Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 7, No. 3, July 2025, pp: 881-893 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Combination of Gamma Correction and Vision Transformer in Lung Infection Classification on CT-Scan Images

Lucky Indra Kesuma[1], Pipin Octavia[2], Purwita Sari[3], Gracia Mianda Caroline Batubara[4], Karina[4]

[1] Department of Information Technology, Universitas Muhammadiyah Palembang, Palembang, Indonesia
[2] Department of Information Systems, Universitas Sjakhyakirti, Palembang, Indonesia
[3] Department of Information Management, Universitas Sriwijaya, Indralaya, Indonesia
[4] Department of Mathematics, Universitas Sriwijaya, Indralaya, Indonesia

Corresponding author: Lucky Indra Kesuma (email: luckyindra25@gmail.com)

Abstract **Lung infection is an inflammatory condition of the lungs with a high mortality rate. Lung infections can be identified using CT-Scan images, where the affected areas are analyzed to determine the infection type. However, manual interpretation of CT-Scan results by medical specialists is often time-consuming, subjective, and requires a high level of accuracy. To address these challenges, this study proposes an automated classification method for lung infections using deep learning techniques. Convolutional Neural Networks (CNNs) are widely used for image classification tasks. However, CNN operates locally with limited receptive fields, making capturing global patterns in complex lung CT images challenging. CNN also struggles to model long-range pixel dependencies, which is crucial for analyzing visually similar regions in lung CT-Scans. This study uses a Vision Transformer (ViT) to overcome CNN limitations. ViT employs self-attention mechanisms to capture global dependencies across the entire image. The main contribution of this study is the implementation of ViT to enhance classification performance in lung CT-Scan images by capturing complex and global image patterns that CNN fails to model. However, ViT requires a large dataset to perform optimally. To overcome these challenges, augmentation techniques such as flipping, rotation, and gamma correction are applied to increase the amount of data without altering the important features. The dataset comprises lung CT-scan images sourced from Kaggle and is divided into Covid and Non-Covid classes. The proposed method demonstrated excellent classification performance, achieving accuracy, sensitivity, specificity, precision, and F1-Score above 90%. Additionally, the Cohen's kappa coefficient reached 89%. These results show that the proposed method effectively classifies lung infections using CT-Scan images and has strong potential as a clinical decision-support tool, particularly in reducing diagnostic time and improving consistency in medical evaluations.**

Keywords **Lungs; CT-Scan, Classification, Gamma Correction, Vision Transformer**.

I. Introduction

Lung infection refers to an inflammatory condition of the lungs caused by various pathogens, including viruses, bacteria, fungi, and parasites [1][2]. Early detection and diagnosis of lung infections can help reduce the high mortality rate, significantly increasing patient survival rates from 14% to 49% [3]. Lung infections are categorized into two classes: Covid and Non-Covid. The Covid class refers to lung infections caused by the *coronavirus disease* 19 (COVID-19), while the Non-Covid class indicates that the patient's lungs are not infected by the virus [4]. Lung infections are typically analyzed using lung images obtained through Computed Tomography (CT). CT-Scan is a method of visualizing images of human organs using X-rays [5]. These CT scans are then manually interpreted by a pulmonologist [6]. However, manual diagnosis has several disadvantages, such as being time-consuming, multi-interpretative, and requiring a high level of accuracy [7][8]. Classifying lung CT-scan images using deep learning is an effective solution to overcome the limitations of manual diagnosis.

To overcome these limitations, deep learning-based approaches, particularly for image classification, have emerged as effective solutions. One widely adopted deep learning method is the Convolutional Neural Network (CNN), which is capable of extracting local features and gradually building complex representations

Manuscript Received 28 September 2024; Revised 10 March 2025; Accepted 5 July 2025; Available online 9 July 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.588

[9]. Xiaoyi et al. [10] applied MobileNetV2 for lung classification involving three classes and achieved an accuracy of 85%. Toroghi et al. [11] applied VGG16 for lung classification involving three classes and achieved an accuracy of 88%, although additional performance metrics were not evaluated. Ragab et al. [12] used CNNs for a similar classification task and achieved an accuracy of 89%. However, CNNs operate locally due to the use of kernels in a limited area, so it is difficult for CNNs to capture global patterns in an image, especially in complex images. In addition, CNNs process images sequentially, which can result in the loss of relationships between distant pixels. Lung images are complex due to fine anatomical details, which causes the structure of different parts of the lungs to look similar and difficult to distinguish.

Vision Transformer (ViT) is a deep learning method designed to capture global image patterns. In contrast, CNNs operate sequentially based on the kernel size and stride used. ViT operates in parallel by dividing the image into several patches and processing each patch simultaneously using self-attention [13]. ViT architecture comprises multi-head self-attention (MSA) and multi-layer perceptron (MLP) blocks designed to capture global patterns within images. Each pixel in the image is considered to have equal importance, allowing all relationships between pixels to be preserved. Moreover, the complex structures in the image can be more easily recognized globally by ViT [14]. ViT is considered more effective than CNNs for learning from complex image data. Several studies have explored the use of ViT in lung infection classification. Mezina et al. [15] applied ViT for for classifying lung infections into nine classes, and achieved an accuracy of 81.9%. However, the sensitivity and specificity remained below 70%. Ukwuoma et al. [6] applied ViT for for classifying four lung infection classes and achieved an accuracy of 87%. However, the F1 score remained below 75%. Toroghi et al. [11] applied ViT to classify lung infections into three classes, and achieved an accuracy of 83%. However, this study did not measure sensitivity, specificity, and F1-score.

ViT has the disadvantage of requiring significantly more parameters than CNN. The large number of parameters makes ViT less effective on small datasets due to the risk of overfitting [16]. VIT requires a large amount of data. Unfortunately, the availability of medical images, such as lung CT-Scan images is still limited [17]. To address this issue, data augmentation is a commonly used technique for increasing the size of a dataset [18]. Data augmentation is a technique employed to enhance both the quantity and diversity of data by applying modifications to the original images [19]. The most basic augmentation techniques are flipping and rotation [20]. Flipping is performed by reversing the image either horizontally, vertically, or both. In contrast, rotation offers greater variability by allowing images to be transformed through specified angular shifts [21]. Rotation allows the generation of more new images than flipping because the rotation angle can be chosen more variably, thus significantly increasing the diversity of the training data. Flipping and rotation have been widely used to improve classification performance [22]. For example, Teramoto et al. [23] applied flipping and rotation for lung cancer classification using VGG16, involving two classes, and achieved an accuracy of 79.2%. Bushara et al. [24] applied a flipping for lung cancer classification using CNN, involving two classes, and achieved an accuracy of 95%. Yadlapalli et al. [25] applied flipping and rotation for lung cancer classification using DenseNet-169, involving two classes, and achieved an accuracy below 85%. However, these studies employed only basic augmentation techniques, which may limit to generate new data variations. In contrast, rotation enables the generation of more diverse data, as it allows the application of rotation angles ranging from 0° to 360°. However, applying excessively large rotation angles may result in distorted or inaccurate images due to the inclusion of irrelevant features, which can ultimately lead to misclassification [26].

