Automated Diagnosis of COVID-19 and Pneumonia Using Deep Learning Techniques on Radiological Images

Kawther Sameer Ali and Ahmad Shaker Abdalrada
College of Computer Science and Information Technology, Wasit University, 52001 Al-Kut, Wasit, Iraq
kawthers206@uowasit.edu.iq; aabdalra@uowasit.edu.iq

Keywords: COVID-19, Pneumonia, EfficientNet-B3, Grad-CAM, Chest X-ray, Deep Learning.

Abstract: Respiratory disease, such as COVID-19 and pneumonia, are among the leading global causes of morbidity and mortality. Inexpensive yet universally applied chest X-ray (CXR) imaging is still difficult to interpret due to overlapping radiographic findings between diseases. In this paper, we propose an improved deep learning framework based on the EfficientNet-B3 architecture, aided by transfer learning, Grad-CAM visualizations, and data augmentation for autonomous diagnosis of respiratory diseases from CXR images. Two publicly available datasets were merged, cleaned, and balanced to create a heterogeneous training corpus of four classes of diagnostic conditions: COVID-19, bacterial pneumonia, viral pneumonia, and normal conditions. The proposed model was achieved accurate in test set 98.69% and high macro-averaged precision, recall, and F1-scores. The use of Grad-CAM visualizations enhanced the concentration of the model on clinically relevant lung regions, making it more explainable. These findings suggest the model's viability as a reliable clinical decision support system, especially in resource-limited settings, and are an advancement towards explainable AI in medical diagnosis.

1 INTRODUCTION

Respiratory disease is among the most common diseases on the planet [1]. For example, chronic respiratory diseases affected over 454 million people and were the third leading cause of death worldwide, causing approximately 4 million deaths annually [2] Pneumonia and COVID-19 are common types of such diseases [3]. Pneumonia, for instance, infects approximately 450 million individuals yearly and kills approximately 4 million, mostly children below the age of five and the elderly [4]. As much as COVID-19 is concerned, it has caused more than 770 million confirmed cases and over 7 million confirmed deaths worldwide since the start of the pandemic, as indicated by the World Health Organization in 2024 [5]. This raises the need for timely intervention using modern technologies.

The lungs are the body's second most crucial organs, and a key part plays in the all-important oxygen process needed for survival [6]. However, due to their fragile nature, they are prone to a lot of respiratory illnesses such as pneumonia and COVID-19 [7]. Both these diseases are some of the most commonly reported respiratory health hazards worldwide, primarily due to their sheer prevalence and sheer impact on the pulmonary system [2]. Moreover, the high level of similarity between clinical presentation and radiographic features of the two tends to generally complicate correct and prompt diagnosis [8].

Pneumonia is bacterial, viral, or fungal infection of the alveoli, which causes fever, cough, and shortness of breath [9], [10]. Whereas, COVID-19, caused by the novel coronavirus (SARS-CoV-2), directly affects the lungs and can range from mild symptoms to pneumonia, which has the same presentation as pneumonia [11]. That similarity makes it difficult and complicated to distinguish between the two by the conventional methods [12] The virus detection furthermore relies primarily on the RT-PCR test, which is highly specific. Even though its low sensitivity renders it susceptible to false negatives, which degrade the effectiveness of early detection [13]. On the other hand, computed tomography (CT) has been helpful in the determination of pulmonary changes due to COVID-19, particularly those presenting as "ground-glass opacities" in the periphery of the lungs [14]. Although this type of imaging is precise, the wide use of it is hampered by limitations such as it being costly and entails radiation exposure [15].

In contrast, chest x-ray (CXR) is also less costly and easily available and can be found commonly in hospitals and general practitioner centers [16]. Chest X-rays are considered a first-line, cost-effective diagnostic tool that can assist in identifying lung infections, chronic lung diseases, and other thoracic conditions [4], [5]. CXR images of COVID-19-infected patients show some characteristic patterns such as bilateral congestions and peripheral opacities, especially in the lower zones of the lungs [12]. But the sensitivity of this kind of imaging is less compared to CT, thus complicating the diagnosis, particularly in the initial stages of the disease [13] .The early and accurate detection of those illnesses through chest X-ray (CXR) imaging remains a critical clinical priority [17].

In this context, there has been increasing reliance on AI-based computer-aided diagnosis (CADx) systems to improve diagnostic accuracy and analyze medical images more efficiently[18]-[22].Among AI techniques, convolutional neural networks (CNNs) have proven highly effective in classifying lung diseases using CXR and CT images [23]. This research focuses on developing an intelligent diagnostic system capable of distinguishing between COVID-19 cases, pneumonia (both bacterial and viral), and healthy cases, based on chest x-ray images and CNN models

This is how the rest of the paper is structured. Related work is presented in Section 2. The materials and procedures are described in full in Section 3. Section 4 presents the comprehensive experimental setup. In Section 5, the experimental results are described in full. The conclusion in Section 6 follows this.

