International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

# Atrous Convolutional Self-Attention-Based Capsule Network for Lung Disease Classification

[1]Praveena Kakarla, [2]Vimala C*

Department of Electronics and Communication Engineering, College of Engineering and Technology,
SRM Institute of Science and Technology, Kattankulathur, Chengalpattu Dt, Tamil Nadu 603203, India.
[1]pk8112@srmist.edu.in | ORCID: 0000-0001-8044-5648, [2]vimalac@srmist.edu.in | ORCID: 0000-0003-1724-9323

*Corresponding Author: Vimala C*

**Abstract:** Lung diseases remain a major global health concern, affecting the supply of oxygen to other parts of the body. There are several types of lung diseases, including asthma, COPD, and pneumonia. Numerous methods have been developed to identify these lung diseases; however, they still have several shortcomings, including long processing times, complex structures, and poor classification accuracy. To address these issues, a lung disease classification system is developed using the proposed method. First, pre-processing techniques are applied to improve image quality by reducing noise and enhancing contrast using Modified Histogram Equalization and Cross-guided Bilateral Filtering. The images are collected from the NIH ChestX-ray dataset. Next, Extended Lyrebird Optimization is utilized to select the optimal features, and the Squeeze-Excited DenseNet201 (SE-DenseNet201) model is employed for feature extraction. Finally, an Atrous Convolutional Self-Attention-based Capsule Network model is utilized for classification, and the Kookaburra Optimization Algorithm is employed for hyperparameter tuning. The proposed approach is evaluated using the NIH ChestX-ray dataset and achieves an accuracy of 92.70%, with 92.13% precision and 92.99% recall.

**Keywords:** Lung disease classification, Chest X-ray imaging, DenseNet201, Capsule network, Feature selection, Metaheuristic optimization.

## 1 INTRODUCTION

Lung diseases are among the major causes of mortality worldwide and affect people in different ways [1]. These illnesses may be caused by infectious organisms such as viruses, bacteria, and fungi [2][3]. A primary challenge is the effective prevention and treatment of chronic lung diseases such as COPD, asthma, and lung cancer, which place a significant strain on the healthcare system. There are specialized tests that can be used to accurately diagnose and monitor lung diseases [4][5]. Chest X-rays (CXR) and computed tomography (CT) are often used to detect various lung diseases [6]. However, radiologists may experience visual fatigue, which may hinder the detection of small lesions, and manual diagnosis can be time-consuming [7]. Therefore, computerized diagnosis based on artificial intelligence (AI) has become increasingly important for detecting lung and other chest diseases [8][9].

Thus, researchers are exploring deep learning (DL)-based methods to automate the detection of pulmonary lung disease. DL methods are widely applied to analyze CT and other medical imaging data. However, research is ongoing to better understand how these DL methods achieve classification results, and which features are critical for producing accurate outputs [10]. Researchers often use statistical methods and descriptors to extract features from lung regions in MRI and CT images. However, extra processing time may be required if these methods extract improper and superfluous data [11]. Also, many preprocessing strategies are used for lung disease categorization. One of the most efficient methods was the EMDB filter, which combined empirical mode decomposition with bilateral filtering to address challenges in lung CT images [12]. The lung region is automatically deleted based on the pre-processed results using a unique grey correlation-based k-means clustering algorithm [13]. Recently, the DL method has mostly used convolutional neural networks (CNNs). Several lung imaging applications use convolutional neural networks for tasks such as feature extraction and image classification [14].

CNNs are commonly employed to classify generated spectrograms. These networks quickly extract features from lung sounds, learn to detect patterns, and perform exceptionally well in image classification tests. As a result, CNNs work effectively for tasks such as object detection, image segmentation, and categorization [15]. Developing an effective lung disease classification model is important for assisting clinical diagnosis. It analyzes acoustic signals from the lungs to discriminate between respiratory disorders such as asthma, pneumonia, and chronic obstructive pulmonary disease [16][17]. Conventional techniques for classifying lung sounds rely on laborious feature extraction and basic ML algorithms; nevertheless, these methods often yield low classification accuracy. Recent advancements in deep learning technologies have significantly influenced the classification of lung sounds. DL methods have the potential to directly extract features from unprocessed data, leading to highly expressive classification models that improve accuracy and stability [18].

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

Metaheuristic algorithms are computational search strategies that mimic intelligent behaviour to find optimal solutions in large and complex problems [19]. When these algorithms are combined with CNNs, they further enhance CNNs' ability to learn and recognize patterns in data. This improvement can be especially useful for identifying diseases such as lung cancer. Additionally, using techniques such as wavelet decomposition to enhance image quality can make the overall process more effective [20].

### 1.1. Motivation and Problem Statement

Over five million people die every year from lung cancer, which represents a significant global health burden. It's one of the main reasons people die around the world, and it happens to both men and women. Numerous studies on the early prediction of lung cancer by computer vision and soft computing techniques have been reported. Lung diseases have been detected using medical imaging methods for a long time, including CT, magnetic resonance imaging (MRI), and isotopes. Radiologists and doctors commonly use X-ray and CT images to diagnose lung diseases. On chest X-ray images, consolidations, blunted costophrenic angles, widely dispersed nodules, cavitation, and infiltrates were identified as signs of lung infection. Thus, radiologists use X-ray images to diagnose a variety of disorders, including pneumonia, pleurisy, nodules, effusion, infiltration, fractures, pneumothorax, and pericarditis. For radiologists, diagnosing and categorizing lung diseases from chest X-ray images is complex. Numerous computer-aided diagnostic systems have been proposed over the past decade to use X-ray images to diagnose lung diseases.

However, these systems were unable to meet the performance standards needed for the identification and categorization of lung diseases. Pneumonia patients are treated appropriately under this categorization. Several papers have discussed the use of CAD systems for COVID-19. These systems use deep learning and automated image-processing methods to identify pneumonia in chest X-ray images. DL models require longer training time on large datasets because they are fully automated feature-learning and extraction methods. As a result, many methods remain stable and reliable even when applied to increasingly large datasets. Deep learning methods like CNNs are popular for diagnosing lung diseases because they're good at finding patterns and are accurate.

However, a problem remains: these methods are computationally complex and slow. This is because chest X-ray images contain complex visual patterns that increase the computational complexity for CNN-based models. Moreover, most available models perform poorly when dealing with multiple classes in a dataset. The models' adaptability to the problem considered is also inadequate, necessitating an effective technique to better address these issues. To address these issues, this paper proposes a DL-based method for identifying lung diseases with improved classification performance. It incorporates a meta-heuristic technique for feature selection, which helps reduce computation time and enhance the model's performance.

### 1.2. Contributions of Research

The main contributions of this work are summarized as follows:

- A hybrid deep learning framework is proposed for lung disease classification that integrates preprocessing, feature extraction, feature selection, and classification within a unified pipeline.
- An attention-enhanced capsule network (AC-SACNet) architecture is introduced that combines atrous convolution with a self-attention mechanism to capture both local spatial features and global contextual relationships in chest X-ray images.
- An Extended Lyrebird Optimization (ELO) algorithm is applied for feature selection to reduce feature redundancy and improve classification performance.
- The Kookaburra Optimization Algorithm (KOA) is employed to fine-tune key hyperparameters of the classification model, improving training stability and predictive performance.
- Extensive experiments conducted on the NIH ChestX-ray dataset demonstrate the effectiveness of the proposed approach compared with several existing deep learning methods.

The remainder of the paper is organized as follows: Section 2 presents a survey of related techniques for lung disease classification. Section 3 provides a brief description and images of the entire procedure used for lung disease classification in the proposed model. In Section 4, the overall performance of the proposed model is analyzed using existing techniques and presented with graphs. Then, the conclusion of the suggested approach is presented in Section 5.

## 2 Related Work

Anwesh Kabiraj et al. [21] developed a Chest X-ray Ultranet (CX-Ultranet) to categorize thirteen thoracic lung disorders using a multiclass cross-entropy loss function based on a compound scaling framework called EfficientNet. This technique achieved an average prediction accuracy of 88% on the NIH Chest X-ray dataset. This technique addresses class imbalance in datasets but requires significant training time. This technique was applied to single-disease pneumonia to assess the model's reproducibility. Maksim Kholiavchenko et al. [22] developed a ResNet34-based neural network to classify chest disease. It is used to train the input chest X-ray map using the output physician gaze map and the binary pathology label.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

The diagnostic accuracy of the ResNet34-based neural network technique was reported with an area under the receiver operating characteristic curve (AUC) of 0.714. This technique analyzed samples from three datasets, such as RSNA, SIIM-ACR, and CheXpert public databases, with 14 different pathology types. Wang et al. [23] studied chronic obstructive pulmonary disease (COPD) as a global health problem. Its gradual and hidden progression highlights the critical need for early detection and precise diagnosis. CT imaging has become an important diagnostic tool. The advent of AI has the potential to improve the accuracy and efficiency of COPD diagnosis, particularly in CT imaging. This combination of healthcare and technology marks a significant shift in the multiscale residual networks of COPD management. The dispersed nature of electronic health records (EHRs) across various healthcare systems and the potential absence of standard forms might lead to inconsistent data, which in turn can affect the quality of CT data.

Deng et al. [24] presented that COPD is a common lung condition that results in limited airflow and respiratory difficulties. It imposes a substantial impact on health, the economy, and society. Lung segment images are used for autonomous 1,781 × 2 lung radiomics and 13,824 × 2 3D CNNs, segmented from IN and EX chest CT images to extract features. Radiomics and feature selection have been used to diagnose COPD stages using 3D CNN features. Finally, an auto-metric graph neural network is used to diagnose COPD status by combining radiomics, 3D CNN features, and risk factors. However, the datasets utilized were collected from limited clinical sources. However, the meta-learning technique enables continual learning with fresh datasets, reducing the need to practice a new one. The use of an auto-metric graph neural network (AMGNN) should concentrate on the combined input of training and test data.

Souid et al. [25] developed an efficient method, called modified MobileNet V2 classification, for estimating lung pathology from frontal thoracic X-rays. This study utilized the NIH Chest-Xray-14 database and compared the performance of MobileNet V2 with other state-of-the-art methods for pathology organization. However, the model exhibits low sensitivity, which may increase the incidence of false negatives in sample classification. This reduced sensitivity is attributed to bias in the data. Table 1 summarizes the existing related work. Another deep learning-based multiclass lung disease classification model was introduced by Kim et al. [26] that accurately classified lung diseases from chest X-ray images. An end-to-end learning procedure known as EfficientNet v2-M was created. Raw pictures were delivered directly to the model, which extracted and recognized the features. The model was evaluated on three classes, including normal, pneumonia, and pneumothorax, which belong to the NIH dataset, and the approach resulted in an overall accuracy value of 82.15%. The model was also tested on the SCH dataset, achieving an accuracy of 82.20%.

Zaidi et al. [27] presented a customized classification model for TB detection using the NIH-CXR dataset. A new customized CNN architecture was developed to perform the classification task for three crucial experiments, including

i) image classification as either TB-infected or normal,
ii) sub-classification of the infected images based on specific manifestations, and
iii) classification of the image to thoracic disease manifestations.

The approach was tested on the NIH dataset and outperformed existing methods, achieving a higher AUC. Nawaz et al. [28] presented a deep learning-based chest disease detection and classification model, with the primary aim of overcoming chest artifacts and inter- and intra-class similarities. The approach, named EfficientDet, was based on the EfficientNet-B0 model. It was built to classify eight categories of chest abnormalities from the input X-ray images. The model was claimed to offer effective feature-computation power, thereby yielding better results in abnormal recognition. In experiments with the NIH dataset, the model achieved an AUC score of 90.8%.

Wu et al. [29] presented a deep learning-based classification approach for rapid screening of lung disease using digital images. The presented network model, a multilayer fractional-order machine vision classifier, was capable of automatically identifying the class of lung disease within a bounding region of interest (ROI). Initially, texture analysis was performed using the fractional-order convolution (FOC) algorithm, which reduced sounds and improved the symptomatic feature. The dimensions of the feature maps were then reduced using max pooling, thereby accelerating computationally expensive operations. Furthermore, the radial Bayesian network and gray relational analysis were used to screen for typical lung diseases among participants. For digital images, the approach achieved an overall accuracy of 83.57% in experiments.

Existing lung disease classification techniques suffer from several drawbacks, including limited data availability, high computational complexity, misclassification, and relatively high false detection rates. Therefore, an Atrous convolutional self-attention-based capsule network model is introduced to address some of the limitations of existing approaches. Therefore, the proposed approach aims to address the limitations of traditional methods.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

Table 1. Survey of related techniques

| Work | Technique used | Performance | Limitation |
|---|---|---|---|
| Anwesh Kabiraj et al. [21] | CX-Ultranet | 88% average prediction accuracy | This technique was tested only for single-disease pneumonia |
| Maksim Kholiavchenko et al. [22] | ResNet34-based neural network | 0.714 - AUC | It analyzed only a limited portion of the dataset. |
| Fangfei Wang et al. [23] | Multiscale residual network | 91.99% - accuracy | High computation time |
| Xingguang Deng et al. [24] | AMGNN | 89.7% - accuracy | It requires diligent attention to the joint use of training and testing data. |
| Abdelbaki Souid et al. [25] | Modified MobileNetV2 model | 0.811 - AUC | Low sensitivity rate, and it enhances the false negative classification |
| Kim et al. [26] | EfficientNet v2-M | 82.15% - accuracy, 81.40% - sensitivity, 91.65% - specificity | The model resulted in overfitting and also lost a lot of meaningful features |
| Zaidi et al. [27] | Customized CNN | 82% - accuracy, 92% - F1-score | The accuracy of this model was highly influenced by the visual quality of the images. |
| Nawaz et al. [28] | EfficientDet | 90.8% - AUC | Images with a blur effect in the dataset resulted in lower performance |
| Wu et al. [29] | Multilayer machine vision classifier | 82.42% - precision, 98.68% - recall, 89.81% - F1-score, 83.57% - accuracy | The model is affected by the artifacts and visual impacts of chest X-ray images |

## 3 Proposed Methodology

Lung diseases are among the life-threatening conditions worldwide and include pneumonia, chronic obstructive pulmonary disease, asthma, and fibrosis. Timely detection of lung diseases is important for improving patient survival. Thus, the proposed study aims to detect and classify various lung diseases using a hybrid deep learning framework combined with a metaheuristic feature selection algorithm. Fig. 1 illustrates the workflow of the proposed method.

