Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.777673

LAYERED ARCHITECTURE FOR MACHINE LEARNING APPLICATIONS FOR THE LUNG CANCER NODULE DETECTION AND CLASSIFICATION OF COMPUTED TOMOGRAPHY SCAN IMAGES

M. Premchander[1]
Ph. D Research Scholar, Dept of CSE, GITAM Deemed-to-be University, Vishakhapatnam, India.
premchandermunimanda@gmail.com

Dr. M. Venkateshwara Rao[2]
Professor Dept of IT, GITAM Deemed-to-be University, Vishakhapatnam, India.
mandapati9venkat@gmail.com

Dr. T.V. Rajinikanth[3]
Professor CSE Dept, Sreenidhi Institute of Science & Technology, Hyderabad. India.
rajinitv@gmail.com

Dr. B. Subbareddy[4]
Professor & HOD Dept of IT, Malla Reddy Engineering College for Women, Autonomous, Hyderabad. India.
bvsr79@gmail.com

Abstract: Medical imaging technologies are required for both the lung cancer early identification and continued monitoring during treatment. Computed Tomography scan images, Chest X-rays, positron emission tomography images, and other medical imaging procedures have all been extensively studied for the identification of lung cancer, magnetic resonance imaging, computed tomography, and molecular imaging methods. These methods have some drawbacks, such as the inability to automatically categorize cancer images, which makes them unsuitable for patients with other illnesses. The development of a sensitive, exact approach for detecting lung cancer in its earliest stages is desperately needed. Applications for medical image-based and textural data methods are rapidly growing. Deep learning is one of the medical imaging fields that is expanding the accelerated. The detection and classification of lung nodules can be done more rapidly and accurately by physicians using medical imaging that is based on deep learning systems. We used kaggle dataset for finding the appropriate results, and in this paper discuss the most current advancements in imaging methods relies on deep learning for the early diagnosis of lung cancer.
Keywords: Computed Tomography scan Images, Lung cancer, Segmentation, Classification, Convolution Neural Network

1.Introduction

In the US, lung cancer has the greatest deadly and fatality rates and is the most common disease as well. GLOBOCAN predicted that there would be in the year 2018 new cases are registered 2.09 and deaths are 1.76 millions globally has considerably grown. Small cell lung cancer (SCLC) account for 12–15% of lung cancer cases, while non-small cell lung carcinoma (NSCLS) liability for 85-88% of lung cancer cases as mentioned by Zehra Karhan et al [1].

Due to the invasiveness and heterogeneity of early detection and treatment of lung cancer are crucial to improving the survival rate 5 years. Many medical imaging procedures, including chest X-rays, MRIs, CT scan images, PET scans, thoroughly studied to cancer detection of lung nodules over the past 25 years. This type of diagnosis has a high false-positive rate,

The purpose of detecting lung cancer, LDCT has been proposed Ada et al [2]. However, participants undergoing LDCT accounted for the majority of cancer-related mortality. In the last ten years, classification of lung nodules has increased to unprecedented heights, in the development of CAD methodology. D. Vinitha[3]. In recent years, a number of deep learning-based lung nodule classification techniques have been put out, with continuously rising state-of-the-art performance. In order to extract features, Chip M et al.[4] created a multiscale convolution neural network (CNN), which they later used a supervised random forest classifier.

K. V. Bawane et al. [5] presented an accuracy 86.5% by combining deep learning data with developed characteristics to distinguish between normal and malignant nodules. The CAD for the lung was developed by combining the LUNA16 dataset and the National Lung Screening Trial (NLST) subset. Thereafter, a CT picture was automatically segmented into likely nodules using a pretrained U-Net. In a test that was randomly separated from the aforementioned data, a 3D CNN was able to recognize early-stage lung cancer using the segmented nodules with an AUC of 0.83, Sukhjinder Kaur et al. [6] built an end-to-end collection of 3D CNN modules to compute the overall risk of lung malignancy based on auto-detection of nodules using the full-size, publicly available NLST dataset. CAD systems can do cancer identification better than trained radiologists. According to Sathyan H and colleagues [7], the CAD-based lung cancer detection approach typically entails four steps: image processing, ROI extraction, feature selection, and classification. The feature selection, accuracy sensitivity, and classification stages are the most crucial. Classifying benign and malignant nodules is difficult. For the detection and classification of lung cancer, numerous computer-aided detection (CAD) systems have undergone substantial research,

Himali Dalvi et al[8]. Though, many academics have utilized deep learning techniques to aid radiologists in making more accurate diagnosis. Past research have demonstrated that CAD systems based on deep learning can significantly increase the effectiveness and accuracy of medical diagnosis, particularly when identifying several common malignancies like breast and lung tumors, Jin H, et al[9]. Deep learning-based CAD solutions use different network architectures than conventional systems, are the core topics of research on deep learning-based lung imaging approaches. To improve deep learning model performance, researchers mostly concentrate on creating new network architectures and loss functions. Review papers on deep learning approaches have lately been published by a number of research groups, Radhika et al [10]. Yet, deep learning techniques have advanced quickly, and every year, numerously new approaches and applications appear. The idea behind this paper is to study the topics that earlier studies are unable to cover, lung cancer detection. Recent developments and techniques are discussed in this work. The most cutting-edge Strategies for detecting lung cancer using deep learning are highlighted in this article. This paper also emphasizes current successes, pertinent research road blocks, and upcoming research directions, D. Moitra et al[11]. The major focuses of the research on deep learning-based lung imaging algorithms include pulmonary nodule identification, segmentation, prospects.

2. Related Work

Radiologists use medical imaging equipment to identify lung conditions. The CT scan images have greater advantages over these medical imaging methods, including details on location, size of the nodule, extension and characterization which are used to detect lung cancer and nodules. More accurate radiation targeting is made possible by 3D CT, which has a substantial impact on lung health, management of cancer. To categorize lung cancer in CT lung pictures, Dr. T. V. RajiniKanth et al. [12] created a system for automatic detection using the best deep neural network and linear discriminate analysis (LDA). The LDA decreased the extracted image features to minimise the feature dimension. A modified gravity search technique was used to apply and optimise the ODNN to produce a more precise categorization result, A. Abdulla et al[13].Compared to CT, LDCT is more capable of detecting cancer and early-stage lung nodules. It does not, however, aid in lowering lung cancer mortality. For high-risk smokers between the ages of 55 and 74, LDCT is advised to be performed every year, Swati Mukherjeeet al [14]. When compared to a CT scan image, PET has significantly higher sensitivity and specificity for detecting lung nodules because of reactive nodule disease., Syed Saba Raoof et al [15]. Increased overall survival rates and progression durations are well correlated with PET. Solitary pulmonary nodules have been diagnosed with 18F-FDG PET, KameliaRoy et al[16]. More precisely, PET-assisted radiation treats roughly. As per Sumathipala et al[17] discussed that Stage IIIA lung cancer affects 32% of people. The PET provides a useful response assessment for patients with NSCLC receiving induction chemotherapy. According to Zhang et al.[18], the nodule binary masks were included in the data collection.

