DOI: https://doi.org/10.52756/ijerr.2023.v30.001 *Int. J. Exp. Res. Rev.*, Vol. 30: 1-14 (2023)

Review Article | Peer Reviewed | Open Access

International Journal of Experimental Research and Review (IJERR)

ISSN: 2455-4855 (Online)
www.iaph.in

Study on segmentation and prediction of lung cancer based on machine learning approaches

Check for updates

Anindita Saha[1] and Rakesh Kumar Yadav[2]*

[1]Department of Computer Science & Engineering, IFTM University, Moradabad, UP, India;
[2]Department of Computer Science & Engineering, Maharishi School of Engineering & Technology, Maharishi University of Information Technology, Lucknow, India

E-mail/Orcid Id:

AS, anindita.cse.kec@gmail.com, https://orcid.org/0000-0002-5780-9252; *RKY*, rkymuit@gmail.com, https://orcid.org/0000-0002-0151-4981

Article History:

Received: 06th Jan., 2023
Accepted: 20th Feb., 2023
Published: 30th Apr., 2023

Keywords:
Deep Learning (DL), Machine Learning (ML), clustering, fuzzy, segmentation.

Abstract: Lung cancer is a dangerous disease in human health. At the early stage, lung cancer detection provides a way to save human life. As a result, improvements in Deep Learning (DL), a technique, a branch of Machine Learning (ML), have helped to identify and classify lung cancer in clinical photographs. DL technology has also outperformed traditional methods in a variety of fields. Researchers are exploring various DL techniques for disease detection to improve the accuracy of the CAD systems in CT lung cancer detection. In this experiment, cutting-edge ML and DL methods for lung disease have been recommended as CAD systems after thoroughly analysing existing frameworks. It can be separated into FP reduction systems and system to detect nodule. The primary characteristics of various approaches are analyzed. The CT lung datasets existing for examination and evaluation with the various approaches are also presented and discussed.

Introduction

Lung cancer is considered a destructive disease around the globe as the early stage of disease detection may increase the survival rate in humans. There are two variants of cancer cells: Benign and Malignant. Where benign is observed as non-cancerous; on the other hand, malignant causes cancerous cells that grow in a small amount inside human lungs. The crucial part here is that we need to detect malicious cancer cells early for survival. However, the benign and malicious modules are more similar but, in some cases, differ from each other in terms of cancer cell location, the shape & structure of the cells (Nair et al., 2018). The challenging task here is to calculate the probability of the level of malignancy in the earlier stage (Silvestri et al., 2018). As a result, various lung diagnosis methods are established: CT (Computed Tomography), Isotope, X-ray and MRI (Magnetic Resonance Image). Apart from this, CT & X-ray chest radiography are considered to be the significant scanning modules for the early diagnosis of cancer cells. They utilized anatomic image modalities which detect the different cancer tissues. Among these two methods, CT is considered to investigate lung diseases suitably. However, medical practitioners utilised invasive approaches to differentiate the malicious and the benign cancer cells. Furthermore, lung diseases can be detected via scan examining, blood tests, chest X-ray, skin test, CT, and sputum sample tests (Setio et al., 2017). Identifying what is malicious cancer and what is benign cancer is still a challenging problem as they both have similarities. To resolve this issue machine learning comes into the picture as a part of the machine learning procedures SVM (Support Vector Machine) plays a significant role in identifying the benign and malicious cancer modules (Choi and Choi, 2014; Peña et al., 2016; Camarlinghi et al., 2012; Teramoto and Fujita, 2013; Santos et al., 2014). The drawback faced by this system is its need to be handcrafted; as a result, it seems hard to reproduce the advanced results. Moreover, the subset of machine learning is deep learning which works similarly to the neurons in the human brain. Deep learning evolving fast in recent times which greatly improves the numerous performances in medical applications (Wu et al., 2018). The advantage of the CAD systems deep

***Corresponding Author:** rkymuit@gmail.com

learning approach is that it helps detect the salient end-to-end features during training time. This capacity from the different computed tomography scans allows deep learning to identify cancer cells. By utilizing the training set which is considered to be rich in variations, the system could find the features of benign and malicious modules, yielding better execution of the result. After the training is accomplished, the generalization of learning and detection of cancer modules could have happened.

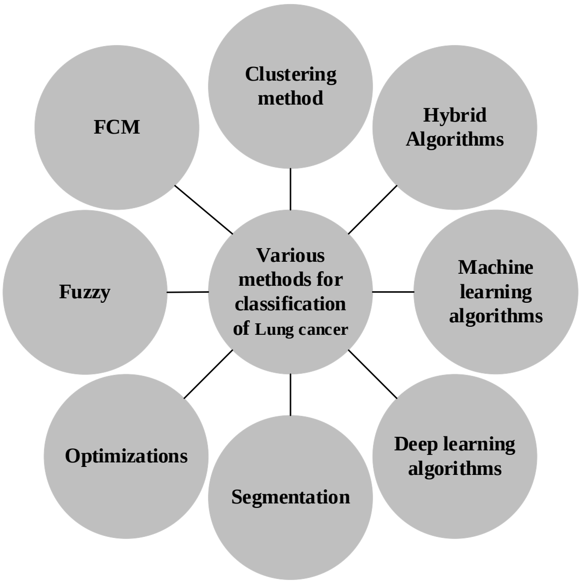

Figure 1. Various methods for lung cancer classification

Deep learning applications in lung cancer include pneumonia, pulmonary nodule diseases, interstitial lung disease, and pulmonary embolism. In addition to that, some common deep-learning neural network analyses utilized in the medical image industry have also been presented in this paper. Also, this paper discusses the trends analysis of the previous research happened. This research article discussed lung cancer classification and detection utilizing DL techniques and trend analysis.

This paper is categorized as follows, section 3 tells about the deep learning processes while detecting lung cancer in the medical image process, and Section 4 describes the taxonomy with a clear explanation that explains every subtopic in the taxonomy. Section 5 discusses the research gap, analytical trends, and future scope of lung cancer. In section 6, the survey's limitations are narrated, and then the conclusion is summarized in section 7.

Various Methods For Classification of Lung Cancer

Researchers and scientists have experimented with and devised various techniques and models to assess and classify lung cancer over time. Researchers and experts have predicted a variety of techniques for identifying lung cancer. The researchers are experimenting with several deep-learning approaches to enhance the efficiency in systematic CAD in lung cancer identification by utilizing the CT. Deep learning is a strategy for reducing computational complexity while also increasing efficiency.

Some models are built using a mix of the methods mentioned above and their correlation. On several occasions, hybrid and combination methods outperformed standalone ones. We have discussed deep learning and the use of recent research in lung cancer diagnosis utilizing clustering, fuzzy systems, machine learning, and segmentation extensively in this study.

