

Multi-Scale Adaptive Attention Framework for Improved Lung Disease Classification

Oladosu Oyebisi Oladimeji^{1,*}, Ayodeji Olusegun Ibitoye²

¹Atlantic Technological University, Ireland, ror.org/0458dap48

²University of Greenwich, United Kingdom, ror.org/00bmj0a71

Corresponding author:

Oladosu Oladimeji, Atlantic Technological
University, Ireland
oladosu.oladimeji@research.atu.ie



Article History:

Received: 10.02.2025

Revised: 03.07.2025

Accepted: 12.07.2025

Published Online: 24.09.2025

ABSTRACT

Lung diseases are a leading cause of morbidity and mortality, underscoring the need for accurate diagnostic tools. Chest X-ray imaging is commonly used for diagnosis, but existing deep learning methods often focus on binary classification and struggle with the complexity of lung diseases. To address these challenges, we developed the Multi-Scale Adaptive Attention Fusion Network (MSAAF-Net), a novel framework designed for enhanced lung disease classification. MSAAF-Net integrates multi-scale feature extraction with self-attention, spatial attention, and channel attention mechanisms, dynamically weighted through a class-aware module. This approach enables the model to capture both fine-grained and large-scale pathological features, improving classification across multiple disease classes. Evaluated on a publicly available chest X-ray dataset using five-fold cross-validation, MSAAF-Net achieved a classification accuracy of 93.53%, an F1-score of 93.76%, and an AUC of 98.33%, surpassing state-of-the-art models. These results demonstrate MSAAF-Net's ability to effectively manage the complexity of multi-class lung disease classification. The framework enhances automated diagnostic accuracy, supporting better clinical decision-making and advancing AI's role in lung healthcare.

Keywords: Deep learning, Lung disease, Attention Fusion, Medical Image Analysis

1. Introduction

Lung diseases represent a significant global health challenge, accounting for a substantial proportion of morbidity and mortality worldwide. According to the World Health Organisation, respiratory and lung diseases, including COVID-19, pneumonia, and tuberculosis (TB), are among the leading causes of death worldwide [1]. Around 450 million people are affected by these diseases, with a higher impact on children, accounting for approximately 657 cases out of 1000 [2]. Chest X-ray (CXR) imaging is a widely utilised and effective diagnostic tool for detecting lung pathologies, offering valuable insights to support clinical decision-making. With the rapid advancements in deep learning, automated analysis of X-ray images has demonstrated great potential in improving diagnostic accuracy and efficiency. Hence, the detection and classification of respiratory diseases and determining their severity are crucial for examining their spread, investigating their characteristics, and developing effective public health interventions [3]. Consequently, research studies on identifying and treating lung disease infections have garnered significant interest worldwide [4]. However, it is challenging to distinguish these respiratory diseases as they share common symptoms such as coughing, sneezing, fever, fatigue, and shortness of breath [5]. Therefore, there are risks of overlapping challenges.

Several studies, including [6]–[8], have indicated that various symptoms of lung diseases can be observed in X-ray and CT images of the lung and complete blood count (CBC). Nonetheless, CXR is one of the most cost-effective and conventional methods used in clinical diagnostics [9]. Furthermore, it can be used to promptly and efficiently identify the disease [3]. Despite this, detecting and classifying lung diseases using CXR images is still challenging for radiologists [10]. However, deep learning has indeed made substantial strides across several industries, with healthcare being no exception [11]. It is effective in developing models capable of diagnosing diseases from medical imaging. Hence, it has been used in diagnosing Tuberculosis [12][13], Pneumonia [14][15], and Covid-19 [16].

The current limitation of these existing solutions is that they are designed for binary classifications like COVID-19 versus Normal, Pneumonia versus COVID-19 [17], TB versus Normal, and Pneumonia versus Normal [18] and Lung Opacity versus normal [19]. However, complexity arises when attempting to incorporate additional classes while preserving optimal performance. Furthermore, the heterogeneity of pathological patterns in lung X-rays, ranging from subtle texture anomalies to large-scale structural changes, presents significant challenges for feature extraction and disease differentiation. To address these gaps, this study explores the following research questions: (1) How can multi-scale feature extraction enhance the detection of diverse lung pathologies in X-ray images? (2) How can adaptive attention mechanisms improve sensitivity to disease-specific features for accurate multi-class classification? To investigate these questions, the Multi-Scale Adaptive

Attention Fusion Network (MSAAF-Net), which combines multi-scale feature extraction with adaptive attention mechanisms to emphasise disease-specific features dynamically, was proposed. This approach enables the model to capture both fine-grained and large-scale pathological variations while maintaining robust performance across multiple disease categories. By addressing these challenges, this research seeks to advance AI-driven diagnostic systems, ensuring improved accuracy and reliability in detecting diverse lung diseases. The proposed framework holds the potential to reduce diagnostic errors, enhance clinical decision-making, and eventually improve patient outcomes, marking a significant step forward in the integration of artificial intelligence into lung healthcare. The remaining sections of this article are structured as follows: the second section describes sample-related works, then the methodology used in this research is discussed in section three. Then the results obtained and the discussion were documented in section four before the conclusions were drawn in the fifth section.

2. Literature Review

This section provides an overview of existing models for liver disease classification. Beginning with an overview in 2.1 to numerous research studies have recently investigated deep learning methods to diagnose liver disease using clinical data, including CT scans and X-ray images, in 2.2.

2.1 Lung Disease Classification Using Chest X-rays

Lung diseases, such as pneumonia, tuberculosis (TB), COVID-19, and lung cancer, remain among the leading causes of disease and death worldwide, underscoring the need for early and accurate diagnosis. CXR imaging has long been a cornerstone of clinical practice for diagnosing lung pathologies due to its widespread availability, rapid turnaround, and cost-effectiveness [20]. As a non-invasive imaging technique, CXRs offer valuable insights into a range of conditions, from infections and fluid accumulation to structural abnormalities and malignancies, making them an essential tool in routine healthcare. Despite their utility, the manual interpretation of chest X-rays poses significant challenges for radiologists. The variability in disease presentation, coupled with overlapping visual features between different lung conditions, often leads to misinterpretation and diagnostic errors. For instance, diseases like pneumonia and COVID-19 can exhibit similar patterns of lung opacity, while early-stage tuberculosis or lung cancer may present with subtle and easily overlooked abnormalities [21]. These complexities necessitate robust and reliable diagnostic methods to aid clinicians in distinguishing between various conditions with high precision.

