International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Classification of Pulmonary Disease (Tuberculosis and Pneumonia) from Chest X-Rays Using Gray Level Co-Occurrence Matrix Feature Extraction and EfficientNet

Eka Mala Sari Rochman[1*], Rika Yunitarini[1], Fifin Ayu Mufarroha[1], Miswanto[2], Yuli Panca Asmara[3], Aeri Rachmad[4]

[1] Department of Informatics, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan 69162, Indonesia
[2] Department of Mathematics, Faculty of Science and Technology, University of Airlangga, Surabaya 60115, Indonesia
[3] Faculty of Engineering and Quantity Surveying, INTI International University, Nilai 71800, Malaysia
[4] Department of Information System, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan 69162, Indonesia

Corresponding Author Email: em_sari@trunojoyo.ac.id

ABSTRACT

Pulmonary diseases such as tuberculosis and pneumonia remain serious health concerns that require accurate and early diagnosis. Chest X-ray imaging combined with deep learning methods offers a promising solution for disease classification. This study proposes a combination of texture feature extraction via Gray Level Co-occurrence Matrix (GLCM) and EfficientNet architecture to enhance the performance of Convolutional Neural Networks (CNNs) in lung disease identification. The models were trained on a merged dataset of chest X-ray images using the Adam optimizer, with a learning rate of 0.00001 and 50 epochs. Experimental results demonstrate that the EfficientNetV2-S architecture achieves the optimal balance with an accuracy of 96.67% and a training time of 27 minutes 59 seconds. Meanwhile, EfficientNetV2-M also achieved 96.67% accuracy but required a longer training duration of 47 minutes and 54 seconds. The application of GLCM was shown to improve the quality of input features across all tested models, thereby enhancing the representation of image textures. These findings highlight the potential of the GLCM-EfficientNet approach for supporting the reliable and efficient early detection of pulmonary diseases in clinical practice.

1. INTRODUCTION

Image classification is one of the primary tasks in image processing, aiming to categorize images based on specific visual characteristics [1]. This process plays a critical role in numerous fields, including the medical sector, where medical images are utilized for accurate and efficient disease diagnosis. In the context of pulmonary disease detection, chest X-ray classification helps differentiate between healthy lungs and those affected by conditions such as tuberculosis and pneumonia.

Abnormalities in lung tissue can be detected early through chest X-ray examinations. These images are typically captured by radiologists, and the final diagnosis is made by specialists through manual assessment [2]. However, due to the similarity of radiographic patterns between pneumonia and tuberculosis, the diagnostic process is often difficult and prone to errors, which can potentially lead to fatal complications. Therefore, image processing techniques can support radiologists by revealing fine details that might otherwise be missed [3].

In recent years, deep learning-based classification techniques have demonstrated high accuracy in medical imaging tasks. Deep learning, particularly Convolutional Neural Networks (CNNs), has emerged as a powerful tool in medical image analysis [4]. CNNs autonomously learn to extract and prioritize relevant features from images through layers such as convolution, pooling, and fully connected layers.

Several studies have shown that the EfficientNetV2 architecture outperforms traditional CNNs in terms of accuracy and efficiency. For example, research [5] on acute lymphoblastic leukemia classification using EfficientNetV2-S and EfficientNetB3 achieved accuracies of 99.73% and 99.25%, respectively. Another study [6] applied EfficientNetV2-B0 for cervical cancer cell prediction, showing the fastest computation time (198 seconds) and a high classification accuracy of 99.4%. In the classification of respiratory diseases using chest X-ray images, EfficientNetV2-B0 achieved a sensitivity of 98.66%, specificity of 99.51%, and accuracy of 99.4% [7]. Although modern CNN architectures such as EfficientNetV2 have demonstrated high accuracy, these approaches are still limited in capturing fine texture details between visually similar lung diseases, such as tuberculosis and pneumonia.

In addition to model architecture, feature extraction is crucial in image classification. One of the most effective feature extraction methods is the Gray Level Co-occurrence Matrix (GLCM), which captures texture patterns based on the spatial relationship between pixels [8]. Previous studies have demonstrated that GLCM features such as contrast, homogeneity, energy, and correlation significantly enhance classification performance. Research [9] using GLCM and

Histogram of Oriented Gradients (HOGs) features for classifying four types of skin lesions showed higher accuracy with GLCM (93.5%) compared to HOG (78.1%). Another study [10] combined CNN with GLCM features for brain tumor classification, achieving 99.5% accuracy—surpassing traditional models like AlexNet, VGG19, and GoogLeNet.

These research results demonstrate that GLCM can enrich image texture representation and improve medical classification accuracy. Therefore, in the context of pulmonary diseases such as tuberculosis and pneumonia, which have visual similarities on chest X-ray images, the application of GLCM is highly relevant because it can help capture subtle texture differences that are difficult to detect with conventional methods.

Tuberculosis, caused by Mycobacterium tuberculosis, primarily affects the lungs and is transmitted through airborne droplets [11]. This disease causes inflammation and fluid buildup in the alveoli, which can reduce lung function and lead to death if left untreated [12]. Pneumonia is an acute respiratory infection caused by bacteria, viruses, or fungi, which fills the alveoli with pus or fluid, impairing oxygen absorption [13, 14].

Both remain global health problems. Tuberculosis caused 1.3 million deaths in 2022, ranking second among infectious disease deaths worldwide [11]. In Indonesia, tuberculosis cases continue to rise, reaching 821,200 cases in 2023, placing Indonesia as the country with the second-highest tuberculosis burden after India [15]. Pneumonia is also a leading cause of death in children under five, accounting for approximately 14% of all global infant deaths [13].

Building on previous research, this study contributes by combining GLCM feature extraction and CNN feature extraction architectures, namely EfficientNet-B0, EfficientNetv2-B0, EfficientNetV3-S, and EfficientNetv2-M, for tuberculosis and pneumonia classification from chest X-ray images. The use of several EfficientNet architecture variants aims to evaluate the performance of each model in combining GLCM feature extraction, thus identifying the optimal architecture. This approach is designed to overcome the limitations of conventional methods, which often struggle to capture subtle texture differences in lung images. It is hoped that it will provide a more accurate and computationally efficient solution to support lung disease diagnosis.

2. PRELIMINARIES

This study focuses on the classification of pulmonary diseases—specifically tuberculosis and pneumonia—through chest X-ray image analysis using deep learning techniques. The approach involves a combination of feature extraction using GLCM and classification using CNNs with the EfficientNetV2 architecture. The evaluation of the model is conducted using a confusion matrix, accuracy, precision, recall, and F1-score metrics.

