International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03443
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

AUTOMATED LUNG CANCER DETECTION USING NAS: A HIGH-PERFORMANCE DEEP LEARNING APPROACH

NETHRASHRUTHI R
Computer Science and Engineering
Jai ShriRam Engineering College
Tiruppur, India
nethraharithra@gmail.com

KIRUTHIGA T
Computer Science and Engineering
Jai ShriRam Engineering College
Tiruppur, India
stkiruthi@gmail.com

Dr. GOKILAVANI A
Computer Science and Engineering
Jai ShriRam Engineering College
Tiruppur, India
gokilavanics005@gmail.com

Abstract – Lung cancer remains one of the leading causes of mortality worldwide, necessitating early and accurate detection for effective treatment. This study explores the application of deep learning techniques, specifically Convolutional Neural Networks (CNNs) and Neural Architecture Search (NAS), for automated lung cancer detection from CT scan images. CNNs, while effective, often require manual architecture tuning, leading to suboptimal performance. NAS, on the other hand, optimizes network architecture automatically, resulting in improved accuracy. Experimental results demonstrate that CNN achieves an accuracy of 84.38%, whereas NAS significantly outperforms it with an accuracy of 96.35%. The superior performance of NAS is attributed to its ability to discover the most efficient network structure tailored to lung cancer detection. These findings highlight the potential of automated deep learning approaches in medical image analysis, contributing to more reliable and precise diagnostic tools.

Keywords: Lung Cancer Detection, Deep Learning, Neural Architecture Search (NAS), Convolutional Neural Networks (CNN).

1. INTRODUCTION

Lung cancer is still one of the most prevalent and deadly types of cancer worldwide at this point. It accounts for a significant percentage of deaths that are credited to cancer annually. Since lung cancer that is diagnosed in its early stages is more treatable than lung cancer that is diagnosed in advanced stages, early detection and proper diagnosis are very much integral in assisting in enhancing survival rates. In traditional diagnostic techniques, including chest X-rays and computed tomography (CT) scans, a lot of dependency is put upon the skills of radiologists, which may be time-consuming as well as subjective. Also, because of the fragile nature of lung nodules and tumour shape variability, human interpretation of medical images poses several challenges. The identification of symptoms is the first step to diagnose the lung cancer. The most typical signs of lung cancer are persistent chest discomfort and a persistent cough. Shortness of breath, a weak sensation, weight loss, a bloody cough, and exhaustion are among the additional symptoms that lung cancer patients experience most frequently.

The scientific community has not yet created a screening method that might detect lung cancer at an early stage in order to improve survival rates. Chest X-rays are a widespread screening technique, although they are not yet trustworthy enough. The creation of screening tools is urgently needed since several studies have found that early-stage tumours are easily curable. Low-dose computed tomography (LDCT) screening is advised once a year for smokers and for those who have given up smoking within the past 15 years. As per the American Society of Clinical Oncologists, people who smoke 30 years or more than 30 years and in the age group of 55-74 are at more risk of lung cancer. With advancements in Artificial Intelligence (AI) and deep learning, automated lung cancer detection systems have emerged as promising tools to assist radiologists in early diagnosis. CNNs, a class of DL models, have shown remarkable success in image analysis and medical imaging applications. CNN-based models can effectively extract features from medical scans, aiding in accurate classification of malignant and benign lung nodules. However, designing an optimal DL model for lung cancer detection remains a challenge, as conventional CNN architectures may not always provide the best performance for a specific dataset.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03443
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

To address this, NAS has been proposed as an automated technique to optimize deep learning architectures. NAS explores different network topologies and hyperparameter configurations to identify the most effective model for a given task. By leveraging NAS, we can improve detection accuracy while reducing the need for manual model design. This study focuses on comparing the performance of existing CNN architectures with a NAS-based approach for lung cancer detection. By evaluating different models and their ability to classify lung cancer from medical images, we aim to enhance diagnostic accuracy and contribute to the development of efficient computer-aided detection (CAD) systems in the healthcare sector. This figure 1 is a representation of lung cancer images, perhaps a CT scan that shows lung nodules that are characteristic of the growth of cancerous tissue. It may have the potential to reveal the differences between normal, benign, and malignant lung tissue.

