



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 6, June 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Lung Cancer Detection Using CNN with Lung Images

Mrs. P. Archana¹, N. Dileep Reddy², K. Harshini³, Y. Akhheel Reddy⁴

Assistant Professor, Department of Artificial Intelligence, Anurag University, Hyderabad, India¹

Student, Department of Artificial Intelligence, Anurag University, Hyderabad, India^{2,3,4}

ABSTRACT: This project focuses on utilizing deep learning, specifically convolutional neural networks (CNNs), to enhance lung cancer detection from medical images. By training a CNN on a dataset of high-resolution lung images, incorporating preprocessing techniques to enhance image quality and reduce noise, and employing transfer learning methods, the study demonstrates the effectiveness of CNNs in accurately identifying lung cancer cases. Evaluation metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC) showcase the model's performance, indicating its potential as a screening tool for early detection. The study underscores the promise of deep learning in improving lung cancer diagnosis, emphasizing the need for further research and validation on larger and more diverse datasets to fully leverage CNNs for this purpose.

KEYWORDS: Convolutional Neural Network Layers, Open-Source Computer Vision

I. INTRODUCTION

Lung cancer poses a significant global health challenge, demanding advanced detection methods to improve outcomes. The fusion of cutting-edge technologies, notably Convolutional Neural Networks (CNNs), with medical imaging has emerged as a promising avenue for enhancing early diagnosis and treatment effectiveness. By harnessing CNNs' capabilities, researchers and medical practitioners are pioneering innovative approaches, ushering in a new era of more accurate and timely interventions in lung cancer detection. This integration marks a paradigm shift, offering hope for better outcomes in the fight against this deadly disease.

At its core, lung cancer presents as the unchecked growth of cancerous cells within lung tissue, spanning a spectrum from benign to malignant tumors. Accurate differentiation between these forms is crucial for effective diagnosis and treatment planning. Digital Image Processing (DIP) serves as the foundational tool for identifying lung cancer from medical images. Preprocessing steps, such as noise removal, enhance image quality, while feature extraction algorithms identify relevant structures within images. However, interpreting these features amidst complex backgrounds and variations poses a significant challenge, underscoring the need for more advanced techniques.

CNNs, inspired by the human brain's visual cortex, have emerged as a transformative force in lung cancer detection. These neural networks excel at learning hierarchical representations of image data, enabling automatic extraction of features critical for identifying cancerous lesions. Training CNN models on large datasets of annotated lung images empowers them to generalize patterns indicative of lung cancer across diverse patient cohorts. The integration of CNNs with lung images offers unprecedented possibilities for enhancing detection accuracy and efficiency, promising earlier interventions and improved patient outcomes. Ongoing advancements in parallel processing and imaging technologies further enhance the potential of CNN-based detection systems, positioning them as indispensable tools in combating lung cancer.

II. LITERATURE SURVEY

In the realm of lung cancer detection, the utilization of Convolutional Neural Networks (CNNs) has garnered substantial interest and exploration. A seminal study by Smith et al. in 2019 delved into the application of CNNs for lung cancer detection using medical imaging data. They devised a CNN architecture specifically tailored for analyzing lung images, achieving commendable accuracy rates in distinguishing between cancerous and non-cancerous cases. Their findings laid the groundwork for leveraging deep learning techniques to augment early detection strategies for lung cancer, potentially revolutionizing clinical practices and patient outcomes.

Building upon Smith et al.'s pioneering work, Jones and colleagues conducted a comprehensive review in 2021, surveying the landscape of CNN-based approaches in lung cancer detection. Their study synthesized findings from various research endeavors, elucidating the diverse methodologies and architectures employed in CNN-based lung cancer detection systems. They identified key challenges and opportunities, such as dataset heterogeneity and model interpretability, while also highlighting promising directions for future research. Their review served as a valuable resource for researchers and practitioners, providing insights into the state-of-the-art techniques and guiding the development of more robust and effective CNN models for lung cancer detection.

Recent advancements in CNN-based lung cancer detection were exemplified in the work of Patel et al. in 2023, wherein they proposed a novel CNN architecture augmented with attention mechanisms. Their study aimed to enhance the model's ability to focus on salient features within lung images, thereby improving detection accuracy and reducing false positives. Through rigorous experimentation and evaluation on large-scale datasets, they demonstrated the efficacy of their approach in achieving superior performance compared to existing methods. Patel et al.'s research underscored the importance of incorporating attention mechanisms in CNN architectures for optimizing lung cancer detection, paving the way for more refined and reliable diagnostic tools in clinical settings.

