

# Trans-CAMNet: A Transformer-Based Grad-CAM Network for Lung Disease Classification

Kajal Kansal<sup>1\*</sup>, Akansha Singh<sup>1</sup>, Krishna Kant Singh<sup>2</sup> and Kanika Kansal<sup>3</sup>

<sup>1</sup>Bennett University, Greater Noida, India

<sup>2</sup>Delhi Technical Campus, Greater Noida, India

<sup>3</sup>ABES Engineering College, Ghaziabad, India

## Abstract

Accurate medical imaging analysis has become crucial in diagnosing and managing pulmonary diseases, especially considering the global prevalence of respiratory disorders. Chest X-ray classification has become one of the most effective diagnostic approaches in diagnosing pulmonary diseases and is valuable in offering clinicians a fast, noninvasive diagnostic solution. However, classifying thoracic abnormalities is challenging because of the variability of the pathological patterns and the lack of large annotated medical image datasets. To tackle these challenges, in this study, we introduce a novel approach that integrates fine-tuned deep learning-based frameworks, including CNNs and transformers. Further, to address the issues associated with deep learning models as black boxes, we employ the Grad-CAM as an interpretability technique to enhance clinical decision-making. It displays the areas that significantly contribute to the model's prediction of the lung regions. The proposed Trans-CAMNet framework, evaluated using the publicly available COVID-19 radiography dataset, achieves an accuracy of 98.33%, out-competing the traditional CNN architectures. These results highlight the possibility of transformer-based architectures in medical imaging tasks, with better classification accuracy and interpretability. These results provide a strong rationale for combining sophisticated deep learning architectures and interpretability methods to meet diagnostic performance and explainability in medical image analysis, especially for challenging pulmonary diseases.

## Keywords

CNN, COVID-19, Grad-CAM, Deep Learning, Pulmonary Diseases

## 1. Introduction

The development of medical imaging procedures has been relatively fast and has contributed significantly to diagnosing and treating pulmonary diseases [1]. Chest X-ray (CXR) is still frequently used as a simple, inexpensive, and safe tool for diagnosing lung diseases, including pneumonia, tuberculosis, and COVID-19. [2] However, the identification and accurate interpretation of CXR findings still pose a significant problem because of lung disease's many and varied pathological patterns [3]. This challenge is magnified by the scarcity of well-annotated large medical image datasets for training deep-learning models [4]. Hence, the development of dependable and generalized models is a challenge.

Recently, deep learning, especially CNN, has proved to be a potential tool for automatically detecting and diagnosing medical images with high accuracy [5]. Nonetheless, CNNs have inherent drawbacks in expressing long-distance relations and global context in images, essential for detecting intricate and minor lung pathologies. [6]. To overcome these issues, transformer-based models have been introduced, which are very efficient in handling sequential data and capturing the global context [7]. Due to self-attention mechanisms, transformers can capture the interactions within an image and improve upon image classification tasks. Though CNNs and transformers have shown outstanding performance in medical imaging, their black-box nature is a significant issue for clinicians [8]. Explaining the predictions made about medical images is essential to prevent the usage of unreliable and untrustworthy models in clinical decision-making. Grad-CAM (Gradient-weighted Class Activation Mapping) is one of the most popular methods to explain the decisions made by deep learning models. It underscores the areas of an image that are more important in predicting a model

<sup>1</sup>CMIS-2024: Seventh International Workshop on Computer Modeling and Intelligent Systems, May 3, 2024, Zaporizhzhia, Ukraine

✉ e22soep0003@bennett.edu.in (K.Kansal); akanshasingh@gmail.com (A.Singh); krishnaiitr2011@gmail.com (K.K.Singh); kanika.kansal2000@gmail.com (K.Kansal)

🆔 0009-0002-2214-5553 (K.Kansal); 0000-0002-5520-8066 (A.Singh); 0000-0002-6510-6768 (K.K.Singh); 0009-0003-0504-2365 (K.Kansal)



© 2024 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

and provides clinicians with more insights into the decision-making [9].

In this paper, we propose Trans-CAMNet, a new framework that integrates the benefits of transformer-based structures with Grad-CAM interpretability for more accurate and transparent lung disease categorization [10]. The proposed model uses CNN and transformers to improve feature extraction and context modeling; Grad-CAM enables the visualization of the model's decision-making process. The performance of Trans-CAMNet is assessed using the COVID-19 radiography dataset, and it is shown that Trans-CAMNet outperforms conventional CNN structures in terms of accuracy and explainability.

The objectives of this work are as follows:

- This study introduces Trans-CAMNet, a novel hybrid architecture that integrates transformer-based models and Grad-CAM for improved classification and interpretability in pulmonary disease diagnosis.
- This study compares the proposed architecture with state-of-the-art CNNs.

The following study sections are discussed: Section 2 describes the related studies. Section 3 discusses the materials and methods used in the study. Section 4 presents the results and discussion, and Section 5 concludes the study.