2 RELATED WORKS

Over the past few years, there has been considerable work in utilizing deep learning techniques for the detection of COVID-19 and pneumonia based on chest X-ray (CXR) images. The utilized models have varied from pre-trained deep networks to sophisticated segmentation and image enhancement techniques, but most have been faced with challenges of data balancing, accuracy in discriminating among the various forms of pneumonia, and limited generalizability of results.

For instance, In a recent study, by Sharma and Guleria [24] proposed VGG16-based model to classify pneumonia from CXR images. The model achieved accuracy of 95.4% on the second dataset used. While the model surpassed baseline classifiers such as SVM and Random Forest, it can be compromised by the imbalance of the dataset, especially considering the higher proportion of pneumonia cases with the rise in the pneumonia image data compared to both normal cases as well as to COVID-19. The failure to differentiate between viral pneumonia and bacterial also compromises its usability at a clinical level.

Similarly, Zhang [25] used a CAAD model based on EfficientNet-B0 to detect viral pneumonia with acceptable AUC results (87.57% for X-VIRAL and 83.61% for X-COVID). Their application of anomaly detection instead of direct classification, however, limited the potential of the model to distinguish between disease types.

In another approach, Goyal and Singh [26] were also proposed a technique that involves image enhancement and region-of-interest extraction, along with deep classifiers RNN and LSTM. The model accuracy was reported to be 94.31% for the CXIP dataset and 95.04% for the C19RD dataset, although the data taken into account was unbalanced, particularly in terms of the number of COVID-19 images and other cases, which can affect the performance of the model in real-life situations.

Furthermore, Hamdi [27] utilized the VGG-16 architecture to differentiate between COVID-19, viral pneumonia, and normal cases with 92.72% accuracy following the application of data augmentation techniques. Bacterial pneumonia was not considered in the study, making the model diagnostically incomplete.

In a similar vein, Hasan and et al. [28]also utilized a VGG-16 model and transfer learning to identify cases as COVID-19, pneumonia, and normal cases with accuracy of 91.69%. While these have been achieved results, discrimination among bacterial and viral infections, a critical aspect to correct diagnosis, was not targeted by the research.

Additionally, Militante, Dionisio [29] relied on VGG-16 to classify 20,000 CXR images including COVID-19, pneumonia, and normal cases, achieving 95% accuracy. However, distinguishing between bacterial and viral infections remained challenging due to the similarity in radiographic patterns, indicating the need to develop more specialized architectures or integrate other techniques to improve discrimination.

Moreover, Hammoudi [30] used the DenseNet169 model and achieved 95.72% accuracy. However, the model did not take COVID-19 as a separate class, but as a viral pneumonia, which can lead to discrimination errors. Besides, the dataset of COVID-

19 was limited (only 145 images), which compromises the validity of the findings.

In contrast, Khan and Bhat [31] proposed the CoroNet model, which is based on the Xception architecture, to identify normal cases, COVID-19, and pneumonia (variants). The model had a high recall for COVID-19 instances (98.2%) and an overall accuracy of 89.6%. However, the limited sample size, especially COVID-19 images, may impact the generalisability of the model.

Likewise, Manickam, Jiang [32] employed pre-trained models such as ResNet50 and InceptionV3 to classify pneumonia images as normal and abnormal (bacterial and viral) through transfer learning. Despite the study achieving 93.06% accuracy, it had no data on COVID-19, which reduces the comprehensiveness of the results.

Finally, Ieracitano, Mammone [33] suggested a model called CovNNet that integrated fuzzy logic techniques for edge detection with CNN structure to improve COVID-19 and pneumonia diagnosis from portable x-ray images. The model achieved an accuracy of 81% with great capacity to handle low-quality images. Nevertheless, the lack of discrimination between pneumonia types and the small sample size were study limitations.

In spite of dramatic advances in the application of deep learning methods in identifying COVID-19 and pneumonia from X-rays, some common challenges and constraints remain shared in most abovementioned works. Most research works have performed binary or ternary classification tasks, without sufficiently differentiating viral and bacterial pneumonia, which is essential in medical environments for prescribing proper treatment measures.

Besides, relying on imbalanced datasets – where COVID-19 images are outnumbered with regard to other classes – can bias model outputs and diminish their generalizability in real-world applications. While architectures such as VGG16, DenseNet, and Xception have exhibited remarkable performance, few works have explored the full potential of more scalable and efficient models such as EfficientNet, particularly under a multi-class classification setting comprising COVID-19, bacterial pneumonia, viral pneumonia, and healthy controls.

Also, the restricted interpretability of model output and lack of focus on clarifying predictive decision-making undermine physicians' confidence in such systems and impede effective implementation in clinical practice.

Thus, a definite research gap is identified: the need to develop a robust, interpretable, and data-balanced diagnostic model that can efficiently distinguish COVID-19 from other types of pneumonia based on CXR images. Bridging this gap – utilizing efficient architectures, transfer learning techniques, attention mechanisms, and explainable AI methods – is the primary motivation for this research.

3 MATERIALS AND METHODS

This section presents the evaluation and interpretation of the proposed predictive model. It also outlines the dataset, experimental parameters, and the performance metrics employed.

