Early Diagnosis Small Cell Lung Cancer and Pneumonia Risk Level Prediction using Optimized Deep Learning Approach

Rajeshwar Prasad
Scholar, Department of Computer Science & Information Technology, Guru Ghasidas Vishwavidyalaya (C.U.), Koni, Bilaspur, (C.G.), Chhattisgarh, 495009, India

Amit Kumar Saxena
Professor & Head, Department of Computer Science & Information Technology, Guru Ghasidas Vishwavidyalaya (C.U.), Koni, Bilaspur, (C.G.), Chhattisgarh, 495009, India

Suman Laha
Scholar, Department of Computer & System Sciences, Visva-Bharati University, Santiniketan, West Bengal, 731235, India

ABSTRACT
Lung cancer is a type of cancer that starts in the lungs. It can cause different symptoms, such as coughing, chest pain, shortness of breath, and weight loss. The objective of this study is to develop and validate a robust and reliable predictive model that can accurately differentiate between small cell lung cancer and pneumonia based on medical imaging data, such as chest X-rays or CT scans. The data collection phase is to identify reliable sources of medical images that encompass both pneumonia cases and healthy individuals without pneumonia. From the collected dataset, the image is pre-processed using Principal Component Analysis (PCA) which is a transformation technique that reduces the size of the p-dimensional dataset. A sophisticated approach is adopted, combining t-distributed stochastic neighbour embedding (t-SNE) and an Ant Lion-Based Autoencoder (ALBAE) technique. The findings of the proposed technique, however, achieved the highest accuracy of 99 %. The study seeks to generate an integrated framework for the diagnosis and risk assessment of lung cancer and pneumonia using Python software. The future scope lies in further refining and optimizing the deep learning models to improve their accuracy and reliability in clinical settings.

Keywords
Early Diagnosis, Small Cell Lung Cancer, Pneumonia Risk Level Prediction, Deep Learning, Principal Component Analysis, and Coati Optimization Algorithm.

1. INTRODUCTION
Lung cancer is a type of cancer that originates in the tissues of the lungs, particularly in the cells lining the air passages. It is one of the most prevalent and deadly forms of cancer worldwide, responsible for a significant number of cancer-related deaths. Lung cancer is a malignant condition characterized by the uncontrolled growth of abnormal cells within the lung tissues [1]. It is a multifaceted disease, often attributed to various factors, but most commonly linked to long-term exposure to carcinogens, primarily found in tobacco smoke [2]. This exposure can lead to genetic mutations in lung cells, triggering the development of cancer. Lung cancer comprises several distinct types, with non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) being the most prevalent [3]. These types differ in their cellular characteristics and treatment approaches. Further, lung cancer patients are particularly vulnerable to pneumonia due to compromised immune systems and diminished lung function, which often happens because of a history of smoking [4]. This heightened susceptibility underscores the importance of preventive measures, early detection, and vigilant management to minimize the risks associated with pneumonia in this at-risk population, ultimately aiming to enhance both the quality of care and patient outcomes [5]. Moreover, understanding an individual's risk level for developing lung cancer is pivotal for targeted prevention strategies, emphasizing the importance of smoking cessation programs and lung cancer screenings, particularly for those at higher risk due to smoking history or occupational exposures [6]. In this exploration of lung cancer detection and risk level prediction, the study will delve into the intricacies of this disease, its various types, and the vital role played by early detection and risk assessment in the fight against lung cancer [7]. The diverse risk factors associated with this condition shed light on the complex interplay between genetics, lifestyle, and environmental influences [8].

Pneumonia is a common complication in lung cancer patients and can significantly impact their prognosis and overall health. Early detection and accurate assessment of pneumonia severity are crucial for timely intervention and effective management of the condition [9]. In recent years, machine learning and deep learning techniques have shown promising results in predicting risk levels and assessing the severity of pneumonia [10]. One innovative approach proposed in recent research involves combining t-distributed Stochastic Neighbour embedding (t-SNE) and an ALBAE technique. By utilizing the advantages of both algorithms, researchers aim to improve the accuracy and efficiency of pneumonia severity assessment in lung cancer patients [11]. The research suggests integrating the long short-term memory network with a graph neural network (LSTM-GNN) model with the Coati Optimization Algorithm for predicting risk levels and assessing pneumonia severity. This cutting-edge method leverages the strengths of both models to enhance the performance and reliability of the predictions [12]. This novel approach represents a significant advancement in the field of medical imaging and computational biology. By incorporating state-of-the-art machine learning techniques and optimization algorithms, researchers aim to provide healthcare professionals with a valuable tool for early detection and precise assessment of pneumonia in lung cancer patients [13]. This integrated approach has the potential to improve patient outcomes, enhance treatment strategies, and ultimately save lives. The problem statement revolves around the early diagnosis of Small Cell Lung Cancer (SCLC) and predicting the risk level for Pneumonia. This is a critical issue in the medical field as early detection and accurate risk assessment can significantly improve patient outcomes and treatment plans

[14]. To address this problem, the motivation is to develop an optimized deep-learning approach that can efficiently diagnose SCLC in its early stages and predict Pneumonia risk levels. This will contribute to better patient care, improved treatment strategies, and more effective resource allocation in healthcare settings [15]. The Contribution of this is more smoke and the earlier in life you begin smoking, the greater your risk for SCLC. Other risk factors include exposure to second-hand smoke, workplace carcinogens, radiation and/or environmental pollution, as well as family history of lung cancer and previous HIV infection. While imaging tests indicate that lung cancer may be present, only a biopsy can confirm it. In a biopsy, pathologists who study diseases in a laboratory examine tissue under a microscope to determine the size and shape of the cells and whether they look like small-cell lung cancer. The following section has a Literature survey in section 2, a proposed research methodology in section 3, an Experimental and results discussion in section 4, and a Research conclusion in section 5.

2. LITERATURE SURVEY

Lung cancer can also increase the risk of getting infections, especially pneumonia, which is a serious infection of the lungs. Guo, *et al* [16] developed and validated a simple and non-invasive model which could assess and stratify lung cancer risk in non-smokers in China. A large sample size, population-based study was conducted under the framework of the Cancer Screening Program in Urban China (CanSPUC). Related risk factors were identified through multivariable Cox regression analysis, followed by the establishment of a risk prediction nomogram. The AUC (Area under the curve) was 0.753, 0.752, and 0.755 for the 1-, 3- and 5-year lung cancer risk in the training set, respectively. In the validation set, the model showed moderate predictive discrimination, with the AUC being 0.668, 0.678, and 0.685 for the 1-, 3- and 5-year lung cancer risk. Hou, *et al* [17] aimed to construct a deep learning model combining both radio mic and clinical features to predict the overall survival of patients with non-small-cell lung cancer (NSCLC). To improve the reliability of the proposed model, radiomics analysis complying with the Image Biomarker Standardization Initiative and the compensation approach to integrate multicentre datasets was performed on contrast-enhanced computed tomography (CECT) images. The C-index values of the combined model achieved 0.74, 0.75, and 0.75, respectively, and AUC values of 0.76, 0.74, and 0.73, respectively, 8, 12, and 24 months after diagnosis. Mikhael, *et al* [18] developed a model called Sybil using Low-dose computed tomography (LDCT) from the National Lung Screening Trial (NLST). Sybil requires only one LDCT and does not require clinical data or radiologist annotations; it can run in real-time in the background on a radiology reading station. Sybil achieved an area under the receiver-operator curves for lung cancer prediction at 1 year of 0.92 (95% CI, 0.88 to 0.95) on NLST, 0.86 (95% CI, 0.82 to 0.90) on Mass-Gathering Health (MGH), and 0.94 (95% CI, 0.91 to 1.00) on Chang Gung Memorial Hospital (CGMH) external validation sets. Oh, *et al* [19] proposed a deep-learning-based solution to classify four lung diseases (pneumonia, pneumothorax, tuberculosis, and lung cancer) and healthy lungs using chest X-ray images. To achieve high performance, the Efficient Net B7 model with the pre-trained weights of ImageNet trained by Noisy Student was used as a backbone model, followed by the proposed fine-tuned layers and hyperparameters. This study achieved an average test accuracy of 97.42%, sensitivity of 95.93%, and specificity of 99.05%.

