Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Segmentation Based Technique for Detection of Lung Cancer using Deep Learning

Dileep Kumar Verma[1*], Prateek Singh[2]

[1*,2]Sam Higginbottom University of Agriculture, Technology and Sciences, India

ARTICLE INFO

Received: 08 Dec 2024

Revised: 15 Dec 2024

Accepted: 21 Dec 2024

ABSTRACT

The identification of lung cancer has emerged as a significant problem in medical research in the past few decades. This study used 8000 CT lung pictures of lung images, categorized into four separate classes: adenocarcinoma, Large-cell carcinoma, normal and squamous cell carcinoma. Deep learning algorithms designed for lung cancer detection are evaluated against five classifier methods employing the SKlearn (Scikit-learn) framework: Single Layer CNN, Multi-layer CNN, VGG16, ResNet50, and Artificial Neural Network (ANN). The outcomes of the suggested five classifiers have been analysed. A model based on a convolutional neural network (CNN) was developed using the ReLU activation functions. Construct a fundamental CNN model and evaluate it with regularization and augmentation methods to enhance accuracy. The accuracy of the Multi-layer CNN is 99%. Logistic regression yields 80%, whereas Single Layer CNN, VGG16 and ResNet50 achieved 97%. We got the least accuracy with 39% from ANN. In the comparison of deep learning algorithms with Multi Layer CNN, the latter had a superior accuracy of 99% in the detection of cancer in CT lung pictures.

Keyword: Lung Cancer, Deep Learning Algorithm, Segmentation, Feature Extraction, Classification

1. Introduction

1. Cancer is categorised as a hazardous disease with elevated fatality rates. Lung cancer has the highest fatality rate among all malignancies and is considered the most dangerous malignancy worldwide. As a result, several researchers are concentrating on methodologies for identifying lung cancer nodules employing digital imaging, particularly Computed Tomography (CT) [1]. CT scans use X-rays to generate several images, presenting a challenge for radiologists in detecting small nodules among these visuals [2]. The primary duty of the radiologist in detecting lung cancer was the examination and decoding of nodules. A plethora of scientists and researchers have invented technology to aid doctors in optimising time and resources. The dimensions and morphology of nodules often provide preliminary evidence of malignancy, facilitating their categorisation as benign or malignant. Lung nodules shorter than three centimetres are often deemed benign, however those over 3 cm are considered malignant or categorised as lung masses [3]. Figures 1(a) and 1(b) depict images of benign and malignant lung carcinoma [4].

Cancer likelihood can be evaluated based on nodule categorization and further results. AI methodologies are significantly contributing to the initial diagnosis and categorization of diverse cancer kinds. In recent years, deep learning (DL) models have been employed across several domains, including medical, agriculture, gaming, and others [5].

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

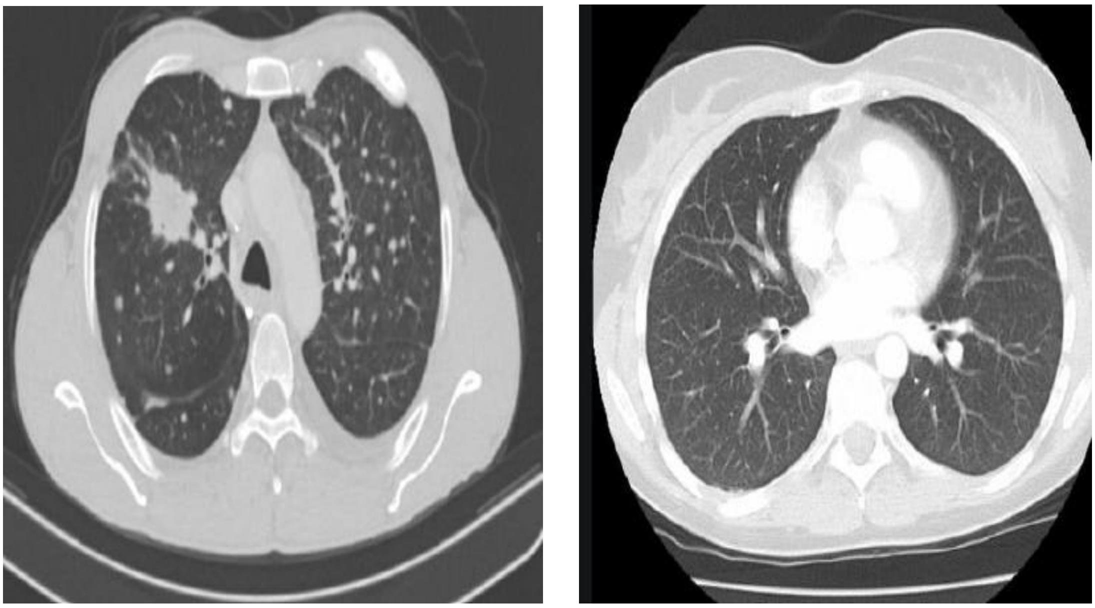

Figure:1 (a)Benign Lung Image **Figure:1 (b) Malignant Lung Image**

Deep learning models have commendable performance across several domains, especially in specialized applications such as picture classification, object recognition, and image segmentation [6]. Deep Learning is an area of Artificial Intelligence that utilizes linked nodes to execute complex tasks. DL algorithms can learn from training data rather than relying on pre-programmed instructions. A number of scholars have already investigated cancer diagnosis with deep learning.

1.1 Categorization of Lung Cancer

Lung cancers are categorized into four main types: adenocarcinoma, Large-cell carcinoma, normal and squamous cell carcinoma. The differentiation among both of these kinds would signify where they come from, growth rate, and the suitable treatment regimen to be applied [7].

- **Adenocarcinoma**

Adenocarcinoma is a kind of cancer that develops in the glandular cells, which are the cells that make mucus and other things. It generally starts in the outside (peripheral) parts of the lungs and can propagate to lymph nodes along with other organs.

- **Large Cell Lung Carcinomas**

Large cell lung cancer constitutes between eighty and eighty percent of the cases of lung cancer. This particular form of lung cancer frequently happens in non-smokers, women, and younger individuals. Large cell lung cancer originates in the larger, main bronchi that possesses the capacity for local spread. Moreover, it has the ability to metastasize across greater distances than the other patterns. The growth rate at its location of origin is reported to be superior to that of other. It could be classified into the following subtypes: squamous cell carcinoma, adenocarcinoma, and large cell lung cancer. Therapies for each of the specified kinds of cancer often include a combination of surgery, chemotherapy, and radiation therapies.

