An Enhanced AlexNet-VGG Hybrid with SiLU Activation and Multi-dimensional Slicing for Lung Infection Classification from 3D CT-scans

Bambang Suprihatin [1], Yuli Andriani [1], Anita Desiani [1,*], Siti Rusdiana Puspa Dewi [2], Muhammad Arhami [1], Deshinta Arrova Dewi [3], and Silfani Cahaya Putri [1]

[1] Mathematics, Faculty of Mathematics and Natural Science, Universitas Sriwijaya, Inderalaya, Indonesia
[2] Faculty of Medicine, Universitas Sriwijaya, Inderalaya, Indonesia
[3] Faculty of Data Science, International University, Malaysia
Email: bambangs@unsri.ac.id (B.S.); yuliandriani@unsri.ac.id (Y.A.); anita_desiani@unsri.ac.id (A.D.); sitirusdiana@fk.unsri.ac.id (S.R.P.D.); Muhammad.arhami@pln.co.id (M.A.); desinta.ad@newinti.edu.my (D.A.D.); 08011382025089@student.unsri.ac.id (S.C.P.)
*Corresponding author

***Abstract*—The use of deep learning for automatic classification has expanded rapidly, particularly through Convolutional Neural Network (CNN) architectures. AlexNet, a widely used CNN, offers a simple and stable structure. However, its performance on Computed Tomography (CT) scan images is limited by challenges in processing high-dimensional data and the scarcity of 3D CT scan datasets, often resulting in suboptimal classification outcomes. The Multi-Dimensional Slicing (MDS) method addresses this issue by converting 3D CT scan images into 2D representations along the axial, coronal, and sagittal planes. This technique enables comprehensive analysis of lung structures using more manageable datasets while preserving critical information. In this study, MDS is integrated with the ALVS-Net architecture, an improved version of AlexNet that incorporates a Visual Geometry Group (VGG) based structure in the final layer to reduce parameter complexity and enhance training stability. The Sigmoid-weighted Linear Unit (SiLU) activation function replaces the traditional Rectified Linear Unit (ReLU) function to improve nonlinear learning capabilities. The results of this study show that the proposed combination of MDS and ALVS-Net achieves excellent performance in classifying lung infections into four classes. This model achieves an accuracy, precision, recall, and F1−Score, all reaching 94%, a G-mean value of 93%, and a Cohen's kappa value of 92%. These results suggest that the proposed method is effective and reliable for classifying lung infections based on CT scan images.**

***Keywords*—lung infection, Computed Tomography (CT) scan, health risk, public health**

I. Introduction

Lung infections are recognized as one of the leading causes of mortality worldwide [1]. These infections vary in type and severity and are generally classified into four categories: mild, moderate, severe, and normal. The detection of lung infections can be performed using a Computed Tomography (CT) scan [2]. A CT scan is a medical imaging tool that provides a three-dimensional visualization of the internal structure of human organs. Compared to X-ray images, CT scan images offer more detailed and complex information. Through CT scans, all parts of an organ can be displayed, ensuring no information is lost during the examination [3]. CT scan images can be utilized to classify the types of lung infections experienced by patients. However, this classification process requires specialized expertise and is highly subjective, with factors such as fatigue and individual experience potentially influencing the results [4]. Recently, automatic classification methods based on deep learning have been rapidly developed. One commonly applied deep learning architecture is the Convolutional Neural Network (CNN) [5].

CNN is a deep learning method composed of several convolutional layers, which enhance the model's ability to learn data in detail [6]. One of the simpler and widely used CNN architectures is AlexNet, which typically consists of five convolutional layers followed by max-pooling layers and three fully connected layers [7]. This structure allows for quick dimensionality reduction and a lighter training process, making it efficient on low-power devices. The direct application of 3D CT scan images to the AlexNet architecture has been considered less effective due to increased network complexity and resource demands. The simple layer structure of AlexNet is insufficient to extract deep, complex features from 3D CT scan images, which are inherently more intricate than 2D X-ray images [8]. CT scan images in 3D consist of three dimensions-depth, length, and width-resulting in more complex input data. AlexNet was originally designed for 2D image processing, so modifications are needed to apply it effectively to 3D CT scan images. Using AlexNet in this context can increase memory usage and computational load. The performance of the AlexNet-based model also requires a

Manuscript received February 9, 2026; revised April 2, 2026; accepted April 14, 2026; published July 28, 2026.

large amount of training data, which remains limited for 3D CT scanners [9]. Slicing techniques can convert 3D images into 2D representations by cutting images along certain dimensions to facilitate analysis [10].

The Multi-Dimensional Slicing (MDS) enables 3D images to be sliced into 2D images by considering three axes: axial, coronal, and sagittal [11]. This technique provides the advantage of capturing images by multiple perspectives-axial slices from top to bottom, coronal from front to back, and sagittal from right to left, thus preserving more comprehensive information [12]. MDS slices images based on the number of voxels in each dimension of the CT scan image, enabling thorough feature representation from various viewpoints.

Besides the data availability, the model performance largely depends on architectural complexity [13]. Shallow and simple networks may fail to achieve optimal performance, particularly in multiclass classification tasks [14]. A more complex and deeper architecture is the Visual Geometry Group (VGG). This architecture is well-suited for large and multiclass datasets, allowing detailed feature extraction and improved classification performance [15]. Both AlexNet and VGG commonly employ the Rectified Linear Unit (ReLU) activation function in their hidden layers. ReLU is simple and widely used; it deactivates neurons with negative values, turning them into zeros, which inhibits learning during training and affects weight updates [16]. The ReLU activation function can be replaced with the Sigmoid-weighted Linear Unit (SiLU) activation function, which allows negative values to be learned and has been shown to improve the model [17].

This study proposes a two-stage method: the first stage involves implementing the MDS technique, and the second stage applies the ALVS-Net architecture. The MDS method is used to slice 3D CT scan images into 2D images along axial, coronal, and sagittal axes while preserving the full informational content of the original images. This technique also helps augment the training dataset, which is essential for accurate classification of lung infections. In the second stage, the ALVS-Net architecture is implemented. This architecture modifies AlexNet by placing it in the initial block, followed by the VGG architecture in the final block. The ReLU activation function is replaced with SiLU in the hidden layers to improve the model's ability to learn from negative values. The use of AlexNet in the initial block helps reduce parameters due to VGG's depth, while VGG in the final block enables deep and detailed feature extraction. Replacing ReLU with SiLU enables the learning of negative neurons and improves classification performance. The performance of this method is evaluated using accuracy, precision, recall, F1−Score, Cohen's kappa, and g-means. The classification is conducted across four categories-mild, moderate, severe, and normal-to provide a detailed assessment of lung infection conditions.

The novelty of this study lies in a unified framework that explicitly leverages multi-view representations (axial, coronal, and sagittal) from 3D CT scans. The framework features a hybrid deep architecture integrating AlexNet and VGG within ALVS-Net. Additionally, the SiLU activation function further enhances the feature learning and improves classification performance.

The main contributions of this study are as follows:

- Proposing a two-stage framework that combines the MDS technique and the combination of AlexNet and VGG (ALVS-Net) architecture for lung infection classification;
- Utilizing MDS to convert 3D CT scan images into multi-view 2D slices in axial, coronal, and sagittal views. This approach preserves the original spatial information and increases data availability.
- Developing the ALVS-Net architecture by integrating AlexNet in the initial block and VGG in the final block to balance computational efficiency and deep feature extraction;
- Replacing the ReLU activation function with SiLU to improve the model's ability to learn from negative feature values and enhance classification performance.
- Evaluating the proposed model using multiple performance metrics, including accuracy, precision, recall, F1−Score, Cohen's kappa, and g-means, across four classes: normal, mild, moderate, and severe.

This paper is organized into six sections. Section I presents the introduction. Section II reviews related work. Section III describes the dataset, the proposed methodology, and the training and testing procedures. Section IV presents the results of the proposed approach, including the system's response to forearm orientation. Section V provides a discussion of the results, comparisons with related studies, and identified limitations. Finally, Section VI concludes the paper by summarizing the objectives, key findings, and directions for future research.

II. Literature Review

A. 3D CT Scan Image

A 3D CT scan image is a medical image produced by a CT scanner. A CT scanner is a medical imaging device that uses rotating X-rays and digital detectors to capture data from various angles around the patient's body. The data of internal organs is reconstructed by a computer system into a three-dimensional image, also known as a 3D CT scan image [18]. 3D CT scan images are widely used as medical images to detect various abnormalities such as lung nodules, pneumonia, tuberculosis, lung cancer, and others. 3D CT scan images visualize the internal structure of organs clearly and comprehensively [19]. The main difference between 3D CT scan images and other medical images, such as X-rays, lies in the visualization approach and the level of information provided, where CT scans offer three-dimensional volumetric data with higher spatial resolution, higher contrast, and clearer depth information. In contrast, X-ray images only produce two-dimensional projections with overlapping anatomical structures, which limits the level of detail in the observed objects and consequently affects diagnostic accuracy.

B. CNN Architectures

CNN is a deep learning method based on artificial neural networks that is currently being developed in various architectures for classification, detection, and segmentation. Some CNN architectures for classification are:

1) Alexnet architecture

AlexNet is a CNN architecture developed by Alex Krizhevsky and colleagues that won the ImageNet competition. This architecture is designed to handle large-scale image classification tasks. AlexNet can extract visual features efficiently and utilizes GPUs to accelerate the training process. AlexNet is composed of five convolutional layers combined with ReLU and max pooling, as well as three fully connected layers, and the application of dropout to minimize the risk of overfitting with the final layer using the softmax function for class grouping in classification. The AlexNet architecture is illustrated in Fig. 1.

