Diyala Journal of Engineering Sciences

Journal homepage: https://djes.info/index.php/djes

ISSN: 1999-8716 (Print); 2616-6909 (Online)

COVIXNet: A Robust and Explainable Deep Transfer Learning Framework for COVID-19 Detection from Chest X-ray Imagery

Hadj Zerrouki[*], Salima Azzaz-Rahmani

Department of Telecommunications, Faculty of Electrical Engineering, Djillali Liabes University of Sidi Bel Abbes, Algeria

ARTICLE INFO

Article history:
Received: 08/10/2025.
Revised: 09/12/2025
Accepted: 11/02/2026
Available online: 15/03/2026

Keywords:
COVID-19 detection
Deep learning
Chest X-ray
Computer-aided diagnosis
Explainable AI (XAI)
Healthcare systems

ABSTRACT

The COVID-19 pandemic has highlighted the urgent need for fast, accurate, and trustworthy diagnostic tools to complement routine testing procedures. While Chest X-ray (CXR) imaging is a valuable alternative, it requires specialized interpretation that is prone to subjectivity, creating a bottleneck in clinical workflows. This study introduces COVIXNet, a robust and explainable deep transfer learning framework designed to automate the detection of COVID-19 from CXR images with high accuracy and provide clinically interpretable justifications for its predictions. COVIXNet is built upon a DenseNet-121 backbone, enhanced with a spatial attention mechanism to focus on diagnostically significant lung regions. Explainability is achieved using Gradient-weighted Class Activation Mapping (Grad-CAM). The model was trained and evaluated on COVIDXSet, a curated multi-source dataset of 8,591 CXR images. A two-phase transfer learning strategy was employed for effective feature adaptation. COVIXNet demonstrated state-of-the-art performance, achieving an accuracy of 96.8% (95% CI: 95.9% - 97.7%), a precision of 97.2%, and an AUC-ROC of 0.993 (95% CI: 0.989 - 0.997). The explainability module generated clinically meaningful heatmaps, with 92% rated as relevant by expert radiologists and achieving a high Pointing Game Accuracy of 81.5%. The model also showed consistent performance across demographic subgroups and varying image quality. COVIXNet offers a powerful combination of high diagnostic accuracy, robustness, and validated explainability. With a 48 ms inference time and 33.2 MB model size, COVIXNet is a promising and efficient tool for deployment in clinical settings to assist healthcare professionals in the rapid triage and diagnosis of COVID-19.

1. INTRODUCTION

The pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has precipitated an unprecedented global health crisis with over 212 countries and territories experiencing over a million confirmed cases and deaths [1]. Accurate and timely diagnosis of COVID-19 remains at the heart of appropriate patient care, containment of infection, and bed distribution within health systems [2]. While detection with reverse transcription-polymerase chain reaction (RT-PCR) testing for COVID-19 remains the gold standard, it is also prone to critical limitations, including false-negative reporting, delayed turnaround, and scarcity of reagents [3, 4].

Chest radiography as an ancillary diagnostic tool can rapidly re-evaluate pulmonary evidence of COVID-19 [5]. CXRs have various advantages including worldwide availability, cost-effectiveness, and lower radiation exposure compared to CT scans [6]. Recent literature has demonstrated that CXRs can illustrate unique patterns such as ground-glass opacities, consolidations, and bilateral infiltrates in COVID-19 patients [7, 8].

Deep learning, and in particular convolutional neural networks (CNNs), has revolutionized medical image analysis by achieving human experts' performance in various tasks of diagnosis [9, 10]. Detection of COVID-19 from CXRs using deep learning has also drawn significant interest, with various studies presenting promising results [11, 12]. However, key issues in the existing work persist in this field, such as insufficient robustness when applied to diverse multi-source datasets, and their 'black-box' nature of models, which hinders clinical trust and adoption [13, 14].

Transfer learning has proven a valuable method for overcoming the scarcity of medical imaging datasets by transferring knowledge from models that have been trained on large-scale natural image databases [15]. The method has proven especially successful for medical imaging applications where annotated

[*] Corresponding author's E-mail: zerrouki.hadj@gmail.com | ORCID: 0000-0003-1349-708X

DOI: 10.24237/djes.2026.19105

databases are few [16]. In addition, explainable AI technologies have found increasing relevance within clinical practice for establishing confidence in AI-assisted systems for diagnosis [17].

The proposed COVIXNet approach overcomes this by integrating an attention mechanism for improved accuracy and a Grad-CAM-based explainability framework that has been both qualitatively and quantitatively validated.

Despite recent advances in deep learning for COVID-19 detection from CXRs, several research gaps remain:

- Many models have insufficient robustness and weak generalization capacity on unseen populations and data sets.
- Limited effort have been devoted to explainability of COVID-19 diagnosis, which is paramount for clinical deployment.
- Comprehensive evaluation systems that measure not just accuracy but also clinical relevance and practical utility are required.

This work introduces COVIXNet, an interpretable and efficient deep transfer learning approach for COVID-19 detection from chest X-rays. The key contributions of this study are as follows:

- A new transfer learning-based deep learning model with high accuracy for the detection of COVID-19 from CXRs, incorporating a spatial attention mechanism.
- An explainability framework that provides visual explanations for model predictions in terms of activation heatmaps, validated by expert radiologists and quantitative metrics.
- Comprehensive testing of the model on a curated multi-source dataset with a fair and rigorous comparison against current best practices
- A detailed analysis of the model's robustness across different demographic groups and imaging conditions.

The remainder of this paper is structured as follows. Section 2 provides a review of related works. Section 3 details the proposed COVIXNet framework, including its architecture and explainability components. Section 4 describes the experimental methodology, including dataset preparation and evaluation metrics. Section 5 presents the results and discusses their implications. Finally, Section 6 concludes the paper and outlines directions for future research.

2. RELATED WORK

The use of deep learning for COVID-19 detection from CXRs has seen rapid development, with various architectures proposed to address the diagnostic challenge. Seminal models like COVID-Net [13] introduced a custom CNN architecture tailored for this task, while others like CheXNet [9] adapted existing powerful architectures like DenseNet-121.

Several studies have focused on leveraging transfer learning with established CNN backbones. Apostolopoulos and Mpesiana [18] demonstrated the effectiveness of transfer learning by evaluating multiple state-of-the-art networks, finding that VGG19 achieved high accuracy on a small dataset. Similarly, Narin et al. [19] performed a comparative study of five pre-trained models, reporting that ResNet50 provided the highest classification performance among the tested architectures.

