IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Prediction and Classification of Lung Cancer Using Machine Learning Techniques

To cite this article: Pragya Chaturvedi *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1099** 012059

Prediction and Classification of Lung Cancer Using Machine Learning Techniques

Pragya Chaturvedi[1], Anuj Jhamb[1], Meet Vanani[1] and Varsha Nemade[1]

[1]Department of Computer Engineering Mukesh Patel School of Technology Management and Engineering Shirpur, Maharashtra, India

E-mail: anuj.jhamb23@gmail.com

Abstract In all the disease that have existed in mankind lung cancer has emerged as one of the most fata one time and again. Also, it is one of the most common and contributing to deaths among all the cancers. Cases of lung cancer are increasing rapidly. There are about 70,000 cases per year in India. The disease has a tendency to be asymptomatic mostly in its earlier stages thus making it nearly impossible to detect. That's why early cancer detection plays an important part in saving lives. An early detection can give a patient a better chance to cure and recover. Technology plays a major role in detecting cancer efficiently. Many researchers have proposed different methods based on their studies. In recent times, to use computer technology to solve this problem, several computer-aided diagnosis (CAD) techniques as well as system have been proposed, developed as well as emerged. Those systems use various Machine learning techniques as well as deep learning techniques, there also have been several methods based off of image processing-based techniques to predict the malignancy level of cancer. Here, in this paper, the aim will be focussed onto list, discuss, compare and analyse several methods in image segmentation, feature extraction as well as various techniques to classify and detect lung cancer in there early stages.

1. Introduction

In 2018 it was estimated that approximately 9.6 million deaths were claimed by lung cancer. Lung cancer tops the list if a person talks about the types and their shares. Estimated cases of lung cancer are around 2.09 million with 1.76 million deaths which account for around 84% deaths [1]. Due to this reason lung cancer has been entitled as one of the most fatal diseases. Tumor is made by multiplication of abnormal cells in lung cancer. Cancer cells tend to spread really fast due to blood streams ans lymph fluid that is present in lung tissue. In general, due to normal lymph flow, cancer cells frequently migrate to the middle of the chest. As cancer cells migrate to other tissues, metastasis occurs. It is important that cancer be detected as early as possible as it tends to spread and is beyond curable in case of a larger spread. It is difficult to diagnose lung cancer since it shows symptoms in the final stage and it is nearly impossible to save a person's life in the final stage. Images of lungs for examination are captured by imaging techniques such as Computed Tomography (CT), Positron Emission Tomography (PET), Magnetic resonance imaging (MRI) and X-ray. CT image technique is the most common out of the mentioned methods due to its ability to give a view excluding overlapping structures. Interpreting and recognizing cancer is complicated for doctors. CT photographs are accurate for the diagnosis of lung cancer. To identify lung cancer, image processing, and deep learning methods will be used. Accuracy can be improved using these approaches. Tumour detection and determination of its form, size, and location is a tough task. Timely detection helps in saving a lot of time. And this time can be used in providing early treatment to the patient. In this project, pre-processing (removing noise if any), post-processing (segmentation) and classification techniques will be used to classify tumors into one of the two groups i.e. Malignant and Benign. Benign refers to a non-cancerous tumor and it doesn't spread to other parts. Abnormal cells divide without control in malignant and may invade surrounding tissues. Exploring different methods to diagnose lung cancer will be a prime aim in this paper. Computed tomography can be used to capture images of lungs across various dimensions so that a 3D image of the chest can be formed. This 3D image can be used to detect tumors present. Normally a doctor or any field expert uses a CT image to detect cancer. Due to the large number of CT images, it is difficult for a doctor or radiologist to detect cancer quickly and accurately. But with the advancement in technology, Computer-Aided Diagnosis (CAD) can be utilized to complete this duty efficiently and in considerably less time. This process has two separate processes i.e. first to identify all the nodules present in the CT image and second to classify the detected lung nodules. In general, a CAD system comprises the following steps which are shown below in figure 1.

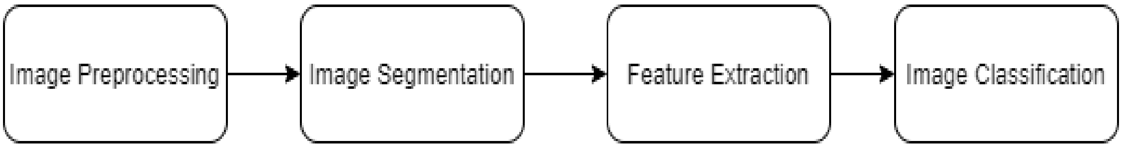

Figure 1. Basic Steps Involved in a CAD System

1.1. Image Pre-Processing

A CAD system cannot directly use CT images. They need to be well pre-processed before the actual use. Various Image pre-processing techniques are used to discard noise and to make images suitable for use. This helps in the betterment of the performance of the whole system and hence the accuracy. Various image pre-processing techniques are listed down in table 1.

Table 1. Comparison of Image Pre-Processing Techniques

Reference	Technique	Use
S. S. Kanitkar et al. [2]	Gaussian and Gabor filtering	Gaussian is used to blur the image to reduce noise and Gabor is used for texture analysis
M. Vas et al. [3]	Median filtering	Removal of salt & pepper noise
K. Punithavathy et al. [4]	CLAHE	Enhancement of the local contrast
D. Sharma et al. [5]	Wiener filtering	Uses linear time-invariant (LTI) of an observed process to generate an approximation of a desired random process.
A. Asuntha et al. [6]	Adaptive bilateral filtering	Sharpness enhancement & noise removal
A. Teramoto et al. [7]	Gaussian and Convolutional edge enhancement filtering	Enhances the local discontinuities in the picture at the boundaries of various objects (edges)
O. Ozdemir et al. [8]	Adaptive Gaussian Filtering	Removal of Gaussian noise

1.2. Image Segmentation

The method of partitioning an image into several segments is known as image segmentation. Segmentation of image is done majorly to find boundaries in the given image. The process of analyzing the image becomes easier as segmentation reduces the image complexity [9]. Table 2 contains various segmentation techniques along with the no. of sample images used.

Table 2. Comparison of Image Segmentation Techniques

Reference	Segmentation Technique	No. of sample images
S. S. Kanitkar et al. [2]	Watershed Transform Marked-Controlled watershed transform Thresholding & Marker Controlled	14 CT Images
M. Vas et al. [3]	Morphological Operations	216 CT Images (128-train & 88-test)
D. Sharma et al. [5]	Sobel Edge Detection	1000 CT Images
A. Asuntha et al. [6]	K-Means, FCM, and Ant Colony algorithms	1000 CT Images
M. Saric et al. [10]	Region of Interest	33 CT Images (25-train & 8-test)
J. Alam et al. [11]	Watershed transform	500 CT Images
T. Aggarwal et al. [12]	Region of Interest	240 CT Images (90-train & 150-test)
F. V. Farahani et al. [13]	Thresholding & Region Growing	60 CT Images
M. S. Rahman et al. [14]	Otsu Thresholding	1000 CT Images

1.3. Feature Extraction

Feature Extraction is a method by which we aim at reducing the number of dimensions that our raw data contains so that it is easier to process and is in a form of manageable classes. Variables in a huge number requiring computational resources in order to process and produce results is characteristic for the massive amounts of data. Feature Extraction techniques deal with simplifying the data while at the same time ensuring that no data is lost. These techniques are responsible for picking and merging the features to minimize the amount of data. Table 3 contains various feature extraction techniques along with features considered while applying that technique.