Another technique for increasing the amount of data involves modifying image contrast. Contrast modification is a form of data augmentation, as it generates new images with varying lighting and contrast characteristics [21]. One common method for adjusting contrast is gamma correction [27]. This technique improves image quality by applying non-linear adjustments to pixel values, including exposure adjustments that enhance contrast and detail [28]. Gamma correction is employed both to adjust contrast and to fine-tune the transformation function's intensity [29]. The effect of gamma correction can be observed through the image intensity distribution histogram, which show the distribution of pixel values [30]. By comparing the histograms before and after augmentation, it can be seen that the intensity distribution changes, indicating that a new and different image has been generated [30][31]. Gamma correction has been used in several studies as an augmentation technique. Maiyanti et al. [21] used gamma correction on soil images, while Rahman et al. [32] applied it to chest X-ray images. Sun et al. [33] employed it retinal images. These studies show that gamma correction contributes to improved classification performance across diverse domains.

This study proposes data augmentation techniques and the Vision Transformer architecture for classifying lung infections in lung CT scan images. Augmentation is applied during the preprocessing stage to enhance both

Manuscript Received 28 September 2024; Revised 10 March 2025; Accepted 5 July 2025; Available online 9 July 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.588

the quantity and variability of the dataset. The augmentation techniques used include rotation, flipping, and gamma correction. Rotation is applied randomly within a range of 1°–15°, while flipping consists of both horizontal and vertical flips. Gamma correction is performed using a random gamma value between 0.5 and 2.

Following augmentation, the classification process is conducted using the Vision Transformer architecture. This architecture utilizes Multi-Headed Self-Attention (MSA) and Multi-Layer Perceptron (MLP) blocks for better feature learning. Lung infections are categorized into two classes, which are Covid and Non-Covid. The performance of the Vision Transformer in classifying lung CT-Scan images is evaluated using accuracy, sensitivity, specificity, F1-score, and Cohen's kappa. This study aims to provide a robust and accurate lung infection classification model based on lung CT-Scan images. The proposed system is intended to support the development of automated diagnostic tools in the medical field, assisting medical personnel in early detection, and treatment of lung infections in patients.

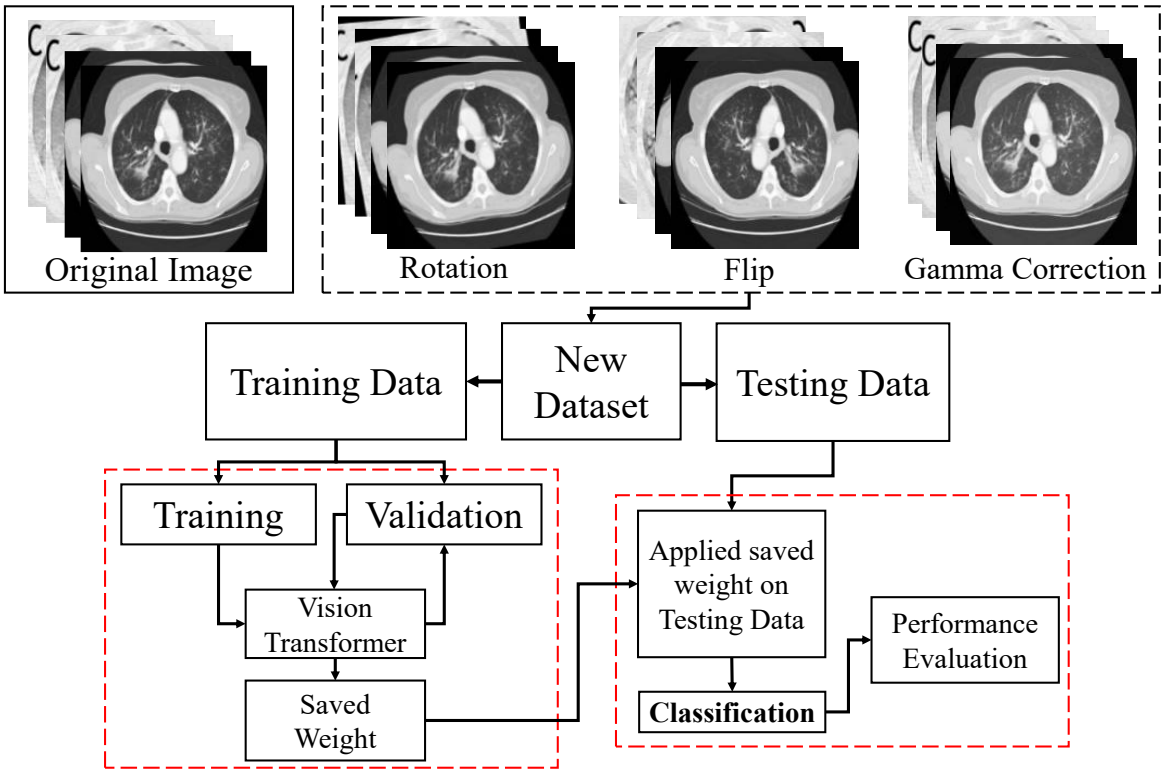
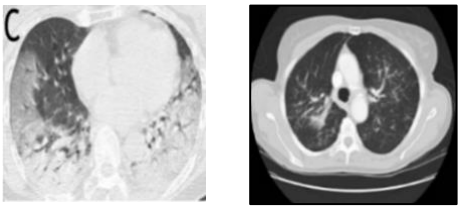


Fig. 1. Proposed method in lung classification on CT-Scan images using Vision Transformer architecture

II. Materials and Methods

This study consists of several stages, including data collection, data augmentation, Covid and Non-Covid classification, and performance evaluation. The data augmentation stage is applied to the dataset using methods, such as flipping, rotation, and gamma correction. In the classification stage, the augmented dataset is used with the Vision Transformer architecture. The classification process with the Vision Transformer architecture involves several steps, including patch embedding, CLS token embedding, transformer encoder, self-attention mechanism, multi-head attention, and multilayer perceptron. This stage includes both training and testing phases. The performance evaluation of the Vision Transformer architecture in classifying Covid and Non-Covid is measured by calculating accuracy, sensitivity, specificity, precision, F1-Score, and Cohen's kappa. An overview of the workflow is illustrated in Fig. 1.



(a) (b)

Fig. 2. Image sample of lung dataset (a)Covid (b)non-Covid

A. Source Data

This study uses a dataset of CT-Scan images of the lungs sourced from Kaggle, accessible at https://www.kaggle.com/code/travishong/covid-19-lung-ctsegmen-

Manuscript Received 28 September 2024; Revised 10 March 2025; Accepted 5 July 2025; Available online 9 July 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.588

tation-classification. This dataset was selected because it contains representative images that highlight key features relevant to distinguishing between Covid and Non-Covid lung infections. The dataset consists of two classes: Covid and Non-Covid. Sample images from datasets is shown in Fig. 2.

B. Augmentation

Data augmentation is employed to enhance the size and variability of the dataset by applying specific transformations to existing images. This process generates new image variations that maintain the essential characteristics of the originals. The augmentation techniques used in this study include rotation, flipping, and gamma correction.

1. Rotation

Rotation is an augmentation technique that adjusts the image by rotating it within a defined angular range. Smaller rotation angles produce minimal changes to the image, while larger angles create more noticeable differences. Empty areas resulting from the rotation are filled with a black background to preserve the original image size. In this study, CT-Scan lung images are randomly rotated between 1° and 15°. This range ensures that the augmented images remain similar to the original and preserve key features of the lung images.