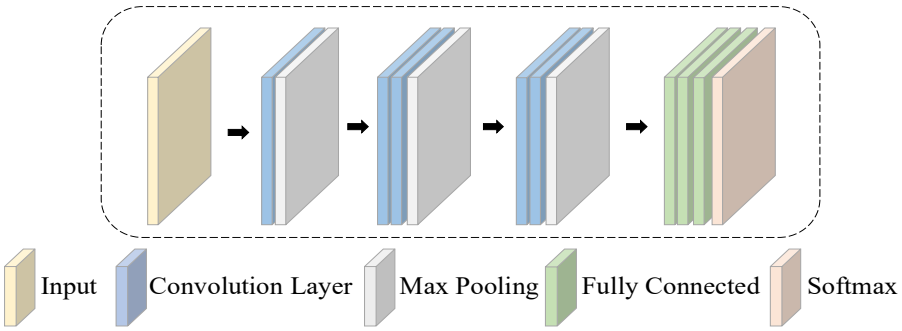


Fig. 1. Alexet architecture for image classification.

2) VGG architecture

The VGG architecture is a sequentially structured, deep CNN architecture that uses small 3×3 convolutional kernels. The purpose of using small kernels is to speed up the training process [20]. The most common variants, VGG-16 and VGG-19, consist of 16 and 19 layers, respectively. The operations in the VGG architecture are the same as those in AlexNet: convolution operations, ReLU activation, Maxpooling, dropout, and softmax in the last layer. The differences lie in the kernel size and layer depth. The VGG architecture is illustrated in Fig. 2.

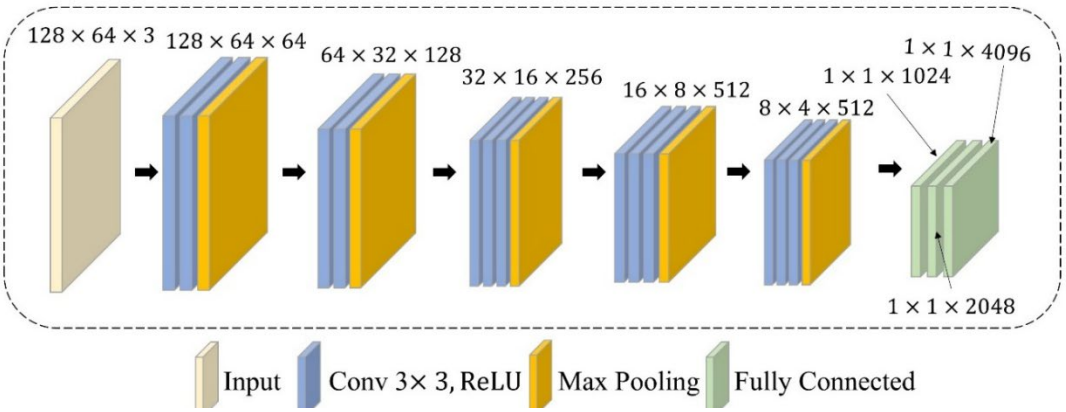


Fig. 2. VGG architecture for image classification.

3) Operation on CNN method

CNN methods have been developed with various architectural variations, but basically have the same operating standards, namely:

a) Convolution layer

Convolution layers are the primary components of a CNN that extract features from input data. A 2.5-Dimensional CNN (2.5D CNN) combines several adjacent CT slices as a multi-channel input. The use of a 2.5D CNN aims to preserve the volumetric information of 3D images, so that relationships between slices can still be represented without requiring a 3D CNN architecture, which is computationally more complex.

The convolution operation follows the standard formulation in CNN, where a kernel slides over the input feature maps to produce output feature maps through a weighted summation between the input and the convolution kernel [21]. The convolution process in a 2.5D CNN can be expressed using Eq. (1).

$$a_{i,j,q} = \sum_{c=1}^{k} \sum_{u=0}^{m-1} \sum_{v=0}^{n-1} \left(c_{u+i,v+j,c} \times k_{u,v,c,q} \right) + b_q \quad (1)$$

where $a_{i,j,q}$ is the value of the convolution output located at the i-th row and j-th column of the q-th channel. While m and n are defined as the length and width of the input matrix and kernel, respectively, $c_{u+i,v+j,c}$ represents the number of entries of the input matrix in the $(u+i)$-th row

and $(v+j)$-th column in the c-th channel, where each channel corresponds to an adjacent CT slice, $k_{u,v,c,q}$ denotes the number of entries of the kernel matrix in the $(u+i)$-th row and $(v+j)$-th column, i is identified as the row of the convolution matrix, j is identified as the n-th column of the convolution matrix, and b_q is recognized as the bias in the q-th kernel. The variable k denotes the number of input channels.

b) Batch normalization

Batch Normalization (BN) is a technique used to normalize intermediate layers in a neural network, stabilizing and speeding up the training process while improving model accuracy.

where m is the number of rows in the matrix, μ_j is the entry normalization process uses the following Eq. (2):

$$n_{i,j} = \frac{a_{i,j}-\mu_j}{\sqrt{\sigma_j^2+\varepsilon}} \quad (2)$$

Eq. (3) shows that the μ_j is the average of each mini-batch for the j-th column, σ_j^2 is the variance in the hidden layer for the j-th column, and $a_{i,j}$ is the matrix element resulting from the convolution of matrix A, $n_{i,j}$ is an entry of the matrix that has undergone the normalization process in the i-th row and j-th column, and ε is the error.

c) Activation function

Some types of activation functions include SiLU and softmax. SiLU is an activation function that uses a sigmoid function multiplied by its own input. The SiLU activation function is calculated as the softmax function multiplied by its input [17]. The softmax activation function is used for multi-class classification, converting values to a range between 0 and 1. The SiLU calculation is shown in Eq. (3)

$$z_{i,j} = n_{i,j}\frac{1}{1+e^{-n_{i,j}}} \quad (3)$$

Here, n is obtained from the calculation in BN, $n_{i,j}$ refers to the entry of the n-matrix located at the i-th row and j-th column, and e represents the base of the natural logarithm. The output produced by the SiLU activation function is subsequently substituted into the softmax activation function, as described in Eq. (4) [22].

$$s_{i,j} = \frac{e^{z_{i,j}}}{\sum_{i=1}^{m}\sum_{j=1}^{n} e^{z_{i,j}}} \quad (4)$$

where $s_{i,j}$ denotes the softmax output and z is the result of the SiLU process, $z_{i,j}$ refers to the entry of the matrix z at the i-th row and j-th column, for $i,j = 1,2,3 \ldots n$, where n is the number of classes.

d) Loss function: Categorical cross-entropy

Categorical cross-entropy is used as a loss function in classification tasks. Its calculation is shown in Eq. (5) [23].

$$L(y,s) = -\sum_{i=1}^{m}\sum_{j=1}^{n} y_{i,j}\log s_{i,j} \quad (5)$$

where $L(y,s)$ is the result of the categorical cross-entropy loss, $y_{i,j}$ represents the entry of the input class matrix at the i-th row and j-th column, and $s_{i,j}$ denotes the classification result at the i-th row and j-th column.

C. Related Works

Militante *et al.* [24] implemented the AlexNet architecture for lung infection classification using X-ray images and achieved an accuracy of 96%, precision of 89%, recall of 86%, and an F1−Score of 89%. Jamil *et al.* [25] also implemented AlexNet for the classification of lung diseases and achieved an accuracy of 98%. However, both studies were limited to binary classification, normal and infection, and only used X-ray data. A similar approach was adopted by Xu *et al.* [26] using three-class classification with the same architecture and achieved an accuracy and F1−Score of 95%, with a recall of 98%.

The AlexNet architecture was also implemented for 3D CT scan images. Hasija *et al.* [27] used AlexNet architecture to classify lung infections into normal, pneumonia, and COVID-19 categories, achieving an accuracy of 95%, a recall of 93%, and an F1−Score of 93%. Agrawal *et al*. [28] applied AlexNet to 3D CT scan images, achieving an accuracy of 95%, a recall of 90%, and a precision of 85%. However, both studies relied on arbitrary cropping techniques. Poorly performed image slicing may result in the loss of important spatial details and restricted viewing angles, compromising classification accuracy and hindering the model's ability to identify useful features [29].

In addition, deeper CNN architectures, such as VGG, have been adopted. Agarwala *et al.* [30] applied the VGG architecture to CT scan images using 2D slicing techniques and achieved an accuracy of 92%, a recall of 90%, and an F1−Score of 90%. However, the study also reported overfitting issues due to the limited dataset and insufficient regularization during preprocessing. Putra *et al.* [31] used VGG-19 and showed impressive performance, with an accuracy rate of 94,6% in classifying lung CT scans. However, their study focused on a three-class classification task, whereas this study addresses a more complex four-class classification problem.

Recent research has also explored more advanced deep learning strategies. Pavel *et al.* [32] proposed a knowledge distillation framework with a teaching assistant to improve lung cancer classification performance while reducing model complexity. The studies achieved accuracies of 90.99% and 94.53% using two different architectures, namely ResNet152 and CNN. Similarly, Irtaza *et al.* [33] investigated multi-label classification of lung diseases using various deep learning architectures, achieving a binary accuracy of 93.4% and an F1−Score of 0.582 after data augmentation.

Recent research has developed activation functions to improve learning capabilities. Kahwachi and Saed [34] applied SiLU in the DenseNet-121 architecture and achieved 91% accuracy, 88% recall, 91% precision, and a 90% F1−Score. Junior *et al.* [35] used the EfficientNet architecture with the SiLU function on X-ray images and achieved 95% accuracy in binary classification.