Table 3. Comparison of Feature Extraction Techniques

Reference	Feature Extraction Technique	Elements/Features Considered
M. Vas et al. [3]	GLCM	Haralick features
J. Alam et al. [11]	GLCM	standard deviation, Mean, fluctuation, entropy, smoothness, IDM, kurtosis, vitality, relationship, differentiate, homogeneity, RMS
T. Aggarwal et al. [12]	GLCM	Contrast, correlation, variance, homogeneity
R. Fakoor et al. [15]	PCA	PCA Features
W. Chen et al. [16]	3D & 2D CNN, Hybrid features fusion model	Volumetric and 2d Features
M. B. Rodrigues et al. [17]	Structural Co-Occurrence Matrix (SCM)	Statistical, Information, Divergence
Y. Xie et al. [18]	Multi-View Knowledge-Based Collaborative (MV-KBC)	Cross-entropy
A. Asuntha et al.	ROI Extraction	Volumetric (Zernike moment, SIFT), texture (Wavelet & LBP), intensity (HOG), geometric (Eccentricity & Curvature descriptor) features

1.4. Image Classification
Classification of images is a basic task that seeks to interpret a picture as a whole. By assigning it to a particular label, the purpose is to identify the image. Image Classification usually refers to images where only one object appears and is examined. Object identification, on the other hand, requires both classification and localization tasks and is used to examine more practical instances in which an image may have several objects. Here the task is to classify lung nodules as malignant or benign. Various classification techniques are listed below in table 4 along with the results obtained.

Table 4. Comparison of Image Classification Techniques

Reference	Classification Technique	Results
A. Asuntha et al. [6]	Fuzzy Particle Swarm Optimisation (FPSO) and CNN	Accuracy - 95.62 Sensitivity - 96.23 Specificity - 95.89
M. B. Rodrigues et al. [17]	MLP, SVM, KNN	Accuracy: MLP - 95.40 SVM - 96.70 KNN - 95.30

Yutong Xie et al. [18]	Knowledge-based Collaborative Deep Learning	Accuracy - 91.60
Gian Son Tran et al. [19]	2D Deep Convolutional Network	Accuracy - 97.20 Sensitivity - 96.00 Specificity - 97.30
Margarita Kirienko et al. [20]	CNN	Accuracy: Validation - 69 Testing - 69 Training - 87 Dice Score: Training - 82 Testing - 80
Moritz Schwyzer et al. [21]	Transfer learning	Accuracy - 97.10 Sensitivity - 95.90 Specificity - 98.10
S. Shanthi et al. [22]	Stochastic diffusion search algorithm & Neural Networks (SDS-NN)	Accuracy - 89.63
Ibrahim M. Nasser et al. [23]	Artificial Neural Network (ANN)	Accuracy - 96.67
Xufeng Huang et al. [24]	Deep Transfer Convolutional Neural Network (DTCNN) and Extreme Learning Machine (ELM)	Accuracy - 94.57
A. Poreva et al. [25]	Decision Tree and SVM	Accuracy: DT - 72 SVM - 75
M. F. Serj et al. [26]	dCNN	Sensitivity - 87 Specificity - 99.1 F1 Score - 95

2. Literature Survey

2.1. CNN

In 2019, Moradi et al. [27] compared different techniques to differentiate lung cancer nodules from non-nodules. To reduce/eliminate the false positive predictions they have come up with 3D Convolutional Neural Network Technique. Nodules exist in different sizes and using just one CNN can result in false detections. So they divided the nodules into four groups according to their size. And they have used four different sizes of 3D CNN. They combined all those 4 classifiers to get better results. Each CNN consists of a nubmber of 3D CNN which are all varying sizes. All 4 classifiers were combined in order to produce results which were better. A combination of Max pooling layer and convolutional layer were used to produce each CNN. The activation function used here is ReLU. Softmax layer accompanied by a fully connected layer is used to produce the output finally. Nodules size varies from 3mm to 3cm so by using just one layer, the prediction could be wrong for either very small nodules or very large values. So they fused all the 4 CNNs and sent their output values (predicted values) to a final classifier. They have chosen a logistic regression classifier that takes inputs from 4 CNNs and produces a final prediction. They have implemented logistic regression by using a decision tree classifier and gradient boosting model. LUNA16 dataset was used in this to train the complete model. LUNA16 is based on the CT images of the LIDC dataset. As a result, they saw that the result by the fused classifier is better than each of the solo classifiers.

In 2018, Bohdan Chapliuk et al. [4] applied neural networks C3D and 3D DenseNet to detect lung cancer using CT images. These Neural networks were applied to whole lung 3D images and two-stage approaches (for segmentation and classification, two different neural networks are trained.) and further compared. Data Science Bowl 2017 dataset containing CT scans of more than 1000 patients was used. For pre-processing all the CT images were converted into Household Units (HU is a unit describing x-ray intensity) by resampling. HU ranges are specific to tumors (-500) so, in the second step, a range for lung tissue that filters out all bones from the image was filtered out by all patient images. The size of the 3D patient image was reduced to 120x120x120. The results for both 3CD and 3D DenseNet are quite similar to 3D DenseNet performing slightly better. The outcome shows that Neural Networks trained on whole lung 3D images performed poorer compared to two-stage approaches.

In 2019, Ruchita Tekade et al. [28]. proposed a method using 2 architectures, one for the segmentation of nodules and the second one to determine the malignancy level. For determining the malignancy level CNN is used for classification as well as for the feature extraction, max pooling is used for sub pooling, ReLU as the activation function, and softmax is the classifier used to perform the classification and assign malignancy level. Adam classifier is used to optimize weight selection in convolutional kernels. For the segmentation of CT scanned images, pre-processing is done using simple thresholding, clear border, morphology erosion, morphology closing, and morphology opening respectively. Using U-Net segmentation masses are generated for lung CT scan images and lung nodules are segmented. This experiment was conducted on LIDC-IDRI, LUNA16, and Data Science Bowl2017 datasets. This approach gives an accuracy of 95.66% and loss 0.09 and dice coefficient of 90%and for predicting log loss 38% using U-Net to segment and further predict malignancy levels.

In 2019, A. Asuntha et al. [6]. indicated a method to detect and classify cancerous tissues using a deep learning approach. CT images were used from LIDC and private datasets as input and Histogram Equalisation was used to enhance the contrast value. The adaptive Bilateral filtering technique was used to denoise the CT images. The artificial Bee Colony segmentation algorithm was used to segment the image to extract the ROI. In total 180 features were extracted (20 Zernike, 1 Curvature, 18 SIFT, 1 Eccentricity, 26 wavelets, and 18 HOG) using Local Binary Pattern and some wavelet techniques. Fuzzy Particle Swarm Optimization was used to select the most important feature and to reduce the complexity of the CNN model which is then used to classify the extracted nodule as benign or malignant. The average accuracy, specificity, and sensitivity of the suggested model are 95.62%, 95.89%, and 96.23% respectively.