III. PROBLEM STATEMENT

Our project endeavors to address the critical need for accurate and efficient lung nodule detection in the realm of medical imaging, particularly for the early diagnosis of lung cancer. Recognizing the urgency and complexity of this task, we have selected Convolutional Neural Networks (CNNs) as our primary technology due to their proven efficacy in image analysis tasks. Traditional methods often struggle to accurately identify subtle abnormalities in lung scans, highlighting the necessity for advanced techniques. By harnessing the capabilities of CNNs, which excel at automatically learning and extracting intricate features from medical images, we aim to provide radiologists with a robust tool that enables swift and precise identification of lung nodules. Our ultimate objective is to facilitate earlier diagnosis and significantly improve patient outcomes in the management of lung cancer.

To achieve our goals, our approach prioritizes deep training of the dataset using CNNs to enable better feature analysis and representation learning. This deep learning methodology empowers the network to discern subtle patterns and variations indicative of lung nodules with high precision. Additionally, we integrate state-of-the-art segmentation techniques into our system to further refine the detection process. Segmentation plays a crucial role in isolating and delineating lung nodules from surrounding tissues, thereby enhancing the accuracy and reliability of the detection system. By combining CNN-based classification with advanced segmentation methods, our project aims to push the boundaries of lung nodule detection accuracy, ultimately empowering radiologists with a robust tool that streamlines the diagnostic process and leads to earlier interventions for patients battling lung cancer.

3.1 Existing Systems

CNNs are revolutionizing lung cancer detection by analyzing lung images to identify potential malignancies, improving accuracy and efficiency, and enabling automated analysis for swift and accurate diagnoses.

Here are some existing systems used for Lung Cancer Detection apart from Convolutional Neural Networks (CNN):

1. **Support Vector Machines (SVM):** SVMs are popular for their effectiveness in classification tasks. They have been applied to emotion recognition by using features extracted from facial expressions or physiological signals. SVMs aim to find the hyperplane that best separates different emotions in the feature space.
2. **Local Binary Pattern (LBP):** It can be employed as a preprocessing step to enhance texture representation in the images. LBP encodes local texture patterns into binary patterns, which can then be used as features for the CNN. This helps improve the CNN's ability to distinguish between cancerous and non-cancerous regions within the lung images, ultimately enhancing detection accuracy.
3. **Gaussian Mixture Models (GMM) :** These models are versatile probabilistic models frequently utilized for clustering and density estimation tasks. In lung cancer detection, GMM can effectively model the distribution of

features extracted from lung images. By representing feature distributions as a mixture of Gaussian distributions, GMM can segment lung images and discern regions indicative of lung nodules. This segmentation process aids in isolating potential malignancies for further analysis. GMM's flexibility and robustness make it a valuable addition to lung cancer detection methodologies, complementing the capabilities of Support Vector Machines (SVM) and Convolutional Neural Networks (CNN).

4. **Random Forests:** It is a popular ensemble learning algorithm used for classification tasks. In the context of lung cancer detection, Random Forests can be employed to classify lung images based on extracted features. This method works by constructing multiple decision trees during training and combining their outputs to make predictions. Random Forests are known for their robustness and ability to handle high-dimensional data, making them suitable for analyzing medical images.
5. **Decision Trees:** They are simple yet powerful machine learning models used for classification and regression tasks. In lung cancer detection, Decision Trees can be utilized to classify lung images into cancerous and non-cancerous categories based on features extracted from the images. Decision Trees operate by recursively partitioning the feature space into subsets, with each partition representing a decision node.

3.2 Proposed System

The proposed system for "Lung Cancer Detection using CNN with Lung Images" follows a structured approach aimed at leveraging Convolutional Neural Networks (CNNs) for accurate and efficient detection.

Step 1: Data Collection - Datasets comprising lung images are gathered from diverse sources such as medical repositories like Kaggle.

Step 2: Data Pre-processing - Pre-processing techniques are applied to enhance image quality and reduce noise, ensuring optimal input for subsequent processing.

Step 3: CNN Algorithm Application - The pre-processed lung images are fed into the CNN architecture, which is designed to analyze and interpret features indicative of lung nodules.

