STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **13**, January 2025, pp 297–308.
Published online in International Academic Press (www.IAPress.org)

Detecting Lung Diseases from X-Ray Images Using Deep Learning

Bao T. Nguyen [1,*], Anh H. Vo [2]

[1] *School of Science, Engineering & Technology, RMIT University, Vietnam*
[2] *Department of Computer Engineering, Sejong University, Seoul, South Korea*

Abstract Lung disease has become one of the most dangerous diseases worldwide after the Covid-19 pandemic. Early diagnosis of lung disease is vital for effective treatment and recovery. In clinical practice, X-ray imaging is currently the most widely used method for diagnosis, and it plays a crucial role as a life-saving factor for individuals suffering from the disease. In recent years, many deep learning approaches have been proposed for the early diagnosis of lung diseases from X-ray images. These approaches have shown high accuracy in predicting the results within a short time. This paper aims to compare different state-of-the-art deep learning models for the task of lung-disease diagnosis. Additionally, we have collected a new dataset of lung disease X-ray images from hospitals in Vietnam to evaluate the performance of each model based on validation loss and validation accuracy. The results show that our proposed deep learning model achieves an accuracy of 98.35% (training) and 86.65% (validation) on the new ChestVN lung disease dataset, which promises to be a good method for applying in daily life. The proposed approach has the potential to assist medical professionals in the early diagnosis of lung diseases, which can lead to better patient outcomes and improved healthcare management.

Keywords Early diagnosis, X-ray images, Lung diseases, Deep learning, Machine Learning, CNNs

DOI: 10.19139/soic-2310-5070-2163

1. Introduction

Lung diseases are one of the most significant health problems, and are considered as the primary cause of death globally, responsible for a significant proportion of all deaths worldwide [2]. Early detection plays a crucial role in improving the chances of survival for individuals with various medical conditions, particularly lung cancer. However, traditional diagnostic methods such as biopsy and bronchoscopy are invasive and have certain limitations. With the advancement of medical imaging technology, IT based automated systems have become a promising tool in medical imaging filed for early detection of lung diseases. Chest X-ray imaging is a commonly used diagnostic tool and is often the first imaging modality used to evaluate patients with suspected lung disease due to its non-invasive nature, low cost, and widespread availability. Interpreting chest X-rays can be a challenging and time-consuming task, even with some experienced radiologists, and misdiagnosis rates can be high [3].

Recent studies have demonstrated that deep learning techniques have shown promising progress in the diagnosing and classifying of various lung diseases [22]. Deep learning is based on artificial neural networks consisting of many hiddern layers to extract complex and compact feature maps or representations of the input data [18]. Through this process, high-level features can be extracted from raw data, enabling the neural network to make accurate predictions or classifications [19]. Deep learning has demonstrated significant potential in medical imaging when identifying patterns and anomalies in medical images, providing valuable diagnostic information for clinicians [8] [10] [11].

*Correspondence to: Nguyen T. Bao (Email: bao.nguyenthien@rmit.edu.vn). School of Science, Engineering & Technology, RMIT University, Vietnam. 702 Nguyen Van Linh, Tan Hung Ward, District 7, Ho Chi Minh, Vietnam.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

Several deep learning architectures have been proposed for diagnosing lung diseases [9] [15], such as CNNs (convolutional neural networks) [15], RNNs (recurrent neural networks), and GANs (generative adversarial networks) [17]. These architectures have demonstrated high accuracy and can significantly reduce the time and cost required for diagnosis compared to traditional methods.

In this paper, we aim to present a thorough and inclusive overview of deep learning techniques employed in the detection of lung diseases. More specifically, we consider to ultilize CNNs due to their success in various image classification tasks. We also present a new dataset of lung disease X-ray images collected from hospitals in Vietnam, which we use to evaluate the performance of different CNN models. We compare the performance of several state-of-the-art CNN architectures, including VGG-19 [16], ResNet-50 [5], and DenseNet-121 [6], based on validation loss and accuracy.

The paper is organized as follows. In Section 2, we present the overview of related works in the field of detecting lung diseases from chest X-ray images using deep learning. In Section 3, we describe our dataset and methodology for training and evaluating the CNN models. In Section 4, we will showcase the experimental findings of our study, where we evaluate and compare the performance of different CNN architectures for detecting lung diseases from chest X-ray images. Finally, in Section 5, we discuss the implications of our findings and outline directions for future research.

2. Related works

Recently, there has been significant interest and attention given to deep learning (DL), which has demonstrated remarkable performance across a wide range of fields, including education, economy, healthcare, and more. One of the most promising areas where DL has demonstrated its potential is medical imaging. With the increasing availability of medical imaging data, DL-based approaches have been exploied more in medical imaging field. The main advantage of DL is its ability to learn hierarchical representations of the raw input image data without the need for handcrafted features. This allows deep learning models to learn complex patterns and representations that are difficult for humans to identify, resulting in better performance and accuracy in various medical image analysis tasks.

Deep learning techniques have also been extensively researched in the field of lung disease diagnosis to improve accuracy and efficiency. Various studies have explored the use of deep learning algorithms for automatically classifying lung diseases. These approaches have shown remarkable performance, demonstrating the potential of DL techniques for medical image analysis, including lung disease diagnosis. In this section, we will review some of the SOTA studies that have addressed this problem.

