Neural Computing and Applications
https://doi.org/10.1007/s00521-018-3518-x

ORIGINAL ARTICLE

Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans

Gur Amrit Pal Singh[1] · P. K. Gupta[1]

Received: 14 August 2017 / Accepted: 27 April 2018

Abstract
Lung cancer is one of the most common causes of death among all cancer-related diseases (Cancer Research UK in Cancer mortality for common cancers. http://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality/common-cancers-compared, 2017). It is primarily diagnosed by performing a scan analysis of the patient's lung. This scan analysis could be of X-ray, CT scan, or MRI. Automated classification of lung cancer is one of the difficult tasks, attributing to the varying mechanisms used for imaging patient's lungs. Image processing and machine learning approaches have shown a great potential for detection and classification of lung cancer. In this paper, we have demonstrated effective approach for detection and classification of lung cancer-related CT scan images into benign and malignant category. Proposed approach firstly processes these images using image processing techniques, and then further supervised learning algorithms are used for their classification. Here, we have extracted texture features along with statistical features and supplied various extracted features to classifiers. We have used seven different classifiers known as *k*-nearest neighbors classifier, support vector machine classifier, decision tree classifier, multinomial naive Bayes classifier, stochastic gradient descent classifier, random forest classifier, and multi-layer perceptron (MLP) classifier. We have used dataset of 15750 clinical images consisting of both 6910 benign and 8840 malignant lung cancer related images to train and test these classifiers. In the obtained results, it is found that accuracy of MLP classifier is higher with value of 88.55% in comparison with the other classifiers.

Keywords Lung cancer · Image processing · Image classification · Machine learning

1 Introduction

Today, cancer has become one of the most common causes of death among youths worldwide. Lung cancer, breast cancer, stomach cancer, and prostate cancer are one of the most frequently diagnosed cancers among men and women that lead severe complexities or in many cases to death, if not detected at early stage. Cancer in human body represents the abnormal growth of cell. According to study, there are nineteen different types of cancer that can affect healthy person. Among all these cancers, lung cancer has the highest mortality rate. It is estimated that about 1.7 million people die annually due to this disease. Principle cause of lung cancer has been attributed to smoking which accounts around 80% of total lung cancer cases worldwide. In its initial stage, it is difficult to detect the lung cancer. As per findings, about 25% of people who were diagnosed with lung cancer in its initial stage experienced no symptoms at all. Unlike other cancers, lung cancer cannot be seen with naked eye and its symptoms are often masked with other disease symptoms such as bronchitis, asthma, and coughing. Lung cancer is usually identified when an X-ray or CT scan of patient's chest is performed for any another good reason [2]. The rest of 75% people are diagnosed when they experience or develop some sort of symptoms. These symptoms may arise due to direct effects of the primary tumor or due to effects of cancer that has spread to other parts of the body (metastases) or due to disturbances of hormones, blood, or other systems. Lung

✉ P. K. Gupta
pradeepkumar.gupta@juit.ac.in

Gur Amrit Pal Singh
guramrit013@gmail.com

[1] Department Computer Science and Engineering, Jaypee University of Information Technology, Solan 173234, India

Published online: 05 May 2018

cancer usually spreads toward the center of the chest cavity; this is because the natural flow of lymph which is outward from the lungs and inward toward the center of the chest [3]. When cancer sets in, a single cell or a bunch of cells abruptly start multiplying in an uncontrolled and disorganized way, which if not stopped can lead to formation of lumps or tumors. Any tumor can be broadly classified into two categories benign and malignant. In contrast to benign tumor, malignant tumor penetrates nearby body cells and can spread to other body parts with its growth. Usually, benign tumors are not considered very dangerous but can become dangerous if they develop vital structures required for their growth such as blood vessels or nerves. As the discussed pattern is followed, the tumor is said to be invasive and radiologists and other healthcare professionals use the various imaging techniques for its detection. In this paper, we have used computed tomography (CT) scans of lungs. Figure 1 depicts a general methodology of lung cancer detection system that consists of five basic stages. The first stage is of image acquisition that represents the collection set of images related to body part. For this work, we have obtained DICOM CT scan images of lungs from an online database [4]. This database was part of the 2015 SPIE Medical Imaging Conference, SPIE with the support of American Association of Physicists in Medicine (AAPM), and the National Cancer Institute (NCI) conducted a Grand Challenge on quantitative image analysis methods for the diagnostic classification of malignant and benign lung nodules. This collection of images is freely available to browse, download, and use for commercial, scientific, and educational purposes, and the same is outlined in the Creative Commons Attribution 3.0 Unported License. The second stage of image pre-processing and segmentation applies various image processing approaches for image enhancement to improve image quality for interpretation and makes it easier for feature extraction. The third stage extracts features from the enhanced segmented images obtained in the previous stage. These features are used by various machine learning approaches for the task of image classification in stage four. Final stage compares the performance of various machine learning approaches used to detect lung cancer in the obtained images.

2 Related work

In recent years many machine learning approaches, especially neural networks, have been widely used for detection of lung cancer using medical images. Many of the proposed approaches have achieved high accuracy rate [5].

2.1 Image processing techniques

Dwivedi et al. [6] have proposed a image pre-processing method known as contrast limited adaptive histogram equalization (CLAHE). They have used gray-level co-occurrence matrix (GLCM) for extracting the image features, which also gives the information regarding the position of those pixels which have similar values of gray level. A GLCM can contain a variety of statistical features that can be extracted from the matrix for analysis purpose. Authors have used automatic feature selection algorithms for determining the best features. Qian et al. [7] have conducted a study to predict the near-term risk of developing breast cancer by using the image dataset of mammograms. Proposed scheme is designed around the four image processing modules like image pre-processing, segmentation, feature extraction, and classification to compute image feature asymmetry. Chaudhary and Singh [8] have processed the image by applying (i) image pre-processing and (ii) feature extraction methods. Image pre-processing step consists of two major segments: image enhancement and image segmentation. For image enhancement, they have tested three algorithms, known as Gabor filter, auto-enhancement, and fast Fourier transform. From the experiments they have found that Gabor filter is the most suitable for image enhancement. For image segmentation, they have used thresholding and watershed segmentation techniques and it is found that watershed segmentation performs better over thresholding technique. Al-Tarawneh [3] has also obtained the similar results by using the same techniques for image enhancement. They have applied binarization and masking approach for feature extraction and extracted four features: average intensity, area, perimeter, and eccentricity from the image. Similarly, Pratap and Chauhan [9] have applied watershed techniques to perform image segmentation. Bhusri et al. [10] have used law's of feature extraction in order to extract features from the region of interest. Many other image processing

Fig. 1 Five basic steps for detection of lung cancer in humans

techniques for image enhancement, pre-processing, segmentation, and feature extraction have been suggested by [5]. Kuruvilla and Gunavathi [11] have used simple Otsu's method for image segmentation, along with morphological opening method, with periodic line as the structuring element of fixed in size. Otsu's method minimizes intra-class variance, whereas at the same time minimizes inter-class variance. They have also identified the major drawback of using simple Otsu's method that even after converting the grayscale image into binary image, some of the gaps are still left in the lungs which correspond to arteries or air present in the lungs and these gaps can cause the classifiers to predict wrong values as they create an illusion of a cancerous mass. In order to overcome this issue, they have suggested that morphological operations must be performed to fill in the gaps left after applying thresholding method on the grayscale CT scan images.