Deep learning models, particularly CNNs, have shown exceptional performance in extracting and analysing features from X-ray images, enabling more accurate and faster disease classification [22]. Unlike traditional machine learning methods, which rely on handcrafted features, CNNs can automatically learn hierarchical representations of image data, making them well-suited for the intricacies of lung disease detection. However, effective classification requires models capable of handling diverse pathological patterns. Lung diseases manifest in varied forms, from localised consolidations and nodules to diffuse opacities and pleural effusions [23]. These variations necessitate techniques that can capture both fine-grained details and large-scale structural changes. Furthermore, the increasing demand for multi-class classification systems that are capable of simultaneously identifying multiple conditions introduces additional complexity. Many existing approaches, while successful in binary classification tasks, such as "COVID-19 vs. Normal" or "Pneumonia vs. Normal" [24], struggle to maintain performance when tasked with distinguishing between multiple diseases in heterogeneous datasets. As disease categories increase, the classification task becomes more complex, requiring models that can generalise well across diverse pathological patterns [25]. Multi-class classification is crucial for real-world clinical applications, where accurate diagnosis involves distinguishing between a variety of lung diseases in a single patient [26].

2.2 Deep Learning in Medical Image Analysis

To diagnose lung disease, several approaches have been proposed. Convolutional Neural Networks (CNNs) are a deep learning technique commonly applied in text and image recognition tasks [27]. The core components of a CNN model include three main layers: convolutional, pooling, and fully connected layers [28]. Convolution and pooling layers play a crucial role in feature extraction, while the fully connected layer commonly functions as the classifier. Ghose et al [29] proposed a deep learning-based CNN for automatically diagnosing lung disease from chest X-rays. The proposed CNN model comprises five convolutional blocks, each containing multiple layers with a Rectified Linear Unit (ReLU) activation function. Additionally, two fully connected (FC) layers are utilised: the first FC layer is paired with a dropout layer, while the final FC layer is connected to a SoftMax classifier for output generation.

CNN-based models often struggle to capture long-range dependencies, which are essential for many medical image analysis tasks [30]. Han et al [31] proposed a method to distinguish COVID-19 pneumonia from other types of viral pneumonia. The method entails an attention-based deep 3D multiple instance learning (AD3D-MIL) framework, where a patient-level label is assigned to a 3D chest CT, treating it as a collection of instances.

However, medical image datasets are typically limited in size, which can lead to a notable amount of training errors. Hence, to mitigate this challenge, transfer learning (TL) has become a popular solution in recent years. Li et al [32] proposed CovXNet, an architecture based on ResNet50 for diagnosing bacterial pneumonia, viral pneumonia, and COVID-19 from CT scans. The features extracted from all slices using ResNet50 are aggregated using a max-pooling operation. The resulting feature map is passed through a fully connected layer with a SoftMax activation function to produce probability scores for

each class type. Their model achieved an accuracy of 89.1%. Similarly, Kavya et al [33] utilised VGG16 and ResNet50 deep learning algorithms to extract features from chest X-ray images, classifying them into three categories: viral pneumonia, normal, and COVID-19. Apostolopoulos and Mpesiana [34] assessed the performance of state-of-the-art CNN architectures to classify lung diseases in X-ray images. Architectures including VGG19, MobileNetV2, Xception, Inception and InceptionResNet50 were accessed. The results showed that MobileNetV2 performed best.

Rajaraman et al [35] presented the application of iteratively pruned deep learning model ensembles for detecting pulmonary manifestations of COVID-19 in chest X-ray images. Using a transfer learning approach, the first model was trained to differentiate between normal and abnormal chest X-rays. The second model leveraged the learning weights from the first model to classify cases of COVID-19 and pneumonia. Abdelhamid et al [36] propose a multi-level diagnostic framework leveraging transfer learning for the accurate diagnosis of lung disease using X-ray scans. The framework employs an Xception pre-trained model for feature extraction from pre-processed images. A global average pooling (GAP) layer is incorporated to mitigate overfitting, followed by an activation layer to minimise losses. The final classification is performed using a softmax layer.

Additionally, attention mechanisms have been shown to improve the performance of deep learning models by enabling them to focus on the most relevant parts of an image. This is particularly useful in medical imaging, where certain regions of an image are more indicative of disease than others [37]. In lung disease classification, attention mechanisms such as self-attention, channel attention, and spatial attention have been applied to improve model interpretability and performance [38]. For instance, models that use spatial attention can focus on specific areas of the lung that are likely to show signs of disease, improving the overall diagnostic accuracy. Despite the proposed approaches, the diverse pathological patterns in lung X-rays, spanning delicate texture irregularities to extensive structural alterations, pose considerable challenges for effective feature extraction and accurate disease differentiation. Hence, this study aimed at alleviating the challenge using the Multi-Scale adaptive attention fusion network.

3. Materials and Methods

This section details the approach used to develop the Multi-Scale Adaptive Attention Fusion Network (MSAAF-Net), designed to enhance lung disease classification from chest X-ray images. The methodology described in Figure 1 encompasses main components like data preparation, network architecture, training process, and performance evaluation. The essence is to tackle the complexities of multi-class classification and improve the robustness of automated diagnostic systems for lung diseases. To begin with, a publicly available chest X-ray dataset is preprocessed to standardise the input images and ensure consistent quality in 3.1.

