Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2024, 2024:68

https://doi.org/10.28919/cmbn/8577

ISSN: 2052-2541

FOCUSING ON CORRECT REGIONS: SELF-SUPERVISED PRE-TRAINING IN LUNG DISEASE CLASSIFICATION

ANDREW TANUBRATA[1,*], GREGORIUS NATANAEL ELWIREHARDJA[1,2], BENS PARDAMEAN[2,3]

[1]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, 11480, Indonesia

[2]Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, 11480, Indonesia

[3]Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, 11480, Indonesia

Abstract. While the current state of lung disease detection using AI relies heavily on specific data types, limiting its real-world applicability, this work explores leveraging Transfer Learning (TL) with VicReg for improved performance. By using a public Chest X-ray dataset, the proposed model employs a ResNet50 model architecture that seamlessly integrates transfer learning and fine-tuned self-supervised Convolutional Neural Networks(CNNs). Can Artificial Intelligence (AI) for lung disease detection be improved to work across different types of medical images? This study addresses this challenge by proposing DOCTOR, a reusable AI model that leverages transfer learning and fine-tuned CNNs. DOCTOR is trained on chest X-rays but designed to be adaptable to other radiology images like CT scans. The results obtained from this proposed model are remarkable, achieving an impressive accuracy rate of97.37%, sensitivity of 96.30, specificity of 97.30% each, and precision of 96.30%. These exceptional performancemetrics demonstrate the proposed model's exceptional competence and efficacy in accurately detecting lung dis- eases. The trained CNN model utilized a ResNet50 backbone pre-trained using VicReg for robust lung disease detection across various modalities, which is referred to in this paper as DOCTOR.

Keywords: lung disease; artificial intelligence; transfer learning; data augmentation; deep learning.

2020 AMS Subject Classification: 68T07, 68T10.

---

*Corresponding author

E-mail address: andrew.tanubrata@binus.ac.id

Received April 01, 2024

1. INTRODUCTION

The timely identification of pulmonary ailments is imperative for efficient intervention and reducing the impact of the disease. Early diagnosis of Chronic Obstructive Pulmonary Disease (COPD), for instance, can slow lung function decline by up to 50%, while early interventionin pulmonary fibrosis can improve quality of life and extend lifespan by several years. Nev- ertheless, achieving precise and prompt diagnosis frequently proves to be a hurdle, especially in locations with limited resources [1]. Chest radiographs, which are easily accessible and cost-effective, play a crucial part in the preliminary diagnosis. However, the interpretation of minor abnormalities within these images necessitates substantial proficiency and is susceptibleto human fallibility.

While current strategies for detecting lung diseases, such as the utilization of chest X-rays, play a pivotal role, they are hindered by inherent limitations. Specifically, their precision dimin-ishes during the initial stages, resulting in the misinterpretation of harmless shadows as malig-nancies. Additionally, these strategies heavily rely on subjective human interpretation, which introduces inconsistency and disparities among healthcare professionals. Moreover, radiolo- gists are burdened with an immense workload, which may potentially impact the promptness and accuracy of their diagnoses. These limitations underscore the pressing necessity for more dependable and objective approaches to detect lung diseases, particularly in their early phases.Even so, Artificial Intelligence (AI) has emerged as a formidable instrument in tackling the aforementioned challenges. Notably, deep learning methodologies, specifically ConvolutionalNeural Networks (CNNs), have exhibited noteworthy aptitudes in the realm of medical image analysis. The capacity of CNNs to discern intricate characteristics and discriminate subtle pat-terns presents a compelling prospect for enhancing the precision and efficiency of lung diseasedetection.

Moreover, acquiring high quantities of labeled chest X-rays to train AI models can present certain difficulties; however, the utilization of Self-Supervised Learning (SSL) uncovers a mul- titude of possibilities. Due to its robustness in handling data imbalance, SSL is often utilized for Transfer-Learning (TL) [3]. The motivation behind the creation of the proposed model, in thiscase, DOCTOR, is to leverage the complementary information gleaned from multimodal data for more accurate disease identification and improved decision-making. Additionally, it servesas a platform for testing and advancing the boundaries of current knowledge in Artificial Intel-ligence (AI) within the healthcare domain. By effectively scrutinizing the immense abundance of unlabeled images, SSL algorithms can progressively enhance their capabilities, even in the absence of

explicit labels [3]. This presents opportunities for extraordinary advancements, suchas enhancing the accuracy of diagnostic procedures for lung diseases in their early stages, en- riching the capabilities of existing AI tools employed in chest X-ray analysis, and even pavingthe way for innovative applications like the monitoring of disease progression and the provi- sion of personalized medicine. By harnessing the knowledge embedded within unlabeled data, SSL provides a future for the improvement of lung disease diagnosis and, ultimately, the care provided to patients.

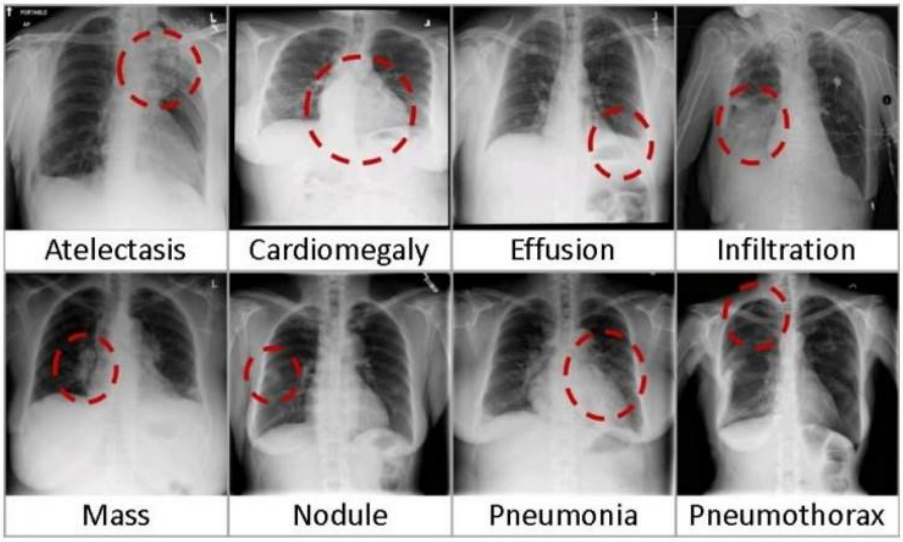

FIGURE 1. Eight Common Thoracic Diseases [2]