Existing studies show that deep learning performs well at classifying lung infections. Several limitations can still be identified as follows:

- Data Representation Limitation: Using 2D images or basic slicing techniques can miss important spatial information in 3D CT scan data. Prior work often lacks clear slicing definitions or uses limited viewing axes. This study clearly defines the slicing process and uses all three anatomical planes, axial, coronal, and sagittal, to preserve spatial information.
- Model Architecture Limitation: Simple architectures like AlexNet may not capture complex features, while deeper ones like VGG can overfit small datasets. This study proposes a hybrid (ALVS-Net) architecture that combines AlexNet in the initial layers and VGG in the final layers for efficient deep feature extraction.
- Classification Scope Limitation: Some studies address only binary or basic multi-class classification, thereby missing the full range of severity. This study uses four categories: normal, mild, moderate, and severe to provide a more comprehensive representation of disease progression.
- Learning Capability Limitation: ReLU activation may restrict learning by ignoring negative values. This study uses the SiLU activation function to enhance nonlinear learning and improve feature representation.

This study proposes a two-stage framework combining MDS with ALVS-Net for robust four-class lung infection classification, preserving spatial information and improving feature learning.

III. Materials and Methods

This research consists of several stages: slicing, architecture modification, training, testing, and evaluation. The overall stages of this research are illustrated in Fig. 3.

The CT scan 3D images will be processed by slicing using the MDS technique, applying slices based on the axial, coronal, and sagittal axes until a new dataset with 2D dimensions is obtained. The training data from this new dataset will be used in the training and validation stages, utilizing the ALVS-Net architecture. The testing process employs the latest dataset and the best weights derived from the training process. The performance evaluation of this method will be conducted by measuring accuracy, precision, recall, Cohen's kappa, and F1−Score.

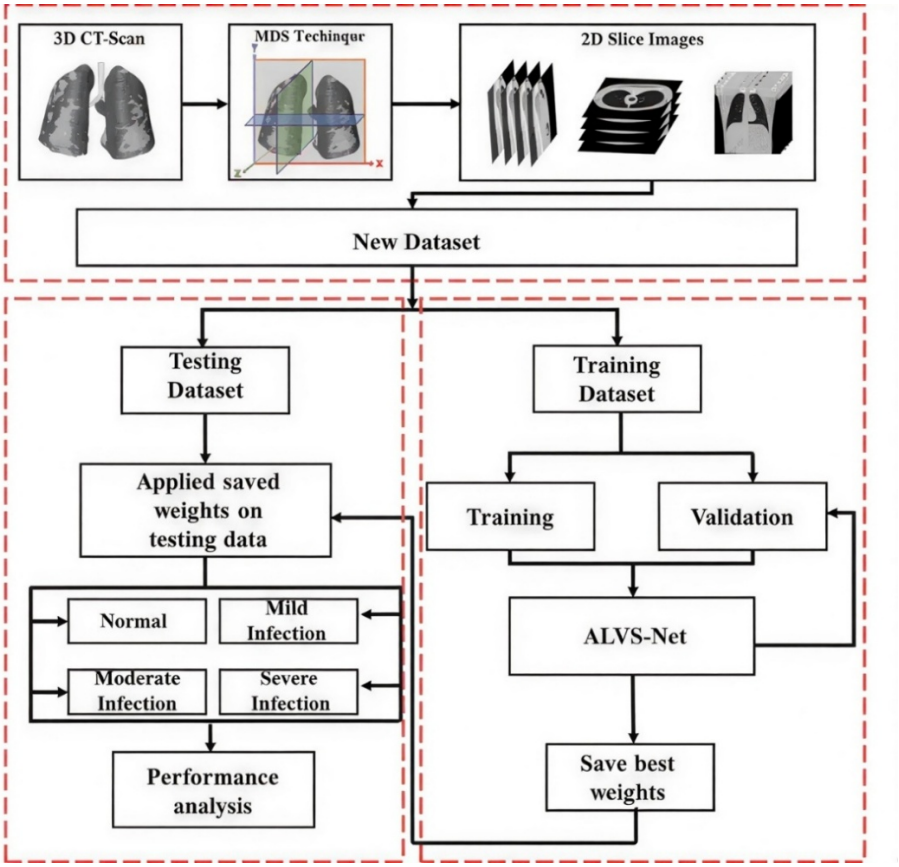
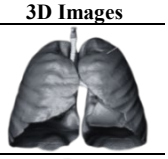
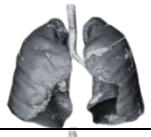
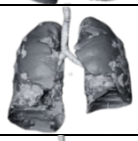
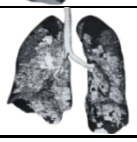


Fig. 3. Proposed method in the implementation of the ALVS-Net architecture for lung infection classification.

A. Data Collection

The data used in this study is publicly available in the MosMedData dataset [36], which can be accessed via the official repository at: https://www.kaggle.com/datasets /mathurinache/mosmeddata-chest-ct-scans-with-covid19. The total number of datasets is 1110, which are divided into four classes: 283 normal class datasets, 277 mild infection class datasets, 269 moderate infection class datasets, and 281 severe infection class datasets, all with the NIfTI file extension. These images have dimensions of 100×100×100 voxels. Several examples of the datasets can be seen in Table I.

TABLE I. DISPLAY OF SEVERAL IMAGES FROM THE CT SCAN DATASET

No	Class	3D Images
1.	Normal	
2.	Mild Infection	
3.	Moderate Infection	
4.	Severe Infection	

B. Training and Testing

During training, the proposed ALVS-Net architecture learns model parameters and generates optimal weights, which are subsequently used in the testing phase. To ensure a comprehensive evaluation of model robustness and generalization, multiple train–test split ratios were employed. Specifically, three configurations: 80:20, 75:25, and 70:30. The use of these split ratios follows common practices in machine learning, where datasets are partitioned into training and testing subsets to evaluate model performance. Ratios such as 80:20, 75:25, and 70:30 are widely adopted to balance sufficient training data and reliable evaluation results [37, 38].

Evaluating multiple split configurations reduces dependency on a single data partition and provides a more robust assessment of model performance, as supported by prior studies on model validation and evaluation [39]. For each configuration, the dataset was split into training and test sets, ensuring consistent class distribution across all categories. The same training settings and hyperparameters were applied across all configurations to ensure a fair and unbiased comparison. During the testing phase, the learned weights were used to evaluate the proposed architecture's performance in classifying lung infections. The prediction results were categorized into four classes: normal, mild infection, moderate infection, and severe infection.

C. MDS Technique

MDS is defined in this study as a technique for converting 3D volumetric images into multiple 2D representations by extracting slices from the volume. The slicing technique cuts images based on three axes, namely the X-axis (coronal plane), which cuts from front to back (or vice versa), the Y-axis (axial plane), which cuts from top to bottom (or vice versa), and the Z-axis (sagittal plane), which cuts from right to left [40]. The MDS technique can be seen in Fig. 4.

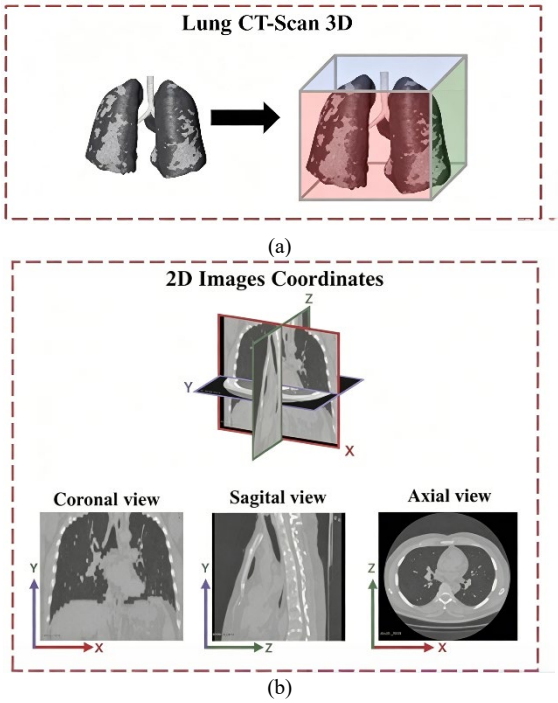


Fig. 4. Illustration of the MDS technique (a) Illustration of a 3D image before slicing (b) Result of slicing the 3D CT scan image into 2D images.

Fig. 4 illustrates the MDS process applied to CT scan images of the lungs. The 3D lung model is generated through image processing and placed within a bounding box representing a 3D volume with dimensions of 100 × 100 × 100 voxels. Each 3D image is sliced into 100 sections along each dimension, resulting in a total of 300 new images: 100 images along the coronal axis, 100 along the axial axis, and 100 along the sagittal axis. The slicing is performed based on the length, width, and depth of each axis to ensure that the information remains intact and can be accurately reconstructed in 2D. This process produces three primary views: coronal, sagittal, and axial, which allow for in-depth analysis from multiple perspectives, aiding in the classification and diagnosis of lung diseases.

D. ALVS-Net Architecture

The ALVS-Net architecture combines AlexNet, VGG, and SiLU activation functions. Generally, the AlexNet architecture consists of five convolutional layers, with each block containing only one layer.

The AlexNet architecture will be used in the initial block of the model, which aims to reduce excess parameters during the training process. In the final block, the VGG architecture will be implemented, as it is capable of learning from large datasets and extracting more complex features. The ReLU activation function in each architecture will be replaced with the SiLU activation function, which aims to improve the performance and stability of the ALVS-Net architecture. The SiLU activation function is illustrated in Fig. 5.

