



Article

An Advanced Deep Learning Framework for Multi-Class Diagnosis from Chest X-ray Images

Maria Vasiliki Sanida ^{1,†}, Theodora Sanida ^{2,*}, Argyrios Sideris ² and Minas Dasygenis ²

¹ Department of Digital Systems, University of Piraeus, 18534 Piraeus, Greece; sanidasilia@gmail.com

² Department of Electrical and Computer Engineering, University of Western Macedonia, 50131 Kozani, Greece; asideris@uowm.gr (A.S.); mdasyg@ieee.org (M.D.)

* Correspondence: thsanida@uowm.gr; Tel.: +30-24610-56534

[†] These authors contributed equally to this work.

Abstract: Chest X-ray imaging plays a vital and indispensable role in the diagnosis of lungs, enabling healthcare professionals to swiftly and accurately identify lung abnormalities. Deep learning (DL) approaches have attained popularity in recent years and have shown promising results in automated medical image analysis, particularly in the field of chest radiology. This paper presents a novel DL framework specifically designed for the multi-class diagnosis of lung diseases, including fibrosis, opacity, tuberculosis, normal, viral pneumonia, and COVID-19 pneumonia, using chest X-ray images, aiming to address the need for efficient and accessible diagnostic tools. The framework employs a convolutional neural network (CNN) architecture with custom blocks to enhance the feature maps designed to learn discriminative features from chest X-ray images. The proposed DL framework is evaluated on a large-scale dataset, demonstrating superior performance in the multi-class diagnosis of the lung. In order to evaluate the effectiveness of the presented approach, thorough experiments are conducted against pre-existing state-of-the-art methods, revealing significant accuracy, sensitivity, and specificity improvements. The findings of the study showcased remarkable accuracy, achieving 98.88%. The performance metrics for precision, recall, F1-score, and Area Under the Curve (AUC) averaged 0.9870, 0.9904, 0.9887, and 0.9939 across the six-class categorization system. This research contributes to the field of medical imaging and provides a foundation for future advancements in DL-based diagnostic systems for lung diseases.



Citation: Sanida, M.V.; Sanida, T.; Sideris, A.; Dasygenis, M. An Advanced Deep Learning Framework for Multi-Class Diagnosis from Chest X-ray Images. *J* **2024**, *7*, 48–71. <https://doi.org/10.3390/j7010003>

Academic Editors: Shunxing Bao and Peng Zhang

Received: 3 December 2023

Revised: 15 January 2024

Accepted: 18 January 2024

Published: 22 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: deep learning framework; convolutional neural network; lung diseases; optimizing; chest X-ray imaging; multi-class diagnosis

1. Introduction

Lung diseases, such as fibrosis, opacity, tuberculosis, and pneumonia (viral and COVID), pose a significant global health burden, impacting the lives of countless individuals worldwide. These diseases are characterized by their detrimental effect on lung function, notably leading to a loss of lung elasticity. This decrease in elasticity results in a reduced total volume of air that the lungs can hold, consequently impairing respiratory function. The ability of some lung diseases to spread rapidly, especially in cases of infectious conditions like tuberculosis and pneumonia, underscores the critical need for prompt and accurate diagnosis. Early identification of these diseases is paramount, as it enables the timely initiation of appropriate treatment, essential in mitigating the spread of the disease and improving patient outcomes. The rapid and accurate diagnosis of lung diseases benefits individual patients by providing them with the necessary treatment. It plays a crucial role in public health by controlling the spread of infectious respiratory conditions [1–4].

In the dynamic and evolving landscape field of medical diagnostics, the integration of cutting-edge technologies has become essential, especially in the area of pulmonary

health. Within this context, chest X-ray imaging emerges as a fundamental tool. It offers a non-invasive and efficient approach to detecting and analyzing lung abnormalities. This imaging technique is of paramount importance for healthcare professionals. It facilitates the rapid and accurate diagnosis of a wide range of lung diseases, making it a key component in managing and treating pulmonary conditions. The strength of chest X-ray imaging lies in its ability to deliver clear and comprehensive images of the chest cavity. These detailed visualizations are crucial for clinicians in accurately detecting, monitoring, and addressing different lung pathologies. As advancements in medical science continue, the role of chest X-ray imaging in diagnosing and managing lung diseases is increasingly significant, underscoring the need for ongoing research and development in this vital area of healthcare [5–8].

In the last few years, the domain of medical image analysis has revolutionized with the introduction of deep learning (DL) methodologies. These approaches are attributable to the inherent capacity of DL to automate and enhance complex analytical processes, thus introducing novel prospects in medical diagnostics. The impact of DL is particularly pronounced in the area of chest radiology, where these advanced computational techniques have shown exceptional proficiency. DL algorithms, characterized by their sophisticated pattern recognition capabilities, have substantially improved the way chest images are interpreted, offering a more nuanced and accurate approach to detecting and diagnosing lung-related diseases. These techniques leverage large datasets of medical images, learning intricate patterns and anomalies that might elude traditional methods, thereby providing a more comprehensive and detailed understanding of pulmonary conditions. As such, DL in chest radiology not only represents a technological advancement but also marks a significant leap in the ability of medical professionals to diagnose and treat lung diseases with greater precision and effectiveness. This integration of DL into chest radiology augments not only the diagnostic accuracy but also fosters the development of personalized treatment strategies. Such advancements are pivotal in improving patient outcomes, marking a significant stride in managing and treating pulmonary health juncture [9,10].

This paper presents an innovative DL framework engineered for the multi-class diagnosis of lung diseases by analyzing chest X-ray images. Our objective is to address the increasing demand for efficient diagnostic tools and to harness the advanced capabilities offered by DL technologies. The proposed framework is a state-of-the-art convolutional neural network (CNN) architecture. This architecture is designed to extract and assimilate discriminative features from chest X-ray images, a process crucial for accurate lung disease identification. The CNN's ability to process and analyze complex visual data from X-rays enables it to identify subtle patterns and anomalies of various lung conditions, which might be challenging to discern through conventional diagnostic methods. The proposed framework provides more accurate, efficient, and reliable diagnostic solutions, and the reliability of the diagnostic outcomes is enhanced by the robustness of the model, which is trained and validated on an extensive dataset, ensuring consistent performance across a wide range of cases.

The most important contributions of this work are as follows:

- We propose an innovative adaptation of the VGG19 to enhance the diagnostic capabilities of this established CNN model, and we introduce the integration of custom blocks into the architecture. These blocks augment the network's capacity to encapsulate crucial image features, paramount in accurately classifying chest X-ray images. The custom blocks are designed to enhance the feature maps generated by the preceding layers through the CNN's normalization, regularization, and spatial resolution enhancement. This process results in a more comprehensive and nuanced representation of the image data, enabling the model to detect and differentiate between subtle and complex patterns indicative of different lung diseases.
- We manage the challenge of dataset imbalance, a common issue in medical imaging studies; our research focuses on a dataset comprising chest X-ray images categorized into six types: opacity, COVID-19, fibrosis, tuberculosis, viral pneumonia, and normal.

The original dataset exhibited a significant imbalance among these categories, a factor that can adversely impact the performance of DL models. To mitigate this issue and enhance the robustness of our model, we employed the data augmentation strategy. This technique involves artificially expanding the dataset by generating new, modified versions of existing images through various transformations such as rotation, scaling, and flipping. So, by applying these augmentations, we were able to transform the imbalanced dataset into a balanced one, ensuring that each class was equally represented.

The structure of this work is organized as follows: Section 2 is devoted to integrating generative AI with chest X-ray imaging. Section 3 reviews relevant studies in the field, providing the context to this topic. Section 4 details the proposed methodology, encompassing the collection used, the algorithm implemented, the data augmentation strategy, and the metrics for evaluating our approach. Section 5 is dedicated to presenting the experimental analysis and the outcomes obtained. In Section 6, the optimization strategies employed in our research are discussed. Finally, Section 7 concludes the article by discussing future research directions.

2. Integrating Generative AI with Chest X-ray Imaging

Integrating Generative AI with chest X-ray imaging for multi-class diagnosis stands at the forefront of medical innovation, creating a critical nexus between state-of-the-art computational methods and medical expertise. At the core of this integration is the strategic use of data augmentation techniques, which encompass a suite of manipulations such as horizontal flipping, brightness adjustments, shifts, rotations, zooms, shears, and changes in fill mode. These classical strategies are not merely tools for image manipulation but are pivotal in crafting a versatile and comprehensive dataset that challenges and refines the learning processes of DL models. These data augmentation techniques enrich the training data with various variations, replicating the diverse array of scenarios that a DL model would encounter in the real world. By systematically altering images through these techniques, the model is exposed to a broad spectrum of variations akin to the range it would need to interpret in a clinical environment. This exposure is critical, as chest X-rays inherently exhibit a degree of variability owing to numerous factors, such as differences in patient anatomy, positioning during the scan, the calibration of imaging equipment, and the intricacies of exposure settings. Each X-ray is a unique confluence of these factors, and a robust DL model must be capable of this variability to provide accurate diagnoses. In a clinical context, the ability of a model to generalize across various conditions and imaging nuances directly translates to its diagnostic utility. The performance of DL models on chest X-rays can significantly affect patient outcomes, as these models assist in early detection, accurate diagnosis, and timely treatment of pulmonary conditions. The multi-class diagnosis capability that Generative AI integration brings is precious in settings where a swift differential diagnosis is critical. Moreover, the variability introduced through data augmentation techniques aids in mitigating overfitting. In overfitting, a model performs exceptionally well on training data but fails to generalize to new, unseen data. By learning from augmented images that reflect a wider range of clinical scenarios, DL models develop a more robust understanding of the features that truly indicate specific diseases rather than artefacts of the dataset they were trained on [11,12].

Generative AI has emerged as a groundbreaking solution in managing one of the most persistent challenges in medical imaging: category disparities in collection. In diagnostic modelling, especially with chest X-ray imaging, collections often exhibit a significant inequality, with a preponderance of common illnesses overshadowing rarer pathologies. This imbalance risks developing biased or underperforming diagnostic models that excel at recognizing frequently occurring conditions but falter with less common ones. Such a skew in data can lead to diagnostic inaccuracies, potentially impacting patient care, especially for those with less common diseases that are underrepresented in the training data. As a result, the diagnostic models trained on these augmented datasets gain a more comprehensive

understanding of a wider array of pathologies, leading to improved identification and classification capabilities across the spectrum of disease [13,14].

