Received: 22 November 2024 | Revised: 11 April 2025 | Accepted: 5 June 2025 | Published online: 18 July 2025

RESEARCH ARTICLE

Artificial Intelligence and Applications
2025, Vol. 00(00) 1–13
DOI: 10.47852/bonviewAIA52024885

BON VIEW PUBLISHING

An Effective ResNet Model for Respiratory Disease Detection: A Case Study on COVID-19 Chest X-ray Images

Bushra Hassan[1], Kaveh Kiani[1, 2, *], and Taha Mansouri[1, 2]

[1] *School of Science, Engineering and Environment, University of Salford, UK*

[2] *Data Science and AI HUB, University of Salford, UK*

Abstract: The COVID-19 pandemic has placed an extraordinary burden on healthcare systems worldwide and underscored the urgent need for faster, more reliable diagnostic methods for respiratory diseases. In this study, we explored how artificial intelligence, particularly deep convolutional neural networks (CNNs), can assist in diagnosing COVID-19 using chest X-ray images. Although new CNN architectures continue to emerge, residual neural network (ResNet) models (ResNet50, ResNet101, and ResNet152) remain popular because of their balance of accuracy and robustness. We compared their performance with other established models such as VGG (VGG11 and VGG16) and AlexNet to examine the trade-offs between architectural complexity and practical effectiveness. Among these, ResNet101 achieved the best results, with an accuracy of 93%, a precision of 91.27%, and an AUC–ROC of 97.41%, making it the leading model in our study. Despite promising outcomes, we also highlighted current limitations in deep learning for medical imaging, such as small sample sizes and limited clinical validation. Our findings aim to support the future development of reliable AI-assisted tools in medical diagnostics and highlight the need for more comprehensive validation and ethical considerations.

Keywords: AlexNet, VGG, ResNet, transfer learning, convolutional neural network, computer vision, computer-aided diagnosis, respiratory disease detection

1. Introduction

Acute respiratory diseases pose significant challenges to the global healthcare community, demanding accurate and effective diagnostic techniques to mitigate their impact on public health. The COVID-19 epidemic and its highly contagious nature have presented an unprecedented challenge to the worldwide healthcare community. Accurate and effective diagnosis techniques are rather important in stopping the virus's spread and lessening its effects on public health as it is highly contiguous. Polymerase chain reaction (PCR) tests and other common laboratory-based diagnostic methods have problems with cost, time, and availability, particularly in resource-constrained settings [1]. AI-based methodologies are designed to enhance, speed up, and support diagnoses, thereby assisting healthcare practitioners in identifying health issues at an earlier stage and providing individuals with superior, more prompt care [2, 3]. X-ray and CT scans of the thorax have become indispensable instruments for formulating classification models. Radiologists can detect and evaluate COVID-19 levels in the lungs using these imaging techniques by looking for ground-glass opacity and their characteristics [4].

This paper compares a number of deep learning models for the detection of COVID-19, such as ResNet50, ResNet101, ResNet152, VGG11, VGG16, and AlexNet, as referenced by Wang et al. [5]. These deep learning models were found to be exceptional in real-world computer vision tasks, and they are extensively used in medical image analysis. Even with the introduction of new architectures, the residual neural network (ResNet) models (such as ResNet50, ResNet101, and ResNet152) are highly relevant and extensively used because of their proven performance and resilience. Comparing the models with older architectures such as VGG and AlexNet sheds important insights into the trade-off between simplicity and complexity in deep learning. This study finds direct relevance in practical use in medical image analysis, where dependable and effective models are indispensable.

To enhance their robustness and precision, researchers consider various techniques in different studies such as transfer learning, pretrained network fine-tuning, hybrid approaches, ensemble methods, and data augmentation [6]. Depth and model structure make deep learning models distinguishable from their counterparts. ResNet models show higher complexity and detail, enabling the capture and representation of complex details [7]. VGG models employ small filter sizes and increased numbers of convolutional layers for distinct feature extraction. Despite its relatively simple structure , AlexNet manifests good performance in tasks involving image classification. Model choice relies on factors such as job complexity, processing resources, and the trade-off between model complexity and precision [4, 6].

This paper consists of several parts. Part one of this paper involves a literature review on the use of deep learning models for the analysis of COVID-19, specifically for the analysis of chest radiographs. The second part discusses the methodology, the preparation of the dataset, and the deep learning models employed. Model performance study and the findings of the assessment metrics are given in this paper afterward. Deep learning models have great potential in COVID-19 diagnosis, but certain challenges and constraints need to be noted. Potential directions for further study are discussed in the conclusion of this paper.

*Corresponding author: Kaveh Kiani, School of Science, Engineering and Environment, University of Salford, United Kingdom. Email: k.kiani@salford.ac.uk

2. Literature Review

Since its emergence in late 2019 and subsequent declaration as a pandemic by the WHO in March 2020, COVID-19 has driven intensive research activities in search of effective diagnostic tools. The AI research community has been in the frontline in the development of models that could boost speed and accuracy in the detection of the disease to effectively control the spread of the virus [8–12].

Conventional detection methods include PCR, which continues to be the benchmark for COVID-19 identification. However, notable drawbacks of testing using PCR exist: very long turnaround times in sample collection, processing, and result delivery. In addition, the resource limitation and high testing cost in developing countries have significantly restricted its widespread use so far [13]. In the context of these challenges, several variants of AI-driven models were considered as viable alternatives that could offer faster and inexpensive detection of COVID-19. Imaging in radiology, namely, CT scans and chest X-rays, has been widely used in COVID-19 diagnosis. CT scans, which generate high-resolution 3D images of lung abnormalities—such as ground-glass opacity—are exceptionally effective at identifying and depicting the extent of an infection. However, CT scanners are expensive and less accessible and expose patients to higher radiation levels, with possible long-term health risks associated [14]. In comparison, chest X-rays are safer, cheaper, and more available, especially in resource-constrained settings. Despite these advantages, some challenges exist in maintaining high-quality X-ray datasets and ensuring that deep learning models effectively generalize across different clinical settings [9].

Deep learning algorithms are becoming increasingly popular in today's world for the detection and identification of COVID-19 from chest X-ray images. Extensive research conducted by Zheng et al. [15] and Krizhevsky et al. [6] provided substantial evidence that these algorithms can successfully complete the COVID-19 classification tests with high levels of sensitivity and accuracy. A well-known example of a deep learning model is ResNet, which is used for image classification. During their research, He and his colleagues [1] developed ResNet50, ResNet101, and ResNet152, which outperformed earlier models. This was accomplished while they were competing in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). These architectures make use of residual connections to train deeper networks to effectively capture intricate information that is present in medical images. The VGG model is a notable deep learning architecture proposed by Simonyan and Zisserman [16] for image classification. VGGNet is a network that is highly accurate but requires a significant number of computational resources because it is constructed using many convolutional layers with relatively small filter sizes. AlexNet, which was presented by Krizhevsky and colleagues [6], was a significant step forward in the field of deep learning because of its success in the ILSVRC competition. High-level features are extracted by AlexNet using a convolutional neural network (CNN) that has many layers. It has demonstrated excellent performance in image classification tests, although it is straightforward in comparison to later architectures [17]. Since the onset of the pandemic, significant interest exists in using deep learning models to detect COVID-19 in CT and chest X-ray images. Researchers have employed various architectures and techniques to enhance classification accuracy, robustness, and interpretability, as outlined below.

Using deep learning architectures, viz., VGG16 and InceptionV3, Apostolopoulos and Mpesiana [8] explored COVID-19 detection in chest X-ray images in 2019. Methods for transfer learning were proposed in their study, allowing fine-tuning of pretrained models for better performance on their dataset. Wang et al. [7] used the ResNet50 model in 2020 for COVID-19 categorization, proving its efficiency for the detection of cases using CXR images. Khan et al. [18] also used the VGG16 model in 2020, which has a transfer learning approach similar to that used by Wang et al. [7]. Making the pretrained model more efficient on their dataset was the primary focus. Li et al. [19] further added value by combining the ResNet50 and VGG16 architectures in a hybrid model, leveraging their strengths for better overall performance. In 2022, the model DeepCOVIDExplainer was proposed by Elaziz et al. [20], which was a ResNet50 model with an explanation mechanism for visualizing and understanding its decision-making, thus increasing transparency and trust. With the progression of the field, in 2023, Aslani and Jacob [21] performed a critical review of 30 experiments using deep learning approaches for the diagnosis of COVID-19. With the review spanning transfer learning, ensemble learning, and GAN-based augmentation methods, the maximum obtained classification accuracies were up to 99.05%. Even with such achievements, the authors noted challenges of limited accessibility of information, lack of interpretability, and overfitting [21]. Recent developments in 2024 depicted continued optimization and deployment as a focus area. Abdulahi et al. [22] proposed PulmoNet, a new 26-layer CNN model optimized for multiclass and binary classifications. PulmoNet obtained stellar rates of accuracy, including 99.4% for COVID-19 vs. healthy [22]. Chauhan et al. [23] developed a lightweight model of CNN for edge deployment with high accuracy, low complexity, and a low risk of overfitting. Kaur et al. [24] developed a short-form CNN model optimized for low-end devices using image enhancement, data augmentation, and hyperparameter optimization, allowing it to outperform top transfer learning models. New techniques continued to surface, including the use of automated machine learning (AutoML) by Yin et al. [25] in the identification of asymptomatic COVID-19 patients. Among the six models developed, the deep neural network (DNN) performed the most optimally with an AUC of 0.898 and an accuracy of 83.7%. Okada et al. [26] developed an AI diagnosis system for COVID-19 based on CT imagery using binary-classification deep learning architectures. Their system was found to perform steadily with fast inference rates for clinical deployment.