2. Flip

Flipping is an augmentation method that mirrors the original image vertically or horizontally. This technique is used to effectively increase the f dataset size. The reversal will generate an image that differs from the original

3. Gamma Correction

Gamma correction is an image enhancement technique that adjusts the brightness and darkness of an image. Gamma correction uses a non-linear operation to modify the contrast, especially for dark images, by applying a power-law transformation. Increasing the gamma value enhances the image brightness, while decreasing the gamma value reduces the brightness, resulting in a darker image. This study uses a random gamma value between 0.5 and 2. This range ensures the augmented images are neither too bright nor too dark, which could cause important features to be lost during the learning process. The image value after gamma correction can be calculated using Eq. (1) [34].

$$G_{i,j} = 255 \left(\frac{h_{i,j}}{255} \right)^{\gamma} \quad (1)$$

In Eq. (1), $G_{i,j}$ is the gamma corrected pixel result at-(i,j), $h_{i,j}$ is the input image value at pixel (i,j). γ is the gamma value. If γ is greater than 1, then the output result will be dark. Otherwise, if γ is smaller than 1 then the output result will be lighter. An illustration of the rotation and flip technique can be shown in Fig. 3.

Fig. 3(a) and Fig. 3(b) illustrate the image augmentation techniques used in this study. As shown in Fig. 3(a), a single image can be rotated at multiple angles to produce different variations, thereby increasing the model's ability to generalize. Fig. 3(b) demonstrates the flipping technique, where an image is mirrored either horizontally or vertically. The blue rectangles in the figure indicate the axis of reflection to the left or right for horizontal flips and above or below for vertical flips.

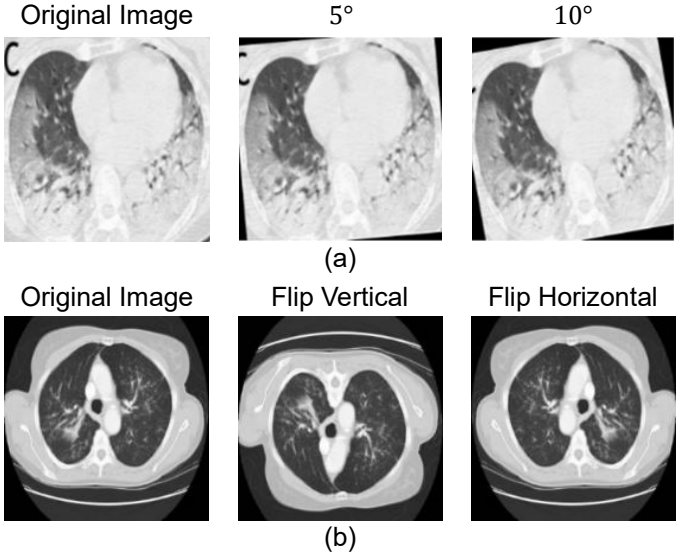


Fig. 3. Illustration of data augmentation (a) rotation technique (b) flip technique

Manuscript Received 28 September 2024; Revised 10 March 2025; Accepted 5 July 2025; Available online 9 July 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.588

C. Vision Transformer

The Vision transformer is one of the most architectures designed for image classification tasks. This architecture consists of several components: patches, CLS embedding, encoder transformer, self-attention, Multi-Head Attention, and Multilayer Perceptron. Fig. 4. illustrates the Vision Transformer process using nine image patches.

Based on Fig. 4, Vision Transformer architecture is broadly divided into three stages. The first stage is patch embedding. The image is split into several equally sized sections, which are then combined to form an embedding sequence. The second stage is the transformer encoder, which is the core of the Vision Transformer architecture. This stage aims to accept the embedding sequence as input and generate the encoding sequence. In this part, the encoder transformer mainly consists of MSA and MLP. The output sequence is transformed back to its original position. The final stage involves the MLP head, which is responsible for producing the final classification predictions [35].

1. Patch Embedding

A patch is a sub-region of the input image obtained by partitioning the image into fixed-size, non-overlapping segments. In the Vision Transformer architecture, the input image is decomposed into these uniform patches, which are subsequently flattened and arranged sequentially to form a one-dimensional input sequence. Eq. (2) is used to divide the image into *n* patches [36].

$$n = \frac{hw}{h_p w_p} \quad (2)$$

Where *n* is the number of patches to be formed, (*h*, *w*) denotes the resolution of the original image, *h* is the image height, *w* is the image width, and $h_p w_p$ represents the resolution of each image patch, assuming a square patch size of *m* x *m* pixels.

2. Position and Class Embedding

After patch embedding, a embedding matrix performs a linear projection of the patches into a vector space compatible with the model. The resulting embedded representations are subsequently combined with a learnable classification token $\boldsymbol{X_{class}}$ which serves as a representative feature for the entire image during classification. The resulting input sequence $\boldsymbol{Z_0}$ can be calculated as shown Eq. (3) [16].

$$Z_0 = [X_{class}; X_1E; X_2E, \dots, X_nE] + E_{pos} \quad (3)$$

where X_{class} is one hot encoding of the class label or a token matrix built by the computer, X_n is the n-th patch in matrix form, E is the embeddings matrix or patch encoding and E_{pos} is the encoding position in the form of an embedding position matrix or position encoding.

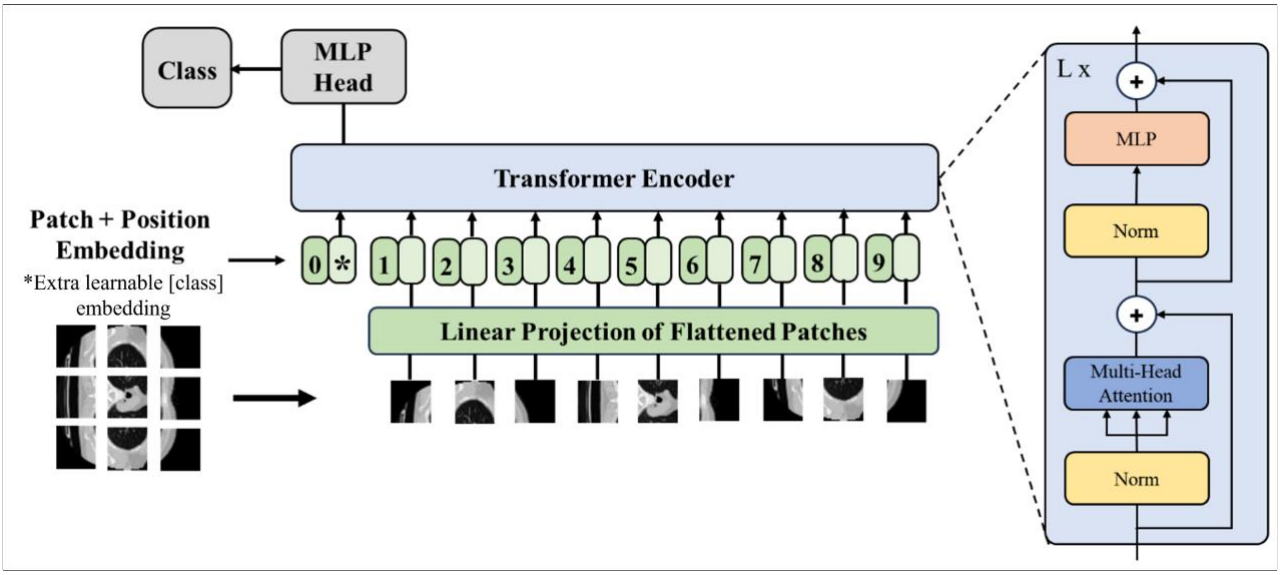


Fig. 4. Vision Transformer Architectures

3. Transformer Encoder

The output from the embedding process serves asthe input in the Transformer Encoder. Transformer Encoder consists of two main connected parts: MSA and MLP. Calculations for both parts can be seen in Eq. (4) and Eq. (5) [16].