AI-integrated Computer-Aided Diagnosis (AI-CAD) systems are designed to enhance the accuracy, speed, and efficiency of diagnosing diseases, thereby revolutionizing patient care and treatment outcomes. Integrating AI into CAD systems has opened up new frontiers in medical imaging analysis, offering powerful tools for detecting, characterizing, and monitoring various health conditions. At the core of AI-CAD systems is the application of advanced DL algorithms, which enable these systems to analyze complex medical images with a level of detail and accuracy previously unattainable. These AI models are trained on vast datasets of medical images, learning to recognize patterns and anomalies indicative of specific diseases. By doing so, AI-CAD systems assist radiologists and clinicians in making more informed diagnostic decisions, reducing the likelihood of human error and the variability that can occur in image interpretation. Furthermore, AI-CAD systems help to alleviate the workload on medical professionals. AI-CAD systems can rapidly process and analyze these images, highlighting areas of concern for further review by a radiologist. This speeds up the diagnostic process and allows radiologists to focus their expertise on more complex cases, improving overall healthcare delivery. Moreover, AI-CAD systems are continuously evolving. These systems learn and improve as they are exposed to more data, increasing their diagnostic accuracy. This continuous learning process ensures that AI-CAD systems remain at the forefront of medical technology, adapting to new challenges and advancements in healthcare [15,16].

3. Relative Work

The utilization of DL techniques in identifying abnormalities in chest X-ray images has attained significant traction in recent years. This surge in popularity is attributed to the remarkable capabilities of these algorithms in discerning intricate patterns and irregularities that might elude traditional analysis methods. In medical research, the application of artificial intelligence (AI) has become increasingly prominent, particularly in facilitating the diagnosis of various health conditions. Numerous studies leveraging AI in medical diagnostics have reported positive results, demonstrating both the accuracy and efficacy of these technologies [17–23]. This section delves into the strategies employed by previous researchers in this domain.

Sarkar et al. [24] proposed a multi-scale CNN model designed for a six-class categorization task, focusing on identifying tuberculosis, bacterial pneumonia, fibrosis, viral pneumonia, normal lung conditions, and COVID-19 using 5700 chest X-ray images. This study examines the efficacy of the VGG19 and the VGG16 models in their standard form and the VGG16 with multi-scale feature mapping forms. The standard VGG19 model achieved an accuracy of 95.61% and the VGG16 of 95.79%. However, when the VGG16 model was enhanced, the accuracy improved to 97.47%. In [25], the authors proposed a 2D-CNN model designed for a six-class categorization assignment, focusing on determining fibrosis, viral pneumonia, tuberculosis, bacterial pneumonia, normal lung conditions, and COVID-19 employing chest X-ray images. This work analyses the effectiveness of the VGG19 and the VGG16 models in their standard format and the 2D-CNN model. The standard VGG19 model reached an accuracy of 89.51%, the VGG16 of 90.43% and the 2D-CNN model of 96.75%. Also, in [26], the authors suggested a ResNet50 with deep features model designed for a five-class categorization assignment, focusing on determining viral pneumonia, tuberculosis, bacterial pneumonia, normal lung conditions, and COVID-19 using 2186 chest X-ray images. The model gained an accuracy score of 91.60%.

The study [27] proposes a DL model for multi-class categorization aimed at identifying pneumonia, COVID-19, normal, and lung cancer utilizing CT and chest X-ray images. This study examines the efficacy of four distinct architectural combinations, which integrate VGG19 and ResNet152V2 with various neural network models like CNN, GRU (Gated Recurrent Unit), and Bi-GRU (Bidirectional Gated Recurrent Unit). The accuracy achieved by the VGG19, combined with a CNN model, is 98.05%. In work [28], the authors sug-

gested a DL multi-class categorization model to determine COVID-19, viral pneumonia, normal, and lung opacity, employing chest X-ray images. This study analyses the effectiveness of the MobileNetV2 model in its standard format and the modified MobileNetV2 model. The standard MobileNetV2 model reached an accuracy of 90.47%, and the modified MobileNetV2 model of 95.80%. In [29], the authors propose a DL-based diagnostic system specifically designed to detect pneumonia utilizing X-ray images rapidly. This study compares the analysis of two prominent DL methods: VGG19 and ResNet50. These methods were evaluated for their efficacy in diagnosing three distinct conditions: pneumonia, COVID-19, and normal lung health. The findings of this study with the proposed diagnostic system have accuracy with the VGG19 method of 96.60%, while the ResNet50 method recorded an accuracy score of 95.80%. Furthermore, in [30], the authors suggested an altered VGG16 model designed for a three-class categorization assignment, focusing on determining pneumonia, normal lung conditions, and COVID-19 utilizing chest X-ray images. The altered VGG16 model earned an accuracy score of 91.69%.

Sanida et al. [31] proposed a DL model designed for a three-class categorization task, focusing on identifying pneumonia, normal lung conditions, and COVID-19 using chest X-ray images. This study examines the efficacy of the VGG19 model in its standard form and in modified forms that include the integration of inception blocks. The standard VGG19 model attained an accuracy of 98.17%. However, when the VGG19 model was enhanced with two inception blocks, the accuracy increased to 99.25%. Furthermore, incorporating a single inception block into the VGG19 model resulted in an accuracy of 98.59%. These results demonstrate the substantial impact that architectural modifications, such as the addition of inception blocks, can have on the performance of a DL model in medical image analysis. In [32], the authors explore the efficacy of a novel deep CNN method called Decompose, Transfer, and Compose (DeTraC). This technique is specifically developed to address the challenges of identifying anomalies in image datasets pertaining to pneumonia, SARS, and COVID-19. The study uses various established CNN models, including VGG19, GoogleNet, ResNet, AlexNet, and SqueezeNet. Each model is assessed for accurately categorizing anomalies within the dataset. The DeTraC with the VGG19 model attained an accuracy score of 97.35%.

Hemdan et al. [33] focused on binary categorization to differentiate between COVID-19 and healthy cases using chest X-ray scans. The study utilized a small dataset of 50 scans, divided into 25 scans representing COVID-19 cases and 25 from healthy individuals. Central to their research was the development of COVIDX-Net, a diagnostic system that leverages seven different pre-trained models. These models included VGG19, Xception, ResNetV2, InceptionV3, DenseNet201, InceptionResNetV2, and MobileNetV2. VGG-19 emerged as the most effective classifier among the seven models, reaching an accuracy of 90.00% and an F1-score of 0.91. Conversely, InceptionV3 was found to have the lowest accuracy in this study, with a rate of 50.00%. In [34], the authors propose an imaging-based fusion technique to differentiate between COVID-19 and healthy cases employing chest X-ray images. This method combines features extracted from chest X-ray images using two distinct processes: the histogram-oriented gradient (HOG) and the VGG-19 model. The HOG with the VGG19 model achieved an accuracy score of 99.49%.

In the work [35], the authors utilized four different DL models—ResNet50, DenseNet121, VGG16, and VGG19—applying the concept of transfer learning to diagnose X-ray images. The study aimed to differentiate between COVID-19 and normal lung conditions. Transfer learning, a method where a model developed for one assignment is reused as the starting point for a model on a second assignment, is particularly effective in systems where the available data is limited, as is often the case in medical imaging. The performance of the VGG16 and VGG19 models outperformed the other two DL strategies, ResNet50 and DenseNet121. The work reported an overall categorization accuracy of 97.00% for ResNet50, 96.66% for DenseNet121, and 99.33% for VGG16 and VGG19.

Numerous studies in the field of medical imaging have demonstrated impressive accuracy rates in scenarios involving binary or limited-class categorization. However,

a recurrent issue observed in these studies is a notable decline in performance when the number of categories to be classified increases. This decline in accuracy is primarily attributed to the heightened complexity involved in distinguishing between multiple conditions, particularly when these conditions exhibit only subtle differences in their features. Such a challenge becomes increasingly pronounced in multi-class categorization contexts, where the distinction between various lung diseases can be nuanced and complex. This inherent limitation significantly impacts the practical utility of these models in real-world clinical settings, where patients often present with a range of diverse lung conditions. In such scenarios, the ability to accurately categorize multiple lung diseases becomes not just beneficial but essential. Consequently, there is a pressing need for a specially designed and robust DL framework capable of performing multi-class categorization of lung diseases with a high accuracy rate and reliance. Such a framework would be invaluable in real-life clinical applications, enabling healthcare professionals to provide more accurate diagnoses and, therefore, more effective treatments for patients with complex lung conditions. This need underscores the importance of ongoing research and development in the field of DL to create more advanced and capable diagnostic tools that can meet the demands of modern healthcare. Table 1 summarises works for lung disease identification, the number of categories, the model employed, and the accuracy rate attained.

Table 1. A summary of works for lung diseases identification.

Work	Categories	Best Model	Accuracy (%)
[24]	6	VGG16+multi-scale	97.47
[25]	6	2D-CNN	96.75
[26]	5	ResNet50	91.60
[27]	4	VGG19+CNN	98.05
[28]	4	Modified MobileNetV2	95.80
[29]	3	VGG19	96.60
[30]	3	VGG16	91.69
[31]	3	VGG19+two inception blocks	99.25
[32]	3	DeTraC+VGG19	97.35
[33]	2	VGG19	90.00
[34]	2	VGG19+HOG	99.49
[35]	2	VGG19	99.33

4. Methodology

4.1. Chest X-ray Collection

In our work, the primary collection utilized for experimentation was the COVID-19 Radiography Database [36]. This comprehensive collection comprises 21,165 chest X-ray images, encompassing a diverse range of cases: 6012 show lung opacity, 3616 images are of COVID-19-positive cases, 10,192 depict normal lung conditions, and 1345 are of viral pneumonia. Additionally, to broaden the scope of our study, we incorporated images representing fibrosis with 1686 images and tuberculosis with 3500 sourced from the NIH Chest X-ray Dataset [37]. This integration of additional cases enhances the diversity and comprehensiveness of our collection. A representative sample of this extensive chest X-ray collection is illustrated in Figure 1, showcasing the variety of cases and conditions included in our study. This diverse collection is instrumental in training and evaluating our DL model, ensuring it is robust and effective in diagnosing a wide range of pulmonary conditions.

The distribution of lung diseases in the collection utilized is illustrated in Figure 1. Figure 2 displays a sample of a normal instance and five different conditions that may harm the lungs. Figure 2a showcases an opacity, which can be identified by areas of increased radiodensity. In this image, there are regions where the normally transparent appearance of the lung fields is obscured, indicating the presence of fluid, cells, or other substances that impede the passage of X-rays. Figure 2b represents a case of tuberculosis. Tuberculosis often manifests as well-defined nodules or consolidations, primarily in the upper lobes

of the lungs. The image may show scarring or calcification due to the infection, which is denser than the surrounding lung parenchyma. Figure 2c depicts fibrosis, characterized by a reticular pattern, with the lung architecture appearing distorted and retracted due to the fibrotic process. This can lead to a honeycombing appearance, with small cystic spaces surrounded by fibrous tissue. Figure 2d indicates a viral infection of the lungs, which can present as a more diffuse, bilateral interstitial pattern. This can result in haziness across the lung fields without a single dominant focus of consolidation, which is often more pronounced in the peripheral areas of the lungs. Figure 2e indicates COVID-19, which typically presents with bilateral peripheral ground-glass opacities and may include a consolidation pattern characterized by interlobular septal thickening superimposed on the ground-glass opacities. Lastly, Figure 2f is a normal lung X-ray, which serves as a control image against which the pathological images can be compared. It exhibits clear lung fields without abnormal opacification, well-defined diaphragms, and sharp costophrenic angles, which indicate the absence of disease.