Among the various DL techniques, convolutional neural networks (CNNs) have gained significant attention for their ability to extract the high-level and compact feature map from images. In 2017, Rajpurkar et al. [13] proposed a model that could diagnose 14 different chest diseases from X-ray images, including pneumonia, tuberculosis, and lung cancer, with an accuracy of 79.5%. The authors utilized a CNN architecture with 121 layers, commonly referred to as DenseNet-121, for their study. This deep neural network was trained to detect various chest diseases to achieve better performance in detecting lung diseases. They confirmed that their model outperformed four radiologists in diagnosing the diseases.

Later in 2018, Wang et al. proposed another deep learning model using the pre-trained ResNet-50 architecture to diagnose three different types of lung diseases, including pneumonia, tuberculosis, and lung cancer, with an accuracy of 93.2% [21]. The authors used data augmentation techniques and transfer learning to improve the model's performance. They showed that their model outperformed other state-of-the-art models, such as AlexNet and VGGNet, in both accuracy and F1-score.

Latter, Jain et al. developed a different model that could diagnose six different types of lung diseases, including pneumonia, tuberculosis, lung cancer, and COVID-19 with the accuracy of 83.3% [7]. The authors used a multi-scale CNN architecture and a region-based CNN architecture to extract features from input images. In [20] Wang et al. proposed another deep CNN architecture for detecting COVID-19. The model achieved high accuracy and showed potential for early detection of the disease. Another study by Jaiswal et al. [1] used a CNN-based model

to classify lung diseases. The proposed model achieved an accuracy of 95.77% on their dataset and outperformed other traditional machine learning models.

In addition to CNN-based models, some researchers have also explored other DL techniques for lung disease diagnosis. For example, in 2020, Ozturk et al. [11] proposed a DL model based on a pre-trained Inception-v3 architecture to diagnose COVID-19 with the accuracy of 98.08%. They also showed that their model outperformed other SOTA models, such as ResNet-50 and VGG16, in terms of accuracy and F1-score. Podder et al. [12] proposed a two-stage DL framework that combines CNN and RNN for the classification of pulmonary nodules in CT images. The proposed model showed high accuracy and outperformed other methods.

Transfer learning is another popular approach in DL, where a pre-trained model is fine-tuned on a specific task. Shin et al. [15] used a transfer learning approach to classify lung nodules in CT images. The authors fine-tuned a pre-trained CNN model and achieved high accuracy in nodule classification. In 2020, Ozturk et al. [11] proposed a DL model based on a pre-trained Inception-v3 architecture to diagnose COVID-19 with the accuracy of 98.08%. They also showed that their model outperformed other state-of-the-art models, such as ResNet-50 and VGG16, in terms of accuracy and F1-score. In 2022, Hamza et al. proposed a DL model based on a pre-trained EfficientNet-B7 architecture to diagnose COVID-19 with 98.28% of accuracy [4]. Other deep learning techniques, such as generative adversarial networks (GANs) and autoencoders, have also been explored for lung disease diagnosis. For example, Sarki et al. [14] proposed a GAN-based method for detecting COVID-19 in chest X-ray images. The proposed method achieved high accuracy and showed robustness to noise and artifacts in the images.

Overall, deep learning has demonstrated remarkable potential in the field of lung disease diagnosis. Despite of its impressive results, there are still several challenges. One of the main challenges is the requirement of a large amount of high-quality annotated data for training. Another challenge is the need for further research to explore the potential of new DL architectures and optimization techniques for improving the performance of lung disease diagnosis. Furthermore, the development of DL models that can operate in real-world clinical settings and assist radiologists in making accurate and timely diagnoses is still an ongoing area of research. In short, the promising results of DL in lung disease diagnosis, coupled with ongoing research, provide a glimpse of a bright future for the application of DL in the field of medical image analysis.

3. Approach

This section explains the methodology adopted in this research for detecting lung diseases from X-ray images using deep learning. The methodology consists of three main steps: data preprocessing, model selection, and model evaluation.

3.1. Data acquisition

The dataset used in our study, ChestVN, was collected from three major medical centers in Vietnam: Cho Ray Hospital, Hoa Hao Medical Center, and BKMEC Hospital. These hospitals were chosen based on their patient volumes and modern diagnostic imaging equipment, ensuring a diverse and high-quality dataset for lung cancer classification.

- Cho Ray Hospital: X-ray images were collected using a Siemens Multix Fusion DR system. The images were captured at a standard resolution of 2000 x 2500 pixels, with a tube voltage range of 100-120 kVp and automatic exposure control (AEC). These high-resolution images were captured in DICOM format to preserve all diagnostic details necessary for classification tasks.
- Hoa Hao Medical Center: The X-ray data were acquired using a Philips DigitalDiagnost C90 system. The image size from this system was 2048 x 2480 pixels, captured with a tube voltage of 100-110 kVp and exposure settings optimized using AEC. These images were stored in DICOM format, ensuring uniformity and fidelity in the diagnostic details.
- BKMEC Hospital: X-ray images were obtained using a GE Definium 8000 digital radiography system. The image resolution was 2100 x 2500 pixels, with a tube voltage range of 100-125 kVp. Like the other sites,