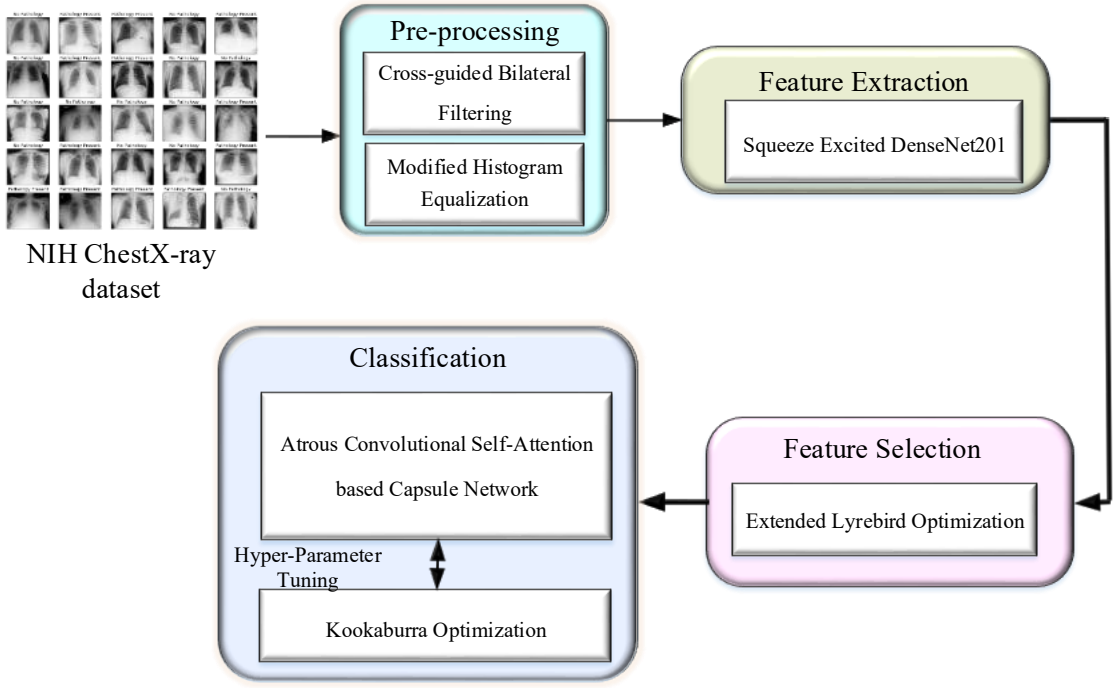


Fig. 1. Architecture of the proposed method

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

Initially, the input images are collected from the NIH Chest X-ray dataset available on Kaggle. These images undergo a preprocessing stage to improve their quality, using Cross-guided Bilateral Filtering (CGBF) and Modified Histogram Equalization (MHE). Once pre-processed, the images are fed into an SEDenseNet201 model to extract essential features. From these extracted features, the most optimal ones are selected using a novel meta-heuristic approach called ELO. Finally, lung diseases are classified based on the selected features through a proposed AC-SACNet model. The constraints of this suggested approach are fine-tuned using the KOA. A detailed description of the proposed method is provided below.

### 3.1. Pre-processing

The initial step in this approach involves collecting lung disease images from the dataset and pre-processing them to enhance their quality. This pre-processing employs CGBF and MHE to remove noise and improve the clarity of image features.

#### 3.1.1. CGBF

CGBF is utilized to remove noise from lung disease images. This method is a combination of the directed bilateral filter [31] and the cross-bilateral filter [30]. It successfully removes artifacts while preserving edge information in the image. This technique uses pixel-level geometric and grayscale similarity to preserve smooth regions, similar to a two-dimensional filter. The kernel is determined by filtering one image in the direction of another image and vice versa. This approach effectively circumvents the primary disadvantage of the bilateral filter: gradient reversal. For the input image $S$, the guide weight $(G_w)$ is multiplied by the guide image $(R)$ in the cross-bilateral filter method. The mathematical procedure for a cross-guided bilateral filter is given by equation (1).

$$R_x = \frac{1}{G}\sum_{y\in c} K_{\sigma c}\left(\|x-y\|\right)K_{\sigma d}G_w|H(x)-H(y)|S_x \tag{1}$$

where, $G$ is defined using the following equation (2).

$$G = \sum_{y\in c} K_{\sigma c}\left(\|x-y\|\right)K_{\sigma d}G_w|H(x)-H(y)| \tag{2}$$

where the two weight parameters such as $K_{\sigma d}$ is calculated over the guided image and $K_{\sigma c}$ is a Cross-guide bilateral filter. Both spatial and range (intensity/colours) kernels are Gaussian–shaped and are defined using equations (3) and (4).

$$K_{\sigma c}(\|x-y\|) = z^{\frac{-\|x-y\|^2}{2\sigma_c^2}} \tag{3}$$

$$K_{\sigma d}(|H(x)-H(y)|) = z^{\frac{-|H(x)-H(y)|^2}{2\sigma_d^2}} \tag{4}$$

where, $\sigma_c$ and $\sigma_d$ are the standard deviations, $H(x)$ and $H(y)$ are the guide image pixel colours of $x$and$y$. Following the Euclidean distance between $x$ and $y$, the weight values of the spatial kernels were computed. $c$ represents the spatial domain, which is the set of all possible image locations.

#### 3.1.2. MHE

A variant of the histogram equalization [32] method, MHE improves image contrast, especially in low-contrast or unevenly illuminated areas. The MHE function is tailored to specific image regions based on local image statistics, whereas classic histogram equalization applies the same equalization algorithm across different regions. The image is divided into small regions or tiles, and each tile's histogram is computed and balanced separately as part of the improved histogram equalization technique. This guarantees that the contrast enhancement is applied only to the areas that need it rather than the entire image. This method allows for MHE, which preserves the image's local contrast while increasing its overall contrast. The following equation represents the MHE technique:

$$M(a) = T(r(a)) = (E-1)\sum_{j=0}^{r(a)} p_r\,(j) \tag{5}$$

where, $M_{(a)}$ represented as the modified pixel value, $T$ represented as the MHE transformation, and $r(a)$ is signified as the original pixel value. $E$ is represented as the number of possible pixel values (256 for 8-bit grayscale images), $h_{k(j)}$ is denoted as the number of pixels in the $k^{th}$ tile and with a value of $j$. The MHE technique has several advantages over traditional histogram equalization. By dividing the image into small tiles, MHE can adapt to local changes in image statistics, resulting in a more uniform contrast enhancement across the entire image. Additionally, the MHE can preserve the local contrast of the image, which is essential for maintaining the image's natural appearance.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

The techniques above efficiently eliminate noise and improve image contrast. CGBF preserves edge integrity while effectively removing artifacts. MHE enhances contrast at targeted locations, resulting in clearer, more detailed images of lung diseases.

### 3.2. Feature Extraction Using SEDenseNet201 Model

The pre-processed image is fed into the feature extraction model, which utilizes SEDenseNet201 to extract various features. Owing to its strong feature reuse and performance in medical image analysis, particularly in chest X-ray analysis, the DenseNet201 model was selected for feature extraction, despite its relatively high computational cost. This is because each layer can access the feature maps of all preceding layers through dense connections, which improves gradient flow, prevents vanishing gradients, and enables the network to capture fine pathological details such as subtle opacities and textures. Conversely, MobileNetV3 and EfficientNet-B0/B1 models are relatively lightweight and primarily designed for mobile platforms and low computational cost, which may lead to reduced representational capability when identifying low-contrast abnormalities and high intra-class variability, which are often common in chest X-ray images. Although it has strong capabilities, ConvNeXt is relatively deeper and requires higher memory, making it less suitable for stable training in the selected experimental setup. Fig. 2 illustrates the structure of the proposed SEDenseNet201 model.

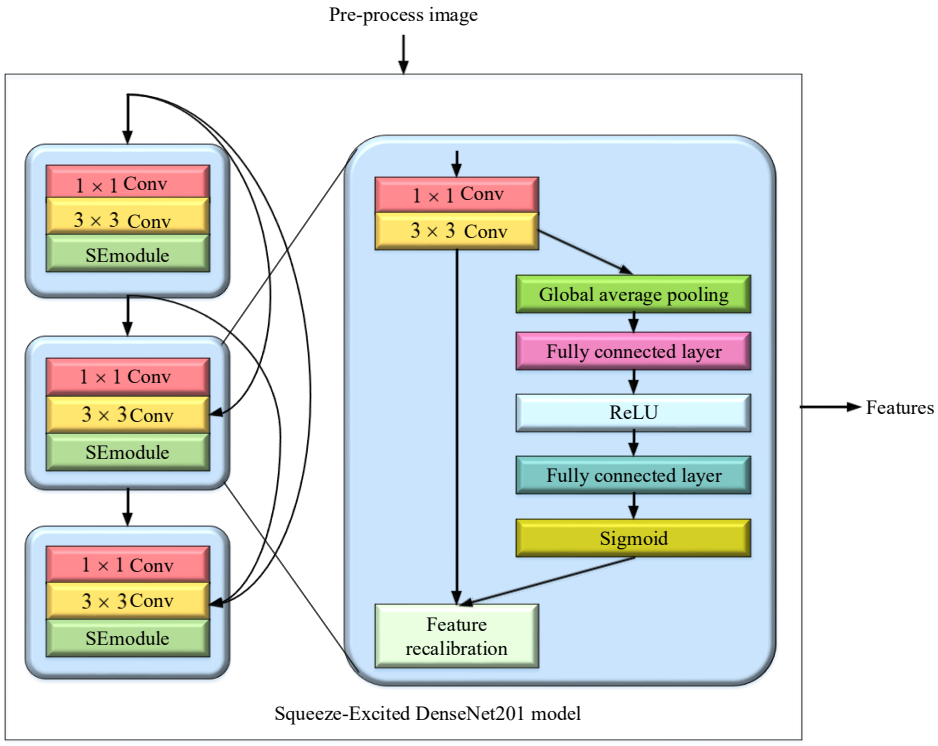


Fig. 2. Architecture of SEDenseNet201

Here, the SEDenseNet201 model combines DenseNet201 [33] and the SE network [34] to leverage multi-dimensional features, thereby enhancing its learning capability. The proposed feature extraction model uses DenseNet201 as the baseline, and squeeze-excitation modules are added to improve feature extraction efficiency. To maintain stable gradient flow, SEDenseNet201 reduces error attenuation in each hidden layer and uses a reverse-conditioning approach to prevent gradient vanishing. Every layer of DenseNet201 is directly connected to all subsequent layers. After the 3×3 convolution layer of every DenseNet block, SEDenseNet201 adds a Squeeze-Excitation module. This module enables lossless transmission or automatically adjusts each channel's weight using the network's global information, allowing adaptive calibration of feature channels. Beneficial features are enhanced, while less useful features are diminished in proportion to their significance.

Each structural block of SEDenseNet201 consists of $n$ sets of feature conversion layers. Each set includes $1 \times 1$ convolutional layer, one SE module for feature recalibration, and $3 \times 3$ convolution layer. The weight of each feature channel can be automatically calibrated by the SE module, which is added after the convolution transformation. Since the pooling layer has no parameters after the first convolutional layer, the feature dimension of the matrix remains constant, making the feature extraction process more stable. A significant architectural change is the removal of average pooling from each neural network block's transfer layer. This deliberate design choice leaves the convolutional layer intact, allowing it to operate unchanged. The neural network's robustness is enhanced by removing average pooling, allowing it to better tolerate perturbations and fluctuations in the input data. Initially, the preprocessed image is provided to the network, where the input layer passes it to the hidden layers.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

The general Densenet201 model is implemented to process the input image. With each dense block, the convolutional layers squeeze and expand the input features, thereby providing enhanced training. The use of additional layers in the dense block resulted in extensive information flow and efficient feature reuse across the entire network. The SE module squeezed and expanded features, thereby improving feature reuse. Moreover, the information flow within the network has been enhanced, resulting in more effective learning and feature extraction.

After every $3 \times 3$ convolution layer, the SE module is utilized to regulate feature channel weights after complication alteration. It is a computational element in which the input feature matrix is $I$, and the nonlinear feature mapping after convolution transformation is$F$. Here, $I \in G^{X' \times Y' \times Z'}$, $F \in G^{X \times Y \times Z}$ where the 3D information of the input feature matrix is presented as $X'$, $Y'$ and $Z'$, the 3D information of the output feature matrix is represented as $X$, $Y$ and $Z$. The $F$corresponding to $R$l ayer network in a single block of the SEDNet model is described in the equation below.

$$F = S^R \otimes f\left(bn\left(S^{(R-1)}\right) \otimes f\left(bn([I_0 I_1 \ldots I_{R-1}])\right)\right) \tag{6}$$

where $\left[I_0, I_1, \ldots I_{R-1}\right]$ indicates merging the $0,1,\ldots,R-1$ layer of the output matrix. The convolution matrix of $1 \times 1$ and $3 \times 3$ are represented as $S^{(R-1)}$ and $S^R$, the batch normalization of every hidden layer's output data is represented as $bn(\cdot)$. The learning ability of nonlinear mapping is represented as the ReLU activation function and $\otimes$ represents the convolution. The signals from each channel's output represent the dependency among feature channels. The global average pooling method, represented by the equation below, compresses spatial data into a channel descriptor to circumvent the challenges of the convolution process within the local receptive field.

$$P_{sq}(a_z) = \frac{1}{X \times Y} \sum_{i=1}^{X} \sum_{j=1}^{Y} a_z(i,j) \tag{7}$$

where, the output maps of $z^{th}$ channel after convolution transformation, and global feature compression, are represented as $a_z$. The function of global feature compression is represented as $P_{sq}(\cdot)$. In compression, the excitation response is used to capture correlations among feature channels and measure the feature-weight aggregation.

$$r = P_{ex}(c, S) = \sigma\left(S_2 f(S_1 c)\right) \tag{8}$$

Here, the weight matrix is represented as $S_1 \in H^{\frac{Z}{Y} \times Z}$and $S_2 \in H^{Z \times \frac{Z}{Y}}$. Rescaling $F$ with activation value $r$ realized as the final output of the SE module, which is described in the equation below.

$$q_z = P_{scale}(a_z, c_z) = c_z \cdot a_z \tag{9}$$

After the final feature is calibrated, the output matrix of $z^{th}$ channel and the function of recalibration are represented as $P_{scale}(.)$. Finally, the SEDenseNet201 model extracts various features from the pre-processed image. By integrating the SE modules with DenseNet201, the model enhances feature learning and stability, ensuring robust feature extraction from lung disease images.

