ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

A NOVEL HYBRID DEEP LEARNING METHOD FOR EARLY DETECTION OF LUNG CANCER USING NEURAL NETWORKS

[1]S.BHARGAVI, [2]V.VAISHNAVI, [3]S.RISHVICK RAO, [4] S.KIRAN KUMAR,[5] Mr.N.SRIKANTH

[1,2,3,4] U.G.Scholor, Department of ECE, Sri Indu College Of Engineering & Technology, Ibrahimpatnam, Hyderabad.

[5]Assistant Professor, Department of ECE, Sri Indu College Of Engineering & Technology, Ibrahimpatnam, Hyderabad.

Abstract— **Using the lung dataset as a basis, this study compares the nodule identification accuracy of Decision Tree (DT) versus Novel Hybrid Learning Method algorithms. With 35 samples split equally between the two groups, a grand total of 70 samples were collected for this experiment. While Group 2 makes use of Decision Trees, Group 1 employs the Novel Hybrid Learning Method (NHLM). As part of the research, we used Google Colab to load the dataset and execute the HLM algorithm. An online statistical analysis tool is used to establish the sample size using data acquired from prior study. The pretest is prepared to go with a power of 80% and an alpha of 0.05. The results of the simulations showed that the HLM approach was more accurate (97.72%) than Decision Tree (85.659%). Testing at significance levels of 0.001 (P<\0.05), the methods show a diverse array of accuracy. If we compare the two datasets for the accuracy of predicting lung nodules, we find that HLM is superior than Decision Tree.**

The current problem is best described by terms like novelty of approach, decision tree, deep learning, computed tomography (CT) pictures, categorization, and mortality rate.

I. INTRODUCTION

Pulmonary/Lung nodules are the small cells that are grown inside the lung. These are classified into two types, such as cancerous (malignant) and non-cancerous (benign) [1]. One of the most fatal malignancies in the world is lung cancer. The lengthy time it takes to discover lung cancer and the poor prognosis for patients after therapy are two key contributors to the high death rate associated with this illness. If cancer could be detected sooner, it might lead to improved patient outcomes and a reduced death rate. However, the survival percentage for lung cancer is considerably increased by early identification. On the basis of minor morphological changes, locations, and clinical indicators, early-stage malignant lung nodules must be differentiated from noncancerous nodules [3]. There are many uses for Computed Tomography (CT) in modern society, but for the sake of this research only their significance in the medical field is discussed. In order to make an early diagnosis of malignant lung nodules, doctors employ a variety of diagnostic techniques, including clinical settings, Scans performed by nuclear medicine specialists utilizing CT and PET scans [4]. Regular screenings for lung nodules are conducted on those who are at a higher risk of developing lung cancer, including current and past smokers. Finally, finding lung nodules is critical for finding lung cancer early and treating it in a way that may enhance patient outcomes and lower death rate.

Almost 17,000 articles have addressed the use of CT scans to diagnose lung nodules since 2022, according to earlier research. There are distinct benefits to each of these research. For the purpose of detecting lung nodules in [5], precise CT lung segmentation is necessary [6]. Lung segmentation has been the target of many deep learning approaches. Algorithms use region-growing techniques to segment the lungs after inserting a small number of seed pixels into the corresponding area of the picture. To identify nodules on CT scans, two popular methods are nodule intensity and color thresholding. To improve the quality of the CT image, one possible preprocessing step is bi-histogram equalization [7]. Lung nodule identification sensitivity is enhanced by improved image registration techniques and the use of more intricate anatomical characteristics (thinner slices). Nevertheless, this greatly increases the datasets. It is possible to get as many as 500 separate pieces or slices of varying thicknesses from a single scan. Even for seasoned radiologists, it might take around 2.5–3.5 minutes to examine only one slice [8]. Radiologists have a lot more work to do when they suspect a nodule on a CT scan. Size, location, shape, adjacent structures, edges, density, and CT slice section thickness are some of the nodule characteristics that may affect the sensitivity of nodule detection [9]. The probability of accurately detecting nodules indicative of lung cancer rises when two radiologists examine the image, as opposed to when just one radiologist does so (82% vs. 68%). Detecting cancerous lung nodules at an early stage is a difficult and time-consuming task for radiologists. No matter how cautious the radiologist is, finding small nodules requires extensive image scanning, which is inherently error-prone.