Du, *et al* [20] aimed to develop a deep neural network model to differentiate pneumonia-type lung carcinoma from pneumonia based on chest CT scanning and evaluate its performance. Lesion areas were extracted and classified by a designed spatial attention mechanism network. The model AUC and diagnostic accuracy were analyzed based on the results of the model. The model accuracy rate, sensitivity, and specificity were 74.20%, 60.37%, and 89.36%, respectively. The model AUC and diagnostic accuracy were analyzed based on the results of the model. Some biomarkers have been proposed by Yin, *et al* [21] to predict the outcomes of immunotherapy and targeted therapy, including programmed cell death-ligand 1 (PD-L1) expression and oncogene mutations. Nevertheless, the detection tests are invasive, time-consuming, and have high demands on tumour tissue. By combining AI methods with radiomics, pathology, genomics, transcriptomics, proteomics, and clinical data, the integrated model has shown predictive value in immunotherapy and targeted therapy, which significantly improves the precision treatment of lung cancer patients. Guo, *et al* [22] aimed to develop and internally validate a risk prediction model for lung cancer. Using data from the Cancer Screening Program in Urban China (CanSPUC) in Henan province, China between 2013 and 2019, the study conducted a prospective cohort study consisting of 282,254 participants including 126,445 males and 155,809 females. The C-index of the model for 1-year lung cancer risk was 0.766 and 0.741 in the training set and validation set, respectively. Li *et al* [23] studied a hybrid approach, merging GLCM with Haralick and autoencoder features. SVM RBF and SVM Gaussian achieved perfect performance, with SVM polynomial at 99.89% accuracy using GLCM and auto encoder features. SVM Gaussian reached 99.56%, and SVM RBF at 99.35% accuracy with GLCM and Haralick features. These results highlight the potential of this method for enhancing lung cancer diagnosis and treatment planning systems. Lin *et al* [24] aimed to develop a combined model integrating deep learning, radionics, and clinical data for classifying lung nodules into benign or malignant categories and further categorizing them into different pathological subtypes and Lung-RADS scores. The model achieved high accuracy in all three classification tasks, with F1 scores ranging from 75.5% to 80.4%. Feng *et al* [25] aimed to develop a prediction model for radiation-induced pneumonia (RP) in early non-small-cell lung cancer (NSCLC) patients undergoing stereotactic body radiation therapy (SBRT). The Hybrid model, which combined all features, showed the best performance with accuracy, sensitivity, specificity, and area under the receiver operator characteristic curve of 0.857, 1, 0.875, and 0.920, respectively.

3. RESEARCH PROPOSED METHODOLOGY

Small cell lung cancer (SCLC) is a type of lung cancer that can be affected by pneumonia, which is a lung infection caused by bacteria, viruses, or fungi [26-28]. Pneumonia can also worsen the symptoms and complications of lung cancer, such as cough, chest pain, haemoptysis, dyspnea, and weight loss [29-30]. Predicting Small Cell Lung Cancer (SCLC) and assessing the risk level of pneumonia affecting lung cancer severity is crucial for early intervention and patient care. SCLC prediction involves analysing medical data, such as patient history and imaging, to detect the presence of SCLC tumours in the lungs. Concurrently, assessing the risk level of pneumonia in lung cancer patients helps determine potential complications and their severity. Combining these predictive approaches enhances patient outcomes by enabling timely diagnosis and personalized treatment plans, ultimately improving the

management of SCLC and its associated pneumonia risks. Parameters of PCA compute metrics, Method impute missing [31-32]. The t-SNE parameters are Perplexity, Early exaggeration**,** Learning rate, and components. Parameters of ALBE Nutrients, Temperature, Salinity, **and** Turbulence. Figure 1 depicts the block diagram of the proposed work [33].

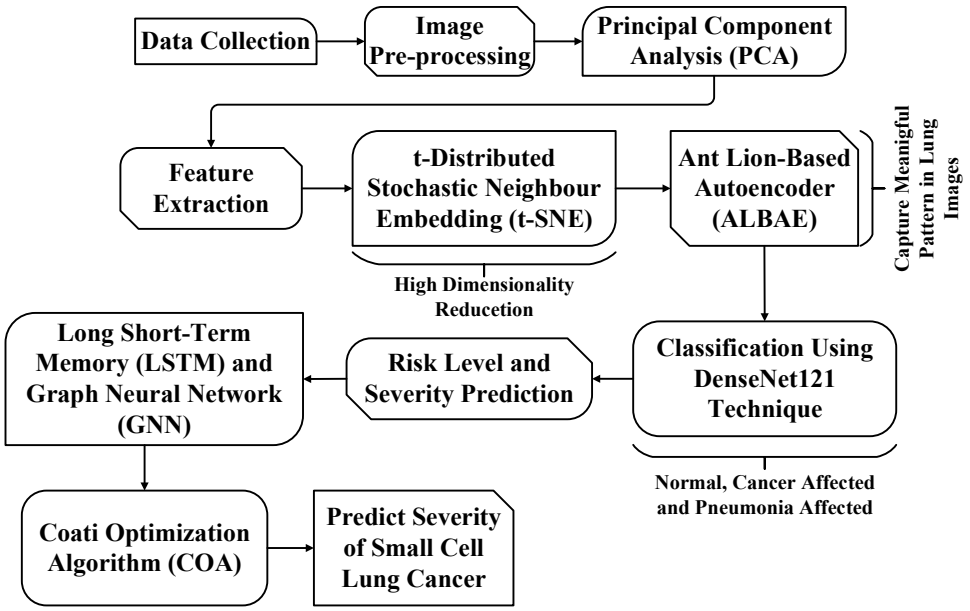


Fig 1: Block diagram of the proposed work

In the initial phase of this research, a diverse dataset of medical images containing both pneumonia and healthy cases was collected while ensuring ethical approvals. Principal Component Analysis (PCA) and adaptive iterative guided filtering were used for image pre-processing. In the feature extraction and classification step, a combination of t-SNE and an ALBAE was employed to extract meaningful attributes from lung cancer images. Finally, a DenseNet121 technique was used to classify the images into normal, cancer-affected, and pneumonia-affected categories. In the risk level severity prediction stage, a novel approach merged LSTM-GNN models with the Coati Optimization Algorithm to predict pneumonia severity in lung cancer patients, considering shared and disease-specific risk factors. This research has the potential to enhance risk assessment and improve patient care for lung cancer and pneumonia management.

3.1 Data collection and image pre-processing

In the data collection phase, the first critical step is to identify reliable sources of medical images encompassing both pneumonia cases and healthy individuals. This involves careful selection of datasets and medical institutions capable of providing high-quality images, while ensuring ethical approvals are obtained to comply with privacy and regulatory standards. The lung cancer dataset used in this study was sourced from Kaggle and includes CT scans of patients at different stages of lung cancer, as well as healthy subjects. Additionally, IQ-OTH/NCCD slides were annotated by oncologists and radiologists from these centers to ensure high-quality labelling. Data cleaning procedures, including imputing missing values and detecting outliers, were applied to maintain dataset integrity. The dataset is designed to be diverse, covering multiple pneumonia types, age groups, and ethnic backgrounds, which supports the development of a model that is broadly applicable across real-world scenarios. Pre-processing involves Principal Component Analysis (PCA) to reduce dimensionality and generate uncorrelated features, followed by an adaptive iterative guided filtering method to suppress noise and enhance image quality. To further strengthen the study, a more extensive evaluation is planned across multiple datasets and varied clinical scenarios. This includes testing on external hospital datasets, different imaging modalities (X-ray and CT scans), and noisy or artifact-affected images to assess model robustness and generalizability. Such a comprehensive evaluation will provide stronger evidence of the model's reliability, practical utility, and applicability in diverse clinical settings.

3.1.1 Principal component analysis

Principal component analysis (PCA) is a commonly used technique in image processing to reduce the dimensionality of data while retaining important information. In the context of early diagnosis of small cell lung cancer, PCA can be used as a pre-processing step to extract relevant features from medical images such as CT scans or X-rays, which can then be used for accurate classification of lung cancer.

The main idea behind PCA is to transform the original image data into a new set of orthogonal variables, called principal components, which capture the maximum amount of variance in the data. Mathematically, PCA involves finding the eigenvectors and eigenvalues of the covariance matrix of the image data. The eigenvectors represent the directions of maximum variance in the data, while the eigenvalues indicate the amount of variance explained by each principal component. Calculate the mean image vector from the input data matrix X, denoted as μ. M observations of an N length random vector $\bar{z}$, the PCA transform is completed by first subtracting the mean from the vector

$$\bar{x} = \bar{z} - E|\bar{z}| \quad (1)$$

The $N \times N$ covariance matrix C_x is computed

$$C_x = \mathrm{E}[\bar{x}\bar{x}'] \quad (2)$$

The principal components $\bar{s}$ of $\bar{x}$ are given in terms of the unit-length eigenvectors ($\overline{e_1}\ldots.\overline{e_N}$), of C_x

$$\bar{s} = W\bar{x} \quad (3)$$

Where the projection matrix W contains the eigenvectors ($\overline{e_1}\ldots.\overline{e_N}$). Usually, the M observations would typically be exemplars taken from any one of C's possible classes. By applying PCA to pre-process medical images for early diagnosis of small cell lung cancer, the study can effectively reduce the dimensionality of the data while preserving important information. The extracted features from the principal components can be used as input for classification algorithms to distinguish between lung cancer patients and healthy individuals with improved accuracy and efficiency. Moreover, PCA helps with the visualization and interpretation of high-dimensional image data, enabling clinicians to identify patterns and abnormalities that may indicate the presence of small-cell lung cancer at an early stage.