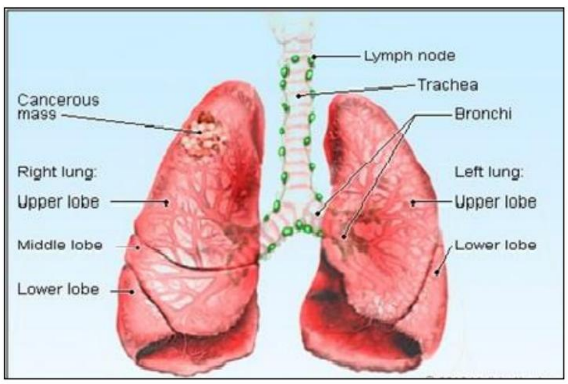

Figure 1 Lung Cancer

2. Related Works

Taher and Sammouda [2011] had the examination presents 2 classification strategies, to be specific-Hopfield Neural Network (HNN) and a Fuzzy c-means (FCM) grouping calculation. For dividing and to identify lung cancer in its beginning times sputum color images are used. The manual examination of the sputum test is tedious, inaccurate and requires intensive prepared people to maintain a strategic distance from analytic blunders. In this examination, 1000 sputum colour images are used to test both HNN and FCM. HNN shows preferred classification result over FCM and HNN is prevailing with regards to removing the nuclei and cytoplasm regions successfully.

SVM is a machine learning technique that builds a framework using data that is already known; it analyses the data and finds patterns, according to M. Gomathi [2012]. Due of its simplicity, the SVM may be used to group the treatment information. In order to conduct the study, live lung pictures are collected. RoI is recovered from lung images using portioning, and these nodules are then used for categorization. For each characteristic, appropriate threshold values are selected, and classification criteria are defined. These criteria are then sent to the SVM classifier. In this study, the categorization of benign and malignant nodules is carried out using several SVM sections, and the manifestation of each kind is calculated.

M Ajina [2017] examined the different strategies that were utilized in removing features, choosing features and classify it. Also, it dissects the new strategies that are utilized at each phase to deliver more execution in classification. It is seen with exactness of 57.5%. At that point KNN classification of ILD is done and it delivers more precision (72.94%) contrasted with ANN classification. It is seen that Deep CNN classification delivers elite regarding precision (84.14%) contrasted with the 2 other classification techniques (ANN and KNN). At long last, characterization is performed utilizing new technique called as Hybrid kernel based SVM classification. The mix of 2 kernel capacities utilized will create more exactness in kernel multiplication. Consequently it delivers the exactness of about 90.52%.

Bhuvaneswari and Therese [2015] proposed beginning time lung malignancy identification. Genetic K-Nearest Neighbor (GKNN) technique is proposed for the recognition. Doctors can identify the nodules in the CT lung pictures that are present in the early stages of lung cancer thanks to this enhancement computation. The human translation of lung illness CT pictures is time-consuming and crucial, therefore to get over this problem, the Genetic Algorithm approach is combined with KNN calculation, which would quickly and successfully organize the malignancy images. With regard to the CT lung pictures, the implementation using the MATLAB image processing toolbox is complete, as are the groups of these images. Examined are the exhibition estimations for categorization rate and false positive rate. In a typical K-NN calculation, the separation between each test and its preparation is first

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03443
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

computed, and then the K-neighbors with the most noticeable separations are selected for categorization.

Lakshmanaprabu and Mohanty [2019] propose Lung cancer is one of the perilous infections that reason immense disease demise around the world. Early identification of lung cancer is the main conceivable approach to identify and diagnosis of nodules, so as to increase the survival rate. To locate a tumour and determine its level of malignancy within the body, a CT scan is used. The current study demonstrates a creative computerised discovering classification strategy for CT of the lungs employing Linear Discriminate Analysis and ODNN. LDA is used to reduce the dimensionality of characteristics extracted from a CT lung picture before classifying the lung as either risky or advantageous.