2.1 Lung Cancer Detection Challenges.

In-depth reviews of the literature from 2014 to 2022 are conducted in this study. Table 1 demonstrate how deep learning-based lung imaging systems have been used to segment, detect, and classify lung nodules utilising current medical images with great efficiency and cutting-edge performance. Deep Learning methods have been used frequently to build convolution neural networks for the detection of lung cancer and lowering false-positive rates. The CT scan is the imaging method that the CAD system uses the most frequently to diagnose lung cancer, according to past studies. There have only been a few published trials employing 3D CNN to identify lung cancer. Segmentation and classification using deep learning algorithms have been successful.

Deep learning for lung cancer detection, however, still a number of issues. Because there are no clear standards for medical image acquisition standards, physicians have not completely embraced deep learning techniques for routine clinical practise. It might be reduced by the unified acquisition techniques. Second, deep learning approaches often need a lot of well annotated medical images to finish training assignments. Even when done by seasoned radiologists, collecting a sizable annotated image dataset is expensive and time-consuming. The lack of labeled data was overcome using a variety of techniques. For instance, transfer learning may be used to address the training issue with limited samples.

Table 1: Different types of methods and their performance metrics

Reference	Year	Model	Type of Image	Dataset used	Results obtained
[22]	2012	Intensity features+SVM	CT scan Images	DLCST	Accuracy:28.23%
[27]	2013	CNN	CT scan Images	LIDC	Sensitivity:74.56%
[33]	2014	FF-BPNN	CT scan Images	LIDC	Sensitivity:92.01%
[35]	2015	DeepBelief Network(DBN)	CT scan Images	LIDC	Sensitivity:75.24%
[43]	2015	Multi-Scale CNN	CT scan Images	LIDC-IDRI	Accuracy:93.84%
[38]	2016	ConvNets	CT scan Images	DLCST	Accuracy:85.43%
[45]	2016	Unsupervised Features+SVM	CT scan Images	DLCST	Accuracy:95.10%
[26]	2017	Deep3D DPN+GBN	CT scan Images	LIDC-IDRI	Accuracy:94.26%
[28]	2018	Feature Representation using Deep Autoencoder	CT scan Images	ELCAP	Accuracy:89.34%
[32]	2018	Multiview Multi-Scale CNN	CT scan Images	LIDC-IDRI+ELCAP	Accuracy:90.76%
[34]	2019	3D Mixnet	CT scan Images	LIDC-IDRI+LUNA16	Accuracy:97.45%
[40]	2019	DCNN	CT scan Images	Interventional Cytology	Sensitivity:99.28%
[42]	2020	2D CNN	CT scan Images	LIDC+ LUNA16	Accuracy:93.62%
[55]	2020	ResNet50	CT scan Images	LIDC	Accuracy:89.30%
[52]	2021	VGG19	CT scan Images	LIDC	Accuracy:95.93%
[57]	2021	DCNN	CT scan Images	Interventional Cytology	Sensitivity:98.53%
[41]	2022	2 pathway Morphology based CNN(2path Morph)	CT scan Images	LIDC-IDRI	Sensitivity:93.67%

[46]	2022	Cat Swarm Optimization based CAD for lung cancer classification(CSO-CAD LCC)	CT scan Images	BenchMark	Sensitivity:87.84%
[54]	2022	CNN based Multi-Task Learning	CT scan Images	LIDC-IDRI	Sensitivity:97.82%

Prediction accuracy and stability will unavoidably be impacted by insufficient data. Consequently, one of the directions for future study is to utilizing limited to improve prediction accuracy. Fourth, creating a prediction model that is robust is a difficult undertaking. Only a single dataset is suitable for the majority of deep learning algorithms. Due to varying acquisition conditions, tools, timing, and other circumstances, the image of the same disease may alter dramatically. Because of this, the generality and resilience of Existing deep learning models suffered from weak robustness and generalization as a result. Instead, because they are independent of the disease's incidence and are easier to apply in reality, researchers should strive to present clinical measurements with greater sensitivity and specificity whenever possible. Researchers should focus more on the following research directions in the future: The following strategies should be used to improve performance: 1. Create new convolution networks and loss functions; 2.Weak supervised learning; 3. Incorporate incorporating existing clinical expertise into model training; and 4. To improve the realism and sensitivity of models and to provide additional context for the study area.

Finally, strong interpretability is needed for deep learning's therapeutic applications, yet the learnt features cannot be well explained by current deep learning algorithms. To explain deep learning models, many researchers have used parameter analysis and visualization techniques. Unfortunately, the interpretable imaging indicators needed for clinical requirements are still some distance away.

In the field of medical pictures, research into the interpretable deep learning approach will therefore be a popular topic. Table1 lists the various classification techniques. Some of the available literature's limited ability to be implemented in clinical practise is partly a results are still used and frequently serve as the sole summary results of some studies, despite their limited clinical importance, including accuracy, AUC, and precision. Instead, because they are unaffected by the prevalence of the disease and researchers should continually measurements, easier to apply in practice. Table 1 compares the CNN's implementation to those of other recent deep learning-based works. Table 1 demonstrates that, with the exception of [35], whose accuracy was 99.1%, our work outperforms most other recent research, utilizing the practice set. The testing set's accuracy is weak, only reaching 64.4% as opposed to the test accuracy of 97.91% found in our work. As compared to earlier studies, the suggested study also shows good results in terms of sensitivity, specificity, accuracy, and AUC.

Contributions

The goal of this work was to develop and validate a Deep Learning model for lung cancer detection on lung radiographs approach, as well as to assess the features of this DL-based model.

The contributions of this article are enumerated in the following sentences:

- By this study, a deep learning-based model for lung cancer segmentation and detection on lung CT scan images was developed.
  .
- All of the nodules and masses were pathologically determined to be lung cancers, and two radiologists annotated these lesions at the pixel level, , our dataset is of high quality, Amrit et al[19].
- The detection or segmentation approaches were less informative than the segmentation method, which is helpful for lung cancer detection as well as follow-up and treatment effectiveness, Krishnaiah et al [20].

2.2 Dataset

The primary source of the data is The Lung Image Database Consortium (LIDC-IDRI), an open access collection that can be accessible at The Cancer Imaging Archive (TCIA) until May 2020. This collection is made available under the terms of the Creative Commons Attribution Non-Commercial 3.0 Unported (CC BY-NC) license. Participant's information from the LIDC-IDRI has previously been provided, but in essence, 1018 clinical chest CT images are included in the collection, which comes from four different institutions. M.Siddardha Kumar et al[21]. There were 2669 consensus nodules among the 7371 annotations that independent radiologists entered. According to existing diagnosis criteria, we removed participants with unreported or unidentified diagnoses and nodules smaller than 3 mm in diameter. 34, 35 As a result, 110 nodules in total 310 different subjects were suitable for analysis, Rani et al[22]. We used Kaggle dataset for this paper implementation and got better results compared to other models.