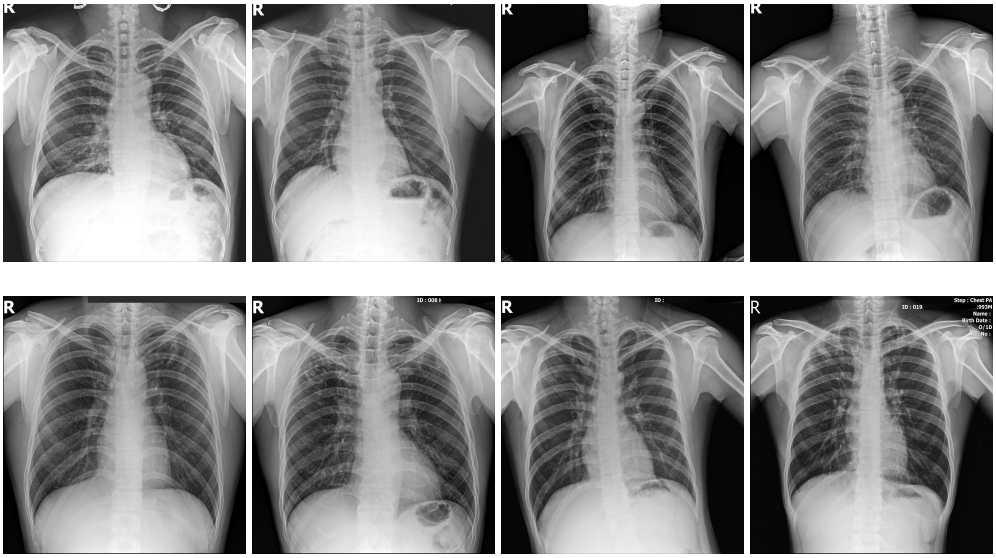


Figure 1. Some samples of the normal (top line) and abnormal (bottom line) images from our X-ray ChestVN dataset.

the images were captured using automatic exposure control to ensure optimal quality and stored in DICOM format.

Across all three centers, the images were collected with consistent protocols, ensuring uniformity in terms of resolution, format, and image quality. All images were anonymized to protect patient privacy, and the standardized high resolution across the dataset makes it suitable for precise lung cancer classification tasks. Finally, the dataset contains a total of 3,000 X-ray images, including 1,500 images of normal lungs and 1,500 images of abnormal lungs with different diseases, such as pneumonia, tuberculosis, and lung cancer.

3.2. Data preprocessing

Data preprocessing is an essential step in machine learning, including deep learning. The images were collected from different X-ray machines, which may result in variations in image quality and size. To ensure uniformity for model training, we resized all images to a fixed size of 224 x 224 pixels using the OpenCV library in Python. In addition to resizing, we applied several preprocessing steps, including normalization and standardization, to improve the convergence of the models and enhance classification performance. More specifically, we normalized the pixel values to a range between 0 and 255 to avoid the influence of different pixel value ranges on model performance. These steps were crucial for maintaining consistency across the dataset and ensuring that variations in image acquisition did not adversely affect the model's ability to generalize. Some samples of the images from the ChestVN are showed in the Figure 1.

Data augmentation is a technique used to artificially expand the size of the training dataset by generating new training data from the existing data. Data augmentation can help prevent overfitting by reducing the model's sensitivity to variations in the data. In the case of lung disease detection using deep learning models, data augmentation techniques can be used to create variations of the lung X-ray images. In our study, we implemented several specific data augmentation techniques, including *rotation, zoom, horizontal flipping, vertical flipping, shearing, and brightness adjustment*. These techniques were selected to enhance the diversity of our training dataset and improve the model's robustness to variations in input data.

The rationale for using data augmentation is to simulate different imaging conditions that a model might encounter in real-world clinical settings. By artificially expanding the dataset, we aimed to reduce overfitting and improve generalization. This is particularly important in medical imaging, where acquiring large, labeled datasets can be challenging.

By applying these data augmentation techniques, we were able to increase the size of the training dataset and generate more diverse variations of the lung X-ray images. This helped to improve the model's ability to generalize and accurately detect lung diseases. Our results indicate that data augmentation had a positive impact on model performance. Specifically, we observed improved accuracy and a more stable training process, as the model became better equipped to handle variations in X-ray images. Finally, we split the dataset into training and validation sets, with the ratio of 80%, and 20% respectively.

3.3. Model selection

In this study, we selected four state-of-the-art deep learning models for detecting lung diseases from X-ray images: VGG19, ResNet, MobileNetV2, and DenseNet. These models have been widely used in computer vision tasks, including medical image analysis, and have achieved outstanding performance on benchmark datasets [8].

VGG19 is a 19-layer convolutional neural network (CNN) architecture that was proposed by Simonyan and Zisserman in 2014 [16]. It has a simple and elegant structure with small convolutional filters (3x3) and max pooling layers. The model has shown excellent performance on image classification tasks, including medical image analysis.

ResNet is a deep CNN architecture proposed by He et al. in 2016 [5]. It has a unique residual learning strategy that allows training of much deeper networks (more than 100 layers) without degradation in performance. ResNet has been widely used in medical image analysis and achieved state-of-the-art results on various benchmark datasets.

DenseNet is another deep CNN architecture proposed by Huang et al. in 2018 [6]. It has a dense connectivity pattern between layers, where each layer is directly connected to every other layer in a feed-forward fashion. DenseNet has achieved excellent performance on various computer vision tasks, including medical image analysis.

