International Journal of Engineering Research and Sustainable Technologies
Volume 2, No.4, Dec 2024, P 1 - 6
ISSN: 2584-1394 (Online version)

DETECTION OF CHEST DISEASES IN RADIO GRAPHS USING DEEP LEARNING

[1]Roselin Sahana Roy V, [2]Pranesh E
[1,2]Department of Computer Science and Engineering, [1,2]Loyola Institute of Technology and Science, Thovalai, India
*** Corresponding author email address**: roselinsahanaroy321@gmail.com

DOI: https://doi.org/10.63458/ijerst.v2i4.92 | ARK: https://n2t.net/ark:/61909/IJERST.v2i4.92

Abstract
One of the most widely used and easily accessible diagnostic methods for identifying a variety of pulmonary disorders is still chest radiography. However, interpreting chest radiographs accurately requires significant experience and time, which can lead to delays in diagnosis and treatment. To enhance the precision and efficiency of diagnosis, we present a deep learning-based method for the automated identification of chest disorders in radiographs. Our approach analyzes chest radiographs and detects anomalies indicative of various pulmonary diseases using convolutional neural networks (CNNs), a type of deep learning model well-suited for image recognition tasks. We utilize a large collection of annotated chest radiographs covering a wide range of pathological conditions, including pulmonary edema, lung cancer, pneumonia, and tuberculosis.

Keywords: Lung Cancer, Deep Learning, Convolutional Neural Network, CT scan, Computer- Aided Diagnosis, Image Preprocessing, Image Segmentation, Feature Extraction, Medical Imaging, Predictive Modelling.

1. Introduction

Lung cancer, the second most prevalent cancer in the world, is a serious public health concern, claiming numerous lives each year. Despite advancements in medical technology, early detection remains a challenge, often resulting in delayed diagnosis and treatment initiation. Symptoms of lung cancer, such as coughing, chest pain, and breathlessness, typically manifest in later stages, complicating timely intervention. However, with approximately 131,880 deaths projected in the United States alone in 2021, there is an urgent need for improved diagnostic methods to reduce mortality rates associated with this disease. Various imaging techniques, including magnetic resonance imaging (MRI), computed tomography (CT), X-ray, positron emission tomography (PET), and single-photon emission computerized tomography (SPECT), offer valuable insights for staging and segmenting lung cancer. Among these, CT scans have emerged as a powerful tool, capable of producing high-resolution 3D images of the chest, facilitating accurate nodule detection. Research indicates that CT scans outperform conventional chest X-rays, particularly in screening individuals at high risk of developing cancer. Despite advancements in imaging technology, the rising prevalence of lung cancer—fuelled by factors such as smoking and environmental pollution—underscores the need for more effective diagnostic approaches. Machine learning methods, particularly deep learning algorithms, have revolutionized cancer detection by enhancing accuracy and efficiency. Leveraging non-small cell lung cancer radio genomic datasets and deep learning techniques, this study aims to improve tumour identification accuracy. By integrating preprocessing, image extraction, binarization, segmentation, and thresholding into a comprehensive framework, the study streamlines the diagnostic process. Specifically, convolutional neural network (CNN) systems are applied to CT chest images from DICOM datasets to predict malignancy at early stages, thereby facilitating timely intervention and improving patient outcomes.
.

2. Literature Review

Investigating the current landscape of automated lung cancer prediction methods, this study explores the efficacy of Convolutional Neural Networks (CNNs) for accurate diagnosis. Through a meticulous evaluation of diverse machine learning algorithms, including CNNs, our research underscores the unmatched potential of deep learning techniques in medical image analysis. Leveraging the VGG-16 architecture, our findings highlight the path toward enhanced predictive accuracy, particularly in discerning various lung cancer types such as Adenocarcinoma, Large Cell Carcinoma, and Squamous Cell Carcinoma. This comprehensive analysis lays a solid foundation for future advancements in lung cancer diagnosis and treatment planning, contributing significantly to the evolving field of medical image analysis. [1]

This paper begins with the acquisition of lung images from MRI and CT scans. Subsequently, advanced image processing techniques are applied to enhance image quality, remove noise, and improve feature extraction. Robot-assisted feature selection plays a crucial role in enhancing system performance. Robots are employed to scan lung images and identify relevant features indicative of potential cancerous regions. This collaborative approach reduces human error and increases system efficiency. [1,2]