A study of lung disease as cancer and classification using fuzzy techniques

Manikandan and Bharathi (2011) have presented the fuzzy system (FIS) to diagnose lung cancer by conducting the region-based segmentation process. Here, the IF-THEN condition was used to analyze the effect of predicted lung nodules. Likewise, Manikandan and Bharathi (2016) have introduced the fuzzy auto-seed cluster means morphological algorithm to classify lung cancer by segmenting the lung nodule from CT images.

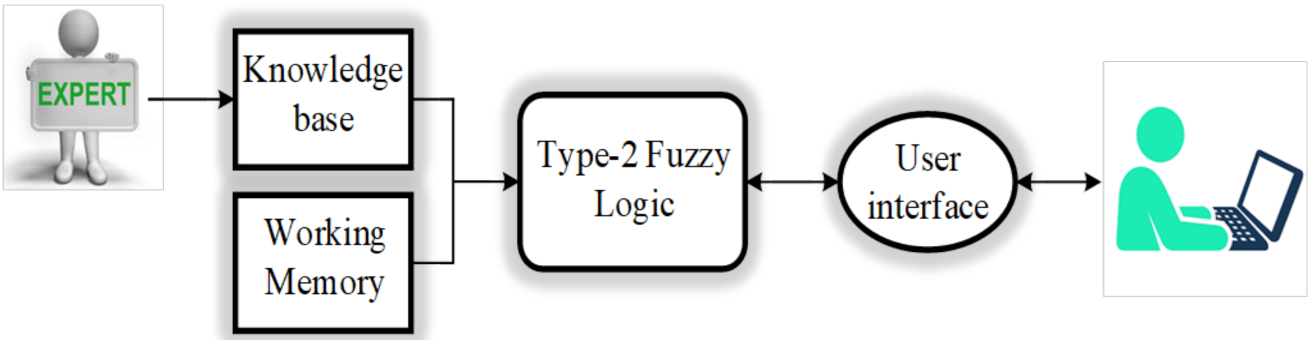

Figure 2. Lung cancer detection using Fuzzy System (Adapted from Manikandan and Bharathi, 2016)

Manickavasagam and Selvan (2019) normal and abnormal images. Where the cuckoo search algorithm-

DOI: https://doi.org/10.52756/ijerr.2023.v30.001

based neuro-fuzzy approach was utilized to predict the cancer portion.

Capizzi et al. (2019) have oeganized a type 1 fuzzy membership function and neural network to examine the input sample for predicting the lung nodules from the X-ray images. Sahu et al. (2019) executed fuzzy c-means methodology. Along with the breathing clustering approach, IIFDL (inter-fraction fuzzy deep learning) spot out lung cancer.

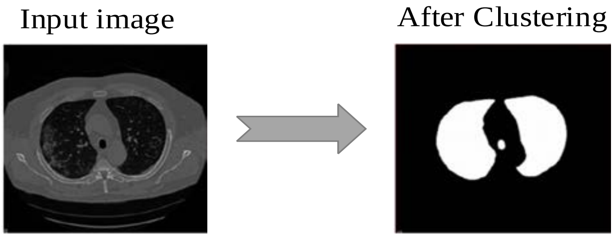

Figure 3. Lung cancer detection using clustering (Sarker et al., 2017)

Table 1. Recent research of lung cancer detection and classification using fuzzy methods.

Sl. No	Author	Finding	Outcome	Dataset	Limitation
1.	Manikandan and Bharathi (2011)	FIS	Accuracy-92%	Own collected Real-time dataset	Not only use hybrid systems.
2.	Manikandan and Bharathi (2016)	Fuzzy auto-seed cluster means morphological algorithm	Accuracy-94%	Own collected Real-time dataset	Computational time is high
3.	Manickavasagam and Selvan (2019)	Neuro-fuzzy with a cuckoo search algorithm	Sensitivity-97%	Own collected Real-time dataset	Could not estimate the location, size and growth rate of the module.
4.	Capizzi et al. (2019)	The type-1 fuzzy system combined with a neural network	Accuracy-92%	Real-time dataset (Radiology Key)	This method does not apply to other pulmonary diseases.
5.	Sahu et al. (2019)	Fuzzy c-means	Accuracy-99%	LIDC-IDRI	Not easy to integration
		fuzzy deep learning	RMSE – 59%	Own collected Real-time dataset	Only find out the lung cancer

Review of the disease of the lung as cancer and its classification using clustering

The chemical-protein, as well as chemical-chemical interaction, were employed by Lu et al. (2016) to detect the drug compound in a particular candidate to predict the lung nodule. Likewise, Using k-means clustering and permutation test algorithms prohibited the candidate's probability rate of curing the cancer Sarker et al. (2017) have presented 'k' means clustering algorithm to segment and predict lung cancer from the CT images.

To forecast the lung modules from computerized images, Shakeel et al. (2019) demonstrated, DITNN (Instantaneously Trained Neural Network) & IPCT (Improved Profuse Clustering Technique). Here, the affected cancer portion was segmented through an IPCT method. It also extracts '0' several spectral features from the affected portion.

Yadav et al. (2013) have presented a Foggy K-means method that adopts a powerful diagnostic feature to prevent lung cancer. The unsupervised data were evaluated through a clustering method. This method provides a better result for a real-world problem.

Review of the disease of the lung as cancer and its classification using segmentation

Based on the Res2Net & 3D-UNet the CT image lung segmentation method was derived and it formed the new CNN 3D-Res2UNet implemented by Xiao et al. (2020). However, this 3D-Res2Net has some strongest capabilities of multiple scale feature extraction, symmetric hierarchical connected networks. As a result of this, this feature provided a greater granular effect while maximizing the received field of every layer in the network. This created network was not appropriate for gradient explosion problems and disappearance. As a

Table 2. Recent research of lung cancer detection and classification using clustering methods.

S. No	Author	Finding	Outcome	Dataset	Limitation
1.	Lu et al. (2016)	K-means clustering algorithm	Determination value 412.93	Realtime dataset	Not effective for other lung diseases
2.	Sarker et al. (2017)	k-means clustering algorithms	Accuracy- 95%	SPIEAAPM	Low efficiency
3.	Shakeel et al. (2019)	Improved profuse clustering technique	Accuracy – 98%	Cancer imaging Archive (CIA) dataset	This method only uses cancer identification.
4.	Yadav et al. (2013)	Foggy K-means approach	Cluster (70,87(20 outlier))	SGPGI	Not used for classification

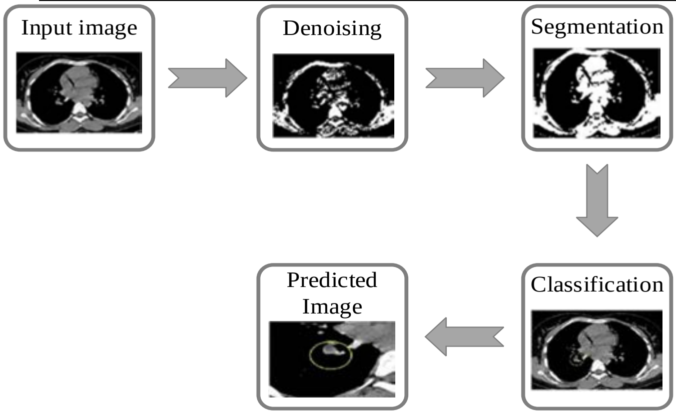

Figure 4. Lung Cancer detection by Segmentation (Skourt et al., 2018)

Table 3. Recent research of lung cancer classification & detection using segmentation methods.