- **Normal**

In lung cancer categorization, the normal class denotes lung tissue that displays no indications of malignancy or atypical cellular proliferation. Normal lung cells maintain a consistent form, size, and activity, facilitating optimal gas exchange and respiratory health. These cells proliferate in a regulated manner and do not aggregate or infiltrate adjacent tissues. Unlike malignant cells, such as those present in adenocarcinoma, squamous cell carcinoma, or small cell lung cancer, normal lung tissues are devoid of genetic abnormalities, uncontrolled proliferation, and spreading capability. Recognizing the normal class is crucial in medical imaging, histology, and machine learning-driven cancer diagnosis, since it establishes a benchmark for differentiating healthy tissue from malignant alterations.

- **Squamous Cell Carcinoma**

Squamous cell carcinoma often begins in the bigger bronchi. Its development is gradual in comparison to other kinds of lung cancer. Adenocarcinomas represent the most rapidly increasing incidence of lung cancer kinds in the USA. This has been often observed among both smokers and non-smokers. These normally arise at the lung periphery and may develop and expand rapidly.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

1.2Requirement for the Automation of Lung Images

Lung imaging is the process of seeing the lungs and the structures that are connected with them, such as the airways, blood vessels, and lung tissue. This is achieved through the application of medical imaging techniques. X-rays, computed tomography (CT) scans, magnetic resonance imaging (MRI) scans, and ultrasound were among the many often recommended methods of imaging for lung assessment. Lung scanning may be used to identify and track several pulmonary disorders [8], it is also capable of diagnosing problems with lung function, such as congestion of the airways or a reduction in lung capacity.

A top apex, three surfaces, and three borders are all characteristics of the conical shape of the lungs. The pulmonary artery is one of two organs located in the chest that are responsible for removing carbon dioxide from the body and sending oxygen to the body. The lungs of a typical adult weigh around one thousand grams [9]. An imaging examination that looks at the lungs and potentially assists in the diagnosis of certain lung illnesses is called a lung scan. There is also the possibility of using a lung scan to evaluate how well the medication is working. One kind of nuclear imaging examination is referred to as a lung scan. This indicates that a minute quantity of radioactive material is used over the whole of the scan.

In today's world, a significant number of individuals are afflicted with lung cancer, and the delayed diagnosis of this illness may occasionally result in death. Cancer will develop as a result of the proliferation of germs in tissues throughout the lung. It is noted that there are several forms of cancer. Detecting lung cancer may be accomplished by the use of several standard procedures. The purpose of this research was to develop a more effective model for detecting cancer by using a variety of Deep learning methods [10].

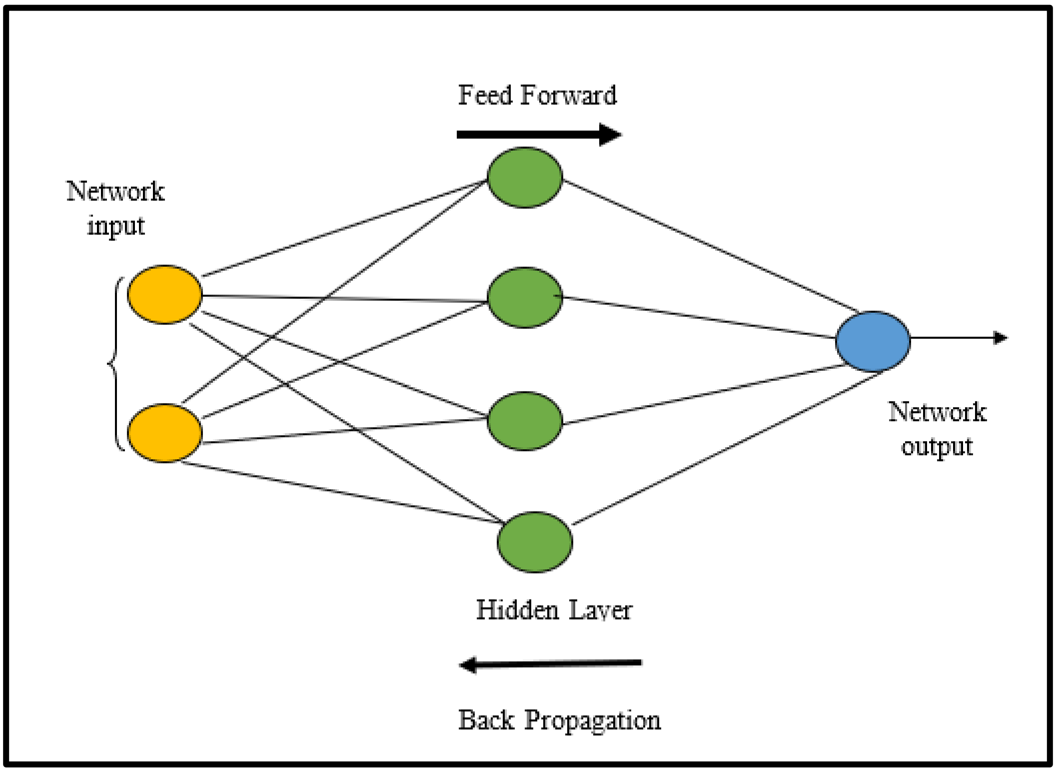

Figure 2 depicts the fundamental structure of these layers (input, hidden, and output) in a conventional neural network model [11]. The input layer acquires the fundamental collection of basic patterns. The hidden layers execute the necessary processing functions to extract the relevant abstract patterns from the raw dataset. The output layer is responsible for synthesizing the patterns necessary to represent the high-level abstract characteristics. A typical deep neural network model consists of over 10 layers and may include several hundred layers in total.

Image detection and identification challenges are structured to identify a specific component inside a medical image. Marginal space learning is a widely used strategy that is both efficient and resilient for organ recognition [12]. The deep learning implementation substitutes probabilistic boosting trees with a neural network-based boosting cascade, enhancing the efficiency of this technique. The efficiency of scanning the full volume to accurately discover anatomical features was enhanced by substituting the search process with an artificial agent [13]. This search methodology employs anatomical principles to identify anatomical landmarks using deep

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

reinforcement learning. The technique rapidly identifies several landmarks throughout an entire CT volume in just seconds.

1.3Role of Deep Learning in Healthcare

A new subset of the machine learning program is known as deep learning, which has been taken into consideration. In this day and age of automation, this has been used as a means of overcoming the limitations that are associated with the conventional methods of machine learning. One thing that has been noted is that the majority of classification systems that are working with regard to the traditional techniques depend only on the feature extraction processes in order to improve their levels of performance. In a broad sense, the process of extracting meaningful characteristics from the data looks to be one that is laborious and time-consuming. Furthermore, in order to develop a feature extractor that is capable of retrieving the necessary information from the data in an effective manner, a significant amount of previous domain experience is required simultaneously. In contrast to the hand-crafted feature engineering, it has been shown that these deep networks are able to acquire the complex hierarchies straight from the raw data, and as a result, they are able to provide the hierarchical data representations that are necessary [14]. The development of a learning model that incorporates many layers of representations has been suggested with the intention of producing a greater number of abstractions.