3.1 Data Description and Preprocessing

The dataset used in this study is the Lung X-Ray Image dataset by Talukder [39]. This dataset comprises high-quality X-ray images, gathered from various sources such as hospitals, clinics, and healthcare institutions. The dataset contains 3,475 X-ray images with 1250 normal, 1125 Lung Opacity, and 1100 Viral Pneumonia instances. The images were pre-processed using a standardised pipeline to ensure consistency across the dataset. Each image was resized to a fixed dimension of 224×224 pixels to maintain uniform spatial dimensions across the dataset. The resize operation was performed using bilinear interpolation to preserve image quality while transforming to the target dimensions. The pixel intensity values were normalised using min-max scaling that transforms the pixel intensities to a bounded range of [0, 1], which is optimal for deep neural network training. This also preserves relative relationships between pixel intensities, which is crucial for medical image analysis.

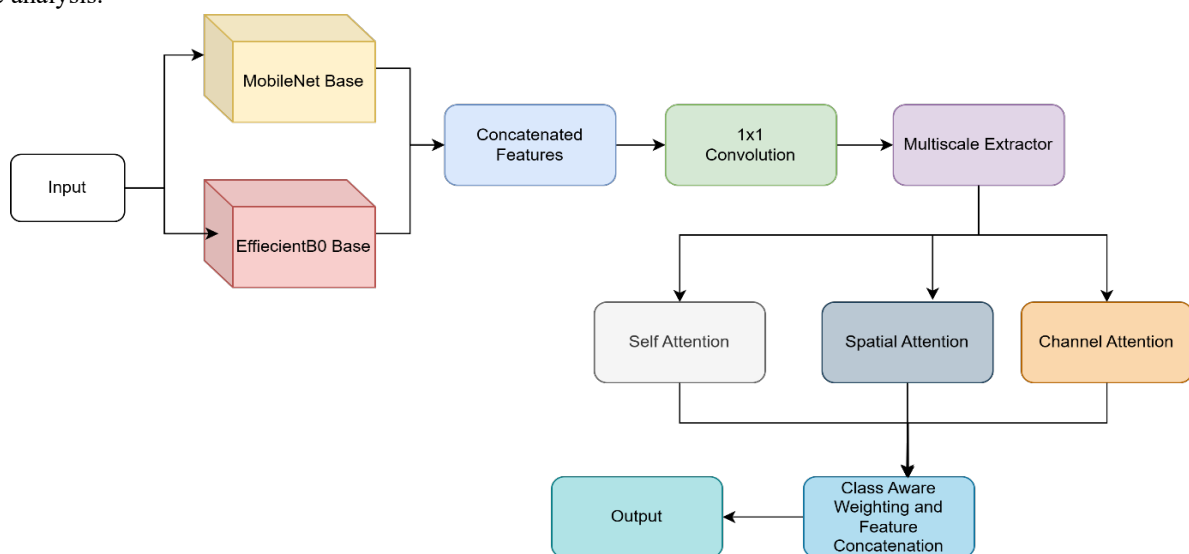


Figure 1. The Proposed Methodology Framework

3.2 Multi-Scale and Adaptive Attention Fusion Network (MSAAF-Net)

The MSAAF-Net utilises two backbones pre-trained on ImageNet MobileNet [40] and EfficientNetB0 [41] as shown in Figure 1, which provides robust low-level features while maintaining computational efficiency. The resulting features are passed through the three attention modules (self-attention, spatial attention, and channel attention). These attention outputs are fused adaptively using the CAWM-generated weights. Finally, global average pooling and a dense layer with ReLU activation are applied, followed by a SoftMax classifier to predict the class probabilities. The model is trained with the Adam optimiser, categorical cross-entropy loss, and accuracy as the evaluation metric.

3.3 Feature Extraction

As shown in Figure 1, two pre-trained convolutional neural networks, MobileNet and EfficientNetB0, were employed as backbone architectures. Both models were initialised with ImageNet weights and used without their classification layers. Feature maps from the final convolutional layers of these networks were extracted and concatenated along the channel dimension to form a unified feature representation. A 1×1 convolutional layer (labelled as '1x1 Convolution' in Figure 1) was applied to the concatenated feature map to reduce dimensionality and enhance semantic coherence. To capture features at different spatial resolutions, the semantic feature map was processed through parallel convolutional layers with kernel sizes 3×3 and 5×5 , producing feature maps that capture finer and coarser patterns. These feature maps are concatenated to improve the model's capability to capture multi-scale information.

3.4 Self-Attention Module

The self-attention module, shown in Figure 1 following the Multiscale Extractor, enhances spatial feature learning by calculating dependencies between different spatial locations in the feature map. Given an input feature map $X \in \mathbb{R}^{H \times W \times C}$, where H, W and C represent the height, width, and channels, respectively. The module constructs three representations from the input using 1×1 convolutions to reduce dimensionality: θ , ϕ , and g [42]. Then attention matrix β is calculated as the normalised dot product $\beta = \text{softmax}(\theta \cdot \phi^T)$. The normalised attention matrix quantifies spatial dependencies by capturing relationships across spatial positions. The output of the module is the sum of the attended features scaled by a learnable weight parameter and the original input, facilitating residual learning.

3.5 Spatial Attention Module

As illustrated in Figure 1, the spatial attention module follows the Multiscale Extractor path. The spatial attention module emphasises relevant spatial information by combining average and maximum pooled feature maps along the spatial dimensions [43]. The module, shown as 'Spatial Attention' in the diagram, first applies global average and maximum pooling operations, followed by 1×1 convolution layers, to capture context [44]. These feature maps are concatenated and passed through a sigmoid activation to generate the spatial attention map. The module multiplies this attention map with the input features, focusing on critical spatial information.

3.6 Channel Attention Module

The channel attention module, depicted in Figure 1 as a parallel path following the Multiscale Extractor, prioritises important channels through global average pooling along spatial dimensions to produce a channel descriptor [45]. This descriptor is then passed through two dense layers, with the first applying ReLU activation and the second applying sigmoid activation, to generate channel-wise weights. These weights are multiplied by the input features to emphasise significant channels selectively.

3.7 Class-Aware Weighting Module (CAWM)

A novel Class-Aware Weighting Module (CAWM) was introduced to assign dynamic weights to the outputs of the three attention mechanisms based on predicted class probabilities, as shown in Figure 2.

