Machine Learning: Science and Technology

PAPER • OPEN ACCESS

CX-Net: an efficient ensemble semantic deep neural network for ROI identification from chest-x-ray images for COPD diagnosis

To cite this article: Agughasi Victor Ikechukwu and Murali S 2023 *Mach. Learn.: Sci. Technol.* **4** 025021

View the article online for updates and enhancements.

MACHINE
LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
15 January 2023

REVISED
5 April 2023

ACCEPTED FOR PUBLICATION
4 May 2023

PUBLISHED
18 May 2023

PAPER

CX-Net: an efficient ensemble semantic deep neural network for ROI identification from chest-x-ray images for COPD diagnosis

Agughasi Victor Ikechukwu[*] **and Murali S**

Department of CSE, Maharaja Institute of Technology Mysore, 571477 Karnataka, India
[*] Author to whom any correspondence should be addressed.

E-mail: victor.agughasi@gmail.com

Keywords: chest x-ray, COPD, CX-Net, ensemble learning, SHAP and Grad-CAM, VinDR-chest x-ray

Abstract

Automatic identification of salient features in large medical datasets, particularly in chest x-ray (CXR) images, is a crucial research area. Accurately detecting critical findings such as emphysema, pneumothorax, and chronic bronchitis can aid radiologists in prioritizing time-sensitive cases and screening for abnormalities. However, traditional deep neural network approaches often require bounding box annotations, which can be time-consuming and challenging to obtain. This study proposes an explainable ensemble learning approach, CX-Net, for lung segmentation and diagnosing lung disorders using CXR images. We compare four state-of-the-art convolutional neural network models, including feature pyramid network, U-Net, LinkNet, and a customized U-Net model with ImageNet feature extraction, data augmentation, and dropout regularizations. All models are trained on the Montgomery and VinDR-CXR datasets with and without segmented ground-truth masks. To achieve model explainability, we integrate SHapley Additive exPlanations (SHAP) and gradient-weighted class activation mapping (Grad-CAM) techniques, which enable a better understanding of the decision-making process and provide visual explanations of critical regions within the CXR images. By employing ensembling, our outlier-resistant CX-Net achieves superior performance in lung segmentation, with Jaccard overlap similarity of 0.992, Dice coefficients of 0.994, precision of 0.993, recall of 0.980, and accuracy of 0.976. The proposed approach demonstrates strong generalization capabilities on the VinDr-CXR dataset and is the first study to use these datasets for semantic lung segmentation with semi-supervised localization. In conclusion, this paper presents an explainable ensemble learning approach for lung segmentation and diagnosing lung disorders using CXR images. Extensive experimental results show that our method efficiently and accurately extracts regions of interest in CXR images from publicly available datasets, indicating its potential for integration into clinical decision support systems. Furthermore, incorporating SHAP and Grad-CAM techniques further enhances the interpretability and trustworthiness of the AI-driven diagnostic system.

1. Introduction

Medical imaging has become increasingly significant in the detection, diagnosis, and prognosis of disease because of advances in imaging technology. Computer-aided diagnostic (CAD), surgical planning, simulation, and robotically-assisted surgery are just a few medical imaging and software tools used in current surgical diagnosis. The importance of biomedical image processing has revolutionized various aspects of medicine, such as the precise diagnosis and staging of diseases [1]. However, medical images are blurry and distorted, making them more challenging to process than other image types [2]. Biomedical image analysis and processing have recently become an important research topic because of the difficulty in doing so quickly and precisely [3, 4]. Emphysema, tuberculosis, edema, aspiration pneumonia, pneumothorax, lung cancer, and recently COVID-19 are all lung diseases that can cause breathing difficulties or ARDS (acute respiratory distress syndrome). Lung diseases can also refer to diseases or disorders affecting the lungs, including the right and left upper, middle, and lower lung regions (ARDS). For example, tobacco abuse is the prime cause of chronic obstructive pulmonary disease (COPD), and a leading risk factor for lung cancer [5]. Over the past two decades, the proportion of individuals diagnosed with lung cancer in developing countries has significantly increased from 31% to 49.9% [6].

In contrast, developed nations like the US have experienced a decline in lung cancer cases [7]. Various environmental factors may contribute to the development of lung disorders such as asthma and cancer. It is crucial to recognize these factors and implement preventive measures to address the growing health concerns related to lung diseases. Airway inflammation, allergens, or pollutants can cause asthma symptoms, including a long-term respiratory illness with difficulty breathing. Asthma is an obstructive lung disease which is curable if diagnosed earlier.

COPD primarily refers to emphysema, chronic bronchitis, and intractable asthma. This family of pulmonary diseases that affect the lungs by limiting air passage to the alveoli threatens both developed and third-world countries, with a fatality rate close to 90% in low-income countries. As per the Global Initiative for Chronic Obstructive Lung Disease findings, COPD is the third major cause of death in 2020 [8, 9].

For medical image processing, various approaches have been used, such as x-rays, magnetic resonance imaging (MRI), endoscopic, ultrasonography, and thermal imaging [10, 11]. CXR is the gold standard for assessing and diagnosing lung diseases with over 2 billion scans performed annually. However, when compared to computed tomography (CT) scans and MRIs, CXR presents distinct issues because of its large dimensions and the frequent absence of labeled data, particularly for identifying regions of interest (ROIs) [12, 13].

A ROI is part of an image chosen for a specific purpose. ROI is typically used in medical imaging, for example, to identify a particular area of a 2D, 3D, or 4D image that requires clinical diagnosis. The lesion area is fascinating to doctors in medical imaging because it contains essential illness information, which doctors use to diagnose and design treatment plans. However, because the ROI includes vital information in the image, its proper extraction is critical and a significant difficulty in medical imaging.

Localization tasks such as detection and segmentation necessitate expensive labeling, a challenge for medical practitioners. For example, in finer-grained segmentation, the goal is to locate a ROI at pixel-level granularity by tracing its path. Deep neural networks (DNNs), which need many data to train, usually need annotated images to do either of these tasks: bounding boxes for detection and pixel-level masks for segmentation. Unfortunately, such annotated data can be prohibitively expensive to obtain. This situation is even worse with medical imaging because the labeler must be an experienced physician, resulting in a costly and time-consuming labeling operation.

Babenko [14] has looked into this task with little supervision in computer vision. While DNNs can categorize images with near-human accuracy, they are frequently considered 'black boxes', with the specific reasoning that the model lacks explanation, a step towards explainable artificial intelligence. For doctors to accept the results of algorithms used in medical imaging, even when vital findings are subtle and hard to spot, it is essential to explain what happened. Other approaches to providing such network explanations involve the use of saliency techniques, for example, gradient class activation mapping (Grad-CAM++) [15, 16] with pixel-wise heatmaps highlighting regions in the image that influenced the predicted classes. While these approaches provide explanations, they do not help with class optimization. Heatmaps are created using low-resolution filters (e.g. 5 × 5) and then propagated back to the input. Often results in occasionally coarse localizations, an area of concern for medical images such as CXR, frequently obtained at extremely high resolution (such as 3000 × 3000). A common convolutional neural network (CNN) strategy is downsampling CXR to the dimensions of some widely used pre-trained networks, such as AlexNet (e.g. 224 × 224). This approach may decrease the localization accuracy, inspired by the development of innovative segmentation and detection methods applicable to various fields, including healthcare.

It is hard to divide the lungs because CXRs often show changes in transparency or consolidation. Areas of the lung that overlap and extreme abnormalities caused by a bacterial infection, fluid buildup, or a lung condition often cause these changes. Since the lungs usually look very different from the rest of the body, it is common for people with pulmonary illnesses to struggle to differentiate between healthy lung tissue and the rest of the body.

To address these challenges, an automated approach that employs a four-fold ensemble to enhance the accuracy of lung region extraction by considering variations in lung structure across different diseases was proposed. Figure 1 illustrates the workflow for the proposed methodology.

In contrast to the standard object detection task, bounding boxes were not predicted. There are three main components to the planned work:

First, a chest x-ray (CXR) image is pre-processed and checked for outliers; then, using a blend of three pre-trained segmentation models (feature pyramid network (FPN), LinkNet, and U-Net) and a deep CNN

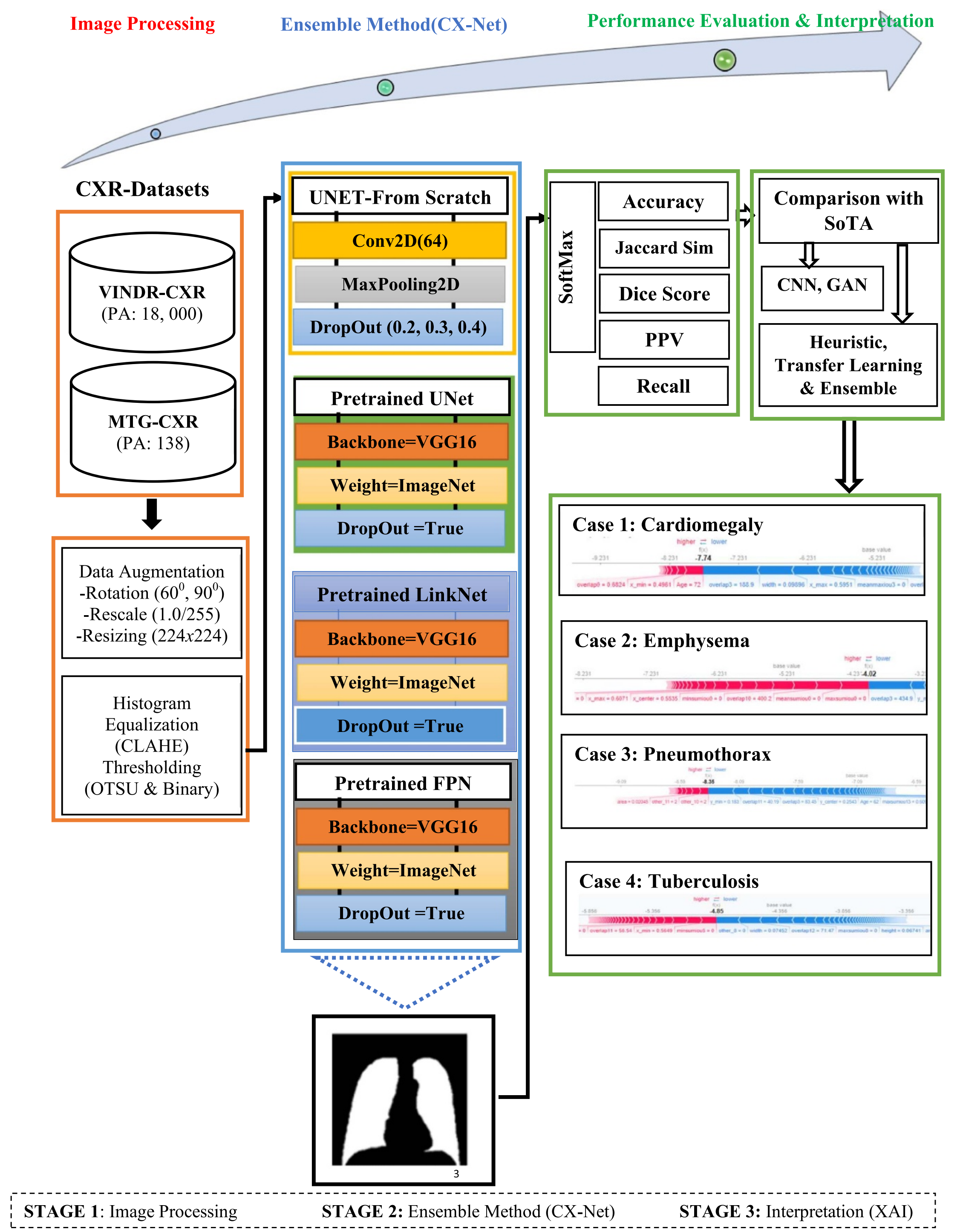


Figure 1. The proposed methodology workflow (details in section 3).

(DCNN) model trained from scratch called CX-Net, two pre-segments are generated. Finally, the pre-segments are merged to generate a final segmented lung region.

Contributions:

- The article presents an ensemble method called CX-Net, which combines four state-of-the-art (SOTA) CNN models (FPN, U-Net, LinkNet, and a locally trained U-Net) for improved lung segmentation in CXRs.
- The use of SHapley Additive exPlanations (SHAP) and gradient-weighted class activation mapping (GRAD-CAM) improves the model's explainability, enhancing the transparency and interpretability of the ensemble model.

- This research is the first to utilize the VinDr-CXR and Montgomery datasets for semantic lung segmentation with semi-supervised localization, demonstrating the model's robustness and adaptability.
- The proposed CX-Net achieved outstanding results in terms of Jaccard overlap similarity (JS), Dice coefficients (DCs), precision (PPV), recall, and accuracy, showcasing its superior performance compared to existing lung segmentation methods.
- Thorough analysis using various pre-processing techniques, highlights the choice of adaptive histogram equalization (HE) and the rationale behind excluding thresholding operations.
- We emphasize the importance of model ensembling in improving the identification of the ROI in CXR images, demonstrating its potential for integration into clinical decision support systems (CDSSs).
- The study contributes valuable insights into the detection and diagnosis of critical lung conditions such as pneumothorax, emphysema, and tuberculosis, demonstrating the potential impact of the proposed method on healthcare outcomes.

This work is organized in six sections. Section 1 highlights the background and motivation of the study. Section 2 presents related works on ROI segmentation using machine and deep learning approaches. Section 3 focused on the dataset and pre-processing techniques. The core methodology is introduced in section 4. Section 5 discusses the results of the experiments and how they stacked up against SOTA. Finally, section 6 concludes the work, stating the limitations and future directions.

2. Related work

The findings focused on the VinDR-CXR dataset from [16]. Seventeen experienced radiologists carefully annotated 18 000 images from the raw data, including 22 local labels for rectangles encircling abnormalities and six universal tags for suspected abnormalities. It consists of more than 100 000 CXR images from two leading Vietnamese hospitals, with a 15 000-person training set and a 3000-person test set in the dataset that has been made public. Three radiologists independently labeled the training set, while five radiologists labeled the test set. Simultaneously with the dataset's release, Nguyen *et al* [17] published the first benchmark for classifying and localizing common thoracic findings. This benchmark, on only the whole label images, shows that 10 606 CXR images had no medical condition.

Notably, lung segmentation is synonymous with the ROI identification from a clinical perspective. Researchers have shown interest in the topic and made significant steps toward better lung segmentation. Here, we present some of the seminal publications. On lung segmentation, the attention of researchers has been drawn to two distinct techniques, namely:

- Traditional or classical and
- Deep learning techniques.

Further, studies in deep learning-based techniques in COPD, COVID-19, and breast cancer research were presented.

