Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-19755-y

Transparent and trustworthy interpretation of COVID-19 features in chest X-rays using explainable AI

Shakti Kinger[1] · **Vrushali Kulkarni**[1]

Received: 12 February 2024 / Revised: 14 May 2024 / Accepted: 23 June 2024

Abstract
The potential of AI-based disease prediction models for assessing COVID-19 patients outperforms conventional methods. However, their black-box nature has limited their applicability. This study explores the approach for COVID-19 identification by integrating Artificial Intelligence (AI) methods with Deep Neural Networks (DNNs), such as EfficientNet and DenseNet, applied for medical imaging. To address the black-box challenge, the study incorporates eXplainable AI (XAI) techniques, including LIME, Grad-CAM, and a novel variant of Grad-CAM++, termed "Modified Grad-CAM++". The modification aims to enhance explanations for COVID-19 predictions, improving the interpretability and transparency of the model's decision-making process. Collaborating with expert radiologists, the study validates the Modified Grad-CAM++ using a separate dataset of radiologist-validated chest X-ray (CXR) images. The Integrated Uncertainty Calculation (IUC) metric is introduced as an evaluation measure for the Modified Grad-CAM++. Remarkably, both EfficientNet and DenseNet achieve high diagnostic precision, with accuracy rates of 98% and 97%, respectively. The integration of XAI algorithms enhances the interpretability and transparency of predictions, ensuring clinical relevance and validity. These precise AI models offer valuable support to healthcare professionals in decision-making and resource allocation.

Keywords COVID-19 diagnosis · Deep neural networks · eXplainable Artificial Intelligence (XAI) · Interpretability · Medical imaging · EfficientNet · DenseNet · LIME · Grad-CAM · Grad-CAM++ · Modified Grad-CAM++

1 Introduction

In recent years, Computer-Aided Diagnostic (CAD) technologies and medication research have both benefited from the application of artificial intelligence (AI) methods aimed at replicating human neural activity and judgment. These methods hold promise for generating new data to enhance clinical outcomes and patient care [15, 37, 38, 51, 58, 66, 76, 77].

✉ Shakti Kinger
shakti.kinger@mitwpu.edu.in

[1] Computer Engineering & Technology, School of Computer Science & Engineering, Dr. Vishwanath Karad MIT World Peace University, Pune, India

DNNs serve diverse purposes, including image recognition, natural language processing, and sensor data analysis, marking a significant stride in deep learning's evolution [41, 49, 78]. These developments have led to considerable advancement in the field of deep learning [36].

The urgent need for accurate, efficient, and rapid COVID-19 diagnosis, especially in overwhelmed healthcare systems during the pandemic, underscores the significance of employing advanced technologies like DNNs. These networks offer a powerful tool for automatic feature extraction and classification, enabling the differentiation of COVID-19 from other respiratory diseases based on CXR images. Leveraging DNNs makes it possible to analyze a large volume of CXR images quickly, aiding in the early detection and subsequent containment of COVID-19, particularly in economically disadvantaged and rural areas where access to advanced medical facilities and equipment may be limited.

AI-assisted solutions have the potential to aid general practitioners in triaging patients, especially in resource-constrained settings where confirmatory testing or skilled radiologists may be lacking. By identifying crucial chest regions for automated diagnosis, these solutions can offer valuable support.

While DNNs excel in both single and multi-modal tasks [47], comprehending their decision-making process proves challenging due to their opaque, black-box nature. This challenge is particularly pronounced in high-stakes circumstances where failure could have serious repercussions. It can be difficult to trust the results of these models due to their lack of interpretability, especially when they are used for high-stakes decisions such as financial or medical ones [1]. To address this issue, the development of more transparent and understandable AI methods may be necessary. These methods would enable greater human monitoring and involvement in the decision-making process, fostering trust and reliability in AI-driven systems [25, 53, 54, 67]. Researchers have been exploring ways to make DNNs more transparent and understandable to address this issue, including the development of techniques that offer valuable insights into the model's workings. Enhancing the explainability of DNNs can encourage their adoption in a range of applications and instill user confidence in their use [17, 40].

Many medical applications, including the diagnosis and treatment of COVID-19, have benefited significantly from the deployment of artificial intelligence (AI) and deep learning techniques. DNNs have shown promising results in detecting COVID-19 from medical photos.

AI is revolutionizing various aspects of combating COVID-19. Medical imaging analysis involves using AI algorithms to detect COVID-19 pneumonia from chest X-rays and CT scans, aiding radiologists in making accurate diagnoses. Symptom analysis tools utilize AI to assess the likelihood of infection, helping prioritize testing. Predictive analytics models forecast the spread of the virus, guiding resource allocation and public health interventions. Genomic analysis tracks the virus's evolution and identifies drug targets. AI-driven drug discovery expedites the search for COVID-19 treatments. Natural language processing (NLP) extracts valuable insights from vast amounts of textual electronic health records, clinical notes, and scientific literature. AI-driven remote monitoring solutions enable real-time tracking of patient health status, reducing the burden on healthcare facilities.

A team of researchers conducted an investigation into the use of chest X-rays for automatic COVID-19 detection, recognizing the potential of AI and deep learning in diagnostic imaging. Given the substantial amount of training data required for deeplearning-based solutions, academics have begun to merge multiple datasets to expand their knowledge base and improve the efficacy of AI-driven diagnostic tools [26, 44, 46, 61, 69].

Challenges arise in training accurate AI models due to limited, incomplete, and heterogeneous data coupled with regular updates to datasets due to the rapid evolution of the pandemic. The black-box nature of deep learning models hampers interpretability and transparency, necessitating explanations for AI-generated predictions. Ensuring diagnostic accuracy and reliability entails generalizing AI models to new variants and addressing ethical and regulatory considerations. Moreover, resource constraints and integration challenges into clinical workflows impede widespread adoption. Effective assessment of AI models' performance requires rigorous validation and standardized benchmarks. Achieving seamless integration necessitates collaboration among AI developers, healthcare providers, and IT professionals to optimize workflow processes.

We utilized the COVID-QU-Ex collection, one of the largest available datasets for COVID-19 CXR images, to train our models. This dataset consists of 33,920 total lung images, including approximately 11,956 COVID-19 images, 11,263 Non-COVID (Pneumonia disease) images, and 10,701 Normal CXR images.

For our analysis, we applied advanced deep neural network (DNN) detection models such as DenseNet121 and EfficientNetB7, known for their superior performance in medical imaging tasks.

Furthermore, we introduced an enhanced version of the Grad-CAM++ algorithm, termed Modified Grad-CAM++, to provide better explanations. To assess its effectiveness, we proposed the Integrated Uncertainty Calculation (IUC) metric, which we used alongside established eXplainable Artificial Intelligence (XAI) methods such as Grad-CAM, and Grad-CAM++. This allowed us to compare the performance of Modified Grad-CAM++ with these existing techniques.

Our research contributes significantly to several key areas:

1. **Preprocessing and Image Segmentation**: We establish a robust preprocessing pipeline for COVID-19 medical images to ensure high-quality data for analysis. Moreover, we employ advanced segmentation algorithms to precisely define regions of interest and extract relevant data, thus enhancing the performance of our models.
2. **Collection of Radiologists' Marked Datasets**: We curate a large dataset of COVID-19 medical images annotated by qualified radiologists. This dataset serves as a valuable resource for developing and testing our models, affirming the quality and reliability of our findings.
3. **DenseNet and EfficientNet Implementation**: We evaluate the performance of state-of-the-art deep learning architectures, specifically DenseNet and EfficientNet, for COVID-19 identification. By leveraging these powerful models, we generate accurate and reliable predictions for both binary and multi-class scenarios.
4. **Proposed Modifted Grad-CAM++ Explainer**: We propose and implement modified explanation techniques to further illuminate the model's logic and improve interpretability. We compare the interpretability of our explainer with existing explainers like LIME, GradCAM, and GradCAM++ by applying them to highlight the decision-making process of AI model's for COVID-19 detection.
5. **Results Validation Using Radiologist-Marked Data**: We validate the effectiveness of our proposed Modified Grad-CAM++ explainer by comparing their explanation with annotations provided by radiologists. The Integrated Uncertainty Calculation (IUC) metric is introduced as an evaluation measure for the Modified Grad-CAM++. This validation stage underscores the potential of our models as reliable tools for COVID-19 diagnosis, ensuring alignment with professional evaluations.

Our study contributes to the development of reliable and understandable COVID-19 detection systems by integrating image preprocessing, dataset curation, sophisticated deep learning architectures, explainability methodologies, and validation against expert annotations. These contributions represent significant advancements in medical imaging analysis and offer invaluable support to medical practitioners in accurately diagnosing COVID-19.

2 Literature Review

2.1 Explainable AI(XAI)

Explainable Artificial Intelligence (XAI) is a rapidly evolving field aimed at enhancing our comprehension of AI systems and their inner workings [68] Its significance is particularly pronounced in industries such as healthcare, finance, and transportation, where AI-driven decision-making profoundly impacts individuals' lives [48]. By employing XAI techniques, predictive accuracy can be improved, and trust in AI system outputs can be bolstered [21, 42, 64].

Moreover, the "Right to explanation" mandated by the European Union's General Data Protection Regulation (GDPR) has elevated the importance of explainability to a legal imperative, ensuring compliance with regulatory frameworks [18]. Explainability encompasses developers' strategies to elucidate the workings of an AI system, while interpretability focuses on end-users' ability to comprehend and interpret algorithmic outputs [22].

Integrating explainability into various stages of AI system development, from pre-modeling to post-implementation, is paramount for fostering responsible AI practices. Understanding the rationales behind AI model successes or failures enables iterative improvements and refinement using explainable methodologies [3].

This paper [35] introduces an explainable AI model for detecting lung disease from chest X-ray images. It not only accurately identifies abnormalities but also provides transparent explanations for its decisions, enhancing trust and clinical utility.

XAI techniques aim to make the decisions and processes of AI systems more understandable and interpretable to humans. Some of the commonly used XAI techniques [34] that help increase the transparency of AI systems include.

- Feature Importance Techniques, like SHAP (SHapley Additive exPlanations), help identify which features or regions in the chest X-ray are most influential in predicting lung diseases. They can highlight specific abnormalities such as opacities, nodules, or consolidations, providing clinicians with valuable insights into the diagnostic process.
- Local explanations techniques, like LIME (Local Interpretable Model-agnostic Explanations), help highlight relevant regions or features in the image providing the model's reasoning and hence gaining confidence in its decision-making process.
- Attention Mechanisms, like Guided Grad-CAM, highlight pathological features of CXR images in the form of attention map visualizations aiding in the interpretation of pathological abnormalities or patterns within the image.
- Counterfactual explanations help understand changes in the chest X-ray affecting the model's predictions. For example, the presence or absence of certain abnormalities/features and its impact on the model's prediction help gain insights into the model's robustness.

Overall, XAI techniques help increase transparency and trust in AI systems for CXR-based lung disease predictions by providing a better understanding of pathological symptoms like.

1. Presence of abnormalities such as nodules, infiltrates, masses, opacities, consolidations, or pleural effusions.
2. Visibility of anatomical structures such as the lungs, bronchi, trachea, diaphragm, ribs, or cardiac silhouette on CXR image.
3. Patterns or distribution of pixel intensities like reticular, nodular or honeycomb patterns.
4. Symmetry and Asymmetry between the left and right lung or between different regions of the same lung.

2.2 AI for COVID-19

COVID-19, a highly contagious viral illness, has triggered widespread outbreaks globally, profoundly impacting public health, the economy, and society. Governments worldwide have implemented various measures to curb its spread, underscoring the importance of adhering to guidelines from health authorities to safeguard oneself and others [20].