When PET and MRI are combined, structural data from MRI and functional imaging from PET are integrated. However, due to interference between the modalities, numerous technical challenges arise. These issues must be addressed using different strategies to correct system flaws and improve image quantification. This review details the various challenges associated with the current PET/MRI system and discusses the future directions of hybrid imaging modalities. Additionally, different PET/MRI data collection and analysis methods are examined in greater depth using specialized software tools. [4]

To categorize lung cancer and assess its severity, as well as to identify malignant lung nodules from input lung images, this study employs innovative deep learning techniques. The best feature extraction methods used in this work include the Zernike Moment, along with the extraction of texture, geometric, volumetric, and intensity features. The optimal features are selected using the Fuzzy Particle Swarm Optimization (FPSO) algorithm. Finally, deep learning techniques are used to classify these features. [5]

By reducing misclassification, this approach enhances lung image quality and improves lung cancer diagnosis. The weighted mean histogram equalization technique is used to effectively reduce image noise and enhance image quality. Additionally, the Improved Profuse Clustering Technique (IPCT) is employed for segmenting affected regions. [6]

"Early Detection of Lung Cancer Using Machine Learning Techniques" by B. Devananda Rao and Dr. Mahammad Arshad presents a comprehensive framework employing deep neural networks and AI for the early detection and staging of lung cancer from CT images. Drawing from prior research, this study focuses on automating detection processes and reducing diagnosis time and costs. Leveraging techniques such as convolutional neural networks (CNNs), image processing, and feature extraction, the study demonstrates promising results, laying a strong foundation for future advancements in medical diagnostics. [10]

Machine learning-based approaches have revolutionized medical image analysis, particularly in segmenting, detecting, and classifying anomalies. These models excel in image classification, detection, and segmentation, facilitating effective image fusion. Key factors contributing to their success include neural networks' ability to extract features from high-dimensional data and GPU acceleration, which enables processing speeds 10–30 times faster than traditional CPUs.

(Deepika Gupta, 2023) In a different approach, authors collected lung image datasets from the Cancer Imaging Archive (CIA) to predict lung cancer. The CIA dataset includes images from various modalities. A total of 5,043 lung images were obtained for analysis. Lung cancer prediction was conducted using two methods: the Improved Profuse Clustering Technique (IPCT) and deep learning with instantaneously trained neural networks (DITNNs). Lung images were preprocessed using the weighted mean function, which improves image description by reintroducing pixels through the cumulative distribution function and probability. An enhanced profuse clustering method was then applied for segmentation. With the lowest classification error of 0.038, a notable cancer detection rate was achieved. [3,11]

3. System Methodology

Step1: Image Pre-processing:

In the initial phase of lung malignancy detection from CT scans, image pre-processing is essential for enhancing image quality and improving detection accuracy. Captured images undergo noise removal and quality enhancement using techniques such as Gaussian and median filters, which effectively smooth speckle noise and remove salt-and-pepper noise from grayscale CT scan images. Additionally, methods like histogram equalization and adaptive bilateral filtering (ABF) may be employed to further enhance image contrast and reduce noise. Enhanced discrete wavelet transform (DWT) denoising techniques also contribute to optimizing image quality. These pre-processing steps (Figure 2) ensure that the images are well-prepared for subsequent lung cancer detection algorithms, improving the reliability and performance of the overall detection system.

Step2: Image Segmentation:

Image segmentation is a critical step in lung cancer detection, as it isolates lung tissue from unwanted noise, enabling accurate tumor examination. Various techniques, such as thresholding, clustering, and watershed segmentation, have been explored for this purpose. While K-means and mean shift clustering have yielded suboptimal results, watershed segmentation has shown promise despite its time-consuming nature. However, thresholding, although employed in some studies, often leaves residual noise. To address these challenges, researchers have explored innovative approaches such as marker-based watershed segmentation to improve accuracy and noise removal. Additionally, techniques like the enhanced profuse clustering technique (IPCT) have been utilized, leveraging pixel resemblance to identify afflicted regions. The U-Net convolutional network has emerged as a popular and efficient tool for image segmentation, featuring shortcut connections that preserve high-resolution features between layers.

Step3: Feature Extraction:

At this stage, features of the lung tumor are extracted to determine whether the lump is primary, malignant, or suspicious

Step4: Detection and Classification:

Deep learning techniques play a crucial role in precisely locating and classifying tumor types in lung cancer detection. Classification—distinguishing between benign and malignant tumors—is a fundamental aspect of this process. Researchers often refer to tumor detection using the primary tumor (T), regional lymph node involvement (N), and metastasis (M) terminology. Previous studies have explored various classification methods, including FPSO-CNN, which significantly reduces the computational complexity of CNNs. This technique effectively identifies cancerous nodules in lung images and determines their precise locations.

