2013), <https://doi.org/10.1002/14651858.CD001991.pub3> CD001991.
- [6] M. Usman Ali, J. Miller, L. Peirson, D. Fitzpatrick-Lewis, M. Kenny, D. Sherifali, P. Raina, Screening for lung cancer: a systematic review and meta-analysis, *Preventive Med.* 89 (August 2016) 301–314, <https://doi.org/10.1016/j.ypmed.2016.04.015>.
- [7] A. Gutierrez, R. Suh, F. Abtin, S. Genshaft, K. Brown, Lung cancer screening, *Semin. Interventional Radiol.* 30 (2) (June 2013) 114–120, <https://doi.org/10.1055/s-0033-1342951>. PMC 3709936.
- [8] P.B. Bach, J.N. Mirkin, T.K. Oliver, et al., Benefits and harms of CT screening for lung cancer: a systematic review, *JAMA: J. Am. Med. Assoc.* 307 (22) (June 2012) 2418–2429, <https://doi.org/10.1001/jama.2012.5521>.
- [9] D.R. Aberle, F. Abtin, K. Brown, Computed tomography screening for lung cancer: has it finally arrived? Implications of the national lung screening trial, *J. Clin. Oncol.* 31 (8) (March 2013) 1002–1008, <https://doi.org/10.1200/JCO.2012.43.3110>.
- [10] L. Frank, L.E. Quint, Chest CT incidentalomas: thyroid lesions, enlarged mediastinal lymph nodes, and lung nodules, *Cancer Imaging* 12 (1) (March 2012) 41–48, <https://doi.org/10.1102/1470-7330.2012.0006>.
- [11] N. Murray, A.T. Tursi, A review of first-line treatment for small-cell lung cancer, *J. Thorac. Oncol.* 1 (3) (March 2006) 270–278, [https://doi.org/10.1016/s1556-0864\(15\)31579-3](https://doi.org/10.1016/s1556-0864(15)31579-3).
- [12] Emre Dandil, A computer-aided pipeline for automatic lung cancer classification on computed tomography scans, *J. Healthcare Eng.* 2018 (2018) 12, <https://doi.org/10.1155/2018/9409267>, Article ID 9409267.
- [13] G. Manogaran, P.M. Shakeel, A.S. Hassanein, M.K. Priyan, C. Gokulnath, Machine-learning approach based gamma distribution for brain abnormalities detection and data sample imbalance analysis, *IEEE Access* (2018 Nov 9), <https://doi.org/10.1109/ACCESS.2018.2878276>.
- [14] P.M. Shakeel, A. Tolba, Zafer Al-Makhadmeh Al-Makhadmeh, Mustafa Musa Jaber, Automatic detection of lung cancer from biomedical data set using discrete AdaBoost optimized ensemble learning generalized neural networks, *Neural Comput. Appl.* (2019) 1–14, <https://doi.org/10.1007/s00521-018-03972-2>.
- [15] N. Gupta, D. Gupta, A. Khanna, P.P. Rebouças Filho, V.H.C. de Albuquerque, Evolutionary algorithms for automatic lung disease detection, *Measurement* 140 (2019) 590–608, <https://doi.org/10.1016/j.measurement.2019.02.042>.
- [16] E.O. Bursalioğlu, F.A. Alkan, Ü.B. Barutçu, M. Demir, Y. Karabul, B. Balkan, O. İçelli, Prediction of electron density and trace element concentrations in human blood serum following radioiodine therapy in differentiated thyroid

- cancer patients, *Measurement* 100 (2017) 19–25, <https://doi.org/10.1016/j.measurement.2016.12.035>.
- [17] S. Dutta, S.K. Pal, R. Sen, Progressive tool flank wear monitoring by applying discrete wavelet transform on turned surface images, *Measurement* 77 (2016) 388–401, <https://doi.org/10.1016/j.measurement.2015.09.028>.
 - [18] J. Valença, D. Dias-da-Costa, E.N.B.S. Júlio, H. Araújo, H. Costa, Automatic crack monitoring using photogrammetry and image processing, *Measurement* 46 (1) (2013) 433–441, <https://doi.org/10.1016/j.measurement.2012.07.019>.
 - [19] H. Cui, X. Wang, D. Feng, Automated localization and segmentation of lung tumor from PET-CT thorax volumes based on image feature analysis, in: *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '12)*, San Diego, Calif, USA, September 2012, pp. 5384–5387.
 - [20] N. Zhang, S. Ruan, S. Lebonvallet, Q. Liao, Y. Zhu, Kernel feature selection to fuse multi-spectral MRI images for brain tumor segmentation, *Comput. Vision Image Understanding* 115 (2) (2011) 256–269.
 - [21] Yu Guo, Yuanming Feng, Jian Sun, et al., Automatic lung tumor segmentation on PET/CT images using fuzzy markov random field model, *Comput. Math. Methods Med.* 2014 (2014) 6, <https://doi.org/10.1155/2014/401201>, Article ID 401201.