These works outline the development of deep learning methods for COVID-19 identification, with progressive developments in model structure, transfer learning techniques, and deployment optimizations. Although promising results were obtained, issues with interpretability and availability of data remain as a subject for further study.

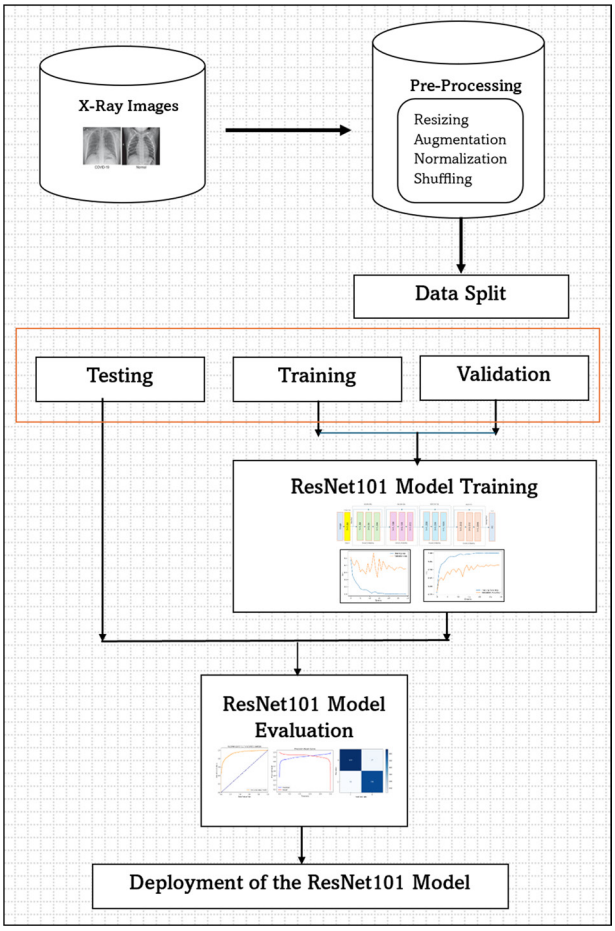
3. Proposed Methodology

In this paper, the authors have chosen and assessed few popular and practical deep learning models for the detection of COVID-19, one of the most severe respiratory diseases, from chest X-ray images. The primary objective was to assess how the original ResNet models, in comparison with other popular deep learning architectures, performed on the COVID-19 X-ray datasets. That is, we were interested in determining whether any of the original ResNet models can be suggested as a valid model for the detection of COVID-19 after our comparative analysis.

This study suggested the ResNet101 model as the core methodology for the detection of COVID-19. ResNet101 is a deep CNN with 101 layers. It was designed with the sole purpose of avoiding the vanishing gradient issue while adding residual connections. This allowed the training of a network with increased efficiency, although with a larger depth [17]. The architecture diagram of the proposed methodology of this study is presented in Figure 1.

ResNet models, introduced by He et al. [1], represent a seminal advancement in deep learning, most notably in training DNNs. ResNet's structure is defined by the presence of residual connections, an innovative feature allowing the network to learn residual mappings as opposed to the direct modeling of the original function. This approach overcomes

Figure 1
Architecture diagram of the proposed methodology



the vanishing gradient problem, allowing the training of extremely deep networks and making ResNet a popular choice in medical image analysis [17]. ResNet's structure consists of many residual blocks, each with convolutional layers, a batch norm, and ReLU activations. Residual connections in the blocks provide a direct shortcut between the block's output and input by bypassing one or more convolutional layers. This structure supports the effective preservation of residual information, maximizing gradient flow under backpropagation and leading to better model performance. ResNet models come in various configurations, such as ResNet50, ResNet101, and ResNet152, distinguished by their respective layer numbers. For instance, ResNet50 consists of 50 layers, ResNet101 consists of 101 layers (Figure 2), and ResNet152 consists of 152 layers, allowing it to learn more complex features. Each model follows a hierarchical structure, with the starting block as the first convolutional layer, followed by a sequence of residual blocks grouped into stages. Each stage increases the depth and number of filters, allowing the network to learn more and more complex hierarchical features [17]. ResNet's fault tolerance and hierarchical structure have established it as a building block for medical image analysis, with accurate feature extraction and categorization a necessity. Its scalability and performance in generalizing over many tasks further support its use in deep learning tasks [7].

The first part of this network is a convolutional layer, conv1, using a 7 × 7 kernel and a stride of 2. Then, batch normalization is applied, followed by a ReLU activation function to introduce nonlinearity into the model. The max-pooling layer, configured with a stride of 2, effectively reduces the spatial dimensions of the feature maps, thereby reducing computational costs. The architecture of this network consists of four distinctive stages (layer1, layer2, layer3, and layer4), each comprising many residual blocks. Each of these blocks consists of a convolutional layer that is integrated with batch normalization and an activation function (ReLU). Furthermore, skip connections enable the network to bypass some layers, thereby enhancing residual learning and easing the problem of flow in gradients. The spatial features are combined into one vector through the global average pooling layer. The final output layer is a fully connected one that makes use of a softmax function to classify the images into two distinct classes, either COVID or normal. ResNet101 was trained using the Adam optimizer with an initial learning rate of 0.001 to facilitate efficient updates of the weights. This model uses the cross-entropy loss function as an objective function to minimize classification errors. To prevent overfitting, dropout regularization was used in the fully connected layers of the proposed model. This model was trained up to a total of 30 epochs, and early stopping, considering the validation loss, prevents overfitting.

VGG models, developed by Simonyan and Zisserman [16], are noted for their simple architecture and effectiveness in image classification tasks [27]. The VGG models follow a consistent architectural structure, comprising multiple convolutional layers along with max pooling layers. The dimensions of convolution layers are defined as 3 × 3, whereas the filter dimensions of max-pooling layers are established at 2 × 2. The VGG architecture comprises two well-known variants: VGG11 and VGG16. As the name suggests, VGG11 comprises 11 layers (Figure 3), from which 8 are identified as convolutional layers and 3 are categorized as fully connected layers. VGG16 consists of 16 layers in total, in which 13 are convolutional layers and 3 are fully connected layers [27]. The extra layers of VGG16 enable the acquisition of more intricate information. Nonetheless, this improvement leads to heightened computational complexity [14].

Figure 2
Architecture diagram of ResNet101

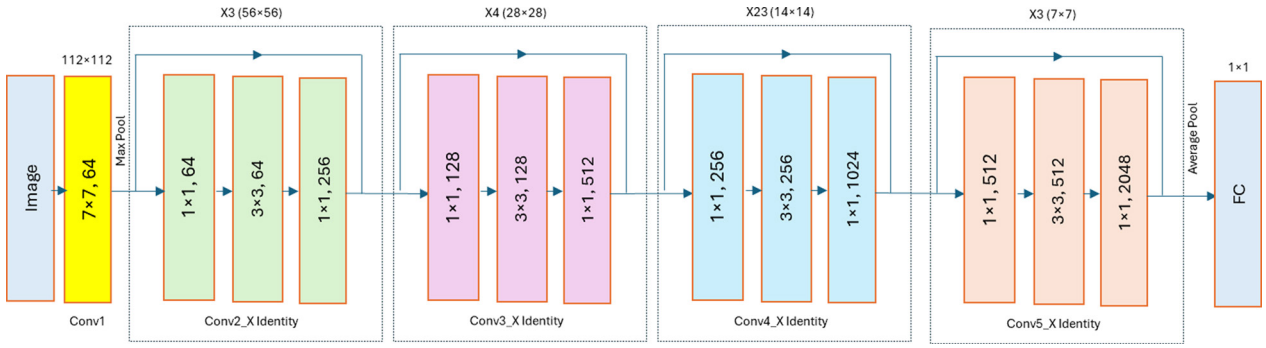


Figure 3
Architecture diagram of VGG

In the context of the ILSVRC in 2012, AlexNet (Figure 4), which Krizhevsky and his colleagues developed, emerged as a groundbreaking model [28]. The work that they did represents a significant milestone in the development of deep learning, particularly in the area of image classification. The model that was developed demonstrated the capability of deep CNNs to outperform conventional methods, thereby establishing CNNs as a foundational approach for computer vision tasks [28].

The architecture of AlexNet uses both convolutional and pooling layers to obtain detailed, hierarchical features from the images that are fed into it. Although it is significantly less complicated than ResNet and VGG or other contemporary architectures, AlexNet continues to be very effective in image classification tasks and has contributed to the development of deeper models that display a greater degree of complexity. It is a significant step toward deep learning methods, and it is important for popularizing CNNs in the process of solving large-scale visual recognition problems according to Tartaglione et al. [14] and Michalska-Ciekańska and Boyko [17].