2.1. Classical machine learning-based segmentation approaches in COPD

Wan Ahmad *et al* [18] proposed an approach to lung refinement based on the hybrid of oriented Gaussian derivatives, thresholding, and Fuzzy C-Means (FCM) clustering. On the JSRT dataset, their system achieved an accuracy of over 90%, except for the overlap, which stood at 87%. When it comes to lung segmentation, deformable model-based techniques include scale-dependent shape and appearance data by employing a joint shape and appearance sparse learning-based framework [19–24]. Lung segmentation is a challenging problem; several authors have developed hybrid approaches incorporating active shape models with other techniques [25, 26]. However, due to the variability in lung field forms, the lung boundaries obtained using these traditional segmentation methods may not be optimal. Furthermore, these algorithms perform poorly when dealing with pulmonary disorders that alter lung texture.

2.2. Deep learning-based segmentation approaches in COPD/emphysema

More than 35 deep learning-based studies on the ChestX-ray8 dataset [27], with 13 potentially generating an emphysema annotation label [28]. Nonetheless, most of these investigations use automatically derived tags that are noisy and unsuitable for clinical assessment [29]. The most recent study [30] of this type offered an extension of DenseNet121 [31] and reported an area under the curve (AUC) of 93.3% localizing emphysema. Moreso, the acknowledged problems with emphysema labeling in that dataset [32] make it difficult to interpret their findings unbiasedly. Using an encoder–decoder CNN (ED-CNN), Kalinovsky and Kovalev [33] proposed lung segmentation and achieved a maximum Dice score of 97.4% on a dataset of 354 CXRs.

The encoder in ED-CNN gradually decreases the spatial dimension of the input image. The ED-CNN decoder recovers object features and position and provides an output image with the lung probability of each pixel. Using lung shapes, Coppini *et al* [34] and Miniati *et al* [35] achieved 90% accuracy, with an AUC of 0.96, in detecting emphysema. These two research projects employ neural networks with custom-made features on well-curated datasets. Wanchaitanawong *et al* [36] recently suggested that AI-based emphysema scores from CXRs might be helpful for ailments where spirometry tests are not feasible and produce equivalent results in diagnosing COPD on a dataset that included 80 patients. Another effective strategy for raising segmentation precision is the use of ensemble approaches. For instance, by modifying the basic structure of the U-Net and InvertedNet models, Gomez *et al* [37] presented four distinct convolutional models. Instance normalization and atrous convolution are only underutilized methods incorporated into this approach. Souza *et al* [38] also used a similar strategy. They combined two deep learning networks, one of which can classify the CXR patches and the other of which may be able to reconstruct infected regions. This innovative method depends on how well the mask reconstruction stage works. However, in CXR images with significant abnormalities, the reconstruction stage often increases the false-positive rate due to an incorrect mask reconstruction.

2.3. Deep learning-based segmentation approaches in COVID-19 research

The outbreak of the Novel Corona Virus in December 2019 has shifted all the attention to COVID-19, churning out many breakthrough research articles. Sabre *et al* [39] proposed a deep learning architecture for CXR classification after screening COVID-19 patients, with protocols emphasizing ROI 'Hide-and-Seek' to validate the deep learning architecture. Inference from the test results validates the effectiveness of their approach. With approximately 16 000 CXR images of COVID-19 patients, including standard and infected cases, Karim *et al* [16] used an explainable DNN called 'DeepCOVIDExplainer' that could discriminate regions using gradient-guided class activation maps (Grad-CAM++). Based on data from hold-outs, the method could find COVID-19 with a higher positive predictive value (PPV). Alam *et al* [40] suggested a hybrid approach using a histogram of oriented gradient and a pre-trained CNN (VGGNet) from CXR images. Modified anisotropic diffusion filtering made it possible to remove noise, and an accuracy of 99.49% showed that the proposed method was better than other methods. Tahir *et al* [15] used a robust approach to discriminate ROI from plain CXRs images using Score-CAM (class activation mapping) visualization. Their system produced a sensitivity of 96.94%, a promising result for AI generalizability. Chaddad *et al* [41] proposed a DCNN model that successfully identified regions of interest that correspond to ground-glass opacities and pleural effusions from CT and CXR images. The segmented areas performed better using six pre-trained neural networks such as DenseNet, NASNet-Mobile, and DarkNet. They yielded an area under the receiver operating characteristic curve of almost 100%. The proposed method will help radiologists improve their diagnostic accuracy and manage COVID-19 in less time. Apostolopoulos and Mpesiana [42] employed ensemble CNNs and reported that MobilleNet v2 [43] produced better accuracy than InceptionV2. Their study on Covid-19 disease biomarker extraction using deep learning and x-ray imaging, with a sensitivity of 98.66%, shows the feasibility of using CXRs for clinical diagnosis. Using pre-trained networks with adequate fine-tuning for differentiating standard CXR images from COVID-19 infected was investigated by Narin *et al* [44] who employed three CNN models [43, 45, 46]. The results show that the pre-trained ResNet50 model has the best classification performance, with an accuracy of 96.1% compared to other models. Similarly, Horry *et al* [47] used a transfer learning approach for pre-processing and proposed a multi-modal classification model to identify COVID-19-infected CXR and CT images. Their findings reveal that VGG19 [48] better distinguished COVID-Pneumonia and standard images. A fine-tuned CNN-based model for pneumonia classification in CXR images was proposed by [49]. It showed that training a DNN from scratch on a low-end PC is doable, although it is computationally expensive. They explored the effects of hyperparameter tuning with dropout variations and achieved better accuracy than most standard methods.

2.4. Deep learning-based segmentation approaches in breast cancer research

Early detection of breast cancer improves the patient's survival rate to a greater degree. Ragab *et al* [50] proposed two approaches to ROI detection: manual and a DCNN, with an accuracy of 71.01%. Using samples from DCNN yielded an accuracy of 73.6%, an improvement of 2.05% over the manual approach. Similarly [51], Wei *et al* proposed a method for breast tumor classification on ultrasound images in which a professional radiologist provided the ground truth regions of interest. The photos were denoised using speckle-reducing anisotropic diffusion for better feature extraction. In their survey paper, Kwong and Mazaheri [52] effectively highlighted the research to identify regions of interest. Most researchers reported DCs and intercept over union (IoU) of more than 80%. Deep learning feature fusion and extreme learning machine (ELM) were investigated by Wang *et al* [53]. The fusion of CNN and ELM to cluster features of sub-regions was able to pinpoint the location of a breast tumor. Using CNN with unsupervised ELM

Table 1. Summary of selected related works on CXR datasets for disease localization.

Year/Authors/References	Dataset images	Image position (AP, PA)	Image format	Labeling method	Gold standard data
(2011) National Lung Screening Trial Research Team *et al* [56]	26 732	Not reported	Not reported	Nil	Nil
(2019) CXR14-Rad-Labels, Majkowska *et al* [57]	4374	AP: 3244 PA: 1132	PNG	Radiologist cohort study	All (4374)
(2019) CheXpert, Irvin *et al* [58]	224 000	AP: 162 000 PA: 29 000	JPEG	Radiologist cohort study	235
(2019) MIMIC-CXR, Johnson *et al* [59]	372 000	AP + PA: 250 000	DICOM, JPEG	Report parsing	Nil
(2020) PadChest, Bustos *et al* [60]	160 000	AP: 20 000 PA: 96 000	DICOM	Radiologists interpretation	27 593
(2020) COVID-CXR, Cohen *et al* [61]	866	AP:344 PA: 438	JPEG, PNG	Varies	Nil
(2021) VinDR-CXR, Nguyen *et al* [17]	18 000	AP:0 PA:18 000	DICOM	Radiologist interpretation	All (18 000)

(US-ELM) clustering, they present a method for mass identification that combines deep, morphological, and density characteristics simultaneously. Dewangan *et al* [54] employed a hybrid optimization technique for early breast cancer diagnosis with a hybrid herd African buffalo approach to discriminate between cancerous and normal MRI images. Chouhan *et al* [55] used taxonomic indices and local binary patterns for their analysis. A DCNN built on top of a highway network was employed to extract mammograms' fourth set of features dynamically. Emotional learning-inspired ensemble classifiers and support vector machine (SVM) were used alongside cross-validation to confirm the system's reliable performance. In summary, we present related literature on select datasets that influence the research work in table 1.

From table 1, the MIMIC-CXR dataset has the highest clinically annotated CXR images, followed by CheXpert. MIMIC-CXR is a dataset that includes 371 920 CXRs from patients and 64 588 patients. The CXRs obtained from patients admitted to Beth Israel Deaconess Medical Center's emergency department between 2011 and 2016. Later, an updated version of MIMIC-CXR known as V2 was made available [62], which featured both the anonymized radiology reports in the digital imaging and communications in medicine (DICOM) format. The PadChest dataset is a comprehensive collection of 160 868 CXR images from 109 931 examinations performed on 67 000 individuals. These images were acquired from San Juan Hospital in Spain between 2009 and 2017 [60]. The dataset offers a rich variety of grayscale images, each having a 16-bit depth and retaining their full resolution, which provides valuable insights for researchers and practitioners working with medical imaging. The attending physicians categorized 27593 of the reports manually. Per the medical standard practices, all the patient's data was stored using the DICOM format in VinDR-CXR, explained further in section 3 (Dataset description).

Literary findings: A review of existing literature reveals that numerous studies have explored datasets from sources such as KAGGLE, ChestX-ray14, CheXpert, MIMIC-CXR, PadChest, and COVID-CXR datasets. The traditional SoTA models, methods like K-Means, FCM clustering, and SVMs have shown significant improvements in accuracy. In addition, deep learning models consistently surpass classical machine learning approaches, owing to their multiple hidden layers and optimal hyperparameter selection. Overall, deep learning has achieved outstanding validation accuracies in semantic segmentation tasks. However, the performance in COPD region segmentation from CXR images remains less impressive, mainly when using the recently published, clinically-annotated VinDR-CXR dataset.

3. Materials and methods

Informed by the background information in sections 1 and 2, the objective is to employ a variety of intelligent models to extract lung regions from pulmonary images accurately. This study strives to develop a fully automated computational model, incorporating fine-tuning approaches to achieve high accuracy.

The proposed methodology combines multiple approaches and models for fine-tuned segmentation of lung regions in CXR images. The fully automatic model integrates various pre-trained networks with CX-Net, a DCNN model trained from scratch after image processing to generate highly relevant results.

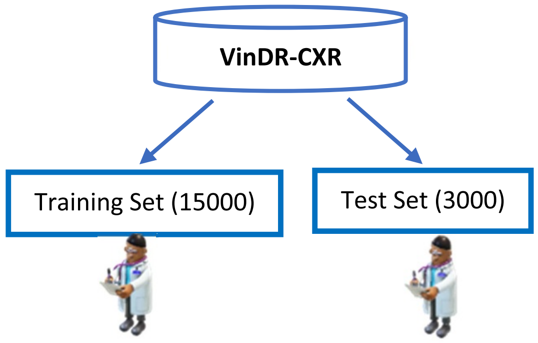


Figure 2. Distribution of the dataset.

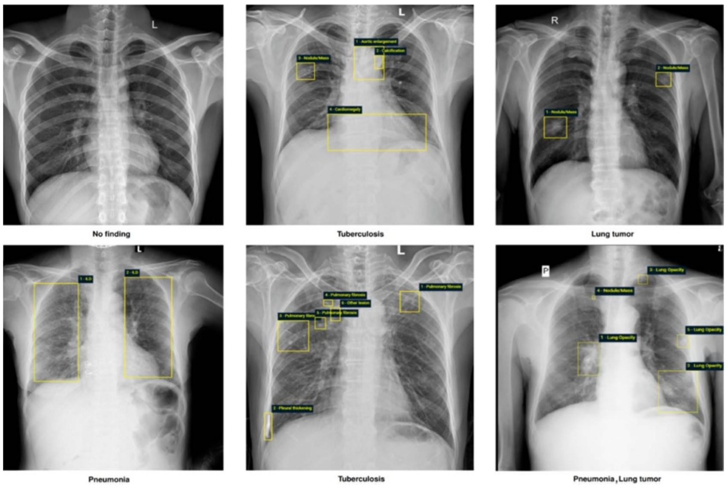


Figure 3. Samples of the dataset with abnormalities enclosed in yellow-rectangular bounding boxes. Reproduced from [17]. CC BY 4.0.

Figure 1 illustrates the proposed framework's primary stages. It includes image pre-processing, leveraging pre-trained CNN segmentation models, initial segmentation, and refinement leading to the final segmentation. This structured approach allows for a systematic and robust process, facilitating accurate and reliable lung segmentation results.

3.1. Dataset description

This study uses the clinically validated CXR dataset VinDR-CXR from [17]. This database includes over 100 000 CXRs from two of Vietnam's largest medical hospitals. It has about 18 000 images and a team of 17 experienced radiologists carefully labeled each one with 22 local labels of rectangles that enclose abnormalities and six global labels of suspected diseases. The public dataset comprises 30 000 records, 15 000 of which serve as training data and 3000 as test data, as depicted in figure 2. Selected samples of the images with bounding-box annotation are illustrated in figure 3. During the training phase, three radiologists labeled each scan independently, whereas a panel of five agreed to identify each scan in the test phase. Labels for the training and validation sets and all de-identified images are freely available in the DICOM format to ensure conformity to medical standards.

In figure 3, it is evident that the CXR images lacked segmented ground truth. The Montgomery dataset [24], which provided validated ground truth, was incorporated to overcome this challenge. We employed a semi-supervised approach to generate ground truth effectively. The dataset comprises 138 posteroanterior (PA) CXR images, including 80 normal and 58 abnormal cases that exhibit signs of tuberculosis. These PA CXR images were sourced from the tuberculosis control program in Montgomery, USA, and are collectively known as the Montgomery dataset. All images have been anonymized, removed identifying features, and are available in DICOM format. This dataset covers a range of abnormalities, such as effusions and miliary patterns.

Table 2. Parameters used for data augmentation.

Method	Default	Augmented
HorizontalFlip	None	True ($p = 0.5$)
VerticalFlip	None	True ($p = 0.5$)
Rescale (Normalization)	—	1./255
Zoom range	—	0.25
Rotation (^{o})	—	60, 90 & 120
x-Shift, y-Shift	None	[−0.1, +0.1]
x-Scale, y-Scale	None	[0.75, 1.25]
Adjusted image	1024 × 1024	224 × 224

The semi-supervised approach encompasses a two-stage process:

(a) Unsupervised clustering approaches

The distinctions and commonalities between two SoTA unsupervised clustering techniques were investigated. First, the generic Voronoi method (a broader term for K-Means) aims to discover k-partitions among data points that are well-shaped and uniformly distributed [63]. With each iteration, points are reassigned based on their distance from the calculated centroid. Second, we utilize an unsupervised neural network called a self-organizing map (SOM) to create a two-dimensional input data representation. Developed by Professor Teuvo Kohonen in the early 1980s [64], SOMs employ an unsupervised learning neural network trained with a competitive learning algorithm to generate subspaces. Neuron weights are adjusted based on their proximity to declared winner cells (i.e. neurons most closely resemble a sample input). Multiple input datasets during training form clusters of similar neurons while different neuron clusters are eliminated.