2.1 CNN

CNN is a deep learning architecture designed to process visual data, especially images. It works by learning hierarchical features through multiple layers that are optimized during training. CNN was first introduced by Yann LeCun through the LeNet-5 architecture, which laid the groundwork for modern deep learning applications in image recognition [16].
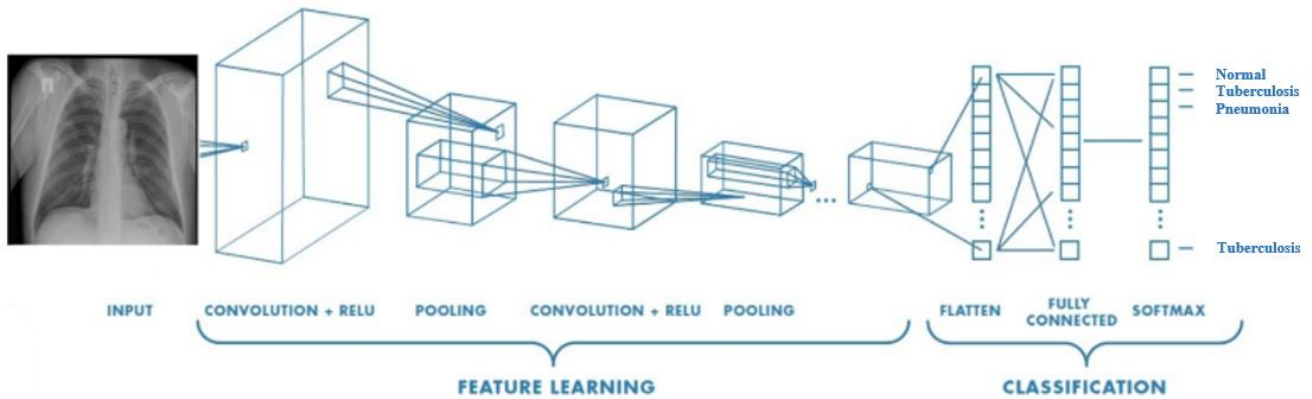

Figure 1. Architecture of the CNN

The CNN architecture comprises three main layers, as shown in Figure 1:
- Convolutional layer: Responsible for extracting spatial features from the input image using filters or kernels. The operation is defined as:

$$Sq(g,h) = \left(\sum_{p=1}^{P} \sum_{u=1}^{U} \sum_{v=1}^{U} I_p(g+u, h+v) \cdot K_{pq}(u,v) \right) + b_q \quad (1)$$

where, I_p is the input pixel, K_{pq} is the kernel matrix, and b_q is the bias [11].

The size of the resulting feature map is calculated as:

$$feature\ map = \frac{C - F + 2P}{s} + 1 \quad (2)$$

where, C is the input dimension, F is the filter size, P is padding, and SSS is the stride [17].
- Activation layer Rectified Linear Unit (ReLU): Applies the ReLU function to add non-linearity:

$$f(x) = max(0, x) \quad (3)$$

- Pooling layer: Reduces the spatial size of the feature maps,

typically using max pooling or average pooling. Global Average Pooling (GAP) is often used in modern CNNs to avoid overfitting and reduce model complexity by averaging the values in the feature map.

- Fully connected layer: After flattening the feature maps into a 1D vector, a series of dense layers is used to produce the final classification. The hidden layer operation is:

$$y_j = \sigma_{yj}\left(\sum_{i=1}^{n} x_i w_{ij} + b_j\right) \quad (4)$$

and the final output before activation is:

$$net_k = \sum_{j=1}^{m} y_j w_{jk} + b_k \quad (5)$$

The Softmax function is applied at the output layer to generate class probabilities:

$$y_i = \frac{e^{zi}}{\sum_j e^{zj}} \quad (6)$$

The loss function used is sparse categorical cross-entropy:

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N} \log(p(y_i)) \quad (7)$$

2.2 EfficientNetV2 architecture

EfficientNetV2, introduced in 2021, is a convolutional network that offers faster training and improved parameter efficiency compared to its predecessor, EfficientNet. It incorporates MBConv and Fused-MBConv blocks to optimize both speed and accuracy, as shown in Figure 2 [18].
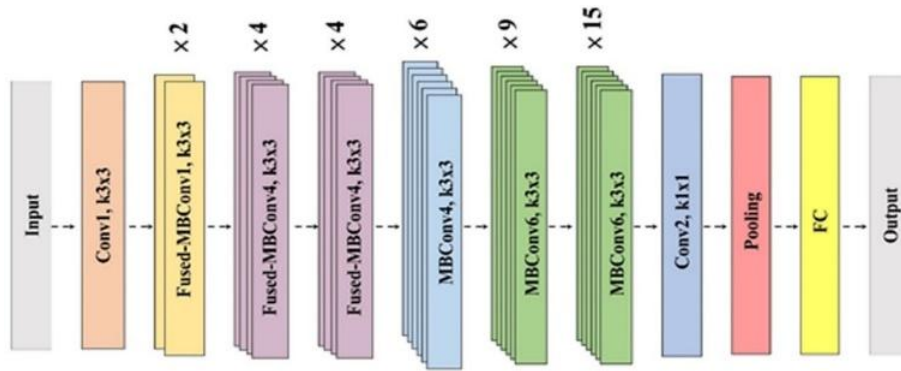

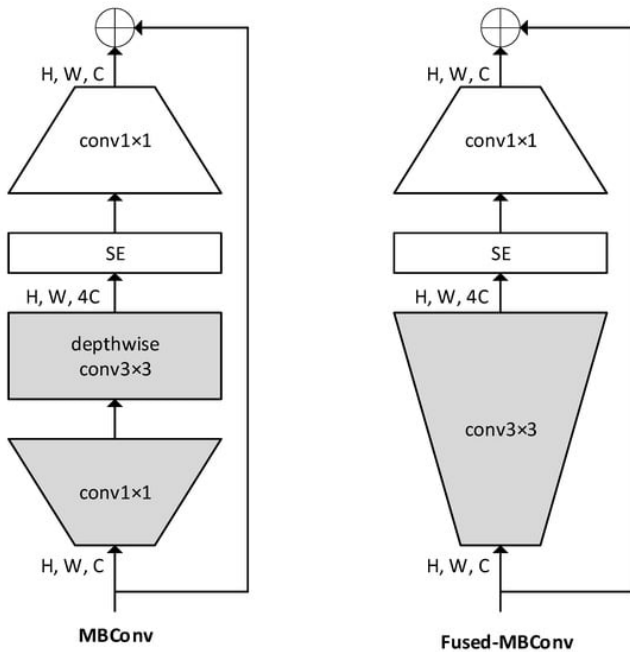**Figure 2.** EfficientNetV2 architecture [19]

Figure 3. Structure of MBConv and fused-MBConv [3]

Figure 3 explains the core components of EfficientNetV2, including:

- **Mobile inverted residual bottleneck convolution (MBConv)**: Used for its efficiency in mobile and embedded environments.

- **Fused-MBConv**: A faster variant introduced for early layers with high resolution.
- **Squeeze-and-excitation (SE) blocks**: Enhance channel interdependencies by recalibrating channel-wise feature responses.

The SE block operates in three stages:
1. Squeeze: Applies global average pooling to condense feature maps into a single channel descriptor.
2. Excitation: Passes the squeezed descriptor through fully connected layers and activation functions to generate weights.
3. Scale: Multiplies the original feature map channels by the generated weights.

EfficientNetV2 also introduces compound scaling, which jointly optimizes depth, width, and resolution. On the ImageNet dataset, EfficientNetV2-S achieves high accuracy with only 24 M parameters and 8.8B FLOPs [18].

2.3 GLCM

GLCM is a statistical method used to analyze texture by considering the spatial relationship between pixel pairs. It quantifies how often a pixel with a specific intensity value occurs in relation to another pixel at a specific distance and orientation. GLCM calculates features such as contrast, correlation, energy, and homogeneity [8].

In the field of medical imaging, GLCM is widely applied due to its ability to capture more detailed texture patterns [9]. This is very important for detecting lung diseases such as

tuberculosis and pneumonia, which often display infiltration or fine patches in lung tissue. These texture patterns are usually difficult to distinguish visually or by conventional CNN methods without additional feature extraction. Therefore, the integration of GLCM can enrich image representation with texture information, thereby improving the accuracy of the model in distinguishing lung conditions that have high visual similarity.