3. System Methodology

The overall structure of the NAS-based lung cancer detection system is illustrated in figure 2, which can be viewed here. Data preprocessing, feature extraction, NAS model training, and classification are some of the modules that are incorporated in this system since it is intended to detect lung cancer from CT images.

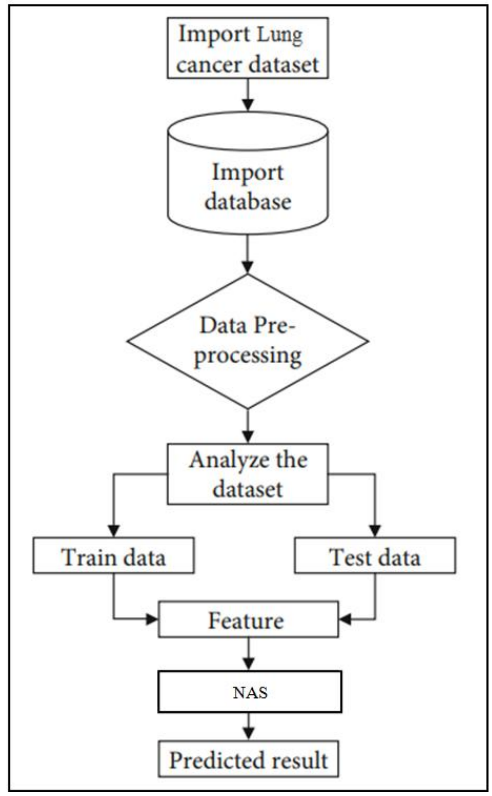

Figure 2 System Architecture

3.1 Data Collection

This research work has taken Lung Cancer CT Scan Dataset, which a valuable dataset for training and evaluating models that are dedicated to identifying lung cancer. It has a total of 1,190 CT scan images from 110 cases, all of which have been well curated for the identification and classification of lung cancer. The number of normal cases is 55, 15 benign cases, and 14 malignant cases are in this dataset. There are numerous CT scan slices per case, the number of images is between 80 and 200 per case. The dataset, collected from Iraqi hospitals, is now available for use by the general public for research purposes.

3.2 Preprocessing and Augmentation

In an effort to prepare the dataset for training deep learning models, preprocessing is a key process that should be incorporated. Since CT scans vary in resolution, intensity, and noise, preprocessing plays a vital role in standardizing the images and increasing the accuracy of the model. Initially, it's required to perform image resizing due to the fact that CT scan images occasionally arrive in a number of different resolutions. In order to keep things uniform and compatible with deep learning models, one has to resize all photographs to a constant size, e.g., 224×224 pixels. Then, image normalisation is applied in a way that normalises pixel values to the interval of [0, 1] or [-1, 1] for stabilising training as well as improving convergence. Also, greyscale CT scan pixel intensities vary from 0 to 255. Noise is often encountered in CT scans due to artefacts of scanning. Due to this, the Gaussian Filter was used in this research to suppress noise without sacrificing details, and the dataset was split into 3 sets: the Validation Set (10%), the Training Set (80%), and the Test Set (10%). DL models must have a diverse range of training examples in order to generalise well. Augmentation is a procedure that samples the dataset and then artificially grows it by making changes like Horizontal or vertical orientation changes, Rotation (±15 degrees), Zooming (photos that are resized), and Brightness Adjustment (this aids in enhancing the generalisation of the model).

3.3 Feature Extraction and Classification

NAS can automate the procedure of creating an optimal deep model for the intention of extracting informative information from CT scans of lung cancer. NAS determines the best feature extraction architecture automatically through a search algorithm, unlike typical CNNs, where the layers, filters, and hyperparameters have to be chosen manually. The technique of finding meaningful patterns in CT scan images that differentiate benign, malignant, and normal cases is known as feature extraction. To optimize feature extraction, NAS looks for optimal CNN layer structures, selects the best kernel sizes, pooling methods, and activation functions, and optimizes efficiency by reducing the number of layers and parameters that are redundant.