4. Materials and Methods

We utilized the **LIDC-IDRI lung dataset**, which encompasses a total of **10,000 images**. To develop an effective diagnostic system for determining the stage of lung tumors, we employed deep learning techniques, specifically **Convolutional Neural Networks (CNNs)**. CNNs are a specialized type of deep learning architecture well-suited for image classification tasks.

The **LIDC-IDRI dataset** provides a comprehensive resource for training and evaluating our model, enabling accurate diagnosis of different stages of lung tumors. Through extensive experimentation and fine-tuning of the CNN architecture, we aim to achieve **high accuracy and robust performance** in lung tumor stage classification. Ultimately, this research contributes to improved medical diagnostics and patient care.

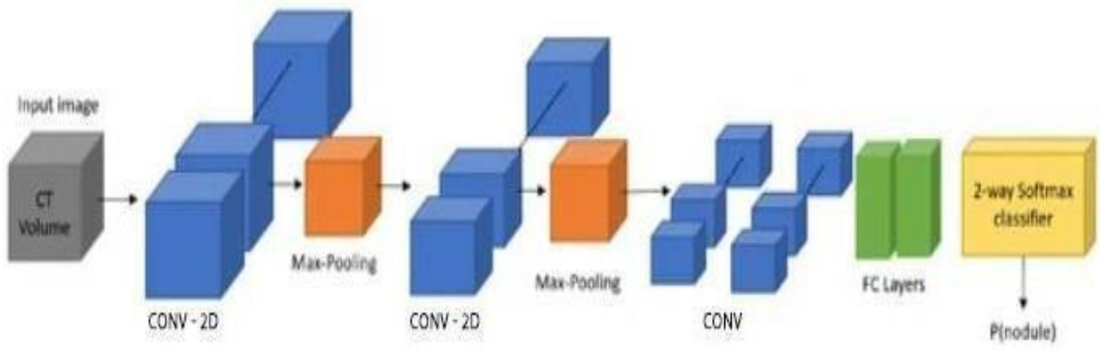

Fig.1. Architecture of Proposed CNN model

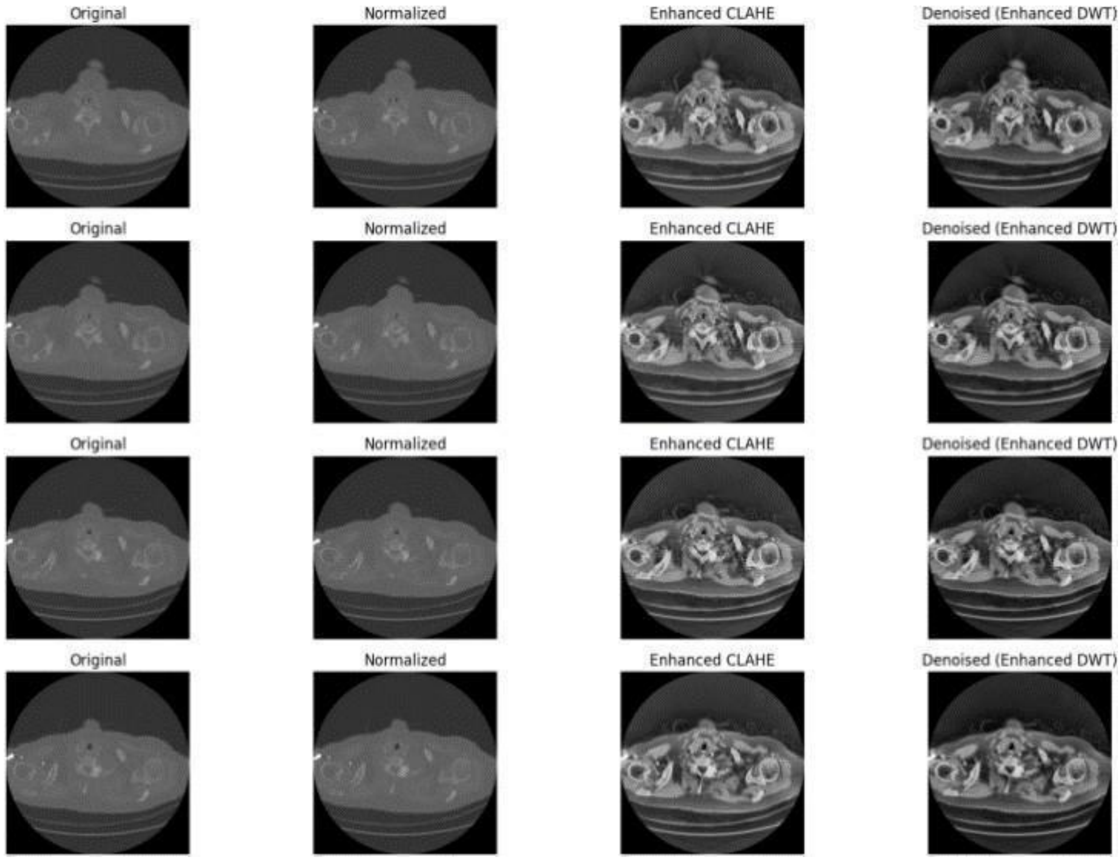

Fig.2 Image Pre-processing

4.1 ***Proposed CNN***

Each Each successive intermediate layer uses the output from the previous layer as its input (Figure. 1). The input layer, which is the first layer, contains a number of neurons equal to the total number of pixels in the image—256 pixels in this case. It is directly connected to the input image. The next stage consists of convolutional layers, which extract features from the input data by applying a set of filters. The kernel configuration determines the filter sizes, which, in the proposed system, are 3×3. Each patch represents a distinct area of the previous layer and corresponds to the receptive field.

Implementation Steps:

1. Input: Lung CT scanned DICOM images (512 × 512) are preprocessed and resized to 256 × 256.
2. Augmentation: Imbalance in classes is addressed through augmentation techniques.
3. Normalization: Image data is normalized to values between 0.0 and 1.0.
4. Labeling: Dataset images are labeled into 4 classes (zero- three) based on their stage.
5. Data Splitting: The dataset is split into training and testing sets with an 80-20 ratio.
6. Convolutional Layers: 2D convolutional layers and batch normalization are applied.
7. Dropout Regularization: To prevent over fitting, 20% of the layers are randomly dropped out.
8. Additional Layers: More convolutional, batch normalization, and max-pooling layers are added, with another 20% dropout.
9. Flattening: Data is converted into a 1 Dim array for inputting into the next layer.
10. Fully Connected Layer: A dense layer is applied for feature extraction.
11. Output Layer: Softmax function is used for multiclass distribution prediction.
12. Develop a website for lung cancer detection and staging using a frontend interface for image upload, backend logic for CNN-based classification, model integration, testing, and deployment with consideration