The shifts in architecture and the model's depth are the primary discriminating features of the chosen models. ResNet50, ResNet101, and ResNet152 are more complex in comparison with VGG11, VGG16, and AlexNet. For intricate tasks such as the identification of COVID-19, the addition of layers in the ResNet models improves the potential of the model in grasping complicated features and representations. VGG11, VGG16, and AlexNet have a consistent architecture characterized by a low number of layers. Although they lack the complexity of ResNet, their performance in image classification tasks remains commendable [17]. VGG models are designed for extracting complicated features in the images using small-sized filters and higher numbers of convolutional layers. One characteristic feature of ResNet is the utilization of residual links. These links boost the flow of information and the gradient, resulting in the ease of deep network training [4]. ResNet models are thus suitable for deep learning tasks. Overall, the ResNet50, ResNet101, ResNet152, VGG11, VGG16, and AlexNet models are effective in the identification of COVID-19 based on chest X-ray examination images. Some features such as the level of task complexity, the processing potential, and the trade-off between model complexity and the level of accuracy are considered in the selection of the most suitable model [15].

4. Experimental Results and Analysis

The chest X-ray dataset provided on Kaggle was used in this study. This dataset was specially prepared for COVID-19 detection, with various X-ray images categorized as either "normal" or "COVID" (Figures 5 and 6). "Normal" images were taken from people who were free of COVID-19 infection, whereas "COVID" images were images from infected patients. Each class consists of 3,616 images, thus being balanced and ready for analysis. Splitting the dataset into subsets such

Figure 4
Architecture diagram of AlexNet

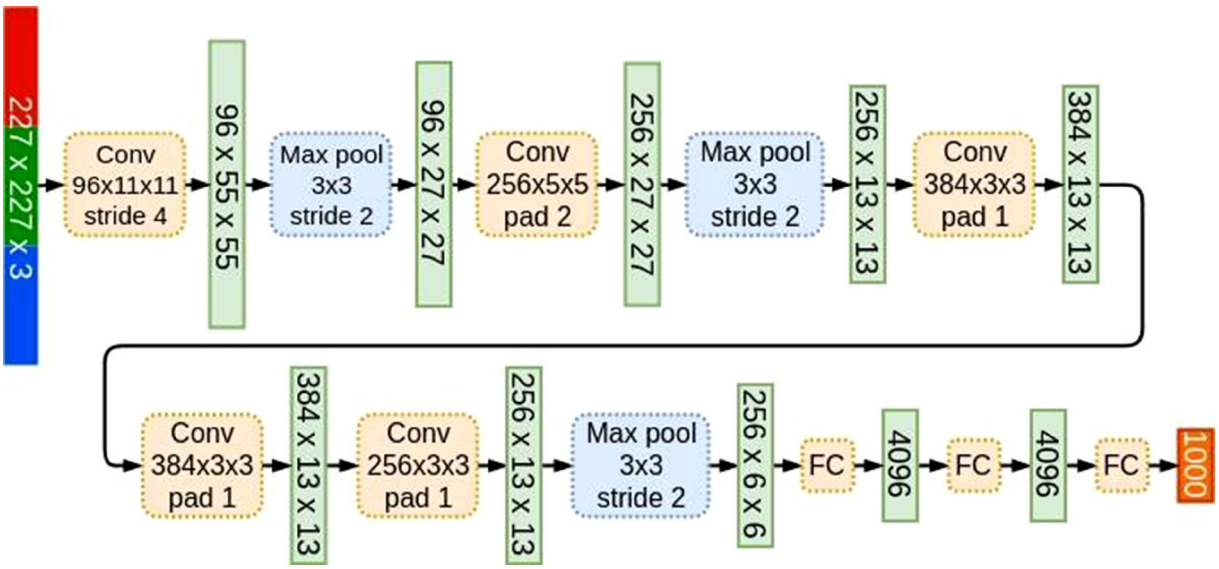


Figure 5
Chest X-ray images from COVID-19 patients

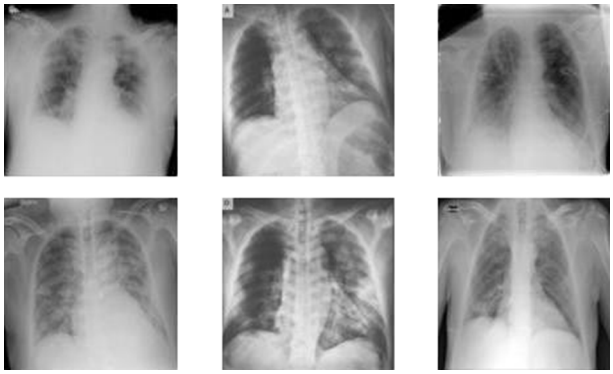
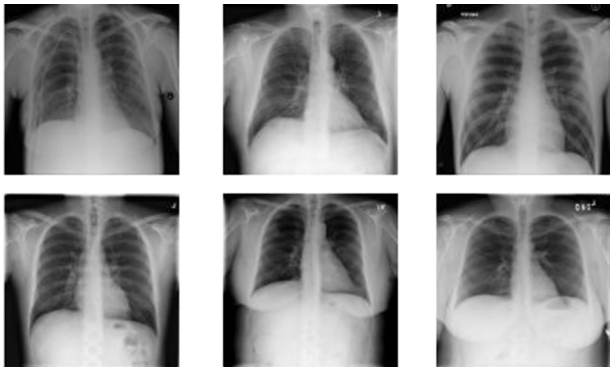


Figure 6
Chest X-ray images from normal patients



as training, testing, validation, and train–validation makes it easier to train and test models. Deep learning models for COVID-19 detection can be developed, evaluated, and generalized using this structured split.

4.1. Dataset and evaluation metrics

Preprocessing is a sequential process that obtains the data ready for training and testing, ensuring that the input images are appropriately formatted to deep learning models.

Data Loading: The dataset will be loaded using the "torchvision.datasets" library, specifically the "ImageFolder" class. This automatically organizes images into their respective folders based on class labels. In this step, necessary image transformations are applied to ensure uniformity and compatibility with the model input requirements.

Data Splitting: After loading the data, images are split into train, validation, and test subsets. This is accomplished by utilizing the "torch.utils.data" library, which systematically partitions the dataset into subsets such as train, test, train–valid, and valid. Then, these subsets are managed using the "DataLoader" class, which will enable easy batch loading, shuffling, and processing. A batch size of 32 is used to minimize computation during the training phase.

Data Normalization: Normalization of the images' pixel values is performed through "torchvision.transforms." This step normalizes the input data to have a mean value of zero, which allows the model to converge faster during training. The mean and standard deviation values for normalization—precomputed on the training dataset—are [0.485, 0.456, 0.406] and [0.229, 0.224, 0.225], respectively. These represent the three RGB channels for the images.

Data Augmentation: Various augmentation techniques have been used to increase the quantity and diversity of the training data. Data augmentation is used to enhance the generalization capability of the model on unseen data [17]. All images were resized to a uniform dimension of 32 × 32 pixels, converted to tensor format, and normalized using values consistent with those in the training set. Augmentation techniques ensure that the model learns from slightly varied versions of the original data, thereby reducing overfitting and improving performance.

By implementing these preprocessing steps, the dataset is systematically organized, standardized, and enhanced, ensuring that it is well suited for the effective training and evaluation of deep learning models.

We have analyzed performance using various performance metrics. The confusion matrix provides several performance metrics to evaluate the model. Accuracy (eq. 1) measures model prediction accuracy. It gives a simple correct prediction-to-total prediction ratio. Precision (eq. 2) is the ratio of correctly predicted positive cases to model-labeled positive cases. Recall or sensitivity (eq. 3) measures how well the model identifies positive cases; it is especially useful when missing positives is costly. On the contrary, specificity is concerned with the proportion of true negative cases that are accurately identified, indicating the efficiency of the model in steering clear of false positives. Finally, F1-score (eq. 4) is a balanced metric that combines precision and recall through their harmonic mean. It becomes particularly useful when a model's performance must be evaluated in situations where positive and negative cases are highly imbalance. These metrics together would give a full understanding of the strengths and weaknesses of the classifier [9, 17].

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{FN}+\text{TN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{3}$$

$$\text{F1 Score} = \frac{2}{\frac{1}{\text{Precision}}+\frac{1}{\text{Recall}}} \tag{4}$$

Confusion Matrix: The confusion matrix is a fundamental tool for assessing the performance of classification models, offering a detailed breakdown of how well predictions align with actual outcomes. It categorizes results into four key groups: true positives (TP), which reflect correctly identified positive cases; false positives (FP), where negative cases are incorrectly labeled as positive; false negatives (FN), which occur when positive cases are missed; and true negatives (TN), representing accurate identification of negative cases. This framework is particularly useful in evaluating models analyzing chest X-ray and CT scan images [9, 29].

AUC–ROC Curve: The area under the receiver operating characteristic curve (AUC–ROC) serves as a visual representation of a model's effectiveness in distinguishing the positive class, such as COVID. It maps the relationship between recall and the false positive rate (FPR), offering insight into the trade-offs at various threshold settings. As noted by Jin et al., the area under this curve, which ranges from 0 to 1, reflects the classifier's overall quality—with higher values indicating stronger performance [9, 29, 30].

Precision–Recall Curve: The precision–recall (PR) curve illustrates the interaction between precision and recall across different thresholds set by the classifier. The ideal model is represented by the point (1,1), and strong-performing models generate curves that approach this optimal intersection. This metric becomes especially useful in scenarios involving class imbalance, such as differentiating between COVID and normal cases. The PR curve's focus on the minority class makes it particularly effective for evaluating imbalanced binary classifiers [9, 29].