As a result of the fact that each of the distinct representation levels in the model would extract a specific feature from the data, the higher-level features are consequently derived from the lower-level features in order to provide interpretation to the data [15].

1.4 Motivation

The precise etiology of lung illness remains ambiguous, prompting the need for early localization of the condition using supplementary diagnostic methods. Despite the existence of different components identified as carcinogens, the precise role of their contribution to tumor proliferation remains unclear. The primary objective of this project is to create a very precise classification architecture specifically for digital lung pictures, aimed at improving lung cancer detection methods. The research aims are enumerated below:

- To create an optimum deep learning model for lung image classification, using feature selection as a key parameter, governed by a modified Brainstorming optimization approach.
- To present an effective lung cancer categorization model that employs the Hybrid Feature Extraction (HFE) model in conjunction with Deep Neural Network (DNN) classification methods.

1.5: Objective of Research

This project aims to develop a thorough system for detecting abnormal lung images utilizing methods for deep learning and methods of image processing. The following are the main objectives of the study:

- **Data Acquisition and Annotation**
  - To compile an extensive collection of pulmonary pictures from credible medical imaging sources (e.g., CT scans or X-rays).
  - To annotate lung photos with the aid of medical specialists, delineating areas of interest and recognizing malignant tendencies to properly train the model.
  - To preprocess the photos for noise removal, normalization, and enhancement to guarantee high-quality input for the deep learning model.
- **Feature Selection and Dimensionality Reduction Utilizing the Brain Storm Optimization Algorithm**
  - To extract pertinent characteristics from pulmonary pictures via sophisticated image processing and convolutional methodologies.
  - To use the Brain Storm Optimization (BSO) method for intelligent feature selection and dimensionality reduction, enhancing the model's performance and efficiency.
- **Classification of Models Utilizing Deep Neural Networks (DNN)**
  - To develop and train a Deep Neural Network (DNN) model proficient in differentiating between normal and abnormal (cancerous) lung pictures.
  - To enhance the architecture of the DNN for superior accuracy, precision, recall, and F1-score in lung cancer diagnosis.

2. Literature Survey

A revolutionary development in medical technology, deep learning models use intricate layered neural networks, allowing for the interpretation of data with heightened complexity and abstraction. With the help of this innovative technology, computer-aided technologies have made great strides in healthcare, particularly in

the area of automated lung cancer diagnosis using radiological images. The use of several image enhancing methods involves a pre-processing phase that enhances diagnostic accuracy. This article analyses several strategies for lung cancer diagnosis using Deep Learning (DL) methods., highlighting their efficacy in identifying lung carcinoma along with their limitations.
Singh, H. et al. (2025), published in Cognitive Computation, offers a prospective analysis of the impact of deep learning (DL) on lung cancer diagnosis. It examines a diverse array of deep learning models, including convolutional neural networks, vision transformers, and hybrid architectures, used across modalities like as computed tomography, positron emission tomography, and chest radiographs. The document emphasizes topics such as self-supervised learning, explainable artificial intelligence, and the integration of multimodal data. It delineates significant obstacles to clinical translation—specifically model generalizability, insufficient annotated datasets, and deficiencies in explain ability—while suggesting ways for future research to address these challenges.

Zhao, F. et al. (2024) examined deep learning models used to evaluate the malignancy risk of lung nodules seen in CT images. It methodically examines and integrates findings from 29 peer-reviewed research, contrasting parameters like accuracy, sensitivity, and AUC. The evaluation indicates that the majority of models exhibit strong performance (AUC ~0.92), although highlights considerable diversity influenced by dataset size, image quality, and model design. It emphasizes that models integrating image data with clinical or radiomic characteristics often surpass those dependent only on imaging. The authors emphasize the need for external validation and uniform assessment methodologies.
Patel, D. et al. (2024) published in Expert Systems with Applications, assesses 153 papers using deep learning for radiology-related lung cancer tasks. It classifies research into three domains: tumour categorization, lesion segmentation, and clinical prediction (e.g., survival and recurrence). The review encompasses many imaging modalities, including CT, PET, MRI, and histology. The primary models examined are CNNs, 3D U-Nets, and transformers. It recognizes improvements in performance since 2021 attributable to expanded datasets and pre-trained networks, while also acknowledging difficulties related to data imbalance, interpretability, and implementation in practical clinical environments.

3. Methodology

The widespread usage of Convolutional Neural Networks (CNNs) for medical image analysis is a direct result of advancements in deep learning and AI. With remarkable accuracy, these deep learning algorithms can examine complicated trends in medical images and spot abnormalities. This research classified chest CT scan images using a CNN-based lung cancer detection algorithm [19]. The AI model is engineered to ascertain the presence of lung cancer in a patient and, if confirmed, to specify its particular kind. The categorization has four categories: Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma, and Normal Lung Scans. Each cancer kind has unique features, and accurate identification is essential for choosing the suitable treatment approach.
A substantial dataset was amassed from various sources, meticulously cleaned, and pre-processed to guarantee optimum performance in constructing an efficient deep learning model. The dataset comprises chest CT scan pictures that were transformed into JPG and PNG formats to ensure interoperability with our deep learning architecture. Due to the prevalence of inconsistencies in real-world medical datasets, including noise, fluctuations in picture quality, and class imbalances, comprehensive pre-processing was conducted to improve image quality and facilitate the model's acquisition of significant patterns from the data. The objective was to develop a dataset that enables the CNN model to accurately distinguish among various cancer kinds and provide exact predictions.
This research seeks to enhance lung cancer diagnosis by merging deep learning with medical imaging using AI-driven automation. The primary objective is to create a dependable and efficient diagnostic system that aids radiologists and medical professionals in the early and more precise detection of lung cancer. AI-based models may significantly enhance healthcare outcomes and save lives by diminishing reliance on human analysis and improving diagnostic efficiency.

3.1 Dataset Collection

The dataset used in this study was enabled by the efforts of several sources and researchers who have made medical imaging data publically accessible. The endeavor to gather, sanitize, and organize this information was crucial in developing a dependable AI-based lung cancer detection system. Subsequently, further datasets and empirical clinical data may augment the model's efficacy, guaranteeing its robust generalization across diverse patient scenarios and imaging settings. The collection of datasets is done from the following online source: https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images [20].

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

3.2 Clustering Analysis

The clustering analysis was carried out in order to investigate the structure of the dataset and to categorize photos that were comparable based on the feature representations of those images [21]. In this stage, the objective was to determine whether or not the dataset contains natural categories that may be of use in the classification process or in subsequent preprocessing processes. The ability of deep learning models, particularly pre-trained convolutional neural networks (CNNs) like VGG16, to extract rich feature representations from photos has been shown. High-level features, such as textures, forms, and patterns, are extracted by a pre-trained model rather of manually picking features. These features are more useful for clustering than the features that are manually selected.