$$Z'_\ell = MSA\left(LN(Z_{\ell-1})\right) + Z_{\ell-1}, \ell = 1,2,3, \dots, n \quad (4)$$

$$Z_\ell = MLP\left(LN(Z'_\ell)\right) + Z'_\ell, \ell = 1,2,3, \dots, n \quad (5)$$

where Z_ℓ is the embedding at the *n*-th layer, $Z_{\ell-1}$ is the embedding at the *(l-1)*-th layer, Z'_ℓ is the result of Multi-Head Self-Attention, and LN is Layer Normalization.

4. Self-Attention

Scaled Dot-Product Attention input consists of query (Q), key (K), and value (V). The calculation is performed through matrix multiplication (MatMul) operation between Q and K, followed by dividing the result by the scaling factor $\sqrt{\boldsymbol{d_k}}$ and applying the

Manuscript Received 28 September 2024; Revised 10 March 2025; Accepted 5 July 2025; Available online 9 July 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.588

softmax function. The formulas for calculating the Q, K, and V are provided in Eq. (6), Eq. (7), and Eq. (8), while the output matrix can be seen in Eq. (9) [16].

$$Q = W_q Z \quad (6)$$

$$K = W_k Z \quad (7)$$

$$V = W_v Z \quad (8)$$

$$Attention(Q,K,V) = softmax\left(\frac{1}{\sqrt{d_k}} QK^T\right)V \quad (9)$$

where W_q, W_k, W_v are the linear transformation weights for Q, K, and V, are usually small numbers and randomly initialized using a random distribution.

5. Multi-Head Attention

The Q, K, and V in self-attention are linearly projected h times. At each projection, the attention function is executed in parallel. Multi-head attention enables the model to focus on information from different representation subspaces at different positions simultaneously. The calculation can be performed using Eq. (10) [37].

$$MultiHead(Q,K,V) = Concat\,(head_1, \dots, head_1)W^0 \quad (10)$$

with $head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right)$, Q is query, K is key, V is value, and W^0 is the applied linear projection matrix.

6. Multilayer Perceptron (MLP)

MLP is part of an artificial neural network commonly used to model functional patterns in non-linear systems [38]. MLP consists of three layers: input, hidden, and output layers. Data from the input layer is passed through a perceptron to the next layer, continuing until it reaches the output. The final output is processed using a SoftMax function. The operation of MLP is shown in Eq. (11) [39].

$$Y_k = f\left\{b_k + \sum_{i=1}^{n}(m_{ki}\, x\, Multihead\,(Q,K,V)\right\} \quad (11)$$

where Y_k is the result of the k-th perceptron, d is the activation function, b_k is the perceptron bias.

D. Training and Testing

The dataset is divided into two parts: 80% for training data and 20% for testing data. The training process uses the Vision Transformer architecture to enable the model to learn and recognize feature patterns. The training data is further divided into training and validation data. Initially, the model is first trained on the training data, and then the results are validated using the validation data. The validation process evaluates the model's classification performance. In the training process, the best-performing weights are saved and later used to evaluate the model on the unseen test set.

E. Evaluation

The model's performance is evaluated using a confusion matrix, which quantifies the number of correctly and incorrectly classified samples. The evaluation metrics include accuracy, sensitivity, specificity, F1-Score, Cohen's kappa, and ROC. Accuracy reflects overall correctness, sensitivity and specificity measure the model's ability to predict positives and negatives, precision assesses positive prediction accuracy, Cohen's Kappa corrects for chance agreement, ROC shows class separation ability, and F1-Score balances precision and sensitivity.

III. Result

A. Augmentation

In this stage, the lung images from the CT-Scan image dataset were resized to 224×224 pixels. To increase data variation, data augmentation techniques such as flipping, rotation, and gamma correction methods were applied. The image results obtained from the augmentation process can be seen in Fig. 5. The histogram comparison of the original images and gamma correction images can be seen in Fig. 6.

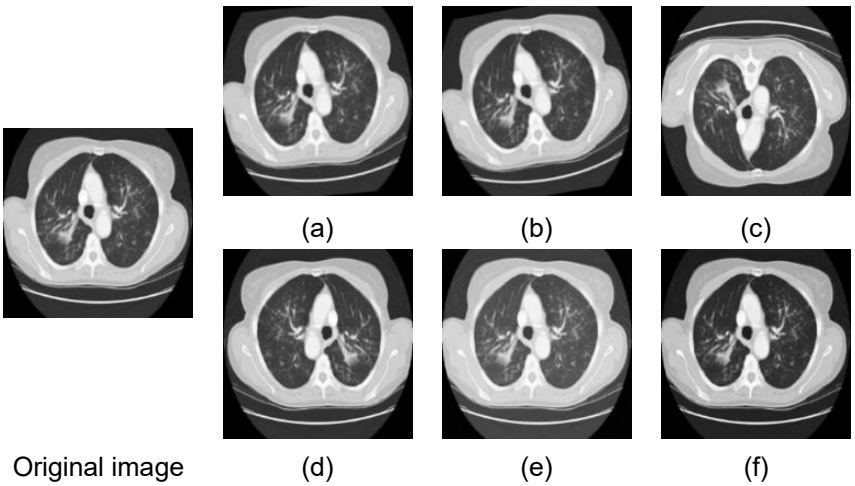


Fig. 5. Illustration of Augmentation Technique, (a) Rotation 5°, (b) Rotation 10°, (c) Vertical Fip, (d) Horizontal Flip, (e) Gamma Correction 0.8, and (f) Gamma Correction 1.2

Manuscript Received 28 September 2024; Revised 10 March 2025; Accepted 5 July 2025; Available online 9 July 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.588

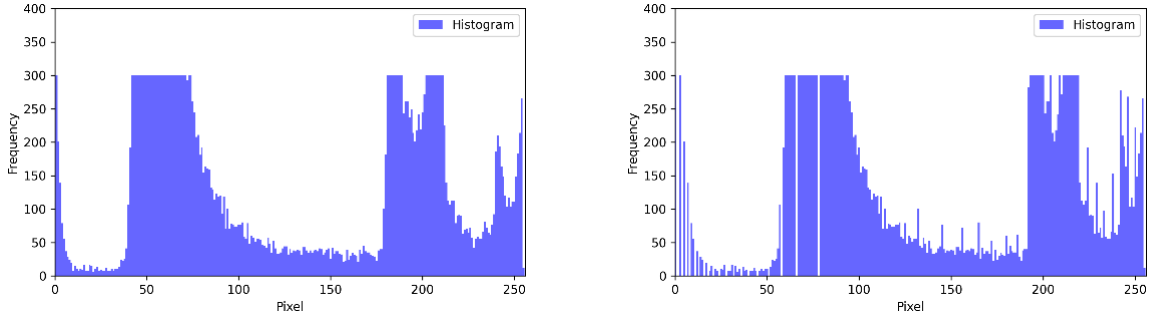
Fig. 5(a) and 5(b) illustrate examples of CT-Scan images rotated by 5° and 10°. Rotation was applied 10 times on each class, resulting in 3,490 images for the COVID class and 3,970 images for the non-Covid class. Fig. 5(c) and 5(d) illustrate examples of original CT-Scan images on each class that were flipped vertically and horizontally, resulting 698 images for the COVID class and 794 images for the non-Covid class. Fig. 5(e) and 5(f) illustrate examples of the original CT-Scan image applied gamma correction with gamma values of 0.8 and 1.2. The gamma correction results in 698 images for the Covid class and 794 for the non-Covid class. Based on the augmentation results, the total CT-Scan images used for the following process are 5,235 for the Covid class and 5,955 for the non-Covid class. This augmentation technique produces images that are different from the original images, thereby increasing the diversity of the data.

