IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Automated Lung Cancer Detection Using Image Processing

Dr. Shridevisoma [1], kaniyar fatima [2], misba anjum [3], vijaylaxmi [4]

Dept. Of Computer Science and Design

Poojya Doddappa Appa College of Engineering, Kalabuargi, India

ABSTRACT

The domain of medical image processing focuses on the early detection of lung cancer using advanced image processing techniques. Currently, lung cancer diagnosis primarily depends on manual analysis of chest X-rays or CT scans by radiologists. While experienced professionals can identify abnormalities, the process is time-consuming, prone to human error, and often limited by subtle signs that are difficult to interpret in early stages. Existing computer-aided detection systems are available, but many suffer from low accuracy, high false-positive rates, and limited generalization to diverse datasets. The system is designed to process standard datasets of lung images, performing key steps such as image preprocessing to enhance quality, feature extraction to identify relevant patterns, and classification to accurately distinguish between cancerous and non-cancerous images. The dual-phase approach consists of a training phase, where the CNN model learns from labeled data, and a testing phase, where the model is validated with unseen images to measure its performance. The application of CNNs allows the model to learn complex features that are difficult to capture through traditional handcrafted techniques, leading to improved accuracy and reduced false positives. Overall, integrating advanced image processing techniques with machine learning models such as CNN offers promising improvements in lung cancer diagnosis. This approach not only accelerates the diagnostic workflow but also supports healthcare professionals by providing consistent and reliable analysis of medical images. The proposed system has the potential to improve patient outcomes by facilitating timely detection and treatment, ultimately contributing to more effective management of lung cancer on a broader scale.

Keywords-

Lung Cancer Detection, CT Imaging, Image Processing, CNN, Medical Diagnosis

I. INTRODUCTION

Lung cancer remains one of the most common causes of cancer-related deaths globally. Its early-stage symptoms are often subtle, making timely and accurate detection challenging. Lung cancer is one of the leading causes of cancer-related deaths worldwide due to its aggressive nature and late-stage diagnosis. It occurs when abnormal cells in the lungs grow uncontrollably, forming malignant tumors that interfere with normal lung function. Lung cancer can be broadly categorized into small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC), with NSCLC being the most common type. Recent advancements in medical imaging and artificial intelligence offer promising avenues to assist radiologists with computer-aided diagnosis tools. CT imaging is widely used for lung cancer screening due to its high-resolution detail. However, interpreting CT scans requires expert knowledge and is subject to variability. This study proposes an automated system combining image processing and deep learning techniques to identify lung nodules and classify malignancy effectively.

Lung cancer accounts for a significant number of cancer-related fatalities globally. The complexity and subtlety of early-stage symptoms often make it difficult for medical professionals to diagnose the disease in its early phases. This paper introduces a system that leverages advanced image processing and deep learning techniques to accurately detect and classify lung nodules in CT scan images. By using CNN-based feature extraction and classification methods, and integrating the model within a web-based platform via Flask, we aim to offer a deployable, scalable, and user-friendly solution for hospitals and diagnostic labs.

1.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) have emerged as one of the most powerful deep learning architectures for analyzing visual data, particularly in the field of medical imaging. Designed to automatically and adaptively learn spatial hierarchies of features from input images, CNNs are especially well-suited for detecting patterns, textures, and anomalies that may not be immediately obvious to human observers. In recent years, CNNs have demonstrated significant success in diagnosing various diseases through radiographic image analysis, including diabetic retinopathy, brain tumors, and most notably, lung cancer. Lung cancer, being one of the leading causes of cancer-related mortality worldwide, necessitates accurate and early diagnosis for effective treatment and improved survival rates. Traditional diagnostic approaches, while reliable, are time-consuming and require expert interpretation. The integration of CNN-based systems into medical diagnostics offers a means to enhance accuracy, reduce workload, and provide consistent analysis.

This paper explores the application of CNNs in detecting lung cancer from CT scan images. The model architecture is designed to automatically extract and learn discriminative features from medical images, distinguishing between cancerous and non-cancerous tissues. The work includes preprocessing, model training, and evaluation phases aimed at demonstrating the potential of CNNs in automated lung cancer diagnosis.