Steps Involved:

- The feature extractor that was used was VGG16, which had been pre-trained on ImageNet.
- All of the fully connected layers were eliminated (include_top=False), and only convolutional features were used in this analysis.
- The dimensions of each picture were adjusted to 224×224 pixels in order to conform to the specifications of the VGG16 input.
- Additionally, the extracted features were flattened into a one-dimensional feature vector for each image.

When compared to raw pixel values, the use of a deep learning model for feature extraction resulted in the generation of meaningful feature representations, which led to an improvement in the accuracy of clustering [22].

For the purpose of locating innate patterns within the dataset, clustering is used. Clustering the photos of lung cancer helps determine whether or not the various types of lung cancer naturally segregate in feature space. For example, adenocarcinoma, squamous cell carcinoma, big cell carcinoma, and normal instances are all examples of lung cancer photographs that belong to distinct categories. Here are the primary steps:

- The K-Means clustering algorithm was used with four clusters, which corresponded to the anticipated number of classes.
- The Elbow approach was used in an effort to ascertain the ideal number of clusters; however, the findings that it produced were not entirely transparent.
- The Silhouette Score was computed in order to assess the quality of the clustering.
- A Silhouette Score of 0.64 was achieved, which represented clusters that were relatively well separated from one another.

The Elbow technique did not give obvious insight, despite the fact that K-Means performed quite well. This suggests that more tuning or alternate techniques for clustering (such as DBSCAN or hierarchical clustering) might be investigated.
It is vital to see how effectively the groups are divided once the clustering process has been completed. For the sake of visualization, the feature space was reduced to three dimensions using t-SNE as shown in figure 3, which stands for t-distributed stochastic neighbour embedding. This was done since feature representations are high-dimensional.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

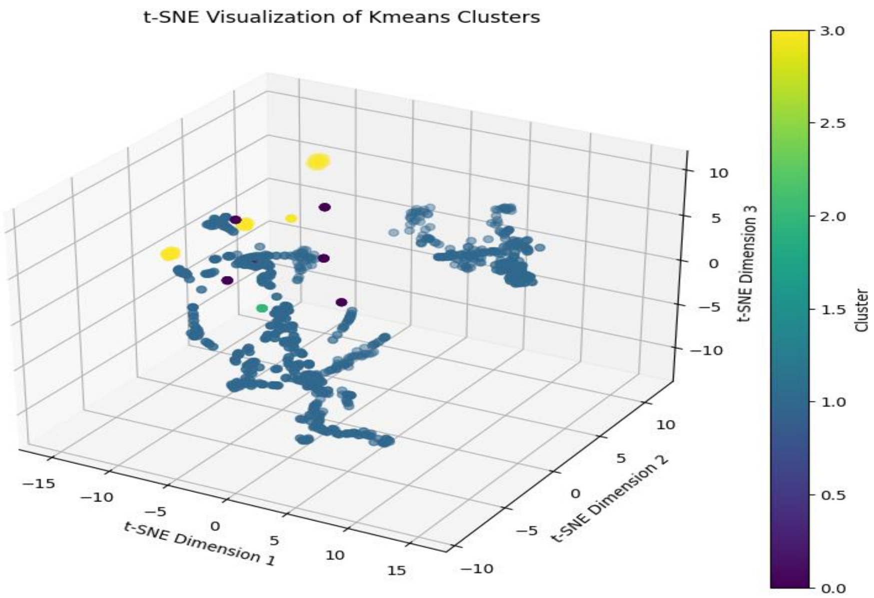

Figure 3: Visualize High-Dimensional Data by Shrinking It Down To 3D

t-SNE (t-Distributed Stochastic Neighbour Embedding) was applied to reduce the 50-dimensional UMAP-transformed data to 3D. By using the graph, we were able to get a better idea of how effectively the data set was clustered and to what extent there were areas that overlapped across the various categories.

3.3 Data Loading and Data Preparation

The process of growing the dataset via data augmentation and conducting an examination of the dataset through exploratory data analysis (EDA) are the primary foci of this stage. The completion of these procedures is essential in order to guarantee that the dataset is well-balanced, diversified, and appropriate for the training of a deep learning model for the classification of lung cancer. One method that may be used to artificially enhance the size of a dataset is known as data augmentation. This method involves performing a variety of alterations to photos that already exist. The prevention of overfitting, the improvement of model generalization, and the optimization of class distributions are all facilitated by this.

In the beginning of this study, the dataset had an imbalance in the number of photographs that were assigned to each category. Each of the classes was enlarged to include an equal amount of two thousand photographs in order to guarantee that the learning process would be fair. For the purpose of the augmentation process, the Albumentations library was used, which offers a wide range of transformation functions that are both effective and versatile.

Techniques of Augmentation that are employed-
Step-1 Flipping the Horizontal (p = 0.4)
- For the purpose of randomly flipping pictures along the horizontal axis, a probability of forty percent was specified.
- The model is able to generalize more effectively as a result of this modification, which introduces varied orientations for the lung pictures.

Step-2 Adjustments to the brightness and contrast (p = 0.1)
- The brightness and contrast of the photographs were subtly changed, with a chance of ten percent.
- Since of this, the model is more resilient since it mimics a variety of illumination situations that may be encountered throughout the process of picture collecting.

Step-3 Rotation (with a standard deviation of ±5°) (p=0.2)
- The rotation of the images was performed within a narrow range of ±5 degrees, with a probability threshold of 20%.
- This takes into consideration any little misalignments that may take place throughout the process of medical imaging.

Step-4 Gaussian Noise when p equals 0.1

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

- A little amount of random noise was applied to the photos with a chance of ten percent.
- With the aid of this transformation, the model is able to become more resistant to fluctuations in picture quality and noise that are present in scans taken in the actual world.

In order to guarantee that the augmented pictures maintained its realistic characteristics while also providing variety, each alteration was executed in a random fashion depending on the probabilities that were established. An additional directory was used to hold the enhanced photos, and inside that directory was an organized folder format that corresponded to each individual class label.

A successful expansion and balancing of the dataset were achieved by the data augmentation process, which also addressed possible problems of class imbalance and overfitting difficulties. For the model to be able to generalize more well to data that it has not before seen, the changes that were done guaranteed that the dataset had adequate variability.

Through the use of exploratory data analysis, it was determined that the dataset was well-structured, error-free, and prepared for pre-processing and model training. The results from EDA served as the basis for the subsequent stage, which consisted of resizing and normalizing the photos before feeding them into the deep learning model for categorization.

Immediately after the completion of data augmentation and exploratory data analysis, the subsequent stage consisted of loading the dataset into a format that was appropriate for the training of deep learning models. Adenocarcinoma, large cell carcinoma, normal, and squamous cell carcinoma were the four categories that were used to classify the CT scan pictures that were included in the dataset.