3.1 Dataset Description

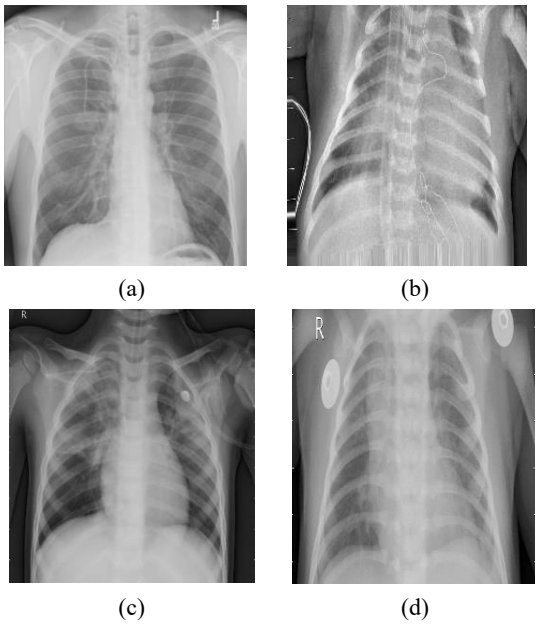
In this study, a combined dataset was constructed by integrating two publicly available chest X-ray (CXR) datasets to support a robust multiclass classification framework. The first source was the COVID-19 Radiography Database [34] and its expanded version, the COVID-QU-Ex Dataset, developed collaboratively by Qatar University, the University of Dhaka, and their medical partners. This dataset is composed of frontal CXR images which are marked as COVID-19 (11,956 images), non-COVID pneumonia 11,263 images, and normal lungs (10,701 images). Lung segmentation masks are part of this dataset but were not used in this work.

The second dataset was gathered from paediatric cohorts (ages 1–5) at the Guangzhou Women and Children's Medical Centre in China and is known as Chest X-Ray Images (Pneumonia) [35], The dataset contains 5,863 anterior-posterior (AP) X-ray images of two classes: pneumonia and normal. Expert radiologists assessed and validated all the scans in terms of diagnostic performance and image quality. From those two datasets, a highly balanced and curated dataset was created with approximately 32,900 images, divided into four diagnostic classes: COVID-19, bacterial pneumonia, viral pneumonia, and normal. Figure 1: shows representative samples from each of the four diagnostic categories.

3.2 Performance Metrics

This study employed a set of widely used benchmark evaluation metrics to assess the effectiveness of the proposed prediction model, including sensitivity (recall), specificity, precision, F1-score, accuracy, and the confusion matrix. These metrics are standard

in medical image classification and diagnostic decision-support systems, where both false positives and false negatives may have serious clinical implications [44], [45].



(a) (b)

(c) (d)

Figure 1: Sample images: a) COVID19 images; b) Normal images; c) Bacterial pneumonia images; d) Viral pneumonia images.

In addition, micro- and macro-averaged scores were reported to provide a comprehensive assessment of the model's performance across all diagnostic classes. Such averaging strategies are commonly adopted in multi-class medical classification tasks, especially when class imbalance is present [44].

Sensitivity (also referred to as recall) reflects the model's ability to correctly identify positive cases, which is critical in medical screening applications to reduce the risk of missed diagnoses. Specificity measures the model's capability to correctly identify negative cases and is essential for minimizing false alarms and unnecessary clinical follow-up procedures. Accuracy provides an overall measure of correct predictions across both positive and negative classes, while precision evaluates the reliability of positive predictions. The F1-score combines precision and recall into a single metric, offering a balanced evaluation that accounts for both false positives and false negatives. These metrics are widely recognized and extensively applied in medical imaging and machine learning-based diagnostic studies [44], [45].

For multi-class classification scenarios, macro averaging computes evaluation metrics independently for each class and then averages them equally, ensuring that all classes contribute uniformly to the final score. This approach is particularly useful for assessing the model's ability to handle each diagnostic category independently. In contrast, micro averaging aggregates true positives, false positives, and false negatives across all classes before computing the metrics, providing a global performance perspective that is especially beneficial for imbalanced datasets [44].

Both macro- and micro-averaged metrics were included in this study to ensure a balanced and holistic evaluation of the model's classification performance.

The confusion matrix was used to provide a detailed summary of the model's predictions for each class, including true positives, true negatives, false positives, and false negatives. In the context of this work, the confusion matrix facilitated an in-depth analysis of the model's ability to distinguish among four diagnostic categories: COVID-19, bacterial pneumonia, viral pneumonia, and normal cases. Correct classifications are represented along the diagonal of the matrix, where higher values indicate stronger classification performance.

3.3 Experiment

The experiment work for this study was carried out in two main stages on different computing platforms. Development, debugging, and early testing were done locally using Jupyter Notebook from the Anaconda package on an individual notebook laptop with a 2.5 GHz Intel Core i7 processor and 12 GB of RAM under Windows 10. Implementation was under the Python computer programming language. For end-to-end model training and performance evaluation, the experiments were subsequently migrated to Kaggle cloud platform, which supports powerful GPU setup. The models were trained on an NVIDIA Tesla P100 GPU in Kaggle's hosted Jupyter Notebook platform. The software stack used Python 3.10 and PyTorch 2.0.