EXISTING SYSTEM

The existing lung cancer detection systems in healthcare heavily rely on manual diagnostic methods such as visual examination of CT scans and other imaging modalities by trained radiologists. While these methods are clinically accepted and widely used, they are time-consuming, prone to human error, and often limited by the subjective judgment of the medical expert. In regions with limited healthcare infrastructure or shortage of specialists, early and accurate diagnosis becomes even more challenging. Moreover, traditional machine learning models, if used, require manual feature extraction and often lack the precision needed for real-world applications.

III. PROPOSED METHODOLOGY

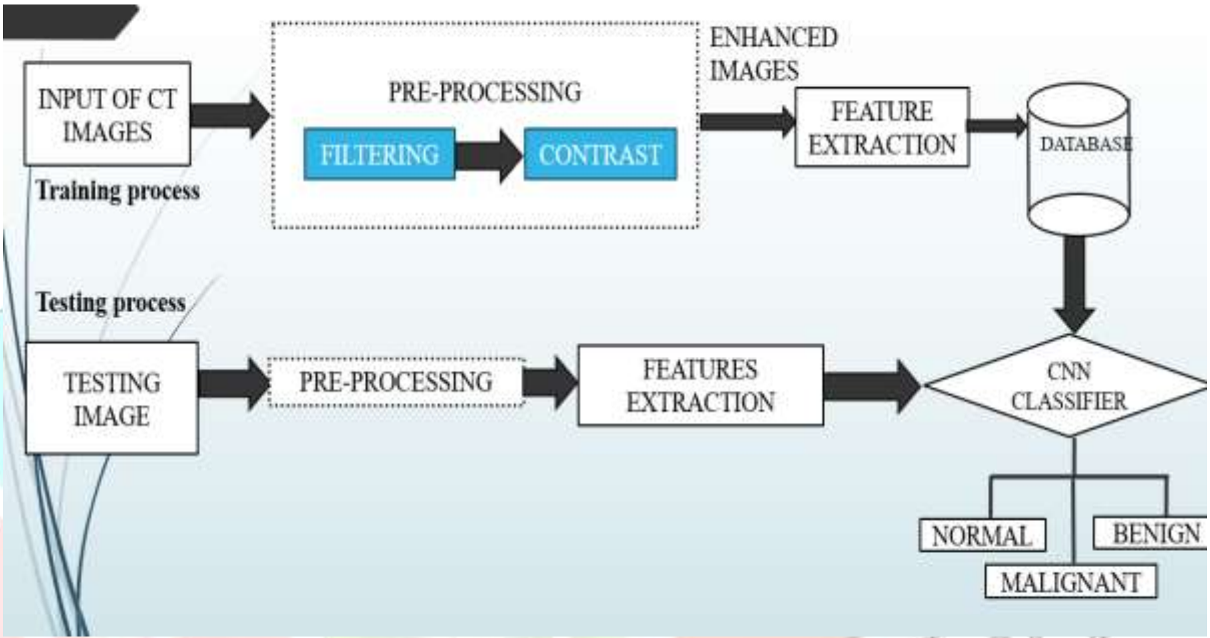

Fig 2.1 System Architecture

Training process

The training process is a critical phase in the development of a CNN model for lung cancer detection. It involves preparing medical image data, extracting meaningful features, and using these features to train a classification model that can distinguish between cancerous and non-cancerous cases. Below is a comprehensive explanation of each phase involved:

1. Input Stage:

The first step in the lung cancer detection process begins with the input stage, where CT (Computed Tomography) scan images of the lungs are collected. These images are typically acquired using medical imaging equipment in hospitals or diagnostic centers. CT scans provide detailed cross-sectional views of lung tissues, making them ideal for identifying abnormalities such as nodules or tumors. The system accepts these CT scan images as input, either in raw form or in a standard format such as DICOM. This stage is crucial because the accuracy of the entire system depends heavily on the quality and clarity of the input images. Once collected, the images are fed into the system where they undergo pre-processing for enhancement and noise reduction to prepare them for deeper analysis.

a) CT (Computed Tomography) scan images of the lungs are collected.
b) The system starts by taking CT scan images as input.
c) These images can be from a medical imaging system and will be processed for further analysis.