3.4. Model evaluation metrics

We evaluated the performance of the models using several metrics, including accuracy, precision, recall, and F1-score. We also plotted the training and validation loss and accuracy curves to analyze the model's performance during training. Finally, we compared the performance of the three models and selected the best-performing model for detecting lung diseases from X-ray images.

4. Experiments & Results

We conducted our experiments on the ChestVN dataset of lung disease X-ray images using the Keras deep learning framework with a TensorFlow backend for implementing the models. The four deep learning models, VGG-19, ResNet-50, MobileNetV2, and DenseNet-121 were trained and evaluated on the dataset. To initialize the models, we used pre-trained weights on the ImageNet dataset and fine-tuned the models on our dataset.

During the training phase, we used the categorical cross-entropy loss function. We experimented with different training epochs and batch sizes to optimize the model performance. Specifically, we trained each model for 10 epochs, 50 epochs, and 100 epochs with a batch size of 32.

To optimize the performance of lung disease classification, we conducted a series of experiments with different learning rates, including 0.001, 0.0001, and 0.00001. The results, presented in Table 1, indicate that the model achieved the highest performance with a learning rate of 0.0001, reaching an accuracy of 87% on the ChestVN dataset.

In addition to adjusting the learning rate, we experimented with different optimizers such as Adam, SGD, and RMSprop. The combination of Adam with a learning rate of 0.0001 produced the most optimal results, ensuring faster convergence and better generalization during training. These tuning efforts demonstrate that carefully selecting hyperparameters significantly improves the overall performance of the model for lung disease detection.

To improve the generalization capability of the models and to prevent overfitting, we applied various data augmentation techniques during the preprocessing step. These techniques include rotation, zooming, horizontal and vertical flipping, and width and height shifting. These techniques enable the models to learn from a more diverse set of data, leading to better performance in classifying lung diseases from X-ray images.

Learning Rate	Validation			
	Accuracy	Presision	Recall	F1-score
0.001	0.75	0.72	0.81	0.76
0.0001	**0.87**	**0.84**	**0.84**	**0.84**
0.00001	0.81	0.76	0.94	0.84

Table 1. Results of lung disease classification using different learning rates with DenseNet-121 model. This table summarizes the accuracy, precision, recall, and F1-score achieved by the model with varying learning rates, highlighting the optimal performance at a learning rate of 0.0001.

Model	No. Parameters	Training	Validation			
		Accuracy	Accuracy	Presision	Recall	F1-score
10 epochs						
VGG-19	23,845,186	0.91	0.73	0.74	0.78	0.76
RestNet-50	30,554,242	0.95	0.79	0.71	0.8	0.75
MobileNetV2	7,651,650	0.96	0.75	0.66	0.9	0.76
DenseNet-121	8,784,066	**0.97**	**0.81**	0.77	0.83	**0.80**
50 epochs						
VGG-19	23,845,186	0.92	0.81	0.78	0.83	0.80
RestNet-50	30,554,242	0.97	0.82	0.82	0.86	0.84
MobileNetV2	7,651,650	0.96	0.81	0.75	0.85	0.80
DenseNet-121	8,784,066	**0.98**	**0.87**	0.82	0.8	**0.83**
100 epochs						
VGG-19	23,845,186	0.93	0.81	0.785	0.75	0.75
RestNet-50	30,554,242	0.967	0.81	0.80	0.88	0.83
MobileNetV2	7,651,650	0.97	0.84	0.85	0.8	0.83
DenseNet-121	8,784,066	**0.98**	**0.87**	**0.84**	**0.84**	**0.84**

Table 2. Results of lung disease classification using three different deep learning models, VGG-19, RestNet-50, MobileNetV2 and DenseNet-121, with different numbers of training epochs. The table shows the number of parameters, training and validation accuracy, as well as precision, recall, and F1-score for each model. The results are reported for training with 10 50, and 100 epochs.

We used a train-validation split of 80-20 to evaluate the performance of the models. The performance of each model was evaluated based on the validation loss and accuracy. The models were trained using an NVIDIA GeForce GTX 1080 Ti GPU for faster processing.

To evaluate the performance of the three deep learning models, we conducted validation step on the independent test set. We used various evaluation metrics such as accuracy, precision, recall, ad F1-score. Accuracy represents the ratio of the number of correctly classified samples to the total number of samples. Precision is the proportion of true positive samples to the total number of samples predicted as positive. Recall, also known as sensitivity, is the proportion of true positive samples to the total number of actual positive samples. F1-score is the harmonic mean of precision and recall. Overall, these evaluation metrics and analysis help us determine the effectiveness and limitations of each model in detecting lung diseases from X-ray images.

Table 2 shows the results of our experiments. It shows the number of parameters of each model, training accuracy and loss, validation accuracy, precision, recall, and F1-score for each model after training for 10, 50 and 100 epochs.

The results of the lung disease classification experiment are summarized in Table 2. Four models VGG-19, ResNet-50, MobileNetV2, and DenseNet-121 were each trained for 10, 50, and 100 epochs.

