EAI Endorsed Transactions
on Industrial Networks and Intelligent Systems

Research Article/Editorial

Predicting the Severity of COVID-19 Pneumonia from Chest X-Ray Images: A Convolutional Neural Network Approach

Thien B. Nguyen-Tat[1,2,*], Viet-Trinh Tran-Thi[1,2], Vuong M. Ngo[3]

[1]University of Information Technology, Ho Chi Minh City, Vietnam
[2]Vietnam National University, Ho Chi Minh City, Vietnam
[3]Ho Chi Minh City Open University, Ho Chi Minh City, Vietnam

Abstract

This study addresses significant limitations of previous works based on the Brixia and COVIDGR datasets, which primarily provided qualitative lung injury scores and focused mainly on detecting mild and moderate cases. To bridge these critical gaps, we developed a unified and comprehensive analytical framework that accurately assesses COVID-19-induced lung injuries across four levels: Normal, Mild, Moderate, and Severe. This approach's core is a meticulously curated, balanced dataset comprising 9,294 high-quality chest X-ray images. Notably, this dataset has been made widely available to the research community, fostering collaborative efforts and enhancing the precision of lung injury classification at all severity levels. To validate the framework's effectiveness, we conducted an in-depth evaluation using advanced deep learning models, including VGG16, RegNet, DenseNet, MobileNet, EfficientNet, and Vision Transformer (ViT), on this dataset. The top-performing model was further enhanced by optimizing additional fully connected layers and adjusting weights, achieving an outstanding sensitivity of 94.38%. These results affirm the accuracy and reliability of the proposed solution and demonstrate its potential for broad application in clinical practice. Our study represents a significant step forward in developing AI-powered diagnostic tools, contributing to the timely and precise diagnosis of COVID-19 cases. Furthermore, our dataset and methodological framework hold the potential to serve as a foundation for future research, paving the way for advancements in the detection and classification of respiratory diseases with higher accuracy and efficiency.

Received on 03 06 2024; accepted on 18 11 2024; published on 25 11 2024

Keywords: Pneumonia Severity, COVID-19, Convolutional Neural Networks, Classification

doi:10.4108/eetinis.v12i1.6240

1. Introduction

COVID-19 is an infectious disease caused by the novel coronavirus strain known as SARS-CoV-2. The first human cases of COVID-19 were first reported in December 2019 [1]. Since then, it has erupted and spread at an astonishing rate worldwide. This pandemic not only poses a threat to human lives but also exerts immense pressure on healthcare systems, impacting every facet of daily life, the economy, and society at large. The rapid transmission of the virus has made controlling and containing its spread extremely challenging. In 2020, the exponential spread of COVID-19 compelled the World Health Organization (WHO) to declare it a global pandemic. This marked a significant shift in healthcare management procedures, requiring nations to implement new regulations to address the situation, including the deployment of infection control systems [2],[3]. COVID-19 has caused millions of confirmed cases and hundreds of thousands of deaths across the globe. The latest data from WHO [4] as of September 2023 reveals approximately 760 million confirmed cases of COVID-19 worldwide, with over 6.8 million fatalities. In the face of a pandemic that threatens both humanity and society, we all must come together and support one another. The efforts of

*Corresponding author. Email: thienntb@uit.edu.vn

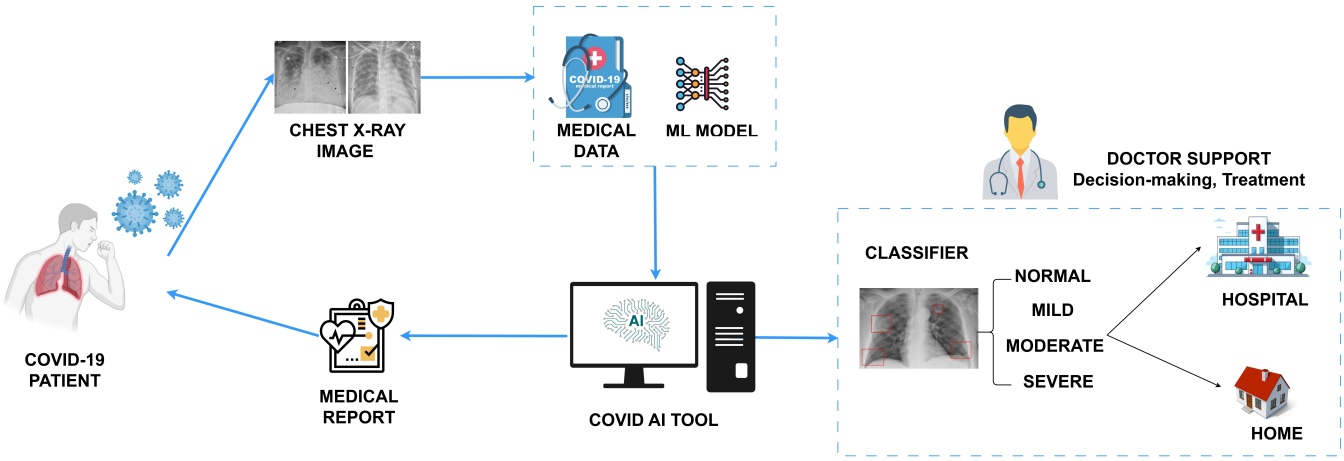

Figure 1. AI-Assisted Workflow for Diagnosing COVID-19-Induced Lung Damage.

healthcare professionals and medical experts, coupled with the use of technology and science to combat the disease, play a pivotal role in controlling and ultimately defeating COVID-19.

Most COVID-19 symptoms can manifest in various forms, ranging from fever, dry cough, and shortness of breath to loss of smell and taste. However, the most common and critical symptom of this virus is lung infection. In the field of COVID-19 diagnosis and monitoring, chest X-rays and CT scans are the most widely examined imaging methods [5]. To determine the status of a COVID-19 infection, doctors often must manually interpret these images. They must identify specific image patterns to confirm pneumonia caused by COVID-19. Manual interpretation can be time-consuming and labor-intensive, especially when dealing with a large volume of images. This presents a challenge in promptly and effectively tracking the patients' conditions. Furthermore, the diagnostic and treatment costs for COVID-19 are steadily increasing, placing a significant economic and financial burden on nations.

A survey conducted by X-ray doctors in Australia has revealed that mobile X-ray services, general X-ray scans, and CT scans are experiencing increased demand, adding to the strain on healthcare professionals and imaging diagnosticians [6]. This overload poses difficulties in patient care, prognosis, and disease control. In the context of limited healthcare facilities such as specialized care units and the shortage of mechanical ventilators for COVID-19 treatment, patient classification based on the severity of the disease becomes exceedingly crucial. The classification of COVID-19 patients into four severity levels: Normal, Mild, Moderate, and Severe, using X-ray images plays an essential and irreplaceable role in diagnosis and treatment. This not only helps identify the seriousness of the patient's condition but also determines whether they should be treated in the hospital or can isolate themselves at home. This plays a vital role in preventing the spread of the virus and ensuring that the healthcare system does not become overwhelmed.

In the healthcare field, Machine Learning (ML) is widely used because of the computer's ability to assist experts in performing specific tasks [7], [8], [9]. For most ML models, collecting X-ray or chest CT data from patients is essential since this data is the primary input for training and building models. Furthermore, establishing a large and diverse database is imperative to ensure an adequate amount of data during the training of high-performing machine learning models. This becomes particularly critical when we aim to develop models that can generalize across various situations and disease severity levels. Thus, these models provide robust decision support tools for the medical and healthcare sectors.

This study provides key advancements in the automated diagnosis and severity assessment of COVID-19-induced lung damage using chest X-ray images. The workflow illustrated in Figure 1 demonstrates the integration of AI-driven analysis in diagnosing COVID-19-induced lung damage, highlighting the sequential steps from patient chest X-ray imaging to AI-based severity classification and medical decision-making for hospitalization or home treatment. Our main contributions are as follows:

- **Development of a Novel and Balanced Dataset:** We constructed a comprehensive dataset of 9,294 high-quality chest X-ray images categorized into four severity levels: Normal, Mild, Moderate, and Severe. This dataset addresses limitations in existing datasets and is publicly

available (https://github.com/trinhttv/predicting-covid-19) to facilitate further research on COVID-19 diagnosis and severity classification.

- **Evaluation of Multiple Deep Learning Models:** We benchmarked the performance of six state-of-the-art deep learning architectures (VGG16, RegNet, DenseNet, MobileNet, EfficientNet, and ViT) on the newly developed dataset. This systematic comparison identified the most effective model for severity classification, demonstrating its applicability in real-world scenarios.
- **Proposal of an Enhanced Classification Model:** Building on the best-performing architecture, we proposed an optimized model with additional fully connected layers and weight adjustments. This enhanced model achieved high sensitivity (94.43%) and demonstrated its potential as a reliable decision-support tool in healthcare applications.
- **Comprehensive Validation and Reliability Assessment:** To ensure the generalizability of our approach, we conducted extensive cross-validation experiments. These evaluations provide solid evidence of the consistency of the proposed model and its ability to accurately classify the severity of COVID-19 lung damage.
- **Facilitation of Resource-Limited Healthcare Settings:** By introducing an automated framework for severity assessment, this study addresses the challenges posed by limited healthcare resources, particularly in managing COVID-19 cases. The proposed solution enables faster, cost-effective diagnosis and efficient patient management, alleviating the burden on overstrained medical facilities.

The rest of the paper is organized as follows: Section 2 will evaluate related works on the topic. Section 3 introduces the datasets and scoring systems for evaluating lung damage levels. The detailed methodology will be presented in Section 4. Section 5 presents experimental results and discusses the study. The conclusion and feature work of the paper will be presented in Section 6.

2. Related works

While there have been numerous studies on using deep learning for detecting and diagnosing COVID-19, most of them have focused on disease detection rather than assessing severity [10], [11],[12],[13],[14],[15],[16]. These studies have demonstrated the potential of employing CNN models for diagnosing COVID-19 from X-ray and CT chest images to identify cases of COVID-19 early or to prioritize COVID-19 testing when testing resources are limited [16].

Deep learning methods for COVID-19 detection have garnered significant attention; however, given the current state of disease control, evaluating the extent of lung damage caused by COVID-19 is crucial for devising appropriate treatment plans, monitoring the patient's condition, and assessing the risk of other diseases for the patient. Currently, the number of studies related to this topic is still limited and insufficient to draw accurate and comprehensive conclusions regarding the reliability and practical applicability of these methods in healthcare practice. In deep learning-based studies of COVID-19, some research has classified the severity of the disease into two groups: severe and non-severe.