## 2. Related Work

Deep learning has advanced in recent years and enhanced the ability to analyze CXR images for diagnosing and comprehending thoracic diseases, including COVID-19 [11]. Recent work has explored strong and deep neural networks, ensemble models, and explainability methods like Grad-CAM, Grad-CAM++, and LRP to improve classification and explainability [12]. When used in different datasets, these approaches demonstrate the increasing role of AI-based instruments in enhancing diagnostic accuracy and aiding clinical management decisions. In this direction, Degerli et al. [13] used five deep neural networks (DNNs) to jointly localize the COVID-19-affected region and estimate the severity level of the infection based on CXR images. The approach used infection maps to explain the areas involved in the disease. QaTa-COV19 dataset was used in the study to offer annotated CXR images for COVID-19 diagnosis. By integrating multiple DNNs, the model performed reasonably well in detecting infected regions and severity levels, which is essential for clinical applications. Similarly, Mahmud et al. [14] used a convolutional neural network (CNN) for the multiclass classification of thoracic diseases, including COVID-19. For details, it could extract the hierarchical features with the help of depth-wise convolution, where the convolution layers were applied with different dilations. The model's performance was tested on three different data sets to demonstrate that it applies to different imaging sources. Chetoui et al. [15] used EfficientNet B7 as a CNN architecture to analyze CXR images from datasets such as BIMCV COVID-19+, RSNA, NIH, Montfort, and others. For explainability, Grad-CAM was used to explain the model's decision-making by visualizing regions of interest in the CXRs. The study also noted that the model could achieve high classification accuracy because of EfficientNet's specified network scaling method and feature extraction. Further proving its real-life capability, it could simultaneously operate on different datasets to increase efficiency.

Karim et al. [16] proposed a model comprising four CNN base learners and a Naïve Bayes as a meta-learner. In this work, four CNN architectures are used as base classifiers, where a Naïve Bayes meta-classifier is used to classify multiple classes of thoracic diseases, including COVID-19. The approach built upon integrating multiple CNNs took advantage of the synergistic learning capability and eliminated the overfitting problem. This model was applied to the Kaggle RSNA dataset to prove its ability to classify and interpret the CXR images accurately. In another study, Lee et al. [17] proposed and implemented an explanatory clustering framework called DeepSHA with a VGG-19-based model. DeepSHA offered explainable AI to cluster similar CXRs and then interpret the clustering to help diagnose. The framework was applied to public datasets, and its advantage was in providing interpretable clusters of similar cases, which would help study diseases and make clinical decisions. Altogether, these works highlight the possibility of using modern deep-learning methods with CXR images to diagnose COVID-19 and other thoracic pathology. Therefore, all the review articles show how deep learning techniques can help analyze CXR images to classify and diagnose thoracic

diseases, including COVID-19. The demonstrated high performance on various benchmarks also highlights the great promise of deep learning for transforming medical imaging into highly accurate, reliable, and explainable tools to enhance diagnostics and treatment of patients.

### **3. Methods Used**

#### **3.1 VGG-16**

VGG-16 is a deep convolutional neural network structure on the Visual Geometry Group initiative of the University of Oxford [18]. This model attracted much attention due to its excellent performance and the simplicity of its model when it was crowned the winner of the ILSVRC. Its design principles have become a vital architectural concept in deep learning, even in the case of image classification [19]. VGG-16 consists of 16 weight layers: A model with 13 convolutional layers and three fully connected layers. The architecture is uniform, with 3 x 3 convolutional filters used throughout the system, with a filter stride of one [20]. These filters allow for preserving the input dimensions when extracting local spatial patterns in the convolutional layers. The network has twice as many filters at the deeper layers (for example, 64, 128, 256, 512) to learn features at successive levels [21]. Max pooling is done using a filter of size 2x2 and a stride of 2 after every few convolutional layers to decrease the spatial size and work at a more abstract level. The last part of the network consists of three fully connected layers, where the previous layer implements the SoftMax activation to output the class probability [22]. At its release, it offered one of the highest performances for large datasets like ImageNet. In addition, specific pre-trained versions of VGG-16 are being introduced in various transfer learning projects [23]. Researchers have used the learned features for other computer vision applications, such as object detection, medical imaging, and style transfer [24].

#### **3.2 ResNet50**

The ResNet-50 model is a well-known deep convolutional neural network devised by Microsoft researchers in their paper "Deep Residual Learning for Image Recognition," published in 2015. This model is from the ResNet family, which proposed residual learning to overcome the problem, including vanishing gradients and performance degradation, that might be encountered when training intense networks [25]. ResNet-50 is a full-residual 50-layer model and is one of the most frequently used networks because of its depth and computational complexity [26]. The main advancement of ResNet-50 is the use of residual blocks. A residual block is built from the shortcut connections through which the model can skip one or several layers during the forward and backpropagation computation [27]. These are often known as skip connections, which endow the network with an ability to learn residual mapping rather than direct mapping [28].

