Volume: 5, Issue: 3
Page: 159-173
2021

International Journal of
Science and Business
Journal homepage: ijsab.com/ijsb

IJSAB International

A Review of most Recent Lung Cancer Detection Techniques using Machine Learning

Dakhaz Mustafa Abdullah & Nawzat Sadiq Ahmed

Abstract:
Lung cancer is a sort of dangerous cancer and difficult to detect. It usually causes death for both gender men & women therefore, so it is more necessary for care to immediately & correctly examine nodules. Accordingly, several techniques have been implemented to detect lung cancer in the early stages. In this paper a comparative analysis of different techniques based on machine learning for detection lung cancer have been presented. There have been too many methods developed in recent years to diagnose lung cancer, most of them utilizing CT scan images and some of them using x-ray images. In addition, multiple classifier methods are paired with numerous segmentation algorithms to use image recognition to identify lung cancer nodules. From this study it has been found that CT scan images are more suitable to have the accurate results. Therefore, mostly CT scan images are used for detection of cancer. Also, marker-controlled watershed segmentation provides more accurate results than other segmentation techniques. In Addition, the results that obtained from the methods based deep learning techniques achieved higher accuracy than the methods that have been implemented using classical machine learning techniques.

IJSB
Literature Review
Accepted 6 February 2021
Published 12 February 2021
DOI: 10.5281/zenodo.4536818

Keywords: *Lung Cancer Detection, Machine Learning, Deep Learning, SCLC, and NSCLC.*

About Author (s)

Dakhaz Mustafa Abdullah (corresponding author), Information Technology, Technical College of Informatics, Akre Information Technology Management, Duhok Polytechnic University, Iraq. Email: dakhaz.abdullah@dpu.edu.kud
Nawzat Sadiq Ahmed, Information Technology Management**,** Technical College of Administration, DPU, Iraq. Email: nawzat.ahmed@dpu.edu.krd

1. Introduction:

From the windpipe or the main airway, or the lungs Lung cancer can start (Yu et al., 2020). It is caused by of the spread of certain cells in the lungs and uncontrolled growth cells. Peo`ple who suffer from chest diseases or emphysema are more likely to with lung cancer (Radhika et al., 2019). Small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma are the two main kinds of lung cancer. SCLC is nearly linked to smoking and is growing faster. Lung cancer of non-small cells is more common and spreads more slowly. It is called mixed small cell/large cell cancer (Hussain et al., 2019). If cancer has characteristics of both types. According to recent statistics in 2018, approximately 234,030 new lung cancer cases were expected to be diagnosed with approximately 85 percent of non-small cell lung cancer (NSCLC) (Siegel et al.,2018). The most significant factor in making this disease so deadly is the progression of lung cancer without symptoms. Nearly a quarter of people had no signs of cancer. Many people know that lung cancer has another condition that causes lung X-rays. Early diagnosis is extremely critical, as lung cancer can also spread quickly. Today, lung cancer can be diagnosed in the early stages by imaging technology developments such as low-dose computed tomography (Günaydin et al., 2019).

A tumor called nodule from the cells in the airways of the respiratory system causes lung cancer. These cells in chest X-rays are always in direct contrast and take the form of a spherical object. If lung nodules can be reliably detected at an early stage, the patient survival rate can be dramatically improved. Since lung nodules using raw chest X-ray imaging cannot be detected quickly, however, interpreting these diagnostic photographs has become repetitive and a very complicated task (Jothilakshmi and SV, 2020). The presence of a small nodule from a large 3D lung CT scan would have to be detected by the computer-aided diagnosis (CAD) system (Zeebaree et al., 2019; Zebari et al., 2020). Fig.1. gives an example of an early-stage lung cancer nodule seen in a 2D slice of a CT scan. Noise from the tissues around it, air, and bone fills the CT scan, so this noise would first have

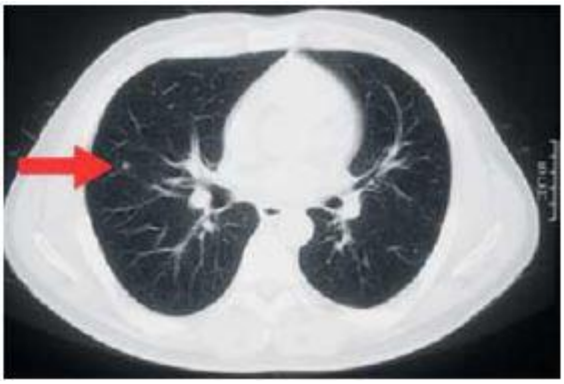

Figure 1: 2D CT scan slice containing a small (5mm) early-stage Lung Cancer nodule (Alakwaa et al., 2017).

to be pre-processed for the CAD systems to search effectively. Image pre-processing, malignancy classification detection, and nodule candidate are our classification pipeline (Alakwaa et al., 2017). By analyzing CT images built by artificial intelligence approaches, machine learning techniques help in the early diagnosis and evaluation of lung nodules (Zeebaree et al., 2021). These systems are referred to as decision support systems that examine the images through the process of preprocessing, segmentation, feature extraction, and classification as shown in Fig. 2 (Saba, 2020).

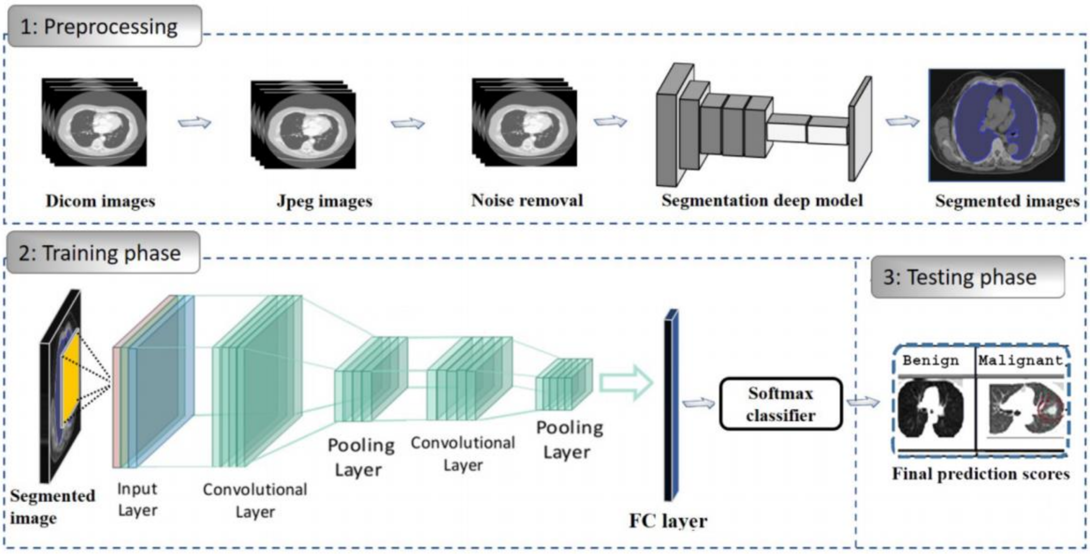

Fig. 2. Machine supported framework for lung nodules prediction (Saba, 2020)