Kelei He and colleagues proposed an automated method to classify COVID-19 patients into severe and non-severe groups with an accuracy of 98.5% [17]. Signoroni [18] developed the BS-Net model to predict the extent of lung damage in X-ray images, achieving higher accuracy and consistency compared to other methods. Jocelyn Zhu [19] utilized deep convolutional neural networks to classify the severity of COVID-19 lung disease from chest X-ray images. The results indicated that the MobileNet model achieved high accuracy in classifying chest X-ray images of COVID-19 patients into different severity groups.

Zhenyu Tangand and colleagues [20] employed chest CT images to assess severity and COVID-19-related characteristics. They developed a model based on 63 quantitative features from CT images, achieving an accuracy of 87.%. Another study [21], the authors developed a method to support COVID-19 diagnosis based on CT images, with an overall accuracy of 82%. Both methods can be valuable in assisting with the diagnosis and treatment of COVID-19 patients. In a study [22], Emrah Irmak presented a machine learning model to classify COVID-19 patients into four different severity levels based on X-ray images. The model achieved an average accuracy of 95.52% on a dataset of 3260 X-ray images evaluated by two radiologists. However, this dataset is not publicly available and needs validation on other datasets. This study proposes a useful approach to support the diagnosis and treatment of COVID-19 patients.

Recently, Zaid Albataineh and colleagues [23] introduced a severity classification system for COVID-19 into three stages: mild, moderate, and severe, based on CT images. This system utilizes segmentation methods and feature extraction from CT images for classification. The results show that the proposed model achieves high accuracy, ranging from 98.24% to 99.9%, depending on the stage of COVID-19 infection. Physician scoring is the standard approach for assessing

the severity of COVID-19 lung disease. However, studies have pointed out the limitations of this method, as it relies on either prognostic scoring of lung damage from CT or X-ray images of COVID-19 patients to predict the severity. This leads to time-consuming annotation for injuries, making the severity scoring method less efficient.

Efficiently, our study classifies lung damage into four stages: no damage, mild damage, moderate damage, and severe damage using advanced machine learning models. This method saves time and assists clinical physicians in dealing with severe cases of COVID-19 lung disease quickly and effectively. Furthermore, we used chest X-ray images as a common, relatively inexpensive, rapid, and accessible diagnostic method in our study. Although it has lower sensitivity compared to CT scans, chest X-ray images are still considered a primary imaging option for assessing the severity of COVID-19 and monitoring patients in many healthcare facilities. This will help improve access and diagnostic efficiency for COVID-19 patients while supporting cost and resource optimization in healthcare.

3. Dataset and scores

During the process of searching for appropriate datasets for classifying the severity of COVID-19 lung damage using CXR images, collecting, and utilizing suitable data sources may encounter difficulties due to variations in their availability and accessibility. In our study, we used Brixia [24] and COVIDGR [25] as two main datasets for classifying the severity of COVID-19 lung damage into different levels using CXR images. The Brixia dataset provides Brixia scores along with corresponding images, while COVIDGR includes severity labels based on the modified RALE index for quantifying COVID-19.

The scoring system used to evaluate the degree of lung damage is an important tool in diagnosing and monitoring respiratory diseases, including COVID-19. The Brixia and RALE scores are used to assess the extent of lung damage in hospitalized patients with COVID-19. In the early months of the pandemic, studies utilized pre-existing scoring systems such as the Radiographic Assessment of Lung Edema (RALE) to assess the severity of CXR. Therefore, the use of Brixia and RALE scores is an effective way to evaluate the extent of lung damage in hospitalized COVID-19 patients [26], [27], [28].

3.1. The Brixia dataset and Brixia score

The dataset was collected from two hospitals in the city of Brescia, Northern Italy, including 4,695 CXR images of COVID-19 patients. The data was collected using both CR and DX methods, with either AP or PA projection used. All the images in this dataset are

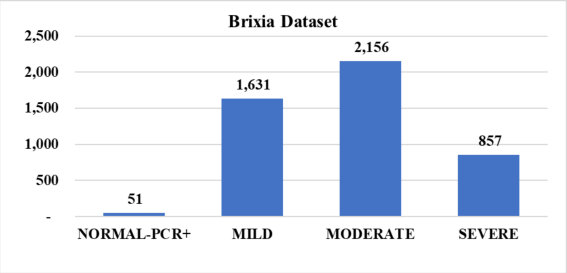

Figure 2. Samples in the Brixia dataset.

provided in the form of anonymized DICOM files and annotated with Brixia scores for the severity of lung injury and relevant metadata in CSV files. The use of two different hospitals in the data collection process increases the diversity of the dataset and improves the accuracy of the classification results. Figure 2 illustrates the distribution of samples in the Brixia dataset.

The Brixia dataset utilizes the Brixia scoring system to assess the extent of lung damage in COVID-19 patients using CXR images. This system was developed by researchers in Brescia, Italy. Scores are assigned to each imaging parameter, including white spot density, lung edema, airway cavities, pleural density, and pleural density. The total score of these parameters is calculated to evaluate the severity of COVID-19 lung disease. The Brixia severity scoring system [29] divides the lungs into 6 zones, 3 zones for each lung when observed in anteroposterior (AP) or posteroanterior (PA) view as shown in Figure 3(a). Specifically:

- Upper zones (A and D): located above the inferior border of the aortic arch.
- Middle zones (B and E): located below the inferior border of the aortic arch and above the inferior border of the right lower pulmonary vein (i.e., right hemidiaphragm).
- Lower zones (C and F): located below the inferior border of the right lower pulmonary vein (i.e., lung bases).

When anatomical landmarks cannot be accurately identified for technical reasons (such as X-rays taken at the bedside of critically ill patients), each lung can be divided into three equal zones.

In assessing the health status of COVID-19 patients, using a consistent scoring system to evaluate the characteristics and extent of lung abnormalities is crucial. Accordingly, six scores are used to assess the characteristics and extent of lung abnormalities, with scores ranging from 0 to 3 assigned to each zone, including:

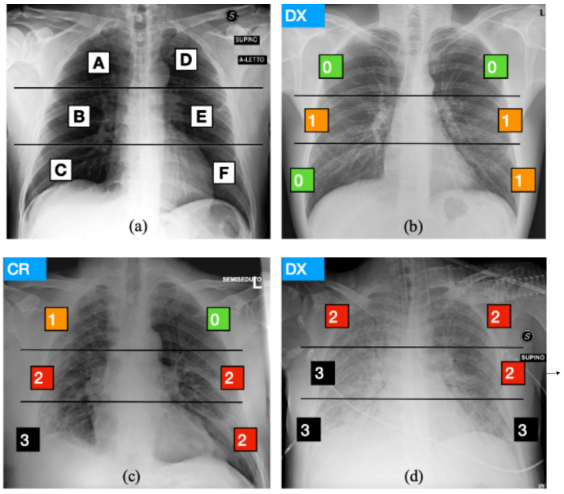

Figure 3. Classification method of lungs based on the Brixia scoring system: (a) definition of zones for each lung; (b-d) examples of scoring for each lung zone.

- 0 points: no lung abnormality
- 1 point: interstitial infiltrate
- 2 points: interstitial infiltrate and alveolar consolidation (predominantly interstitial)
- 3 points: interstitial infiltrate and alveolar consolidation (predominantly alveolar)

Other lung injuries, such as pleural effusion or abdominal aortic aneurysm, are not included in the scoring system.

From these six scores, a global score can be synthesized to fall within the range of [0,18]. The global score provides information about the patient's health status, helping healthcare experts make more effective diagnostic and treatment decisions. Examples of scores assigned to different cases are presented in Figure 3(b-d). Specifically, Figure 2b illustrates a case with a global score of 3, indicating primarily peripheral lung involvement. Figure 2c depicts a case with a global score of 10, indicating dispersed and evenly distributed lung abnormalities, with a concentration in the lower region, scoring 2 and 3. Finally, Figure 2d describes a case with a global score of 15, indicating evenly distributed lung abnormalities between scores 2 and 3.

3.2. The COVIDGR dataset and RALE score

COVIDGR is an X-ray image dataset developed by X-ray experts in Spain to support the diagnosis of COVID-19. This dataset includes X-ray images that are positive for COVID-19 corresponding to patients with positive RT-PCR test results and an error margin of less than 24 compared to the X-ray. In addition, the dataset

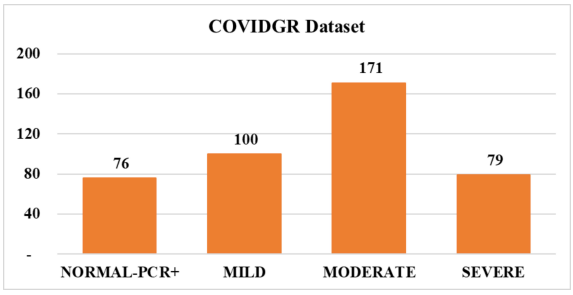

Figure 4. Samples in the COVIDGR dataset.

also includes images that do not show symptoms but are positive according to the RT-PCR test, represented as the severity level of normal PCR+, along with other severity levels, including mild, moderate, and severe. The close collaboration of X-ray experts in the construction of this dataset ensures the reliability of classification results.

The dataset consists of a total of 852 CXR images, divided into two parts: 426 positive X-ray images and 426 negative X-ray images. Among the positive cases include all severity levels assessed by the RALE score, including 76 images of Normal severity of PCR+, 100 images of Mild severity, 171 images of Moderate severity, and 79 images of Severe severity. Figure 4 shows the distribution of sample counts in the COVIDGR dataset.

The RALE score is a tool used to assess the severity of abnormalities in the lungs on CXR. It's also a scoring system used to evaluate the extent of lung injury for COVID-19 patients in the COVIDGR dataset. The RALE score has been shown to be highly accurate in diagnosing acute respiratory distress syndrome (ARDS) [30], and initial changes in the RALE score can indicate the prognosis of patients on mechanical ventilation. This system is based on the density of white spots on the X-ray image to assess the severity of lung congestion. As the RALE score increases, it may indicate the presence of respiratory mechanical issues and the concentration of abnormal biological substances in the lungs.