Lung cancer has been detected in the patients and lung nodules in the dataset. Binary masks for the nodules were collected in the form of an XML file.Figure 1 lists the numbers of patients and nodules that were omitted, along with the explanation, Dendi Gayathri Reddy et al [23].

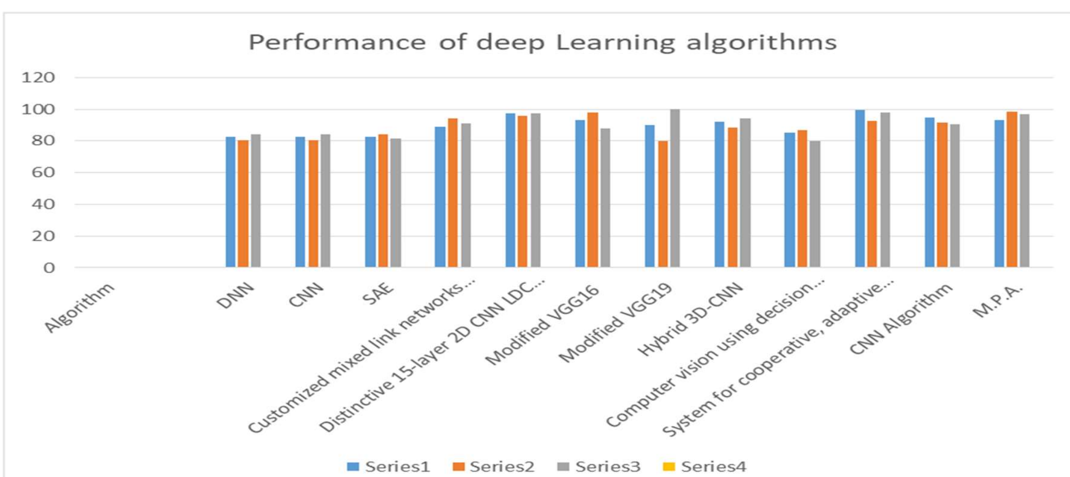


Figure: 1 Deep learning algorithms

Different Deep Learning Algorithms and their percentages of accuracy, sensitivity, and specificity are displayed in Fig. 1 above. We can contrast several algorithms as well. The following table 2 compares the CNN's application with other current deep learning-based works. It demonstrates that, with the exception of Abdul razak Yahya et al [58], whose accuracy was 99.1%, our work outperforms most other recent research. utilising the practise set, compared to the 97.91% test accuracy discovered in our research.

Table 2 Comparison of different Lung cancer Techniques

Published Work	**Dataset**	**Accuracy**	**sensitivity**	**Specificity**	**Precision**	**AUC**
CNN+SVM	LIDC	78.32%	72.80%	79.24%		83.80 %
CNN	LIDC	80.86%.	-	-	-	
DBNs	LIDC	84.30%	-	-	-	
SDAE	LIDC	81.33%				
Deep CNN	LUNA 16	Training:99.24%	80.72%	59.37%	55.46 % -	
		Testing:93.40%				
Multi-view multi-scale CNN	LIDC/IDRI	93.50%	-	91.34%		
Proposed CNN-SVM	CT-Scan Image	98.72%	98.80%	99.64%	98.97 %	98.89 %
		Training:97.73				
		Testing:98.92%				

It can be shown from Table 2 that by applying the suggested approach CNN-SVM and CT scan pictures, we were able to reach Accuracy97.8%.

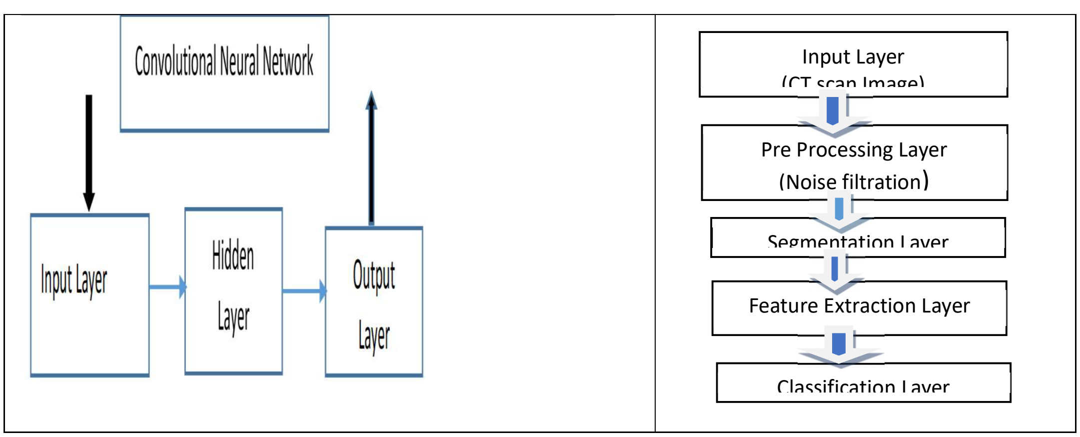


Figure: 2 Basic block diagram of CNN and Architecture of Proposed Lung Cancer Detection System Diagram

3. Proposed Approach

In this section, we go over how to extract region-based features and an ensemble classifier to accurately identify lung cancer from CT scan pictures. The flowchart for the procedure is shown in Figure 2.

3.1.Input Layer

To build a visual depiction of some tissues and organs, the initial phase of pre-processing is the collecting of CT scan images of the human lung. An image is a DE-laminate artifact of visual perception that appears to some people, typically as a person or a tangible thing, Palani D et al[24].

3.2 Pre-Processing

The cumulative distribution function of the noise, which is known as Gaussian noise, is equivalent to the pixel standard of images. It must forecast the average difference in pixel value. By utilizing the entropy capacity, Gaussian noise is presented, Alacrity et al [25]. Auto encoders will operate using an unsupervised learning methodology. The encoder's goal is to combine the input and send it to the system. The image will be recreated by the decoder using the code, Alacrity et al [26]. The model occasionally fills the supplied data, to make a difference in this issue and discover a condensed form of the input data as shown in fig. 2. When Gaussian noise is present, a Gaussian noise assessed auto encoder can guarantee accurate and dependable lung nodule detection.

3.3 Segmentation layer

Threshold is the simplest method of image segmentation. This is based on gray scale image into a binary image. Threshold values will be between 0 and 1, the images are segmented based on these values only. The original property of an image is based a portion of pixels on its natural intensity, as we e used here Ostu's threshold Method, Ozge Gunaydin [27]. A subfield of computation and digital image processing called "image segmentation" tries to combine similar neighborhoods or segments of an image under the relevant class labels, WHO-2018 [28].