In 2018, Margarita Kirienko et al. [29] suggested a CNN-based approach with 69%, 69%, and 87% accuracy invalidation, test, and training sets respectively. Tumour, Node, Metastasis (TNM) staging was used to stage lung cancer from 1 to 4. Fluorodeoxyglucose positron emission tomography (FDG-PET)/ Computed Tomography (CT) images were used as input. These images were classified into either T1-T2 or T3-T4 using CNN. The system was developed using two networks - a classifier and a feature extractor. The feature extractor was used for relevant features that are to be extracted and a classifier was used to classify the patch. The experiment was performed on 472 patients (T1-T2 = 353 and T3-T4 = 119).

In 2020, QINGHAI ZHANG et al. [30] proposed a method for designing of Lung nodule detection system which is automatic. The dataset used for the proposed method is LIDC-IRDI public dataset. The proposed method used for this study is Multi-Scene Deep Learning Framework which contains several steps. CT images are given as input and the probability distribution of distinct gray levels is obtained by threshold segmentation that is Histogram. Correcting the smooth lung outlines is the main aim for the lung parenchyma segmentation process. The replacement of the vein system in the lung helps to identify the nodule structure. Vessel filters are used for removing the vessels which reduce the number of false positive. The design of CNN contains a pooling layer, a convolutional layer, and a fully integrated layer. Segmentation and classification identify Class 1 and Class2 that are two class of image data and discrete images which are separated from the lung images respectively [31]. Segmentation is done to identify cancerous tumor cells in lungs. The accuracy of the determined nodules is determined by four different

IOP Conf. Series: Materials Science and Engineering 1099 (2021) 012059 doi:10.1088/1757-899X/1099/1/012059

types of CNN architecture. In 2020, Mesut Togacar et al. [32] proposed a CNN-based technique to detect lung cancer. They have taken in a total of 100 images (50 cancerous and 50 others) belonging to 69 different patients. Due to less number of images, augmentation was used to get a healthy dataset. AlexNet, LeNet, and VGG-16 CNNs were used as a part of the study. Stochastic Gradient Descent was used as an optimization method (for AlexNet and VGG-16) to update the weights for each training set. Other than this, RMSProp and ADAM were also used as the optimization methods (for LeNet). mRMR algorithm was used to extract the features. Some traditional machine learning models such as LR, LDA, SVM, KNN, and DT, are also used after the CNN architectures. The performance was improved by using the Principal Component Analysis method. 99.51 accuracy was obtained by choosing KNN with CNN & mRMR.

In 2019, Samaiya Dabeer et al. [33] proposed the diagnosis of cancer in a histopathological image using CNN based approach. Original Data Set (UC Irvine Machine Learning Repository), MITOS- ATYPIA-14, and BreakHis. The BreakHis database network has been utilized. The model was trained using 2480 benign and 5429 malignant samples belonging to the RGB color model. Therefore, the proposed system depicted in Fig. 2 classifies breast tissue as being either benign or malignant by an effective classification model. To begin with, the implementation of the deep net by processing the images in the dataset is done. Redundancy has to be reduced in the data as it contributes to complexities in networks and is obsolete. The precision for the benign and malignant classes are found to be 90.55% and 94.66%, respectively.

In 2019 Pouria Moradi et al. [27]. specified the use of 3D CNN in order to reduce the false possitives. The network weights are started with Xavier weight initialization. To train network weights with learning rate 0.01 and 10^{-5} decay per epoch with 0.9 momenta, stochastic gradient descent is used. For a Meta classifier combination of three decision trees that were trained. The dataset used to train and evaluate the system is LUNA 16. The system gives 91.23% accuracy for 3.09 false positives.

In 2017, Qi Dou et al. [31] proposed a method so as to decrease the number of false possitives that were detected in pulmonary nodule detection. The dataset used for the proposed method is LUNA 16. The proposed method used in this study is 3D CNN with Multilevel Contextual and it has several layers. For extraction of a stack of high-level representation, the 3D convolutional layer is used by sweep over the input image. Subsampling of 3D features and filling the invariance to local translation in 3D space is done by a 3D max-pooling layer. The fully connected layer has the neurons which have a denser connection that benefits like a stronger representation of capability of extracted representation. 87% of sensitivity with 4 false positives is achieved.

In 2018, Mehdi Fatan Serj et al. [26]. proposed a method to detect lung cancers efficiently by using deep CNN-based techniques. They have developed a network that is composed of 2 max-pooling layers, 3 convolutional layers, a softmax layer (binary), and a fully connected layer. The method was tested on the dataset available by Kaggle for Kaggle Data Science Bowl 2017 Competition. Deep CNN based model worked better than other CNN based models [34][35]. For the loss function, cross-entropy was used to maximize the multinomial logistic regression objective and hence maximizing the probability of patients with lung cancer. They have achieved 87% sensitivity and 99.1% specificity.

In 2018, Anum Masood et al. [36] proposed a method to detect symptoms and lung cancer in the early stages by using IoT and CNN based approach. They have proposed an IOT based system that comprises smart wearable devices and some symptoms charts which can be used to check if the patient is showing any relevant symptoms and hence can alarm the doctor. CT images of such patients were then put as an input to the CNN model. Gabor filter was used as a pre-processing method. Thresholding was used to get the Region of Interest. DFCNet was used as the main classification model. The proposed model gave 86.02% accuracy, 83.91% sensitivity, and 80.59% specificity on the LIDC-IDRI dataset. This experiment was also conducted on some other datasets as well including a real-time dataset from a hospital.

Albert Chon et al. [37]. Proposed a method consisting of deep neural network techniques to detect cancer in its earlier stages. The datasets used were Kaggle's Data Science Bowl 2017 dataset and LUNA16. In the pre-processing phase, pixel values of the CT images are first converted into Hounsfield units and then thresholding was used for segmentation. After segmentation, normalization of the 3D image was carried out to map values between 0 and 1. Down sampling of 0.5 units in all three dimensions was done.

Finally, zero-centering was done by subtracting the mean value of all the images from the training dataset. Instead of directly inputting the segmented images into the classifier, a U-Net was trained by using the LUNA16 dataset, and then it was used for effective segmentation by detecting the exact location of nodules. Linear classifiers, 3D-CNN, and 3D Googlenet models were used as additional classifiers to decrease the false-positive values. 3D Googlenet performed best out of three with an accuracy of 75.1%, Sensitivity of 77%, Specificity of 74.1%, and AUC of 75.7%. The main point that concludes was that the model was trained on less number of the labeled dataset so it can further be generalized to all forms of cancers.
In 2017, Wafaa Alakwaa et al. [38] proposed a 3D CNN-based approach to detect lung cancers. Kaggle Data Science Bowl and LUNA16 datasets were used. LUNA16 one was to train the U-Net model to detect lung nodules as lung nodules were not labeled in the Kaggle dataset. Segmentation, downsampling, Normalisation, and Zero Centering were performed in the image pre-processing phase. The pixel values of the CT images were first translated to Hounsfield units and then for segmentation, thresholding was used. After segmentation, 3D image normalization was carried out with the goal of mapping values between 0 and 1. Downsampling of 0.5 units has been performed in all three dimensions. Finally, zero-centering was achieved by subtracting the mean value of the images from the training dataset. A U-Net was trained using the LUNA16 dataset instead of entering the segmented images directly into the classifier, to detect the exact position of nodules. Accuracy, false-positive rate, Mis-Classification rate and false-negative rate were found to be 86.6%, 11.9%, 13.4% and 14.7% respectively. The basic architecture of 3D CNN is shown in figure 3.