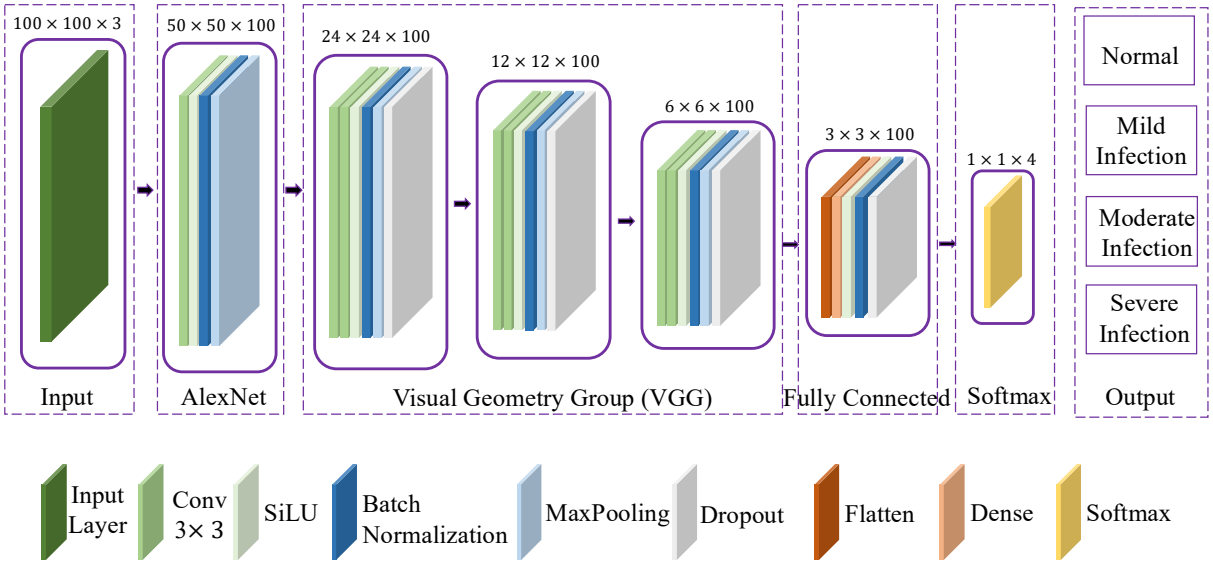


Fig. 5. Modification of VGG and alexnet architecture by replacing the ReLU activation function with SiLU.

Fig. 5 illustrates the ALVS-Net architecture, which consists of five main blocks incorporating max pooling and dropout layers, with an input image size of 100 × 100 pixels. The first block adopts the AlexNet architecture, including a 3×3 convolution layer, SiLU activation, batch normalization, and max pooling, to reduce the number of model parameters. The second to fourth blocks apply the VGG architecture, incorporating two 3×3 convolutional layers, SiLU activation, batch normalization, max pooling, and dropout, which enables deeper feature extraction while maintaining structural consistency. The final block consists of a fully connected layer with a softmax activation function for classification.

E. Evaluation

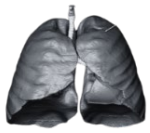
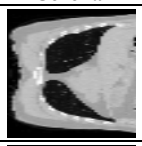
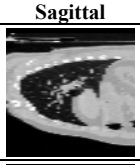
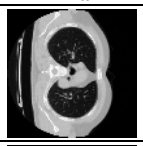
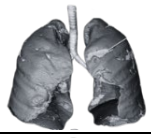
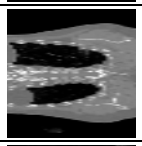
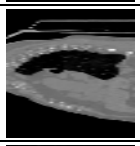
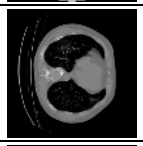
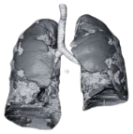
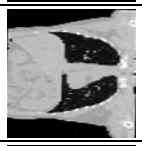
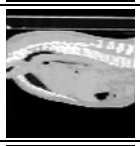
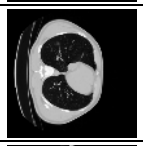
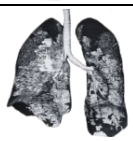
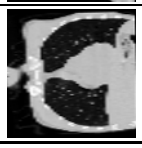
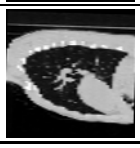
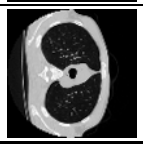
In the classification of lung infections, each image will be categorized into four categories: normal, mild infection, moderate infection, and severe infection. In this study, model performance evaluation was based on the accuracy value, which measures the precision of model predictions, to provide an overview of the model's overall performance [41]. Accuracy measures the model's ability to accurately classify true positive labels, aiming to assess the accuracy in detecting data from various classes [42]. Recall measures the model's ability to correctly classify true positive labels, to identify all relevant cases [43]. F1−Score measures the balance between precision and recall, aiming to assess the model's ability to detect true positives while minimizing classification error [44]. Cohen's kappa calculates the level of agreement between the prediction results produced [45]. G-Means assesses the balance between sensitivity and specificity, aiming to evaluate the imbalance of positive and negative classes [46].

IV. Result

A. MDS Technique

The MDS technique aims to convert 3D CT scan images into 2D images by slicing along the axial, coronal, and sagittal axes. Each image generates 300 new slices: 100 from the coronal axis, 100 from the axial axis, and 100 from the sagittal axis, adjusted according to the number of voxels in the original image. The goal of this method is to enhance the diversity and volume of training data. The 3D CT scan image was sliced along three axes: coronal, axial, and sagittal. The number of slices was determined based on the voxel dimensions of each 3D image. Specifically, slices were generated along the coronal axis, sagittal axis, and axial axis of the 3D image. All parts of the 3D image were fully visualized without loss of information. A total of 297,900 images was produced using the MDS technique, generated by slicing each 3D image into 100 images along the coronal, axial, and sagittal axes. The resulting dataset consisted of 74,550 normal images, 74,400 mild infection images, 74,475 moderate infection images, and 74,475 severe infection images. The class distribution is relatively balanced across all categories, with only minor differences in sample counts. Therefore, the dataset does not exhibit significant class imbalance, and all classes are adequately represented during training. Based on this balanced distribution, standard evaluation metrics are sufficient to assess model performance without the need for weighted metrics. Table II presents the results of the MDS technique for each class.

TABLE II. RESULTS OF THE MULTI-DIMENSIONAL SLICING (MDS) TECHNIQUE

No	Class	3D Images	2D MDS Image Views		
			Coronal	Sagittal	Axial
1.	Normal				
2.	Mild Infection				
3.	Moderate Infection				
4.	Severe Infection				

B. Experimental Setup

1) Hardware and software environment

All experiments were conducted on a workstation equipped with an Intel Core i3-10100 CPU, 32 GB RAM, and an NVIDIA RTX 3090 GPU with 24 GB GDDR6X VRAM running on Windows 11 64-bit. The software environment used in this study includes Python 3.10.0, TensorFlow 2.10, Jupyter Notebook 7.5.4, NumPy 1.23.0, OpenCV 4.7.0, and Matplotlib 3.10.7. All model development, training, and evaluation processes were implemented using these libraries.

2) Hyperparameter configuration

The hyperparameters used in this study were carefully selected based on a combination of standard practices in deep learning and empirical observations obtained from preliminary experiments. The selection process involved controlled trials to achieve a balance between training stability, convergence speed, and model generalization. The hyperparameter configuration for the classification models includes a learning rate of 0.00001, batch size of 64, 100 training epochs, weight decay of 0.001, and an input image size of 128×128 pixels.

This study uses a 2.5D CNN that sequentially combines multiple CT slices obtained from the MDS process. This combination enables the model to extract features not only from a single slice but also from relationships between slices, thereby capturing depth-based information without requiring a 3D CNN architecture, which is computationally more complex. The model is designed to classify four classes and utilizes the categorical cross-entropy loss function. No advanced hyperparameter optimization methods were applied. Instead, the hyperparameters were determined through empirical tuning based on controlled experimental trials. All hyperparameter settings were applied consistently across all models, including ALVS-Net, AlexNet, and VGG, to ensure a fair and unbiased comparison. The detailed architectural parameters and training configurations for each model are presented in Table III.

TABLE III. ARCHITECTURAL PARAMETERS AND TRAINING CONFIGURATION

Architecture	Total Params (Millions)	Trainable (Millions)	Non-Trainable (Millions)	Training Time (Min)
VGG + ReLU	168.94	168.94	0.0000	1144
VGG + SiLU	168.94	168.94	0.0000	1199
AlexNet + ReLU	24.74	24.74	0.0007	458
AlexNet + SiLU	24.74	24.74	0.0007	491
ALVS-Net + ReLU	1.80	1.80	0.0010	218
ALVS-Net + SiLU	1.80	1.80	0.0010	229

Table III shows that the proposed ALVS-Net model has significantly fewer parameters and requires substantially shorter training time compared to AlexNet and VGG. ALVS-Net contains approximately 1.81 million parameters, which is considerably lower than AlexNet's 24.74 million and VGG's 168.94 million. In terms of training efficiency, ALVS-Net also demonstrates faster training time, ranging from 218 to 229 minutes, compared to AlexNet, which requires 458 to 491 minutes, and VGG, which requires 1144 to 1199 minutes.

The architectures using the SiLU activation function generally require slightly longer training time than those using ReLU, which can be attributed to its higher computational complexity. These results indicate that the proposed ALVS-Net achieves a better balance between

model complexity and computational efficiency while maintaining strong classification performance.

C. Training

The training process for the ALVS-Net architecture used a dataset of 297,900 samples. To evaluate the model's robustness and generalization, three train–test split ratios were used: 80:20, 75:25, and 70:30. For each configuration, the dataset was partitioned to maintain a consistent class distribution. The model was trained for 100 epochs with a batch size of 64. Fig. 6 presents the model's training and validation performance for the 80:20, 75:25, and 70:30 splits. The blue curves represent training performance, while the orange curves indicate validation performance.

In Fig. 6(a), both training and validation accuracy increase rapidly during the early epochs and gradually stabilize at values close to 1.0, with both curves remaining closely aligned. In Fig. 6(b), the training and validation losses decrease steadily and converge to a low range of approximately 0.02-0.04, with only minor fluctuations, indicating a stable optimization process and minimal overfitting. In Fig. 6(c), the 70:30 split exhibits the most stable behavior, with the training and validation curves for both accuracy and loss remaining closely aligned throughout training, indicating consistent performance across the dataset. Overall, among the evaluated configurations, the 70:30 split provides the most balanced performance, characterized by stable convergence and minimal discrepancy between the training and validation curves, resulting in more reliable classification performance.