According to the World Health Organization (WHO), as of July 1, 2022, there have been a total of 545,226,550 confirmed cases of COVID-19 and 6,334,728 fatalities reported globally. A large number of vaccine doses have also been distributed. However, the reopening of commercial operations and easing of social distancing measures have made it more difficult for countries to control the spread of the virus, particularly as testing kits remain scarce. The COVID-19 incubation period, during which an infected person may not show symptoms but can still spread the virus to others, can range from 5–6 days on average to as long as 14–21 days in some cases [30, 65]. COVID-19 is typically diagnosed using reverse transcription-polymerase chain reaction (RT-PCR) tests and chest X-rays. Although RT-PCR assays are the gold standard for detecting the virus, their sensitivity can be low; some studies have shown that it can be as low as 71% [14, 72]. The diagnosis of COVID-19 can also be made via chest X-rays and CT scans, which may reveal distinguishing features such as symmetrical lung involvement and anomalies in the pictures. However, these traits may not always be discernible with the unaided eye and can be challenging to identify from those of other types of pneumonia [13, 26]. Medical imaging can be a helpful tool for determining how COVID-19 affects the lungs, but it's crucial to use a variety of techniques for an appropriate diagnosis. X-rays are used in chest radiography (CXR) to create images of internal chest structures, such as the heart, lungs, and blood arteries. It should, however, be read in conjunction with additional clinical and diagnostic data [9, 26]. According to professionals in the industry, deep learning has the ability to develop an autonomous sickness classifier for radiography. Chest X-rays have been used to investigate the application of deep learning and artificial intelligence in the diagnosis of COVID-19. Using chest X-rays, convolutional neural networks (CNN) have been used to detect TB, pneumonia, right pleural effusion, cardiomegaly, aberrant mediastinum, and pneumonia [4, 7, 19, 29, 31, 39, 79]. The paper [43] presents an explainable AI framework for interpreting pulmonary diseases from chest radiographs. Utilizing deep learning techniques, such as CNNs and attention mechanisms, the framework accurately identifies abnormalities and provides transparent explanations for predictions. This approach aims to enhance trust and understanding among healthcare professionals, potentially improving diagnosis and patient care. The paper [28] introduces a hybrid DCNN-ViT-GRU model with explainable AI

to enhance the diagnosis of lung abnormalities. By integrating convolutional neural networks (DCNN), Vision Transformers (ViT), and Gated Recurrent Units (GRU), the model achieves better performance in detecting lung abnormalities from medical images.

Convolutional neural networks (CNNs) and other artificial intelligence (AI) techniques have been employed in a number of studies to create deep learning models for the detection of COVID-19 utilizing chest radiography (CXR) pictures. In one study [2], five pre-trained CNN models were tested on CXR images, with the VGG19 [59] and MobileNetv2 [55] models producing the highest accuracy scores of 0.93 and 0.92, respectively. Another study proposed a hybrid approach using CNNs for feature extraction and support vector machines (SVMs) for classification, achieving an accuracy of 0.95 on a test set of 50 COVID-19 images [16]. A two-stage strategy was also suggested, in which CXR patches were used to identify the outlines of the lungs and heart, which were then used for classification, with an accuracy of 88.9% and a sensitivity of 96.4% [45],The Xception network architecture was used in another study to develop a deep transfer learning-based method for COVID-19 identification [11] and the COVID-ResNet architecture, a pre-trained ResNet-50 model [57], was used in a different study for the detection of COVID-19 from chest X-rays through incremental downsizing of input images and fine-tuning of the network. The DenseNet-121, ResNet50, Inception-V3, InceptionResNetV2, Xception, and EfficientNet-B2 models were also used in a study to develop the DeepCOVID-XR composite of CNNs for binary prediction of COVID-19 versus non-COVID-19 cases, achieving an accuracy of 0.83 and an AUC of 0.90 on a test set of 300 images [71]. In another study, a modified version of the CheXNet deep learning model was used to create COVID-CXNet, which was then used to identify COVID-19-based pneumonia from chest radiographs [24]. This study presents an explainable AI model that uses deep learning methods like CNN and RNN as well as machine learning algorithms including random forests, SVM, logistic regression, and SVM. The model uses genetic data analysis to pinpoint certain genes or biomarkers linked to COVID-19, giving clear insights into the genetic factors affecting COVID-19 status. This method offers early disease detection and individualised treatment plans [74].

This [32] research presents a MobileNet-based CNN model with an innovative fine-tuning mechanism for detecting COVID-19 infections, combining the efficiency of MobileNet architecture with a novel fine-tuning approach. Experimental results confirm the model's effectiveness in accurately identifying COVID-19 cases from medical imaging data, suggesting its potential for real-world applications in healthcare settings.

3 Material and Methods

3.1 Setup Details

The DNN models presented in this research were trained using a robust computing setup. Hardware included an Intel(R) Xeon(R) Silver 4208 CPU @ 2.10 GHz, accompa- nied by an RTX A6000 GPU boasting 48 GB RAM, and a 1 TB SanDisk SSD G5B for efficient storage management. The system operated on Ubuntu 20.04 as the operating system. TensorFlow 2.9.2 served as the primary deep learning framework, supplemented by CUDA 11 for GPU acceleration. Keras, OpenCV, and seaborn were also utilized for model development, data manipulation, and visualization, respectively. This setup provided a powerful and versatile environment for training and evaluating complex DNN architectures, ensuring the reliability and reproducibility of our experimental results.

3.2 Proposed Framework

A comprehensive pipeline for COVID-19 diagnosis is proposed as shown in Fig. 1.

1. **Dataset Preparation:** To create a thorough dataset for COVID-19 diagnosis research, a diverse collection of chest X-ray images is meticulously collected and expertly annotated by collaborating with radiologists.
2. **Data Preprocessing:** To improve the quality of the acquired dataset and make it easier to analyze, preprocessing techniques like image scaling, normalization, and noise reduction are applied.
3. **Utilizing UNet for Data Segmentation:** To precisely identify COVID-19 related anomalies, the UNet model, a potent segmentation method, is used to effectively segment the lung regions in the chest X-ray images.
4. **DNN-based Multi-class Classiftcation:** To perform multi-class classification, a deep neural network model is trained on the segmented data to distinguish between COVID-19 positive, negative, and maybe other classes for thorough diagnosis.
5. **Use of the Interpretability Model for Justiftcations:** The classification decisions made by the DNN are explained using an interpretability model, such as Grad-CAM or LIME, giving details about the characteristics and regions that went into the diagnosis.
6. **Veriftcation of Explanations Using Radiologist-Annotated Pictures:** To ensure the accuracy and alignment of the interpretability model's explanations with professional knowledge, an iterative process is undertaken. This involves testing the model's explanations against radiologist-annotated images, comparing highlighted regions to radiologist-identified features, and iteratively refining the model, dataset, inputs (segmented/non-segmented) until consistency in the model's interpretability is achieved.
7. **Analysis and Evaluation:** The suggested approach enables comprehensive analysis and evaluation of the COVID-19 diagnosis pipeline, evaluating the precision, interpretability, and clinical applicability of the findings. The overall diagnostic procedure is improved and refined as a result of this analysis.

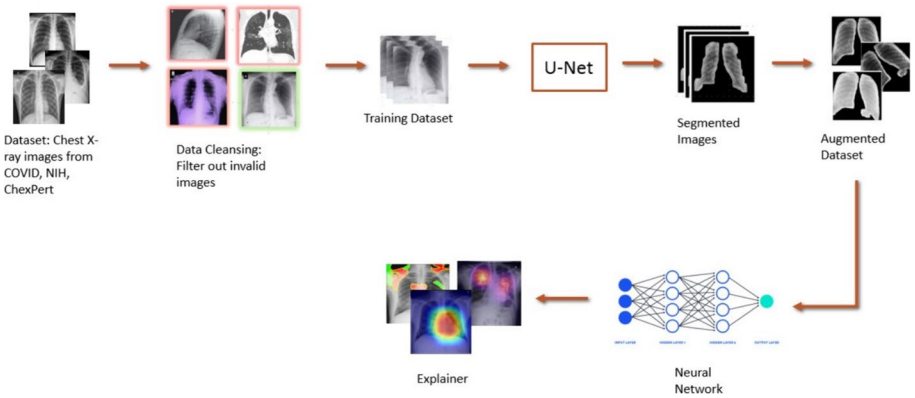

Fig. 1 Proposed Architecture for COVID-19 Detection

Our complete framework incorporates these processes in an effort to improve the precision, explicability, and dependability of COVID-19 diagnosis while offering insightful information to researchers and medical professionals fighting the epidemic.

3.2.1 Dataset Preparation

In this study, multiple datasets are employed for various purposes such as classification, segmentation, and weight initialization. The datasets utilize open data sources, as specified below, for classification. Several image data repositories were utilized to compile publicly accessible COVID-19 Chest X-ray pictures. Additionally, samples for both normal and pneumonia cases were extracted from the NIH chest X-ray dataset, which was also used in the Kaggle RSNA pneumonia detection competition. The authors provide information regarding the number of distinct samples in each class for the COVID-19 datasets, as some images were included in several datasets.

The COVID-QU-Ex collection Fig. 2, assembled by Qatar University researchers, contains 33,920 chest X-ray (CXR) pictures [9, 50, 62], including:

- 11,956 COVID-19 [9, 10, 12, 24, 70, 73, 75] CXR Images
- 11,263 viral or bacterial pneumonia illnesses that are not COVID [5, 33, 60].
- 10,701 Normal [5, 33, 60] (healthy) CXR images.

The authors of the above dataset collected images from various sources to tackle the class imbalance challenge. For example, to address the deficiency of 'Normal' class images, 1128 CXR normal images from the RSNA dataset were added. Similarly, 1228 COVID-19 images were replaced with Non-COVID (Pneumonia) class CXR images. This substitution ensured that COVID-19 cases were appropriately represented across the datasets while maintaining the integrity of the original normal class. By sourcing data from multiple datasets, each with its unique characteristics and demographics, authors aimed to capture a diverse range of cases, thereby mitigating the risk of bias towards specific demographic groups or disease severities.

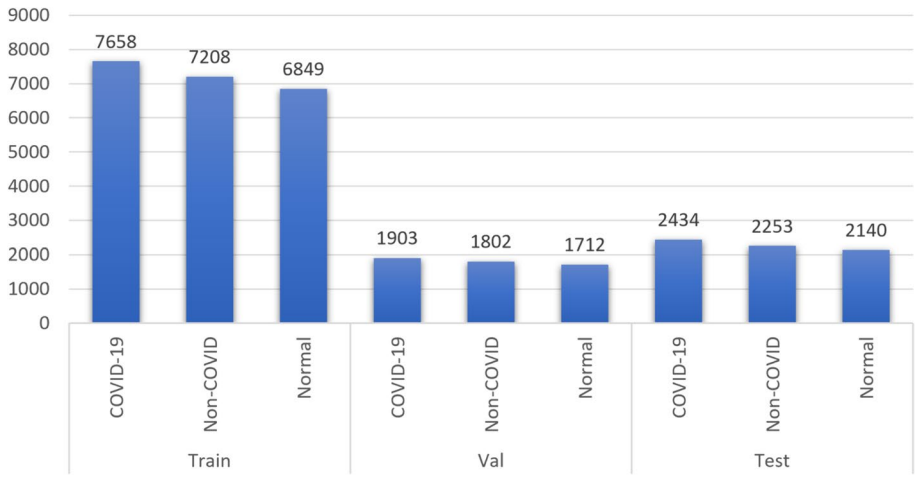

Fig. 2 Dataset Distribution for Training, Validation, and Testing

The full dataset includes ground-truth lung segmentation masks, making it a comprehensive resource for lung segmentation tasks. It is noted as one of the most comprehensive lung mask datasets ever produced.