(b) Supervised classification

Fix and Hoges [65] introduced one of the earliest supervised classification methods, the K-nearest neighbor (KNN) classifier. When utilizing an annotated dataset, most of an unknown data sample's KNN is employed to determine its class label. The KNN classifier, for $k = 1$, is a specific instance of the classifier mentioned above. This approach excels in classification problems and possesses several attractive qualities, such as simplicity, efficiency, nonparametric nature, and speed [66]. Nevertheless, significant challenges arise in adjusting its parameters, like the neighborhood size (k) used in the topological representation quality, which can profoundly impact the result [67].

(i) Data augmentation: In computer vision, data augmentation plays a pivotal role in enhancing the performance of deep learning models by artificially expanding the training dataset. This technique effectively mitigates the challenges posed by limited data availability, a prevalent issue in implementing machine learning solutions. Examples of data augmentation include image augmentation for classification tasks and mask augmentation for segmentation tasks. By employing these methods, models can achieve better generalization and performance, addressing the constraints associated with time-consuming and costly data collection processes. The parameters chosen are in table 2.

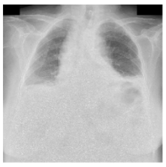
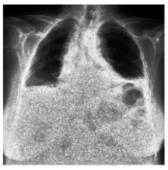
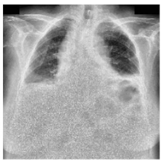
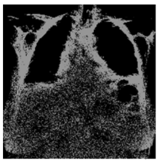
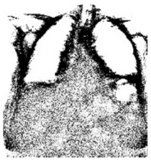
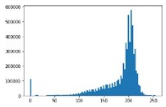
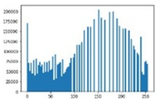
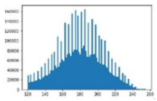
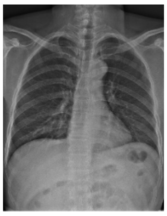
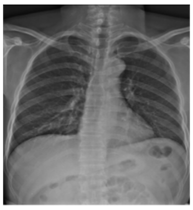
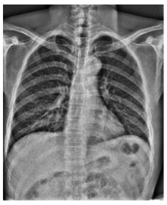
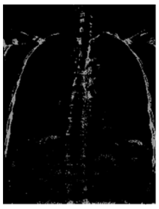
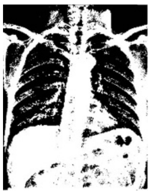
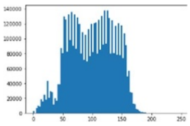
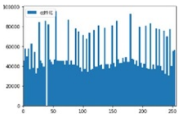
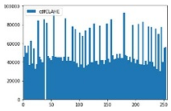
Initially, the images underwent flipping along the x-axis and were rescaled accordingly. Following this, each unique patch was rotated by 60, 90, and 180 degrees as part of the data augmentation technique, as depicted in table 2. The image dimensions were normalized to accommodate most pre-trained networks that perform optimally with 256 × 256 inputs. Multiples of 30 degrees were employed to avoid interpolation requirements. Allowing for rotations helped prevent overfitting. The aim is to elucidate the functioning of rotations within the CX-Net architecture, as they are frequently utilized to improve data. In total, 18 138 CXR images from the two datasets were analyzed.

3.2. Image pre-processing and enhancement

It is worth noting that the Montgomery CXR images had fixed dimensions of 4982 × 4020. At the same time, the VinDr-CXR dataset exhibited varying dimensions, with most images around 2500 × 2500. Given that all images were in DICOM format, a Python script was developed to convert DICOM to the corresponding photographic network group (PNG) format, as shown in table 3. To expedite training, all images were downscaled to 224 × 224, an approach that proved successful in segmenting microscopic white blood cells (WBCs) [68] for clinical diagnosis after normalization. This process ensures pixel intensity lies between 0 and 1, where 0 represents an entirely dark image, and 1 denotes an utterly bright image. The range of numbers between 0 and 1 denotes distinct tones of gray. This step is crucial when merging datasets with varying pixel value ranges from different sources. Subsequently, contrast-limited adaptive HE (CLAHE) was applied to enhance lung regions further.

The pre-processing techniques considered are crucial for improving the quality of the input images, as highlighted in table 3. For example, in the case of HE, the images appear darker and exhibit added noise,

Table 3. The result of our pre-processing approach.

Original input image (2525 × 2508)	Resized image (224 × 224)	Histogram equalization (HE)	CLAHE	Binary thresholding	OTSU thresholding
					
					
					
					

with an increased file size due to the consideration of global contrast rather than local contrast. As a result, adaptive HE was employed to address these issues.

Consider an image (f) that Xr and Yc represent with intensity values between 0 to $N-1$. Where N is the possible intensity value (max. of 256). Let Q denote a standardized histogram of f with a bin size of intensity values. Then,

$$Qn = \frac{\#\ \text{of pixel values}\ (n)}{\text{sum of pixel values}}\ \text{where}\ n\ =\ 0,1,2,\ \ldots,\ N-1. \tag{1}$$

The histogram-normalized image, denoted by the letter H, will be defined as:

$$Hi,j = \text{floor}\left((N-1)\sum_{n=0}^{fi,j} pn\right) \tag{2}$$

where the floor() function rounds down to the nearest integer. Then, transforming the intensities, m, of f by the function yields equation (3):

$$J(m) = \text{floor}\left((N-1)\sum_{n=0}^{m} pn\right). \tag{3}$$

Considering the intensities of f and H as continuous random variables X, β on the interval $[0, L-1]$ provides the impetus for this transformation. β is the sum of the intensities of f and g according to equation (4),

$$\beta = J(m) = (L-1)\int_{0}^{X} px(x)\,\mathrm{d}x, \tag{4}$$

where $px =$ the probability density function of f.

J is the cumulative distribution function of the product of X and $(L-1)$.

Factoring in the probabilities and partial derivatives, equation (4) becomes:

$$\frac{d}{dy}\left(\int_{0}^{y} py(z)\,\mathrm{d}z\right) = py(y) = \frac{px\left(T^{-1(y)}\right)d}{dy\left(T^{-1(y)}\right)}. \tag{5}$$

Equation (5) ensures that the distribution of pixel intensities is even across the input image.

In CLAHE, the transformation function's slope governs local contrast amplification. Consequently, the redistribution process may push some bins above the clip limit again (the region highlighted in blue in the histogram), leading to a practical clip limit that exceeds the desired limit. If this outcome is undesirable, the redistribution procedure can be iterated until the excess becomes negligible.

The need for a thresholding operation was deemed unnecessary, as the primary emphasis was on deep learning techniques. Furthermore, these advanced approaches possess the inherent ability to adapt and effectively manage variations within the input data, thereby optimizing the model's overall performance.

4. Proposed methodology

This section presents the modified U-Net model utilized for lung segmentation alongside the variants of pre-trained models comprising the ensemble approach. As per the literature, the U-Net architecture has found widespread application in segmenting medical images.

The input for the initial design of U-Net was a complete image, and the output was a predicted mask. Input images typically come in various resolutions, including 128 by 128, 256 by 256, and 512 by 512 pixels. These images go through four to five layers of convolutional down-sampling and up-sampling to be shrunk and expanded. However, this study used three different pre-trained segmentation models with a fully connected U-Net model trained from scratch. The input to the U-Net model was modified to 224 × 224, which invariably reduced the training time by 1%, and then applied batch normalization and dropout regularization variations. Throughout the literature, the model will be referred to as CX-Net, to imply that modifications to the original U-Net model were made. The block diagram for the proposed ensemble method is in figure 4.

4.1. Ensemble methods

Ensemble techniques encompass machine learning strategies that merge multiple base models into a single, more accurate predictive model. Bagging, stacking, and boosting represent the three primary methods within ensemble learning. Understanding and considering each approach when working on predictive modeling projects is essential. The term 'meta-learner' in this research refers to combining four distinct segmentation models into an 'ensemble', resulting in a well-segmented output. Generally, CNN architectures employ one of two ensemble techniques:

(a) Utilizing various CNN algorithms for feature extraction from medical images [69], and
(b) A hybrid approach that merges using a mathematical formula [70].

The benefit of using this method is that the ensemble technique accurately identifies the ROI from the images, resulting directly from precise predictions made using previous CNN models' outcomes. As shown in figure 5, the image is simultaneously provided to four functional layers (CX-Net, U-Net, Linknet, and FPN). The global max-pooling layer accepts each functional layer's learned representations. After merging the outputs of several parallel flows, the results are flattened and combined to create image feature vectors. Subsequently, layers are stacked to fine-tune the network, prevent overfitting, and enhance segmentation accuracy. Next, the four models' weights are standardized. Finally, the ensemble technique delivers the result using a fully connected layer.

4.2. CX-Net model

A detailed description of CX-Net, a modified version of the U-Net model, is presented in table 4, utilizing an input image of size 224 × 224 instead of the standard 256 × 256. CX-Net comprises five blocks each for both up-sampling and down-sampling. This study demonstrates the effectiveness of CX-Net, an ensemble of a network trained from scratch alongside U-Net, Linknet, and FPN, in segmenting lung fields in CXR images using an encoder–decoder network. During the training process, the network generates a mask for lung segmentation and learns higher-order structures, guiding the segmentation model through semi-supervised learning. It is achieved by leveraging images from the Montgomery dataset and transferring the learned features to images from VinDr-CXR without ground truth masks. This approach aids in identifying and addressing chest anomalies, such as pneumothorax and other structural deformities, while also proving invaluable in detecting chronic diseases like COPD.

The initial step in the CNN model involves extracting 64 × 64 pixel squares from the input image. Before creating the segmentation mask, the cropped areas from the original x-ray are classified into lung and non-lung regions. Patches are then successively clipped from the precise location in the original CXR and mask images. The ratio of lung pixels to non-lung pixels in the original image can be determined by

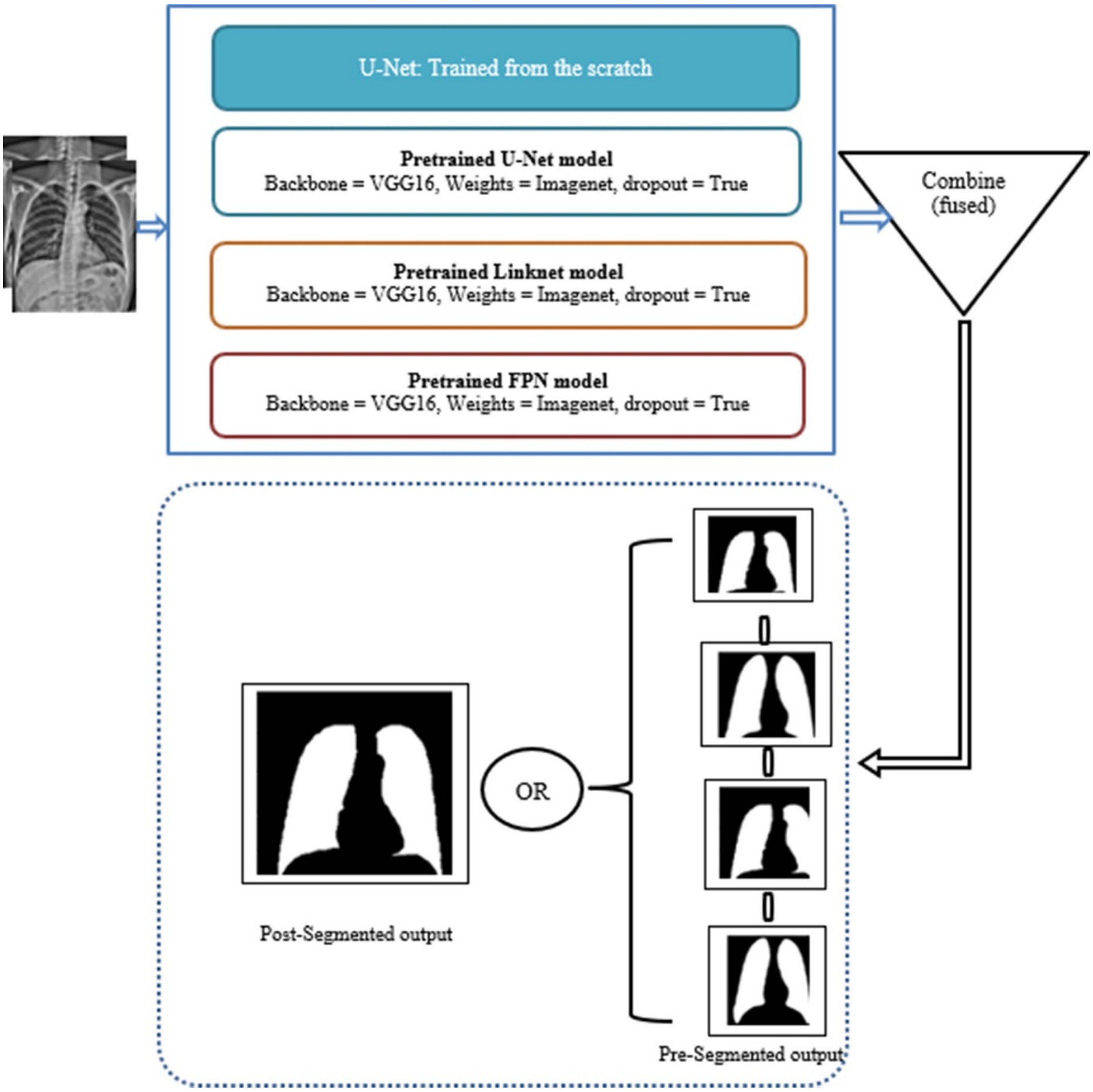

U-Net: Trained from the scratch
Pretrained U-Net model
Backbone = VGG16, Weights = Imagenet, dropout = True
Pretrained Linknet model
Backbone = VGG16, Weights = Imagenet, dropout = True
Pretrained FPN model
Backbone = VGG16, Weights = Imagenet, dropout = True
Combine (fused)
OR
Post-Segmented output
Pre-Segmented output

Figure 4. The proposed ensemble methodology (CX-Net) for CXR image segmentation.

comparing the cropped patches. For example, a specific area containing 40% or more lung tissue is classified as a lung patch, while a non-lung patch receives a different label. Extensive research has shown that a 40% cutoff indicates a lung patch. Once the CXR data patches are cropped and labeled, a CNN model is employed for classification.