It was decided to make use of the KerasImageDataGenerator class in order to import and pre-process the photos in an effective manner. Through the use of this tool, the process of importing photos from directories, automatically performing pre-processing procedures, and creating batches of data for training, validation, and testing may be simplified.

3.3.1 Image Processing

Pre-processing the photos was done in order to guarantee consistency and enhance the effectiveness of the learning process before they were input into the model [23]. Pre-processing techniques that were performed while loading included the following:

- **Resizing**

- In order to ensure consistency, all of the photos were scaled to a set dimension of the same size.
- This guarantees that all of the photos have the same input form, which is necessary for CNN models to function properly.

- **Normalization**

- The normalization process included dividing the pixel values by 255 in order to bring them into the range of [0,1].
- Maintaining values inside a narrow and constant range is one of the ways that normalization makes the training process for the model more effective.

- **Batch Processing**

- The dataset was loaded in batches, which enabled a reduction in the amount of memory that was used and an improvement in the computing performance.
- During the training phase, each batch consisted of a predetermined number of photos that were processed concurrently.

- **Shuffling**

- During the training process, the dataset was randomized in order to eliminate any likelihood of learning bias occurring.
- One of the benefits of shuffling is that it prevents the model from learning the sequence of the data, which ultimately results in improved generalization.

- **Class Mode Selection**

- A categorical label format was used since the dataset was divided into four separate classes throughout the classification process.
- Because of this, the output labels were structured appropriately for multi-class categorization, which was a significant achievement.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Immediately after the loading of the dataset, the subsequent stage consisted of preparing it for training by using further pre-processing approaches.

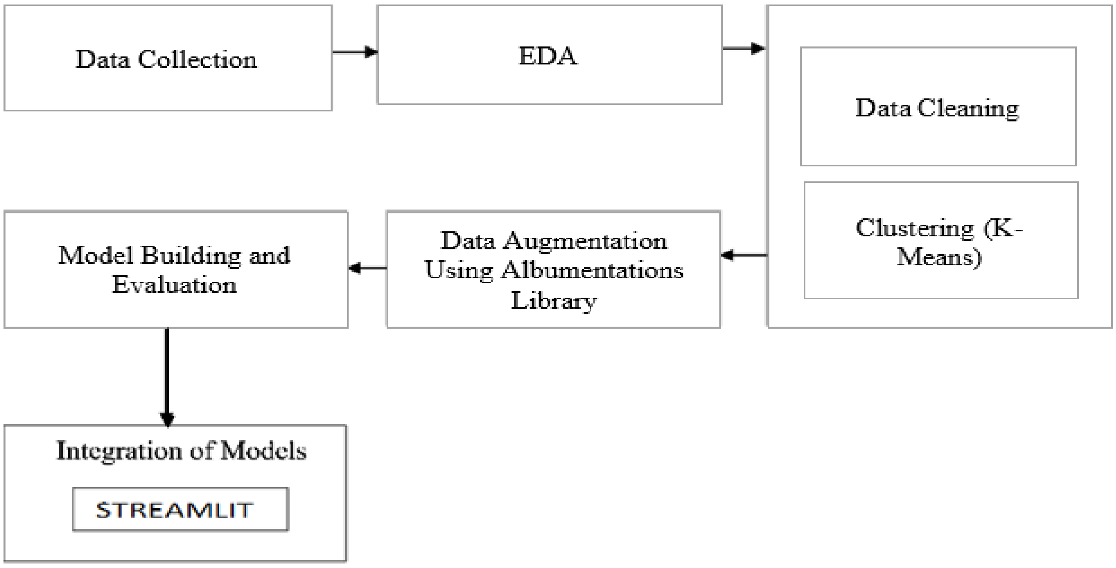

Figure 4: Working Mechanism to Identify Lung Cancer

In order to improve both performance and generalization, it is essential to properly prepare the data in order to guarantee that the deep learning model gets input in the appropriate format and range. For the purpose of ensuring consistency across all inputs, the photos were scaled to a resolution of 128 × 128 pixels. In order to strike a compromise between memory efficiency and training speed, a batch size of sixteen was used. With these decisions, we were able to guarantee that the dataset was processed effectively throughout the training of the model while also preserving an adequate amount of information in each picture [24]. Following the completion of this stage, the dataset was successfully partitioned into training and validation sets, with each respective set having pictures from the four different types of lung cancer. In figure 4, we can show the working mechanism.

4. Result

Following the algorithms were trained and adjusted, the efficacy of the models was evaluated using a variety of metrics, such as accuracy, precision, recall, F1-score, and confusion matrices. These metrics were employed to evaluate the models' overall efficacy. The assessment of models is very important since it offers insights into how well each model works on data that has not yet been observed. The respective comprehensive assessment reports for each model are shown below.
The findings from a number of different models for the identification of lung cancer show that there is a discernible pattern of performance improvement with increasing model complexity. The Single_Layer_CNN model produced an accuracy of 97%, whereas the Multi_Layer CNN model considerably improved the findings, attained 99% accuracy, and achieved a 99% accuracy rate. A 97% accuracy was achieved by the VGG16 model, which used transfer learning and benefited from extensive pre-trained feature extraction. This model performed even better than earlier models. With an astonishing 97% accuracy, ResNet50 earned the greatest performance of all the networks included. On the other hand, the ANN model, which relied only on thick layers, fell behind with an accuracy of 39%, underlining the significance of convolutional architectures for various medical image-based applications. In table 3 we can see the performance of models-

Table 1: Comparatively Analysis of Deep Learning Classification

Proposed Model	Class	Accuracy	Precision	Recall	F1-Score
Single Layer CNN	Adenocarcinoma	0.97	0.97	0.94	0.95
	Large Cell Carcinoma	0.97	0.96	0.98	0.97
	Normal	0.97	1.0	1.0	1.0
	Squamous Cell Carcinoma	0.97	0.95	0.95	0.95

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Multi-Layer CNN	Adenocarcinoma	0.99	0.99	0.98	0.98
	Large Cell Carcinoma	0.99	0.98	0.99	0.98
	Normal	0.99	1.0	1.0	1.0
	Squamous Cell Carcinoma	0.99	0.99	0.98	0.99
VGG16	Adenocarcinoma	0.97	0.98	0.94	0.96
	Large Cell Carcinoma	0.97	0.96	0.95	0.96
	Normal	0.97	0.99	0.99	0.99
	Squamous Cell Carcinoma	0.97	0.94	0.97	0.96
ResNet50	Adenocarcinoma	0.97	0.93	0.98	0.96
	Large Cell Carcinoma	0.97	0.98	0.95	0.97
	Normal	0.97	1.0	1.0	1.0
	Squamous Cell Carcinoma	0.97	0.98	0.95	0.97
ANN	Adenocarcinoma	0.39	1.0	0.0	0.0
	Large Cell Carcinoma	0.39	0.0	0.0	0.0
	Normal	0.39	0.99	0.57	0.73
	Squamous Cell Carcinoma	0.39	0.29	1.0	0.45

