An Integrated Hybrid Neural Network Framework for Early Prediction of Lung Cancer

Aseena Shaik Babu[1], Oruganti Aravindu[2], Komatireddy Sai Vardhan[3], Dupati Abhilash[4], Empati Harshit[5]

[1] Assistant Professor, Department Of Computer Science And Engineering(AI& ML), Samskruthi College of Engineering And Technology , Kondapur(V), Ghatkesar(M), Medchal(D),Telangana

[2,3,4,5]BTech Students ,Department Of Computer Science And Engineering(AI& ML), Samskruthi College of Engineering And Technology , Kondapur(V), Ghatkesar(M), Medchal(D),Telangana

Abstract—**Lung cancer is one of the most life-threatening diseases, mainly because it is often detected at a late stage. Identifying the disease early and determining its stage accurately can greatly improve a patient's chances of survival. Although various techniques such as image processing, biomarker analysis, and automated systems are available, achieving high accuracy in early diagnosis still remains a major challenge for medical professionals. In this study, CT scan images obtained from the LIDC-IDRI dataset are used to develop an improved detection system. Unlike traditional approaches that depend on manual observation, this research introduces a Hybrid Neural Network-based method called Cancer Cell Detection using Hybrid Neural Network (CCDC-HNN). The model uses deep learning techniques to extract important features from CT images, which helps in identifying cancer at an early stage. Additionally, a 3D Convolutional Neural Network is applied to enhance the accuracy of detection and to clearly distinguish between benign and malignant tumors. The performance of the proposed model is evaluated using standard statistical measures, and the results show that it provides reliable and accurate predictions, making it a promising approach for early lung cancer diagnosis.**

Keywords—**Lung Cancer Detection, Hybrid Neural Network, Deep Learning, 3D Convolutional Neural Network (3D-CNN), CT Scan Imaging, Early Diagnosis, Feature Extraction, Tumor Classification, Benign and Malignant Detection, Medical Image Analysis, LIDC-IDRI Dataset, Computer-Aided Diagnosis (CAD).**

I. INTRODUCTION

Lung cancer is a serious and life-threatening disease caused by the uncontrolled growth of abnormal cells in the lungs. Its exact cause is not fully known, but factors such as smoking, environmental exposure, and genetic conditions increase the risk. The global rise in cancer cases has made it a major public health concern, requiring improved diagnostic methods [2]. Early detection is essential to increase survival rates, yet identifying lung cancer at an initial stage remains challenging. Recent advancements in deep learning, especially convolutional neural networks, have shown promising results in analyzing medical images and identifying cancer-related patterns more effectively than traditional approaches [5]. These technologies provide new opportunities to enhance diagnostic accuracy and support healthcare professionals in making better clinical decisions.

Conventional lung cancer diagnosis mainly depends on radiologists examining CT scan images, which can be time-consuming and sometimes inaccurate. Even experienced professionals may struggle to detect small nodules or early-stage abnormalities. To overcome these limitations, researchers have explored automated systems based on machine learning and image processing techniques [10]. These systems aim to assist doctors by improving detection speed and accuracy. Deep learning-based classification methods have been successfully applied to CT images, enabling better identification of lung nodules and reducing human error [7]. Therefore, integrating intelligent systems into medical diagnostics can significantly enhance the reliability of early cancer detection.

Lung nodules are small masses of tissue that may indicate the presence of cancer, and their accurate detection is crucial for diagnosis. These nodules vary in size, shape, and density, making them difficult to identify using traditional methods. Advanced segmentation techniques have been developed to detect small nodules in CT images more precisely [4]. However, distinguishing between benign and malignant nodules remains a complex task due to similarities in their appearance. Deep learning-based approaches have been introduced to improve classification and management of pulmonary nodules by learning detailed features from imaging data [6]. These methods help reduce false detections and improve diagnostic performance.