Sl. No	Author	Finding	Outcome	Dataset	Limitation
1.	Xiao et al. (2020)	Segmentation method based on 3D-Unet	Accuracy-95	LUNA16	Yields lower efficiency.
2.	Kamal et al. (2020)	Recurrent 3D-DenseUNet	dice score of 0.7228	NSCLC-Radiomics	Rather than improving the model design, it sees how it performs on additional medical imaging tasks.
3.	Skourt et al. (2018)	U-net	Dice-Coefficient index of 0.9502	LIDC	Offers Lower efficiency
4.	Jiang et al. (2018)	Adversarial domain adaptation method & tumour-aware	Accuracy of 80% is achieved.	NSCLC	Lower computational time.
5.	Senthil et al. (2019)	five optimization algorithms used.	Accuracy-95.89%.	Cancer Imaging Archive (TCIA)	Only used optimization methods.

result, it builds up the accuracy in segmentation and detection. For the volumetric lung cancer detection from that of computer tomography results, Kamal et al. (2020) executed the method 3D-DenseUNet which was considered a deep learning-based architecture. To pull out the nicely crushed spatial-temporal information, the recommended approach has a 3D kind of encoder block to pull out the finely grained coarse-grained temporal features and spatial features and recurrent block of ConvLSTM, then eventually, to rebuild the precise volumized segmentation masks from that of the new featured space, a 3D kind of decoder block had been utilized.

Computed Tomography segmentation by utilizing U-net architecture is the promising DL in the image classification approaches by Skourt et al. (2018). This architecture embodies a path to pull out the high-end data and also it symmetrically expands the path. This network needed to be trained from end-to-end images and works great. In the Dice-Coefficient index, the experimental results demonstrated that it obtained an accurate segmentation of 0.9502.

Jiang et al. (2018) have presented an adversarial domain adaptation and a tumor-aware technique for producing MRI segmentations utilizing unpaired CT and MR images. On the testing phase, by using our methods, the combined MRI yielded an accuracy of 0.74, while on the other hand, utilizing the identical method and it was trained in the semi-supervised occurrences provided 0.80

rates of accuracy. Through the very few amounts of MRI information, their results recommended that tumour-awarded, adversarial domain adaptation and bigger computer tomography datasets have been supported to accomplish reasonably accurate cancer detection. Similarly, to reduce physicians' interpretation of CT scan results, a quick image segmentation technique for medical imaging was developed by Senthil et al. (2019). Implementing the five optimization algorithms assisted in extracting the lung images, which were later to be executed and analysed in the current research.

Review on machine learning techniques for prediction lung disease as cancer

Nanglia et al. (2021) executed ML classifier. The SURF did the feature extraction process with a genetic algorithm. On the other hand, Xie et al. (2021) have employed the naïve Bayes approach to find the plasma metabolites for earlier detection of lung cancer. Furtherly, an interdisciplinary approach was used to detect lung cancer. From the source of MR and PET images, Bębas et al. (2021) implemented HOG and SVM parameters to forecast the ADC (adenocarcinoma) & SCC (squamous cell carcinoma) kind of lung cancer earlier.

lung cancer boundaries with the help of Laser-induced breakdown spectroscopy (LIBS).

Likewise, Chen et al. (2021) have presented SVM and PCA approaches to detect the miRNAs features to analyze the Lung squamous cell carcinoma (LUSC) disease. On the other hand, Shanthi and Rajkumar (2021) have presented decision tree approaches to categorize the lung tumor from the histopathological images. Likewise, Yu et al. (2020) have presented the Extreme Gradient Boosting (XGBoost) approach to detect lung cancer. Here, the probability for chromosomal arm-level CNV from cfDNA was analyzed to diagnose lung cancer.

Uthoff et al. (2019) have employed ANN to variate lung cancer as cancerous and non-cancerous lung nodules through the CT scan images. Alzubi et al. (2019) have presented ANN-based ensemble approach to diagnose lung disease as cancer. The classification period gets minimised by selecting the attributes via the MLMR technique and Newton–Raphsons Maximum Likelihood. Junior et al. (2018) have presented the three machine learning approaches to predict lung cancer such as artificial neural network based on radial basis function-based (RBF), (KNN), and Naive Bayes via CT images.

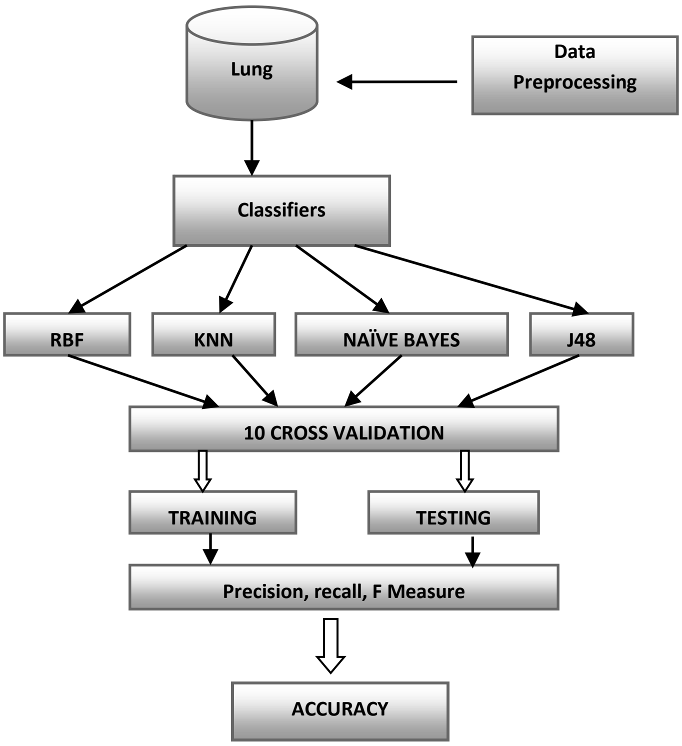

Figure 5. Lung cancer prediction using machine learning (Adapted from Lin et al., 2021)

Hsu et al. (2021) have employed the multilayer perceptron approach with the neural network to predict a different kind of cancer. For the training operations, the ten-Fold cross-validation techniques have been utilized here. Similarly, Lin et al. (2021) have presented the SVM-based RF – boosting tree approach to know the

Maleki et al. (2021) have presented the k-Nearest-Neighbors with a genetic algorithm to detect lung cancer. Here, a genetic algorithm was utilized for the feature selection process. Yuan et al., (2020) have employed the support vector machine (SVM) approach to identify the type of lung cancer. Here, the Monte Carlo approach was utilized for feature selection. Similarly, Asuntha et al.