As outlined in Section 3.2, multiple evaluation criteria were used to measure predictive performance of model under proposal. In addition to conventional accuracy and loss criteria, confusion matrix statistics were computed to gain more insight into classification performance and uncover patterns of misclassification.

To evaluate the model's generalization performance, the dataset was assembled by integrating two publicly available chest X-ray (CXR) datasets. Initially, the data were divided into training,

validation, and testing subsets using an approximate 70%, 10%, and 20% split, respectively. At this stage, the training set comprised 23,687 images across four diagnostic categories: COVID-19, bacterial pneumonia, viral pneumonia, and normal. Due to significant class imbalance, data augmentation techniques were applied to the training set to equalize class representation, resulting in a balanced set of 43,028 images. Combined with the validation (2,635 images) and testing (6,578 images) subsets, the total dataset used in this study consisted of 52,241 chest X-ray images.

Key performance indicators – including accuracy, sensitivity, and specificity – were calculated to assess the effectiveness of the predictive model. These metrics provided a comprehensive understanding of the model's ability to distinguish between the targeted pulmonary conditions.

4 THE PROPOSED MODEL

This section illustrates the proposed model for the prediction of the lung disease (COVID or Pneumonia). The proposed deep learning-based classification framework is divided into several stages: dataset preparation, preprocessing and augmentation, feature extraction, model training, evaluation, and explainability analysis using Grad-CAM, as shown in Figure 2. Preprocessing operations were carried out before training to make the data uniform and improve the generalization performance of the model. Since the original images were of varying sizes, all images were resized to 300×300 pixels to ensure consistency in input dimensions across the model. Also, the images were saved in PNG format to maintain visual integrity and prevent compression artifacts of lossy formats like JPEG [36]

Preprocessing operations were carried out before training to make the data uniform and improve the generalization performance of the model. Since the original images were of varying sizes, all images were resized to 300×300 pixels to ensure consistency in input dimensions across the model. Also, the images were saved in PNG format to maintain visual integrity and prevent compression artifacts of lossy formats like JPEG [36] .To enhance model robustness and avoid overfitting, And to fight the class imbalance issue, we applied data augmentation to the training set to obtain a perfectly balanced set with exactly 10,757 images in each class (COVID, bacterial_pneumonia, viral_pneumonia, and normal) for a total of 43,028 images. The validation and test sets were kept unbalanced to keep realistic evaluation as well as to avoid leakage therefore were not augmented. The final balance of images in each split is presented in Table 1.

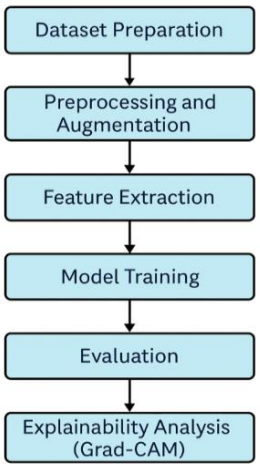


Figure 2: A flowchart illustrating the proposed deep learning-based classification pipeline.

Table 1: Final distribution of images per class in each dataset split.

Class	Train (Balanced)	Validation	Test
BACTERIAL_ PNEUMONIA	10,757	670	1,672
VIRAL_ PNEUMONIA	10,757	479	1,195
NORMAL	10,757	1,196	2,988
COVID	10,757	290	723
Total	43,028	2,635	6,578

Data augmentation techniques were applied to the training set like random rotation, horizontal flip, brightness, and translation. These augmentations simulate real-world variability and help the model learn invariant features [37].

The training set was fully balanced via augmentation, ensuring equal representation of each class. This balancing approach kept the model from being biased toward any specific class during training, whereas the unbalanced validation and test sets provided a more realistic real-world performance estimate. The training set was employed for tuning model parameters, and the validation set was utilized for observing training progress and model selection. To avoid overfitting and increase generalization, two regularization techniques were applied: early stopping and Stochastic Weight Averaging (SWA). Early stopping was employed to terminate training once the optimal validation performance ceased to improve, thus preventing it from learning noise or

spurious patterns in the training dataset[38]. In contrast, SWA is employed to average model weights over multiple epochs close to convergence, yielding a flatter and stable solution that enhances the model's robustness and generalization performance. In the same vein, the test set was held strictly aside for final evaluation, in order to have an unbiased assessment of how well the model generalizes to new data[39].

For feature extraction and classification, the EfficientNet-B3 model was employed. The model was initialized with ImageNet-pretrained weights, leveraging transfer learning to retain low-level visual features while adapting the final layers to the target task. The final classification head was modified to include a dropout layer (p=0.4) and a fully connected linear layer with four output neurons, corresponding to the four diagnostic classes. The EfficientNet-B3 model was chosen due to its balance between computational efficiency and high accuracy[40, 41]. Transfer learning was also used, leveraging weights pre-trained on large datasets such as ImageNet, to improve the model's generalization ability[40]. To increase the transparency of interpretation and understand the areas affecting the prediction process, Grad-CAM (Gradient-weighted Class Activation Mapping) technology were combined into the methodology[42]. This approach aims to address the challenges associated with the similarity of visual features across different diseases, as well as the limited diversity of available training data[43].