### 3.3. Feature Selection using ELO

After feature extraction, the next crucial step involves selecting the most relevant features for lung disease classification. The ELO technique is introduced for this purpose, incorporating two optimization strategies: the Lyrebird Optimization Algorithm (LOA) [35] and Levy Flight Optimization (LFO) [36].

#### 3.3.1. Initialization

LOA is a population-based optimization strategy that iteratively searches for the best solution in the problem space. Each lyrebird in the LOA population represents a possible solution, and its position in the search space is randomly assigned at the start of the process. However, in ELO, the initialization process is enhanced by using Levy flight-distributed values instead of random numbers to initialize agents' positions in the search space. This mechanism increases the wide diversity of ELO, as shown in equation (10).

$$R_x = ELOA\_Levy(dim)' \cdot * (high - low) + low \tag{10}$$

Here, the $LOA\_Levy(dim)$ is denoted as levy flight function, and $dim$ is denoted as the problem's dimensionality, providing a broader search distribution for the ELO algorithm.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

### 3.3.2. Mathematical modelling of LOA

In each iteration, the population members' positions are updated based on the mathematical modeling of the lyrebird's reaction to threats. The LOA updating process consists of two stages: exploration and exploitation. Equation (11) simulates the lyrebird's decision-making process in the LOA design when it must choose between hiding and escaping from danger. Hence, only one of these phases is used to update each LOA member's position during an iteration.

$$Updateprocessfor R_x = \begin{cases} basedonphase1, & w_f \le 0.5 \\ basedonphase2, & else \end{cases} \tag{11}$$

Here, $w_f$ is signified as the random number from the interval [0, 1].

### 3.3.3. Exploration Phase

This phase models the lyrebird's movement from a hazardous location to a safe area; the population member's position is updated in the search space during this phase. This stage enables the LOA algorithm to perform a global search. Equation (12) is used to find each LOA member's set of safe places.

$$MN_x = \left\{R_{g,}E_g < E_x and g \in \{1,2,,..I\}\right\} \tag{12}$$

where, $MN_x$ is denoted as the set of safe areas for$x^{th}$ lyrebird and $R_g$ is denoted as the $g^{th}$ row of $R$ matrix, which has a better objective function value than the $x^{th}$ LOA member. The LOA design assumes that the lyrebird will randomly escape to one of these safe zones. A new location is obtained for every LOA member utilizing equation (13), and the prior position is changed if it increases the goal function value, as determined by equation (14).

$$R_{x,y}^{F1} = R_{x,y} + w_{x,y} \cdot (MMN_{x,y} - S_{x,y} \cdot R_{x,y}), \tag{13}$$

$$R_x = \begin{cases} R_x^{F1}, & E_x^{F1} \le E_x, \\ R_x, & else, \end{cases} \tag{14}$$

where, $MMN_x$ is denoted as the selected safe area for $x^{th}$ lyrebird, $MMN_{x,y}$ is denoted as the $y^{th}$ dimension, $R_x^{F1}$ is denoted as the novel location calculated for the $x^{th}$ lyrebird based on the escaping strategy of the proposed LOA, $R_{x,y}^{F1}$ is denoted as the$y^{th}$ dimension, $E_x^{F1}$ is denoted as the objective function value, $w_{x,y}$is denoted as the random number from the interval [0,1], and $S_{x,y}$ is denoted by random numbers selected from 1 or 2.

### 3.3.4. Exploitation Phase

During this phase, the location of each population member in the search space is updated based on the lyrebird's behavior while it hides within its surrounding safe area. This phase illustrates LOA's capability for local search. A new location is obtained for each LOA member utilizing equation (15) and replaces the prior location if it increases the objective function value, as determined by equation (16).

$$R_{x,y}^{F2} = R_{x,y} + (1 - 2w_{x,y}) \cdot \frac{az_y - bz_y}{d} \tag{15}$$

$$R_x = \begin{cases} R_x^{F2}, & E_x^{F2} \le E_x, \\ R_x, & else, \end{cases} \tag{16}$$

where, $R_x^{F2}$ is denoted as the new calculated for the $x^{th}$ lyrebird based on the hiding strategy of the suggested LOA, $R_{x,y}^{F2}$ is denoted as the $y^{th}$dimension, $E_x^{F2}$ is represented as the objective function value, $w_{x,y}$ is signified as the random numbers from the interval [0,1], and $d$ is denoted as the iteration counter.

### 3.3.5. Repetition Process

The LOA iteration is complete once all lyrebirds' positions are updated. The algorithm moves on to the next phase, where it will continue to improve the LOA population by updating the algorithm. This process continues until the algorithm reaches the final iteration. At that point, the best LOA population is found. After applying ELO, the method returns the best candidate solution as the optimal solution, having converged on the most promising conclusion. The best candidate solution is revised and saved throughout the iterations. In addition, the objective function for selecting the optimal feature is given in equation (17).

$$Fitness = \alpha(1 - Accuracy) + \beta \frac{N_{sel}}{N_{total}} \tag{17}$$

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

where accuracy is classification accuracy, $N_{sel}$ is the number of selected features, $N_{total}$ is the total number of features, $\alpha$ and $\beta$ are weighting coefficients respectively. To strike a balance between accuracy and dimensionality reduction, the fitness function is constructed as a weighted mixture of feature subset size and classification error. The pseudo-code of the ELO algorithm is presented in Table 2.

Table 2. ELO algorithm

***Start ELO***
**Input:** dim_size, population_size, max_iterations, and objective_function
**Initialize** the population of lyrebirds using Levy Flight distribution values using equation (10)
Assess the objective function
Determine the best candidate solution
***For*** $d = 1\ to\ D$
  ***For*** $x = 1\ to\ I$
    Calculate the decision–making probability $R_x$ using equation (11).
    ***If*** $w_f \leq 0.5$(**chose exploration phase**)
      Determine the candidate's safe area for $x^{th}$ lyrebird based on an equation (12)
      Calculate the new location of the $x^{th}$ LOA member utilizing equation (13)
      Update $x^{th}$ LOA member utilizing equation (14)
    ***Else*** **(chose exploitation phase)**
      Calculate the new location of the $x^{th}$LOA member utilizing equation (15)
      Update $x^{th}$ LOA member utilizing equation (16)
    ***End (if)***
  ***End (For*** $x = 1\ to\ I$)
  Save the best candidate solution so far
***End (For*** $d = 1\ to\ D$)
**Output** the best quasi-optimal solution obtained with ELOA
***End ELO***

Finally, ELO selects the optimal features for lung disease classification.

### 3.4. Lung disease classifications using AC-SACNet model

The selected features are fed into the AC-SACNet model, which classifies lung diseases using components including Atrous Convolutional Networks (ACNN), self-attention mechanisms, and capsule networks. The parameters of the proposed approach are fine-tuned using the Kookaburra Optimization Algorithm. The architecture of the AC-SACNet model is depicted in Fig. 3.

#### 3.4.1. ACNN

The ACNN plays an important role in the classification process. It comprises convolutional modules and fully connected layers. The atrous convolution method is utilized to process incoming features that often contain large amounts of image data. Atrous refers to the introduction of gaps (or dilation) into the standard convolution layer, which enhances the receptive field without increasing the number of parameters. This method increases the receptive field of convolutional layers while maintaining computational efficiency. Atrous convolution is employed to maintain spatial details and incorporate localized inductive bias, which is essential for learning visual features, even as self-attention methods can capture global contextual relationships. High-resolution feature maps are preserved while multi-scale feature extraction is enabled through atrous convolution, which increases the receptive field without spatial downsampling. In medical imaging tasks with limited data, where Transformers alone might overfit or fail to capture the precise boundary of structural representations, this complementary integration helps to alleviate these issues. The procedure for attaining the output $b(p)$ after passing through an Atrous convolution layer $w(f)$ of length $f$can be expressed as follows:

$$b(p) = \sum_{f=1}^{f} a(p + ft \cdot f)w(f). \tag{18}$$

where, the $a(p)$is denoted as the input features, $f$is denoted as the parameter corresponding to the convolution. $t$ is denoted as the distance between adjacent nodes in the kernel given by the dilation rate. Finally, atrous convolution increases the receptive field and improves the resolution of the output feature map. This leads to better accuracy training with the goal of maximizing efficiency.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

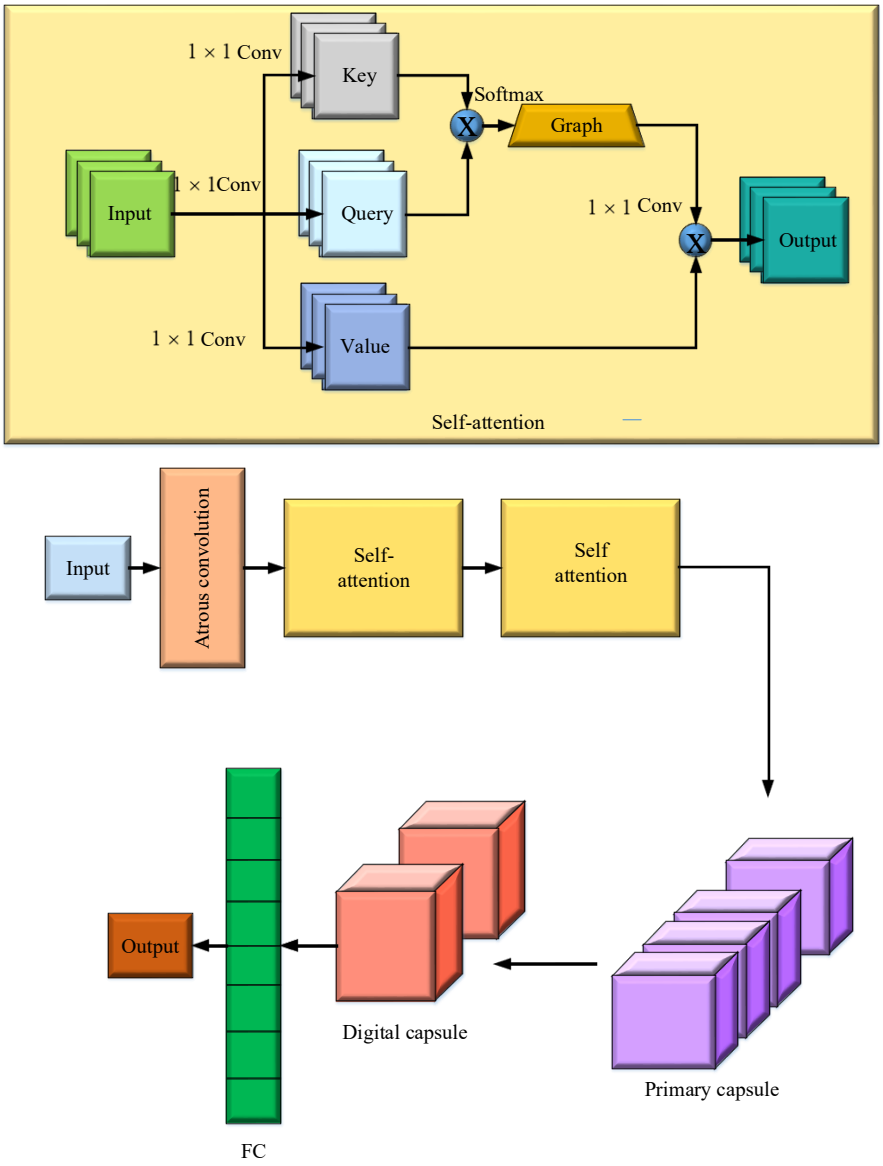


Fig. 3. AC-SACNet model

### 3.4.2. Self-Attention Mechanism

The self-attention layer [37] is incorporated to enhance the network's ability to focus on important features at crucial moments, reducing noise through the location attention mechanism. Using the output key matrix and value feature matrix, this system learns the self-attention weights at different times and applies various weights to different parts of the input sequence. Two distinct pooling processes are used to combine channel information from two feature maps. To enhance the model's capacity for spatial feature extraction, the initial key feature matrix and value matrix are weighted independently, resulting in output feature maps by self-attention. The following formula can be used to characterize the self-attention process:

$$Attention(X,Y,Z) = soft\,max\left(\frac{XY^{d}}{\sqrt{t_y}}\right)Z \tag{19}$$

where, $X$ is denoted as the query, $Y$ is denoted as the key, and $Z$ is denoted as the value. The architecture of the self-attention block in AC-SACNet is different from the regular Transformer attention mechanism. The proposed self-attention mechanism is designed to operate directly on CNN feature maps and aims to be computationally efficient and localized in space, unlike the Transformer attention mechanism, which uses query-key-value (QKV) projections and multi-head global token-wise attention. To reduce the full token-to-token quadratic complexity, it typically uses convolutional projections to compute attention weights along the spatial and/or channel dimensions. Furthermore, instead of being a sequence modeling approach like Transformers, it is incorporated before the capsule layers to optimize feature maps. Rather than replacing the Transformer encoder in its entirety, this makes the attention block computationally efficient, more suitable for dense visual features, and a complement to atrous convolution and capsule routing.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

### 3.4.3. Capsule Network

The output of the self-attention mechanism serves as the input to the capsule network. In a capsule network [38], the initial layer consists of primary capsules that process the output of convolutional layers to extract low-level features like edges, corners, and orientations. Each primary capsule produces an 8-dimensional vector representing the properties of the identified feature. The subsequent layer, known as "digit capsules," aims to identify numbers from 0 to 9. Each digit capsule outputs a 16-dimensional vector that encodes the presence, position, deformation, and other relevant information of each digit. In capsule networks, the vectorial activation function (squashing) is defined as:

$$w_i = \frac{\|P_i\|^2}{1+\|P_i\|^2} \frac{P_i}{\|P_i\|} \tag{20}$$

where, $w_i$represent the output vector of the capsule$j$and $P_i$ indicates the total input to the capsule$j$. Capsule networks generate vector outputs as opposed to scalar ones, which are the result of standard neural networks. More complex and reliable representations are possible because of the length of these vectors, which indicates the likelihood that the entity will be present in the input. The direction of the vector encodes the distinctive features of the object. If an object is present in the image, short vectors are squashed towards 0, and long vectors are compressed towards 1. The prediction vector$o_{i|j}$is obtained by multiplying the weight matrix of $V_{ji}$ and output $U_j$ of the capsule in the lower layer:

$$P_i = \sum_j a_{ji} o_{i|j} \tag{21}$$

$$o_{i|j} = V_{ji} U_j \tag{22}$$

where, $a_{ji}$ represent the coefficient that the dynamic routing procedure determines, and the equation is written by,

$$a_{ji} = \frac{exp(b_{ji})}{\sum l\, exp(b_{jl})} \tag{23}$$

where, $b_{ji}$ is denoted as the log prior probability of the aggregation of a capsule $i$and the capsule $j$. The weights are updated through dynamic routing between two successive capsule layers, and the allocated weight factors dictate how feature information is transmitted from the low-level to the high-level capsule.