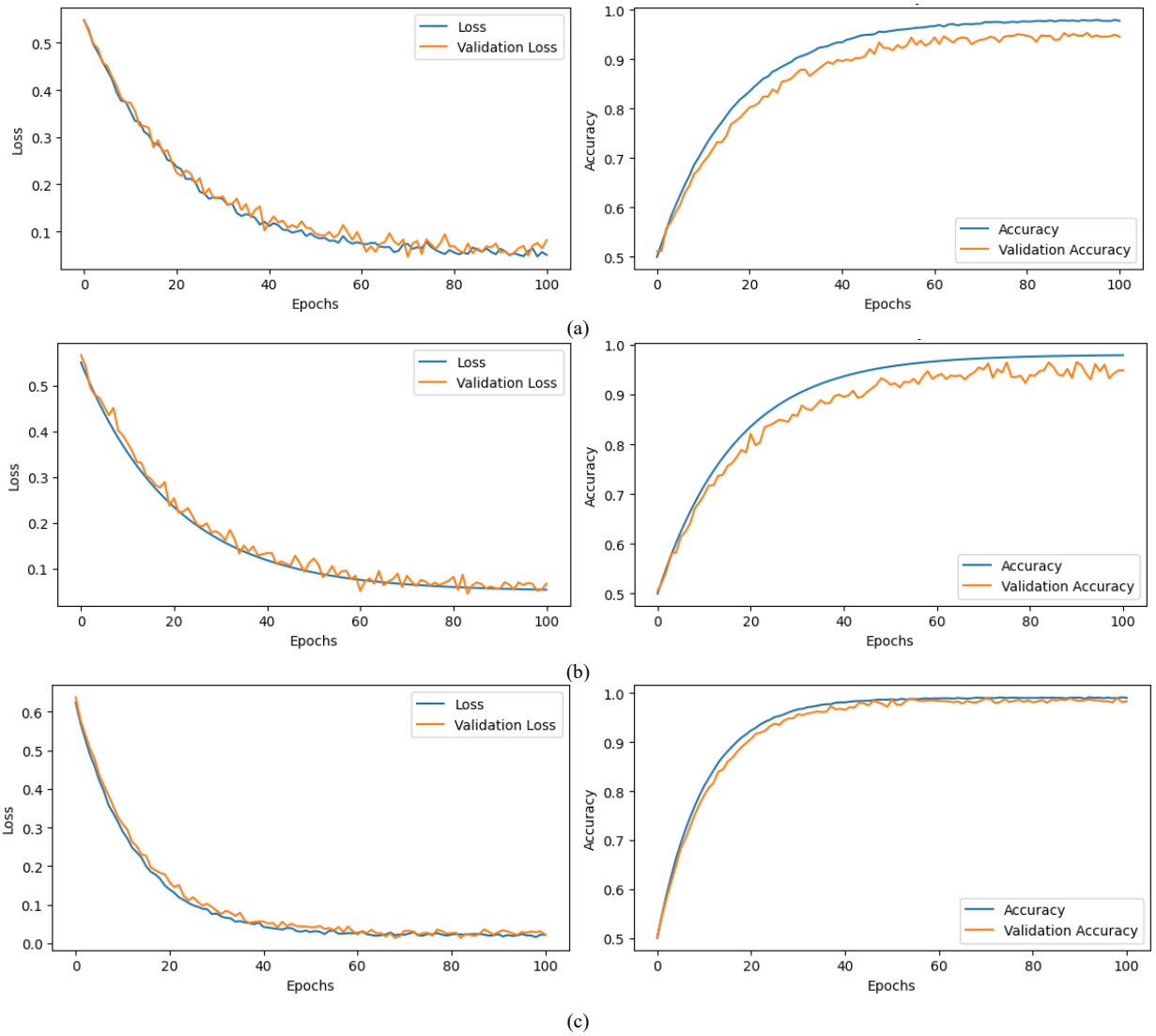


Fig. 6. Training and validation loss and accuracy for lung infection classification under different data split configurations: (a) 80:20 split, (b) 75:25 split, and (c) 70:30 split.

Additional evaluation metrics, including precision, recall, F1−Score, and Cohen's kappa, were calculated to provide a more comprehensive assessment of the model's performance. Relying solely on accuracy and loss is not sufficient, as these metrics do not fully capture the balance between classification errors and class-wise performance.

These additional metrics are presented for the 70:30 data split, which demonstrates the best overall

performance among the evaluated configurations. The corresponding graphs for precision, recall, F1−Score, and Cohen's kappa are shown in Figs. 7 and 8.

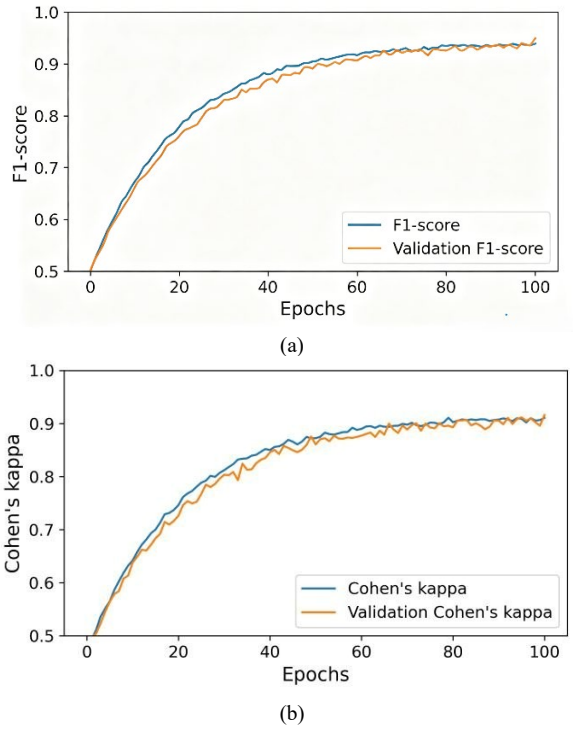


Fig. 7. Training results of the ALVS-Net architecture for lung infection classification with 70:30 data split: (a) F1−Score of training and validation data (b) Cohen's kappa of training and validation data.

In Fig. 7(a), the F1−Score increases steadily from early epochs and gradually stabilizes at a high value near 0.95. The F1−Score balances precision and recall, indicating that the model can reduce both false positives and false negatives simultaneously. The close alignment between the training and validation curves shows that this balance is maintained consistently on unseen data, indicating stable and reliable classification performance.

In Fig. 7(b), the Cohen's kappa values also increase smoothly and stabilize at a high level. Cohen's kappa measures the agreement between predicted labels and the ground truth while accounting for random chance. The high, stable values indicate that the model's predictions are not random but are based on meaningful, consistent feature patterns. The minimal gap between training and validation curves further shows that this agreement remains consistent across different data samples.

In Fig. 8(a), the precision curves increase gradually and stabilize at a high value close to 0.95. Precision represents the proportion of correctly predicted positive samples among all predicted positives. The close alignment between the training and validation curves indicates that the model produces very few false positives, and that this behavior is consistent not only on training data but also on unseen validation data. The smooth upward trend shows that the model progressively improves its confidence in positive predictions as training progresses.

In Fig. 8(b), the recall curves also rise steadily and reach similarly high values. Recall measures the proportion of correctly identified positive samples among all actual positives. The similarity between the training and validation curves indicates that the model can detect most true positives without significant performance degradation on the validation data. This indicates that important features related to positive classes are successfully learned and generalized.

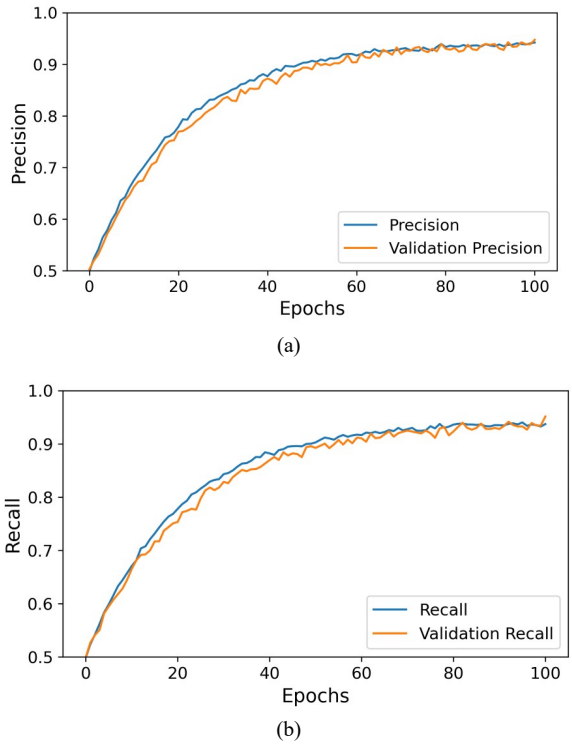


Fig. 8. Training results of the ALVS-net architecture for lung infection classification with 70:30 data Split: (a) Precision of training and validation data (b) Recall of training and validation data.

D. Testing

The evaluation was conducted using the best-performing weights obtained during training of the proposed ALVS-Net architecture. Model performance was assessed using a confusion matrix along with several evaluation metrics, including accuracy, precision, recall, F1−Score, and Cohen's kappa. Fig. 9 presents the confusion matrix of the best-performing ALVS-Net architecture using a 70:30 split.

For Mild Infection, 20,612 samples are correctly identified. However, 1,212 samples are misclassified as Moderate and 640 as Severe. This indicates that mild cases share visual similarities with higher severity levels, making the boundary between these classes less distinct. For Moderate Infection, 18,236 samples are correctly classified, while 2,505 samples are misclassified as Mild and 289 as Severe. The higher number of misclassifications toward Mild shows that Moderate cases overlap with adjacent categories, making intermediate severity levels more difficult to distinguish. For Severe Infection, 22,423 samples are correctly classified, with only a small number misclassified into other classes. This reflects that severe cases have more prominent and

distinctive features, allowing the model to identify them more reliably.

