International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol. 13, No. 2, 2026, pp. 62-77.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Fusing Imaging and Clinical Data for COPD Analysis: A Multimodal DenseNet Approach with Grad-CAM

R. Hendra Kumar [a*], Gurram Sunitha [b]

[a] Research Scholar, Department of Computer Science and Engineering, School of Computing, Mohan Babu University, Tirupati, India. E-mail: 22202R010037@mbu.asia; hendra.sagar@gmail.com, Orcid: https://orcid.org/0009-0007-5154-7099*

[b] Department of Computer Science and Engineering, School of Computing, Mohan Babu University, Tirupati, India. E-mail: sunitha.g@mbu.asia; gurramsunitha@gmail.com, Orcid: https://orcid.org/0000-0002-3305-8167

Abstract

In the analysis and diagnosis of lung diseases through chest X-rays with deep learning models, the presence of bones and tissues, as well as burrs and blurs that may be present on the X-rays, can cause complications. COPD, or chronic obstructive pulmonary disease, is a respiratory disease that progresses steadily. It has been a challenging disease to classify and grade. It is important that the severity and grading of the disease are done correctly and on time. This paper aims to present a method that can classify the severity of COPD using MDenseNet-Grad-CAM. This method combines Multimodal DenseNet and Gradient-weighted Class Activation Mapping. It combines X-rays with clinical parameters that can indicate the progression of the disease. This combination can help the model learn better and provide a more holistic approach. It also incorporates Grad-CAM for better interpretability. This method can provide a better approach than other models that only rely on X-rays or data. It has a better classification accuracy than the reference models. It has an accuracy rate of 98.55%. Ablation studies confirmed the contribution of each component. DenseNet alone achieved 87.55% accuracy. Adding clinical data increased accuracy to 92.11%. Integrating Grad-CAM without clinical data reached 93.65%. Combining DenseNet, clinical data, and Grad-CAM produced the best accuracy of 98.55%. Findings confirm its value for early COPD diagnosis and severity classification. The method demonstrates strong potential for clinical decision support and patient care applications.

Keywords: Chronic Obstructive Pulmonary Disease, Gradient-Weighted Class Activation Mapping, Model Interpretability, Multimodal Densenet, Severity Classification.

DOI: 10.71086/IAJSE/V13I2/IAJSE1354
Received: 24 February 2026/Revised: 03 April 2026/Accepted: 07 May 2026/Published: 30 June 2026

Introduction

Pulmonary disease refers to any condition that impacts lung function. These disorders are common health concerns and occur in smokers and non-smokers of all ages and genders. Thoracic radiography, also known as chest X-ray (CXR), is a projection-based thoracic imaging technique. It is used for the diagnosis of lung and cardiovascular diseases. It also helps detect disorders affecting the rib cage or the spinal column. In addition, it identifies cancers that originate in the chest region or spread from other organs.

Medical professionals prefer thoracic radiography due to its ability to produce high-quality diagnostic images. Its main role is to monitor chest-related symptoms, detect lung abnormalities, and assess the patient's condition. It is also used to evaluate the effectiveness of treatment interventions. Thoracic radiography is widely available, non-invasive, and more affordable than many advanced imaging techniques. These characteristics make it an important tool in pulmonary disease diagnosis and management.

Thoracic radiographs can reveal various conditions, including heart enlargement, pneumonia, bronchitis, pleurisy, nodules, consolidations, infiltrations, elevated costal angles, and cavities. However, radiologists often find it challenging to classify thoracic illnesses using thoracic radiography due to the overlap of these characteristics. To save time and money, address a shortage of specialists, and help make an accurate diagnosis, it may be necessary in some difficult situations to supplement a radiographic examination with additional, more reliable and sophisticated diagnostic procedures (AlEisa et al., 2022). Computer-assisted diagnosis (CAD) systems assist doctors and radiologists in reading radiographs. These systems aim to detect diseases through automated analysis. Current CAD methods lack the required reliability for confirming pathological conditions in radiographs (Guefrechi et al., 2021). They are used only as visual aids to support decision-making.

Deep Learning (DL) methods show high performance in image classification and feature extraction (Ben Jabra et al., 2021). Several advanced deep networks achieve strong accuracy in these tasks. Researchers have adapted such networks for medical image classification. Findings indicate that deep networks capture features that distinguish different disease categories in medical images (Waleed et al., 2022; Bennour et al., 2024). Recent DL advancements have improved disease detection in medical imaging, particularly in lung infection diagnosis.

This research proposes a Multimodal DenseNet with Gradient-weighted Class Activation Mapping (Grad-CAM) for Chronic Obstructive Pulmonary Disease (COPD) severity classification. DenseNet uses dense connections to improve feature propagation and gradient flow. However, the use of multimodal data combines all these elements, providing a more complete image of the patient’s condition, including COPD.

This helps to analyze the condition in different ways. The Grad-CAM algorithm also helps to understand the areas of the image that are affecting the predictions the most, providing a better explanation of the predictions made by the models. This helps to make predictions regarding the severity of COPD with high accuracy while providing reliable explanations, thus making the predictions more reliable through the use of DenseNet, which helps to extract features from the different types of input data provided.

The end result is to create models for predicting the severity of COPD that are accurate and interpretable. The use of the multimodal framework also helps to use chest X-rays, including the clinical data provided, such as the patient’s age, smoking history, and spirometry results, to help detect both the visual signs and medical indicators of the progression of the disease. The use of clinical data helps the models learn more and provide a complete assessment of the patient’s condition regarding COPD.

The paper is organized in the following way: The introduction provides an overview of COPD and the issues with diagnosing and grading the disease, which has led to the introduction of the proposed methodology of combining Multimodal DenseNet with Grad-CAM to understand the disease better in terms of its severity. The literature review examines available techniques in the classification of the severity of COPD and its limitations. The offered methodology section presents the way of adding the DenseNet and Grad-CAM to the clinical data in order to achieve better classification performance. The experimental findings and discussions incorporate the descriptions of datasets, the performance of the model, and the comparison with the existing models, which provide the benefits of the proposed approach. The paper ends with a conclusion summarizing the findings, limitations, and recommendations for future research.

Literature Survey

An examination of GOLD and STAR categories for COPD staging in clinical practice was performed (Nishimura et al., 2024). Data from an ambulatory cohort study included 141 COPD patients between 2015 and 2023. GOLD 1–4 and STAR 1–4 groups were compared for mortality and COPD-related health status. Seventy-five patients (53.2%) had matching GOLD and STAR severity ratings. Using linear weights, Bangdiwala's B value was 0.775. Maximum follow-up was 95 months, with a median of fifty-four months. Twenty-nine participants (20.5%) had confirmed deaths.

A dataset with 148 COPD patients and 279 subjects with normal lung function was used (Mou et al., 2024). VCap, a quantitative analysis method, was proposed. Gramian Angular Field (GAF) translation converted volumetric capnograms into two-dimensional greyscale images. The CapnoNet model was used for feature extraction and classification. Two data augmentation methods improved model performance. F1-score, precision, and recall were 95.45%, 95.21%, and 95.70%. Diagnosis accuracy for COPD was 95.83%.

A multi-class COPD severity diagnostic system was designed (Khanaghavalle et al., 2024). Feature extraction included spectrogram, Mel spectrogram, and chromogram analysis. Pre-processing and augmentation methods were applied. The training was done with RESNET50 model. The system was highly sensitive in categorizing COPD levels of severity. Results supported the significance of sound-based prognoses in the diagnosis of COPD at an early stage.

The deep learning model with lung sounds was suggested (Sahu et al., 2024). Existing methods of determining the severity of COPD e.g. spirometry are intrusive and time consuming. The model involved statistical analysis of the lung sound samples in the form of early detection and classification of the severity. Attributes were obtained in the Open Respiratory Sound Database and the Respiratory@TR data. These were Chroma Short-Term Fourier Transform (CSTFT), Mel-Frequency Cepstral Coefficients (MFCCs), Mel-Spectrogram (mSpec), Melspectrogram-snippet and integrated features.