This method divides the lungs into 4 regions by drawing a horizontal line starting from the first branch of the left lung and a vertical line through the center of the vertebrae. Then, each region is assigned a concentration score ranging from 0 to 4, based on the degree of lung opacity (0: none, 1: minimal <25%, 2: mild 25-50%, 3: moderate 50-75%, 4: severe >75%), and a thickness score of lung opacity (1: hazy, 2: moderate, 3: dense). The RALE score for each lung region is calculated by multiplying the concentration score and the thickness score, and then assigning a score to that region (from 0 to 12). The total RALE score for the entire lungs is calculated by summing the scores of

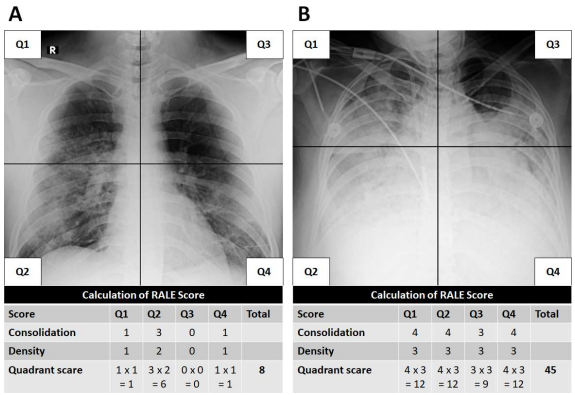

Figure 5. The calculation of the Radiographic Assessment of Lung Edema (RALE) score in CXR.

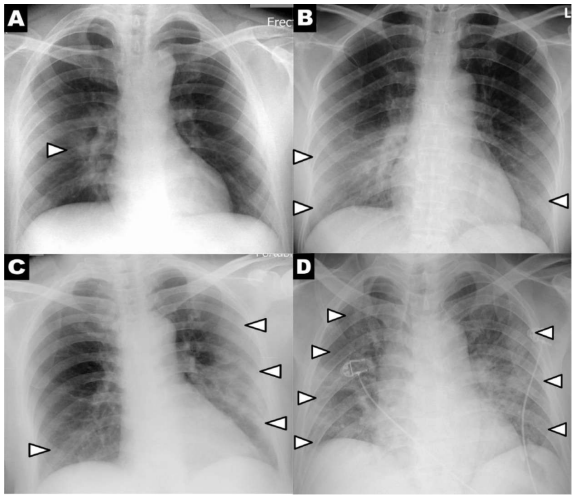

Figure 6. An example of how to calculate the severity score on a CXR in patients with COVID-19 is presented as follows (right lung score + left lung score = total score used): (A) 1 + 0 = 1; (B) 2 + 1 = 3; (C) 1 + 3 = 4; (D) 4 + 3 = 7.

all regions (from 0 to 48). Figure 5 illustrates the calculation of scores for each region and the final total lung score.

To assess the severity of lung involvement in patients with COVID-19, a study [31] has proposed using the RALE score and adjusting it accordingly. The new score is calculated by assigning values from 0-4 to each lung based on the degree of imaging characteristics such as consolidation and ground-glass opacities in four quadrants of each lung. The cumulative scores of all lungs will generate a final severity score. Based on this score, experts can determine the severity of infection in four different stages: normal (0), mild (1-2), moderate (3-5), and severe (6-8). An example of how to calculate the severity score on a CXR in COVID-19 patients is illustrated in Figure 6.

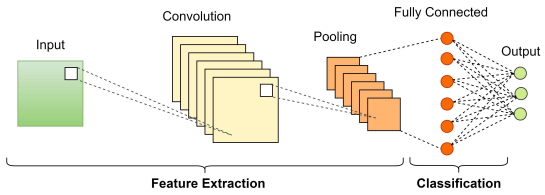

Figure 7. Schematic diagram of a basic convolutional neural network (CNN) architecture.

4. Proposed methods

4.1. System architecture

Classification is one of the important applications of Convolutional Neural Networks (CNNs) in the field of deep learning. With the advancement of technology, image classification has become more accurate through the use of deep learning algorithms such as CNNs. This makes image classification easier for researchers and experts in various fields, including healthcare, technology, science, and the arts. CNN is a popular convolutional neural network architecture used for image and sound analysis. It operates by combining convolutional layers, pooling layers, and fully connected layers to create an accurate classification model. Figure 7 details some of the basic components of the CNN architecture.

When it comes to applying new technologies for analyzing large-scale data, deep learning demonstrates its power by requiring less effort in data cleaning and less human intervention. Therefore, systematically analyzing problems and applying appropriate CNN models to improve performance is a good approach, especially for classifying the severity of lung damage caused by COVID-19.

Figure 8 details our approach to applying deep learning to the classification problem. To achieve this, we employed an efficient combination of neural networks. Additionally, we demonstrate an end-to-end approach to ensure accurate and comprehensive results. After inputting the images into the network, we perform normalization and use data augmentation techniques to increase the size of the two datasets. We preprocess the images through five neural network models efficiently to leverage the knowledge learned from large data and rapidly build and refine models for specific tasks. This step allows the networks to extract data features and minimize loss functions automatically. Finally, each input image is classified into one of four classes: NORMAL-PCR+, MILD, MODERATE, and SEVERE.

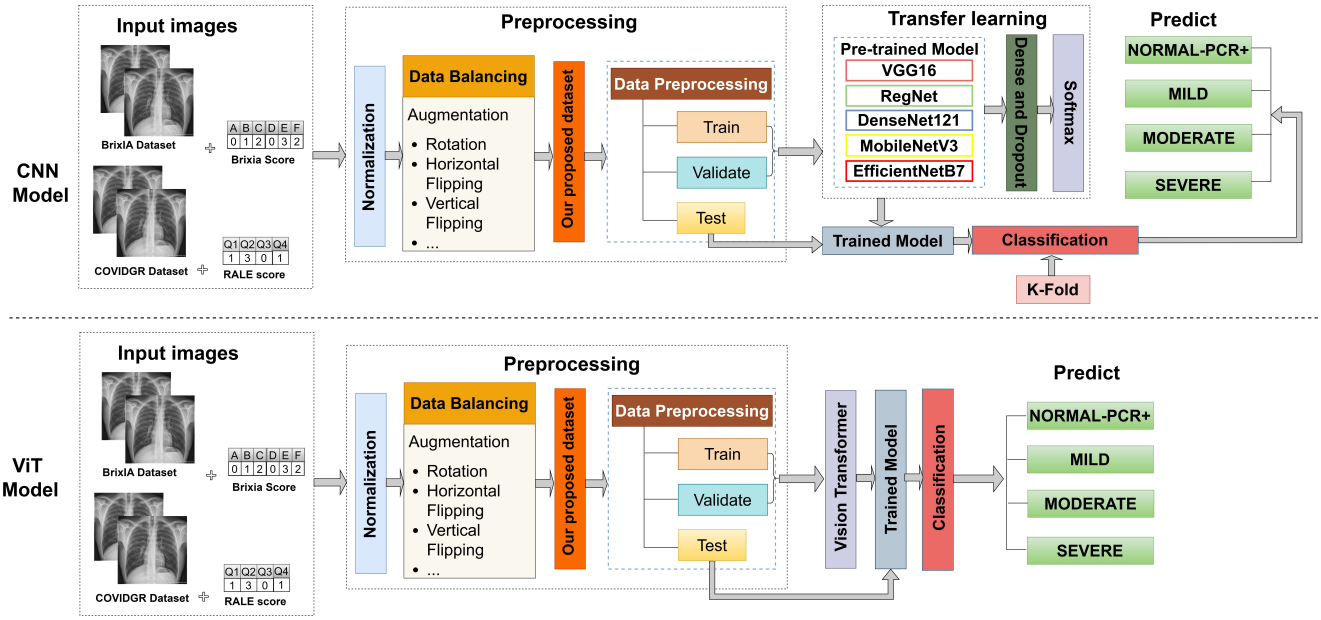

Figure 8. Overall Architecture of Chest X-Ray Classification Method.

4.2. Our proposed dataset

The process of training the model using Convolutional Neural Network (CNN) techniques requires many labeled images to be used as data for the necessary stages. In addition, the input images must have an appropriate resolution for the development of deep learning architectures. Furthermore, the accuracy of the model also depends on ensuring that all attributes of the image are consistent with each other, including brightness, contrast, resolution, and frame rate. Ensuring the homogeneity of these attributes can help improve the accuracy of the classification model.

In addition to CNN, Vision Transformer (ViT) is a cutting-edge neural network model widely used in computer vision tasks, particularly image classification. A key advantage of ViT is its reduced reliance on factors like input image resolution, enabling the model to maintain high performance even when processing images of varying resolutions.

The Brixia scoring system is a good method for stratifying the risk of COVID-19 patients based on the severity of their cases. However, this system requires expertise and experience to interpret CXR, while the RALE scoring system can predict the need for supplemental oxygen and ICU admission to support mechanical ventilation but lacks detailed and more complex indices in evaluating chest X-ray images for the diagnosis of COVID-19 pneumonia. According to the study by Roberto Maroldi [28], RALES and Brixia scores have a significant and reliable correlation in the diagnosis of COVID-19 pneumonia.

Based on the similarity between the two systems for assessing the severity of lung damage caused by COVID-19 introduced above, we conducted an analysis and comparison of two different datasets regarding the degree of lung injury severity. Subsequently, we adjusted and merged them to create a new data set with more comprehensive information. This new dataset encompasses four levels of lung injury severity, summarized as follows:

- No damage: X-ray images of lungs show no signs of damage or only minor signs that do not affect the patient's respiratory function.
- Mild damage: X-ray images of lungs show signs such as mild haziness, patchiness, or some signs of inflammation. These mild damages do not affect the patient's respiratory function and can be self-treated or closely monitored.
- Moderate damage: X-ray images of lungs show signs such as increased haziness, more signs of inflammation, or some fibrosis. These moderate damages can affect the patient's respiratory function and require treatment with anti-inflammatory drugs and respiratory support.
- Severe damage: X-ray images of lungs show severe signs such as increased haziness, extensive signs of inflammation, significant fibrosis, lung consolidation, or lung edema. These severe damages seriously affect the patient's respiratory function and require immediate hospital treatment,

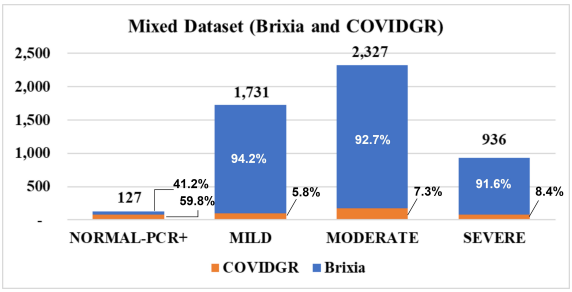

Figure 9. The combined dataset from the Brixia and COVIDGR datasets.