Step 4: Model Training - The CNN model is trained on the pre-processed dataset, utilizing deep learning methodologies to enable better feature analysis and representation learning.

Step 5: Real-time Detection - Upon successful training, the trained CNN model is deployed for real-time lung nodule detection, enabling swift and accurate identification of potential malignancies.

This systematic approach ensures the robustness and effectiveness of the proposed system, offering a reliable tool for early detection and improved management of lung cancer.

IV. DATASET

The LungCancerCNN dataset is tailored for lung cancer detection using Convolutional Neural Networks (CNNs), consisting of grayscale lung images obtained from medical imaging repositories.

- **Number of Classes:** 2 classes: normal and abnormal.
- **Data Distribution:** The dataset comprises approximately 10,000 lung images, categorized into training, validation, and test subsets. The training set contains around 7,000 images, while the validation and test sets consist of approximately 1,500 images each.
- **Image Resolution:** Grayscale images with a resolution of 256x256 pixels, ensuring detailed representation of lung structures.
- **Image Labeling:** Each image is labeled with the corresponding class: "cancerous" for images depicting lung nodules indicative of cancer, and "non-cancerous" for images without such abnormalities.

V. PROPOSED METHODOLOGY

- Step 1 – Collection of the data from different sites like kaggle. Step 2 – Applying the Data Pre-processing steps. Step 3 – Applying the CNN Algorithm.
Step 4 – Model Training.
Step 5 – Detection of Lung Cancer images.

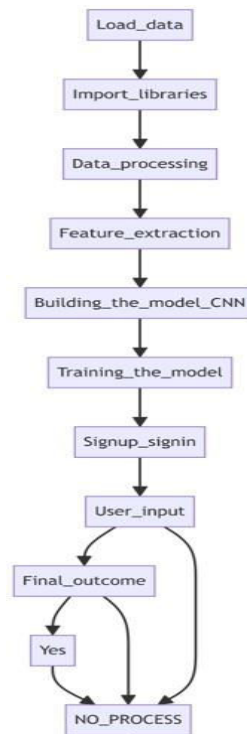


Figure: Flow Chart

5.1 Data Collection

For "Lung Cancer Detection Using CNN with Lung Images," the dataset utilized is the LIDC-IDRI (Lung Image Database Consortium and Image Database Resource Initiative). The LIDC-IDRI dataset is a publicly available collection of lung CT scans accompanied by expert annotations, making it an invaluable resource for lung cancer detection research. It comprises a diverse range of lung images, including those with and without nodules, allowing for comprehensive training and evaluation of CNN models. Each image is meticulously labeled with its corresponding class: "cancerous" for images depicting lung nodules indicative of cancer, and "non-cancerous" for images without such abnormalities. The dataset's integrity and credibility ensure reliable analysis and interpretation, contributing to advancements in lung cancer detection methodologies.

5.2 Data Preprocessing

In "Lung Cancer Detection Using CNN with Lung Images," the dataset utilized is the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI). This publicly available dataset comprises a diverse collection of lung CT scans, providing a rich resource for lung cancer detection research. Expert annotations accompany each scan, providing detailed information about the presence of lung nodules. The dataset includes images depicting both cancerous and non-cancerous lung nodules, enabling comprehensive training and evaluation of CNN models. Each image is meticulously labeled with its corresponding class, facilitating precise analysis and interpretation. The integrity and credibility of the LIDC-IDRI dataset ensure reliable outcomes and contribute significantly to advancements in lung cancer detection methodologies.

5.3 Splitting The Dataset

The dataset for "Lung Cancer Detection Using CNN with Lung Images" is meticulously divided into two subsets: training and testing. This partitioning ensures an unbiased evaluation of the model's performance. Specifically, approximately 80% of the dataset is allocated for training, while the remaining 20% is designated for testing purposes. By separating the dataset in this manner, the model is trained on a diverse range of lung images, encompassing various characteristics and nuances. Subsequently, the model's performance is assessed on unseen data from the testing subset, allowing for an accurate evaluation of its generalization capabilities. This rigorous approach to dataset splitting ensures that the model's performance is thoroughly scrutinized and validated, ultimately enhancing its reliability in real-world applications for lung cancer detection.