2. RELATED WORKS

Recent advancements in Artificial Intelligence (AI) have brought about a new era in the di- agnosis and prognosis of pulmonary diseases. In the early studies, Wang et al [2] conducteda significant study that employed a CNN to categorize chest X-rays into three distinct groups (see Figure 1): normal pneumonia, and tuberculosis where its model achieved an accuracy of 92.6% based on a dataset comprising 1,387 chest X-rays. In the domain of tuberculosis, recent advancements in deep learning for tuberculosis screening, as reviewed by Santosh et al [4], have demonstrated significant progress in the automated analysis of chest X-ray images, leading to enhanced precision in disease detection. Likewise, Chen et al [5] proved the capability of deep learning models, particularly TransUNet, which incorporates channel attention mechanism, thereby enhancing the diagnostic process for lung cancer. In the context of advanced diagnostic methods, Serte and Demirel [6] introduced an innovative AI system designed to enhance the

accuracy of COVID-19 detection through the analysis of 3D CT Scan images. This trend extends to Chronic Obstructive Pulmonary Disease (COPD), where Kumar et al [7] developed a novel multimodal framework that integrates CT scan images and lung sound samples for early diagnosis and accurate prediction of COPD. As the role of AI in pulmonary medicine continues to grow, Jalaber et al [8] highlight the essential role of chest CT examinations in COVID-19 patient triage and underscore the potential of AI in assisting radiologists with diagnosis and prognosis evaluation. Collectively, these endeavors demonstrate the immense potential of AI in revolutionizing the management of pulmonary diseases, spanning for early detection and precise diagnosis to personalized risk assessment and targeted interventions.

AI is rapidly revolutionizing the field of lung disease diagnosis, as evidenced by the remark-able findings of Wang et al [2] in their study, where they achieved an exceptional F1 score of 92.5% in the detection of pneumonia. Similarly, recent advancements in deep learning for tuberculosis screening have shown significant progress, with studies utilizing CNNs to analyze chest X-ray images, leading to promising methods and challenges in the accurate detection of TB-related abnormalities. Furthermore, Chen et al [4] demonstrated the effectiveness of AI in segmenting lung nodules for the early detection of lung cancer, with TransUNet achieving a DICE index of up to 0.887, indicating a high level of accuracy in the analysis. These studies collectively highlight the tremendous potential of AI in classifying chest X-rays, detecting early-stage lung cancer, and guiding personalized treatment. The integration of AI into the field of lung diseasediagnosis holds great promise for expediting the diagnostic process, enhancing the detection of abnormalities, and ultimately improving patient outcomes. This trend encompasses a wide range of lung diseases, including pneumonia, tuberculosis, lung cancer, and COVID-19.

The ongoing exploration of deep learning frameworks, explainable AI techniques, and the integration of multimodal data underscores the immense promise of AI in revolutionizing lung disease diagnosis and enhancing patient outcomes. Additionally, the usage of DL models trained using SSL for TL purposes can bring out significantly better results in cases wherethe used dataset is small or imbalanced [3]. Observing this wave of innovation, the authors feel compelled to contribute to this burgeoning field by developing a unique model. The di- verse challenges and ever-evolving nature of AI in lung disease detection necessitate ongoing creativity and determination.

3. METHODOLOGY

3.1. Dataset Discussion. This study utilized two publicly accessible datasets of chest X-ray images, namely the NIH Chest X-ray dataset (CXR-8) [9] and the Open-i dataset [10] (as detailed in Table 1). The CXR-8 dataset encompasses an extensive collection of more than 112,000 chest X-ray imagesderived from over 30,000 distinct patients, exhibiting a diverse range of lung diseases and abnor- malities. Similarly, the Open-i dataset comprises nearly 2,000 chest X-ray images from unique patients. Both datasets are meticulously annotated to encompass various clinical observations, encompassing the presence, type, and severity of lung diseases. For training, a portion of the CXR-8 dataset was carefully chosen for the purpose of training, with the distribution of 80% for training, 10% for validation, and 10% for testing. On the other hand, the Open-i dataset wasemployed as the validation dataset. The training dataset was meticulously curated and balanced to ensure its representativeness of the real-world population as well as to guarantee an adequate number of images for each lung disease and abnormality.

TABLE 1. Data distribution for each class in the Open-i (Testing Dataset) [9]and CXR-8 Dataset (Training Dataset) [10]

Directory	**Open-i**	**CXR-8**
Atelectasis	315	5,834
Cardiomegaly	345	9,298
Effusion	153	7,820
Infiltration	60	16,047
Mass	15	12,492
Nodule	106	25,259
Pneumonia	40	28,697
Pneumothorax	22	1,933
Normal	1,379	34,436

3.2. DOCTOR. The capacity of DOCTOR to analyze various types of data and excel in the detection of lung diseases is rooted in its innovative integration of state-of-the-art artificial intelligence techniques. At the core of this integration lies a pre-trained ResNet50 architecture. This modified ResNet backbone serves as the fundamental basis for extracting meaningful representations from a wide range of data modalities. To further enhance the learning process of DOCTOR, the incorporation of the VicReg algorithm plays a crucial role. VicReg leverages a regularization approach based on variance, invariance, and covariance, which guides the model towards the discovery of robust and informative features [11]. This approach facilitates a comprehensive understanding of patterns in lung scans, thereby enabling DOCTOR to discern subtle variations associated with different lung diseases. The combined strength of pre-trained feature extraction and the regularization provided by VicReg enables DOCTOR to surpass the limita- tions of models that rely on a single modality, resulting in superior accuracy and generalizabilityin the detection of lung diseases.