Composite Function

A composite function is built by the NAS-optimized model, which consists of many deep learning operations in every layer of the network. The following constitute this composite function, which are achieved through the use of convolutional layers, while batch normalisation is used to normalise activations and improve stability. In addition, the convolutional layers are used in a way that hierarchical lung nodule features are produced. In order to achieve non-linearity, the Rectified Linear Unit (ReLU) activation is used. All these processes are performed sequentially by every layer, which ensures that maximum effective learning of lung CT scan features takes place.

Pooling Layers

In order to preserve only the most necessary information, feature maps are minimized by pooling layers together. The NAS-based network is divided into many highly interconnected blocks in a bid to limit the complexity of the involved computations and to downsample feature maps effectively. Reshaping the dimensionality of the features with the help of convolution layers and performing pooling operations for spatial reduction are two among the tasks helped by transition layers, which are inserted between these blocks. Effective feature aggregation in this work is achieved with a 22-layer average pooling layer. An eleven-layer convolutional block to extract the vital lung nodule patterns, and a batch normalisation layer which is inserted between transitions to maintain numerical stability. Ratio of Growth. It is the rate of growth of the network that is denoted by the growth rate (k) in NAS.

Growth Rate

It achieves this by regulating the quantity of information that is propagated from the lower layers to the higher layers above it. An increased growth rate will increase the capacity of the model, but it may also lead to overfitting. The model's complexity can be decreased, though feature extraction can be impacted by a reduced growth rate. The NAS model achieves to deliver only informative features in forward passing due to its capability of establishing a tradeoff between the layer connection and the growth rate. This gives the better lung cancer classification performance.

Loss Function

In terms of overall model performance and generalisation, the loss function is a very critical aspect to be considered. A Categorical Cross-Entropy Loss training technique is used in an attempt to train the NAS-chosen classification model. Through penalising wrong predictions, this loss function can effectively deal with multi-class classification problems, which consequently leads to higher accuracy in lung cancer identification.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03443
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

4. Result and Discussion

The confusion matrix of the standard CNN model used for lung cancer classification is shown in figure 3. Precisely, it illustrates how accurately the CNN model classified instances as normal, benign, or malignant, and it can also indicate where the model incorrectly classified cases.

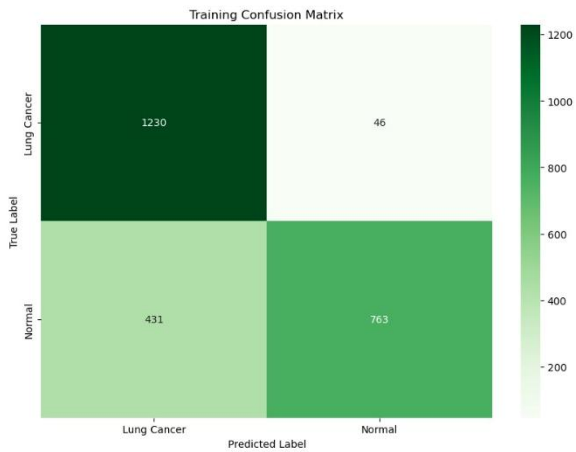

Figure 3 CNN Confusion Matrix

Similar to Figure 3, the confusion matrix presented in Figure 4 describes the classification results for the NAS-optimized model. For comparison, in order to find out whether NAS improves accuracy, sensitivity, and specificity for lung cancer detection or not, a comparison is done with CNN.

Figure 4 NAS Confusion Matrix

Figure 5 shows how the accuracy of the standard CNN model changes over the course of many epochs, considering both the training set and the validation set. A rise in accuracy that is either slow or irregular can be a sign of either overfitting or inefficiencies in feature extraction.