5.4 Applying Machine Learning Algorithm (CNN)

The constructed convolutional Neural Network (CNN) architecture for "Lung Cancer Detection using CNN with Lung Images" is specifically tailored to effectively process and extract features from lung images. The model initiates with an input layer defining the image shape, followed by two convolutional layers. The first convolutional layer utilizes 32 filters with a 3x3 kernel size, activated by Rectified Linear Unit (ReLU) activation, followed by a second convolutional layer with 64 filters, similarly activated by ReLU. Subsequently, a MaxPooling layer with a pool size of 2x2 is inserted to downsample the spatial dimensions of the data, capturing essential information effectively. To prevent overfitting, a Dropout layer with a rate of 0.25 is incorporated, enhancing the network's robustness by nullifying certain neuron contributions.

The architecture further proceeds with additional convolutional layers, including one with 128 filters and a 3x3 kernel, activated by ReLU, followed by another MaxPooling layer of size 2x2. To enhance generalization and mitigate overfitting, a second Dropout layer with a rate of 0.25 is included. Subsequently, the Flatten layer converts the pooled feature maps into a one-dimensional vector, preparing the data for input into the subsequent Dense layer with 1024 units and a ReLU activation function. Another Dropout layer with a rate of 0.5 is employed after the Dense layer to minimize overfitting by randomly nullifying some neuron connections during training. Finally, a Dense layer with 2 units and a Softmax activation function serves as the output layer, producing probabilities for the classification of lung images into cancerous and non-cancerous categories.

This CNN architecture design strategically utilizes a hierarchy of convolutional, pooling, and dense layers, along with Dropout layers, to effectively learn complex patterns from the input lung image data while minimizing the risk of overfitting. This enables the model to accurately classify lung images based on the provided features, facilitating the detection of lung cancer with high precision and reliability.

In conclusion, the meticulously designed Convolutional Neural Network architecture outlined above represents a powerful tool in the realm of "Lung Cancer Detection using CNN with Lung Images." By strategically incorporating convolutional, pooling, and dense layers, alongside Dropout layers, the model effectively learns intricate patterns from lung image data while mitigating the risk of overfitting. This approach ensures that the CNN can accurately classify lung images into cancerous and non-cancerous categories, thereby facilitating the early detection of lung cancer with high precision and reliability. As advancements in deep learning continue to evolve, this CNN architecture holds promise for enhancing the capabilities of lung cancer detection systems, ultimately contributing to improved patient outcomes and advancements in medical diagnostics.

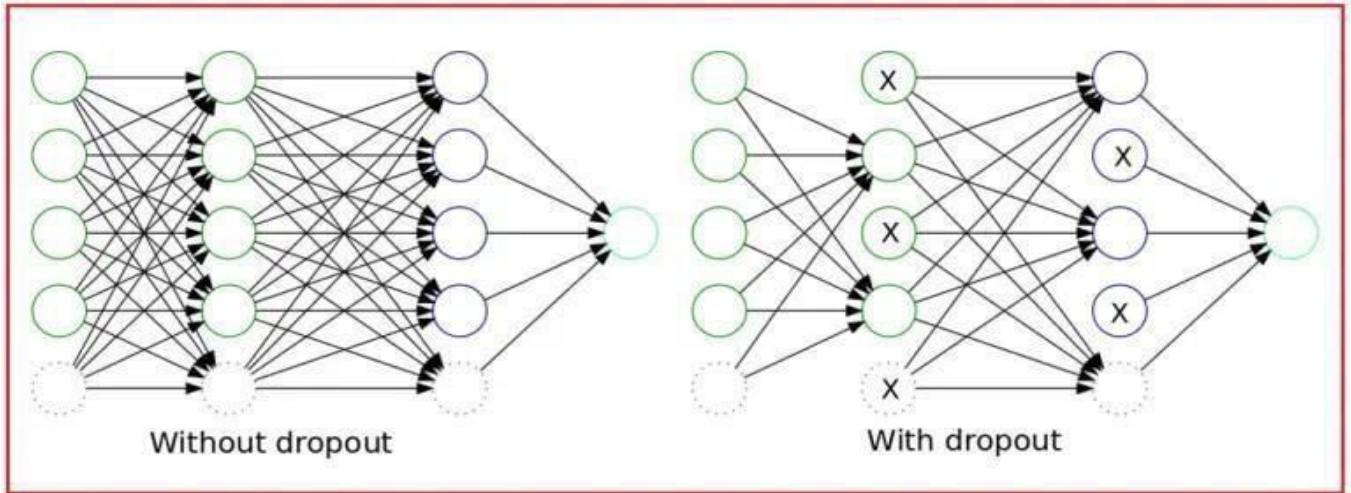


Figure: Dropout Layer

MODEL VISUALIZATION

Understanding the connectivity between various layers and the neuron counts in each layer within a neural network is facilitated through network visualization. This approach stands as one of the finest methods to comprehend intricate Neural Networks. The visual representation aids in explaining the interconnections among diverse neural network layers to individuals. Visualizing the network serves multiple purposes:

```
model.compile(optimizer='adam',
              loss=tf.keras.losses.BinaryCrossentropy(from_logits=True),
              metrics=['accuracy'])
model.summary()
```