3.3. Training Configurations. The configuration of the model underwent a series of iterative experiments to optimize its performance. Initially, the trials focused on selectively freezing certain layers for transfer learning. However, in the final configuration, all layers within the architecture are unfrozen, allowing for a thorough fine-tuning process with a learning rate of
0.001 and momentum of 0.9 for the task of lung disease detection. In Addition, a total of100 epochs were implemented in DOCTOR for its training and testing setup. DOCTOR also utilizes the VicReg algorithm, which is a self-supervised learning approach that promotes the development of robust feature representations by imposing constraints on variance, invariance,and covariance on the learned embeddings. From a mathematical perspective, VicReg mini- mizes a loss function that consists of a contrastive term for aligning positive pairs, as well as two regularization terms: one that ensures each embedding dimension has a variance above 0.1and another that decorrelated embedding pairs to reduce redundancy [11] (as shown in Figure 2 and Figure 3). The combination ofcomprehensive fine-tuning and the regularization strategies employed by VicReg facilitates the extraction of informative and discriminative features to classify lung diseases.

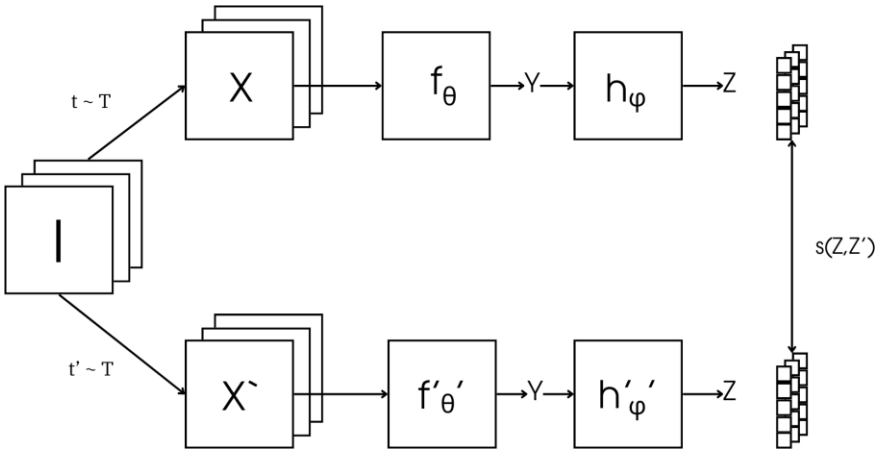

FIGURE 2. VicReg Algorithm: Variance-Invariance-Covariance Regularizationfor Self-Supervised Learning [11]

$$Variance\ regularization \mid y = f_\theta(x),\ and\ y' = f_\theta(x')$$
$$Invariance\ Regularization \mid z = h_\phi(y)\ and\ z' = h_\phi(y')$$
$$CovarianceRegularization \mid Z = [Z_1\ ,\ \ldots\ ,\ Z_n]\ \epsilon\ IR^{n \times d}$$

FIGURE 3. VicReg Algorithm Formula[11]

4. RESULTS

4.1. Model Evaluation. The performance of the fine-tuned models were evaluated on the testing set, and metrics such as precision (P), recall (R), and F1-Score (F) were employed. These metrics can be calculated by leveraging the quantities of True Positive (TP), False Positive (FP),and False Negative (FN), as illustrated in the following equation:

(1) $$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

(2) $$Sensitivity(Recall) = \frac{TP}{TP + FN}$$

(3) $$Specificity = \frac{TN}{TN + FP}$$

(4) $$Precision = \frac{TP}{TP + FP}$$

(5) $$F1\_Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The evaluation of the distinct capabilities possessed by each class was conducted through the utilization of a Confusion Matrix, which is a statistical tool that quantifies the accuracy of a classification model. This matrix provides a comprehensive summary of the model's performance by comparing the actual and predicted class labels. The Confusion Matrix offers valuable insights into the true positives, true negatives, false positives, and false negatives, al-lowing for a thorough analysis of the model's performance. Furthermore, the assessment of the model's performance is also facilitated by the Receiver Operating Characteristics (ROC) curve. This graphical representation showcases the trade-off between the true positive rate and the false positive rate at various classification thresholds. The ROC curve provides a visual interpretation of the model's performance and helps in determining the optimal threshold for classification. In this study, the Confusion Matrix can be observed in Figure 5, while Figure 6 depicts the presentation of the ROC curve, which serves as an important tool for evaluating theperformance of the DOCTOR model.

Testing Set | OpenL Dataset										
TARGET / UTPUT	Atelectasis	Cardiomegaly	Effusion	Infiltration	Mass	Nodule	Pneumonia	Pneumothorax	Normal	SUM
Atelectasis	308 12.65%	1 0.04%	0 0.00%	1 0.04%	2 0.08%	1 0.04%	0 0.00%	2 0.08%	0 0.00%	315 97.78% 2.22%
Cardiomegaly	0 0.00%	342 14.05%	2 0.08%	0 0.00%	1 0.04%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	345 99.13% 0.87%
Effusion	0 0.00%	0 0.00%	150 6.16%	0 0.00%	0 0.00%	0 0.00%	2 0.08%	1 0.04%	0 0.00%	153 98.04% 1.96%
Infiltration	0 0.00%	0 0.00%	0 0.00%	53 2.18%	0 0.00%	1 0.04%	5 0.21%	1 0.04%	0 0.00%	60 88.33% 11.67%
Mass	0 0.00%	0 0.00%	0 0.00%	0 0.00%	10 0.41%	5 0.21%	0 0.00%	0 0.00%	0 0.00%	15 66.67% 33.33%
Nodule	0 0.00%	0 0.00%	0 0.00%	0 0.00%	6 0.25%	100 4.11%	0 0.00%	0 0.00%	0 0.00%	106 94.34% 5.66%
Pneumonia	0 0.00%	0 0.00%	0 0.00%	2 0.08%	0 0.00%	0 0.00%	38 1.56%	0 0.00%	0 0.00%	40 95.00% 5.00%
Pneumothorax	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	2 0.08%	20 0.82%	0 0.00%	22 90.91% 9.09%
Normal	0 0.00%	29 1.19%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	1350 55.44%	1379 97.90% 2.10%
SUM	308 100.00% 0.00%	372 91.94% 8.06%	152 98.68% 1.32%	56 94.64% 5.36%	19 52.63% 47.37%	107 93.46% 6.54%	47 80.85% 19.15%	24 83.33% 16.67%	1350 100.00% 0.00%	2371 / 2435 97.37% 2.63%