2.4 Confusion matrix and evaluation metrics

To evaluate the classification model, a confusion matrix is used. It visualizes the performance of a classification algorithm by showing the counts of correct and incorrect predictions categorized as shown in Figure 4.
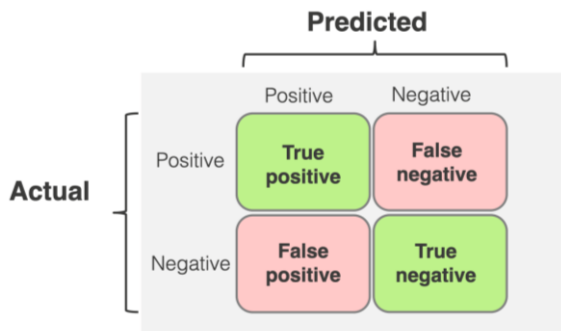

Figure 4. Confusion matrix

From these, the evaluation metrics are derived [20-22]:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \times 100 \qquad (8)$$

$$Precision = \frac{TP}{TP + FP} \times 100 \qquad (9)$$

$$Recall = \frac{TP}{TP + FN} \times 100 \qquad (10)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (11)$$
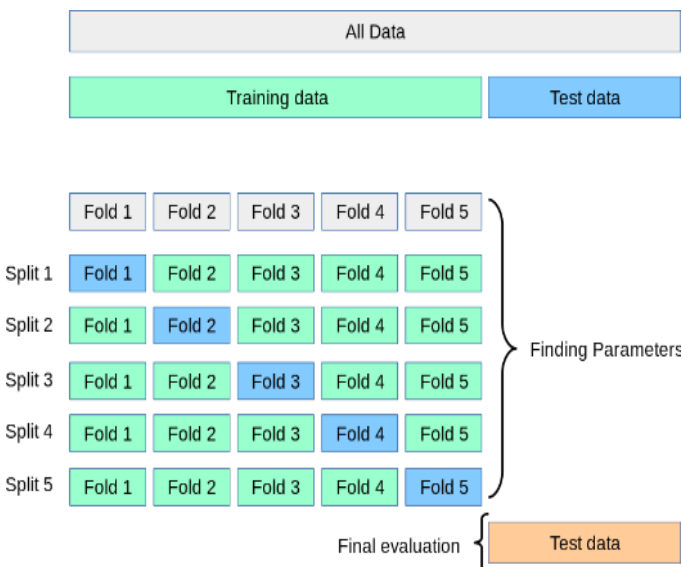

Figure 5. 5-fold cross-validation

2.5 K-fold cross-validation

This study employs 5-fold cross-validation as shown in Figure 5 to ensure robust and generalized model performance. The dataset is divided into five equal parts. In each iteration, four parts are used for training and one for validation. This process is repeated five times, so each subset is used exactly once for validation, and the final performance is the average across all iterations [23].
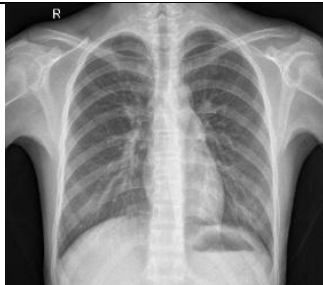
3. MAIN RESULTS

In general, the implementation of this research uses hardware in the form of a laptop with Windows 11 Pro as the operating system. The detailed specifications of the laptop include Intel® Core™ i7-106110U CPU @ 1.80GHz (8 CPUs), 16 GB RAM, and GPU NVIDIA L4 with 23 GB memory. The software used in this study is Python 3.12.8 with Google Colab as the development environment.

3.1 Data gathering

The dataset used in this study consists of digital chest X-ray images obtained from Kaggle. The tuberculosis dataset can be accessed via https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset, and the pneumonia dataset via https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia. Both datasets were combined and restructured to create three classification categories: normal, tuberculosis, and pneumonia, as shown in Table 1.

Table 1. Description of image classes in the dataset

Dataset	Description
	Chest X-ray images with clear lung structure and no abnormalities.
	Chest X-ray images showing mid-lower lung infiltrates and unclear textures.
	Chest X-ray images with widespread white patches indicating infection.

Total image distribution:
- Normal: 5037 images
- Tuberculosis: 700 images
- Pneumonia: 4283 images

To enhance the tuberculosis dataset, data augmentation techniques were applied, including zoom, rotation, horizontal and vertical shifting, and shearing. Table 2 lists the augmentation parameters.

Table 2. The augmentation parameters

Augmentation	Value
Zoom Range	0.1
Rotation Range	15°
Width Shift Range	0.05
Height Shift Range	0.05
Shear Range	0.05

After augmentation, the data distribution of each class becomes more balanced, with each category (normal, tuberculosis, pneumonia) having approximately 2700 image data in the training data and 300 image data in the testing data, with a total of all image data for all categories being 9000 image data. The normalization process is also carried out by dividing the pixel value by 255.0 to standardize the image intensity from the two source datasets, so that the quality of feature extraction is more consistent. All images were resized to 224 × 224 pixels to match the input requirements of EfficientNetV2-B0. All images were resized to 224 × 224 pixels to match the input of EfficientNetV2-B0 while maintaining patch consistency in the GLCM calculation, although this process may reduce fine texture details.

3.2 Analysis

In this section, we describe the process flow in classifying chest X-ray images of lung diseases using the EfficientNetV2-B0 architecture and GLCM texture features. The overall process is illustrated in the IPO diagram in Figure 6.

The lung disease classification process includes several steps, namely: input, process, and output. The following are the explanations of each stage:

a. Input
- The system starts by inputting a total of 9000 chest X-ray images, classified into three categories: normal, tuberculosis, and pneumonia. Each category includes 2700 training images and 300 testing images. Feature extraction using GLCM is also performed on each image.

b. Process
- The images undergo preprocessing, including resizing and augmentation for the tuberculosis class.
- GLCM feature extraction is calculated at four orientations (0°, 45°, 90°, 135°) and three spatial distances (1, 3, 5). The resulting features (energy, correlation, homogeneity, contrast) [15-17] are then averaged to obtain a stable texture representation before being combined with the CNN feature vector.
- The concatenation method is used in the fully connected layer [18], where the GLCM feature vector is added to the feature vector generated by EfficientNetV2-B0, then further processed using the Softmax activation function to produce three-class classification probabilities.

- Data is split using a 90:10 ratio for training and testing [19].
- To avoid overfitting and ensure more robust results, this study uses 5-fold cross-validation, where the training dataset is divided into five parts, which are used alternately as validation data.

c. Classification
- The classification model is built using the EfficientNet-B0, EfficientNetV2-B0, EfficientNetV2-S, and EfficientNetV2-M architecture, integrated with GLCM features. The CNN model processes both visual patterns and texture information to enhance classification accuracy.

d. Output
- The system outputs a classification result predicting one of the three possible categories: Normal, tuberculosis, or pneumonia based on the processed chest X-ray input.
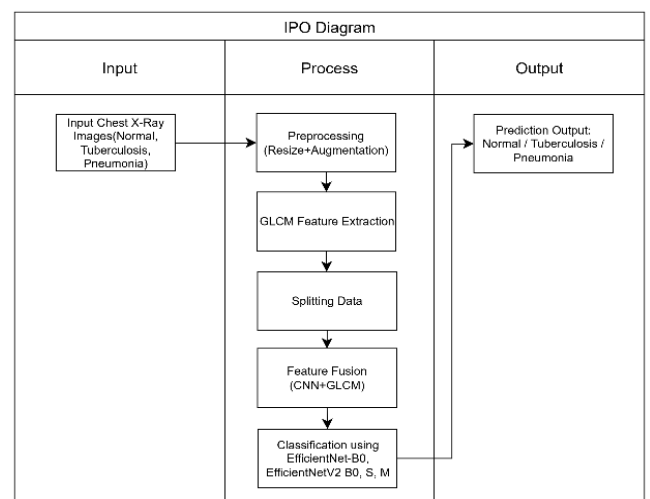

Figure 6. IPO diagram of the classification system

4. RESULTS AND DISCUSSION

In this section, we describe the classification results of chest X-ray images for detecting pulmonary diseases, specifically tuberculosis and pneumonia, using the CNN method. The dataset used consists of merged public datasets from Kaggle, with data preprocessing carried out using GLCM feature extraction. Each model was trained using the Adam optimizer with a learning rate of 0.00001, 50 epochs, and a batch size of 32. This study evaluates the performance of four different CNN architectures. The classification accuracy obtained by each model is described as follows.