During the training step, the models were fed with the training dataset and optimized using the backpropagation algorithm. The training accuracy, loss, and validation accuracy were monitored and recorded for each epoch. As

shown in the table, all models had high training accuracy and low training loss values after 10 and 50 epochs. The DenseNet-121 model achieved the highest training accuracy of 0.97 after 10 epochs, which improved to 0.98 after training for 50 epochs. In comparison, the VGG-19 model had the lowest training accuracy, with 0.91 and 0.92 after 10 and 50 epochs, respectively. Meanwhile, ResNet-50 achieved an accuracy of 0.95 and 0.97 after training for 10 and 50 epochs, respectively.

During the validation step, the models were evaluated on a separate dataset from the training set, which was not used during training. This was done to assess the generalization capability of the models, i.e., how well they can classify lung diseases on unseen data. The validation accuracy, precision, recall, and F1-score were computed and recorded for each model after 10 and 50 epochs. The accuracy is the proportion of correctly classified samples out of all samples in the validation set. The precision is the proportion of true positive samples out of all samples classified as positive. The recall is the proportion of true positive samples out of all actual positive samples. The F1-score is the harmonic mean of precision and recall, which provides a balanced measure of both metrics. As shown in Table 2, the DenseNet-121 model achieved the highest validation accuracy, precision, recall, and F1-score across all epochs. This indicates that the model is able to generalize well to unseen data and can accurately classify lung diseases. The VGG-19 and ResNet-50 models also achieved good performance, but were outperformed by DenseNet-121.

The results of our experiments, as seen in Table 2, demonstrate that DenseNet-121 is the best performing model for classifying lung diseases. This is evident from the highest accuracy, precision, recall, and F1-score achieved by the model compared to the other two models, VGG-19 and ResNet-50 for both training and validation stage. In Figure 2, we highlight some key regions in the lung X-ray images that significantly influence the final classification. This not only enhances transparency in the model's decision-making process but also increases trust in the predictions made by the model. Additionally, we compared the performance of various methods, including VGG-19, ResNet-50, MobileNetV2, and DenseNet-121, to support the diagnostic decision-making process. The results of these comparisons are illustrated in Figure 2, which provides a visual representation of the important features recognized by each model. This visualization aids radiologists in understanding the basis for the model's predictions and fosters collaboration between AI systems and clinical professionals.

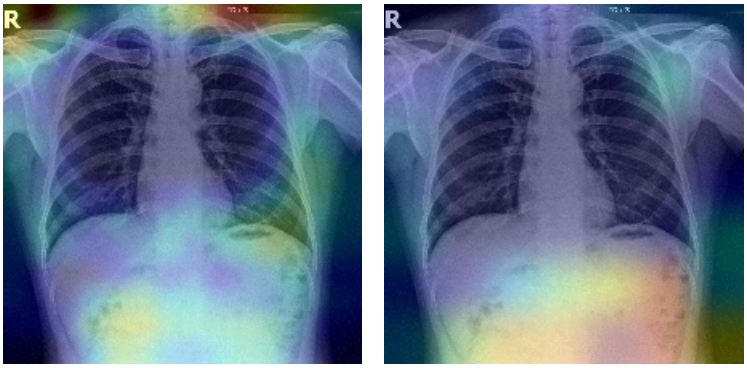
We also observed that increasing the number of training epochs improved the performance of all models. For example, the validation accuracy of VGG-19 increased from 0.73 to 0.80 when trained for 50 epochs, and the same trend can be seen in the other models as well. The improved performance metrics, such as precision, recall, and F1-score, also demonstrate the effectiveness of training the models for a higher number of epochs. Furthermore, the bold values in Table 2 indicate the better performing metrics, which further emphasize the superiority of the DenseNet-121 model. The model achieved the highest F1-score of 0.80 after training for 10 epochs and 0.83 after training for 50 epochs.

To mitigate overfitting, we employed several techniques during the training process, including dropout and early stopping.

- Applying data augmentation techniques during the preprocessing step. These techniques enable the models to learn from a more diverse set of data, leading to better performance in classifying lung diseases from X-ray images.
- Dropout was applied to randomly deactivate a fraction of neurons during training, which helps prevent the model from relying too heavily on any particular feature.
- Early stopping was implemented to halt training when the validation loss stopped improving, preventing the model from continuing to learn patterns specific to the training set.

The Figure 3 illustrates the training and validation accuracy/loss curves, showing the impact of these techniques on improving the generalization performance of the model and ensuring a more balanced fit across training and validation datasets.

We also conducted a comprehensive analysis of false positives and negatives to identify common misclassification patterns and investigate potential causes, such as image quality and disease stage. This detailed examination is crucial for understanding the limitations of our model and for guiding future improvements. To support our findings, we included a comparison of confusion matrices between DenseNet121 and other methods on the ChestVN dataset, as illustrated in Figure 4.



(a) VGG-19 (b) ResNet 50

(c) MobileNetV2 (d) Densenet-121

Figure 2. Visualization of the important regions that influence the predicted classification of "abnormal" cases for DenseNet-121 compared to other methods. This figure highlights the areas identified by each model, illustrating their focus on key features within the lung X-ray images.