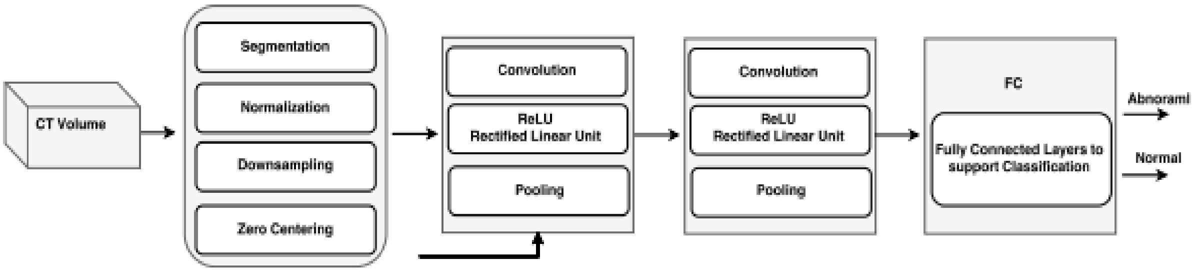

Figure 2. 3D convolutional neural network architecture [17]

In 2017, Atsushi Teramoto et al. [7] proposed a method to classify Lung Cancer Types which is automated from Cytological Images using Deep CNN. Image dataset is used which contains Seventy-six (76) cases of cancer cells. Data augmentation is done on the images which are obtained by microscope and have the sharpness of the targeted cells which varies and are direction-invariant. Gaussian filter and convolutional edge enhancement filters are used for filtering. All the details like filter size, the stride of each layer are specified. The architecture has three layers that are two fully connected layers, a convolutional layer, and three pooling layers.70% of the classification is done correctly using DCNN.

2.2. Others

In 2019, Aicha Majda et al. [39] proposed four different feature extraction methods namely CNN, PCA, Restricted Boltzmann Machines (RBM), and 2D-DFT with which they did a comparative study. For further evaluation of which method gave the best performance on three hidden layers of a neural network was used. LIDC-IDRI dataset was used to train these neural networks. Lung nodule regions are extracted in patches from the CT scan using a descriptive file followed by data augmentation to enhance the volume of the data set. CNN proved to be achieving better results compared to other methods in this experiment. Other than CNN 2D-DFT closely resembled the results in terms of accuracy, it suffered a case of high variance and bias. This bias and variance in 2D-DFT eventually rose the amount of overfitting due to correlations being missed in between features extracted and outputs kept as target.

In 2015, K.Punithavathy et al. [4] . explained lung cancer detection based on texture features and Fuzzy C means. The paper mainly concentrates on the image pre-processing parts using different techniques to get better results and a clustering method to generate the outcome. In the pre-processing part, to increase the contrast present in the Computed Tomography Images (CT images), Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied. Instead of applying this technique to the whole image, it is applied to small regions of the images known as tiles. Bilinear interpolation is used to combine the different enhanced parts/regions of the image. Wiener filters are used to reduce the noise by a significant amount. Region extraction plays an important role to get the desired region. Morphological operations such as closing were used to get the desired region i.e. region having lung lobes and leaving behind the blood vessels, bronchi, and all other internal parts. The structuring element of disc shape was used in the closing operation. While in the feature extraction process, texture-based features were concentrated as intensity value is not the right parameter to extract features. The classification of the pre-processed image is done using FCM. FCM is chosen as it retains important features of the image. FCM classifier is based on unsupervised learning. Figure 4 represents the process flow of the technique proposed in this paper [40].

Figure 3. Process Flow of the system [9]

In 2019, P. Mohamed Shakeel et al. [41] proposed two methods for the detection of lung cancer from CT images. The dataset used for this study is Cancer imaging Archive (CIA) dataset. Deep Learning trained neural network and Improved Profuse clustering are used in this study. CT images contain low-quality images and it has noise, so to remove all this, CT image pre-processing is done. For improving the image quality, Image histogram techniques are used as it is a very efficient method on different images Segmentation of cancer affected regions is done with the help of improved CT image using IPCT. The improved profuse clustering technique is applied to segment cancer influenced parts from the improved lung CT image. For detecting inconsistency in the image pixels, two procedures of improved profuse techniques work as it checks the image pixel and puts the similar superpixel in the same group. Predicting the similitude of data using the pixel eigenvalue is done during the process of segmentation when the pixels are continuously examined. Different features of spectral that are standard deviation, 3rd-moment skewness, mean, and 4th-moment kurtosis are derived from the region which are segmented and which is forwarded for the feature extraction stage as it is very effective to spot lung cancer which has connected features. 98.42% accuracy is ensured by the system with minimum classification error to be 0.038. The Deep Learning network is shown in figure 5.

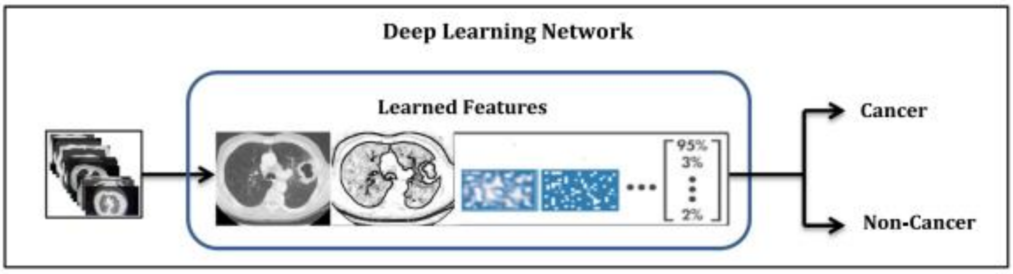

Figure 4. Deep learning training process structure [42]

In 2020, S. Shanthi et al. [22] proposed a system consisting of a stochastic diffusion search algorithm (SDS) and classification algorithms such as Neural networks, Decision trees, and Naive Bayes to detect lung cancer. 270 images (140 normal and 130 abnormal) from a dataset named TCGA were acquired and used. Grey level co-occurrence matrix (GLCM) was applied so as to extract the features of texture. The Gabor filter was used for shape-based features. SDS algorithm was used for feature selection. It has mainly 4 phases - Initialisation phase (assignment of agents to some random hypotheses), Evaluation phase (evaluate the fitness value to find the maximum), Test Phase (Active Agent if: current agent's fitness value> random agent's fitness value, in any other case Inactive Agent) and Diffusion phase (select a random agent if the current agent is inactive else copy the hypothesis of the current agent and offset it). After applying SDS, different classification methods were applied. After observing the accuracies of all the classification models, Neural Network along with the SDS algorithm (SDS-NN) proved to perform better as compared to others. An observation was made implying that the classification of images improves with improved feature selection.
In 2015, N.D Thombore et al. [2]. proposed the Marker-Controlled Watershed Transform method for the detection of Lung cancer. On the input image, a Gaussian filter is applied and it helps remove noise which is a very effective method. The Gaussian filter also removes high-frequency components from the image thresholding and marker-controlled watershed segmentation is used to convert a grayscale image into a binary image. Below or above the particular threshold value is assigned by the two levels to the pixel. 100% accuracy is achieved compared to the other thresholding algorithm.