3.4 Feature Extraction Layer

This is the vital role in image processing techniques for image feature extraction stage, Which utilizing algorithms and methods for identifying forms or separating sections of an image is accomplished on the lung region, the feature will be obtained from the better diagnosis, and the overall characteristics of the improved segmented image created utilizing the banalization approach, W. Alakwaa et al[29].

3.5 Classification Layer

We create a classifier using the five machine learning models by using an ensemble method called max-voting. The favoured ensemble technique may be a typical example of the multi-professional approach, which enables the lateral integration of many classifiers, S. Bhatia et al[30]. The contrast in variance part of the prediction errors produced by the contribution models is frequently the mechanism for improved performance with ensembles, T.L. Chaunzw et al [31]. A classification method called Random Forest produces a large number of decision trees from a seed that are permitted to split up in any way. The below figure 3 shows that it contains normal lung image and preprocessing of lung cancer detection, image, which is initially the CT image is given as Input data and the image will be forwarded under various types of layers like fully connected, Xie H et al [32].

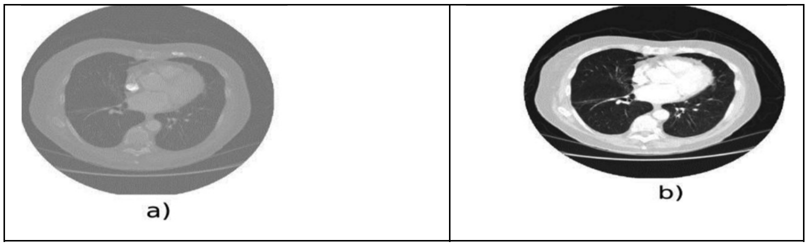


Figure: 3 a) CT scan original images, b) Pre-processed image

4. CNN Architecture/ working model

We used mean squared error in our model to determine the loss between the input of the encoder and the output of the decoder. The model was trained using the original CT scan picture, and the auto encoder was utilized to reduce the quality disparity between the decoder's output and D. Kumar, et al[33], the original noiseless images. At present the Neural network with its three hidden layers are provided for a lot of research in medical domain and in academia for this task, but also used in actual life also, detection, J. Kuruvilla et al [34], object detection models are broadly divided in to two standards, namely traditional object detection models based on image processing and object detection algorithms based on Convolution Neural Networks. In the Neural network model, first layer consists of fully connected layer the CT scan Images are split in to 80:20 for training and validation objective, P. Monkam et al [35].

4.1 Max Pooling

Equations (1), (2), and (3) were used to determine the performance combining the recall, accuracy, and precision measures along with the confusion matrix plot (3). The output metrics TP, FP, FN, and TN represent
values for the models' training and validation images that are true positive, false positive, false negative, and true negative developed by D. Moitra et al and D. Moitra et al, respectively, as shown in tables 1 and 2[36–37]. After being stored in the required file format, the trained model weights were fed into the model architecture and utilised to make predictions about the future.

$$Accuray = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (1)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

	Lung Parenchyma segmented	Lung Parenchyma not segmented
Lung Parenchyma Present	True Positive	True Negative
Lung Parenchyma not Present	False Positive	False Negative

Segmentation Algorithm	Accuracy	Sensitivity	Specificity
Region Based	76.50	65.40	87.62
Cluster Based	89.35	88.25	91.64
Watershed Based	96.54	97.45	88.84

Table: 3 Measure of Objective Evaluation Parameter, and Table: 4 Comparative analysis of the objective evaluation of the segmentation Techniques

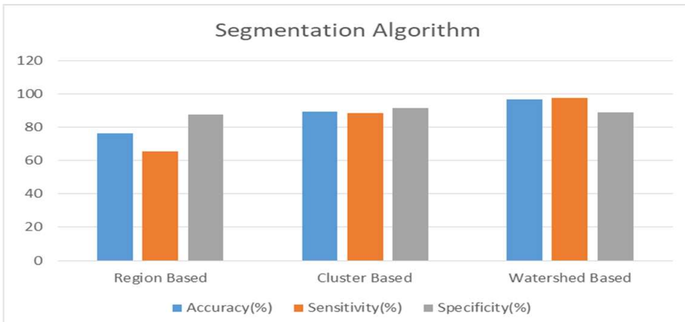


Figure 4: Segmentation algorithm

4.2 Fully connected layer

This is used to image recognition, image size with 28*28*3. CNNs are a sort of feed-forward artificial neural network where the organization and neural connection patterns are modeled after the human visual vertex. Only edges performing in a specific orientation will cause this visual vertex's case-sensitive brain cells to respond. When exposed to vertical, horizontal, and diagonal edges, some neurons started to fire. The four layers of CNN are Pooling, ReLU layer, Convolution layer, and fully connected Layer, Different types segmentation algorithms are shown in fig. 4, G. Silva el al and Chen et al et.al[38-39]

4.3 Convolution Layer

Convolution neural network compares the image pixel by pixel, the pixels that it looks for features in the image. By using CNN model we can compare the features of the Computed Tomography (CT) images, A. Teramoto et al[40]. We can mark the features in the image multiply each image pixels by the corresponding feature pixel. We create feature map and put the values of the filter at the place, using the same features move to another location and perform the filtering, X. Wang et al [41].

4.4 ReLU layer

Every negative value from the filtered photos can be eliminated in this layer and replaced with zeros. When the input value rises beyond a given threshold, it also becomes active as a node. When the input value is below zero, the output is also zero, but when the input value rises above a particular threshold, it develops a relationship with the dependent variable, Z. Zainudin et al [42].

4.5 Pooling Layer

We reduce the size of the image in this layer, and it can select a stride of a window size, with the help of new window slide on the filtered images; from each window we can select maximum value from the window, Zhang et al and P. Gupta et al[43-44].

4.6 Fully Connected Layer

In this last layer, actual classification takes place. Here, we may take a look at our filtered and condensed photographs and group them into a list. The lung cancer detection process is carried out by designing a GUI by using CT images and various filters, Initially it can be modified by Input image to gray scale image, in this GUI we also changing that gray scale image to threshold image which are shown in fig. 5,S. Jena et al and S.M. Ashhar [45-46].