The data on twelve-channel lung sounds obtained as a part of Respiratory Database@TR were applied to package severe instances of COPD (Roy & Satija, 2022). The significant technique in feature extraction and classification was Variational Mode Decomposition (VMD).

The researchers took recourse to blood examination, testing of ventilation functionality, and other clinical facts to come up with predictions on the severity of the disease. The participants in the study consisted of 141 patients with COPD and they were divided into mild-to-moderate and severe groups (Guo, 2023). The baseline data were analyzed using univariate and multivariate logistic regression analysis and a nomogram created to predict severe COPD with the help of the significant factors.

The other hybrid approach was aimed at predicting the values of pulmonary functions (Zhou et al., 2022). They employed data on volumetric capnography and added an error correction layer with an improved application of KNN, regression layer with a mixture of extreme gradient boosting and support vectors and a sequence regression with one-dimensional convolutional neural networks.

Nevertheless, the extent of the disease is a major issue to be classified correctly despite the significant progress made. The complexity of the disease process is not reflected adequately with the help of spirometry and symptom scores. The majority of approaches do not involve biomarkers, deep learning, and individual patient data, and this is one of the gaps in application of multidimensional models, and more specifically, the application of advanced technologies to enhance the accuracy and usefulness of the approaches.

The literature review in this paper identifies the past research on the severity classification of COPD through deep learning and medical imaging. Although there are a lot of other techniques that use single modalities such as X-ray and clinical information, they are usually not accurate and readable. Not many methods involve both modalities, and this makes these methods not very effective in offering a holistic evaluation. In addition, the available models are not transparent in the decisions they make, thus preventing their usage in a clinical setting. The paper closes these gaps by combining MDenseNet, which is used to extract features, and Grad-Cam, which is used to provide visual explanations; it is a more precise and explainable approach than previously applied.

Proposed Methodology

In this segment, a novel learning model, MDenseNet-Grad-CAM, is proposed to evaluate the severity of COPD by integrating Multimodal DenseNet and Grad-CAM, as represented in fig.1. In this proposed model, X-ray images and clinical data, including age, smoking history, and spirometry, are integrated to improve the understanding of visual

and clinical features of COPD progression. DenseNet, a deep learning architecture that has been reported to be effective in feature extraction with dense connections, is utilized to extract features from X-ray images. Clinical data is integrated to refine the learning process of the proposed model to provide a comprehensive evaluation of COPD severity. To provide a clear understanding of the proposed model, Grad-CAM is integrated to highlight key features of X-ray images that influence COPD severity in different levels.

MDenseNet-Grad-CAM Network Architecture & Preprocessing

The MDenseNet-Grad-CAM is a hybrid of DenseNet and Grad-CAM, which improves COPD severity classification. DenseNet, with its many layers of connections, enhances feature propagation and vanishing gradient, thus it is best suited for medical image classification. This network has multiple dense blocks and then downsampling transition layers.

Grad-CAM is employed to achieve interpretability through the generation of heatmaps that identify the most important parts of the chest X-ray images, which play the biggest part in the severity classification. Multimodal data fusion is also used in the architecture, consisting of integrating clinical information (e.g., age, smoking history, spirometry results) with X-ray characteristics to give a multidimensional assessment of COPD progression. The classification of severity is projected by a final output layer through a softmax activation.

Prior to resizing the input photos, the Median Filter (MF) technique removes noise (Femy et al., 2024). Pixel windowing's MF notation is expressed in equation (1).

$$MF\big(j(a,b)\big) = Median\big(j(a+r,b+t)\big) \qquad (1)$$

Here, $j(a,b)$ is the value of a pixel.

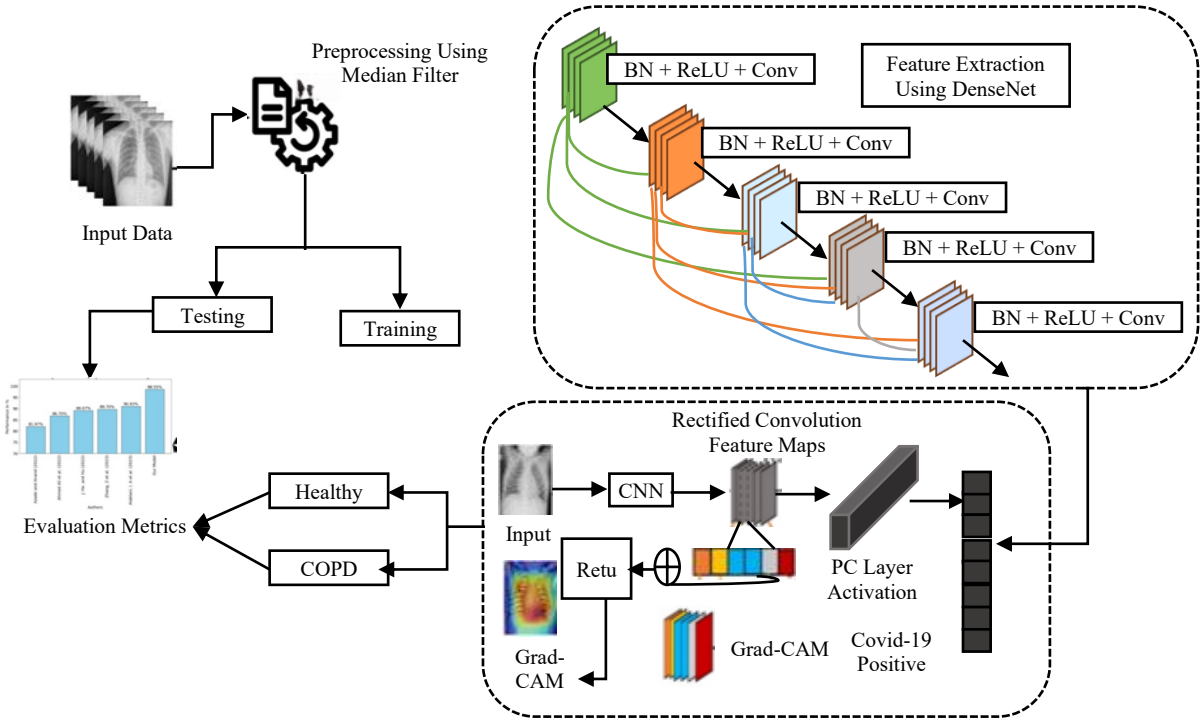

Fig. 1: Proposed MDenseNet-Grad-CAM Learning Model for COPD Severity Detection

Feature Extraction for DenseNet

Early detection and monitoring of COPD is important to control the progression of the disease and improve patient outcomes. This is where Multimodal DenseNet comes into play, a highly efficient deep learning technology that can be used to interpret different medical data, such as radiological images, medical records, and physiological data, to detect COPD at early stages and classify the disease. DenseNet is particularly useful for this task because the densely connected structure of the network facilitates the learning of a diverse set of features, as illustrated in fig. 2. By using

different data modalities, DenseNet utilizes the advantages of each data type to make accurate predictions for COPD diagnosis and prognosis (Yin et al., 2022).

In the context of the Multimodal DenseNet for COPD, there are some important equations that describe the working mechanism of the DenseNet framework. The output of the container of DenseNet is given by equation (2):

$$\mathrm{Y} = \mathrm{f}(\mathrm{H}_1, \mathrm{H}_2, \ldots, \mathrm{H}_\mathrm{k}) \qquad (2)$$

where $H_1, H_2, \ldots, H_k$ *are the inputs from prior layers or modalities*. This high level of interconnectedness enables all layers to have access to inputs from all previous layers, thereby increasing feature extraction. In multimodal fusion, when data from multiple sources such as clinical records, radiology, and physiological signals is fused together, typically a fusion strategy of early or late fusion is used.