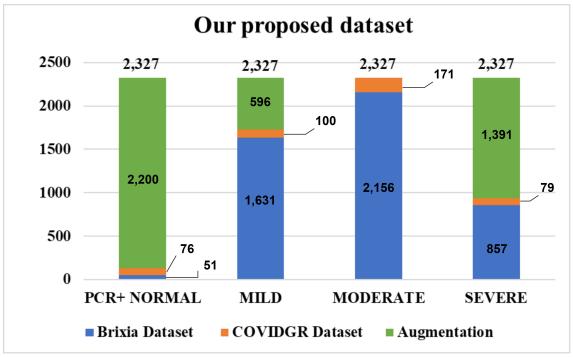

Figure 10. Samples in our proposed dataset.

including the use of ventilators and specific treatment drugs.

To ensure that the Brixia dataset is compatible with the training model and consistent with the COVIDGR dataset, we have performed several pre-processing steps. Firstly, we converted all DICOM format images to JPEG format to standardize the image format in the dataset. Then, we manually labeled the images in the dataset into 4 groups: normal, mild, moderate, and severe, based on the corresponding Brixia scores for each image in the CSV description file of the dataset.

Figure 9 illustrates the number of samples in the new dataset after merging the two datasets, which totals 5,121 samples. This has resulted in a new data set with a larger and more diverse set of data samples.

It can be observed that the dataset has a significant class imbalance, with a noticeable difference in the size of the minority and majority classes. Any dataset with uneven distribution between majority and minority classes can be considered as having class imbalance. Class imbalance can cause several issues in the model-building process, including poor training effectiveness, low accuracy, and reduced generalization ability of the model. Therefore, to enhance the effectiveness of the model, addressing the issue of class imbalance is crucial in classification tasks and can be achieved by using data balancing techniques or utilizing classification algorithms designed specifically to handle imbalanced data [32].

Data pre-processing is an extremely important step for any machine learning system or algorithm. In the process of processing image data for training, validation, and testing, we use the image_dataset_from_directory method to sequentially read the image data. To speed up the processing, we divide the input images into smaller batches with a batch_size of 128. By sequential training on each batch and computing the final values, we ensure the accuracy and reliability of the model. With the use of this method, we ensure that each data set mentioned above is processed accurately and efficiently.

4.3. Data Augmentation

To improve the compatibility with the model, we have enhanced the size of the images in the dataset to a resolution of 224x224 and used RGB color mode for the images. However, due to the imbalance and significant difference in the number of images among different labels in the datasets, we have combined the two datasets and applied typical data augmentation techniques. We chose to augment the minority labels with the same number of samples as the majority label in the dataset. This approach helps to improve the classification performance of the model on minority samples, thereby enhancing the accuracy and reliability of the model [33]. These augmentation techniques include image flipping, zooming, and resizing to enhance the model's training capability, as follows:

- Rotate the image at a random angle from -5 to 5 degrees.
- Scale the image up or down by a random factor from 1.1 to 1.5 times.
- Flip the image horizontally (swap the positions of pixels from left to right and vice versa).
- Flip the image vertically (swap the positions of pixels from top to bottom and vice versa).
- Perform random transformations on the image, including stretching, squeezing, and skewing.

With the use of this method, we have ensured that the model is trained on a diverse and sufficiently large dataset to optimize its accuracy and reliability. The result after applying these techniques is an expanded dataset with a total of 9,294 images. Data augmentation helps to supplement the original dataset with additional data, thereby increasing the accuracy and generalization ability of the model. Balancing the labels also plays an important role in ensuring the reliability of the classification results in our study.

Table 1. Details about the separation of COVID-19 lung x-ray data

	Normal-PCR+	Mild	Moderate	Severe
Train	1,628	1,628	1,628	1,628
Validate	350	350	350	350
Test	349	349	349	349
Total	2,327	2,327	2,327	2,327

The new dataset consists of a total of 9,294 X-ray images, including 2,327 images for each data sample (Normal-PCR+, Mild, Moderate, Severe). Figure 10 shows the distribution of sample counts in our proposed dataset. The dataset has been divided into different sets for the training, validation, and testing stages of the model, as shown in Table 1.

- The training set consists of 6,498 X-ray images belonging to 4 different classes, including Normal-PCR+, Mild, Moderate, and Severe. This dataset is used to train the model and learn the features of different classes.
- The validation set includes 1,400 X-ray images belonging to the same 4 classes as the training set. This dataset is used to evaluate the performance of the model on new data and assess its generalization ability.
- Finally, the test set consists of 1,396 X-ray images belonging to the 4 classes of Normal-PCR+, Mild, Moderate, and Severe. This dataset is used to evaluate the final performance of the model and assess the model's accuracy in classifying new data.

The division of the dataset into training, validation, and test sets helps to evaluate the performance of the model on different data and ensures the generalization of the classification results. Table 2 presents the training results of the EfficientNetB7 model on the original data (the synthesized dataset before data augmentation and label balancing) and on the augmented and balanced dataset. Noticeable differences in accuracy can be observed when applying data augmentation to the training dataset.

4.4. Model training

Evaluating the extent of lung damage caused by COVID-19 is a crucial task in ensuring effective diagnosis and treatment for patients. To accomplish this task, we chose to use models such as VGG16, RegNet, Densenet, MobileNet, EfficientNet, and ViT for the classification and quantification of lung damage. These models have been trained on large datasets and have demonstrated excellent feature extraction capabilities from images. This is essential in detecting the manifestations and characteristics of lung injuries. Moreover, these models have been trained on various large datasets, including medical images. Utilizing pre-trained models can enhance the overall capability and performance of the lung damage assessment system. In particular, ViT excels at learning global features from images, enhancing lung lesion detection accuracy. Additionally, models like MobileNet and EfficientNet are specifically designed to perform well on devices with limited computational resources, making them especially valuable for scenarios where rapid and efficient lung lesion assessment is required.

The VGG16 model. The VGG16 model [34] is a significant model in the field of deep learning that has achieved success in image classification tasks. It has a simple architecture, using convolutional layers and fully connected layers to learn features from images. This model exhibits high performance but requires significant computational resources. We chose VGG16 as the baseline model in this study because it is one of the fundamental models in computer vision. While VGG16 may not be the latest model available, it still delivers relatively good performance in classifying lung damage severity on our proposed dataset.

The RegNet model. RegNet is a unique deep learning architecture designed for high-performance image classification. One notable feature of RegNet is its creation of multiple parallel paths with varying depths and widths, improving accuracy without significantly increasing computational complexity [35]. RegNet has demonstrated performance improvements over traditional architectures and meets the requirements of various image datasets. In this study, we utilized the RegNet0Y40 model, integrated into TensorFlow, to evaluate the newly constructed dataset for the purpose of classifying COVID-19 lung damage. Employing RegNet0Y40 is a crucial step in assessing the performance of models on this new dataset, ensuring the objectivity and reliability of research results.

The DenseNet121 model. DenseNet121 is a popular deep learning architecture within the DenseNet family of networks [36]. This architecture tightly connects layers, efficiently reusing features and achieving higher performance in image recognition. DenseNet121 has been trained on large datasets and can be applied in various applications, including image classification, and identifying pathology in medical images. We employed this model to classify the severity of COVID-19 lung damage based on X-ray images. Alongside other models, DenseNet121 contributes to this research by providing a foundation for the development of effective prediction and classification methods.

Table 2. Training results of EfficientNetB7 model on the dataset before data augmentation and after performing data augmentation and label balancing

	Without Augmentation	With Augmentation
Training Loss	0.935	0.264
Validation Loss	1.114	0.732
Training Accuracy	0.471	0.922
Validation Accuracy	0.455	0.738

The MobileNetV3 model. MobileNetV3 [37] is a deep learning architecture designed to optimize performance and integration for mobile and embedded devices. This model utilizes depth wise separable convolution layers to improve feature reuse and minimize information loss. A study [38] has employed MobileNetV3 for classifying patients with pulmonary tuberculosis based on X-ray image datasets. The results demonstrated that MobileNetV3 achieved the highest accuracy compared to other models, indicating its suitability for classifying patients with pulmonary tuberculosis based on X-ray image datasets. We leveraged this excellence and fine-tuned the MobileNetV3 model to fit the task of classifying the severity of COVID-19 lung damage. The application of MobileNetV3 in this research provides benefits in terms of computational resource efficiency and accelerated classification processes, which is crucial in situations requiring rapid and efficient lung damage assessment.

The EfficientNetB7 model. EfficientNetB7 is a specific variant of the EfficientNet [39] model, an advanced deep learning architecture designed to achieve higher accuracy and efficiency than other models. EfficientNetB7 is characterized by its large size, with high numbers of layers and parameters. It uses a combination of scaling methods, including depth, width, and resolution, to create a network that is both computationally efficient and highly accurate. EfficientNetB7 has been widely used in various applications, including image classification, object detection, and medical imaging. It has shown outstanding performance in tasks related to COVID-19 detection, especially in processing CT and X-ray images of patients' lungs. The model has been extensively evaluated and has demonstrated superior results in accuracy and efficiency compared to other deep learning models [40]. To address the problem of classifying the severity of lung damage caused by COVID-19, we employed the EfficientNetB7 neural network architecture to implement the feature extraction process. This architecture is designed with a total of 7 main blocks to optimize lung damage classification and detection. These blocks are configured to detect various features of lung damage, thereby enabling a more accurate assessment of lung damage severity. The detailed architecture of the EfficientNetB7 neural network is described in Figure 11.