2.2 Classification using machine learning approaches

Neural networks form the basis for many machine learning-based approaches. It is the task for classifying an image into various classes based on the labels of the input training dataset. There are various machine learning algorithms which can be used for the task of image classification and can be classified into two categories known as supervised and unsupervised learning approaches. Mitra and Pal [12] and Amato et al. [13] have obtained the similar results. Karabatak and Ince [14] have used a combination of association rules and neural network to provide efficient computer-aided diagnostic system. The use of association rules reduces the dimensions of feature vector, without greatly compromising the accuracy of the overall system. Dwivedi et al. [6] have used multinomial multivariate Bayesian references for classification of image obtained after image processing stage, into normal (non-cancerous) and abnormal (cancerous) images. Adi et al. [15] have developed a system based on digital image processing techniques for identification of cancer cells through the stages of feature extraction using GLCM and classification using a naive Bayes algorithm. In their results they have achieved the accuracy of 88.57% in detecting the lung cancer. Joachims [16] has presented an improved algorithm for training support vector machine (SVM) on large-scale datasets and problems and described its effective way of implementation in SVM. Joachims has also introduced the technique for shrinking the size of problem during its optimization. Tidke and Chakkarwar [17] have presented a CAD system for early detection of lung cancer from CT images and to classify whether tumor is benign or malignant. They have presented the five-stage model and used GLCM for textural feature extraction and SVM classifier for image classification. In their experiments they have used small size of dataset which consists of only 25 JPEG images and obtained results show 96% accuracy using SVM classifier. Touw et al. [18] have listed the key features of random forest and mentioned that this algorithm is the most widely used in life sciences for both regression and classification of tasks, for example the prediction of disease state of patients. It also allows the extraction of additional relevant knowledge from omics data. Shi et al. [19] have presented a random forest clustering strategy for tumor profiling based on tissue microarray data. They have used this method to detect and analyze the renal cell carcinoma a type of kidney cancer in adults. Ramos-Gonzalez et al. [20] have proposed a novel case-based reasoning framework for diagnosis of lung cancer subtypes. They have used gradient boosted regression trees (GBRT) feature selection method to achieve high predictive accuracy. In their experiments they have used *k*-nearest neighbors (kNN), naive Bayes classifier (NB) and support vector machine (SVM) methods.

2.3 Deep learning techniques

Sakamoto et al. [21] have presented cascaded multi-stage neural networks with single-sided classifiers to reduce the false positives of lung nodule classification in CT scan images. They have trained convolutional neural network with balanced dataset of 888 images and achieved sensitivity of 92.4 and 94.5% in their experiments. Demyanov et al. [22] have presented the method for automatic detection of dermoscopic patterns using deep convolutional neural network and other image classification algorithms. In their experiments, they have generated more than 2000 samples for each class of dermoscopic patterns with correct classification rate around 83–88%. Lopez et al. [23] have also focused on detection of skin cancer and presented the deep learning-based approach for its detection. In their proposed solution they have designed VGGNet convolutional neural network architecture. Bewal et al. [24] have reviewed various neural network techniques like convolutional neural networks (CNN), artificial neural network (ANN) with feed forward. for the detection of breast cancer and found that neural networks can greatly help in providing second opinion to healthcare professionals and greatly reduce the time of treatment for the patients. Havaei et al. [25] have presented a fully automatic brain tumor segmentation method based on deep neural networks (DNNs). They have considered the convolutional neural network architecture above all as it exploits both local features and more global contextual features. Wong et al. [26] have demonstrated the feasibility of applying the deep learning approaches for automatically differentiating between normal and cancerous lung. They have used

coherent anti-Stokes Raman scattering (CARS) algorithm for diagnosing lung cancer. They have employed partial least squares regression and support vector machine classification algorithm to extract 35 different features from the image and to make prediction about the lung cancer. In the obtained results they have achieved the accuracy of 89.2% in classifying the various lung cancer images. Qian et al. [27] have designed and implemented three deep structured algorithms based on multichannel region of interest. They have used convolutional neural network (CNN), deep belief network (DBN), and stacked denoising auto-encoder (SDAE) for designing and implementing auto-generated features. For comparison purpose, they have used conventional computer-aided diagnosis (CADx) systems. In their obtained results, though the performance of the proposed system is somewhat comparable with that of the conventional CADx system, the performance of the proposed system needs to be tested further. Mahbod et al. [28] have considered convolutional neural network over existing classical machine learning methods for object detection and classification. They have presented the solution for skin cancer detection by employing pre-trained convolutional neural networks that ensembles learning from 2000 skin lesion images. These skin images are categorized into three categories known as melanoma, seborrheic keratosis, and benign nevi images. In their experiments, they have achieved the competitive results with 84.8 and 93.6% for melanoma and seborrheic keratosis binary classification problem, respectively. Similarly, Esteva et al. [29] have also proposed a deep convolutional neural network for detection and classification of skin cancer. Maass et al. [30] have proposed architecture based on deep convolutional networks and used the deep learning for classification of tumor. By using deep learning they have combined the feature extraction and classification into single model. The objective of their study is to begin the research on advanced end-to-end learning methods. In their obtained results, proposed model displayed better results with deep learning approaches.

Difficulty arises in using above-mentioned techniques for segmenting to such an extent so that the task of classifier could be ease down. However, most of the discussed techniques can find the presence or absence of cancer but classifying cancer into two benign and malignant is still a difficult task and most of the above-mentioned studies failed to do. Proposed methodology overcomes the difficulties faced by these discussed techniques. It can detect the presence, absence of cancerous mass and classify the mass into benign or malignant tumor if the cancerous mass is present.

3 Methodology

For obtaining better results we have broadly divided the various stages of Fig. 1 into two different categories known as image analysis and image classification.

1. *Image analysis* is the process of working on images in order to improve the image quality for human readability and for removing any noise present in images for its efficient classification. Image analysis can be performed in the following steps:

   – *Image pre-processing and segmentation* The first step of image analysis is image pre-processing. The acquired image is converted into a grayscale image as shown in Fig. 2; afterward, we apply image denoising methods to remove the noise in the image. Here, we have considered three image denoising methods known as median blur, Gaussian blur, and bilateral blur, as shown in Fig. 3a–c, respectively. After analyzing the results it is found that Gaussian blur outperforms other methods.