The model was trained with the AdamW optimizer, a learning rate of 3e-4, and cross-entropy loss with label smoothing (0.05). To enhance training stability and performance, CosineAnnealingWarmRestarts was used as a learning rate scheduler, along with early stopping and Stochastic Weight Averaging (SWA). Upon training completion, the model's performance was assessed using the unseen test dataset.

5 RESULT AND DISCUSSION

As shown in Table 2, the performance of the proposed deep learning model was thoroughly assessed in this section using a number of important classification metrics, such as accuracy, precision, recall, specificity, F1-score, and macro average. These metrics provide a thorough grasp of the diagnostic capabilities of the model for each of the four target classes: Normal, bacterial, viral, and COVID-19 pneumonia [44] .

Given the clinical significance of minimizing misdiagnoses (especially for COVID-19 and pneumonia cases), these complementary metrics are critical to evaluating the model's robustness. In this study, after training over 30 epochs, the model across the entire test dataset achieved 98.69% accuracy, 0.987 macro-averaged precision, 0.982 recall, 0.985 F1-score, 0.995 specificity, and 0.99 AUC, demonstrating robust classification across the four diagnostic classes. Detailed per-class results are reported in Table 2, and illustrated by the confusion matrix in Figure 3. These results demonstrate consistently high precision, recall, Specificity, and F1-score across all categories, highlighting the robustness of the model [45] .

Grad-CAM Visualization Results: Figure 4 shows Grad-CAM (Gradient-weighted Class Activation Mapping) visualizations of example chest X-ray (CXR) images from each of the four diagnostic classes: Viral Pneumonia, Normal, Bacterial Pneumonia, and COVID-19. Each row shows the original CXR image (left), the exaggerated Grad-CAM heatmap (center), and the overlay of the heatmap on the original image (right).

These visualizations provide a sense of the regions of the lungs that the EfficientNet-B3 model is paying attention to in order to make its predictions:

Table 2: Illustrates the model's performance.

Class	Precision	Recall (Sensitivity)	Specificity	F1-Score	AUC	Support
BACTERIAL_PNEUMONIA	0.990	0.985	0.995	0.990	–	1672
COVID	0.990	0.968	0.999	0.980	–	723
NORMAL	0.990	0.996	0.991	0.990	–	2988
VIRAL_PNEUMONIA	0.980	0.978	0.994	0.980	–	1195
Macro Average	0.987	0.982	0.995	0.985	0.99	6578
Micro Average	0.986	0.986	–	0.986	–	6578
Overall Accuracy	–	–	–	0.987	–	6578

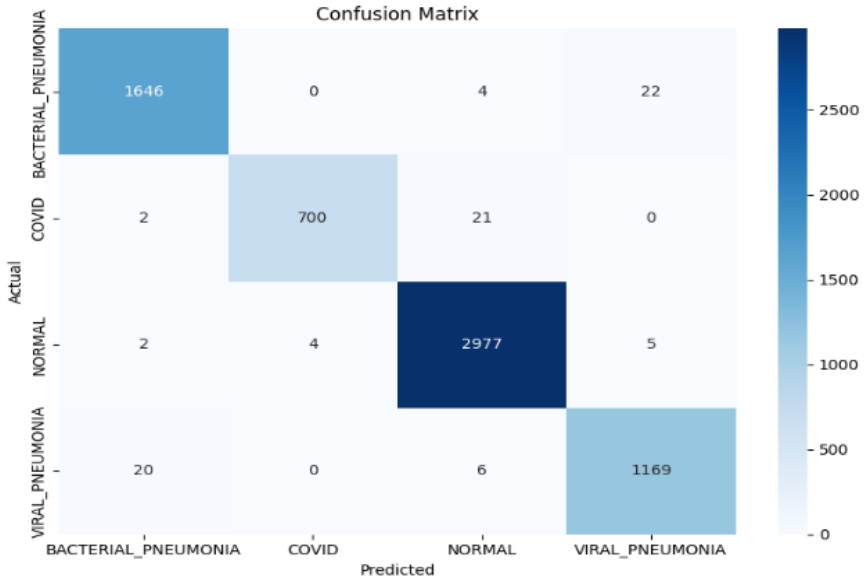
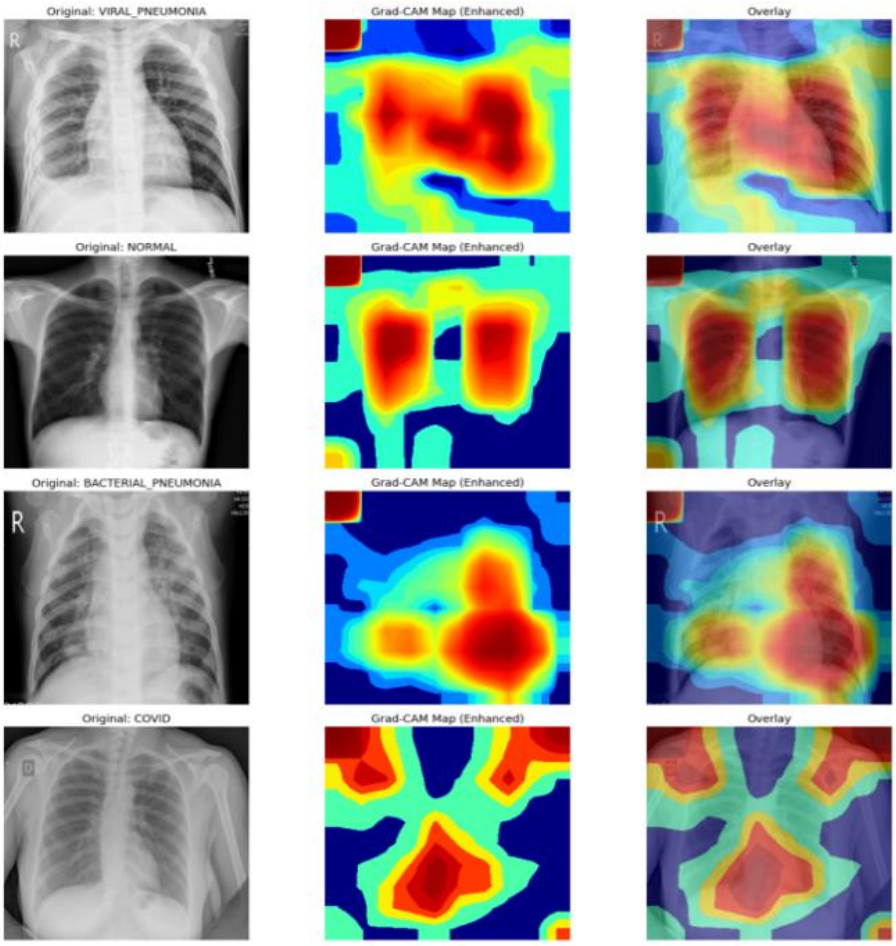