4.1 Accuracy graph of EfficientNetV2-B0 architecture

In Figure 7, the application of the CNN method with the EfficientNetV2-B0 architecture shows good classification performance, with accuracy reaching 96.00% and testing loss of 0.1252 on the best-performing fold. The accuracy curve is relatively stable across training epochs, indicating that the learning process converged well without significant overfitting. The loss graph also shows a decreasing trend toward zero, which supports the model's consistent performance.
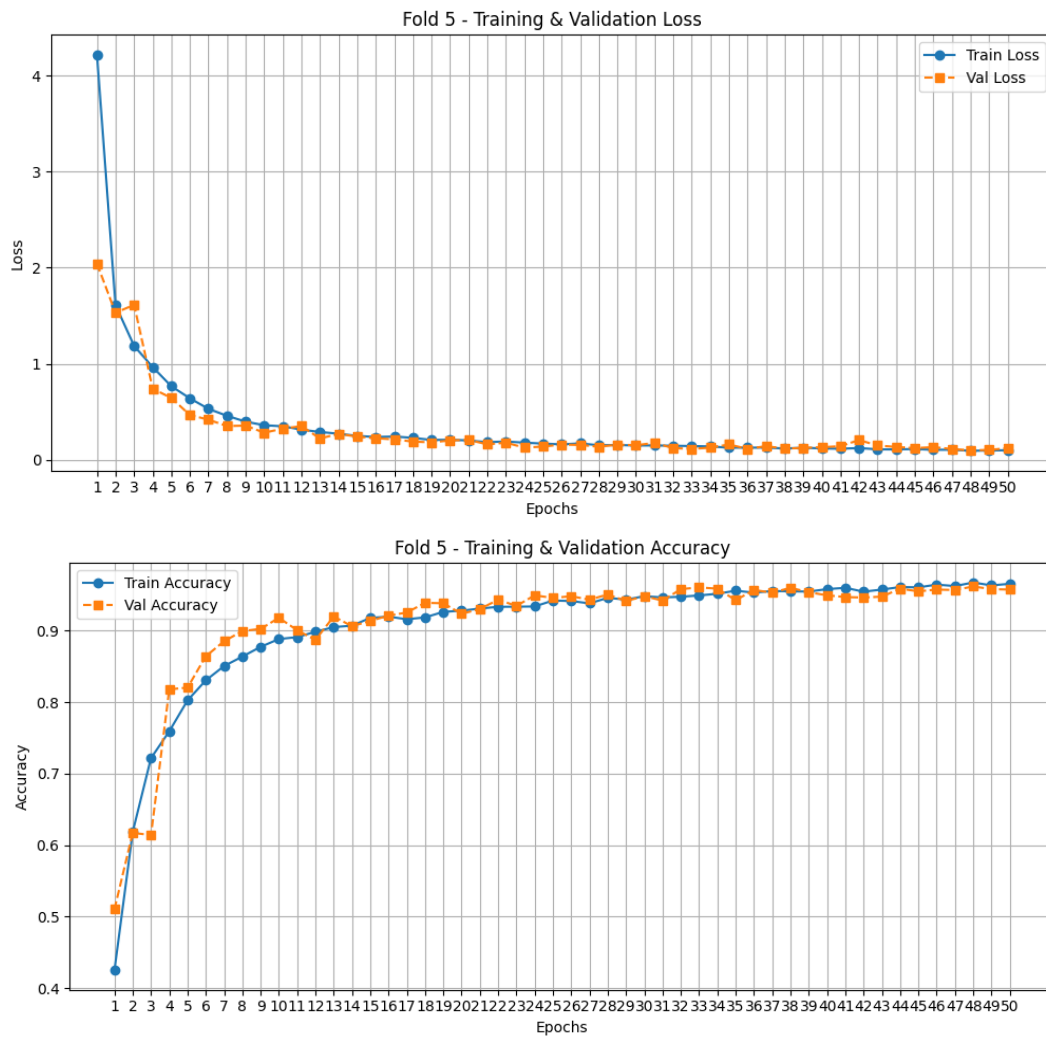
Figure 7. Accuracy and loss graph of EfficientNetV2-B0

```
Classification Report Fold 5:
              precision    recall  f1-score   support

           0       0.93      0.96      0.94       300
           1       0.96      0.96      0.96       300
           2       1.00      0.96      0.98       300

    accuracy                           0.96       900
   macro avg       0.96      0.96      0.96       900
weighted avg       0.96      0.96      0.96       900
```

Figure 8. Classification report of EfficientNetV2-B0

In Figure 8, the normal class achieved a precision of 0.93, a recall of 0.96, and an F1-score of 0.94. The pneumonia class performed fairly well with a precision of 0.96, a recall of 0.96, and an F1-score of 0.96. Meanwhile, the tuberculosis class achieved a precision of 1.00, a recall of 0.96, and an F1-score of 0.98. Based on Figure 9, the most misclassification errors occurred between the normal and pneumonia classes (11 normal cases were predicted as pneumonia, and 13 pneumonia cases were predicted as normal). This clearly indicates that the model is not performing well, even though both categories often display similar infiltrate and opacity patterns in X-ray images. In contrast, the tuberculosis class had only a few mispredictions (10 cases were incorrectly predicted as normal, and 2 cases were incorrectly predicted as pneumonia).
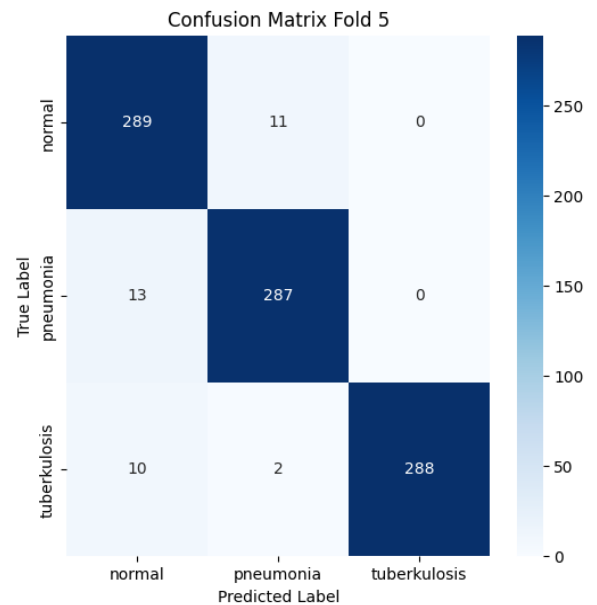

Figure 9. Confusion matrix of EfficientNetV2-B0

4.2 Accuracy graph of the EfficientNet-B0 architecture

As shown in Figure 10, the EfficientNet-B0 architecture achieved a maximum accuracy of 95.22% with a loss of