EfficientNetB7 is a very powerful model with numerous parameters, which can make it prone to overfitting, especially when dealing with limited data. To address this issue, we attempted to fine-tune the model to fit our proposed dataset better. By introducing Dropout, we aim to reduce the model's tendency to "over remember" the training data and help it generalize better. In our case, Dropout values ranging from 0.2 to 0.5 are considered suitable for our task. EfficientNetB7 models typically end with several convolutional layers, and the entire image is represented as a 4D tensor. To use Dense layers (or fully connected layers) afterward, we employed a Flatten layer to transform this 4D tensor into a 1D vector. A Dense (128) layer is used with 128 neurons because it aligns with the number of output classes required for the classification task. In the case of classifying into four severity levels (Normal, Mild, Moderate, and Severe), we need an output layer with the same number of neurons as the input classes, which is 4. We utilize the SoftMax Activation function to convert the output values into probabilities, making it easier to interpret and use the classification results. Additionally, using Dense (128) does not heavily increase computational resources or significantly complicate the model's complexity.

Vision Transformer model. The Vision Transformer (ViT) is a neural network model that leverages the Transformer architecture for computer vision tasks, unlike traditional CNN models that rely on convolution. ViT breaks the input image into small patches, typically 16x16 pixels, and maps them into feature vectors. These vectors and positional encoding are then passed through multiple self-attention layers to learn global spatial relationships between the patches. This allows ViT to process and synthesize information from the entire image efficiently. ViT performs exceptionally well when trained on large datasets like ImageNet, surpassing CNNs thanks to its ability to capture global features and independence from input image resolution. However, ViT requires substantial data and computational resources to be trained effectively. According to research by Dosovitskiy et al. [41], ViT has shown its capability

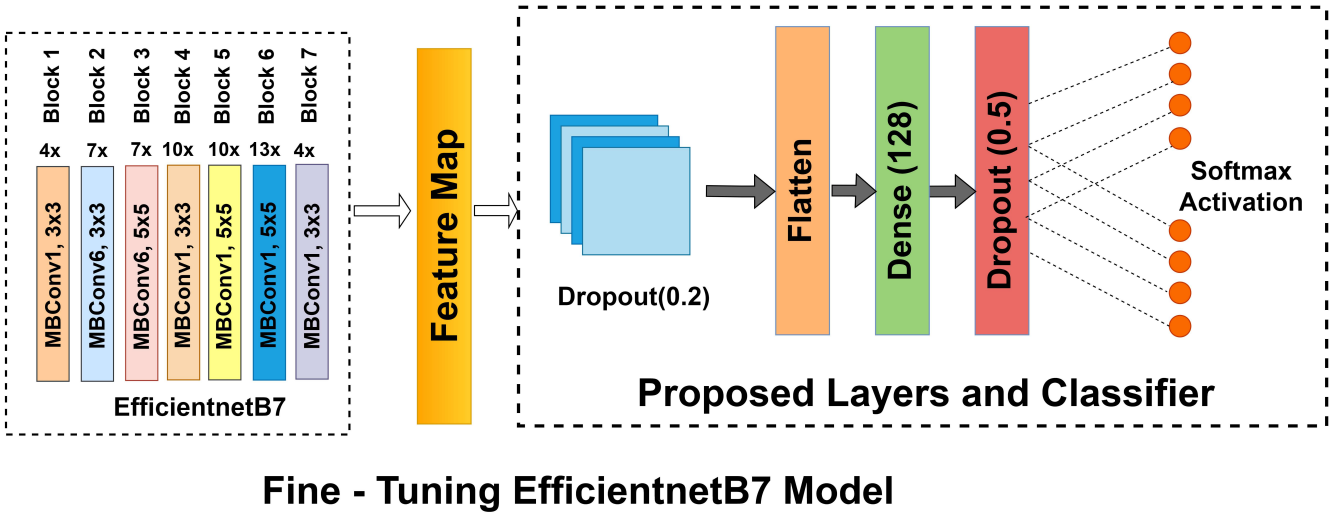

Figure 11. Proposed fine tuning method using EfficientNetB7 Neural Network.

to compete with state-of-the-art CNN models across various image classification tasks.

5. Experimental results

5.1. Performance

To evaluate the quality of predicted classes compared to ground truth classes, we used two commonly used metrics: accuracy and F1 score. These metrics are derived from four types of values in the confusion matrix: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). From these values, we can calculate evaluation indices to assess the accuracy and effectiveness of the model in classifying lung images as either pathological or non-pathological. In multi-class classification, the values of TP, TN, FP, and FN can be calculated in a similar way as in binary classification as follows:

- TP (True Positive): The number of images predicted as abnormal and are abnormal.
- TN (True Negative): The number of images predicted as normal and are normal.
- FP (False Positive): The number of images predicted as abnormal but are normal.
- FN (False Negative): The number of images predicted as normal but are abnormal.

The authors in the study [42] have pointed out that based on the values of TP, TN, FP, and FN, we can calculate the performance evaluation indices of the model. In multi-class classification, we need to calculate accuracy, precision, and sensitivity for each class. To calculate the performance evaluation values of the model in multi-class classification, Table 3 describes the criteria used to evaluate COVID-19 classification models.

Evaluating a model's performance using these metrics is very useful in selecting the most suitable model for lung injury prediction. During the training and testing process on 6,498 images, we assessed the performance of several networks and found that EfficientNetB7 is one of the top-performing networks with an accuracy of 75% and a specificity of 92% for classifying 4 levels of injury severity. In addition, Vision Transformer (ViT) demonstrated higher specificity, reaching 94%, highlighting its ability to classify lesion levels accurately. Figure 12 is used to illustrate the confusion matrix of correct and incorrect predictions for the models on our augmented dataset. Throughout the model evaluation process, EfficientNetB7 (10e) has demonstrated superior predictive capabilities compared to other models, especially in classifying images with mild and moderate injury levels. In total, 1,396 images were tested, of which 250 images were accurately labeled as non-injured, 288 images were accurately labeled as a minor injury, 310 images were accurately labeled as moderate injury, and 269 images were accurately labeled as severe injury.

Figure 13 provides a summary of various metrics used to assess the models, including Accuracy, Precision, Sensitivity, F1-score, and Specificity displayed for each model. These metrics are crucial in evaluating the performance of the models, and help determine which model should be prioritized for specific applications. In the future, researching and developing machine learning models for COVID-19 lung injury assessment will continue to bring significant value to the medical field. The variation in classification capabilities among the

Table 3. The criteria used to evaluate the quality of predicted classes in the study

Metric	Formula	Description
Accuracy	$\frac{1}{m}\sum_{i=1}^{m}\frac{TP+TN}{TP+TN+FP+FN}$	This is the average accuracy of the model in classifying m classes
Precision	$\frac{1}{m}\sum_{i=1}^{m}\frac{TP}{TP+FP}$	The macro average precision of m classes
Sensitivity	$\frac{1}{m}\sum_{i=1}^{m}\frac{TP}{TP+FN}$	The macro average recall of m classes
F1-score	$\frac{1}{m}\sum_{i=1}^{m}\frac{2P*R}{P+R}$	The macro average F1-score of the model in classifying m classes, in which: P (Precision) and R (Sensitivity)
Specificity	$\frac{1}{m}\sum_{i=1}^{m}\frac{TN}{TP+FN}$	The macro average specificity of the model in classifying m classes

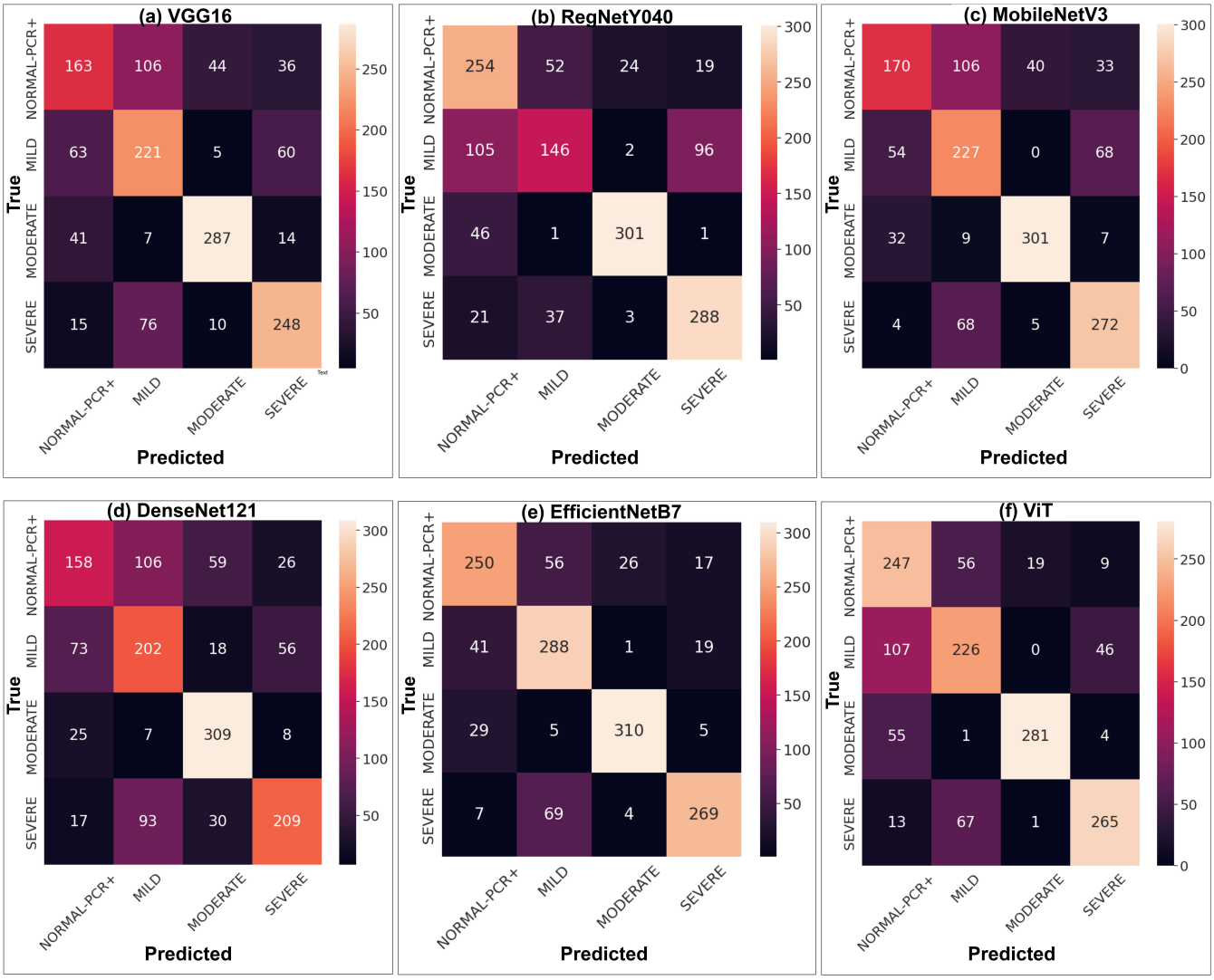

Figure 12. Confusion matrices for the models applied to our proposed dataset.

models will help meet the specific needs of patients and provide essential information to support physicians in making effective diagnostic and treatment decisions.