Data encountered by current models or approaches is contaminated with noise, or erroneous information included in the dataset. The end goal of this proposed study is to use the provided data to identify lung nodules in CT scans by comparing the results produced using Decision Tree and employing Novel HLM to improve the accuracy value. In order to reduce mortality rates, this study aims to enhance the accuracy of identifying lung nodules, especially tiny or early-stage lesions.

II. MATERIALS AND METHODS

Researchers from the Saveetha Institute of Medical and Technical Sciences in Chennai worked on the project that resulted in this paper. We address two varieties. There are 35 people in each group. One may use clincalc.com to get the sample size by plugging in the F-score from previous research. The assumptions of an 80% pretest power and a constant alpha of 0.05 are used in all calculations.

Each category in this dataset used 35 samples. Separate the test data into two sets: one set uses 35 pieces trained using the Decision Tree classifier that is currently in use, and the other set uses 35 pieces trained with the Novel Hybrid Learning Method.

A. Algorithm for Novel Hybrid Learning Method

Step 1: Include all modules applied to the program.

Step 2: Install the information source you downloaded through Kaggle.

Step 3: Then, execute a research study on the database.

Step 4: Perform data visualization.

Step 5: Perform data Pre-Processing.

Step 6: Divide the dataset into two halves for testing and training.

Step 7 involves training a new hybrid learning algorithm on the dataset using Google Colab.

Step 8: Use accurate analytical techniques to verify the prediction.

B. Algorithm for Decision Tree

Step 1: Include all modules applied to the program.

Step 2: Install the information source you downloaded through Kaggle.

Step 3: Then, execute a research study on the database.

Step 4: Perform data visualization.

Step 5: Perform data Pre-Processing.

Sixth Step: Separate the data set into testing and training processes.

Step7: Use Google Colab's Decision Tree to train the data set.

Step 8: Confirm the forecast using reliable analytical methods.

This task required a 64-bit CPU, 8 GB of RAM, and a display resolution of 1024 x 768 pixels. Google Colab was used for the compilation of HLM code in that. The training and testing of the lung nodule data set occurs after the program has been launched. The current classifier, Decision Tree, is compared to the Novel Hybrid Learning Method's acquired accuracy. The acquired accuracy values are used for performance analysis. After the analysis was finished, the data was visualized. After that, the incorrect data is removed from the dataset by data preparation. After the sounds have been eliminated, the accuracy is next tested. Size, form, and other characteristics may be used to identify the existence of nodules in the lungs. The HLM method is implemented in Google Colab, the software used for the algorithm.

To increase precision, researchers employ the Hybrid Learning Method, which combines two algorithms. Both regression and classification may be accomplished using CNN and LGBM, a supervised deep learning method. A convolutional neural network's (CNN) most distinctive characteristic is the way it automatically extracts useful characteristics from input data using a succession of convolution layers. In order to create feature maps that emphasize patterns and edges in the input data, these convolution layers usually use a collection of filters or kernels.

The supervised learning technique known as Decision Tree can handle both classification and regression problems; however, it is often used for the former. The classifier is designed like a tree, with the dataset's attributes represented by the core nodes, the classification rules by the branches, and the result at the end of the process at each leaf node.

The workflow procedure is shown in Figure 1. Google Cola is used to implement the data set with code stimulation. Following data input, data visualization may take place. In the data preprocessing step, which follows visualization, the mounted code checks the Google Colab error numbers from the disk. After getting the lung nodule's accuracy using the Novel Hybrid Learning Method, we compare it to the accuracy of the current classifier, Decision Tree (DT).

C. Analyzing Data Statistically

In order to assess the veracity of the study project and its methodology, the SPSS software is used. This study uses CT scans of the lungs as its independent variable and the average accuracy as its dependent variable. The testing was conducted using an independent sample test.