Many detection systems for lung cancer have been developed. However, certain systems lack adequate detection accuracy, and some systems must also be developed in order to reach the highest accuracy of 100%. pulmonary cancer identification and classification were based on machine learning techniques and image processing techniques (Zebari et al., 2020). We studied recent cancer detection systems developed to select the latest best systems, and analysis was carried out on them.

2. Machine Learning

Machine learning is the skill of learning computers, where a machine is created with such algorithms from which it can make its own choices and give the user the result (Khalaf et al., 2019). It is essentially known to be the Artificial Intelligence subfield (Ramos-Lima et al., 2020; Elassad et al., 2020). For complicated data classification and decision making, Machine Learning is used today. In general terms, the development of algorithms helps the machine to learn and make the required decisions It has close connections with mathematical optimization, which provides the field with tools, theory and implementation domain, and is used in a number of computational activities where explicit algorithms cannot be planned and programmed (Somvanshi et al., 2016; Maione et al., 2019; Zeebaree et al., 2019). The techniques and tasks of Machine Learning are broadly classified into three categories:

Supervised learning: this type solves regression problems, such as forecasting weather, population growth prediction and, life experience predicting using algorithms of Linear Regression or Random Forest (Moujahid et al., 2018; Abdulqader et al., 2020). In addition, supervised learning solves classification problems such as voice recognition, digit recognition, diagnostics, and identity fraud detection by using algorithms in many fields, such as Support Vector Machines, Random Forest, Nearest Neighbor, and others that are utilized in many areas (Ahmed and Sadiq 2018; Zeebaree et al., 2018). In supervised learning, there are two levels. The training phase and the testing phase. There must be known labels in the data sets used for the training process. The algorithms study the relationship between the input values and the labels and attempt to predict the testing data values (Kubat, 2017; Zantalis et al., 2019).

Unsupervised learning: this type deals with topics relating to the reduction in dimensionality used for visualization of large data, feature elaboration, or discovering secret structure. It is also used for clustering concerns such as recommendation frameworks, customer segmentation, and targeted marketing (Sulaiman et al.,2019). In comparison to supervised learning, no labels are available in this method. In this type, algorithms aim to recognize patterns on testing data and predict future values or cluster the data (Kubat, 2017; Zantalis et al., 2019).

Reinforcement Learning: in this type, based on a collection of tuning parameters the algorithms attempt to predict the output for a problem. Then the output becomes an input parameter, and then a new output is found once the optimum output is found. The Dep Learning and Artificial Neural Networks (ANN) used this style (Al-jaboriy et al., 2019). the applications which primarily used Reinforcement learning such as robot navigation, skill acquisition, real-time decisions and AI gaming (Kubat, 2017; Zantalis et al., 2019).

3. Related Works

As in other fields, researchers have effectively implemented statistical and machine learning techniques in the background of such diseases such as lung cancer, build prediction models. The detection of lung cancer has previously been considered using techniques of image processing as the work implemented by (Abdillah et al., 2017) in initiation with deep learning and neural networks techniques, domain of medical image has been used recently. Several researchers (Chauhan and Jaiswal, 2016; Nasser and Abu-Naser, 2019; Sharif et al., 2020) have attempted to classify and detect lung cancer by using techniques of classical neural networks and machine learning. Recently, some researchers have tried to use deep learning techniques for lung cancer detection (Alakwaa et al., 2017; Shakeel et al., 2019; Bhatia et al., 2019; Shen et al., 2017; Gao et al., 2018; van and de Bruijne, 2016; Bhandary et al., 2020; Gang et al., 2018; Li et al., 2020; Ausawalaithong et al., 2018).

Using machine learning technique and chemical sensor array, (Huang, et al., 2018) developed a breath test to detect lung cancer. A prospective research to record cases of lung cancer and non-tumor controls between 2016 and 2018 is performed, and 1 alveolar air samples are analyzed using arrays of carbon nanotube sensors. For the model derivation and inner validation, subjects enrolled in 2016 and 2017 were used. In subjects recruited in 2018, the model was validated externally. Using the pathological records as the reference standard, the diagnostic accuracy was evaluated. Using the pathological records as the reference standard, the model was validated externally. then the diagnostic accuracy was evaluated. in the external validation were the areas under the receiver operating characteristic curve (AUCs) 0.91 (95 percent CI = 0.79-1.00) by linear discriminant analysis, and by the technique of the supporting vector machine 0.90 (95 percent CI = 0.80-0.99). The authors concluded that lung cancer could be identified with high precision by incorporating the sensor array technique and machine learning. Chauhan and Jaiswal (2016) suggested the automatic classification of diseases with machine learning based on a practical approach to detecting lung cancer principle. In addition, some benchmark sets showed that the proposed work model was compared to other conventional approaches. The authors have shown that the algorithm proposed is successful over other methods such as SURF and ICA. Nasser and Abu-Naser (2019) have improved an Artificial Neural Network (ANN) model to confirm detection lung cancer, Symptoms such as anxiety, yellow fingers, respiratory illness, exhaustion, allergies, wheezing, coughing, shortness of breath, chest pain and trouble swallowing have been used lung cancer diagnose. The proposed ANN model is developed, educated, and validated using a data set called "survey lung cancer" as input variables for the proposed ANN model and other information about the individual as input variables for the proposed ANN model. The model evaluation revealed that, with 96.67 percent accuracy, the proposed ANN model was able to detect the absence or presence of lung cancer.