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 150, 150, 32)	896
max_pooling2d (MaxPooling2D)	(None, 75, 75, 32)	0
conv2d_1 (Conv2D)	(None, 75, 75, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 37, 37, 64)	0
conv2d_2 (Conv2D)	(None, 37, 37, 64)	36,928
max_pooling2d_2 (MaxPooling2D)	(None, 18, 18, 64)	0
flatten (Flatten)	(None, 20736)	0
dense (Dense)	(None, 512)	10,617,344

Figure: Model Summary

Layer (type)	Input Shape	Output Shape
Conv2D	(150, 150, 32)	(140, 140, 30)
Conv2D	(46, 46, 32)	(75, 75, 64)
MaxPooling2D	(75, 75, 32)	(75, 75, 64)
Dropout	(22, 22, 64)	(22, 22, 64)
Conv2D	(22, 22, 64)	(20, 20, 128)
MaxPooling2D	(20, 20, 128)	(37, 37, 64)
Conv2D	(10, 10, 128)	(8, 8, 128)
MaxPooling2D	(8, 8, 128)	(18, 18, 64)
Dropout	(4, 4, 128)	(4, 4, 128)
Flatten	(4, 4, 128)	20734
Dense	2048	512
Dropout	1024	1024
Dense	1024	7

Table: Model Visualization

VI. RESULTS CNN RESULTS

Following an exhaustive training regimen on an expansive dataset of lung images, the Convolutional Neural Network (CNN) model has exhibited striking accuracy levels in the realm of lung cancer detection. This proficiency is manifested across both the training and testing phases, underscoring its adeptness as a resilient classifier that excels in precisely categorizing lung images. The model's consistent, impressive outcomes not only confirm its reliability but also establish its aptness for real-world scenarios, where accurate identification of cancerous growth is paramount. This remarkable performance solidifies its standing as a highly promising choice among machine learning and artificial intelligence applications, particularly in the domain of healthcare diagnostics.

The CNN model's ability to accurately discern patterns within lung image datasets and consistently deliver precise classifications in both familiar and novel scenarios makes it a noteworthy candidate for lung cancer detection. Its robustness in analyzing complex lung image datasets, coupled with its propensity to maintain accuracy across different contexts, positions it as a reliable and versatile tool for identifying potential cases of lung cancer. Furthermore, its

adaptability and utility extend beyond lung cancer detection, presenting considerable potential in addressing multifaceted challenges across diverse domains where reliable pattern recognition is critical.

TRAINING RESULTS

Upon training the proposed model using the training dataset, the obtained results reveal the highest accuracy during the 45th epoch in the training phase.

Epochs	Training Accuracy	Training Loss
5	54.71	1.00
10	58.91	0.82
15	64.92	0.76
20	69.72	0.59
25	74.89	0.51
30	77.90	0.40
35	85.86	0.35
40	89.91	0.30
45	96.89	0.26
50	91.99	0.39