Additionally, the test set comprises 39 radiologist-verified COVID-19 images. It's important to mention that no post-processing is performed on these images, ensuring the integrity and authenticity of the data for analysis and evaluation purposes.

The dataset was split into Training, Validation and Testing sets with the Training set including 7658 COVID-19, 7208 Non-COVID and 6849 Normal CXR images. Similarly, test set included 2434 COVID-19, 2253 Non-COVID and 2140 Normal CXR images. The Validation set included 1903 COVID-19, 1802 Non-COVID and 1712 Normal CXR images.

3.2.2 Data Preprocessing

The images were resized to dimensions of 224×224x3, significantly reducing the number of parameters to enhance the neural network's efficiency. Additionally, a thresholding method was employed to remove overly bright pixels that could potentially disrupt the image processing procedure. Any missing portions were filled in to maintain focus on the relevant visual characteristics, minimizing the impact of text annotations present in some images. Data enrichment techniques, including rotating training set images by up to 10 degrees, were utilized to enrich the training dataset and increase its diversity. These preprocessing and data augmentation techniques collectively improve the model's ability to generalize and effectively learn from the available data, ultimately enhancing the performance and robustness of COVID-19 classification.

3.2.3 Data Segmentation using UNet

In order to improve model precision, we narrowed down its focus exclusively to the lung regions. The model demonstrates increased precision in identifying specific patterns and anomalies associated with COVID-19 virus, consequently enhancing its accuracy. This targeted approach effectively eliminates unnecessary noise from non-lung areas, thereby improving the model's ability to detect subtle signs of infection. Furthermore, it optimizes computational resources by reducing the input size, thus enhancing the efficiency of the model. The accuracy of this approach heavily relies on the quality of the masking process. Improper execution of the masking process can lead to information loss, thereby impacting the effectiveness of the model.

The UNet model was selected based on specific criteria tailored to the requirements of our data segmentation task. The UNet architecture's simplicity and effectiveness facilitate the seamless integration of low-level features with high-level features, making it adept at handling segmentation tasks. Its ability to perform well with limited training data, preserve contextual information, and adapt to imbalanced datasets, in addition to its flexibility and comparative performance, make it the preferred choice over other segmentation models such as FCN or DeepLab.

Figure 3 provides tabular architecture diagram of UNet model. Each row represents a layer in the UNet architecture for image segmentation. Here's an explanation of the components and their functionalities.

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 224, 224, 3)]	0	[]
conv2d (Conv2D)	(None, 224, 224, 32)	896	['input_1[0][0]']
conv2d_1 (Conv2D)	(None, 224, 224, 32)	9248	['conv2d[0][0]']
max_pooling2d (MaxPooling2D)	(None, 112, 112, 32)	0	['conv2d_1[0][0]']
conv2d_2 (Conv2D)	(None, 112, 112, 64)	18496	['max_pooling2d[0][0]']
conv2d_3 (Conv2D)	(None, 112, 112, 64)	36928	['conv2d_2[0][0]']
max_pooling2d_1 (MaxPooling2D)	(None, 56, 56, 64)	0	['conv2d_3[0][0]']
conv2d_4 (Conv2D)	(None, 56, 56, 128)	73856	['max_pooling2d_1[0][0]']
conv2d_5 (Conv2D)	(None, 56, 56, 128)	147584	['conv2d_4[0][0]']
max_pooling2d_2 (MaxPooling2D)	(None, 28, 28, 128)	0	['conv2d_5[0][0]']
conv2d_6 (Conv2D)	(None, 28, 28, 256)	295168	['max_pooling2d_2[0][0]']
conv2d_7 (Conv2D)	(None, 28, 28, 256)	590080	['conv2d_6[0][0]']
max_pooling2d_3 (MaxPooling2D)	(None, 14, 14, 256)	0	['conv2d_7[0][0]']
conv2d_8 (Conv2D)	(None, 14, 14, 512)	1180160	['max_pooling2d_3[0][0]']
conv2d_9 (Conv2D)	(None, 14, 14, 512)	2359808	['conv2d_8[0][0]']
conv2d_transpose (Conv2DTranspose)	(None, 28, 28, 256)	524544	['conv2d_9[0][0]']
concatenate (Concatenate)	(None, 28, 28, 512)	0	['conv2d_transpose[0][0]', 'conv2d_7[0][0]']
conv2d_10 (Conv2D)	(None, 28, 28, 256)	1179904	['concatenate[0][0]']
conv2d_11 (Conv2D)	(None, 28, 28, 256)	590080	['conv2d_10[0][0]']
conv2d_transpose_1 (Conv2DTranspose)	(None, 56, 56, 128)	131200	['conv2d_11[0][0]']
concatenate_1 (Concatenate)	(None, 56, 56, 256)	0	['conv2d_transpose_1[0][0]', 'conv2d_5[0][0]']
conv2d_12 (Conv2D)	(None, 56, 56, 128)	295040	['concatenate_1[0][0]']
conv2d_13 (Conv2D)	(None, 56, 56, 128)	147584	['conv2d_12[0][0]']
conv2d_transpose_2 (Conv2DTranspose)	(None, 112, 112, 64)	32832	['conv2d_13[0][0]']
concatenate_2 (Concatenate)	(None, 112, 112, 128)	0	['conv2d_transpose_2[0][0]', 'conv2d_3[0][0]']
conv2d_14 (Conv2D)	(None, 112, 112, 64)	73792	['concatenate_2[0][0]']
conv2d_15 (Conv2D)	(None, 112, 112, 64)	36928	['conv2d_14[0][0]']
conv2d_transpose_3 (Conv2DTranspose)	(None, 224, 224, 32)	8224	['conv2d_15[0][0]']
concatenate_3 (Concatenate)	(None, 224, 224, 64)	0	['conv2d_transpose_3[0][0]', 'conv2d_1[0][0]']
conv2d_16 (Conv2D)	(None, 224, 224, 32)	18464	['concatenate_3[0][0]']
conv2d_17 (Conv2D)	(None, 224, 224, 32)	9248	['conv2d_16[0][0]']
conv2d_18 (Conv2D)	(None, 224, 224, 3)	99	['conv2d_17[0][0]']

Total params: 7760163 (29.60 MB)
Trainable params: 7760163 (29.60 MB)
Non-trainable params: 0 (0.00 Byte)

Fig. 3 UNet Architecture

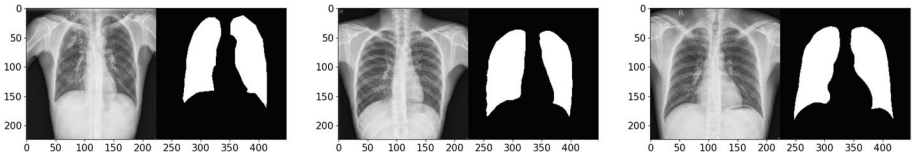

Fig. 4 Input Image and Mask to train UNet model

1. Input Layer (input 1): This is the input layer where the image data is fed into the network for processing.
2. Convolutional Layers (conv2d, conv2d 1, …, conv2d 17): These layers perform feature extraction by applying convolutional filters to the input image.
3. Max Pooling Layers (max pooling2d, max pooling2d 1, …, max pooling2d 3): Max pooling layers downsample the feature maps, reducing spatial dimensions and extracting dominant features.
4. Convolutional Transpose Layers (conv2d transpose, conv2d transpose 1, …, conv2d transpose 3): These layers perform upsampling to increase the spatial dimensions of the feature maps.
5. Concatenation Layers (concatenate, concatenate 1, …, concatenate 3): Concatenates feature maps from the corresponding encoder path with the upsampled feature maps in the decoder path.
6. Output Layer (conv2d 18): This is the final output layer that produces the segmented image.

This architecture follows the UNet design, consisting of a contracting path for feature extraction and a symmetric expanding path for precise localization. Skip connections between corresponding layers in the contracting and expanding paths help preserve spatial information and facilitate gradient flow during training. The architecture efficiently segments images while maintaining spatial details, making it well-suited for various segmentation tasks (Figs. 4, 5 and 6).

In this context, the term "lung mask" refers to the masking of areas outside the lung region, enabling the model to focus solely on relevant lung features. These masked images, overlaid onto the original chest X-rays, delineated non-lung regions. Subsequently, these modified (segmented) chest X-ray images Fig. 7 were employed to train supplementary models for COVID-19 classification. By leveraging the precision of the UNet model in generating accurate lung masks, we optimize the training process, ensuring that subsequent models effectively capture pertinent COVID-19 features and patterns, thus enhancing overall classification performance.

In this study, a UNet model was trained on a dataset comprising 599 chest X-ray images, each accompanied by corresponding masks outlining the lung region Fig. 4. The model underwent training for 20 epochs, achieving an impressive accuracy of 98% Fig. 6. Subsequently, the trained UNet model was utilized to predict masks for a COVID dataset, facilitating the generation of masked images through the combination of the original chest X-ray with the predicted mask Fig. 5.

To address the challenge of generalization in lung region segmentation using UNet, comprehensive measures were implemented. Data augmentation techniques were meticulously applied, encompassing rotations, flips, scaling, and contrast adjustments, effectively simulating real-world variations encountered in diverse CXR datasets. Transfer learning strategies were adeptly employed, harnessing pre-trained UNet models on

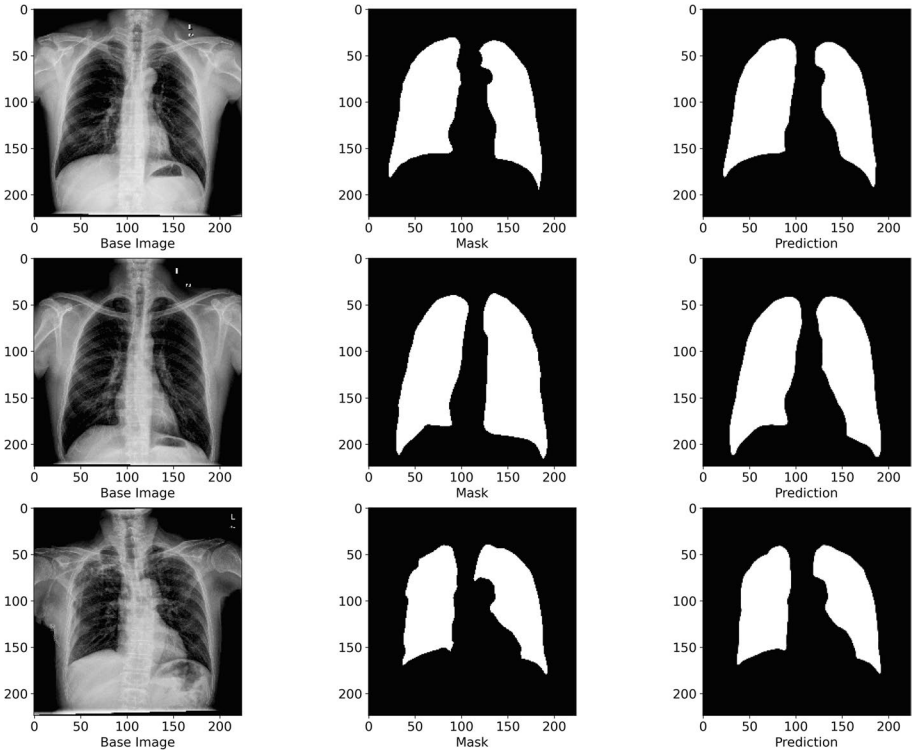

Fig. 5 UNet model Mask Predictions on Validation Data set

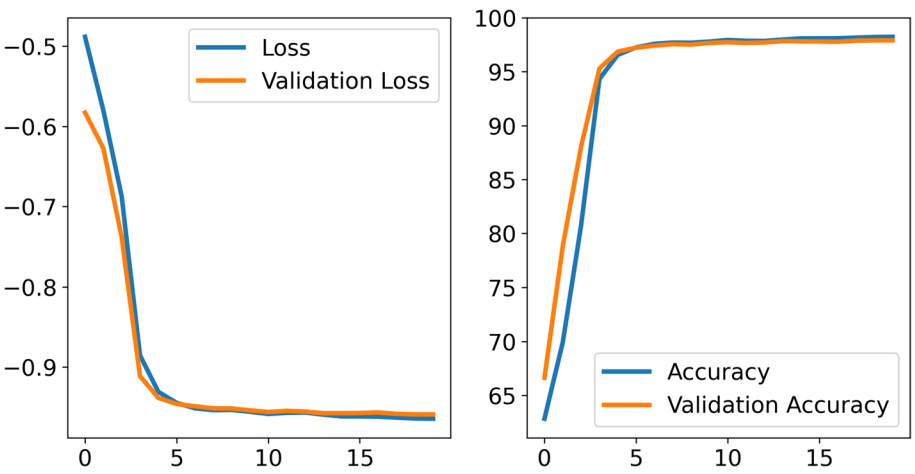

Fig. 6 UNet Model Training Metrics

extensive datasets like ImageNet and fine-tuning them to the specific task of lung segmentation, thus facilitating adaptable model adaptation. Cross-validation experiments