DOI: https://doi.org/10.52756/ijerr.2023.v30.001

Table 4. Various machine learning techniques for lung cancer prediction.

Sl. No.	Author	Findings	Performance Metrics	Dataset	Limitations
1.	Nanglia et al. (2021)	SVM and NN	Accuracy-98%	ELCAP lung image dataset	Here we utilized lung cancer classification of above 500 sample of computed tomography only.
2.	Xie et al. (2021)	Naive Bayes	98% of accuracy	Own Collected real time dataset	Not incorporated the optimization algorithms.
3.	Bębas et al. (2021)	Texture parameters histogram of Oriented Gradients & SVM	75% of accuracy	Own Collected real time dataset	For the tumor formation mechanisms, this method doesn’t seem to be specify the physical process.
4.	Hsu et al. (2021)	MLP -NN	98% of accuracy	UCI repository	This method does not support complex and high-dimensional, real-time datasets.
5.	Lin et al. (2021)	RF- boosting tree	98% of accuracy	US NIST atomic spectroscopy database	Not easily integrated.
6.	Ye et al. (2020)	SVM and PCA	98% of accuracy	Gene Expression Omnibus database & TCGA	Only use SVM.
7.	Shanthi et al. (2020)	Stochastic Diffusion Search algorithm with NN	Accuracy-89%	TCGA dataset	Less improved accuracy.
8.	Yu et al. (2020)	XGBoost classifier	Specificity - 99%	Real-time dataset from Beijing Chest Hospital	This method does not support large datasets.
9.	Uthoff et al., (2019)	ANN	98% of accuracy	LungX dataset	It was not that flexible.
10.	Alzubi et al. (2019)	ANN-based WONN-MLB	93% of accuracy	Thoracic Surgery Dataset	The WONN-MLB approach does not test all the data points in the dataset.
13.	Junior et al. (2018)	Naive Bayes, ANN, and KNN	Accuracy-89%, 87% and 88%	Retrospective datasets.	In this method, performing training, testing, and validation using a large dataset is a complex process
14.	Maleki et al. (2021)	Nearest-Neighbors with genetic algorithm	99% of accuracy	Pima Indians Dataset	It was improved less by the genetic-based classification
15.	Yuan et al. (2020)	SVM	96% of accuracy	Gene Expression Omnibus	Only use the same features.

DOI: https://doi.org/10.52756/ijerr.2023.v30.001

16.	Asuntha and Srinivasan (2016)	SVM with PSO and GA	89.5%	Own Collected real time dataset	To identify the cancer affected region, this method did not seem to support the Spearman & Pearson algorithm.
17.	Valluru, and Jeya (2020)	SVM with GWO and GA	96.54%	Own Collected real time dataset	By the incorporation of the DL methods, the proposed method could be enhanced.

(2016) have presented the hybrid approach with SVM for detecting lung cancer. Here, canny filter, Superpixel Segmentation, and Gabor filter were used for preprocessing stages. Likewise, Valluru and Jeya (2020) have presented the SVM for classifying the lung tumor portion. Furtherly, grey whole optimization (GWO) and genetic algorithm (GA) were utilized for selecting the features.

Review Lung prediction as cancer by deep learning

Lung modules are considered to be the early indicator of lung cancer detection. The Deep Learning and predominant role in the classification and detection of the cancer, though it has lots of advantages towards the cancer detection, it considered to be the time taking process. So as a result of this CAD (Computer-Aided Diagnosis) is improved for the several DL methods. The DL technique utilized lung cancer detection was discussed in this section.

An immovable texture CNN was shown to produce diverse results on a distinct dataset (Masood et al., 2019; Sori et al., 2019; Lin et al., 2020; Xie et al., 2018; Coudray et al., 2018). Authors used various techniques for better performance and discussed their limitations.

An Adaptive Hierarchical Heuristic Mathematical Model (AHHMM) was presented by Yu et al. (2020) for the evaluation of Cell Lung Cancers progression. This method provides an accuracy of 96.67%. On the other hand, Asuntha and Srinivasan (2020) have employed the Fuzzy Particle Swarm Optimization-CNN (FPSOCNN) identification of cancerous lung modules. Pang et al. (2019) have introduced an automated classification method based on the AdaBoost algorithm and DenseNet lung disease problems effective technique for lung disease type pathology diagnosis. Likewise, Nguyen et al. (2021) have introduced a R-CNN method to identify the lung nodule. Here, the size of the adaptive anchor box was created through ground-truth nodule sizes. Additionally, Chen et al. (2021) have employed a Lung Dense Neural Network (LDNNET) method to classify the lung nodules, which is an adaptive structure based on convents. This method is for classification purposes and also incorporates the SoftMax classifier to mitigate the problems of training deep convnets. Similarly, Teramoto et al. (2017) have employed deep CNN to categorize various lung cancer .

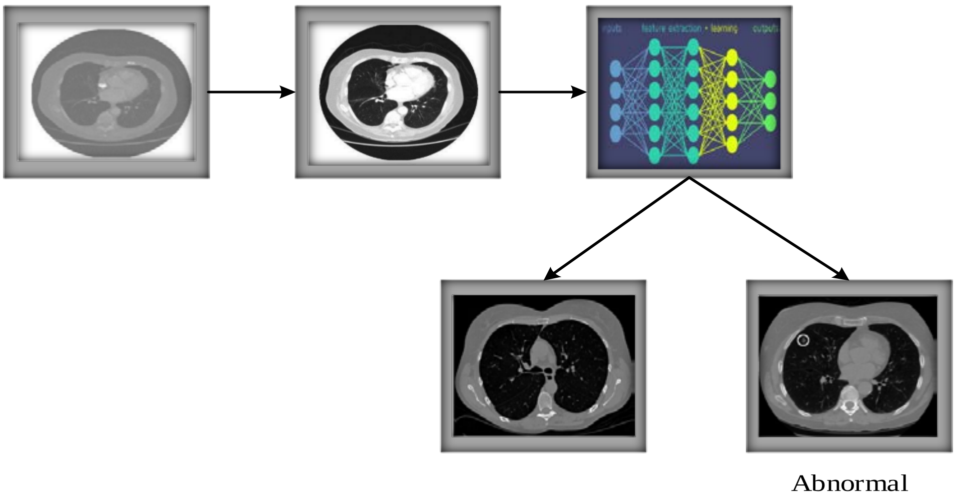

Figure 6. Lung cancer classification based on deep learning (Masood et al., 2019)

Table 5. Various deep learning (DL) approaches for lung cancer diagnosis.