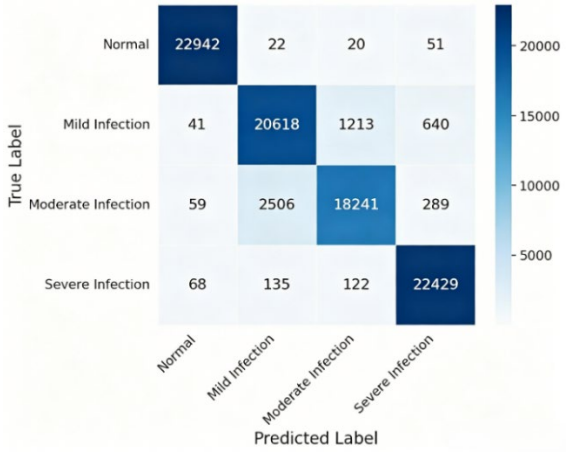


Fig. 9. Confusion matrix of ALVS-Net architecture using 70:30 data split for lung infection classification.

The dominance of diagonal values over off-diagonal elements demonstrates that the model effectively separates most classes. The misclassification patterns, particularly between Mild and Moderate, indicate that the model's performance is influenced by feature similarity across neighboring classes. In contrast, classes with more distinct characteristics, such as Normal and Severe, are more easily recognized. To analyze classification performance across classes, Fig. 10 presents the comparison of testing results for each class using the ALVS-Net architecture.

Fig. 10 presents the evaluation of model testing results based on accuracy, precision, recall, F1−Score, and Cohen's kappa for each class. The Normal class achieves the highest performance across all metrics, with accuracy of 99.75%, precision of 99.60%, recall of 99.43%, F1−Score of 99.51%, and Cohen's kappa of 99.35%. These near-perfect values demonstrate that the model identifies normal cases with very high confidence and minimal misclassification. The high recall means that almost all actual normal samples are correctly detected, while the high precision means that predictions labeled as normal are highly accurate. The Cohen's kappa value of 0.99 indicates very strong agreement between the predicted and actual labels, beyond random chance.

For the Mild Infection class, the model maintains strong performance, with accuracy above 95% and precision, recall, and F1−Score above 90%. However, Cohen's kappa is 87%, which is lower than that of other classes. This reflects a slight inconsistency in prediction agreement, suggesting that some samples are more difficult to classify correctly, likely because they are similar to neighboring classes.

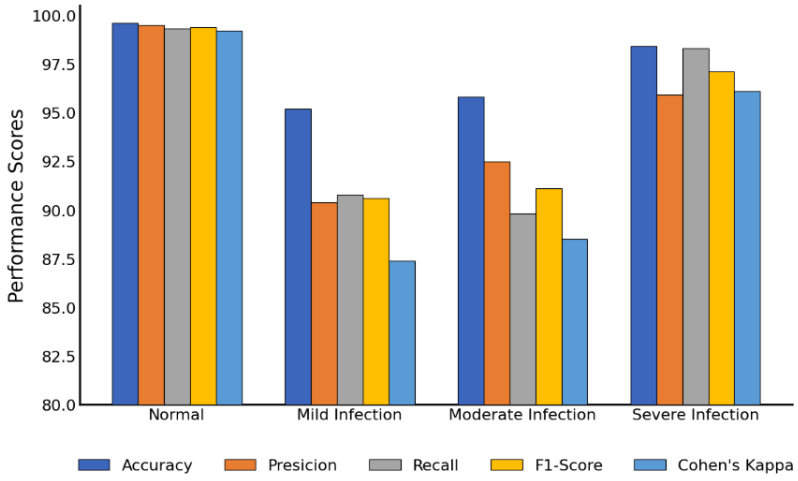


Fig. 10. Comparison of testing results for each class using the ALVS-Net architecture for lung infection classification.

For the Moderate Infection class, accuracy, precision, and F1−Score remain high, but recall and Cohen's kappa fall below 90%. The lower recall shows that a portion of Moderate cases is not correctly detected and is instead classified as other classes, particularly Mild. The reduced Cohen's kappa reflects weaker agreement between predictions and actual labels, indicating that classification results for this class are less stable. This pattern corresponds to overlapping feature characteristics between Moderate and adjacent classes, making separation more challenging.

In contrast, the Severe Infection class shows consistently high performance, with all metrics exceeding 95%. The high recall means that most severe cases are successfully detected, while the high precision indicates that predictions for this class are highly accurate. The strong Cohen's kappa confirms consistent and reliable predictions across samples. These results indicate that severe cases exhibit more prominent and distinguishable features, enabling clearer separation from other classes.

An additional performance metric evaluated in this study is the Receiver Operating Characteristic (ROC) curve. The ROC curve is a two-dimensional plot of the True Positive Rate (TPR) and False Positive Rate (FPR), with the Area Under the Curve (AUC) serving as a quantitative indicator of model performance. A higher AUC value signifies better classification capability. Fig. 11 presents the ROC curve and AUC values for lung infection classification using the ALVS-Net architecture.

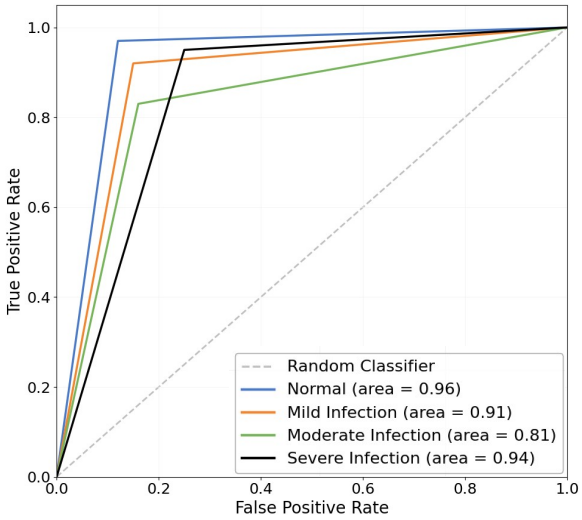


Fig. 11. ROC curve for lung infection classification using the ALVS-net architecture.

Fig. 11 presents the ROC curves and AUC values for each class in the lung infection classification task using the ALVS-Net architecture. The ROC curve illustrates the relationship between the true positive rate (sensitivity) and the false positive rate (1 − specificity), showing how well the model distinguishes between the classes at different threshold settings.

For the Normal class (blue line), the AUC value of 0.96 indicates a very strong ability to separate normal samples from infected classes. The curve rises sharply toward the top-left corner, meaning that the model achieves a high true positive rate while maintaining a low false positive rate. This reflects that normal cases are clearly distinguishable and rarely confused with infection classes. For the Mild Infection class (orange line), the AUC value of 0.91 shows strong classification performance. The curve remains close to the top-left region, indicating that most mild cases are correctly identified, although there is still some overlap with other classes. This aligns with the earlier observation that mild cases share similarities with neighboring severity levels.

For the Moderate Infection class (green line), the AUC value of 0.81 is lower compared to other classes. This indicates that the model has more difficulty separating Moderate cases from other categories. The curve lies further away from the top-left corner, meaning that increases in true positive rate are accompanied by relatively higher false positive rates. This reflects overlapping feature characteristics, which make Moderate cases more challenging to distinguish from Mild Infection. In contrast, the Severe Infection class (black line) achieves an AUC value of 0.94, indicating strong discriminative performance. The curve is close to the top-left corner, showing that the model can identify severe cases with high sensitivity while keeping false positives low. This reflects that severe cases contain more prominent and distinguishable features.

The differences in AUC values across classes reflect variations in feature separability. Higher AUC values, such as those for Normal and Severe classes, correspond to clearer class boundaries and more confident predictions. The lower AUC value for the Moderate class reflects overlapping feature distributions, which increases classification difficulty and aligns with the misclassification patterns observed in the confusion matrix.

V. Discussion

The performance of the ALVS-Net architecture for lung infection classification was evaluated through an ablation study to assess the contribution of each component. The experiments were conducted using multiple train-test split configurations, where the 70:30 split achieved the best overall performance and was therefore selected for detailed analysis. Table IV presents the comparative results of different architectural combinations, including VGG and AlexNet with ReLU and SiLU activation functions, as well as the proposed ALVS-Net model for lung infection classification.

TABLE IV. Ablation Study Results Based on Accuracy, Precision, and Recall for Lung Infection Classification

Architecture	Accuracy (%)	Precision (%)	Recall (%)
VGG + ReLU	91.22	91.14	91.02
VGG + SiLU	92.04	91.95	91.85
AlexNet + ReLU	87.53	87.41	87.29
AlexNet + SiLU	89.13	89.01	88.91
ALVS-Net + ReLU	92.94	92.87	92.76
ALVS-Net + SiLU	94.22	94.18	94.06

Table IV shows that the ALVS-Net model with the SiLU activation function achieves the highest performance among all evaluated configurations, with an accuracy of 94.22%, precision of 94.18%, and recall of 94.06%. The high accuracy indicates that the model correctly classifies the majority of samples. The high precision reflects that most predicted positive cases are correct, resulting in a low number of false positives. Meanwhile, the high recall indicates that the model identifies most of the actual positive cases, meaning that only a small number of true positives are missed during classification.

The baseline models, including VGG and AlexNet with both ReLU and SiLU activation functions, also demonstrate competitive performance across all metrics. However, the proposed ALVS-Net consistently provides improved results, particularly in maintaining a balance between precision and recall. This balanced performance reflects the model's ability to minimize both false positives and false negatives simultaneously.

In addition, the comparison of activation functions shows that both ReLU and SiLU provide stable performance, but using SiLU leads to a noticeable improvement of approximately 2% across the evaluated metrics. This improvement reflects more effective feature learning and contributes to better overall classification performance. Additional evaluation metrics are presented in Table V, including F1−Score, G-means, and Cohen's kappa, which provide a more comprehensive assessment of the model's balance, consistency, and reliability for lung infection classification.