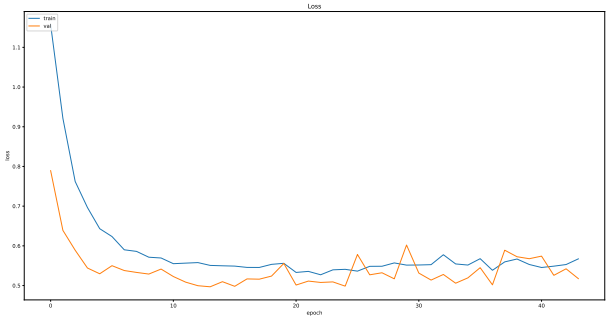
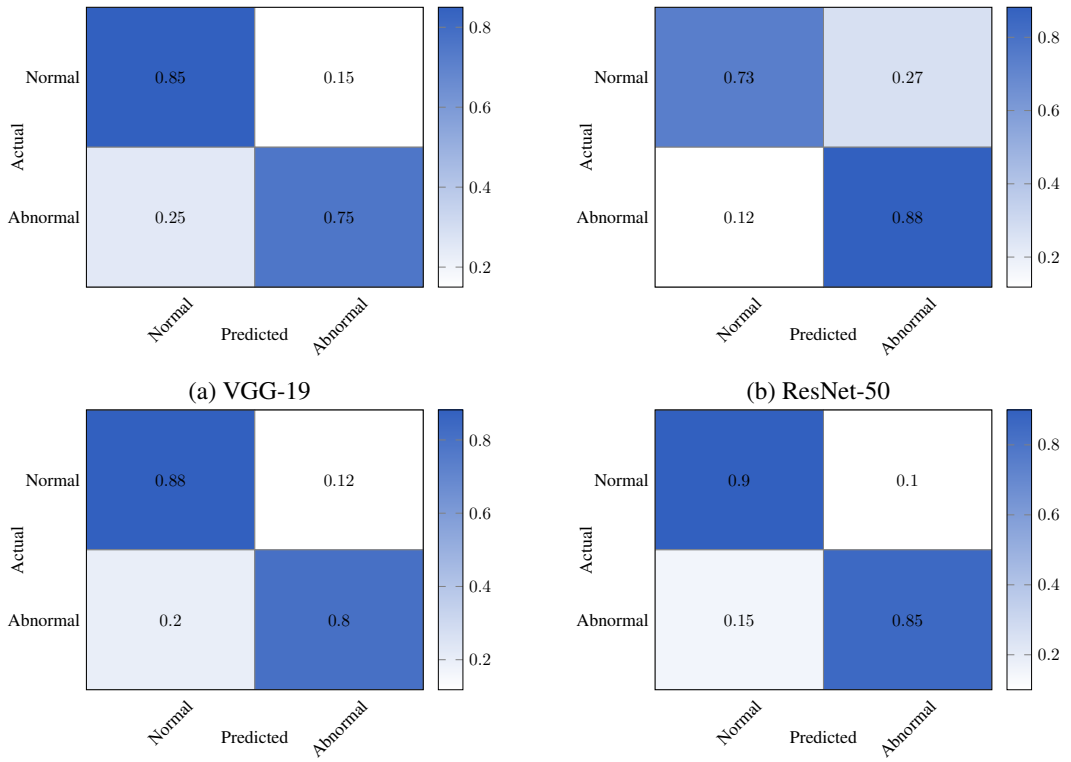


Figure 3. The loss values during the training and validation process. This figure demonstrates the model's convergence, showing how techniques such as dropout and early stopping were applied to reduce overfitting and improve generalization.

The results demonstrate that DenseNet121 achieves better performance than the other methods, with an accuracy of 90% for normal cases and 85% for abnormal cases. This performance underscores the model's ability to correctly identify a significant proportion of cases, but it also highlights specific areas where misclassifications occur. For instance, we observed that certain image quality issues, such as blurriness or inconsistent exposure, contributed to false negatives, particularly in advanced disease stages.

By analyzing these patterns, we can pinpoint the root causes of misclassifications and develop targeted strategies for improvement. This may involve refining our data preprocessing techniques, enhancing image acquisition protocols, or even incorporating additional data augmentation strategies to simulate various conditions. Overall, this analysis not only reinforces the efficacy of DenseNet121 but also lays the groundwork for future enhancements aimed at maximizing the model's clinical applicability.



(a) VGG-19 (b) ResNet-50

(c) MobileNetV2 (d) DenseNet121

Figure 4. Comparison of confusion matrices for DenseNet121 and other methods on the ChestVN dataset. This figure illustrates the model's performance in distinguishing between normal and abnormal cases, highlighting areas of strength and identifying common misclassification patterns.

Overall, the results of our experiments demonstrate the effectiveness of deep learning models in classifying lung diseases. Experimental results suggest that DenseNet-121 is the best performing model for classifying lung diseases based on the metrics evaluated. However, it is important to note that the choice of model ultimately depends on the specific problem and the available resources. In addition, our results indicate that increasing the number of epochs can significantly improve the performance of the models. Moreover, we ackowledge that the dataset used in this study was indeed collected from medical centers in Vietnam, which may introduce certain population or regional biases. While our results demonstrate strong performance on this dataset, we recognize that further validation on datasets from other regions is necessary to confirm the broader applicability of our findings. To mitigate this limitation, other test our model on some publicly available X-ray datasets must be carried out. This will allow us to assess the generalizability of our model across diverse populations and healthcare settings.