Sl. No	Author	Findings	Performance Outcome	Dataset	Limitation
1.	Masood et al. (2019)	Cloud-Based 3D Deep Convolutional Neural Network	Accuracy is 98.51%	LUNA16, ANODE09, and LIDC-IDR datasets	This method detects only micronodules with a diameter less than 3 mm
2.	Sori et al. (2019)	Multi-Path Convolutional Neural Network	87.8% accuracy	KDSB 2017 challenge dataset	Less improved accuracy.
3.	Lin et al. (2020)	optimization using 2D CNN.	98.83% & 99.97% accuracy using SPIE-AAPM attributes	LIDC-IDRI and SPIE-AAPM dataset	Computational efficiency is low
4.	Xie et al. (2018)	Multi-View model	91.60% accuracy	LIDC-IDRI dataset	This method uses the nodules with an uncertain level of malignancy
5.	Coudray et al. (2018)	CNN	85.6%	The Cancer Genome Atlas (TCGA) dataset	deep-learning models visualization tools 30 281 are not included in identifying the cancer features.
6.	Yu et al. (2020)	Adaptive Hierarchical Heuristic Mathematical Model	96.67 % accuracy	Diagnostic Image Analysis Group dataset	The presented approach was not effective therefore, hybridization is required.
7.	Asuntha and Srinivasan (2020)	FPSO with CNN.	The accuracy obtained as 94.97%	A real-time data set from Arthi Scan Hospital	Not included the pulmonary nodules and optimize this method.
8.	Pang et al. (2019)	Automated classification model based on DenseNet network and AdaBoost algorithm.	89.85% accuracy	A real-time data set from Shandong Provincial Hospital	To proceed the classification of lung cancer, this method did not seem to be supportive for high quality.
9.	Nguyen et al. (2021)	Adaptive anchor box with fastest R-CNN model.	95.7% accuracy	LUNA16 dataset	In the trade-off enhanced performance, this method improvises the network complexity.
10.	Chen et al. (2021)	LDNNET	98.8396% on LUNA 16 and 99.9480% on Kaggle DSB 2017	Kaggle DSB 2017 databases.	Only used different datasets and not used others methods.

DOI: https://doi.org/10.52756/ijerr.2023.v30.001

11.	Tiwari et al. (2021)	MU based 3FCM and TWEDLNN algorithm.	96% of accuracy	LIDC-IDRI database	Its utilized in computed tomography images only.
12.	Teramoto et al. (2017)	DCNN	71.1% of accuracy	Real-time dataset from interventional cytology.	This method could not able to classify the cells and array of cells comprehensively.
13.	Sahu et al. (2018)	multiple view sampling-based multi-section CNN	93.18% of accuracy	IDRI & LIDC dataset have been utilized.	Lung nodules in the large CT scan images were not effectively predicted
14.	Wang and Charkborty (2021)	3D-CNN, RNN based nodule detector	AUC obtained as 0.86	Lung images 2016 dataset (LUNA16 (LIDC), (NDSB)	To access the probability of cancer, this supported only one of the models along with the higher malignancy.
15.	Polat and Danaei (2019)	3D-CNN module	The accuracy obtained as 91.81%	Kaggle dataset & data science bowl.	The method does not support more complicated CT scan images.
16.	Nasrullah et al. (2019)	Classification system & multi-strategy-based lung module detection.	Specificity 91%	LUNA16 and LIDC-IDRI dataset	Low efficiency.
17	Lakshmanaprabu et al. (2019)	ODNN -LDA with MGSA	Accuracy 94.56%	ELCAP Public Lung Image Database	This method does not support high dosage CT lung images.

Tiwari et al. (2021) and Sahu et al. (2018) identify lung cancer. CNN framework for attaining the structural data of nodules from CT images. The cross-section of the nodule was obtained from multiple view angles. The volumetric data of nodules were encoded into a compact presentation using the aggregated data. Likewise, Wang and Charkborty (2021) demonstrated a 3D-CNN based module and RNN based detector to offer an entire lung cancer identification system. This model incorporates two cascaded modules: the nodule detection module and the risk evaluation module. Based on the lung cancer support vector machine, Polat and Danaei (2019) expressed the 3D convolutional neural network along with convolutional Soft Max & hybrid 3D-CNN and RBF. Based on the lung module classification and detection system, Nasrullah et al. (2019) implemented a strategy-based system. Here the detection was performed via fastest R-CNN. The module classification was performed via a GBM (gradient boosting machine). Similarly, Lakshmanaprabu et al. (2019) presented ODNN with LDA and MGSA for detecting cancer through computerized images.

Evaluation and Analysis

This analysis aims to investigate several methods for fuzzy, clustering, segmentation, ML, and DL. Various kinds of lung cancer detection and classification and metrics have been introduced. The review paper is analyzed lung cancer detection and classification.

The term of accuracy is described as the amount of prediction made correctly. The mathematical representation is below,

$$\text{Accuracy} = \frac{tp+tn}{tp+fp+fn+tn} \qquad \text{-(1)}$$

$$\text{Precision} = \frac{tp}{tp+fp} \quad \text{-(2)}$$

$$\text{Recall} = \frac{tp}{tp+fn} \quad \text{-(3)}$$

$$\text{F} - \text{measure} = \frac{2\times(\text{recall}\times\text{precision})}{\text{recall}+\text{precision}} \quad \text{-(4)}$$

True Positive, True Negative, False Positive, False Negative represented by TP, TN, FP, FN in these equations. The total number of classifications using lung cancer in various techniques taken according to this survey paper is pictorially shown in figure 7.

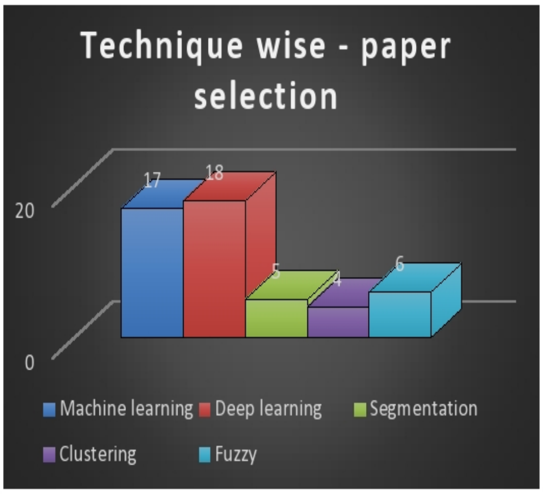

Figure 7. Total number of lung cancer classifications used

The analysis explains that lung cancer classification based on fuzzy, clustering, segmentation, ML, and deep learning was utilized widely and is also applied in the present research. Around twenty-one research articles are reviewed from the various methods: fuzzy, clustering, segmentation, etc., taken from 2011-2021. Previous work on lung cancer classification taken according to this review paper is pictorially shown in fig 8, the pie chart.