Lung cancer is generally categorized into Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC), with NSCLC being more common. Accurate classification of these types is important for selecting appropriate treatment strategies. Computer-Aided Diagnosis (CAD) systems have been widely adopted to support this process by analyzing medical images and detecting abnormalities [15]. Compared to traditional X-ray imaging, CT scans provide more detailed and comprehensive information about lung structures. Advanced decision support systems and deep learning frameworks have further improved the detection and classification of lung tumors, enabling more precise and efficient diagnosis [12]. The final stage of this work focuses on evaluating the trained model using new CT scan images to verify its performance in real-world scenarios. The system processes the input images and predicts the presence of lung cancer, along with identifying whether the detected nodules are benign or malignant using the proposed hybrid neural network model. This step highlights the practical usefulness of the developed approach in assisting medical professionals with faster and more accurate diagnosis. The ability to automatically analyze complex medical images can support early detection and help in timely treatment decisions, which is crucial for improving patient survival rates [7][12]. Overall, this study demonstrates the importance of deep learning techniques in modern healthcare and their capability to solve critical diagnostic challenges. It also opens the way for future enhancements, such as incorporating larger datasets, improving model efficiency, or integrating more advanced architectures to achieve even better accuracy and reliability.

II. LITERATURE SURVEY

R. Yamashita et al. (2018) [5] discussed the growing importance of convolutional neural networks in the field of radiology and medical imaging. Their study explains how CNN models are capable of automatically learning complex features from medical images, reducing the need for manual feature extraction. The authors highlighted that CNNs have been successfully applied in various diagnostic tasks such as tumor detection, classification, and segmentation. They also emphasized that deep learning models can improve diagnostic accuracy when trained on large datasets. However, the study pointed out challenges such as the need for high computational power and large labeled datasets. Overall, their work provides a strong foundation for applying deep learning techniques in healthcare, particularly in lung cancer detection, where accurate image interpretation is essential for early diagnosis and effective treatment planning.

F. Ciompi et al. (2017) [6] focused on the use of deep learning techniques for automatic pulmonary nodule management in lung cancer screening. Their research proposed a system that can classify lung nodules based on CT scan images using advanced neural network models. The study demonstrated that deep learning methods can effectively differentiate between different types of nodules by learning their unique characteristics. This approach reduces the dependency on manual interpretation and improves the consistency of results. The authors also highlighted the importance of large annotated datasets for training accurate models. Their findings showed that automated systems can support radiologists by improving detection accuracy and reducing workload. This work plays a significant role in advancing computer-aided diagnosis systems for lung cancer and supports the use of hybrid models for improved performance.

Q.Z. Song et al. (2017)[7] presented a deep learning-based method for classifying lung nodules using CT scan images. Their study utilized neural networks to automatically extract relevant features from images, eliminating the need for manual processing. The proposed method achieved high accuracy in distinguishing between different types of lung nodules, demonstrating the effectiveness of deep learning in medical image classification. The authors emphasized that early and accurate classification is critical for improving patient outcomes. They also discussed the challenges of handling complex image data and the importance of optimizing model performance. Their research contributes to the development of reliable automated systems that assist in early detection of lung cancer. It highlights how deep learning can enhance diagnostic precision and reduce human error in clinical practice.

A. Masood et al. (2020)[12] developed an advanced decision support system for lung cancer detection and classification using enhanced deep learning techniques. Their approach combined region-based convolutional neural networks with multilayer feature fusion to improve detection accuracy. The system was designed to identify lung nodules and classify them more effectively compared to traditional methods. The authors demonstrated that integrating multiple features from different layers of the network leads to better performance in detecting complex patterns in medical images. Their results showed significant improvements in accuracy and reliability, making the system suitable for clinical applications. This study highlights the importance of combining different deep learning techniques to build robust diagnostic models. It also supports the idea of hybrid frameworks for achieving higher precision in lung cancer detection.

J. Talukdar et al. (2018) [15] provided a comprehensive survey of various image processing techniques used for lung cancer detection from CT scan images. The study reviewed multiple methods, including traditional image processing, machine learning, and deep learning approaches. The authors discussed the advantages and limitations of each technique, emphasizing that deep learning models offer better performance in terms of accuracy and automation. They also highlighted challenges such as noise in images, variability in nodule appearance, and difficulties in segmentation. The survey concluded that combining different techniques can improve overall detection performance. This work is valuable for understanding the evolution of lung cancer detection methods and the need for advanced systems. It supports the development of hybrid models that integrate multiple approaches for more accurate and reliable diagnosis.

III. DATASET DETAILS

The dataset used in this project is based on lung CT scan images collected from the LIDC-IDRI repository. It consists of a large collection of medical images that are useful for detecting lung cancer at early stages. The dataset contains more than 14,000 CT scan images along with corresponding mask images, which help in identifying the exact region of cancerous cells. These mask images play an important role in segmentation, where regions with white pixels indicate the presence of malignant cells, while regions without such markings are considered benign. The data is organized in image format, making it suitable for deep learning models that process visual information. This dataset provides a strong foundation for training models to detect, analyze, and classify lung nodules accurately, which is essential for improving early diagnosis and supporting medical decision-making.