FIGURE 4. The Confusion Matrix of DOCTOR on the testing set using theOpen-i Dataset [9]

As seen above in Figure 4, The depiction of the confusion matrix denotes the efficacy of DOCTOR in identifying various ailments of the lungs. DOCTOR exhibits a consistently high level of precision, as the majority of diseases showcase a precision value exceeding 0.9. The most

noteworthy precision value is observed for the category labeled as “Normal” at 0.979, closely followed by “Atelectasis” at 0.978 and “Cardiomegaly” at 0.991. Conversely, the cat- egory labeled “Mass” exhibits the lowest precision value at 0.667 meaning it is still not ableto confidently diagnose the mass disease. The overall accuracy of the model is quantified as 0.9737.

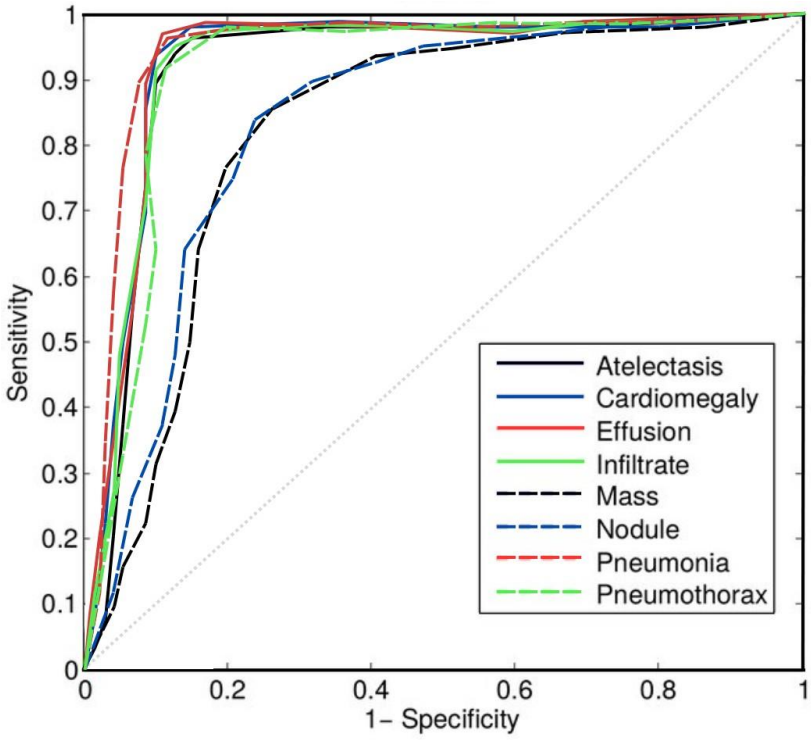

FIGURE 5. DOCTOR’s ROC performance on Open-i Dataset with Pre-trainedCustomized CNN

On the other hand, DOCTOR’s ROC Curve in Figure 5 indicates the performance of DOC- TOR at all classification thresholds. It plots the True Positive Rate (TPR) on the y-axis and theFalse Positive Rate (FPR) on the x-axis. The TPR is the proportion of positive cases that are correctly classified, and the FPR is the proportion of negative cases that are incorrectly classi-fied as positive. The area under the ROC curve (AUC) is a measure of the overall performanceof a model. An AUC of 1 represents perfect classification, while an AUC of 0.5 represents random guessing [12]. The AUC of DOCTOR attains a desirable value of 0.98, signifying its exceptional performance in correctly discerning between lung disease cases and normal cases.This remarkable AUC value is a result of DOCTOR’s ability to identify 6 out of 8 diseases (Atelectasis, Cardiomegaly, Effusion, Infiltrate, Pneumonia, Pneumothorax) with greater preci-sion. Such an achievement underscores the proficiency of DOCTOR in differentiating betweenvarious pathological conditions in the lungs and cases that are considered normal, thereby con- solidating its reputation as an effective

diagnostic tool.

4.2. Loss Plots. Loss Plots serve as a valuable tool for researchers, enabling them to compare and analyze the loss landscapes of multiple models. Moreover, these plots also allow for the comparison of different loss landscape projections, providing a comprehensive understanding of the various models under consideration. By utilizing Loss Plots, researchers can effectively construct more accurate loss projections, thereby enhancing their ability to gain valuable in- sights into the impact of training parameters and architectural choices on these projections [13].