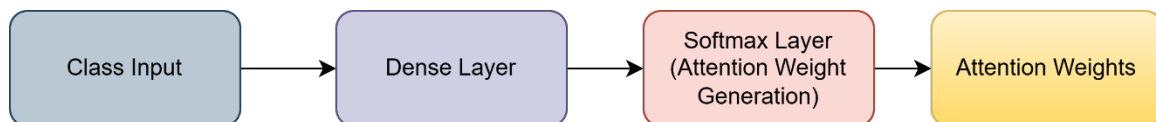


Figure 2. Class-Aware Weighting Module Diagram

The class-aware weighting module generates weighting coefficients for each attention type (Self, spatial and channel attentions) based on the class-specific input as shown in Figure 1. This module consists of a dense layer with softmax activation, producing a weight vector for adaptive attention fusion. These weights dynamically balance the contributions of self-attention, spatial attention, and channel attention based on the input class. Hence, for an input vector $y \in \mathbb{R}^C$ the CAWM generates weights $\alpha_{self}, \alpha_{spatial}, \alpha_{channel}$.

3.8 Adaptive Attention Fusion

The adaptive attention fusion module combines the outputs of self-attention, spatial attention, and channel attention using weights generated by the CAWM. Each attention output is scaled by its corresponding weight, then summed to produce the fused feature representation, enabling adaptive attention mechanisms tailored to each class.

$$F_{fused} = \alpha_{self} \cdot F_{self} + \alpha_{spatial} \cdot F_{spatial} + \alpha_{channel} \cdot F_{channel}$$

In summary, the proposed methodology outlines a comprehensive approach to lung disease classification using the Multi-Scale Adaptive Attention Fusion Network (MSAAF-Net). By integrating multi-scale feature extraction with adaptive attention mechanisms, the framework effectively addresses the challenges of capturing diverse pathological patterns and enhancing sensitivity to disease-specific features. The use of a five-fold cross-validation strategy ensures robust model evaluation, and the performance metrics, as discussed in section four, provide a clear assessment of the model's accuracy and reliability.

4. Results and Discussion

The model in this research was built upon TensorFlow 2.10.0, Keras 2.10.0, Python 3.9.16, and OpenCV, an open-source computer vision library. The computing environment was based on Windows 11 (64-bit) with an Intel(R) Core (TM) i7-12850HX @ 2.10GHz processor. The dataset was divided into 80% training and 20% testing, resulting in 2780 and 695 images for the training and test sets, respectively. The model was trained using five-fold cross-validation on the training set, while the remaining 20% served for independent testing. The reason for the choice of five-fold cross-validation is to report both mean performance and standard deviation across folds, providing insight into model stability and generalisability, which is crucial for clinical applications where consistent performance is paramount. Additionally, the approach aligns with best practices in medical AI research, where robust validation is essential for clinical translation, and has been widely adopted in similar lung disease classification studies [46,47]

In the experiment, the Adam optimizer with the default learning rate of 1e-3 was used with a batch size of 16. This value was selected based on initial training experiments showing stable convergence and no significant benefit from manual tuning. Categorical cross-entropy was chosen as the loss function. The model was trained for 20 epochs for each fold.

This section presents the performance results of the model. Evaluation metrics included accuracy, precision, recall, and Area Under the Receiver Operating Characteristic Curve (AUC). Figure 3. displays the ROC curve, AUC scores across the training folds, and the mean and standard deviation of the model's performance.

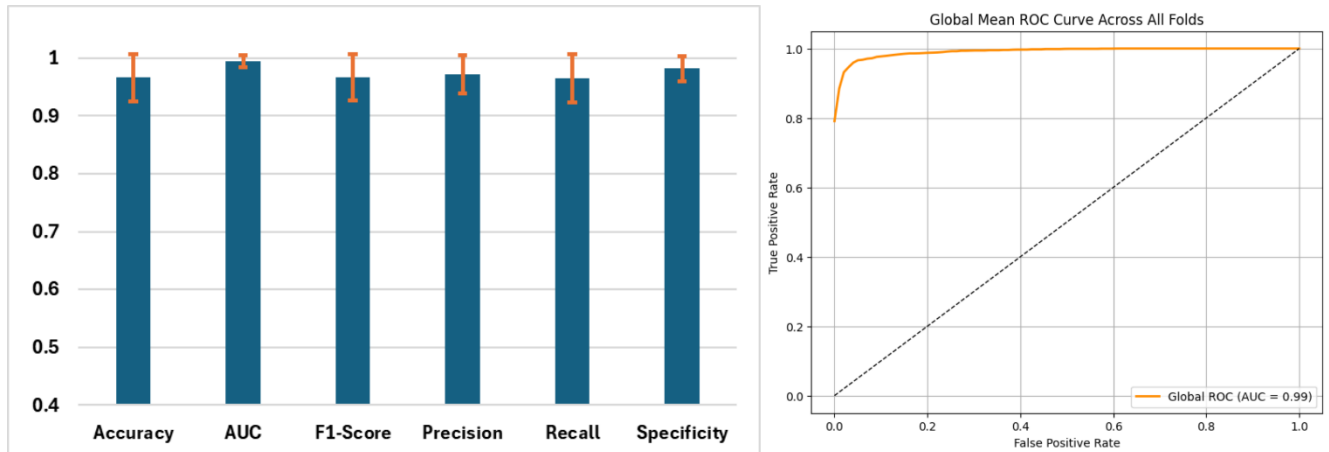


Figure 3. The five-fold training ROC Curves and Average performance metrics (Accuracy, F1-score, Precision, Recall, Specificity) across five-fold cross-validation. Error bars represent standard deviation, showing consistent performance across folds.

The ROC curves, shown in Figure 3, provide an overview of the model's ability to distinguish between the classes. The AUC of the mean ROC curves for the five folds was 99%, indicating the excellent discriminative ability of the model. The high AUC values and well-balanced precision and recall metrics indicate that the model can be effectively deployed in real-world applications. The AUC values across all three classes in the dataset are nearly 1, signifying that the model demonstrates exceptional efficiency in classifying lung images within each category.

The bar chart in Figure 3 illustrates the average performance metrics across all five validation folds, along with error bars representing standard deviation. The model's average performance is promising, with error bars (standard deviation) below 0.15, highlighting its stability and reliability.