ResNet-50 architecture has 48 convolution layers, one max pooling layer, and only one fully connected layer. It uses bottleneck residual blocks, where each block has three convolutional layers: As for the convolutional layers, there's always one 1x1 layer for downsampling, one 3x3 layer for feature extracting, and the third 1x1 layer for upsampling [29]. This design helps reduce computational costs, although it results in high representational power. In addition, performing batch normalization after each convolutional layer helps stabilize the training process and accelerate the convergence speed. ResNet-50 has performed well on many benchmarks, including the ILSVRC [30]. The pre-trained ResNet-50 model is commonly used for transfer learning, and researchers can further modify it as per the application domain for analyzing X-ray images, detecting tumors, or classifying satellite images [31].

#### **3.3 Inception-V3**

Inception-V3 is a deep convolutional neural network, a third version of Inception architecture proposed by Google in 2015. This was pointed out in a paper by Christian Szegedy et al. titled 'Rethinking the Inception Architecture for Computer Vision.' Compared to the previous models, the model under consideration expands on the existing algorithms and brings new methods for increasing the speed and accuracy of computations [32]. Inception-V3 is one of the most used architectures in computer vision tasks, especially image classification. The structure of Inception-V3 architecture is such that it performs well on large-scale image classification problems. It uses inception modules to extract features at various scales due to parallel 1x1, 3x3, and 5x5 convolutions [33]. These outputs are concatenated to cover a variety of spatial features efficiently. To enhance computational efficiency, the model proposes factorized convolutions or using two consecutive and smaller kernels (e.g., 5x5)

instead of one large one (e.g.,  $3 \times 3$ ) with a predictable decrease in accuracy and size of the parameters. Furthermore, batch normalization is used heavily across the layers for training purposes and to prevent overfitting [34]. It is also important to note that the Inception-V3 network is computationally efficient yet has achieved high levels of accuracy. The model does this by including auxiliary classifiers as part of the training process to assist the training in case of vanishing gradients. In addition, label smoothing applied to the loss function enhances the generalization because the model stops making nearly specific predictions [35]. Together with the Inception modules developed with much care, these techniques make Inception-V3 work efficiently and accurately on benchmarks such as ImageNet and more efficiently than deeper networks [36]. Inception-V3 has shown great versatility in many applications, from image classification transfer learning to feature extraction [37]. It is typically used in object detection, diagnosing medical images or images in general, and even art-related tasks such as transferring style [38].

### 3.4 DenseNet169

DenseNet-169 is a type of deep convolutional neural network of the DenseNet family, which was presented by Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger in their work "Densely Connected Convolutional Networks" in 2017. DenseNet architectures were created to overcome the shortcomings of traditional deep learning structures utilizing the dense connectivity method, which implies the direct connection of every layer of the neural network to any other layer in a feedforward manner [39]. This unique approach has distanced DenseNet as one of the most efficient architectures for recognition of images [40]. The DenseNet-169 model combined 169 layers, including the convolutional, pooling, and fully connected layers [41]. In DenseNet, the idea is to feed all the feature maps of a layer to the subsequent layers and take in all the previous layer's feature maps [42]. This is accomplished through dense blocks where feature maps are concatenated instead of summed, as in ResNet networks [43]. Transition layers are employed between these blocks of high density for feature maps down sampling and dimensionality reduction [44]. The growth rate, another hyperparameter in DenseNet, determines the number of new feature maps each layer in the network creates to balance the computational complexity and model capability [45].

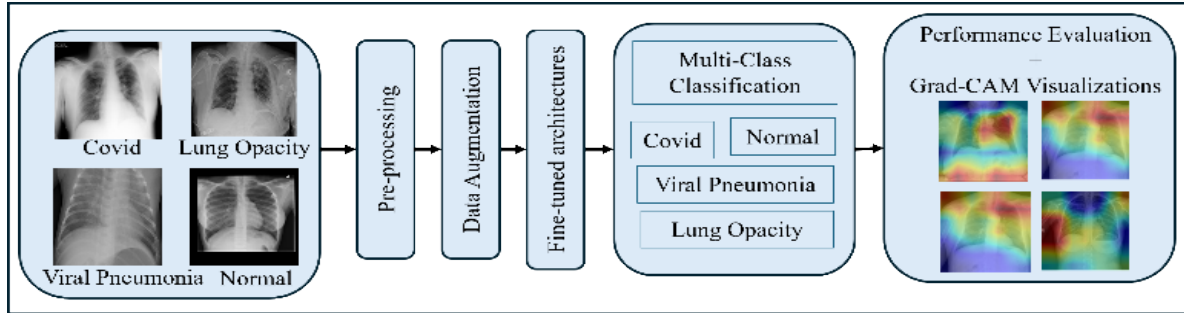
Another favorable point that can be identified with DenseNet-169 is the utilization of parameters [46]. Compared to the conventional architectures in which many parameters are needed to pursue high accuracy, DenseNet adopts dense connectivity to keep feature reuse low [47]. This leads to better gradient flow during backpropagation and easier model training, even with fewer parameters. Compared to other architectures, such as ResNet and ResNeXt, DenseNet-169 is best suited for learning detailed features in datasets and is, therefore, well suited for image classification, segmentation, and other vision-based tasks [48].