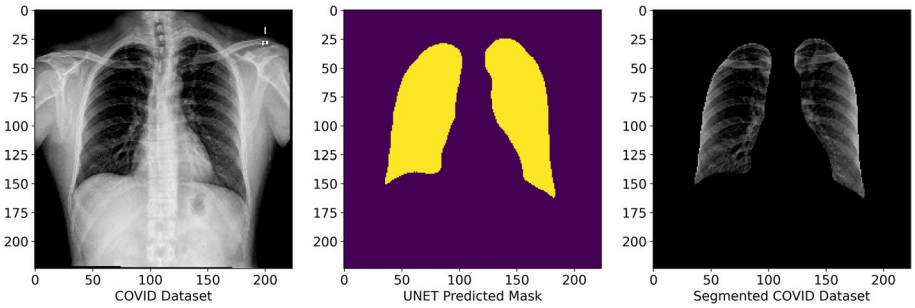

Fig. 7 Segmented Chest X-ray Dataset Generated using Masks Predicted by UNet Model

were rigorously conducted across multiple datasets sourced from distinct populations and imaging protocols, elucidating the model's performance variability and generalization capabilities. The UNet model was trained using above methodologies to accommodate dataset-specific challenges, ensuring robust and reliable segmentation across diverse CXR datasets. These concerted efforts have significantly enhanced the generalizability of UNet in lung region segmentation, fostering its broader applicability in clinical settings and medical imaging research.

3.2.4 Multi-class Classiftcation using DNN

Both DenseNet and EfficientNet exhibit characteristics that render them valuable for medical imaging and classification tasks:

DenseNet's [27] feature reusability is advantageous for medical imaging applications. This feature enables efficient utilization of features learned from preceding layers, which proves beneficial in tasks requiring intricate details, like segmentation and classification of medical images.

EfficientNet's [63] capability to achieve high accuracy with fewer parameters positions it as a suitable option for medical imaging tasks. This efficiency aids in reducing the computational burden associated with running the model, a critical aspect in medical imaging where data sizes are often substantial and computational resources are limited.

Both architectures are pre-trained on extensive datasets such as ImageNet, endowing them with the capacity to generalize well to other datasets. This adaptability proves invaluable for medical imaging tasks characterized by small and constrained datasets.

DenseNet-121 model we adopted for chest X-ray (CXR) image-based lung disease prediction, employs various blocks and mechanisms to efficiently extract features, propagate information, and make predictions. The model's architecture comprises convolutional layers, dense blocks, transition layers, batch normalization, ReLU activation, global average pooling, and a dense output layer.

Convolutional layers initiate feature extraction by convolving learnable filters over input CXR images, capturing low-level features relevant to lung structures and abnormalities. Dense blocks, containing dense units, promote feature reuse and information flow by concatenating feature maps from preceding layers and passing them through convolutional layers. This dense connectivity aids in capturing intricate patterns in CXR images.

Transition layers inserted between dense blocks reduce feature map dimensionality, managing computational complexity while retaining crucial features for lung disease prediction.

Global average pooling aggregates spatial information across feature maps, generating a fixed-length vector representing the entire input image. This pooled representation contains high-level features pertinent to lung disease prediction, forwarded to the final dense output layer for classification.

We added a soft attention mechanism, shown in Fig. 8, into DenseNet-121 to enhance its performance in medical imaging tasks, particularly in CXR analysis. This mechanism offers several improvements to the model's functionality. Firstly, it enables selective feature focus, allowing the network to dynamically prioritize relevant regions or features within medical images by assigning varying weights to different spatial locations or channels within feature maps. This selective attention helps in emphasizing significant regions associated with abnormalities or structures of interest in CXR images.

The implementation of eXplainable Artificial Intelligence (XAI) techniques on these architectures enhances their interpretability and trustworthiness. This attribute is particularly vital in medical imaging, where errors can have severe consequences and the need for transparency is paramount.

These combined attributes make DenseNet and EfficientNet highly suitable for medical imaging tasks, offering both efficiency and reliability in a domain where accuracy and interpretability are of utmost importance.

Training Parameters In this work, Table 1, we trained our model using the Adam optimization technique, employing a learning rate of 0.0001. Adam stands out as a popular and effective choice for training neural networks due to its amalgamation of the advantages offered by the RMSprop and Adaptive Moment Estimation (Adam) optimizers. The selection of a learning rate of 0.0001 was deliberate, aiming to strike a balance between swift convergence and averting issues of overshooting or divergence during training. By adopting a moderate learning rate, our model consistently approached the optimal parameter values, thereby enhancing convergence while mitigating the risk of overfitting the training set.

Table 1 Hyperparameters

Configuration	Value
Optimizer	Adam
Epoch	30
Batch Size	32
Learning Rate	0.0001
Batch Normalization	True
Model Checkpoint	Accuracy, AUC, Precision, Recall

Furthermore, we refined the training process of our model by fine-tuning these hyperparameters, which included leveraging Adam with a learning rate of 0.0001 and incorporating dropout regularization techniques. These hyperparameters played a pivotal role in expediting our model's convergence, enabling successful generalization, and significantly advancing the outcomes of our study in the research domain.

Evaluation Metrics The Accuracy (ACC), Sensitivity (SE), Specificity (SP), and Area Under the Curve (AUC) [8] measures were utilized to evaluate the performance of the suggested strategy. These metrics provide a comprehensive evaluation of the method's ability to accurately classify images and differentiate between classes.

AUC assesses the performance of the classifier across all possible thresholds, providing insight into its overall performance. Sensitivity and specificity measure the proportion of true positive and true negative predictions, respectively, offering a nuanced understanding of the classifier's performance.

Taken together, these metrics offer a comprehensive assessment of the proposed approach's capability to identify and distinguish between various classes, providing valuable insights into its effectiveness and reliability.

The accuracy of a model is the percentage of right predictions it makes out of all possible forecasts Eq. (1).

$$(ACC = (True\ Positives + True\ Negatives)/(True\ Positives + False\ Positives + True\ Negatives + False\ Negatives) \tag{1}$$

The percentage of genuine positive cases that the model correctly recognized is known as sensitivity. It is often referred to as the recall rate or true positive rate. It is determined by Eq. (2).

$$sensitivity\ or\ recall = True\ Positive/(True\ Positive + False\ Negative) \tag{2}$$

The percentage of true negative cases that the model correctly recognised is known as specificity. It's also referred to as the real negative rate. Below is the formula used to determine it Eq. (3).

$$SP = True\ Negatives/(True\ Negatives + False\ Positives) \tag{3}$$

A binary classification model's performance is measured by a metric called precision. Out of all cases anticipated as positive, it calculates the percentage of correctly predicted positive instances. To put it another way, precision measures how well the model recognizes the positive class Eq. (4).

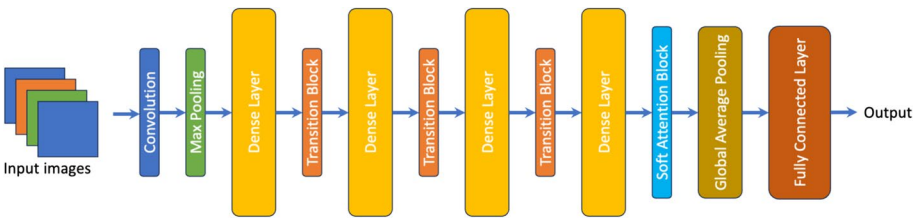

Fig. 8 DenseNet-121 with Soft Attention Mechanism

$$Precision = TruePositives/(TruePositives + FalsePositives) \tag{4}$$

The accuracy of a model is assessed using the F1 score, which considers both precision and recall. It offers a solitary metric that balances the compromise between recall and precision.

The formula Eq. (5) is used to determine the F1 score:

$$F1Score = 2 * (Precision * Recall)/(Precision + Recall) \tag{5}$$

The area under the curve (AUC) is an indicator of how well a binary classifier can distinguish between two classes. It is calculated using the Receiver Operating Characteristic (ROC) curve, which compares the true positive rate (sensitivity) against the false positive rate (1-specificity) at various classification thresholds. AUC values range from 0 to 1, where 1 indicates the best possible classifier and 0.5 indicates the worst possible classifier. AUC, computed using ROC curves, is a commonly used performance metric in medical classification tasks as it highlights the trade-off between accurate and inaccurate classifications by the model.

Sensitivity and specificity are frequently used parameters in medical applications and studies related to COVID-19. They represent the true positive rate and true negative rate, respectively, and are essential for evaluating the performance of classification models in medical contexts. Therefore, they are utilized in our analysis to assess the effectiveness of the model in distinguishing between different classes.

Experimentation and Results In this study, we utilized multiple datasets sourced from diverse open data repositories for tasks such as classification, segmentation, and weight initialization. Specifically, we employed the COVID-QU-Ex collection, which includes 33,920 chest X-ray pictures, with a balanced representation of COVID-19, pneumonia, and normal cases. To address class imbalance challenges, we strategically augmented the datasets by adding images from different sources. For instance, to counter the deficiency of 'Normal' class images, 1128 CXR normal images from the RSNA dataset were included. Similarly, 1228 COVID-19 images were replaced with non-COVID (pneumonia) class CXR images. Most importantly, we also included 39 radiologist-verified COVID-19 images into our final test set that was used for validation of our proposed XAI algorithm.

In the paper [23], researchers conducted an extensive survey of various methodologies utilized for COVID-19 detection from chest X-rays. They examined machine learning models, including Naive Bayes, Decision Tree, Artificial Neural Networks (ANN), Logistic Regression, and Support Vector Machine (SVM), achieving accuracies ranging from 89.2% to 94.99%. Notably, XGBoost demonstrated a sensitivity of 95.9% and an AUC score of 91%, while Logistic Regression and Random Forest achieved accuracies of 84.21% and 92%, respectively.