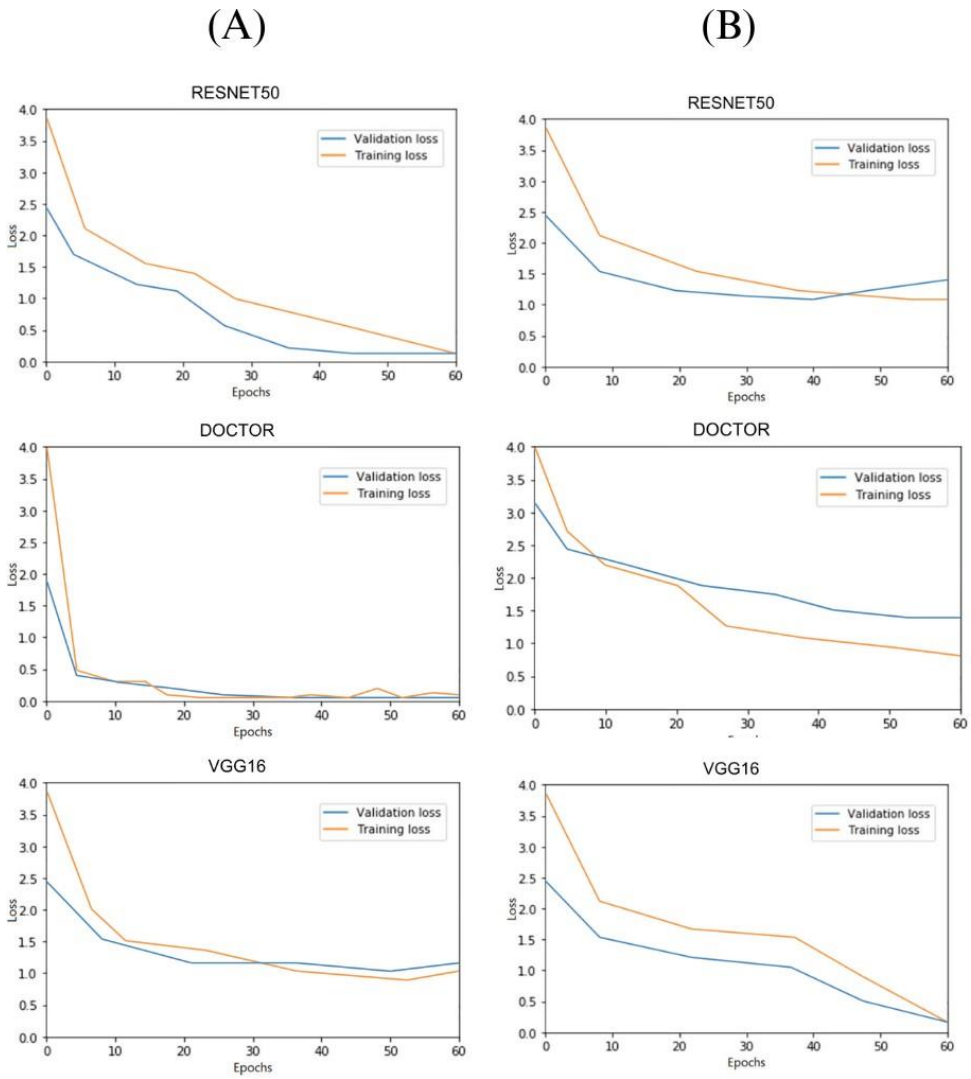

FIGURE 6. The training and validation loss plots of the three models: (A) loss plots that belong to three different types of models (ResNet50, DOCTOR, and VGG16) with no layers frozen and (B) loss plots belonging to three models butwith all layers frozen.

This particular section of the research paper delves into a comprehensive and thorough anal-ysis of the loss plots of three different models (ResNet50, VGG16, and DOCTOR) that can beobserved within these loss plots. These trajectories, in turn, offer invaluable insights and pro- vide a deeper understanding of the training process employed by the three models. In the initialconfiguration, the models are allowed to adapt to all the layers, leading to a gradual decrease inboth training and validation losses. This gradual decrease in losses demonstrates a consistent training and steady learning process. Furthermore, it is worth noting that the two model losses(ResNet50 and the VGG16) eventually stabilize at a relatively low value at around 0.7, while DOCTOR stabilizes below 0.5. This indicates that DOCTOR's trend of stabilization at a low loss value indicates successful learning, as it suggests that the model has not been subject to overfitting during the training process. This finding highlights the efficacy and effectiveness of the training process for the three models. The meticulous analysis of these loss plots con- tributes significantly to the understanding of the training process and its impact on the model'sperformance.

On the contrary, the second configuration encompasses the act of freezing every layer of the three models. At the beginning stage, the configuration's loss values exhibit a consistent de- crease. Nevertheless, the validation loss swiftly deviates from this consistent trend, insinuatingthat the layers that have been frozen cannot effectively adjust to the training data.

TABLE 2. Comparison Table of DOCTOR, Pre-trained ResNet50, and the VGG16's result on the Open-i Testing Dataset [9] and the CXR-8 Dataset[10]

Method	**DOCTOR**		**Pre-trained ResNet50**		**Pre-trained VGG16**	
Image Dataset	CXR-8	Open-i	CXR-8	Open-i	CXR-8	Open-i
Accuracy %	96.30	**97.37**	96.40	96.70	95.10	95.30
Sensitivity %	95.30	**96.30**	96.20	96.20	95.50	95.50
Specificity %	95.30	**97.30**	95.60	96.10	95.50	96.40
Precision %	94.60	**96.30**	95.70	95.80	95.10	95.35
F1 Score %	94.80	**98.80**	96.00	96.00	93.35	95.15

In Table 2, it can be seen that DOCTOR outperformed the pre-trained ResNet-50 and VGG16 model on all metrics on the Open-i testing dataset where configurations of each model with no layers frozen, with the highest accuracy of 97.37% indicating DOCTOR's ability to correctly classify most classes, Sensitivity of 96.30% where it correctly identified a very high true posi-tive cases which is crucial for avoiding missed diagnoses, Specificity 97.30% where it demon-strates

the effectiveness in correctly identifying true negatives to minimize false alarms, Preci-sion 96.30% to ensure high confidence in positive diagnoses, and F1-Score 98.8% reflecting astrong balance between precision and recall to signify the overall reliability of DOCTOR. Theresults in the table imply DOCTOR is a promising tool for medical image classification tasks within the domain of the Open-i dataset. DOCTOR's strong performance stems from its archi-tecture, where it builds upon the strong feature extraction capabilities of a pre-trained ResNet50and an added customized block tailored to the Open-i dataset. This block constraints additionalconvolutional layers carefully tuned to extract even more nuanced and domain-relevant featuresfrom the images. This allows DOCTOR to better discriminate between healthy and diseased cases, ultimately leading to its superior accuracy, sensitivity, and overall reliability.