Fig. 6 shows the histogram comparison of the original images and gamma correction images. In the original image, the histogram displays an uneven pixel distribution, with concentrations in both low and high-intensity ranges, indicating suboptimal contrast. The histogram becomes more evenly distributed in the gamma-corrected image, especially in the medium to high intensity range. This demonstrates that gamma correction successfully enhances the image contrast and brightness, making previously obscured details and enhancing the overall visual quality.

B. Training

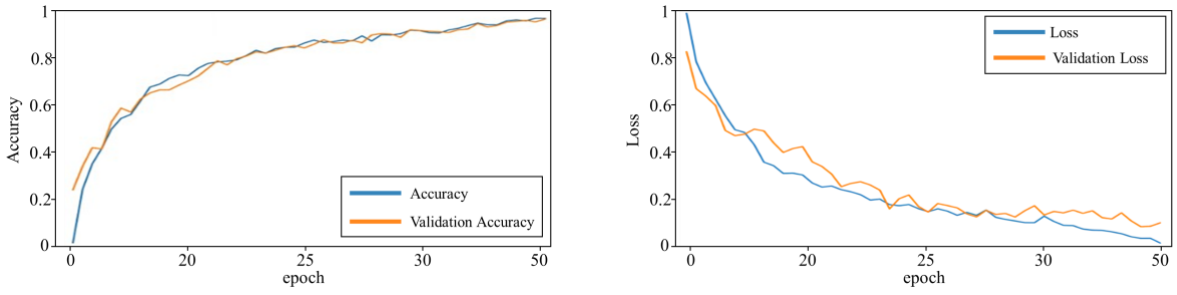
During the training stage, the augmented CT-Scan dataset consisting of 10,920 images was split into 80% training (8,736 images) and 20% testing (2,184 images). The training data was further divided into 75% actual training data (6,552 images) and 25% validation data (2,184 images). The model was trained using a Vision Transformer with 50 epochs and a batch size of 32, resulting in 274 weight updates per epoch. Weight parameters were iteratively updated to minimize prediction error. If validation loss decreases, the corresponding weights are retained; otherwise, updates continue until the optimal weights are achieved. Accuracy and loss curves for both training and validation data are shown in Fig. 7.

Based on Fig. 7(a), the training and validation data loss graph has decreased and increased at several epochs. The training accuracy improved from 0.6376 at the first epoch to 0.9117 at epoch 25, while the validation accuracy increased from 0.6367 to 0.9038 at epoch 24. In the following epochs, both values remained stable above 0.90, indicating that the Vision Transformer model achieved a high level of classification accuracy. Based on Fig. 7(b), the training and validation loss values fluctuated across epochs. The training loss decreased from 0.9765 to 0.0815, while the validation loss decreased from 0.8267 to 0.1596 by the final epoch. Although slight fluctuations were observed, both loss values remained low, indicating that the ViT model has a low prediction error.



(a) (b)

Fig. 6. Histogram Comparison (a) Original Image (b) Gamma Correction Image



(a) (b)

Fig. 7. Results Graph (a) Accuracy and (b) Loss in the Training Stage

Manuscript Received 28 September 2024; Revised 10 March 2025; Accepted 5 July 2025; Available online 9 July 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.588

C. Testing

At this stage, the performance of the Vision Transformer model was measured using the test dataset. The results of the lung infection classification are contained in the confusion matrix, which serves to evaluate the performance of the model. The model performance is calculated, including accuracy, sensitivity, specificity, precision, and F1-score. These metrics are presented in Fig. 8, which illustrates the results for accuracy, sensitivity, and specificity metrics.

Based on Fig. 8, Covid and non-Covid classes achieved a high accuracy of 94.01%, indicating strong reliability in distinguishing between COVID and Non-COVID cases. The sensitivity metric shows a slight variation between the two classes, with the model performing better in detecting the COVID class, 97.10% compared to the non-Covid class, 91.67%. This suggests a potential risk of misclassifying certain non-Covid cases. Conversely, the specificity is higher for the non-Covid class, 97.20%, than for the COVID class, 91.66%, indicating that the model is more effective at correctly identifying non-Covid cases. This difference in specificity reflects the possibility that some COVID cases may be incorrectly classified as non-Covid.

The highest precision of 97.10% is achieved for the COVID class, demonstrating that the model produced few false positives in this class. However, the slightly lower precision for the Non-COVID class indicates occasional misclassification as COVID. The F1 score was 93.33% for COVID and 94.56% for Non-Covid, with slightly better performance predicting non-Covid cases. The model balances excellent classification, minimizing misclassifications. A high F1-score for COVID is useful for avoiding unnecessary interventions. Cohen's Kappa for Non-Covid is 91.07%, indicating little difference in performance, with better results for COVID classification.

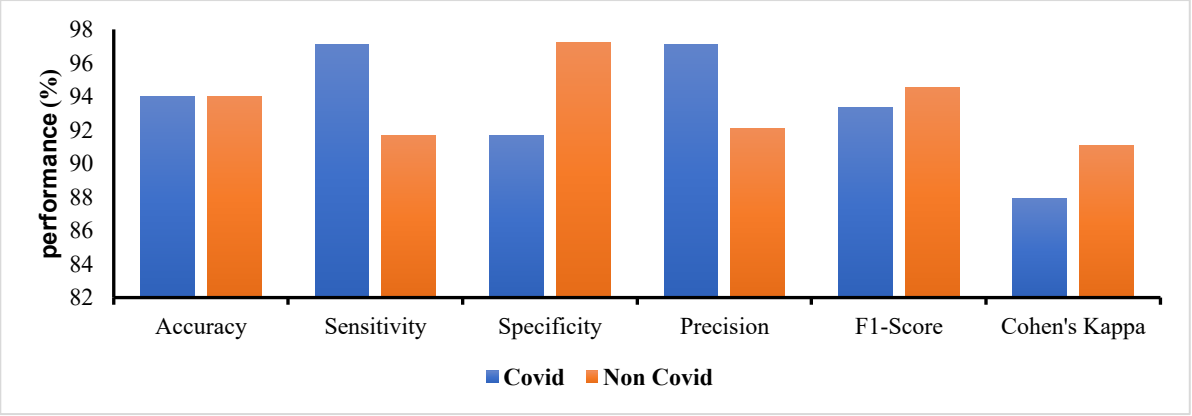


Fig. 8. Model Performance Graph Based

The performance of the vision transformer model can also be seen based on the Receiver Operating Characteristic (ROC) Curve graph, which describes the trade-off relationship between True Positive Rate (TPR) and False Positive Rate (FPR). The proposed model's ROC results are shown in Fig. 9.