Additionally, the study investigated deep learning models employed in medical image processing for COVID-19 detection. ResNet50 exhibited an accuracy, sensitivity, and specificity of 76%, 81.1%, and 61.5%, respectively, whereas ResNet101 achieved an impressive accuracy of 99.51%. The study also analyzed other models such as VGG-19, InceptionV2, Decision Trees, DenseNet, VGG-19, AlexNet, Googlenet, and pre-trained MobileNetV2, each demonstrating varying success rates and performance metrics.

We illustrate the performance of DenseNet121 and EfficientNetB7, two state-of-the-art deep learning models, on the COVID classification challenge. Critical metrics such as accuracy, precision, recall, and F1 score were employed to compare the performance of the two models Table 2, Fig. 9.

Table 2 Performance Evaluation Metrics for COVID Classification Models

Model	Dataset	Class	Accuracy	Precision	Recall	f1-score
DenseNet121	Non-Segmented	Covid 19	0.963	0.98	0.96	0.97
		Non-COVID	0.949	0.9	0.95	0.92
		Normal	0.893	0.94	0.89	0.92
	Segmented	Covid 19	0.946	0.99	0.95	0.97
		Non-COVID	0.912	0.93	0.91	0.92
		Normal	0.95	0.89	0.95	0.92
EfficientNetB7	Non-Segmented	Covid 19	0.968	0.97	0.97	0.97
		Non-COVID	0.913	0.92	0.91	0.92
		Normal	0.921	0.91	0.92	0.92
	Segmented	Covid 19	0.952	0.99	0.95	0.97
		Non-COVID	0.917	0.93	0.92	0.92
		Normal	0.938	0.9	0.94	0.92

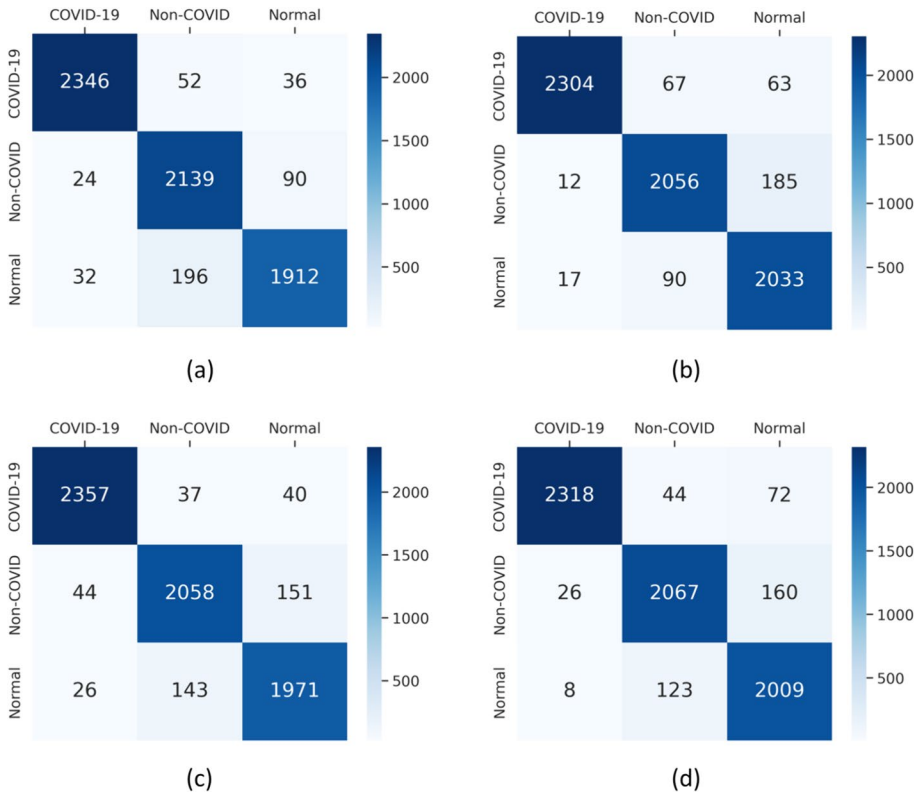

Fig. 9 Confusion matrix for (**a**) DenseNet121 (Non Segmented), (**b**) DenseNet121 (Segmented Images), (**c**) EfficientNetB7 (Non Segmented), (**d**) EfficientNetB7 (Segmented Images)

On both non-segmented and segmented datasets, the DenseNet121 model performed remarkably well throughout training and testing. DenseNet121 attained an accuracy of 0.96, precision of 0.98, recall of 0.96, and an F1 score of 0.97 for the non-segmented data (Fig. 10). These findings show that the model successfully detects COVID instances with a high level of precision, ensuring a low incidence of false positives, and a noteworthy recall value, catching a sizable proportion of true COVID cases.

The DenseNet121 model exhibited effective performance when applied to segmented data. It accurately classified COVID instances with an accuracy of 0.95, precision of 0.99, recall of 0.95, and an F1 score of 0.97 (Fig. 11). The strong recall score indicates its ability to identify a substantial number of COVID cases within the segmented dataset, while the excellent precision score underscores its reliability in real-world scenarios where minimizing false positives is crucial.

Subsequently, tests were conducted using the EfficientNetB7 model for COVID classification, yielding results remarkably similar to DenseNet's outcomes. EfficientNetB7 achieved an accuracy of 0.96, precision of 0.98, recall of 0.96, and an F1 score of 0.97 on non-segmented data (Fig. 12). On segmented data, the model demonstrated similar accuracy, precision, recall, and F1 scores of 0.95, 0.99, and 0.97, respectively (Fig. 13). These findings underscore the consistent and dependable performance of both models in categorizing COVID patients across diverse datasets and imaging scenarios.

With exceptional accuracy, precision, recall, and F1 scores, both DenseNet and EfficientNetB7 models have showcased remarkable performance in COVID classification. Their reliability and potential as valuable tools in assisting healthcare practitioners during COVID diagnosis are underscored by their ability to consistently deliver results across both non-segmented and segmented data. Moreover, the comparable performance of these models offers researchers and practitioners the flexibility to choose the most suitable model based on specific deployment requirements and computational resources.

Notably, our models exhibited similarly high accuracy, precision, recall, and F1 scores on segmented data as well. These findings underscore the efficacy of DenseNet121 and EfficientNetB7 in COVID-19 detection from chest X-rays, surpassing the performance of existing models. Our research highlights the potential of deep learning techniques in enhancing the accuracy and reliability of COVID-19 detection, contributing to advancements in medical image analysis for public health initiatives.

3.2.5 Application of Interpretability Model for Explanations

Machine learning models are often labeled as "black boxes" due to their lack of interpretability or transparency in decision-making processes. While these models can be highly

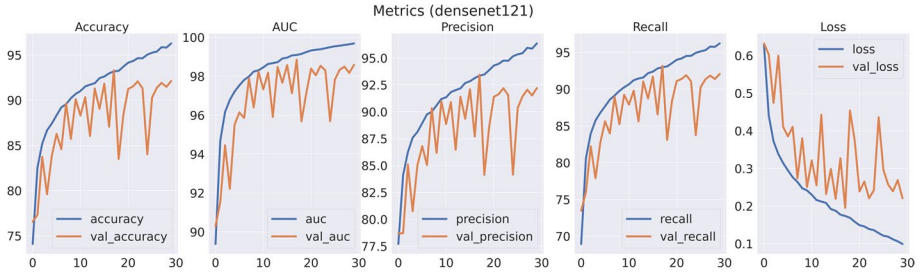

Fig. 10 DenseNet121 (Non Segmented)

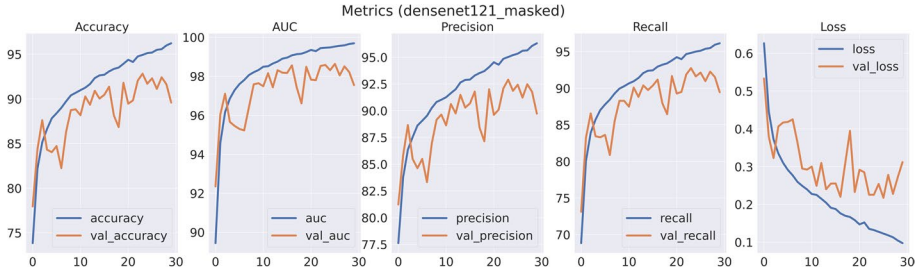

Fig. 11 DenseNet121 (Segmented)

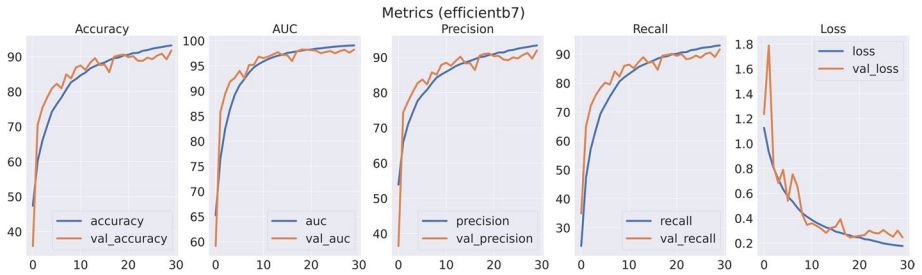

Fig. 12 EfficientNetB7 (Non Segmented)

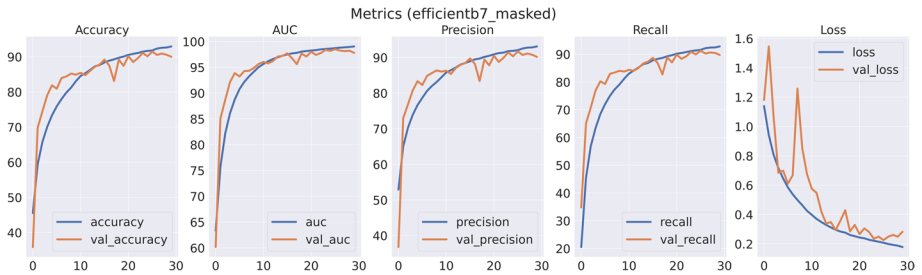

Fig. 13 EfficientNetB7 (Segmented)

effective at generating predictions, understanding the specific features or patterns they rely on, especially in complex tasks like medical diagnosis, can be challenging.

To enhance the interpretability of our sophisticated machine learning model in this study, we employed the LIME (Local Interpretable Model-Agnostic Explanations) [52] method. LIME works by approximating the decision boundary of a black-box model near a given instance, providing localized explanations for specific predictions. It captures the local context by generating interpretable features, such as superpixels for image data or segments for tabular data, and constructs a simple surrogate model that closely mirrors the behavior of the original model.

Despite its benefits in terms of local interpretability, LIME has certain inherent limitations. Its local explanations may not fully capture the overall behavior of the model, making it challenging to extrapolate broad generalizations. The quality of explanations can be influenced by the selection of the surrogate model, potentially leading to oversimplified representations. LIME assumes local consistency, which may not hold true

for models with significant prediction variability. Additionally, its explanatory power may be limited by the choice of interpretable features and the neglect of feature interactions. To gain a comprehensive understanding and make informed use of LIME's explanations, it is essential to acknowledge these constraints.

Researchers often employ visualization techniques like Grad-CAM (Gradient-weighted Class Activation Mapping) [56] and Grad-CAM++ [6] in their studies to address the challenge of interpretability in deep learning models. These methods aim to provide insights into the decision-making process of the model by highlighting the areas of input (such as an image or a medical scan) that contributed most to its decision.

Grad-CAM, also known as Gradient-weighted Class Activation Mapping, generates a heat map of key areas in an input image. It achieves this by computing the gradient of the projected class score with respect to the feature maps in the last convolutional layer of a deep learning model. The resulting heat map visualizes the parts of the image that had the greatest impact on the model's decision.