3.2 Feature extraction and classification

In the domain of lung cancer analysis, the process of feature extraction is pivotal for discerning pertinent information from medical images. This involves the extraction of crucial attributes from the images, including edges delineating structures, pixel intensity rates conveying texture and density, and the identification of blood clots. To achieve this, a sophisticated approach is adopted, combining t-distributed stochastic neighbour embedding (t-SNE) and an ALBAE technique. T-SNE aids in the reduction of high-dimensional image data into a lower-dimensional representation, preserving essential data relationships. Auto encoders are neural network models used for feature extraction and dimensionality reduction. Finally, in the classification stage, the lung CT images are classified into three different categories such as normal lung, cancer-affected lung, and pneumonia-affected lung using the DenseNet121 technique.

3.2.1 T-distributed Stochastic neighbour embedding (t-SNE) with ant lion-based auto encoder (albae) technique

The combination of t-distributed stochastic neighbour embedding (t-SNE) and the ALBAE technique for feature extraction in early diagnosis of small cell lung cancer can provide an advanced approach for identifying patterns and clusters within high-dimensional data. t-SNE is a powerful technique for visualizing and understanding high-dimensional data by mapping it to a lower-dimensional space while preserving local structure. It is commonly used for dimensionality reduction and visualization of complex datasets.

$$C = KL((P||Q) = \sum_J \sum_J p_{IJ} log \frac{P_{ij}}{Q_{ij}} \quad (4)$$

Where again, p_{IJ} and Q_{ij} is set as zero. This type of SNE is symmetric SNE for $\forall i, j$ the low-dimensional map Q_{ij} are given by in equation (5),

$$Q_{ij} = \frac{\exp(-||Y_i - Y_j||^2)}{\sum_{K \neq 1} \exp(-||Y_k - Y_1||^2)} \quad (5)$$

The pairwise similarities in the high-dimensional space P_{ij} in equation (6),

$$P_{ij} = \frac{\exp(-||x_i - x_j||^2/2\sigma^2)}{\sum_{K \neq 1} \exp(-||x_k - x_1||^2/2\sigma^2)} \quad (6)$$

When a high-dimensional data point x_i is an outlier (i.e., all pairwise distances $||x_i - x_j||^2$ are large for xi). For such an outlier, the values of P_{ij} are extremely small for all j, so the location of its low-dimensional map point yi has very little effect on the cost function. As a result, the position of the map point is not well determined by the positions of the other map points. This study circumvents this problem by defining the joint probabilities pi j in the high-dimensional space to be the symmetrized conditional probabilities, that is, $P_{ij} = \frac{P_{i|j} + P_i|^j}{2n}$. This ensures that $\sum_j P_{ij} > \frac{1}{2n}$ for all data points x_i , as a result of which each data point x_i makes a significant contribution to the cost function.

$$q_{ij} = \frac{(1+||y_i - y_j||^2)^{-1}}{\sum_{K \neq 1}(1+||y_i - y_j||^2)^{-1}} \quad (7)$$

The gradient of the Kullback-Leibler divergence between P and the Student-t-based joint probability distribution Q (computed using equation 4) is derived, and is given in equation (8),

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (P_{ij} - q_{ij})(y_i - y_j)(1 + ||y_i - y_j||^2)^{-1} \quad (8)$$

Euclidean distances in the high-dimensional and the low-dimensional space (i.e., as a function of$(P_{ij} - q_{ij})$ and $(y_i - y_j)$ for the symmetric versions of SNE, UNI-SNE, and t-SNE. positive values of the gradient represent an attraction between the low dimensional data points y_i and y_j , whereas negative values represent a repulsion between the two data points.

Table 1: Algorithm for t-distributed stochastic neighbour embedding

Algorithm 1: T-Distributed Stochastic Neighbour Embedding
Data: data set $X = \{x1, x2, \ldots, xn\}$,
Cost function parameters: perplexity Perp, optimization
parameters: number of iterations T, learning rate η, momentum α(t).
Result: low-dimensional data representation $Y(T) = \{y1, y2, \ldots, yn\}$.
begin to compute pairwise affinities P_{ij} with perplexity Perp (using Equation (1) set $P_{ij} = \frac{P_{i\|j} + P_i\|^j}{2n}$.
sample initial solution $Y(0) = \{y1, y2, \ldots, yn\}$ from N (0,10−4 I) for $t = 1\ to\ T$
do compute low-dimensional affinities q_{ij} (using Equation 4) compute gradient $\delta C\ \delta Y$ (using Equation 5)
set $Y(t) = Y(t-1) + \eta\ \delta C\ \delta Y + \alpha(t)\ Y(t-1) - Y(t-2)$
end
end

The ALBAE technique is inspired by the foraging behaviour of ant lions, which are known for their efficient and effective

hunting strategies (table 1). This technique can be applied to auto encoders, which are neural networks designed for learning efficient representations of input data.

$$\lambda_p = \frac{\Delta q}{\mu(Q)} \times \tau \tag{9}$$

Where, $\lambda_p$refers to the parameter used for the pre-processing function, s defines the total number of images given to pre-processing. The parameter Δq defines the actual quality of the pre-processed image. However, the parameter $\mu(Q)$ defines the enhanced image quality.

$$Z = f_\theta(A) = S(wtA + b_a) \tag{10}$$
$$A' = g_\theta(Z) = S(wt'A + b_Z) \tag{11}$$

Where $f_\theta(A)\ and\ g_\theta(Z)$ denotes encoded and decoded functions. The parameters wt and wt' are weight matrices of the encoder and decoder, b is biased for input data A and S is an activation function. The decoder function g reconstructs A using the hidden representation of Z. For the A dataset, the auto-encoder learning technique is used to set the parameters b_a and b_z for minimizing reconstructive loss. The target function is shown in equation (12).

$$\theta = minL(A, A') = minL(A, g\left(\left(f(A)\right)\right) \tag{12}$$

The reconstructive loss is obtained from the square errors for linear reconstruction $(L_1)$is derived as,

$$L1(\theta) = \sum_{i=1}^{N} ||a_i - a_i'||^2 \tag{13}$$

The reconstructive loss is obtained from cross entropy for nonlinear reconstruction $(L_2)$is derived as follows in equation (14),

$$L2(\theta) = \sum_{i=1}^{N} [a_i \log(Z_i) + (1 - a_i)\log(1 - Z_i)] \tag{14}$$

Scattered moments of an ant's area,

$$X(t) = 0, c(t_1)2r - 1, C\left((t_1)2r - 1\right), \dots, C((t_1)2r - 1)] \tag{15}$$

$$a(t) = \begin{cases} 0\ if\ rand \leq 0.5 \\ 1\ if\ rand\ > 0.5 \end{cases} \tag{16}$$

Preserving the ant placements in the resulting matrix is the next stage in the optimization process,

$$M_{ant} = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,d} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n,1} & X_{n,2} & \cdots & X_{n,d} \end{bmatrix} \tag{17}$$

Where n indicates the number of ants, the M_{ant} is where the ant is located, the term $k_{i,j}$ which includes the j the variable's value of the ith ant, and the d refers to the adjustable number. It is possible to compare the particles in both this technique and the PSO approach broadly. A fitness term is depicted in equation (18),

$$M_{AL} = \begin{bmatrix} XL_{1,1} & XL_{1,2} & \cdots & XL_{1,d} \\ XL_{2,1} & XL_{2,2} & \cdots & XL_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ XL_{n,1} & XL_{n,2} & \cdots & XL_{n,d} \end{bmatrix} \tag{18}$$

The position of each antlion is suggested to be stored in the MAL matrix in equation (18), and the MOA will preserve the ants' physical condition. Nonetheless, this is evidenced by the fact that ants constantly recalculate their positions,

$$Y_i^t = \frac{(Y_i^t - a_i)(b_i - C_i^t)}{d_i^t - a_i} + C_i \tag{19}$$

$$C^t = \frac{C^t}{g} \tag{20}$$

$$Ant_j^t = Ant_i^t; if\ f(Ant_i^t) > f(Antlion_j^t) \tag{21}$$

$$\lambda_C = \begin{cases} P_n(\Delta\pi) = 1; normal \\ P_n(\Delta\pi) > 1; Lung\ cancer \\ P_n(\Delta\pi) < 1; Pneumonia \end{cases} \tag{22}$$

Classifying the three different categories of images among the data set. This approach is developed for detecting lung cancer effectively. So, the parameter $P_n(\Delta\pi) > 1;$ defines the image data that contains lung cancer.

Table 2: Algorithm for ALBAE technique

Algorithm 2: ALBAE Technique
Input: CT image of lungs **Output:** classification output of the lung disease pre-processing function of the proposed model can be declared through the equation. (22), $\lambda_C = \Delta\alpha / \mu(Q) \times \chi\tau$ The reconstructive loss obtained from cross entropy for non-linear reconstruction (L2) is derived as, $Y_i^t = \frac{(Y_i^t - a_i)(b_i - C_i^t)}{d_i^t - a_i} + C_i$ The constant ratio g is represented as follows: $C^t = \frac{C^t}{g}$ adaptive mechanism increases the possibility of a fresh hunt, as seen by Equation (21) $Ant_j^t = Ant_i^t; if\ f(Ant_i^t) > f(Antlion_j^t)$ $\lambda_C = \begin{cases} P_n(\Delta\pi) = 1; normal \\ P_n(\Delta\pi) > 1; Lung\ cancer \\ P_n(\Delta\pi) < 1; Pneumonia \end{cases}$ Evaluate and test the network Classify the signals and return O_{cs}

Table 2 takes a computed tomography (CT) image of the lungs as input and aims to classify any lung disease present. The first step in the algorithm involves pre-processing the input data using a function which calculates the parameter λ_C. Next, the algorithm calculates the reconstructive loss using cross entropy for non-linear reconstruction (L2), as shown in the equation.