Before training the model, several preprocessing steps are applied to improve the quality and usability of the dataset. Initially, all CT scan images are loaded into the system and resized to a consistent format to ensure uniformity. After resizing, normalization is performed on pixel values to bring them within a specific range, which helps in stabilizing and improving the performance of deep learning algorithms. The dataset is then divided into training and testing sets, where 80% of the images are used to train the model and the remaining 20% are used for evaluation. This split ensures that the model is tested on unseen data, allowing for a reliable assessment of its performance. Proper preprocessing helps the model learn important features effectively and contributes to achieving accurate and consistent results in lung cancer detection.

IV. PROPOSED METHODOLOGY

The proposed system follows a systematic approach to detect and classify lung cancer using a hybrid deep learning model. Initially, CT scan images are collected from the LIDC-IDRI dataset. The dataset is then preprocessed to ensure consistency and quality of input data. This includes resizing all images to a uniform size and normalizing pixel values to improve model performance. After preprocessing, the dataset is divided into training and testing sets, where 80% of the images are used for training and 20% for testing. To enhance detection accuracy, a UNET-based segmentation technique is applied to identify and isolate the regions containing potential cancer cells. This segmentation step helps the model focus on relevant areas in the images, improving the efficiency of feature extraction and reducing noise from irrelevant regions.

Once the segmented images are obtained, a hybrid deep learning model called CCDC-HNN is applied for classification. This model combines the strengths of 3D Convolutional Neural Networks (3D-CNN) and Long Short-Term Memory (LSTM) networks. The 3D-CNN is responsible for extracting deep spatial features from CT scan images, while the LSTM captures sequential patterns and variations across image slices. The trained model is then evaluated using performance metrics such as accuracy, precision, recall, and confusion matrix. Visualization tools like ROC curves and training graphs are also used to analyze model performance. Finally, the system allows users to input new CT scan images, where the model detects cancer cells and classifies them as benign or malignant, demonstrating its practical application in real-world medical diagnosis.

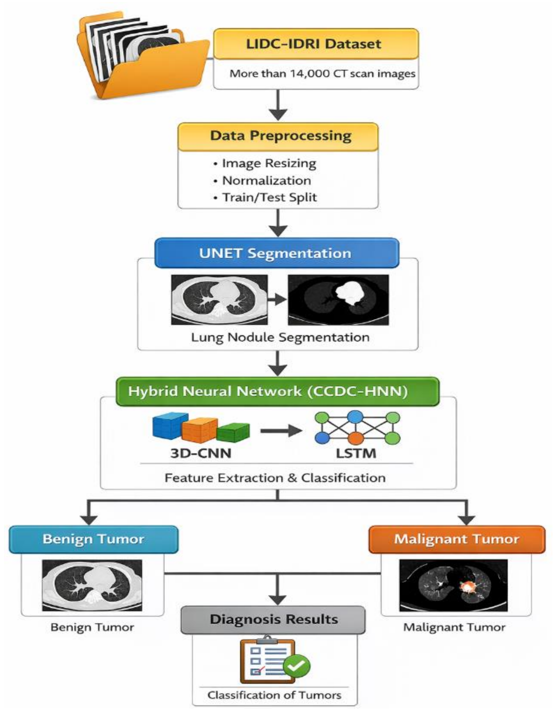

Figure [1] : Hybrid Deep Learning-Based Lung Cancer Detection and Classification System

Figure [1] presents the workflow of the proposed lung cancer detection system. It starts with CT scan images from the LIDC-IDRI dataset, followed by preprocessing steps like resizing and normalization. The images are then processed using UNET for segmentation of lung nodules. Next, the hybrid CCDC-HNN model (3D-CNN + LSTM) extracts features and performs classification. Based on the analysis, the system produces diagnosis results and finally classifies tumors as benign or malignant.

V. RESULT AND DISCUSSION

The results obtained from this study show that the proposed hybrid deep learning model (CCDC-HNN) performs effectively in detecting and classifying lung cancer using CT scan images. After preparing the dataset through preprocessing and dividing it into training and testing sets, the model was trained on 80% of the data and evaluated on the remaining portion. The system achieved an accuracy of approximately 91%, which indicates its strong capability in identifying cancerous and non-cancerous cases. Other performance measures such as precision, recall, and F1-score also produced stable and satisfactory values, showing that the model is consistent in its predictions. The confusion matrix clearly shows that most of the predictions are correct, with very few errors observed. In addition, the ROC curve demonstrates good separation between benign and malignant classes, confirming the model's reliability. The use of UNET for segmentation along with the combination of 3D-CNN and LSTM has helped improve overall performance. These results suggest that the proposed method can be a useful tool for supporting early lung cancer diagnosis.