All selected models were trained using the Adam optimizer with a learning rate of 0.001 for 30 epochs. To improve generalization and mitigate overfitting, the training incorporated both data augmentation and dropout regularization techniques. A brief overview of the models is provided below:

1) AlexNet: A relatively shallow CNN with core convolutional and pooling layers, followed by fully connected layers. Although computationally lightweight, its accuracy tends to lag behind deeper models [29].
2) ResNet Variants (ResNet50, ResNet101, and ResNet152): These architectures utilize residual connections to overcome vanishing gradient issues, enabling effective training of very deep networks. Their primary distinction lies in their depth, with 50, 101, and 152 layers, respectively [29].
3) VGG11 and VGG16: Known for their uniform design, these models stack convolutional layers with small filters, followed by pooling and batch normalization. Although computationally intensive, they are effective in hierarchical feature extraction and are widely used in image classification tasks [29].

4.2. Experiments on disease detection

The training progress across the six models is illustrated in Figures 7–12. The training and validation performances of the models, as observed through loss and accuracy trends, provide insights into their learning behavior. ResNet50 showed a steady decrease in loss, with validation accuracy improving consistently over epochs, achieving a final accuracy of 91%. The minimal gap between training and validation losses indicates effective generalization. ResNet101 displayed a similar behavior, with training and validation losses decreasing smoothly, leading to the highest accuracy (93%) and strong convergence without overfitting. ResNet152 achieved stable learning but exhibited a slightly larger training-validation loss gap, reflecting some tendency toward overfitting. Nonetheless, it achieved a recall of 94.77%, excelling at identifying COVID-positive cases. VGG11 and VGG16 demonstrated fluctuations in training loss, particularly during early epochs. VGG16 achieved 84% accuracy, outperforming VGG11's 82%, although both struggled with higher false positives. AlexNet showed slower learning, with its validation accuracy plateauing at 74%, indicating limitations in capturing complex features in medical images.

AUC–ROC (Figures 13–18) and precision–recall (PR) curves (Figures 19–24) further illustrate the models' performance. ResNet101 and ResNet50 achieved the highest AUC scores of 97.41% and 97.59%, respectively. These values demonstrate their exceptional ability to discriminate between COVID and normal cases, even in imbalanced datasets. ResNet152 followed with an AUC of 96.17%, excelling in sensitivity (high recall). VGG16 recorded a moderate AUC of 92.45%, reflecting its reasonable balance between false positives and false negatives. AlexNet, with an AUC of 86.6%, displayed the weakest discrimination ability, consistent with its overall lower performance.