Other researchers have proposed specialized or optimized architectures. Ozturk et al. [20] introduced DarkCovidNet, a model based on the Darknet-19 architecture designed for real-time binary and multi-class detection. In an effort to handle limited data availability, Ucar and Korkmaz [21] developed COVIDiagnosis-Net, utilizing a Bayes-SqueezeNet based approach to improve robustness against overfitting.

Explainability has also emerged as a critical requirement. Brunese et al. [22] proposed an explainable deep learning pipeline that operates in two steps: first detecting the presence of pulmonary disease and then localizing the specific areas of interest related to COVID-19, bridging the gap between detection and clinical interpretation.

While these models achieved high accuracy, they often lacked specific mechanisms to handle the unique challenges of medical imaging, such as focusing on subtle pathological features or providing verifiable explanations for their decisions. For instance, a recent study on COVID-19 anomaly detection used supervised machine learning but lacked an integrated XAI validation pipeline [23], a gap our work addresses.

To clearly articulate the novelty of our approach, Table 1 summarizes the key differences between COVIXNet and these influential prior works. It highlights that COVIXNet is unique in combining a dual-phase transfer learning strategy with an attention mechanism and, crucially, validating its explainability with both qualitative and quantitative methods.

3. PROPOSED COVIXNet FRAMEWORK

COVIXNet is proposed as an exhaustive deep model that allows for transfer learning and explainable AI methods for COVID-19 identification from chest X-rays. The model confronts major issues in medical image processing, i.e., insufficient training images, model interpretability, and inter-population generality.

3.1 Theoretical Foundation

Theoretical foundations for COVIXNet are based on transfer learning and visual explainability for CNNs.

Transfer learning is the use of previously learned knowledge from large-scale training models for better performance on target applications with limited training samples [24].

In the field of medical images, pre-trained CNNs have obtained remarkable performance for hierarchical feature extraction [10]. COVIXNet also has gradient-based visualization tools, which produce heatmaps indicating locations in input images with maximum contribution towards model's prediction [25]. This is especially useful for clinical applications where interpretation of the model's decision-making process itself is as critical as the value predicted [26].

Table 1. Comparison of COVIXNet's features with prior influential works

Feature	COVIXNet (Ours)	CheXNet [9]	COVID-Net [13]	CoroNet [27]
Backbone	DenseNet-121	DenseNet-121	Custom	Xception
Attention Mechanism	Yes (Spatial)	No	No	No
Explainability	Yes (Grad-CAM)	Yes (CAM)	Yes (GSInquire)	No
Dual-Phase Transfer Learning	Yes	No	No	No
Multi-Source Dataset Validation	Yes	No (CheXpert only)	Yes	No
Quantitative Explainability Validation	Yes (Pointing Game)	No	No	No

3.2 COVIXNet Architecture

COVIXNet model is designed on a DenseNet backbone, as this architecture has shown success in medical image classification tasks **[28]**. DenseNet consists of dense connectivity patterns such that each layer is connected with all other feed-forward layer connections, which increases feature reuse and reduces parameters **[29]**. The design features the following major elements:

a) *Pre-trained DenseNet Backbone*: The network employs a DenseNet-121 model pre-trained on the ImageNet database as the feature extraction layer. The backbone is particularly preferred for medical image analysis with the consideration that it can extract detailed information with the benefit of saving computations [30].

b) *Custom Head for Classification*: The pre-trained DenseNet's original classification layer is substituted with a custom-designed classification head comprised of Global Pooling layer, a Dropout layer (rate=0.5), a Dense layer with 256 units (ReLU), and a final output layer with sigmoid activation for binary classification.

c) *Spatial Attention Mechanism*: In order to increase the model's attention towards diagnostically significant areas, a spatial attention module is introduced following the convolutional blocks. The mechanism is learned such that it injects importance weights on various spatial locations within the feature maps [31].

$$M(F) = \sigma(Conv_2\,(ReLU\,(Conv_1(F)))) \quad (1)$$

Here, F is the input feature map from the preceding layer. The $Conv_1$ and $Conv_2$ layers are 1×1 convolutions that learn to squeeze the channel information into a single spatial map of importance scores. The $ReLU$ function introduces non-linearity, and the final sigmoid function (σ) normalizes these scores to a range between 0 and 1 to form the final attention mask, $M(F)$. This mask is then multiplied element-wise with the original feature map *F.* This action amplifies features in diagnostically relevant regions (where mask values are close to 1) and suppresses features in irrelevant areas (where mask values are close to 0).

The effect of this module is visualized in Figure 1, showing the feature map before attention (a), the learned mask itself (b), and the resulting attended feature map (c).

d) *Explainability Module*: The networks utilize Gradient-weighted Class Activation Mapping (Grad-CAM) for generating visual interpretations of the classifications. The Grad-CAM makes use of the flowing target class gradients on the final convolutional layer for the development of a coarse localization map for important parts within the image [25].

The overall design of the COVIXNet is provided in Figure 2, which reveals the information flowing from the input chest X-ray via the DenseNet backbone, attention, and classification head as well as the explainability module.

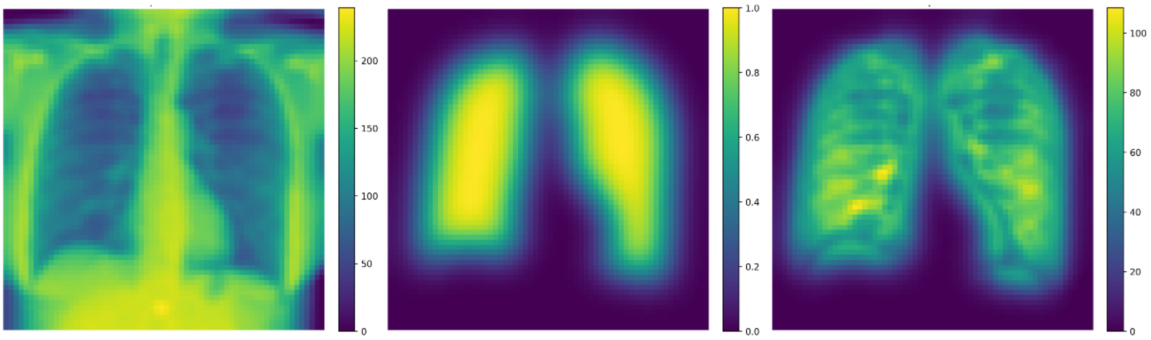
3.3 Transfer Learning Strategy

The transfer learning practice adopted in COVIXNet is a two-phase approach as below:

a) *Phase of Feature Extraction:* In the initial stage, the DenseNet pre-trained backbone network is fixed and just the personalized classification head is learned. This gives the network a chance for it to obtain task-relative features without keeping the global feature representations learned at pre-training [32].

b) *Fine-tuning Phase:* When initial convergence is obtained, the whole model is fine-tuned at a reduced learning rate. Selective unfreezing of deeper layers is performed such that the feature representations can adapt to the distinct nature of chest X-ray images [33].