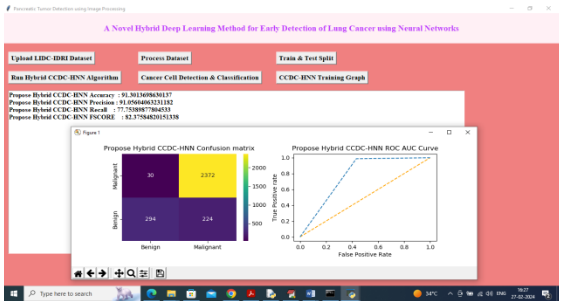

Figure [2]: Model Performance and Confusion Matrix

Figure [2] displays the performance of the proposed hybrid model using metrics such as accuracy, precision, and recall. The confusion matrix is also shown, where most values are concentrated along the diagonal, indicating correct predictions. This visual representation helps in understanding how well the model classifies benign and malignant cases.

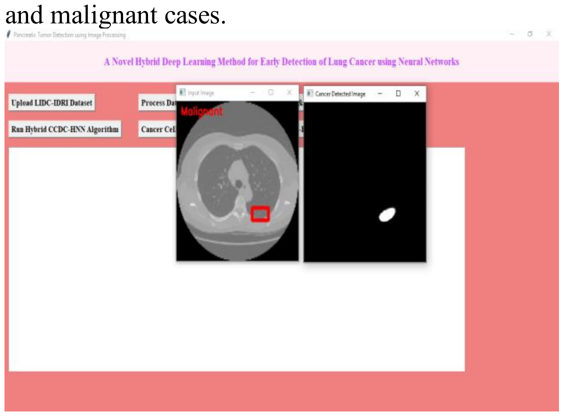

Figure [3]: Cancer Cell Detection and Classification Output

Figure [3] shows the final output of the system where a test CT image is analyzed. The first image represents the original scan, while the second image highlights the segmented region of the detected cancer cells. The system then classifies the result as either benign or malignant. This screen demonstrates the practical working of the model in real-time diagnosis.

DISCUSSION

The results of this work highlight the value of using hybrid deep learning models for medical image analysis, especially in lung cancer detection. The proposed CCDC-HNN model shows strong performance because it combines the advantages of

3D-CNN and LSTM networks. The 3D-CNN is effective in capturing detailed spatial information from CT scan images, while the LSTM helps in understanding patterns and variations across image sequences. This combination makes the system more capable of handling complex data compared to single-model approaches. In addition, the use of UNET segmentation improves the overall accuracy by isolating important regions of interest, allowing the model to focus only on relevant features. Proper preprocessing steps, including normalization and data splitting, also contributed to stable and reliable results. Visualization methods such as confusion matrices and ROC curves provided a clear understanding of model performance. Overall, the study demonstrates that integrating multiple deep learning techniques can enhance diagnostic accuracy and provide useful support in early detection of lung cancer.

VI. CONCLUSION

This project presents an effective approach for early lung cancer detection using a hybrid deep learning model applied to CT scan images. The dataset was carefully prepared through preprocessing steps such as resizing, normalization, and splitting into training and testing sets, ensuring reliable and consistent model performance. These preprocessing techniques helped in reducing noise and improving the quality of input data, which directly contributed to better learning outcomes.

The use of UNET segmentation enabled the system to accurately detect and isolate regions of interest containing potential cancerous nodules. This step plays a crucial role in improving classification performance, as it allows the model to focus only on the relevant areas of the lung. In addition, the hybrid combination of 3D-CNN and LSTM within the proposed CCDC-HNN model enhances the system's capability to capture both spatial and sequential features from CT scan data. This combination improves the model's understanding of complex patterns, leading to more precise classification of lung nodules.

The model achieved strong accuracy and showed stable performance across different evaluation metrics. It was able to effectively distinguish between benign and malignant cases, reducing the chances of misclassification. The use of performance metrics such as accuracy, precision, recall, and F1-score, along with visualization tools, confirmed the robustness and reliability of the proposed approach.