The prediction of lung cancer-prone patients will aid doctors in their treatment decision-making. In this respect, in the report, (Faisal et al., 2018) attempted to measure the discriminative capacity of multiple predictors to improve the efficacy of the diagnosis of lung cancer by their symptoms. A variety of classifiers are tested on a benchmark dataset retrieved from the UCI repository, including (SVM), Multi-Layer Perceptron, Naïve Bayes, C4.5 Decision tree, Neural Network, as seen in Figure 3. Familiar ensembles such as Majority Voting and Random Forest are also compared with the output. It is noted that Gradient-boosted Tree exceed other individuals as well as ensemble classifiers on the basis of performance assessments and achieved 90 percent precision.

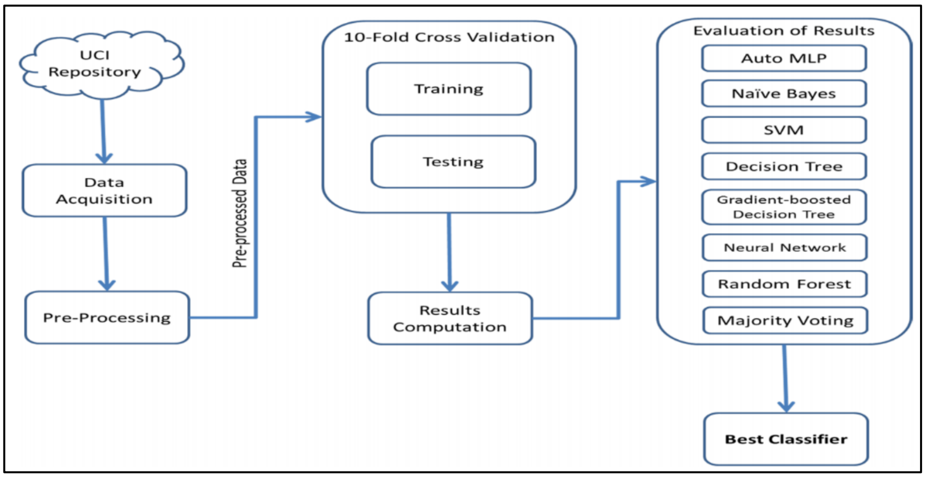

Figure 3: Overview of the Proposed Approach (Faisal et al., 2018).

Wu and Zhao (2017) suggested a new, Entropic Degradation Method (EDM) algorithm for the detection of Small Cell Lung Cancer (SCLC) for Computed Tomography (CT) images. The early diagnosis of lung cancer can be encouraged by this study. The training and test results are lung CT scans given by the National Cancer Institute in high resolution. The authors picked 12 lung CT scans from the library, six for safe lung and 6 for SCLC patients. Then, 5 random scans were taken from each party to practice their model and the remaining two scans were used for research. The suggested algorithm hit 77.8 percent accuracy. Reddy et al. (2019) suggested a structure consisting of various means such as pre handling, image securing, thresholding, binarization, extraction of attributes, division, and recognition of the neural system. The suggested strategy pursues approaches in which parallel thresholding is the initial stage, and then feature extraction, and then these highlights are used to train the fuzzy neural system for machine learning approaches and validate the neural system. From CT scan images, the proposed system accurately defines the lung condition. The proposed framework tested 150 types of pulmonary CT images and achieved the result where 96.67 percent of the framework's overall achievement rate met the framework desire. The significance of radiomics in the prediction of pathological stages of non-small cell lung cancer (NSCLC) has been studied for the first time by (Yu et al., 2019). Nine optimal image characteristics were identified as predictive and diagnostic biomarkers for NSCLC pathological phases. The prediction model has been verified using various machine learning algorithms to efficiently predict the tumor stages of NSCLC, particularly for lung adenocarcinoma (LUAD). The results not only extend the use of machine learning algorithms in the prediction of CT image features for pathological staging, but also recognizing possibility imaging biomarkers that can be used in NSCLC for diagnosis and prediction of the pathological stage. Singh and Gupta, (2019) have demonstrated an effective approach to the classification and detection of lung cancer-linked CT scan images into malignant and benign categories. First, these images are processed using techniques of image processing in the proposed approach, and then for their classification are used supervised learning algorithms. Here, along with statistical features, texture characteristics are extracted and different extracted characteristics are supplied to classifiers. In addition, are used for seven different classifiers. Multinomial naive Bayes classifier, support vector machine classifier, k-nearest neighbors' classifier, decision tree, stochastic gradient descent classifier, multi-layer perceptron (MLP) classifier and random forest classifier. In addition, 15750 dataset clinical images are used to train and test these classifiers, consisting of both 8840 malignant and 6910 benign images related lung cancer. In the results obtained, the accuracy of the MLP classifier is higher compared to the other classifiers, with a value of 88.55 percent.

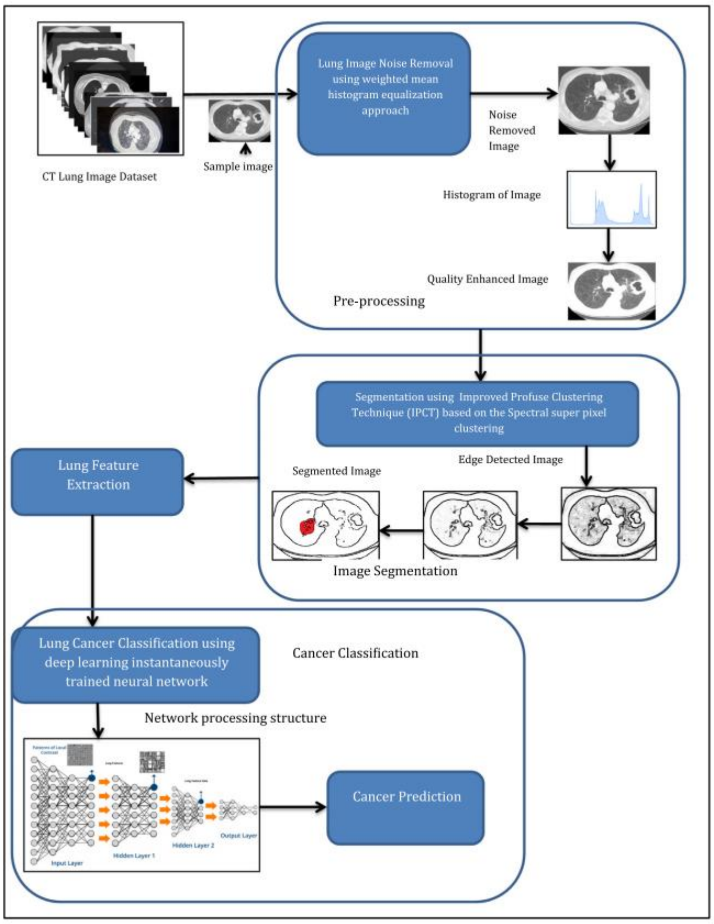

Figure 4: Lung cancer classification structure (Shakeel et al., 2019).