These data clearly demonstrate that deeper convolutional models, notably those employing transfer learning like Multi-layer CNN, VGG16 and ResNet50, provide considerable benefits in lung cancer detection tasks. High diagnostic accuracy demands architectures that are capable of extracting and learning complicated hierarchical information. While simpler models give a decent starting point, these architectures are necessary for high diagnostic accuracy. When compared to other models, Multi-Layer CNN emerged as the most promising option for the diagnosis of lung cancer that is both robust and dependable. The graphical representation of outcomes is below-

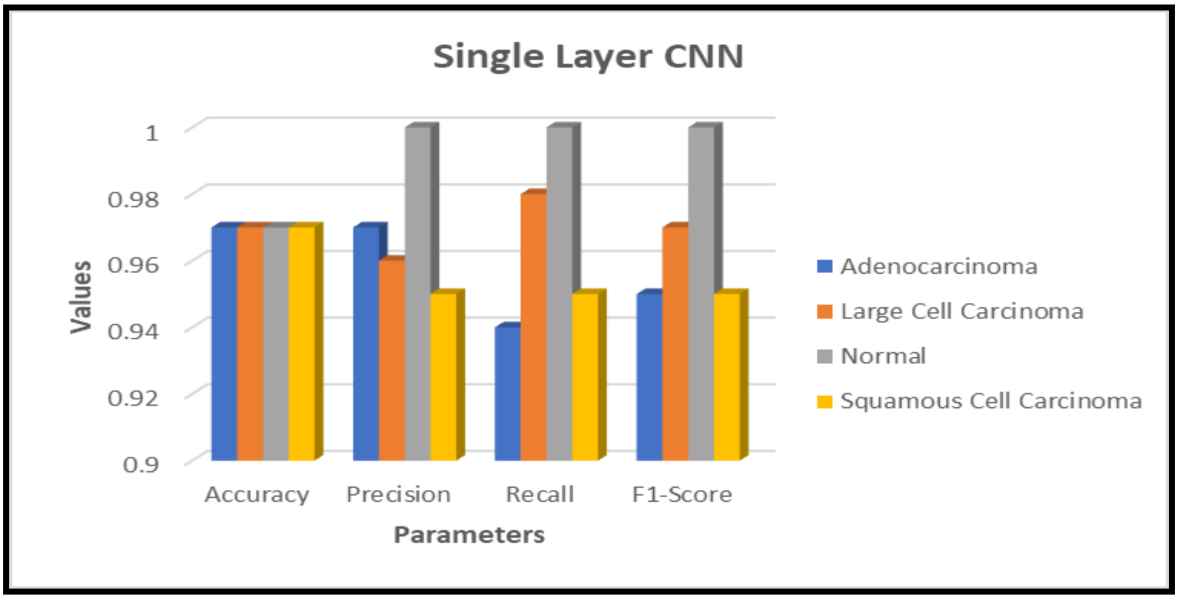

Figure 5: Comparing Outcome for Single Layer CNN for Four Different Classes

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

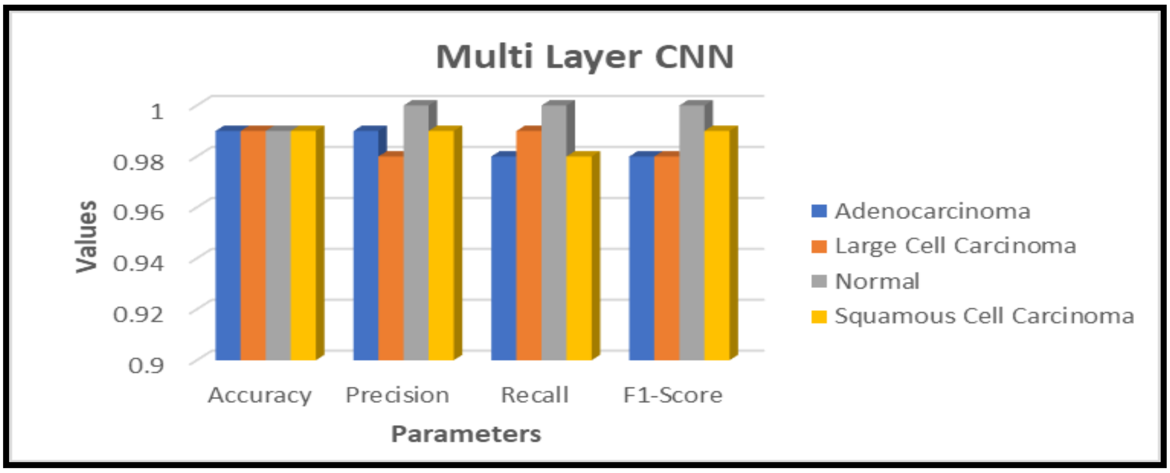

Figure 6: Comparing Outcome for Multi-Layer for Four Different Classes

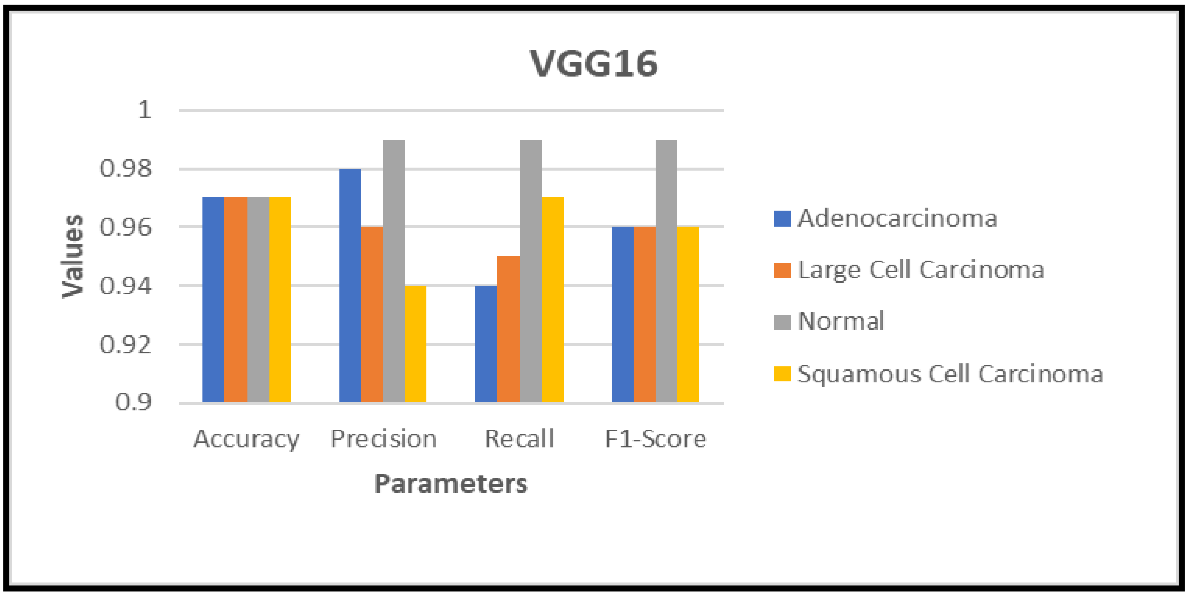

Figure 7: Comparing Outcome for VGG16 for Four Different Classes

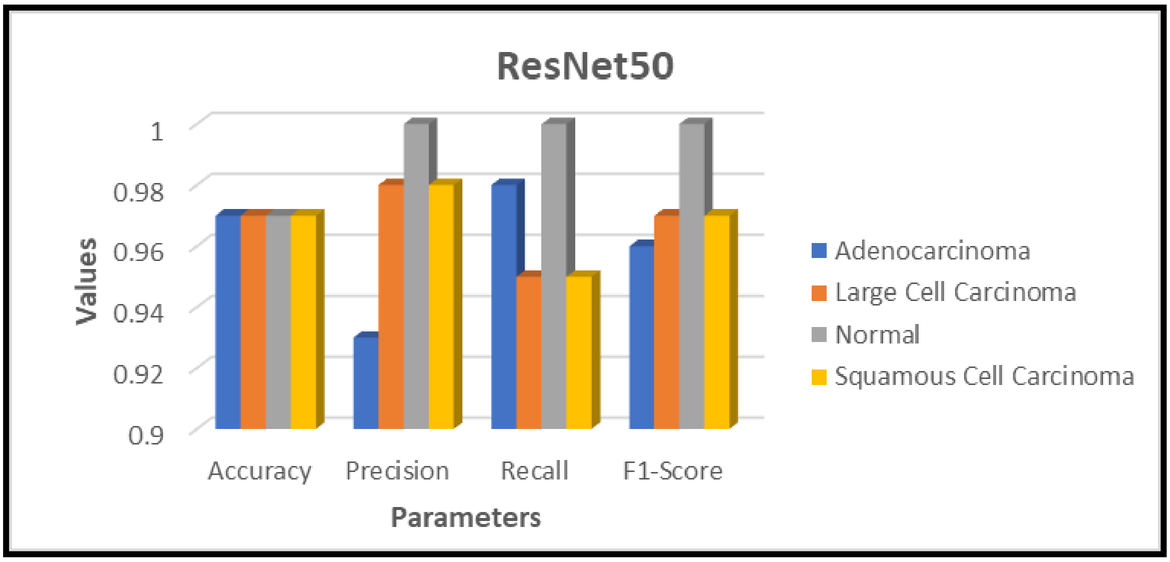

Figure 8: Comparing Outcome for ResNet50 for Four Different Classes

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

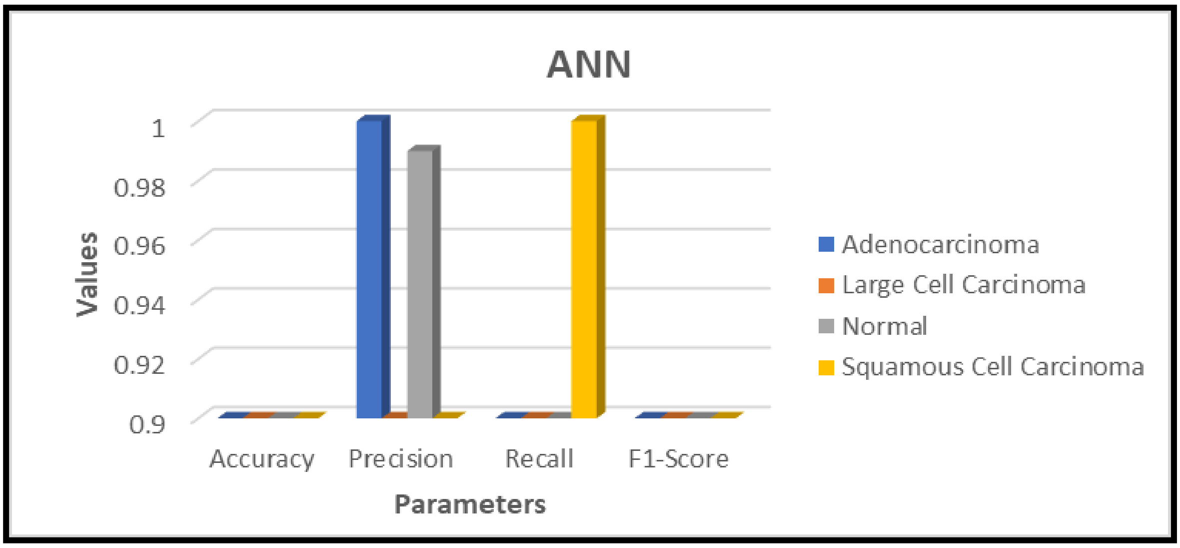

Figure 9: Comparing Outcome for ANN for Four Different Classes

This section delineates the results of the lung cancer detection system created via an extensive pipeline including data preprocessing, clustering, deep learning model training, and the deployment of the final models via a Streamlit online interface. The efficacy of five models—Single Layer CNN, Multi-layer CNN, VGG16, ResNet50, and Artificial Neural Network (ANN)—was assessed for its capability in categorising lung cancer pictures. The dataset underwent