0.1877. However, the training curve exhibited more fluctuation compared to other models. Although the model performed acceptably, the relatively higher validation loss and testing loss suggest that it may be more prone to overfitting. Nonetheless, the model still provides reliable classification outcomes.
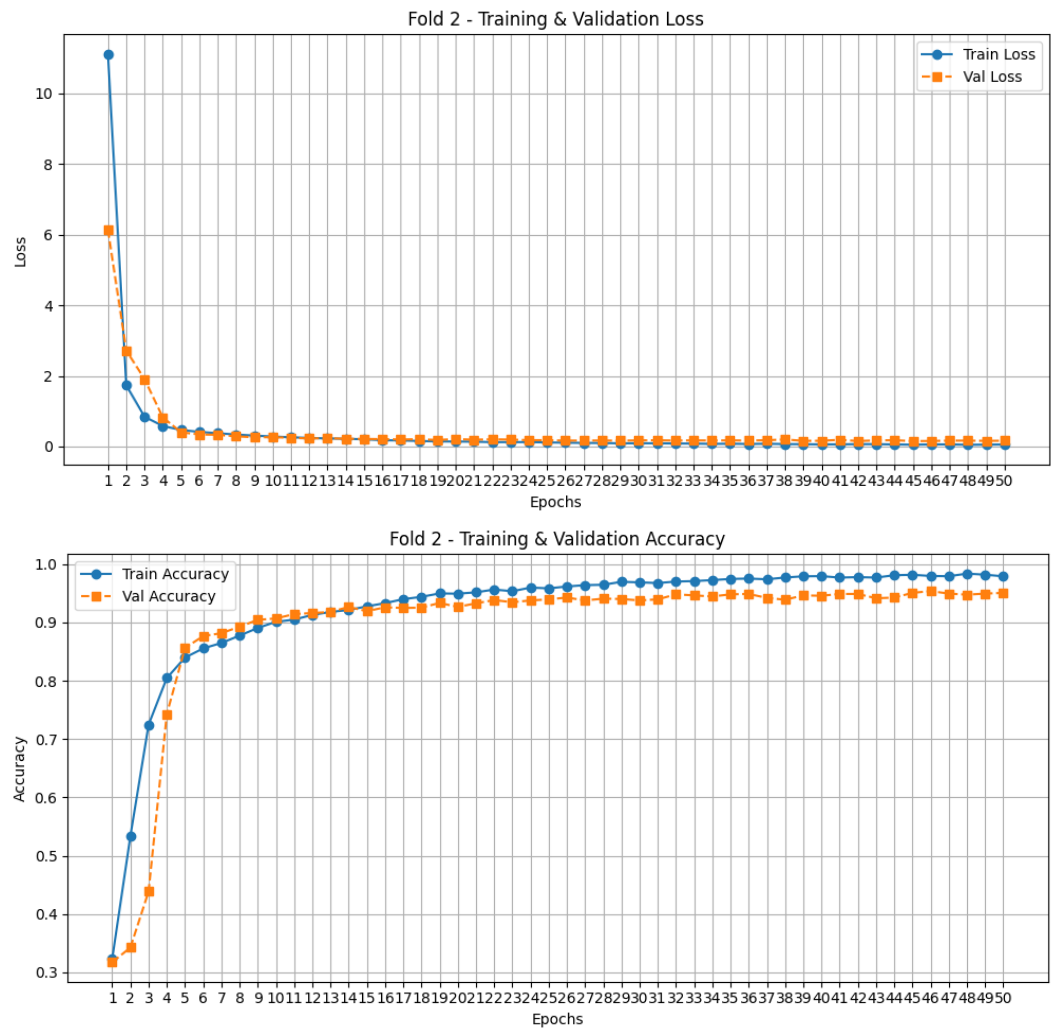

Figure 10. Accuracy and loss graph of EfficientNet-B0

Figure 11 shows that the normal class achieved a precision of 0.95, a recall of 0.91, and an F1-score of 0.93. The pneumonia class performed quite well with a precision of 0.92, a recall of 0.96, and an F1-score of 0.94. Meanwhile, the tuberculosis class achieved the highest performance with a precision of 0.99, a recall of 0.99, and an F1-score of 0.99. Figure 12 shows that the highest number of misclassification errors occurred between the normal and pneumonia classes: 22 normal cases were incorrectly predicted as pneumonia, and 13 pneumonia cases were incorrectly predicted as normal.

Figure 12. Confusion matrix of EfficientNet-B0

```
Classification Report Fold 2:
              precision    recall  f1-score   support

           0       0.95      0.91      0.93       300
           1       0.92      0.96      0.94       300
           2       0.99      0.99      0.99       300

    accuracy                           0.95       900
   macro avg       0.95      0.95      0.95       900
weighted avg       0.95      0.95      0.95       900
```
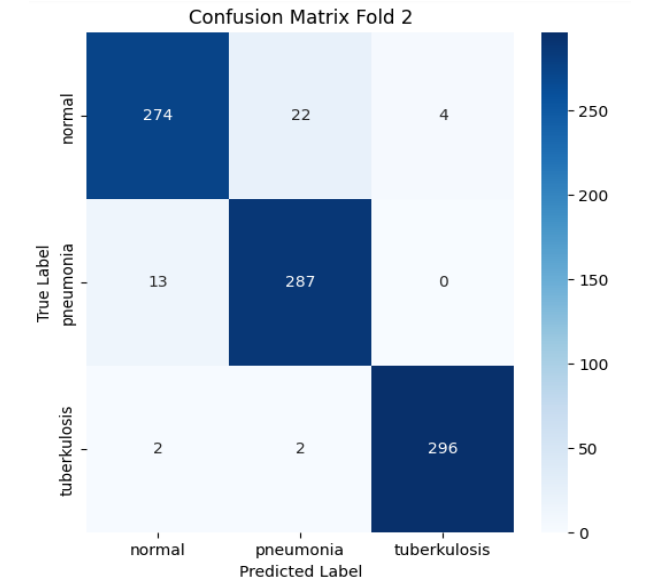
Figure 11. Classification report of EfficientNet-B0

This indicates that despite the high overall accuracy, the model still struggled to distinguish the patterns of light

infiltrates and faint opacities that often appear in pneumonia X-ray images, mimicking normal lung conditions. In contrast, the tuberculosis class performed very well with significantly fewer mispredictions (2 cases were incorrectly predicted as normal, and 2 cases were incorrectly predicted as pneumonia). The high error rate between the normal and pneumonia classes may be caused by the limitations of the EfficientNet-B0 architecture, which has fewer parameters than the EfficientNetV2 variant, so its ability to capture subtle texture differences is less than optimal.