5. Conclusion & Future Works

In this paper, we have made some contributions in the field of lung disease classification. Firstly, we have collected a new dataset named ChestVN from various hospitals in Vietnam. The dataset consists of high-quality chest X-ray

images of both patients diagnosed with various lung diseases and healthy person. Due to privacy concerns, the data will not be published, but it will be made available to researchers upon request to the authors. The ChestVN dataset has a large number of samples with high-quality annotations, making it a valuable resource for researchers in the field of lung disease classification. Secondly, we have performed experiments on the ChestVN dataset using three different models, namely VGG-19, ResNet-50, and DenseNet-121. The experiments were conducted to compare the performance of these models in classifying different types of lung diseases. The models were trained using a transfer learning approach, where the pre-trained models were fine-tuned on the ChestVN dataset. We evaluated the models based on various metrics, including accuracy, precision, recall, and F1-score, to determine their performance. Finally, our experimental results have shown that DenseNet-121 outperformed the other models in classifying different types of lung diseases. The model achieved an accuracy of 0.81 after training for 10 epochs and 0.87 after training for 50 epochs. The precision, recall, and F1-score of DenseNet-121 were also better than the other models. Our results indicate that DenseNet-121 is a promising model for lung disease classification, and it could be used in future research to improve the accuracy of lung disease diagnosis.

In conclusion, our contributions in this paper include the collection of a new lung disease dataset, the performance of experiments on the ChestVN dataset using three different models, and the determination that DenseNet-121 outperforms the other models in lung disease classification. The proposed approach for lung disease classification using X-ray images has significant potential for clinical application. By providing a reliable and efficient method for early detection, our model can assist radiologists in diagnosing lung diseases, including lung cancer, pneumonia, and other pulmonary conditions. This can lead to timely interventions and improved patient outcomes. However, there are limitations that need to be addressed before clinical implementation:

- Dataset Diversity: Our model was trained on a dataset collected from hospitals in Vietnam, which may limit its generalizability to other populations and healthcare settings. Further validation on diverse datasets is essential to ensure robustness across different demographics.
- Real-World Variability: Variability in imaging conditions, such as differences in X-ray machines and patient positioning, could impact the model's performance in real-world clinical environments.
- Interpretability: While our model achieves high accuracy, it is crucial to develop interpretability tools to help clinicians understand the model's decisions, enhancing trust and facilitating integration into clinical workflows.

To translate our findings into clinical practice, several steps are necessary. First, validation studies should be conducted across multiple hospitals to assess the model's performance in diverse settings. Additionally, collaboration with healthcare professionals is crucial to ensure smooth integration of the model into existing radiology workflows. Pursuing regulatory approval through clinical trials will help demonstrate the model's safety and efficacy. Lastly, training clinicians on how to use and interpret the model is essential to ensure its effective application in practice. By addressing these aspects, our approach can significantly enhance diagnostic accuracy and efficiency in clinical settings. These contributions provide valuable insights into the field of lung disease classification and could have a significant impact on the accuracy of lung disease diagnosis in the future.

In future work, we aim to explore several key areas to enhance our research. One significant direction is the integration of additional imaging modalities, such as CT scans or MRI, alongside X-rays, which could improve diagnostic accuracy through a multimodal approach. To improve the classification results, one possibility is to further explore the use of other deep learning models for lung disease classification on the ChestVN dataset. This could include testing newer models that have been developed since the completion of this project, or exploring the use of different architectures altogether. It is also possible to investigate the potential of transfer learning to improve the performance of the models. Transfer learning involves using pre-trained models as a starting point for training on a new dataset. This approach has been shown to be effective in many computer vision tasks and may be useful for improving the performance of the models on the ChestVN dataset. The rationale for using these pretrained models lies in their extensive training on large datasets, such as ImageNet, which allows them to leverage learned representations and feature extraction capabilities. This significantly reduces the need for large amounts of labeled data in our specific domain, thereby accelerating the training process and improving model performance.
Another potential direction is to explore the use of explainable AI techniques to gain insight into how the models are making their predictions. This could involve techniques such as saliency maps, which highlight the regions

of the input image that are most important for the model's prediction, or gradient-weighted class activation mapping (Grad-CAM), which highlights the regions of the image that are most relevant to the predicted class. Last but not least, it may be worth exploring the use of ensemble methods to combine the predictions of multiple models to improve overall classification performance. This could involve using a combination of different models with complementary strengths to produce more robust predictions. In future work, we plan to implement and evaluate different ensemble strategies, such as majority voting and weighted averaging, to assess their potential for enhancing the classification accuracy not only on the ChestVN dataset but also on some other available lung dataset too.

We also plan to focus on real-world deployment considerations, including field testing to evaluate model performance in diverse clinical environments and developing intuitive user interfaces that facilitate integration into existing workflows. Continuous feedback mechanisms from clinicians will be implemented to refine the model further. Additionally, conducting longitudinal studies to assess the model's effectiveness in tracking disease progression over time is essential. Addressing ethical considerations, such as bias and patient privacy, will be crucial for responsible deployment. Finally, engaging with healthcare professionals will ensure that our model meets clinical needs and enhances its practical application. By pursuing these avenues, we aim to significantly improve diagnostic capabilities and facilitate successful implementation in various healthcare settings.

Acknowledgement

We would like to express our sincere gratitude to the medical practitioners and doctors from Cho Ray Hospital, Hoa Hao Medical Center, and BKMEC Hospital for their generous support in providing us with the necessary resources and expertise to conduct this research. We also want to extend our gratitude to the students from RMIT who helped refine the ChestVN dataset. Without their invaluable contribution, this research would not have been possible.