4.1 Comparison with State-of-the-Art

A comparative study was conducted to compare the proposed method with the existing state-of-the-art approaches. To the best of our knowledge, these state-of-the-art methods have not utilised the Class-Aware Weighting Module, which is a critical factor for real-world applicability. The results of the comparative study in Table 1 show that the proposed approach outperforms existing state-of-the-art approaches. Four methods were selected for the comparative study, and the results were evaluated using the dataset used in this research based on overall accuracy.

Table 1. Comparison with the existing three-class classification methods on lung disease diagnosis

| Ref. | Methods | Accuracy | Recall | Precision | F1-Score | AUC |
|-------------|------------------|---------------|---------------|---------------|---------------|---------------|
| [29] | Custom CNN | 0.3540 | 0.3540 | 0.1180 | 0.1783 | 0.5 |
| [33] | ResNet50 | 0.8964 | 0.8964 | 0.9055 | 0.9005 | 0.9709 |
| [34] | MobileNetV2 | 0.8849 | 0.8849 | 0.8868 | 0.8877 | 0.9723 |
| [36] | Xception | 0.8590 | 0.8590 | 0.8690 | 0.8690 | 0.9594 |
| Ours | MSAAF-Net | 0.9353 | 0.9353 | 0.9380 | 0.9376 | 0.9833 |

State-of-the-art methods such as simple custom CNN [29], ResNet50 [33], MobileNetV2[34] and XceptionNet [36] were implemented using similar experimental settings as described earlier, employing an 80:20% split ratio for training and testing.

The results align with previous findings demonstrating the limitations of simple custom architectures in handling complex medical imaging data. The significant performance gap between the Custom CNN and the other models highlights the importance of leveraging advanced architecture and transfer learning for improved accuracy and reliability. The superior performance of MSAAF-Net demonstrates several key advantages of the proposed approach. Firstly, it provides complementary feature representations that capture both efficiency-optimised features and compound scaling benefits. Hence, better than the best single-backbone architecture, ResNet [33]. In addition, unlike traditional methods that apply uniform attention weights, the class-aware weighting enables the model to adaptively emphasise different attention mechanisms based on disease-specific characteristics, thereby improving sensitivity and specificity.

One of the key innovations of MSAAF-Net is its ability to capture both fine-grained details and large-scale structural patterns through multi-scale convolutional layers. Traditional models often struggle with varying pathological appearances ranging from small nodules to diffuse opacities due to their reliance on single-scale feature maps. By combining features extracted at multiple resolutions, MSAAF-Net enhances its robustness to diverse visual manifestations of lung diseases.

Ablation Study

To validate the contribution of each attention module in the proposed MSAAF-Net architecture, an ablation study was conducted by evaluating the model's performance when each attention mechanism was used independently, as well as when no attention was applied. Specifically, the following configurations were compared: Baseline Model (MobileNet + EfficientNetB0) without any attention modules, baseline model with only self-attention applied, baseline model with only spatial attention applied, baseline model with only channel attention applied and the full architecture including all three attention mechanisms fused via the Class-Aware Weighting Module (CAWM). Each configuration was trained using five-fold cross-validation under identical hyperparameters (batch size = 16, learning rate = 5e-4, epochs = 20). The results are summarised in Table 2.

Table 2: Results of the Ablation Study

| Methods | Accuracy | Recall | Precision | F1-Score | AUC |
|-----------------------|---------------|---------------|---------------|---------------|---------------|
| Baseline | 0.9094 | 0.9094 | 0.9151 | 0.9126 | 0.9756 |
| Baseline+Self Attn | 0.9108 | 0.9108 | 0.9155 | 0.9137 | 0.9757 |
| Baseline+Spatial Attn | 0.8662 | 0.8662 | 0.9030 | 0.8711 | 0.9611 |
| Baseline+Channel Attn | 0.9094 | 0.9094 | 0.9170 | 0.9128 | 0.9818 |
| MSAAF-Net | 0.9353 | 0.9353 | 0.9380 | 0.9376 | 0.9833 |

As shown in Table 2, the baseline model achieved a classification accuracy of 91%, indicating the strong foundation provided by the dual backbone feature extractors. However, adding individual attention mechanisms improved performance across all metrics. The spatial and channel attention modules contributed more significantly than the self-attention module. Finally, combining all three attention mechanisms using the CAWM further enhanced performance, achieving the highest accuracy of 93%, demonstrating the effectiveness of adaptive fusion based on class-specific features.

Interpretability via Grad-CAM Visualisations

To enhance the interpretability of the proposed MSAAF-Net, Grad-CAM visualisations for a subset of test images were generated. Grad-CAM highlights the image regions most influential in the model's decision-making process, offering insights into how attention mechanisms focus on disease-specific features such as opacities, consolidations, and interstitial changes.

Grad-CAM was implemented by computing gradients of the predicted class output concerning the final convolutional layer outputs. These gradients were globally averaged and used to generate a heatmap that was overlaid on the original input image using the viridis colourmap for clarity.

Figure 4 presents example Grad-CAM visualisations for normal, Lung Opacity, and Viral Pneumonia cases. In each case, the model focuses on anatomically relevant regions such as lung fields and bronchial structures, aligning with clinical observations. This indicates that the attention-based fusion effectively directs the model toward diagnostically meaningful areas of the chest X-ray. These visualisations support the model’s reliability and transparency, which is essential for clinical adoption.