Figure 5 illustrates the subsequent layers in the network, with kernel sizes of 2 × 2 for all convolution layers and 3 × 3 for all pooling layers. The four convolution layers saw progressive feature increases, with 64, 128, 256, and 256 features applied respectively. The use of ensembling on the validation sets resulted in effective lung region-wide segmentation. Extracting patches using an overlapping approach is crucial to obtain a more precise mask for segmenting lungs in test samples. This implies that a stride of 1 or 2 should be utilized for removing patches. The stride value directly correlates with the quality of the segmented lung contour. For instance, for a 224 × 224 image with a stride of 2, the CNN model must predict 28 800 patches to generate a complete segmentation mask for the entire CXR image.

Suppose an image patch corresponds to a lung. In that case, the central four pixels of the associated mask will be painted white (1 for each pixel), allowing the CNN model to yield a comprehensive x-ray segmentation. It is important to note that using a stride of 3 results in a slight degradation in smoothness compared to a stride of 2. The subsequent section demonstrates that the U-Net layer can restore the original contour's smoothness.

Table 4 outlines some layers and hyperparameters associated with the proposed model, while algorithm 1 explains the procedure. It is crucial to understand that the selection of hyperparameter tuning allowed us to obtain all trainable parameters. Consequently, the model was optimized for computational time, achieving a Dice score surpassing that of most SoTA models, as further elaborated in the discussion of the results.

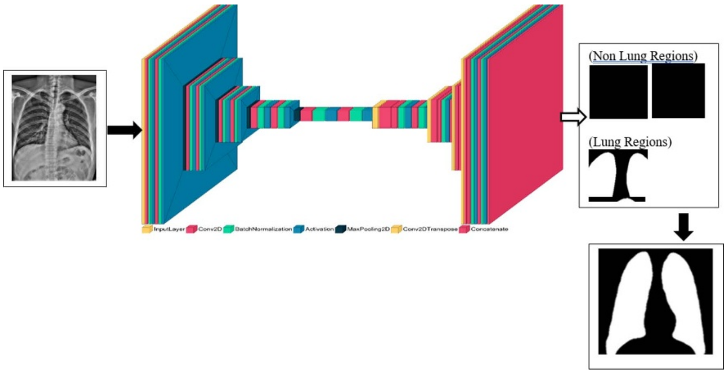


Figure 5. A U-Net model trained for CXR image segmentation.

Table 4. Summary of the model layers and parameters.

Layer (type)	Output shape	No. of parameters	Connected to
Input_3 (InputLayer)	(None, 224, 224, 3)	0	[]
Conv2d_38 (Conv2D)	(None, 224, 224, 64)	1792	[‘input_3[0][0]’]
Batch_normalization_18 (BatchNormalization)	(None, 224, 224, 64)	256	[‘conv2d_38[0][0]’]
Activation_18 (Activation)	(None, 224, 224, 64)	0	[‘batch_normalization_18[0][0]’]
Dropout_12 (Dropout)	(None, 224, 224, 64)	0	
Conv2d_39 (Conv2D)	(None, 224, 224, 64)	36 938	[‘activation_18[0][0]’]
Batch_normalization_19 (BatchNormalization)	(None, 224, 224, 64)	256	[‘conv2d_39[0][0]’]
Activation_19 (Activation)	(None, 224, 224, 64)	0	[‘batch_normalization_19[0][0]’]
Dropout_13 (Dropout)	(None, 224, 224, 64)	0	
Max_pooling2d_8 (MaxPooling2D)	(None, 128, 128, 64)	0	[‘activation_19[0][0]’]
Conv2d_40 (Conv2D)	(None, 128, 128, 12)	73 856	[‘max_pooling2d_8[0][0]’]
Batch_normalization_20 (BatchNormalization)	(None, 128, 128, 12)	512	[‘conv2d_40[0][0]’]
Activation_20 (Activation)	(None, 128, 128, 12)	08	[‘batch_normalization_20[0][0]’]
…	…	…	…
…	…	…	…
Conv2d_54 (Conv2D)	(None, 224, 224, 64)	73 792	[‘concatenate_11[0][0]’]
Batch_normalization_34 (BatchNormalization)	(None, 224, 224, 64)	256	[‘conv2d_54[0][0]’]
Activation_34 (Activation)	(None, 224, 224, 64)	0	[‘batch_normalization_34[0][0]’]
Conv2d_55 (Conv2D)	(None, 224, 224, 64)	36 928	[‘activation_34[0][0]’]
Batch_normalization_34 (BatchNormalization)	(None, 224, 224, 64)	256	[‘conv2d_55[0][0]’]
Activation_35 (Activation)	(None, 224, 224, 64)	0	[‘batch_normalization_35[0][0]’]
Conv2d_56 (Conv2D)	(None, 224, 224, 1)	65	[‘activation_35[0][0]’]

4.2.1. Pretrained U-Net model

The U-Net architecture is a well-employed CNN model for semantic segmentation that has been heavily used in biomedical image segmentation [71–73]. The original U-Net model, proposed by Ronnberger *et al* [71], accepts an entire input image and returns the masked image as an output. These can be 128 × 128, 256 × 256, or 512 × 512 pixels in size; however, this varies depending on the application. Images are encoded with four or five convolution layers and then decoded to the desired resolution. With 23 convolutional layers, two 3 × 3 convolutions are applied, followed by a rectified linear unit (ReLU) and a down-sampling 2 × 2 max pooling operation with stride 2. For the feature map to be up-sampled, the number of feature channels is cut in half using a 2 × 2 convolution, then expanded to the cropped feature map from the contracted path, and finally, a ReLU is applied to the combined map.

The entire image gets a pre-segmented mask. As a result, the initial U-Net design was reduced to keep the scaling while downsampling the image using VGG16 as the backbone and ImageNet as the encoder weights. Sigmoid was used for the activation as it outputs the probability of the images being of a lung region. Figure 6 displays the architecture of the U-Net model.

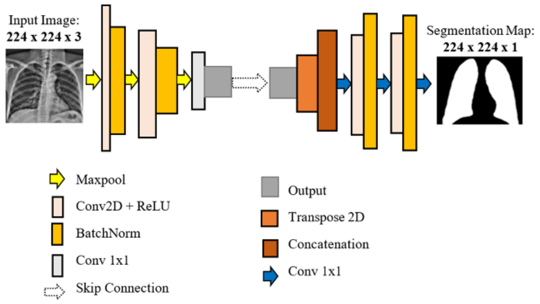


Figure 6. U-Net model for semantic segmentation.

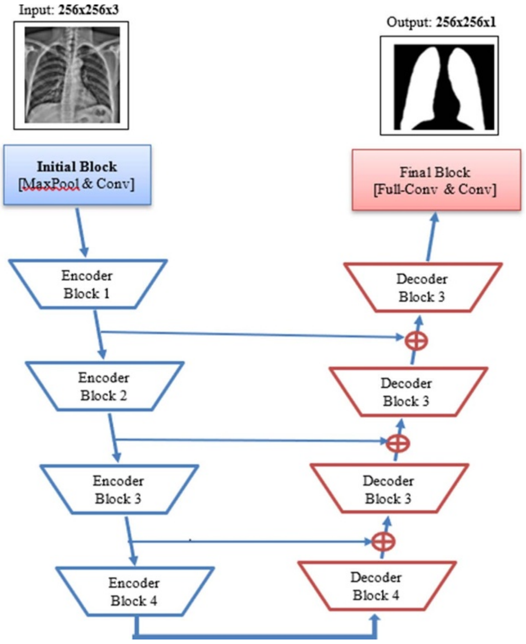


Figure 7. LinkNet model for semantic segmentation.

4.2.2. Pretrained LinkNet model

The second pre-trained network explored is the LinkNet, dubbed a full CNN for fast semantic segmentation [74]. It consists of 4 blocks each for up-sampling and down-sampling. The max-pool operation uses a 3 × 3 filter with a 7 × 7 convolutional size with batch normalization between each convolutional layer followed by a non-linearity function, ReLU [75, 76].

Input images usually convolve with a 7 × 7 kernel and a stride of 2 in the encoder's first block. This block likewise performs spatial max-pooling in a 3 × 3 area with a stride of 2. Further along, encoder blocks are residual blocks that make up the encoder. Figure 7 explains the internal layers of these encoder blocks. Newer segmentation methods rely on neural networks like VGG16 (with 138 million parameters and tens of thousands of floating-point operations per second) as their encoder. The unique thing about this architecture is how the links between each encoder and decoder were linked. It differs from how neural network models work for segmentation. Successive down-sampling processes in the encoder lead to the loss of some spatial information. Using the down-sampled output of the encoder, which does not have a trainable parameter, to retrieve this lost information is challenging.

The backbone networks were initially designed for classification tasks. The global average pooling layer and all fully connected layers were removed to make them suitable for semantic segmentation. The decoder components each have five blocks. The nearest-neighbor up-sampling layer is used for each level of the first four blocks to enlarge the image size by replicating a nearby pixel's value. Then, the nearest-neighbor up-sampling output is combined with the encoder's output. The feature map is fed into 3 × 3 convolutions, batch normalization, and ReLU activation layers. In the last block of the decoder, up-sampling is followed by

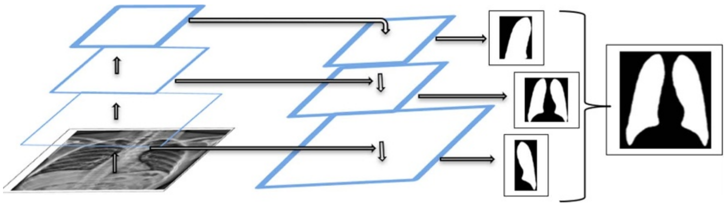


Figure 8. FPN model for CXR image segmentation.

two 3 × 3 convolutional layers. The number of channels reduces from 256 to 16 during the decoding process. The output is a mask that is constructed pixel by pixel and specifies the category of each pixel.

4.2.3. Pretrained FPN model

Besides the two pre-trained networks, a third pre-trained network, a FPN, was explored. FPN, as a feature extractor, accepts a single-scale image of arbitrary size as input and generates correspondingly sized feature maps at various layers in an entirely convolutional manner. This method is independent of the convolutional backbone architectures. As a result, it acts as a versatile approach for constructing feature pyramids within deep convolutional networks, which can be applied to problems such as object detection. The feedforward computation of the backbone ConvNet constitutes the bottom-up pathway. This pathway computes a feature hierarchy consisting of multiple-scale feature maps with a scaling step of 2. Each step of the feature pyramid corresponds to a single level of the pyramid. The output of the final layer of each stage serves as a reference set of feature maps, as depicted in figure 8.

To improve the functionality of the U-Net architecture, the original encoder was swapped for a 50-layer ResNeXt encoder that had been pre-trained using the ImageNet database [77]. The corresponding decoder was also modified to work with the new encoder. The ResNeXt50 encoder introduces a new building block that unifies transformations with a similar network, utilizes residual connections to enhance blocks of multiple convolution layers, and creates gradient shortcuts to lessen the likelihood of vanishing gradient problems. As a result, it makes it possible to train deeper network architectures, as highlighted in algorithm 1.

The CX-Net algorithm is designed to segment lung regions in CXR images. The algorithm consists of several key steps:

Image loader: Pre-processing and data augmentation are applied to all images in the VinDr-CXR and Montgomery datasets. Images are resized to 224 × 224 × 3, and CLAHE is performed. Then, data augmentation with 60°, 90°, and 120° rotations is applied, followed by normalizing pixel values.

Grouping lung and non-lung regions: The algorithm identifies whether a given image patch belongs to the lung or non-lung region. If the image patch is classified as a lung, it returns 1; otherwise, it returns 0.

Computing patches: The algorithm extracts image patches using stride lengths of 1 and 2. Feature maps are selected from these patches and forwarded to the proposed CX-Net model.

CX-Net model: The algorithm constructs a Conv2D model and adds activation functions, learning rate, and tuned hyperparameters. The image patches are then forwarded through pre-trained U-Net, LinkNet, and FPN models with VGG16 as the backbone and ImageNet weights.

Post-segmentation: After the segmentation process, the algorithm selects pre-segmented images, performs binary conjunction, and outputs merged images to complete the segmentation process. The purpose of this approach is to improve the precision of the segmentation. The CNN model proposed in figure 4 has proven effective in segmenting the lung region fairly. However, CXR images with severe anomalies produce an overall worse segmentation result. Combining the results of these four segmentations can help restore some of the lung tissue considered to be lost.

The overall algorithm effectively segments lung regions in CXR images, leveraging the power of an ensemble model and various pre-processing techniques to improve the accuracy of lung segmentation.

5. Results and discussion

This section presents the results of the proposed lung segmentation approach, beginning with the outcomes of both the initial and final segmentation stages. Performance metrics of various methods were compared using publicly available CXR imaging datasets, ultimately selecting the top three. Table 5 demonstrates that FPN, Linknet, and pre-trained U-Net models achieved the highest validation accuracy.

On the VinDr-CXR and Montgomery datasets, we evaluate the segmentation results of the proposed ensemble networks. The pre-trained models incorporating the ImageNet encoder (i.e. U-Net, FPN, and

Algorithm 1. CX-Net algorithm for CXR segmentation.

```
Begin-Procedure imageLoader(sources)
    for all images in sourcess (VinDr-CXR, Montgomery):
        resize ←(224 * 224 * 3)
            do:
                 pre-processing: CLAHE
                apply data augmentation ($60^\circ$, $90^\circ$, $120^\circ$)
                resize ← (imgPixels/max(255))
            end do
            find class(lung, non-lung)
                if 'lung' in class:
                    return 1 (lung)
                else:
                    return 0 (non-lung)
                end if
            end find
        computePatches(sources)
                for all images in normalized sources:
                    extract patches using strides of 1, 2
                        select feature map
                end for
            forward to proposed model (CX-Net):
            do:
                   I.   create conv2D model
                  II.   add activation, learning rate
                 III.   tune hyperparameters

            end do
            forward to model (U-Net):
            do:
                   I.   call the pretrained U-Net
                  II.   backbone = VGG16
                 III.   weight = ImageNet

            end do
            forward to model (LinkNet):
            do:
                   I.   call the pretrained U-Net
                  II.   backbone = VGG16
                 III.   weight = ImageNet

            end do
            forward to the next model (FPN):
            do:
                   I.   call the pretrained U-Net
                  II.   backbone = VGG16
                 III.   weight = ImageNet

           end do
     end computePatches
     procedure postSegmentation (testImages)
           do:
                   I.   select pre-segmented images
                  II.   perform binary conjunction
                 III.   output merged images

           end do
     end procedure
 end Begin-Procedure
```

Liknet) outperformed the model trained from scratch. We observed the best performance metric on the Montgomery dataset with ground truth (i.e. JS = 0.992, DC = 0.994, PPV = 0.993, and recall = 0.980) for lung field segmentation. However, compared to the U-Net model trained from scratch, some lung regions were improperly segmented (highlighted in red circles).