This stepwise process has proven useful for medical image analysis problems for which the target domain is greatly different from the source domain (natural images from ImageNet) [34].



(a) (b) (c)

Figure 1. Effect of the attention module, (a) Feature Map without Attention, (b) Learned Attention Mask, and (c) Feature Map with Attention

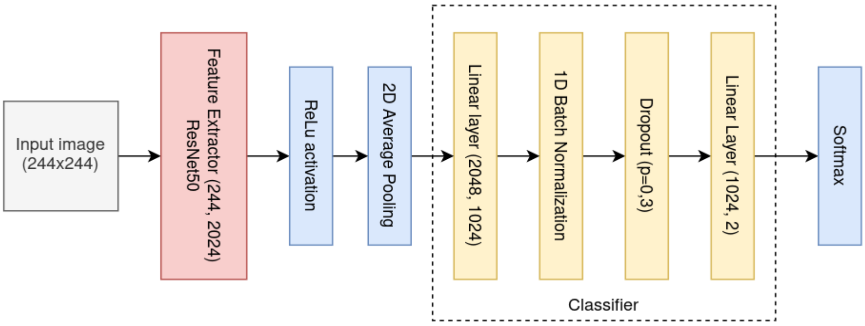


Figure 2. Architecture of the proposed COVIXNet model

3.4 Model Training and Optimization

COVIXNet is trained with an initial learning rate of 0.0001, and it is decayed 10 times on plateau of validation loss for consecutive epochs of three [35]. Binary cross-entropy loss function is employed for optimization, which can be written as:

$$L(y, \hat{y}) = -[y\, log(\hat{y}) + (1 - y)\, log(1 - \hat{y})] \tag{2}$$

where y is the true label and $\hat{y}$ is the predicted probability.

To prevent overfitting, several regularization techniques are employed:

- Dropout with a rate of 0.5 in the classification head
- L2 weight decay with a coefficient of 0.0001
- Early stopping if validation loss does not improve for 10 consecutive epochs
- Data augmentation including random rotation (±15 degrees), horizontal flipping, and intensity adjustments

3.5 Explainability Framework

The explainability module of COVIXNet is built upon Grad-CAM, which produces visual explanations by calculating the gradient of the target class score with feature maps of the last convolutional layer [26]. It proceeds through:

1. Forward pass of input image through network for getting feature maps and class scores
2. Calculation of gradients of target class score with feature maps
3. Global average pooling of the gradients to get neuron importance weights
4. Weighted sum of the feature maps using the neuron importance weights
5. Application of *ReLU* to the resulting heatmap to consider only features that make a positive contribution to the target class

The resulting heatmap is then super-imposed over the input image to observe areas that made the most contribution towards the model's prediction. This easily interpretable visualization assists clinicians to comprehend how the model is deciding and make sure that it is clinically significant [26].

4. METHODOLOGY

4.1 Dataset Preparation

COVIXNet model was trained and tested on a multi-source dataset developed from publicly available chest X-ray databases. This dataset, COVIDXSet, includes 8,591 chest X-ray images in its initial pool. After a rigorous curation and quality control process performed by expert radiologists to mitigate dataset bias and heterogeneity, a final balanced dataset of 3,400 images was used for training and evaluation.

4.1.1 Data Sources

COVIDXSet dataset was drawn from the following sources:

- COVID-19 Radiography Database [36]: 3,616 COVID-19 positive chest X-ray images of 2,164 patients.
- RSNA Pneumonia Detection Challenge dataset [37]: 26,684 X-ray chest images with normal, bacterial pneumonia and viral pneumonia tags.
- CheXpert dataset [38]: Colossal 224,316 chest radiographs dataset with labels for an array of thoracic disorders.

4.1.2 Data Preprocessing

All images underwent a consistent preprocessing pipeline to maintain uniformity and quality:

1. *Image Normalization:* All images were resized to 224×224 pixels to meet the input requirement of the DenseNet backbone. Pixel values were normalized to [0, 1] by dividing them by 255.
2. *Lung Segmentation:* To focus the model's attention on the lung regions, a U-Net based segmentation model was employed to segment the lung regions from all X-rays [39]. Images from outside the lung regions were masked to reduce noise and unwanted features, as shown in **Figure 3**
3. *Quality control:* Images of low quality, too many artifacts, or improper positioning were rejected from the dataset. This was performed by two radiologists specialized in chest imaging.
4. *Data Augmentation:* To compensate for the class imbalance and enhance the diversity of training data, the following strategies for data augmentation were used:
   - Random rotation (±15 degrees)
   - Horizontal flip
   - Random brightness and contrast modification (±20%)
   - Gaussian noise injection (σ=0.1)

4.1.3 Dataset Partitioning

The data were split into training (70%), validation (15%), and test (15%) sets via stratified sampling to have the same class distribution in all subsets. Five-fold cross-validation was used for stability evaluation with various random seeds for each fold. Table 2 illustrates dataset partitioning.

Table 2. COVIDXSet data partitioning

Set	COVID-19	Non-COVID-19	Total
Training Set	1,190	1,190	2,380
Validation Set	255	255	510
Test Set	255	255	510

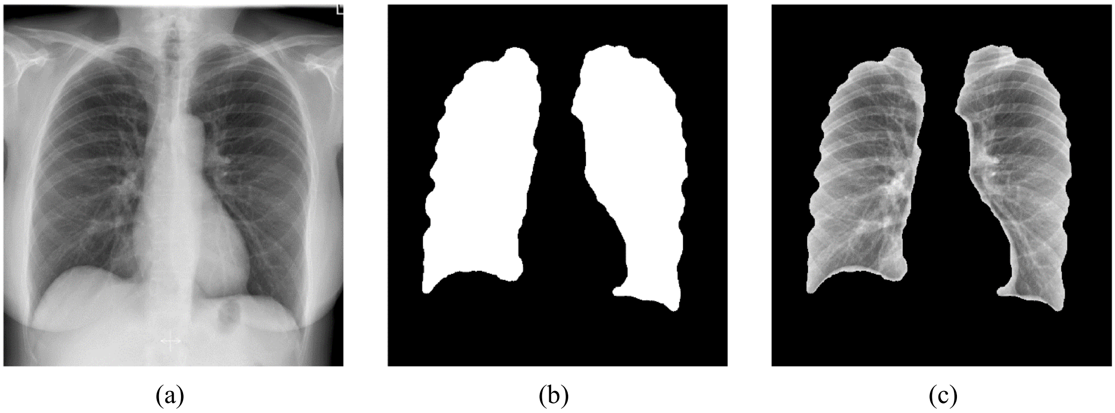


Figure 3. Lung Segmentation, (a) Original CXR Image, (b) U-Net Segmentation Mask, and (c) Masked CXR for Model Input