### 3.5 Proposed Trans-CAMNet

In this research, we proposed a fine-tuned Vision Transformer (ViT) model to classify chest radiograph images from the COVID-19 Radiography Dataset. Vision Transformer architecture is appropriate for medical imaging tasks since it uses a self-attention mechanism to capture long-range dependencies and global contextual information [49]. In the proposed model, the deep neural network is trained on a large-scale dataset to obtain the general features and then trained on the COVID-19 Radiography Dataset for tailoring the CXR images. The Vision Transformer takes an input image and partitions the input image into fixed spatial regions such as  $16 \times 16$ . The patch is then flattened into a vector and mapped into the fixed-dimensional embedding space [50]. These are supplemented by a learnable class token and position-specific positional encodings to feed to the transformer encoder. The encoder, implemented as a stack of several instances of the multi-head self-attention mechanism and feedforward neural networks, can learn global relations between patches. This approach helps the model to determine regions in chest radiographs that are important to distinguish between COVID-19, lung opacity, pneumonia, and normal cases [51].

We use a transfer learning approach to implement the Vision Transformer for the COVID-19 Radiography Dataset. A labeled chest radiograph is used to fine-tune the pre-trained ViT, thereby enabling it to modify the learned features for the distribution of the dataset [52]. Fine-tuning is working on the model's weights, and this can be done using a supervised learning approach where the loss can be optimized to get better results with the classes. Also, data augmentation and regularization are used, with the data size relatively small in this project, to avoid overfitting. The fine-tuned Vision Transformer shows substantial performance enhancements in diagnosing chest radiographs, using its capability to model global dependency and recognize the subtle differences in the radiographic

features of COVID-19 [53]. Additionally, the attention maps of the model also make interpretation easier since they point out the areas that are most relevant to the prediction in the obtained CXR images. These attention-based visualizations are consistent with the radiological diagnosis, making the model accurate and clinically usable. The concept of the proposed fine-tuned Vision Transformer model indicates that transformer-based models can be used to solve issues in medical image analysis. By incorporating external knowledge and learning the characteristics of chest radiographs, the model provides high accuracy on the COVID-19 Radiography Dataset and advances the research of AI approaches to COVID-19 detection and diagnosis. Figure 1 describes the workflow used in the study.



**Figure 1:** Workflow used in the study

## 4. Experiment

### 4.1 Dataset Used

The dataset used for the study consists of four disease categories: COVID-19, Normal, Viral Pneumonia, and Lung Opacity. The training and testing split is 70:30. The COVID-19 category consists of 3,616 images; from them, 2,531 images are utilized for training, and 1,085 images are used for testing. The most extensive files, containing 10,200 images in the Normal category, have been split between 7,140 images for training and 3,060 for testing. The Viral Pneumonia category contains 1345 images; of them, 941 are used for training, and 404 are used for testing. For the Lung Opacity, the category comprised 6,012 images, with 4,208 for training and 1,804 for testing. This means the model addresses various diseases and is trained and tested equally for all disease groups, making it reliable and accurate.

### 4.2 Evaluation Metrics

All standard measures were used to assess the outcome of the proposed models, such as Accuracy, Precision, Recall, and the F1-score. Accuracy calculates the ratio of the total number of instances correctly predicted to the total number of cases. Recall measures the model's capability of correctly identifying positive samples without counting false samples, and it is essential in reducing wrong classification. Recall measures how many positive actual cases the model identified. The F1-score, the measure of precision and recall in equal proportion, is helpful in the case of an unbalanced set of data. All these metrics, taken together, present a strong framework by which one can perform a comparative analysis of the strengths and weaknesses of each model to determine their ability to predict.

### 4.3 Results

The ability of the models to perform in terms of features such as the accuracy, precision, recall, and F1-score of the identified models, VGG-16, ResNet50, Inception-V3, DenseNet-169, and Trans-CAMNet is valuable information regarding each model's suitability. As each model corresponds to a different architectural complexity and ingenuity tier, the experiment (Table 1) shows how the performance differs on the given dataset. VGG-16, the oldest architecture among the architectures under comparison, has a test accuracy of 94.06%, precision of 84.62%, recall of 79.70%, and F1-measure of 82.09%. This can be attributed to its inability to perform residual or dense connections, preventing it from learning deeper hierarchical features excellently. The precision and recall are somewhat lower, implying that several images are misclassified, and VGG-16 is not suited for complex patterns of a given dataset. ResNet50 yields a much better result of 95.70 % accuracy, 93.61 % precision, 97.74 % recall, and a f1- score of 95.63%. The high recall suggests that ResNet50 has excellent actual