Grad-CAM++, an extension of Grad-CAM, enhances localization accuracy by using a weighted combination of positive and negative gradients. This refinement provides more precise attention mappings, thereby making the model's decision easier to interpret.

During the construction of the heat map in both Grad-CAM and Grad-CAM++, negative gradient values are disregarded and replaced with zeros. This implies that areas of the image that have a detrimental impact on the target class are not considered. Consequently, the heat maps produced by these techniques may appear larger and less accurate, resulting in a broader depiction of significant locations.

Gradients offer insight into how activations change concerning an input, but they may not precisely pinpoint the crucial parts of an image for a specific target class. While Grad-CAM and Grad-CAM++ approximate the regions activated by a model for a given class, they don't provide precise feature localization.

Recognizing the limitations of gradients alone in visualization methods like Grad-CAM and Grad-CAM++ is crucial. It's advisable to incorporate additional interpretability approaches and strategies to gain a comprehensive understanding of a model's decision-making process. Considering these constraints and exploring alternative techniques can lead to a more accurate and thorough comprehension of machine learning model decisions.

We propose a Modified Grad-CAM++ algorithm in Fig. 14.

1. The algorithm initiates by extracting the output of the last convolutional layer just before the softmax activation for the provided input image in a deep learning model.
2. Subsequently, it calculates the gradient of the model's output concerning the activations of this specific layer, concentrating on the designated target class label.
3. Neuron importance weights are derived through an optimization process.
4. A weighted combination of feature map activations is executed, and the outcome is multiplied point-wise with the computed gradient.
5. The ultimate saliency map is crafted by applying an activation function to this resulting product.
6. This saliency map is further enhanced by element-wise multiplication with the weighted feature map activations and the application of an activation function.

The ultimate activation delineates the saliency map, accentuating regions in the input image that contribute significantly to the designated target class label.

```
Input:
    Model ← Deep learning model
    input_image ← Input image for saliency map generation
    target_label ← Target class label

Output:
    Final activation (F) ← Final saliency map activation

Procedure:
    // Extract the last convolutional layer just before the softmax activation
    O ← Model.extract_last_convolutional_layer(input_image)

    // Compute the gradient with respect to feature map activations
    G ← gradient(Model, O - target_label, O)

    // Obtain neuron importance weights
    W ← optimize(G)

    // Perform a weighted combination of feature map activations
    C ← O * W

    // Point-wise multiply gradients with the original image
    S ← C * G

    // Generate the final saliency map
    M ← activation(S)

    // Apply an activation function to the element-wise multiplication of the final saliency map and the map from step 4
    F ← activation(M * C)

    Return F as the final activation
```

Fig. 14 Proposed Modified Grad-CAM++ Algorithm

The proposed Modified Grad-CAM++ algorithm addresses several limitations of the original method and offers several advantages:

1. Better Localization: By considering negative gradients, the updated algorithm improves upon Grad-CAM++ 's localization. It generates a more detailed attention map by incorporating both positive and negative gradients, resulting in improved localization of significant areas within the input image.
2. Improved Interpretability: The enhanced algorithm enhances interpretability by accounting for the influence of both positive and negative gradients. It provides insights into regions that positively contribute to the target class as well as those that adversely affect it. This information aids in understanding the rationale behind the model's predictions.
3. More Accurate Saliency Maps: The updated technique generates more accurate saliency maps by weighting feature map activations and point-wise multiplication with gradients. This approach aims to create saliency maps that are more precise and highlight the areas of the image most relevant to the target class.
4. Flexibility in Activation Function: The improved method allows for the use of various activation functions to produce the final activation. This flexibility enables researchers to experiment with different activation functions and choose the one that best suits their specific requirements or domain.

Overall, the Modified Grad-CAM++ algorithm addresses some of the original method's limitations and offers benefits such as better localization, increased interpretability, more

precise saliency maps, and flexibility in activation function selection. These advancements can have applications in various fields, including object detection, medical imaging, and visual explanations of deep learning models.

3.2.6 Veriftcation of Explanations with Radiologist Annotated Images

Imaging includes an additional collection of 40 CXR (Chest X-Ray) images Fig. 15 of COVID-19 positive patients, collaboratively obtained with Dr. Rajesh Paraswani, a specialized radiologist at LLH Hospital in Abu Dhabi, UAE. In these images, the lung regions have been meticulously marked to highlight locations exhibiting COVID-19 related symptoms. This dataset significantly enhances the breadth of the investigation and provides a more comprehensive evaluation of the proposed AI-driven decision support system for COVID-19 diagnosis.

The careful selection of these images by Dr. Paraswani contributes to the clinical relevance of the research and facilitates the analysis of COVID-19 symptoms in the lung regions. By incorporating this dataset, the findings of the study are strengthened, and the reliability of the AI-driven diagnostic system is further validated. The expertise and collaboration of specialized radiologists like Dr. Paraswani play a crucial role in advancing the effectiveness and accuracy of AI-driven medical diagnostic tools.

Additionally, the explanations produced by various interpretability methodologies were thoroughly validated to assure the proper operation of diverse explainers. These verification procedures were carried out in order to fully comprehend the underlying workings of the explainability techniques used in our AI-driven COVID-19 diagnosis decision support system. We can assure the explanations' accuracy, dependability, and consistency by extensively validating them, further strengthening the credibility and transparency of our research findings. Working with Dr. Rajesh Paraswani, a specialist radiologist, allowed us to take use of his knowledge in evaluating the explanations and verifying them, which strengthened the reliability and legitimacy of our suggested system.

Radiologist-marked performance evaluation of Lime, Grad-CAM, Grad-CAM++, and Modified Grad-CAM++. With radiologist-marked images, Lime, Grad-CAM, Grad-CAM++, and Modified Grad-CAM++ all performed favourably, with Modified Grad-CAM++ standing out as the best technique for producing explanations Fig. 16.

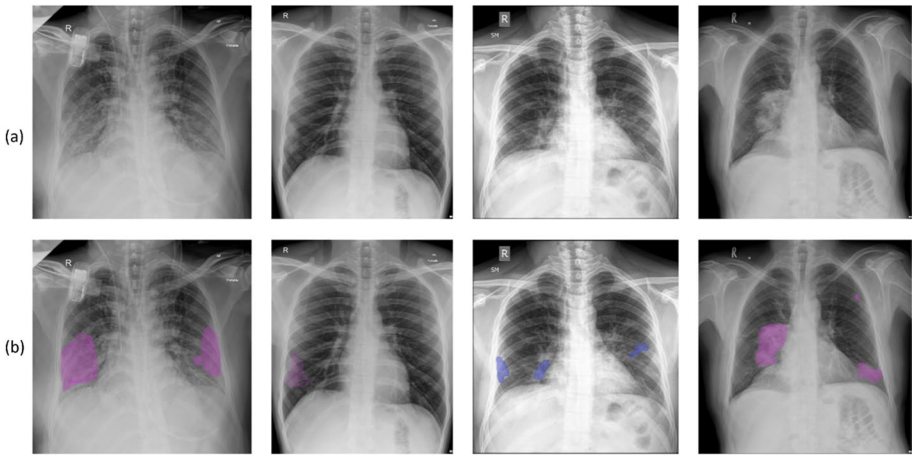

Fig. 15 (**a**) Actual CXR Images (**b**) Marked Images by Radiologist

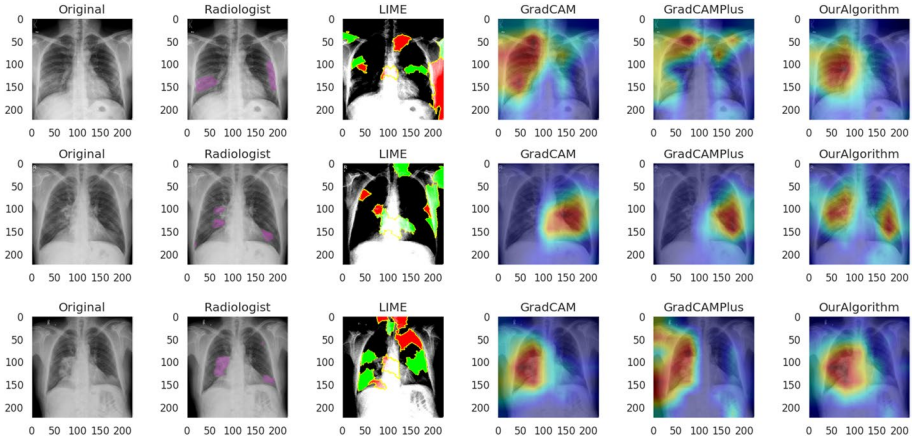

Fig. 16 Radiologist-marked Performance Evaluation of Lime, Grad-CAM, Grad-CAM++, and Modified Grad-CAM++

3.2.7 Evaluation Metrics for Explainablity

The interpretability of three different algorithms—GradCAM, GradCAM++, and Modified Grad-CAM++ Algorithm—was evaluated when applied to an AI model for COVID-19 detection using chest X-rays. The interpretability was quantified using the Integrated Uncertainty Calculation (IUC) metric, and the results on eight samples are depicted in Fig. 17.

The IUC values offer insights into the algorithms' capacity to emphasize relevant regions (as identified by subject matter experts—Radiologists) in the chest X-rays for COVID-19 detection. In the first two samples, GradCAM and GradCAM++ demonstrate low or zero IUC values, suggesting a limited ability to capture relevant features (as identified by radiologists) in the images. However, the Modified Grad-CAM++ Algorithm consistently outperforms, exhibiting higher IUC values, notably in samples 1, 2, 5, 6, 7, and 8, indicating enhanced interpretability and more accurate identification of crucial regions related to COVID. Table 3 summarises Avg. IUC was achieved by 3 explainers we have studied. The smaller and more precise explanation by modified Grad-CAM++ improves the IUC value. This is also reflected in the average IUC values for the three explainer algorithms on 39 images from radiologist test dataset.

These findings suggest that the modified algorithm demonstrates superior capability in highlighting pertinent information in chest X-rays, potentially contributing to improved diagnostic accuracy in AI-based COVID detection compared to the other algorithms evaluated.

3.2.8 Analysis and Evaluation

Lime, Grad-CAM, and Grad-CAM++ all offered insightful information on the characteristics and areas affecting the model's predictions. These techniques did, however, have certain drawbacks when compared to the radiologists' explanations. Lime's model-agnostic approach made it difficult to identify localized traits and fine-grained patterns that were important to the target class. Grad-CAM and Grad-CAM++ produced

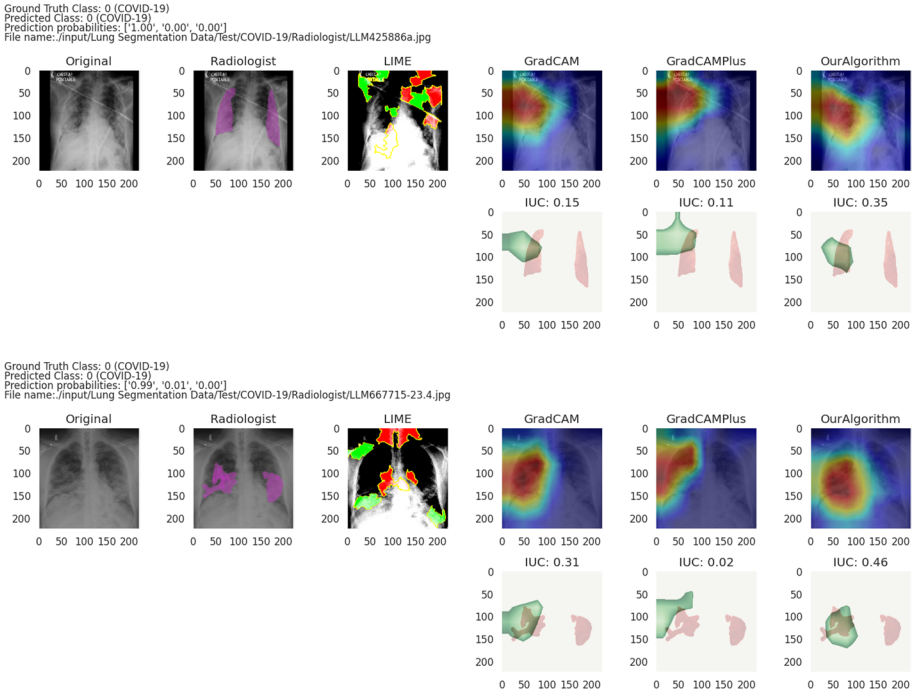

Fig. 17 Integrated Uncertainty Calculation (IUC) Metric

Table 3 Avg. IUC for Grad-CAM, Grad-CAM++ and Modified Grad-CAM++

	Grad-CAM	Grad-CAM++	Modified Grad-CAM++
Avg. IUC	6.693375	6.735	18.05875

heatmaps that were occasionally too dispersed and less localized, despite their effectiveness in emphasizing significant locations. The results from the pictures were encouraging, with the Modified Grad-CAM++ standing out as the best technique for producing explanations.