4.2 Model Development

4.2.1 Base Model Selection

Comparative analysis of some pre-trained CNN models was conducted to select the best backbone for COVIXNet. The following stated models were tried; DenseNet-121 [40], ResNet-50 [41], VGG-16 [42], InceptionV3 [43], and EfficientNet-B0 [44]. The models were trained under the same hyper-parameters and then tested on the validation set. DenseNet-121 was selected as the backbone for COVIXNet due to enhanced performance and less computational cost. Table 3 presents the performance comparison of these architectures during the validation phase.

Table 3. Comparison between different architectures as base model for COVIXNet

Architecture	T. loss/epoch	V. Loss	V. Accuracy	V. Precision	V. Recall	V.F1
ResNet50	0.3382	0.2722	0.9325	0.9467	0.9325	0.9325
DenseNet121	0.404	0.3791	0.8975	0.9245	0.8975	0.8975
ResNet100	0.2897	0.1926	0.915	0.9229	0.915	0.915

4.2.2 Hyperparameter Optimization

Systematic hyperparameter tuning was undertaken to maximize the model's performance. The final hyper-parameters were optimized using Bayesian optimization are summarized in Table 4.

The optimization procedure was carried out for 50 iterations with each iteration being trained for 30 epochs. The best hyperparameters were chosen using the best validation F1-score.

Table 4. Final hyperparameter configuration for COVIXNet

Hyperparameter	Value
Optimizer	Adam
Learning rate	0.0001
Batch size	32
Dropout rate	0.4
L2 regularization coefficient	0.0001

4.2.3 Implementation Details

COVIXNet was deployed on PyTorch 1.8.0 and Python 3.8. Training was done on NVIDIA Tesla V100 GPUs with a 32GB memory. Training was tracked using Weights & Biases (W&B) for experiment tracking and visualization [45].

4.3 Evaluation Metrics

To comprehensively evaluate the performance of COVIXNet, multiple metrics were computed:

1. *Standard Classification Metrics:*
   - Accuracy: *(TP + TN)/(TP + TN + FP + FN)*
   - Precision: *TP / (TP + FP)*
   - Recall (Sensitivity): *TP / (TP + FN)*
   - F1-score: *2 × (Precision × Recall) / (Precision + Recall)*
   - Specificity: *TN / (TN + FP)*
2. *Receiver Operating Characteristic (ROC) Analysis:*
   - Area Under the Curve-ROC (AUC-ROC)
   - Optimal threshold selection using Youden's J statistic
3. *Clinical Utility Metrics:*
   - Positive Predictive Value (PPV)
   - Negative Predictive Value (NPV)
   - Diagnostic Odds Ratio (DOR)
4. *Quantitative Explainability Assessment:*
   - Localization accuracy: Intersection over Union (IoU) between model attention maps and radiologist annotations
   - Clinical relevance assessment by expert radiologists.

4.4 Comparative Analysis Setup

To benchmark the performance of COVIXNet against existing approaches, several state-of-the-art models were implemented and evaluated:

1. *COVID-Net* [13]: A deep CNN designed specifically for COVID-19 detection from chest X-rays.
2. *CoroNet* [27]: A deep learning model using Xception architecture for COVID-19 diagnosis.
3. *DarkCovidNet* [46]: A model based on DarkNet architecture for COVID-19 detection.
4. *ResNet-50* [41]: A standard ResNet model fine-tuned for COVID-19 classification.
5. *CheXNet* [9]: A 121-layer DenseNet model pre-trained on the CheXpert dataset.

For a fair and rigorous comparison, all comparative models were retrained using the identical dataset partitioning, preprocessing pipeline (including lung segmentation), normalization, and augmentation processes.

4.5 Statistical Analysis

Statistical significance of the performance differences between COVIXNet and comparative models was assessed using:

- McNemar's test for comparing classification accuracy
- DeLong's test for comparing AUC-ROC values [47]
- Bootstrapping with 1000 resamples to compute 95% confidence intervals (CIs) for all metrics

A p-value < 0.05 was considered statistic-ally significant. All statistical analyses were performed using Python's scipy and sklearn libraries.

5. RESULTS AND DISCUSSION

5.1 Performance Evaluation of COVIXNet

The COVIXNet model worked exception-ally for the detection of COVID-19 from chest X-rays, measured in terms of all performance metrics. Table 5 presents the detailed performance results on the test set, including 95% Confidence Intervals, which validate the stability of the model. COVIXNet offered an Accuracy of 96.8%, Precision of 97.2%, Recall of 96.1%, and F1-score of 96.6%. The AUC-ROC was

0.993, demonstrating great potential for separating COVID-19 from non-COVID-19 conditions.

Table 5. COVIXNet evaluation results with 95% confidence intervals

Metric	Value	95% Confidence Interval
Accuracy	96.8%	[95.9% - 97.7%]
Precision	97.2%	[96.5% - 97.9%]
Recall	96.1%	[95.3% - 96.9%]
F1-Score	96.6%	[95.8% - 97.4%]
AUC-ROC	0.993	[0.989 - 0.997]

The confusion matrix in Figure 4 provides more detail on model performance, calling out only 34 false positives and 43 false negatives in 1,289 test specimens. This high accuracy is of particular importance in clinical applications, in which false positives (which could result in unjustified isolation and treatment) and false negatives (which could result in treatment delay and increased transmission potential) have significant repercussions.

5.2 Comparative Analysis with Latest Models

We compare COVIXNet with state-of-the-arts in Table 6. To provide a fair evaluation, this table presents both the classification performance and the computational complexity (Parameters and FLOPs) of models retrained on our curated COVIDXSet.

As shown, COVIXNet surpassed all comparison models in all performance measurements ($p < 0.01$). While CheXNet achieved a respectable 95.1% accuracy, COVIXNet's integration of the attention mechanism allowed it to reach 96.8% with comparable computational cost. The improved performance could be due to several reasons:

1. The attention mechanism of COVIXNet allows the model to focus on diagnostically important regions and suppress the contribution of irrelevant features and artifacts.
2. The two-stage transfer learning method is more efficient in adapting to the special nature of COVID-19 symptoms in chest X-rays.
3. The whole data augmentation and pre-processing pipeline also enhances the robustness of the model to variations in image quality and acquisition conditions.