### 3.4.4. Fully Connected Layer and Final Classification

The digital capsule layer transmits the features to the fully connected layer, where ReLU is used as the activation function. The output of the fully connected layer is determined by:

$$m(d) = [m_1^d, m_2^d, m_3^d, \ldots, m_N^d]. \tag{24}$$

where, $m(d)$ is denoted as the fully connected layers. Finally, the proposed AC-SACNet model classifies lung diseases. Further improvements may be explored to enhance classification accuracy and reduce errors.

### 3.4.5 Hyper-parameter tuning using KOA

KOA [39] is a population-based optimization technique used to fine-tune the parameters of the proposed classifier for lung disease classification. When optimizing parameters for DL models, KOA has been applied to various optimization problems, including those with high-dimensional search spaces. In the proposed approach, KOA is utilized to fine-tune the weight values of the classification model to improve classification performance. Generally, weight values are randomly selected by the classification model during each iteration, and the value is chosen to be within the limit of [0, 1]. When these parameters are selected randomly, the error value may increase, resulting in higher training loss. To address this issue, the KOA algorithm is used to determine suitable weight values during each iteration, resulting in improved accuracy.

This is achieved by initializing the weight values as a population, and the algorithm selects the most ideal individual (weight value) based on fitness evaluation. The iterations are repeated until the most optimal value is obtained to boost the classification process. KOA was selected because it can optimize both discrete architectural parameters and continuous training hyperparameters within a single population-based search, which is difficult for traditional tuning methods. KOA performs a global search and helps avoid premature convergence, unlike Bayesian Optimization, which becomes inefficient for high-dimensional and mixed discrete-continuous spaces. KOA performs a search on architectural parameters such as capsule size and dilation factors directly, unlike Hyperband, which mainly focuses on early stopping and resource allocation and not on the search of structural parameters. Routing iterations and capsule architectures cannot be tuned using methods such as Adam with schedulers because they are optimization algorithms for training weights and not for architecture or hyperparameters.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

Due to its flexibility in global search, robustness to non-convex fitness spaces, and suitability for simultaneous fine-tuning of deep architecture and capsule-related parameters, KOA was selected. The KOA population matrix encapsulates the positions of the kookaburras during KOA implementation. The kookaburras' positions are initially randomized using the following equation:

$$P = \begin{bmatrix} P_1 \\ \vdots \\ P_k \\ \vdots \\ P_n \end{bmatrix}_{n\times m} = \begin{bmatrix} p_{1,1} \cdots p_{1,D} \cdots p_{1,m} \\ \vdots \ddots \vdots \iddots \vdots \\ p_{k.1} \cdots p_{k,D} \cdots p_{k,m} \\ \vdots \iddots \vdots \ddots \vdots \\ p_{n,1} \cdots p_{n,D} \cdots p_{n,m} \end{bmatrix}_{n\times m} \tag{25}$$

$$p_{k,D} = lo_D + ran.(up_D - lo_D) \tag{26}$$

In this context, $P$represents the population matrix, $P_k$denotes the $k$th kookaburra, $p_{k,D}$indicates the $D$th dimension in the search space. $n$and$m$ specify the number of kookaburras and the number of decision values, respectively. $ran$simplifies to a random number between 0 and 1 while $(up_D - lo_D)$ signify lower and upper bounds of $D$, correspondingly. The objective function of the problem is then assessed, with the resulting values forming a vector representation as per equation (27).

$$O = \begin{bmatrix} O_1 \\ \vdots \\ O_k \\ \vdots \\ O_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} O(P_1) \\ \vdots \\ O(P_k) \\ \vdots \\ O(P_n) \end{bmatrix}_{n\times 1} \tag{27}$$

where $O$ denotes the overall objective function while $O_k$ represents the objective function associated with the $k$th kookaburra.

**Exploration:** To mimic the hunting behavior of kookaburras, the KOA design considers the positions of other kookaburras with superior objective function values as potential prey locations for every individual kookaburra. The possible prey set for each kookaburra is determined by the following equation:

$$PC_k = \{P_i : O_i < O_k and i \neq k\}, \text{where} k = 1,2,\ldots n \text{and } i \in \{1,2,\ldots,n\} \tag{28}$$

where, $PC_k$ denotes the candidate prey set for $k$th kookaburra, $P_i$ signifies that kookaburra has a better objective function value than the $k$th kookaburra, and $O_i$ represents the objective function value. Equation (29) computes a fresh position for the kookaburra. This novel location will take the place of the consistent kookaburra if it improves the objective function value as calculated by equation (30).

$$p_{k,D}^{C1} = p_{k,D} + ran.\left(CPC_{k,D} - r.p_{k,D}\right), k = 1,2,\ldots,n \text{and } D = 1,2,\ldots m \tag{29}$$

$$P_k = \begin{cases} P_k^{C1}, O_k^{C1} < O_k \\ P_k, else \end{cases} \tag{30}$$

where, $P_k^{C1}$ symbolizes the new proposed location of the $k$th kookaburra, $p_{k,D}^{C1}$ denotes $D$th dimension, $O_k^{C1}$ indicates the objective function value, $ran$ represents the random number with a normal distribution in the range of [0, 1], $CPC_{k,D}$ denotes the $D$th dimension of the chosen prey for $k^{\text{th}}$ kookaburra, $r$indicates the random number from the set $\{1,2\}$, $n$ indicates the number of the kookaburra, and $m$ signifies the number of decision variables.

**Exploitation:** To simulate the performance of kookaburras based on their movement near the hunting place, a random location is designed utilizing the following equation:

$$p_{k,D}^{C2} = p_{k,D} + (1 - 2ran)\frac{(up_D - lo_D)}{t}, k = 1,2,\ldots,n, t = 1,2,\ldots T \text{and } D = 1,2,\ldots m \tag{31}$$

$$P_k = \begin{cases} P_k^{C2}, \ O_k^{C2} < O_k \\ P_k, \ else \end{cases} \tag{32}$$

where, $P_k^{C2}$ symbolizes the new proposed location of the $k$th kookaburra, $p_{k,D}^{C2}$ denotes $D$th dimension, $O_k^{C2}$ indicates the objective function value, $t$ denotes the iteration counter of the algorithm, and $T$ represents the maximum number of algorithm reiterations.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

**Iterative Process:** The KOA approach iteratively updates the kookaburras' positions in two phases: exploration and exploitation. The method for updating the KOA population within the search space is outlined in the previous sections. By refining and retaining the most effective solution after each iteration, ongoing improvement is ensured. The algorithm progresses to the next iteration until the final iteration of the algorithm, as governed by Equations (25)-(32). Table 3 presents the algorithm of KOA.

Table 3. Algorithm of KOA

| **Pseudo code of KOA** |
|---|
| **Start** KOA.<br>**Input:** variables, objective function, and constraints.<br>**Fix** KOA population size $(n)$ and iterations$(T)$.<br>**Create** the initial $p$at random using $p_{k,D} = lo_D + ran.(up_D - lo_D)$<br>Estimate the objective function.<br>**For** $t = 1toT$<br>**For** $k = 1ton$<br>**Exploration**<br>Define the candidate prey set using $PC_k = \{P_i: O_i < O_k and i \neq k\}$,<br>**Select** the prey for the $k$th KOA member at random.<br>Compute the new location of $k$th KOA member utilizing $p_{k,D}^{C1} = p_{k,D} + ran.\left(CPC_{k,D} - (C_{k+1}).p_{k,D}\right)$<br>Upgrade $k$th KOA member using $P_k = \begin{cases} P_k^{C1}, O_k^{C1} < O_k \\ P_k, else \end{cases}$<br>**Exploitation**<br>Evaluate the new location of $k$th KOA member utilizing $p_{k,D}^{C2} = p_{k,D} + (1 - 2ran)\frac{(up_D - lo_D)}{t}$<br>Upgrade $k$th KOA member using $P_k = \begin{cases} P_k^{C2}, O_k^{C2} < O_k \\ P_k, else \end{cases}$ , else<br>**End**<br>Store the best candidate solution so far.<br>**End**<br>**Output** the best quasi-optimal solution attained with the KOA.<br>**End** KOA. |

The KOA approach is used to fine-tune the parameters of the suggested classifier for lung disease classification. The optimal parameters obtained using KOA are used to train the classification and are evaluated using a test dataset. The effectiveness of the classifier using the KOA approach is demonstrated through experimental results.

## 4 RESULTS

The performance of the proposed technique was evaluated using several related techniques, such as CNN, MobileNet, EfficientNet (Eff_Net), and Autoencoder (AE) models. The model is evaluated based on several performance metrics such as accuracy, precision, recall, and other standard evaluation measures. The hyperparameters of the proposed model and their values are given in Table 4. Table 5 lists the system configuration parameters and their values.

### 4.1. Dataset Description

The NIH ChestX-ray dataset includes 112,120 frontal-view X-ray images of 30,805 unique patients with fourteen disease labels extracted from radiology reports. There are 15 classes included in the dataset: Cardiomegaly, Hernia, No finding, Mass, Infiltration, Effusion, Emphysema, Atelectasis, Pleural_thickening, Pneumothorax, Fibrosis, Nodule, Consolidation, Edema, and Pneumonia. The dataset size is approximately 45 GB, which can be downloaded through this link: https://www.kaggle.com/datasets/nih-chest-xrays/data.

Minority classes were supplemented with rotation, flipping, and contrast adjustment to improve representation balance and boost generalization in order to lessen class imbalance in the NIH ChestX-ray14 dataset. Additionally, per-class performance indicators were assessed to make sure the majority of classes did not dominate overall accuracy. Weak labels produced by automated natural language processing of radiological reports can be found in the NIH ChestX-ray14 dataset. False positives and annotation noise may be introduced by such labels. The suggested attention-enhanced capsule architecture is made to concentrate on spatially significant locations, which increases robustness against label discrepancies, even when manual relabeling was not done. Class disparity, demographic distribution, and the co-occurrence of several diseases can all lead to dataset bias. Per-class precision, recall, and F1-score were examined in order to lessen the influence of bias. The dataset was split into training, validation, and testing sets in a 70:15:15 ratio.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

Although the NIH ChestX-ray14 dataset allows multi-label annotations, the classification task in this study considers each image with its dominant pathology label. Fig. 4 represents the sample image classes in the dataset.

Table 4. Hyperparameters and their values

| Module | Hyperparameter | Value / Setting |
|---|---|---|
| **Pre-processing** | MHE | Clip limit = 2.0 |
| | CGBF | Kernel size = 5×5, σs = 3, σr = 0.1 |
| **SEDenseNet201 (Feature Extractor)** | Backbone network | ImageNet weights |
| | SE reduction ratio | r = 16 |
| | Input image size | 224 × 224 × 3 |
| **Extended Lyrebird Optimization (ELO)** | Population size | 30 |
| | Maximum iterations | 100 |
| | Feature selection objective | Maximize Accuracy – λ·Feature_Ratio (λ = 0.01) |
| **AC-SACNet (Classifier)** | Capsule vector dimension | 16 |
| | Number of primary capsules | 32 |
| | Routing iterations | 3 |
| | Atrous dilation rates | {1, 2, 4, 8} |
| | Self-attention type | Hybrid attention |
| **Training Hyperparameters (KOA Optimized)** | Learning rate | 1e-4 |
| | Batch size | 32 |
| | Optimizer | Adam |
| | Weight decay | 1e-5 |
| | Dropout rate | 0.5 |
| | Number of epochs | 100 |
| **KOA Optimization Parameters** | Population size | 20 |
| | Maximum iterations | 50 |
| | Search space | LR, batch size, capsule dim, dilation rate, routing iterations |

Table 5. System configuration

| Category | Parameter | Value / Description |
|---|---|---|
| **Training Time** | Time per epoch | 145 seconds |
| | Total training time (100 epochs) | 4.02 hours |
| **Inference Time** | Per-sample inference latency | 18.4 ms |
| | Throughput | 54 samples/second |
| **Hardware Platform** | CPU | Intel(R) Core(TM) i7.4790 CPU @ 3.60 GHz |
| | GPU | NVIDIA GeForce RTX 3060 (12 GB) |
| | RAM | 64 GB |
| | Storage | 1 TB NVMe SSD |
| | System type | 64-bit operating system |
| **Model Complexity** | Number of trainable parameters | 43.7 million |
| | FLOPs | 28.6 GFLOPs |
| **Software Environment** | Python Framework | Python 3.8 programming language |

### 4.2. Performance Metrics and Formulation

The performance of the suggested method is evaluated based on the performance metrics, and its formulation is described as follows.