The constant ratio g is then determined, and the algorithm adapts the mechanism based on this ratio to increase the chances of accurate classification. The adaptive mechanism guides the algorithm towards making a fresh prediction. The classification output is based on the evaluation and testing of the neural network, and the final output is O_{cs}, which corresponds to the classification of the signals indicating the presence of lung disease (such as normal, lung cancer, or pneumonia). Overall, this combined approach can enhance the accuracy and efficiency of early diagnosis in small cell lung cancer by uncovering hidden patterns and relationships in the data that may not be apparent through traditional methods. Transforming data using learning can have many motivations. The most common motivations are visualization, compressing the data, and finding a representation that is more informative for further processing. One of the simplest and most widely used algorithms for all of these is principal component analysis. The two algorithms: Linear Discriminant Analysis, commonly used for feature extraction in supervised learning, and t-SNE, which is commonly used for visualization using 2-dimensional scatter plots. Feature extraction is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing. A characteristic of these large data sets is a large number of variables that require a lot of computing resources to process. Feature extraction is the name for methods that select and /or combine variables into features, effectively reducing the amount of data that must be processed, while still accurately and completely describing the original data set.

3.2.2 Densenet121 technique

DenseNet121 is a convolutional neural network architecture that has shown strong performance in image classification tasks. In the context of early diagnosis of small cell lung cancer and pneumonia risk, DenseNet121 can be used in the feature extraction stage to extract relevant features from medical imaging data, such as chest X-rays. The layer generates the feature maps such as $x0$, …, $xl-1$from the upcoming layers, and it is explained in equation (23)

$$X_l = H_l([X_0, X_1, \dots, X_{l-1}]) \quad (23)$$

Where, $Re_i^l$is called the ReLU layer, the output layer is represented as h, FC_i^L is called the FC layer followed by the convolutional layers that evaluate the activation function. The DenseNet is used for feature reuse which is a central concept that results in extremely compact versions. Overall, leveraging DenseNet121 for feature extraction in the context of early diagnosis of small cell lung cancer and pneumonia risk can help improve the accuracy and efficiency of diagnosis, allowing for earlier detection and intervention to improve patient outcomes. DenseNet121 is one of the implementations of the Dense Net network with four dense blocks, and each dense block consists of 6, 12, 24, and 16 dense layers sequentially. Dense Net, short for Dense Convolutional Network, is a deep learning architecture for convolutional neural networks (CNNs). DenseNet revolutionized the field of computer vision by proposing a novel connectivity pattern within CNNs, addressing challenges such as feature reuse, vanishing gradients, and parameter efficiency. Unlike traditional CNN architectures where each layer is connected only to subsequent layers, DenseNet establishes direct connections between all layers within a block. This dense connectivity enables each layer to receive feature maps from all preceding layers as inputs, fostering extensive information flow throughout the network.

Where, $[X_0, X_1, \dots, X_{l-1}]$ is known as the feature maps that are concatenated and produced with layers ranging from 0, ..., $l-1$. An activation function like ReLU is applied to increase the non-linearity at the pooling layer when the feature maps are fed. The Hl function generates the k level-based mapping features that are followed by $l\ th$ the layer which is evaluated using equation (24)

$$H_l = K_0 + k \times (L-1) \quad (24)$$

Where, K_0 is known as the overall number of channels that are present in the input layer. K is known as the hyperparameter having a better growth rate in the network. K Feature maps are added at every layer having their state. The obtained activation functions are fed to the convolutional layers which are expressed as shown in equation (25),

$$g_i^L = b_i^L + \sum_{j-1}^{m_1(L-1)} \Psi_{i,j}^L \times h_j^{L-1} \quad (25)$$

From the above equation (25) g_i^L is known as the output layer represented as L, b_i^L represents the base value, $\Psi_{i,j}^L$ is called the filter connection with feature map, $i\ th$ level feature maps, and j level features. $h_j$is the output layer having $L-1$ features. The model has reduced the unwanted features which solved the overfitting problem,

$$m_1^L = m_1^{L-1} \quad (26)$$

$$m_2^L = \frac{m_2^{L-1} - F(L)}{S^L} + 1 \quad (27)$$

$$m_3^L = \frac{m_3^{L-1} - F(L)}{S^L} + 1 \quad (28)$$

Where, S^L is known as the neural network parameters which change the image movements that are expressed as m_1^L , m_2^L , m_3^L are feature maps that are obtained from the filter. The ReLU and FC are the other layers that present the equation (29)

$$Re_i^l = \max(h, h_i^{L-1}) \quad (29)$$

]

$$FC_i^L = f\left(Z_i^L\right) with\ Z_i^L = \sum_{j-1}^{m_1(l-1)} \sum_{s-1}^{m_2(l-1)} \sum_{s-1}^{m_3(l-1)} w_{i,j,r,s}^l (FC_I^{L-1})_{r,s} \quad (30)$$

3.3 Risk level severity prediction

The combination of the LSTM-GNN model with the Coati Optimization Algorithm offers a cutting-edge approach for predicting pneumonia severity in lung cancer patients. By optimizing hyperparameters and architecture, this model effectively assesses risk levels and severity, enhancing patient care and treatment decisions. Through training on a diverse dataset of medical images, the LSTM-GNN model accurately predicts pneumonia severity, with potential implications for improving outcomes in lung cancer and pneumonia management. This innovative research holds promise for advancing risk assessment and patient care in the context of lung cancer and pneumonia.

3.3.1 LSTM-GNN model with coati optimization algorithm:

The Long Short-Term Memory-Graph Neural Network (LSTM-GNN) model combines the strengths of both LSTM and GNN to achieve better performance in various machine learning tasks. LSTM-GNN, a hybrid model consisting of temporal and graph encoding components the input of LSTM-GNN through the network as follows,

$$L = L_{LSTM-GNN} + \alpha L_{LSTM} \quad (31)$$

Where LLSTM-GNN is the loss on the full model prediction, LLSTM is the loss on the prediction made by the LSTM component LSTM (computed by passing through a distinct fully-connected layer), and α is treated as a hyperparameter. The Coati Optimization Algorithm is a novel optimization technique inspired by the behaviour of coatis, which are social mammals known for their cooperative foraging strategies. This algorithm mimics the hunting behaviour of coatis in search of optimal solutions to complex optimization problems.

$$X_i: x_{ij} = lb_j + r.\left(ub_j - lb_j\right), i = 1,2,\dots,N,\ \ j = 1,2,\dots,m \quad (32)$$

Where, X_i is the position of the ith coati in the search space, x_{ij} is the value of the jth decision variable, N is the number of coatis, m is the number of decision variables, r is a random real number in the interval [0, 1], and lb_j and ub_j.$X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,m} \\ x_{i,1} & \cdots & x_{1,j} & \cdots & x_{i,m} \\ \vdots & & \vdots & & \vdots \\ x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,m} \end{bmatrix}$ (33)

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix} = \begin{bmatrix} F(x_1) \\ \vdots \\ F(x_i) \\ \vdots \\ F(x_N) \end{bmatrix} \quad (34)$$

$$X_i^{p1}: x_{ij}^{p1} = x_{ij} + r.\left(Iguana_j - I.x_{i,j}\right), for\ i = 1,2,\dots,[\frac{N}{2}] \quad (35)$$

Where, $j = 1,2,\dots,m$. Based on this random position, coatis on the ground move in the search space, which is simulated using,

$$Iguana^G: Iguana_j^G = lb_j + r.\left(ub_j - lb_j\right), j = 1,2,\dots m \quad (36)$$

$$X_i^{p1}: x_{i,j}^{p1} = \begin{cases} X_{i,j} + r.\left(Iguana_j^G - I.X_{i,j}\right), F_{Iguana^G} \\ X_{i,j} + r.\left(X_{i,j} - Iguana_j^G\right), else, \end{cases} < F_i \quad (37)$$

For $i = [N/2] + 1, [N/2] + 2, \dots, N$ and $j = 1,2,\dots,m$ (38)

The new position calculated for each coati is acceptable for the update process if it improves the value of the objective function, otherwise, the coati remains in the previous position. This update condition is for $i = 1,2,\dots,N$,

$$X_i = \begin{cases} X_i^{p1}, F_i^{p1} < F_i \\ X_i, \quad else \end{cases} \quad (39)$$

The new position calculated for the ith coating, X_i^{p1} is its jth dimension, F_i^{p1} is its objective function value, $Iguana$ represents the iguana's position in the search space, which refers to the position of the best member, Iguanaj is its jth dimension, I is an integer, which is randomly selected from the set $\{1, 2\}$, $Iguana_j^G$ is the position of the iguana on the ground, which is randomly generated, $Iguana_j^G$is its jth dimension, $Iguana_j^G$ is its value of the objective function, and $\lfloor\cdot\rfloor$ is the floor function (also known as the greatest integer function). The Coati Optimization Algorithm is the fundamental idea of COA is the simulation of the two natural behaviours of coatis: their behaviour when attacking and hunting iguanas and their escape from predators. The implementation steps of COA are described and mathematically modelled in two phases of exploration and exploitation.