4.3 Accuracy graph of EfficientNetV2-S architecture

Figure 13 presents the training result for EfficientNetV2-S. This model achieved the highest accuracy among all, at 96.67%, with a low loss value of 0.0968. The graph indicates a smooth and increasing accuracy trend over epochs, with minimal fluctuations. This implies that EfficientNetV2-S effectively captures the complex features of chest X-ray images enhanced with GLCM, while maintaining training stability.
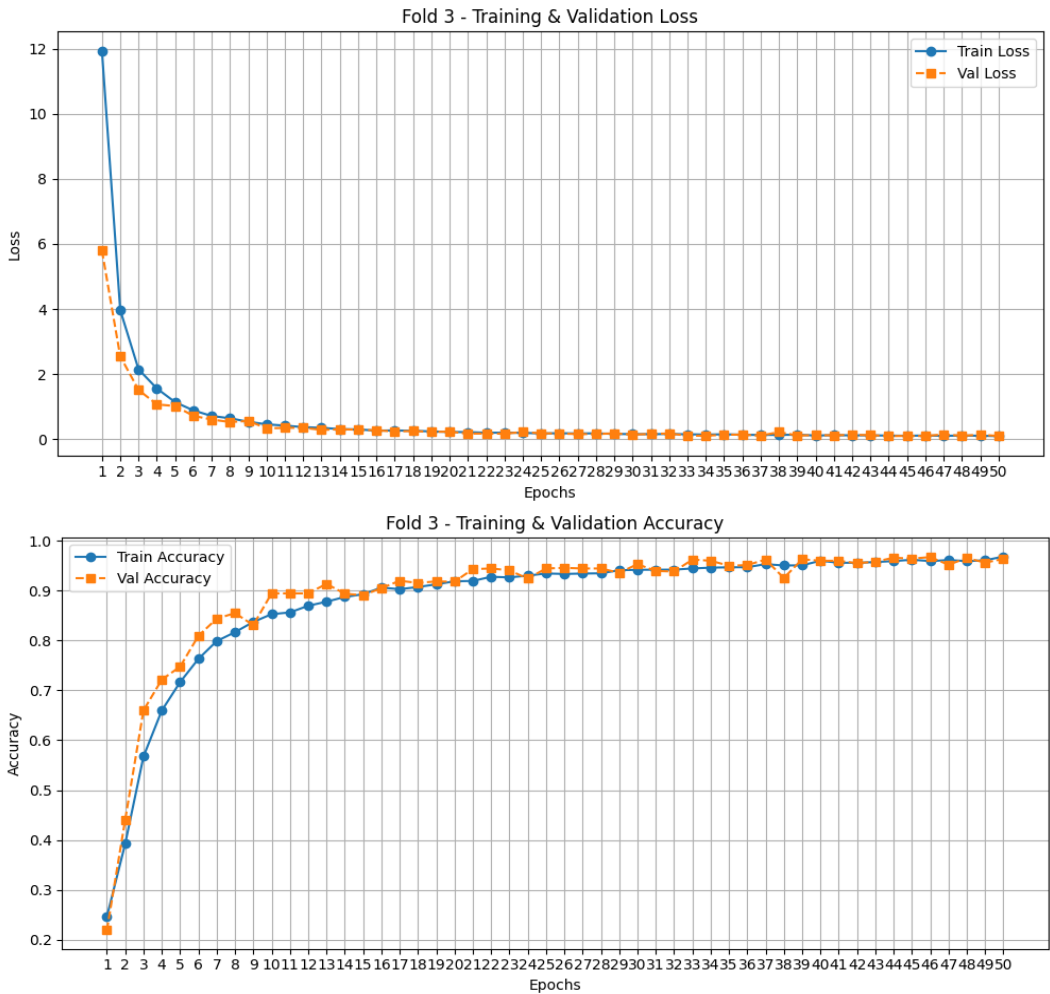
Figure 13. Accuracy and loss graph of EfficientNetV2-S

In Figure 14, the normal class achieved a precision of 0.96, a recall of 0.94, and an F1-score of 0.95. The pneumonia class performed very well with a precision of 0.95, a recall of 0.98, and an F1-score of 0.97. Meanwhile, the tuberculosis class achieved a precision of 0.99, a recall of 0.98, and an F1-score of 0.98. As can be seen in Figure 15, misclassification errors still occurred, particularly between the normal and pneumonia classes, with 14 normal images incorrectly predicted as pneumonia and 7 pneumonia images incorrectly predicted as normal. However, these errors were relatively small compared to the EfficientNet-B0 and EfficientNetV2-B0 architectures. For the tuberculosis class, mispredictions occurred in only 6 images incorrectly predicted as normal, while no tuberculosis cases were incorrectly predicted as pneumonia. This indicates that the model is capable of recognizing the typical texture patterns of tuberculosis quite well.

The superior performance of EfficientNetV2-S can be explained by the combination of GLCM features with CNN in this architecture, making the model more sensitive in identifying differences in infiltrate texture and opacity, which are usually subtle when using only CNN without additional feature extraction. Furthermore, the stable accuracy and loss trends indicate that the model does not experience overfitting, resulting in more consistent classification results.

```
Classification Report Fold 3:
              precision    recall  f1-score   support

           0       0.96      0.94      0.95       300
           1       0.95      0.98      0.97       300
           2       0.99      0.98      0.98       300

    accuracy                           0.97       900
   macro avg       0.97      0.97      0.97       900
weighted avg       0.97      0.97      0.97       900
```

Figure 14. Classification report of EfficientNetV2-S
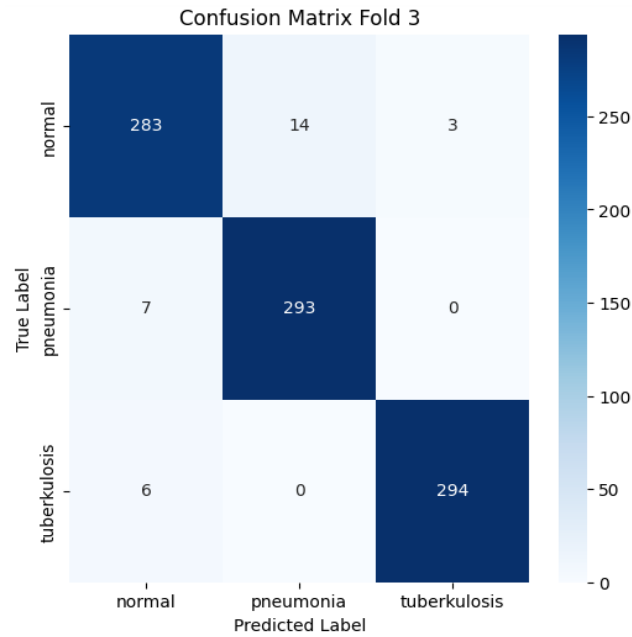
Figure 15. Confusion matrix of EfficientNetV2-S
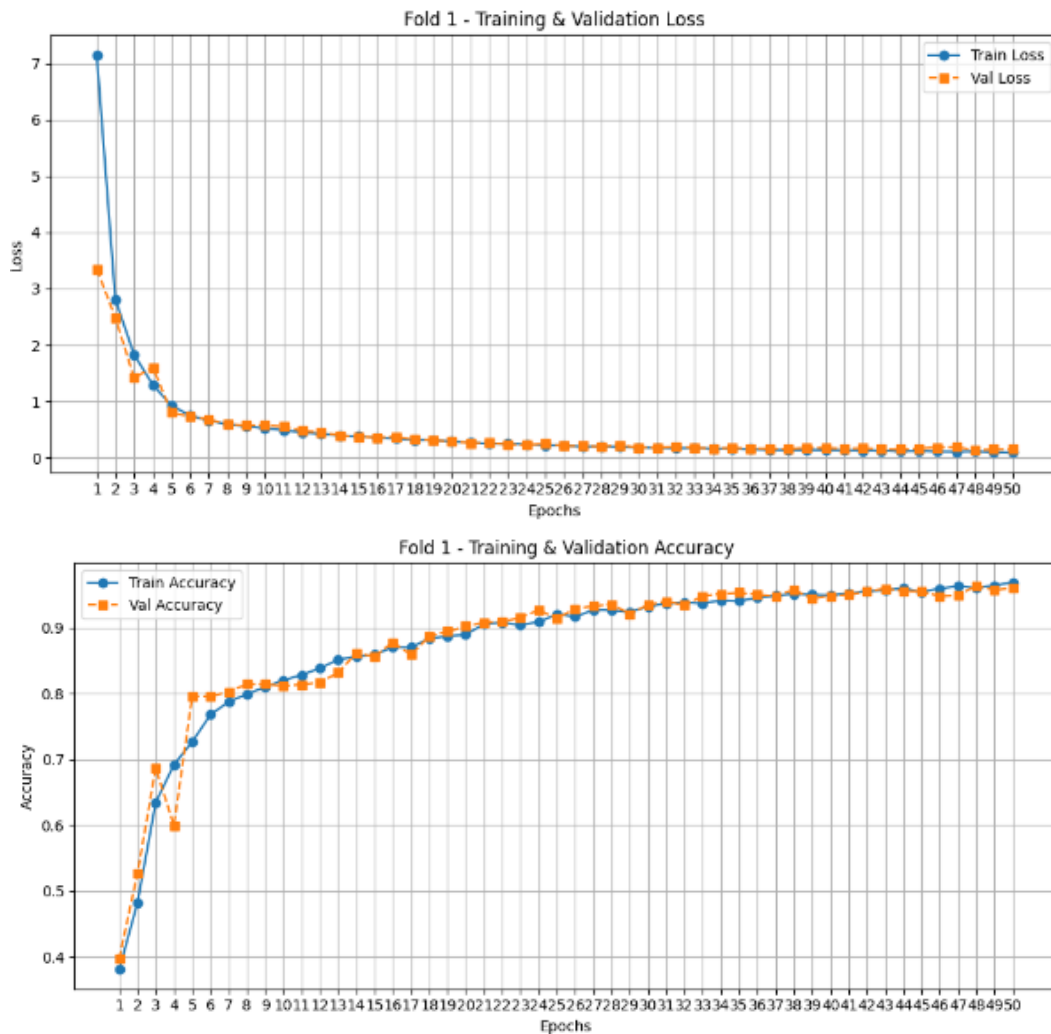
Figure 16. Accuracy and loss graph of EfficientNetV2-M