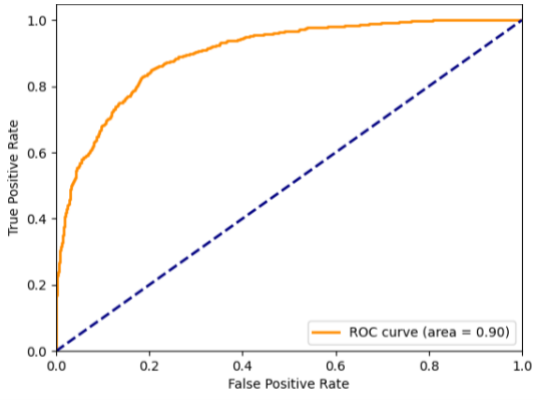


Fig. 9. Receiver Operating Characteristic (ROC) Graphs

Based on Fig. 9, the orange ROC curve approaches the upper left corner, meaning that the model has a high TPR and low FPR. The high TPR value indicates that the vision transformer model can distinguish between the two classes. The Area Under Curve (AUC) value is excellent at 90%. The high AUC value indicates that the vision transformer model can distinguish between each class very well. This suggests that applying data augmentation techniques such as rotation, flipping, and gamma correction can significantly enhance the performance of the Vision Transformer model.

IV. Discussion

In this study, the Vision Transformer architecture is used to classify lung infections based on two classes, namely Covid and non-Covid. The model's performance was then compared with results from several previous studies. This comparative analysis helps determine whether the performance of the Vision Transformer model used is good. Comparison of the classification results of this study with other studies with

Manuscript Received 28 September 2024; Revised 10 March 2025; Accepted 5 July 2025; Available online 9 July 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.588

datasets consisting of Covid and Non-Covid classes can be seen in Table 1 for accuracy, sensitivity, and specificity, and Table 2 for precision, F1-Score, and Cohen's kappa.

Based on Table 1, the proposed method achieved the highest accuracy and sensitivity (the bolded values represent the highest scores). This indicates that the proposed method provides highly accurate predictions that closely correspond to the actual class labels. The high sensitivity value suggests that the proposed method can correctly detect samples that are truly Covid class. The specificity value obtained in the proposed method is still lower than research [39]. However, the specificity value obtained is already very good because it exceeds 90%. This value shows that the proposed method also has an excellent ability to identify non-COVID classes that are truly non-COVID classes.

Based on Table 2, the highest precision, F1-Score, and Cohen's kappa were obtained by the proposed method. Shafi et al. [40] applied M-Segnet and Hybrid SqueezeNet with rotation augmentation, obtained a precision of 88.05% and F1-Score of 88.00%. Meanwhile, Krit et al. [41] applied ResNet50 which also uses rotation only achieved of precision 79.34% and F1-Score of 79.28%. Alshazly et al. [42] applied SqueezeNet with horizontal flip, cropping, and Gaussian noise augmentation, obtained F1-Score of 69.70%. Wang et al. [43] applied M-Inception did not use any augmentation techniques and only obtained Cohen's kappa of 77%. Meanwhile, He et al. [44] applied DenseNet-169 with horizontal flip, cropping, and color jittering, obtained an F1-score of 85%. In comparison, the proposed method with flip, rotation, and gamma correction augmentation obtains a precision of 94.38%, an F1-score of 93.95%, and a Cohen's kappa of 89.49%, showing excellent performance in improving the accuracy and consistency of the results.

Tabel 1. Comparison of Accuracy, Sensitivity, and Specificity with Other Studies

Method	Augmentation Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
M-Segnet and Hybrid Squeezenet [40]	Rotation	93.99	87.96	**96.01**
Resnet50 [41]	Rotation	79.43	79.24	-
SqueezeNet [42]	Horizontal Flip, Cropping, and Gaussian Noise	73.89	62.24	84.76
M-Inception [43]	-	89.5	88	87
DenseNet-169 [44]	Horizontal Flip, Cropping, and Color Jittering	86	-	-
Proposed Method	Flip, Rotation, and Gamma Correction	**94.01**	**94.38**	94.43

Tabel 2. Comparison of Precision, F1-Score, and Cohen's Kappa with Other Studies

Method	Augmentation Method	Precision (%)	F1-Score (%)	Cohen's Kappa (%)
M-Segnet and Hybrid Squeezenet [40]	Rotation	88.05	88.00	-
Resnet50 [41]	Rotation	79.34	79.28	-
SqueezeNet [42]	Horizontal Flip, Cropping, and Gaussian Noise	79.22	69.70	-
M-Inception [43]	-	-	77	69
DenseNet-169 [44]	Horizontal Flip, Cropping, and Color Jittering	-	85	-
Proposed Method	Flip, Rotation, and Gamma Correction	**94.38**	**93.95**	**89.49**

These findings confirm the effectiveness of the proposed method for potential application in the medical field for classifying lung infections. The application of this method can assist medical personnel in improving the speed and accuracy of Covid and non-Covid diagnoses and support better decision-making in patient care. However, collaboration between medical personnel, data availability, security, and technology compatibility with existing systems is essential to apply this method in medical practice. The combination of augmentation and classification techniques proposed in this study shows good performance, although the research is limited to two classes and uses specific augmentation techniques. These limitations can be

Manuscript Received 28 September 2024; Revised 10 March 2025; Accepted 5 July 2025; Available online 9 July 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.588

addressed by modifying or improving the proposed method. Further research on lung infection classification using augmentation techniques and the Vision Transformer architecture is needed to support accurate and reliable diagnosis.

V. Conclusion

This study aimed to classify lung infections using the Vision Transformer (ViT) architecture combined with data augmentation techniques to improve performance on CT-Scan image data. The main finding demonstrates that the proposed model achieved strong classification results, with accuracy, sensitivity, specificity, precision, and F1-score all above 90%, and a Cohen's Kappa score of 89%, indicating high agreement with the ground truth. An additional finding is that the application of augmentation techniques: rotation, flipping, and gamma correction, not only increased the dataset size from 746 to 11,190 images but also enhanced data variability, allowing the model to generalize better and learn distinguishing features between Covid and Non-Covid cases. Further research is necessaryto explore the integration of more diverse datasets from different sources to improve robustness, and to evaluate the model's performance in real-time clinical environments.

References

[1] J. Oliva and O. Terrier, "Viral and bacterial co-infections in the lungs: Dangerous liaisons," *Viruses*, vol. 13, no. 9, 2021, doi: 10.3390/v13091725.

[2] Y. Gao *et al.*, "Size and charge adaptive clustered nanoparticles targeting the biofilm microenvironment for chronic lung infection management," *ACS Nano*, vol. 14, no. 5, pp. 5686–5699, May 2020, doi: 10.1021/acsnano.0c00269.

[3] K. Dimilier, B. Ugur, and Y. K. Ever, "Tumor Detection on CT Lung Images Using Image Enhancement," *J. Sci. Technol.*, vol. 7, no. 1, pp. 133–136, 2017.