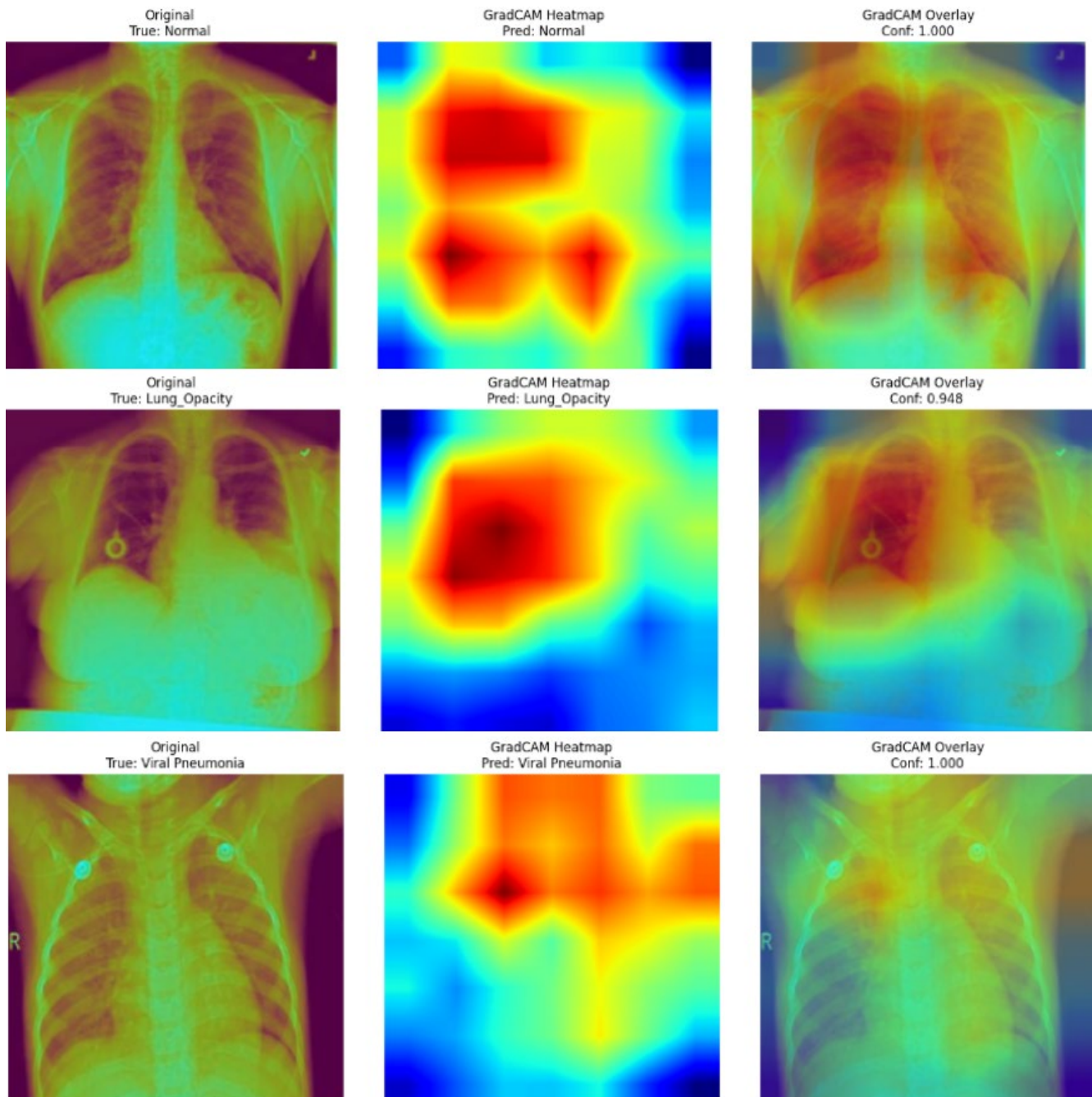


Figure 4: Examples of Grad-CAM visualizations from the proposed MSAAF-Net model showing the regions of interest in chest X-rays for Normal (Row 1), Lung Opacity (Row 2), and Viral Pneumonia (Row 3) classes

Lung disease is one of the most prevalent and impactful health conditions worldwide. This research assessed the diagnosis of lung diseases, focusing on conditions such as normal, Lung Opacity, and Viral Pneumonia. This task was performed using an MSAAF-Net to overcome the challenge of the heterogeneous nature of pathological patterns of chest X-rays in lung disease diagnosis. Unlike conventional attention models that rely on a single or fixed attention type, MSAAF-Net introduces a Class-Aware Weighting Module (CAWM). This module dynamically adjusts the contribution of self-attention, spatial

attention, and channel attention based on class relevance. As a result, the model outperformed existing state-of-the-art techniques, including a custom CNN with five convolution blocks, and two fully connected and dropout layers [29]. Heavyweight models with a ResNet50 base [33] and XceptionNet [36], as well as a lightweight model that utilises MobileNetV2 [34]. The superior performance of this model is primarily attributed to the integration of the Class-Aware Weighting Module. However, one of the major limitations of this study is that the model was validated on a single public dataset without external validation on independent or private hospital datasets. External validation is essential for assessing real-world applicability and robustness. In addition, the dual-backbone architecture and multiple attention mechanisms result in higher computational overhead compared to lightweight methods like MobileNetV2 [34]. While the proposed model outperforms benchmarks, future research could explore testing the model on bigger and more diverse datasets would help validate its generalizability across different populations and imaging modalities.

5. Conclusions

Lung disease is widely recognised as a leading cause of mortality worldwide. This paper proposes a novel approach, MSAAF-Net, for multiclass lung disease classification from X-ray images. The model was applied to predict the three classes- normal, Lung Opacity, and Viral Pneumonia, using the Lung X-ray image dataset. The validation results achieved were as follows: Accuracy of 95.53%, Recall of 93.53%, and Specificity of 99.38%. This model differs significantly from existing detection approaches by focusing exclusively on lung-specific features. It employs multi-scale feature extraction with self-attention, spatial attention, and channel attention, dynamically weighted by a novel class-aware module to enhance sensitivity to disease-specific features. The proposed method achieves high accuracy in prediction and classification tasks, outperforming other deep-learning classifier models and demonstrating its potential for enhanced diagnostic performance. This method demonstrates significant ability as a reliable tool for assisting in the stratification of lung disease patients, thereby enhancing diagnostic and treatment strategies. In future research, explainability frameworks will be used for the model explanations.