#### 4.2.1. Accuracy

Accuracy is a performance metric used to evaluate the model's correct detection rate, as represented in the equation below.

$$accuracy = \frac{t_{pos}+t_{neg}}{t_{pos}+f_{pos}+t_{neg}+f_{neg}} \quad (33)$$

here the rate of true positive and true negative is represented as $t_{pos}$ and $t_{neg}$. Then, the false positives and false negatives are signified as $f_{pos}$ and $f_{neg}$.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

#### 4.2.2. Precision

Precision is defined as the ratio of correctly predicted positive samples to the total predicted positive samples, and it is calculated using the equation below.

$$precision = \frac{t_{pos}}{t_{pos}+f_{pos}} \tag{34}$$

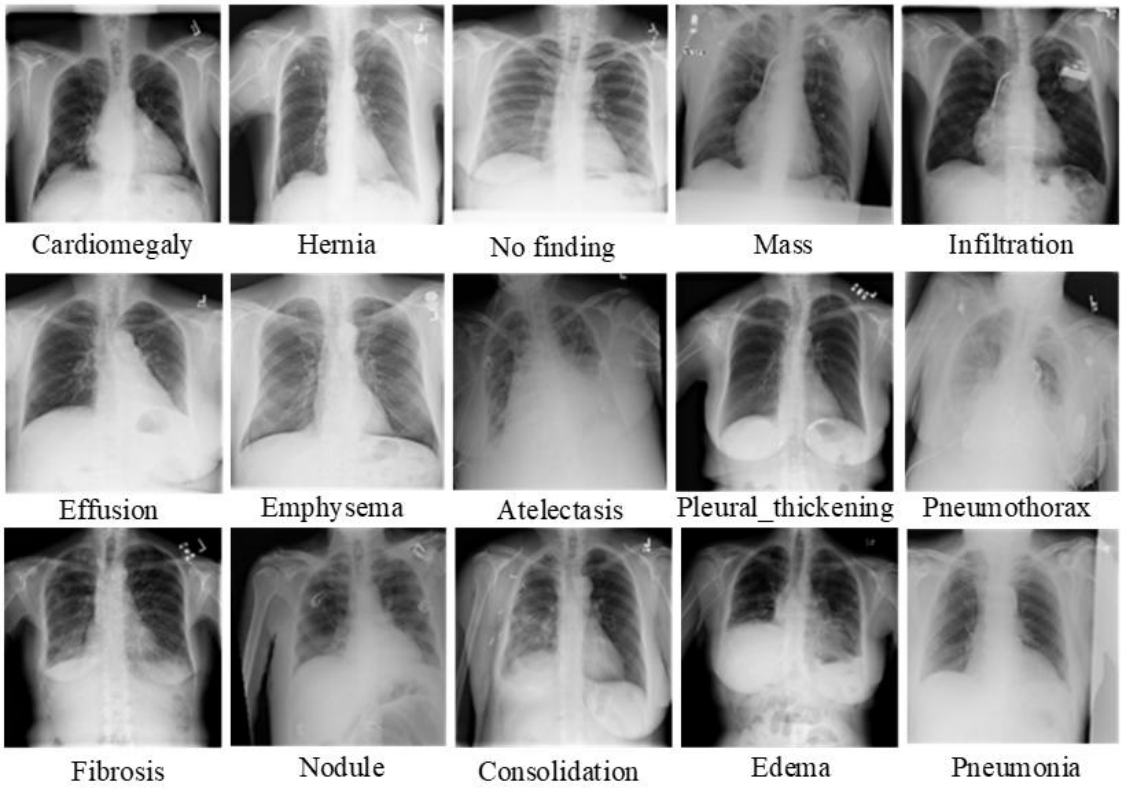


Fig. 4. Sample images of NIH ChestX-ray dataset

#### 4.2.3. Specificity

The specificity is the performance metric utilized to detect the negative sample correctly, which is described in the equation below.

$$specificity = \frac{t_{neg}}{f_{pos}+t_{neg}} \tag{35}$$

#### 4.2.4. Sensitivity/recall

Recall (or sensitivity) represents the proportion of correctly detected positive samples.

$$sensitivity = \frac{t_{pos}}{t_{pos}+f_{neg}} \tag{36}$$

#### 4.2.5. F1-score

The harmonic mean of sensitivity and precision is the F1-score, which is evaluated using the equation described below.

$$F1 = \frac{2\times precision\times sensitivity}{precision+sensitivtiy} \tag{37}$$

#### 4.2.6. False Positive Rate (FPR)

The FPR is utilized to evaluate the probability of falsely rejecting the null hypothesis, which is described in the equation below.

$$FPR = \frac{FP}{FP+TN} \tag{38}$$

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

#### 4.2.7. True Positive Rate

The TPR measures the proportion of true positive outcomes that are correctly given and classified by the model, and it is described in the equation below.

$$TPR = \frac{TP}{TP+FN} \tag{39}$$

### 4.3. Performance Analysis

The performance of the AC-SACNet technique is evaluated using several performance metrics and compared with existing techniques such as CNN, MobileNet, Eff_Net, and AE. The proposed approach involves pre-processing, feature extraction, feature selection, and classification steps. After extracting the features, the best features are selected using the ELO algorithm, and the selected features are highlighted in Table 6.

Table 6. Features selected by the ELO algorithm

| Features extracted | Features selected |
|---|---|
| Feature Extraction..<br><br>1/1 [==============================] - 1s 846ms/step<br><br>Extracted Features shape: (1, 27, 27, 32)<br><br>Features: [[[[0. 0. 0. ... 0.<br>0. ]<br>[0. 0.09440371 0. ... 0. 0.<br>0. ]<br>[0. 0. 0. ... 0. 0.<br>0. ]<br>...<br>[0. 0.7492356 0. ... 0.06516054 0.<br>0. ]<br>[0. 0.34361097 0. ... 0.01373772 0.<br>0. ]<br>[0. 0.12568142 0. ... 0. 0.<br>0. ]] | Feature Selection..<br><br>Selected Features shape: (1, 27, 27, 16)<br><br>Selected Features: [[[[0. 1.8110467 0.5835577 ... 0.<br>0. ]<br>[0. 2.6563206 1.0096316 ... 0. 0.<br>0. ]<br>[0. 2.5338607 0.8444793 ... 0. 0.<br>0. ]<br>...<br>[0. 2.0944006 0.9498317 ... 0.06516054 0.<br>0. ]<br>[0. 2.124631 0.9972961 ... 0.01373772 0.<br>0. ]<br>[0. 1.7300473 0.7302919 ... 0. 0.<br>0. ]] |

The features extracted and the features selected from the extracted subset are displayed in Table 6. Since the number of extracted features is large, they cannot be completely displayed in the table. However, it can be understood that the shape feature has been extracted, and it can be provided as (1, 27, 27, 32), and the selected shape feature is (1, 27, 27, 16). This represents the number of features extracted and the number of features selected from the image. In the proposed work, the feature extraction phase extracts complex textures, object structures, and meaningful patterns, including object boundaries, contours, and key shapes from the images. In the case of feature selection, the most relevant and informative features are selected, while the redundant and irrelevant feature values are removed. The comparison of accuracy, recall, and F1-score of the AC-SACNet method is represented in Fig. 5.

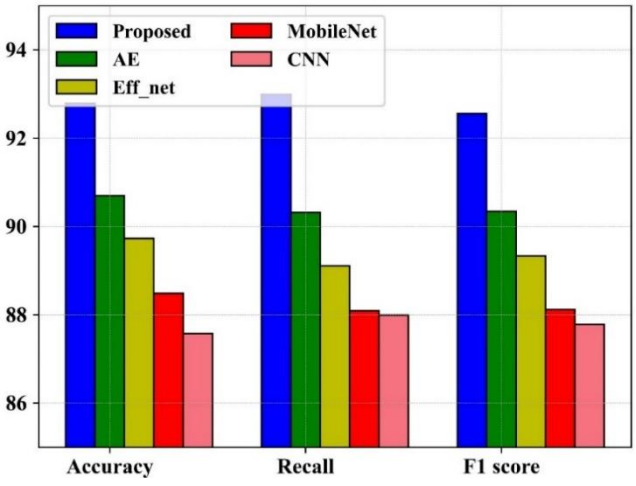


Fig. 5. Comparison of Accuracy, Recall, and F1 score

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

The accuracy of the AC-SACNet method is reported as 92.79%, which is higher than other related techniques such as CNN (87.58%), MobileNet (88.49%), Eff_Net (89.73%), and AE (90.7%). Similarly, the recall and F1 scores of the AC-SACNet technique were analyzed with the existing techniques. The AC-SACNet technique achieves 92.99% recall, which is higher than CNN (87.99%), MobileNet (88.09%), Eff_Net (89.11%), and AE (90.32%). Then, the F1-score of the AC-SACNet technique is obtained to be 92.56%, which is higher than CNN (87.78%), MobileNet (88.12%), Eff_Net (89.33%), and AE (90.34%). The performance analysis of the AC-SACNet mode compared with other related techniques determines its efficient performance. The Comparison of Specificity, Precision, and TPR is represented in Fig. 6.

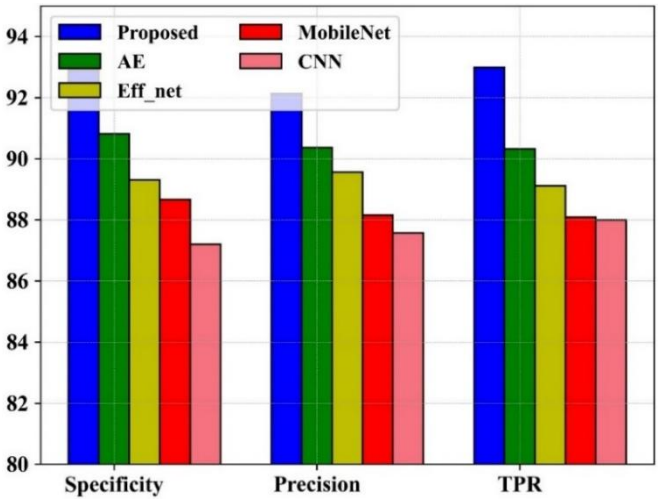


Fig. 6. Comparison of Specificity, Precision, and TPR

The specificity of the AC-SACNet technique is obtained to be 92.89%, which is higher than other related techniques such as CNN (87.2%), MobileNet (88.67%), Eff_Net (89.31%), and AE (90.81%). Similarly, the precision and TPR of the AC-SACNet technique were analyzed in conjunction with the existing techniques. The AC-SACNet technique achieves 92.13% precision, which is higher than CNN (87.57%), MobileNet (88.15%), Eff_Net (89.56%), and AE (90.36%). Then, the TPR of the AC-SACNet technique is obtained to be 92.99%, which is higher than CNN (87.99%), MobileNet (88.09%), Eff_Net (89.11%), and AE (90.32%). The performance of both proposed and related techniques is compared to determine their efficiency. The comparison of FPR is given in Fig. 7.

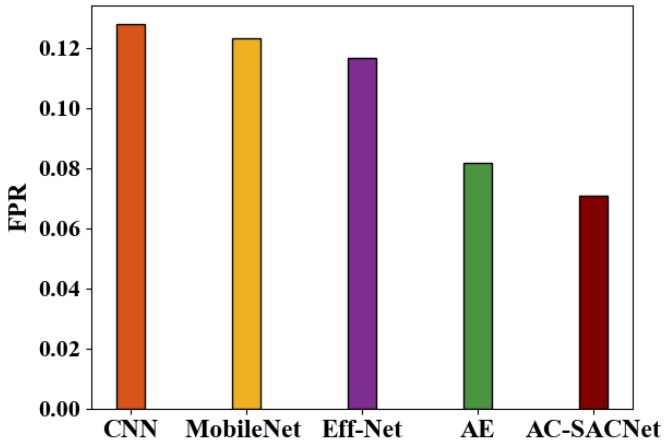


Fig. 7. Comparison of FPR

The FPR of the AC-SACNet technique is obtained to be 0.0711, which is lower than CNN (0.128), MobileNet (0.1233), Eff_Net (0.1169), and AE (0.0819). The comparison of the ROC curve is denoted in Fig. 8. The performance of the ROC curve was analyzed using various related techniques such as CNN, MobileNet, Eff_Net, and AE. It was analyzed based on the AUC value in False Positive Rate and True Positive Rate. The AUC of the AC-SACNet technique is obtained to be 0.921, which is higher than other related techniques such as CNN (0.877), MobileNet (0.881), Eff_Net (0.897), and AE (0.908). The ROC curve indicates that the true positive rate of the proposed approach is higher than that of the other compared methods. It also proves that the model is capable of efficiently distinguishing the features belonging to different groups in the dataset. Therefore, the proposed approach is capable of providing a better true positive rate compared to the other models. The accuracy and loss analysis of training and testing are characterized in Fig. 9.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