4.3. DOCTOR Visualizations. To further assess the model's performance, gradCAM was also used to verify whether the model focused on the correct regions of the image. In the present study, DOCTOR produces highly informative visual representations of the thorax and pulmonary region. Within Figure 7, a detailed and all-encompassing examination is presented, which serves to highlight and present various aspects.

GradCAM is a methodology that can be effectively employed in order to gain insight into the specific aspects of attention that a deep neural network is focusing on during the decision- making process. In the present scenario, the utilization of GradCAM visualizations serves the purpose of showcasing the specific regions of the Chest X-ray DOCTOR that are being utilizedto determine the presence or absence of a lung disease [14].

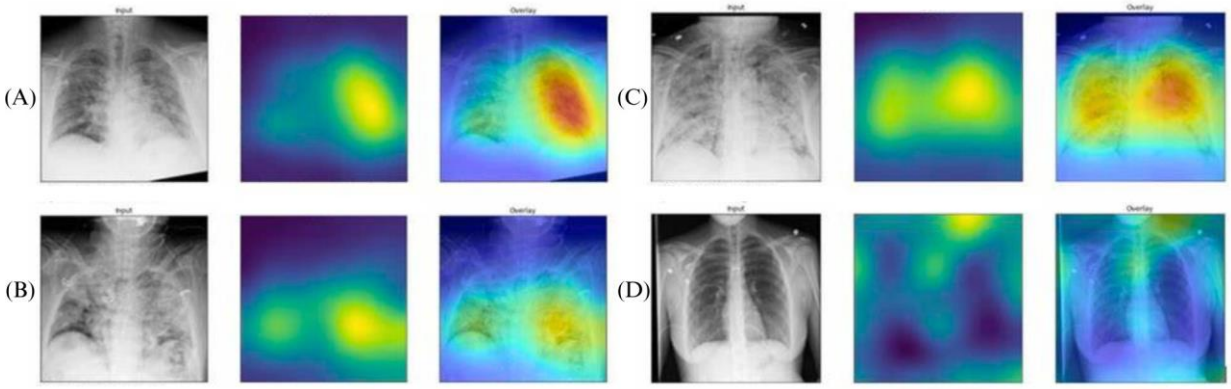

FIGURE 7. 4 GradCam to visualize DOCTOR's output on Lung Disease Detection. Visualization (7A), (7B), and (7C) show a possibility of lung anomaly by DOCTOR, whereas Visualization (7D) indicates a healthy andnormal lung.

Visualizations (7A), (7B), and (7C) exhibit regions of the pulmonary system that have been accentuated by the GradCAM, thereby signifying that the medical professional directs its atten-tion towards these specific regions in order to ascertain the presence or absence of a pulmonaryailment. In these instances, the medical professional has accurately detected anomalies within the pulmonary system. The visualization denoted as (7D) portrays a radiographic image of thethoracic cavity that does not display any regions that have been accentuated by the GradCAM.This observation implies that the medical professional does not allocate its attention to any par-ticular segment of the pulmonary system when making a determination regarding the presenceor absence of a pulmonary ailment. In this specific case, the medical professional ascertained the absence of any pulmonary ailment. The absence of the GradCAM-highlighted regions, which are observable on the pulmonary system in Visualization (7D), indeed suggests that the modeldid not detect any irregularities within these particular areas. This corroborates the notion that the medical professional can accurately identify the absence of pulmonary ailments.

4.4. Discussion. While DOCTOR attains a remarkably commendable area under the curve (AUC) value, it is important to note that there are still possibilities for further optimizations. Specifically, in the case of disease classifications such as Mass and Nodule, there exists room for improvement where DOCTOR could enhance its performance. By addressing these limitations,it is plausible to envision a scenario where DOCTOR's AUC value could be elevated to a higherlevel. The realization of such advancements would not only serve to consolidate the already impressive performance of DOCTOR but also open up avenues for the development of even more precise and robust solutions in future studies.

Through the evaluation conducted by the medical professional, it has been observed that the utilization of transfer learning has experienced a substantial augmentation in its level of ac-ceptance within the realm of medical imaging, specifically in the domain of identifying lung ailments [15]. Transfer learning encompasses the utilization of a model that has undergone ex-tensive training with a vast dataset consisting of images to address specific tasks. This particular approach has proven to be highly efficacious as it not only saves a considerable amount of timebut also mitigates the effort that would otherwise be required to train models from scratch [16].Overall, the training procedure of both the VGG and ResNet models was executed with re- markable smoothness, owing to their well-established and pre-trained nature. These modelsdisplayed an impressive capacity to acquire the distinctive characteristics observed in lung im-ages, resulting in

commendable levels of accuracy when tested on the training dataset. On thecontrary, the training process of the DOCTOR model encountered more obstacles, primarilydue to its customized nature. The successful training of this model necessitated the explorationof various code and hyperparameter configurations, as well as the need to rectify numerous errors that were encountered along the way. A notable hurdle in training DOCTOR was its heightened susceptibility to overfitting. Overfitting is a problem that occurs when a machinelearning model learns the training data too well, and as a result, it is unable to generalize to new data [17]. To address this issue, Several trials of different approaches were done such as:

- Regularization: This technique penalizes complex models, discouraging them from memorizing the training data and promoting simpler, more generalizable models [18]. A technique called Early Stopping terminates training before the model memorizes the training data. Monitoring validation errors during training can indicate the optimal stop-ping point [19].
- Data Augmentation: This artificially increases the diversity of the training data, re-ducing the dependence on the specific examples presented during training. Geometric transformations were used to randomly rotate, crop, or flip images to create new training samples without requiring additional data collection [20].
- Ensemble Methods: Combining predictions from multiple, diverse models often leadsto more robust and generalizable results compared to a single model [21].