Table 3: Coati optimization algorithm

Algorithm 3: Coati Optimization Algorithm
Input: Initialize the population of coatis with random solutions to the optimization problem.
Output: The fitness of each coati in the population based on the objective function of the optimization problem.
Set parameters of N and T.
Set $i = t = 1$. Update the position of each coati in the subgroup using a combination of exploration and exploitation strategies.
Evaluate the fitness of each coati in the subgroup by X_i^{p1}.
Replace the worst coatis in the population with the best coatis from the subgroup.
$Iguana_j^G$ is it the value of the objective function?
Return the best solution found by the coatis as the optimal solution to the optimization problem.
end

The algorithm starts by initializing a population of coatis with random solutions to the optimization problem (Table 3). The fitness of each coati in the population is evaluated based on the objective function of the optimization problem. Parameters N and T are set, and two variables i and t are initialized to 1. The algorithm then updates the position of each coati in a subgroup using a combination of exploration and exploitation strategies. The fitness of each coati in the subgroup is evaluated using X_i^{p1}. The algorithm then replaces the worst coatis in the population with the best coatis from the subgroup. The fitness of each coati in the population is evaluated using the objective function $Iguana_j^G$. Overall, this algorithm uses a population of coatis to explore and exploit solutions to an optimization problem, gradually improving the solutions over iterations.

Overall, the LSTM-GNN model combined with the Coati Optimization Algorithm offers a promising approach for early diagnosis of small-cell lung cancer and pneumonia risk. By leveraging the strengths of both models and algorithms, healthcare providers can improve diagnostic accuracy and facilitate timely interventions for better patient outcomes.

4. EXPERIMENTATION RESULT AND DISCUSSION

The objective is to evaluate the effectiveness of the optimized deep learning approach in accurately predicting small-cell lung cancer and pneumonia risk levels. The Python-based model, trained on a dataset of lung cancer and pneumonia images, demonstrates promising results in accurately predicting the risk levels of these conditions. The model's ability to detect small cell lung cancer and pneumonia early on can significantly impact patient outcomes and treatment decisions. The discussion delves into the potential applications of this approach in clinical settings, highlighting its potential to

improve patient care and contribute to advancements in medical imaging analysis.

Table 4: Simulation system configuration

Python	Version 3.8.0
Operation System	Ubuntu
Memory Capacity	4GB DDR3
Processor	Intel Core i5 @ 3.5GHz

Table 4 provides details about the simulation system configuration used for the study. The system was equipped with Python version 3.8.0, running on an Ubuntu operating system. The system's hardware included a 4GB DDR3 memory capacity and an Intel Core i5 processor clocked at 3.5GHz. This configuration ensured that the system had sufficient resources to run the simulations efficiently and accurately. The use of Python as the programming language allowed for flexibility and ease of implementation, while the Ubuntu operating system provided a stable and reliable environment for conducting the experiments.

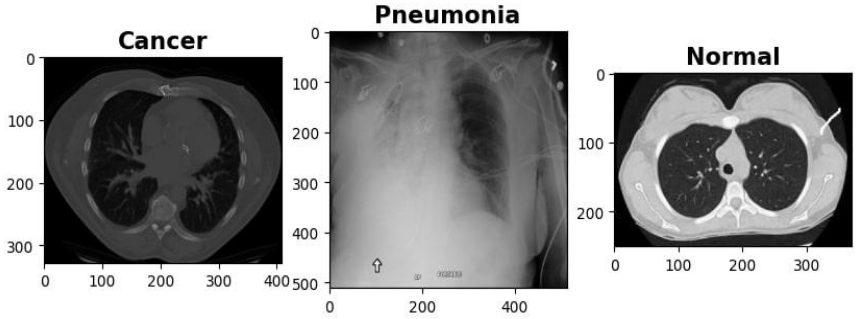


Fig 2: Input images for cancer, pneumonia, and normal classes

Figure 2 depicts the input images for cancer, pneumonia, and normal classes. Each image represents a sample from the respective class, with the colour of each pixel corresponding to the intensity of the image. The images provide a visual representation of the features and characteristics of the samples in each class, which can help understand the differences between the classes and inform the development of classification models.

4.1 Pre-processed results for lung cancer

In the collected dataset, the image undergoes pre-processing using Principal Component Analysis (PCA), a transformation technique that reduces the size of the p-dimensional dataset containing associated variables to a lower-dimensional space with uncorrelated variables. PCA is particularly useful in medical imaging analysis as it can help to reduce the dimensionality of the input data while retaining the most important information. This pre-processing step aids in simplifying the dataset, making it more manageable for subsequent analysis and improving the performance of predictive models by reducing noise and enhancing the visibility of critical features.

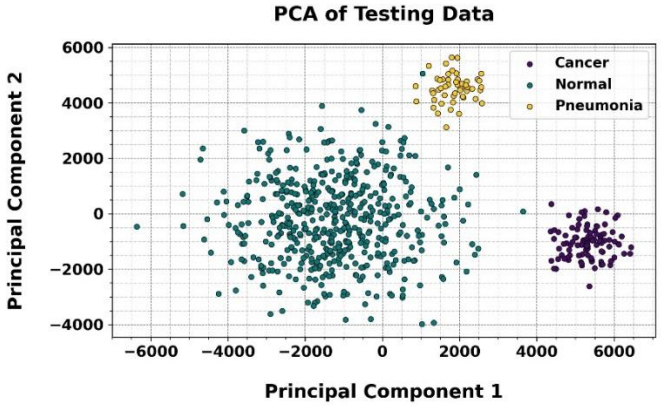


Fig 3: PCA of testing data for cancer, normal, and pneumonia cases

Figure 3 illustrates the PCA of the testing data for cancer, normal, and pneumonia cases. PCA is a dimensionality reduction technique that transforms data into a lower-dimensional space, highlighting the most significant components. To illustrate the novelty of the study, the PCA plot shows the distribution of cancer, normal, and pneumonia cases in a reduced-dimensional space, allowing for visual inspection of the data's structure and potential separation between the different classes. This analysis can help identify patterns and relationships within the data and provide insights into the

effectiveness of the classification model in distinguishing between the different classes.

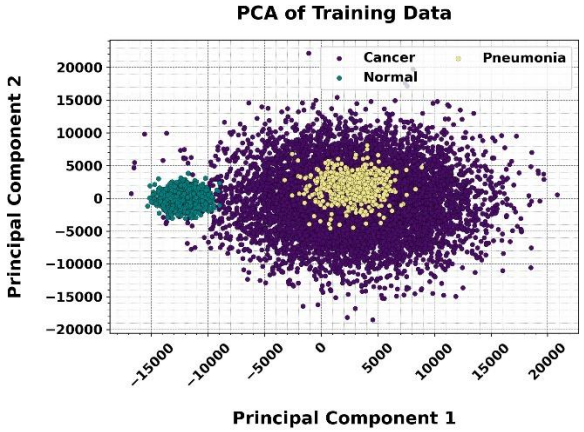


Fig 4: PCA visualization of training data for cancer, normal, and pneumonia classes

Figure 4 illustrates the PCA of the training data for cancer, normal, and pneumonia classes. PCA is a dimensionality reduction technique that transforms the original dataset into a lower-dimensional space while preserving as much of the variance as possible. In this context, PCA is applied to the training data to visualize the distribution of the data points in the reduced space. Each point in the plot represents a sample from the training dataset. The colour of each point corresponds to the class label (cancer, normal, or pneumonia). To illustrate the novelty of the study, provides insights into the separability of the classes in the reduced space, which can aid in understanding the underlying structure of the data and inform the development of classification models.

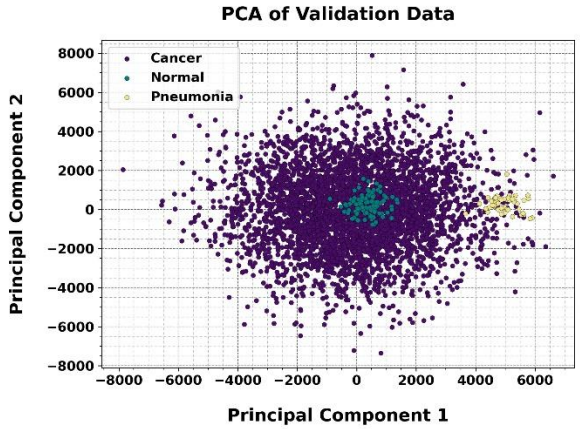


Fig 5: t-SNE **visualization of training, testing, and validation sets**

Figure 5 illustrates the t-distributed Stochastic Neighbour Embedding (t-SNE) visualization of the training, testing, and validation sets. t-SNE is a dimensionality reduction technique that is particularly useful for visualizing high-dimensional data in a lower-dimensional space. The colour of each point corresponds to the class label (cancer, normal, or pneumonia). To illustrate the novelty of the study, insights into the distribution and reparability of the samples in the reduced space can help understand the underlying structure of the data and inform the development of classification models.