2. Pre-Processing Stage:

In this stage, the input CT images undergo a series of **pre-processing techniques** aimed at improving image quality and enhancing the visibility of important features. The main goal is to prepare the images for accurate feature extraction and classification in later stages. Pre-processing involves two primary operations:

a) Filtering
b) Contrast Enhancement

a) **Filtering**: Filtering is applied to reduce noise and improve image clarity. In medical imaging, noise can arise from various sources, including the imaging device and patient movement. To address this, a **Gaussian filter** is commonly used. This method smooths the image by reducing high-frequency noise without significantly blurring the important structures like lung nodules. The Gaussian filter helps retain the edges of anatomical features while suppressing irrelevant background variations.

- Purpose: Reduce noise and improve image clarity.
- Method: Gaussian Filter to smooth the image and remove high-frequency noise.

b) Contrast Enhancement:
follows filtering to improve the distinction between different tissues in the lung. One effective method used is **Histogram Equalization**, which enhances the global contrast of the image by redistributing pixel intensity values. This makes subtle features, such as early-stage nodules or lesions, more visible. As a result, both normal and abnormal areas in the lung scan become more easily identifiable to both the algorithm and medical professionals.

- Purpose: Enhance contrast to make critical features more distinguishable.
- Method: Histogram Equalization , Improves global contrast by redistributing pixel intensity values.

3. Feature Extraction Stage:

The feature extraction stage is a critical step where the system identifies and extracts meaningful information from the pre-processed CT images. This stage involves analyzing the texture, shape, intensity, and structural patterns within the lung images to distinguish between healthy and abnormal tissues. For example, cancerous nodules often have irregular shapes, uneven edges, and distinctive texture characteristics that set them apart from benign or normal lung tissue.

Feature extraction can be performed manually using traditional image processing techniques, or automatically using deep learning methods such as Convolutional Neural Networks (CNNs). CNNs are especially effective as they learn to extract features hierarchically—from simple edges in early layers to complex patterns in deeper layers. The extracted features are then used as input for classification algorithms that predict whether a given image is cancerous or non-cancerous. By focusing on the most relevant aspects of the image, the feature extraction process significantly enhances the model's ability to make accurate and reliable diagnoses.

a) Extracts relevant features from the pre-processed images.
b) This step involves analyzing the image's texture, shape, and intensity patterns.

These features help in distinguishing between normal and abnormal tissues.

4. Database :

After the images are pre-processed and enhanced, they are stored in a centralized **Image Database**. This database serves as a repository of standardized images that are ready for analysis and feature extraction. Storing the enhanced images in a structured database allows for efficient data retrieval, comparison, and scalability. It also ensures that image data is preserved in a consistent format, which is important for training machine learning models, performing statistical evaluations, and maintaining a record of patient scans. Additionally, the image database can be linked with metadata such as patient ID, scan date, and diagnostic outcomes for comprehensive study and validation.

5. CNN CLASSIFIER

CNN Model Architecture

The CNN model includes:

- Input Layer: Accepts 224×224 grayscale images.
- Convolutional Layer 1: 32 filters, 3×3 kernel, ReLU activation
- MaxPooling Layer 1: 2×2 window
- Convolutional Layer 2: 64 filters, ReLU activation
- MaxPooling Layer 2: 2×2
- Flatten Layer
- Dense Layer 1: 128 neurons, ReLU
- Dropout Layer: 0.5 (for regularization)
- Output Layer: Softmax for multi-class classification (normal, benign, malignant)

Testing process

The testing process evaluates the performance of the trained lung cancer detection model on unseen CT scan images. It follows a similar pipeline as the training phase but with the focus on model inference and accuracy assessment rather than learning. Below is a comprehensive breakdown of each stage involved in testing:

1. Input of Testing Image

a) CT (Computed Tomography) scan testing images of the lungs are collected.
b) The system starts by taking CT scan testing images as input.
c) These images can be from a medical imaging system and will be processed for further analysis.

2. Pre-Processing Stage

input testing CT images undergo pre-processing to improve quality and enhance features.