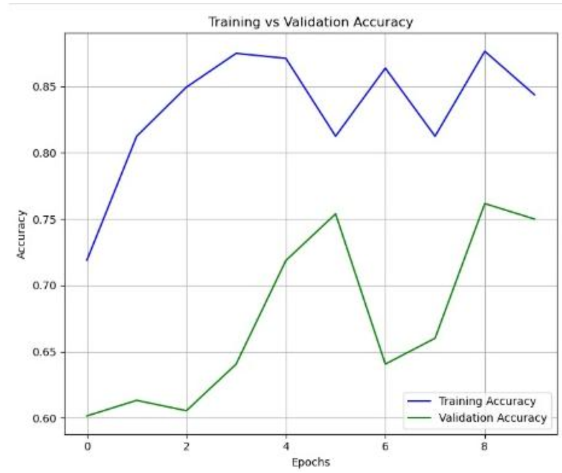

Figure 5 CNN Training & Validation Accuracy

The CNN model loss curves are illustrated in figure 6, where they are presented during both training and validation processes. There is a relationship between a declining loss trend and better learning, but large differences between training and validation losses can suggest that over fitting has taken place.

Figure 6 CNN Training & Validation Loss

Over the course of several epochs, the training and validation accuracy of the NAS-based model is depicted in figure 7, which serves as an illustration. In comparison to CNN, NAS demonstrates a stable and better level of

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03443
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

accuracy, which shows that it is able to select a more effective architecture for lung cancer diagnosis.

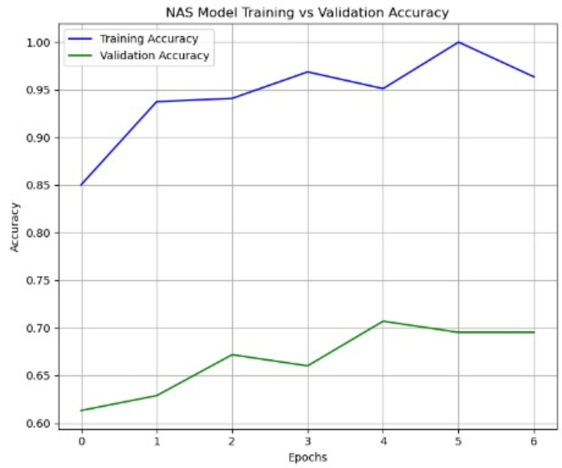

Figure 7 NAS Training & Validation Accuracy

The NAS model's training and validation loss is shown in figure 8, which can be accessed here. If the loss values were lower and steadier compared to CNN, it would mean that the NAS-optimized architecture would perform with greater convergence and generalization.

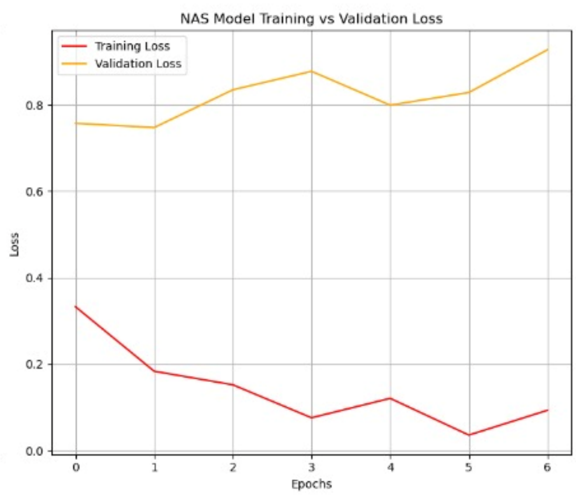

Figure 8 NAS Training & validation Loss

Figure 9 depicts the comparison between the performance parameters that were utilized by the deep learning models employed in the analysis and study. The findings identify that NAS is better than standard CNN designs with respect to identifying lung cancer.