Another important outcome of this system is its practical applicability in real-world medical scenarios. The model can analyze new CT scan images quickly and provide early predictions, which can assist doctors in making timely decisions. Early detection of lung cancer significantly increases the chances of successful treatment and patient survival, making this system highly valuable in healthcare settings.

Moreover, the proposed framework is flexible and can be further improved by incorporating larger and more diverse datasets, which would help in increasing generalization and reducing bias. The integration of more advanced architectures and optimization techniques could also enhance performance and reduce computational complexity.

Overall, this work demonstrates the significant potential of deep learning techniques in medical image analysis. It provides a reliable and efficient method for early lung cancer detection and establishes a strong foundation for future research aimed at improving accuracy, speed, and real-world deployment in clinical environments.

REFERENCES

[1] M. Kaur, S.R. Sakhare, K. Wanjale, et al., Earlystroke prediction methods for prevention of strokes, Behav. Neuro. (2022) 7725597, 1–9.

[2] R.L. Siegel, et al., Cancer statistics, Cancer J. Clin. 72 (2022) 7–33.

[3] M.D. Martin, J.P. Kanne, L.S. Broderick, et al., Lung-RADS: Pushing the limits, RadioGraphics 37 (7) (2017) 1975–1993.

[4] S. Diciotti, G. Picozzi, M. Falchini, et al., 3-D segmentation algorithm of small lung nodules in spiral CT images, IEEE Trans. Inf. Technol. Biomed. 12 (1) (2008) 7–19.

[5] R. Yamashita, M. Nishio, R.K.G. Do, et al., Convolutional neural networks: An overview and application in radiology, Insights Imaging 9 (2018) 611–629.

[6] F. Ciompi, et al., Towards automatic pulmonary nodule management in lung cancer screening with deep learning, Sci. Rep. 7 (1) (2017) 1–11.

[7] Q.Z. Song, L. Zhao, X.K. Luo, et al., Using deep learning for classification of lung nodules on computed tomography images, J. Healthc. Eng. (2017) 8314740, 2017.

[8] S. Khan, N. Islam, Z.I. Jan, et al., A novel deep learning-based framework for detecting and classifying breast cancer using transfer learning, Pattern Recognit. Lett. 125 (2019) 1–6.

[9] Y.A. Hamad, K. Simonov, M.B. Naeem, Breast cancer detection and classification using artificial neural networks, in: Proc. 1st Annual Int. Conf. on Info. and Sc. (AiCIS), 2018, pp. 51–57.

[10] S. Bhatia, N. Mittal, S.K. Sonbhadra, et al., Lung cancer detection: a deep learning approach, in: Soft Comp. for Prob. Solving, 2019, pp. 699–705.

[11] D. Palani, K. Venkatalakshmi, An IoT based predictive modelling for predicting lung cancer using fuzzy cluster based segmentation and classification, J. Med. Syst. 43 (2) (2019) 21.

[12] A. Masood, et al., Automated decision support system for lung cancer detection and classification via enhanced RFCN with multilayer fusion RPN, IEEE Trans. Ind. Inform. 16 (12) (2020) 7791–7801.

[13] V. Fredriksen, et al., Teacher-student approach for lung tumor segmentation from mixed-supervised datasets, PLoS One 17 (1) (2022) e0261917.

[14] T. Saba, et al., Cloud-based decision support system for detecting and classifying malignant cells in breast cancer using breast cytology images, Micro. Res. Tech. 82 (6) (2019) 775–785.

[15] J. Talukdar, et al., A survey on lung cancer detection in CT scans images using image processing techniques, Int. J. Curr. Trends Sci. Technol. 8 (3) (2018) 20136–20140.

[16] R. Krithiga, P. Geetha, Deep learning-based breast cancer detection and classification using fuzzy merging techniques, Mach. Vis. Appl. 31 (7) (2020) 1–18.

[17] N. Shrivastava, J. Bharti, et al., Breast tumour detection and classification based on density, Multimedia Tools Appl. 79 (35) (2020) 26467–26487.

[18] R. Suresh, A.N. Rao, B.E. Reddy, et al., Detection and classification of normal and abnormal patterns in mammograms using a deep neural network, Concurr. Comput.: Pract. Exper. 31 (14) (2019) e5293.

[19] A.T. Saba, et al., Lung nodule detection based on ensemble of hand crafted and deep features, J. Med. Syst. 43 (12) (2019) 332.