 - [22] Ayman El-Baz, Garth M. Beache, Georgy Gimelfarb, et al., Computer-aided diagnosis systems for lung cancer: challenges and methodologies, *Int. J. Biomed. Imaging* 2013 (2013) 46, <https://doi.org/10.1155/2013/942353>, Article ID 942353.
 - [23] Prajwal Rao, Nishal Ancelette Pereira, Raghuram Srinivasan, Convolutional neural networks for lung cancer screening in computed tomography (CT) scans, 2nd International Conference on Contemporary Computing and Informatics (IC3I) in IEEE, 2016.
 - [24] QingZeng Song, Lei Zhao, XingKe Luo, XueChen Dou, Using deep learning for classification of lung nodules on computed tomography images, *J. Healthcare Eng.* 2017 (2017) 7, <https://doi.org/10.1155/2017/8314740>.
 - [25] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC), Radiology data from the clinical proteomic tumor analysis consortium lung squamous cell carcinoma [CPTAC-LSCC] collection, *Cancer Imaging Arch.* (2018), <https://doi.org/10.7937/k9/tcia.2018.6emub512>.
 - [26] I.R. Valente, P.C. Cortez, E.C. Neto, J.M. Soares, V.H. de Albuquerque, J.M. Tavares, Automatic 3D pulmonary nodule detection in CT images: a survey, *Comput. Methods Programs Biomed.* 124 (1) (2016) 91–107.
 - [27] Angalaparameswari Rajasekaran, P. Senthil kumar, Image denoising using median filter with edge detection using canny operator, *Int. J. Sci. Res. (IJSR)* 3 (2) (February 2014) 30–34.
 - [28] Yogesh Rao, Nisha Sarwade, Roshan Makkar, Denoising and enhancement of medical images using wavelets in LabVIEW, *IJ. Image, Graphics Signal Process.* (2015).
 - [29] C. Li, X. Wang, S. Eberl, et al., A likelihood and local constraint level set model for liver tumor segmentation from CT volumes, *IEEE Trans. Biomed. Eng.* 60 (10) (2013) 2967–2977.
 - [30] K.P. Sridhar, S. Baskar, P.M. Shakeel, V.S. Dhulipala, Developing brain abnormality recognize system using multi-objective pattern producing neural network, *J. Ambient Intell. Hum. Comput.* (2018) 1–9, <https://doi.org/10.1007/s12652-018-1058-y>.
 - [31] P.M. Shakeel, S. Baskar, V.S. Dhulipala, M.M. Jaber, Cloud based framework for diagnosis of diabetes mellitus using K-means clustering, *Health Inf. Sci. Syst.* 6 (1) (2018 Dec 1) 16, <https://doi.org/10.1007/s13755-018-0054-0>.
 - [32] P. Mohamed Shakeel, Tarek E.El. Tobely, Haytham Al-Feel, Gunasekaran Manogaran, S. Baskar, Neural network based brain tumor detection using wireless infrared imaging sensor, *IEEE Access* (2019) 1.
 - [33] P.M. Shakeel, G. Manogaran, Prostate cancer classification from prostate biomedical data using ant rough set algorithm with radial trained extreme learning neural network, *Health Technol.* (2018) 1–9, <https://doi.org/10.1007/s12553-018-0279-6>.
 - [34] Z. Zhang et al., TextCC: New feedforward neural network for classifying documents instantly, *Adv. Neural Networks* 3497 (2005) 232–237.
 - [35] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, F. Prior, The cancer imaging archive (TCIA): maintaining and operating a public information repository, *J. Digital Imaging* 26 (6) (December 2013) 1045–1057 (paper).
 - [36] I. Jasmine Selvakumari Jeya, S.N. Deepa, Lung cancer classification employing proposed real coded genetic algorithm based radial basis function neural network classifier, *Comput. Math. Methods Med.* 2016 (2016) 15, <https://doi.org/10.1155/2016/7493535>, Article ID 7493535.
 - [37] Margarita Kirienko, Martina Sollini, Giorgia Silvestri, et al., Convolutional neural networks promising in lung cancer T-parameter assessment on baseline FDG-PET/CT, *Contrast Media Mol. Imaging* 2018 (2018) 6, <https://doi.org/10.1155/2018/1382309>, Article ID 1382309.
 - [38] D.S. Rizzuto, M.J. Kahana, An autoassociative neural network model of paired-associate learning, *Neural Comput.* 13 (9) (2001) 2075–2092, <https://doi.org/10.1162/089976601750399317>.
 - [39] Mhd Saeed Sharif, Maysam Abbod, Abbes Amira, Habib Zaidi, Artificial neural network-statistical approach for PET volume analysis and classification, *Adv. Fuzzy Syst.* 2012 (2012) 10, <https://doi.org/10.1155/2012/327861>, Article ID 327861.