Table 5. Results of the lung segmentation on both VinDR-CXR and Montgomery datasets.

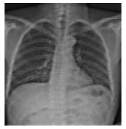
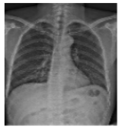
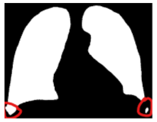
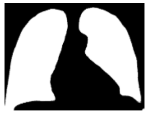
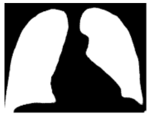
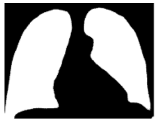
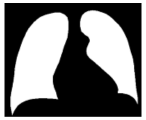
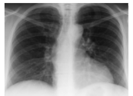
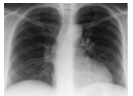
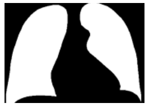
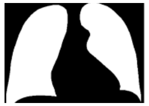
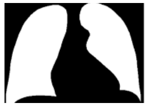
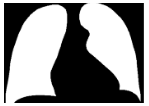
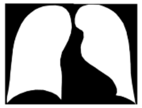
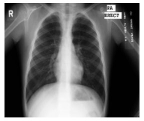
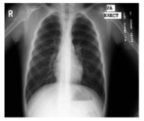
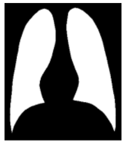
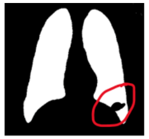
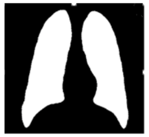
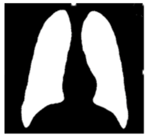
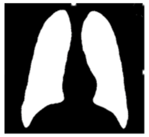
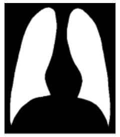
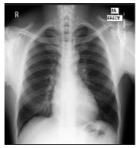
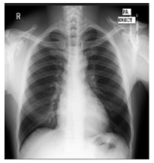
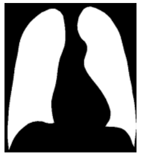
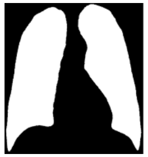
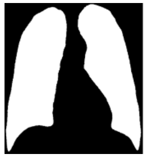
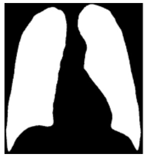
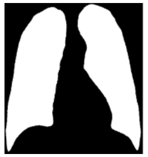
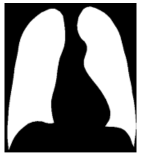
Dataset	Resized input image	After CLAHE	Ground truth	U-Net (from scratch)	U-Net (pre-trained)	FPN (pre-trained)	Linknet (pre-trained)	Final segmentation
VinDr-CXR			Nil					
			Nil					
Montgomery								
								

Table 8 compares the outcomes of incorporating ensembling to other methods using the VinDR-CXR dataset and a standard evaluation technique [78]. The segmentation result are displayed in table 5.

We then compared the results of various segmentation techniques to demonstrate the efficacy of the proposed approach. Additionally, we present a comparative analysis with chosen case studies from various authors to test the robustness of the proposed methodology.

5.1. Model explainability using SHAP and GRAD-CAM

SHAP is a suitable measure of feature importance that provides a robust and interpretable method for explaining individual predictions of complex machine learning models [79]. It is based on the concept of Shapley values from cooperative game theory and enables a consistent way to attribute the output of a model to its input features. By using SHAP, we can understand the influence of each feature on the model's predictions, enabling better transparency and interpretability. In the context of deep learning models for medical imaging, such as the CX-Net algorithm for CXR segmentation, SHAP can be employed to evaluate the contribution of each feature, such as specific regions or patterns within the image the final decision [80]. This information can help to identify potential biases, improve model performance, and instill trust in the model's predictions among practitioners.

Grad-CAM is another widely used technique for visualizing the interpretability of deep learning models, especially CNNs [81, 82]. Grad-CAM visually explains model predictions by highlighting essential regions in the input image, contributing to the final classification or segmentation decision. It does this by computing the gradients of the target class to the feature maps of the last convolutional layer, followed by a weighted combination of these gradients to produce a heatmap. This heatmap can be superimposed on the input image to reveal the salient regions that the model focuses on during its decision-making process. For example, applying Grad-CAM to the CX-Net algorithm for CXR segmentation would allow a deeper understanding of which regions within the x-ray images the model deems essential for accurate lung segmentation. Thus, providing valuable insights into the model's behavior and aiding in the interpretability of its predictions [83].

5.2. Select case studies on COPD segmentation

Lung segmentation is an important use case for a high-quality segmentation algorithm that generalizes well to various images. This section evaluates segmentation results for several COPD CXR images. We concentrated on prevalent illnesses requiring medical treatment, such as cardiomegaly, emphysema, pneumothorax, and tuberculosis.

Cardiomegaly (Case 1): This can be seen plainly on a CXR. A cardiothoracic window on a PA film of more than 50% is considered diagnostic of cardiomegaly. The enlarged heart's chamber of origin can be distinguished with the help of other x-ray findings in the chest. With the aid of the annotations file, a few samples that represented a generalization of the ailments are shown in figures 9–12. Experimental findings reveal that the model was robust to outliers with precision, recall, and Dice scores of more than 96%. The images are highlighted in three variants of colors that represent the ground truths (blue), the result of segmentation from a U-Net model (red), and the proposed model (orange).

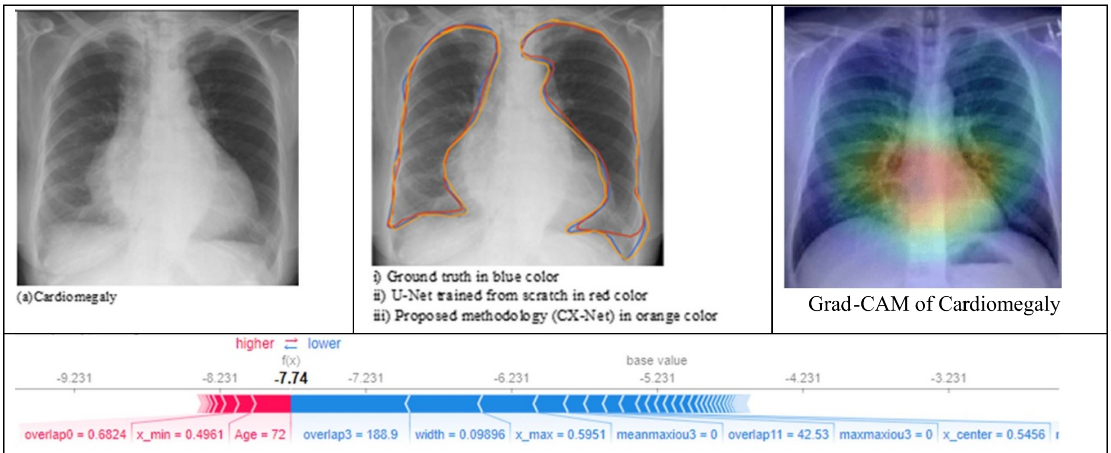
SHAP values indicate the contribution of each model feature to a prediction but do not reveal how the features contributed to the target variable. This is because a model may not accurately represent reality, and predictions can be incorrect. The values predicted for $f(x)$ were obtained using the equation $f(x) = \frac{\exp(x)}{1+\exp(x)}$. This is a sigmoid function with its logit function $f(x) = \log(y) - \log(1-y)$. Given that $f(x) = -8.85$, the probability expressed as a percentage is 13.4%

Emphysema (Case 2): Radiographic changes in the chest of a patient with moderate to severe emphysema include bilaterally hyperlucent lungs of considerable capacity, flattened hemidiaphragm with expanded costophrenic angles, horizontal ribs, and a narrow mediastinum.

Pneumothorax (Case 3): It is the medical word describing the presence of gas (usually air) in the pleural space (plural: pneumothoraces). A condition known as tension pneumothorax develops when the gas collection continually swells and presses on mediastinal tissues. Initial imaging, often a supine or semi-recumbent chest radiograph, may fail to detect an occult pneumothorax.

Tuberculosis (Case 4): Tuberculosis, also known as (TB) due to its frequent abbreviation, is a spectrum of diseases caused by *Mycobacterium tuberculosis* that can affect virtually every body part. *Mycobacterium bovis* can also cause a subset of cases.

Overall, the proposed approach performs better than existing methods on various datasets, illness types, and individual patient instances. There are occasions where the method's overall performance diminishes, yet it can still segment a sizable portion of the lung region. Although there is some variation in the results across clinical experts as regards the segmentation efficacy, the overall performance is relatively consistent, a step that validates the integration into a CDSS.



Explanation: In the given results, the SHAP values provide insight into the prediction made for the input image with the ID "d60854afebf749f87a8c95a07cb30d48". The model identifies the presence of "Cardiomegaly" with a class ID of 3. The radiologist associated with this prediction is **R9**, and the probability of the prediction being cardiomegaly is **0**.0204429. The first and second predicted classes are 0 and 3, with probabilities of 0.956254 and 0.0204429, respectively. The alternative prediction for this image is class 0, with the same probability as the first prediction (0.956254). The logit value, which represents the raw model output before applying the sigmoid function, is -3.86947. The prediction is flagged as accurate, indicating potential issues or attention points for further examination.

These findings suggest that the model's primary prediction is class 0, with a high probability of 0.956254, while cardiomegaly (class 3) is the second most probable prediction with a considerably lower probability of 0.0204429. The flagged status implies that this particular case may require additional scrutiny, possibly due to the complexity or uncertainty of the prediction. By analyzing the SHAP values and related information, clinicians and researchers can better understand the model's decision-making process, leading to improved interpretability, trustworthiness, and identification of potential areas for model improvement.

Grad-CAM addresses the issue of understanding the model's prediction by generating heatmaps that highlight the regions in the input image responsible for the predicted class, enabling better interpretability of the model's decision-making process. By using the gradients of the target class flowing into the final convolutional layer, Grad-CAM computes the importance of each feature map for a specific class prediction. This technique helps to identify which parts of the image contribute the most to the model's decision and assists in verifying if the model is focusing on the correct regions of interest, ultimately improving the trustworthiness and interpretability of the model.

Figure 9. Recall = 98.2, PPV = 97.4, DC = 96.9.

5.3. Parameters for training

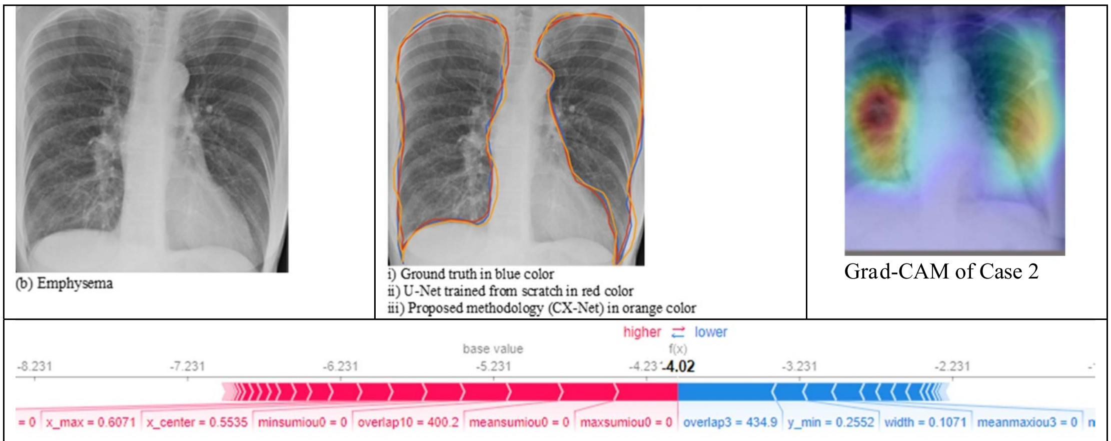
The model learning rate was 0.000 05 but fine-tuned to 0.000 01 after extensive experimentation. The size of each batch was 64, a good fit since we have more than 10 000 images to train. The number of epochs stood at 50. After the 21st epoch, the model did not improve further, and the learning rate automatically remained the same throughout the training, as presented in table 6. Adaptive moment estimation (ADAM) [84], a first-order gradient-based optimization of stochastic function, was used as an optimizer. The input image is further downsampled by a factor of 2 during the encoding phase. Figures 5–8 show the model with each part of the decoder sub-block. Data augmentation [85], a well-known regularization technique, was used to improve performance by reducing overfitting. The framework for augmentation in table 2 shows that various methods like vertical flip, normalization, and others improve the model's generalizability.

Utilizing the widely accepted '70:30 rule', the model allocated 70% of the images for training, with each image featuring dimensions of 224 × 224 pixels.

The remaining 30% of the data was reserved for testing the methodology. The training was conducted using the most recent version of Tensorflow (2.10.0) on a hardware setup consisting of an NVIDIA GeForce GTX 1060 GPU, an Intel Core i7-8750H processor operating at 2.2 GHz, and 16 GB of RAM. This arrangement allowed for efficient processing and optimal performance during the model's training and testing phases.

5.4. Evaluation of performance metrics

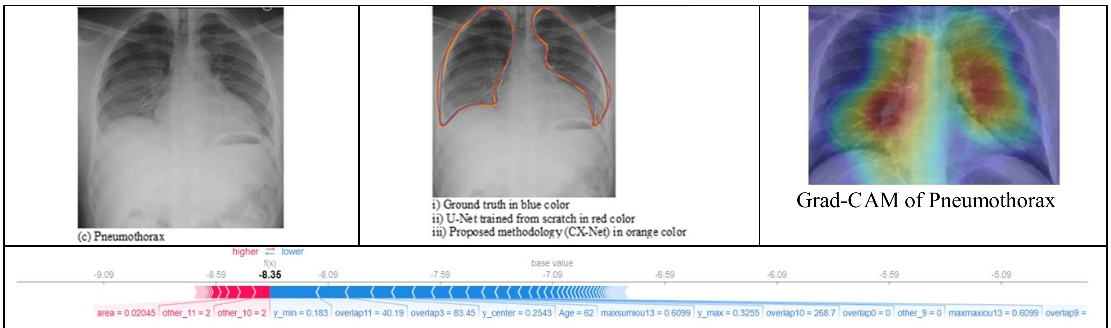
Our proposed methodology used the Jaccard similarity (JS) and the DC. These metrics have found useful applications in accessing the quality of segmentation models. Similarly, for the losses, Jaccard loss (JL) was used, Dice loss (DL), and binary cross entropy loss (BCE), respectively.