Early fusion involves merging features from different modalities before passing them into the DenseNet. Mathematically, this can be expressed as in equation (3).

$$\mathrm{X}_{\mathrm{fused}} = \mathrm{concat}(\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_\mathrm{m}) \qquad (3)$$

where $X_1, X_2, \ldots, X_m$ represents features from the different modalities, and X_{fused} is the fused feature vector that is then passed through the DenseNet.

For the prediction of COPD severity, a final output layer with a softmax activation function will be employed to predict the class label *c* as shown in equation (4).

$$\mathrm{P}(\mathrm{c}|\mathrm{X}_{\mathrm{fused}}) = \frac{\mathrm{e}^{\mathrm{w}_\mathrm{c}^\mathrm{T}\mathrm{X}_{\mathrm{fused}}+\mathrm{b}_\mathrm{c}}}{\sum_{\mathrm{c}'=1}^{\mathrm{C}} \mathrm{e}^{\mathrm{w}_{\mathrm{c}'}^\mathrm{T}\mathrm{X}_{\mathrm{fused}}+\mathrm{b}_{\mathrm{c}'}}} \qquad (4)$$

Equation (4) represents the probability for each class based on the fused multimodal feature vector X_{fused}. Categorical cross-entropy loss function is used for continuous evaluation of the DenseNet model, as given by equation (5).

$$\mathrm{L} = -\sum_{\mathrm{c}=1}^{\mathrm{C}} \mathrm{y}_\mathrm{c} \log\big(\mathrm{P}(\mathrm{c}|\mathrm{X}_{\mathrm{fused}})\big) \qquad (5)$$

Where $\boldsymbol{y_c}$ is the ground truth label for the class c, and equation (4) is the forecast likelihood for class c. This loss function penalizes incorrect predictions and guides the model to minimize the error across all classes during training.

Grad-CAM

Grad-CAM is a technique that helps visualize and interpret the parts of an image that are crucial to the classification task (Fig. 3). It functions by utilizing the gradients of the target class score in connection with the activations of a CNN feature map (Tsai & Lee, 2025). Grad-CAM creates a CAM, which highlights the most significant areas of the image for a specific class prediction, by calculating these gradients and averaging them across the spatial dimensions.

Grad-CAM computes average gradients to find the most important feature maps and then uses those weights to create a visual map of where the model is focusing its attention, as shown in equation (6). This gradient shows how the score for class *c* changes in response to small variations in activation of the feature map A^k.

$$\frac{\partial y^c}{\partial A^k} \qquad (6)$$

To measure significance for each feature map, gradients are averaged over all spatial locations *(i, j)* in the map. This produces $\alpha^c{}_k$ for each feature map, reflecting its contribution to the target class as in equation (7).

$$\alpha_\mathrm{k}^\mathrm{c} = \frac{1}{\mathrm{Z}} \sum_{\mathrm{i,j}} \frac{\partial \mathrm{y}^\mathrm{c}}{\partial \mathrm{A}_{\mathrm{ij}}^\mathrm{k}} \qquad (7)$$

Here, *Z* signifies the total spatial locations *(height X weight)* in the feature map, whereas A_{ij}^k is activation at spatial position (i, j).

CAM for the class *c* is then generated by a weighted sum of feature maps A^k using computed weights α_k^c as in equation (8).

$$\mathrm{CAM}^\mathrm{c} = \sum_\mathrm{k} \alpha_\mathrm{k}^\mathrm{c} \mathrm{A}^\mathrm{k} \qquad (8)$$

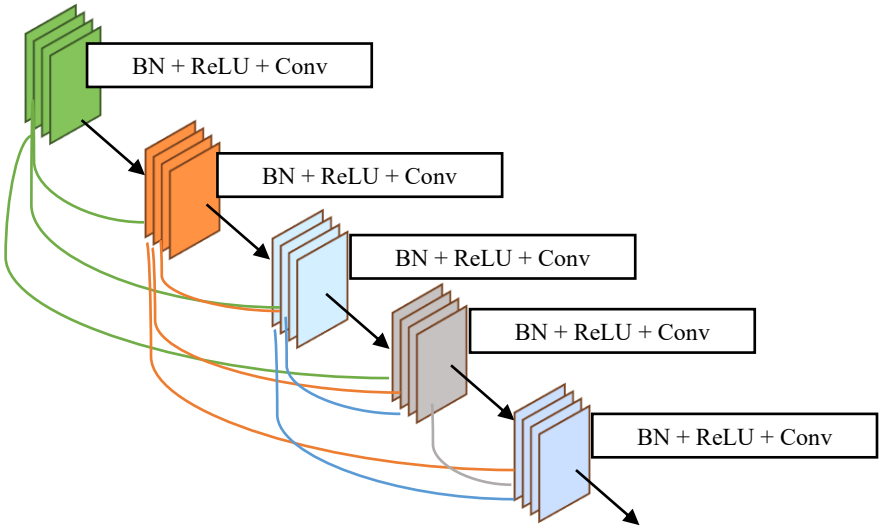

Fig. 2: Architecture of DenseNet Learning Model

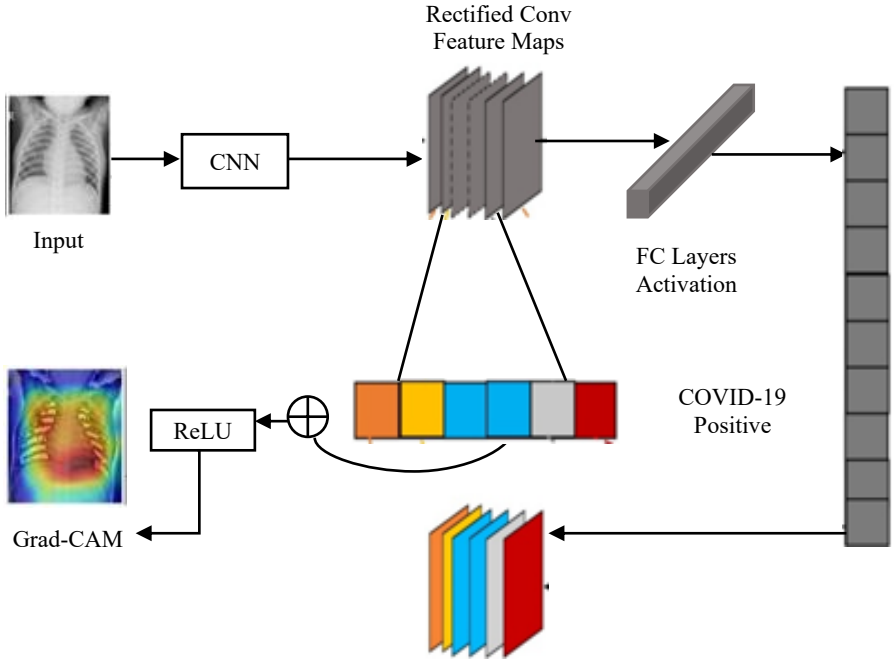

Fig. 3: Architecture of Grad-CAM

This high level of interconnectedness enables all layers to have access to inputs from all previous layers, thereby increasing feature extraction. In multimodal fusion, when data from multiple sources such as clinical records, radiology, and physiological signals is fused together, typically a fusion strategy of early or late fusion is used equation 9.

$$Re\,L\,U(CAM^c) = max(0, CAM^c) \qquad (9)$$

- CAM^c: Class Activation Map (CAM) for class c.
- ReL_U: ReLU (Rectified Linear Unit) activation.
- max $(0, CAM^c)$: Keeps positive values of $CAM^c$and sets negative values to zero.

It applies ReLU to the CAM, keeping only positive activations.