Figure 14 shows that the ViT(13f) model excels at classifying the severity of lung injury, especially in correctly classifying severe (SEVERE) lesions. Its ability to learn global features within images allowed it to perform exceptionally well in recognizing SEVERE labels. Similarly, the EfficientNetB7 (13e) model proved highly effective, especially in classifying mild to moderate (MID) lesions. Other models, such as VGG16 (13a), RegNet (13b), MobileNet (13c), and DenseNet (13d), were also evaluated. While each model has its advantages, they often struggle with accurately classifying mild lesions, with many errors resulting in misclassification or confusion with other classes. The differences in classification performance among these models highlight the importance of selecting the most suitable model for assessing and predicting COVID-19 lung injuries. Specifically, choosing the appropriate model is crucial for effectively detecting and managing mild injuries that require close monitoring and severe injuries that demand prompt treatment.

5.2. Cross-validation

When considering ensuring the accuracy and reliability of a model, several factors can impact its performance, including the size and diversity of the dataset, model architecture, hyperparameters, and optimization methods. To address these issues, we employed cross-validation as a method to assess the performance of our model across multiple runs. This approach allows us to obtain more accurate and reliable results. Furthermore, we researched and selected this cross-validation method to ensure its suitability for our research purposes. Cross-validation is a common method for assessing the performance of a model on a specific dataset. It involves dividing the dataset into several parts, or "folds," and training the model on each fold while evaluating its performance on the remaining folds. This process is repeated multiple times, with each iteration using a different fold for evaluation. As a result, we can obtain a more accurate estimate of the model's performance on previously unseen data.

Our dataset was split into 5 and 10 folds for the purpose of training and testing the classification dataset. To measure classification performance, we used a confusion matrix to calculate the average accuracy across all the folds used, and the corresponding results are displayed in Table 5 and Table 4. The result is the average accuracy of the model across all the folds used. Through cross-validation techniques, it can be observed that the model achieves a specificity of 92.89% and an accuracy of 75.09%. This is a noteworthy and promising result regarding the model's capability to detect and classify medical conditions.

The specificity of 92.89% indicates the model's ability to identify negative cases accurately and reliably. This is particularly important in the medical field, showing the model's capacity to minimize errors by excluding unnecessary or inappropriate cases. The accuracy of 75.09% demonstrates that the model is performing consistently and exhibits good overall classification performance. Although this figure may be lower than the specificity, it can be explained by the trade-off between detecting positive cases and handling many negative cases.

5.3. Run Time

The run time graph in Figure 15 illustrates the variations in prediction time and training time for the models. Prediction time is measured from when the model receives input data to when it returns prediction results, while training time is the time the model spends learning from training data.

From the graph, it is evident that the RegNetY040m model offers shorter training and prediction times compared to the other models. However, it's important to note that execution time can vary based on computing resources and input data size. Thus, performance and runtime must be considered to ensure the model performs effectively on large datasets. Figure 16 illustrates the variations in runtime among the evaluated models. We recorded the execution time for a single run of each model on a test dataset comprising 1,369 images. Additionally, Figure 17 shows the detailed average values from 8 runs and the average accuracy across 4 data labels. While ViT has a longer training time, it excels in managing large datasets and operates independently of input image resolution, making it advantageous for handling complex data scenarios. Conversely, EfficientNetB7 is better suited for situations requiring fast results or deployment on resource-constrained devices due to its optimal balance between performance and resource usage.

The accuracy and loss plots offer valuable insights into the performance of each model during training and testing. The accuracy plot tracks the model's predictive performance across epochs, highlighting improvements or declines in accuracy as training progresses. In parallel, the loss plot represents the reduction in the loss function over each epoch-a consistent downward trend on the loss plot signals effective learning and optimization. For instance, Figure 18 shows that the EfficientNetB7 model maintains a steadily decreasing loss curve, indicating robust optimization and efficient learning, particularly in classifying mild to moderate lesions. Comparing these plots across models helps to pinpoint which ones perform best. While specific

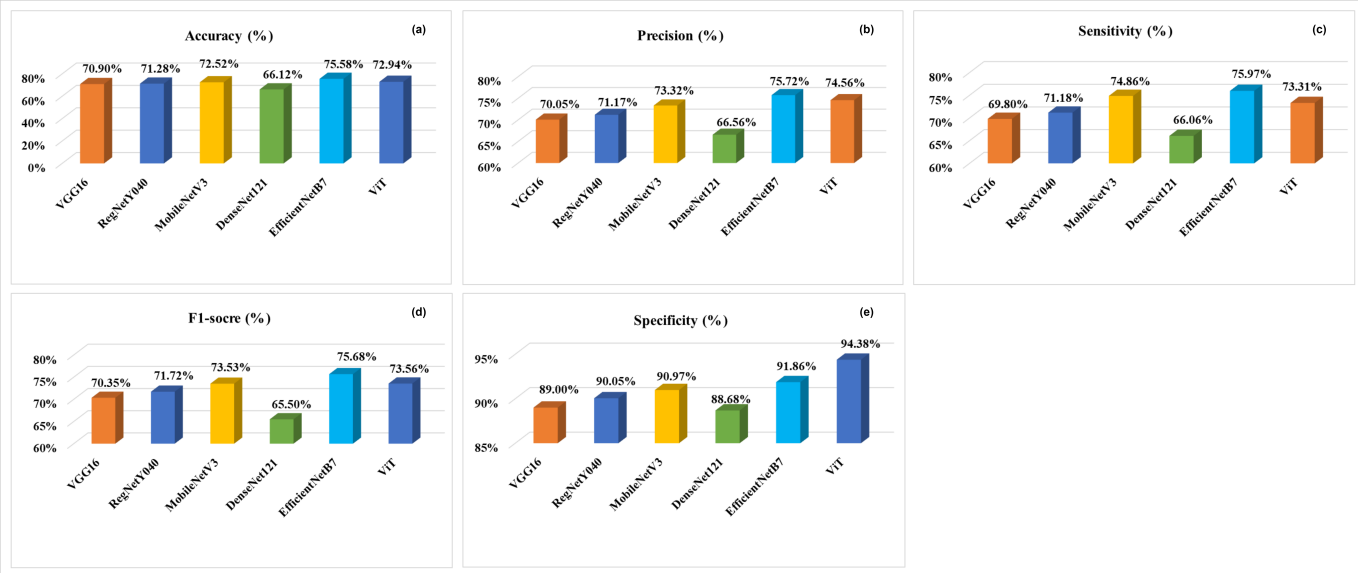

Figure 13. The performance metrics for the evaluated ML models on our proposed dataset.

Table 4. Performance Evaluation and Accuracy Metrics for 5 Folds for the EfficientNetB7 Model Average Classification

Metrics	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	**Average**
Accuracy(%)	72.9	73.8	74.2	73.0	73.3	**73.46**
Precision(%)	72.9	74.4	74.3	73.6	73.6	**73.79**
F1-score(%)	72.4	73.2	73.8	72.5	72.9	**72.98**
Sensitivity(%)	73.1	73.5	74.1	73.2	73.2	**73.44**
Specificity(%)	90.9	91.2	91.4	91.1	91.1	**91.15**

Table 5. Average Classification Performance Evaluation and Accuracy Metrics for 10 Folds for the EfficientNetB7 Model

Metrics	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6
Accuracy(%)	75.4	73.7	74.5	75.9	77.8	73.8
Precision(%)	75.7	73.6	74.8	75.7	77.8	74.2
F1-score(%)	75.9	73.1	74.1	75.2	77.2	73.3
Sensitivity(%)	75.5	73.5	74.3	75.0	77.8	74.0
Specificity(%)	95.1	91.2	91.5	93.3	95.9	91.3

Metrics	Fold-7	Fold-8	Fold-9	Fold-10	**Average**
Accuracy(%)	73.4	76.7	73.0	76.3	**75.09**
Precision(%)	74.8	76.8	73.3	76.6	**75.37**
F1-score(%)	73.1	76.1	72.6	76.9	**74.78**
Sensitivity(%)	73.4	76.8	73.0	76.2	**74.99**
Specificity(%)	91.1	94.2	91.0	94.1	**92.89**

models might achieve higher accuracy, an uneven reduction in loss may suggest challenges in optimizing and learning the data's underlying features. Overall, the accuracy and loss plots provide essential insights for evaluating and selecting the most suitable model for classifying COVID-19-related lung lesions.