Explanation: The SHAP values provide insight into the prediction made for the input image with the ID "6ed2ab0e4ba47734a7b2a47deff243c1". The model identifies the presence of "Emphysema" with a class ID of 7. The radiologist associated with this prediction is R9, and the probability of the prediction being emphysema is 0.0127829. The first and second predicted classes are 0 and 3, with probabilities of 0.977964 and 0.0127829, respectively. The alternative prediction for this image is class 0, with the same probability as the first prediction (0.977964). The logit value, which represents the raw model output before applying the sigmoid function, is -4.34678. The prediction is flagged as true, indicating potential issues or attention points for further examination.

These findings suggest that the model's primary prediction is class 0, with a high probability of 0.977964, while cardiomegaly (class 3) is the second most probable prediction with a significantly lower probability of 0.0127829. The flagged status implies that this particular case may require additional scrutiny, possibly due to the complexity or uncertainty of the prediction.

Figure 10. Recall = 98.6, PPV = 97.6, DC = 96.3.

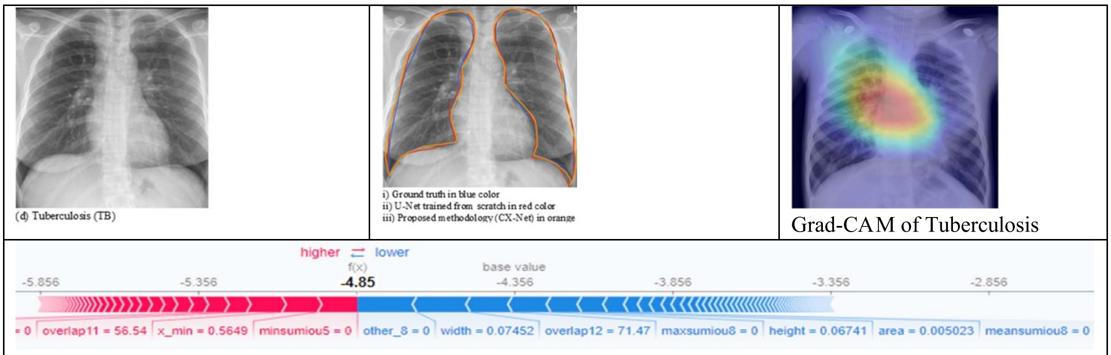


Explanation: The results obtained from the SHAP values for the image with ID 6e4391555899c8474c4d32f42b2ba21b provide insights into the prediction process for the class Pneumothorax (class_id 12). The radiologist ID associated with the prediction is **R10**. The probability of Pneumothorax presence in the image is **0.00319737**, which is relatively low. The model has predicted two other classes as more likely for this image, with class 13 (first_pred) having the highest probability of 0.861297, followed by class 6 (second_pred) with a probability of 0.0530703. The alternative class considered for this image is also class 13, with an alternative probability of 0.861297, which matches the first_prob value.

The logit value associated with the class Pneumothorax prediction is -5.74222, indicating that the model's confidence in this prediction is low. The negative logit value implies that the model leans more towards other classes as its prediction. The flagged attribute is set to "True," which highlights that the prediction for the Pneumothorax class may not be accurate or reliable, and further investigation or review may be needed. Overall, the SHAP values suggest that the model's prediction for pneumothorax in this particular image is not strong. It is more inclined to predict other classes, such as class 13, as more probable outcomes.

Figure 11. Recall = 98.9, PPV = 98.6, DC = 96.2.

Jaccard similarity (JS): To quantify the degree to which two samples are alike and dissimilar, statisticians use the JS coefficient, also called the Jaccard index or the IoU. It assesses the similarity between finite sample sets by dividing intersection by union. For example, given two sets (x, y), the JS can be expressed as:



Explanation: The results obtained from the SHAP values indicate a case where a chest X-ray image with the ID “0d30dc1e0070e7a934f39452e3ad3b83” was analyzed for the presence of Tuberculosis(TB). The image was reviewed by radiologist R8, and the probability of a TB in this image was 0.00475181. However, the first prediction (class ID 0) and second prediction (class ID 9) do not correspond to the TB class. The first prediction had a high probability of 0.944056, while the second prediction had a probability of 0.0228744. Therefore, the SHAP value’s logit for this case is -5.34447, and the case is flagged as valid.

These results imply that the model might have misclassified the presence of a TB in the analyzed image. The first and second predictions are inconsistent with the actual class of interest (TB), suggesting that the model’s predictions may not accurately represent the presence of this particular feature in the image. The fact that the case is flagged as true indicates that it might require further investigation to determine the presence or absence of a TB, potentially by a human expert or another more accurate model

Figure 12. Recall = 98.2, PPV = 97.7, DC = 96.2.

Table 6. Parameters for training.

Early stoppage (ES)	True
Learning rate (LR)	10^{-5} (0.000 01)
Patience	3
Batch size (BS)	64
Weight-decay	0.1
Momentum	0.9
Optimizer	Adam
Epochs	50
*Best epoch	21
System configuration	Nvidia GTX 1060, 6GB GPU

$$\mathrm{JS}_{(x,y)} = {}^{A\,n\,B}\!/_{A\,u\,B}, \tag{6}$$

where $\boldsymbol{n}$ = intersection and $\boldsymbol{u}$ = union of the two sets, respectively.

Equation (6) illustrates the operation:

$$\frac{\sum_{i=1}^{n}\left(x\left[i\right]*y\left[i\right]\right)}{\sum_{i=1}^{n}x\left[i\right]+\sum_{i=1}^{n}y-\ \sum_{i=1}^{n}\left(x\left[i\right]*y\left[i\right]\right)}. \tag{7}$$

For simplicity, JS(x,y) is:

$$\mathrm{JS}_{(x,y)} = \mathrm{TP}/\mathrm{FP}+\mathrm{TP}+\mathrm{FN}, \tag{8}$$

where TP, FP, and FN denote samples of true positives, false positives, and false negatives.

From figure 8, JS can be expressed as the ratio of the IoU, a commonly used metric in deep learning models.

Dice coefficient (DC): The DC, also known as ‘dice score or F-score,’ also quantifies the similarity like JS(x,y), but with a different weight on true positive as the best value is reported at 1 (00%), and the worst, a zero (0) score as shown in equation (9),

$$\mathrm{DC}_{(x,y)} = {}^{2\left|A\,n\,B\right|}\!/_{\left|A\right|+\left|B\right|}, \tag{9}$$

where $|A|$ and $|B|$ are the cardinalities of the two sets (i.e. the number of elements in each set).

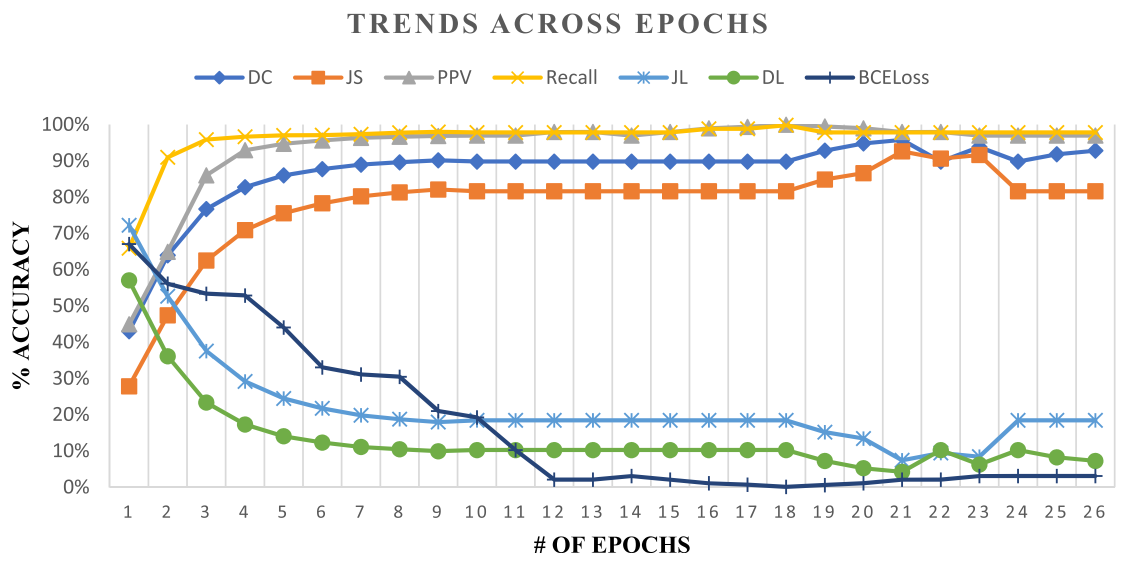


Figure 13. Results of ensemble segmentation methods for CX-Net on the CXR dataset.

When applied to Boolean data, using the definition of true positive (TP), false positive (FP), and false negative (FN), as in equation (10):

$$\xi_{\mathrm{DS}} = 2\mathrm{TP}/2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}. \tag{10}$$

Jaccard loss (JL): The JL or IoU loss optimizes the segmentation metric like DL. The Tversky loss gives FN and FP differing weights, while DL gives them equal weights. A criterion to measure the loss is:

$$\mathrm{JL}_{(x,y)} = 1 - {}^{|A\,n\,B|}/_{|A\,u\,B|} \tag{11}$$

Dice loss (DL): The criterion to measure the DL is:

$$\mathrm{DL}_{(x,y)} = 1 - {}^{2\,|A\,n\,B|}/_{|A| + |B|} \tag{12}$$

Binary cross entropy loss (BCE_{Loss}): In the context of lung segmentation, $\mathrm{BCE}_{\mathrm{Loss}}$ provides the following definition of the generic loss function; it creates a criterion that compares the ground truth (gt) with the prediction (p),

$$\mathrm{BCE}_{\mathrm{loss}} = -\sum \left[\mathrm{Agt}\,(x) \log\left(P\sim(x)\right) + \left(1 - \mathrm{Agt}\,(x)\right) \log\left(1 - P\sim(x)\right)\right], \tag{13}$$

where Agt(x)$\in \{0{,}1\}$ = the actual ground truth segmentation label of the pixel x and, $P \sim (x)$ = the predicted probability of x being the lung regions.

During training, we tracked the 'precision and recall' amongst other hyperparameters to see how it improved over time. Figure 13 captures the trends across various epochs.

Precision, also known as the PPV, addresses the question: What proportion of identifications was correct? For example, a model that produces no false positives has a precision of 1.0. The criterion is as follows:

$$\text{Precision (PPV)}_{(x,y)} = \frac{\mathrm{TP}_{(x,y)}}{\mathrm{TP}_{(x,y)} + \mathrm{FP}_{(x,y)}} \tag{14}$$

where TP = true positive.

Sensitivity, also known as recall, accounts for the actual positives identified correctly. It is defined mathematically as:

$$\text{Sensitivity (Recall)}_{(x,y)} = \frac{\mathrm{TP}_{(x,y)}}{\mathrm{TP}_{(x,y)} + \mathrm{FN}_{(x,y)}} \tag{15}$$

where FN = false positive.

Figure 13 represents the performance metrics across 50 epochs. For clarity's sake, we present only the first 26 epochs with the Dice score, Jaccard score, precision, and recall that stood above 90%.

Table 7. Segmentation accuracy (evaluated by Jaccard, Dice, and BCE) of models using two different datasets.

		Metrics				Losses		
Model	Dataset	JS	DC	PPV	Recall	JL	DL	$\text{BCE}_{(\text{Loss})}$
CX-Net	VinDr-CXR	0.926	0.958	0.978	0.978	0.074	0.042	0.020
	Montgomery	0.925	0.955	0.976	0.976	0.075	0.045	0.020
FPN	VinDr-CXR	0.980	0.974	0.969	0.975	0.020	0.026	0.020
	Montgomery	0.982	0.970	0.963	0.971	0.018	0.030	0.020
LinkNet	VinDr-CXR	0.982	0.970	0.963	0.971	0.018	0.030	0.020
	Montgomery	0.986	0.974	0.983	0.961	0.014	0.026	0.020
U-Net	VinDr-CXR	0.982	0.984	0.983	0.980	0.008	0.006	0.030
	Montgomery	0.992	0.994	0.993	0.980	0.008	0.006	0.041

Note: Only the best evaluation metrics are in table 4—the best metrics obtained at the 21st epoch.

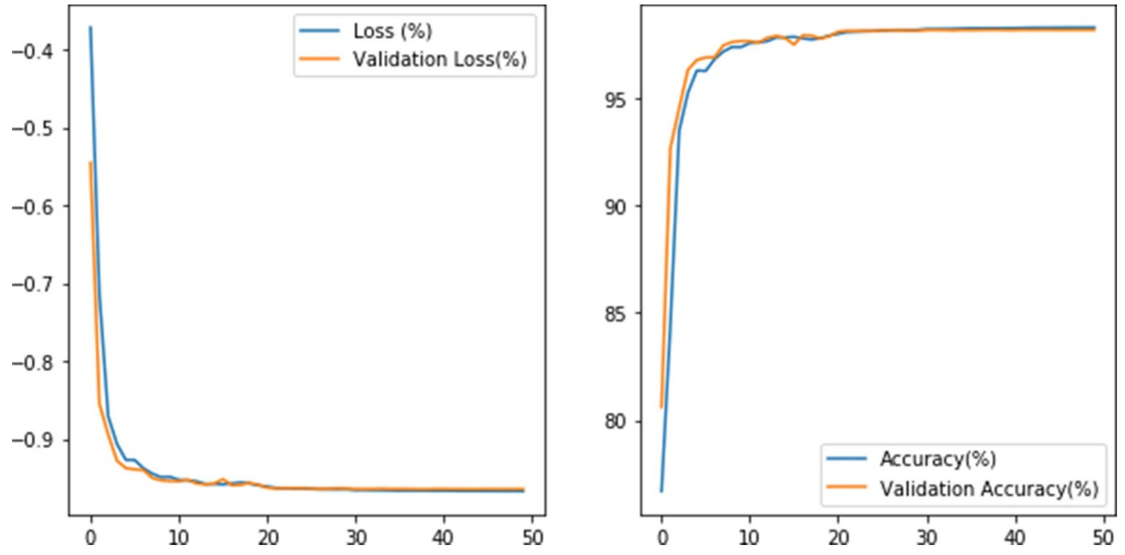


Figure 14. Loss vs. accuracy curve.

As illustrated in figure 13, the Recall metric improved significantly for the training, from 67% to 99.8%. This is a promising result, as it suggests that the model is increasingly improving at identifying the desired ROIs from the CXR images. The Dice score, IoU, and PPV metrics also showed similar improvements, supporting this conclusion. It is worth noting that the Recall metric decreased slightly at the 26th epoch, but this is likely because the model was starting to overfit the training data. Overall, these results are very encouraging, and they suggest that the model has the potential to be very effective at lung segmentation.