In order to understand the impact of the image that is fed into the model, the gradients are calculated with respect to the image. This is because the class score is sensitive to different parts of the image, as given by equation (10).

$$\frac{\partial y^c}{\partial X} \tag{10}$$

- y^c: The predicted score for the class c.
- X: The input image or feature map.
- $\frac{\partial y^c}{\partial X}$: The gradient showing how changes in the input image X affect the predicted score for the class c, highlighting which image regions influence the model's decision.

Equation (11) calculates a weighted sum of activations at each spatial point to determine the contribution of each feature map to the class score.

$$CAM_{ij}^c = \sum_k \alpha_k^c A_{ij}^k \tag{11}$$

- CAM_{ij}^c: CAM value at pixel $(i, j)$for class c.
- α_k^c: Weight for feature map $k$for class c.
- A_{ij}^k: Activation value at pixel (i, j)in the feature map k.
- Σ_k: Summing over all feature maps to compute CAM.

It combines weighted feature map activations to highlight important image regions for the class $\boldsymbol{c}$.

As shown in equation (12), CAM is normalized to the range [0,1] by dividing by its greatest value.

$$\hat{CAM}^c = \frac{CAM^c}{max(CAM^c)} \tag{12}$$

- $\boldsymbol{CAM^c}$: Original Class Activation Map (CAM) for class c.
- $\mathbf{max}$ $(\boldsymbol{CAM^c})$: Maximum value in the CAM for class c.
- $\widehat{\boldsymbol{CAM^c}}$: Normalized CAM, scaled between 0 and 1 for class c.

ReLU activation and normalization are applied to create a heatmap. Equation (13) provides an expression for this.

$$\mathrm{Heatmap} = \mathrm{ReLU}\left(\widehat{\mathrm{CAM}}^c\right) \tag{13}$$

The output from equation (13) indicates the important spatial locations that affect the class decision made by the model. It can also be superimposed on the input image and hence acts as a visualization tool that shows where the model is attending. This can be very important, especially in deep learning models for applications such as medical imaging and object recognition.

The learning model proposed incorporates the advantages of both DenseNet and Grad-CAM. DenseNet can effectively utilize feature propagation and reuse while handling complex multimodal cues for COPD severity classification. Grad-CAM can effectively identify the important locations within medical images and help medical practitioners identify disease-specific patterns. With the ability to identify the lung locations affected by COPD, the disease can be better localized. The proposed model can help avoid overfitting during COPD classification.

Pseudocode for MDenseNet-Grad-CAM Model

Step 1: Load and Preprocess Data

- Load X-ray images and clinical data (age, smoking history, spirometry).
- Normalize and resize images.
- Apply Median Filter to remove noise.

Step 2: Feature Extraction with DenseNet

- Initialize DenseNet with specified layers and growth rate.
- Apply convolution and ReLU activation to extract features.

- Use transition layers for downsampling.

Step 3: Data Fusion

- Combine DenseNet features with clinical data (age, smoking, spirometry).
- Pass the fused data through DenseNet layers.

Step 4: Apply Grad-CAM for Visualization

- Calculate gradients of the target class score with respect to the last convolutional layer.
- Average gradients to generate weights.
- Create Class Activation Map (CAM) and normalize.
- Superimpose the CAM on the X-ray image for visualization.

Step 5: Make Prediction

- Use Softmax activation to predict COPD severity class.

Step 6: Evaluate Model

- Calculate accuracy, precision, recall, F1-score.
- Perform ablation study to assess contributions of DenseNet, clinical data, and Grad-CAM.

MDenseNet-Grad-CAM model is a machine that is trained to diagnose the severity of COPD using both X-ray images and clinical data (age, smoking history, and spirometry outcomes). The model will involve the use of DenseNet to extract the features, use of convolutional layers of ReLU activation and incorporate clinical data into the process by refining the learning process. Grad-CAM gives interpretability through heatmaps which indicate areas in the X-ray images that are important in the prediction. The model works as predictions with the help of a Softmax function and is tested on the accuracy, precision, recall, and F1-score, and an ablation study is conducted to determine the effect of each issue.

Experimental Results and Discussion

Dataset Description

The data that is being used in this research paper comprises 6,751 Pulmonary Function Test (PFT) images, of which include 1,222 testing images, 755 validation images, and 4,774 training images Pulmonary Chest X-Ray Abnormalities, (www.kaggle.com). Such images are utilized to train, validate, and test the MDenseNet-Grad-CAM model of COPD severity classification. The learning of the model also relies on the training set; the validation set is employed to improve the hyperparameters, and the testing set is employed to examine the performance of the model properly. The data set will enhance the accuracy and understanding of the model in relation to predicting the COPD severity. Fig. 4 shows a few of the sample images of the PFT dataset.

Environmental Setup

Experiments in the paper use an Intel Core i9-10900K processor running at 3.70 GHz, with 64 GB of RAM and an RTX 3080 GPU. In the paper, the following methods are used: CoroNet (Khan et al., 2020), CVDNet (Ouchicha et al., 2020), CapsNet (Toraman et al., 2020), CNN (Sarki et al., 2022).

The software to be used in this work comprises Python to implement the MDenseNet-Grad-CAM model as well as significant libraries, including TensorFlow and Keras to develop the deep learning models. Data is manipulated and pre-processed using NumPy and Pandas and the results are visualized using Matplotlib and seaborn. The model is trained and tested on a machine that has the Intel Core i9-10900K processor, 64 GB of RAM and RTX 3080 graphics cards and this is the minimum computing power to properly train and make inferences.

Parameter Initialization

The table 1 below is a summary of the five important hyperparameters of the MDenseNet-Grad-CAM model. The parameters are learning rate, batch size, DenseNet layers, growth rate, and dropout rate all of which are crucial to optimize the model to classify COPD severity.

Table 1: Key Parameter Settings for MDenseNet-Grad-CAM

Parameter	Value/Initialization
Learning Rate (η)	0.001
Batch Size	64
DenseNet Layers	3 layers
Growth Rate	32
Dropout Rate	0.2

Results and Discussion

The table 2 and Fig. 5(a-e) reveal the comparative performance in relation to the five methods. MDenseNet-Grad-CAM leads the pack having the highest figures in all the performance measures which include: accuracy of 98.55 %, recall rate of 92.66 %, and precision of 91.66%, F1-score of 93.33%, and 92.91% specificity. In comparison, CNN has impressive performance metrics, including an accuracy of 95.87%, 87.33% precision, and 91.44% recall. CapsNet is also very competitive, with an accuracy of 93.44% and recall of 89.33%. However, its precision is lower, at 70.44%.