[4] G. Muhammad and M. Shamim Hossain, "COVID-19 and Non-COVID-19 Classification using Multi-layers Fusion From Lung Ultrasound Images," *Inf. Fusion*, vol. 72, pp. 80–88, Aug. 2021, doi: 10.1016/j.inffus.2021.02.013.

[5] M. Mazonakis and J. Damilakis, "Computed tomography: What and how does it measure?," *Eur. J. Radiol.*, vol. 85, no. 8, pp. 1499–1504, Aug. 2016, doi: 10.1016/j.ejrad.2016.03.002.

[6] C. C. Ukwuoma *et al.*, "Automated Lung-Related Pneumonia and COVID-19 Detection Based on Novel Feature Extraction Framework and Vision Transformer Approaches Using Chest X-ray Images," *Bioengineering*, vol. 9, no. 11, p. 709, Nov. 2022, doi: 10.3390/bioengineering9110709.

[7] Y. Guo and Y. Peng, "BSCN: Bidirectional Symmetric Cascade Network for retinal vessel segmentation," *BMC Med. Imaging*, vol. 20, no. 1, p. 20, Dec. 2020, doi: 10.1186/s12880-020-0412-7.

[8] X. Hua *et al.*, "WSC-Trans: A 3D network model for automatic multi-structural segmentation of temporal bone CT," *arXiv Prepr. arXiv2211.07143*, 2022.

[9] G. I. Okolo, S. Katsigiannis, and N. Ramzan, "IEViT: An enhanced vision transformer architecture for chest X-ray image classification," *Comput. Methods Programs Biomed.*, vol. 226, p. 107141, 2022, doi: 10.1016/j.cmpb.2022.107141.

[10] X. Liu, Z. Yu, and L. Tan, "Deep Learning for Lung Disease Classification Using Transfer Learning and a Customized CNN Architecture with Attention," in *2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE)*, Aug. 2024, pp. 341–346. doi: 10.1109/ICSECE61636.2024.10729291.

[11] M. N. Toroghi, U. U. Sheikh, and S. S. Irani, "Classification of COVID-19 and Lung Opacity using Vision Transformer on Chest X-ray Images," *J. Phys. Conf. Ser.*, vol. 2622, no. 1, pp. 1–8, 2023, doi: 10.1088/1742-6596/2622/1/012016.

[12] M. Ragab, S. Alshehri, N. A. Alhakamy, R. F. Mansour, and D. Koundal, "Multiclass Classification of Chest X-Ray Images for the Prediction of COVID-19 Using Capsule Network," *Comput. Intell. Neurosci.*, vol. 2022, no. 6185013., pp. 1–8, May 2022, doi: 10.1155/2022/6185013.

[13] Y. Bazi, L. Bashmal, M. M. Al Rahhal, R. Al Dayil, and N. Al Ajlan, "Vision transformers for remote sensing image classification," *Remote Sens.*, vol. 13, no. 3, pp. 1–20, 2021, doi: 10.3390/rs13030516.

[14] C. F. Chen, Q. Fan, and R. Panda, "CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 347–356, 2021, doi: 10.1109/ICCV48922.2021.00041.

[15] A. Mezina and R. Burget, "Detection of Post-COVID-19-Related Pulmonary Diseases in X-ray Images using Vision Transformer-based Neural Network," *Biomed. Signal Process. Control*, vol. 87, p. 105380, Jan. 2024, doi: 10.1016/j.bspc.2023.105380.

[16] A. Souid, N. Sakli, and H. Sakli, "Classification and Predictions of Lung Diseases from Chest X-Rays Using MobileNet V2," *Appl. Sci.*, vol. 11, no. 6, pp. 1–16, 2021, doi: 10.3390/app11062751.

[17] X. Zhai *et al.*, "An Image Is Worth 16 X 16

Manuscript Received 28 September 2024; Revised 10 March 2025; Accepted 5 July 2025; Available online 9 July 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.588

Words :," *Int. Conf. Learn. Represent.*, 2021.

[18] A. Desiani, M. Erwin, B. Suprihatin, S. Yahdin, A. I. Putri, and F. R. Husein, "Bi-Path Architecture of CNN Segmentation and Classification Method for Cervical Cancer Disorders Based on Pap-smear Images," *Int. J. Comput. Sci.*, vol. 48, no. 3, pp. 1–9, 2021.

[19] Erwin, A. Safmi, A. Desiani, B. Suprihatin, and Fathoni, "The Augmentation Data of Retina Image for Blood Vessel Segmentation Using U-Net Convolutional Neural Network Method," *Int. J. Comput. Intell. Appl.*, vol. 21, no. 01, p. 2250004, 2022, doi: 10.1142/S1469026822500043.

[20] C. Shorten and T. M. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[21] S. I. Maiyanti *et al.*, "Rotation-Gamma Correction Augmentation on CNN-Dense Block for Soil Image Classification," *Appl. Comput. Sci.*, vol. 19, no. 3, pp. 96–115, 2023, doi: 10.35784/acs-2023-27.

[22] X. Liu, G. Karagoz, and N. Meratnia, "Analyzing the Impact of Data Augmentation on the Explainability of Deep Learning-Based Medical Image Classification," *Mach. Learn. Knowl. Extr.*, vol. 7, no. 1, pp. 1–28, 2025, doi: 10.3390/make7010001.

[23] A. Teramoto *et al.*, "Automated classification of benign and malignant cells from lung cytological images using deep convolutional neural network," *Informatics Med. Unlocked*, vol. 16, p. 100205, 2019, doi: 10.1016/j.imu.2019.100205.

[24] B. A. R. and V. K. R. S., "Deep Learning-based Lung Cancer Classification of CT Images using Augmented Convolutional Neural Networks," *ELCVIA Electron. Lett. Comput. Vis. Image Anal.*, vol. 21, no. 1, Sep. 2022, doi: 10.5565/rev/elcvia.1490.

[25] P. Yadlapalli, D. Bhavana, and S. Gunnam, "Intelligent classification of lung malignancies using deep learning techniques," *Int. J. Intell. Comput. Cybern.*, vol. 15, no. 3, pp. 345–362, Jul. 2022, doi: 10.1108/IJICC-07-2021-0147.

[26] K. Alomar, H. I. Aysel, and X. Cai, "Data Augmentation in Classification and Segmentation: A Survey and New Strategies," *J. Imaging*, vol. 9, no. 2, p. 46, Feb. 2023, doi: 10.3390/jimaging9020046.

[27] P. Thanapol, K. Lavangnananda, P. Bouvry, F. Pinel, and F. Leprévost, "Reducing Overfitting and Improving Generalization in Training Convolutional Neural Network (CNN) under Limited Sample Sizes in Image Recognition," in *2020 - 5th International Conference on Information Technology (InCIT)*, 2020, pp. 300–305. doi: 10.1109/InCIT50588.2020.9310787.

[28] M. A. Khan *et al.*, "Lungs cancer classification from CT images: An integrated design of contrast based classical features fusion and selection," *Pattern Recognit. Lett.*, vol. 129, pp. 77–85, 2020, doi: 10.1016/j.patrec.2019.11.014.

[29] A. Desiani, Erwin, B. Suprihatin, M. Adrezo, and A. M. Alfan, "A Hybrid System for Enhancement Retinal Image Reduction," in *Proceedings - 3rd International Conference on Informatics, Multimedia, Cyber, and Information System, ICIMCIS 2021*, 2021, pp. 80–85. doi: 10.1109/ICIMCIS53775.2021.9699259.