The pre-processing includes:

c) **Filtering**: It enhances the image clarity.
d) **Contrast Enhancement**: Adjusts the brightness and contrast of the image to highlight important features.

a) **Filtering**

- Purpose: Reduce noise and improve image clarity.
- Method: Gaussian Filter to smooth the image and remove high-frequency noise.

b) Contrast Enhancement

Purpose: Enhance contrast to make critical features more distinguishable.

Method: Histogram Equalization, Improves global contrast by redistributing pixel intensity values.

3. Feature Extraction

a) The system extracts key features from the testing image, similar to the training images.
b) These extracted features are then sent to the classification model.

4. Classification Using Trained Model

a) The classifier (which has been trained on previous CT images) analyzes the extracted features.
b) It compares the features of the testing image with the learned patterns.
c) The classifier then predicts the category of the image:

- Normal – No abnormality detected.
- Benign – A non-cancerous tumor.
- Malignant – A cancerous tumor requiring medical intervention.

The testing process is a vital phase in validating the accuracy, robustness, and reliability of the lung cancer detection system. By applying the same pre-processing and feature extraction techniques used during training, the system ensures consistency and minimizes errors due to data discrepancies. Once processed, the testing images are evaluated by the trained classification model, which effectively distinguishes between normal, benign, and malignant cases.

MODEL ARCHITECTURE

The model architecture for lung cancer detection using CNN is designed to automatically learn and extract spatial features from medical images for accurate classification. Typically, the architecture starts with multiple convolutional layers that apply filters to the input brain scans to detect features such as edges, textures, and patterns associated with hemorrhages. These layers are followed by activation functions like ReLU and pooling layers (e.g., max pooling) to reduce spatial dimensions and computational complexity while retaining important features. The deeper layers of the network capture more abstract representations of the data. After feature extraction, the output is flattened and passed through fully connected (dense) layers, which combine the features to make a final prediction. The last layer uses a softmax or sigmoid activation function depending on whether the task is multi-class or binary classification. Alternatively, pre-trained models like ResNet, VGG, or DenseNet can be used for transfer learning, where the final layers are fine-tuned for hemorrhage classification, leveraging the rich feature representations learned from large-scale image datasets.

The below Figure 2.2 shows a simple structure of a Convolutional Neural Network (CNN) used for detecting lung cancer from CT or MRI scans. The process starts with the input image, which is passed through a series of convolution layers that automatically learn important features, like edges or patterns, from the image. These features help the model understand where a cnacer might be present. Next, the image goes through a pooling layer, which reduces the size of the data while keeping the most important information, making the model faster and more efficient. After that, the important features are flattened and passed through fully connected layers, which help the model make a decision based on the learned patterns.

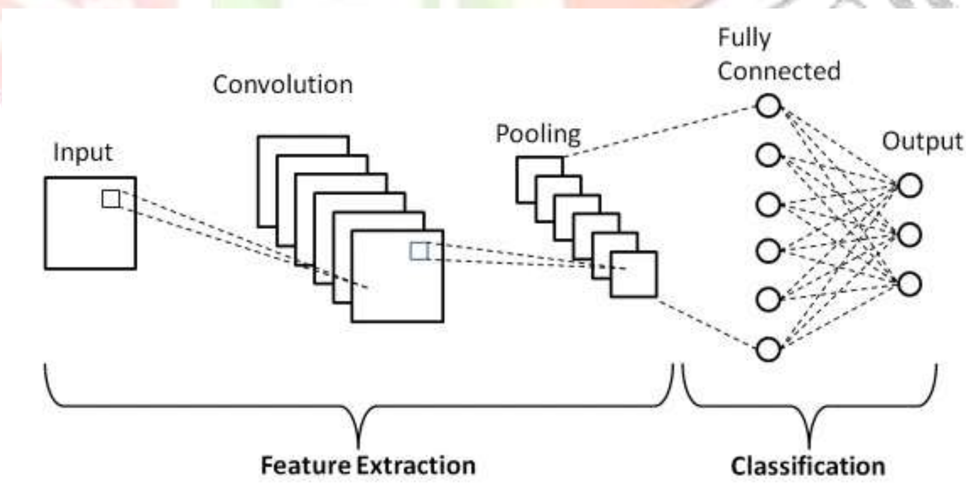

Figure 2.2 Model Architecture

Finally, the output layer gives the result, such as whether a lung cancer is present and what type it is. This whole process allows the CNN to detect lung cancer accurately by learning directly from medical images.