identification capacity. Its residual architecture helps reduce the vanishing gradient problem; thus, the model can train deeper networks. The high percentage of true positives and true negatives focuses on the stability of the measure between precision and recall. Similar performance is improved in Inception-V3 by attaining an accuracy of 97.13%, precision of 97.98%, recall of 91.79%, and F1-score of 94.79%. The inception modules mean multiple-scale filtering, allowing the model to get high-level features efficiently. This leads to better precision than ResNet50, meaning it has fewer false positives. DenseNet-169 achieved an accuracy of 97.96%, precision of 92.83%, recall of 95.48%, and F1-score of 94.13%. Due to its condensed network connections, this architecture entails reusing features and gradients, making learning extraordinary. Its high recall means it is good at identifying true positives.

**Table 1**  
**Performance of different models.**

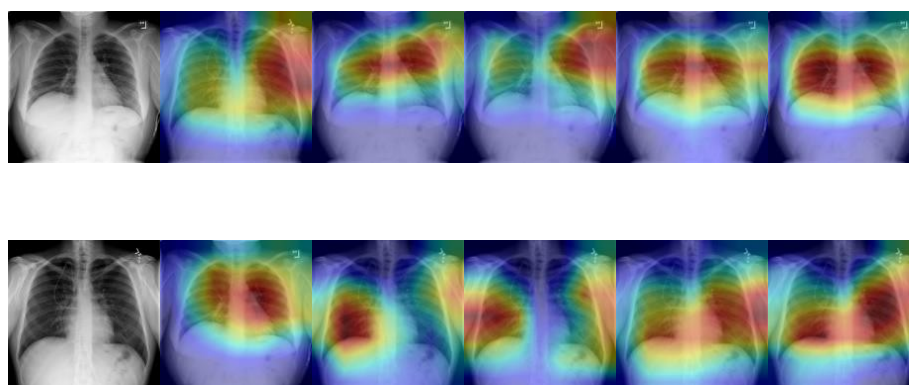
Model	Accuracy	Precision	Recall	F1score
VGG-16	94.06	84.62	79.70	82.09
ResNet50	95.70	93.61	97.74	95.63
Inception-V3	97.13	97.98	91.79	94.79
DenseNet169	97.96	92.83	95.48	94.13
Trans-CAMNet	98.33	97.98	98.56	98.27

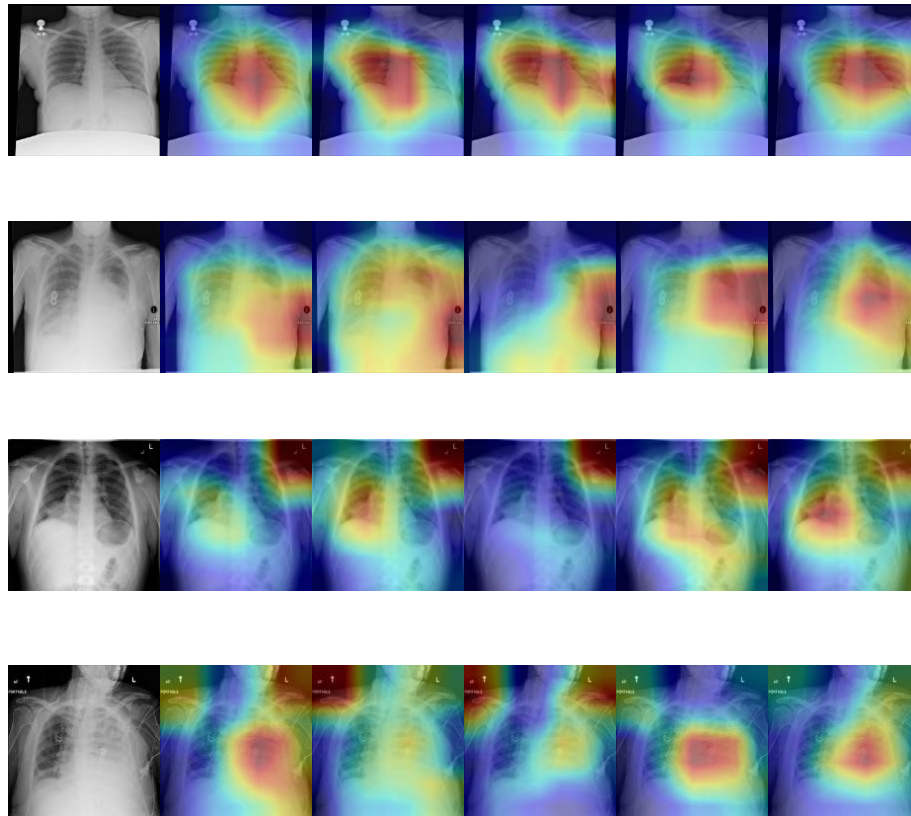
The proposed Trans-CAMNet has the highest overall accuracy of 98.33%, precision of 97.98%, recall of 98.56%, and F1 score of 98.27%. The nearly optimal values of precision and recall demonstrate excellent reliability, which is especially valuable for tasks where false positive and false negative results need to be avoided. The choice of model depends on the specific application requirements, as Trans- CAMNet is the best solution for critical cases with the highest level of needed accuracy and favoring balanced precision and recall values. This model may be improved by developing vision transformer architectures with attention to focus on the most essential objects while preserving overall context. Figure 2 depicts the Grad-CAM visualizations of different models.