4.4 Accuracy graph of EfficientNetV2-M architecture

In Figure 16, the EfficientNetV2-M architecture also achieved an accuracy of 96.67%, but with a slightly higher loss of 0.1141. This model shows stable performance; however, it requires the longest training time, averaging around 48

minutes per fold. Although it performs comparably to EfficientNetV2-S in terms of accuracy, the higher computational cost makes it less efficient for large-scale or real-time applications.

Figure 17 shows that the normal class achieved a precision of 0.97, a recall of 0.93, and an F1-score of 0.95. The pneumonia class achieved a precision of 0.94, a recall of 0.99, and an F1-score of 0.97. Meanwhile, the tuberculosis class achieved a precision of 0.99, a recall of 0.98, and an F1-score of 0.98. As can be seen in Figure 18, most misclassification errors still occurred in the normal class, which was incorrectly predicted as pneumonia (17 cases), and four normal images were incorrectly predicted as tuberculosis. Minor errors also occurred in the tuberculosis class, with six tuberculosis images incorrectly predicted as normal and one tuberculosis image incorrectly predicted as pneumonia. However, the pneumonia class was detected very well with only two errors (two pneumonia images predicted as normal).

```
Classification Report Fold 1:
              precision    recall  f1-score   support

           0       0.97      0.93      0.95       300
           1       0.94      0.99      0.97       300
           2       0.99      0.98      0.98       300

    accuracy                           0.97       900
   macro avg       0.97      0.97      0.97       900
weighted avg       0.97      0.97      0.97       900
```
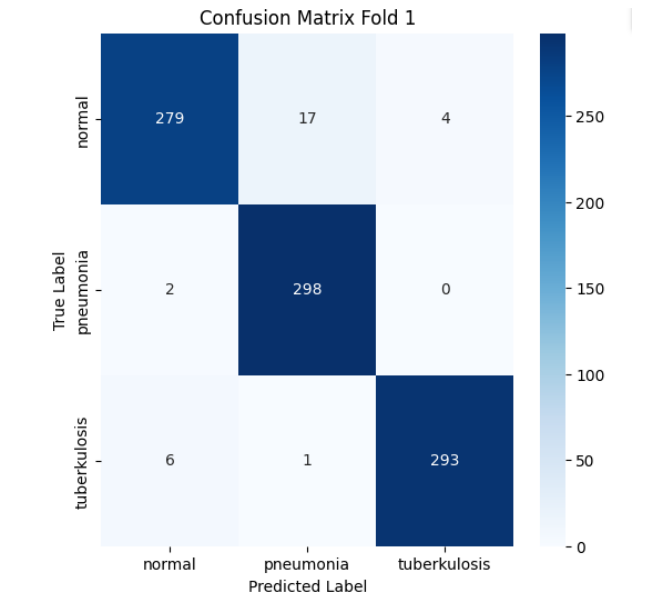
Figure 17. Classification report of EfficientNetV2-M

Figure 18. Confusion matrix of EfficientNetV2-M

When compared to the EfficientNetV2-S architecture, this model achieved the same accuracy (96.67%) but with a slightly higher loss value. This indicates that although the overall predictions are correct, the prediction probability distribution in EfficientNetV2-M is not as robust as in EfficientNetV2-S, meaning the model still presents higher uncertainty in some cases.

Furthermore, the training time for EfficientNetV2-M is significantly longer (47 minutes 54 seconds) compared to EfficientNetV2-S (27 minutes 59 seconds). This difference is due to the larger architecture size of EfficientNetV2-M, with a

higher number of parameters and layer complexity. While this allows the model to learn more complex feature representations, in this study, the additional complexity did not significantly improve accuracy compared to EfficientNetV2-S.

Based on the results from the four CNN architectures, it can be observed that although all models achieved high classification accuracy, the number of parameters, training time, and loss values varied significantly. Among them, EfficientNetV2-S demonstrated the best trade-off between performance and efficiency, achieving the highest accuracy (96.67%) with moderate training time compared to the larger EfficientNetV2-M, which required substantially longer training (47 minutes 56 seconds) without notable performance gains. In contrast, EfficientNetV2-B0 achieved slightly lower accuracy (96.00%) but was the fastest to train, making it a practical option when computational resources are limited.

Comparisons between architectures in terms of accuracy, testing loss, and training time are summarized in Table 3.

Table 3. Comparative performance of CNN architectures

CNN Architecture	Accuracy	Loss	Running Time
EfficientNet-B0	95.22%	0.1877	14 minutes 32 sec
EfficientNetV2-B0	96.00%	0.1252	10 minutes 58 sec
EfficientNetV2-S	96.67%	0.0968	28 minutes 5 sec
EfficientNetV2-M	96.67%	0.1141	47 minutes 56 sec

The comparative results are summarized in Table 3, highlighting that while all architectures are suitable for pulmonary disease classification, EfficientNetV2-S provides the most balanced choice due to its superior accuracy, relatively low loss, and manageable training time. This balance indicates that EfficientNetV2-S, especially when enhanced with GLCM feature extraction, can serve as an efficient and reliable model for real-world medical image classification tasks, supporting more accurate and timely pulmonary disease diagnosis.