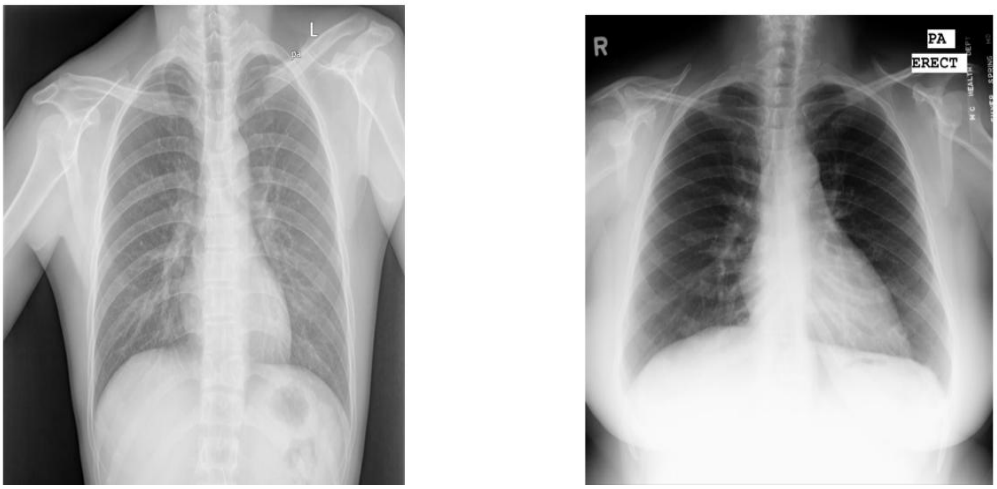

Fig. 4: Sample of the Dataset

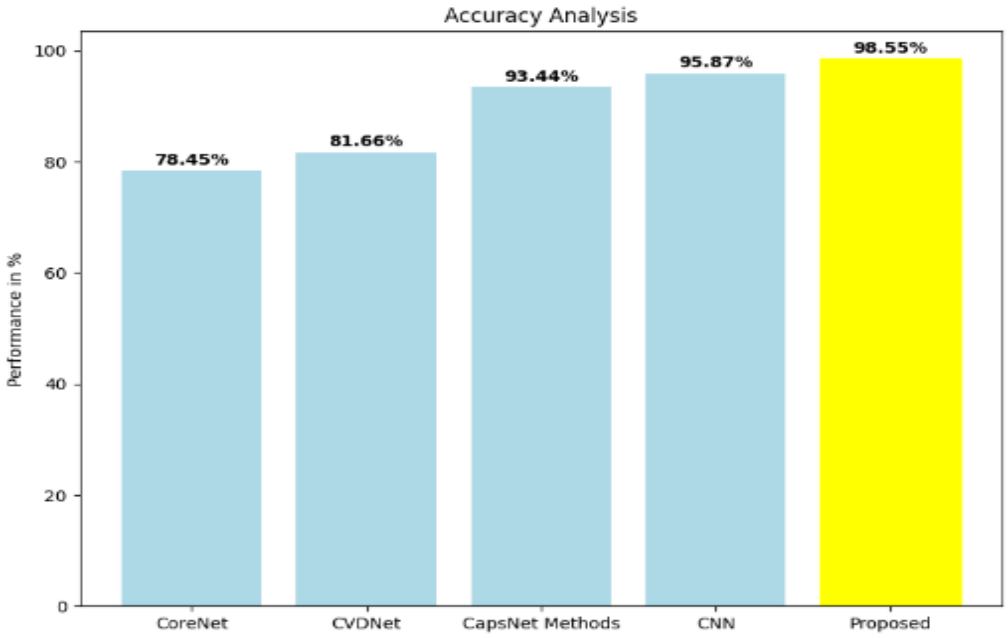

Fig. 5(a): Accuracy Analysis Comparison of Different Models

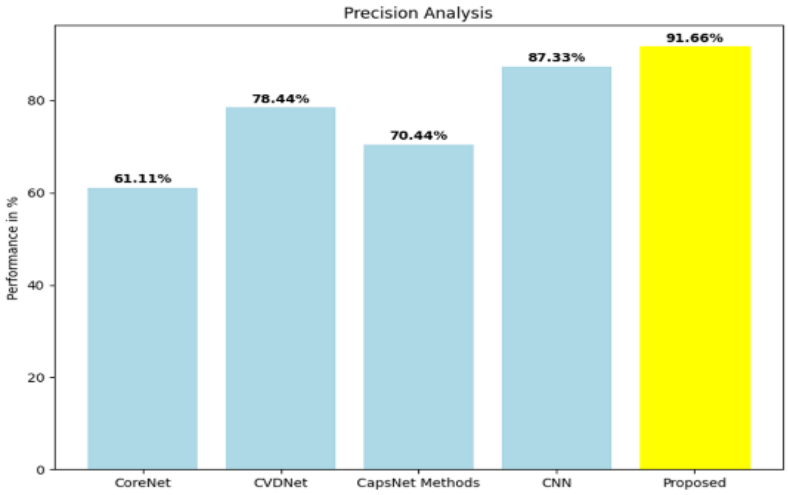

Fig. 5(b): Precision Analysis Comparison of Different Models

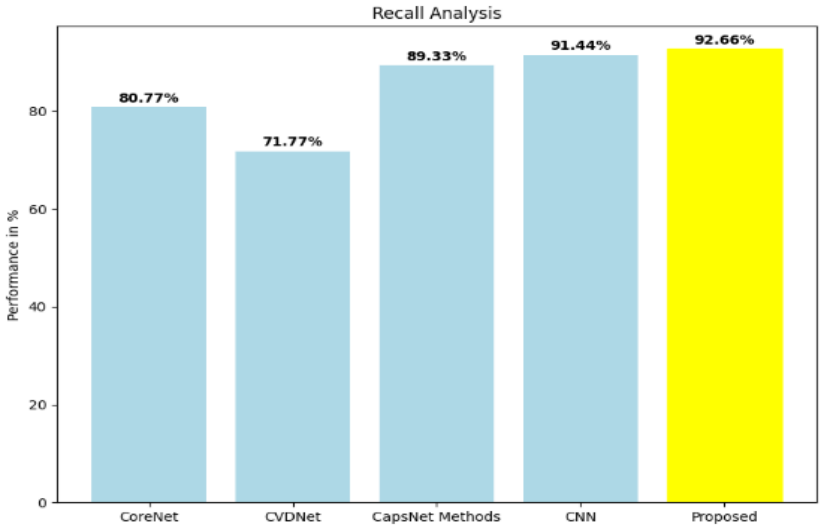

Fig. 5(c): Recall Analysis Comparison of Different Models

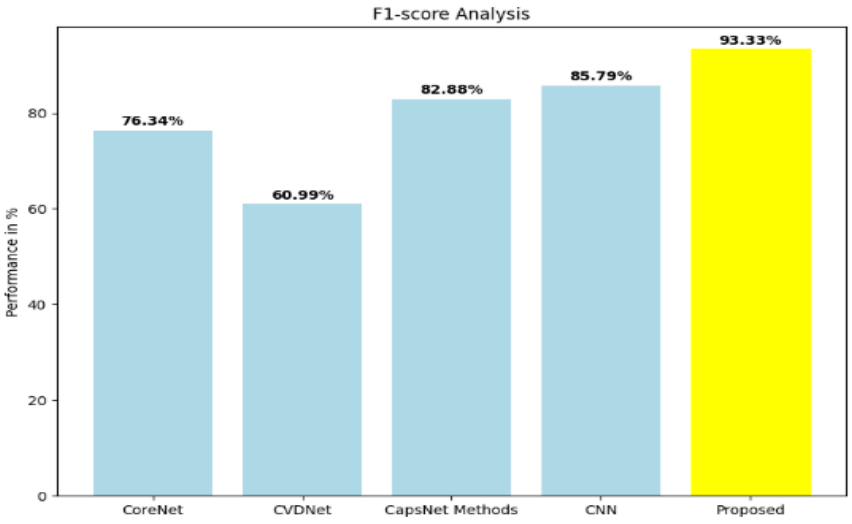

Fig. 5(d): F1-score Analysis Comparison of Different Models

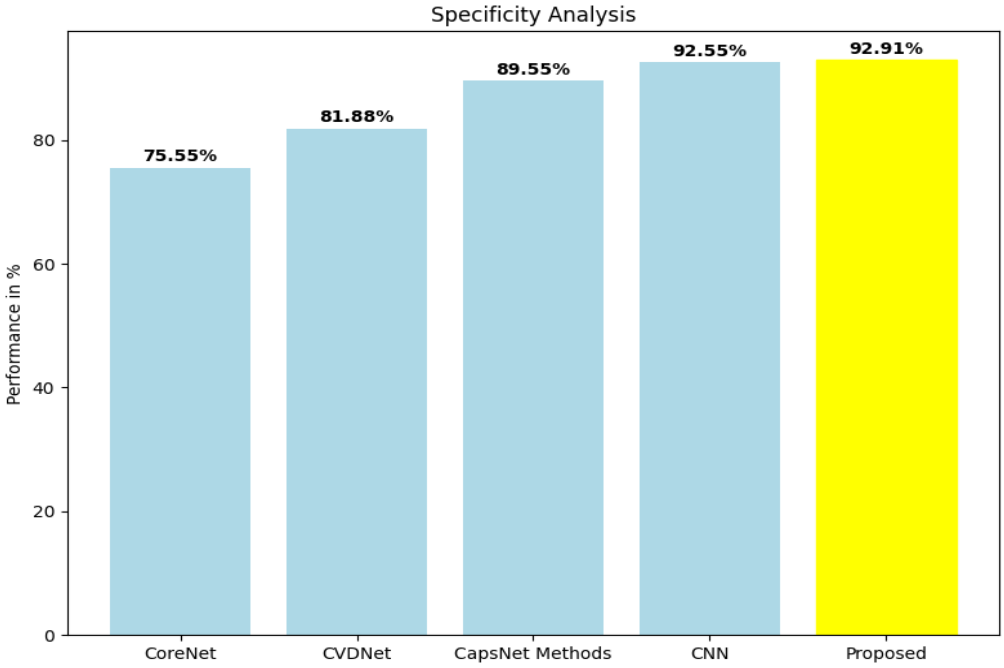

Fig. 5(e): Specificity Analysis Comparison of Different Models

Table 2: A Comparison of MDenseNet-Grad-CAM

Learning Model	Accuracy	Precision	Recall	F1-score	Specificity
CoroNet	78.45	61.11	80.77	76.34	75.55
CVDNet	81.66	78.44	71.77	60.99	81.88
CapsNet	93.44	70.44	89.33	82.88	89.55
CNN	95.87	87.33	91.44	85.79	92.55
MDenseNet-Grad-CAM	98.55	91.66	92.66	93.33	92.91

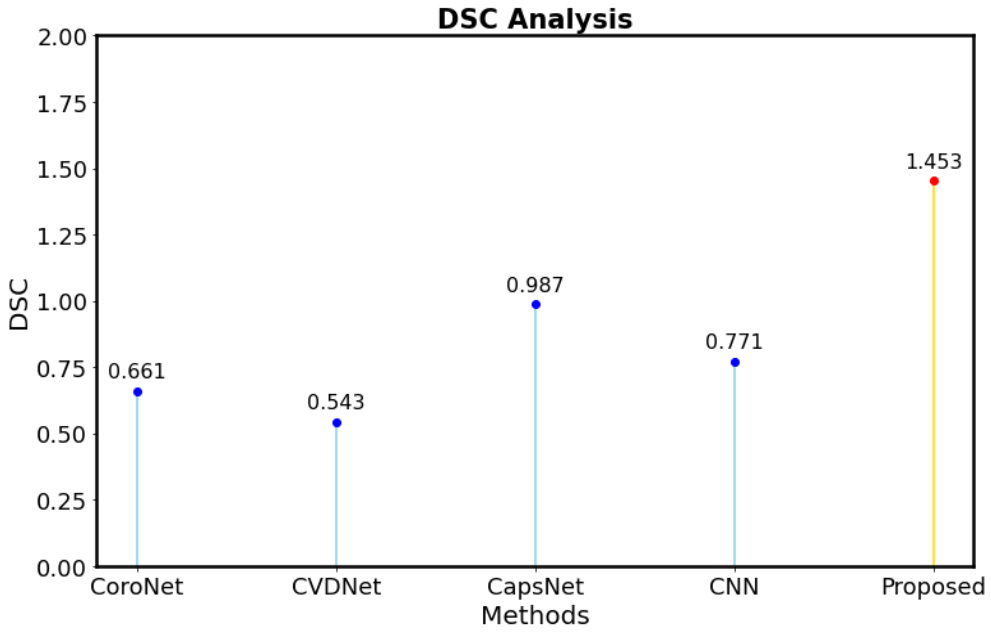

Fig. 6: DSC Analysis for MDenseNet-Grad-CAM

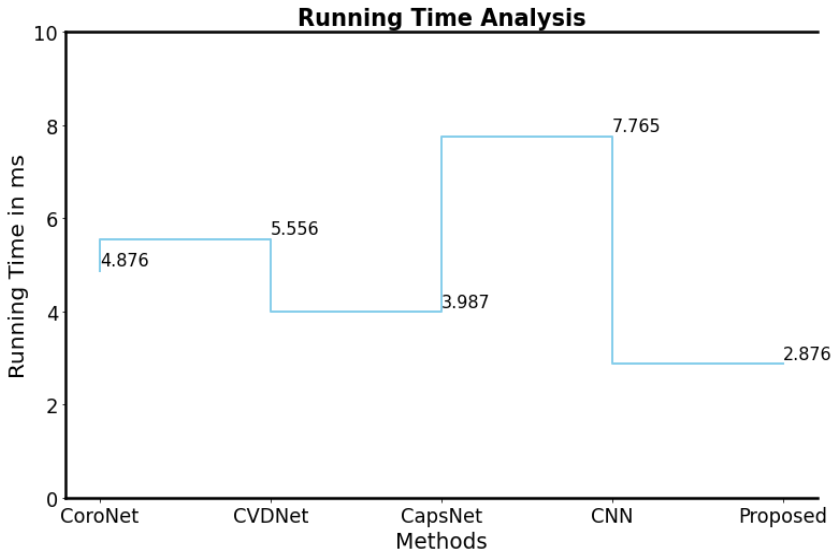

Fig. 7: Run Time Evaluation of MDenseNet-Grad-CAM