Figure 7
Performance of VGG11 during training

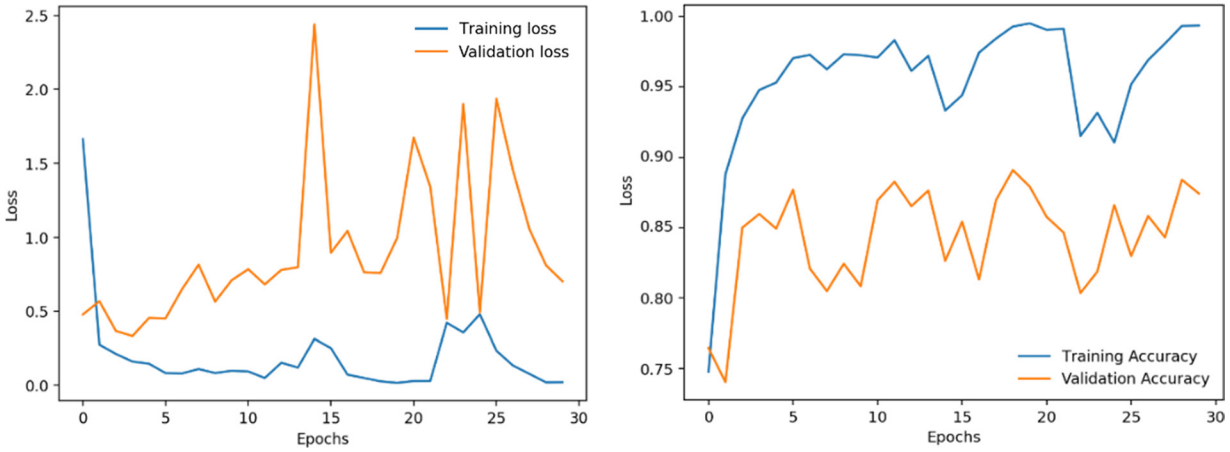


Figure 8
Performance of VGG16 during training

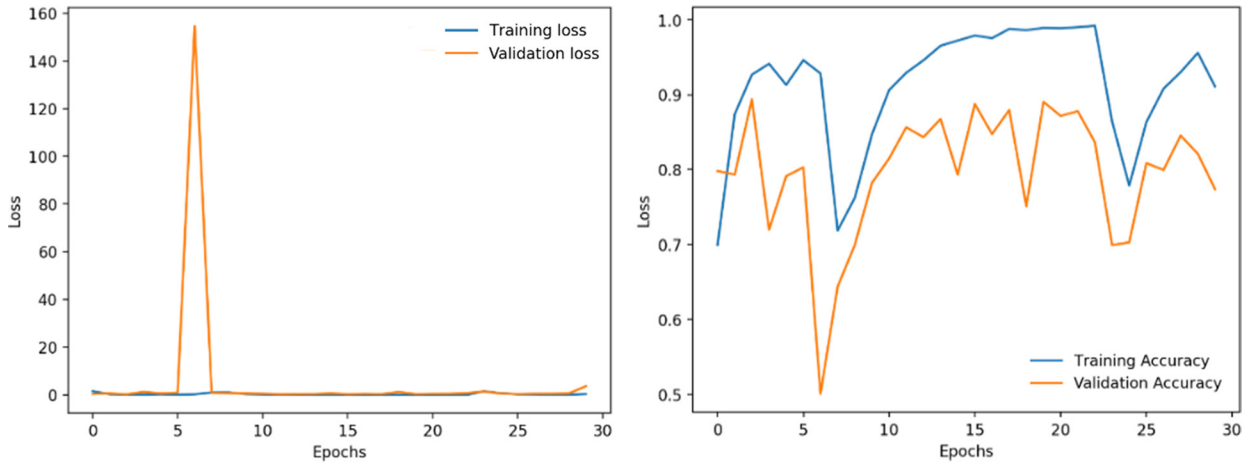


Figure 9
Performance of ResNet50 during training

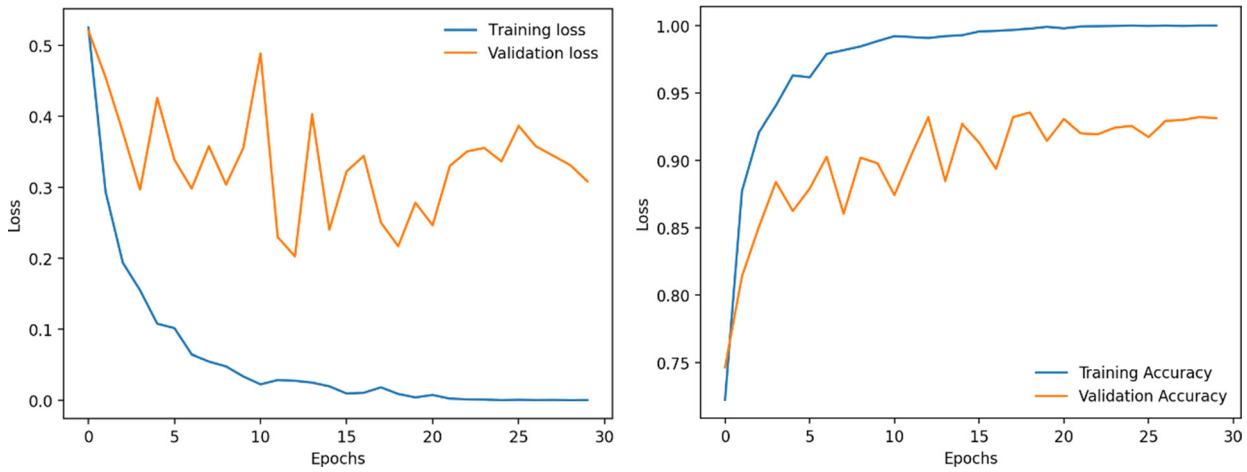


Figure 10
Performance of ResNet101 during training

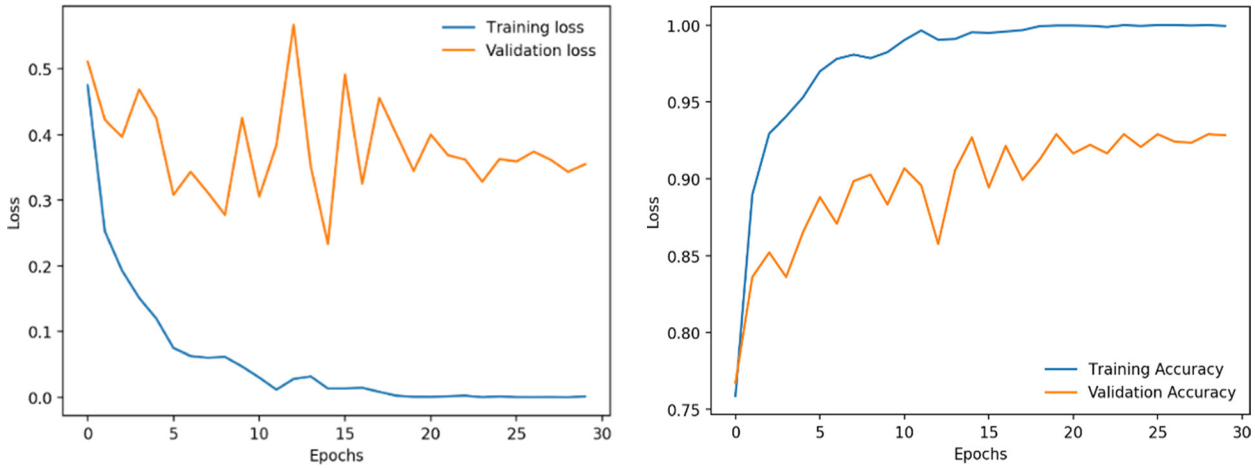


Figure 11
Performance of ResNet152 during training

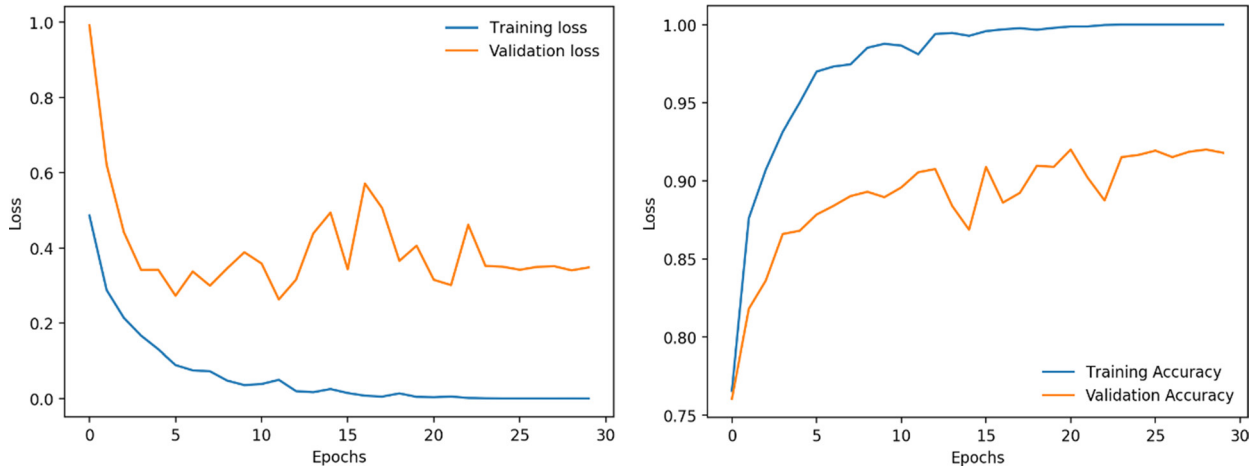


Figure 12
Performance of AlexNet during training

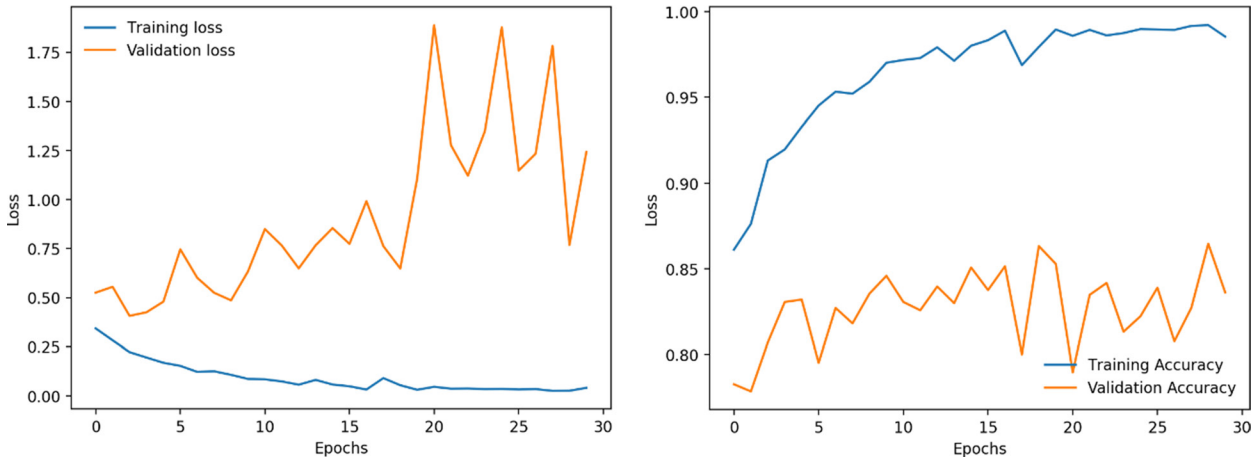


Figure 13
ResNet50

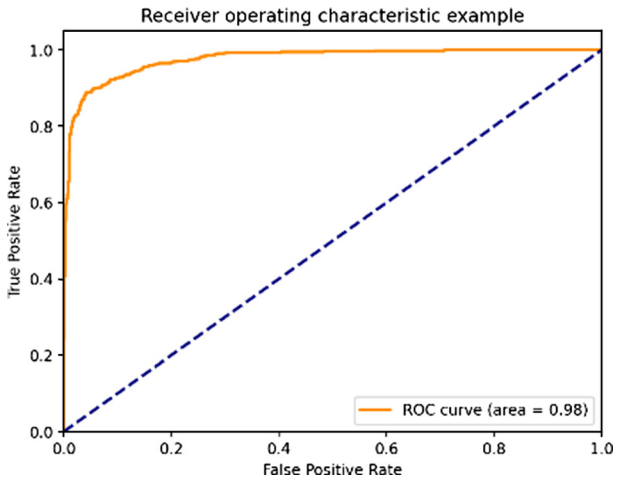


Figure 14
ResNet101

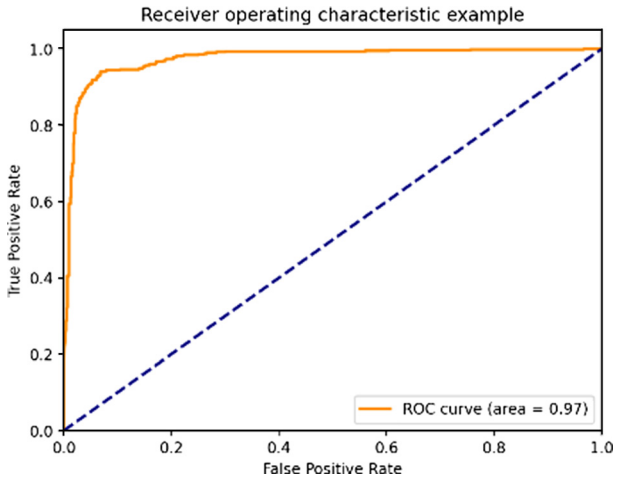


Figure 15
ResNet152

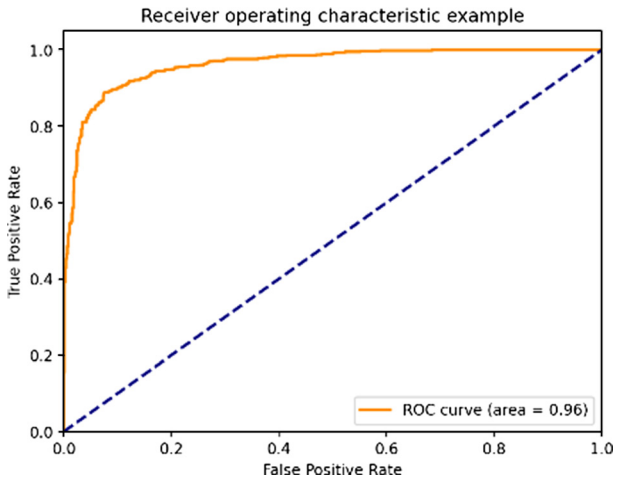


Figure 16
AlexNet

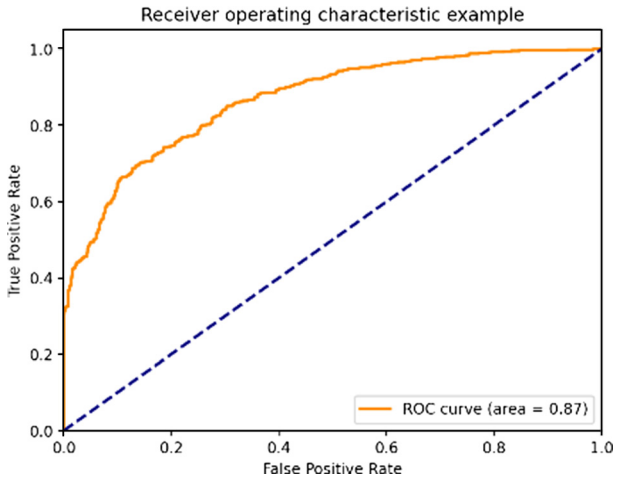


Figure 17
VGG11

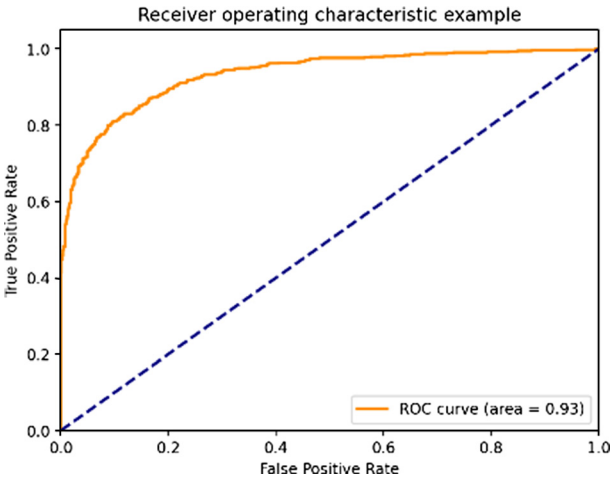


Figure 18
VGG16

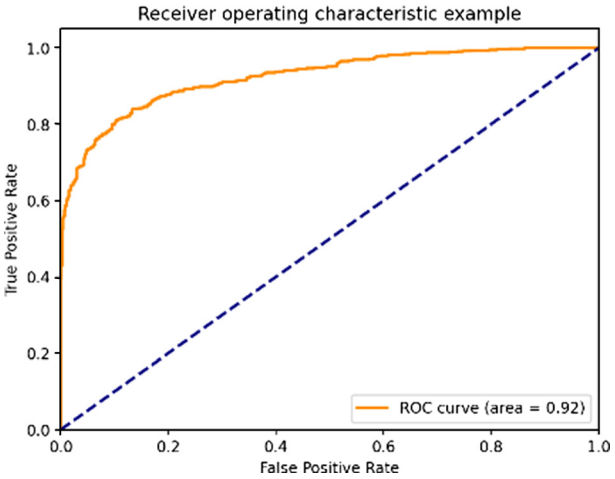


Figure 19
PR curve—ResNet50

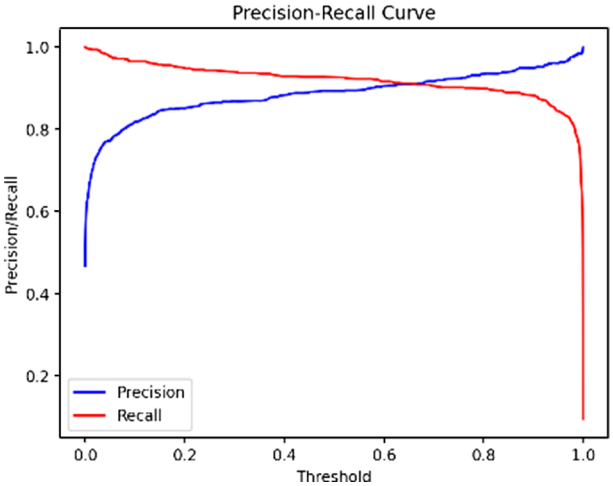


Figure 20
PR curve—ResNet101

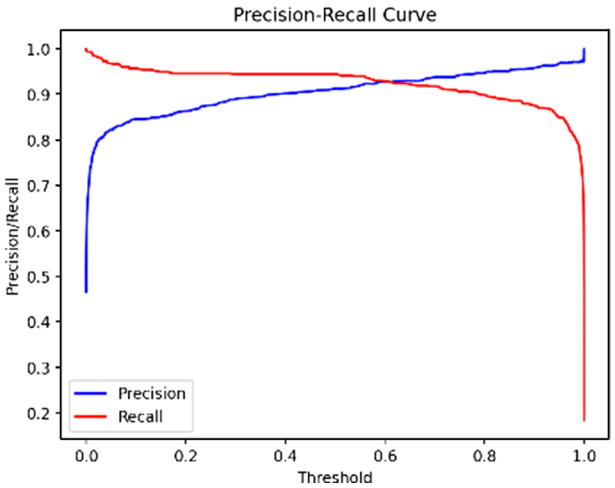


Figure 21
PR curve—ResNet152

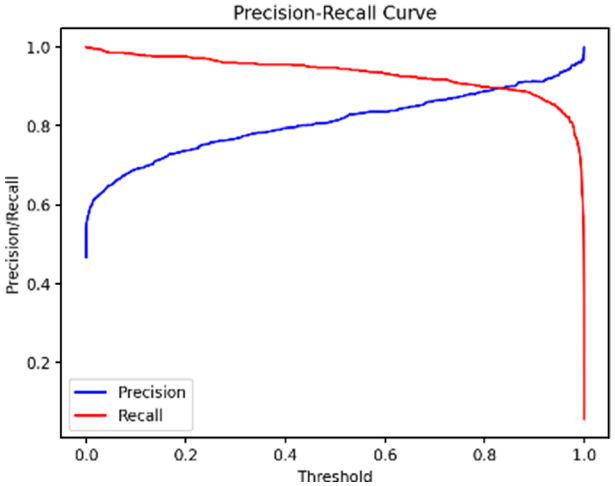


Figure 22
PR curve—AlexNet