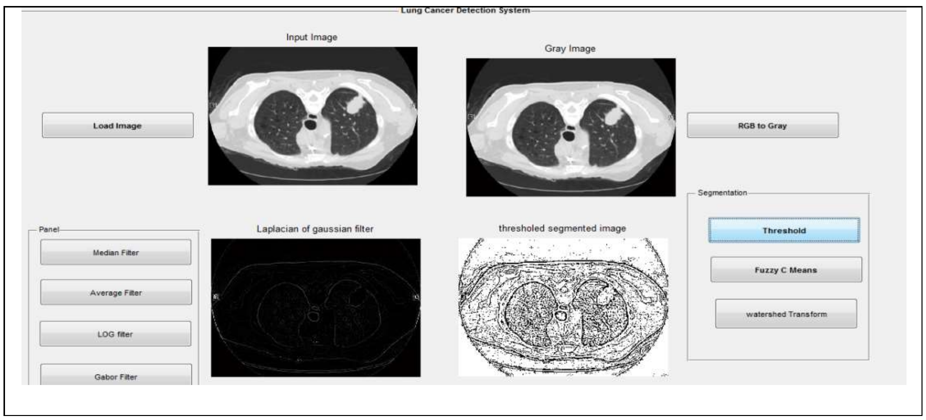


Figure: 5 Pre-processing of images using various filters

With a large dataset, the Stochastic gradient descent (SGD) optimization algorithm can successfully address real-world deep learning problems. This paper provides lung CT images into adenocarcinoma, large cell carcinoma, normal tissue, or squamous cell carcinoma using a hybrid CNN-SVM approach, according to A. M. Schilham et al and G. oppini [47–49]. The approach was utilised with the lung image dataset, which is a common and open collection of CT images. The objective was to boost accuracy because that is what any computer-aided detection system strives for. Tan et al. and J. Kuruvilla et al.[50–55] tested the method on a total of 5103 images, and a classification accuracy rate of 97.91% was achieved.

It demonstrates the value and potential uses of the suggested hybrid CNN-SVM. Two stages are involved. In the first step, a CNN made up of several volumetric convolution, rectified linear units (ReLU), and max pooling layers is used to extract pertinent volumetric features from the input data. The classifier comes in at step two. After a number of FC and threshold layers, which carry out the neural network's high-level reasoning, comes a SoftMax layer. To retain as many of the original values of the DICOM images as is practical, no scaling was applied to the CT scans in the collection, produced by G.S. Tran et al. and Abdulrazak et al. [56-58]. The randomsub-volumes obtained during training from the training set's CT scans are normalized using estimation of the normal distribution is presented by Q. Z. Song, et al [59-60].

5. Results and Discussion

Model Assessment

Now that our model is prepared, let's use various metrics to assess how well it performed on the validation data. To achieve this, we will first use this model to predict the class for the validation data, and then we will compare the results with the actual labels. The following output screen shows the accuracy is 97.9%, as shown in fig. 6.

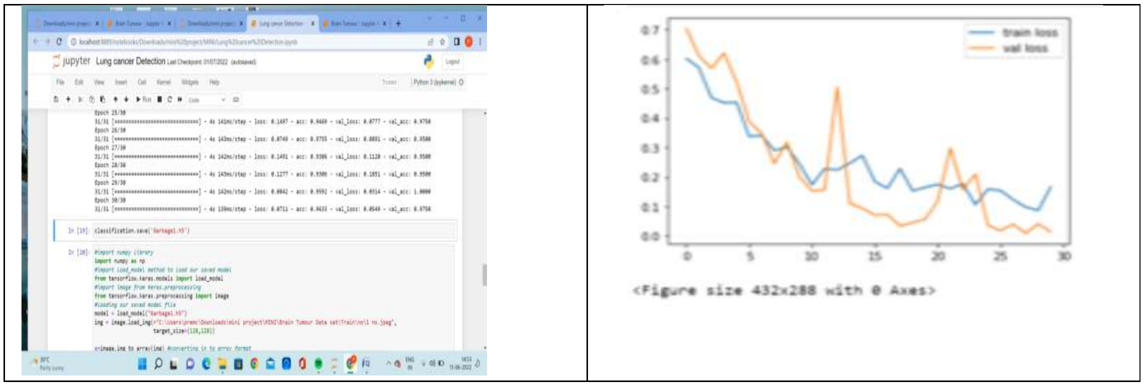


Figure: 6 a) output screen of Accuracy b) Output screen of Training and Testing

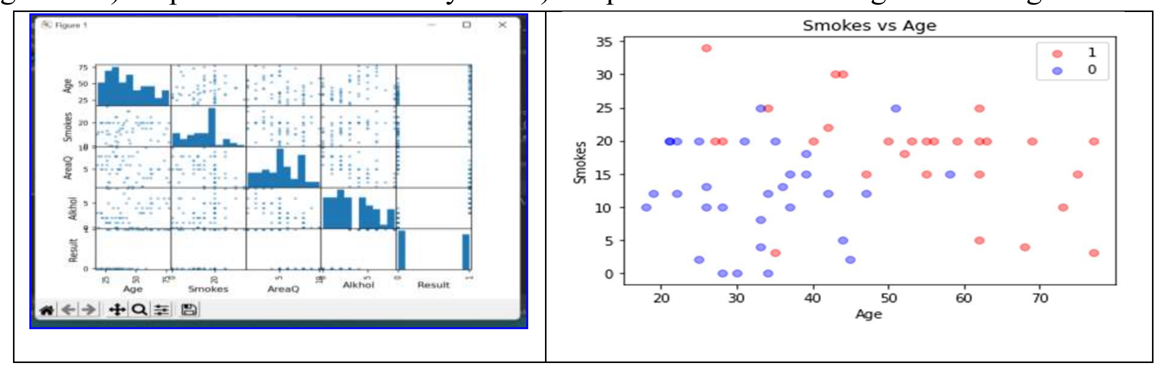
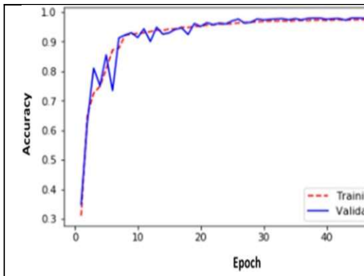
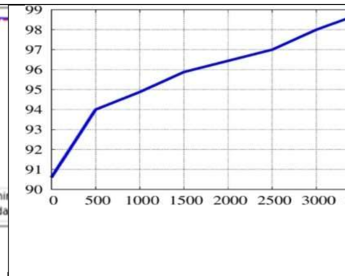


Figure: 7 a) output screen of Accuracy b) Lung cancer detection

The crucial element that causes a model to converge is the presence of a differentiable transformation, which is necessary in order to use gradient descent to learn the model's parameters. The ideal model parameters are initially unknown. Thus, theoretically, you can determine the optimal parameter values by minimizing the cost function in Eq. (1). This can be achieved by feeding a training dataset into the model and iteratively changing the parameters to minimize the cost function. For the first 2000 iterations, as shown in Fig. 7, we experiment with the various optimizers (such as RMSprop, Adam, and SGD) to minimize the negative log-likelihood loss in Eq. (1). The figure displays how well the SGD optimizer works better than Adam and RMSprop. The impact of the training epochs' (iterations') number on the generator training accuracy is depicted in Fig 9. The image also demonstrates that after 2000 iterations, the model's accuracy reaches saturation. Figure 10 also shows how selecting the learning rate affects the generator's accuracy. At 0.0001, the generator model has the highest accuracy. To optimize the loss in relation to CNN network parameters φ and θ, we use stochastic gradient descent.