TABLE. MODEL TRAINING SUMMARY

Component	Details
Datasets	LIDC-IDRI, Kaggle Lung Cancer Dataset
Tools	Python, OpenCV, TensorFlow/Keras, NumPy, Matplotlib
Models	CNN (Convolutional Neural Network), ResNet-based Architecture
Framework	Supervised Learning using Deep CNN with Image Preprocessing
Runtime Support	GPU/CPU Training Environment

EXPERIMENTAL RESULTS AND DISCUSSION

Initial training was conducted on a subset of annotated CT images. The following performance metrics were recorded:

- **Training Accuracy:** ~92%
- **Validation Accuracy:** ~88%
- **Observations:**
  - Proper normalization and data augmentation improved generalization.
  - Misclassifications were mainly due to unclear tumor boundaries or poor image contrast.

IV. CONCLUSION AND FUTURE WORK

This research demonstrates the feasibility of using a CNN-based image processing pipeline for lung cancer detection. The current implementation covers model design and training, showing encouraging classification performance on benchmark datasets. Future work will expand to include segmentation techniques, larger datasets, real-time diagnostics, and integration into clinical workflows. The model's deployment as a decision support tool could significantly enhance diagnostic accuracy and early detection rates.

Future Work

Future enhancements may include:

- Testing with larger and multi-institutional datasets
- Incorporating segmentation techniques (e.g., U-Net)
- Experimenting with transfer learning using more recent models like ResNet or EfficientNet
- Deploying the model in cloud-based environments for scalability
- Integration with hospital systems for clinical use

ACKNOWLEDGMENTS

The authors would like to thank the Department of Computer Science and Design, PDA College of Engineering, for providing guidance and resources for the completion of this project. Special thanks to faculty mentors and peers for their continuous encouragement and support throughout the development of this system.

REFERENCES

[1] M. R. Hasan and M. A. Kabir, "Lung cancer detection and classification based on image processing and statistical learning," Int. J. Med. Imaging Comput. Vis., vol. 1, no. 1, pp. 1–7, 2019.

[2] R. Y. Bhalerao, H. P. Jani, R. K. Gaitonde, and V. Raut, "A novel approach for detection of lung cancer using digital image processing and convolution neural networks," in Proc. 5th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS), Coimbatore, India, 2019, vol. 2, pp. 577–583.

[3] J. L. Causey et al., "Lung cancer screening with low-dose CT scans using a deep learning approach," IEEE Trans. Med. Imaging, vol. 38, no. 4, pp. 899–909, Apr. 2019.

[4] J. L. Causey et al., "Lung Cancer Screening with Low-Dose CT Scans Using a Deep Learning Approach," IEEE Transactions on Medical Imaging, vol. 38, no. 4, pp. 899–909, 2019.

[5] V. J. Pawar, K. D. Kharat, S. R. Pardeshi, and P. D. Pathak, "Lung cancer detection system using image processing and machine learning techniques," Int. J. Adv. Trends Comput. Sci. Eng. (IJATCSE), vol. 9, no. 4, pp. 5956–5963, 2020.

[6] A. Gupta, V. L. K. Manda, and B. I. Seraphim, "Lung cancer detection using image processing and CNN," Int. J. Adv. Res. Innov. Ideas Technol. (IJARIIT), vol. 7, no. 3, pp. 1232–1236, 2021.

[7] V. Shariff, C. Paritala, and K. M. Ankala, "Advances in lung cancer diagnosis with medical imaging and deep learning techniques: Challenges and opportunities," J. Theor. Appl. Inf. Technol. (JATIT), vol. 101, no. 17, pp. 7083–7095, 2023.

[8] M. Rane, P. Chawla, U. Dandekar, A. Dhakane, P. Dattawade, and A. Daberao, "Lung cancer detection," J. Chem. Health Risks, vol. 14, no. 3, pp. 177–186, 2024.

[9] A. Akbayeva, A. Shilikbay, A. Toleubekova, and Zhumagaliyeva, "Lung cancer detection on CT images using image processing," Int. J. Med. Imaging, vol. 12, no. 2, pp. 55–60, 2024.