preprocessing procedures such as scaling, normalisation, and augmentation to improve model generalisability. Furthermore, unsupervised K-means clustering was used as an exploratory measure to comprehend inherent groups inside the data. Although not used directly for classification, clustering offered insights into feature distribution and facilitated the identification of probable outliers.

Among all the models, the Multi-layer CNN got the greatest classification performance across all criteria, indicating its appropriateness for intricate picture recognition tasks, such as lung cancer diagnosis. ResNet50 had strong performance, marginally lagging behind VGG16. Basic models such as Single ANN exhibited worse accuracy and recall, highlighting the need for more complex architectures in this problem space.

The resulting model was included into an intuitive Streamlit web application to guarantee usability and real-time inference capabilities. This service enables users to submit lung CT scan pictures and get immediate predictions. The interface presents model confidence ratings and a visual representation of the uploaded picture. The application underwent testing on several browsers and devices to guarantee responsiveness and dependability.

5. Conclusion and Future Scope

This paper presents a complete deep learning system for lung cancer diagnosis using CT scan data, including sophisticated picture pre-processing, unsupervised clustering, model training across several architectures, and deployment via an intuitive Streamlit web application. The objective was to evaluate and contrast the efficacy of several deep learning models in identifying lung cancer, while also illustrating the feasibility of using these models in real-time clinical settings.