$$L(\varphi; \theta) = L1 + L2 \qquad \text{Eq---1}$$





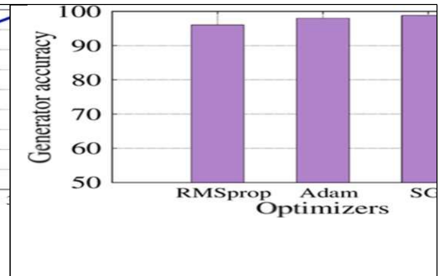


Figure 8.a) Proposed framework accuracy	b) Number of generated samples	c) Different optimizers of proposed model

The output screens of proposed lung cancer detection framework accuracy in fig. 8a, In fig 8b number of generated samples are shown, and in fig. 8c different optimizers of proposed lung cancer detection. A new thorough assessment and meta-analysis evaluating the performance of existing deep learning methods for detecting lung cancer was conducted. Deep learning techniques had a combined sensitivity and specificity of 93% for identifying lung cancer respectively 68%. The outcomes revealed that AI is essential for medical imaging, but there are still a lot of open research questions.

Train test split

To guarantee model robustness and unbiased performance evaluation, train-test split was performed on the dataset. Tiles were divided between training and testing sets at a ratio of 8:2 for each tile size. To guarantee that each class was distributed equally over the training and testing sets, a stratified train-test split was used. Each training set was further divided in an 8:2 ratio into training data and validation data. On the basis of training data, models were developed, and validation data were utilized to track the development of the models. In order to handle class imbalance and prevent the CNN model's inherent bias towards the class with the greatest population, an oversampling method was used on both the training and validation datasets.

6. Conclusion

In the contemporary years with the active development of machine learning automation, a comprehensive object detection technology has advanced rapidly and made improvement this analysis compared up to the minute and utmost leading edge CNN is best situated for object detection method. On the other hand there is a speechless and tremendous inconsistency between the productivity and rapid of the disclosure model and the civilized fulfillment. Object detection, it effects an unreasonable to analyze the thousands of appearances that are transfer to the web every day. Researchers must think that larger data sets and more evenly distributed data can result in better results. Out of all other object unmasking Convolution Neural Networks considered is the best inclusive attainment. The objective of this disclosure technique is to resolve where the objects are positioned in a given computed Tomography image. The model was able to generate accuracy for both training and validation was 96.11% and 97.98%, respectively. Future work can be improved by routinely increasing the model's precision by training it on larger and more thorough datasets.

Also, combining several machine learning models allows for comparison.

Acknoledgements:

I thank **Dr. M. Venkateshwaar Rao,** Professor in IT Dept, GITAM UNIVERSITY-Visakhapatanam,

Dr.T.V. Rajinikanth, Dean & Professor in CSE Dept SNIST-Hyderabad, **Dr. M. Madhavee Latha**, **Principal of MRECW, Mrs.A. RadhaRani**, **Professor** in CSE Dept, MRECW, **Dr. B. Subba Reddy**, **Professor & HOD** Dept of IT, MRECW, and **Dr. CVPR Prasad, Professor & HOD** in CSE Dept, MRECW for their knowledge and assistance in all areas of our study as well as for helping to write this paper.

Conflict of Interest: The authors declare no conflicts of interests.

Author Contribution Statement:

The authors confirm the contribution to the paper as follows: study and Design, data collection, analysis and interpretation of results and final version of the manuscript was approved by all authors, M. Premchander, Dr. M. Venkateshwara Rao, Dr. T.V Rajinikanth, Dr. B. Subbareddy after they had evaluated the findings.

REFERENCES

[1] Zehra Karhan and Taner Tunç2, ”Lung Cancer Detection and Classification with Classification Algorithms” *IOSR Journal of Computer Engineering* (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 22788727, Volume 18, Issue 6, Ver. III (Nov.-Dec. 2016), PP 71-77.

[2] Ada and RajneetKaur, ” A Study of Detection of Lung Cancer Using Data Mining Classification Techniques ” *International Journal of Advanced Research in Computer Science and Software Engineering*”, Volume 3, Issue 3, March 2013

[3] D. Vinitha, Dr.Deepa Gupta, and Khare, S., “Exploration of Machine Learning Techniques for Cardiovascular Disease”, *Applied Medical Informatics*, vol. 36, pp. 23–32, 2015.

[4] Chip M and etal. "Prediction of lung cancer patient survival via supervised machine learning classification techniques. *"International journal of medical informatics* 108 (2017):1-8.

[5]] K. V. Bawane and A. V. Shinde“Diagnosis Support System for Lung Cancer Detection Using Artificial Intelligence”-*International Journal of Innovative Research in Computer and Communication Engineering*,Vol. 6, Issue 1, January 2018

[6] Sukhjinder Kaur “ComparativeStudy Review on Lung Cancer Detection Using Neural Network and Clustering Algorithm*”, International Journal of Advanced Research in Electronics and Communication Engineering* (IJARECE) Volume 4, Issue 2, February 2015

[7] Sathyan H and Panicker,J.V„"Lung Nodule Classification Using Deep ConvNets on CT Images",9th *International Conference on Computing, Communication and Networking Technologies*, ICCCNT 2018

[8]] Himali Dalvi ."Lung Cancer Detection Using Image Processing and Machine Learning Health Care "IEEE International Conference on Current Trends toward Converging Technologies,Coimbatore,India,IEEE2018

[9] Jin H, Li Z, Tong R and Lin L. A deep 3D residual CNN for false-positive reduction in pulmonary nodule detection. *Medical Physics* 2018.

[10] Radhika PR and Rakhi.A.S.Nair."A Comparative Study ofLung Cancer Detection using Machine Learning Algorithms "*International Conference on Electrical ,Computer andCommunicationTechnologies* (ICECCT)IEEE 2019.

[11] D. Moitra and R. Mandal, "Classification of Non-Small Cell Lung Cancer using One-Dimensional Convolutional Neural Network*", Expert Systems with Applications*, vol. 159.

[12] Dr. T. V. Rajini Kanth and Tarun Rao 2014, "A Hybrid Random Forest based support vector machine classification supplemented by boosting," *International Research Journal Publisher, Global Journals Inc*. (USA), volume 14, issue 1, pp. 43-53.

[13] A. Abdullah and S. M. Shaharum, "Lung cancer cell classification method using artificial neural network", *Information Engineering Letters*, vol. 2, no. 1, pp. 49-59, 2012.

[14] Swati Mukherjee and Prof. S. Bohra. "Lung Cancer Disease Diagnosis Using Machine Learning Approach, *"Third International Conference on Intelligent Sustainable Systems*, ICISS2020.

[15] Syed Saba Raoof, MA.Jabbar and Syed Aley Fathima."Lung Cancer Prediction using Machine Learning: A Comprehensive Approach" *Second International Conference on Innovative Mechanisms for Industry Applications* "ICIMIA 2020.