The JS metric peaked at the 21st epoch but suddenly dropped to 81.5% at the 24th epoch. The Dice and Jaccard losses followed a similar trajectory, with minima at the 21st epoch (2%). However, the BCE loss showed abnormal behavior, sharply increasing at the 24th epoch. This increase is likely because the model was beginning to learn to predict the ground truth labels rather than the actual labels. Early stoppage with a patience factor of 3 ensured the model did not overfit during the training phase.

For an adequate comparison, each model is trained and evaluated on the VinDR-CXR and Montgomery CXR datasets. The VinDR-CXR dataset uses the contrast enhancement method (CLAHE), as explained in table 3. The results are in table 7.

Table 7 shows a similar trend across all three pre-trained segmentation models. Thus, we evaluate the accuracy and loss over 50 epochs, as presented in figure 14.

Accuracy, also known as the error rate, is a metric that measures the number of correct predictions to the sum of predictions. It is expressed mathematically as:

$$\text{Accuracy} = \frac{\text{True}_P + \text{True}_N}{\text{True}_P + \text{True}_N + \text{False}_P + \text{False}_N} \tag{16}$$

where p = positive; n = negative.

The accuracy ranges from 0 to 1 and is a fraction of the percentage (i.e. accuracy * 100).

Table 8. Comparison with SoTA.

Authors/Year	Dataset	Problem scope	Technique	Methodology	Novelty	JS (%)	DC (%)	PPV (%)	Recall (%)	Accuracy (%)
Rajpurkar *et al* (2017) [86]	ChestX-ray14	Lung segmentation	Deep CNN	CNN	Custom architecture for lung segmentation	NR	95.0	NR	NR	85.0
Novikov *et al* (2018) [87]	JSRT	Lung segmentation	Deep CNN	CNN	Custom architecture for lung segmentation	95.0	97.0	NR	NR	NR
Ayan and Unver (2019) [88]	Pneumonia CXR Dataset	Lung segmentation	Transfer learning	Pre-trained CNN	Transfer learning on pneumonia CXR dataset	NR	87.0	88.0	94.0	87.0
Chen *et al* (2020) [89]	Kinetics-400 & Kinetics-600	Lung segmentation	CNN & dilated CNN	CNN with dilation	Dilation in CNN for improved performance	NR	NR	NR	NR	96.3
Munawar *et al* (2020) [90]	JSRT	Lung segmentation	GAN	GAN	GAN-based approach for lung segmentation	94.3	97.4	NR	NR	NR
Salehi *et al* (2021) [91]	Pediatric CXR dataset	Lung segmentation	Transfer learning	Pre-trained CNN	Transfer learning on pediatric CXR dataset	NR	89.0	NR	91.0	83.0
Khan *et al* (2021) [92]	LIDC-IDRI & Lung-PET-CT	Lung nodules segmentation	Transfer learning	VGG19	VGG19 with handcrafted SegNet model	82.7	90.5	98.4	84.1	**99.7**
Rehman *et al* (2021) [93]	NIH-CXR & COVID-19 x-ray	Lung segmentation	Deep CNN	Transfer learning	32-layered CNN with transfer learning	NR	NR	**99.3**	**99.7**	**99.7**
Jaszcz *et al* (2022) [94]	Chest x-ray	Lung segmentation	Heuristics	RFOA	Heuristic red fox optimization algorithm (RFOA)	94.4	97.1	NR	NR	97.2
Ours [CX-Net] (2023)	VinDR-CXR & Montgomery	Lung segmentation	Ensemble learning	Ensemble of CNN models	Combining multiple CNN models for segmentation	**99.2**	**99.4**	**99.3**	98.0	97.6[b]

[a] Where NR = not reported; JSRT = Japanese Society of Radiological Technology; GAN = generative adversarial networks.

[b] *Note*: Bold represent the better performance metric.

Similarly, the loss captures the rate at which error rates plummet *t* the minimum. It is inversely proportional to the accuracy. For a good (learnable model), the accuracy tends toward 100%, while the loss tends toward zero.

The use of callbacks' early stoppage with a patience of 3 ensures that the model stopped training when the metric did not improve beyond the best-recorded value. In our experiment, the best accuracy was at the 21st epoch. The accuracy was 97.6%, while the loss was around 2.4%.

5.5. Comparison with relevant literature

The proposed methodology is studied carefully and compared to other SoTA methodologies in this section. In addition, the outcomes of this investigation to those of successful prior studies using different methods and algorithms are in table 8. Segmentation techniques have matured to where accuracy rates consistently exceed 92% across all performance measures. The proposed method uses four DNNs to detect bacterial infections like tuberculosis, pneumothorax, and emphysema.

Rajpurkar *et al* [86] proposed a dense CNN model with 121 layers for the discrimination of various thoracic diseases. On the Pneumonia CXR image, an accuracy of 85% was achieved, which was on par with clinical validation.

Multiclass segmentation of CXR images was investigated by Novikov *et al* [87] using the publicly available Japanese Society of Radiological Technology (JSRT), which achieved commendable results in differentiating the lungs and heart from clavicles.

Two popular CNN methods, Xception and VGG16, were used by Ayan and Unver [88]. Experimental results show that VGG16 was better than the Xception model by a margin of 5%.

Chen *et al* [89] employed an orthogonal model with complementary CNNs to reduce channel-wise redundancy and achieved an accuracy of 83% on the ImageNet challenge.

Munawar *et al* [90] employed generative adversarial networks for the segmentation of lung regions for disease diagnosis. An IoU of 94.3 outperformed most of the literature that used the same dataset (JSRT).

Using four pre-trained models (such as DenseNet121), Salehi *et al* [91] investigated an approach for the detection and classification of pediatric pneumonia with a classification accuracy of 86.8%.

Table 8 compares the proposed method, CX-Net, with SoTA techniques in lung segmentation. The table includes information on the dataset used, problem scope, technique, methodology, novelty, JS, DC, PPV, Recall, and Accuracy. Various approaches are employed, ranging from DCNN, transfer learning, and GAN, to heuristics. The novelty of each method are highlighted, such as custom architectures, dilation in CNN, and the use of heuristic algorithms.

The proposed method, CX-Net, utilizes an ensemble learning approach on the VinDR-CXR and Montgomery datasets. It combines multiple CNN models for segmentation, resulting in impressive performance metrics: JS (99.2%), DC (99.4%), PPV (99.3%), Recall (98.0%), and Accuracy (97.6%). In addition, the table demonstrates that the proposed approach performs better than other techniques in lung segmentation, indicating its potential for practical applications in medical imaging and COPD diagnosis.

5.6. Limitation

Focusing primarily on segmenting PA-CXR images, this study does not consider methods for AP-CXR; as a result, lateral CXR images are excluded from the research. It is acknowledged that various modalities, such as CT scans, can effectively capture lung regions, providing a quantitative view of the lungs and being regarded as the 'gold standard' by radiologists and physicians. However, real-time data are often messy and heterogeneous, making it uncertain how well the proposed model generalizes to these datasets, as with many deep learning methods. In addition, other clinically validated datasets, such as JSRT, MC, and University of Texas Medical Branch (UTMB), were not investigated. Consequently, future research will expand the model to include these modalities and datasets.

6. Conclusion and future work

In this work, a deep convolutional network (CNN) was trained to generate regions of interest, commonly known as segmented regions, from CXR images with better accuracy than most existing SoTA ones. To further enhance the precision of segmentation in PA-CXR images, a method using was proposed using deep learning techniques (CX-Net). The approach used four parallel deep-learning models to generate the pre-segmentation masks. Therefore, the overall accuracy of the segmentation increased by using the pre-segmented masks when processing CXRs from patients with pulmonary disease. Notably, this framework can train CNNs well even with relatively modest data. Furthermore, various case studies and datasets extensively examined the proposed model's performance. As a result, the framework achieves better results than most SoTA currently explored. Across all validation sets, we saw averages of 99.2% Jaccard score, 99.4% Dice similarity score, 99.3% precision, 98.0% recall, and 97.6% accuracy.

This work contributes primarily in four ways. First, it introduced a strategy for reducing the space required to store images in datasets. Second, the risk of over-fitting is reduced by employing variations of augmentation, batch normalization, and dropout. In particular, it shows that the proposed stages of contrast enhancement and image binarization improve faster convergence while requiring less data storage, resulting in only a 0.9% decrease in prediction accuracy (99.1%). Third, a novel computational model (CX-Net) that improved lung region segmentation using an adaptive U-Net model was proposed. Fourth, based on relevant literature, we are the first to employ the proposed approach to the VinDR-CXR dataset. Finally, to validate the efficacy of the proposed pre-processing strategy, experiments was conducted on the VinDR-CXR and Montgomery datasets using four popular CNN-based segmentation models. From experiments, we conclude that using the pre-processed version of the dataset (VinDR-CXR) improves training convergence by 20.5% and reduces storage space utilization by 75% on average compared to the original dataset (VinDR-CXR). Finally, this work is tailored towards developing a functional CAD module that will aid radiologists in diagnosing COPD and other related ailments.

Data availability statements

The two datasets used for this research were obtained from the clinically validated PhysioNet Database (https://physionet.org/content/vindr-cxr/1.0.0/), and https://data.lhncbc.nlm.nih.gov/public/Tuberculosis-Chest-X-Ray-Datasets/Montgomery-County-CXR-Set/MontgomerySet/index.html.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions

Agughasi Victor Ikechukwu: Conceptualization, Investigation, Data collection, Design, Writing—original draft. Writing—review and editing, Analysis and Interpretation of results.

Murali S: Study Conception, Supervision, Investigation on challenges and Draft manuscript verification.

Conflict of interest

The authors declare no conflict of interest.

Funding

This research received no external funding.

Appendix

S. No.	Abbreviation	Description
1.	ALL	Acute lymphoblastic leukemia
2.	BMP	Bitmap
3.	CADS	Computer-aided diagnostic system
4.	CBC	Complete blood count
5.	CDSS	Clinical decision support system
6.	CNN	Convolutional neural network
7.	CT	Computed tomography
8.	CXR	Chest x-ray images
9.	$\mathrm{FN}_{\mathrm{all}}$	False negative of ALL
10.	$\mathrm{FP}_{\mathrm{all}}$	False positive of ALL
11.	FPR	False positive rate
12.	GAN	Generative adversarial network
13.	GCS	Global contrast stretching
14.	Grad-CAM	Gradient-weighted class activation mapping
15.	GPU	Graphics processing unit
16.	HSI	Hue, saturation, and intensity
17.	HSV	Hue, saturation, and value
18.	HVN	Hypercomplex-valued network
19.	ILSVRC 2015	ImageNet large-scale visual recognition
20.	JSRT	Japanese Society of Radiological Technology
21.	LIDC-IDRI	Lung image database consortium and image database resource initiative
22.	LR	Learning rate
23.	M	Momentum
24.	MM	Multiple myeloma
25.	NIH-CXR	National Institute of Health Chest X-Rays
26.	PET-CT	Positron emission tomography computed tomography
27.	PPV	Positive predictive value
28.	RFOA	Red fox optimization algorithm
29.	RBC	Red blood corpuscles
30.	ReLU	Rectified linear unit
31.	ResNet	Residual networks
32.	ROC	Receiver operating curve
33.	SHAP	SHapley Additive exPlanations
34.	SOTA	State-of-the-art
35.	SVM	Support vector machines
36.	TCIA	The Cancer Imaging Institute
37.	TN	True negative
38.	TP	True positive
39.	TPR	True positive rate
40.	TPU	Tensor processing unit
41.	2D	Two dimensional
42.	VGG	Visual geometry group networks
43.	WBC	White blood corpuscles

ORCID iD

Agughasi Victor Ikechukwu https://orcid.org/0000-0002-1175-3089

References

[1] Antonelli M *et al* 2021 The medical segmentation decathlon

[2] Lehman T M and Brendo J 2005 Strategies to configure image analysis algorithms for clinical usage *J. Am. Med. Inform. Assoc.* **12** 497–504

[3] Geldermann I, Grouls C, Kuhl C, Deserno T M and Spreckelsen C 2013 Black box integration of computer-aided diagnosis into PACS deserves a second chance: results of a usability study concerning bone age assessment *J. Digit. Imaging* **26** 698–708

[4] Deserno T M, Handels H, Maier-Hein K H, Mersmann S, Palm C, Tolxdorff T, Wagenknecht G and Wittenberg T 2013 Viewpoints on medical image processing: from science to application *Curr. Med. Imaging Rev.* **9** 79–88

[5] Cheng T Y D, Cramb S M, Baade P D, Youlden D R, Nwogu C and Reid M E 2016 The international epidemiology of lung cancer: latest trends, disparities, and tumor characteristics *J. Thoracic Oncol.* **11** 1653–71

[6] Lim J U and Yoon H K 2022 Narrative review: association between lung cancer development and ambient particulate matter in never-smokers *J. Thoracic Dis.* **14** 553–63

[7] Subramanian J and Govindan R 2007 Lung cancer in never smokers: a review *J. Clin. Oncol.* **25** 561–70

[8] Agustí B, Beasley A and Celli R 2019 Pocket guide to COPD diagnosis, management, and prevention: a guide for health care professionals (Atlanta, GA: Global Initiative Chronic Obstructive Lung Disease) pp 1–43

[9] Moran I *et al* 2023 Deep transfer learning for chronic obstructive pulmonary disease detection utilizing electrocardiogram signals *IEEE Access* **11** 40629–44

[10] Niazi M K K, Parwani A V and Gurcan M N 2019 Digital pathology and artificial intelligence *Lancet Oncol.* **20** e253–61

[11] Bejnordi B E, Litjens G, Timofeeva N, Otte-Holler I, Homeyer A, Karssemeijer N and van der Laak J A 2016 Stain specific standardization of whole-slide histopathological images *IEEE Trans. Med. Imaging* **35** 404–15

[12] Rozenberg E, Freedman D and Bronstein A 2019 Localization with limited annotation for chest x-rays (arXiv:1909.08842)

[13] Rozenberg E, Freedman D and Bronstein A A 2021 Learning to localize objects using limited annotation, with applications to thoracic diseases *IEEE Access* **9** 67620–33

[14] Babenko B, Yang M H and Belongie S 2011 Robust object tracking with online multiple instance learning *IEEE Trans. Pattern Anal. Mach. Intell.* 1619–32

[15] Tahir A M, Qiblawey Y, Khandakar A, Rahman T, Khurshid U, Musharavati F, Islam M T, Kiranyaz S, Al-Maadeed S and Chowdhury M E H 2022 Deep learning for reliable classification of COVID-19, MERS, and SARS from chest x-ray images *Cogn. Comput.* **14** 1752–72