Table V presents the proposed ALVS-Net with the SiLU activation function, which achieves the highest performance, with an F1−Score of 94.08%, G-means of 93.90%, and Cohen's kappa of 92.29%. The high F1−Score reflects a strong balance between precision and recall, indicating that both false positives and false negatives are effectively minimized.

TABLE V. ABLATION STUDY RESULTS BASED ON F1−SCORE, COHEN'S KAPPA, AND G-MEANS EVALUATIONS FOR LUNG CLASSIFICATION

Architecture	F1−Score (%)	G-Means (%)	Cohen's kappa (%)
VGG + ReLU	91,03	90,75	88,28
VGG + SiLU	91,86	91,62	89,38
AlexNet + ReLU	87,31	86,93	83,37
AlexNet + SiLU	88,92	88,60	85,50
ALVS-Net + ReLU	92,77	92,55	90,59
ALVS-Net + SiLU	94,08	93,90	92,29

The G-means result of 93.90% reflects balanced performance across classes, despite minor variations observed in class-wise results. This shows that the model maintains overall balance without significantly favoring any particular class. The Cohen's kappa value of 92.29% indicates strong agreement between predicted labels and ground truth beyond chance, confirming that the model produces reliable, consistent predictions across different samples.

These results demonstrate that the proposed model achieves stable and well-balanced performance across multiple evaluation metrics. The consistency across F1−Score, G-means, and Cohen's kappa reflects the model's ability to capture meaningful and representative features for each class, resulting in reliable classification outcomes. The effectiveness of the proposed model is evaluated by comparing it with existing studies on lung infection classification, as presented in Table VI.

TABLE VI. COMPARISON OF ALVS-NET ARCHITECTURE PERFORMANCE WITH OTHER STUDIES

Method	Slicing Technique	Accuracy (%)	Precision (%)	Recall (%)	Label
ResNet18 [47]	Wavelet-based 2D	**99**	**98**	**100**	2
ResNet50 [48]	2D-based	96	**98**	96	4
AlexNet [28]	Random	95	85	90	2
Super-xDnn [49]	Superpixels	**98**	**98**	98	2
ResNet50-MV [50]	Axial Slicing	84	80	**100**	**2**
ALVSNet-MDS (proposed)	Multi-Dimensional Slicing	94	94	94	4

Table VI shows that the proposed ALVS-Net-MDS achieves balanced performance across accuracy, precision, and recall, with each metric reaching 94%. Although several existing methods report higher values in specific metrics, it is important to note that many of these approaches are designed for binary classification tasks. In contrast, the proposed method addresses a more challenging multi-class problem with four labels.

This increased complexity requires the model to distinguish subtle variations in infection severity, making direct comparison with binary classification methods less straightforward. Unlike models that achieve high performance in only one or two metrics, the proposed method demonstrates consistent, uniform performance across all evaluation metrics, indicating stable, reliable classification behavior.

The integration of MDS enables the model to leverage information from the axial, coronal, and sagittal planes. This multi-plane representation preserves richer spatial context, which contributes to more consistent performance across different classes.

Table VII presents additional evaluation metrics, including F1−Score, Cohen's kappa, and G-means, which provide a more comprehensive assessment of model performance in terms of balance, agreement, and class-wise stability. The proposed ALVS-Net-MDS achieves an F1-score of 94%, a Cohen's kappa of 92%, and a G-mean of 93%, indicating strong and consistent performance across all evaluation aspects. While some methods achieve higher values in specific metrics, they are primarily evaluated on simpler binary classification tasks. In contrast, the proposed model maintains stable and well-balanced performance across multiple metrics in a more complex multi-class setting, demonstrating its robustness. The high F1−Score indicates an effective balance between precision and recall, while the Cohen's kappa value reflects strong agreement between predicted and true labels beyond chance. The G-means score further confirms that the model performs consistently across all classes without bias toward any particular category.

These results highlight that the proposed method offers a significant advantage in converting 3D images into 2D slices by segmenting from coronal, axial, and sagittal planes, and further dividing each plane into 100 2D image slices. This approach ensures that the entire structural information is preserved, thereby enhancing the comprehensive representation of lung anatomy. By maintaining the complete image characteristics from various viewpoints, this technique contributes to improved model accuracy in detecting and classifying lung infections with greater precision. In the future, ALVS-Net may be applied in automated diagnostic systems to detect various pulmonary diseases and to enhance the efficiency of medical data analysis.

TABLE VII. COMPARISON OF AVERAGE RESULTS BASED ON F1−SCORE, COHEN'S KAPPA, AND G-MEANS EVALUATIONS

Method	Slicing Technique	F1−Score (%)	Cohen's kappa (%)	G-Means (%)	Label
Super-xDnn [49]	SLIC Superpixels	**98**	**97**	**98**	**2**
ResNet-152 [51]	Superpixels	89	−	89	2
ResNet50-MV [50]	Axial Slicing	−	−	80	2
ALVS-Net-MDS (proposed)	Multi-Dimensional Slicing	94	92	93	4

VI. Conclusion

The lung infection classification process in this study was carried out through three main stages: pre-processing, training, and testing. In the pre-processing stage, techniques were applied to Multi-Dimensional Slicing (MDS) on 3D CT scan images, which produced 297,900 2D images from 1,110 3D CT scan images by slicing along three axes: axial, coronal, and sagittal. The MDS technique has proven to be effective in handling high data complexity, enabling it to meet training data needs and optimize model performance. At the training and testing stages, the ALVS-Net architecture was utilized, with a modification of the ReLU activation function to SiLU. The evaluation results show an accuracy of 97%, reflecting the model's ability to correctly classify all classes. Precision and recall each reached 97%, indicating that the model was able to accurately classify true positives and true negatives. The F1−Score value of 97% demonstrates an excellent balance between precision and recall in identifying true positives and minimizing classification errors. The Cohen's kappa value of 96% indicates a very good level of agreement between the model predictions and the actual results, while the G-Means is 97%. These results show that the ALVS-Net architecture effectively assesses how well the model predicts the correct classification area and determines the optimal number of clusters. The evaluation results for each class/label showed very good outcomes, with scores above 90% for the normal class, moderate infection, and severe infection. In the case of mild infections, the evaluation results indicated that Cohen's kappa remained at 87%, while other evaluations were above 90%. Overall, the evaluation results demonstrate that the combination of the MDS technique and the ALVS-Net architecture provides excellent performance in the accurate classification of lung infections.

Despite these promising results, several limitations should be noted. The model shows slightly lower consistency when distinguishing classes with similar visual characteristics, particularly between Mild and Moderate infections, due to the subtle differences between these categories. In addition, the current approach is limited to classification tasks and does not provide localization of infection regions within CT scan images.

Future work may focus on integrating segmentation techniques to enable more precise localization of infection areas. Additionally, expanding the dataset and conducting cross-dataset validation can help improve the model's robustness and generalizability.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Bambang Suprihatin conceptualized and designed the study, curated the data, performed the analysis, and drafted the manuscript. Yuli Andriani contributed to the critical review and substantive revisions. Anita Desiani was involved in data visualization and manuscript drafting. Siti Rusdiana Puspa Dewi conducted the literature search. Muhammad Arhami reviewed the manuscript. Deshinta Arrova Dewi contributed to the data curation process, and Silfani Cahaya Putri developed the study framework. All authors had approved the final version.

Funding

The publication of this article was funded by Universitas Sriwijaya 2025. In Accordance with the Rector's Decree No. 0014/UN9/SK.LPPM/2025.

References

[1] E. Yu and C. Li, "Global trends and attributable risk factors in the disease burden of lower respiratory infections," *Trop. Med. Infect. Dis.*, vol. 10, no. 180, pp. 1–15, 2025.

[2] D.-P. Fan *et al.*, "Inf-Net: Automatic COVID-19 Lung Infection Segmentation From CT Images," *IEEE Trans. Med. Imaging*, vol. 39, no. 8, pp. 2626–2637, Aug. 2020.

[3] V. Shah, R. Keniya, A. Shridharani, M. Punjabi, J. Shah, and N. Mehendale, "Diagnosis of COVID-19 using CT scan images and deep learning techniques," *Emerg. Radiol.*, vol. 28, no. 3, pp. 497–505, 2021.

[4] W. Tan *et al.*, "Classification of COVID-19 pneumonia from chest CT images based on reconstructed super-resolution images and VGG neural network," *Heal. Inf. Sci. Syst.*, vol. 9, no. 10, pp. 1–12, 2021.

[5] J. A. Ayeni, "Convolutional Neural Network (CNN): The architecture and applications," *Appl. J. Phys. Sci.*, vol. 4, no. 4, pp. 42–50, 2022.

[6] A. Garg and N. Karimian, "Leveraging deep CNN and transfer learning for side-channel attack," *Int. Symp. Qual. Electron. Des.*, vol. 22, pp. 91–96, 2021.

[7] X. Zhang, "The AlexNet, LeNet-5 and VGG NET applied to CIFAR-10," in *Proc. 2nd International Conference on Big Data & Artificial Intelligence and Software Engineering (ICBASE)*, 2021, pp. 414–419.

[8] 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. 915, pp. 1–24, Apr. 2025.

[9] A. M. A. El-shazli, S. M. Youssef, and A. H. Soliman, "Intelligent computer-aided model for efficient diagnosis of digital breast tomosynthesis 3D imaging using deep learning," *Appl. Sci.*, vol. 12, no. 5736, pp. 1–19, 2022. doi: 10.3390/app12115736

[10] J. Yang, X. Huang, Y. He, J. Xu, C. Yang, and G. Xu, "Reinventing 2D convolutions for 3D images," *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 8, pp. 3009–3018, 2021.