III. RESULTS

Figure 1 shows the flow diagram for HLM algorithm to detect lung nodules and the flow is importing data set, Data Visualization, Data Pre-Processing, Trained Data, HLM Implementation (Google colab), Finding Accuracy.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

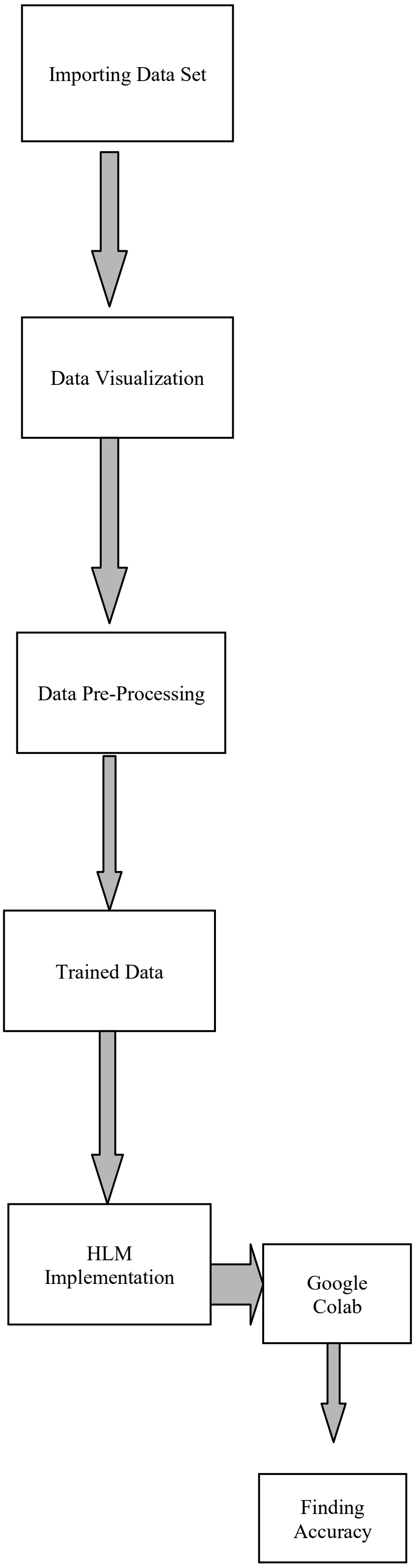

Fig. 1. Flow diagram for HLM algorithm to detect lung nodules.

In Figure 2, we can see a bar chart comparing the accuracy numbers of HLM and DT. HLM provides a higher level of accuracy, at around 97.723, while DT comes in at about 85.809. Here, we may use the bar chart to get the error bar as 95%(+/-2%).

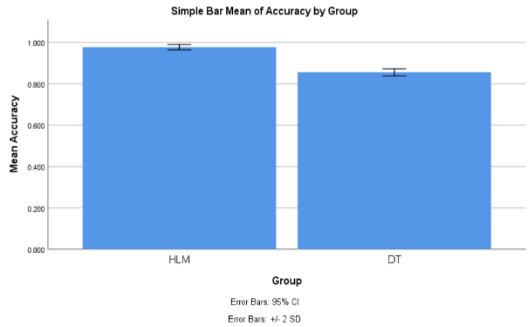

Figure 2 shows the average accuracy per group in the form of a simple bar chart. After making note of the accuracy for every image in the dataset using HLM and DT tests, we used SPSS Analytics to compare the results and get the mean average accuracy. The Mean Accuracy +/- 2 SD group is on the opposing side of the HLM vs. DT group.

Using 10 samples each, Table 1 displays the findings for DT and HLM in diagnosing lung nodules.

Groups 1 and 2's Percentage-Based Lung Nodule Detection Performance Comparison: Table I. As can be seen from the data in the table, Group 1 much outperforms Group 2. The highest accuracy achieved by novel HML is 97.23%, whereas that of DT is 85.659 percent.

ITERATIONS	ACCURACY	
	HML	DT
1	97.895	84.82
2	97.963	85.61
3	97.836	86.25
4	96.975	84.44
5	97.809	84.22
6	97.880	85.61
7	97.863	86.23
8	96.735	86.51
9	97.808	85.20
10	97.982	84.20

Table 2 displays the coefficient's Bayesian estimate. The data with a variance of 0.001 is shown in this table together with the values of the mode and means. The dependent variable here is accuracy, the model is groups, and the standard reference priors are supposed to be c. For each

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

group, we take the credible interval with a 95% confidence interval around the top and lower bounds, respectively.

