Article:
Elakkiya, R, Chandra, Harshiv, Pears, N. E. orcid.org/0000-0001-9513-5634 et al. (2 more authors) (2025) Lung image quality assessment and diagnosis using generative autoencoders in unsupervised ensemble learning. Biomedical signal processing and control. 107268. ISSN 1746-8094

https://doi.org/10.1016/j.bspc.2024.107268

Lung Image Quality Assessment and Diagnosis Using Generative Autoencoders in Unsupervised Ensemble Learning

Elakkiya Rajaseker[1], Harshiv Chandra[2], Nick Pears[3], Ketan Kotecha[4]

Subramaniyaswamy Vairavasundaram[5, *]

[1,2]Department of Computer Science, BITS Pilani, Dubai Campus, Dubai 345055, United Arab Emirates; {elakkiyaceg@gmail.com, f20200085@dubai.bits-pilani.ac.in}

[3]Department of Computer Science, University of York, York YO10 5DD, United Kingdom. nick.pears@york.ac.uk

[4]School of Computing, SASTRA Deemed University, Thanjavur-613401, India. vsubramaniyaswamy@gmail.com

[5]Symbiosis Centre for Applied Artificial Intelligence, Symbiosis International (Deemed University), Pune, India, director@sitpune.edu.in

Corresponding Author: Subramaniyaswamy Vairavasundaram (vsubramaniyaswamy@gmail.com)

Abstract

The lung is a critical organ for blood gas exchange. Early lung disease detection is often hindered by subtle symptoms. The diagnosis typically requires expert analysis of lung image scans, where both conventional methods and advanced machine learning (ML) and deep learning (DL) techniques are employed for scan segmentation and disease detection. However, it is a challenge to develop reliable computational models, due to the scarcity of high-quality data. With a major emphasis on lung disease classification and segmentation performance, this work presents a novel data quality assessment technique specifically designed for lung image datasets. The proposed pipeline combines ensemble learning, unsupervised learning, and generative autoencoders with attention mechanisms (GAME). By including attentional mechanisms in the generative autoencoders, we improve the tool's ability to locate and prioritise areas of interest in lung images and extract the right features, which increases the accuracy of our algorithms. The pipeline automates quality checks and reduces the need for high-quality data, while reducing human oversight. Because it's crucial to validate the robustness of the algorithms in our tool, we tested it using lung scans from the NIH-ChestXray and CheXpert datasets, and obtained IOU scores of 0.88 and 0.86, F1 scores of 0.95 and 0.95, and accuracy scores of 0.96 and 0.95, respectively, which shows they can be used as a classification tool as well as a segmentation tool. Given the challenging conditions in which the early diagnosis of lung disease is made, this is a significant step forward in the development of self-sustaining and accurate disease detection systems. The proposed model is made available to the public and the same is accessible via https://github.com/harshivchandra/LungDataQualityAssessment.

Keywords: Generative Autoencoders, Ensemble Learning, Lung Segmentation, Quality Assessment, Unsupervised Learning

1. Introduction

Lungs are an important organ in the human body, responsible for the task of gas exchange in the blood. A healthy lung ensures that its ability to adapt, correct and rebuild its structure is intact [1]. Medical advances in body imaging over the past decades have been able to significantly improve the detection of various lung diseases. However, because of increased stress and pollution in the modern-day lifestyle, most lung diseases have become harder to

detect solely via visual inspection by human experts, often necessitating the use of modern computational techniques. These are typically based on deep learning and machine learning.

1.1.Lung Imaging Techniques

Many advanced techniques have been developed in the field of lung imaging, and to produce high-fidelity reconstructions of the human lung, complex machinery is needed. Wilhelm Roentgen's discovery of X-rays, which facilitated the growth of the emerging discipline of medical imaging—which uses non-invasive techniques to inspect the body—is frequently credited with sparking the development of lung imaging. These days, lung imaging can be done in a variety of methods, some of which are listed below. The lung images produced by a few of these methods are shown in Figure 1. Table 1 provides a brief comparison between each technique, with respect to its advantages and disadvantages.