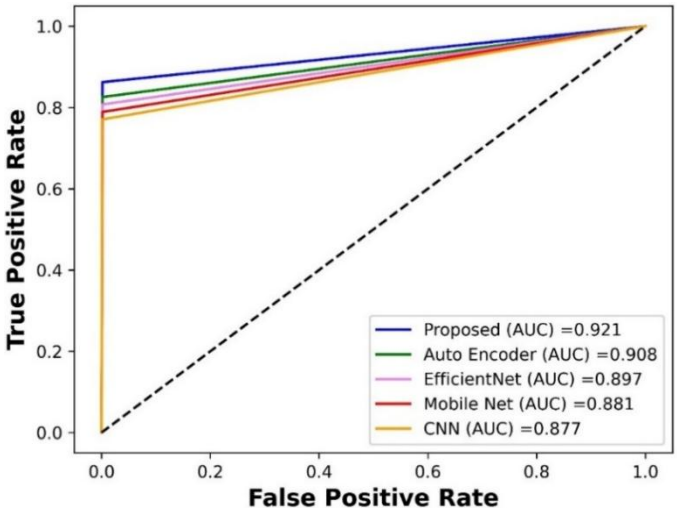


Fig. 8. Comparison of the ROC curve

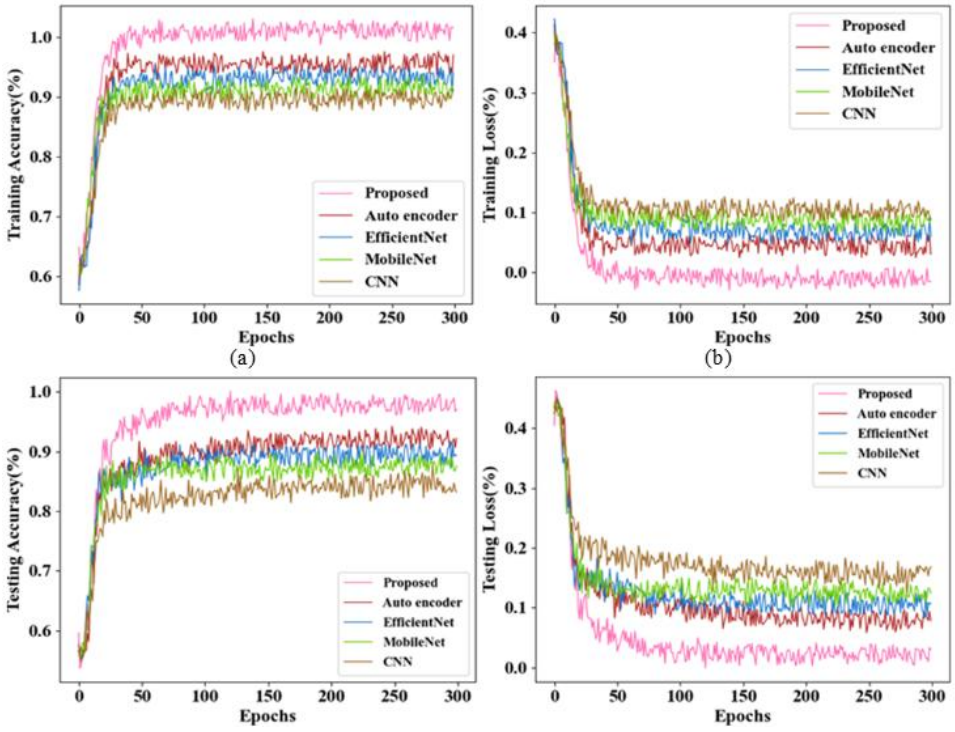


Fig. 9. Analysis of (a) Training Accuracy, (b) Training Loss, (c) Testing Accuracy and (d) Testing Loss

The accuracy of the AC-SACNet technique in both training and testing has been increased compared to the other related techniques. The loss of the AC-SACNet technique in both existing and related techniques is reduced compared to other related techniques. It is evaluated by splitting the data into an 80:20 ratio; 80% of the data from the dataset was utilized for training, and 20% of the data was utilized for testing, which determines the efficient performance of the AC-SACNet technique. The pre-processed image and output classification for different input images are given in Fig. 10.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

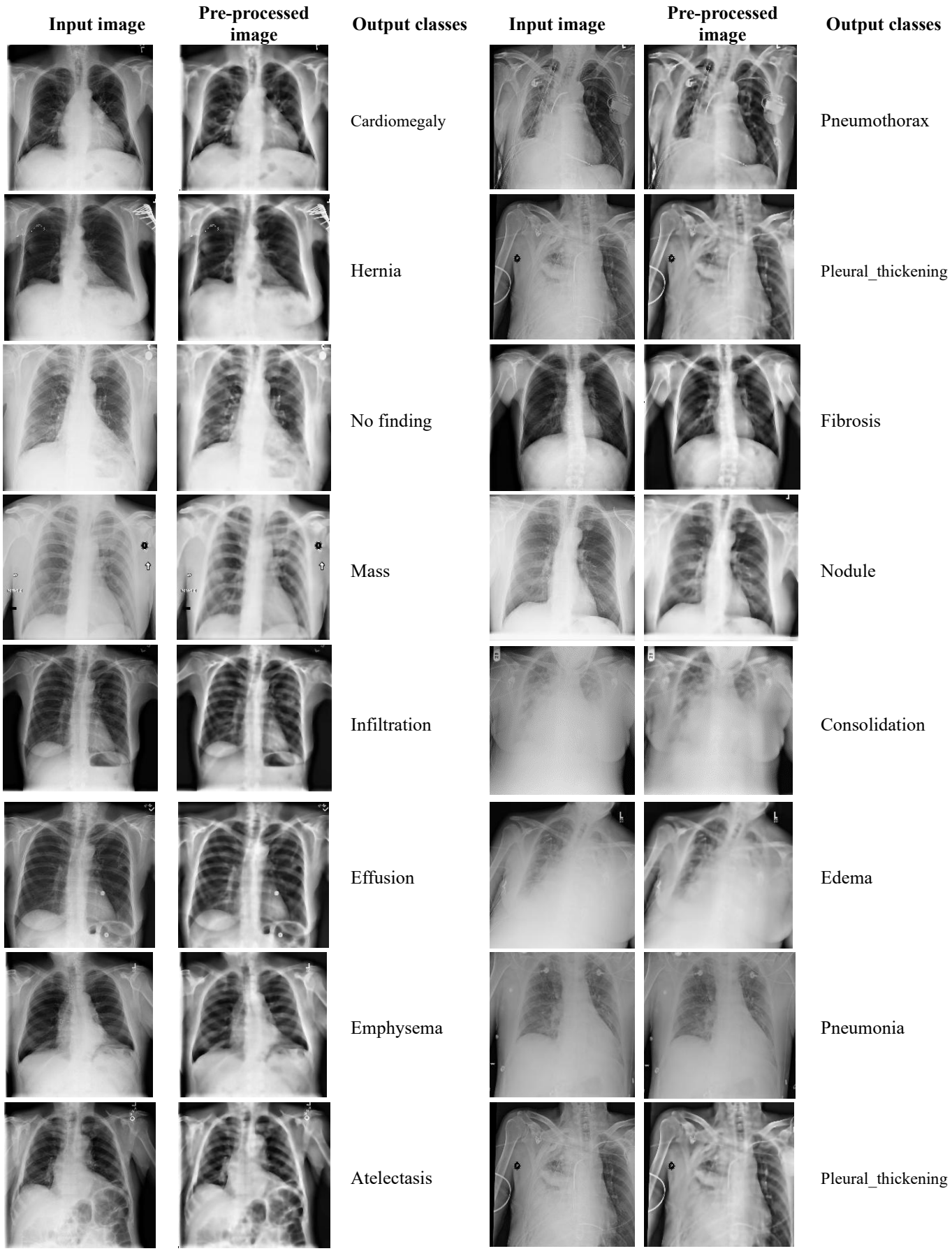


Fig. 10. Pre-processed images of various classes in NIH ChestX-ray dataset

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

The overall performance of the AC-SACNet technique with the related techniques is described in Table 7. Table 7 presents various performance metrics with existing techniques. The AC-SACNet technique achieves 92.79% accuracy for efficient lung disease classification.

Table 7. Performance analysis on NIH ChestX-ray dataset

| Performance | Techniques used | | | | |
|---|---|---|---|---|---|
| | CNN | MobileNet | Eff_Net | AE | AC-SACNet |
| **Accuracy (%)** | 87.58 | 88.49 | 89.73 | 90.7 | **92.79** |
| **Precision (%)** | 87.57 | 88.15 | 89.56 | 90.36 | **92.13** |
| **Recall (%)** | 87.99 | 88.09 | 89.11 | 90.32 | **92.99** |
| **F1 score (%)** | 87.78 | 88.12 | 89.33 | 90.34 | **92.56** |
| **Specificity (%)** | 87.2 | 88.67 | 89.31 | 90.81 | **92.89** |
| **TPR** | 0.8799 | 0.8809 | 08911 | 0.9032 | **0.9299** |
| **FPR** | 0.128 | 0.1233 | 0.1169 | 0.0819 | **0.0711** |
| **AUC** | 0.877 | 0.881 | 0.897 | 0.908 | **0.921** |

For comparison, some of the popular methods are chosen, such as AMGNN [24], modified MobileNet v2 [25], EfficientNet v2-M [26], customized CNN [27], EfficientDet [28] and multilayer machine vision classifier [29]. Fig. 11 depicts a comparison graph of the proposed method with the existing state-of-the-art methods in terms of precision. From the graph, it can be identified that the proposed approach provided satisfactory performance. It can also be observed that the precision value of the existing customized CNN is better than that of the proposed approach. The visual impacts of the images in the dataset led to a reduction in the precision performance of the proposed approach. However, the results are satisfactory when compared with the existing methods in the domain of lung disease classification.

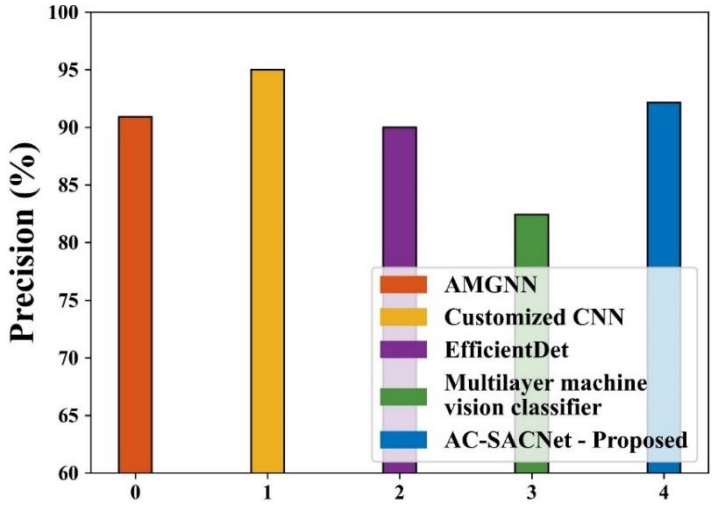


Fig. 11. Comparison of precision with the existing state-of-the-art methods

Fig. 12 presents a comparison between the proposed and existing state-of-the-art methods in terms of recall. It is seen that the recall value attained by the existing multilayer machine vision classifier is better than that of the proposed approach. This is because the proposed method is slightly affected by the visual impacts of the images in the dataset. This reduced the recall value of the proposed approach; however, the accuracy of a model is independent of recall, and the model can produce better results. Fig. 13 depicts a comparison of F-measure between the proposed and existing state-of-the-art methods. From the graph, the f-measure of the proposed approach is greater than that of the existing approaches. This proves that the proposed approach can provide the best positive predictions compared to the other methods. Among the compared approaches, the customized CNN and EfficientDet yielded better scores, proving that those methods can provide better classification results.

A comparison between the proposed and existing state-of-the-art methods in terms of accuracy is depicted in Fig. 14. From Fig. 14, it can be concluded that the proposed approach is capable of producing better and more accurate performance results. It is seen that the EfficientDet and the proposed classifiers have produced more or less similar results. The performance of EfficientDet is found to be slightly better than that of the proposed approach. This is because of the capability of this method to provide better positive predictions than the proposed approach. One of the problems with the proposed approach is that the visual impacts of the input images affected the overall performance. Also, the presence of biases in the dataset contributed to the reduction in the accuracy value. However, the accuracy of the proposed approach is found to be satisfactory in the chosen domain. Moreover, the performance of all these models would considerably change while dealing with real-world data.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

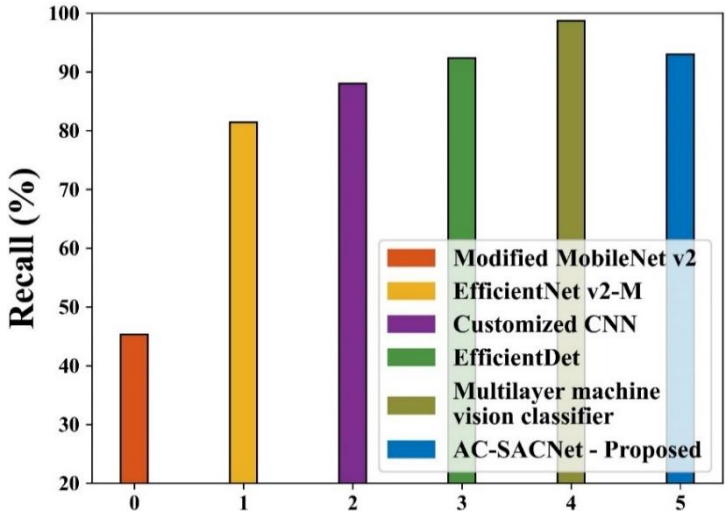


Fig. 12. Comparison of recall with the existing state-of-the-art methods

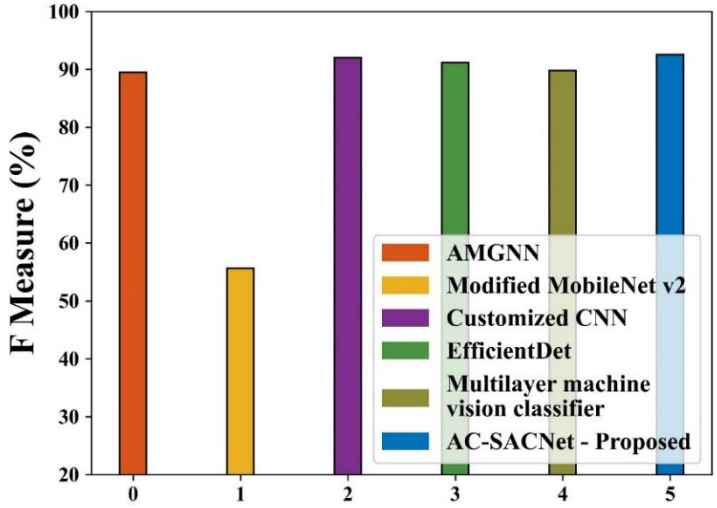


Fig. 13. Comparison of F-measure with the existing state-of-the-art methods

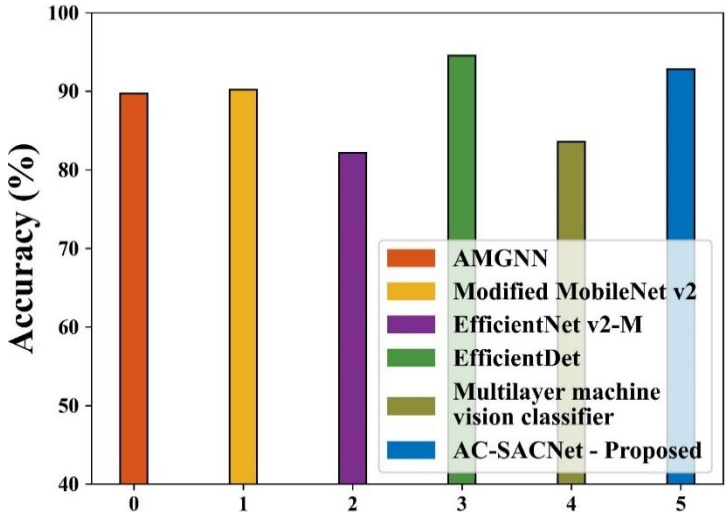


Fig. 14. Comparison of accuracy with the existing state-of-the-art methods

The proposed model's class-wise performance analysis on the NIH Chest X-ray dataset is shown in Table 8. Across all 14 illness categories, the results show consistently good accuracy, precision, and recall. Notably, the model performs well in detecting emphysema (97.2% accuracy) and hernia (98.3% accuracy), demonstrating efficient feature discrimination for distinct illnesses.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