2.3. SVM
In 2019, Sanjukta Rani Jena et al. [42] proposed a method that focuses on texture analysis based on feature extraction of images and then classifying them. In image pre-processing, several filters are used to remove the unnecessary noise and stabilize the image. In the feature extraction part, shaped based FETs (Area, Perimeter, Median, Mean, and Variance) and intensity-based FETs (Contrast, Uniformity, Homogeneity) are used. Then the local binary pattern (LBP) is used for texture matching. The performance of LBP is better than other available textual patterns. Then an SVM classifier is used for classification. A hyperplane is chosen such that it maximizes the margin (the distance between a few close points and the hyperplane).
In 2019 Nidhi S. Nadakarni et al. [43] proposed an automated system for lung cancer detection at an early stage. CT images from the Cancer Image Archive Database were used in DICOM format. These images were then pre-processed using various image enhancement techniques such as Median Filtering, Smoothening, and Contrast Adjustment to remove noise and improve image quality. Further Morphological opening operations were performed after transforming the grayscale image into a binary image for image segmentation. In the feature extraction method features like area, perimeter, and eccentricity (roundness) are evaluated. Using these features classification of images is done into normal and abnormal using SVM supervised learning classifier. The proposed methodology as said by the authors detects cancer in the early stages accurately. Figure 6 shows the cancerous lung CT image and its histogram representation.

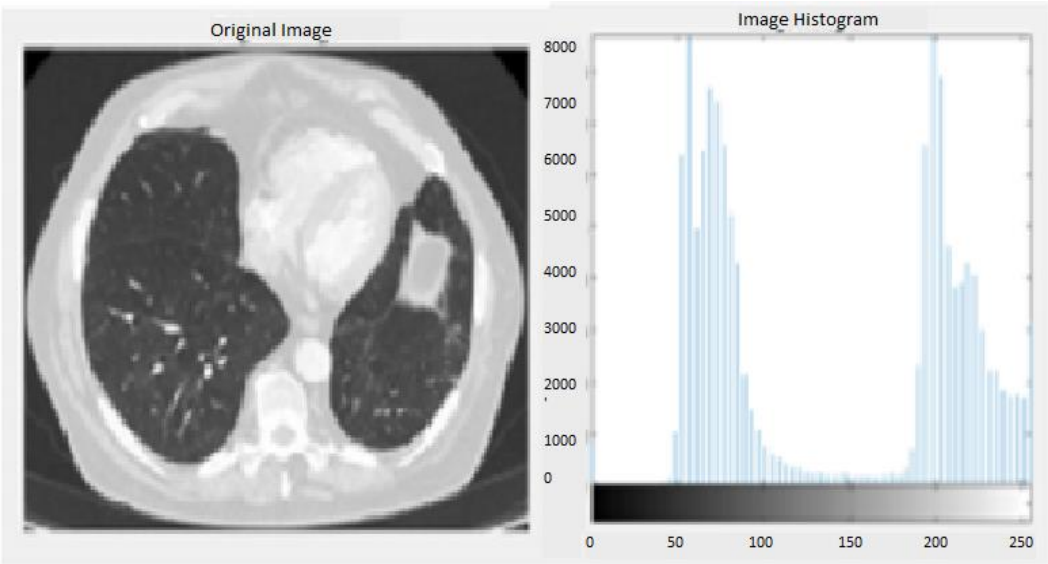

Figure 5. Cancerous lung CT image and histogram of the image [3].

2.4. ANN

In 2017 Amita Desai et al. [3] adopted a method using an artificial neural network (ANN) classifier to predict lung cancer. Images from Manipal Hospital in Goa, V.M.Salgaocar Hospital and SMRC were used. The cropped image is converted into binary which reduces the coputational complexity that arises and also decreases the storage issues. it also prepares the image for further morphological operation. After successful pre-processing feature extraction is done by resizing and applying the HAAR wavelet transform, then GLCM is calculated in different directions extracting 7 features from them. The next seven Haralick features are extracted. Then a forward neural network is fed using a backpropagation algorithm. The training accuracy attained is 96% while for testing was 92%. They also achieved 88.7% sensitivity and 97.1% specificity.

In 2019, Ibrahim M. Nasser et al. [23] proposed to detect the absence or presence of lung cancer using ANN. To diagnose the disease, symptoms were used. The dataset used is as described in the Table. The ANN model predicted the presence of lung cancer with 96.67% accuracy, and with less than 1% training error rate after 1418105, it also deduced that the factor that has the highest impact on results was “Age”. Table 5 contains the importance factor of various attributes of the dataset on the presence of lung cancer.

Table 5. Dataset Description with their respective importance [39]

Reference	Input Name	Importance
1	Age	123.2382
0	Gender	26.2635
11	Coughing	26.2357
9	Wheezing	23.2983
2	Smoking	22.4438
6	Chronic Disease	21.4716
3	Yellow Fingers	20.5510

10	Alcohol Consuming	20.1778
7	Fatigue	19.7445
5	Peer Pressure	18.2220
4	Anxiety	17.9241
8	Allergy	16.0747
14	Chest Pain	14.5559
13	Swallowing Difficulty	10.9188
12	Shortness of Breath	10.4047

In 2018, Moritz Schwyzer et al, [21]. proposed the Deep Neural Networks method using ultralow dose PET/CT of detection of lung cancer which is automated. Data used here contains a total of 100 patient's entries 50 of which are having cancer while the other 50 are not patients of lung cancer. The binary classification was performed on slices where the lung tumor of patients are present visually and slices of patients which does not have any lung cancer. The residual neural network was performed for training purposes by classifying lung cancer. 97.1% accuracy, 95.9% sensitivity, 98.1% specificity were obtained. Table 6 shows a comparison of various methodologies and performance results.

Table 6. Comparison of Different Methodologies and their Results

Reference	**Dataset Used**	**Methodology**	**Results**
Rebecca L et al. [8]	2017 Data Science Bowl on Kaggle, LUNA 16	3D Probabilistic Deep Learning, V-Net architecture	CADe sensitivity-96.5% average false positives-19.7 CADx AUC-0.87
Maja Stella et al. [10]	ACDC LUNGH	VGG16, ResNet50, CNN	Accuracy 97.9 93
Janee Alam et al. [11]	UCI machine learning database	Watershed Transform, GLCM	Identification- 97 Cancer Prediction- 87
M. S. Rahman et al. [14]	The Cancer Imaging Archive (TCIA)	Gaussian Blur, Otsu Threshold, MobileNet, Inception-V3, VGG-8	Best Achieved amongst three Neural Networks Accuracy-97%, Specificity-97.85%, Sensitivity-96.26%