Alam et al., (2018) are proposing an efficient lung cancer detection and prediction algorithm using the SVM multi classifier. For cancer detection, multi-stage classification was used. The probability of lung cancer can also be predicted by this system. Improvement of the image and segmentation have been done individually in each classification stage. Image scaling, transformation of color space and improved contrast were used to improve image. For segmentation threshold and marker-controlled watershed-based segmentation was used. SVM binary classifier was used for classification purposes. The proposed algorithm provides a precise 97% for the identification of cancer and 87% for the prediction of cancer. Makaju et al., (2018) suggested a system that uses watershed segmentation for detection to detect the cancerous nodule from the lung image of CT scan and SVM for classification nodule as benign or malignant. the model accuracy 92% for detects cancer. Where compared model accuracy was 86.6% Shakeel et al. Shakeel et al., (2019) performed an assessment to prediction of lung cancer by CT images using Deep Learning and improved profuse clustering technique (IPCT) with Instantaneously Trained Neural Networks (DITNN) approach. Initially, from the Cancer Imaging Archive (CIA) dataset consisting of 5043 images DICOM format divided into 2043 testing images and 3000 training images, lung CT images were collected. The image quality was then improved by computing the weighted mean function that replaced the pixel was computed using the distribution process of probability and cumulative, as shown in Figure 4. The portion affected was segmented by the pixel similarity value measurement after enhancing the representation of the image. For the extraction of spectral related characteristics Clusters were formed on the basis of the similarity measure. The characteristics were trained and classified by classifier techniques that effectively predict up to 98.42 percent accuracy of cancer with a minimum classification error of 0.038. Alakwaa et al. (2017) developed an architecture of deep convolutional neural network to detect nodules in lung cancer patients and predict interest points using the U-Net architecture. This move is a 3D CNN preprocessing step. In the initial scans of Kaggle CT, an updated U-Net trained in LUNA16 (CT scans with labelled nodules) was used to detect nodule candidates. U-Net nodule detection produced several false positive results, which means that CT-regions with segmented lungs, where the most likely candidate nodules were located in 3D convolutional Neural Networks (CNNs), ultimately classed for lung cancer as positive or negative as shown in figure 5. The U-Net nodule detection produced many false positives. The 3D CNNs produced a test set of 86.6 percent accuracy. The efficiency of the proposed CAD method outdoes the existing literature CAD systems that have a multitude of preparation and test phases, each involving a lot of etiquette details, whereas the proposed CAD system only has three main phases (segmentation, applicant identification and malignancy classification nodule), allowing improved training and detection.

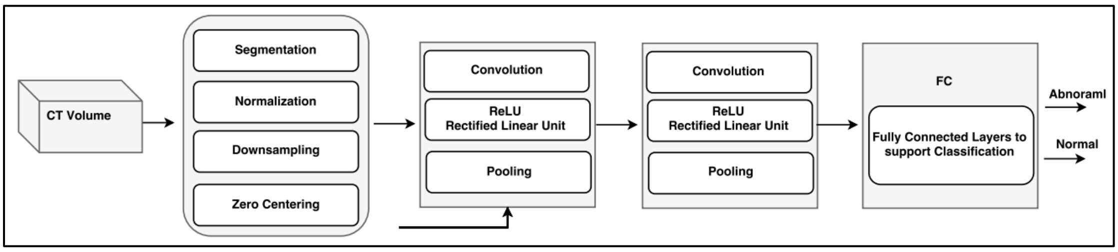

Figure 5: 3D convolutional neural networks architecture (Alakwaa et al. 2017).

Bhatia et al., (2019) used deep residual learning and CT scans to detect lung cancer. The authors described a pipeline of pre-processing techniques for highlighting cancer sensitive lung regions and extracted features using UNet and ResNet models. The feature set is fed to several categories, i.e., XGBoost and Random Forest are combined to predict how likely a CT scan is to be cancerous. The accuracy achieved with this work was 84 percent above previous attempts by LIDC-IRDI. A successful fuzzy auto-seed cluster means that (Manikandan & Bharathi, 2016) have developed a morphological algorithm to Divided lung nodules to detect lung cancer from successive slices of CT images. Averaging the minimum and maximum pixel values in each row of an image automatically selected the original cluster values. In order to identify the actual malignant nodules, the derived centroid change and texture features were used as the inputs to the Support Vector Machine (SVM) kernel classifier. This work was conducted on 50 normal cases and 56 malignant (cancer) (pulmonary infections) with a total of 745 benign nodules and 56 malignant nodules, respectively. In these, 60 percent of subjects were used for training (30 non-cancerous & 34 cancerous). For research, the remaining 40% (20 non-cancerous & 22 cancerous) subjects were included. This work produced 100 percent, 93 percent and 94 percent respectively of good sensitivity, specificity and precision. The per patient False Positive (FP) was estimated as 0.38.

Different methods of diagnosis of lung cancer used image segmentation were proposed and analyzed by (Abdillah et al. 2017). As explained in Figure 6, three segmentation image approaches for evaluating lung cancer are used, such as Area Growing, Marker Controlled Watershed, and Marker Controlled Watershed with Masking. Where certain techniques are used to segment the image of the CT scan. Image enhancement using the image segmentation, Gabor filter, and extraction of features is accompanied by detection phases. The investigators proved the feasibility of their method from the experimental findings. Furthermore, the findings revealed that the best way for the identification of main features was watershed with a highly accurate and robust masking process.