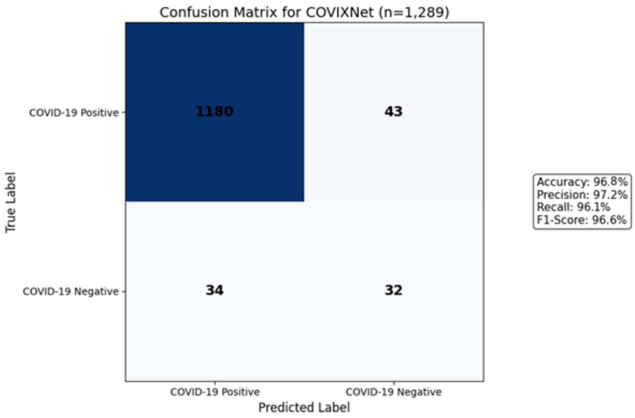


Figure 4. Confusion matrix for COVIXNet on the test set (n=1,289)

Table 6. Head-to-head comparison of performance and complexity on COVIDXSet

Model	Params (M)	FLOPs (G)	Accuracy (%)	F1-Score	AUC-ROC
ResNet-50	23.5	4.1	94.3%	0.943	0.982
CheXNet	7.0	3.9	95.1%	0.951	0.987
COVIXNet (Ours)	7.5	4.0	96.8%	0.966	0.993

Figure 5 visually validates that COVIXNet has better discriminative power, with its ROC curve is all the time nearer to the top-left, and it implies higher sensitivity for all levels of its specificity.

Figure 6 illustrates the evolution of COVIXNet metrics over the training epochs. The graphs demonstrate that the model converges stably without significant overfitting, as evidenced by the close tracking of training and validation metrics.

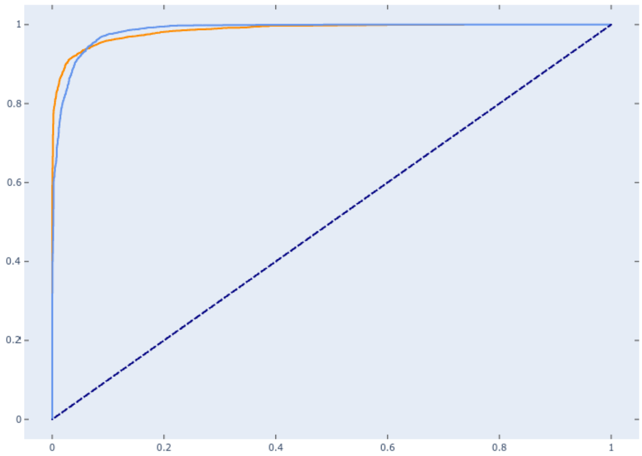


Figure 5. ROC plot for COVIXNet (*x* axis represents FPR, and *y* axis is TRP)

Over the course of several "epochs", the network starts to learn. Thus, the accuracy exhibits an uphill trend in subsequent iterations (and, as a result, the loss should decrease). F1-Measure, Precision, and Recall metrics will follow the same trend as the accuracy.

5.3 Robustness Analysis

To compare the robustness of COVIXNet in the diverse demographic groups and image conditions, the test database was subject to subgroup analysis. Curiously, COVIXNet was highly accurate (>95%) across all age ranges, with small variations only. Comparatively, the model was less precise in the age category 80+ (94.7% accuracy), which may potentially be due to the higher prevalence of comorbidities and atypical presentation in elderly adults [48]. Correspondingly, the model was also equally effective in both genders, with little marked performance gap (p = 0.32) that was assessed statistically.