5.4. Discussion

In the battle against the COVID-19 pandemic, classifying the extent of lung damage plays a crucial role. It allows for a clear determination of the severity of the disease, ranging from mild cases that can be managed with home isolation to severe lung damage that requires

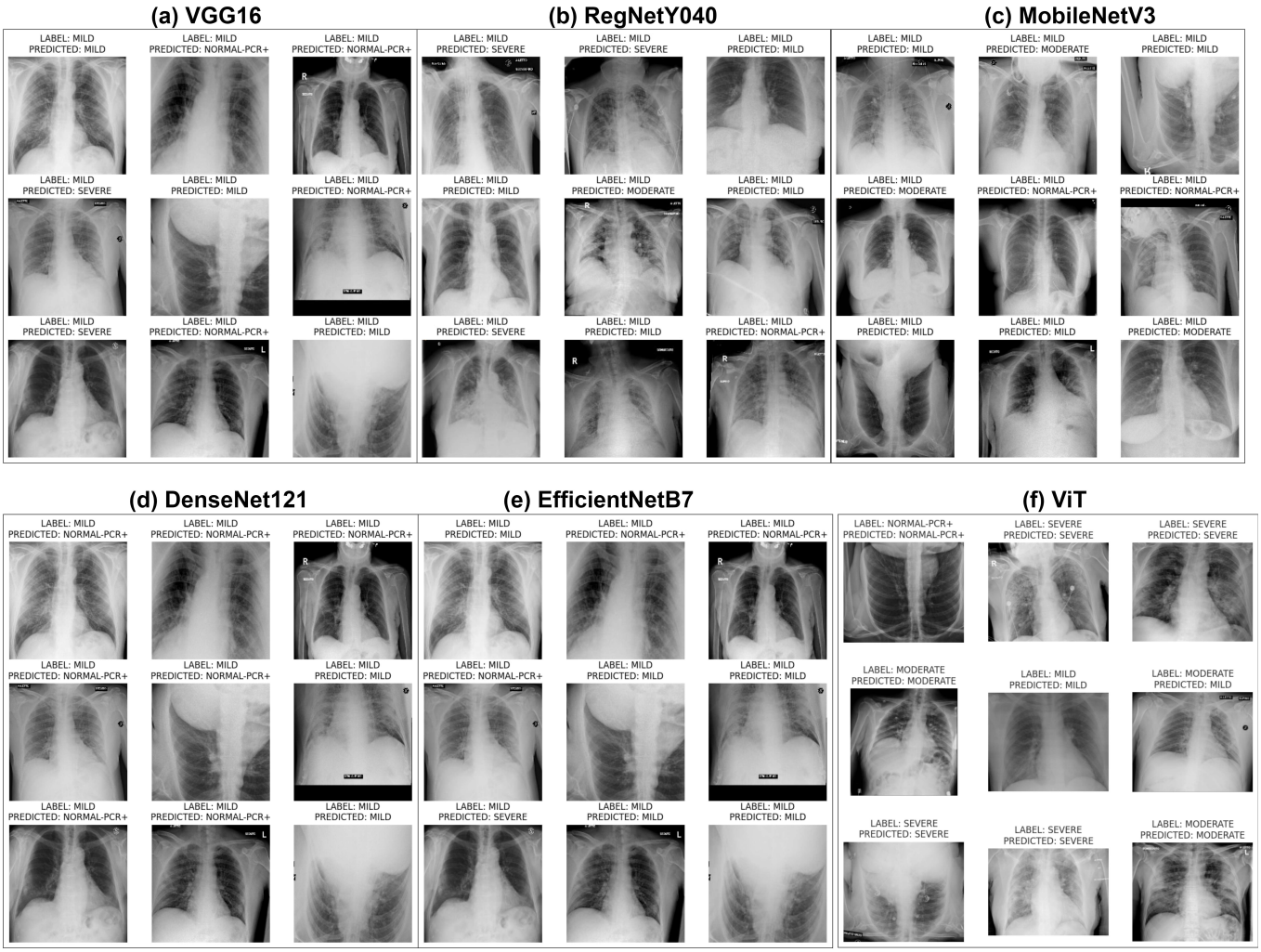

Figure 14. An example of the classification performance of the EfficientNetB7 model compared to the other models for the MILD class on the augmented dataset.

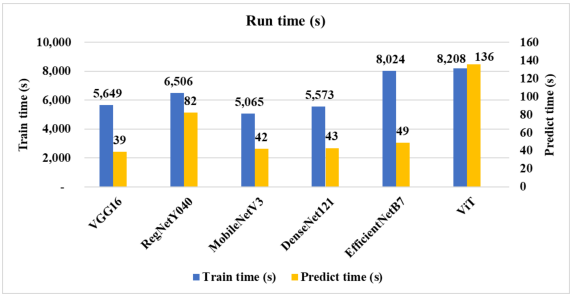

Figure 15. Training and Prediction Times of the Models.

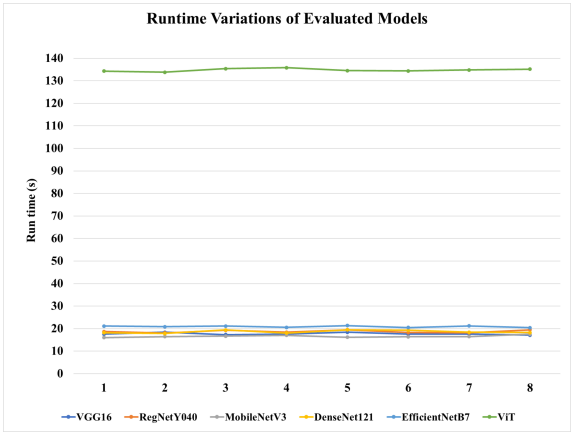

Figure 16. Run time Variations in 8 Runs.

hospital treatment. This classification also helps doctors monitor disease progression, predict outcomes, and adjust treatment methods promptly. Notably, the use of chest X-ray images is a common method for performing

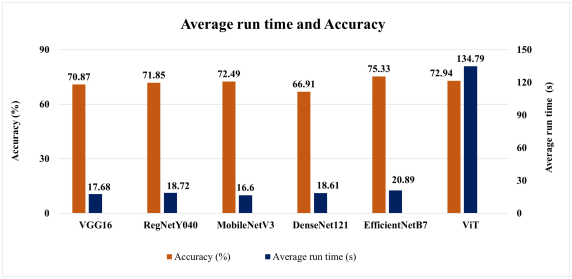

Figure 17. Average Run Time of 8 Runs and Accuracy of Models on the Test Dataset.

this classification. While not as sensitive as CT images, X-ray images are still considered a reasonable choice for assessing the extent of lung damage and monitoring the progress of COVID-19 in many healthcare facilities. Additionally, machine learning models have the potential to create powerful tools for doctors and healthcare researchers, improving the diagnostic and treatment processes for patients. By utilizing machine learning technology, doctors can rely on automated diagnostic information to assess risks and make initial treatment decisions. However, this process still requires intervention and careful consideration from the doctor to ensure that the final decisions are based on their extensive medical knowledge and the specific circumstances of the patient.

In this study, we utilized two publicly available datasets that were labeled to correspond to different levels of lung damage, including no damage, mild damage, moderate damage, and severe damage. However, due to the early stages of the pandemic, the availability of labeled data was limited, resulting in a significant imbalance in the number of images for each category. Furthermore, overfitting to the training models could potentially occur, impacting the accuracy of our models. Our data augmentation approach has proven effective in improving the accuracy of the models. However, it is essential to collect more data for lung damage assessment. This would enable us to access a larger and more diverse data set, ultimately enhancing the results of our research.

6. Conclusions and Future Works

Compared to other studies evaluating COVID-19 based on deep learning methods, our research focuses on classifying the severity of lung damage. As part of the study, we also contribute to the public dataset by constructing a new dataset that can support the assessment of COVID-19-induced lung damage. The proposed COVID-19 dataset contains 9,294 images (2,327 NORMAL-PCR+, 2,327 MILD, 2,327 MODERATE, and 2,327 SEVERE). After evaluating five models in the task of classifying COVID-19-induced lung damage on the proposed dataset, the EfficientNetB7 model achieved a specificity of 92.89% and an accuracy of 75.09%. This demonstrates the potential of the EfficientNetB7 model in developing predictive methods for assessing the severity of COVID-19-induced lung damage. In particular, ViT model excels with a specificity of up to 94%, showcasing its capability to classify severe lesions accurately. While ViT requires more training time, it is exceptionally adept at handling large datasets and operates independently of the input image resolution. With the second-highest accuracy among the evaluated models, ViT's robust data processing and high classification precision underscore its potential in assessing lung lesions caused by COVID-19. Nonetheless, selecting the most suitable model should be based on specific needs, whether for in-depth analysis of severe lesions or deployment on resource-constrained systems.

In the future, to improve the accuracy in assisting doctors in diagnosing the four levels of lung damage caused by COVID, we may explore and implement several additional solutions. First, we will collect more lung damage data based on the RALE/Brixia scoring system to enrich the training dataset for our models. Leveraging the ideas from the research by Viacheslav V Danilov et al. [43], we will carry out a preprocessing step with segmentation to separate the two lungs from the surrounding organs, eliminating irrelevant information and enabling deep learning models to focus more on specific lung regions or the entire lungs in future lung-related studies. We will also apply self-supervised learning techniques such as MAE, MoCo, and SimMIM to enhance classification accuracy. Furthermore, in the medical field, extensive experimentation and validation are crucial before applying any approach in real-life scenarios. Our study represents an initial step in assessing the effectiveness of our proposed methods. We recognize the need for further experiments and validations to ensure robustness and reliability in practical applications. Future work will focus on additional testing and collaboration with medical professionals to refine and validate our approach.

Acknowledgements

This research was supported by The VNUHCM-University of Information Technology's Scientific Research Support Fund.

Ethical Approval

It is not applicable. The study does not include any medical tests, treatments, or interventions.