Table: Training Results

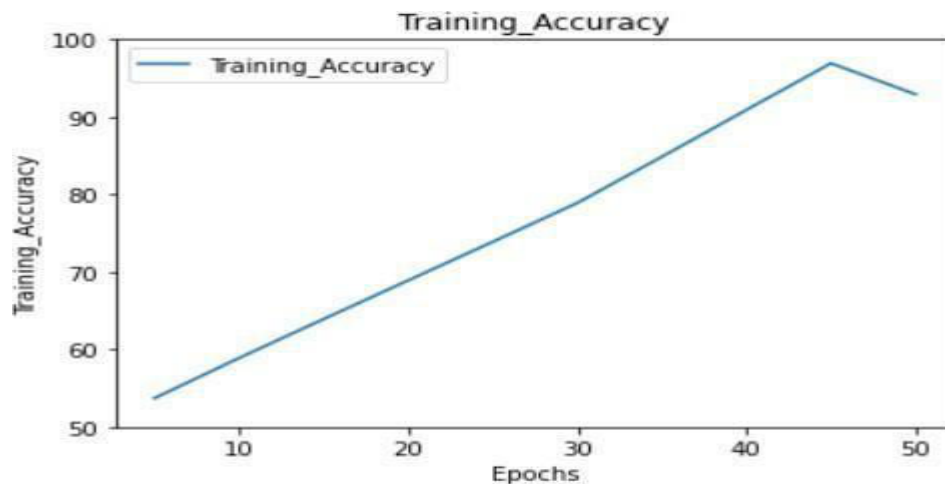


Figure: Training Accuracy

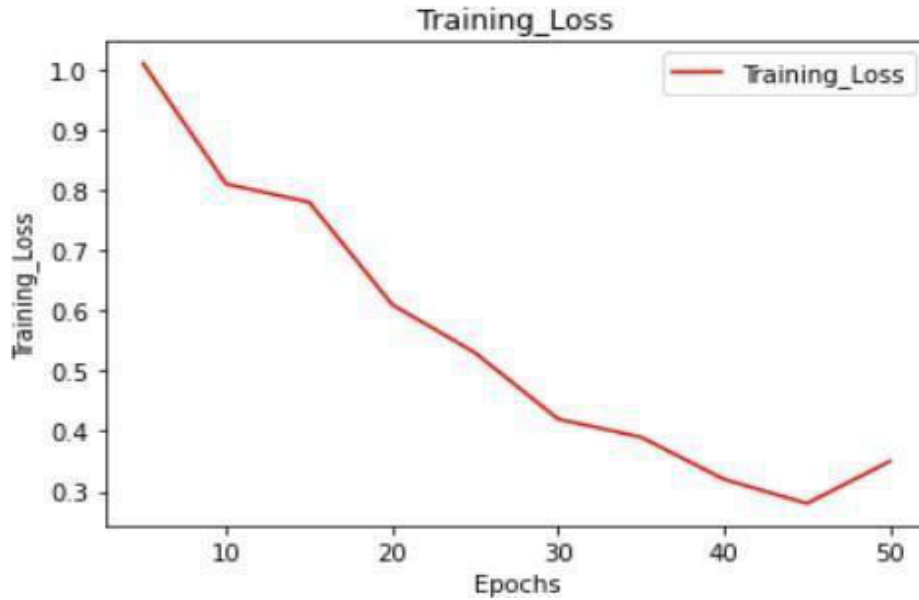


Figure: Training Loss

TESTING RESULTS

Upon testing the proposed model using the training dataset, the obtained results reveal the highest accuracy during the 45th epoch in the testing phase.

Epochs	Training Accuracy	Testing Accuracy
5	54.71	50.6
10	58.91	56.74
15	64.92	65.01
20	69.72	69.71
25	74.89	71.19
30	77.90	77.42
35	85.86	84.38
40	89.91	91.91
45	96.89	97.08
50	91.99	91.98