[30] W. Setiawan, M. M. Suhadi, Husni, and Y. D. Pramudita, "Histopathology of Lung Cancer Classification Using Convolutional Neural Network With Gamma Correction," *Commun. Math. Biol. Neurosci.*, vol. 2022, pp. 1–17, 2022, doi: 10.28919/cmbn/7611.

[31] S.-E. Weng, S.-G. Miaou, and R. Christanto, "A Lightweight Low-Light Image Enhancement Network via Channel Prior and Gamma Correction," pp. 1–23, 2024.

[32] T. Rahman *et al.*, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Comput. Biol. Med.*, vol. 132, p. 104319, 2021, doi: https://doi.org/10.1016/j.compbiomed.2021.104319.

[33] X. Sun *et al.*, "Robust Retinal Vessel Segmentation from a Data Augmentation Perspective," in *Ophthalmic Medical Image Analysis*, 2021, pp. 189–198.

[34] Erwin, H. K. Putra, B. Suprihatin, and F. Ramadhini, "A Hybrid CLAHE-GAMMA Adjustment and Densely Connected U-NET for Retinal Blood Vessel Segmentation using Augmentation Data," *Eng. Lett.*, vol. 30, no. 2, pp. 485–493, 2022.

[35] X. Zhu, Y. Jia, S. Jian, L. Gu, and Z. Pu, "ViTT: Vision Transformer Tracker," *Sensors*, vol. 21, no. 16. 2021. doi: 10.3390/s21165608.

[36] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," 2020.

[37] A. Vaswani *et al.*, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, 2017, vol. 30.

[38] J. C. Ye, "Artificial Neural Networks and Backpropagation," *Math. Ind.*, vol. 37, no. August, pp. 91–112, 2022, doi: 10.1007/978-981-16-6046-7_6.

[39] Y. Chen, X. Gu, Z. Liu, and J. Liang, "A Fast Inference Vision Transformer for Automatic

Manuscript Received 28 September 2024; Revised 10 March 2025; Accepted 5 July 2025; Available online 9 July 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.588

Pavement Image Classification and Its Visual Interpretation Method," *Remote Sens.*, vol. 14, no. 8, pp. 1–20, 2022, doi: 10.3390/rs14081877.

[40] S. M. Shafi and S. K. Chinnappan, *Segmenting and classifying lung diseases with M-Segnet and Hybrid Squeezenet-CNN architecture on CT images*, vol. 19, no. 5 May. 2024. doi: 10.1371/journal.pone.0302507.

[41] K. Sriporn, C.-F. Tsai, C.-E. Tsai, and P. Wang, "Healthcare Analyzing Lung Disease Using Highly Effective Deep Teqhniques," *MDPI Healthc.*, vol. 8, no. 107, pp. 1–21, 2020.

[42] H. Alshazly, C. Linse, M. Abdalla, E. Barth, and T. Martinetz, "COVID-Nets: Deep CNN Architectures for Detecting COVID-19 using Chest CT Scans," *PeerJ Comput. Sci.*, vol. 7, pp. 1–40, 2021, doi: 10.7717/peerj-cs.655.

[43] S. Wang *et al.*, "A Deep Learning Algorithm Using CT Images to Screen for Corona Virus Disease (COVID-19)," *IMAGING INFORMATICS Artif. Intell. A*, vol. 31, pp. 6096–6104, 2021.

[44] X. He *et al.*, "Sample-Efficient Deep Learning for COVID-19 Diagnosis Based on CT Scans," *IEEE Trans. Med. Imaging*, pp. 1–10, 2020.

Authors' Biography

Lucky Indra Kesuma was born in Palembang on September 25, 1990. He earned his associate degree in Computerized Accounting in 2011 at Sriwijaya University. He then pursued his bachelor's degree in information systems at the same university, graduating in 2014. Subsequently, he obtained a master's degree in computer science, specializing in Informatics Engineering, completing his studies in 2017 at Bina Darma University. In 2023, he finished his doctoral studies (Ph.D.) in Engineering Science (Informatics). The author currently serves as a permanent lecturer in the Department of Information Technology, Universitas Muhammadiyah Palembang; Information Systems Lecture program at the Faculty of Computer Science, Sjakhyakirti University, Palembang; and is actively engaged in the Tri Dharma of Higher Education, encompassing education, research, and community service.

Pipin Octavia was born in Palembang on October, 12th 1990. She graduated from the information systems study program at Sriwijaya University in 2013 and earned her master's degree in informatics engineering in 2018. Since 2021, she has been working as a lecturer in the informatics program at Sjakhyakirti University. In addition to her teaching role, she serves as an editor for the research journals published by Politeknik Darussalam and the Faculty of Computer Science, Sjakhyakirti University, Palembang, Indonesia. She has authored fifteen academic articles between 2022 and 2024. Her research interest lie primarily in the field of image processing. In recognition of her research contributions, she received a Beginner Lecturer Research Grant in 2023 and 2024.

Purwita Sari was born in Palembang on June 9, 1992. She is a permanent lecturer in the Informatics Management Study Program, Faculty of Computer Science, Sriwijaya University. She completed her undergraduate studies in the Department of Information Systems at Sriwijaya University and earned her master's degree in the Department of Informatics Engineering at Bina Darma University. The author pursues research in Information Systems, focusing on Databases, Information Systems Analysis and Design, and Data Structures. She actively performs the primary duties of the Tri Dharma of Higher Education, which include Teaching, Research, and Community Service. Her dedication to these areas helps advance the academic community and contribute to societal development.

Gracia Mianda Caroline Batubara was born in Palembang in 2002. She graduated from Mathematics, Sriwijaya University in 2024. In 2023, she served as a Python Training Assistant at the Computational Laboratory Faculty of Mathematics and Natural Science Sriwijaya University. Her interests include various areas of the Computational Mathematics field, such as Machine Learning, Image Processing, Artificial Intelligence, Computer Programming, Data Mining, and

Manuscript Received 28 September 2024; Revised 10 March 2025; Accepted 5 July 2025; Available online 9 July 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.588

Databases. She actively participated in student activities at Sriwijaya University, involving learning, research, organization, and community service. She has earned several medals and achievements from national and international competitions. In 2023, she authored article on the application of Artificial Intelligence, and in 2024 she was awarded an incentive grant under the Program Kreativitas Mahasiswa (PKM).

Karina was born in OKI on November 21, 2001. She graduated from the Mathematics Study Program at Sriwijaya University in 2024. Her interests include Computational Mathematics, particularly in Machine Learning, Image Processing, Artificial Intelligence, Computer Programming, and Database. In 2022, she joined the Startup Campus' Certified Internship & Independent Study (MSIB) under the Data Science pathway program. In 2023, she authored an article on the application of Artificial Intelligence using K-NN and C.45 algorithms. She has been actively engaged in various student activities at Sriwijaya University, including academic, research, organizational, and community service. In 2024, she developed a Sign Language Translator Application, which was implemented as part of a community service project in Palembang. She has also earned several medals and achievements in national and international competitions, reflecting her dedication to academic and social impact.

Manuscript Received 28 September 2024; Revised 10 March 2025; Accepted 5 July 2025; Available online 9 July 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.588