MDenseNet-Grad-CAM demonstrates a good balance between specificity, with a figure of 81.88%, and precision, with a figure of 78.44%. The recall is also quite good, at 71.77%. In comparison, CoroNet maintains a good recall, 80.77%, and an F1 score of 76.34%. However, its accuracy is lower, at 78.45%. In conclusion, the results demonstrate the advantage of MDenseNet-Grad-CAM over other methods, which is evident through its high precision, recall, and accuracy.

In fig. 6 demonstrates the comparison of the Dice Similarity Coefficient (DSC) between different models. MDenseNet-Grad-CAM has the highest DSC of 1.453, which is quite higher than the other models. CapsNet has a good DSC of 0.987, while CNN has a moderate DSC of 0.771. In comparison, CoroNet and CVDNet have lower DSCs of 0.661 and 0.543, respectively. This presents the advantage of the proposed method over other methods, which is evident through its higher DSC.

In fig. 7 demonstrates the run-time analysis of different models. MDenseNet-Grad-CAM has the least run time, at 2.876 ms, which demonstrates its advantage over other methods through its faster run time. CapsNet has the second least run time, at 3.987 ms, followed by CoroNet and CVDNet, with run times of 4.876 ms and 5.556 ms, respectively. CNN has the highest run time, at 7.765 ms