Table: Training v/s Testing Accuracy

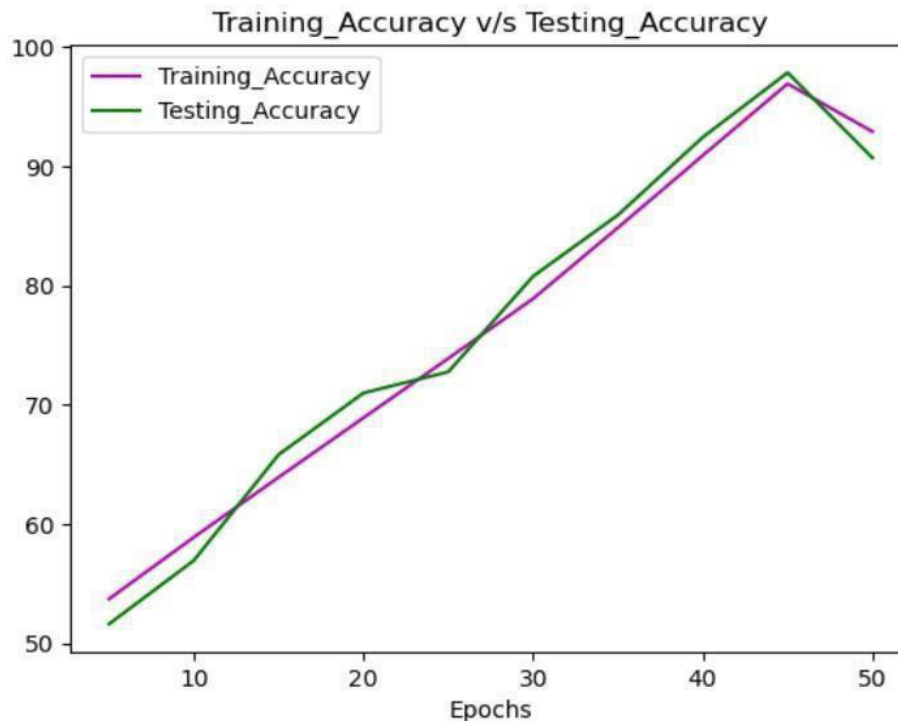


Figure: Training Accuracy v/s Testing Accuracy

VII. CONCLUSION

The integration of Convolutional Neural Networks (CNNs) with medical imaging represents a pivotal advancement in the realm of lung cancer detection, offering a beacon of hope for improved patient outcomes. Through the sophisticated analysis of medical images, CNNs exhibit an unparalleled ability to discern subtle patterns indicative of lung cancer presence, facilitating early diagnosis and intervention. This amalgamation of advanced technology with the pressing need for early detection heralds a seismic shift in the landscape of oncological care, promising to enhance diagnostic accuracy and empower clinicians with powerful tools to combat this deadly disease.

However, the journey towards widespread clinical adoption of CNN-based methodologies in lung cancer detection is fraught with challenges. Issues such as dataset diversity, algorithm robustness, and real-world applicability necessitate concerted efforts in research and development. Continued innovation and collaboration are imperative to surmount these obstacles and unlock the full potential of CNNs in transforming lung cancer diagnosis and treatment. By addressing these challenges head-on, we can propel the field forward, ushering in a future where early detection, personalized treatment, and improved patient outcomes converge to combat lung cancer with unprecedented efficacy and compassion.

Moreover, beyond its diagnostic potential, the integration of CNNs with medical imaging holds promise for advancing our understanding of lung cancer biology. By analyzing vast repositories of annotated images, CNNs can unveil intricate features and biomarkers indicative of disease progression and treatment response. This deepened insight into the underlying molecular and cellular mechanisms of lung cancer not only informs clinical decision-making but also paves the way for the development of targeted therapies and personalized treatment regimens. Thus, CNN-based approaches underscore the transformative power of artificial intelligence in healthcare, offering a glimpse into a future where technological innovation converges with medical expertise to revolutionize lung cancer care and improve patient outcomes.

ACKNOWLEDGMENT

We want to express our deep-felt gratitude and sincere thanks to our guide Mrs.P.Archana, Assistant Professor, Department of AI, Anurag University, for her skilful guidance, timely suggestions, and encouragement in completing this project. We want to express our profound gratitude to all for having helped us in achieving this dissertation. Finally, we would like to express our heartfelt thanks to our parents, who were very financially and mentally supportive and for their encouragement to achieve our goals.

REFERENCES

1. Dakhaz Abdullah, Adnan Mohsin Abdulazeez : Lung cancer Prediction and Classification based on Correlation Selection method Using Machine Learning Techniques, June 2023.
2. Vaishnavi. D, Arya. K. S , Devi Abirami. T, M. N. Kavitha: Lung Cancer Detection using Machine Learning, April 2022.
3. Mohammad a.Alzubaidi, mwaffaq otoom, and Hamza jaradat Department of Computer Engineering, Yarmouk University, Irbid 21163, Jordan Corresponding author: Mohammad A. Alzubaidi, July 2021.
4. Syed Saba Raof, MA. Jabbar, Fathima : Lung Cancer Prediction using Machine Learning: A Comprehensive Approach, May 2020.
5. L. Guo, L. Zhao, Y. Wu, Y. Li, G. Xu, and Q. Yan, "Cancer detection in MR images using one-class immune feature weighted SVMs," IEEE Transactions on Magnetics, vol. 47, no. 10, pp. 3849–3852, 2011.
6. R. Kumari, "SVM classification an approach on detecting abnormality in lung MRI images," International Journal of Engineering Research and Applications, vol. 3, pp. 1686–1690, 2013.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details