W. Chen et al. [16]	134 CT scans from Shandong Cancer Hospital	3D and 2D CNN, Hybrid features fusion module (HFFM)	Dice score - 88.8 Sensitivity - 87.2 Precision - 90.9
M. B. Rodrigues, et al. [17]	LIDC-IDRI	Laplace, Gaussian & Sobel filtering, multilayer perceptron, SVM, KNN, SCM Mean HU	Accuracy (SCM Mean HU) - 96.70
Yutong Xie et al. [18]	LIDC-IDRI	Knowledge-based Collaborative Deep Learning, U-Net, 3D-GLCM-SVM	Accuracy-91.60% Sensitivity-86.52% Specificity-94% AUC-95.70%
Gian Son Tran et al. [19]	LIDC-IDRI	2D Deep Convolutional Network	Accuracy-97.2 Sensitivity-96.0 Specificity-97.3
Xufeng Huang et al. [24]	LIDC-IDRI, First Affiliated Hospital of Guangzhou Medical University in China(FAM-GMU)(Number of entries=115)	Extreme Learning Machine (ELM) andDeep Transfer Convolutional Neural Network (DTCNN)	Accuracy-94.57%
Zhou Liu et al. [30]	Private Dataset	DQN, H-DQN, CNN	-
Qi Dou et al. [35]	LUNA16, Kaggle Data science Bowl2017	3D CNN,	Sensitivity-87% Specificity- 99.1%
P. Mohamed Shakeel [41]	deep learning instantaneously trained neural networks Improved profuse clustering (IPCT)	Cancer Imaging Archive (TCIA)	Accuracy-98.42%
H. Xie et al. [44]	LUNA16(Testing),	2DCNN R-CNN(Detection Of Nodules)	AUC-0.954
X. Li et al. [45]	LIDC-IDRI, General Hospital Of Guangzhou Military Command	Anisotropic nonlinear diffusion filter, Random Walker(RW), Random Forest(RF), GLCM, LBP, Gabor Filter	Sensitivity- 0.92 Specificity-0.83 Accuracy-0.90 AUC-0.95
W. Sun et al. [46]	LIDC-IDRI	CNN, Deep Belief Network, Restricted Boltzman Machine, Stacked Denoising Autoencoder	Accuracy-0.822 AUC-0.818
S. Wang et al. [47]	LIDC-IDRI, Guangdong General Hospital	CF-CNN	Mean DSC%-81.66±0.05

M. Nishio et al. [48]	The Cancer Imaging Archive(TCIA)	SVM or XGBoost	AUC-0.850 Accuracy-0.797
W. Zhu et al. [49]	LUNA16	3D DPN, 10 fold cross-validation, 3D Faster R-CNN	Accuracy-81.42%
L. M. Pehrson et al. [50]	LIDC-IDRI	Feature-Based Framework, Support Vector Machine, GLMR, ELM, PNN, ANN, DBN, D Architecture	Accuracy-90%
H. Wei et al. [51]	SCLC patients Shandong Cancer Hospital (dataset of 134 patients)	Neighborhood gray-tone difference matrices, Spatial gray-level dependence matrices, Gray Level Histogram Analysis	AUC-0.797
X. Huang et al. [52]	LIDC-IDRI	Faster R-CNN	Accuracy - 91.4
P. P. R. Filho et al. [53]	Walter Cantidio Hospital	Spatial Interdependence Matrix, Visual Information Fidelity Optimum-path forest (OPF) classification	Accuracy - 98.2 F-score - 95.2
W. Shen et al. [54]	LIDC-IDRI	Multi-crop Convolutional Neural Network (MC-CNN)	Accuracy - 87.14
J. Jiang et al. [55]	TCIA, MSKCC, LIDC	 U-Net SegNet FRRN Increment MRRN Desne MRRN	Sensitivity: 0.80 0.77 0.76 0.85 0.82
P. P. R. Filho et al. [56]	40 chest CT images	3D Adaptive Crisp Active Contour Method (3D ACACM)	F-measure - 99.22 ± 0.14
K. Yu et al. [57]	Kaggle Science Bowl dataset	Lung mask, lung segmentation 2d & 3D ResNet, U-Net, VGG-Net CNN, tree-based classifiers	-
N. Khosravan et al. [58]	LUNA16	Morphological operations, ADAM optimizer Semi-Supervised Multi-Task Learning	DSC - 91 Sensitivity - 98
S. Makaju et al. [59]	LIDC-IDRI	Median Filter, Gaussian Filter Watershed Segmentation SVM	Accuracy-92%, Specificity-50%, Sensitivity-100%
S. Baek et al. [60]	96 PET/CT images of	U-Net	-

	NSCLC patients		
Bohdan Chapliuk et al. [61]	Data Science Bowl 2017	C3D, 3D DenseNet	-
Rachid Sammouda et al. [62]	A database consisting of 3D CT images	Unsupervised Modified Hopfield Neural Network Classifier	-
Hongyoon Choi et al. [63]	NCBI GEO(Gene Expression Data)(11 microarray dataset)	Weighted Gene Coexpression Network Analysis, Cox regression, CNN, Kaplan Meier Method,	<C-index- 0.709±0.042>
Chip M. Lyncha et al. [64]	SEER database	Gradient Boosting Machines (GBM), Supervised Machine Learning Techniques like Decision Trees, Support Vector Machines (SVM)	-
Dipanjan Moitra et al. [65]	Cancer Imaging Archive (TCIA)	1D CNN	Accuracy-96 ± 3%
QingZeng Song et al. [66]	LIDC-IDRI	CNN, DNN, SAE	Accuracy-84.15%
P.K Gupta et al. [67]	Private	GLCM, SVM, KNN, Decision Tree, MLP, SGD, Stochastic Gradient, RF classifier, Bayes Classifier	-
Brahim AIT SKURT et al. [68]	LIDC-IDRI	U-Net	Dice Coefficient- 0.9502
Mingyang Lu et al. [69]	Dataset of patients from Third Affiliated Hospital of Soochow University.	Min-Redundancy Max-Relevance (mRMR), Risk Ovarian Malignancy Algorithm, Logistic Regression, Decision Tree	-

3. Conclusion

One of the most fatal diseases to have existed is lung cancer. This disease unfortunately is extremely tough to treat after having spread upto an extent or reaching a serious stage. Computer-Aided Detection (CAD) is one of the constantly growing technologies that help detect cancer by feeding in certain inputs containing patient-related information such as scans like CT-Scan, X-Ray, MRI Scan, unusual symptoms in patients or biomarkers, etc. SVM, CNN, ANN, Watershed Segmentation, Image enhancement, Image processing are a few methods used to improve the accuracy and aid the process. For training, the most popular datasets used are LUNA16, Super Bowl Dataset 2016, and LIDC-IDRI. By the means of this review paper, we aim to list out all the major researches that have been done over the past years and can be improved upon to achieve better results.