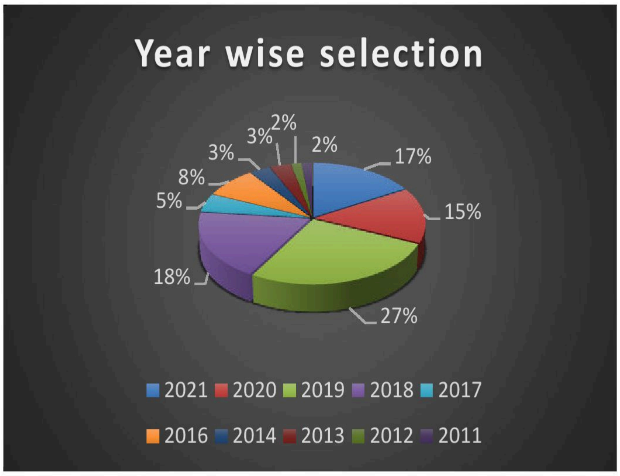

Figure 8. Year wise analysis of lung cancer classification

The analysis describes lung cancer classification with previous to the current period. There are around twenty research articles taken from the lung cancer classification.

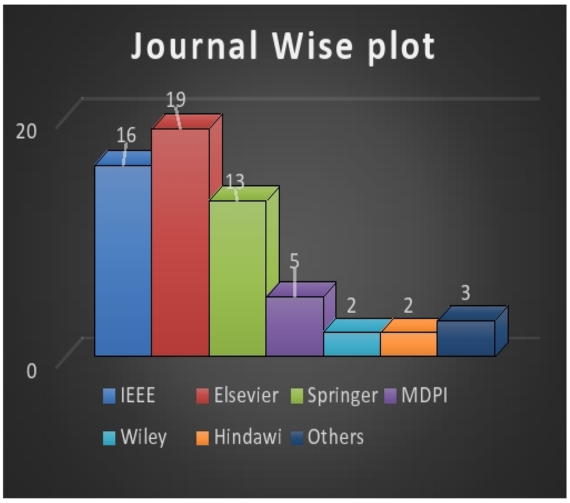

Figure 9. Journal wise selection of lung cancer classification

The journal-wise selection of classifications using lung cancer in various techniques taken according to this survey paper is pictorially shown in figure 9.

Conclusion and future Scope

Various strategies have been reviewed for identifying and classifying lung cancer. This article aims to focus on the progression and historical progress review of various machine learning, clustering, segmentation, deep learning, and fuzzy approaches for classifying lung cancer, as well as their accuracy along with the conceptual frameworks from these techniques. Intelligibly combining deep learning with other methods performed much better than uncombined ways for classifying Lung cancer detection, according to research findings. Deep learning-based automated procedures using computer-aided diagnosis systems have just been presented in this review. The accuracy rate of using morphological and time-frequency-based features was more than 99 percent, indicating that traditional deep learning classifiers are efficient for lung cancer detection. According to a new study, deep learning methodologies outperformed regular classifiers in computing complexity, efficiency, and accuracy, all of which are significant in real-time applications.

The validation process would comprise testing various models on the same datasets for direct comparisons, and testing in various population sets. The multi-omics approach will help to integrate the various immune cells of biomarkers. In that view, systems cancer immunology may guide clinical decision-making earlier. The

DOI: https://doi.org/10.52756/ijerr.2023.v30.001

limitation of non-invasive biosensors comprehends the possibility of binding in patient's biomarker level, samples, and smaller target size. For molecule detection, the investigation marks the trends of electrodes which are made up of nano-biosensor. For preprocessing and segmentation process, effective techniques need to be employed. This leads to accurately predicting the cancer portion.

Conflict of Interest

Nil

References

Alzubi, J. A., Bharathikannan, B., Tanwar, S., Manikandan, R., Khanna, A., & Thaventhiran, C. (2019). Boosted neural network ensemble classification for lung cancer disease diagnosis. *Applied Soft Computing*, *80*, 579-591. https://orcid.org/10.1016/j.asoc.2019.04.031.

Asuntha, A., & Srinivasan, A. (2020). Deep learning for lung Cancer detection and classification. *Multimedia Tools and Applications*, *79*, 7731-7762. https://orcid.org/10.1007/s11042-019-08394-3.

Asuntha, A., Singh, N., & Srinivasan, A. (2016). PSO, genetic optimization and SVM algorithm used for lung cancer detection. *Journal of Chemical and Pharmaceutical Research*, *8*(6), 351-359.

Bębas, E., Borowska, M., Derlatka, M., Oczeretko, E., Hładuński, M., Szumowski, P., & Mojsak, M. (2021). Machine-learning-based classification of the histological subtype of non-small-cell lung cancer using MRI texture analysis. *Biomedical Signal Processing and Control*, *66*, 102446. 10.1016/j.bspc.2021.102446.

Camarlinghi, N., Gori, I., Retico, A., Bellotti, R., Bosco, P., Cerello, P., Gargano, G., Lopez, T. E., Megna, R., Peccarisi, M., & Fantacci, M. E. (2012). Combination of computer-aided detection algorithms for automatic lung nodule identification. *International Journal of Computer Assisted Radiology and Surgery*, *7*, 455-464. 10.1007/s11548-011-0637-6.

Capizzi, G., Sciuto, G. L., Napoli, C., Połap, D., & Woźniak, M. (2019). Small lung nodules detection based on fuzzy-logic and probabilistic neural network with bioinspired reinforcement learning. *IEEE Transactions on Fuzzy Systems*, *28*(6), 1178-1189. https://orcid.org/10.1109/TFUZZ.2019.2952831.

Chen, Y., Wang, Y., Hu, F., Feng, L., Zhou, T., & Zheng, C. (2021). LDNNET: towards robust classification of lung nodule and cancer using lung dense neural network. *IEEE Access*, *9*, 50301-50320. https://orcid.org/10.1109/ACCESS.2021.3068896.

Choi, W. J., & Choi, T. S. (2014). Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor. *Computer methods and programs in Biomedicine*, *113*(1), 37-54. https://orcid.org/10.1016/j.cmpb.2013.08.015.

Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A.L., Razavian, N., & Tsirigos, A. (2018). Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nature medicine*, *24*(10), 1559-1567. https://orcid.org/10.1038/s41591-018-0177-5.

Hsu, C. H., Chen, X., Lin, W., Jiang, C., Zhang, Y., Hao, Z., & Chung, Y. C. (2021). Effective multiple cancer disease diagnosis frameworks for improved healthcare using machine learning. *Measurement*, *175*, 109145. https://orcid.org/10.1016/j.measurement.2021.109145.

Jiang, J., Hu, Y. C., Tyagi, N., Zhang, P., Rimner, A., Mageras, G. S., Deasy, J.O., & Veeraraghavan, H. (2018). Tumor-aware, adversarial domain adaptation from CT to MRI for lung cancer segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*, Springer International Publishing, pp. 777-785. https://orcid.org/10.1007/978-3-030-00934-2_86.

Junior, J. R. F., Koenigkam-Santos, M., Cipriano, F. E. G., Fabro, A. T., & de Azevedo-Marques, P. M. (2018). Radiomics-based features for pattern recognition of lung cancer histopathology and metastases. *Computer Methods and Programs in Biomedicine*, *159*, 23-30. https://orcid.org/10.1016/j.cmpb.2018.02.015.