   Once image denoising is done, we apply thresholding methods for converting the grayscale image into a binary image. A binary image consists of only two pixel values either "0" or "1," whereas in a grayscale image each pixel can acquire any value between 0 and 255. In this paper, we have compared three thresholding methods known as global thresholding, Otsu's method with adaptive mean thresholding, and Otsu's method with adaptive Gaussian thresholding. While comparing these three thresholding methods it is found out that Otsu's method with adaptive Gaussian

Fig. 2 Input image

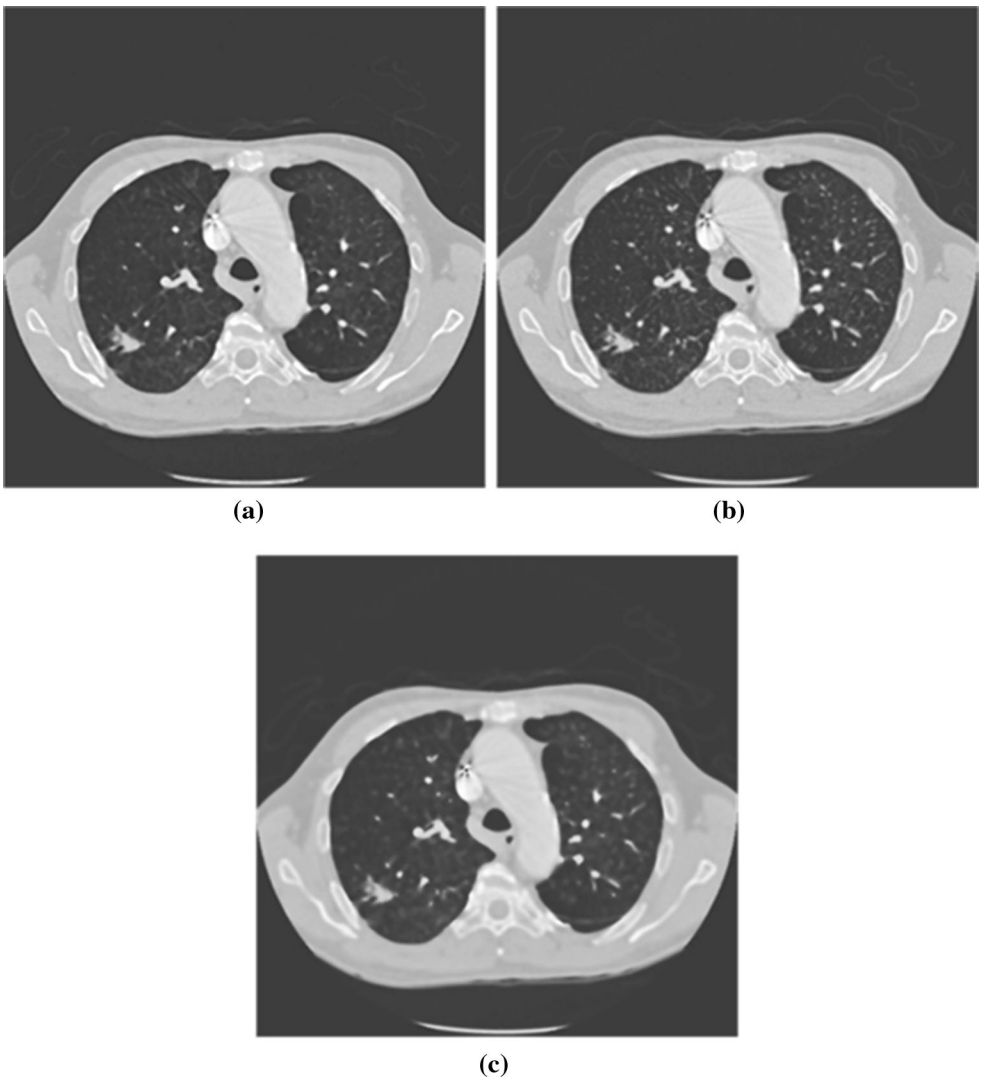

Fig. 3 Applying image denoising methods on input image **a** median blur, **b** Gaussian blur and **c** bilateral blur

thresholding provides the suitable results for further analysis. Here, Fig. 4a–c represents the various results of these thresholding methods, respectively, on the image obtained after applying Gaussian blur operation. After thresholding, morphological opening operation is applied on the image to fill in the gaps left after thresholding. Morphological operations are some of the simplest operations that can be performed on an image based on its shape. There are two basic morphological operations: erosion and dilation, which are generally used on binary images. These morphological operations are implemented using a structuring element, also known as kernel. As soil erosion erodes the soil, similarly erosion operation also erodes the boundaries of the foreground object. Here, kernel slides over the image and a pixel value in the original image is considered as "1" if every pixel value inside the kernel window is "1"'; otherwise, the pixel value is considered as '0'. On the other hand, dilation is opposite of erosion. A pixel value in the original image is considered to be "1" if any pixel value inside the kernel is "1." So, it increases the size of foreground object. Here, Fig. 5 shows image obtained after applying morphological opening operation on the image obtained after Gaussian blur and Otsu's adaptive Gaussian thresholding method.

– *Feature extraction* Texture refers to an object's appearance and the characteristic of its surface which arises from the proportion of the elementary parts, arrangement, shape, density, and size of the object. Texture feature extraction is the collection of such features through texture analysis process. In this paper, we have considered texture feature