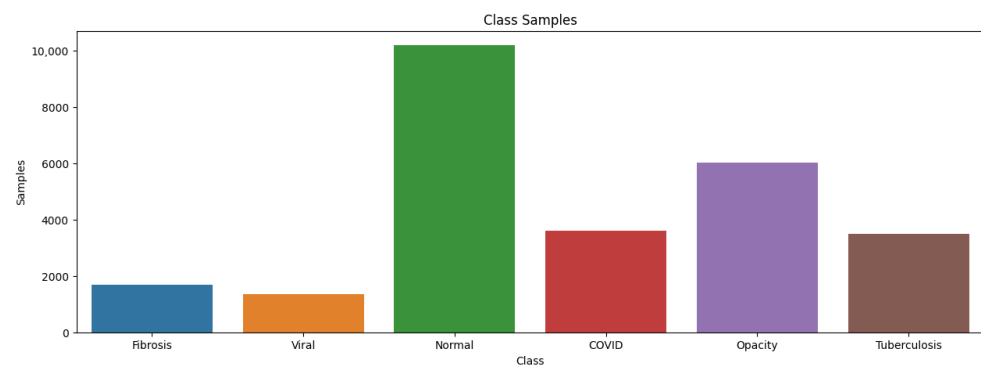


Figure 1. The distribution of lung diseases in the collection [36,37].

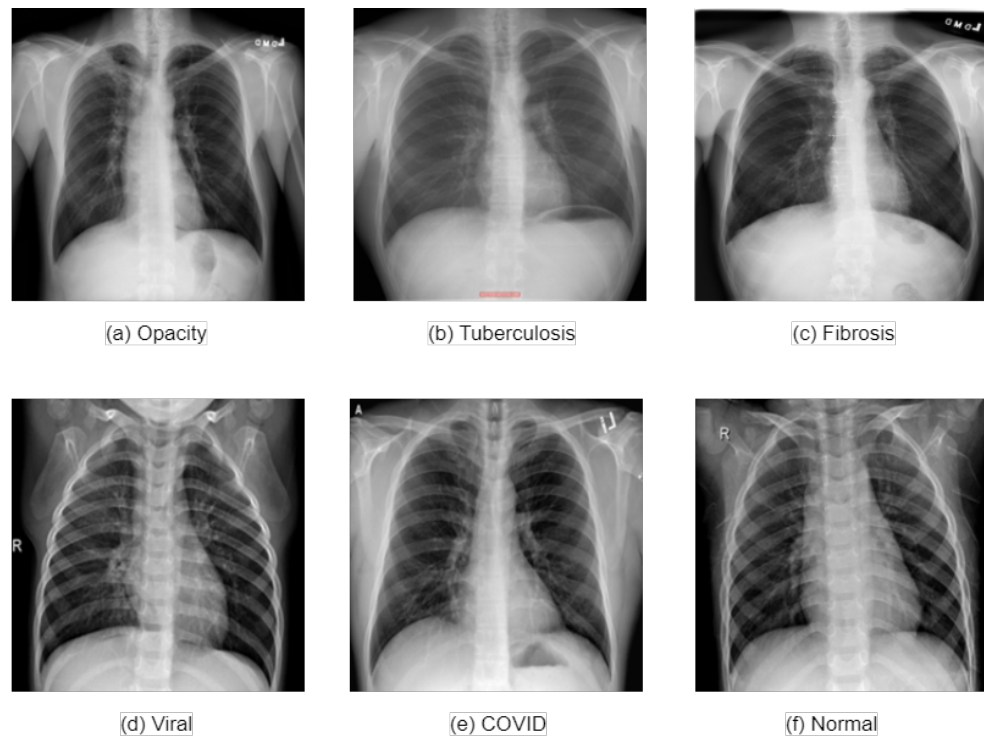


Figure 2. Samples by category from the collection (a–f) [36,37].

4.2. Collection Splitting

In our study, 80% of the collection is designated for the training phase. Within this portion, we further divided the collection into two segments: 60% was used for direct training purposes, while the remaining 20% was set aside for validation. This validation segment is crucial in DL models and ensuring their accuracy and reliability. The remaining 20% of the collection was reserved for the testing phase, which is critical for evaluating the performance of DL models under conditions that simulate real-world scenarios [38]. The collection distribution into these training, validation, and testing segments is comprehensively illustrated in Table 2.

Table 2. The collection distribution.

Condition	Training Collection	Validation Collection	Testing Collection
Opacity	3833	958	1221
Tuberculosis	2264	566	670
Fibrosis	1094	274	318
Viral pneumonia	867	216	262
COVID	2291	572	753
Normal	6516	1629	2047
Total	16,865	4215	5271

4.3. Image Preprocessing and Augmentation

Image preprocessing is a critical step to ensure compatibility with the CNN architecture. Initially, all images in the collection are resized to a uniform dimension of 224×224 pixels. This standardization is essential as larger images might obscure the critical traits necessary for accurate diagnosis. Following the resizing process, pixel values in all images are normalized to a range between 0 and 1. This normalization step is crucial for harmonizing the input data and enhancing the CNN's ability to process the images effectively [39].

Moreover, image augmentation techniques [12] are employed to tackle the challenge posed by the limited quantity of images in the training collection and enhance the efficiency of the training process. These techniques include various transformations for expanding the training collection and introducing a diversity of image orientations and scales. This diversity is vital in preventing the model from overfitting to the training collection, thereby ensuring that the model generalizes well to new, unseen data.

In our study, we have employed a series of data augmentation techniques (except normal lung conditions) to enhance the quality of the training collection, as detailed in Table 3. So, each category of training collection has 6516 images, such as normal lung conditions. These techniques, pivotal in increasing the diversity and robustness of the training collection, include height shift, zoom, random rotation, brightness range, horizontal flipping, shearing transformation, fill mode, and width shift. The combination of resizing, normalization, and augmentation [40,41] strategies is pivotal in optimizing the training process and improving the overall performance and reliability of the CNN model in diagnosing lung conditions.

Figure 3 provides visual examples of these data augmentation techniques applied to the training collection. Each technique introduces specific changes that challenge the model to learn the essential features of the anatomical structures, regardless of these variations, ensuring that the model's performance is robust across a wide range of imaging conditions. The horizontal flip technique mirrors the original image along the vertical axis. This simulates the scenario where an image could be oriented differently, and the model must recognize structures regardless of their left–right orientation. The brightness range technique adjusts the intensity of the pixels in the image, simulating variations in exposure levels that can occur during X-ray image acquisition. This ensures that the model is not overly reliant on specific brightness levels for feature identification. Width shift and height

shift techniques involve translating the image horizontally or vertically, respectively. This simulates slight patient movements or different positioning that can result in shifts in the anatomical structures in the images. The rotation augmentation applies a slight angular rotation to the image, which helps the model learn to identify features that may not be perfectly aligned due to variations in patient positioning during the X-ray procedure. Zoom applies a uniform scaling to the image, enlarging or shrinking it. This can be reflective of the varying sizes of patients or the distance between the X-ray source and the patient during image capture. The shear transformation applies a shearing effect, skewing the image. This can mimic the effect of angled perspectives, where the anatomy appears distorted due to the imaging angle relative to the body. Lastly, fill mode deals with how to handle newly introduced pixels in transformations that change the geometry of the image, such as rotation or width/height shifts. This technique ensures that the model is not confused by artificial pixel values that do not represent true anatomical structures.

Table 3. Details of data augmentation strategies in the training collection.

Parameter	Value
Horizontal flip	True
Brightness Range	[0.5, 1.30]
Width shift	[0.7, 1.25]
Rotation	[+30, −30]
Zoom	[0.4, 0.9]
Height shift	0.25
Sheare	0.35
Fill mode	Nearest

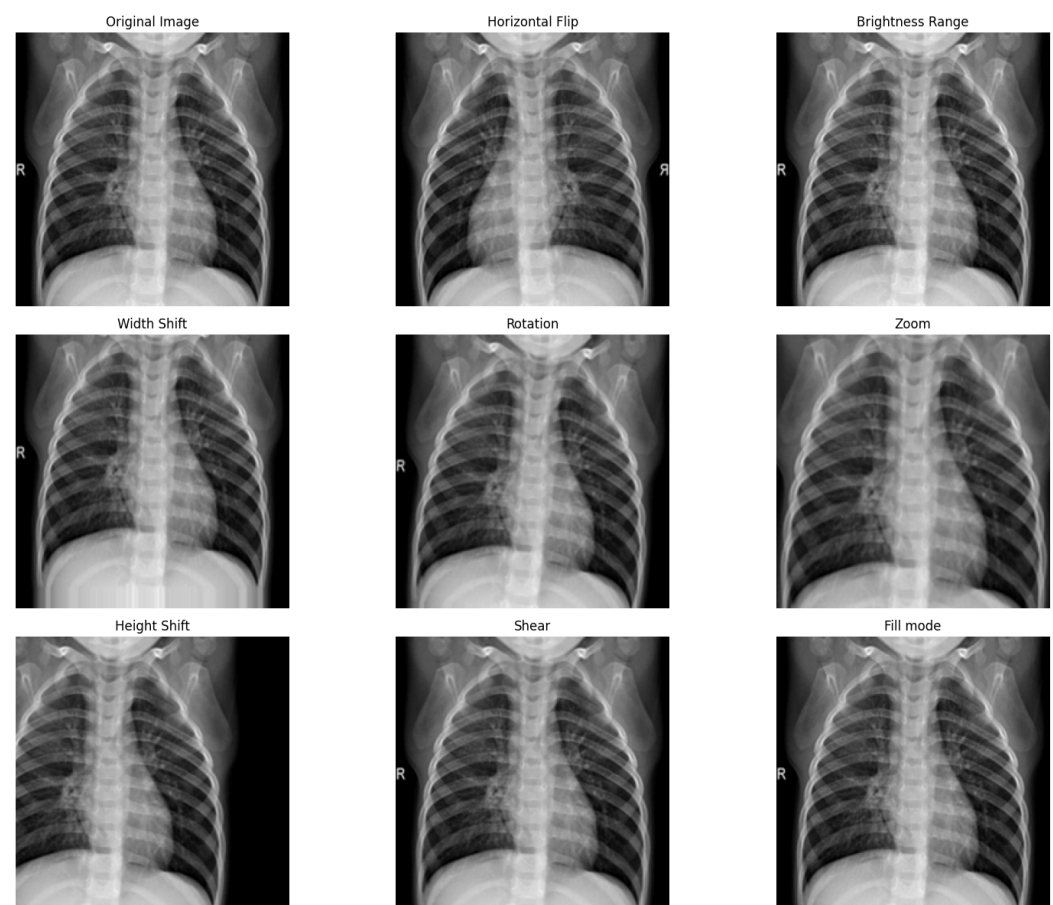


Figure 3. Example of data augmentation techniques for viral pneumonia.