Infiltration performs worse (86.5% accuracy), most likely as a result of its inter-class overlap and diffuse radiographic features. The model's robustness against class imbalance is demonstrated by its macro-average precision (92.13%) and recall (92.99%), which both attest to its balanced performance across frequent and rare classes.

Table 8. Class-wise evaluation of NIH ChestX-ray dataset

| Class | Accuracy (%) | Precision (%) | Recall (%) |
|---|---|---|---|
| **Atelectasis** | 90.8 | 91.2 | 92.4 |
| **Cardiomegaly** | 95.4 | 93.8 | 94.6 |
| **Consolidation** | 91.6 | 90.7 | 92.1 |
| **Edema** | 94.9 | 92.6 | 93.8 |
| **Effusion** | 89.7 | 90.4 | 91.3 |
| **Emphysema** | 97.2 | 94.1 | 95.5 |
| **Fibrosis** | 93.6 | 91.8 | 92.9 |
| **Hernia** | 98.3 | 95.6 | 96.2 |
| **Infiltration** | 86.5 | 88.2 | 89.7 |
| **Mass** | 94.1 | 92.0 | 93.5 |
| **Nodule** | 92.7 | 91.3 | 92.4 |
| **Pneumonia** | 93.9 | 92.5 | 94.2 |
| **Pleural Thickening** | 92.8 | 90.9 | 92.1 |
| **Pneumothorax** | 94.4 | 92.6 | 93.9 |
| **Macro Averages** | **92.79** | **92.13** | **92.99** |

### 4.4. Ablation Study

This section deals with the ablation study of the proposed approach, which is partitioned into multiple units to ensure the contribution of each component involved in the method. In this work, the proposed approach is partitioned into four different units: Unit 1, Unit 2, Unit 3, and Unit 4. Unit 1 indicates the inclusion of pre-processing and DenseNet201 + Softmax Layer, Unit 2 signifies the inclusion of pre-processing and DenseNet201 + Capsule (No Attention); Unit 3 signifies the inclusion of pre-processing DenseNet201 + Attention + Softmax, and Unit 4 presents the complete proposed approach without any elimination. The results obtained through this evaluation are presented in Table 9.

Table 9. Ablation study of the proposed approach

| Model Variant | Accuracy (%) | Sensitivity (%) | F1-score (%) | Specificity (%) |
|---|---|---|---|---|
| **NIH ChestX-ray dataset** | | | | |
| **DenseNet201 + Softmax (Unit 1)** | 89.84 | 89.62 | 89.10 | 89.71 |
| **DenseNet201 + Capsule (No Attention) (Unit 2)** | 91.32 | 91.05 | 90.88 | 91.14 |
| **DenseNet201 + Attention + Softmax (unit 3)** | 91.87 | 92.10 | 91.76 | 91.52 |
| **Proposed AC-SACNet (Attention + Capsule) (Unit 4)** | **92.79** | **92.99** | **92.56** | **92.89** |

The role played individually and collectively by the capsule classifier and attention mechanism on both datasets is adequately proven through its ablation study. Using the baseline model DenseNet-201 with softmax (Unit 1), the accuracy for the NIH ChestX-ray14 dataset is 89.84%. Increasing this to using a capsule classifier instead of softmax (Unit 2) results in an accuracy of 91.32%, which shows capsule routing to be more consistent with spatial and part-whole relationships preserved in a chest radiograph image. Further improving this to using attention with softmax but replacing it (Unit 3) results in an accuracy of 91.87%, which shows how discriminative area attention helps to improve results. The best results are obtained by the complete network of proposed model AC-SACNet, which has an accuracy of 92.79%, a recall of 92.99%, an F1-score of 92.56%, and a specificity of 92.89%.

Table 10 presents a well-designed ablation study, where only the pre-processing steps are varied while keeping the AC-SACNet model and training process fixed. Also, other feature extraction, feature section processes were not carried out. Experiments E1 and E3 individually employ CGBF (for denoising) and MHE (for contrast enhancement) to evaluate their individual effects, whereas Experiment E1 serves as the baseline with no pre-processing. Based on the results, MHE and CGBF individually improve the accuracy, sensitivity, specificity, and F1-score values on the NIH ChestX-ray dataset, although MHE generally provides a greater improvement due to its enhanced contrast representation. The best performance is achieved in Experiment E4, which combines CGBF and MHE and demonstrates the joint effect of the two pre-processing techniques. These controlled results quantitatively confirm that the proposed pre-processing pipeline significantly improves feature quality and makes a substantial contribution to the overall performance benefits of the framework.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

Table 10. Isolated contribution of pre-processing through controlled experiment

| Experiments | Pre-processing | Accuracy (%) | Specificity (%) | Sensitivity (%) | F1-Score (%) |
|---|---|---|---|---|---|
| **E1** | None | 87.21 | 80.22 | 81.00 | 82.22 |
| **E2** | CGBF only | 88.54 | 82.55 | 82.79 | 83.54 |
| **E3** | MHE only | 88.56 | 83.65 | 84.56 | 85.64 |
| **E4** | CGBF + MHE | 89.23 | 85.22 | 86.35 | 87.54 |

To determine the effects of feature selection strategy on the performance of the proposed classifier, Table 11 compares different feature reduction methods quantitatively while keeping the proposed AC-SACNet classifier constant. With the full feature set, the accuracy of the proposed model on the NIH ChestX-ray dataset is 91.20%, without feature selection. Although Mutual Information and RFE achieve higher accuracy with a reduced feature set, PCA achieves better accuracy by reducing the feature set to 40% of the original features. Despite selecting only 18% of the features, the proposed ELO feature selection strategy always achieves the highest accuracy and F1-score on both datasets, thus confirming its superiority in identifying the most discriminative features and eliminating redundancy.

Table 11. Quantitative comparison of the proposed ELO with different feature reduction methods

| Method | Accuracy (%) | F1-score (%) | Selected Features (%) |
|---|---|---|---|
| **No Feature Selection** | 91.20 | 90.85 | 100% |
| **PCA** | 92.10 | 91.74 | 40% |
| **Mutual Information** | 92.45 | 92.01 | 35% |
| **RFE** | 92.54 | 92.00 | 28% |
| **ELO (Proposed)** | **92.79** | **92.56** | **18%** |

Table 12. Variance analysis of the ELO-based feature selection

| Run | Accuracy (%) | F1-score (%) |
|---|---|---|
| **1** | 92.68 | 92.44 |
| **2** | 92.73 | 92.51 |
| **3** | 92.81 | 92.59 |
| **4** | 92.77 | 92.54 |
| **5** | 92.85 | 92.62 |
| **6** | 92.69 | 92.47 |
| **7** | 92.74 | 92.53 |
| **8** | 92.79 | 92.56 |
| **9** | 92.83 | 92.60 |
| **10** | 92.72 | 92.49 |
| **Mean** | 92.76 ± 0.06 | 92.54 ± 0.06 |

The optimization process was performed ten times independently with different random seeds to test the robustness of the ELO-based feature selection. On the NIH ChestX-ray dataset, the average accuracy is 92.76% ± 0.06. The robustness of the ELO algorithm in making feature subset selections is verified by the small standard deviation values, which indicate good stability and convergence properties.

### 4.5. Statistical Validation

Statistical validation is necessary to demonstrate that the observed performance improvements of a proposed model are real, reliable, and not due to random chance or data variability. In machine learning experiments, performance metrics (accuracy, precision, recall, F1, etc.) vary slightly across folds because of different train–test splits. Simply reporting higher mean values is not sufficient for scientific rigor, as small differences may arise from randomness rather than genuine model superiority. Statistical tests (such as t-tests and confidence intervals (CIs)) quantify whether the difference between models is significantly larger than the natural variability in the data. If the p-value is below a predefined threshold (e.g., 0.05) and CIs do not overlap zero, the null hypothesis can be confidently rejected, and the improvement is statistically significant.

This strengthens the credibility, reproducibility, and acceptance of the research, particularly in high-impact journals where reviewers require proof that performance gains are not accidental. The statistical significance analysis using Welch's independent t-test (based on the reported Mean ± SD, assuming n = 5 per model) is shown in Table 13, which demonstrates that the performance improvements of AC-SACNet over CNN, MobileNet, EfficientNet, and AE are highly significant rather than marginal fluctuations due to fold variability. The computed t-values are extremely large (ranging from 11.188 to 23.079), indicating that the difference between group means is many standard errors away from zero. Corresponding p-values are well below the conventional threshold of 0.05 (all $p < 0.00001$), strongly rejecting the null hypothesis that the models perform equally well.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

Moreover, the 95% CIs for the mean accuracy difference between AC-SACNet and each baseline are entirely positive and do not cross zero (e.g., 1.658%–2.522% vs AE and 4.676%–5.744% vs CNN), confirming that AC-SACNet consistently outperforms the competing architectures with high statistical confidence. The narrow width of these CIs further suggests low variance across folds and stable model generalization. Therefore, the observed performance gains are statistically robust, practically meaningful, and unlikely to be caused by random variation in cross-validation splits, thereby strengthening the empirical validity of the proposed approach.

Table 13. Statistical Validation

| Comparison | t-value | p-value | 95% CI (Mean Difference) | Significant? |
|---|---|---|---|---|
| **AC-SACNet vs CNN** | 23.079 | < 0.000001 | [4.676, 5.744] | Yes |
| **AC-SACNet vs MobileNet** | 20.370 | < 0.000001 | [3.806, 4.794] | Yes |
| **AC-SACNet vs EffNet** | 15.003 | 0.000001 | [2.585, 3.535] | Yes |
| **AC-SACNet vs AE** | 11.188 | 0.000004 | [1.658, 2.522] | Yes |

### 4.6. Computational Overhead Analysis of KOA

The computational complexity of KOA can be expressed as $O(P.T.C\_training)$, with $P$as the population size, $T$ as the number of iterations for optimization, and $C_training$as the cost of training and validating the AC-SACNet model for a given candidate solution. Compared to the training cost of the deep model, the KOA update processes introduce a negligible complexity of $O(P.D)$, where $D$ represents the dimensionality of the search space of the hyperparameters. The storage complexity for the fitness and potential solutions is $O(P.D)$. Most importantly, this complexity is only introduced during the offline hyperparameter optimization process and does not contribute to the inference process once the optimal hyperparameters have been determined.

### 4.7. Discussion

The efficacy of the suggested technique is assessed utilizing numerous techniques to determine efficiency based on its performance. The comparative analysis of the proposed technique with various related techniques is described in Table 14.

Table 14. Comparative analysis based on their performance

| Work | Technique used | Performance |
|---|---|---|
| **Kabiraj et al. [21]** | CX-Ultranet | 88% average prediction accuracy |
| **Kholiavchenko et al. [22]** | ResNet34-based neural network | 0.714 – AUC |
| **Wang et al. [23]** | Multiscale residual network | 91.99% - accuracy |
| **Deng et al. [24]** | Auto-metric graph neural network (AMGNN) | 89.7% - accuracy |
| **Souid et al. [25]** | Modified MobileNetV2 model | 0.811 – AUC |
| **Proposed** | **AC-SACNet model** | **92.79% accuracy** (NIH ChestX-ray dataset) |

From the performance comparison in Table 14, it can be justified that the proposed approach is more accurate than most of the existing systems. Five different techniques are compared with the proposed method, and different performance metrics are utilized to make the comparison. Compared to the technique presented by Kabiraj et al. [21], the proposed approach is 4.79% more accurate in lung disease classification. This improvement is achieved with the help of different techniques, including the feature selection algorithm and the tuning mechanism followed in the classification phase. Among the compared models, the closest accuracy value is gained by the multiscale residual network introduced by Wang et al. [23], with an overall accuracy of 91.99%. Similarly, the approach presented by Deng et al. [24] also produced better accuracy values compared to other techniques.