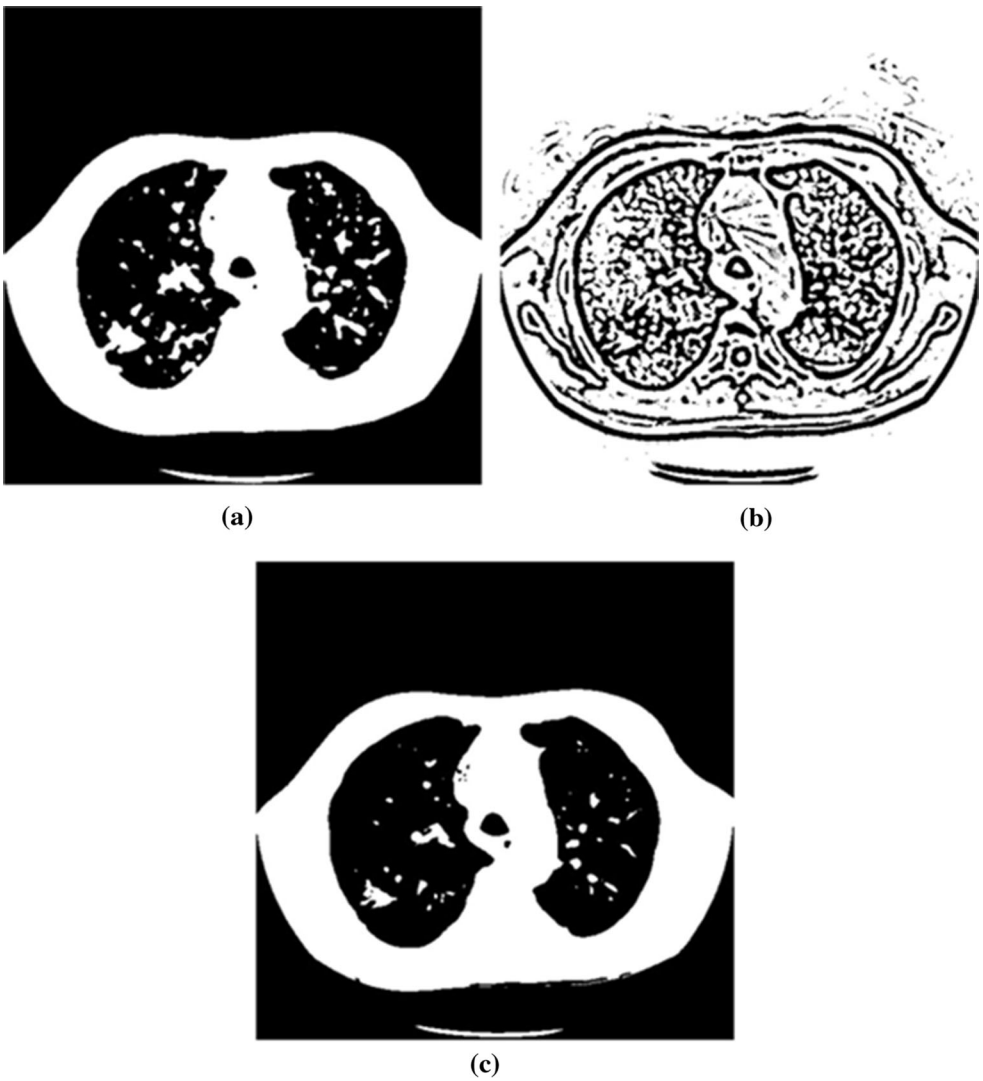

Fig. 4 Applying thresholding methods on obtained image after Gaussian blur operation **a** global thresholding, **b** Otsu's method with adaptive mean thresholding, and **c** Otsu's method with adaptive Gaussian thresholding

extraction using two methods: (1) Using gray-level co-occurrence matrix and (2) using statistical parameters.

Here, GLCM can be applied to different texture feature analysis. It can be easily used for extracting second-order texture information from the images. The texture feature calculation provides a measure for the variation in intensity at the pixel of interest and can be calculated using GLCM. Two parameters are used for computing the GLCM known as (1) relative distance between the pixel pair "d" measured in pixel number and (2) the relative orientation of the pixel pair "d." The value of the parameter Θ is quantized in the following four directions (e.g., 0°, 45°, 90°, and 135°).

– *GLCM features* GLCM has a total of fourteen different features, but among them the most useful features are: contrast, dissimilarity, homogeneity, correlation, angular second moment (ASM), and energy are considered in this paper.

Contrast is the measure of the intensity contrast between a pixel and its neighbor in the entire image [11]. It favors the contributions from $p(i, j)$ away from the diagonal, i.e., $i \neq j$ [11] and can be represented as:

$$\text{CONTRAST} = \sum_{n=0}^{G-1} n^2 \left\{ \sum_{i=1}^{G} \sum_{j=1}^{G} p(i,j) \right\}, \quad |i-j| = n \tag{1}$$

Dissimilarity is defined as the sum of pixel values where $p(i, j)$ is the absolute difference between i and j and can be represented as:

Fig. 5 Image obtained after morphological opening operation

$$\text{DISSIMILARITY} = \sum_{i=0}^{M}\sum_{j=0}^{N} p(i,j)|i-j| \tag{2}$$

Homogeneity or inverse difference moment (IDM) Inverse difference moment because of the denominator in the following equation (i.e., $(1+(i-j)^2)$) will receive small contributions from a homogeneous area (i.e., where $i \neq j$) and can be expressed as:

$$\text{HOMOGENEITY} = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} \frac{1}{1+(i-j)^2)} p(i,j) \tag{3}$$

This results in high value of IDM for homogeneous images and relatively low IDM value for inhomogeneous images.

Correlation is a measure of gray-level linear dependence between the pixels at the specified positions relative to each other and can be expressed as:

$$\text{CORRELATION} = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} \frac{\{i \times j\} \times p(i,j) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y} \tag{4}$$

Angular second moment (ASM) is the measure of homogeneity of an image. Any homogeneous scene contains very few gray levels; thus, the corresponding GLCM will have only a few but high values of $p(i, j)$. Thus, the sum of squares will be high. ASM is expressed as:

$$\text{ASM} = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} p(i,j))^2 \tag{5}$$

Energy of the image after applying GLCM is represented as:

$$\text{ENERGY} = \sqrt{\text{ASM}} \tag{6}$$

Statistical features From the region of interest we have extracted six statistical parameters, namely standard deviation, skewness, kurtosis, fifth and sixth central moments, root-mean-square, and mean.

Mean It is the average of all pixel intensity values and can be expressed as [11]:

$$\mu = \frac{1}{MN}\sum_{i=0}^{M}\sum_{j=0}^{N} p(i,j) \tag{7}$$

where $p(i, j)$ is the value of pixel intensity at the point (i, j), and $M \times N$ is the size of the image.

Standard Deviation: is the estimation of mean square deviation of the gray pixel value $p(i, j)$ from its mean value μ [11].

$$\mu = \sqrt{\frac{1}{MN}\sum_{i=0}^{M}\sum_{j=0}^{N}(p(i,j)-\mu)^2} \tag{8}$$

Skewness characterizes the degree of asymmetry of a pixel distribution in the specified window around its mean. It is a pure number, and it characterizes only the shape of the distribution [11].

$$S = \frac{1}{MN}\sum_{i=0}^{M}\sum_{j=0}^{N}\left[\frac{p(i,j)-\mu}{\sigma}\right]^3 \tag{9}$$

Kurtosis measures the peakness or flatness of a distribution relative to a normal distribution and can be represented as [11]:

$$K = \frac{1}{MN}\sum_{i=0}^{M}\sum_{j=0}^{N}\left[\frac{p(i,j)-\mu}{\sigma}\right]^4 \tag{10}$$

Fifth and sixth central moment is a moment of a probability distribution of a random variable about the random variable's mean and measures the deviation of the value associated with that particular variable from the mean value of its probability distribution. The fifth and sixth central moments can be presented as follows [11]:

$$\text{Fifth Central Moment} = \frac{1}{MN}\sum_{i=0}^{M}\sum_{j=0}^{N}\left[\frac{p(i,j)-\mu}{\sigma}\right]^5 \tag{11}$$

$$\text{Sixth Central Moment} = \frac{1}{MN}\sum_{i=0}^{M}\sum_{j=0}^{N}\left[\frac{p(i,j)-\mu}{\sigma}\right]^6 \tag{12}$$

Root-mean-square error It measures the error between the predicted value and the known value. Root-mean-square error can be calculated as:

$$rms = \sqrt{\frac{p(i,j)^2}{M \times N}} \tag{13}$$

All the above-mentioned features are extracted for every patient in the dataset along with the class label for each patient and are stored in a matrix format. Once the image analysis gets over, image classification is carried out using various machine learning algorithms. In the following step we have employed various machine learning algorithms to perform image classification.