Figure II. Bayesian Coefficient Estimation. Mean and mode values of the provided data with a 0.001 degree of dispersion. In this case, we choose the credible interval using the 95% upper bound and lower bound for the groups.

Groups	Mode	Mean Posterior	Variance	95% Credible Interval Lower Bound	95% Credible Interval Upper bound
HLM	97.723	97.723	0.001	97.5	98.0
DT	85.859	85.859	0.001	85.3	85.8

Table 3 contains the T-test tables. The HLM classifier obtained an average value of 97.723% when tested with 35 N per group, whereas the DT classifier achieved an average value of 85.859%, as shown in Table 1. We also find the standard deviation when the two classifiers' means are different.

T-Test Results Set (Table III). Taking into account the 35 N as a whole, we can see that the DT classifier's mean value is lower than the HLM classifier's (97.723%). The STD deviation was derived in different ways by the two classifiers present in the room.

	Groups	N	Mean	Std.Deviation	Std.Mean Error
Accuracy	HLM	35	97.723	0.06445	0.001089
	DT	35	85.859	0.08546	0 .001445

Both sets of data were subjected to an independent sample T-test using the data from Table 4. The findings showed a reliability of 67.400, a validity of 0.121943, and a dispersion of 0.001809. There is a notable disparity in the accuracy of the approaches at a significance level of 0.001 ($P<0.05$).

Independent sample test is shown in Table IV. A two-group independent sample t-test revealed an accuracy of 67.40, a mean difference of 12.1943, and a standard error difference of 0.001809.

Leven's Test for Equality of Variance				T-test for equality of Variance					95% Confidence Interval of the difference	
Accuracy		F	sig.	t	dif	sig(2-tailed)	Mean diff	Std.Error Difference	Lower	upper
	Equal Variance Assumed	3.265	0.001	67.400	68	0.001	12.194 3	0.001809	11.8333	121.55 53
	Equal Variance Not assumed			67.400	63.22 1	0.001	12.194 3	0.001809	11.8328	12.555 8

IV. DISCUSSION

A 97.723% greater accuracy rate in data prediction was seen using the HLM algorithm in comparison to the DT classifier, according to the study ($P<0.05$, Independent variable test, SPSS IBM tool). Testing at significance levels of 0.001 (P<\0.05), the methods show a diverse array of accuracy. Prior research has shown that HLM gives superior accuracy in detecting CT-scan lung nodules compared to state-of-the-art classifiers.

This further demonstrates that the same conclusions have been reached by other writers. To put the project into action, Python is used. The LIDC/IDRI database's data pre-processing offers more control over individual building parts than deep learning approaches. Because of this, other researchers may evaluate the findings using the same data set, which encourages studies to be reproduced. We put our CAD system through its paces on the LUNA16 Challenge. The LUNA16 Nodule Detection Challenge's Nodule Detection Track (NDET) [10] found that the proposed CAD system had the highest average FROC-score of 0.891, placing it first. Selected studies demonstrated a false positive rate ranging from 0.138 to 38.8 per scan and a sensitivity ranging from 68.9% to 100% for the diagnosis of lung nodules. Firmino et al. and a few of other selective experiments achieved sensitivity levels over 90%. What feeds into the nuero fuzzy classifier are feature vectors. This deep learning approach combines neural networks with fuzzy classifiers. The fuzzy layer makes use of the feature vector to create a pre-classification vector before passing it on to MLP for test sample classification. In trials using noisy nodule data, the suggested fuzzy neural network outperforms the state-of-the-art. In terms of pulmonary nodule classification models, the LUNA16 datasets were spot on. In a bid to save resources, the Faster R-CNN model may quickly eliminate superfluous CT images by ignoring them. This improves its ability to identify the pulmonary nodule in relevant images. This study proposes a computer-aided detection (CAD) method that partitions images to locate lung nodules. Our CAD system outperforms previously published systems with a state-of-the-art performance, with an increased detection sensitivity of 92.8% and 8 FPs/scan [18]. The suggested model's near-perfect performance could also be useful for users of real-time computer-aided systems in radiology.