5. CONCLUSIONS

Based on the discussion in the previous section, conclusions can be drawn as follows:
a. From the results obtained, it can be concluded that the CNN algorithm combined with GLCM feature extraction provides good classification performance for detecting pulmonary diseases—tuberculosis and pneumonia—based on chest X-ray images. The use of GLCM significantly improves feature richness and model generalization, enhancing classification accuracy.
b. The use of different CNN architectures significantly influences both classification accuracy, loss, and training time. As explained in the results and discussion section, the EfficientNetV2-S architecture achieved the best performance with an accuracy of 96.67%, a loss of 0.0968, and a training time of 27 minutes 59 seconds per fold, making it the most efficient and effective model compared to EfficientNet-B0, EfficientNetV2-B0, and EfficientNetV2-M.
c. Research limitations: The sample size for tuberculosis is relatively small, and the dataset has limited diversity. Furthermore, the hyperparameter selection (learning rate 0.00001 and 50 epochs) was based on preliminary experiments and previous literature, allowing for

performance improvement through fine-tuning in future research.

d. The developed model has the potential to be applied to hospital diagnostic systems and mobile devices for early lung disease screening. However, its implementation is still limited by the dataset used, requiring further exploration. Future research could expand the model's application by testing it on CT images, integrating attention mechanisms to improve important feature detection, and using Grad-CAM or other explainability methods to support the interpretation of predictions by medical professionals.

ACKNOWLEDGMENT

REFERENCES

[1] Nurcahyati, I., Saragih, T.H., Farmadi, A., Kartini, D., Muliadi, M. (2024). Classification of lung disease in X-Ray images using gray level co-occurrence matrix method and convolutional neural network. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 6(4): 332-342. https://doi.org/10.35882/jeeemi.v6i4.457

[2] Achmad, A., Achmad, A.D. (2019). Backpropagation performance against support vector machine in detecting tuberculosis based on lung X-ray image. In First International Conference on Materials Engineering and Management-Engineering Section (ICMEMe 2018), pp. 84-88. https://doi.org/10.2991/icmeme-18.2019.19

[3] Verma, D., Bose, C., Tufchi, N., Pant, K., Tripathi, V., Thapliyal, A. (2020). An efficient framework for identification of tuberculosis and pneumonia in chest X-ray images using neural network. Procedia Computer Science, 171: 217-224. https://doi.org/10.1016/j.procs.2020.04.023

[4] Udayana, I.P.A.E.D., Indrawan, I.G.A., Prawira, I.M.K.S. (2024). Comparison of deep learning methods for detecting tuberculosis through chest X-rays. Journal of Computer Networks, Architecture and High Performance Computing, 6(3): 1290-1299. https://doi.org/10.47709/cnahpc.v6i3.4345

[5] Saeed, A., Shoukat, S., Shehzad, K., Ahmad, I., Eshmawi, A.A., Amin, A.H., Tag-Eldin, E. (2022). A deep learning-based approach for the diagnosis of acute lymphoblastic leukemia. Electronics, 11(19): 3168. https://doi.org/10.3390/electronics11193168

[6] Sidik, D.P., Utaminingrum, F., Muflikhah, L. (2023). Penggunaan variasi model pada arsitektur EfficientNetV2 untuk prediksi sel kanker serviks. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 7(5): 2116-2121.

[7] Jalehi, M.K., Albaker, B.M. (2023). Highly accurate multiclass classification of respiratory system diseases from chest radiography images using deep transfer learning technique. Biomedical Signal Processing and Control, 84: 104745. https://doi.org/10.1016/j.bspc.2023.104745

[8] Mohanaiah, P., Sathyanarayana, P., GuruKumar, L. (2013). Image texture feature extraction using GLCM approach. International Journal of Scientific and Research Publications, 3(5): 1-5.

[9] Rachmad, A., Hapsari, R.K., Setiawan, W., Indriyani, T., Rochman, E.M.S., Satoto, B.D. (2023). Classification of tobacco leaf quality using feature extraction of gray level co-occurrence matrix (GLCM) and K-nearest neighbor (K-NN). In 1st International Conference on Neural Networks and Machine Learning 2022 (ICONNSMAL 2022), pp. 30-38. https://doi.org/10.2991/978-94-6463-174-6

[10] Gurunathan, A., Krishnan, B. (2022). A hybrid CNN-GLCM classifier for detection and grade classification of brain tumor. Brain Imaging and Behavior, 16(3): 1410-1427. https://doi.org/10.1007/s11682-021-00598-2

[11] Rochman, E.M.S., Suprajitno, H., Kamilah, I., Rachmad, A., Santosa, I. (2023). Tuberculosis classification using random forest with K-prototype as a method to overcome missing value. Communications in Mathematical Biology and Neuroscience, 2023: 11. https://doi.org/10.28919/cmbn/7873

[12] Anwar, F., Yunianto, M., Aisha Putri, R.F. (2023). Tuberculosis detection using gray level co-occurrence matrix (GLCM) and K-Nearest Neighbor (K-NN) algorithms. Aceh International Journal of Science & Technology, 12(3): 402-410. https://doi.org/10.13170/aijst.12.3.33241

[13] Afkar, A.N.D., Rachmad, A., Rochman, E.M.S. (2025). Klasifikasi pneumonia dengan metode convolutional neural network. JATI (Jurnal Mahasiswa Teknik Informatika), 9(4): 5821-5828. https://doi.org/10.36040/jati.v9i4.13938

[14] Praskatama, V., Sari, C.A., Rachmawanto, E.H., Yaacob, N.M. (2023). Pneumonia prediction using convolutional neural network. Jurnal Teknik Informatika (JUTIF), 4(5): 1217-1226. https://doi.org/10.52436/1.jutif.2023.4.5.1353

[15] Rachmad, A., Chamidah, N., Rulaningtyas, R. (2020). Mycobacterium tuberculosis identification based on colour feature extraction using expert system. Annals of Biology, 36(2): 196-202.

[16] Yi, N., Li, C., Feng, X., Shi, M. (2018). Research and improvement of convolutional neural network. In 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), Singapore, pp. 637-640. https://doi.org/10.1109/ICIS.2018.8466474

[17] Ibrahim, N.U.R., Saâ, S.O.F.I.A., Hidayat, B., Darana, S. (2022). Klasifikasi grade telur ayam negeri secara non-invasive menggunakan convolutional neural network. ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika, 10(2): 297. https://doi.org/10.26760/elkomika.v10i2.297

[18] Tan, M., Le, Q. (2021). Efficientnetv2: Smaller models and faster training. In International Conference on Machine Learning, pp. 10096-10106.

[19] Islam, N., Shin, S. (2023). Robust deep learning models for OFDM-based image communication systems in intelligent transportation systems (ITS) for smart cities. Electronics, 12(11): 2425. https://doi.org/10.3390/electronics12112425

[20] Yen, L.W., Thinakaran, R., Somasekar, J. (2025).

Machine learning-based denoising techniques for Monte Carlo rendering: A literature review. Machine Learning, 16(2): 0160259. http://doi.org/10.14569/IJACSA.2025.0160259

[21] Rachmad, A., Hutagalung, J., Hapsari, D., Hernawati, S., Syarief, M., Rochman, E.M.S., Asmara, Y.P. (2024). Deep learning optimization of the EfficienNet architecture for classification of tuberculosis bacteria. Mathematical Modelling of Engineering Problems, 11(10): 2664-2670.

[22] Rachmad, A., Syarief, M., Hutagalung, J., Hernawati, S., Rochman, E.M.S., Asmara, Y.P. (2024). Comparison of CNN architectures for Mycobacterium tuberculosis classification in sputum images. Ingenierie des System, 29(1): 49-56. https://doi.org/10.18280/isi.290106

[23] Sunil, C.K., Jaidhar, C.D. (2021). Cardamom plant disease detection approach using EfficientNetV2. IEEE Access, 10: 789-804. https://doi.org/10.1109/ACCESS.2021.3138920

https://doi.org/10.18280/mmep.111008