.

4.2 Classified and feature-extracted results

In a sophisticated approach, t-SNE and an ALBAE technique are combined. T-SNE is used to reduce high-dimensional image data into a lower-dimensional representation while preserving essential data relationships. This combination leverages the strengths of both methods, allowing for more effective and efficient data analysis and interpretation. In the classification stage, the lung CT images are classified into three different categories such as normal lung, cancer-affected lung, and pneumonia-affected lung using the DenseNet121 technique

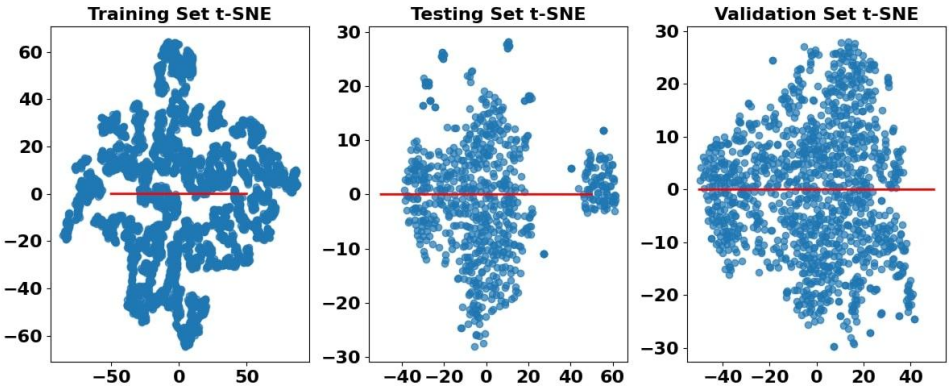


Fig 6: t-SNE visualisation of training, testing, and validation sets for lung cancer classification

Figure 6 illustrates the t-distributed Stochastic Neighbour Embedding (t-SNE) visualization of the training, testing, and validation sets. t-SNE is a dimensionality reduction technique that is particularly useful for visualizing high-dimensional data in a lower-dimensional space. Each point in the plot represents a sample from the respective dataset, with the x-axis and y-axis representing the first two principal components. The colour of each point corresponds to the class label (cancer, normal, or pneumonia). To illustrate the novelty of the study, insight into the distribution and separability of the samples in the reduced space can help understand the underlying structure of the data and inform the development of classification models.

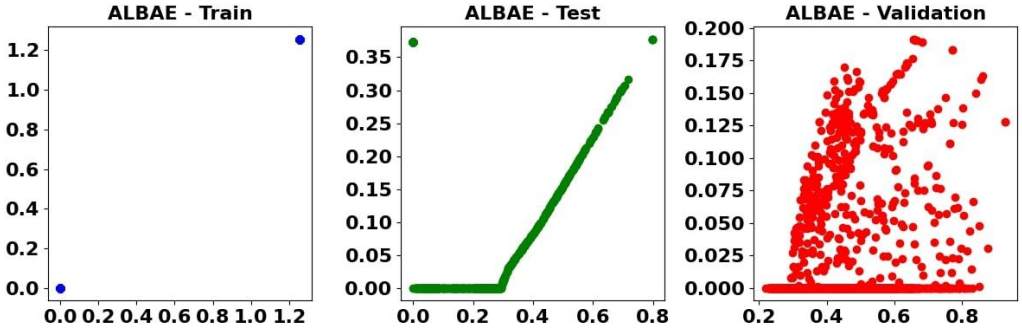


Fig 7: Albae performance on training, test, and validation datasets

Figure 7 illustrates the performance of the ALBAE model on the training, test, and validation datasets. In the training dataset, the ALBAE model achieved an MSE of 0.0438 and an RMSE of 1.1719. On the test dataset, the model demonstrated a stratified MSE of 0.0641 and an end-to-end RMSE of 1.1804. These results indicate that the ALBAE model performed well on both the training and test datasets, with relatively low errors and high accuracy. To illustrate the novelty of the study, validation dataset, which was not used during model training or testing, will provide further insight into the model's generalization and robustness in real-world scenarios.

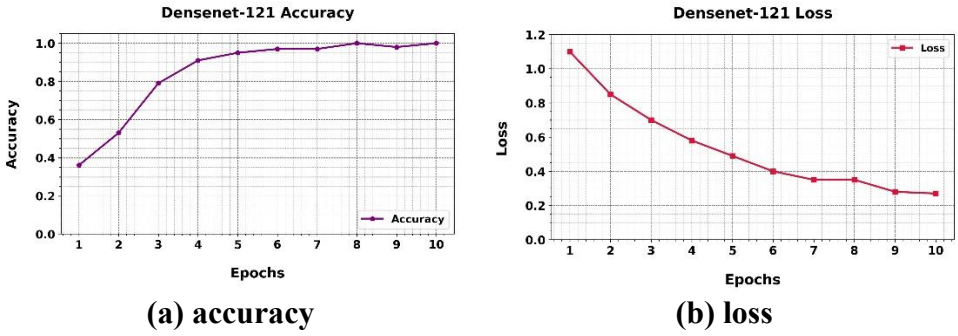
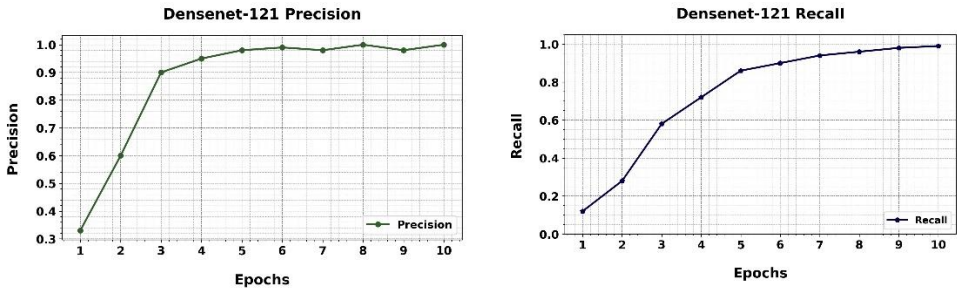


Fig 8: Accuracy and loss performance of the dense net 121 model over epochs

Figure 8 illustrates the performance of the dense net 121 model in terms of accuracy and loss over epochs. The accuracy of the model steadily increases over epochs, reaching a high of 0.9758, indicating that the model is effective in correctly classifying images. The loss of the model decreases over epochs, reaching a low of 0.2364, which suggests that the model is learning to minimize errors and make more accurate predictions. To illustrate the novelty of the study, results indicate that the DenseNet 121 model is performing well and is capable of accurately classifying images, making it a valuable tool for various applications in computer vision and image recognition.



(a) Precision **(b) Recall**

Fig 9: Precision and recall values for dense net 121 model

Figure 9 illustrates the precision and recall values for the dense net 121 model. The precision value of 0.9739 indicates that 97.39% of the positive predictions made by the model were correct, while the recall value of 0.9703 indicates that 97.03% of the actual positive cases were correctly identified by the model. To illustrate the novelty of the study, high precision and recall values suggest that the DenseNet 121 model is effective in accurately identifying positive cases, making it a reliable tool for diagnosing lung cancer and pneumonia.

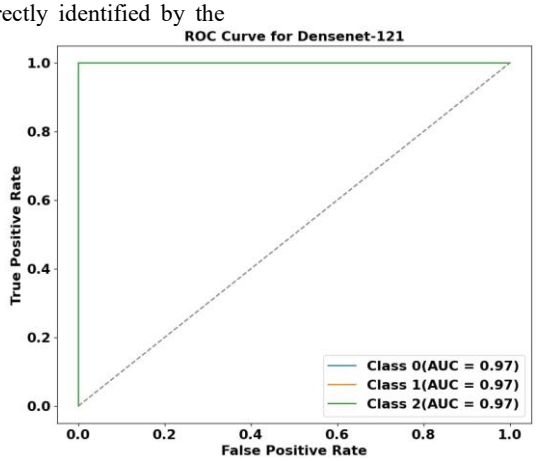


Fig 10: ROC curve for dense net 121 model with AUC values for each class

Figure 10 presents the ROC curve for the dense net 121 model, with the AUC values for each class. For Class 0, the AUC is 0.97, indicating a high level of accuracy in distinguishing between Class 0 and other classes. Similarly, for Class 1 and Class 2, the AUC values are also 0.97, suggesting that the model performs well in classifying these classes. To illustrate the novelty of the study, results demonstrate the robustness and effectiveness of the dense net 121 models in accurately predicting the target variable across different classes.