[11] T. Zhang *et al.*, "S3-Slicer : A general slicing framework for multi-axis 3D printing," *ACM Trans. Graph*, vol. 41, no. 6, pp. 1–15, 2022.

[12] C. Li, Y. Mao, Y. Guo, J. Li, and Y. Wang, "Computer methods and programs in biomedicine multi-dimensional cascaded net with uncertain probability reduction for abdominal multi-organ segmentation in CT sequences," *Comput. Methods Programs Biomed.*, vol. 221, pp. 1–14, 2022.

[13] T. Rahman *et al.*, "Transfer learning with deep Convolutional Neural Network (CNN) for pneumonia detection using chest x-ray," *Appl. Sci.*, vol. 10, no. 9, pp. 1–17, 2020.

[14] A. Ullah, N. Javaid, M. Asif, M. U. Javed, and A. S. Yahaya, "AlexNet, adaboost and artificial bee colony based hybrid model for electricity theft detection in smart grids," *IEEE Access*, vol. 10, pp. 18681–18694, 2022.

[15] A. Desiani, Erwin, B. Suprihatin, Y. Wahyudi, E. S. Cahyono, and M. Arhami, "BVU-Net: A u-net modification by vgg-batch normalization for retinal blood vessel segmentation," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 6, pp. 303–314, 2022.

[16] C. C. Nworu, E. J. Ekpenyong, J. Chisimkwuo, and C. N. Onyeukwu, "The effects of modified ReLU activation functions in image classification biomedical engineering and medical devices the effects of modified ReLU activation functions in image classification," *Biomed. Eng. Med. Devices*, vol. 7, no. 9, pp. 1–5, 2022.

[17] S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural Networks*, vol. 107, pp. 3–11, 2018.
[18] G. Ji *et al.*, "Application value of CT three-dimensional reconstruction technology in the identification of benign and malignant lung nodules and the characteristics of nodule distribution," *BMC Med. Imaging*, vol. 25, no. 1, 7, 2025.
[19] M. T. Vervoorn, M. Wulfse, F. A. A. M. Hoesein, M. Stellingwerf, N. P. V. Kaaij, and L. M. Heer, "Application of three-dimensional computed tomography imaging and reconstructive techniques in lung surgery: A mini-review," *Front. Surg.*, vol. 9, 2022.
[20] A. Rachmad, M. Syarief, J. Hutagalung, S. Hernawati, E. Mala, and S. Rochman, "Comparison of CNN architectures for mycobacterium tuberculosis classification in sputum images," *Int. Inf. Eng. Technol. Assoc.*, vol. 29, no. 1, pp. 49–56, 2024.
[21] M. Sahu and R. Dash, "A survey on deep learning : Convolution Neural Network (CNN)," in *Proc. Intelligent and Cloud Computing*, 2020, vol. 2.
[22] B. Asadi and H. Jiang, "On approximation capabilities of ReLU activation and softmax output layer in neural networks," arXiv preprint arXiv:2002.04060, 2020.
[23] T. Meeradevi, S. Sasikala, L. Murali, N. Manikandan, and K. Ramaswamy, "Lung cancer detection with machine learning classifiers with multi-attribute decision-making system and deep learning model," *Sci. Rep.*, vol. 15, no. 8565, pp. 1–19, 2025.
[24] S. V. Militante, N. V. Dionisio, and B. G. Sibbaluca, "Pneumonia detection through adaptive deep learning models of convolutional neural networks," in *Proc. Control and System Graduate Research Colloquium (ICSGRC)*, 2020, pp. 88–93.
[25] H. Jamil, M. Mohamed, and A. Joret, "Chest X-ray pneumonia classification using AlexNet," *Evol. Electr. Electron. Eng.*, vol. 2, no. 2, pp. 191–198, 2021.
[26] Y. Xu, Y. Wang, and N. Razmjooy, "Lung cancer diagnosis in CT images based on alexnet optimized by modified Bowerbird optimization algorithm," *Biomed. Signal Process. Control*, vol. 77, pp. 1–10, 2022.
[27] S. Hasija, P. Akash, M. Bhargav Hemanth, A. Kumar, and S. Sharma, "A novel approach for detection of COVID-19 and pneumonia using only binary classification from chest CT-scans," *Neurosci. Informatics*, vol. 2, no. 4, pp. 1–9, Dec. 2022.
[28] A. Agarwal, K. Patni, and R. D, "Lung cancer detection and classification based on alexnet CNN," in *Proc. 2021 6th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, Jul. 2021, pp. 1390–1397.
[29] K. Sugiura *et al.*, "Big-Volume SliceGAN for improving a synthetic 3D microstructure image of additive-manufactured TYPE 316L steel," *J. Imaging*, vol. 9, no. 5, 2023.
[30] S. Agarwala, A. Kumar, A. K. Dhara, S. B. Thakur, A. Sadhu, and D. Nandi, "Special convolutional neural network for identification and positioning of interstitial lung disease patterns in computed tomography images," *Pattern Recognit. Image Anal.*, vol. 31, no. 4, pp. 730–738, 2021.
[31] A. Z. Putra, D. V. M. Situmorang, J. P. K. Giawa, and R. A. Tarigan, "Pneumonia classification based on lung CT scans using VGG-19," *J. Dan Penelit. Tek. Inform.*, vol. 7, no. 4, pp. 2458–2466, 2023.
[32] M. Afser *et al.*, "Non-small cell lung cancer detection through knowledge distillation approach with teaching assistant," *PLoS One*, vol. 19, no. 11, pp. 1–22, 2024.
[33] M. Irtaza, A. Ali, M. Gulzar, and A. Wali, "Multi-label classification of lung diseases using deep learning," *IEEE Access*, vol. 12, pp. 124062–124080, 2024.
[34] W. T. Kahwachi and K. A. Saed, "A novel architecture of convolutional neural network to diagnose COVID-19 disease," *Math. Stat. Eng. Appl.*, vol. 71, no. 4, pp. 4831–4856, 2022.
[35] M. A. Junior, P. Appiahene, Y. M. Missah, and V. A.-Adjepong, "The use of Efficientnet_b0 network to identify COVID-19 in chest X-ray images," *J. Electr. Syst. Inf. Technol.*, vol. 11, no. 23, pp. 1–15, 20243.
[36] С. П. Морозов *et al.*, "MosMedData : Data set of 1110 chest CT scans performed during the COVID-19 epidemic," *Digit. Diagnostics*, vol. 1, no. 1, pp. 49–59, 2020.
[37] Q. H. Nguyen *et al.*, "Influence of data splitting on performance of machine learning models in prediction of shear strength of soil," *Hindawi*, vol. 26, no. 29, pp. 1–15, 2021.
[38] A. Nazarkar, H. Kuchulakanti, C. S. Paidimarry, and S. Kulkarni, "Impact of various data splitting ratios on the performance of machine learning models in the classification of lung cancer," in *Proc. the Second International Conference on Emerging Trends in Engineering (ICETE 2023)*, 2023, pp. 96–104.
[39] H. R. Maier *et al.*, "On how data are partitioned in model development and evaluation: Confronting the elephant in the room to enhance model generalization," *Environ. Model. Softw.*, vol. 167, pp. 1–8, 2023.
[40] Y. Zhang, Q. Liao, L. Ding, and J. Zhang, "Computerized medical imaging and graphics bridging 2D and 3D segmentation networks for computation-efficient volumetric medical image segmentation: An empirical study of," *Comput. Med.* Imaging *Graph.*, vol. 99, pp. 1–8, 2022.
[41] L. Hussain *et al.*, "Machine-learning classification of texture features of portable chest X-ray accurately classifies COVID-19 lung infection," *Biomed. Eng. Online*, vol. 19, no. 88, pp. 1–18, 2020.
[42] T. Pranata, A. Desiani, B. Suprihatin, H. Hanum, and F. Efriliyanti, "Segmentation of the lungs on X-Ray thorax images with the U-Net CNN architecture," *Comput. Eng. Appl. J.*, vol. 11, no. 2, pp. 101–111, 2022.
[43] A. Ponraj *et al.*, "A multi-patch-based deep learning model with VGG19 for breast cancer classifications in the pathology images," *Digit. Heal.*, vol. 11, 2025.
[44] S. Goyal and R. Singh, "Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 4, pp. 3239–3259, 2023.
[45] D. Yamada, S. Ohde, R. Imai, K. Ikejima, M. Matsusako, and Y. Kurihara, "Visual classification of three computed tomography lung patterns to predict prognosis of COVID-19: A retrospective study," *BMC Pulm. Med.*, vol. 22, no. 1, pp. 1–9, 2022.
[46] E. Diz-mellado, S. Rubino, S. Fern, and G. Macarena, "Applied Machine Learning Algorithms for Courtyards Thermal Patterns Accurate Prediction," *Mathematics*, vol. 9, no. 1142, pp. 1–19, 2021, doi: doi.org/10.3390/math9101142.
[47] S. Ahuja, B. K. Panigrahi, N. Dey, V. Rajinikanth, and T. K. Gandhi, "Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices," *Appl. Intell.*, vol. 51, no. 1, pp. 571–585, 2021.
[48] S. A. Sukanya and K. Kamalanand, "Deep learning-assisted efficient staging of SARS-CoV-2 lesions using lung CT slices," *Math. Probl. Eng.*, vol. 2022, pp. 1–12, 2022.
[49] E. Soares, P. Angelov, and Z. Zhang, "An explainable approach to deep learning from CT-scans for Covid identification," *Evol. Syst.*, pp. 1–8, 2020.
[50] S. Serte and H. Demirel, "Deep learning for diagnosis of COVID-19 using 3D CT scans," *Comput. Biol. Med.*, vol. 132, pp. 1–12, 2021.
[51] E. C. Tetila *et al.*, "System for quantitative diagnosis of COVID-19 associated Pneumonia based on Superpixels with deep learning and chest CT," *Obs. LA Econ. Latinoam.*, vol. 21, no. 9, pp. 10883–10905.