## 5. Conclusion

This research shows that the proposed approach of fine-tuning the CNN- -transformer can effectively classify pulmonary diseases from CXR images. The Trans-CAMNet proposed in this study yields impressive results with an accuracy of 98.33%, thereby out-competing traditional CNN-based models. When used as an interpretability tool, Grad-CAM enlightens the model's decision-making process and increases its suitability for clinical use. These results highlight the opportunity to incorporate transformer-based medical imaging architectures that increase diagnostic performance and interpretability. The proposed approach can serve as a basis for future work combining deep learning models with interpretability methods and ensure more accurate and explainable machine learning-based diagnostics of pulmonary diseases.

**CXR**s      VGG-16      ResNet50      Inception-V3      DenseNet169      Trans-CAMNet





**Figure 2:** Grad-CAM visualizations of different models.

## Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

## References

- [1] A. Kumar, R. Manikandan, U. Kose, D. Gupta, and S. C. Satapathy, "Doctor's Dilemma: Evaluating an Explainable Subtractive Spatial Lightweight Convolutional Neural Network for Brain Tumor Diagnosis," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 17, no. 3s, pp. 1–26, Oct. 2021, doi: 10.1145/3457187.
- [2] K. Kansal, T. B. Chandra, and A. Singh, "Advancing differential diagnosis: a comprehensive review of deep learning approaches for differentiating tuberculosis, pneumonia, and COVID-19," *Multimed Tools Appl*, May 2024, doi: 10.1007/s11042-024-19350-1.
- [3] R. K. Singh, R. Pandey, and R. N. Babu, "COVIDScreen: explainable deep learning framework for differential diagnosis of COVID-19 using chest X-rays," *Neural Comput Appl*, vol. 33, no. 14, pp. 8871–8892, Jul. 2021, doi: 10.1007/s00521-020-05636-6.
- [4] R. Wu, C. Liang, Y. Li, X. Shi, J. Zhang, and H. Huang, "Self-supervised transfer learning framework driven by visual attention for benign–malignant lung nodule classification on chest CT," *Expert Syst Appl*, vol. 215, p. 119339, Apr. 2023, doi: 10.1016/j.eswa.2022.119339.
- [5] 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, Nov. 2020, doi: 10.1016/j.cmpb.2020.105608.
- [6] L. V. de Moura, C. Mattjie, C. M. Dartora, R. C. Barros, and A. M. Marques da Silva, "Explainable Machine Learning for COVID-19 Pneumonia Classification With Texture-Based Features Extraction in Chest Radiography," *Front Digit Health*, vol. 3, Jan. 2022, doi: 10.3389/fdgth.2021.662343.