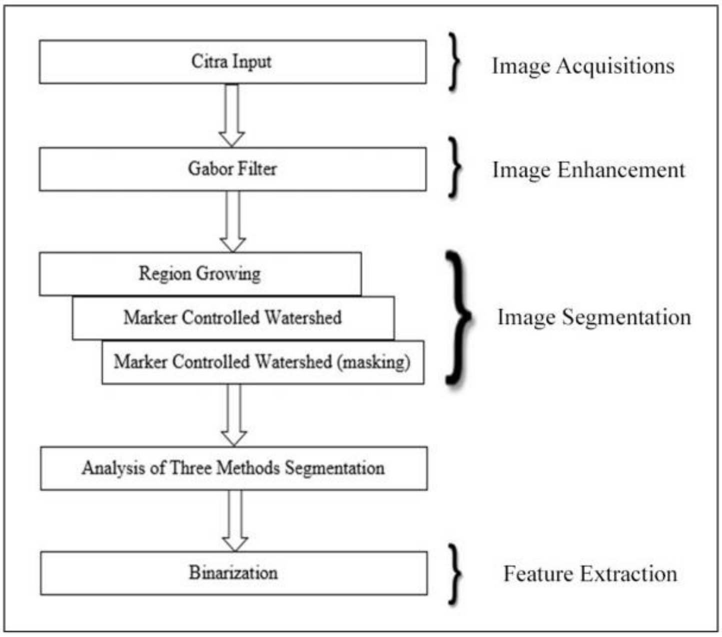

Figure 6. Image Processing for Lung Cancer Detection Stages (Wu and Zhao 2017).

For the detection of lung cancer, (Tripathi et al., 2019) present a comparative analysis of different image segmentation techniques. Thresholding techniques, Marker Controlled Watershed Segmentation, Edge detection and PDE dependent segmentation techniques are used in these techniques. Combined with multiple segmentation algorithms, there are different classifier strategies for the identification of lung cancer nodules using image recognition. The SVM classifier for Partial Differential Equation (PDE) dependent segmentation is defined in this report. It was discovered from this research that the marker-controlled segmentation of the watershed produces more detailed outcomes than other techniques of segmentation. Shakeel et al., (2020) have introduced a new, streamlined machine learning and image processing technique to predict Non-small cell lung cancer images CT scan dataset are obtained for lung cancer identification. By applying the multilevel brightness-preserving approach that efficiently analyses each pixel, removes the noise and also improves the quality of the lung image, the collected images are examined. The affected region is segmented from the noise-removed lung CT image by using an enhanced deep neural network that segments the region in terms of the use of network layers and different features. Then the effective characteristics are chosen using the intelligent-generalized rough set approach to hybrid spiral optimization, and those characteristics are categorized using the ensemble classifier. The method discussed improves the prediction rate of lung cancer, which is analyzed using MATLAB-based findings such as recall logarithmic loss, accuracy, F-score, and mean absolute error. To evaluate lung abnormalities in considered pictures, a changed AlexNet (MAN) is proposed by (Bhandary et al., 2020). Two modal images are taken into consideration: chest X-Rays and lung CTs. On these two image datasets, the proposed MAN is checked separately. In the initial exam process, the X-Ray thoracic is graded as normal and the pneumonia class as explained in Figure 7. In comparison with the other DL techniques considered in this review, the proposed DL method provides an accuracy of 96 percent.

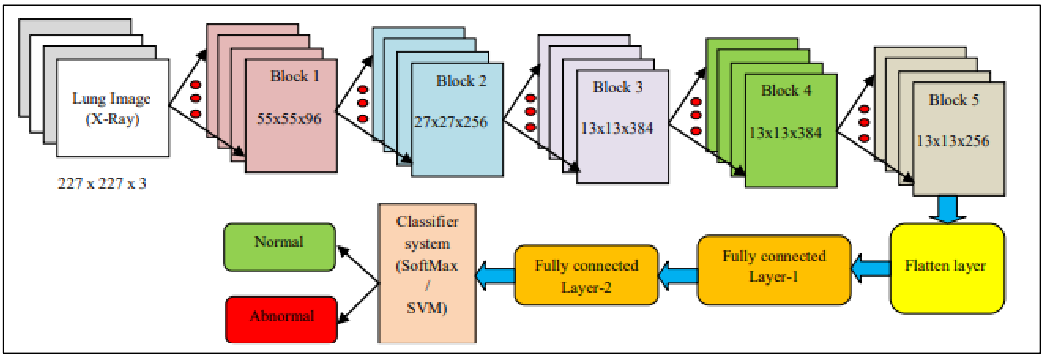

Figure 7: Modified AlexNet architecture proposed to classify the chest X-Ray radiographs (Makaju et al., 2018).

In addition, as shown in Figure 8, the MAN architecture is used with and without Ensemble-Feature-Technique (EFT) to classify lung CT images into benign and malignant. The proposed SVM with MAN classifier fulfill an accuracy of 86.47 percent and has an accuracy of 97.27 percent along with EFT, a similar DL system. This study confirms that the MAN system proposed works well on the image datasets considered. In addition, a comparative study of current state-of-the-art DL techniques indicates that the proposed DL method provides greater precision compared to existing systems.

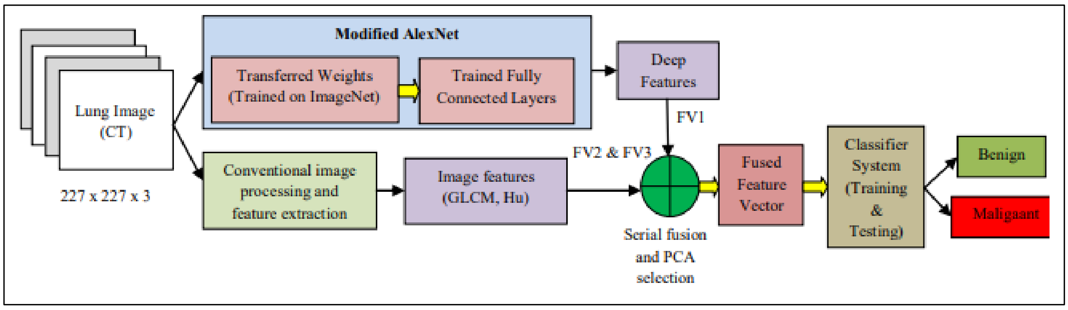

Figure 8. Ensemble-Feature-Technique implemented to examine the LIDC-IDRI database (Bhandary et al., 2020).