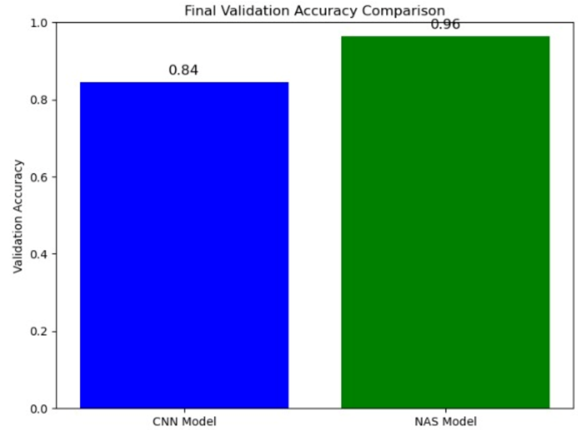

Figure 9 Comparative analysis using DL models

A CT scan classification result of the NAS model is presented in figure 10, showing if the input had been appropriately labeled as Normal, Benign, or Malignant. It also includes a sample classification result.

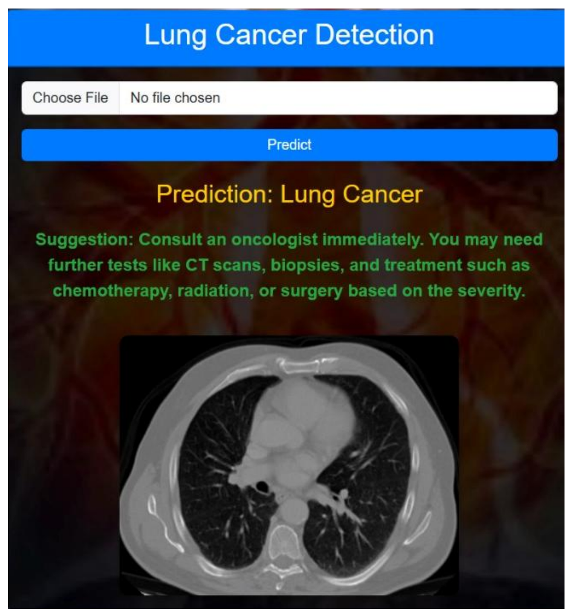

Figure 10 NAS Result Outcome 1

In a similar fashion to figure 10, the figure 11 represents another example of the NAS model's classification output, thus giving further evidence that NAS is effective at identifying lung cancer.

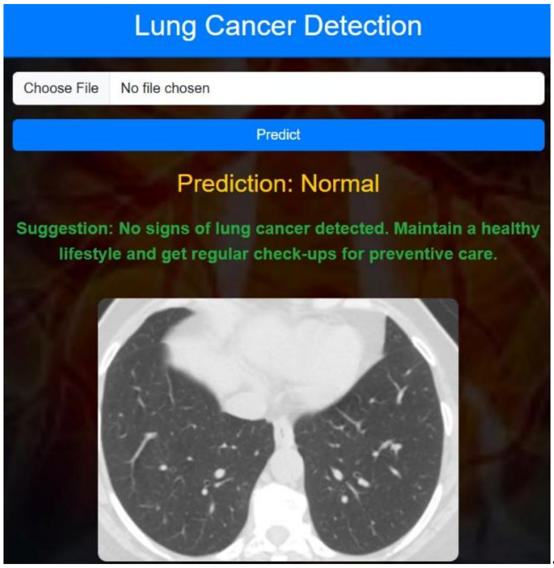

Figure 11 NAS Result Outcome 2

5. Conclusion

The prime motive of this work is to detect lung cancer using deep learning model. By using automated architecture selection, Neural Architecture Search (NAS) can outperform typical Convolutional Neural Networks (CNNs) when it comes to lung cancer detection. This brings about higher accuracy, better feature extraction, and better generalisation. NAS dynamically optimises filter sizes, connections, and layers, thus significantly reducing overfitting and computing complexity involved. This is in contrast to standard CNNs, which are dependent on structures that are created manually. Because of this, the classification performance of NAS is much improved, which makes it a solution that is both more efficient and scalable for medical imaging applications. By the use of NAS, models for lung cancer detection improve to be more reliable, responsive, and effective in the end, leading to the early diagnosis of the disease and improving outcomes for patients.