References

- [1] S. H. Karaddi and L. D. Sharma, "Automated multi-class classification of lung diseases from CXR-images using pre-trained convolutional neural networks," *Expert Syst. Appl.*, vol. 211, 2023, doi: 10.1016/j.eswa.2022.118650.
- [2] W. H. O. OMS, "Coronavirus disease (COVID-19) Situation Report – 193," *Coronavirus Dis.*, no. June, 2022.
- [3] G. V. E. Rao, R. B., P. N. Srinivasu, M. F. Ijaz, and M. Woźniak, "Hybrid framework for respiratory lung diseases detection based on classical CNN and quantum classifiers from chest X-rays," *Biomed. Signal Process. Control*, vol. 88, p. 105567, Feb. 2024, doi: 10.1016/j.bspc.2023.105567.
- [4] M. H. Al-Sheikh, O. Al Dandan, A. S. Al-Shamayleh, H. A. Jalab, and R. W. Ibrahim, "Multi-class deep learning architecture for classifying lung diseases from chest X-Ray and CT images," *Sci. Rep.*, vol. 13, no. 1, p. 19373, Nov. 2023, doi: 10.1038/s41598-023-46147-3.
- [5] C. Huang et al., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, 2020, doi: 10.1016/S0140-6736(20)30183-5.
- [6] M. O. Wielpütz, C. P. Heußel, F. J. F. Herth, and H.-U. Kauczor, "Radiological Diagnosis in Lung Disease," *Dtsch. Arztebl. Int.*, 2014, doi: 10.3238/arztebl.2014.0181.
- [7] L. A. Rousan, E. Elobeid, M. Karrar, and Y. Khader, "Chest x-ray findings and temporal lung changes in patients with COVID-19 pneumonia," *BMC Pulm. Med.*, vol. 20, no. 1, 2020, doi: 10.1186/s12890-020-01286-5.
- [8] E. Saad, B. Maamoun, and A. Nimer, "Increased Red Blood Cell Distribution Predicts Severity of Chronic Obstructive Pulmonary Disease Exacerbation," *J. Pers. Med.*, vol. 13, no. 5, 2023, doi: 10.3390/jpm13050843.
- [9] S. Resnick et al., "Clinical relevance of the routine daily chest X-Ray in the surgical intensive care unit," *Am. J. Surg.*, vol. 214, no. 1, 2017, doi: 10.1016/j.amjsurg.2016.09.059.
- [10] S. Goyal and R. Singh, "Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 4, 2023, doi: 10.1007/s12652-021-03464-7.
- [11] G. M. M. Alshmrani, Q. Ni, R. Jiang, H. Pervaiz, and N. M. Elshennawy, "A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images," *Alexandria Eng. J.*, vol. 64, 2023, doi: 10.1016/j.aej.2022.10.053.
- [12] T. B. Chandra, K. Verma, B. K. Singh, D. Jain, and S. S. Netam, "Automatic detection of tuberculosis related abnormalities in Chest X-ray images using hierarchical feature extraction scheme," *Expert Syst. Appl.*, vol. 158, 2020, doi: 10.1016/j.eswa.2020.113514.
- [13] R. Hooda, A. Mittal, and S. Sofat, "Automated TB classification using ensemble of deep architectures," *Multimed. Tools Appl.*, vol. 78, no. 22, 2019, doi: 10.1007/s11042-019-07984-5.

- [14] J. E. Luján-García, C. Yáñez-Márquez, Y. Villuendas-Rey, and O. Camacho-Nieto, "A transfer learning method for pneumonia classification and visualization," *Appl. Sci.*, vol. 10, no. 8, 2020, doi: 10.3390/AP10082908.
- [15] O. Stephen, M. Sain, U. J. Maduh, and D. U. Jeong, "An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare," *J. Healthc. Eng.*, vol. 2019, 2019, doi: 10.1155/2019/4180949.
- [16] N. Khasawneh, M. Fraiwan, L. Fraiwan, B. Khassawneh, and A. Ibnian, "Detection of covid-19 from chest x-ray images using deep convolutional neural networks," *Sensors*, vol. 21, no. 17, 2021, doi: 10.3390/s21175940.
- [17] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2," *Informatics Med. Unlocked*, vol. 19, 2020, doi: 10.1016/j.imu.2020.100360.
- [18] S. Vasamsetti, G. S. S. Shreyas, V. Chemboli, and S. Thota, "Comparative Performance Analysis of Deep Learning Models for Lung Disease Prediction using Chest X-Ray Images," in *6th International Conference on Inventive Computation Technologies, ICICT 2023 - Proceedings, 2023*. doi: 10.1109/ICICT57646.2023.10134132.
- [19] M. Singla, K. S. Gill, D. Upadhyay, and S. Devliyal, "Optimizing Lung Opacity Classification in Chest X-ray Images through Transfer Learning on VGG19 CNN Model," in *2024 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES), IEEE, May 2024*, pp. 1–4. doi: 10.1109/ICSSES62373.2024.10561338.
- [20] P. Rajpurkar et al., "CheXaid: deep learning assistance for physician diagnosis of tuberculosis using chest x-rays in patients with HIV," *npj Digit. Med.*, vol. 3, no. 1, p. 115, Sep. 2020, doi: 10.1038/s41746-020-00322-2.
- [21] M. A. A. Al-qaness et al., "Chest X-ray Images for Lung Disease Detection Using Deep Learning Techniques: A Comprehensive Survey," *Arch. Comput. Methods Eng.*, vol. 31, no. 6, 2024, doi: 10.1007/s11831-024-10081-y.
- [22] Y. K. Mali, L. Sharma, K. Mahajan, F. Kazi, P. Kar, and A. Bhogle, "Application of CNN Algorithm on X-Ray Images in COVID-19 Disease Prediction," in *2023 IEEE International Carnahan Conference on Security Technology (ICST), IEEE, Oct. 2023*, pp. 1–6. doi: 10.1109/ICST59048.2023.10726852.
- [23] S. Kordnoori, M. Sabeti, H. Mostafaei, and S. Seyed Agha Banihashemi, "Advances in medical image analysis: A comprehensive survey of lung infection detection," *IET Image Process.*, vol. 18, no. 13, pp. 3750–3800, Nov. 2024, doi: 10.1049/ipr2.13246.
- [24] A. U. Ibrahim, M. Ozsoz, S. Serte, F. Al-Turjman, and P. S. Yakoi, "Pneumonia Classification Using Deep Learning from Chest X-ray Images During COVID-19," *Cognit. Comput.*, vol. 16, no. 4, 2024, doi: 10.1007/s12559-020-09787-5.
- [25] Z. Tariq, S. K. Shah, and Y. Lee, "Lung Disease Classification using Deep Convolutional Neural Network," in *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019, 2019*. doi: 10.1109/BIBM47256.2019.8983071.
- [26] A. Ali, Y. Wang, and X. Shi, "Detection of multi-class lung diseases based on customized neural network," *Comput. Intell.*, vol. 40, no. 2, Apr. 2024, doi: 10.1111/coin.12649.
- [27] J. G. Melekoodappattu, A. S. Dhas, B. K. Kandathil, and K. S. Adarsh, "Breast cancer detection in mammogram: combining modified CNN and texture feature based approach," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 9, 2023, doi: 10.1007/s12652-022-03713-3.
- [28] F. F. Ting, Y. J. Tan, and K. S. Sim, "Convolutional neural network improvement for breast cancer classification," *Expert Syst. Appl.*, vol. 120, pp. 103–115, 2019, doi: 10.1016/j.eswa.2018.11.008.
- [29] P. Ghose, M. A. Uddin, U. K. Acharjee, and S. Sharmin, "Deep viewing for the identification of Covid-19 infection status from chest X-Ray image using CNN based architecture," *Intell. Syst. with Appl.*, vol. 16, 2022, doi: 10.1016/j.iswa.2022.200130.
- [30] A. Hatamizadeh et al., "UNETR: Transformers for 3D Medical Image Segmentation," in *Proceedings - 2022 IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2022, 2022*. doi: 10.1109/WACV51458.2022.00181.
- [31] Z. Han et al., "Accurate Screening of COVID-19 Using Attention-Based Deep 3D Multiple Instance Learning," *IEEE Trans. Med. Imaging*, vol. 39, no. 8, 2020, doi: 10.1109/TMI.2020.2996256.
- [32] L. Li et al., "Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy," *Radiology*, vol. 296, no. 2, 2020, doi: 10.1148/radiol.2020200905.
- [33] N. Sri Kavya, T. shilpa, N. Veeranjanyulu, and D. Divya Priya, "Detecting Covid19 and pneumonia from chest X-ray images using deep convolutional neural networks," in *Materials Today: Proceedings, 2022*. doi: 10.1016/j.matpr.2022.05.199.