[16] Karim M R, Döhmen T, Rebholz-Schuhmann D, Decker S, Cochez M and Beyan O 2020 DeepCOVIDExplainer: explainable COVID-19 diagnosis based on chest x-ray images (arXiv:2004.04582)

[17] Nguyen H Q *et al* 2022 VinDr-CXR: an open dataset of chest x-rays with radiologist's annotations (arXiv:2012.15029)

[18] Wan Ahmad W S H M, Zaki W M D W and Ahmad Fauzi M F 2015 Lung segmentation on standard and mobile chest radiographs using oriented Gaussian derivatives filter *Biomed. Eng. OnLine* **14** 20

[19] Shao Y, Gao Y, Guo Y, Shi Y, Yang X and Shen D 2014 Hierarchical lung field segmentation with joint shape and appearance sparse learning *IEEE Trans. Med. Imaging* **33** 1761–80

[20] Xu T, Mandal M, Long R, Cheng I and Basu A 2012 An edge-region force guided active shape approach for automatic lung field detection in chest radiographs *Comput. Med. Imaging Graph.* **36** 452–63

[21] Cootes T F, Edwards G J and Taylor C J 2001 Active appearance models *IEEE Trans. Pattern Anal. Mach. Intell.* **23** 681–5

[22] Seghers D, Loeckx D, Maes F, Vandermeulen D and Suetens P 2007 Minimal shape and intensity cost path segmentation *IEEE Trans. Med. Imaging* **26** 1115–29

[23] Tianli Y, Jiebo L and Ahuja N 2005 Shape regularized active contour using iterative global search and local optimization *2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05)* (*San Diego, CA, USA*) vol 2 pp 655–62

[24] Candemir S, Jaeger S, Palaniappan K, Musco J P, Singh R K, Xue Z, Karargyris A, Antani S, Thoma G and McDonald C J 2014 Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration *IEEE Trans. Med. Imaging* **33** 577–90

[25] Iakovidis D K, Savelonas M A and Papamichalis G 2009 Robust model-based detection of the lung field boundaries in portable chest radiographs supported by selective thresholding *Meas. Sci. Technol.* **20** 104019

[26] van Ginneken B and ter Haar Romeny B M 2000 Automatic segmentation of lung fields in chest radiographs *Med. Phys.* **27** 2445–55

[27] Wang X, Peng Y, Lu L, Lu Z, Bagheri M and Summers R M 2017 ChestX-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases *2017 IEEE Conf. on Computer Vision and Pattern Recognition* (CVPR) pp 3462–71

[28] Çallı E, Sogancioglu E, van Ginneken B, van Leeuwen K G and Murphy K 2021 Deep learning for chest x-ray analysis: a survey *Med. Image Anal.* **72** 102125

[29] Calli E, Scholten E T, Murphy K, van Ginneken B and Sogancioglu E 2019 Handling label noise through model confidence and uncertainty: application to chest radiograph classification *Medical Imaging 2019: Computer-Aided Diagnosis* (*San Diego, USA*) p 41

[30] Wang H, Wang S, Qin Z, Zhang Y, Li R and Xia Y 2021 Triple attention learning for classification of 14 thoracic diseases using chest radiography *Med. Image Anal.* **67** 101846

[31] Huang G, Liu Z, van der Maaten L and Weinberger K Q 2017 Densely connected convolutional networks *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (*Honolulu, HI*) pp 2261–9

[32] Oakden-Rayner L 2020 Exploring large-scale public medical image datasets *Acad. Radiol.* **27** 106–12

[33] Kalinovsky A and Kovalev V 2016 Lung image segmentation using deep learning methods and convolutional neural networks *Int. Conf. on Pattern Recognition and Informtion Processing* (*Minsk, Belarus State University, 3–5 October*) p 21–24

[34] Coppini G, Miniati M, Monti S, Paterni M, Favilla R and Ferdeghini E M 2013 A computer-aided diagnosis approach for emphysema recognition in chest radiography *Med. Eng. Phys.* **35** 63–73

[35] Miniati M, Coppini G, Monti S, Bottai M, Paterni M and Ferdeghini E M 2011 Computer-aided recognition of emphysema on digital chest radiography *Eur. J. Radiol.* **80** e169–175

[36] Wanchaitanawong J 2021 A predictive model using artificial intelligence on chest radiograph in addition to history and physical examination to diagnose chronic obstructive pulmonary disease *J. Med. Assoc. Thai.* **104** 79–87

[37] Gómez O, Mesejo P, Ibáñez O, Valsecchi A and Cordón O 2020 Deep architectures for high-resolution multi-organ chest x-ray image segmentation *Neural Comput. Appl.* **32** 15949–63

[38] Souza J C, Bandeira Diniz J O, Ferreira J L, França da Silva G L, Corrêa Silva A and de Paiva A C 2019 An automatic method for lung segmentation and reconstruction in chest x-ray using deep neural networks *Comput. Methods Programs Biomed.* **177** 285–96

[39] Sadre R, Sundaram B, Majumdar S and Ushizima D 2021 Validating deep learning inference during chest x-ray classification for COVID-19 screening *Sci. Rep.* **11** 16075

[40] Alam N-A-A, Ahsan M, Based M A, Haider J and Kowalski M 2021 COVID-19 detection from chest x-ray images using feature fusion and deep learning *Sensors* **21** 1480

[41] Chaddad A, Hassan L and Desrosiers C 2021 Deep CNN models for predicting COVID-19 in CT and x-ray images *J. Med. Imaging* **8** 014502

[42] Apostolopoulos I D and Mpesiana T A 2020 Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks *Phys. Eng. Sci. Med.* **43** 635–40

[43] Sandler M, Howard A, Zhu M, Zhmoginov A and Chen L-C 2018 MobileNetV2: inverted residuals and linear bottlenecks *2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 4510–20

[44] Narin A, Kaya C and Pamuk Z 2021 Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks *Pattern Anal. Appl.* **24** 1207–20

[45] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (*Las Vegas, NV, USA*) pp 770–8

[46] Tsiknakis N *et al* 2020 Interpretable artificial intelligence framework for COVID-19 screening on chest x-rays *Exp. Ther. Med.* **20** 727–35

[47] Horry M J, Chakraborty S, Paul M, Ulhaq A, Pradhan B, Saha M and Shukla N 2020 COVID-19 detection through transfer learning using multi-modal imaging data *IEEE Access* **8** 149808–24

[48] Simonyan K and Zisserman A 2015 Very deep convolutional networks for large-scale image recognition (arXiv:1409.1556)

[49] Victor Ikechukwu A and Murali S 2021 ResNet-50 vs. VGG-19 vs. training from scratch: a comparative analysis of the segmentation and classification of pneumonia from chest x-ray images *Glob. Transit. Proc.* **2** 375–81

[50] Ragab D A, Sharkas M, Marshall S and Ren J 2019 Breast cancer detection using deep convolutional neural networks and support vector machines *PeerJ* **7** e6201

[51] Wei M, Du Y, Wu X, Su Q, Zhu J, Zheng L, Lv G and Zhuang J 2020 A benign and malignant breast tumor classification method via efficiently combining texture and morphological features on ultrasound images *Comput. Math. Methods Med.* **2020** 1–12

[52] Kwong T and Mazaheri S 2022 A survey on deep learning approaches for breast cancer diagnosis (arXiv:2109.08853)

[53] Wang Z, Li M, Wang H, Jiang H, Yao Y, Zhang H and Xin J 2019 Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features *IEEE Access* **7** 105146–58

[54] Dewangan K K, Dewangan D K, Sahu S P and Janghel R 2022 Breast cancer diagnosis in an early stage using novel deep learning with hybrid optimization technique *Multimed. Tools Appl.* **81** 13935–60

[55] Chouhan N, Khan A, Shah J Z, Hussnain M and Khan M W 2021 Deep convolutional neural network and emotional learning based breast cancer detection using digital mammography *Comput. Biol. Med.* **132** 104318

[56] The National Lung Screening Trial Research Team 2011 Reduced lung-cancer mortality with low-dose computed tomographic screening *N. Engl. J. Med.* **365** 395–409

[57] Majkowska A *et al* 2020 Chest radiograph interpretation with deep learning models: assessment with radiologist-adjudicated reference standards and population-adjusted evaluation *Radiology* **294** 421–31

[58] Irvin J *et al* 2019 CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison *Proc. AAAI Conf. on Artificial Intelligence* vol 33 pp 590–7

[59] Johnson A E W *et al* 2019 MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs (arXiv:1901.07042)

[60] Bustos A, Pertusa A, Salinas J-M and de la Iglesia-vayá M 2020 PadChest: a large chest x-ray image dataset with multi-label annotated reports *Med. Image Anal.* **66** 101797

[61] Cohen J P *et al* 2020 Predicting COVID-19 pneumonia severity on chest x-ray with deep learning *Cureus* **12** e9448

[62] Johnson A E W, Pollard T J, Berkowitz S J, Greenbaum N R, Lungren M P, Deng C-Y, Mark R G and Horng S 2019 MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports *Sci. Data* **6** 317

[63] Jain A K 2010 Data clustering: 50 years beyond K-means *Pattern Recognit. Lett.* **31** 651–66

[64] Kohonen T 1990 The self-organizing map *Proc. IEEE* **78** 1464–80

[65] Fix E and Hodges J L 1989 Discriminatory analysis. Nonparametric discrimination: consistency properties *Int. Stat. Rev*. **57** 238

[66] Torralba A, Fergus R and Freeman W T 2008 80 million tiny images: a large data set for nonparametric object and scene recognition *IEEE Trans. Pattern Anal. Mach. Intell.* **30** 1958–70

[67] Dudani S A 1976 The distance-weighted k-nearest-neighbor rule *IEEE Trans. Syst. Man Cybern.* **SMC-6** 325–7

[68] Victor Ikechukwu A and Murali S 2022 i-Net: a deep CNN model for white blood cancer segmentation and classification *Int. J. Adv. Technol. Eng. Explor.* **9**

[69] Datta Gupta K, Sharma D K, Ahmed S, Gupta H, Gupta D and Hsu C-H 2021 A novel lightweight deep learning-based histopathological image classification model for IoMT *Neural Process. Lett.* **55** 205–28

[70] Kassani S H, Kassani P H, Wesolowski M J, Schneider K A and Deters R 2019 Classification of histopathological biopsy images using ensemble of deep learning networks (arXiv:1909.11870)

[71] Ronneberger O, Fischer P and Brox T 2015 U-Net: convolutional networks for biomedical image segmentation (arXiv:1505.04597)

[72] Liu W, Luo J, Yang Y, Wang W, Deng J and Yu L 2022 Automatic lung segmentation in chest x-ray images using improved U-Net *Sci. Rep.* **12** 8649

[73] Mique E and Malicdem A 2020 Deep residual U-Net based lung image segmentation for lung disease detection *IOP Conf. Ser. Mater. Sci. Eng.* **803** 012004

[74] Chaurasia A and Culurciello E 2017 LinkNet: exploiting encoder representations for efficient semantic segmentation *2017 IEEE Visual Communications and Image Processing (VCIP)* pp 1–4

[75] Ioffe S and Szegedy C 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift (arXiv:1502.03167)

[76] Nair V and Hinton G E Rectified linear units improve restricted Boltzmann machines p 8

[77] Deng J, Dong W, Socher R, Li L-J, Kai L and Fei-Fei L 2009 ImageNet: a large-scale hierarchical image database *2009 IEEE Conf. on Computer Vision and Pattern Recognition* (*Miami, FL*) pp 248–55

[78] van Ginneken B, Stegmann M B and Loog M 2006 Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database *Med. Image Anal.* **10** 19–40

[79] Lundberg S M, Erion G and Lee S I 2021 Consistent individualized feature attribution for tree ensembles *Nat. Commun.* **12** 1–9

[80] Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D 2021 Grad-CAM: visual explanations from deep networks via gradient-based localization *Int. J. Comput. Vis.* **128** 336–59

[81] Victor Ikechukwu A, Sreyas P, Sena A, Preetham H and Raksha K 2022 Explainable deep learning model for Covid-19 diagnosis *IRJMETS* **04** 3051–9

[82] Datta S, Bhole A, Ghosh S and Chakraborty C 2021 Explainable AI and ML in image analysis for COVID-19 detection: a review *J. Ambient Intell. Humaniz. Comput.* **123** 1–23

[83] Li H, Wu X, Zhang L and Sun J 2021 Explainable medical image recognition based on joint gradient-weighted class activation mapping *Neural Comput. Appl.* **33** 4619–31

[84] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization (arXiv:1412.6980)

[85] Pham V T and Nguyen T P 2023 Identification and localization COVID-19 abnormalities on chest radiographs *3rd Int. Conf. on Artificial Intelligence and Computer Vision (AICV2023)* (*5–7 March 2023*) vol 164 pp 251–61

[86] Rajpurkar P *et al* 2017 CheXNet: radiologist-level pneumonia detection on chest x-rays with deep learning (arXiv:1711.05225)

[87] Novikov A A, Lenis D, Major D, Hladuvka J, Wimmer M and Buhler K 2018 Fully convolutional architectures for multiclass segmentation in chest radiographs *IEEE Trans. Med. Imaging* **37** 1865–76

[88] Ayan E and Unver H M 2019 Diagnosis of pneumonia from chest x-ray images using deep learning *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)* (*Istanbul, Turkey*) pp 1–5

[89] Chen Y *et al* 2019 Drop an octave: reducing spatial redundancy in convolutional neural networks with octave convolution (arXiv:1904.05049)

[90] Munawar F, Azmat S, Iqbal T, Gronlund C and Ali H 2020 Segmentation of lungs in chest x-ray image using generative adversarial networks *IEEE Access* **8** 153535–45

[91] Salehi M, Mohammadi R, Ghaffari H, Sadighi N and Reiazi R 2021 Automated detection of pneumonia cases using deep transfer learning with pediatric chest x-ray images *Br. J. Radiol.* **94** 20201263

[92] Khan M A, Rajinikanth V, Satapathy S C, Taniar D, Mohanty J R, Tariq U and Damaševičius R 2021 VGG19 network assisted joint segmentation and classification of lung nodules in CT images *Diagnostics* **11** 2208

[93] Rehman N, Zia M S, Meraj T, Rauf H T, Damaševičius R, El-Sherbeeny A M and El-Meligy M A 2021 A self-activated CNN approach for multi-class chest-related COVID-19 detection *Appl. Sci.* **11** 9023

[94] Jaszcz A, Połap D and Damaševičius R 2022 Lung x-ray image segmentation using heuristic red fox optimization algorithm *Sci. Program.* **2022** 1–8