4. References

[1]. World Health Organisation‘s Official website https://www.who.int/news-room/fact-sheets/detail/cancer#:~:text=The%20most%20common%20causes%20of,Lung%20(1.76%20 million%20deaths)
[2]. Kanitkar SS, Thombare ND and Lokhande SS 2015 Detection of lung cancer using marker-controlled watershed transform *Int. Conf. on Pervasive Computing* (ICPC) pp. 1-6.
[3]. Vas M and Dessai A 2017 Lung cancer detection system using lung CT image processing *Int. Conf. on Computing, Communication, Control and Automation* (ICCUBEA) pp. 1-5.
[4]. Punithavathy K, Ramya MM and Poobal S 2015 Analysis of statistical texture features for automatic lung cancer detection in PET/CT images *Int. Conf. on Robotics, Automation, Control and Embedded Systems* (RACE) pp. 1-5.
[5]. Sharma D and Jindal G 2011 Identifying lung cancer using image processing techniques *Int. Conf. on Computational Techniques and Artificial Intelligence* (ICCTAI) vol. 17 pp. 872-880.
[6]. Asuntha A, and Srinivasan A 2020 Deep learning for lung Cancer detection and classification. Multimedia Tools and Applications pp 1-32.
[7]. Teramoto A, Tsukamoto T, Kiriyama Y and Fujita H 2017 Automated classification of lung cancer types from cytological images using deep convolutional neural networks *BioMed research International*.
[8]. Ozdemir O, Russell RL and Berlin AA 2019 A 3D Probabilistic Deep Learning System for Detection and Diagnosis of Lung Cancer Using Low-Dose CT Scans *IEEE Transactions on Medical Imaging* 1419-29.
[9]. Huang PW, Lin PL, Lee CH and Kuo CH 2013 A classification system of lung nodules in CT images based on fractional brownian motion model *Int. Conf. on system science and engineering* (ICSSE) pp. 37-40.
[10]. Šarić M, Russo M, Stella M and Sikora M 2019 CNN-based method for lung cancer detection in whole slide histopathology images *Int. Conf. on Smart and Sustainable Technologies* (SpliTech) pp. 1-4.
[11]. Alam J, Alam S and Hossan A 2018 Multi-stage lung cancer detection and prediction using multi-class svm classifier *Int. Conf. on Computer, Communication, Chemical, Material and Electronic Engineering* (IC4ME2) pp. 1-4.
[12]. Aggarwal T, Furqan A and Kalra K 2015 Feature extraction and LDA based classification of lung nodules in chest CT scan images *Int. Conf. on Advances in Computing, Communications and Informatics* (ICACCI) pp. 1189-1193.
[13]. Farahani FV, Ahmadi A and Zarandi MF 2015 Lung nodule diagnosis from CT images based on ensemble learning *IEEE Conf. on Computational Intelligence in Bioinformatics and Computational Biology* (CIBCB) pp. 1-7.
[14]. Rahman MS, Shill PC and Homayra Z 2019 A New Method for Lung Nodule Detection Using Deep Neural Networks for CT Images *Int. Conf. on Electrical, Computer and Communication Engineering* (ECCE) pp. 1-6.
[15]. Fakoor R, Ladhak F, Nazi A and Huber M 2013 Using deep learning to enhance cancer diagnosis and classification. *In Proceedings of the Int. Conf. on machine learning* vol. 28 (New York, USA).
[16]. Chen W, Wei H, Peng S, Sun J, Qiao X and Liu B 2019 HSN: hybrid segmentation network for small cell lung cancer segmentation. *IEEE Access* 75591-603.
[17]. Rodrigues MB, Da Nóbrega RV, Alves SS, Rebouças Filho PP, Duarte JB, Sangaiah AK and De Albuquerque VH 2018 Health of things algorithms for malignancy level classification of lung nodules *IEEE Access* 6 18592-601.
[18]. Xie Y, Xia Y, Zhang J, Song Y, Feng D, Fulham M and Cai W 2018 Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT. *IEEE transactions on medical imaging*. 38(4) 991-1004.
[19]. Tran GS, Nghiem TP, Nguyen VT, Luong CM and Burie JC 2019 Improving accuracy of lung nodule classification using deep learning with focal loss. *Journal of Healthcare Engineering*.

[20]. Kirienko M, Sollini M, Silvestri G, Mognetti S, Voulaz E, Antunovic L, Rossi A, Antiga L and Chiti A 2018 Convolutional neural networks promising in lung cancer T-parameter assessment on baseline FDG-PET/CT *Contrast Media & Molecular Imaging*.
[21]. Schwyzer M, Ferraro DA, Muehlematter UJ, Curioni-Fontecedro A, Huellner MW, Von Schulthess GK, Kaufmann PA, Burger IA and Messerli M 2018 Automated detection of lung cancer at ultralow dose PET/CT by deep neural networks–initial results. *Lung Cancer* 170-3.
[22]. Shanthi S and Rajkumar N 2020 Lung Cancer Prediction Using Stochastic Diffusion Search (SDS) Based Feature Selection and Machine Learning Methods. *Neural Processing Letters* 1-4.
[23]. Nasser IM and Abu-Naser SS 2019 Lung Cancer Detection Using Artificial Neural Network. *International Journal of Engineering and Information Systems* (IJEAIS) 17-23.
[24]. Huang X, Lei Q, Xie T, Zhang Y, Hu Z and Zhou Q 2020 Deep Transfer Convolutional Neural Network and Extreme Learning Machine for Lung Nodule Diagnosis on CT images *arXiv preprint arXiv* 2001.01279.
[25]. Poreva A, Karplyuk Y and Vaityshyn V 2017 Machine learning techniques application for lung diseases diagnosis *5th IEEE Workshop on Advances in Information, Electronic and Electrical Engineering* (AIEEE) pp. 1-5.
[26]. Serj MF, Lavi B, Hoff G and Valls DP 2018 A deep convolutional neural network for lung cancer diagnostic. *arXiv preprint arXiv* 1804.08170.
[27]. Moradi P and Jamzad M 2019 Detecting Lung Cancer Lesions in CT Images using 3D Convolutional Neural Networks *4th Int. Conf. on Pattern Recognition and Image Analysis* (IPRIA) pp. 114-118.
[28]. Tekade R and Rajeswari K 2018 Lung cancer detection and classification using deep learning. *Fourth Int. Conf. on Computing Communication Control and Automation* (ICCUBEA) pp. 1-5.
[29]. Sasikala S, Bharathi M, Sowmiya BR 2018 Lung Cancer Detection and Classification Using Deep CNN *International Journal of Innovative Technology and Exploring Engineering* (IJITEE) 2278-3075.
[30]. Liu Z, Yao C, Yu H and Wu T 2019 Deep reinforcement learning with its application for lung cancer detection in medical Internet of Things *Future Generation Computer Systems* pp 1-9.
[31]. Zhang Q and Kong X 2020 Design of Automatic Lung Nodule Detection System Based on Multi-Scene Deep Learning Framework. *IEEE Access*.
[32]. Toğaçar M, Ergen B and Cömert Z 2020 Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks *Biocybernetics and Biomedical Engineering* 40(1) 23-39.
[33]. Dabeer S, Khan MM and Islam S 2019 Cancer diagnosis in histopathological image: CNN based approach *Informatics in Medicine Unlocked* 16:100231.
[34]. Rossetto AM and Zhou W 2017 Deep learning for categorization of lung cancer ct images *IEEE/ACM Int. Conf. on Connected Health: Applications, Systems and Engineering Technologies* (CHASE) pp. 272-273.
[35]. Dou Q, Chen H, Yu L, Qin J and Heng PA 2016 Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection *IEEE Transactions on Biomedical Engineering* 64(7) 1558-67.
[36]. Masood A, Sheng B, Li P, Hou X, Wei X, Qin J and Feng D 2018 Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images *Journal of biomedical informatics* 79 117-28.
[37]. Chon A, Balachandar N and Lu P 2017 Deep convolutional neural networks for lung cancer detection. *Stanford University*.
[38]. Alakwaa W, Nassef M and Badr A 2017 Lung cancer detection and classification with 3D convolutional neural network (3D-CNN) *Lung Cancer* 8(8):409.
[39]. Skourt BA, Nikolov NS and Majda A 2019 Feature-Extraction Methods for Lung-Nodule Detection: A Comparative Deep Learning Study *Int. Conf. on Intelligent Systems and Advanced Computing Sciences* (ISACS) pp. 1-6.