During DOCTOR's training, the medical practitioner encountered an additional obstacle. Specifically, a frequent occurrence of errors transpired wherein the image analysis software misinterpreted the anatomical structure of the shoulder as that of the lung within chest images.The rationale behind this misinterpretation was attributable to the similarities in texture and shape exhibited by both the shoulder and lung within certain images. Some measures taken to address these issues include:

- Increasing the size and diversity of the training dataset: This phenomenon may be at-tributed to the reduced overfitting caused by the inclusion of larger and more diverse training samples [17]. In other words, the huge amount of images exposed the model toidentify and extract the key features that distinguish these structures.
- Adjusting the hyperparameters of the model: Adjusted the hyperparameters of the model,such as the learning rate and the optimizer [22]. Using a lower learning rate and a more robust optimizer helped improve the model's ability to distinguish between shoulder and lung.

- Added penalties on the model's output for misclassifying: Added a penalty to themodel's output for misclassifying shoulders as lungs. This helps to encourage the model to focus on learning the features that are specific to the lung.

Current AI models that are currently being used to detect lung diseases often face a consid- erable predicament in their inclination to rely heavily on a sole modality, which can be either Chest X-rays or CT scans. This reliance on a single modality inevitably imposes limitations on the accuracy and generalizability of these models, thereby hindering their ability to effectively detect and diagnose lung diseases.

TABLE 3. Comparison Table between result methods from other journals andDOCTOR

Method	**Accuracy %**	**Sensitivity (Recall) %**	**Specificity %**	**Precision %**
A.S. et al. [23]	98.60	98.40	98.50	98.60
Ucar et al. [24]	98.60	-	99.10	-
Nour et al. [25]	98.97	89.39	99.75	-
Wang et al. [26]	93.40	93.30	95.76	-
Roy, and Das. [27]	97.80	-	-	-
Zahid et al. [28]	97.22	96.87	99.12	95.54
DOCTOR (Proposed Method)	97.37	96.30	97.30	96.30

In Table 3, It can be analyzed that DOCTOR achieves a competitive accuracy of 97.37%, comparable to other recent works in the literature. Notably, Zahid et al. [28] achieves the high-est sensitivity of 99.12% but at the cost of lower accuracy (97.22%). Conversely, Nour et al. [25] has the highest accuracy of 98.97% but suffers from lower sensitivity (89.39%). DOC- TOR's balanced performance across all metrics suggests it's a promising approach. However, some methods lack data for certain metrics, making direct comparisons difficult. To solidify these initial findings, further investigation into missing values and the methods' performance on specific datasets is warranted.

5. CONCLUSION

While several cutting-edge methodologies claim to possess a high degree of accuracy, these methods often demonstrate a compromise between sensitivity and specificity or lack critical evaluation metrics. However, the proposed model, known as DOCTOR, successfully overcomes these limitations and achieves a comparable outcome akin to the methodologies put forth by Za-hid et al. [28] and Roy, and Das. [27]. It is worth noting that DOCTOR consistently maintains an exceptional level of precision, surpassing the performance of established competitors as pro-posed by Zahid et al. [28]. It can be inferred that one major advantage of DOCTOR is its correct Rol identification as visualized by the GradCAM, also its very strong results on its F1-score at 98.80% support the claim that DOCTOR is highly reliable in detecting lung anomalies.This well-balanced performance opens up avenues for earlier diagnoses, improved treatment planning, and the potential for wider applicability across various lung diseases. Additionally, computer-based diagnostic systems can achieve high accuracy and consistency, which further emphasizes the need for robust solutions like DOCTOR. Consequently, there arises a strong desire for a dedicated system that can aid radiologists in managing their workload, particularlythe one that leverages the advanced capabilities of deep learning models [29,30]. In the upcominginvestigation, despite the success of the dual dataset technique, a more sophisticated deep learn- ing technique will be introduced for the identification of lung diseases, in which a larger numberof datasets will be gathered to enhance effectiveness. The extent of deep learning achievement is significantly impacted by the abundance of available data [31,32].

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] D.M. Parkin, F. Bray, J. Ferlay, et al. Estimating the world cancer burden: Globocan 2000, Int. J. Cancer 94 (2001), 153–156. https://doi.org/10.1002/ijc.1440.

[2] X. Wang, Y. Peng, L. Lu, et al. ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2097-2106, 2017.

[3] S. Elvan, Y.P. Aedentrisa, G.N. Elwirehardja, et al. Evaluating self-supervised pre-trained vision transformers on imbalanced data for lung disease classification, ICIC Express Lett. Part B: Appl. 15 (2023), 83-89. https://doi.org/10.24507/icicelb.15.01.83.

[4] K. Santosh, S. Allu, S. Rajaraman, et al. Advances in deep learning for tuberculosis screening using chest X-rays: The last 5 years review, J. Med. Syst. 46 (2022), 82. https://doi.org/10.1007/s10916-022-01870-8.

[5] W. Chen, Y. Wang, D. Tian, et al. CT lung nodule segmentation: A comparative study of data preprocessing and deep learning models, IEEE Access 11 (2023), 34925–34931. https://doi.org/10.1109/access.2023.3265170.

[6] S. Serte, H. Demirel. Deep learning for diagnosis of COVID-19 using 3D CT scans, Computers Biol. Med. 132 (2021), 104306. https://doi.org/10.1016/j.compbiomed.2021.104306.

[7] S. Kumar, V. Bhagat, P. Sahu, et al. A novel multimodal framework for early diagnosis and classification of COPD based on CT scan images and multivariate pulmonary respiratory diseases, Computer Methods Progr. Biomed. 243 (2024), 107911. https://doi.org/10.1016/j.cmpb.2023.107911..