The system's effectiveness was significantly influenced by the comprehensive data pre-processing. Five models were developed and assessed: Single Layer CNN, Multi-layer CNN, VGG16, ResNet50, and Artificial Neural Network (ANN). The Multi-layer CNN surpassed all other designs, attaining the best accuracy and F1-score, thereby exhibiting exceptional proficiency in feature extraction and classification of lung cancer patterns. This result highlights the efficacy of advanced bespoke convolutional networks designed for medical imaging applications, as they effectively balance complexity with domain-specific learning.

Conversely, the ANN model had the worst performance, underscoring its deficiencies in managing the spatial hierarchies intrinsic to picture data. Although it has a relatively straightforward architecture, it was deficient in depth and convolutional layers required to capture complex information in CT scans, leading to decreased accuracy, precision, and recall in comparison to other models.

In conclusion, the amalgamation of a sophisticated deep learning model with a user-friendly online application presents a potential approach for the early and precise identification of lung cancer. The system attains superior predictive performance while also addressing usability issues, making it appropriate for use in healthcare environments.

Future efforts will concentrate on:

- Augmenting the dataset to include a broader range of instances and picture categories,
- Integrating explainable AI (XAI) methodologies for improved interpretability,

6.References

1. Javed, R., Abbas, T., Khan, A. H., Daud, A., Bukhari, A., &Alharbey, R. (2024). Deep learning for lungs cancer detection: a review. *Artificial Intelligence Review*, *57*(8), 197.
2. Bhuiyan, M. S., Chowdhury, I. K., Haider, M., Jisan, A. H., Jewel, R. M., Shahid, R., &Siddiqua, C. U. (2024). Advancements in early detection of lung cancer in public health: a comprehensive study utilizing machine learning algorithms and predictive models. Journal of Computer Science and Technology Studies, 6(1), 113-121.
3. Hosseini, S. H., Monsefi, R., &Shadroo, S. (2024). Deep learning applications for lung cancer diagnosis: a systematic review. Multimedia Tools and Applications, 83(5), 14305-14335.
4. Zhang, Y., Yang, Z., Chen, R., Zhu, Y., Liu, L., Dong, J., ... & Zhou, M. (2024). Histopathology images-based deep learning prediction of prognosis and therapeutic response in small cell lung cancer. NPJ digital medicine, 7(1), 15.
5. Prasad, U., Chakravarty, S., &Mahto, G. (2024). Lung cancer detection and classification using deep neural network based on hybrid metaheuristic algorithm. Soft Computing, 28(15), 8579-8602.
6. Sivasankaran, P., &Dhanaraj, K. R. (2024). Lung Cancer Detection Using Image Processing Technique Through Deep Learning Algorithm. Revue d'IntelligenceArtificielle, 38(1).
7. Shariff, V. A. H. I. D. U. D. D. I. N., Chiranjeevi, P., & Km, A. (2023). An Analysis on Advances In Lung Cancer Diagnosis With Medical Imaging And Deep Learning Techniques: Challenges And Opportunities. Journal of Theoretical and Applied Information Technology, 101(17), 7083-7095.
8. Primakov, S. P., Ibrahim, A., van Timmeren, J. E., Wu, G., Keek, S. A., Beuque, M., ... &Lambin, P. (2022). Automated detection and segmentation of non-small cell lung cancer computed tomography images. *Nature communications*, *13*(1), 3423.
9. Primakov, S. P., Ibrahim, A., van Timmeren, J. E., Wu, G., Keek, S. A., Beuque, M., ... &Lambin, P. (2022). Automated detection and segmentation of non-small cell lung cancer computed tomography images. Nature communications, 13(1), 3423.
10. Singh, G. A. P., & Gupta, P. K. (2019). Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. Neural Computing and Applications, 31(10), 6863-6877.
11. Negahdar, M. (2022, September). Automatic Grading of Emphysema by Combining 3D Lung Tissue Appearance and Deformation Map Using a Two-Stream Fully Convolutional Neural Network. In *International Workshop on Machine Learning in Medical Imaging* (pp. 181-190). Cham: Springer Nature Switzerland.
12. Jacob, I. J., &Darney, P. E. (2021). Design of deep learning algorithm for IoT application by image-based recognition. *Journal of ISMAC*, *3*(03), 276-290.
13. Harrison, P. T., Vyse, S., & Huang, P. H. (2020, April). Rare epidermal growth factor receptor (EGFR) mutations in non-small cell lung cancer. In Seminars in cancer biology (Vol. 61, pp. 167-179). Academic Press.
14. Asuntha, A., & Srinivasan, A. (2020). Deep learning for lung Cancer detection and classification. Multimedia Tools and Applications, 79(11), 7731-7762.
15. Bhatia, S., Sinha, Y., & Goel, L. (2019). Lung cancer detection: a deep learning approach. In Soft Computing for Problem Solving: SocProS 2017, Volume 2 (pp. 699-705). Springer Singapore.
16. Singh, H., & Rao, M. (2025). Deep Learning Innovations in the Detection of Lung Cancer: Advances, Trends, and Open Challenges. Cognitive Computation. https://doi.org/10.1007/s12559-025-10408-2
17. Zhao, F., Kumar, A., & Lin, W. (2024). Deep learning models for predicting malignancy risk in CT-detected pulmonary nodules: A systematic review and meta-analysis. Lung, 202(3), 335–349.
18. Patel, D., Yadav, R., & Thomas, M. (2024). Deep learning in radiology for lung cancer diagnostics: A systematic review of classification, segmentation, and predictive modeling. Expert Systems with Applications, 239, 122237. https://doi.org/10.1016/j.eswa.2024.122237
19. Klangbunrueang, R., Pookduang, P., Chansanam, W., &Lunrasri, T. (2025, February). AI-Powered Lung Cancer Detection: Assessing VGG16 and CNN Architectures for CT Scan Image Classification. In *Informatics* (Vol. 12, No. 1, p. 18). MDPI.
20. https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images.
21. Wang, Y., Zhang, W., Liu, X., Tian, L., Li, W., He, P., ... & Pan, X. (2025). Artificial intelligence in precision medicine for lung cancer: A bibliometric analysis. *Digital Health*, *11*, 20552076241300229.

22. Yin, S., Yu, Y., Wu, N., Zhuo, M., Wang, Y., Niu, Y., ... & Xi, J. J. (2024). Patient-derived tumor-like cell clusters for personalized chemo-and immunotherapies in non-small cell lung cancer. *Cell stem cell*, *31*(5), 717-733.
23. Tan, S. L., Selvachandran, G., Paramesran, R., & Ding, W. (2025). Lung Cancer Detection systems Applied to Medical images: a state-of-the-art survey. *Archives of Computational Methods in Engineering*, *32*(1), 343-380.
24. Harini, D., Rajesh, U., Vyshnavi, K. S., Supraja, S., & Reddy, B. H. P. Lung Cancer Detection Using Image Processing. *IJSAT-International Journal on Science and Technology*, *16*(2).