2. *Image classification* It is the task of classifying the input image into two classes where "0" represents benign tumor, and '1' represents malignant tumor. If the whole available dataset is used as the training set, the classifier will simply memorize the dataset and provides 100% accuracy on the dataset, but as soon as new images are included in the dataset the classifier performs poorly. If we divide the dataset into training and testing set and continuously test on the testing dataset while tweaking the parameters, without reshuffling both the training and testing set, then the test set will bleed into the training set and is memorized by the classifier again which will result in over-fitting of data; hence, the performance of the system on a completely new dataset would be poor. In order to resolve these issues we have divided the dataset into three different datasets for the purpose of training, testing, and validation. Training dataset is used to train the classifiers, whereas the testing dataset is used to test classifiers performance and for tweaking various parameters in order to achieve good performance results. Both training and testing datasets are randomly shuffled after each tweaking of parameters. Validation dataset is kept aside and is only used for final system performance analysis. This is done in order to avoid the problem of overfitting, as there is a chance of training dataset bleeding into testing dataset, even after continuous shuffling. This also avoids the problem of classifier memorizing the dataset and using that to predict the class of new instances, as the classifier would never have seen the validation dataset before it is provided to it for prediction. The training set consists of 70% ($\approx$ 15,750 images) of the total images in the dataset, whereas both the test set and the validation set each consist of 15% ($\approx$ 3373 images) of the total images in the dataset. Here, we have used seven supervised machine learning algorithms to perform classification tasks, known as *K*-nearest neighbors (KNN) classifier, support vector machine (SVM) classifier, decision tree (DT) classifier, multinomial naive Bayes classifier, stochastic gradient descent (SGD) classifier, random forest classifier, and multilayer perceptron (MLP) classifier.

4 Proposed algorithm

A new image processing technique has been proposed for extracting various features from the images of lung cancer dataset. These features were then used by various supervised machine learning algorithms to detect and classify cancerous mass present. The proposed image pre-processing pipeline helps the machine learning algorithms to better predict the presence of cancerous mass. The results presented in Fig. 6 show the classification of accuracy of each of the seven machine learning algorithms along with dependency of each of the classifiers on the hand-crafted features. As presented in Algorithm 1, input to the proposed algorithm consists of three sets, namely cancerous image set (I), label set (L), and position set (P_x, P_y). The cancerous image set consists of CT scan images of patient's lungs, the label set contains labels for each image in the set "I" marked as either benign (0) or malignant (1), and the position set contains the (x, y) coordinates of cancerous mass present in each image of the set "I."

Algorithm 1 Detection and classification of lung cancer

Input:
I - Cancerous Image Set
L - Label set of each image identifying it as benign or malignant
Position set $P(P_x, P_y)$ - coordinates of cancer location in each image
Output:
Classified Image (O)- classification of cancer either as benign or malignant
Procedure $Select\_ROI(I,\ L,\ P)$
new_image: empty folder containing cropped images
$count \leftarrow 0$
for Image in set (I, L, P) **do**
  $x \leftarrow P_x, y \leftarrow P_y$
  while $(x \leq P_x + 130)$ and $y \leq P_y + 130)$ **do**
    $x \leftarrow x + 1$;
    $y \leftarrow y + 1$;
    $new\_image \leftarrow crop\_image(image, x, y, P_x, P_y)$;
    $count \leftarrow count + 1$;
  end while
end for
end Procedure
Procedure $Extract\_features(new\_image, count, L))$
$features\_matrix[]$: empty list containing features of all cropped images
$train\_set \leftarrow unique.new\_image[0.8 \times count]$;
$labels\_train \leftarrow unique.L[0.8 \times count]$;
$validation\_set \leftarrow unique[new\_imagetrain\_set]$;
$labels\_validation \leftarrow unique[L - labels\_train]$;
for Image in new_image) **do**
  $de\_noise \leftarrow Gaussian\_Blur(new\_image)$;
  $thresh \leftarrow Otsu\_Adaptive\_Thresholding(de\_noise)$;
  $Open \leftarrow Morph\_Opening(thresh)$;
  Append $feature\_extraction(open)tofeatures\_matrix$;
end for
end Procedure
Procedure $Classification(features\_matrix[], labels\_train, labels\_validation))$
$features\_train\_set \leftarrow features\_matrix$;
$features\_validation\_set \leftarrow features\_matrix$;
for feature in $feature\_matrix.row\_countandlabel\_train.row\_count$ **do**
  Append $labels\_train(feature)tofeatures\_train\_set(feature)$;
  $train\_classifier(features\_train\_set)$;
  $classifer\_accuracy \leftarrow test\_classifier(features\_train\_set, labels\_train).score$;
  $classifier\_prediction \leftarrow test\_classifier(features\_validation\_set, labels\_validation)$;
end for
end Procedure
return $classifier\_accuracy, classifier\_prediction$

The first step for detection and classification of the cancerous mass present in the lungs is to extract the region of interest (ROI) present in the image. It is that region or area in the input image where the cancerous mass is present. Every image in the set "I" consists of variable size of cancerous mass. Therefore, a uniform area of 130×130 pixels is selected from each input image and ROI is extracted using the Select_ROI method. The coordinates for the center of ROI in the image are taken from the position set corresponding to that image. Once ROI is extracted for each image in set "I," three image processing steps are used to enhance the quality and visibility of the extracted ROI using Extract_features method. These steps include Gaussian blur for denoising the image, followed by Otsu's method with adaptive thresholding, and morphological opening operation. The output of the morphological opening operation "open" is given to the feature_extraction function, and a total of fourteen different texture and statistical features are calculated for each image. The output of the Extract_features method is stored in a matrix called features_matrix and provided to the classifiers for the task of detection and classification. The classification method takes three inputs, namely features_matrix (a matrix consisting of features extracted for the set "I"), labels_train (containing labels corresponding to train_set), labels_validation (containing labels corresponding to validation_set) and returns, classifier_accuracy and classifier_prediction variables. Firstly, the features_matrix is copied to two matrices called