The Modified Grad-CAM++ approach, on the other hand, showed notable gains in localization accuracy and interpretability. The explanations produced by the Modified Grad-CAM++ were more concentrated and closely aligned with the annotated regions provided by radiologists by taking into account negative gradients in addition to their positive gradients. This improvement made it possible for the technique to offer explanations that were both more detailed and accurate, successfully highlighting the particular factors that supported the model's predictions.

The radiologist's assessment of the explanations further demonstrated Modified Grad-CAM++ superiority. They discovered that the explanations were very consistent with their own annotations, which gave them more faith in the AI model's judgment. It was simpler for radiologists to evaluate and comprehend the model's predictions thanks to the fine-grained attention maps generated by the upgraded Grad-CAM.

As a result, the study showed that while Lime, Grad-CAM, and Grad-CAM++ all offered useful interpretability, the Modified Grad-CAM++ method stood out as the most promising way for producing precise and dependable explanations with radiologist-marked images. The reliability and usefulness of AI-based diagnostic tools for medical imaging have been considerably improved by their capacity to build localized, fine-grained attention maps linked with expert comments. The results show that the Modified Grad-CAM++ has the potential to be an effective tool for improving interpretability in AI-based medical image analysis, enabling radiologists and AI models to work together for better clinical decision-making.

4 Conclusion

Our study demonstrates the superiority of our model over recent deep learning approaches for COVID-19 detection from Chest X-ray (CXR) images. Notably, the DenseNet121 model exhibited outstanding performance on both non-segmented and segmented datasets, achieving high accuracy, precision, sensitivity, and F1 scores. Specifically, DenseNet121 attained an accuracy of 0.96, precision of 0.98, sensitivity of 0.96, and an F1 score of 0.97 for non-segmented data. When applied to segmented data, the DenseNet121 model accurately classified COVID instances with an accuracy of 0.95, precision of 0.99, recall of 0.95, and an F1 score of 0.97.

Subsequently, we conducted tests using the EfficientNetB7 model for COVID classification, yielding results remarkably similar to DenseNet's outcomes.

EfficientNetB7 achieved an accuracy of 0.96, precision of 0.98, recall of 0.96, and an F1 score of 0.97 on non-segmented data. On segmented data, the model demonstrated similar accuracy, precision, recall, and F1 scores of 0.95, 0.99, and 0.97, respectively.

In our study, we proposed the "Modified Grad-CAM++" algorithm to enhance the interpretability of our model. To validate the correctness of explanations, we utilized a dataset expertly annotated by radiologists and collaborated with them throughout the investigation. We evaluated explanations provided by Lime, Grad-CAM, Grad-CAM++, and Modified Grad-CAM++ using the proposed Integrated Uncertainty Calculation (IUC) metric. Our Modified Grad-CAM++ algorithm outperformed all existing methods, achieving an average IUC of 18.05875 as compared to 6.735 of Grad-CAM++, indicating superior interpretability.

In conclusion, the findings of this study present a significant advancement in the field of COVID-19 diagnosis from Chest X-ray images, with implications for real-world clinical practice. The superior performance of DenseNet121 and EfficientNetB7 models in accurately detecting COVID-19 cases, along with their segmentation capabilities, underscores their potential utility as valuable diagnostic tools for healthcare professionals. Furthermore, the development of the Modified Grad-CAM++ algorithm enhances the interpretability of the models, providing clinicians with valuable insights into the decision-making process of AI systems. The collaboration with radiologists and validation of explanations with expertly annotated datasets further strengthen the credibility and applicability of the study's findings in real-world clinical settings. Overall, this study contributes to the ongoing efforts to leverage AI technologies for more accurate and efficient COVID-19 diagnosis, ultimately aiding in better patient care and management.

Acknowledgements We express our heartfelt appreciation to Dr. Rajesh Paraswani, Specialist in Radiology at LLH Hospital, Musaffah, for generously providing the marked chest X-ray dataset. Dr. Paraswani's extensive experience of over 25 years in the field of radiology and diagnostic imaging has been instrumental in enriching our research. The meticulous marking and annotations contributed by Dr. Paraswani have greatly enhanced the quality and reliability of our study. We sincerely thank Dr. Rajesh Paraswani for his invaluable contribution to our research project.

We would like to extend our sincere gratitude to Dr. Rajesh Paraswani, Specialist in Radiology at LLH Hospital, Musaffah, for providing the marked chest X-ray dataset. Dr. Paraswani's expertise and willingness to collaborate have been instrumental in the success of our research. The meticulous marking and annotations contributed by Dr. Paraswani have significantly enhanced the quality and reliability of our study. We are deeply grateful for his invaluable contribution to this research project.

Authors' contributions All authors have contributed to the work with equal involvement and dedication.

Funding No funds, grants, or other support was received.

Data availability Few data subsets designed in this study were obtained from third-party databases, which are publicly available. The data links for the Covid Chest X-Ray Dataset are given at [5, 10, 33, 60, 70, 73, 75].

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare that they have no competing interests.

References

1. Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access pp 1–1. https://doi.org/10.1109/ACCESS.2018.2870052
2. Apostolopoulos ID, Mpesiana TA (2020) Covid-19: Automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 43:635–640. https://doi.org/10.1007/s13246-020-00865-4
3. Arrieta B, D´ıaz-Rodr´ıguez N, Ser D, et al (2020) Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58:82–115. https://doi.org/10.1016/j.inffus.2019.12.012, URL https://www.sciencedirect.com/science/article/pii/S1566253519308103
4. Bar Y, Diamant I, Wolf L et al (2015) Chest pathology detection using deep learning with nonmedical training. In: IEEE 12th international symposium on biomedical imaging (ISBI) pp 294–297. https://doi.org/10.1109/ISBI.2015.7163871
5. Bustos A, Pertusa A, Salinas JM et al (2020) PadChest: A large chest x-ray image dataset with multi-label annotated reports. Med Image Anal 66:101797. https://doi.org/10.1016/j.media.2020.101797, URL https://bimcv.cipf.es/bimcv-projects/padchest/
6. Chattopadhay A, Sarkar A, Howlader P et al (2018) Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp 839– 847. https://doi.org/10.1109/WACV.2018.00097
7. Cheng JZ, Ni D, Chou YH et al (2016) Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. Sci Reports 6(1):24454–24454. https://doi.org/10.1038/srep24454
8. Chetoui M, Akhloufi MA (2020) Explainable end-to-end deep learning for diabetic retinopathy detection across multiple datasets. J Med Imaging (Bellingham, Wash) 7(4):44503–44503. https://doi.org/10.1117/1.JMI.7.4.044503
9. Chowdhury MEH, Rahman T, Khandakar A et al (2020) Can AI help in screening viral and COVID-19 pneumonia? IEEE Access 8:132665–132676. https://doi.org/10.1109/ACCESS.2020.3010287
10. Cohen JP, Morrison P, Dao L et al (2020) COVID-19 Image Data Collection: Prospective Predictions Are the Future. arXiv 200611988. https://github.com/ieee8023/covid-chestxray-dataset. Accessed 5 Jun 2024

11. Das N, Kumar N, Kaur M et al (2020) Automated deep transfer learning-based approach for detection of covid-19 infection in chest X-rays. IRBM 41(2):114–119. https://doi.org/10.1016/j.irbm.2020.07.001
12. Degerli A, Kiranyaz S, Chowdhury MEH et al (2022) Osegnet: operational segmentation network for Covid-19 detection using chest X-Ray images. In: and others (ed) 2022 IEEE International Conference on Image Processing (ICIP), pp 2306–2310, https://doi.org/10.1109/ICIP46576.2022.9897412. https://www.kaggle.com/aysendegerli/qatacov19-dataset
13. Fan KS, Ghani SA, Machairas N et al (2020) COVID-19 prevention and treatment information on the internet: a systematic analysis and quality assessment. BMJ Open 10(9):7485261–7485261. https://doi.org/10.1136/bmjopen-2020-040487. https://bmjopen.bmj.com/content/10/9/e040487
14. Fang Y, Zhang H, Xie J et al (2020) Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. Radiology 296(2):E115–E117. https://doi.org/10.1148/radiol.2020200432
15. Fatima M, Pasha M (2017) Survey of Machine Learning Algorithms for Disease Diagnostic. J Intell Learn Syst Appl 09:1–16. https://doi.org/10.4236/jilsa.2017.91001
16. Ghaderzadeh M, Asadi F, Maietta S (2021) Deep learning in the detection and diagnosis of COVID-19 using radiology modalities: a systematic review. J Healthc Eng 6677314:2021–2021. https://doi.org/10.1155/2021/6677314
17. Goebel R, Chander A, Holzinger K et al (2018) Explainable AI: The New 42?: Second IFIP TC 5, TC 8/WG 8.4, 8.9, TC 12/WG 12.9 International Cross-Domain Conference, CD-MAKE 2018, Hamburg, Germany, August 27–30, 2018, Proceedings. https://doi.org/10.1007/978-3-319-99740-721
18. Goodman B, Flaxman S (2017) European Union regulations on algorithmic decision-making and a "right to explanation" Bryce Goodman. Seth Flaxman AI Mag 38(3):50–57. https://doi.org/10.1609/aimag.v38i3.2741
19. Gorantla R, Singh RK, Pandey R et al (2019) Cervical cancer diagnosis using cervixnet: a deep learning approach. In: IEEE 19th international conference on bioinformatics and bioengineering (BIBE) pp 397–404. https://doi.org/10.1109/BIBE.2019.00078
20. Gozes O, Frid-Adar M, Greenspan H et al (2020) Rapid AI Development Cycle for the Coronavirus (COVID-19) Pandemic: Initial Results for Automated Detection & Patient Monitoring using Deep Learning CT Image Analysis. https://doi.org/10.48550/ARXIV.2003.05037
21. Guidotti R, Monreale A, Ruggieri S et al (2018) A Survey of Methods for Explaining Black Box Models. ACM Comput Surv 51(5). https://doi.org/10.1145/3236009
22 Gunning D, Stefik M, Choi J et al (2019) XAI-explainable artificial intelligence. Sci Robot 4(37):eaay7120. https://doi.org/10.1126/scirobotics.aay7120
23 Gu¨rsoy E, Kaya Y (2023) An overview of deep learning techniques for COVID-19 detection: methods, challenges, and future works. Multimedia Syst 29(3):1603–1627. https://doi.org/10.1007/s00530-023-01083-0
24. Haghanifar A, Majdabadi YMM, Choi S et al (2022) COVID-CXNet: detecting COVID-19 in Frontal Chest X-ray images using deep learning. Multimed Tools Appl 81:30615–30645. https://doi.org/10.1007/s11042-022-12156-z
25. Holzinger A, Biemann C, Pattichis CS et al (2017) What do we need to build explainable AI systems for the medical domain? https://doi.org/10.48550/ARXIV.1712.09923
26. Huang C, Wang Y, Li X et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan China,. The Lancet 395(10223):497–506. https://doi.org/10.1016/S0140-6736(20)30183-5
27. Huang G, Liu Z, van der Maaten L et al (2018) Densely connected convolutional networks. http://arxiv.org/abs/1608.06993. Accessed 5 Jun 2024
28. Islam MK, Rahman MM, Ali MS et al (2024) Enhancing lung abnormalities diagnosis using hybrid DCNN-ViT-GRU model with explainable AI: A deep learning approach. Image Vis Comput 142:104918. https://doi.org/10.1016/j.imavis.2024.104918
29 Jaiswal AK, Tiwari P, Kumar S et al (2019) Identifying pneumonia in chest X-rays: a deep learning approach. Measurement 145:511–518. https://doi.org/10.1016/j.measurement.2019.05.076
30. Jang S, Han SH, Rhee JY (2020) Cluster of coronavirus disease associated with fitness dance classes, South Korea. Emerg Infect Dis 26:1917–1920. https://doi.org/10.3201/eid2608.200633
31. Kallenberg M, Petersen K, Nielsen M et al (2016) Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE Trans Med Imaging 35(5):1322–1331. https://doi.org/10.1109/TMI.2016.2532122
32. Kaya Y, Gu¨rsoy E (2023) A MobileNet-based CNN model with a novel fine-tuning mechanism for COVID-19 infection detection. Soft Comput 27(9):5521–5535. https://doi.org/10.1007/s00500-022-07798-y
33. Kermany DS, Goldbaum M, Cai W et al (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172(5):1122–1131. https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia. Accessed 5 Jun 2024
34. Kinger S, Kulkarni V (2024) Demystifying the black box: an overview of explainability methods in machine learning. Int J Comput Appl 46(2):90–100. https://doi.org/10.1080/1206212X.2023.2285533