4.3 Prediction of Risk Level Severity

The LSTM-GNN model is a powerful tool that can help predict pneumonia severity in lung cancer patients, a critical factor that significantly impacts their overall health and treatment decisions. By accurately assessing the severity of pneumonia, healthcare professionals can better tailor treatment plans and interventions, leading to improved patient outcomes and quality of life. The LSTM-GNN model's ability to predict pneumonia severity can also help identify high-risk patients who may require more intensive monitoring or intervention, ultimately contributing to more effective and personalized patient care.

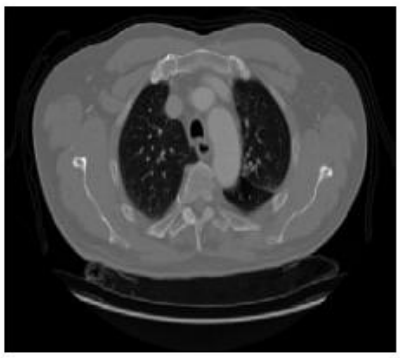


Fig 11: Predicted image for small cell lung cancer

Figure 11 showcases the predicted image for small-cell lung cancer. This image is generated by a predictive model that has been trained on a dataset of lung cancer images. The model has learned to identify features indicative of small-cell lung cancer and has applied this knowledge to predict the input image. To illustrate the novelty of the study, the predicted image provides

a visual representation of the model's output, which can be useful for understanding the model's performance and for further analysis of the features that contribute to the prediction

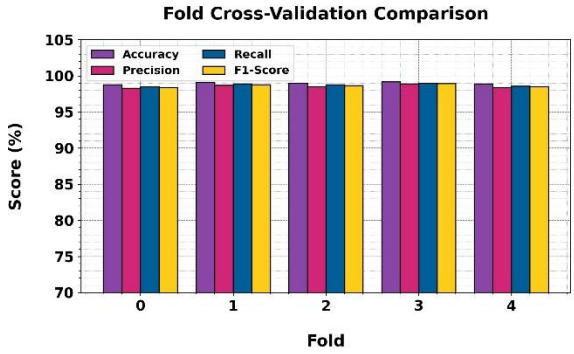


Fig 12: Consistent and robust performance of the model across five-fold cross-validation

Figure 12 shows the fold cross-validation results demonstrating the model's consistent and robust performance across five folds. Accuracy remains very high, averaging 99.0%, while precision (98.58%), recall (98.76%), and F1-score (98.66%) all show strong balance, indicating the model reliably identifies true positives with minimal false positives or false negatives. The small variation across folds highlights the stability and generalization capability of the model.

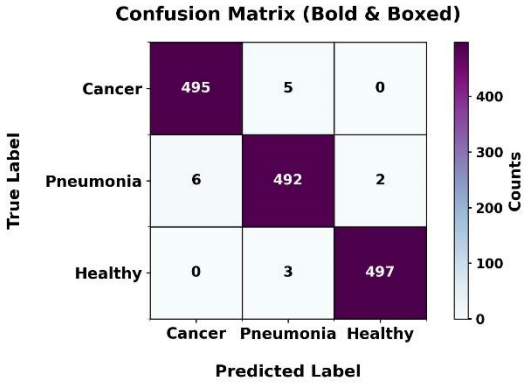


Fig 13: Confusion matrix analysis of cancer, pneumonia, and healthy classification

Figure 13 displays the confusion matrix shows the classification performance of a medical diagnosis model across three classes: Cancer, Pneumonia, and Healthy. Out of 500 cancer cases, the model correctly identified 495 and misclassified 5 as pneumonia, with none labelled healthy. For pneumonia cases, it correctly detected 492 out of 500, misclassifying 6 as cancer and 2 as healthy. For healthy cases, it achieved very high accuracy, correctly predicting 497 out of 500, with only 3 mislabeled as pneumonia and none as cancer. Overall, the model demonstrates excellent accuracy across all categories, with minimal misclassifications, indicating strong reliability in distinguishing between the three conditions.

\

4.4 Comparison analysis

The comparison analysis aims to evaluate the diagnostic accuracy among various techniques for lung cancer and pneumonia detection. This involves examining the performance of different models, such as deep learning-based models like Densenet-121 and CNN, along with other techniques like feature-based methods or traditional machine learning algorithms. By comparing the accuracy, sensitivity, specificity, and other metrics of these techniques, the study can identify the most effective approach for accurately detecting lung cancer and pneumonia. This analysis is critical for improving early diagnosis and treatment outcomes for these conditions.

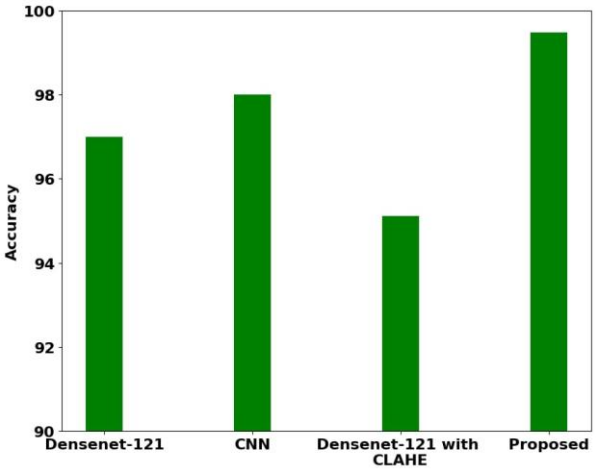


Fig 14: Comparison of diagnostic accuracy among various techniques for lung cancer and pneumonia detection

Figure 14 compares the accuracy of different techniques for diagnosing lung cancer and pneumonia. Densenet-121 achieved an accuracy of 96.9676%, while the CNN model achieved 97.9897%. When Densenet-121 was combined with CLAHE, the accuracy dropped to 95.1107%. To illustrate the novelty of the study, the proposed technique achieved the highest accuracy of 99.4718%. These results suggest that the proposed technique outperforms the other techniques in accurately diagnosing lung cancer and pneumonia.

5. RESEARCH CONCLUSION

This study demonstrates the effectiveness of an optimized deep learning approach for the early diagnosis of small cell lung cancer and pneumonia risk prediction using medical imaging data, including chest X-rays and CT scans. By integrating advanced machine learning techniques such as convolutional neural networks (CNNs), Cox regression models, and the LSTM-GNN architecture, the proposed framework achieved a maximum accuracy of 99%, with the LSTM-GNN model reporting a loss of 0.2349 and an accuracy of 0.9711. These results indicate that the model can reliably differentiate between small cell lung cancer and pneumonia, with predictions closely aligned with ground truth, thereby enhancing the precision and reliability of clinical decision-making. The findings have significant implications for patient care, as early and accurate diagnosis can lead to timely treatment interventions, improved prognosis, and better healthcare outcomes. Furthermore, the proposed framework provides a scalable and adaptable platform for automated medical image analysis, potentially reducing the workload of healthcare professionals while maintaining high diagnostic accuracy. For future work, the study can be extended by incorporating larger and more diverse datasets across multiple hospitals and imaging modalities to improve generalizability. Integration of explainable AI techniques, real-time prediction capabilities, and hybrid models combining imaging data with clinical and genomic information can further enhance the model's accuracy and clinical applicability. Additionally, deployment in real-world clinical settings and continuous performance monitoring will be crucial for validating the model's effectiveness and ensuring its practical utility in routine healthcare workflows. Analysis and lays the groundwork for future research in this field.

Conflict of interest

The authors declare no Conflict of Interest.

Funding statement

The authors received no funding from an external source.

Data availability statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

6. REFERENCES

[1] Forouzannezhad, P., Maes, D., Hippe, D.S., Thammasorn, P., Iranzad, R., Han, J., Duan, C., Liu, X., Wang, S., Chaovalitwongse, W.A. and Zeng, J. Multitask learning radiomics on longitudinal imaging to predict survival outcomes following risk-adaptive chemoradiation for non-small cell lung cancer. Cancers. 14(5), (2022), 1228.

[2] Chao. Y., Zhou, J., Hsu, S., Ding, N., Li, J., Zhang Y., Xu, X., Tang, X., Wei, T., Zhu, Z. and Chu, Q. Risk factors for immune checkpoint inhibitor-related pneumonitis in non-small cell lung cancer. Translational Lung Cancer Research. (2022), 11(2), 295.

[3] Matsuura, S., Morikawa, K., Ito, Y., Kubota, T., Ichijo, K., Mochizuki, E., Akiyama, N., Uehara, M., Harada, M., Tsukui, M, and, Koshimizu, N. The geriatric nutritional risk index and prognostic nutritional index predict the overall survival of advanced non-small cell lung cancer patients. Nutrition and Cancer. 74(5), (2022), 1606-1613.

[4] Wang, S., Yu, H., Gan, Y., Wu, Z., Li, E., Li, X., Cao, J., Zhu, Y., Wang, L., Deng, H, and, Xie, M. Mining whole-lung information by artificial intelligence for predicting EGFR genotype and targeted therapy response in lung cancer, a multicohort study. The Lancet Digital Health. 4(5), (2022), 309-e319.

[5] Yang, Y., Xu, L., Sun, L., Zhang, P, and, Farid, SS. Machine learning application in personalized lung cancer recurrence and survivability prediction. Computational and Structural Biotechnology Journal. 20, (2022), 1811-1820.