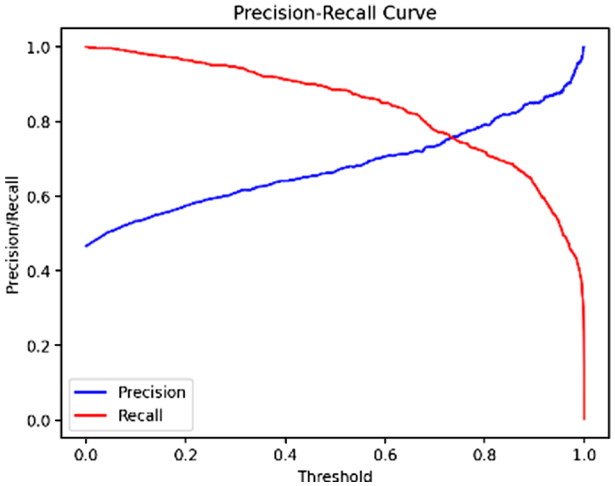


Figure 23
PR curve—VGG11

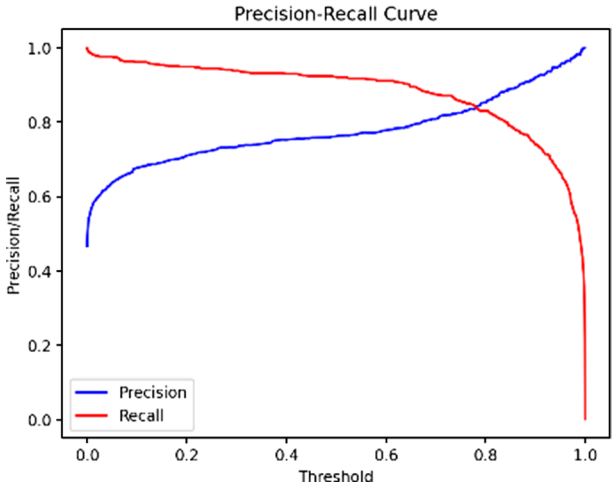
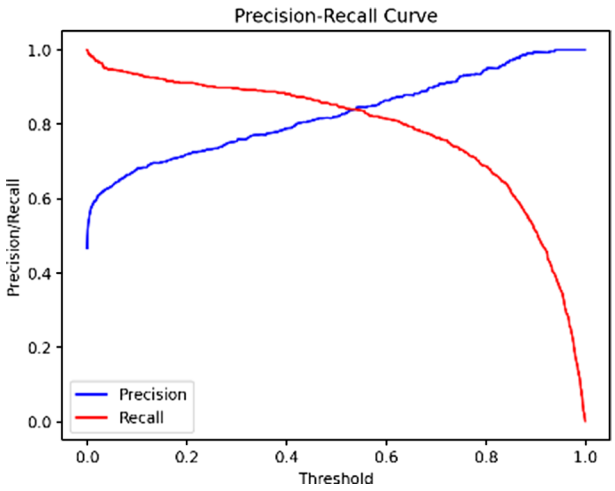


Figure 24
PR curve—VGG16



The PR curves reveal similar trends, where ResNet101 and ResNet50 maintained high precision and recall across all thresholds, confirming their robustness.

The confusion matrices for all six models (Figures 25–30) offer a comprehensive view of how each model performs in classifying COVID-positive versus normal cases, shedding light on their respective strengths and shortcomings. Among them, ResNet101 stood out, correctly identifying 664 COVID cases (true positives) and 596 normal cases (true negatives). It kept misclassifications relatively low, with 57 false positives, normal cases incorrectly labeled as COVID, and just 35 false negatives. This balanced accuracy helped the model achieve high precision (91.27%) and an impressive recall (94.45%), suggesting it handles both types of errors well.

ResNet50 followed closely, albeit with slightly reduced performance. It correctly classified 651 COVID cases and 584 normal ones, although it made 70 false positive and 47 false negative errors.