4.4. Proposed Modified VGG19 Model

In this study, we introduce a tailored modification to the VGG19 architecture to enhance its capability to extract deep features. VGG19 stands out as a DL model for its notable depth, featuring 19 layers, which positions it among the deeper architectures in the VGG [42] series. Implementing small convolutional filters across the network is instrumental in capturing intricate details within image data. This structural design enables VGG19 to learn complex features at multiple levels of abstraction, a characteristic that has proven highly beneficial in a variety of image recognition assignments. The standard VGG19 model comprises 19 layers, including 16 convolutional layers and 3 fully connected layers, supplemented with 4 max-pooling layers as shown in Figure 4.

The modification we propose focuses on the outputs of the last 3 max-pooling layers, redirecting them through a newly incorporated series of layers. This series comprises (1) batch normalization, (2) dropout, and (3) up-sampling. Batch normalization [43] improves the neural network's stability and performance, normalizing the previous layer's output to a standard scale. This process is critical in addressing the internal covariate shift, whereby the distribution of network activations varies significantly during training, thus impeding the model's learning efficiency. By applying batch normalization, we actively recalibrate the outputs from the network's previous layers to adhere to a standard scale. This recalibration is a profound transformation that normalizes the input for each mini-batch, ensuring consistency in the input data distribution as it flows through successive layers. Such normalization proves instrumental in accelerating the training process, as it allows for higher learning rates and reduces the sensitivity to the initial weights. This is particularly advantageous in the medical imaging domain, where the heterogeneity and complexity of the data can be a substantial barrier. For instance, lung X-ray images present many subtle variances due to differences in patient anatomy, the position during X-ray capture, and the inherent characteristics of various lung pathologies. These subtle variances can manifest as minute pixel intensity and contrast differences, challenging conventional neural network models without batch normalization. By implementing batch normalization, our model gains increased stability during training, manifesting in enhanced performance and generalizability when processing lung X-ray images. This stability is pivotal in ensuring that the internal dynamics of the network do not overshadow the subtle nuances of lung pathologies. Instead, they are captured, retained, and emphasized throughout the learning process, leading to a model that is robust and remarkably sensitive to the intricacies of lung disease presentations.

Following batch normalization, our model employs a dropout [44] layer, a regularization method to prevent overfitting. Overfitting occurs when a neural network model becomes excessively complex, capturing noise and spurious details in the training data that do not generalize to new, unseen data. In medical imaging, where the diversity of presentations and the subtlety of pathological features are vast, overfitting can drastically undermine the model's utility in clinical settings. The dropout layer addresses this issue by introducing a form of controlled randomness during the training process. With each iteration, a specified proportion of the neuron outputs is randomly set to zero, effectively dropping out those units from the network. This random omission of neuron outputs compels the network to develop a more distributed data representation. The underlying principle is to prevent the network from becoming overly reliant on any particular set of features, promoting the learning of robust and redundant feature representations. The dropout layer encourages the neural network to learn to classify lung pathologies using a diverse array of features from across the entirety of its architecture. By doing so, the network becomes adept at recognizing patterns indicative of disease, even when some data points are missing or obscured. This is common in real-world medical imaging due to factors such as patient movement or variable imaging conditions. Moreover, the utilization of dropout in neural networks mirrors the principles of ensemble learning within a single model architecture, akin to having multiple models contribute to the outcome. Each pass through the network during training uses a different version of the network, resulting in

a model that is less sensitive to the idiosyncrasies of the training data and more capable of generalizing from the learned patterns to accurately diagnose new and unseen X-ray images. This aspect of dropout is particularly salient in medical imaging analysis, ensuring that the model retains its diagnostic accuracy not just on the data it was trained on but also on future clinical cases.

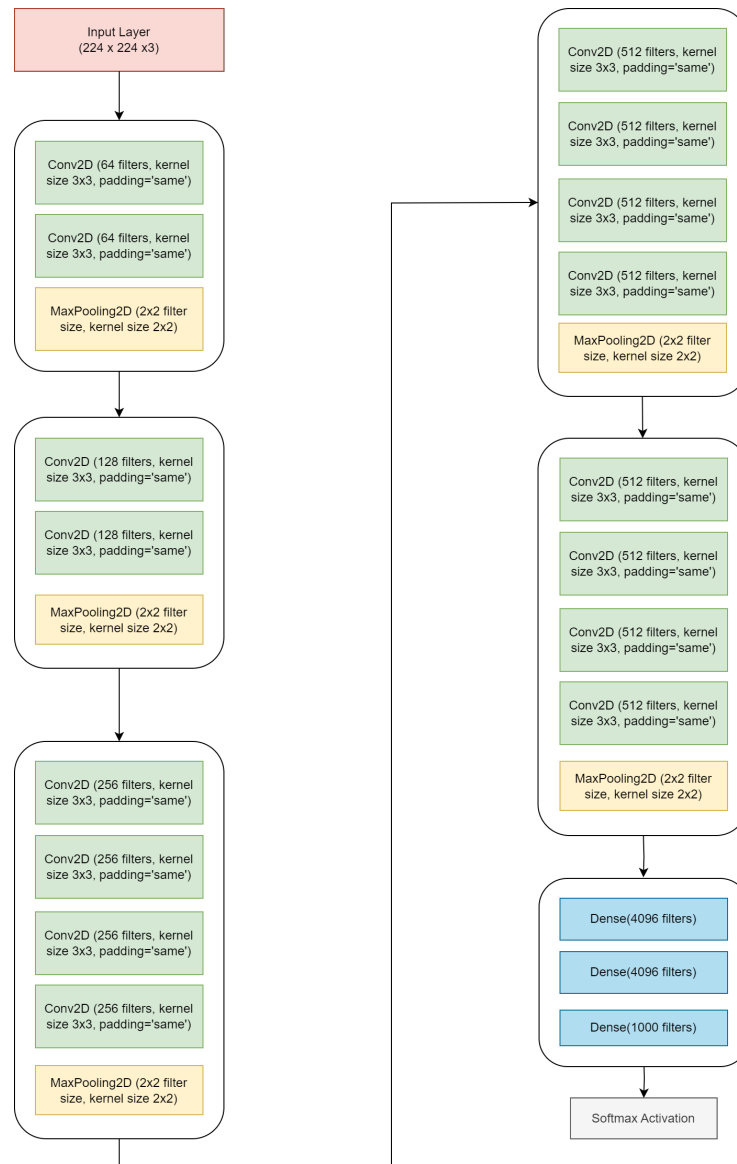


Figure 4. The block diagram of the basic VGG19 model.

Finally, the up-sampling [45] layer plays a crucial role in this modified architecture. The up-sampling layer's fundamental role emerges after the VGG19 convolutional layers have successfully extracted the depth and complexity of features from the input images. While adept at feature extraction, these convolutional layers inevitably reduce the spatial dimensions of the input due to max pooling. Max pooling results in losing finer spatial details crucial for precise medical diagnostics. The up-sampling process is specifically engineered to counteract this reduction in spatial resolution. It works by effectively increasing the size of the feature maps, thereby restoring the spatial dimensions that were compressed during max pooling. This restoration is not merely a scaling up of the feature map; it is a process that aims to reconstruct the critical spatial details that are often lost in lower resolutions. In lung X-ray analysis, this spatial detail reconstruction is paramount. Lung pathologies, such as nodules, opacities, or other anomalies, often manifest as subtle and

minute variations in the X-ray images. The ability of the up-sampling layer to enlarge the feature maps without losing these vital details allows the model to maintain the integrity of such critical information. This maintenance and enhancement of spatial resolution ensure that the model can accurately identify and analyse these pathologies, which might be missed or inaccurately represented in lower resolutions. Furthermore, the up-sampled feature maps allow the model to understand better the spatial relationships and structures within the lung images. This understanding is crucial when distinguishing between various lung conditions, where abnormalities' location, shape, and size can indicate specific diseases. By providing a more detailed and nuanced view of the lung's internal structures, the up-sampling layer significantly contributes to the model's diagnostic accuracy, making it an invaluable tool in the early detection and analysis of lung diseases.

Our approach ensures that the detailed, nuanced features identified early in the feature extraction process are not lost or overly processed in subsequent stages. As a result, the modified VGG19 model strikes a balance between maintaining depth and enhancing efficiency. It is specifically designed to optimize the extraction of deep features, rendering it highly suitable for advanced image analysis tasks that require a nuanced understanding of complex image data. The detailed architecture of the modified VGG19 model is presented in Table 4, and Figure 5 illustrates the block diagram of the model.

Table 4. Detailed architecture of modified VGG19 model.

Layer (Type)	Output Shape	Param #
input_1 (Input Layer)	(None, 224, 224, 3)	0
conv2d (Conv2D)	(None, 224, 224, 64)	1792
conv2d_1 (Conv2D)	(None, 224, 224, 64)	36,928
max_pooling2d (MaxPooling2D)	(None, 112, 112, 64)	0
conv2d_2 (Conv2D)	(None, 112, 112, 128)	73,856
conv2d_3 (Conv2D)	(None, 112, 112, 128)	147,584
max_pooling2d_1 (MaxPooling2D)	(None, 56, 56, 128)	0
conv2d_4 (Conv2D)	(None, 56, 56, 256)	295,168
conv2d_5 (Conv2D)	(None, 56, 56, 256)	590,080
conv2d_6 (Conv2D)	(None, 56, 56, 256)	590,080
conv2d_7 (Conv2D)	(None, 56, 56, 256)	590,080
batch_normalization (BatchNorm)	(None, 56, 56, 256)	1024
dropout (Dropout)	(None, 56, 56, 256)	0
up_sampling2d (UpSampling2D)	(None, 112, 112, 256)	0
conv2d_8 (Conv2D)	(None, 112, 112, 512)	1,180,160
conv2d_9 (Conv2D)	(None, 112, 112, 512)	2,359,808
conv2d_10 (Conv2D)	(None, 112, 112, 512)	2,359,808
conv2d_11 (Conv2D)	(None, 112, 112, 512)	2,359,808
batch_normalization_1 (BatchNorm)	(None, 112, 112, 512)	2048
dropout_1 (Dropout)	(None, 112, 112, 512)	0
up_sampling2d_1 (UpSampling2D)	(None, 224, 224, 512)	0
conv2d_12 (Conv2D)	(None, 224, 224, 512)	2,359,808
conv2d_13 (Conv2D)	(None, 224, 224, 512)	2,359,808
conv2d_14 (Conv2D)	(None, 224, 224, 512)	2,359,808
conv2d_15 (Conv2D)	(None, 224, 224, 512)	2,359,808
batch_normalization_2 (BatchNorm)	(None, 224, 224, 512)	2048
dropout_2 (Dropout)	(None, 224, 224, 512)	0
up_sampling2d_2 (UpSampling2D)	(None, 448, 448, 512)	0
global_average_pooling2d (GlobalAvg)	(None, 512)	0
dense (Dense)	(None, 256)	131,328
dense_1 (Dense)	(None, 128)	32,896
dense_2 (Dense)	(None, 6)	774
Total params:		20,194,502
Trainable params:		20,191,942
Non-trainable params:		2560