The current setup has a few flaws, one of which being the lengthy training period required to train the image, which might lead to poorer accuracy rates on occasion. Our access to massive datasets allowed us to streamline the process of training for database picture classification by focusing on the most important features. The anticipated improvements

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

in AI and ML bode well for the future of early lung nodule diagnosis.

V. CONCLUSION

Since lung cancer is so common and often diagnosed too late, this deep learning method might prove useful at the first stages of the scanning process. Our system may accelerate the evaluation of a lung CT scans long sequence of images, reducing the likelihood of human error and the mortality rate. When compared to prior results, the suggested approach outperformed the current classifier and the accuracy prediction with less computing time. In comparison to DT's 85.859% accuracy, the suggested Novel Hybrid Learning Method algorithm achieved 97.723%. With significance levels of 0.001 (P<0.05), the methods greatly vary in accuracy.

REFERENCES

[1] F. H. Silver, T. Deshmukh, N. Ryan, A. Romm, and H. Nadiminti, "'Fingerprinting' Benign and Cancerous Skin Lesions Using Vibrational Optical Coherence Tomography: Differentiation among Cancerous Lesion Types Based on the Presence of New Cells, Blood Vessels, and Fibrosis," *Biomolecules*, vol. 12, no. 10. p. 1332, 2022. doi: 10.3390/biom12101332.

[2] C. Printz, "American Society of Clinical Oncology says 5% to 6% of new cancers and cancer deaths are attributable to alcohol consumption," *Cancer*, vol. 124, no. 8. pp. 1642–1642, 2018. doi: 10.1002/cncr.31363.

[3] L. Zhao, C. Bai, and Y. Zhu, "Preliminary Study on Diagnostic Value of Artificial Intelligence in Early-Stage Lung Cancer," *D99. ADVANCING RISK ASSESSEMENT FOR PULMONARY NODULES*. 2020. doi: 10.1164/ajrccm-conference.2020.201.1_meetingabstracts.a7688.

[4] G. B. Coura-Filho, M. T. S. de Oliveira, and A. L. M. de Campos, "Positron Emission Tomography-Computed Tomography (PET-CT) in Thyroid Nodules," *Nuclear Medicine in Endocrine Disorders*. pp. 59–68, 2022. doi: 10.1007/978-3-031-13224-7_8.

[5] G. Liang, Z. Diao, and H. Jiang, "Uncertainty Analysis Based Attention Network for Lung Nodule Segmentation from CT Images," *2022 the 6th International Conference on Virtual and Augmented Reality Simulations*. 2022. doi: 10.1145/3546607.3546615.

[6] J. Sousa, T. Pereira, I. Neves, F. Silva, and H. P. Oliveira, "The Influence of a Coherent Annotation and Synthetic Addition of Lung Nodules for Lung Segmentation in CT Scans," *Sensors*, vol. 22, no. 9. p. 3443, 2022. doi: 10.3390/s22093443.

[7] G. Noyel and M. Jourlin, "Region Homogeneity In The Logarithmic Image Processing Framework: Application To Region Growing Algorithms," *Image Analysis & Stereology*, vol. 38, no. 1. p. 43, 2019. doi: 10.5566/ias.2038.

[8] S. Wu, M. Nakao, K. Imanishi, M. Nakamura, T. Mizowaki, and T. Matsuda, "Computed Tomography slice interpolation in the longitudinal direction based on deep learning techniques: To reduce slice thickness or slice increment without dose increase," *PLOS ONE*, vol. 17, no. 12. p. e0279005, 2022. doi: 10.1371/journal.pone.0279005.

[9] J. Wang *et al.*, "Pulmonary Nodule Detection in Volumetric Chest CT Scans Using CNNs-Based Nodule-Size-Adaptive Detection and Classification," *IEEE Access*, vol. 7. pp. 46033–46044, 2019. doi: 10.1109/access.2019.2908195.

[10] A. Rey, B. Arcay, and A. Castro, "A hybrid CAD system for lung nodule detection using CT studies based in soft computing," *Expert Systems with Applications*, vol. 168. p. 114259, 2021. doi: 10.1016/j.eswa.2020.114259.