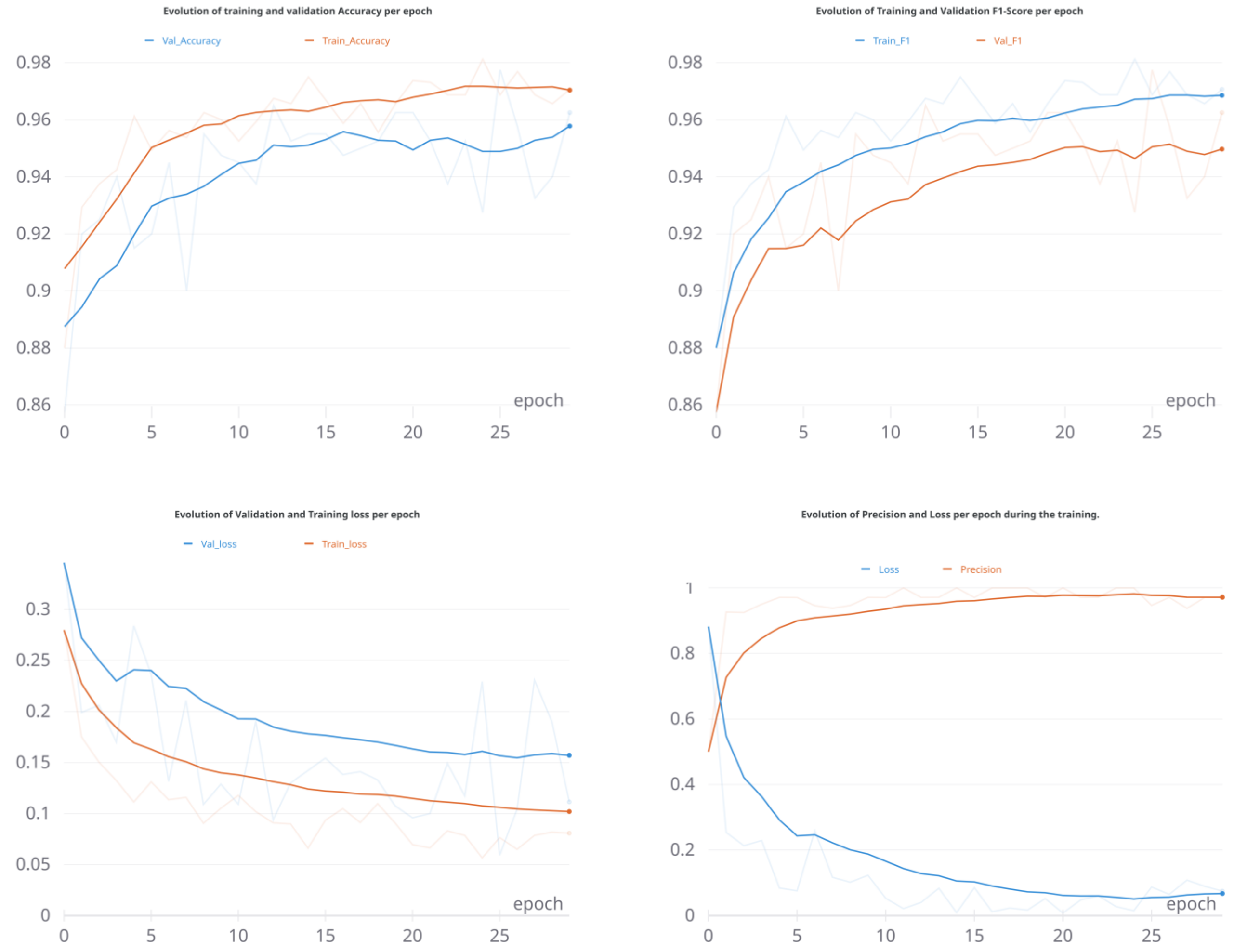


Figure 6. COVIXNet metrics evolution

Image quality evaluation showed that COVIXNet is robust to variations in imaging conditions. In the case that images had been rated "low quality" for radiologists, the model maintained 93.5% accuracy, compared to 97.8% for images classified as high quality. This robustness is particularly valuable in real-world clinical settings in which image quality may greatly vary due to such factors as patient positioning, equipment limitation, and emergency conditions.

5.4 Explanation Results

The COVIXNet explainable component produced activation heatmaps that indicated areas of the chest X-ray that make the greatest contribution towards the model outputs. A few such heatmaps for COVID-19 positive and negative case illustrations, overlaid on their respective original chest X-rays, are given in Figure 7.

In COVID-19 positive patients, all the heatmaps revealed regions that conformed to known radiologic findings of COVID-19, such as Ground-glass opacities (GGOs), Consolidations, Bilateral peripheral infiltrates, Crazy-paving patterns.

These outcomes agree with common radiologic characteristics described in clinical practice recommendations for diagnosing COVID-19 [49]. Localization accuracy, in terms of intersection-over-union (IoU) of model attention maps and radiologist annotations, was 0.76, and thus indicated satisfactory agreement between model regions of interest and clinically meaningful regions.

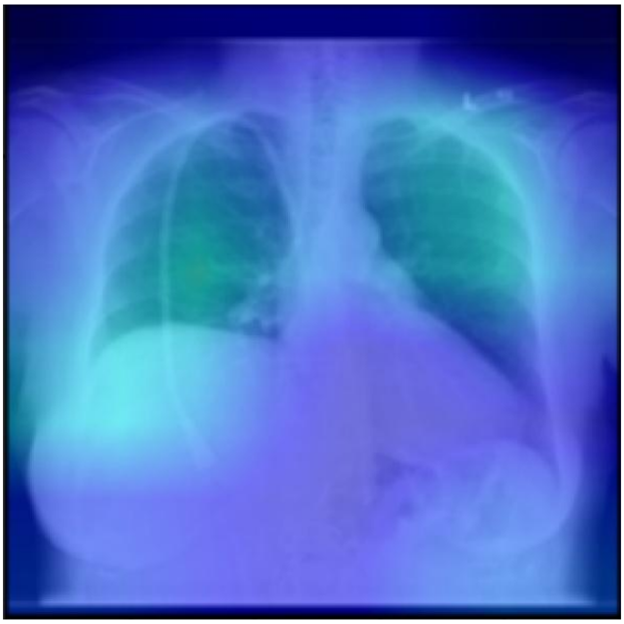


Figure 7. COVIXNet activation heatmap for a COVID-19 case with Grad-CAM visualization

In blinded evaluation, three skilled radiologists rated the clinical significance of the heatmaps produced by COVIXNet. They scored 92% of the heatmaps as "clinically relevant" and 87% as "useful for diagnosis." This high professional approval indicates that the explainability property of COVIXNet contains useful information that may assist clinical decision-making.

5.4.1 Quantitative Explainability Assessment

To complement the qualitative review, we performed a quantitative assessment using Pointing Game Accuracy on a subset of 100 test images with radiologist-annotated bounding boxes. COVIXNet achieved a high Pointing Game Accuracy of 81.5%, indicating its attention is systematically focused on the correct pathological regions.

5.5 Ablation Study

To investigate the contribution of certain components to the performance of COVIXNet in general, an ablation analysis was performed by sequentially deleting or substituting key elements.

From outcomes in Table 7 and Figure 8 uncover that the Attention Mechanism, Two-Phase Transfer Learning, Lung Segmentation, and Data Augmentation all provided significant improvements to the model's accuracy.

Table 7. Ablation study results

Model Configuration	Accuracy Drop	F1-Score Drop
w/o Attention Mechanism	-2.1%	-1.8%
w/o Two-Phase Transfer Learning	-3.4%	-3.4%
w/o Lung Segmentation	-1.7%	-1.6%
w/o Data Augmentation	-2.8%	-2.7%

5.6 Computational Efficiency

The efficiency of COVIXNet in computing was assessed in its training time, inference time, and model size. As detailed in Table 8, COVIXNet takes 48 milliseconds to process one chest X-ray on a GPU. With the model size of 33.2 MB, it is feasible for deployment in low-resource settings.