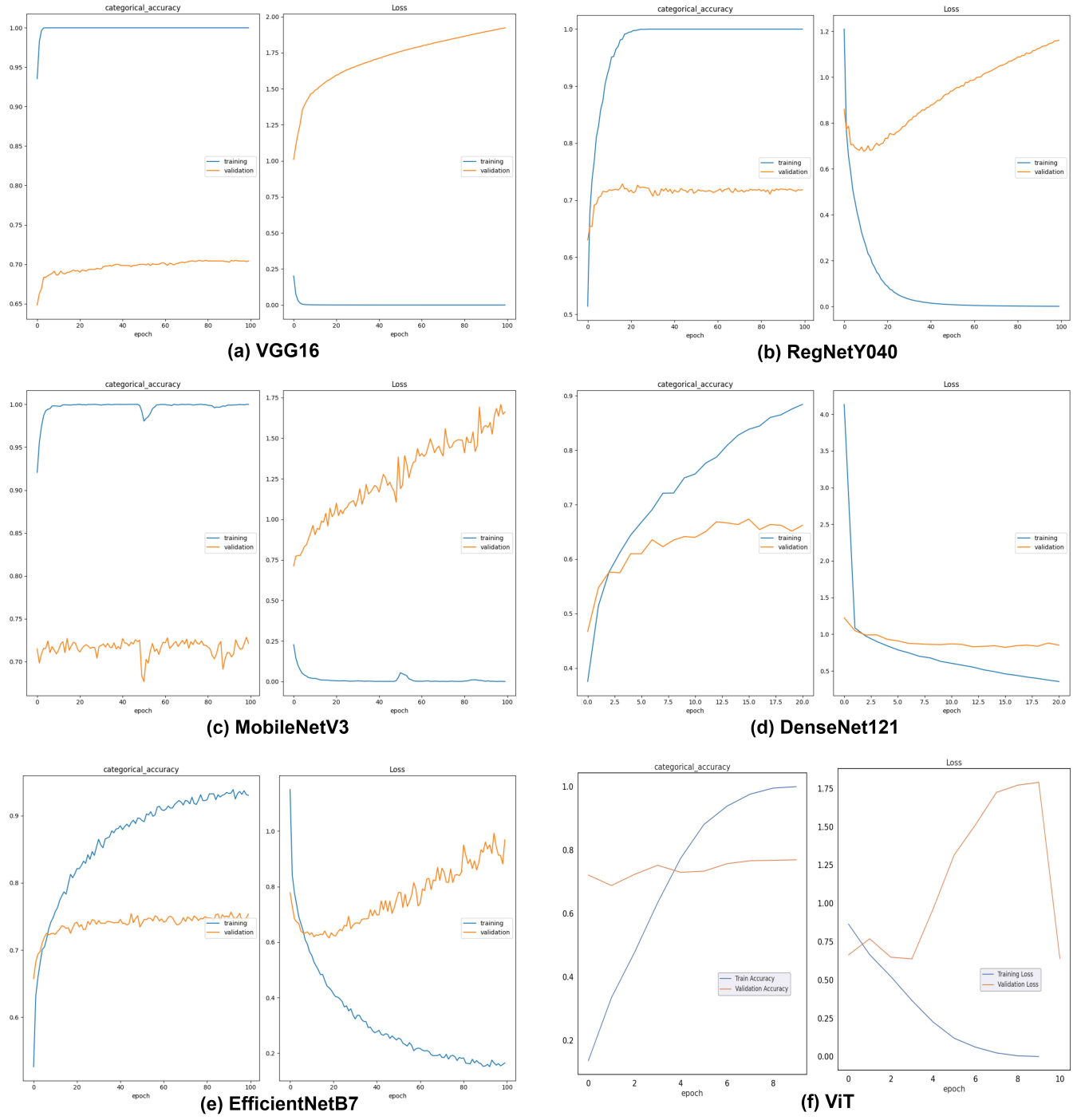

Figure 18. Accuracy and Loss Across Epochs

Statements of Data Availability

The data that support the findings of this study are openly available on GitHub at https://github.com/trinhttv/predicting-covid-19 (our proposed dataset).

References

[1] World Health Organization. COVID-19 Situation Report 94 [Internet]. World Health Organization. 2020. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200423-sitrep-94-covid-19.pdf

[2] Maulanza H, Abidin TF, Mubarak Z, Abdullah A. Model and simulation to reduce COVID-19 new infectious cases: A survey. 2020 International Conference on Electrical Engineering and Informatics (ICELTICs). 2020. doi:10.1109/iceltics50595.2020.9315493
[3] Loey M, Manogaran G, Taha MHN, Khalifa NEM. Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustainable Cities and Society. 2020;102600
[4] World Health Organization. WHO COVID-19 Dashboard [Internet]. World Health Organisation. 2023. Available from: https://covid19.who.int
[5] Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). Radiology. 2020;295(1):200230.
[6] Shanahan MC, Akudjedu TN. Australian radiographers' and radiation therapists' experiences during the COVID-19 pandemic. Journal of Medical Radiation Sciences. 2021;68(2).
[7] Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. European Respiratory Journal. 2020;56(2):2000775.
[8] Duong LT, Le NH, Tran TB, Ngo VM, Nguyen PT. Detection of tuberculosis from chest X-ray images: Boosting the performance with vision transformer and transfer learning. Expert Systems with Applications. 2021;184:115519.
[9] Thien B. Nguyen-Tat and Thien-Qua T. Nguyen and Hieu-Nghia Nguyen and Vuong M. Ngo. Enhancing brain tumor segmentation in MRI images: A hybrid approach using UNet, attention mechanisms, and transformers. Egyptian Informatics Journal. 2024; 27:100528.
[10] Allioui H, Mohammed MA, Benameur N, Al-Khateeb B, Abdulkareem KH, Garcia-Zapirain B, et al. A Multi-Agent Deep Reinforcement Learning Approach for Enhancement of COVID-19 CT Image Segmentation. Journal of Personalized Medicine. 2022;12(2):309.
[11] Khan M, Alhaisoni M, Tariq U, Hussain N, Majid A, Damaševičius R, et al. COVID-19 Case Recognition from Chest CT Images by Deep Learning, Entropy-Controlled Firefly Optimization, and Parallel Feature Fusion. Sensors. 2021;21(21):7286.
[12] Abayomi-Alli OO, Damaševičius R, Abbasi AA, Maskeliūnas R. Detection of COVID-19 from Deep Breathing Sounds Using Sound Spectrum with Image Augmentation and Deep Learning Techniques. Electronics. 2022;11(16):2520.
[13] Haghanifar A, Majdabadi MM, Choi Y, Deivalakshmi S, Ko S. COVID-CXNet: Detecting COVID-19 in frontal chest X-ray images using deep learning. Multimedia Tools and Applications. 2022;
[14] Bhattacharyya A, Bhaik D, Kumar S, Thakur P, Sharma R, Pachori RB. A deep learning based approach for automatic detection of COVID-19 cases using chest X-ray images. Biomedical Signal Processing and Control. 2022;71:103182.
[15] Loey M, El-Sappagh S, Mirjalili S. Bayesian-based optimized deep learning model to detect COVID-19 patients using chest X-ray image data. Computers in Biology and Medicine. 2022;105213.
[16] Zoabi Y, Deri-Rozov S, Shomron N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. npj Digital Medicine [Internet]. 2021;4(1):1–5.
[17] He K, Zhao W, Xie X, Ji W, Liu M, Tang Z, et al. Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT images. Pattern Recognition. 2021;113:107828.
[18] Signoroni A, Savardi M, Benini S, Adami N, Leonardi R, Gibellini P, et al. BS-Net: Learning COVID-19 pneumonia severity on a large chest X-ray dataset. Medical Image Analysis. 2021;71:102046.
[19] Zhu J, Shen B, Abbasi A, Hoshmand-Kochi M, Li H, Duong TQ. Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs. Singh D, editor. PLOS ONE. 2020;15(7):e0236621.
[20] Tang Z, Zhao W, Xie X, Zhong Z, Shi F, Liu J, et al. Severity Assessment of Coronavirus Disease 2019 (COVID-19) Using Quantitative Features from Chest CT Images [Internet]. arXiv.org. 2020.
[21] Carvalho AR, Ranieri A, Gabriel Madeira Werberich, Nery S, Joana Carneiro Pinto, Schmitt W, et al. COVID-19 Chest Computed Tomography to Stratify Severity and Disease Extension by Artificial Neural Network Computer-Aided Diagnosis. 2020;7.
[22] Irmak E. COVID-19 disease severity assessment using CNN model. IET Image Processing. 2021;
[23] Albataineh Z, Aldrweesh F, Alzubaidi MA. COVID-19 CT-images diagnosis and severity assessment using machine learning algorithm. Cluster Computing. 2023;
[24] https://brixia.github.io/ [Internet]. Brixia. Available from: https://brixia.github.io/
[25] Tabik S. COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on Chest X-Ray images. IEEE Journal of Biomedical and Health Informatics. 2020;1–1.
[26] A. Abo-Hedibah S, Tharwat N, H. Elmokadem A. Is chest X-ray severity scoring for COVID-19 pneumonia reliable? Polish Journal of Radiology. 2021;86(1):432–9.
[27] Marques P, Fernandez-Presa L, Carretero A, Gómez-Cabrera MC, Viña J, Signes-Costa J, et al. The radiographic assessment of lung edema score of lung edema severity correlates with inflammatory parameters in patients with coronavirus disease 2019—Potential new admission biomarkers to predict coronavirus disease 2019 worsening. Frontiers in Medicine. 2022;9.
[28] Maroldi R, Paolo Rondi, Giorgio Maria Agazzi, Ravanelli M, Borghesi A, Farina D. Which role for chest x-ray score in predicting the outcome in COVID-19 pneumonia? 2021;31(6):4016–22.
[29] Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. La radiologia medica. 2020;125(5):509–13.
[30] Warren MA, Zhao Z, Koyama T, Bastarache JA, Shaver CM, Semler MW, et al. Severity scoring of lung oedema on the chest radiograph is associated with clinical outcomes in ARDS. Thorax. 2018;73(9):840–6.
[31] Valk CMA, Zimatore C, Mazzinari G, Pierrakos C, Sivakorn C, Dechsanga J, et al. The Prognostic Capacity

of the Radiographic Assessment for Lung Edema Score in Patients With COVID-19 Acute Respiratory Distress Syndrome—An International Multicenter Observational Study. Frontiers in Medicine. 2022;8.

[32] Leevy JL, Khoshgoftaar TM, Bauder RA, Seliya N. A survey on addressing high-class imbalance in big data. Journal of Big Data. 2018;5(1).

[33] Krawczyk B. Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence. 2016;5(4):221–32.

[34] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition [Internet]. arXiv.org. 2015.

[35] Xu J, Pan Y, Pan X, Hoi S, Yi Z, Xu Z. RegNet: Self-Regulated Network for Image Classification. IEEE Transactions on Neural Networks and Learning Systems [Internet]. 2022;1–6.

[36] MHe K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016;770–8.

[37] Michele A, Colin V, Santika DD. MobileNet Convolutional Neural Networks and Support Vector Machines for Palmprint Recognition. Procedia Computer Science [Internet]. 2019;157:110–7.

[38] Le TM, Nguyen-Tat BT, Ngo VM. Automated evaluation of Tuberculosis using Deep Neural Networks. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems. 2022;8(30):e4.

[39] Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks [Internet]. arXiv.org. 2019.

[40] Mozaffari J, Amirkhani A, Shokouhi SB. A survey on deep learning models for detection of COVID-19. Neural Computing and Applications [Internet]. 2023;1–29.

[41] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. International Conference on Learning Representations. 2021.

[42] Ngo VM, Duong TVT, Nguyen TBT, Nguyen PT, Conlan O. An Efficient Classification Algorithm for Traditional Textile Patterns from Different Cultures Based on Structures. Journal on Computing and Cultural Heritage. 2021;14(4):1–22.

[43] Danilov, V.V., Litmanovich, D., Proutski, A. et al. Automatic scoring of COVID-19 severity in X-ray imaging based on a novel deep learning workflow. Sci Rep 12, 12791 (2022). https://doi.org/10.1038/s41598-022-15013-z