Figure 3: Confusion matrix showing the true vs. predicted labels for each class.



Figures 4: Illustrate Grad-CAM visualizations for each class with the Correct prediction.

In Viral Pneumonia, the model is highlighting bilateral mid-to-lower lung fields, in keeping with the clinical presentation of viral infiltration. In Normal cases, attention is diffusely distributed without any notable areas of activation, in keeping with the absence of abnormal radiographic findings. For Bacterial Pneumonia, the model highlights focal areas intensely, mirroring the lobar consolidation patterns of bacterial infections. For COVID-19, the model emphasizes peripheral and basal lung regions, consistent with known radiological features of COVID-19 such as ground-glass opacities. These Grad-CAM visualizations confirm that the model's predictions are grounded in clinically sound features, rendering the model more interpretable and more reliable for medical diagnosis.

The model was optimized with several methods including advanced augmentation techniques, label smoothing, AdamW optimizer, SWA, and GradScaler mixed precision training .Early stopping were also employed to boost performance and prevent overfitting.

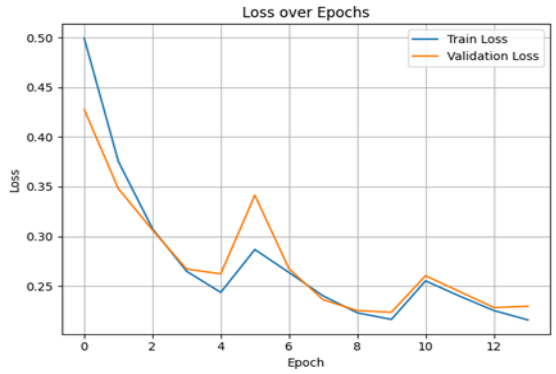


Figure 5: Training and validation loss.

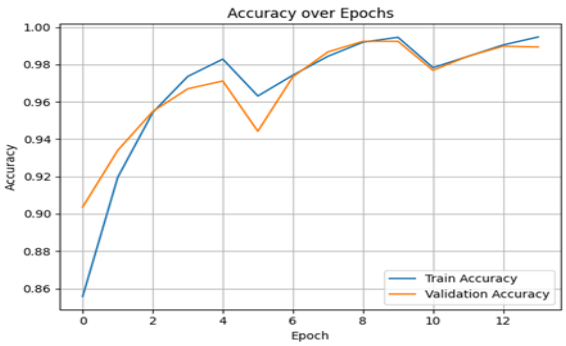


Figure 6: Training and validation accuracy.

A detailed breakdown by class is shown in Table 2, and the confusion matrix is visualized in Figure 3, clearly illustrating the model's exceptional ability to differentiate between four categories: COVID, Bacterial_Pneumonia, Viral_Pneumonia, and Normal.

In Figures 5 and 6. The left panel illustrates the loss curves, where both validation and training loss decrease consistently, indicating good convergence with no harsh overfitting. The right panel illustrates the accuracy curves, where both validation and training accuracy increase consistently across epochs. The relatively narrow gap between the two curves guarantees that the model is generalizing well and is stable while learning.