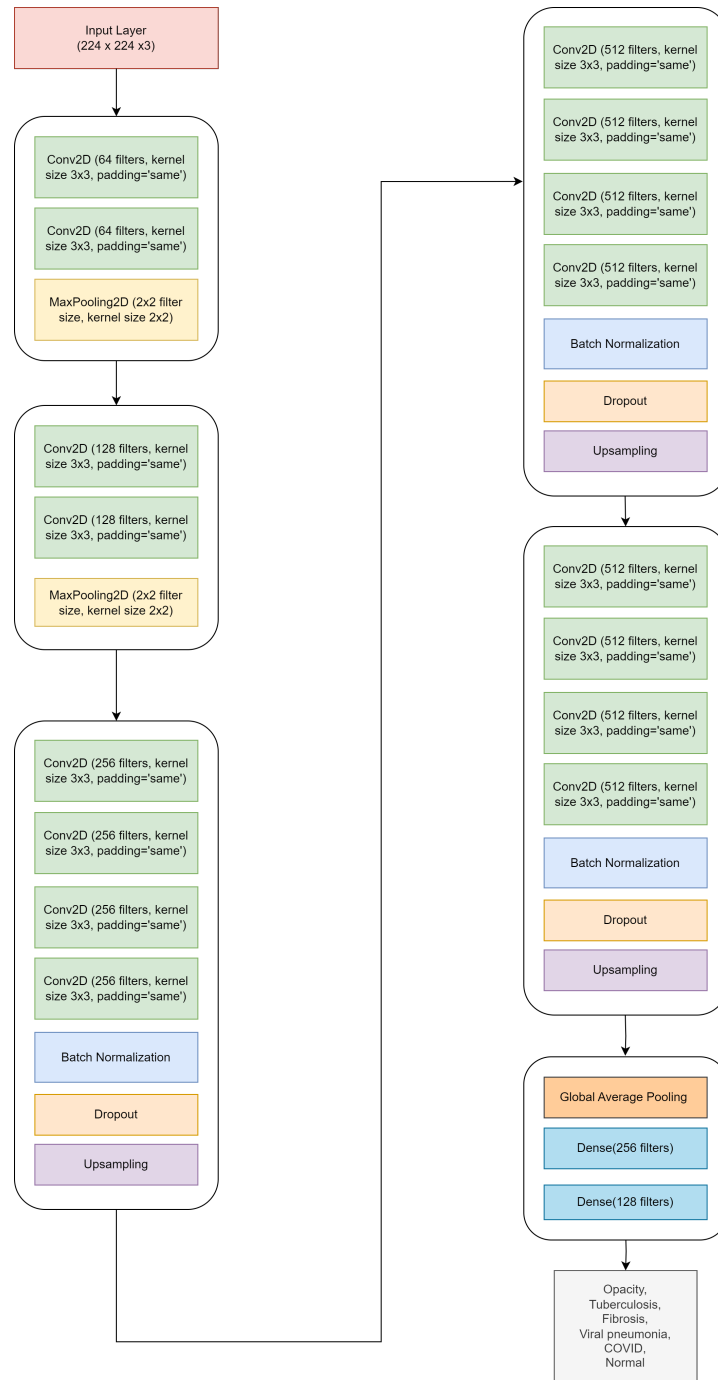


Figure 5. The block diagram of the modified VGG19 model.

4.5. Implementation Settings

Table 5 offers a detailed account of the computational environment and the hyperparameters chosen to execute our experiments. The experiments were performed on a system running Windows 10 Pro, with 16 GB of RAM. We used the NVIDIA RTX 3050 GPU model with 8 GB of onboard memory. Python language we used, and the back-end framework was the Keras package with TensorFlow, which provides a high-level, user-friendly API for constructing and training models. Our model was trained for 30 epochs to ensure sufficient training without excessive overfitting. The Adam optimizer was chosen for its adaptiveness in updating network weights. This optimization algorithm is known for its efficiency with large collections and high-dimensional parameter spaces, often in image categorization tasks. A mini-batch size of 32 was selected for the better convergence properties, and a

learning rate of 0.0001 was set, allowing the model to converge gradually. The loss function used was cross-entropy, a standard choice for categorization problems.

Table 5. Implementation settings for our experiments.

Name	Values/Types
Operating system	Windows 10 Pro
RAM	16 GB
GPU	NVIDIA RTX 3050 8 GB
Language	Python
Backend	Keras package with TensorFlow
Number of epochs	30
Optimizer	Adam
Mini batch size	32
Learning rate	0.0001
Loss function	Cross-entropy

4.6. Quality Measures

In the domain of DL, the assessment of a model's aptitude for a given task is contingent upon certain evaluative quality measures [46–48]. These measures are indispensable in ascertaining the model's capacity to generalize from the training collection to unseen data—essentially, its predictive power in real-world systems. Accuracy, precision, recall, and F1-score offer insights into various aspects of model performance, such as its overall correctness, ability to minimize false positives, effectiveness in identifying all relevant examples, and balance between precision and recall, respectively.

As delineated by Equation (1), accuracy quantifies the percentage of images the model correctly categorized to the total number of images assessed. As specified by Equation (2), precision assesses the proportion of true positives within the subset of samples labelled as positive by the model. Recall, also known as sensitivity and detailed in Equation (3), gauges the model's ability to identify all relevant samples within the actual positive category accurately. Additionally, the f1-score is a composite metric that harmonizes precision and recall, providing a singular measure of the classifier's exactitude, as shown by Equation (4).

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

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (3)$$

$$\text{F}_1\text{-score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (4)$$

where TNs, or true negatives, denotes the count of negative examples the model accurately categorizes as negative. FPs, or false positives, represent instances where the model erroneously categorizes negative examples as positive. TPs, or true positives, corresponds to the number of positive examples the model correctly recognizes. Conversely, FNs, or false negatives, indicates examples where the model incorrectly categorizes positive examples as negative.

5. Investigation Outcomes

Our methodology was evaluated using the testing collection composed of images completely unseen during training to ensure an unbiased and objective assessment of the performance of the basic and the modified VGG19 model. Additionally, the testing collection was isolated from image augmentation techniques that might influence the

model's learning in a way that could give an advantage to this specific collection. With testing collection, we can simulate the model's deployment in a real-world scenario with new patient images. The performance of these unseen images provides a measure of the model's true diagnostic ability and indicates how it would function when applied in a clinical setting.

5.1. Accuracy and Loss Curves

Figure 6 illustrates that for the modified VGG19 model, the training accuracy curve maintains a consistently high trajectory, which points to the model's ability to fit the training data well. The validation accuracy curve, although slightly lower, follows a similar trend, reflecting that the model generalizes well to new image data. On the loss curve side, the model effectively minimizes the error between its predictions and the actual labels. The validation loss demonstrates some variability but generally follows the downward trend of the training loss, which corroborates the model's capacity to generalize without overfitting.

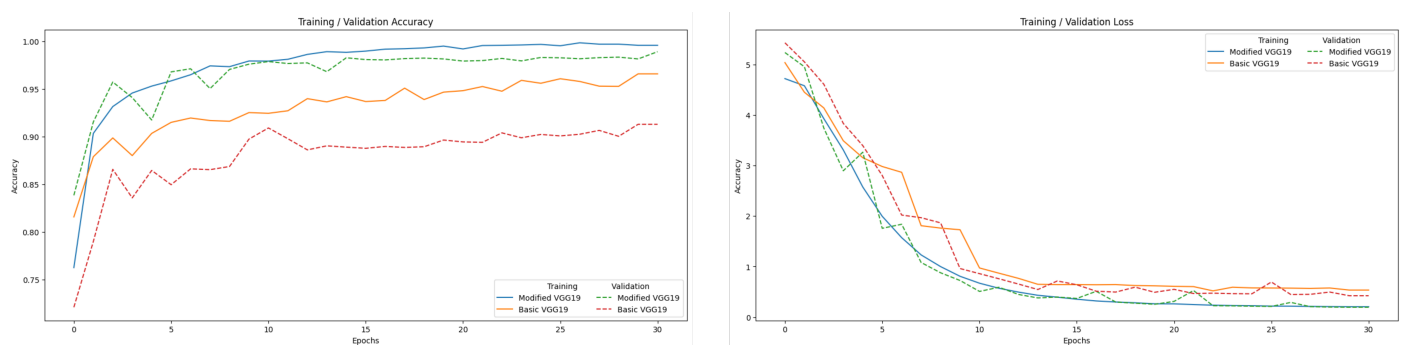


Figure 6. Accuracy and loss curves for the basic and the modified VGG19 model.

5.2. Classification Report

The classification report shown in Table 6 for the modified VGG19 model offers a comprehensive view of its performance metrics across different categories in the context of lung X-ray image analysis. The precision metric reflects the accuracy of positive predictions for each category. It signifies the proportion of true positives among all instances categorized as positives by the model. For instance, the precision 1.0000 for COVID indicates that the model perfectly identifies COVID cases without any false positives. Similarly, high precision values in other classes, like 0.9933 for opacity, denote a high level of reliability in the model's positive predictions for these conditions. Recall, another critical metric, measures the model's capability to correctly identify all samples of a given category. It is the proportion of actual positives that were correctly identified. The recall of 1.0000 for the sample indicates that the model successfully recognized all tuberculosis cases in the test collection. High recall values across categories suggest that the model is highly sensitive and effective in detecting the presence of these lung conditions. The F1-score, a harmonic mean of precision and recall, provides a single measure that balances both the precision and recall. It is particularly useful when the distribution of class instances is uneven or when the cost of false positives and false negatives varies. For example, the F1-score of 0.9973 for COVID implies a near-perfect balance between precision and recall, indicating the model's exceptional performance in identifying COVID cases. Similarly, high F1-scores for other classes like fibrosis 0.9752 and normal 0.9878 highlight the model's overall robustness and effectiveness.

Table 6. Classification report for the modified VGG19 model.

Category	Precision	Recall	F1-Score
Opacity	0.9933	0.9738	0.9835
Tuberculosis	0.9955	1.0000	0.9978
Fibrosis	0.9604	0.9906	0.9752
Viral	0.9886	0.9924	0.9905
COVID	1.0000	0.9947	0.9973
Normal	0.9845	0.9912	0.9878

The classification report shown in Table 7 for the basic VGG19 model shown for opacity, the precision is 0.9696, suggesting that the model is highly accurate in its predictions for this category, the recall for tuberculosis is 0.9985, indicating that the model is exceptionally good at identifying all cases of tuberculosis, fibrosis has an F1-score of 0.9158, indicating a solid balance between precision and recall.