References

[1] Abbas M., Precise Feature Selection in Predictive Genetic Models Using Grey Wolf Optimization Algorithm, In 13th International Conference on Computer Engineering and Systems. IEEE, 2018, 63–68.

[2] Akram S., (2017), Intelligent Detection and Classification Methodology for Multidimensional Lung Tumor Cells, 1(2).

[3] Bakhbakhi Y., (2011), Phase Equilibria Prediction of Solid Solute in Supercritical Carbon Dioxide with and Without a Cosolvent: The Use of Artificial Neural Network, Expert Systems with Applications, 38(9), 11355–11362.

[4] Bhuvaneswari P. and Therese A. B., (2015), Detection of Cancer In Lung with K-NN Classification using Genetic Algorithm, Procedia Materials Science, 10, 433–440.

[5] Cong Xue W. J.Zhihuang Hu and Zhao Y., (2012), National Survey of the Medical Treatment Status for Non-Small Cell Lung Cancer (Nsclc) in China, Lung Cancer, 77(2), 371–375.

[6] Durai M. S. and Iyengar N., (2010), Effective Analysis and Diagnosis of Lung Cancer using Fuzzy Rules, International Journal of Engineering Science and Technology, 2(6), 2102–2108.

[7] Horsthemke W. and Varutbangkul, Predictive Data Mining for Lung Nodule Interpretation, In Seventh IEEE International Conference on Data Mining Workshops. IEEE, 2007, 157–162.

[8] Hyder S. A. and Sukanesh R., An Efficient Algorithm for Denoising MR And CT Images using Digital Curvelet Transform, In Software Tools and Algorithms for Biological Systems, volume 696, Springer, 2011, 471–480.

[9] Ignatious S. and Joseph, Computer Aided Lung Cancer Detection System, In Global Conference on Communication Technologies. IEEE, 2015, 555–558.

[10] Lakshmanaprabu S. and Mohanty, (2019), Optimal Deep Learning Model for Classification of Lung Cancer on CT Images, Future Generation Computer Systems, 92, 374–382.

[11] M Ajina L. M., (2017), Diagnosis of Interstitial Lung Disease by Pattern Classification, Procedia Computer Science, 115, 195–208.

[12] M.Gomathi D., (2012), An Effective Classification of Benign and Malignant Nodules using Support Vector Machine, Journal of Global Research in Computer Science, 3(7), 6–9.

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

[14] Rendon-Gonzalez E. and Ponomaryov, Automatic Lung Nodule Segmentation and Classification in CT Images based on SVM, In 9th International Kharkiv Symposium on Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves. IEEE, 2016, 1–4.

[15] Sang J. and Alam, Automated Detection and Classification for Early Stage Lung Cancer on CT Images Using Deep Learning, In Pattern Recognition and Tracking XXX, volume 10995. International Society for Optics and Photonics, 2019, 109950S.

[16] Sharma D. and Jindal G., Identifying Lung Cancer Using Image Processing Techniques, In International Conference on Computational Techniques and Artificial Intelligence, volume 17. 2011, 872–880.

[17] Taher F. and Sammouda R., Lung Cancer Detection by using Artificial Neural Network and Fuzzy Clustering Methods, In IEEE GCC Conference and Exhibition. IEEE, 2011, 295–298.

[18] Tekade R. and Rajeswari K., Lung Cancer Detection and Classification using Deep Learning, In 4th International Conference on Computing Communication Control and Automation. IEEE, 2018, 1–5.

[19] Kiruthiga T and Vijaya Kumar (2016), Review on Association Rule Mining Techniques with Machine Learning Approach for Cancer Treatment, Asian Journal of Research in Social Sciences and Humanities Volume 6, Issue 8, Pages 482-488.

[20] Zhang W. and Wang, (2017), Global Optimal Hybrid Geometric Active Contour for Automated Lung Segmentation on CT Images, Computers in Biology and Medicine, 91, 168–180.