[40]. Wu Q and Zhao W 2017 Small-cell lung cancer detection using a supervised machine learning algorithm *Int. Symp. on Computer Science and Intelligent Controls* (ISCSIC) pp. 88-91.
[41]. Shakeel PM, Burhanuddin MA and Desa MI 2019 Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks *Measurement* 145 702-12.
[42]. Jena SR, George T and Ponraj N 2019 Texture Analysis Based Feature Extraction and Classification of Lung Cancer *IEEE Int. Conf. on Electrical, Computer and Communication Technologies* (ICECCT) pp. 1-5.
[43]. Nadkarni NS and Borkar S 2019 Detection of Lung Cancer in CT Images using Image *Processing 3rd Int. Conf. on Trends in Electronics and Informatics* (ICOEI) pp. 863-866.
[44]. Xie H, Yang D, Sun N, Chen Z and Zhang Y 2019 Automated pulmonary nodule detection in CT images using deep convolutional neural networks *Pattern Recognition* 85 109-19
[45]. Li XX, Li B, Tian LF and Zhang L 2018 Automatic benign and malignant classification of pulmonary nodules in thoracic computed tomography based on RF algorithm IET Image Processing 12(7) 1253-64.
[46]. Sun W, Zheng B and Qian W 2017 Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis Computers in biology and medicine 89 530-9.
[47]. Wang S, Zhou M, Liu Z, Liu Z, Gu D, Zang Y, Dong D, Gevaert O and Tian J 2017 Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation Medical image analysis 40 172-83.
[48]. Nishio M, Nishizawa M, Sugiyama O, Kojima R, Yakami M, Kuroda T and Togashi K 2018 Computer-aided diagnosis of lung nodule using gradient tree boosting and Bayesian optimization PloS one 13(4):e0195875.
[49]. Zhu W, Liu C, Fan W and Xie X 2018 Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification *IEEE Winter Conf. on Applications of Computer Vision* (WACV) pp. 673-681.
[50]. Pehrson LM, Nielsen MB and Ammitzbøl Lauridsen C 2019 Automatic pulmonary nodule detection applying deep learning or machine learning algorithms to the LIDC-IDRI database: a systematic review Diagnostics 9(1):29.
[51]. Wei H, Yang F, Liu Z, Sun S, Xu F, Liu P, Li H, Liu Q, Qiao X and Wang X 2019 Application of computed tomography-based radiomics signature analysis in the prediction of the response of small cell lung cancer patients to first-line chemotherapy Experimental and therapeutic medicine 17(5):3621-9.
[52]. Huang X, Sun W, Tseng TL, Li C and Qian W 2019 Fast and fully-automated detection and segmentation of pulmonary nodules in thoracic CT scans using deep convolutional neural networks *Computerized Medical Imaging and Graphics* 74 25-36.
[53]. Rebouças Filho PP, da Silva Barros AC, Ramalho GL, Pereira CR, Papa JP, de Albuquerque VH and Tavares JM 2019 Automated recognition of lung diseases in CT images based on the optimum-path forest classifier *Neural Computing and Applications* 31(2):901-14.
[54]. Shen W, Zhou M, Yang F, Yu D, Dong D, Yang C, Zang Y and Tian J 2017 Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification Pattern Recognition 61 663-73.
[55]. Jiang J, Hu YC, Liu CJ, Halpenny D, Hellmann MD, Deasy JO, Mageras G and Veeraraghavan H 2018 Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images *IEEE transactions on medical imaging* 38(1) 134-44.
[56]. Rebouças Filho PP, Cortez PC, da Silva Barros AC, Albuquerque VH and Tavares JM 2017 Novel and powerful 3D adaptive crisp active contour method applied in the segmentation of CT lung images Medical image analysis 35 503-16.
[57]. Yu KH, Lee TL, Yen MH, Kou SC, Rosen B, Chiang JH and Kohane IS 2020 Reproducible Machine Learning Methods for Lung Cancer Detection Using Computed Tomography Images: Algorithm Development and Validation *Journal of medical Internet research*. 22(8): e16709.

[58]. Khosravan N and Bagci U 2018 Semi-supervised multi-task learning for lung cancer diagnosis *40th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* (EMBC) pp. 710-713.
[59]. Makaju S, Prasad PW, Alsadoon A, Singh AK and Elchouemi A. Lung cancer detection using CT scan images *Procedia Computer Science* 125 107-14.
[60]. Baek S, He Y, Allen BG, Buatti JM, Smith BJ, Plichta KA, Seyedin SN, Gannon M, Cabel KR, Kim Y and Wu X 2019 What does ai see? deep segmentation networks discover biomarkers for lung cancer survival *arXiv preprint arXiv*: 1903.11593.
[61]. Chapaliuk B and Zaychenko Y 2018 Deep Learning Approach in Computer-Aided Detection System for Lung Cancer *IEEE First Int. Conf. on System Analysis & Intelligent Computing* (SAIC) pp. 1-4.
[62]. Sammouda R. Segmentation and analysis of CT chest images for early lung cancer detection. *Global Summit on Computer & Information Technology* (GSCIT) pp. 120-126.
[63]. Choi H and Na KJ 2018 A risk stratification model for lung cancer based on gene Coexpression network and deep learning *BioMed research Internationa*l 2018.
[64]. Lynch Chip M, Behnaz A, Fuqua Joshua D, de Carlo Alexandra R, Bartholomai James A, Balgemann Rayeanne N, van Berkel Victor H and Frieboes Hermann B 2017 Prediction of lung cancer patient survival via supervised machine learning classification techniques. *International journal of medical informatics* 108 1-8.
[65]. Moitra D and Mandal RK 2020 Classification of Non-Small Cell Lung Cancer using One-Dimensional Convolutional Neural Network 2020 *Expert Systems with Applications* 113564.
[66]. Song Q, Zhao L, Luo X and Dou X 2017 Using deep learning for classification of lung nodules on computed tomography images *Journal of healthcare engineering*.
[67]. Singh GA and Gupta PK 2019 Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans *Neural Computing and Applications* 31(10) 6863-77.
[68]. Skourt BA, El Hassani A and Majda A 2018 Lung CT image segmentation using deep neural networks *Procedia Computer Science* 109-13.
[69]. Lu M, Fan Z, Xu B, Chen L, Zheng X, Li J, Znati T, Mi Q and Jiang J 2020 Using Machine Learning to Predict Ovarian Cancer *International Journal of Medical Informatics* 104195.