Table 7. Classification report for the basic VGG19 model.

Category	Precision	Recall	F1-Score
Opacity	0.9696	0.9394	0.9542
Tuberculosis	0.9941	0.9985	0.9963
Fibrosis	0.8635	0.9748	0.9158
Viral	0.9353	0.9924	0.9630
COVID	0.9893	0.9867	0.9880
Normal	0.9674	0.9580	0.9627

The comparison between the classification reports in Tables 6 and 7 of the modified and basic VGG19 models reveals significant improvements in the modified model across all categories. The modified VGG19 demonstrates superior precision, recall, and F1-score in the opacity category, indicating a more accurate and reliable performance in the tuberculosis category, where the modified model not only improves precision but also achieves a perfect recall, leading to a higher F1-score. The improvements are particularly striking in the fibrosis category, where the modified VGG19 shows a substantial increase in precision while maintaining high recall, resulting in a markedly better F1-score. Similarly, for the viral category, the modified model exhibits a notable enhancement in precision without compromising the recall and an improved overall F1-score. In the case of COVID, the modified VGG19 achieves perfect precision and slightly better recall and F1-score, underscoring its enhanced diagnostic accuracy. Finally, the modified model outperforms the basic model in all metrics for the normal category, indicating a more consistent and reliable performance. These outcomes suggest that the modifications made to the VGG19 model have significantly bolstered its capabilities, making it a more robust tool for accurate image categories of lung diseases.

5.3. Overall Performance

Table 8 encapsulates the overall performance of the basic and the modified VGG19 model as applied to the categorization of lung diseases utilizing chest X-ray images. The accuracy of the modified VGG19 stands at 98.88%, a substantial increase from the already high 96.57% of the basic model. This high level of accuracy indicates the modified model's robustness and potential as a reliable diagnostic aid. Precision shows a remarkable improvement in the modified model, 0.9870, compared to the basic model, 0.9532. This precision is crucial in medical diagnostics to avoid misdiagnosis and to ensure appropriate treatment planning. In terms of recall, the modified VGG19 achieves 0.9904, surpassing the basic model's 0.9750. This means the modified VGG19 successfully identified 99.04% of all actual cases of lung diseases, a crucial capability in medical imaging where missing a positive diagnosis can have significant consequences. The F1-score, which balances precision and

recall, is higher in the modified model 0.9887 than in the basic model 0.9633, reinforcing the model's balanced performance in correctly identifying cases and minimizing false positives or negatives. Lastly, the AUC metric is significantly higher in the modified VGG19 0.9939 than in the basic model 0.9836. This high AUC value indicates the modified model's ability to distinguish between different lung conditions, affirming its utility in a clinical setting where such differentiation is critical to appropriate patient care.

Table 8. Overall performance of the basic and the modified VGG19 model.

Average	Basic VGG19	Modified VGG19
Accuracy	96.57%	98.88%
Precision	0.9532	0.9870
Recall	0.9750	0.9904
F1-score	0.9633	0.9887
AUC	0.9836	0.9939

5.4. Confusion Matrix

The confusion matrix displayed in Figure 7 illustrates the performance of the modified VGG19 model on the test collection of lung disease images. Each row of the matrix corresponds to the actual category, while each column represents the model predictions. The model correctly identified 749 cases as COVID, with 4 cases that were actually COVID incorrectly categorized as something else. There were no non-COVID cases miscategorised as COVID. All 315 fibrosis cases were correctly identified, with 2 recategorizations as normal and 1 as opacity. Most normal cases were correctly categorized (2029 out of 2047), with a small number of subcategorization as other diseases—6 as fibrosis, 7 as opacity, 2 as tuberculosis, and 3 as viral pneumonia. The model correctly identified 1189 out of 1221 opacity cases, with some confusion where 7 cases were miscategorised as normal and 24 as fibrosis. Only 1 case of opacity was miscategorised as tuberculosis. The model identified all 670 tuberculosis cases with no subcategorization. The model also performed exceptionally well in categorizing viral pneumonia cases, with 260 out of 262 cases correctly identified and only 2 cases subcategorization as normal. From the outcomes, the confusion matrix suggests that the modified VGG19 model exhibits high accuracy in categorizing different lung diseases, which is essential for developing reliable AI-assisted diagnostic tools in healthcare.

The confusion matrix for the basic VGG19 model in Figure 8 for COVID, identified 743 true positives. While well identified with 310 correct predictions, fibrosis is confused with normal, suggesting some feature overlap between the two categories. The normal category has the highest number of instances and is identified correctly with 1961 true positives, although there are notable confusions with opacity and fibrosis. Opacity is predominantly identified correctly with 1147 true positives, but some confusion with normal and fibrosis could indicate similar visual features that the model struggles to differentiate. Tuberculosis is accurately identified with 669 true positives and almost no recategorisations, demonstrating the model's effectiveness for this category. Viral infections are perfectly identified with 260 true positives, indicating that the model has learned distinguishing features for this category very well. The basic VGG19 model is highly effective at diagnosing various conditions, with particularly strong performance in distinguishing tuberculosis and viral infections. However, there is room for improvement in differentiating between normal, opacity, and fibrosis cases, as evidenced by the misclassifications between these categories.

		Proposed CNN - Confusion Matrix					
Actual	COVID	749	0	4	0	0	0
	Fibrosis	0	315	2	1	0	0
	Normal	0	6	2029	7	2	3
	Opacity	0	7	24	1189	1	0
	Tuberculosis	0	0	0	0	670	0
	Viral	0	0	2	0	0	260
		COVID	Fibrosis	Normal	Opacity	Tuberculosis	Viral
		Predicted					

Figure 7. The confusion matrix of lung diseases in the test collection for the modified VGG19 model.

		VGG19 - Confusion Matrix					
Actual	COVID	743	0	7	1	1	1
	Fibrosis	0	310	3	5	0	0
	Normal	6	31	1961	30	2	17
	Opacity	1	18	54	1147	1	0
	Tuberculosis	1	0	0	0	669	0
	Viral	0	0	2	0	0	260
		COVID	Fibrosis	Normal	Opacity	Tuberculosis	Viral
		Predicted					

Figure 8. The confusion matrix of lung diseases in the test collection for the basic VGG19 model.

5.5. Receiver Operating Characteristic (ROC) Curves

The ROC curves displayed in Figure 9 are a powerful tool for evaluating the diagnostic ability of the modified VGG19 model for lung disease categorization employing X-ray images. The curve for COVID, with an AUC of 0.9973, indicates a high true positive rate and a low false positive rate, which shows excellent model performance in COVID detection. Similarly, the AUC for tuberculosis stands at 0.9997, which suggests that the

model can distinguish between X-ray images with tuberculosis and those without it. The ROC curves for fibrosis, normal, opacity, and viral pneumonia also show high AUC values (all above 0.98), indicating that the model performs very well across these conditions.

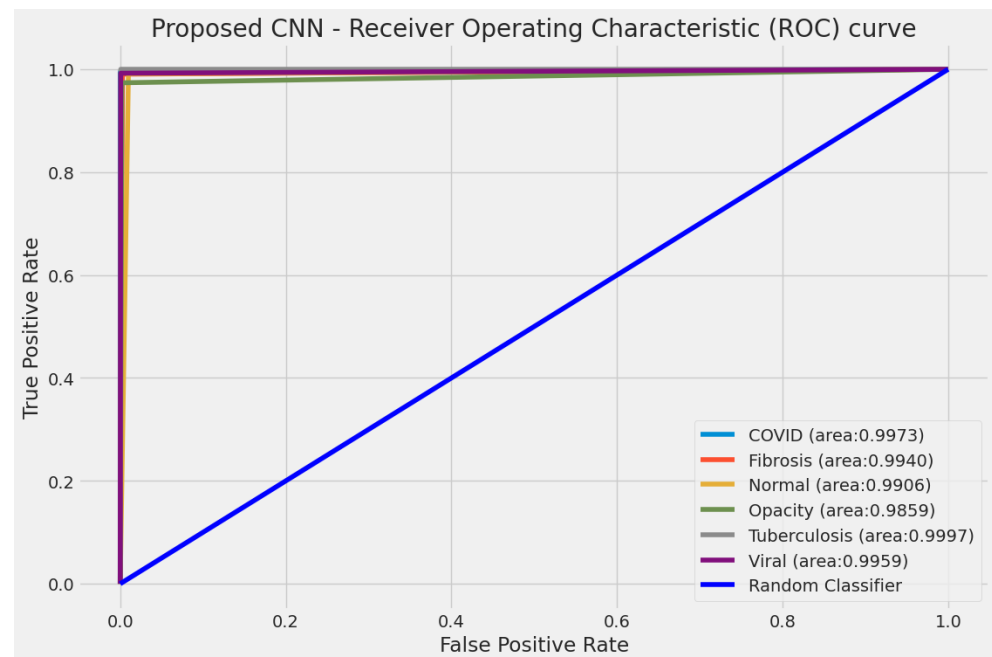


Figure 9. The ROC curves of lung diseases in the test collection for the modified VGG19 model.

The ROC curves for the basic VGG19 model are depicted in Figure 10. The curve for COVID, with an AUC of 0.9925, displays a high true positive rate and a low false positive rate, indicating the superior basic model performance in COVID identification. Also, the AUC for tuberculosis stands at 0.9988, which implies that the basic model can differentiate between X-ray images with tuberculosis and those without it. Furthermore, the ROC curves for fibrosis, normal, opacity, and viral pneumonia display high AUC values (above 0.96), suggesting that the basic model manages these situations well.

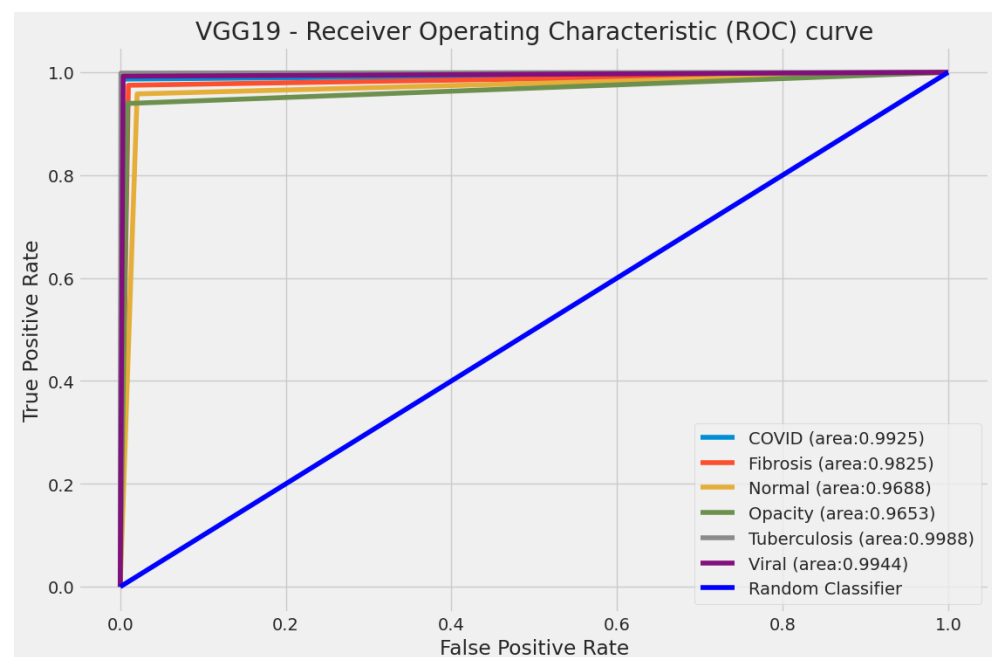


Figure 10. The ROC curves of lung diseases in the test collection for the basic VGG19 model.