Kamal, U., Rafi, A. M., Hoque, R., Wu, J., & Hasan, M. K. (2020). Lung cancer tumor region segmentation using recurrent 3d-denseunet. In *Thoracic Image Analysis: Second International Workshop, TIA 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 2*, Springer International Publishing, pp. 36-47. https://orcid.org/10.1007/978-3-030-62469-9_4.

Lakshmanaprabu, S. K., Mohanty, S. N., Shankar, K., Arunkumar, N., & Ramirez, G. (2019). Optimal deep learning model for classification of lung cancer on CT images. *Future Generation Computer Systems*, *92*, 374-382.

https://orcid.org/10.1016/j.future.2018.10.009.

Lin, C. J., Jeng, S. Y., & Chen, M. K. (2020). Using 2D CNN with Taguchi parametric optimization for lung cancer recognition from CT images. *Applied Sciences*, *10*(7), 2591. 10.3390/app10072591

Lin, X., Sun, H., Gao, X., Xu, Y., Wang, Z., & Wang, Y. (2021). Discrimination of lung tumor and boundary tissues based on laser-induced breakdown spectroscopy and machine learning. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *180*, 106200. https://orcid.org/10.1016/j.sab.2021.106200.

Lu, J., Chen, L., Yin, J., Huang, T., Bi, Y., Kong, X., Zheng, M., & Cai, Y. D. (2016). Identification of new candidate drugs for lung cancer using chemical–chemical interactions, chemical–protein interactions and a K-means clustering algorithm. *Journal of Biomolecular Structure and Dynamics*, *34*(4), 906-917. https://orcid.org/10.1080/07391102.2015.1060161

Maleki, N., Zeinali, Y., & Niaki, S. T. A. (2021). A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection. *Expert Systems with Applications*, *164*, 113981. https://orcid.org/10.1016/j.eswa.2020.113981.

Manickavasagam, R., & Selvan, S. (2019). Automatic detection and classification of lung nodules in CT image using optimized neuro fuzzy classifier with cuckoo search algorithm. *Journal of Medical Systems*, *43*(3), 77. 10.1007/s10916-019-1177-9.

Manikandan, T., & Bharathi, N. (2016). Lung cancer detection using fuzzy auto-seed cluster means morphological segmentation and SVM classifier. *Journal of Medical Systems*, *40*, 1-9. https://orcid.org/10.1007/s10916-016-0539-9.

Manikandan, T., & Bharathi, N. (2011). Lung cancer diagnosis from CT images using fuzzy inference system. In *Computational Intelligence and Information Technology: First International Conference, CIIT 2011, Pune, India, November 7-8, 2011*. Springer Berlin Heidelberg. *Proceedings*, pp. 642-647. https://orcid.org/10.1007/978-3-642-25734-6.

Masood, A., Yang, P., Sheng, B., Li, H., Li, P., Qin, J., Lanfranchi, V., Kim, J., & Feng, D. D. (2019). Cloud-based automated clinical decision support system for detection and diagnosis of lung cancer in chest CT. *IEEE journal of Translational Engineering in Health and Medicine*, *8*, 1-13. https://orcid.org/10.1109/JTEHM.2019.2955458.

Nair, M., Sandhu, S. S., & Sharma, A. K. (2018). Cancer molecular markers: A guide to cancer detection and management. Academic Press. In *Seminars in Cancer Biology*, Vol. 52, pp. 39-55. https://orcid.org/10.1016/j.semcancer.2018.02.002.

Nanglia, P., Kumar, S., Mahajan, A. N., Singh, P., & Rathee, D. (2021). A hybrid algorithm for lung cancer classification using SVM and Neural Networks. *ICT Express*, *7*(3), 335-341. https://orcid.org/10.1016/j.icte.2020.06.007.

Nasrullah, N., Sang, J., Alam, M. S., Mateen, M., Cai, B., & Hu, H. (2019). Automated lung nodule detection and classification using deep learning combined with multiple strategies. *Sensors*, *19*(17), 3722. https://orcid.org/10.3390/s19173722.

Nguyen, C. C., Tran, G. S., Burie, J. C., & Nghiem, T. P. (2021). Pulmonary nodule detection based on faster R-CNN with adaptive anchor box. *Ieee Access*, *9*, 154740-154751. https://orcid.org/10.1109/ACCESS.2021.3128942.

Pang, S., Zhang, Y., Ding, M., Wang, X., & Xie, X. (2019). A deep model for lung cancer type identification by densely connected convolutional networks and adaptive boosting. *IEEE Access*, *8*, 4799-4805. 10.1109/ACCESS.2019.2962862.

Peña, D. M., Luo, S., & Abdelgader, A. M. (2016). Auto diagnostics of lung nodules using minimal characteristics extraction technique. *Diagnostics*, *6*(1), 13. https://orcid.org/10.3390/diagnostics6010013.

Polat, H., & Danaei, M.H. (2019). Classification of pulmonary CT images by using hybrid 3D-deep convolutional neural network architecture. *Applied Sciences*, *9*(5), 940. 10.3390/app9050940.

Sahu, P., Yu, D., Dasari, M., Hou, F., & Qin, H. (2018). A lightweight multi-section CNN for lung nodule classification and malignancy estimation. *IEEE Journal of Biomedical and Health Informatics*, *23*(3), 960-968. https://orcid.org/10.1109/JBHI.2018.2879834.

Sahu, S. P., Agrawal, P., Londhe, N. D., & Verma, S. (2019). Lung segmentation of CT images using fuzzy C-means for the detection of cancer in early stages. In *Advances in Data and Information Sciences: Proceedings of ICDIS 2017*, Springer Singapore. Vol. 2, pp. 167-176. https://orcid.org/10.1007/978-981-13-0277-0_14.

Santos, A. M., de Carvalho Filho, A. O., Silva, A. C., De Paiva, A. C., Nunes, R. A., & Gattass, M. (2014). Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM. *Engineering Applications of Artificial Intelligence*, *36*, 27-39. https://orcid.org/10.1016/j.engappai.2014.07.007.

Sarker, P., Shuvo, M. M. H., Hossain, Z., & Hasan, S. (2017). Segmentation and classification of lung tumor from 3D CT image using K-means clustering algorithm. IEEE. In *2017 4th International Conference on Advances in Electrical Engineering* (ICAEE), pp. 731-736. https://orcid.org/10.1109/ICAEE.2017.8255451.

Senthil Kumar, K., Venkatalakshmi, K., & Karthikeyan, K. (2019). Lung cancer detection using image segmentation by means of various evolutionary algorithms. *Computational and Mathematical Methods in Medicine*, *2019*, 4909846. https://orcid.org/10.1155/2019/4909846.