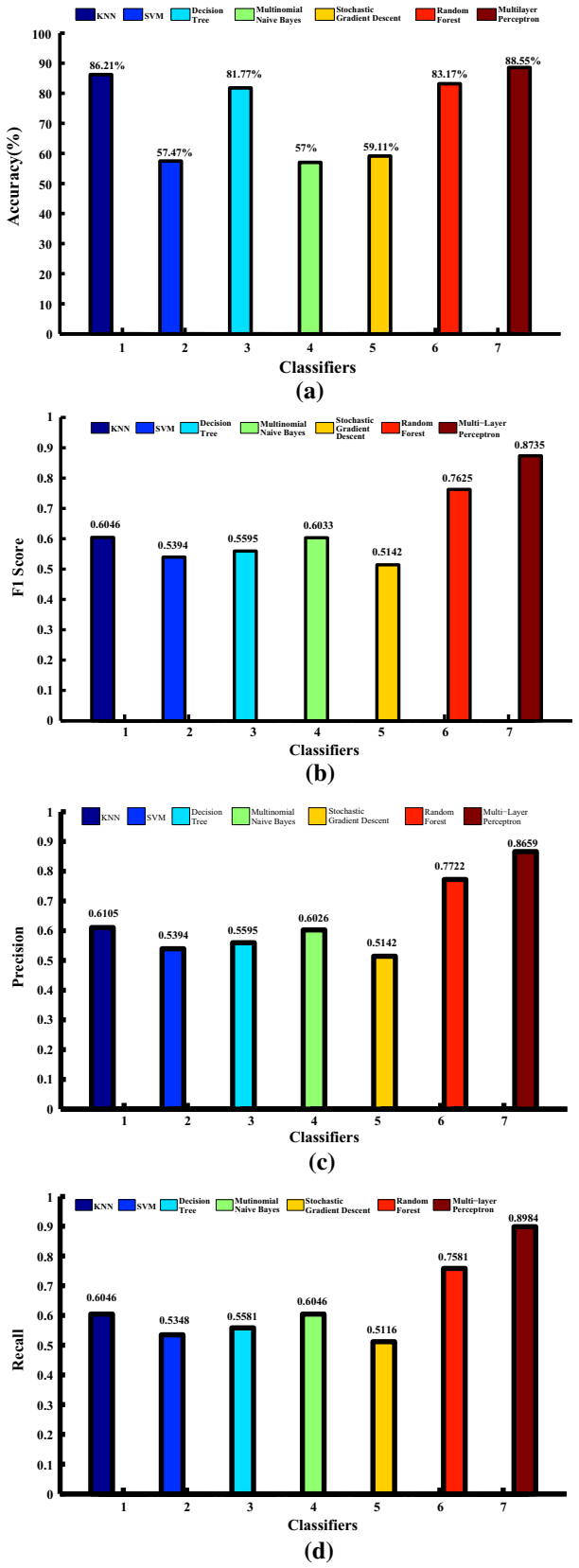

Fig. 6 Performance evaluation of classifiers by using **a** accuracy percentage, **b** *F*1 score, **c** precision, and **d** recall

features_train_set, and features_test_set. For all the features extracted for each of the training image and presented in the features_validation_set matrix, labels from labels_train are added to the features_train_set according to the corresponding image. This matrix is further used for training the classifiers. The function test_classifier (features_train_set, labels_train) calculates the accuracy of each of the trained classifiers and stores the result in the classifier_accuracy. This value is then used for tweaking the parameters of the classifier in order to increase the overall accuracy of the classifier. The function test_classifier(features_validation_set, labels_test) predicts whether cancerous mass present images provided in the test set is benign (0) or malignant (1) by comparing the output of the function with the labels of the corresponding image present in the labels_validation set, and stores the result in the classifier_prediction.

5 Performance evaluation

The dataset used included 512×512 pixels images, categorized over 2 classes—benign and malignant. The total size of the dataset is ≈ 11.2 GB. For each image containing the cancerous portion region of interest was extracted which was a generalized area of 130×130 pixels. The center for this area was obtained from the CSV file provided with the dataset. This complete setup was run using OpenCV for image processing tasks, along with scikit-learn used as a machine learning library written in python. Here, Table 1 represents the details of experimental setup. We have used 64-bit Intel Xeon E5 3.2-GHz processor for extraction and processing of images. Once ROI is extracted from each image, all the above-mentioned seven machine learning algorithms are evaluated at the configured experimental setup on the basis of the following four parameters which in turn is calculated using the confusion matrix.

– *Accuracy* is a statistical measure of how well a classifier correctly identifies or excludes a condition. The accuracy is the percentage of true results (both true positive and true negative) in the given dataset [11].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{14}$$

– *F1 score* is a measure of classifiers accuracy. It uses both precision and recall to calculate the score and can be represented as:

$$F_1\text{Score} = \frac{\text{Precision.Recall}}{\text{Precision} + \text{Recall}} \tag{15}$$

– *Precision* measures the total number of positive cases which the algorithm identifies as per the following [11]:

Table 1 Experimental setup

S. no.	Testbed parameter	Value
1	Operating system	Microsoft Windows 10 Pro
2	Processor	Intel Xeon E5 3.2 Ghz
3	Architecture	64-bit architecture
4	Hard disk	1 TB
5	RAM	8 GB
6	GPU	NVIDIA Quadro P1000
7	GPU memory	4 GB GDDR5
8	CUDA parallel-processing cores	640
9	Programming language used	Python
10	Libraries used	Scikit-learn, Pandas, NumPy

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{16}$$

– *Recall* also known as sensitivity and measures the percentage of actual positive cases which the algorithm correctly identifies. That is the percentage of the images containing a benign or malignant nodule correctly classified by the algorithm as benign or malignant.

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

where

– *True positive (TP)* Images containing benign/malignant, classified as benign/malignant.
– *False negative (FN)* Images containing benign/malignant, classified as malignant/benign.
– *True negative (TN)* Images not containing benign/malignant classified as not containing benign/malignant.
– *False positive (FP)* Images not containing benign/malignant classified as containing benign/malignant.

As obtained after simulations, Fig. 6a–d represents accuracy percentage, F_1 score, precision, and recall of each of various mentioned machine learning algorithms for the given dataset. All the seven classifiers were trained by using the previously discussed GLCM feature set containing all 14 features. Further, performance evaluation using the above-mentioned four parameters for measuring the performance of a classifier was then calculated on subset of the total dataset. This was done so that a uniform result can be obtained, without the possibility of overfitting the dataset. From Fig. 6a–d, we can find that the highest classification accuracy percentage of 88.55% was obtained by using multi-layer perceptron neural network. The multi-layer perceptron mode used consisted of three layers, namely input layer, hidden layer, and output layer, along with highest F_1 score, precision, and recall values of 0.8735, 0.8659, and 0.8984, respectively. Out of 14 GLCM features, extracted in the feature extraction stage the effect of each feature on the overall accuracy percentage, F_1 score, precision, and recall, for each classifier was measured by removing that feature from the feature set and then training the classifier on the remaining features in the feature set.