5.6. Comparisons with Related Work

Table 9 offers a comprehensive synopsis for multi-class lung disease identification through automated diagnosis using DL models. The research by Sarkar et al. [24] employs the VGG16 model enhanced with multi-scale features, achieving an accuracy of 97.47%, an F1-score of 0.9500, and an AUC of 0.9900 on a dataset comprising 5700 images across six categories. Sultana et al. [25] utilize a 2D-CNN model on a significantly larger dataset of 14,948 images spanning the same six categories. Their model achieves an accuracy of 96.75%, an F1-score of 0.9386, and an AUC of 0.9939. The work of Al et al. [26] explores using ResNet50 on a dataset of 2186 images across five categories. While their model achieves a lower accuracy of 91.60% and an AUC of 0.9900, the F1-score is not reported. Ibrahim et al. [27] present an approach using VGG19 combined with a CNN, applied to a large dataset of 33,676 images across four categories. Their model achieves an impressive accuracy of 98.05%, an F1-score of 0.9824, and an AUC of 0.9966, showcasing the strength of VGG19 when enhanced with additional CNN layers. Sanida et al. [28] modified the MobileNetV2 model to analyze a dataset of 21,165 images across four categories, reaching an accuracy of 95.80% and an F1-score of 0.9629. Their study emphasizes the effectiveness of lightweight models like MobileNetV2 in lung disease diagnosis. Our research employs a modified VGG19 model on an extensive dataset of 48,582 images across six categories. Our model outperforms the others in terms of accuracy, reaching 98.88%, and maintains a high F1-score of 0.9887 and an AUC of 0.9939. This demonstrates the superiority of our modified VGG19 in multi-class lung disease identification, balancing high accuracy with robust performance metrics.

Table 9. Comparisons with related work for multi-class lung disease identification.

Work	Categories	Number of Images	Model	Accuracy (%)	F1-Score	AUC
[24]	6	5700	VGG16+multi-scale	97.47	0.9500	0.9900
[25]	6	14,948	2D-CNN	96.75	0.9386	0.9939
[26]	5	2186	ResNet50	91.60	-	0.9900
[27]	4	33,676	VGG19+CNN	98.05	0.9824	0.9966
[28]	4	21,165	Modified MobileNetV2	95.80	0.9629	-
Ours	6	48,582	Modified VGG19	98.88	0.9887	0.9939

6. Discussion

Integrating Generative AI with DL in medical imaging, particularly in diagnosing lung diseases, marks a significant stride in healthcare technology. Despite the advancements achieved using DL techniques in accurately diagnosing lung conditions such as opacity, COVID-19, fibrosis, tuberculosis, viral pneumonia, and normal lung conditions, substantial research gaps and problems remain that present challenges and motivations for further study.

A primary research gap in this field is the lack of highly accurate and robust automated systems for diagnosing various lung diseases from chest X-ray images. While traditional models have shown promise, their effectiveness in distinguishing between different lung conditions, especially those with subtle radiographic differences, is limited. This gap becomes more pronounced with emerging diseases like COVID-19, where rapid and accurate diagnosis is crucial. There is also a need for models that can generalize well across diverse patient populations and varying image qualities, a challenge not fully addressed by existing models.

Current DL models face issues such as overfitting, limited interpretability, and the need for large annotated collections. Overfitting leads to models performing well on training data but failing to generalize to new data. The nature of DL models also poses a challenge, as medical professionals require models that provide interpretable diagnostic insights. Furthermore, the performance of these models is highly dependent on the quantity and

quality of the training data, which can be a limitation in medical imaging due to privacy concerns, data availability, and the labour-intensive process of medical annotation.

The motivation for research in integrating Generative AI with medical imaging stems from the impact of advanced diagnostic tools on patient care and healthcare systems. The potential to enhance the accuracy and efficiency of diagnosing lung diseases is a development that carries far-reaching implications for patients and healthcare providers. So, by automating the diagnostic process, an AI-integrated CAD system can alleviate this burden, enabling quicker times for medical image analysis. The modified VGG19 model, with its deeper architecture, offers a promising avenue for improvement over traditional models.

Our approach utilizes the outputs from the last three max-pooling layers of the VGG19 model. We redirect these through custom layers, including batch normalization, dropout, and up-sampling. This not only refines the feature maps generated by the preceding layers but also enhances the model's capability to identify subtle and complex patterns characteristic of different lung diseases. Moreover, improving the generalizability and interpretability of the modified model significantly increases their clinical applicability, leading to broader adoption in healthcare settings.

The comparison between the basic VGG19 model and the modified reveals significant enhancements in the latter's diagnostic capabilities. The modified VGG19 model's accuracy is markedly superior, registering at 98.88%, which is an increase of nearly 2.5 percentage points over the basic model. The confusion matrix of the modified VGG19 shows a commendable increase in true positive rates for most categories. The basic VGG19 model showcases AUC values exceeding 0.96 across all categories, affirming its robustness as a reliable diagnostic tool. Nevertheless, the modified VGG19 model achieved even higher AUC values that exceeded 0.98 for all categories. These AUC values indicate an exceptionally high true positive rate and a low false positive rate, essential for accurate medical diagnosis. These results suggest that the modifications made to the VGG19 model have substantial practical implications. They indicate a potential for significantly more accurate diagnoses and better patient management strategies. The modified VGG19 model stands out as an improved tool in the medical imaging field, with its enhanced performance likely to contribute to more effective healthcare.

7. Conclusions and Future Work

Our work introduced a novel DL framework, which harnesses the power of VGG19 architecture by integrating custom blocks into the model. We developed a system capable of multi-class diagnosis of a spectrum of lung conditions, including fibrosis, opacity, tuberculosis, normal lung states, viral pneumonia, and COVID-19 pneumonia. This DL framework system is particularly noteworthy given chest X-rays' critical role in promptly and accurately identifying pulmonary abnormalities. Our evaluation process, conducted on an extensive dataset, highlights the framework's remarkable capabilities, surpassing existing state-of-the-art methods in lung disease diagnosis. The achieved accuracy 98.88% is a testament to the framework's precision and reliability. The framework demonstrated superior accuracy and exhibited exceptional performance across critical metrics like precision, recall, F1-score, and AUC, averaging 0.9870, 0.9904, 0.9887, and 0.9939, respectively. In future work, we intend to harness the power of GANs to create a more diverse and representative range of synthetic medical images and extend this study in several directions, including enhancing the model's generalization to different patient demographics, integrating it with other diagnostic tools and electronic health records for more comprehensive patient assessments, and expanding its applicability to other thoracic and respiratory conditions.

Author Contributions: Project administration, T.S.; visualization, M.V.S. and T.S.; investigation, M.V.S. and T.S.; software, M.V.S. and T.S.; conceptualization, M.V.S. and T.S.; resources, M.V.S. and T.S.; writing—review and editing, M.V.S. and T.S.; formal analysis, M.V.S., T.S. and A.S.; methodology, M.V.S. and T.S.; validation, M.V.S., T.S. and A.S.; writing—original draft preparation, T.S., M.V.S. and A.S.; supervision, M.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
AUC	Area under the curve
Bi-GRU	Bidirectional gated recurrent unit
CNN	Convolutional neural network
DeTraC	Decompose, transfer, and compose
DL	Deep learning
GRU	Gated recurrent unit
GPU	Graphics processing unit
HOG	Histogram-oriented gradient
ROC	Receiver operating characteristic
VGG	Visual geometry group

References

1. Khan, A.; Ferrero, J.L. Idiopathic Pulmonary Fibrosis Misdiagnosed as Sputum-Negative Tuberculosis. In *The Misdiagnosis Casebook in Clinical Medicine: A Case-Based Guide*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 479–487. https://doi.org/10.1007/978-3-031-28296-6_65.
2. Ezzahi, M.; Ennasery, Z.; El Malih, S.; Akammar, A.; El Bouardi, N.; Haloua, M.; Lamrani, M.Y.A.; Boubbou, M.; Serraj, M.; Maaroufi, M.; et al. Mediastinal fibrosis as a late and fatal complication of treated tuberculosis mimicking a neoplastic process in a 34-year-old man. *Radiol. Case Rep.* **2023**, *18*, 4287–4293. <https://doi.org/10.1016/j.radcr.2023.09.001>.
3. Lazar, M.; Barbu, E.C.; Chitu, C.E.; Tiliscan, C.; Stratan, L.; Arama, S.S.; Arama, V.; Ion, D.A. Interstitial Lung Fibrosis Following COVID-19 Pneumonia. *Diagnostics* **2022**, *12*, 2028. <https://doi.org/10.3390/diagnostics12082028>.
4. Ali, M.U.; Kallu, K.D.; Masood, H.; Tahir, U.; Gopi, C.V.; Zafar, A.; Lee, S.W. A CNN-Based Chest Infection Diagnostic Model: A Multistage Multiclass Isolated and Developed Transfer Learning Framework. *Int. J. Intell. Syst.* **2023**, *2023*, 6850772. <https://doi.org/10.1155/2023/6850772>.
5. Puram, V.V.; Sethi, A.; Epstein, O.; Ghannam, M.; Brown, K.; Ashe, J.; Berry, B. Central Apnea in Patients with COVID-19 Infection. *J* **2023**, *6*, 164–171. <https://doi.org/10.3390/j6010012>.
6. Kotei, E.; Thirunavukarasu, R. A Comprehensive Review on Advancement in Deep Learning Techniques for Automatic Detection of Tuberculosis from Chest X-ray Images. *Arch. Comput. Methods Eng.* **2023**, *31*, 455–474. <https://doi.org/10.1007/s11831-023-09987-w>.
7. Sanida, T.; Sideris, A.; Chatzisavvas, A.; Dossis, M.; Dasygenis, M. Radiography Images with Transfer Learning on Embedded System. In Proceedings of the 2022 7th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Ioannina, Greece, 23–25 September 2022; pp. 1–4. <https://doi.org/10.1109/SEEDA-CECNSM57760.2022.9932978>.
8. Sanida, T.; Varlamis, I. Application of affinity analysis techniques on diagnosis and prescription data. In Proceedings of the 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS), Thessaloniki, Greece, 22–24 June 2017; pp. 403–408. <https://doi.org/10.1109/CBMS.2017.114>.
9. Irmici, G.; Cè, M.; Caloro, E.; Khenkina, N.; Della Pepa, G.; Ascenti, V.; Martinenghi, C.; Papa, S.; Oliva, G.; Cellina, M. Chest X-ray in Emergency Radiology: What Artificial Intelligence Applications Are Available? *Diagnostics* **2023**, *13*, 216. <https://doi.org/10.3390/diagnostics13020216>.
10. Holfelder, M.; Mulansky, L.; Schlee, W.; Baumeister, H.; Schobel, J.; Greger, H.; Hoff, A.; Pryss, R. Medical device regulation efforts for mHealth apps during the COVID-19 pandemic—An experience report of Corona Check and Corona Health. *J* **2021**, *4*, 206–222. <https://doi.org/10.3390/j4020017>.
11. Koohi-Moghadam, M.; Bae, K.T. Generative AI in medical imaging: applications, challenges, and ethics. *J. Med. Syst.* **2023**, *47*, 94.
12. Kebaili, A.; Lapuyade-Lahorgue, J.; Ruan, S. Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review. *J. Imaging* **2023**, *9*, 81. <https://doi.org/10.3390/jimaging9040081>.
13. Bandi, A.; Adapa, P.V.S.R.; Kuchi, Y.E.V.P.K. The power of generative ai: A review of requirements, models, input–output formats, evaluation metrics, and challenges. *Future Internet* **2023**, *15*, 260.