[8] C. Jalaber, T. Lapotre, T. Morcet-Delattre, et al. Chest CT in COVID-19 pneumonia: A review of current knowledge, Diagn. Interv. Imaging 101 (2020), 431-437. https://doi.org/10.1016/j.diii.2020.06.001.

[9] H.Q. Nguyen, K. Lam, L.T. Le, et al. VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations, Sci. Data 9 (2022), 429. https://doi.org/10.1038/s41597-022-01498-w.

[10] Raddar, Chest X-rays (Indiana University), Kaggle, 2020. https://www.kaggle.com/datasets/raddar/chest-xrays-indiana-university.

[11] R.E. Caraka, S. Shohaimi, I.D. Kurniawan, et al. Ecological show cave and wild cave: Negative binomial Gllvm's arthropod community modelling, Procedia Computer Sci. 135 (2018), 377-384. https://doi.org/10.1016/j.procs.2018.08.188.

[12] A.C.J.W. Janssens, F.K. Martens, Reflection on modern methods: Revisiting the area under the ROC curve, Int. J. Epidemiol. 49 (2020), 1397–1403. https://doi.org/10.1093/ije/dyz274.

[13] R. Bain, M. Tokarev, H. Kothari, et al. LossPlot: A better way to visualize loss landscapes, preprint, (2021). http://arxiv.org/abs/2111.15133.

[14] R.R. Selvaraju, M. Cogswell, A. Das, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618-626.

[15] S. Atasever, N. Azginoglu, D.S. Terzi, et al. A comprehensive survey of deep learning research on medical image analysis with focus on transfer learning, Clinic. Imaging 94 (2023), 18-41. https://doi.org/10.1016/j.clinimag.2022.11.003.

[16] M. Iman, H.R. Arabnia, K. Rasheed, A review of deep transfer learning and recent advancements, Technologies 11 (2023), 40. https://doi.org/10.3390/technologies11020040.

[17] X. Ying, An overview of overfitting and its solutions, J. Phys.: Conf. Ser. 1168 (2019), 022022. https://doi.org/10.1088/1742-6596/1168/2/022022.

[18] Y. Tian, Y. Zhang, A comprehensive survey on regularization strategies in machine learning, Inform. Fusion 80 (2022), 146-166. https://doi.org/10.1016/j.inffus.2021.11.005.

[19] L. Prechelt, Early Stopping — But When?, in: G. Montavon, G.B. Orr, K.-R. Müller (Eds.), Neural Networks: Tricks of the Trade, Springer, Berlin, 2012: pp. 53–67. https://doi.org/10.1007/978-3-642-35289-8_5.

[20] Philips, B. Pardamean, G.N. Elwirehardja, et al. A systematic literature review of deep learning application in

multiclass anomaly detection for chest medical imaging, in: 2023 6th International Conference of Computer and Informatics Engineering (IC2IE), IEEE, Lombok, Indonesia, 2023: pp. 1–6. https://doi.org/10.1109/IC2IE60547.2023.10330925.

[21] A. Mohammed, R. Kora, A comprehensive review on ensemble deep learning: Opportunities and challenges, J. King Saud Univ. - Computer Inform. Sci. 35 (2023), 757-774. https://doi.org/10.1016/j.jksuci.2023.01.014.

[22] T. Agrawal, Hyperparameter optimization in machine learning: Make your machine learning and deep learning models more efficient, Apress, Berkeley, CA, 2021. https://doi.org/10.1007/978-1-4842-6579-6.

[23] M.H. Al-Sheikh, O. Al Dandan, A.S. Al-Shamayleh, et al. Multi-class deep learning architecture for classifying lung diseases from chest X-Ray and CT images, Sci. Rep. 13 (2023), 19373. https://doi.org/10.1038/s41598-023-46147-3.

[24] F. Ucar, D. Korkmaz, COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images, Med. Hypoth. 140 (2020), 109761. https://doi.org/10.1016/j.mehy.2020.109761.

[25] M. Nour, Z. Cömert, K. Polat, A Novel Medical Diagnosis model for COVID-19 infection detection based on Deep Features and Bayesian Optimization, Appl. Soft Comput. 97 (2020), 106580. https://doi.org/10.1016/j.asoc.2020.106580.

[26] L. Wang, Z.Q. Lin, A. Wong, COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images, Sci. Rep. 10 (2020), 19549. https://doi.org/10.1038/s41598-020-76550-z.

[27] S. Roy, A.K. Das, Deep-CoV: An integrated deep learning model to detect COVID-19 using chest X-ray and CT images, Comput. Intell. 39 (2023), 369-400. https://doi.org/10.1111/coin.12568.

[28] Z. Ullah, M. Usman, S. Latif, et al. Densely attention mechanism based network for COVID-19 detection in chest X-rays, Sci. Rep. 13 (2023), 261. https://doi.org/10.1038/s41598-022-27266-9.

[29] Jimmy, T.W. Cenggoro, B. Pardamean, et al. Detection of pulmonary tuberculosis from chest X-Ray images using multimodal ensemble method, Commun. Math. Biol. Neurosci. 2022 (2022), 126. https://doi.org/10.28919/cmbn/7776.

[30] R.E. Caraka, N.T. Nugroho, S.K. Tai, et al. Feature importance of the aortic anatomy on endovascular aneurysm repair (EVAR) using Boruta and Bayesian MCMC, Commun. Math. Biol. Neurosci. 2020 (2020), 22. https://doi.org/10.28919/cmbn/4584.

[31] H.H. Muljo, B. Pardamean, K. Purwandari, et al. Improving lung disease detection by joint learning with COVID-19 radiography database, Commun. Math. Biol. Neurosci. 2022 (2022), 1. https://doi.org/10.28919/cmbn/6838.

[32] T.W. Cenggoro, B. Mahesworo, A. Budiarto, et al. Features importance in classification models for colorectal cancer cases phenotype in Indonesia, Procedia Computer Sci. 157 (2019), 313-320. https://doi.org/10.1016/j.procs.2019.08.172.