Figure 25
ResNet50

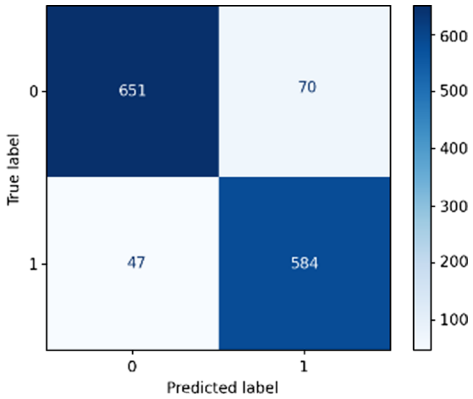


Figure 26
ResNet101

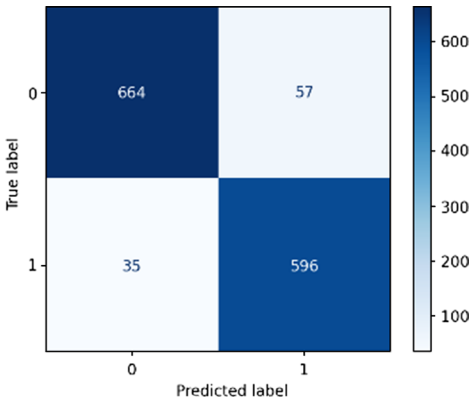


Figure 27
ResNet152

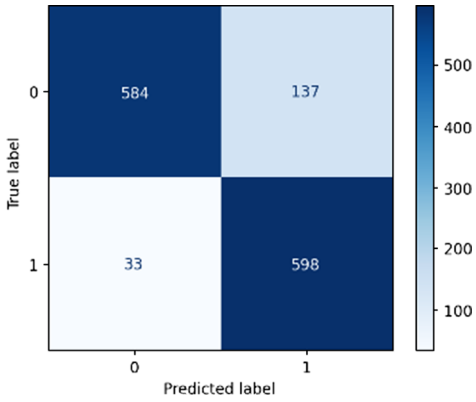


Figure 28
AlexNet

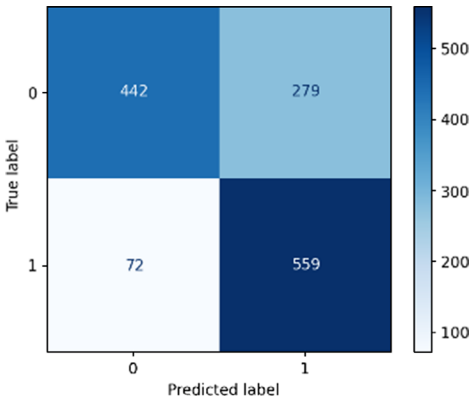


Figure 29
VGG11

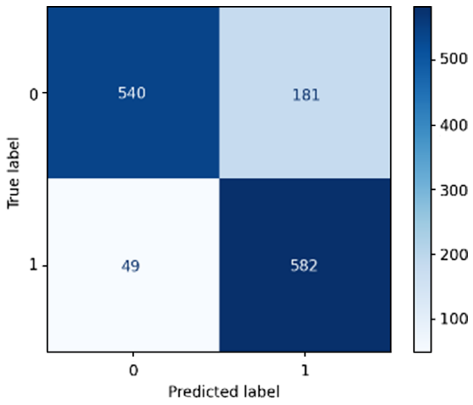
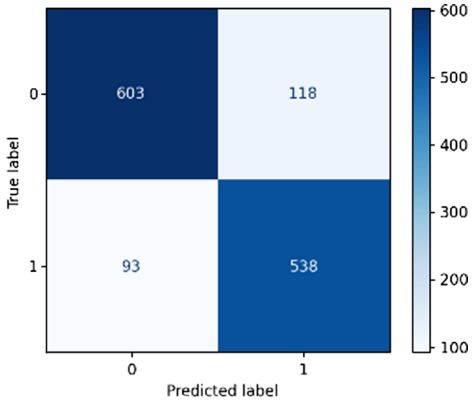


Figure 30
VGG16



Although this reduced its precision a bit, it still maintained solid recall. Conversely, ResNet152 excelled in sensitivity, reaching the highest recall (94.77%) with 584 true positives and 598 true negatives. However, its 137 false positives pulled its precision down to 81.36%, which could be a drawback in clinical settings where false alarms are costly.

Turning to the VGG models, VGG16 achieved 603 true positives and 538 true negatives. Still, it faced challenges, particularly with 118 false positives and 93 false negatives, leading to only moderate precision and recall. VGG11 showed a slightly different error pattern. Despite having 540 true positives and 582 true negatives, it struggled with a high count of 181 false positives, compared to 49 false negatives. This indicates that it had a harder time correctly identifying normal cases.

AlexNet, unfortunately, delivered the weakest performance. With just 442 true positives and 559 true negatives, it produced a substantial number of errors, 279 false positives and 72 false negatives. The result was a lower precision (66.7%) and an accuracy of 74%, pointing to clear limitations in distinguishing between the two classes, especially in minimizing false alarms.

In summary, this confusion matrix analysis clearly highlights ResNet101 as the most dependable model. Its strong balance between precision and recall makes it particularly well suited for medical use, where both false positives and false negatives carry serious implications. By contrast, models such as AlexNet and VGG11 showed notable weaknesses, especially in misclassifying normal cases as COVID, which could pose risks in real clinical environments. On the basis of these findings, ResNet101 emerges as the recommended choice.

The comprehensive evaluation of the models is summarized in Table 1. ResNet101 emerged as the best-performing model, achieving the highest accuracy (93%), precision (91.27%), and F1-score (92.83%) while maintaining a strong AUC of 97.41%. Therefore, the proposed model based on this study is ResNet101.

The key observations are the following:

1) Precision: Critical in healthcare applications to minimize false positives. ResNet101 demonstrated the highest precision, reducing the likelihood of misclassifying normal cases as COVID-positive cases.
2) Recall: ResNet152 achieved the highest recall (94.77%), making it ideal in scenarios where sensitivity to COVID-positive detection is prioritized.
3) F1-Score: ResNet101 and VGG16 balanced precision and recall effectively, with F1-scores of 92.83% and 84.35%, respectively.

Despite achieving a moderate AUC, AlexNet's overall accuracy and precision were significantly lower, highlighting its limitations. The results clearly indicate that ResNet101 is the most effective model for COVID-19 detection from chest radiographs, offering a strong balance of precision, recall, and accuracy. Its ability to minimize false positives and false negatives makes it a reliable choice for real-world clinical applications. ResNet50 also performed exceptionally well, whereas ResNet152 excelled in recall, making it suitable for cases where sensitivity is prioritized. Conversely, VGG models showed moderate performance with some issues in false positives, whereas AlexNet exhibited the weakest results, indicating the need for further optimization.

4.3. Research limitations

This study employed a single publicly available dataset for model comparison. Although the dataset in this study consists of a big set of COVID-19 positive and negative chest radiographs, but with real-time medical information, which contains several errors and missed information, it can be a good option while comparing the performance of various models to eliminate bias and generalizability issues.

Although the models show good accuracy and performance metrics, they have not yet undergone clinical validation. Without testing in real-world healthcare or with the involvement of medical professionals, the clinical usefulness of the models cannot be assured.

During the use of medical-related data, one should consider ethical issues. Although this study relies on publically available data, in the event of an expansion of the work for clinical data, ethical issues relating to AI, including the issues of data privacy, patient permission, and bias in algorithms, need to be considered.

Although this study concentrated on concrete architectures such as ResNet, VGG, and AlexNet, a general assessment of other state-of-the-art architectures may yield more in-depth insights and indicate the most effective methods for COVID-19 detection.

Table 1
Models' performance summary

Model	Overall Accuracy	Precision	Recall	F1-Score	AUC
ResNet101	93%	91.27	94.45	92.83	97.41
ResNet50	91%	89.29	92.55	90.89	97.59
ResNet152	87%	81.36	94.77	87.55	96.17
VGG16	84%	84.32	84.44	84.35	92.45
VGG11	82%	83.97	83.56	82.97	82.97
AlexNet	74%	66.70	88.58	76.10	86.60

This study highlights the need for the inclusion of AI technologies in healthcare systems but suggests no concrete solutions. Pragmatic guidelines and frameworks for the inclusion of AI and ML models in healthcare systems should be developed in follow-up studies. Technical issues have to be addressed with a view to infrastructure compatibility, and healthcare professionals need to be trained and equipped.

5. Conclusion and Future Work

This paper discusses the use of AI and ML methods, specifically deep learning models, in COVID-19 identification based on X-ray chest images. This paper discusses existing literature on the subject and evaluates the performance of deep learning models, such as ResNet, VGG, and AlexNet, in achieving better accuracy and efficiency in comparison with traditional diagnostic techniques for COVID-19 specifically. The experimental outcomes depict the performance of the models in terms of various metrics. ResNet101 showed improved effectiveness in the correct identification of COVID and normal cases in comparison with the alternative models, achieving the highest overall accuracy rate of 93% and maximum precision of 91.27%, evidencing a lowered rate of false positives. ResNet101 presents a capable detection rate for positive COVID patients, with a 94.45% rate of recall. F1-score of 92.83%, measuring precision and recall, reflects a positive balance in lowering false positives and negatives. ResNet101 showed effective discrimination, with 97.41% as the obtained AUC. ResNet101's ability in classifying COVID-19 and normal classes precisely places it as the top model for the assessment of the chosen X-ray images relevant for COVID-19. Deep learning models with advanced architectures can be designed based on hybrid architectures, transformers with attention, and graph neural networks. These methods can further improve diagnostic accuracy in encoding complex medical image patterns. The use of ensemble methods may lower prediction error by exploiting model architectures. For the enhancement of healthcare application specificity, the transfer learning can be pretrained in domain-specific medical imaging sets, such as MRIs or CT scans, during fine-tuning based on full databases. Multimodal integrated data may boost model generalizability. Chest X-rays and CT scans can be combined with electronic health records, lab test results, and patient histories to improve diagnosis. This multimodel will enable the AI models to discover deeper correlations and make more accurate predictions. Clinical validation in a real-world environment also constitutes a crucial aspect. AI model testing in clinics must be emphasized with healthcare professionals in subsequent studies. Clinical trials at a high scale will ensure that such models are accurate, reliable, and usable for everyday medical practice. Ethics are also in high demand for instilling confidence and fairness into AI-based healthcare solutions. Future studies must prioritize patient data privacy, informed consent, and reduction of algorithmic bias for such models to produce equal outputs for varied sets of populations and imaging devices. Explainable AI or XAI enhances model transparency in such a manner that clinicians can comprehend their decision-making and have confidence in it. Future studies should utilize advanced data augmentation using generative adversarial networks to eliminate the scarcity issue of the datasets and enhance the model's robustness. Improved precision, ethics, scalability, and integration of the AI diagnostic solutions into the global healthcare infrastructure will be a result of further studies.