Here, Tables 2, 3, 4, 5, 6, 7, and 8 represents the effect of removing features from the feature set given to each

Table 2 Applying KNN classifier for removing listed features on the accuracy percentage, *F*1 score, precision, and recall

Features removed	Accuracy (%)	*F*1 score	Precision	Recall
Mean	89.71	0.7450	0.7530	0.7441
Standard deviation	86.91	0.5067	0.5068	0.5116
Skewness	89.48	0.5041	0.5248	0.5116
Kurtosis	88.78	0.4872	0.4946	0.4883
Fifth central moment	89.25	0.5707	0.5976	0.5581
Sixth central moment	89.95	0.5081	0.5052	0.5116
Entropy	85.95	0.4400	0.4401	0.4418
Root-mean-square	89.95	0.5079	0.5322	0.5116
Contrast	88.08	0.4677	0.5079	0.4418
Dissimilarity	88.78	0.5100	0.5089	0.5116
Homogeneity	83.87	0.4418	0.4418	0.4418
Energy	87.14	0.5393	0.5471	0.5348
Correlation	87.38	0.5467	0.5671	0.5348
ASM	84.81	0.5771	0.5753	0.5813

Table 3 Applying SVM classifier for removing listed features on the accuracy percentage, $F1$ score, precision, and recall

Features removed	Accuracy (%)	$F1$ score	Precision	Recall
Mean	57.24	0.3906	0.3875	0.3953
Standard deviation	58.87	0.5150	0.5190	0.5116
Skewness	58.64	0.5606	0.5637	0.5581
Kurtosis	57.70	0.3595	0.3835	0.3488
Fifth central moment	53.07	0.3265	0.3287	0.3255
Sixth central moment	55.84	0.4296	0.4489	0.4418
Entropy	57.71	0.6693	0.6673	0.6744
Root-mean-square	56.54	0.5813	0.5813	0.5813
Contrast	54.67	0.6033	0.6080	0.6046
Dissimilarity	55.37	0.3973	0.4149	0.3953
Homogeneity	56.77	0.5348	0.5348	0.5348
Energy	56.54	0.5745	0.5910	0.5813
Correlation	55.60	0.5348	0.5348	0.5348
ASM	55.37	0.5113	0.5111	0.5116

Table 4 Applying decision tree classifier for removing listed features on the accuracy percentage, $F1$ score, precision, and recall

Features removed	Accuracy (%)	$F1$ score	Precision	Recall
Mean	80.84	0.5343	0.5341	0.5348
Standard deviation	78.27	0.6033	0.6026	0.6046
Skewness	79.67	0.6221	0.6274	0.6279
Kurtosis	78.73	0.5174	0.5938	0.5116
Fifth central moment	80.37	0.4905	0.5276	0.4883
Sixth central moment	80.14	0.5566	0.5558	0.5571
Entropy	80.14	0.5556	0.5536	0.5581
Root-mean-square	79.20	0.5122	0.6126	0.4883
Contrast	80.37	0.4331	0.4319	0.4418
Dissimilarity	77.10	0.5516	0.5616	0.5581
Homogeneity	78.50	0.5758	0.6214	0.5581
Energy	78.27	0.5138	0.5165	0.5116
Correlation	74.76	0.5100	0.5134	0.5116
ASM	76.63	0.4723	0.4837	0.4651

Table 5 Applying multinomial naive Bayes classifier for removing listed features on the accuracy percentage, $F1$ score, precision, and recall

Features removed	Accuracy (%)	$F1$ score	Precision	Recall
Mean	51.40	0.5482	0.5537	0.5581
Standard deviation	60.28	0.5348	0.5348	0.5348
Skewness	51.63	0.5323	0.5320	0.5348
Kurtosis	57.24	0.4418	0.4418	0.4418
Fifth central moment	48.83	0.5800	0.5811	0.5813
Sixth central moment	55.14	0.6059	0.6079	0.6046
Entropy	48.59	0.5576	0.5577	0.5581
Root-mean-square	57.00	0.6007	0.6027	0.6046
Contrast	57.47	0.6538	0.6915	0.6511
Dissimilarity	60.28	0.4732	0.5061	0.4651
Homogeneity	53.73	0.5813	0.5813	0.5813
Energy	55.14	0.4878	0.4893	0.4883
Correlation	49.53	0.5230	0.5202	0.5348
ASM	51.16	0.4799	0.4909	0.4883

Table 6 Applying stochastic gradient descent classifier for removing listed features on the accuracy percentage, $F1$ score, precision, and recall

Features removed	Accuracy (%)	$F1$ score	Precision	Recall
Mean	57.71	0.4089	0.4105	0.4186
Standard deviation	58.87	0.5312	0.5299	0.5348
Skewness	50.93	0.4662	0.4679	0.4651
Kurtosis	60.48	0.6069	0.6098	0.6046
Fifth central moment	59.57	0.3754	0.3890	0.3720
Sixth central moment	59.11	0.5586	0.5596	0.5581
Entropy	57.06	0.4571	0.5209	0.4418
Root-mean-square	59.57	0.5827	0.5866	0.5813
Contrast	64.01	0.4963	0.5124	0.4883
Dissimilarity	62.38	0.4883	0.4883	0.4883
Homogeneity	52.10	0.6129	0.6269	0.6046
Energy	55.37	0.6711	0.6707	0.6744
Correlation	54.20	0.4532	0.4716	0.4418
ASM	54.43	0.5981	0.6084	0.6046

Table 7 Applying random forest classifier for removing listed features on the accuracy percentage, $F1$ score, precision, and recall

Features removed	Accuracy (%)	$F1$ score	Precision	Recall
Mean	84.81	0.8492	0.8719	0.8348
Standard deviation	84.57	0.8494	0.8930	0.8581
Skewness	84.57	0.9279	0.9279	0.9279
Kurtosis	82.24	0.8105	0.8436	0.7744
Fifth central moment	82.24	0.8643	0.8904	0.8581
Sixth central moment	85.04	0.6700	0.6716	0.6720
Entropy	80.60	0.7639	0.7632	0.7651
Root-mean-square	80.37	0.8319	0.8354	0.8581
Contrast	84.34	0.7412	0.7425	0.7418
Dissimilarity	81.54	0.7633	0.7620	0.7651
Homogeneity	85.74	0.6893	0.6991	0.6953
Energy	79.67	0.8018	0.7997	0.8046
Correlation	79.67	0.7171	0.7202	0.7348
ASM	85.28	0.8411	0.8497	0.8311

Table 8 Applying multi-layer perceptron classifier for removing listed features on the accuracy percentage, $F1$ score, precision, and recall

Features removed	Accuracy (%)	$F1$ score	Precision	Recall
Mean	88.55	0.8681	0.8695	0.8916
Standard deviation	87.14	0.8116	0.8121	0.8116
Skewness	91.12	0.8116	0.8121	0.8116
Kurtosis	87.61	0.7230	0.7500	0.7186
Fifth central moment	91.58	0.8556	0.8556	0.8581
Sixth central moment	88.78	0.8706	0.8635	0.8916
Entropy	89.01	0.8748	0.8629	0.8916
Root-mean-square	89.95	0.8079	0.8327	0.8116
Contrast	90.65	0.7787	0.7959	0.7651
Dissimilarity	90.42	0.8219	0.8367	0.8116
Homogeneity	87.85	0.8443	0.8626	0.8348
Energy	86.91	0.8856	0.8923	0.8813
Correlation	84.34	0.8836	0.8847	0.8813
ASM	89.71	0.7128	0.4116	0.4186

classifier for classification. Also, Tables 2, 3, 4, 5, 6, 7, and 8 represents that accuracy is calculated in percentage, whereas F_1 score, precision, and recall values are calculated on the scale of 0–1.