Setio, A. A. A., Traverso, A., De Bel, T., Berens, M. S., Van Den Bogaard, C., Cerello, P., Chen, H., Dou, Q., Fantacci, M.E., Geurts, B., Gugten, R.V., Heng, P.A., Jansen, B., de Kaste, M.M.J., Kotov, V., Lin, J.Y., Manders, J.T.M.C., Sóñora-Mengana, A, García-Naranjo, J.C., Papavasileiou, E., Prokop, M., Saletta, M., Schaefer-Prokop, C.M., Scholten, E.T., Scholten, L., Snoeren, M.M., Torres, E.L., Vandemeulebroucke, J., Walasek, N., Zuidhof, G.C.A., Ginneken, B.V., & Jacobs, C. (2017). Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Medical image analysis*, *42*, 1-13. https://orcid.org/10.1016/j.media.2017.06.015.

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. https://orcid.org/10.1016/j.measurement.2019.05.027.

Shanthi, S., & Rajkumar, N. (2021). Lung cancer prediction using stochastic diffusion search (SDS) based feature selection and machine learning methods. *Neural Processing Letters*, *53*, 2617-2630. 10.1007/s11063-020-10192-0.

Silvestri, G. A., Tanner, N. T., Kearney, P., Vachani, A., Massion, P. P., Porter, A., Springmeyer, S.C., Fang, K.C., Midthun, D., & Mazzone, P.J. (2018). Assessment of plasma proteomics biomarker's ability to distinguish benign from malignant lung nodules: results of the PANOPTIC (Pulmonary Nodule Plasma Proteomic Classifier) trial. *Chest*, *154*(3), 491-500. https://orcid.org/10.1016/j.chest.2018.02.012.

Skourt, B.A., El-Hassani, A., & Majda, A. (2018). Lung CT image segmentation using deep neural networks. *Procedia Computer Science*, *127*, 109-113. https://orcid.org/10.1016/j.procs.2018.01.104.

Sori, W. J., Feng, J., & Liu, S. (2019). Multi-path convolutional neural network for lung cancer detection. *Multidimensional Systems and Signal Processing*, *30*, 1749-1768. https://orcid.org/10.1007/s11045-018-0626-9.

Teramoto, A., Tsukamoto, T., Kiriyama, Y., & Fujita, H. (2017). Automated classification of lung cancer types from cytological images using deep convolutional neural networks. *BioMed Research International*, *2017*, 4067832. https://orcid.org/10.1155/2017/4067832.

Teramoto, A., & Fujita, H. (2013). Fast lung nodule detection in chest CT images using cylindrical nodule-enhancement filter. *Int. J. Comput. Assist. Radiol. Surg.*, *8*(2), 193-205. https://orcid.org/10.1007/s11548-012-0767-5.

Tiwari, L., Raja, R., Awasthi, V., Miri, R., Sinha, G. R., Alkinani, M. H., & Polat, K. (2021). Detection of lung nodule and cancer using novel Mask-3 FCM and TWEDLNN algorithms. *Measurement*, *172*, 108882. 10.1016/j.measurement.2020.108882.

Uthoff, J., Stephens, M. J., Newell Jr, J. D., Hoffman, E. A., Larson, J., Koehn, N., De Stefano, F.A., Lusk, C.M., Wenzlaff, A.S., Watza, D., Neslund-Dudas, C., Carr, L.L., Lynch, D.A., Schwartz, A.G., & Sieren, J. C. (2019). Machine learning approach for distinguishing malignant and benign lung nodules utilizing standardized perinodular parenchymal features from CT. *Medical Physics*, *46*(7), 3207-3216. https://orcid.org/10.1002/mp.13592.

Valluru, D., & Jeya, I. J. S. (2020). IoT with cloud based lung cancer diagnosis model using optimal support vector machine. *Health Care Management Science*, *23*, 670-679. https://orcid.org/10.1007/s10729-019-09489-x.

Wang, W., & Charkborty, G. (2021). Automatic prognosis of lung cancer using heterogeneous deep learning models for nodule detection and eliciting its morphological features. *Applied Intelligence*, *51*, 2471-2484. https://orcid.org/10.1007/s10489-020-01990-z.

Wu, C., Luo, C., Xiong, N., Zhang, W., & Kim, T. H. (2018). A greedy deep learning method for medical disease analysis. *IEEE Access*, *6*, 20021-20030. https://orcid.org/10.1109/ACCESS.2018.2823979.

Xiao, Z., Liu, B., Geng, L., Zhang, F., & Liu, Y. (2020). Segmentation of lung nodules using improved 3D-UNet neural network. *Symmetry*, *12*(11), 1787. https://orcid.org/10.3390/sym12111787.

Xie, Y., Meng, W. Y., Li, R. Z., Wang, Y. W., Qian, X., Chan, C., Yu, Z.F., Fan, X.X., Pan, H.D., Xie, C., Wu, Q.B., Yan, P.Y., Liu, L., Tang, Y.J., Yao, X.J., Wang, M.F., & Leung, E. L. H. (2021). Early lung cancer diagnostic biomarker discovery by machine learning methods. *Translational Oncology*, *14*(1), 100907. https://orcid.org/10.1016/j.tranon.2020.100907.

Xie, Y., Xia, Y., Zhang, J., Song, Y., Feng, D., Fulham, M., & 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. https://orcid.org/10.1109/TMI.2018.2876510.

Yadav, A. K., Tomar, D., & Agarwal, S. (2013). Clustering of lung cancer data using foggy k-means. In *2013 International Conference on Recent Trends in Information Technology* (ICRTIT), IEEE, pp. 13-18. https://orcid.org/10.1109/ICRTIT.2013.6844173.

Ye, Z., Sun, B., & Xiao, Z. (2020). Machine learning identifies 10 feature miRNAs for lung squamous cell carcinoma. *Gene*, *749*, 144669. https://orcid.org/10.1016/j.gene.2020.144669.

Yu, D., Liu, Z., Su, C., Han, Y., Duan, X., Zhang, R., Liu, X., Yang, Y., & Xu, S. (2020). Copy number variation in plasma as a tool for lung cancer prediction using Extreme Gradient Boosting (XGBoost) classifier. *Thoracic Cancer*, *11*(1), 95-102. https://orcid.org/10.1111/1759-7714.13204.

Yu, H., Zhou, Z., & Wang, Q. (2020). Deep learning assisted predict of lung cancer on computed tomography images using the adaptive hierarchical heuristic mathematical model. *IEEE Access*, *8*, 86400-86410. 10.1109/ACCESS.2020.2992645.

Yuan, F., Lu, L., & Zou, Q. (2020). Analysis of gene expression profiles of lung cancer subtypes with machine learning algorithms. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1866*(8), 165822. https://orcid.org/10.1016/j.bbadis.2020.165822.

How to cite this Article:
Anindita Saha and Rakesh Kumar Yadav (2023). Study on segmentation and prediction of lung cancer based on machine learning approaches. *International Journal of Experimental Research and Review*, *30*, 1-14.
DOI : https://doi.org/10.52756/ijerr.2023.v30.001