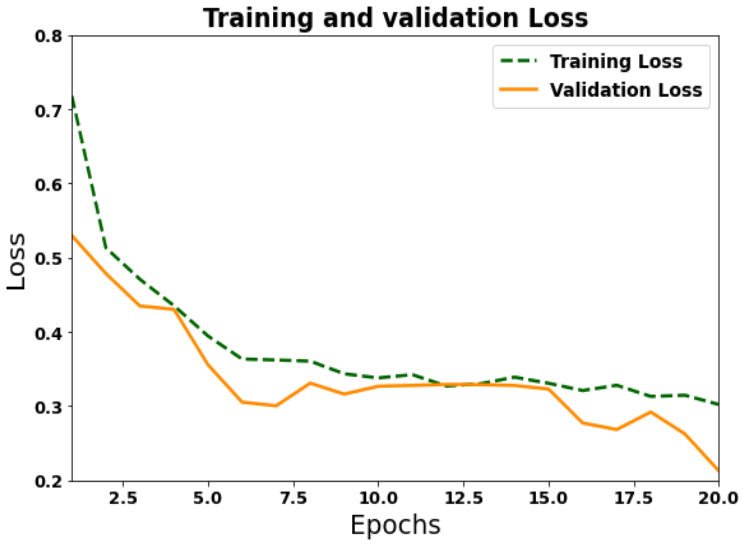

Fig. 8(a): Training and Validation Loss During Training

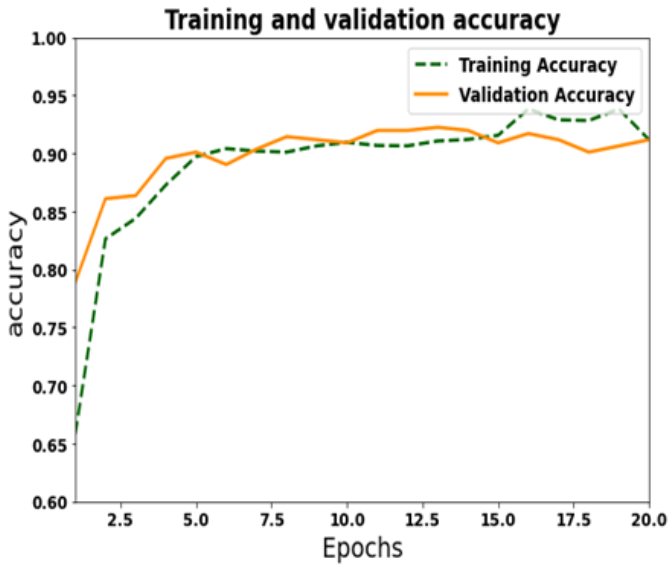

Fig. 8(b): Training and Validation Accuracy During Training

The fig. 8(a & b) demonstrates training and validation loss and the accuracy of the model in the classification of COPD. The overall decrease in the number of losses is a sign of successful learning. The distance between training and validation loss is small indicating high generalization to unknown data. The model is very accurate both during the training and validation stages. This demonstrates good consistency in terms of COPD severity classification. The high percentage of loss and accuracy indicate that the model is stable and can be used in medical diagnostic processes.

The comparison between the performance parameters, DSC, and running time proves the excellence of the proposed method compared to the current ones. The highest accuracy (98.55%), precision (91.66%), recall (92.66%), F1-score (93.33%), and specificity (92.91) are obtained by MDenseNet-Grad-CAM, and it is much higher than the approaches used to compare it [17-20]. Moreover, it has the highest Dice Similarity Coefficient (DSC) of 1.453, which means that it has a high level of presentation of the segmentation. The proposed method is also most efficient in terms of computational efficiency with the minimum running time of 2.876 ms. The proposed technique could provide a balanced solution, which is associated with Toraman et al. with the high DSC of 0.987 and the prolonged running time (3.987 seconds) and with Sarki et al. that provides high accuracy (95.87) yet the longest running time (7.765 seconds). The results prove that the proposed method can be a good choice in the real world with the possibility to offer high accuracy, segmentation precision, and computing efficiency.

Ablation Study

This section presents the results obtained from the ablation tests on MDenseNet-Grad-CAM, comparing the proposed method with the current models CoroNet, CVDNet, CapsNet, and CNN. The development and validation of the MDenseNet-Grad-CAM are achieved by introducing new features of the proposed method gradually. Table 3 illustrates that the ablation study highlights the incremental contributions of various components to the total model presentation. Starting with DenseNet alone, an accuracy of 87.55% was achieved. Incorporating clinical data significantly enhanced the accuracy to 92.11%, showcasing the value of multimodal data integration. In addition, Grad-CAM to DenseNet further enhanced interpretability and presentation, yielding an accuracy of 93.65%. Lastly, the integration of DenseNet, clinical data, and Grad-CAM in the proposed method had the highest accuracy of 98.55, which determined the synergistic effect of the integration of the elements to favorable predictive presentation.

Influence of Grad-CAM

Grad-CAM plays a major role in MDenseNet-Grad-CAM interpretability, which mainly occurs in CNNs. Grad-CAM provides clear insights into the forecast process by generating heatmaps that illustrate the regions of an image that most affect the model's selection. Transparency is essential in medical image analysis, and this technique helps to grasp the inner workings of a complex method. It allows the visualization of important features, creating a model that is additionally interpretable and trustworthy for practitioners, leading to informed decisions and better validation of model predictions. Finally, the DenseNet joined with the Grad-CAM technique achieved an impressive presentation

of 98.55% accuracy for input data. In comparison, the existing models CoroNet, CVDNet, CapsNet, and CNN achieved accuracy presentations of 78.45%, 81.66%, 93.44%, and 95.87%, correspondingly.

Table 3: Experimental Results of the Ablation Study on MDenseNet-Grad-CAM

Method	Accuracy (%)
DenseNet	87.55
DenseNet + Clinical Data	92.11
DenseNet + Grad-CAM	93.65
DenseNet + Clinical Data + Grad-CAM	98.55

Table 4: Comparative Analysis of MDenseNet-Grad-CAM With the Benchmark Models

Learning Model	Learning Model	Accuracy (%)
Ali et al., (2022)	KNN	86.70
Azade & Anand, (2022)	CNN	81.97
He & Hu, (2022)	CNN	89.07
Alablani & Alenazi, (2023)	COVID-ConvNet	90.93
Zhang et al., (2023)	MCENet	89.70
Proposed Model	**MDenseNet-Grad-CAM**	98.55

The table 4 shows a comparative analysis of the MDenseNet-Grad-CAM accuracy against state-of-the-art methods. The findings indicate that the proposed DenseNet-Grad-CAM model is much more effective than other methods, with an accuracy of 98.55, the highest of all the compared models. It is important to note that the models like COVID-ConvNet (90.93) and MCENet (89.70) were good but could not achieve the high accuracy that was demonstrated by MDenseNet-Grad-CAM. The previous approaches, such as CNN (81.97% by Azade and Anand, 2022, and 89.07% by J.), were also studied (89.07% by J. The accuracy of KNN (86.70% by Ahmed Ali et al., 2022) and He and Hu, 2022) was lower. This analogy highlights the effectiveness and the strength of the proposed method in the analysis of accuracy.

Although the Multimodal DenseNet with Grad-CAM is very accurate in the classification of COPD severity, it is associated with a set of limitations and difficulties. Such a model can also be biased in case the population of patients is not sufficiently represented as the effectiveness of such a model relies on the quality of input data. Besides, the computational needs of the DenseNet and the Grad-CAM model can also be an issue in under-resourced environments, e.g., rural clinics. The interpretation of the outcomes of the Grad-CAM can also be a matter of expertise, which might be an issue with non-experts. The models must overcome all these issues in order to be applicable in reality.

Conclusions

In summary, the Multimodal DenseNet, along with Grad-CAM, proves to be a powerful tool for the classification of COPD severity. By using the combination of the Multimodal DenseNet along with Grad-CAM, the system is able to achieve 98.55% accuracy, 91.66% precision, 92.66% recall, 93.33% F1-score, and 92.91% specificity. The Grad-CAM results show clear highlights of the regions that are being used to make the final decision, thereby increasing the precision of the diagnosis and the planning of the treatment. To prove the significance of the DenseNet, the Grad-CAM, and the clinical data, the ablation study shows that DenseNet achieves 87.55% accuracy on its own, the addition of the clinical data increases the accuracy to 92.11%, the Grad-CAM alone achieves 93.65% accuracy, and the combination of DenseNet, Grad-CAM, and the clinical data achieves the highest accuracy of 98.55%. These results are further validated by the benchmark results, where the KNN achieves 86.70%, CNN achieves 81.97%, CNN achieves 89.07%, the COVID-ConvNet achieves 90.93%, the MCENet achieves 89.70%, while the MDenseNet-Grad-CAM achieves the highest accuracy of 98.55%, thereby proving the power of the predictive model while at the same time increasing the interpretability of the results.