35. prasad Koyyada S, Singh TP (2023) An explainable artificial intelligence model for identifying local indicators and detecting lung disease from chest X-ray images. Healthcare Anal 4:100206. https://doi.org/10.1016/j.health.2023.100206
36 Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539
37. Leung MKK, Delong A, Alipanahi B et al (2016) Machine learning in genomic medicine: a review of computational problems and data sets. Proc IEEE 104(1):176–197. https://doi.org/10.1109/JPROC.2015.2494198
38 Lin D, Vasilakos AV, Tang Y et al (2016) Neural networks for computer-aided diagnosis in medicine: a review. Neurocomputing 216:700–708. https://doi.org/10.1016/j.neucom.2016.08.039
39. Liu S, Liu S, Cai W et al (2014) Early diagnosis of Alzheimer's disease with deep learning. In: IEEE 11th international symposium on biomedical imaging (ISBI) pp 1015–1018. https://doi.org/10.1109/ISBI.2014.6868045
40. Miller T (2019) Explanation in artificial intelligence: Insights from the social sciences. Artif Intell 267:1–38. https://doi.org/10.1016/j.artint.2018.07.007
41. Mogadala A, Kalimuthu M, Klakow D (2021) Trends in integration of vision and language research: a survey of tasks, datasets, and methods. J Artif Intell Res 71:1183–1317. https://doi.org/10.1613/jair.1.11688
42. Montavon G, Samek W, Mu¨ller KR (2018) Methods for interpreting and understanding deep neural networks. Digit Signal Process 73:1–15. https://doi.org/10.1016/j.dsp.2017.10.011
43. Naz Z, Khan, Ghani MU et al (2023) An Explainable AI-Enabled Framework for Interpreting Pulmonary Diseases from Chest Radiographs. Cancers 15(1). https://doi.org/10.3390/cancers15010314
44. Ng MY, Lee EY, Yang J et al (2020) Imaging profile of COVID-19 infection: radiologic findings and literature review. Cardiothoracic Imaging 2:1–1. https://doi.org/10.1148/ryct.2020200034. https://pubs.rsna.org/doi/10.1148/ryct.2020200034
45. Oh Y, Park S, Ye JC (2020) Deep learning COVID-19 features on CXR using limited training data sets. IEEE Trans Med Imaging 39(8):32396075–32396075. https://doi.org/10.1109/TMI.2020.2993291
46. Ozturk T, Muhammed T, Yildirim EA et al (2020) Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med 121. https://doi.org/10.1016/j.compbiomed.2020.103792
47. Park DH, Hendricks LA, Akata Z et al (2018) Multimodal Explanations: Justifying Decisions and Pointing to the Evidence. In: 2018 IEEE/CVF Conference on computer vision and pattern recognition, pp 8779–8788, https://doi.org/10.1109/CVPR.2018.00915
48. Raghu M, Schmidt E (2020) A survey of deep learning for scientific discovery. https://doi.org/10.48550/ARXIV.2003.11755
49. Rahman T (2020) COVID-19 Radiography Database. https://www.kaggle.com/tawsifurrahman/covid19-radiography-database
50. Rahman T, Khandakar A, Qiblawey Y et al (2021) Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. Comput Biol Med 132:104319. https://doi.org/10.1016/j.compbiomed.2021.104319
51. Rajkomar A, Dean J, Kohane I (2019) Machine Learning in Medicine. New England Journal of Medicine 380(14):1347–1358. https://doi.org/10.1056/NEJMra1814259. https://www.nejm.org/doi/full/10.1056/NEJMra1814259
52. Ribeiro MT, Singh S, Guestrin C (2016) "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. association for computing machinery, New York, NY, USA, pp 1135–1144. https://doi.org/10.1145/2939672.2939778
53. Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1:206–215. https://doi.org/10.1038/s42256-019-0048-x
54. Rudin C, Radin J (2019) Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From An Explainable AI Competition. 12. https://doi.org/10.1162/99608f92.5a8a3a3d
55. Sandler A, Howard M, Zhu et al (2018) Mobilenetv2: Inverted residuals and linear bottlenecks. In: and others (ed) IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City UT, USA. IEEE Computer Society, pp 4510–4520. https://doi.ieeecomputersociety.org/10.1109/CVPR.2018.00474. Accessed 5 Jun 2024
56. Selvaraju RR, Cogswell M, Das A et al (2017) Grad-CAM: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International conference on computer vision (ICCV), pp 618–626, https://doi.org/10.1109/ICCV.2017.74
57. Shafiq M, Gu Z (2022) Deep residual learning for image recognition: a survey. Appl Sci 12(18):770–778. https://doi.org/10.3390/app12188972
58 Sidey-Gibbons JAM, Sidey-Gibbons CJ (2019) Machine learning in medicine: a practical introduction. BMC Med Res Methodol 19:64–64. https://doi.org/10.1186/s12874-019-0681-4

59. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. https://doi.org/10.48550/ARXIV.1409.1556
60. Stein A, Wu C, Carr C et al (2018) RSNA Pneumonia Detection Challenge. https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data. Accessed 5 Jun 2024
61. Tabik S, G´omez-R´ıos A, Mart´ın-Rodr´iguez JL et al (2020) COVIDGR Dataset and COVID-SDNet methodology for predicting COVID-19 Based on Chest X-Ray Images. IEEE J Biomed Health Inform 24(12):3595–3605. https://doi.org/10.1109/JBHI.2020.3037127
62. Tahir AM, Chowdhury ME, Khandakar A et al (2021) COVID-19 infection localization and severity grading from chest X-ray images. Comput Biol Med 139:105002. https://doi.org/10.1016/j.compbiomed.2021.105002
63. Tan M, Le Q (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. In: Chaudhuri K, Salakhutdinov R (eds) Proceedings of the 36th International Conference on Machine Learning, pp 6105–6114. https://proceedings.mlr.press/v97/tan19a.html. Accessed 5 Jun 2024
64. Tjoa E, Guan C (2021) A survey on explainable artificial intelligence (XAI): toward medical XAI. IEEE Trans Neural Netw Learn Syst 32(11):4793–4813. https://doi.org/10.1109/TNNLS.2020.3027314
65. Tong ZD, Tang A, Li KF et al (2020) Potential presymptomatic transmission of SARS-CoV-2, Zhejiang Province, China, 2020. Emerg Infect Dis 26:7181913–7181913. https://doi.org/10.3201/eid2605.200198
66 Varoquaux G, Cheplygina V (2022) Machine learning for medical imaging: methodological failures and recommendations for the future. npj Digit Med 5:48. https://doi.org/10.1038/s41746-022-00592-y
67. van der Velden BH, Kuijf HJ, Gilhuijs KG et al (2022) Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Med Image Anal 79:102470. https://doi.org/10.1016/j.media.2022.102470
68. Vilone G, Longo L (2020) Explainable artificial intelligence: a systematic review. https://doi.org/10.48550/arXiv.2006.00093
69. Wang L, Wong A (2020) COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-ray Images. Frontiers in Medicine 7. https://doi.org/10.3389/fmed.2020.608525. https://www.frontiersin.org/articles/10.3389/fmed.2020.608525
70. Wang L, Wong A, Lin ZQ et al (2019) Figure 1 COVID-19 Chest X-ray Dataset. https://github.com/agchung/Figure1-COVID-chestxray-dataset. Accessed 5 Jun 2024
71. Wehbe JRM, Sheng D et al (2021) An Artificial Intelligence Algorithm to Detect COVID-19 on Chest Radiographs Trained and Tested on a Large U.S. Clin Data Set Radiol 299:167–176. https://doi.org/10.1148/radiol.2020203511
72. Wiersinga WJ, Rhodes A, Cheng AC et al (2020) Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. JAMA 324(8):782–793. https://doi.org/10.1001/jama.2020.12839
73. Winther HB, Laser H, Gerbel S et al (2020) Covid-19-image-repository. https://github.com/ml-workgroup/covid-19-image-repository/tree/master/png
74. Yagin FH, Cicek IB, Alkhateeb A et al (2023) Explainable artificial intelligence model for identifying COVID-19 gene biomarkers. Comput Biol Med 154:106619. https://doi.org/10.1016/j.compbiomed.2023.106619
75. Yamac M (2021) URL Kaggle. https://www.kaggle.com/aysendegerli/qatacov19-dataset. Accessed 5 Jun 2024
76. Yanase J, Triantaphyllou E (2019) A systematic survey of computer-aided diagnosis in medicine: Past and present developments. Expert Syst Appl 138:112821. https://doi.org/10.1016/j.eswa.2019.112821
77. Yassin NI, Omran S, Houby EME et al (2018) Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. Comput Methods Programs Biomed 156:25–45. https://doi.org/10.1016/j.cmpb.2017.12.012
78. Young T, Hazarika D, Poria S et al (2018) Recent Trends in Deep Learning Based Natural Language Processing [Review Article]. IEEE Comput Intell Mag 13(3):55–75. https://doi.org/10.1109/MCI.2018.2840738
79. Yu P, Zhu J, Zhang Z et al (2020) A Familial Cluster of Infection Associated With the 2019 Novel Coronavirus Indicating Possible Person-to-Person Transmission During the Incubation Period. J Infect Dis 221(11):1757–1761. https://doi.org/10.1093/infdis/jiaa077. https://academic.oup.com/jid/article-pdf/221/11/1757/33202315/jiaa077.pdf