References

1. Anitha, J., Kalaiarasu, M., Kumar, N.S., Sundar, G.R.: Detection and classification of lung diseases using deep learning. AIP Conference Proceedings **2519**(1), 030001 (2022). https://doi.org/10.1063/5.0109980, https://aip.scitation.org/doi/abs/10.1063/5.0109980
2. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A.: Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **68**(6), 394—424 (November 2018). https://doi.org/10.3322/caac.21492, https://doi.org/10.3322/caac.21492
3. Forte, G.C., Altmayer, S., Silva, R.F., Stefani, M.T., Libermann, L.L., Cavion, C.C., Youssef, A., Forghani, R., King, J., Mohamed, T.L., Andrade, R.G.F., Hochhegger, B.: Deep learning algorithms for diagnosis of lung cancer: A systematic review and meta-analysis. Cancers **14**(16) (2022). https://doi.org/10.3390/cancers14163856, https://www.mdpi.com/2072-6694/14/16/3856
4. Hamza, A., Attique Khan, M., Wang, S.H., Alqahtani, A., Alsubai, S., Binbusayyis, A., Hussein, H.S., Martinetz, T.M., Alshazly, H.: Covid-19 classification using chest x-ray images: A framework of cnn-lstm and improved max value moth flame optimization. Frontiers in Public Health **10** (2022). https://doi.org/10.3389/fpubh.2022.948205, https://www.frontiersin.org/articles/10.3389/fpubh.2022.948205
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
6. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks (2018)
7. Jain, R., Gupta, M., Taneja, S., Hemanth, D.J.: Deep learning based detection and analysis of covid-19 on chest x-ray images. Applied Intelligence **51**(3), 1690–1700 (2021). https://doi.org/10.1007/s10489-020-01902-1, https://doi.org/10.1007/s10489-020-01902-1
8. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**, 60–88 (2017). https://doi.org/https://doi.org/10.1016/j.media.2017.07.005
9. Lundervold, A.S., Lundervold, A.: An overview of deep learning in medical imaging focusing on mri. Zeitschrift für Medizinische Physik **29**(2), 102–127 (2019). https://doi.org/https://doi.org/10.1016/j.zemedi.2018.11.002, special Issue: Deep Learning in Medical Physics
10. Nguyen, C.P., Hoang Vo, A., Nguyen, B.T.: Breast cancer histology image classification using deep learning. In: 2019 19th International Symposium on Communications and Information Technologies (ISCIT). pp. 366–370 (2019). https://doi.org/10.1109/ISCIT.2019.8905196
11. Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O., Rajendra Acharya, U.: Automated detection of covid-19 cases using deep neural networks with x-ray images. Computers in Biology and Medicine **121**, 103792 (2020). https://doi.org/https://doi.org/10.1016/j.compbiomed.2020.103792

12. Podder, P., Das, S.R., Mondal, M.R.H., Bharati, S., Maliha, A., Hasan, M.J., Piltan, F.: Lddnet: A deep learning framework for the diagnosis of infectious lung diseases. Sensors **23**(1) (2023). https://doi.org/10.3390/s23010480, https://www.mdpi.com/1424-8220/23/1/480
13. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., et al.: Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225 (2017)
14. Sarki, R., Ahmed, K., Wang, H., Zhang, Y., Wang, K.: Automated detection of covid-19 through convolutional neural network using chest x-ray images. PLOS ONE **17**(1), 1–26 (01 2022). https://doi.org/10.1371/journal.pone.0262052, https://doi.org/10.1371/journal.pone.0262052
15. Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M.: Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE Transactions on Medical Imaging **35**(5), 1285–1298 (2016)
16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2015)
17. Suganyadevi, S., Seethalakshmi, V., Balasamy, K.: A review on deep learning in medical image analysis. International Journal of Multimedia Information Retrieval **11**(1), 19–38 (2022). https://doi.org/10.1007/s13735-021-00218-1, https://doi.org/10.1007/s13735-021-00218-1
18. Vo, A.H., Le, T.B.T., Pham, H.V., Nguyen, B.T.: An efficient framework for outfit compatibility prediction towards occasion. Neural Computing and Applications (2023). https://doi.org/10.1007/s00521-023-08431-1, https://doi.org/10.1007/s00521-023-08431-1
19. Vo, A.H., Nguyen, N., Nguyen, T., Dang, T., Nguyen, D., Nguyen, B.T.: An automatic recommendation system for plant disease treatment. In: Huang, Y.P., Wang, W.J., Quoc, H.A., Le, H.G., Quach, H.N. (eds.) Computational Intelligence Methods for Green Technology and Sustainable Development. pp. 625–637. Springer International Publishing, Cham (2023)
20. Wang, L., Wong, A.: Diagnosis of covid-19 pneumonia using deep learning algorithms on chest x-ray images. Clinical Imaging **72**, 76–83 (2021)
21. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017). https://doi.org/10.1109/cvpr.2017.369
22. Wang, X., Peng, Y., Lu, L., Lu, S., Bagheri, M., Summers, R.M.: Chest x-ray image classification using deep convolutional neural networks. Journal of thoracic imaging **36**(4), 233–241 (2021)