- [7] S. Sah, B. Surendiran, R. Dhanalakshmi, and M. Yamin, "Covid-19 cases prediction using SARIMAX Model by tuning hyperparameter through grid search cross-validation approach," *Expert Syst*, vol. 40, no. 5, Jun. 2023, doi: 10.1111/exsy.13086.
- [8] H. I. Hussein, A. O. Mohammed, M. M. Hassan, and R. J. Mstafa, "Lightweight deep CNN-based models for early detection of COVID-19 patients from chest X-ray images," *Expert Syst Appl*, vol. 223, p. 119900, Aug. 2023, doi: 10.1016/j.eswa.2023.119900.
- [9] K. Kansal, T. B. Chandra, and A. Singh, "ResNet-50 vs. EfficientNet-B0: Multi-Centric Classification of Various Lung Abnormalities Using Deep Learning," *Procedia Comput Sci*, vol. 235, pp. 70–80, 2024, doi: 10.1016/j.procs.2024.04.007.
- [10] K. Kansal and S. Sharma, "Predictive Deep Learning: An Analysis of Inception V3, VGG16, and VGG19 Models for Breast Cancer Detection," 2024, pp. 347–357. doi: 10.1007/978-3-031-56703-2\_28.
- [11] K. Kansal, T. B. Chandra, A. Singh, and K. K. Singh, "E-CNN: ensembled CNN learning approach for pneumonia detection in chest X-ray images," *IET Conference Proceedings*, vol. 2024, no. 7, pp. 80–86, Sep. 2024, doi: 10.1049/icp.2024.2532.
- [12] K. Kansal and S. Sharma, "A Predictive Deep Learning Ensemble-Based Approach for Advanced Cancer Classification," 2024, pp. 335–346. doi: 10.1007/978-3-031-56703-2\_27.
- [13] A. Degerli *et al.*, "COVID-19 infection map generation and detection from chest X-ray images," *Health Inf Sci Syst*, vol. 9, no. 1, p. 15, Dec. 2021, doi: 10.1007/s13755-021-00146-8.
- [14] T. Mahmud, M. A. Rahman, and S. A. Fattah, "CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization," *Comput Biol Med*, vol. 122, p. 103869, Jul. 2020, doi: 10.1016/j.combiomed.2020.103869.
- [15] M. Chetoui and M. A. Akhloufi, "Deep Efficient Neural Networks for Explainable COVID-19 Detection on CXR Images," 2021, pp. 329–340. doi: 10.1007/978-3-030-79457-6\_29.
- [16] Md. R. Karim, T. Dohmen, M. Cochez, O. Beyan, D. Rebholz-Schuhmann, and S. Decker, "DeepCOVIDExplainer: Explainable COVID-19 Diagnosis from Chest X-ray Images," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, Dec. 2020, pp. 1034–1037. doi: 10.1109/BIBM49941.2020.9313304.
- [17] K.-S. Lee, J. Y. Kim, E. Jeon, W. S. Choi, N. H. Kim, and K. Y. Lee, "Evaluation of Scalability and Degree of Fine-Tuning of Deep Convolutional Neural Networks for COVID-19 Screening on Chest X-ray Images Using Explainable Deep-Learning Algorithm," *J Pers Med*, vol. 10, no. 4, p. 213, Nov. 2020, doi: 10.3390/jpm10040213.
- [18] C. Sitaula and M. B. Hossain, "Attention-based VGG-16 model for COVID-19 chest X-ray image classification," *Applied Intelligence*, vol. 51, no. 5, pp. 2850–2863, May 2021, doi: 10.1007/s10489-020-02055-x.
- [19] B. Chinta and Moorthi. M, "EEG-dependent automatic speech recognition using deep residual encoder based VGG net CNN," *Comput Speech Lang*, vol. 79, p. 101477, Apr. 2023, doi: 10.1016/j.csl.2022.101477.
- [20] B. K. Durga and V. Rajesh, "A ResNet deep learning- based facial recognition design for future multimedia applications," *Computers and Electrical Engineering*, vol. 104, p. 108384, Dec. 2022, doi: 10.1016/j.compeleceng.2022.108384.
- [21] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting COVID- 19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2," *Inform Med Unlocked*, vol. 19, p. 100360, 2020, doi: 10.1016/j.imu.2020.100360.
- [22] Y. Chen *et al.*, "Classification of lungs infected COVID-19 images based on inception-ResNet," *Comput Methods Programs Biomed*, vol. 225, p. 107053, Oct. 2022, doi:10.1016/j.cmpb.2022.107053.
- [23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.
- [24] N. N. Prakash, V. Rajesh, D. L. Namakhwa, S. Dwarkanath Pande, and S. H. Ahammad, "A DenseNet CNN-based liver lesion prediction and classification for future medical diagnosis," *Sci Afr*, vol. 20, p. e01629, Jul. 2023, doi: 10.1016/j.sciaf.2023.e01629.
- [25] M. G. Lanjewar, K. G. Panchbhai, and P. Charanarur, "Lung cancer detection from CT scans using modified DenseNet with feature selection methods and ML classifiers," *Expert Syst*