Compared to the traditional models like VGG16-based classifiers (i.e., VGG16+SVM, VGG16+KNN, NB, or RF), our new EfficientNet-B3 model significantly outperformed in terms of accuracy, stability, and interpretability. Unlike these models, our method did not require external handcrafted feature engineering but enjoyed end-to-end feature extraction with fewer parameters and higher efficiency. Therefore, the proposed approach attains state-of-the-art performance in multiclass pneumonia detection and can find application as a reliable diagnostic assistant in clinical environments.

6 CONCLUSIONS

Early and correct detection of pneumonia, COVID-19 and other respiratory tract infections, plays a crucial role in improving the outcome of the treatment. The current article described a top-ranked deep learning algorithm using EfficientNet-B3 for diagnosing chest X-ray images into four disease classes. Data augmentation, transfer learning, and visualization techniques like Grad-CAM helped the algorithm achieve remarkable accuracy and interpretability. In the future, we plan to extend this work with more diverse datasets from different institutions and incorporating clinical metadata to further improve the model's generalization capability. The model was trained and evaluated on merged, preprocessed datasets: the COVID-19 Radiography Database and the Chest X-Ray Images (Pneumonia) dataset. Images were split into 70% train, 10% validation, and 20% test sets to allow for solid and balanced evaluation. Large-scale experiments confirm that the proposed EfficientNet-B3 model has attained a strong test accuracy of 98.69%, with macro-averaged precision, recall, and F1-scores of 0.987, 0.983, and 0.985, respectively, for every class. Moreover, Grad-CAM visualizations confirmed that the model was paying attention to significant pulmonary regions, and this added extra interpretability and clinical confidence to

the predictions Compared to previous models such as VGG16-based CNNs, SVM, KNN, Naïve Bayes, and Random Forest, which achieved test accuracies typically ranging from 92% to 95% and often struggled to distinguish between pneumonia subtypes, our EfficientNet-B3 model achieved a higher accuracy of 98.69% with robust performance across all four classes. This demonstrates clear superiority in multiclass discrimination and clinical reliability.

This demonstrates its ability to be employed as a reliable tool for helping radiologists discriminate COVID-19 and other forms of pneumonia from normal CXR images. In the future, we intend to increase the model's generalizability by merging heterogeneous, cross-center data, exploring multimodal learning through imaging integration with clinical or genomic data, and deploying the model in real-time clinical settings with explainable AI interfaces for clinician feedback and validation. These developments may facilitate more transparent, scalable, and accurate AI-assisted diagnostic systems in routine clinical workflows.

REFERENCES

[1] M. Institute for Health and Evaluation, "Global burden of chronic respiratory diseases and risk factors, 1990–2019," 2020. [Online]. Available: https://www.healthdata.org/research-analysis/library/global-burden-chronic-respiratory-diseases-and-risk-factors-1990-2019.

[2] GBD Chronic Respiratory Disease Collaborators, "Global burden of chronic respiratory diseases and risk factors, 1990–2019: an update from the Global Burden of Disease Study 2019," eClinicalMedicine, vol. 61, p. 101936, Apr. 2023, doi: 10.1016/j.eclinm.2023.101936.

[3] World Health Organization, "WHO Coronavirus (COVID-19) Dashboard," 2024. [Online]. Available: https://covid19.who.int.

[4] Cleveland Clinic, "Chest X-ray," 2022. [Online]. Available: https://my.clevelandclinic.org/health/diagnostics/10228-chest-x-ray.

[5] C. Luo, "Deep learning-based early detection of pneumonia and COVID-19 using chest X-ray," Scientific Reports, vol. 14, Art. no. 60861, 2024. [Online]. Available: https://www.nature.com/articles/s41598-024-60861-6.

[6] J. E. Brinkman and S. S. Brinkman, "Physiology, Pulmonary," in StatPearls [Internet]. Treasure Island, FL, USA: StatPearls Publishing, 2023.

[7] Centers for Disease Control and Prevention, "Underlying Conditions and the Higher Risk for Severe COVID-19," 2023. [Online]. Available: https://www.cdc.gov/covid/hcp/clinical-care/underlying-conditions.html.

[8] H. X. Bai et al., "Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT," Radiology, 2020, doi: 10.1148/radiol.2020200823.

[9] World Health Organization, "Pneumonia Fact Sheet," 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia.

[10] MedlinePlus, "Pneumonia," 2024. [Online]. Available: https://medlineplus.gov/pneumonia.html.

[11] World Health Organization, "Novel Coronavirus – China," 2020. [Online]. Available: https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/.

[12] H. Y. F. Wong et al., "Frequency and distribution of chest radiographic findings in COVID-19 positive patients," Radiology, 2020, doi: 10.1148/radiol.2020201160.

[13] Y. Fang et al., "Sensitivity of chest CT for COVID-19: comparison to RT-PCR," Radiology, vol. 296, no. 2, pp. E115–E117, 2020.

[14] M. Chung et al., "CT imaging features of 2019 novel coronavirus (2019-nCoV)," Radiology, vol. 295, no. 1, pp. 202–207, 2020.

[15] A. Sodickson et al., "Recurrent CT, cumulative radiation exposure, and associated radiation-induced cancer risks from CT of adults," Radiology, vol. 251, no. 1, pp. 175–184, 2009.