This comparison proves that the proposed approach is accurate in classification and can accurately distinguish the features belonging to different groups in the dataset. Accurately differentiating the input features during labeling is crucial for any classification model. Also, the error rate in classification is reduced with the help of fine-tuning, which enables the model to perform better on diverse inputs. It can be assured that the proposed model can be applied in any clinical setting with minor modifications in the parameter settings to attain satisfactory performance in the field of lung disease classification.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

## 5 CONCLUSION

In this paper, a lung disease classification system is proposed. The proposed approach integrates preprocessing, feature extraction, feature selection, and classification techniques for lung disease detection. Modified Histogram Equalization and Cross-guided Bilateral Filtering are used to improve image quality and reduce noise. The SEDenseNet201 model extracts features from the images, and the Extended Lyrebird Optimization algorithm is used for feature selection. The AC-SACNet classifier is used to categorize different lung diseases, while the Kookaburra Optimization Algorithm is employed for hyperparameter tuning. The proposed model achieves an accuracy of 92.70% on the NIH ChestX-ray dataset. The results indicate that the proposed method performs competitively compared with existing techniques. Therefore, the proposed method may be useful for lung disease classification from chest X-ray images. Future work will explore cross-dataset validation and larger multi-center datasets to further evaluate the robustness of the proposed model.

## FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## ETHICS STATEMENT

This study did not involve human or animal subjects and, therefore, did not require ethical approval.

## STATEMENT OF CONFLICT OF INTERESTS

The authors declare no conflicts of interest related to this study.



## REFERENCES

[1] F. J. M. Shamrat, S. Azam, A. Karim, K. Ahmed, F. M. Bui, and F. De Boer, "High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images," *Computers in Biology and Medicine*, vol. 155, p. 106646, Feb. 2023, doi: 10.1016/j.compbiomed.2023.106646.

[2] P. Podder *et al.*, "LDDNET: A deep learning framework for the diagnosis of infectious lung diseases," *Sensors*, vol. 23, no. 1, p. 480, Jan. 2023, doi: 10.3390/s23010480.

[3] M. Murugappan, A. K. Bourisly, N. B. Prakash, M. G. Sumithra, and U. R. Acharya, "Automated semantic lung segmentation in chest CT images using deep neural network," *Neural Computing and Applications*, vol. 35, no. 21, pp. 15343–15364, Apr. 2023, doi: 10.1007/s00521-023-08407-1.

[4] J. Weiss *et al.*, "Deep learning to estimate lung disease mortality from chest radiographs," *Nature Communications*, vol. 14, no. 1, p. 2797, May 2023, doi: 10.1038/s41467-023-37758-5.

[5] V. Ravi, V. Acharya, and M. Alazab, "A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images," *Cluster Computing*, vol. 26, no. 2, pp. 1181–1203, Jul. 2022, doi: 10.1007/s10586-022-03664-6.

[6] P. Taneja, A. Sharma, and M. Singh, "Enhancing Multi-Class Lung Disease Classification with Bagging-Based Deep Learning Ensembles: A Comparative Study of Five Architectures," *Procedia Computer Science*, vol. 260, pp. 209–216, Jan. 2025, doi: 10.1016/j.procs.2025.03.195.

[7] H. Malik, T. Anees, A. S. Al-Shamaylehs, S. Z. Alharthi, W. Khalil, and A. Akhunzada, "Deep Learning-Based Classification of chest diseases using x-rays, CT scans, and cough sound images," *Diagnostics*, vol. 13, no. 17, p. 2772, Aug. 2023, doi: 10.3390/diagnostics13172772.

[8] G. M. M. Alshmrani, Q. Ni, R. Jiang, H. Pervaiz, and N. M. Elshennawy, "A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images," *Alexandria Engineering Journal*, vol. 64, pp. 923–935, Nov. 2022, doi: 10.1016/j.aej.2022.10.053.

[9] H. Malik, T. Anees, M. Din, and A. Naeem, "CDC_Net: multi-classification convolutional neural network model for detection of COVID-19, pneumothorax, pneumonia, lung Cancer, and tuberculosis using chest X-rays," *Multimedia Tools and Applications*, vol. 82, no. 9, pp. 13855–13880, Sep. 2022, doi: 10.1007/s11042-022-13843-7.

[10] Z. Naz, M. U. G. Khan, T. Saba, A. Rehman, H. Nobanee, and S. A. Bahaj, "An Explainable AI-Enabled Framework for Interpreting Pulmonary Diseases from Chest Radiographs," *Cancers*, vol. 15, no. 1, p. 314, Jan. 2023, doi: 10.3390/cancers15010314.

[11] X. Sun, Z. Song, H. Jiang, Y. Ma, and M. Chen, "Image classification of immune checkpoint inhibitor–related pneumonia in lung cancer patients," *Clinical Imaging*, vol. 86, pp. 31–37, Mar. 2022, doi: 10.1016/j.clinimag.2022.03.012.

[12] C. Liu, W. Xie, R. Zhao, and M. Pang, "Segmenting lung parenchyma from CT images with gray correlation-based clustering," *IET Image Processing*, vol. 17, no. 6, pp. 1658–1667, Jan. 2023, doi: 10.1049/ipr2.12744.

[13] K. Wang, Y. An, J. Zhou, Y. Long, and X. Chen, "A novel Multi-Level feature selection method for radiomics," *Alexandria Engineering Journal*, vol. 66, pp. 993–999, Nov. 2022, doi: 10.1016/j.aej.2022.10.069.

[14] J. R. Astley *et al.*, "A hybrid model- and deep learning-based framework for functional lung image synthesis from multi-inflation CT and hyperpolarized gas MRI," *Medical Physics*, vol. 50, no. 9, pp. 5657–5670, Mar. 2023, doi: 10.1002/mp.16369.

[15] T. Wanasinghe, S. Bandara, S. Madusanka, D. Meedeniya, M. Bandara and I. D. L. T. Díez, "Lung Sound Classification With Multi-Feature Integration Utilizing Lightweight CNN Model," in *IEEE Access*, vol. 12, pp. 21262-21276, 2024, doi: 10.1109/ACCESS.2024.3361943.

[16] N. S. Haider and A. K. Behera, "Computerized lung sound based classification of asthma and chronic obstructive pulmonary disease (COPD)," *Journal of Applied Biomedicine*, vol. 42, no. 1, pp. 42–59, Dec. 2021, doi: 10.1016/j.bbe.2021.12.004.

[17] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "A neural Network-Based method for respiratory sound analysis and lung disease detection," *Applied Sciences*, vol. 12, no. 8, p. 3877, Apr. 2022, doi: 10.3390/app12083877.

[18] Y. Zhang *et al.*, "Research on lung sound classification model based on dual-channel CNN-LSTM algorithm," *Biomedical Signal Processing and Control*, vol. 94, p. 106257, Mar. 2024, doi: 10.1016/j.bspc.2024.106257.

[19] M. Kaveh, M. S. Mesgari, and B. Saeidian, "Orchard Algorithm (OA): A new meta-heuristic algorithm for solving discrete and continuous optimization problems," *Mathematics and Computers in Simulation*, vol. 208, pp. 95–135, Jan. 2023, doi: 10.1016/j.matcom.2022.12.027.

[20] A. Kabiraj, T. Meena, K. Tadepalli, and S. Roy, "An explainable weakly supervised model for multi-disease detection and localization from thoracic X-rays," *Applied Soft Computing*, vol. 166, p. 112139, Aug. 2024, doi: 10.1016/j.asoc.2024.112139.

[21] N. W. Asnake, A. O. Salau, and A. M. Ayalew, "X-ray image-based pneumonia detection and classification using deep learning," *Multimedia Tools and Applications*, vol. 83, no. 21, pp. 60789–60807, Jan. 2024, doi: 10.1007/s11042-023-17965-4.

[22] M. Kholiavchenko, I. Pershin, B. Maksudov, T. Mustafaev, Y. Yuan, and B. Ibragimov, "Gaze-based attention to improve the classification of lung diseases," *Medical Imaging 2022: Image Processing*, p. 10, Mar. 2022, doi: 10.1117/12.2612767.

[23] F. Wang, S. Li, S. Li, Y. Gao, S. Li, and S. Li, "Computed tomography-based artificial intelligence in lung disease—Chronic obstructive pulmonary disease," *MedComm – Future Medicine*, vol. 3, no. 1, Feb. 2024, doi: 10.1002/mef2.73.

[24] X. Deng *et al.*, "COPD stage detection: leveraging the auto-metric graph neural network with inspiratory and expiratory chest CT images," *Medical & Biological Engineering & Computing*, vol. 62, no. 6, pp. 1733–1749, Feb. 2024, doi: 10.1007/s11517-024-03016-z.

[25] A. Souid, N. Sakli, and H. Sakli, "Classification and Predictions of Lung Diseases from Chest X-rays Using MobileNet V2," *Applied Sciences*, vol. 11, no. 6, p. 2751, Mar. 2021, doi: 10.3390/app11062751.

[26] S. Kim, B. Rim, S. Choi, A. Lee, S. Min, and M. Hong, "Deep Learning in Multi-Class Lung Diseases' Classification on chest x-ray images," *Diagnostics*, vol. 12, no. 4, p. 915, Apr. 2022, doi: 10.3390/diagnostics12040915.

[27] S. Z. Y. Zaidi, M. U. Akram, A. Jameel, and N. S. Alghamdi, "A deep learning approach for the classification of TB from NIH CXR dataset," *IET Image Processing*, vol. 16, no. 3, pp. 787–796, Nov. 2021, doi: 10.1049/ipr2.12385.

[28] M. Nawaz, T. Nazir, J. Baili, M. A. Khan, Y. J. Kim, and J.-H. Cha, "CXray-EffDet: Chest Disease Detection and Classification from X-ray Images Using the EfficientDet Model," *Diagnostics*, vol. 13, no. 2, p. 248, Jan. 2023, doi: 10.3390/diagnostics13020248.

[29] J. -X. Wu, P. -Y. Chen, C. -M. Li, Y. -C. Kuo, N. -S. Pai and C. -H. Lin, "Multilayer Fractional-Order Machine Vision Classifier for Rapid Typical Lung Diseases Screening on Digital Chest X-Ray Images," in *IEEE Access*, vol. 8, pp. 105886-105902, 2020, doi: 10.1109/ACCESS.2020.3000186.

[30] D. C. Lepcha, B. Goyal, and A. Dogra, "Image Fusion based on Cross Bilateral and Rolling Guidance Filter through Weight Normalization," *The Open Neuroimaging Journal*, vol. 13, no. 1, pp. 51–61, Dec. 2020, doi: 10.2174/1874440002013010051.

[31] S. Samuel, R. S. Ochawar, and M. S. S. Rukmini, "Hybrid deep autoencoder network based adaptive cross guided bilateral filter for motion artifacts correction and denoising from MRI," *The Imaging Science Journal*, vol. 72, no. 1, pp. 76–91, Apr. 2023, doi: 10.1080/13682199.2023.2196494.

[32] K. Verma, G. Sikka, A. Swaraj, S. Kumar, and A. Kumar, "Classification of COVID-19 on Chest X-Ray Images Using Deep Learning Model with Histogram Equalization and Lung Segmentation," *SN Computer Science*, vol. 5, no. 4, Mar. 2024, doi: 10.1007/s42979-024-02695-7.

[33] H. A. Sanghvi, R. H. Patel, A. Agarwal, S. Gupta, V. Sawhney, and A. S. Pandya, "A deep learning approach for classification of COVID and pneumonia using DenseNet-201," *International Journal of Imaging Systems and Technology*, vol. 33, no. 1, pp. 18–38, Sep. 2022, doi: 10.1002/ima.22812.

[34] H. Mkindu, L. Wu, and Y. Zhao, "Lung nodule detection of CT images based on combining 3D-CNN and squeeze-and-excitation networks," *Multimedia Tools and Applications*, vol. 82, no. 17, pp. 25747–25760, Mar. 2023, doi: 10.1007/s11042-023-14581-0.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 1 | pp. 42-68 | Jan-Mar 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 30 Nov 2025 | Revised: 02 Mar 2026 | Accepted: 08 Mar 2026 | Published: 16 Mar 2026

[35] Purba Daru Kusuma, "Focus and Shake Algorithm: A New Stochastic Optimization Employing Strict and Randomized Dimension Mappings," *International Journal of Intelligent Engineering & Systems*, vol. 17, no. 3, pp. 551-562, 2024, doi: 10.22266/ijies2024.0630.43.

[36] D. Dahiya, "COVID-19 disease prediction utilizing dilated convolution neural network based levy flight tunicate swarm optimization," *Wireless Personal Communications*, vol. 131, no. 3, pp. 1515–1528, May 2023, doi: 10.1007/s11277-023-10505-1.

[37] Minki Kim, Byoung-Dai Lee, "Automatic Lung Segmentation on Chest X-rays Using Self-Attention Deep Neural Network," *Sensors*, vol. 21, no. 2, 369, 2021, doi: 10.3390/s21020369.

[38] F. Karim *et al.*, "Towards an effective model for lung disease classification," *Applied Soft Computing*, vol. 124, p. 109077, May 2022, doi: 10.1016/j.asoc.2022.109077.

[39] Mohammad Dehghani, Eva Trojovská, Pavel Trojovský, Om Parkash Malik, "OOBO: A New Metaheuristic Algorithm for Solving Optimization Problems," *Biomimetics*, vol. 8, no. 6, 468, 2023, doi:10.3390/biomimetics8060468.

[40] N. L. Yadav, S. Singh, R. Kumar, and D. K. Nishad, "Transfer learning with fuzzy decision support for multi-class lung disease classification: performance analysis of pre-trained CNN models," *Scientific Reports*, vol. 15, no. 1, p. 35127, Oct. 2025, doi: 10.1038/s41598-025-19114-3.

[41] E. Rajasekar, H. Chandra, N. Pears, S. Vairavasundaram, and K. Kotecha, "Lung image quality assessment and diagnosis using generative autoencoders in unsupervised ensemble learning," *Biomedical Signal Processing and Control*, vol. 102, p. 107268, Nov. 2024, doi: 10.1016/j.bspc.2024.107268.