- Appl, vol. 224, p. 119961, Aug. 2023, doi: 10.1016/j.eswa.2023.119961.
- [26] I. Pacal, "Improved Vision Transformer with Lion Optimizer for Lung Diseases Detection," *Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi*, May 2024, doi: 10.29137/umagd.1469472.
  - [27] P. Rajpurkar et al., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," Nov. 2017.
  - [28] L. Yao, E. Poblens, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels," Oct. 2017.
  - [29] F. Altaf, S. M. S. Islam, and N. K. Janjua, "A novel augmented deep transfer learning for classification of COVID-19 and other thoracic diseases from X-rays," *Neural Comput Appl*, vol. 33, no. 20, pp. 14037–14048, Oct. 2021, doi: 10.1007/s00521-021-06044-0.
  - [30] 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: 10.1007/s13246-020-00865-4.
  - [31] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra 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: 10.1016/j.compbimed.2020.103792.
  - [32] 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: 10.1016/j.cmpb.2020.105581.
  - [33] 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," Mar. 2020.
  - [34] Prabira Kumar Sethy and S. Behera, "Detection of Coronavirus Disease (COVID-19) Based on Deep Features," *Medicine, Computer Science*, 2020.
  - [35] S. Toraman, T. B. Alakus, and I. Turkoglu, "Convolutional capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks," *Chaos Solitons Fractals*, vol. 140, p. 110122, Nov. 2020, doi: 10.1016/j.chaos.2020.110122.
  - [36] H. Panwar, P. K. Gupta, M. K. Siddiqui, R. Morales-Menendez, and V. Singh, "Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet," *Chaos Solitons Fractals*, vol. 138, p. 109944, Sep. 2020, doi: 10.1016/j.chaos.2020.109944.
  - [37] L. Wang and A. Wong, "COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images," Mar. 2020.
  - [38] M. Toğaçar, B. Ergen, and Z. Cömert, "COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches," *Comput Biol Med*, vol. 121, p. 103805, Jun. 2020, doi: 10.1016/j.compbimed.2020.103805.
  - [39] S. Guendel et al., "Learning to recognize Abnormalities in Chest X-Rays with Location-Aware Dense Networks," Mar. 2018.
  - [40] P. Kumar, M. Grewal, and M. M. Srivastava, "Boosted Cascaded Convnets for Multilabel Classification of Thoracic Diseases in Chest Radiographs," 2018, pp. 546–552. doi: 10.1007/978-3-319-93000-8\_62.
  - [41] H. Wang et al., "Detecting thoracic diseases via representation learning with adaptive sampling," *Neurocomputing*, vol. 406, pp. 354–360, Sep. 2020, doi: 10.1016/j.neucom.2019.06.113.
  - [42] S. Sani and H. E. Shermeh, "A novel algorithm for detection of COVID-19 by analysis of chest CT images using Hopfield neural network," *Expert Syst Appl*, vol. 197, p. 116740, Jul. 2022, doi: 10.1016/j.eswa.2022.116740.
  - [43] C. K. Kim et al., "An automated COVID-19 triage pipeline using artificial intelligence based on chest radiographs and clinical data," *NPJ Digit Med*, vol. 5, no. 1, p. 5, Jan. 2022, doi: 10.1038/s41746-021-00546-w.
  - [44] M.-L. Huang and Y.-C. Liao, "A lightweight CNN-based network on COVID-19 detection using X-ray and CT images," *Comput Biol Med*, vol. 146, p. 105604, Jul. 2022, doi: 10.1016/j.compbimed.2022.105604.
  - [45] Md. Nahiduzzaman, Md. R. Islam, and R. Hassan, "ChestX-Ray6: Prediction of multiple diseases including COVID-19 from chest X-ray images using convolutional neural network," *Expert Syst Appl*, vol. 211, p. 118576, Jan. 2023, doi: 10.1016/j.eswa.2022.118576.
  - [46] G. M. M. Alshmrani, Q. Ni, R. Jiang, H. Pervaiz, and N. M. Elshennawy, "A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images,"

- Alexandria Engineering Journal, vol. 64, pp. 923–935, Feb. 2023, doi: 10.1016/j.aej.2022.10.053.
- [47] Y. H. Bhosale and K. S. Patnaik, "PulDi-COVID: Chronic obstructive pulmonary (lung) diseases with COVID-19 classification using ensemble deep convolutional neural network from chest X-ray images to minimize severity and mortality rates," Biomed Signal Process Control, vol. 81, p. 104445, Mar. 2023, doi: 10.1016/j.bspc.2022.104445.
  - [48] Md. Nahiduzzaman et al., "Parallel CNN-ELM: A multi-class classification of chest X-ray images to identify seventeen lung diseases including COVID-19," Expert Syst Appl, vol. 229, p. 120528, Nov. 2023, doi: 10.1016/j.eswa.2023.120528.
  - [49] C. Antunes, J. Rodrigues, and A. Cunha, "CTCovid19: Automatic Covid-19 model for Computed Tomography Scans Using Deep Learning," Intell Based Med, vol. 11, p. 100190, 2025, doi: 10.1016/j.ibmed.2024.100190.
  - [50] S. Sultana, A. B. M. A. Hossain, and J. Alam, "COVID-19 detection from optimized features of breathing audio signals using explainable ensemble machine learning," Results in Control and Optimization, vol. 18, p. 100538, Mar. 2025, doi: 10.1016/j.rico.2025.100538.
  - [51] R. Rajpoot, S. Jain, V. B. Semwal, and D. Singh, "Quantitative Assessment of XAI Methods for COVID-19 Detection: A Comparative Approach," SN Comput Sci, vol. 6, no. 2, p. 122, Jan. 2025, doi: 10.1007/s42979-025-03663-5.
  - [52] N. P., J. Wekalao, A. N., and S. K. Patel, "Design and Analysis of a Plasmonic Metasurface-Based Graphene Sensor for Highly Sensitive and Label-Free Detection of COVID-19 Biomarkers," Plasmonics, Jul. 2024, doi: 10.1007/s11468-024-02442-x.
  - [53] C. J. Ejiyi et al., "ATEDU-NET: An Attention-Embedded Deep Unet for multi-disease diagnosis in chest X-ray images, breast ultrasound, and retina fundus," Comput Biol Med, vol. 186, p. 109708, Mar. 2025, doi: 10.1016/j.compbimed.2025.109708.