- [34] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," *Phys. Eng. Sci. Med.*, vol. 43, no. 2, 2020, doi: 10.1007/s13246-020-00865-4.
- [35] S. Rajaraman, J. Siegelman, P. O. Alderson, L. S. Folio, L. R. Folio, and S. K. Antani, "Iteratively Pruned Deep Learning Ensembles for COVID-19 Detection in Chest X-Rays," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3003810.
- [36] A. A. Abdelhamid, E. Abdelhalim, M. A. Mohamed, and F. Khalifa, "Multi-Classification of Chest X-rays for COVID-19 Diagnosis Using Deep Learning Algorithms," *Appl. Sci.*, vol. 12, no. 4, 2022, doi: 10.3390/app12042080.
- [37] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, 2021, doi: 10.1016/j.neucom.2021.03.091.
- [38] X. Li et al., "Deep Learning Attention Mechanism in Medical Image Analysis: Basics and Beyonds," *Int. J. Netw. Dyn. Intell.*, 2023, doi: 10.53941/ijndi0201006.
- [39] M. A. Talukder, "Lung X-Ray Image," *Mendeley Data*. 2023.
- [40] D. Sinha and M. El-Sharkawy, "Thin MobileNet: An Enhanced MobileNet Architecture," in *2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2019*, 2019. doi: 10.1109/UEMCON47517.2019.8993089.
- [41] L. Gaur, U. Bhatia, N. Z. Jhanjhi, G. Muhammad, and M. Masud, "Medical image-based detection of COVID-19 using Deep Convolution Neural Networks," in *Multimedia Systems*, 2023. doi: 10.1007/s00530-021-00794-6.
- [42] X. Pan et al., "On the Integration of Self-Attention and Convolution," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2022. doi: 10.1109/CVPR52688.2022.00089.
- [43] X. Zhu, D. Cheng, Z. Zhang, S. Lin, and J. Dai, "An empirical study of spatial attention mechanisms in deep networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019. doi: 10.1109/ICCV.2019.00679.
- [44] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020. doi: 10.1109/CVPR42600.2020.01155.
- [45] O.O. Oladimeji and A. O. J. Ibitoye, "Brain tumor classification using ResNet50-convolutional block attention module", *Applied Computing and Informatics*, Vol. ahead-of-print No. ahead-of-print, 2023. <https://doi.org/10.1108/ACI-09-2023-0022>
- [46] K.C. Chen, H.R. Yu, W.S. Chen et al, "Diagnosis of common pulmonary diseases in children by X-ray images and deep learning", *Sci Rep* **10**, 17374, 2020. <https://doi.org/10.1038/s41598-020-73831-5>.
- [46] F. M. J. M., Shamrat, S. Azam, A. Karim, R. Islam, Z. Tasnim, P. Ghosh, and F. De Boer, "LungNet22: A Fine-Tuned Model for Multiclass Classification and Prediction of Lung Disease Using X-ray Images", *Journal of Personalized Medicine*, 12(5), 680, 2022. <https://doi.org/10.3390/jpm12050680>

Article Information Form

Authors Contributions

Oladosu O. Oladimeji: Writing – original draft, Writing – review & editing, Visualisation, Validation, Methodology, Investigation, Data curation, Conceptualization.

Ayodeji O.J. Ibitoye: Writing – original draft, Writing – review & editing, Visualisation, Validation, Methodology, Investigation.

Conflict of Interest Notice

The authors declare that there is no conflict of interest regarding the publication of this paper.

Artificial Intelligence Statement

No artificial intelligence tools were used while writing this article.

Plagiarism Statement

This article has been scanned by iThenticate™.