[6] Ye, Z., Zhang, Y., Liang, Y., Lang, J., Zhang, X., Zang, G., Yuan, D., Tian, G., Xiao., M, and, Yang, J. Cervical cancer metastasis and recurrence risk prediction based on deep convolutional neural network. Current Bioinformatics. 17(2), (2022), 164-173.

[7] Huang, H., Li, L., Luo, W., Yang, Y., Ni, Y., Song, T., Zhu, Y., Yang, Y, and, Zhang, L. Lymphocyte percentage as a valuable predictor of prognosis in lung cancer. Journal

of Cellular and Molecular Medicine. (2022), 26(7),1918-1931.

[8] Dritsas., E, and Trigka, M. Lung cancer risk prediction with machine learning models. Big Data and Cognitive Computing. (2022), 6(4):139.

[9] Guo, LW., Liu, ZY., Meng, QC., Zheng, LY., Chen, Q., Liu, Y., Xu, HF., Kang, RH., Zhang, LY., Cao, XQ, and, Liu, SZ. Construction and validation of a lung cancer risk prediction model for non-smokers in China. Frontiers in Oncology. 11, (2022), 766939.

[10] Hou, KY., Chen, JR., Wang, YC., Chiu, MH., Lin, SP., Mo, YH., Peng, SC, and, Lu, CF. Radiomics-based deep learning prediction of overall survival in non-small-cell lung cancer using contrast-enhanced computed tomography. Cancers. 14(15), (2022), 3798.

[11] Ruan, Y., Han, J., Yang, A., Ding, Q, and, Zhang, T. Impact of Preoperative Inflammatory Indices and Postoperative Pneumonia on Postoperative Atrial Fibrillation in Patients with Non-Small Cell Lung Cancer, A Retrospective Study, (2024).

[12] Wu, Y., Rocha, BM., Kaimakamis, E., Cheimariotis, GA., Petmezas, G., Chatzis, E., Kilintzis, V., Stefanopoulos, L., Pessoa, D., Marques, A, and, Carvalho, P. A deep learning method for predicting the COVID-19 ICU patient outcome fusing X-rays, respiratory sounds, and ICU parameters. Expert Systems with Applications. 235, (2024), 121089.

[13] Chen, Y., Froelich, MF., Tharmaseelan, H., Jiang, H., Wang, Y., Li, H., Tao., M., Gao., Y., Wang., J., Liu., J and Schoenberg, SO. Computed tomography imaging phenotypes of hepatoblastoma identified from radionics signatures are associated with the efficacy of neoadjuvant chemotherapy. Pediatric Radiology. 54(1), (2024), 58-67.

[14] Cervia-Hasler, C., Brüningk, SC., Hoch, T., Fan, B., Muzio, G., Thompson, RC., Ceglarek, L., Meledin, R., Westermann, P., Emmenegger, M, and, Taeschler, P. Persistent complement dysregulation with signs of thromboinflammation in active Long Covid. Science. 383(6680), (2024), 7942.

[15] Jiang, X., Luo, C., Peng, X., Zhang, J., Yang, L., Liu, LZ., Cui, YF., Liu, MW., Miao, L., Jiang, JM, and, Ren, JL. Incidence Rate of Occult Lymph Node Metastasis in Clinical T 1-2 N 0 M 0 Small Cell Lung Cancer Patients and Radiomics Prediction Based on Contrast-enhanced CT Imaging, A Multicentre Study. (2024).

[16] Lakshmi, PN, and, Vedavathi, K. Deep Learning and Probabilistic Neural Networks Based Detection and Classification of Lung Diseases for Pneumonia. International Journal of Intelligent Systems and Applications in Engineering. 12(12s), (2024), 708-713.

[17] Nahiduzzaman, M., Abdulrazak, LF., Ayari, MA., Khandakar, A, and, Islam, SRA. Novel framework for lung cancer classification using lightweight convolutional neural networks and ridge extreme learning machine model with Shapley Additive explanations (SHAP). Expert Systems with Applications. 248, (2024), 123392.

[18] Mikhael, PG., Wohlwend, J., Yala, A., Karstens, L., Xiang, J., Takigami, AK., Bourgouin, PP., Chan., P., Mrah., S., Amayri., W, and, Juan., YH, Sybil. A validated deep learning model to predict future lung cancer risk from a single low-dose chest computed tomography. Journal of Clinical Oncology. 41(12), (2023), 2191-2200.

[19] Oh, J., Park, C., Lee, H., Rim, B., Kim, Y., Hong, M., Lyu, J., Han, S, and, Choi, S. OView-AI Supporter for Classifying Pneumonia, Pneumothorax, Tuberculosis, and Lung Cancer Chest X-ray Images Using Multi-Stage super pixel classification. Diagnostics. 13(9), (2023), 1519.

[20] Du, W., Luo, X, and, Chen, M. A practical deep learning model in differentiating pneumonia-type lung carcinoma from pneumonia on CT images, ResNet added with an attention mechanism. Journal of Oncology. (2022).

[21] Yin, X., Liao, H., Yun, H., Lin, N., Li, S., Xiang, Y, and, Ma, X. Artificial intelligence-based prediction of clinical outcome in immunotherapy and targeted therapy of lung cancer. In Seminars in cancer biology. Academic Press. (2022).

[22] Guo, LW., Liu, ZY., Meng, QC., Zheng, LY., Chen, Q., Liu, Y., Xu, HF., Kang, RH., Zhang, LY., Cao, XQ, and, Liu, SZ. A risk prediction model for selecting high-risk population for computed tomography lung cancer screening in China. Lung Cancer. 163 (2022), 27-34.

[23] Li, L., Yang, J., Por, L.Y., Khan, M.S., Hamdaoui, R., Hussain, L., Iqbal, Z., Rotaru, I.M., Dobrotă, D., Aldrdery, M. and Omar, A. Enhancing lung cancer detection through hybrid features and machine learning hyper parameters optimization techniques. Heliyon. (2024).

[24] Lin, CY., Guo, SM., Lien, JJJ., Lin, WT., Liu, YS., Lai, CH., Hsu., IL., Chang, CC, and, Tseng, YL. Combined model integrating deep learning, radiomics, and clinical data to classify lung nodules at chest CT. La radio logia medica. 129(1), (2024), 56-69.

[25] Feng, A., Huang, Y., Zeng, Y., Shao, Y., Wang, H., Chen, H., Gu, H., Duan, Y., Shen, Z, and, Xu, Z. Improvement of Prediction Performance for Radiation Pneumonitis by Using 3-Dimensional Dosiomic Features. Clinical Lung Cancer. (2024).

[26] Dhruba., Panda, C., Prabhat, K., Patnaik. CAD Modeling of Complex Resonant Frequencies of a Rectangular Microstrip Patch with a Superstrate using Complex Backpropagation Algorithm. (2013).

[27] Prabhat, K., Patnaik., Dhruba., Panda C. Fast Extraction of L & C Parameters of MEMS. Transmission Line using Neural Network. (2013).

[28] Vamshi, Krishna, M, Different Fractal Antenna Structure Analysis using ANN. International Journal of Innovative Technology and Exploring Engineering. 9(5), (2020).

[29] Prabhat, K., Patnaik, Dhruba, C, Panda Santosh Kumar, Pantina. Digital combinational circuit optimization using invasive weed optimization technique. Lat. Am. J. Phys. Educ. Sept. 8(3), (2014).

[30] Patnaik, PK., Malijeddi, M., and, Panda, DC. Wearable Microstrip Patch Antenna for Disease Detection and WiMAX Application. 2021 2nd International Conference on Range Technology (ICORT), Chandipur, Balasore, India. (2021), 1-4. doi, 10.1109/ICORT52730.2021.9582104. 2021.

[31], Sofi, S., Jan, N., Qayoom, H., Alkhanani, M., Almilaibary, A., Ahmad, Mir, M. Elucidation of interleukin-19 as a therapeutic target for breast cancer by computational analysis and experimental validation. Saudi

J Biol Sci. (2023), 30(9), 103774. doi, 10.1016/j.sjbs. (2023).103774. Epub 2023 Aug 11. PMID, 37675062; PMCID, PMC10477739.

[32] Qayoom, H., Alshehri, B., Ul, Haq, B., Almilaibary, A., Alkhanani, M., Ahmad, Mir, M. Decoding the molecular mechanism of stypoldione against breast cancer through network pharmacology and experimental validation. Saudi J Biol Sci. 30(12), (2023), 103848. doi, 10.1016/j.sjbs. (2023).103848. Epub 2023 Oct 21. PMID, 37964781; PMCID, PMC10641555.

[33] Qayoom H, Alkhanani M, Almilaibary A, Alsagaby SA, Mir, MA. Mechanistic elucidation of Juglanthraquinone C targeting breast Cancer, A network Pharmacology-based investigation. Saudi J Biol Sci. 30(7), 103705. doi, 10.1016/j.sjbs. (2023).103705. Epub 2023 Jun 15. PMID, 37425621; PMCID, PMC10329161.