Due to the small number of radiologists available A computer-aided detection scheme to assist radiologists in decision-making should be established, because of the small number of radiologists and the vast number of chest x-ray radiographs (CXR) available for observation. In several computer vision applications deep learning showed state-of-the-art results. A deep-seated neural network (CNN) was then performed on CXR images by (Gang et al., 2018; Li et al., 2020; Ausawalaithong et al., 2018). Specifically, Li et al., (2020) has preprocessed the CXR images with lung segmentation and rib elimination. Patches were then extracted for each pixel in the lung field and three CNNs were trained at various image resolutions as shown in Figure 9. Finally, the fusion function approach was used to integrate all information gathered at various resolutions. The complete fusion method showed the best performance in four fusion methods tested. 99 percent lung nodules can be identified in the JSRT database using the proposed process. The approach proposed was precise, durable and can be used in actual clinical practice.

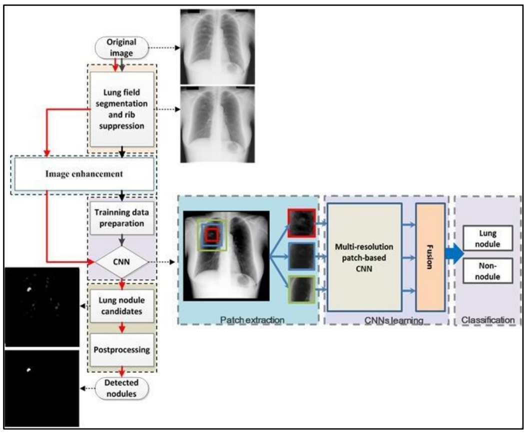

Figure 9. The block diagram of the proposed method (Black and red arrows show the training and testing processes, respectively) (Li et al., 2020).

It is obvious from the state-of-the-art literature review that the science community has paid a lot of attention to lung cancer. Classical machine learning and neural network methods have been used to most approaches. Others used deep learning methods and applied them to images of both kinds (CT and X-Ray). This study therefore focuses on investigating these methods in order to the best technique used for detecting lung cancer.

Table 1: Comparison among state-of-the-art Lung Cancer detection Methods

Ref.	Year	Preprocessing	Methods	Datasets	Results
(Li et al., 2020)	2020	lung field segmentation and rib suppression	multi-resolution patch-based CNNs were trained for lung nodule detection	Japanese Society of Radiological Technology (JSRT) database	The method can detect 99% lung nodules on JSRT database
(Bhandary et al., 2020)	2020	Morphological segmentation and watershed segmentation are used for automated nodule segmentation	MAN is used to classify chest X-Rays images and EFT is used to classify the lung CT images.	Dataset of Chest X-Ray and Lung cancer (LIDC-IDRI)	DL accuracy is 96% for X-Ray images while the accuracy is 97.27% for CT images
(Shakeel et al., 2020)	2020	multilevel brightness-preserving approach	improved deep neural network and ensemble classifier.	Database of cancer imaging archive (CIA) dataset	The proposed system recognized the cancer with maximum accuracy.
Shakeel et al., 2019	2019	The noise is removed using weighted mean histogram equalization approach. In addition, improved profuse clustering technique (IPCT) is applied for segmenting the affected region.	Deep learning instantaneously trained neural network (DITNN) is used.	Image was collected from Cancer imaging Archive (CIA) dataset	accuracy 98.42% & minimum classification error of 0.038.
(Reddy et al. 2019)	2019	Picture securing, pre-handling, binarization, thresholding, division, feature extraction are applied.	The fuzzy neural system is used to test the neural system with machine learning approaches.	Dataset obtained from UCI repository	Accuracy 96.67 %.

(Bhatia et al., 2019)	2019	This step consists of segmentation is followed by normalization and zero centering.	A number of classifiers like XGBoost and Random Forest are used.	Dataset of Lung Image Database Consortium image collection (LIDC-IDRI)	accuracy 84%
(Makaju et al., 2018)	2018	median filter and Gaussian filter are applied on the CT images.	Watershed segmentation for detection and SVM for classification of nodule as Malignant or benign.	Database of Lung Image Database Consortium (LIDC)	accuracy 92%.
(Faisal et al., 2018)	2018	pre-processing for data cleaning is applied	A number of classifiers including: MLP, Neural Network, Decision Tree, Naïve Bayes, Gradient Boosted Tree, and SVM are assessed	dataset obtained from UCI repository	accuracy 90%.
(Singh and Gupta, 2019)	2018	Converting to grayscale image, applying denoising methods such as median blur, Gaussian blur, and bilateral blur, then applying thresholding methods for converting the grayscale image into a binary image.	k-nearest neighbors classifier, support vector machine classifier, decision tree classifier, multinomial naïve Bayes classifier, stochastic gradient descent classifier, random forest classifier, and multi-layer perceptron (MLP) classifier are applied	database of Lung Image Database Consortium (LIDC)	accuracy 88.55%.
(Alam et al., 2018)	2018	image enhancement and segmentation has been done. Image scaling, color space transformation and contrast enhancement	multi-class SVM classifier	The lung cancer dataset used for training is taken from UCI machine learning database	precision of 97% for cancer identification and 87% for cancer prediction
(Alakwaa et al. 2017)	2017	segmentation, normalization, down-sampling, and zero-centering are applied on the 3D Scan image.	3D Convolutional Neural Networks (CNNs).	CT scan dataset from Kaggles Data Science Bowl (DSB)	accuracy 86.6%.
(Abdillah et al. 2017)	2016	image enhancement using Gabor filter is applied	Region Growing, Marker Controlled Watershed, and Marker Controlled Watershed with Masking are applied.	Used Scan CT Image are from VIA and ELCAP database	watershed with masking method has highest accuracy and robustness.

4. Discussion and Analysis

This research work is focusing on lung cancer detection-based machine learning techniques. Where, most of the works that have been explored in the Literature were based on CT Scan images and some of them used X-ray images. And, in both cases, lung cancer detection procedure goes through the following phases: -

Pre-processing: - The pre-processing is the first phase in which CT scan image or the X-ray image is taken as input. Then, some techniques of image Processing will be applied such as de-noising, thresholding, binarization, normalization and zero centering. Followed by Segmentation that will segment the similar and dissimilar regions from the CT scan image. Many segmentation methods such as Region Growing, Marker Controlled Watershed, and Marker Controlled Watershed with Masking have used in the literature. From table 1. It has been demonstrated that watershed with masking method obtained higher results. Finally, the features can be extracted to be prepared for the next step which is represented by classification.