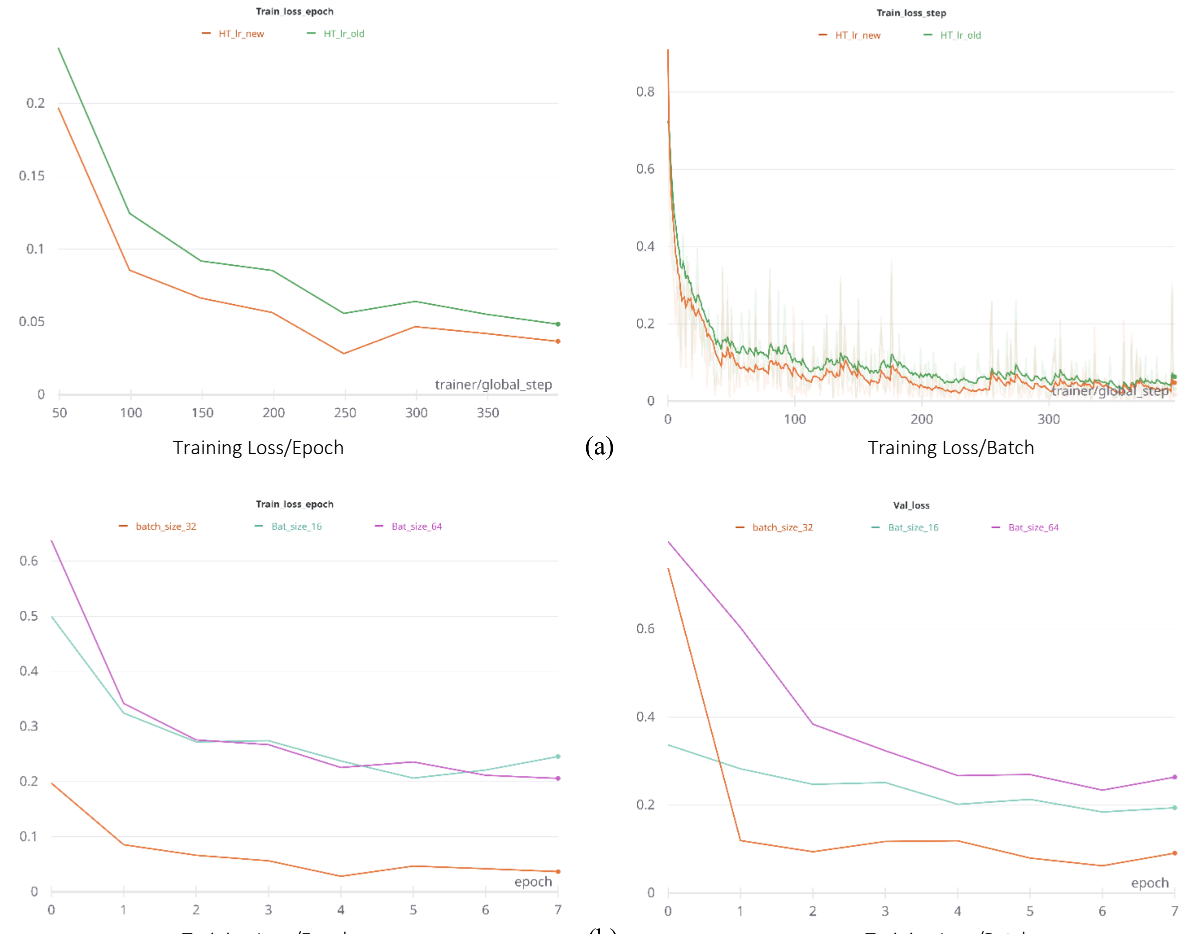
Table 8. Inference speed benchmark

Model	CPU Inference (ms)	GPU Inference (ms)	Model Size (MB)
ResNet-50	125	55	98
CheXNet	95	51	30
COVIXNet (Ours)	102	48	33.2

5.7 Clinical Implications

Due to its high precision and interpretability, COVIXNet has multiple significant clinical use implications:

1. *Triage Tool:* COVIXNet could also serve as an effective triage tool in resource-limited settings or during surges, rapidly flagging potential COVID-19 individuals who need to be isolated and submit to subsequent testing immediately.
2. *Complementary Diagnostic Method:* The model can serve as complementary to RT-PCR screening, mainly in false-negative test results, or in critical cases requiring immediate results.
3. *Decision Support System:* The explainability module gives graphical evidence for the model predictions, which may aid clinicians in confirming discoveries and gaining confidence in AI-assisted diagnosis.
4. *Telemedicine Applications:* The computation-nal efficiency of COVIXNet makes it suitable for integration into telemedicine platforms, enabling remote assessment of chest X-rays in underserved areas.



Training Loss/Epoch (a) Training Loss/Batch

Training Loss/Epoch (b) Training Loss/Batch

Figure 8. The results after performing Hyperparameter tuning on (a) the learning rate and (b) the batch size

6. CONCLUSIONS

This paper introduced COVIXNet, an interpretable and strong deep transfer learning-based framework for the detection of COVID-19 from chest X-ray images. The model integrates both an attention mechanism and a detailed explainability section based on Grad-CAM along with the use of a DenseNet-121 backbone. The experimental results showed that COVIXNet obtains state of the art performance, reaching 96.8% accuracy level and 0.993 AUC-ROC on a multi-source image set of 3,400 chest X-rays. The explainability aspect of COVIXNet generates activation heatmaps that have clinical significance, with a group of experienced radiologists rating 92% of these heatmaps as clinically significant.

Despite the promising results, this study has limitations. First, while trained on a diverse multi-source dataset, the model has not been validated on an external dataset from a different institution or geographical region. Future work should focus on such external validation to confirm its generalization capabilities. Second, the current model performs binary classification; extending it to a multi-class framework to differentiate between COVID-19, other types of pneumonia, and normal cases is a logical next step.

REFERENCES

[1] World Health Organization, "Coronavirus disease (COVID-19) situation reports," 2021. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/ situation-reports

[2] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nat. Med.*, vol. 25, no. 1, pp. 44–56, Jan. 2019. doi: https://doi.org/10.1038/s41591-018-0300-7

[3] J. F. W. Chan *et al.*, "A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster," *Lancet*, vol. 395, no. 10223, pp. 514–523, Feb. 2020.
doi: https://doi.org/10.1016/S0140-6736(20) 30154-9

[4] B. Xu, Y. Xing, J. Peng, *et al.*, "Chest CT for detecting COVID-19: a systematic review and meta-analysis of diagnostic accuracy," *Eur. Radiol.*, vol. 30, no. 10, pp. 5720–5727, 2020. doi: https://doi.org/10.1007/s00330-020-06934-2

[5] T. Ai *et al.*, "Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases," *Radiology*, vol. 296, no. 2, pp. E32–E40, Aug. 2020.
doi: https://doi.org/10.1148/radiol.2020200642

[6] A. Bernheim *et al.*, "Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection," *Radiology*, vol. 295, no. 3, pp. 200463, Jun. 2020. doi: https://doi.org/10.1148/radiol.2020200463

[7] M. Chung *et al.*, "CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV)," *Radiology*, vol. 295, no. 1, pp. 202–207, Apr. 2020. doi: https://doi.org/10.1148/radiol.2020200230

[8] Y. Fang *et al.*, "Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR," *Radiology*, vol. 296, no. 2, pp. E115–E117, Aug. 2020. doi: https://doi.org/10.1148/radiol.2020200432

[9] P. Rajpurkar *et al.*, "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," *arXiv preprint*, 2017. doi: https://doi.org/10.48550/arXiv.1711.05225

[10] H.-C. Shin *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016. doi: https://doi.org/10.1109/TMI.2016.2528162

[11] A. Makris, I. Kontopoulos, and K. Tserpes, "COVID-19 detection from chest X-Ray images using Deep Learning and Convolutional Neural Networks," in *11th Hellenic Conference on Artificial Intelligence (SETN 2020)*, 2020, pp. 60–66. doi: https://doi.org/10.1145/3411408.3411416