for legal and ethical standards.

4.2 Dataset

The LIDC-IDRI lung dataset was utilized for training the nodule detection model, consisting of eighteen thousand CT scans separated into working out and authentication sets, with augmentation applied to balance the dataset, including shearing, zooming, and horizontal flipping. Class imbalance was addressed through augmentation techniques.

a) Preprocess and augment the LIDC-IDRI lung dataset.
b) Organize dataset into class
c) Partition dataset into smaller parts for resource management.
d) Configure CNN model with fixed batch size and learning rate for training.
e) Train model using Adam optimizer on the augmented dataset.
f) Split data into training, testing, and validation sets for evaluation.
g) Validate model performance on validation set to monitor progress.
h) Test trained model on separate testing set for evaluation.
i) Iterate on training process for refinement, adjusting model and augmentation techniques as needed.
j) Document entire process including dataset handling, model configuration, and evaluation results.

Prepare trained model for deployment, ensuring compatibility with target environment.

Table 1. Image Type with Data sets

Image Type	Data Set 1	Data Set 2	Data Set 3	Data Set 4
Benign	601	734	564	762
Primary Lung Cancer	745	626	627	938
Malignant	789	570	570	871
Suspicious malignant	670	670	670	990
Total	2805	2600	2431	3561

5. Result and Discussions

The deep learning-based CNN model developed for lung disease detection was evaluated using the LIDC-IDRI lung CT dataset, which was pre-processed, augmented, and split into training, validation, and testing sets. The performance of the model was measured based on its accuracy, precision, recall, F1-score, and confusion matrix for each class. During the training phase, the model achieved an accuracy of 93% on the training dataset, with a substantial reduction in validation loss and an improvement in classification accuracy over epochs. The model's validation performance was consistent, showing a validation accuracy of 91% after training, demonstrating its ability to generalize well on unseen data. The model's performance on the testing set achieved an overall accuracy of 89%, indicating that it effectively detects and classifies lung conditions from CT images. Precision, recall, and F1-score values were calculated for each class, like Benign Class: Precision of 0.91, Recall of 0.93, and F1-score of 0.92. Primary Lung Cancer Class: Precision of 0.87, Recall of 0.85, and F1-score of 0.86. Malignant Class: Precision of 0.89, Recall of 0.88, and F1-score of 0.88. Suspicious Malignant Class: Precision of 0.83, Recall of 0.81, and F1-score of 0.82.