[16] KameliaRoy,SheliSinha Chaudhury, Madhurima Burman, AhanaGanguly, ChandrimaDutta, SayaniBanik and RaynaBanik. "AComparative Study of Lung Cancer detection using supervised neural network *"International Conference on Opto-ElectronicsandAppliedOptics*, Optronix2019.

[17] Sumathipala and Yohan, et al. "Machine learning to predict lung nodule biopsy method using CT image features: A pilot study."*ComputerizedMedicalImagingandGraphics*71,2019:1-8.

[18] Zhang and Junjie, et al. "Pulmonary nodule detection in medical images: a survey." *Biomedical Signal Processing and Contro*l43, 2018: PP138-147.

[19] Amrit Sreekumar, Karthika Rajan Nair, Sneha Sudheer ,Ganesh Nayar H, and JyothishaJ Nair.Malignant Lung Nodule Detection using DeepLearning "*International Conference on Communication and Signal Processing,*July28-30,2020,India.

[20] Krishnaiah, V.,G.Narsimha, and Dr N.Subhash Chandra."Diagnosis of lung cancer prediction system using data mining classification techniques." *International Journal of Computer Science and Information Technologies* 4.1(2013): 39-45.

[21] M.Siddardha Kumar and Prof.Dr.K.Venkata Rao. "Prediction Of Lung Cancer Using Machine LearningTechnique:A Survey "*International Conference on Computer Communication and Informatics*(ICCCI) 2021,Jan.27–29,2021,Coimbatore,India.

[22] Sanjukta Rani Jena, Dr.Thomas George and Dr.Narain Ponraj."Texture Analysis Based Feature Extraction and Classification of Lung Cancer" *International Conference on Electrical, Computer and Communication Technologies* (ICECCT) IEEE 2019.

[23] Dendi Gayathri Reddy,EmmidiNaga Hemanth Kumar, Desireddy Lohith Sai Charan Reddy and MonikaP "Integrated Machine Learning Model for Prediction ofLung Cancer Stages from Textual data using Ensemble Method".*1st International Conference on Advances in Information Technology* IEEE 2019.

[24] PalaniD and K.Venkatalakshmi."AnIoT-based predictive modeling for predicting lung cancer using fuzzy cluster-based segmentation and classification."*Journal of medical systems*43.2 2019.

[25] Alacrity and Shamballa 2020,"*TheIQ-OTHNCCD lung cancer dataset*" ,Mendeley Data, V1, DOI: 10. 1762

[26] Nikita Banerjee and Subha laxmi Das."PredictionLungCancer–InMachineLearningPerspective"*International Conference on Computer Science, Engineering and Applications*(ICCSEA) 2020.

[27] Ozge Gunaydin, Melike Gunay and Oznur Sengel. "Comparison of Lung Cancer Detection Algorithms "*Scientific Meeting on Electrical-Electronics & Biomedical Engineering, Computer Science* EBBT, IEEE 2019.

[28] World Health Organization" *in Cancer Fact sheet, World Health Organization*, November 2018.

[29] W. Alakwaa, M. Nassef and A. Badr, "Lung Cancer Detection and Classification with 3D Convolution Neural Network (3D-CNN)", *International Journal of Advanced Computer Science and Applications*, vol. 8, 2017.

[30] S. Bhatia, Y. Sinha and L. Goel, "Lung Cancer Detection: A Deep Learning Approach" in *Advances in Intelligent Systems and Computing,* Singapore: Springer, vol. 817, 2019.

[31] T.L. Chaunzw et al., "Using deep-learning radiomics to predict lung cancer histology", *Journal of Clinical Oncology*, vol. 36, pp. 8545-8545, 2018.

[32] Himali Dalvi ."Lung Cancer Detection Using Image Processing and Machine Learning Health Care "*IEEE International Conference on current Trends toward Converging Technologies,*Coimbatore,India, IEEE 2018.

[33] D. Kumar, A. Wong and D. A. Clausi, "Lung Nodule Classification Using Deep Features in CT Images", 2015 12th *Conference on Computer and Robot Vision*, pp. 133-138, 2015.

[34] J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for CT images", *Computer methods and programs in biomedicine,* vol. 113, 2013.

[35] P. Monkam, S. Qi and M. Xu et al., "CNN models discriminating between pulmonary micro-nodules and non-nodules from CT images", *BioMed Eng OnLine* 17, 2018.

[36] D. Moitra and R. Mandal, "Classification of Non-Small Cell Lung Cancer using One-Dimensional Convolutional Neural Network", *Expert Systems with Applications*, vol. 159.

[37] A. M. Rossetto and W. Zhou, "Deep Learning for Categorization of Lung Cancer CT Images", 2017 IEEE/ ACM *International Conference on Connected Health: Applications Systems and Engineering Technologies* (CHASE), pp. 272-273, 2017.

[38] G. Silva, O.S. Neto, A. Silva, A. Paiva and M. Gattass, "Lung nodules diagnosis based

on evolutionary convolutional neural network", *Multimedia Tools and Applications*, vol. 76, pp. 19039-19055, 2017.

[39] Chen and K. Suzuki, "Computerized Detection of Lung Nodules by Means of Virtual Dual-Energy Radiography", *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 2, pp. 369-378, 2013.

[40] A. Teramoto, T. Tsukamoto, Y. Kiriyama and H. Fujita, "Automated Classification of Lung Cancer Types from Cytological Images Using Deep Convolutional Neural Networks", *BioMed research international*, 2017.

[41] X. Wang et al., "Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis", IEEE *transactions on cybernetics*, vol. 50, no. 9, pp. 3950-3962, 2020.

[42] Z. Zainudin., S. M. Shamsuddin and S. Hasan, "IIIOP Conference Series: Materials Science and Engineering", *Joint Conference on Green Engineering Technology & Applied Computing* 2019, vol. 551, 2019.

[43] C. Zhang et al., "Toward an Expert Level of Lung Cancer Detection and Classification Using a Deep Convolution Neural Network", *The oncologist*, vol. 24, no. 9, pp. 1159-1165, 2019.

[44] P. Gupta, S. Garg and A. Garg, "Convolution Neural Network for Lung Cancer", *International conference on Recent Trends in Artificial Intelligence IOT Smart Cities & Applications* (ICAISC-2020).

[45] S. Jena and T. George, "Morphological feature extraction and KNG-CNN classification of CT images for early lung cancer detection", *International Journal of Imaging Systems and Technology*, vol. 30, 2020.

[46] S.M. Ashhar, S.S. Mokri and A.A. Rahni et al., "Comparison of deep learning convolutional neural network (CNN) architectures for CT lung cancer classification", *International Journal of Advanced Technology and Engineering Exploration*, vol. 8, no. 74, pp. 126-134, 2021.

[47] A. M. Schilham, B. Ginneken and M. Loog, "A Computer-aided Diagnosis System for Detection of Lung Nodules in Chest Radiographs with an Evaluation on a Public Database", *Medical Image Analysis*, vol. 10, no. 2, pp. 247-258, 2006.