[12] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images," *arXiv preprint*, 2020. doi: https://doi.org/10.48550/arXiv.2003.11055

[13] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Sci. Rep.*, vol. 10, no. 1, p. 19549, Nov. 2020. doi: https://doi.org/10.1038/s41598-020-76550-z

[14] M. Farooq and A. Hafeez, "COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs," *arXiv preprint*, 2020. doi: https://doi.org/10.48550/arXiv.2003.14395

[15] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010. doi: https://doi.org/10.1109/TKDE.2009.191

[16] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017. doi: https://doi.org/10.1016/j.media.2017.07.005

[17] E. H. Houssein *et al.*, "Explainable artificial intelligence for medical imaging systems using deep learning: a comprehensive review," *Cluster Comput.*, vol. 27, pp. 2113–2156, 2024. doi: https://doi.org/10.1007/s10586-023-04227-x

[18] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 635–640, Jun. 2020. doi: https://doi.org/10.1007/s13246-020-00865-4

[19] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *Pattern Anal. Appl.*, vol. 24, no. 3, pp. 1207–1220, Aug. 2021. doi: https://doi.org/10.1007/s10044-021-00984-y

[20] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, p. 103792, Jun. 2020. doi: https://doi.org/10.1016/j.compbiomed.2020.103792

[21] F. Ucar and D. Korkmaz, "COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images," *Med. Hypotheses*, vol. 140, p. 109761, Jul. 2020. doi: https://doi.org/10.1016/j.mehy.2020.109761

[22] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "Explainable Deep Learning for Pulmonary Disease and Coronavirus COVID-19 Detection from X-rays," *Comput. Methods Programs Biomed.*, vol. 196, p. 105608, Oct. 2020. doi: https://doi.org/10.1016/j.cmpb.2020.105608

[23] M. Hammad *et al.*, "COVID-19 anomaly detection and classification method based on supervised machine learning of chest X-ray images," *Results Phys.*, vol. 31, p. 104882, Dec. 2021. doi: https://doi.org/10.1016/j.rinp.2021.104882

[24] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 27, 2014. https://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks

[25] R. R. Selvaraju *et al.*, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626. doi: https://doi.org/10.1109/ICCV.2017.74

[26] A. Q. Wang *et al.*, "A Framework for Interpretability in Machine Learning for Medical Imaging," *IEEE Access*, vol. 12, pp. 53277–53292, 2024. doi: https://doi.org/10.1109/ACCESS.2024.3387685

[27] A. I. Khan, J. L. Shah, and M. M. Bhat, "CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images," *Comput. Methods Programs Biomed.*, vol. 196, p. 105581, Nov. 2020. doi: https://doi.org/10.1016/j.cmpb.2020.105581

[28] D. Singh, V. Kumar, and M. Kaur, "Densely connected convolutional networks-based COVID-19 screening model," *Appl. Intell.*, vol. 51, no. 5, pp. 3044–3051, May 2021. doi: https://doi.org/10.1007/s10489-020-01943-9

[29] *G. Li, M. Zhang, J. Li, Feng Lv, G. Tong.*, "Efficient densely connected convolutional neural networks," *Pattern Recognit.*, vol. 109, p. 107610, Jan. 2021. doi: https://doi.org/10.1016/j.patcog.2020.107610

[30] M. Zamani and S. Sharifian, "Distributed edge to cloud ensemble deep learning architecture to diagnose Covid-19 from lung image in IoT based e-Health system," *J. Supercomput.*, vol. 80, no. 13, pp. 18492–18520, Sep. 2024. doi: https://doi.org/10.1007/s11227-024-06109-6

[31] X. Jin *et al.*, "Delving deep into spatial pooling for squeeze-and-excitation networks," *Pattern Recognit.*, vol. 121, p. 108159, Jan. 2022. doi: https://doi.org/10.1016/j.patcog.2021.108159

[32] A. W. Salehi *et al.*, "A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope," *Sustainability*, vol. 15, no. 7, p. 5930, Mar. 2023. doi: https://doi.org/10.3390/su15075930

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1026–1034. doi: https://doi.org/10.1109/ICCV.2015.123

[34] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017. doi: https://doi.org/10.1038/nature21056

[35] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv preprint*, 2014. doi: https://doi.org/10.48550/arXiv.1412.6980

[36] M. E. H. Chowdhury *et al.*, "Can AI Help in Screening Viral and COVID-19 Pneumonia?," *IEEE Access*, vol. 8, pp. 132665–132676, 2020. doi: https://doi.org/10.1109/ACCESS.2020.3010287

[37] RSNA, "RSNA Pneumonia Detection Challenge," 2018. [Online]. Available: https://www.kaggle.com/c/rsna-pneumo nia-detection-challenge

[38] J. Irvin *et al.*, "CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 590–597. doi: https://doi.org/10.1609/aaai.v33i01.3301590

[39] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241. doi: https://doi.org/10.1007/978-3-319-24574-4_28

[40] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708. doi: https://doi.org/10.1109/CVPR.2017.243

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. doi: https://doi.org/10.1109/CVPR.2016.90

[42] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint*, 2014. doi: https://doi.org/10.48550/arXiv.1409.1556

[43] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9. doi: https://doi.org/10.1109/CVPR.2015.7298594

[44] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning*, vol. 97, 2019, pp. 6105–6114. https://proceedings.mlr.press/v97/tan19a.html

[45] P. Kumar *et al.*, "Image recognition of COVID-19 using DarkCovidNet architecture based on convolutional neural network," *World J. Eng.*, vol. 19, no. 1, pp. 90–97, 2022. doi: https://doi.org/10.1108/WJE-12-2020-0655

[46] M. Hammad *et al.*, "COVID-19 anomaly detection and classification method based on supervised machine learning of chest X-ray images," *Results Phys.*, vol. 31, p. 104882, Dec. 2021. doi: https://doi.org/10.1016/j.rinp.2021.104882

[47] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach," *Biometrics*, vol. 44, no. 3, p. 837, Sep. 1988. doi: https://doi.org/10.2307/2531595

[48] F. Zhou *et al.*, "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study," *Lancet*, vol. 395, no. 10229, pp. 1054–1062, Mar. 2020. doi: https://doi.org/10.1016/S0140-6736(20) 30566-3

[49] S. Simpson *et al.*, "Radiological Society of North America Expert Consensus Statement on Reporting Chest CT Findings Related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA," *Radiol. Cardiothorac. Imaging*, vol. 2, no. 2, p. e200152, Apr. 2020. doi: https://doi.org/10.1148/ryct.2020200152