Classification: - In this phase, the extracted features are fed to specified classifier to classify them as normal and malignant accordingly. Many classifiers have been used by the researchers in the literature such as: multi-layer perceptron (MLP), SVM, Naïve Bayes, Neural Network, Gradient Boosted Tree, Decision Tree, k-nearest neighbors, multinomial random forest classifier naïve Bayes, stochastic gradient descent, and ensemble classifier. From table 1 it is clear that highest accuracy result was about 97% obtained by (Alam et al., 2018) using multi class SVM classifier as well as adopting marker-controlled watershed-based segmentation for image segmentation. On the other hand, all the works that have been implemented using Deep Learning methods obtained high accuracy results where the highest result was about 99% by (Li et al., 2020) using multi-resolution patch-based CNNs.

5. Conclusion

When lung cancer is diagnosed at an early stage, it would be beneficial because the medication will then be initiated to prevent disease from having a harmful result. Therefore, this paper summarizes a detailed survey on various machine learning approaches to classify lung malignancies using either CT scan images or X-ray images. Also, many classifiers have been used by the researchers in the literature such as: MLP, SVM, Naïve Bayes, Neural Network, Gradient Boosted Tree, Decision Tree, k-nearest neighbors, multinomial random forest classifier naïve Bayes, stochastic gradient descent, and ensemble classifier .Consequently, and based on the extensive survey that have been done in this work, it can be concluded that the methods which utilized deep learning techniques obtained higher results in terms of accuracy than other classical machine learning techniques. Where, the highest result was about 99% using multi-resolution patch-based CNNs.

References

Yu, K. H., Lee, T. L. M., Yen, M. H., Kou, S. C., Rosen, B., Chiang, J. H., & Kohane, I. S. (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.

Radhika, P. R., Nair, R. A., & Veena, G. (2019, February). A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms. In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)* (pp. 1-4). IEEE.

Hussain, L., Rathore, S., Abbasi, A. A., & Saeed, S. (2019, March). Automated lung cancer detection based on multimodal features extracting strategy using machine learning techniques. In Medical Imaging 2019: Physics of Medical Imaging (Vol. 10948, p. 109483Q). International Society for Optics and Photonics.

Siegel, R. L., Miller, K. D. and Jemal, A., “Cancer statistics, 2018,” CA. Cancer J. Clin. 68(1), 7–30 (2018).

Günaydin, Ö., Günay, M., & Şengel, Ö. (2019, April). Comparison of lung cancer detection algorithms. In 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT) (pp. 1-4). IEEE.

Jothilakshmi, R., & SV, R. G. (2020) Early Lung Cancer Detection Using Machine Learning And Image Processing.

Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zebari, D. A. (2019, April). Trainable Model Based on New Uniform LBP Feature to Identify the Risk of the Breast Cancer. In 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 106-111). IEEE.

Zebari, D. A., Zeebaree, D. Q., Abdulazeez, A. M., Haron, H., & Hamed, H. N. A. (2020). Improved Threshold Based and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images. IEEE Access, 8, 203097-203116.

Alakwaa, W., Nassef, M., & Badr, A. (2017). Lung cancer detection and classification with 3D convolutional neural network (3D-CNN). Lung Cancer, 8(8), 409.

Zeebaree, D. Q., Abdulazeez, A. M., Zebari, D. A., Haron, H., & Hamed, H. N. A. (2021) Multi-Level Fusion in Ultrasound for Cancer Detection Based on Uniform LBP Features.

Saba, T. (2020). Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges. Journal of Infection and Public Health, 13(9), 1274-1289.

Zebari, D. A., Zeebaree, D. Qhy7 ., Abdulazeez, A. M., Haron, H., & Hamed, H. N. A. (2020). Improved Threshold Based and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images. IEEE Access, 8, 203097-203116.

Khalaf, B. A., Mostafa, S. A., Mustapha, A., Mohammed, M. A., & Abduallah, W. M. (2019). Comprehensive review of artificial intelligence and statistical approaches in distributed denial of service attack and defense methods. IEEE Access, 7, 51691-51713.

Ramos-Lima, L. F., Waikamp, V., Antonelli-Salgado, T., Passos, I. C., & Freitas, L. H. M. (2020). The use of machine learning techniques in trauma-related disorders: A systematic review. Journal of psychiatric research, 121, 159-172.

Abou Elassad, Z. E., Mousannif, H., Al Moatassime, H., & Karkouch, A. (2020). The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review. Engineering Applications of Artificial Intelligence, 87, 103312.

Somvanshi, M., Chavan, P., Tambade, S., & Shinde, S. V. (2016, August). A review of machine learning techniques using decision tree and support vector machine. In 2016 International Conference on Computing Communication Control and automation (ICCUBEA) (pp. 1-7). IEEE.

Maione, C., Barbosa Jr, F., & Barbosa, R. M. (2019). Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: A review. Computers and Electronics in Agriculture, 157, 436-446.

Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zebari, D. A. (2019, April). Machine learning and Region Growing for Breast Cancer Segmentation. In 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 88-93). IEEE.

Moujahid, A., Tantaoui, M. E., Hina, M. D., Soukane, A., Ortalda, A., ElKhadimi, A., & Ramdane-Cherif, A. (2018, June). Machine learning techniques in ADAS: A review. In 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE) (pp. 235-242). IEEE.

Abdulqader, D. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2020). Machine Learning Supervised Algorithms of Gene Selection: A Review. Machine Learning, 62(03).

Ahmed, N. S., & Sadiq, M. H. (2018, October). Clarify of the random forest algorithm in an educational field. In 2018 International Conference on Advanced Science and Engineering (ICOASE) (pp. 179-184). IEEE.

Zeebaree, D. Q., Haron, H., & Abdulazeez, A. M. (2018, October). Gene selection and classification of microarray data using convolutional neural network. In 2018 International Conference on Advanced Science and Engineering (ICOASE) (pp. 145-150). IEEE.