6 Conclusion

In this paper, we have applied image processing and machine learning approaches for detection and classification of lung cancer. The techniques have been categorized and implemented in five different stages known as image acquisition, image pre-processing and segmentation, feature extraction, image classification, and performance evaluation. Simulation results were obtained using four different parameters known as accuracy, F_1 score, precision, and recall. Here, obtained results represent that multi-layer perceptron or neural networks can be applied for detection and classification of lung cancer CT scan images that claims high accuracy and precision value in comparison with other. The total number of hidden layers present in the multi-layer perceptron model was 3 with 100 hidden units in each layer. Rectified linear unit (ReLU) was used as the activation function for this model along with the L2 regularization ($\alpha = 3.0$), and Limited Memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) as the solver.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Cancer Research UK (2017) Cancer mortality for common cancers. http://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality/common-cancers-compared. Accessed May 2017
2. Dimililer K, Ugur B, Ever YK (2017) Tumor detection on CT lung images using image enhancement. Online J Sci Technol 7(1):133–138
3. Al-tarawneh MS (2012) Lung cancer detection using image processing techniques. Leonardo Electron J Pract Technol 20:147–58
4. Armato III SG, Hadjiiski L, Tourassi GD, Drukker K, Giger ML, Li F, Redmond G, Farahani K, Kirby JS, Clarke LP (2015) SPIE-AAPM-NCI Lung nodule classification challenge dataset. The Cancer Imaging Arch. https://doi.org/10.7937/K9/TCIA.2015.UZLSU3FL
5. Gonzalez RC, Woods RE (2002) Digital image processing. Prentice Hall, Upper Saddle River, NJ, pp 797–800
6. Dwivedi MS, Borse MR, Yametkar MA (2014) Lung cancer detection and classification by using machine learning and multinomial Bayesian. IOSR J Electron Commun Eng (IOSR-JECE) 9(1):69–75
7. Sun W, Zheng B, Lure F, Wu T, Zhang J, Wang BY, Saltzstein EC, Qian W (2014) Prediction of near-term risk of developing breast cancer using computerized features from bilateral mammograms. Comput Med Imaging Graph 38(5):348–357
8. Chaudhary A, Singh SS (2012) Lung cancer detection on CT images by using image processing. In: Proceedings of 2012 IEEE international conference on computing sciences (ICCS). pp 142–146
9. Pratap GP, Chauhan RP (2016) Detection of Lung cancer cells using image processing techniques. In: Proceedings of IEEE international conference on power electronics, intelligent control and energy systems (ICPEICES). pp. 1–6
10. Bhusri S, Jain S, Virmani J (2016) Classification of breast lesions based on laws' feature extraction techniques. In: Proceedings of 3rd international conference on computing for sustainable global development (INDIACom). pp. 1700–1704
11. Kuruvilla J, Gunavathi K (2014) Lung cancer classification using neural networks for CT images. Comput Methods Programs Biomed 113(1):202–209
12. Mitra S, Pal SK (1995) Fuzzy multi-layer perceptron, inferencing and rule generation. IEEE Trans Neural Netw 6(1):51–63
13. Amato F, Lpez A, Pea-Mndez EM, Vahara P, Hampl A, Havel J (2013) Artificial neural networks in medical diagnosis. J Appl Biomed 11:47–58
14. Karabatak M, Ince MC (2009) An expert system for detection of breast cancer based on association rules and neural network. Expert Syst Appl 36(2):3465–3469
15. Adi K, Widodo CE, Widodo AP, Gernowo R, Pamungkas A, Syifa RA (2017) Nave Bayes algorithm for lung cancer diagnosis using image processing techniques. Adv Sci Lett 23(3):2296–2298
16. Joachims T (1998) Making large-scale SVM learning practical (No. 1998, 28). In: Technical Report, SFB 475: Komplexittsreduktion in Multivariaten Datenstrukturen, Universitt Dortmund, pp 1–18
17. Tidke SP, Chakkarwar VA (2012) Classification of lung tumor using sVM. Int J Comput Eng Res 2(5):1254–1257
18. Touw WG, Bayjanov JR, Overmars L, Backus L, Boekhorst J, Wels M, van Hijum SA (2012) Data mining in the life sciences with random forest: A walk in the park or lost in the jungle? Brief. Bioinform. 14(3):315–326
19. Shi T, Seligson D, Belldegrun AS, Palotie A, Horvath S (2005) Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. Mod Pathol 18(4):547–557
20. Ramos-Gonzlez J, Lpez-Snchez D, Castellanos-Garzn JA, de Paz JF, Corchado JM (2017) A CBR framework with gradient boosting based feature selection for lung cancer subtype classification. Comput Biol Med 86:98–106
21. Sakamoto M, Nakano H, Zhao K, Sekiyama T (2017) Multi-stage neural networks with single-sided classifiers for false positive reduction and its evaluation using Lung X-ray CT images. ArXiv preprint arXiv:1703.00311, pp 1–11
22. Demyanov S, Chakravorty R, Abedini M, Halpern A, Garnavi R (2016) Classification of dermoscopy patterns using deep convolutional neural networks. In: Proceedings of 13th international symposium on biomedical imaging (ISBI). pp 364–368
23. Lopez AR, Giro-i-Nieto X, Burdick J, Marques O (2017) Skin lesion classification from dermoscopic images using deep learning techniques. In: Proceedings of 13th IASTED international conference on biomedical engineering (BioMed). pp 49–54
24. Bewal R, Ghosh A, Chaudhary A (2015) Detection of breast cancer using neural networks a review. J Clin Biomed Sci 5(4):143–148
25. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin PM, Larochelle H (2017) Brain tumor

segmentation with deep neural networks. Med image Anal 35:18–31
26. Weng S, Xu X, Li J, Wong ST (2017) Combining deep learning and coherent anti-Stokes Raman scattering imaging for automated differential diagnosis of lung cancer. J Biomed Opt 22(10):106017
27. Sun W, Zheng B, Qian W (2017) Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis. Comput Biol Med 89:530–539
28. Mahbod A, Ecker R, Ellinger I (2017) Skin lesion classification using hybrid deep neural networks. arXiv preprint arXiv:1702.08434, pp 1–5
29. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118
30. Behrmann J, Etmann C, Boskamp T, Casadonte R, Kriegsmann J, Maass P (2017) Deep learning for tumor classification in imaging mass spectrometry. Bioinformatics 1:1–10