Ethical Statement

This study did not require formal ethical approval because it used publicly available data from an open-access dataset on Kaggle, which does not contain identifiable patient information. This exemption is based on the fact that no direct human participants were involved and the data are fully anonymized and openly accessible (Covid19 Detection dataset, https://www.kaggle.com/datasets/donjon00/covid19-detection?select=Data). The dataset has been appropriately cited in this manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/donjon00/covid19-detection?select=Data.

Author Contribution Statement

Bushra Hassan: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Kaveh Kiani:** Conceptualization, Methodology, Investigation, Resources, Writing – review & editing, Visualization, Supervision, Project administration. **Taha Mansouri:** Conceptualization, Methodology, Investigation, Resources, Supervision.

References

[1] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. https://doi.org/10.1109/CVPR.2016.90

[2] Maghded, H. S., Ghafoor, K. Z., Sadiq, A. S., Curran, K., Rawat, D. B., & Rabie, K. (2020). A novel AI-enabled framework to diagnose coronavirus COVID-19 using smartphone embedded sensors: design study. In *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science*, 180–187. https://doi.org/10.1109/IRI49571.2020.00033

[3] Toğaçar, M., Ergen, B., & Cömert, Z. (2020). COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches. *Computers in Biology and Medicine*, *121*, 103805. https://doi.org/10.1016/j.compbiomed.2020.103805

[4] Tao, S., Guo, Y., Zhu, C., Chen, H., Zhang, Y., Yang, J., & Liu, J. (2019). Highly efficient follicular segmentation in thyroid cytopathological whole slide image. In *International Workshop on Health Intelligence*, 149–157. https://doi.org/10.1007/978-3-030-24409-5_14

[5] Wang, S., Zha, Y., Li, W., Wu, Q., Li, X., Niu, M., …, & Tian, J. (2020). A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *European Respiratory Journal*, *56*(2), 2000775. https://doi.org/10.1183/13993003.00775-2020

[6] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *25*.

[7] Wang, L., Lin, Z. Q., & Wong, A. (2020). COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Scientific Reports*, *10*(1), 19549. https://doi.org/10.1038/s41598-020-76550-z

[8] Apostolopoulos, I. D., & Mpesiana, T. A. (2020). COVID-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, *43*, 635–640. https://doi.org/10.1007/s13246-020-00865-4

[9] Wang, B., Jin, S., Yan, Q., Xu, H., Luo, C., Wei, L., …, & Dong, J. (2020). AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system. *Applied Soft Computing*, *98*, 106897. https://doi.org/10.1016/j.asoc.2020.106897

[10] JavadiMoghaddam, S., & Gholamalinejad, H. (2021). A novel deep learning based method for COVID-19 detection from CT image. *Biomedical Signal Processing and Control*, *70*, 102987. https://doi.org/10.1016/j.bspc.2021.102987

[11] Shames, M. A., & Kamil, M. Y. (2025). Lung infection detection via CT images and transfer learning techniques in deep learning. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *47*(1), 206–218. https://doi.org/10.37934/araset.47.1.206218

[12] Marefat, A., Marefat, M., Hassannataj Joloudari, J., Nematollahi, M. A., & Lashgari, R. (2023). CCTCOVID: COVID-19 detection from chest X-ray images using Compact Convolutional Transformers. *Frontiers in Public Health*, *11*, 1025746. https://doi.org/10.3389/fpubh.2023.1025746

[13] Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., …, & Peng, Z. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA*, *323*(11), 1061–1069. http://doi.org/10.1001/jama.2020.1585

[14] Tartaglione, E., Barbano, C. A., Berzovini, C., Calandri, M., & Grangetto, M. (2020). Unveiling COVID-19 from chest X-ray with deep learning: A hurdles race with small data. *International Journal of Environmental Research and Public Health*, *17*(18), 6933. https://doi.org/10.3390/ijerph17186933

[15] Zheng, C., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H., …, & Wang, X. (2020). Deep learning-based detection for COVID-19 from chest CT using weak label. *IEEE Transactions on Medical Imaging*, *39*(8), 2615–2625. https://doi.org/10.1109/TMI.2020.2995965

[16] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv Preprint: 1409.1556*. https://doi.org/10.48550/arXiv.1409.1556

[17] Michalska-Ciekańska, M., & Boyko, O. (2023). COVID-19 diagnosis using deep learning from X-ray and CT images – Overview. *Advances in Cyber-Physical Systems*, *8*(2), 126–132. https://doi.org/10.23939/acps2023.02.126

[18] Khan, A. I., Shah, J. L., & Bhat, M. M. (2020). CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest X-ray images. *Computer Methods and Programs in Biomedicine*, *196*, 105581. https://doi.org/10.1016/j.cmpb.2020.105581

[19] Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., …, & Xia, J. (2020). Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy. *Radiology*, *296*(2), E65–E71. https://doi.org/10.1148/radiol.2020200905

[20] Elaziz, M. A., Hosny, K. M., Salah, A., Darwish, M. M., Lu, S., & Sahlol, A. T. (2020). New machine learning method for image-based diagnosis of COVID-19. *Plos One*, *15*(6), e0235187. https://doi.org/10.1371/journal.pone.0235187

[21] Aslani, S., & Jacob, J. (2023). Utilisation of deep learning for COVID-19 diagnosis. *Clinical Radiology*, *78*(2), 150–157. https://doi.org/10.1016/j.crad.2022.11.006

[22] Abdulahi, A. T., Ogundokun, R. O., Adenike, A. R., Shah, M. A., & Ahmed, Y. K. (2024). PulmoNet: A novel deep learning based pulmonary diseases detection model. *BMC Medical Imaging*, *24*(1), 51. https://doi.org/10.1186/s12880-024-01227-2

[23] Chauhan, S., Edla, D. R., Boddu, V., Rao, M. J., Cheruku, R., Nayak, S. R., …, & Nigat, T. D. (2024). Detection of COVID-19 using edge devices by a light-weight convolutional neural network from chest X-ray images. *BMC Medical Imaging*, *24*(1), 1. https://doi.org/10.1186/s12880-023-01155-7

[24] Kaur, B. P., Singh, H., Hans, R., Sharma, S. K., Kaushal, C., Hassan, M. M., & Shah, M. A. (2024). An augmentation aided concise CNN based architecture for COVID-19 diagnosis in real time. *Scientific Reports*, *14*(1), 1136. https://doi.org/10.1038/s41598-024-51317-y

[25] Yin, M., Xu, C., Zhu, J., Xue, Y., Zhou, Y., He, Y., …, & Fu, C. (2024). Automated machine learning for the identification of asymptomatic COVID-19 carriers based on chest CT images. *BMC Medical Imaging*, *24*(1), 50. https://doi.org/10.1186/s12880-024-01211-w

[26] Okada, N., Umemura, Y., Shi, S., Inoue, S., Honda, S., Matsuzawa, Y., …, & Fujimi, S. (2024). "KAIZEN" method realizing implementation of deep-learning models for COVID-19 CT diagnosis in real world hospitals. *Scientific Reports*, *14*(1), 1672. https://doi.org/10.1038/s41598-024-52135-y

[27] Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., …, & Shen, D. (2020). Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. *IEEE Reviews in Biomedical Engineering*, *14*, 4–15. https://doi.org/10.1109/RBME.2020.2987975

[28] Shadeed, G. A., Tawfeeq, M. A., & Mahmoud, S. M. (2020). Deep learning model for thorax diseases detection. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *18*(1), 441–449. http://doi.org/10.12928/telkomnika.v18i1.12997

[29] Das, S., Ayus, I., & Gupta, D. (2023). A comprehensive review of COVID-19 detection with machine learning and deep learning techniques. *Health and Technology*, *13*(4), 679–692. https://doi.org/10.1007/s12553-023-00757-z

[30] Huang, J., & Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, *17*(3), 299–310. https://doi.org/10.1109/TKDE.2005.50

How to Cite: Hassan, B., Kiani, K., & Mansouri T. (2025). An Effective ResNet Model for Respiratory Disease Detection: A Case Study on COVID-19 Chest X-ray Images. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA52024885