Kubat, M. An Introduction to Machine Learning; Springer: Cham, Switzerland, 2017.

Zantalis, F., Koulouras, G., Karabetsos, S., & Kandris, D. (2019). A review of machine learning and IoT in smart transportation. Future Internet, 11(4), 94.

Sulaiman, D. M., Abdulazeez, A. M., Haron, H., & Sadiq, S. S. (2019, April). Unsupervised Learning Approach-Based New Optimization K-Means Clustering for Finger Vein Image Localization. In 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 82-87). IEEE.

Al-jaboriy, S. S., Sjarif, N. N. A., Chuprat, S., & Abduallah, W. M. (2019). Acute lymphoblastic leukemia segmentation using local pixel information. Pattern Recognition Letters, 125, 85-90.

Huang, C. H., Zeng, C., Wang, Y. C., Peng, H. Y., Lin, C. S., Chang, C. J., & Yang, H. Y. (2018). A study of diagnostic accuracy using a chemical sensor array and a machine learning technique to detect lung cancer. Sensors, 18(9), 2845.

Chauhan, D., & Jaiswal, V. (2016, October). An efficient data mining classification approach for detecting lung cancer disease. In 2016 International Conference on Communication and Electronics Systems (ICCES) (pp. 1-8). IEEE.

Nasser, I. M., & Abu-Naser, S. S. (2019). Lung Cancer Detection Using Artificial Neural Network. International Journal of Engineering and Information Systems (IJEAIS), 3(3), 17-23.

Sharif, M. I., Li, J. P., Naz, J., & Rashid, I. (2020). A comprehensive review on multi-organs tumor detection based on machine learning. Pattern Recognition Letters, 131, 30-37.

Faisal, M. I., Bashir, S., Khan, Z. S., & Khan, F. H. (2018, December). An evaluation of machine learning classifiers and ensembles for early stage prediction of lung cancer. In 2018 3rd International Conference on Emerging Trends in Engineering, Sciences and Technology (ICEEST) (pp. 1-4). IEEE.

Wu, Q., & Zhao, W. (2017, October). Small-cell lung cancer detection using a supervised machine learning algorithm. In 2017 International Symposium on Computer Science and Intelligent Controls (ISCSIC) (pp. 88-91). IEEE.

Reddy, U., Reddy, B., & Reddy, B. (2019). Recognition of Lung Cancer Using Machine Learning Mechanisms with Fuzzy Neural Networks. Traitement du Signal, 36(1), 87-91.

Yu, L., Tao, G., Zhu, L., Wang, G., Li, Z., Ye, J., & Chen, Q. (2019). Prediction of pathologic stage in non-small cell lung cancer using machine learning algorithm based on CT image feature analysis. BMC cancer, 19(1), 1-12.

Singh, G. A. P., & Gupta, P. K. (2019). Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. Neural Computing and Applications, 31(10), 6863-6877.

Alam, J., Alam, S., & Hossan, A. (2018, February). Multi-stage lung cancer detection and prediction using multi-class svm classifie. In 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2) (pp. 1-4). IEEE.

Makaju, S., Prasad, P. W. C., Alsadoon, A., Singh, A. K., & Elchouemi, A. (2018). Lung cancer detection using CT scan images. Procedia Computer Science, 125, 107-114.

Shakeel, P. M., Burhanuddin, M. A., & Desa, M. I. (2019). Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks. Measurement, 145, 702-712.

Bhatia, S., Sinha, Y., & Goel, L. (2019). Lung cancer detection: A deep learning approach. In Soft Computing for Problem Solving (pp. 699-705). Springer, Singapore.

Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. Annual review of biomedical engineering, 19, 221-248.

Gao, M., Bagci, U., Lu, L., Wu, A., Buty, M., Shin, H. C., & Mollura, D. J. (2018). Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 6(1), 1-6.

van Tulder, G., & de Bruijne, M. (2016). Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted Boltzmann machines. IEEE transactions on medical imaging, 35(5), 1262-1272.

Manikandan, T., & Bharathi, N. (2016). Lung cancer detection using fuzzy auto-seed cluster means morphological segmentation and SVM classifier. Journal of medical systems, 40(7), 181.

Abdillah, B., Bustamam, A., & Sarwinda, D. (2017, October). Image processing based detection of lung cancer on CT scan images. In Journal of Physics: Conference Series (Vol. 893, No. 1, p. 012063). IOP Publishing.

Tripathi, P., Tyagi, S., & Nath, M. (2019). A comparative analysis of segmentation techniques for lung cancer detection. Pattern Recognition and Image Analysis, 29(1), 167-173.

Shakeel, P. M., Burhanuddin, M. A., & Desa, M. I. (2020). Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier. Neural Computing and Applications, 1-14.

Bhandary, A., Prabhu, G. A., Rajinikanth, V., Thanaraj, K. P., Satapathy, S. C., Robbins, D. E., ... & Raja, N. S. M. (2020). Deep-learning framework to detect lung abnormality–A study with chest X-Ray and lung CT scan images. Pattern Recognition Letters, 129, 271-278.

Gang, P., Zhen, W., Zeng, W., Gordienko, Y., Kochura, Y., Alienin, O., ... & Stirenko, S. (2018, March). Dimensionality reduction in deep learning for chest X-ray analysis of lung cancer. In 2018 tenth international conference on advanced computational intelligence (ICACI) (pp. 878-883). IEEE.

Li, X., Shen, L., Xie, X., Huang, S., Xie, Z., Hong, X., & Yu, J. (2020). Multi-resolution convolutional networks for chest X-ray radiograph-based lung nodule detection. Artificial intelligence in medicine, 103, 101744.

Ausawalaithong, W., Thirach, A., Marukatat, S., & Wilaiprasitporn, T. (2018, November). Automatic lung cancer prediction from chest X-ray images using the deep learning approach. In 2018 11th Biomedical Engineering International Conference (BMEiCON) (pp. 1-5). IEEE.

Cite this article:

Abdullah, D. M. & Ahmed, N. S. (2021). A Review of most Recent Lung Cancer Detection Techniques using Machine Learning. *International Journal of Science and Business*, *5*(3), 159-173. doi: https://doi.org/10.5281/zenodo.4536818

Retrieved from http://ijsab.com/wp-content/uploads/695.pdf

Published by