Despite promising results of the proposed MDenseNet-Grad-CAM model, it has certain limitations. The model can be biased in that the patient group used in the model is not a representative sample of the entire population, and as such, this could have an impact on the generalizability of the model. Moreover, both the computational needs of DenseNet and Grad-Cam can be a problem in resource-limited settings, such as in rural clinics. Further research will be carried out to increase the size of the dataset to cover more diverse patient populations, make the model more resource-efficient, and consider more different types of data, including biomarkers, to enhance the performance and resilience of the model.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

References

AlEisa, H. N., Touiti, W., Ali ALHussan, A., Ben Aoun, N., Ejbali, R., Zaied, M., & Saadia, A. (2022). Breast cancer classification using FCN and beta wavelet autoencoder. *Computational Intelligence and Neuroscience*, *2022*(1), 8044887. https://doi.org/10.1155/2022/8044887

Guefrechi, S., Jabra, M. B., Ammar, A., Koubaa, A., & Hamam, H. (2021). Deep learning-based detection of COVID-19 from chest X-ray images. *Multimedia tools and applications*, *80*(21), 31803-31820. https://doi.org/10.1007/s11042-021-11192-5

Ben Jabra, M., Koubaa, A., Benjdira, B., Ammar, A., & Hamam, H. (2021). COVID-19 diagnosis in chest X-rays using deep learning and majority voting. *Applied Sciences*, *11*(6), 2884. https://doi.org/10.3390/app11062884

Waleed, J., Azar, A. T., Albawi, S., Al-Azzawi, W. K., Ibraheem, I. K., Alkhayyat, A., ... & Kamal, N. A. (2022). An effective deep learning model to discriminate coronavirus disease from typical pneumonia. *International Journal of Service Science, Management, Engineering, and Technology (IJSSMET)*, *13*(1), 1-16. https://doi.org/10.4018/IJSSMET.313175

Bennour, A., Aoun, N. B., Khalaf, O. I., Ghabban, F., Wong, W. K., & Algburi, S. (2024). Contribution to pulmonary diseases diagnostic from X-ray images using innovative deep learning models. *Heliyon*, *10*(9). https://doi.org/10.1016/j.heliyon.2024.e30308

Nishimura, K., Kusunose, M., Shibayama, A., & Nakayasu, K. (2024). Comparison of disease severity classifications of chronic obstructive pulmonary disease: GOLD vs. STAR in clinical practice. *Diagnostics*, *14*(6), 646. https://doi.org/10.3390/diagnostics14060646

Mou, X., Wang, P., Sun, J., Chen, X., Du, L., Zhan, Q., ... & Fang, Z. (2024). A novel approach for the detection and severity grading of chronic obstructive pulmonary disease based on transformed volumetric capnography. *Bioengineering*, *11*(6), 530. https://doi.org/10.3390/bioengineering11060530

Khanaghavalle, G. R., Rahul, G., Senajith, S. R., Vishnuvasan, T. S., & Keerthana, S. (2024, April). Chronic obstructive pulmonary disease severity classification using lung sound. In *2024 10th international conference on communication and signal processing (ICCSP)* (pp. 428-432). IEEE. https://doi.org/10.1109/ICCSP60870.2024.10543344

Sahu, P., Kumar, S., & Behera, A. K. (2024, July). Soundnet: Leveraging deep learning for the severity classification of chronic obstructive pulmonary disease based on lung sound analysis. In *2024 IEEE international conference on electronics, computing and communication technologies (CONECCT)* (pp. 1-6). IEEE. https://doi.org/10.1109/CONECCT62155.2024.10677193

Roy, A., & Satija, U. (2022, November). Automated severity detection of chronic obstructive pulmonary disease using lung sounds. In *2022 IEEE 19th India council international conference (INDICON)* (pp. 1-6). IEEE. https://doi.org/10.1109/INDICON56171.2022.10040031

Guo, S. (2023). Logistic regression analysis of clinical characteristics for differentiation of chronic obstructive pulmonary disease severity. *Emergency Medicine International*, *2023*(1), 5945191. https://doi.org/10.1155/2023/5945191

Zhou, R., Wang, P., Li, Y., Mou, X., Zhao, Z., Chen, X., ... & Fang, Z. (2022). Prediction of pulmonary function parameters based on a combination algorithm. *Bioengineering*, *9*(4), 136. https://doi.org/10.3390/bioengineering9040136

Femy, F. F., Nallasivan, G., Basarikodi, K., Natarajan, C., Bhuvanesh, A., & Swapna, T. (2024). A system to classify chronic obstructive pulmonary disease using pre-trained-densenet201 with tsa. *Measurement: Sensors*, *33*, 101234. https://doi.org/10.1016/j.measen.2024.101234

Yin, L., Hong, P., Zheng, G., Chen, H., & Deng, W. (2022). A novel image recognition method based on DenseNet and DPRN. *Applied Sciences*, *12*(9), 4232. https://doi.org/10.3390/app12094232

Tsai, C. M., & Lee, J. D. (2025). Dynamic ensemble learning with gradient-weighted class activation mapping for enhanced gastrointestinal disease classification. *Electronics*, *14*(2), 305. https://doi.org/10.3390/electronics14020305

Pulmonary Chest X-Ray Abnormalities, Kaggle Dataset. https://www.kaggle.com/datasets/kmader/pulmonary-chest-xray-abnormalities

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

Ouchicha, C., Ammor, O., & Meknassi, M. (2020). CVDNet: A novel deep learning architecture for detection of coronavirus (Covid-19) from chest x-ray images. *Chaos, Solitons & Fractals*, *140*, 110245. https://doi.org/10.1016/j.chaos.2020.110245

Toraman, S., Alakus, T. B., & Turkoglu, I. (2020). Convolutional capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks. *Chaos, Solitons & Fractals*, *140*, 110122. https://doi.org/10.1016/j.chaos.2020.110122

Sarki, R., Ahmed, K., Wang, H., Zhang, Y., & Wang, K. (2022). Automated detection of COVID-19 through convolutional neural network using chest x-ray images. *Plos one*, *17*(1), e0262052. https://doi.org/10.1371/journal.pone.0262052

Azade, A., & Anand, K. M. (2022, February). Impact of image augmentation in COVID-19 detection using chest X-ray images. In *2022 IEEE Delhi Section Conference (DELCON)* (pp. 1-5). IEEE. https://doi.org/10.1109/DELCON54057.2022.9752785

Ali, H. A., Hariri, W., Zghal, N. S., & Aissa, D. B. (2022, May). A comparison of machine learning methods for best accuracy covid-19 diagnosis using chest x-ray images. In *2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)* (pp. 349-355). IEEE. https://doi.org/10.1109/SETIT54465.2022.9875477

He, A.J., & Hu, H. (2022, March). Identifying Covid-19 Chest X-Rays by Image-Based Deep Learning. In *Proceedings of the 2022 7th International Conference on Machine Learning Technologies* (pp. 112-118). https://doi.org/10.1145/3529399.3529418

Zhang, Z., Ma, Y., & Li, K. (2023, February). An improved X-ray image diagnosis method for COVID-19 pneumonia on a lightweight neural network embedded device. In *Proceedings of the 2023 3rd International Conference on Bioinformatics and Intelligent Computing* (pp. 351-357). https://doi.org/10.1145/3592686.3592749

Alablani, I. A., & Alenazi, M. J. (2023). COVID-ConvNet: A convolutional neural network classifier for diagnosing COVID-19 infection. *Diagnostics*, *13*(10), 1675. https://doi.org/10.3390/diagnostics13101675