[16] P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv, 2017. [Online]. Available: https://arxiv.org/abs/1711.05225.

[17] A. S. Musallam, A. S. Sherif, and M. K. Hussein, "Efficient framework for detecting COVID-19 and pneumonia from chest X-ray using deep convolutional network," Egyptian Informatics Journal, vol. 23, no. 2, pp. 247–257, Jul. 2022, doi: 10.1016/j.eij.2022.01.002.

[18] L. Rasmy et al., "Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction," NPJ Digital Medicine, vol. 4, no. 1, p. 86, 2021.

[19] A. Abdalrada, A. F. Neamah, and H. Murad, "Predicting diabetes disease occurrence using logistic regression: An early detection approach," Iraqi Journal for Computer Science and Mathematics, vol. 5, no. 1, pp. 160–167, 2024.

[20] Z. A. Anwer and A. S. Abdalrada, "Assessing institutional performance using machine learning on Arabic Facebook comments," Engineering, Technology & Applied Science Research, vol. 14, no. 4, pp. 16025–16031, 2024.

[21] A. S. Abdalrada et al., "Machine learning models for prediction of co-occurrence of diabetes and cardiovascular diseases: A retrospective cohort study," Journal of Diabetes & Metabolic Disorders, vol. 21, no. 1, pp. 251–261, 2022.

[22] A. S. Abdalrada et al., "Meta learning ensemble technique for diagnosis of cardiac autonomic neuropathy based on heart rate variability features," in Proc. 30th Int. Conf. Computer Applications in Industry and Engineering (CAINE), 2017, pp. 169–175.

[23] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," Insights into Imaging, 2018, doi: 10.1007/s13244-018-0639-9.

[24] A. Sharma and A. Guleria, "A deep learning based model for the detection of pneumonia from chest X-ray images using VGG-16 and neural networks," Procedia Computer Science, 2023.

[25] Y. Zhang, "Confidence aware anomaly detection (CAAD) model for detecting viral pneumonia from chest X-rays," Computers in Biology and Medicine, 2023.

[26] M. Goyal and D. Singh, "RNN and LSTM-based technique for image enhancement and region-of-interest extraction in pneumonia detection," Journal of Ambient Intelligence and Humanized Computing, 2022, doi: 10.1007/s12652-021-03464-7.

[27] F. Hamdi and T. S. Kandil, "Detection of COVID-19 and pneumonia in chest X-ray images using deep learning techniques," 2024, doi: 10.1109/ICCI61671.2024.10485175.

[28] M. D. K. Hasan et al., "Deep learning approaches for detecting pneumonia in COVID-19 patients by analyzing chest X-ray images," Mathematical Problems in Engineering, 2021, doi: 10.1155/2021/9929274.

[29] S. V. Militante, N. V. Dionisio, and B. G. Sibbaluca, "Pneumonia and COVID-19 detection using convolutional neural networks," in Proc. IEEE ICVEE, 2020, pp. 1–6.

[30] K. Hammoudi, "Deep learning on chest X-ray images to detect and evaluate pneumonia cases," Journal of Medical Systems, 2021, doi: 10.1007/s10916-021-01745-4.

[31] A. I. Khan and M. M. Bhat, "CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest X-ray images," Computers in Biology and Medicine, 2021.

[32] A. Manickam, "Automated pneumonia detection on chest X-ray images," Measurement, 2021, doi: 10.1016/j.measurement.2021.110506.

[33] C. Ieracitano and N. Mammone, "CovNNet: A deep learning model combining fuzzy logic and CNNs for COVID-19 and pneumonia diagnosis," Neurocomputing, 2020.

[34] T. Rahman, "COVID-19 Radiography Database," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database.

[35] P. T. Mooney, "Chest X-Ray Images (Pneumonia)," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia.

[36] R. H. Wiggins et al., "Image file formats: past, present, and future," Radiographics, vol. 21, no. 3, pp. 789–798, 2001.

[37] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of Big Data, vol. 6, no. 1, p. 60, 2019, doi: 10.1186/s40537-019-0197-0.

[38] B. M. Hussein and S. M. Shareef, "An empirical study on the correlation between early stopping patience and epochs in deep learning," in ITM Web of Conferences, vol. 64, p. 01003, 2024, doi: 10.1051/itmconf/20246401003.

[39] H. Guo, J. Jin, and B. Liu, "Stochastic weight averaging revisited," Applied Sciences, vol. 13, no. 5, p. 2935, 2023, doi: 10.3390/app13052935.

[40] C. Tan et al., "A survey on deep transfer learning," in Lecture Notes in Computer Science, vol. 11141, 2018

[41] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Machine Learning, 2019, pp. 6105–6114.

[42] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Computer Vision, 2017, pp. 618–626.

[43] E. Tartaglione et al., "Unveiling COVID-19 from chest X-ray with deep learning: A hurdles race with small data," arXiv, 2020. [Online]. Available: https://arxiv.org/abs/2004.05405.

[44] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information Processing & Management, 2009, doi: 10.1016/j.ipm.2009.03.002.

[45] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," BMC Medical Imaging, 2015, doi: 10.1186/s12880-015-0068-x.