[48] N. Camarlinghi, I. Gori, A. Retico, R. Bellotti, P. Bosco and P. Cerello, et al., "Combinat ion of computer-aided detection algorithms for automatic lung nodule identification", *International journal of computer assisted radiology and surgery,* vol. 7, no. 3, pp. 455-464, 2012.

[49] G. oppini, S. Diciotti, M. Falchini, N. Villari and G. Valli, "Neural Networks for Computer-Aided Diagnosis: Detection of Lung Nodules in Fourth International Conference on Electronics Communication and Aerospace Technology (ICECA-2020)

[50] Tan M, Deklerck R, Jansen B, Bister M and Cornelis J. A novel computer-aided lung nodule detection system for CT images. *Med Phys* 2011.

.[51] S. Chen and K. Suzuki, "Computerized Detection of Lung Nodules by Means of Virtual Dual-Energy Radiography", *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 2, pp. 369-378, 2013.

[52] J. Wei, Y. Hagihara, A. Shimizu and H. Kobatake, "Optimal Image Feature Set for Detecting Lung Nodules on Chest X-Ray Images", *Computer Assisted Radiology and Surgery*, pp. 706-711, 2002.

[53] C. Hardie, S. K. Rogers, T. Wilson and A. Rogers, "Performance Analysis of a new Computer Aided Detection System for Identifying Lung Nodules on Chest Radiographs", *Medical Image Analysis*, vol. 12, no. 3, pp. 240-258, 2008.

[54] A. A. Abdullah and S. M. Shaharum, "Lung cancer cell classification method using artificial neural network", *Information Engineering Letters*, vol. 2, no. 1, pp. 49-59, 2012.

[55] J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for ct images", *Computer methods and programs in biomedicine*, vol. 113, no. 1, pp. 202-209, 2014.

[56] M. Masud, N. Sikder, A. Al Nahid, A. K. Bairagi, and M. A. Alzain, A Machine Learning Approach to Diagnosing Lung and Colon Cancer using DeepLearning-Based Classification Framework,Sensors(Switzerland),21 2021

[57] Choi WJ and Choi TS. Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor. *Comput Methods Programs Biomed* 2014.

[58] Abdulrazak Yahya Saleh a,1,* , Chee Ka Chin b,2 , Vanessa Penshie a,3 , Hamada Rasheed Hassan Al-Absi c,Lung cancer medical images classification using hybrid CNN-SVM, International Journal of Advances in Intelligent Informatics ISSN 2442-6571 Vol. 7, No. 2, July 2021, pp. 151-162

[59] S. Zheng, J. Guo, X. Cui, R. N. J. Veldhuis, M. Oudkerk, and P. M. A. Van Ooijen., Automatic Pulmonary Nodule Detection in C.T. Scans using Convolutional Neural Networks Based on Maximum Intensity Projection, Ieee *Trans. Med. Imaging* .39, 2020.

[60] Q. Z. Song, L. Zhao, X. K. Luo, and X. C. Dou, Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images, J. *Healthc.Eng* 2017.

AUTHORS BIOGRAPHY

M. Prem Chander is a Ph. D. scholar at GITAM UNIVERSITY in Visakhapatnam, Andhra Pradesh, India, in computer science and engineering. He is currently working as Assistant Professor in the dept of CSE at MALLAREDDY ENGINEERING COLLEGE FOR WOMEN(AUTONOMOUS), HYDERABAD. As well as obtained M. Tech.(CSE) in Computer Science an.d Engineering from Jawaharlal Nehru Technological University(JNTU), Anantapurum, AP,INDIA in 2011. Prior to his professional career, he obtained In 2005, Bachelor of Technology (CSIT)degree are given by Sri Datta Institute of Engineering and Science, Hyderabad, Telangana, India, which is affiliated with JNTUH. Image processing and machine learning methods for medical applications are among his current research interests. He has been a teacher for 18 years in total. He oversaw

	about 30 B.Tech. level projects and 10 M.Tech. level projects. Additionally, he served as the organiser of two national conferences at Maheshwara Engineering College in Hyderabad, CONVIVALITY2K11 and CONVIVALITY2K12. He is an IEEE Member
	**Dr. M. Venkateshwara Rao** obtained his Ph.D. In the Department of Information Technology at the Gitam Institute of Technology, GITAM UNIVERSITY, in Visakhapatnam, Andhra Pradesh, India, he holds the position of Associate Professor. He holds M.Sc. in Electronics, and M.Phil. in Optoelectronics, and having a M.Tech (CST). He has 28 years of teaching and industrial experience. His works have been featured in a number of professional journals and conferences, international journals, 15 national conference 20
	**Dr. T. V. RajiniKanth** has obtained his **Ph.D. degree** in C.S.E. branch from Osmania University, Hyderabad in July, 2008 and M. Tech.(C.S.E.) degree from Osmania University, Hyderabad in January, 2001. His specialization area in research is “Spatial Data mining”. He obtained his PGDCS degree from HCU, Hyderabad in 1996. He received his M. Sc. (Applied mathematics) degree in the year 1989 from S.V. University, Thirupati as University Ranker. He is currently working as **Professor & Dean in CSE Dept., at SNIST, Hyderabad.** He worked as a Professor and Head, Department of IT, GRIET, Hyderabad from 2008 June to Oct 2013. Before that he worked as a Professor in CSE department, GRIET, Hyderabad since 2007 November. Prior to that, he worked as Asso. Prof. in VNRVJIET, Hyderabad where he joined in 1996. His total teaching experience is 23 years. Under his extreme guidance 17 Research scholars awarded Ph. D.
	**Dr. SUBBAREDDY BORRA** working as Professor and HOD, in the Department of IT, MRECW. He received B.Tech Degree from BAPATLA ENGINEERING COLLEGE in 2002. He received MASTER OF ENGINEERING (COMPUTER SCIENCE AND ENGINEERING) Degree from SATHYABAMA UNIVERSITY in 2008. He received MASTER OF TECHNOLOGY (NEURAL NETWORKS) Degree from JNTUK in 2011. He obtained Ph.D Degree in the area of Image Processing in the year 2021from JawaharlalNehru Technological University Hyderabad (JNTUH). He is Member of CSI, IETE, IE, ISCA, ISTE. He has

	a Good Academic and Research Experience of more than 18 years. He was previously associated SGIT Where He served as Professor of CSE. He is associated with MRECW from one year. His current area of research includes Machine Learning and Data Science. He has 20 papers that have been published in reputable international journals and has delivered five presentations at reputable national and international conferences. Attended Various Workshops and Faculty Development Programs.He has organized several Programs on Research Development and emerging areas in CSE&IT. He has delivered Guest Lectures at Various Institutes in AP. He conducted several Workshops as a Resource Persons. He also acted as Reviewer for reputed Journals

=