14. Yu, P.; Xu, H.; Hu, X.; Deng, C. Leveraging Generative AI and Large Language Models: A Comprehensive Roadmap for Healthcare Integration. *Healthcare* **2023**, *11*, 2776.
15. Najjar, R. Redefining radiology: A review of artificial intelligence integration in medical imaging. *Diagnostics* **2023**, *13*, 2760.
16. Eltawil, F.A.; Atalla, M.; Boulos, E.; Amirabadi, A.; Tyrrell, P.N. Analyzing barriers and enablers for the acceptance of artificial intelligence innovations into radiology practice: a scoping review. *Tomography* **2023**, *9*, 1443–1455.
17. Ait Nasser, A.; Akhloufi, M.A. A review of recent advances in deep learning models for chest disease detection using radiography. *Diagnostics* **2023**, *13*, 159. <https://doi.org/10.3390/diagnostics13010159>.
18. Santosh, K.; GhoshRoy, D.; Nakarmi, S. A systematic review on deep structured learning for COVID-19 screening using chest CT from 2020 to 2022. *Healthcare* **2023**, *11*, 2388. <https://doi.org/10.3390/healthcare11172388>.
19. Butt, M.J.; Malik, A.K.; Qamar, N.; Yar, S.; Malik, A.J.; Rauf, U. A Survey on COVID-19 Data Analysis Using AI, IoT, and Social Media. *Sensors* **2023**, *23*, 5543. <https://doi.org/10.3390/s23125543>.
20. Mostafa, F.A.; Elrefaei, L.A.; Fouda, M.M.; Hossam, A. A Survey on AI Techniques for Thoracic Diseases Diagnosis Using Medical Images. *Diagnostics* **2022**, *12*, 3034. <https://doi.org/10.3390/diagnostics12123034>.
21. Sailunaz, K.; Özyer, T.; Rokne, J.; Alhaji, R. A survey of machine learning-based methods for COVID-19 medical image analysis. *Med. Biol. Eng. Comput.* **2023**, *61*, 1257–1297. <https://doi.org/10.1007/s11517-022-02758-y>.
22. Alafif, T.; Tehame, A.M.; Bajaba, S.; Barnawi, A.; Zia, S. Machine and deep learning towards COVID-19 diagnosis and treatment: Survey, challenges, and future directions. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1117. <https://doi.org/10.3390/ijerph18031117>.
23. Han, X.; Hu, Z.; Wang, S.; Zhang, Y. A survey on deep learning in COVID-19 diagnosis. *J. Imaging* **2022**, *9*, 1. <https://doi.org/10.3390/jimaging9010001>.
24. Sarkar, O.; Islam, M.R.; Syfullah, M.K.; Islam, M.T.; Ahamed, M.F.; Ahsan, M.; Haider, J. Multi-Scale CNN: An Explainable AI-Integrated Unique Deep Learning Framework for Lung-Affected Disease Classification. *Technologies* **2023**, *11*, 134. <https://doi.org/10.3390/technologies11050134>.
25. Sultana, A.; Nahiduzzaman, M.; Bakchy, S.C.; Shahriar, S.M.; Peyal, H.I.; Chowdhury, M.E.; Khandakar, A.; Arselene Ayari, M.; Ahsan, M.; Haider, J. A real time method for distinguishing COVID-19 utilizing 2D-CNN and transfer learning. *Sensors* **2023**, *23*, 4458. <https://doi.org/10.3390/s23094458>.
26. Al-Timemy, A.H.; Khushaba, R.N.; Mosa, Z.M.; Escudero, J. An efficient mixture of deep and machine learning models for COVID-19 and tuberculosis detection using X-ray images in resource limited settings. In *Artificial Intelligence for COVID-19*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 77–100. https://doi.org/10.1007/978-3-030-69744-0_6.
27. Ibrahim, D.M.; Elshennawy, N.M.; Sarhan, A.M. Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases. *Comput. Biol. Med.* **2021**, *132*, 104348. <https://doi.org/10.1016/j.compbiomed.2021.104348>.
28. Sanida, T.; Sideris, A.; Tsiktisiris, D.; Dasygenis, M. Lightweight neural network for COVID-19 detection from chest X-ray images implemented on an embedded system. *Technologies* **2022**, *10*, 37. <https://doi.org/10.3390/technologies10020037>.
29. Ibrokhimov, B.; Kang, J.Y. Deep Learning Model for COVID-19-Infected Pneumonia Diagnosis Using Chest Radiography Images. *BioMedInformatics* **2022**, *2*, 654–670. <https://doi.org/10.3390/biomedinformatics2040043>.
30. Hasan, M.K.; Ahmed, S.; Abdullah, Z.E.; Monirujjaman Khan, M.; Anand, D.; Singh, A.; AlZain, M.; Masud, M. Deep learning approaches for detecting pneumonia in COVID-19 patients by analyzing chest X-ray images. *Math. Probl. Eng.* **2021**, *2021*, 9929274. <https://doi.org/10.1155/2021/9929274>.
31. Sanida, T.; Tabakis, I.M.; Sanida, M.V.; Sideris, A.; Dasygenis, M. A Robust Hybrid Deep Convolutional Neural Network for COVID-19 Disease Identification from Chest X-ray Images. *Information* **2023**, *14*, 310. <https://doi.org/10.3390/info14060310>.
32. Abbas, A.; Abdelsamea, M.M.; Gaber, M.M. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. *Appl. Intell.* **2021**, *51*, 854–864. <https://doi.org/10.1007/s10489-020-01829-7>.
33. Hemdan, E.E.D.; Shouman, M.A.; Karar, M.E. Covidx-net: A framework of deep learning classifiers to diagnose COVID-19 in X-ray images. *arXiv* **2020**, arXiv:2003.11055.
34. Ahsan, M.; Based, M.A.; Haider, J.; Kowalski, M. COVID-19 detection from chest X-ray images using feature fusion and deep learning. *Sensors* **2021**, *21*, 1480. <https://doi.org/10.3390/s21041480>.
35. Khan, I.U.; Aslam, N. A deep-learning-based framework for automated diagnosis of COVID-19 using X-ray images. *Information* **2020**, *11*, 419. <https://doi.org/10.3390/info11090419>.
36. Kaggle. COVID-19 Radiography Dataset. Available online: <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database/activity> (accessed on 20 November 2023).
37. Deeplake. NIH Chest X-ray Dataset. Available online: <https://datasets.activeloop.ai/docs/ml/datasets/nih-chest-x-ray-dataset/> (accessed on 22 November 2023).
38. Rácz, A.; Bajusz, D.; Héberger, K. Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification. *Molecules* **2021**, *26*, 1111. <https://doi.org/10.3390/molecules26041111>.
39. Baghdadi, N.; Maklad, A.S.; Malki, A.; Deif, M.A. Reliable sarcoidosis detection using chest X-rays with efficientnets and stain-normalization techniques. *Sensors* **2022**, *22*, 3846. <https://doi.org/10.3390/s22103846>.
40. Tasci, E.; Uluturk, C.; Ugur, A. A voting-based ensemble deep learning method focusing on image augmentation and preprocessing variations for tuberculosis detection. *Neural Comput. Appl.* **2021**, *33*, 15541–15555. <https://doi.org/10.1007/s00521-021-06177-2>.

41. Jönemo, J.; Abramian, D.; Eklund, A. Evaluation of augmentation methods in classifying autism spectrum disorders from fMRI data with 3D convolutional neural networks. *Diagnostics* **2023**, *13*, 2773. <https://doi.org/10.3390/diagnostics13172773>.
42. Zhao, D.; Zhu, D.; Lu, J.; Luo, Y.; Zhang, G. Synthetic medical images using F&BGAN for improved lung nodules classification by multi-scale VGG16. *Symmetry* **2018**, *10*, 519. <https://doi.org/10.3390/sym10100519>.
43. Segu, M.; Tonioni, A.; Tombari, F. Batch normalization embeddings for deep domain generalization. *Pattern Recognit.* **2023**, *135*, 109115. <https://doi.org/10.1016/j.patcog.2022.109115>.
44. Choe, J.; Lee, S.; Shim, H. Attention-based dropout layer for weakly supervised single object localization and semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 4256–4271. <https://doi.org/10.1109/TPAMI.2020.2999099>.
45. Zhao, Y.; Li, G.; Xie, W.; Jia, W.; Min, H.; Liu, X. GUN: Gradual upsampling network for single image super-resolution. *IEEE Access* **2018**, *6*, 39363–39374. <https://doi.org/10.1109/ACCESS.2018.2855127>.
46. Menghani, G. Efficient deep learning: A survey on making deep learning models smaller, faster, and better. *ACM Comput. Surv.* **2023**, *55*, 1–37. <https://doi.org/10.1145/3578938>.
47. Sanida, T.; Sideris, A.; Sanida, M.V.; Dasygenis, M. Tomato leaf disease identification via two-stage transfer learning approach. *Smart Agric. Technol.* **2023**, *5*, 100275. <https://doi.org/10.1016/j.atech.2023.100275>.
48. Sanida, T.; Tsiktsiris, D.; Sideris, A.; Dasygenis, M. A heterogeneous implementation for plant disease identification using deep learning. *Multimed. Tools Appl.* **2022**, *81*, 15041–15059. <https://doi.org/10.1007/s11042-022-12461-7>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.