The confusion matrix indicated that the model was particularly effective at distinguishing between benign and malignant classes, with few misclassifications in these categories. There was a slightly higher misclassification rate for the suspicious malignant class, highlighting an area for improvement. The use of data augmentation techniques such as shearing, zooming, and horizontal flipping helped address the class imbalance and significantly improved the model's performance, especially on minority classes like suspicious malignant and primary lung cancer. Regularization techniques, including dropout (20%) and batch normalization, contributed to reducing overfitting and enhancing the model's generalization ability. This was evident in the consistent performance across training, validation, and testing datasets. The model, trained with the specified batch size and learning rate, was optimized for real-time lung cancer detection in clinical settings. The accuracy and robustness of the model, along with the backend integration for processing CT scans, ensure that it can be deployed in a web-based interface for lung disease detection and staging.

6. Conclusion

In conclusion, this study demonstrates the potential of deep learning, specifically convolutional neural networks (CNNs), in automating the detection of pulmonary diseases in chest radiographs. By leveraging large, annotated datasets, our approach effectively identifies various abnormalities such as pneumonia, tuberculosis, lung cancer, and pulmonary edema. CNNs augment diagnostic accuracy, reduce interpretation time, and ultimately improve patient outcomes by enabling faster decision-making. As medical imaging continues to evolve, deep learning can play a pivotal role in advancing computer-aided diagnosis, offering substantial improvements in both efficiency and diagnostic precision in clinical practice. The model is capable of accurately classifying lung conditions into four categories and diagnosing lung cancer and related diseases. Future work will focus on further improving the model's performance on the suspicious malignant class and optimizing its deployment for clinical use..

References

1. R. Manser, A. Lethaby, L. B. Irving, C. Stone, G. Byrnes, M. J. Abramson, D. Campbell, “Screening for lung cancer,” Cochrane Database of Systematic Reviews, vol. 6, 2013.
2. M. Usman Ali, J. Miller, L. Peirson, D. Fitzpatrick-Lewis, M. Kenny, D. Sherifali, P. Raina, “Screening for lung cancer: A systematic review and meta-analysis,” Preventive Medicine, vol. 89, pp. 301–314, 2016.
3. American Cancer Society, “Key statistics about lung cancer,” Available: https://www.cancer.org/cancer/lung-cancer/about/key-statistics.html, Accessed: December 23, 2021.
4. S. Musafargani, K. K. Ghosh, S. Mishra, P. Mahalakshmi, P. Padmanabhan, B. Gulyas, “PET/MRI: A frontier in the era of complementary hybrid imaging,” European Journal of Hybrid Imaging, vol. 2, p. 12, 2018.
5. A. Asuntha, A. Srinivasan, “Deep learning for lung cancer detection and classification,” Multimedia Tools and Applications, vol. 79, pp. 7731–7762, 2020.
6. P. Mohamed Shakeel, M. A. Burhanuddin, M. Ishak Desa, “Lung cancer detection from CT images using improved profuse clustering and deep learning instantaneously trained neural networks,” Measurement, vol. 145, 2019.
7. Q. Z. Song, L. Zhao, X. K. Luo, X. C. Dou, “Using deep learning for classification of lung nodules on computed tomography images,” Journal of Healthcare Engineering, vol. 2017, Article ID 8314740, 7 pages, 2017.
8. G. Jakimovski, D. Davcev, “Using double convolutional neural network for lung cancer stage detection,” Applied Sciences, vol. 9, p. 427, 2019.
9. S. Makaju, P. W. C. Prasad, A. Alsadoon, A. K. Singh, A. Elchouemi, “Lung cancer detection using CT scan images,” Procedia Computer Science, vol. 125, 2018.
10. B. Devananda Rao, M. Arshad, “Early detection of lung cancer using machine learning techniques,” IEEE-ICCI Conference, 2023.
11. D. Gupta, S. Dawn, “Detection and staging of lung cancer from CT scan images by deep learning,” 2023 International Conference on Disruptive Technologies (ICDT), 2023.