- X-ray: The most commonly performed procedure, also known as a chest radiograph. It is inexpensive, easy to perform, quick, and carries very little radiation risk. Digital radiography (DR) systems provide excellent images. An X-ray corrects for all but the smallest lung abnormality but requires additional examination to identify the specific disease since the spatial resolution is much lower than that of CT.
- Computed Tomography (CT): CT is an imaging method that captures multiple X-rays from different planes of the lungs using a radiopaque contrast agent to highlight anomalies and improve image resolution. CT has better lesion detection in any region than X-rays, because of its higher temporal resolution and spatial resolution [4].
- Magnetic Resonance Imaging (MRI): This creates tremendously detailed images using a strong magnetic field and high-frequency radio waves. MRI is better for identifying abnormalities of blood vessels in the chest but is more expensive and time-consuming than CT scans, and has much lower sensitivity for detecting lung abnormalities. As a result, it’s very rarely used for chest imaging.
- Ultrasonography: chest image created by bouncing sound waves off the body; used to diagnose lung cancer by identifying masses or fluid in the pleural space.
- Positron Emission Tomography (PET) scanning: PET images human tissue at the metabolic level through the use of radioisotopes of naturally occurring elements. Metabolic rates in normal versus malignant airspaces in the lung can be distinguished, making PET a common method of cancer identification.

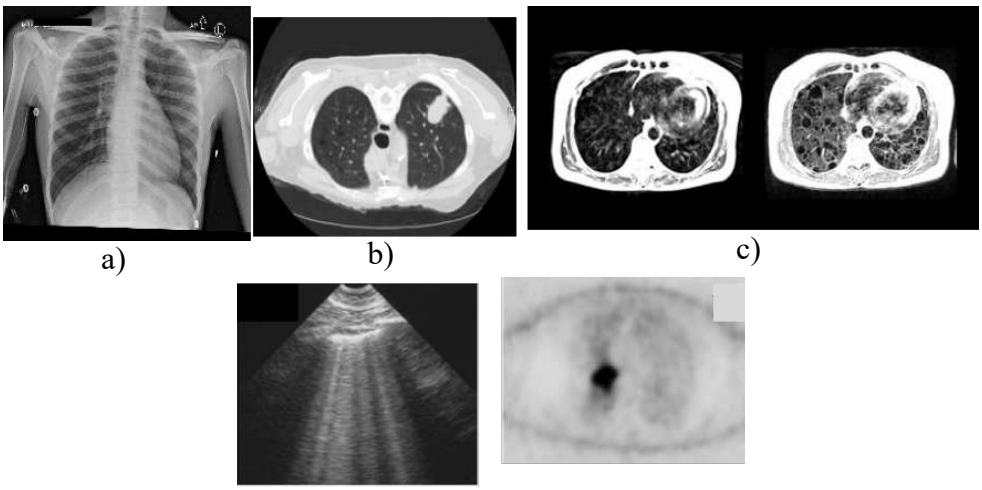

Fig. 1. Lung Imaging using, a) X-Ray [5], b) CT [6], c) MRI [7], d) Ultrasonography [8], e) PET [9]

Table 1: Advantages and Disadvantages of different types of lung imaging techniques

Lung Imaging Type	**Advantages**	**Disadvantages**
X-Ray	1. Quick and cost-effective 2. Good for detecting large abnormalities like tumors	1. Limited detail 2. Less effective for soft tissue differentiation
CT	1. High resolution images 2. Good for detecting detailed abnormalities	1. High radiation exposure 2. Expensive to conduct
MRI	1. No exposure to ionizing radiation. 2. Excellent soft tissue contrast.	1. Expensive to conduct 2. Not ideal for lung tissue due to low signal to noise ratio
Ultrasonography	1. No exposure to radiation 2. Useful for detecting fluid around lungs	1. Limited penetration of sound waves into lung tissue 2. Not useful for detailed imaging of deeper lung structures
PET	1. Excellent for detecting tissue metabolic activity 2. Useful for assessing treatment response	1. Expensive to conduct 2. Involves exposure to radioactive tracers

1.2. Advances in AI for disease diagnosis

Over time, due to rapid strides in the technological domain, there has been a rising interest in the use of advanced computational techniques to simplify difficult tasks. One of the first major applications of Artificial Intelligence (AI) in the medical domain is the MYCIN model [2] , that used a human defined rule based AI system , to provide medical diagnoses to doctors for bacterial infections. Although useful, the lack of improvement over human diagnosis, and the use of manually defined if-then-else rules, made similar rule-based disease diagnosis systems redundant. The development of truly automated computational models began with the introduction of neural networks in the 1980s, which led to the development and deployment of the first machine learning models for ischaemic heart disease detection using ECG scans [3]. Simultaneous breakthroughs in parallel scientific computing and model theory led to the implementation of deep learning models for disease prediction, which have retained prominence due to their ability to handle large datasets, model complex relationships within medical data, and achieve higher predictive accuracy.

Thus, in this paper, we will explore the background behind lung imaging, and understand the segmentation and classification models currently applied in the domain. In Section 2, we will explore related works in the field while understanding both traditional models and modern ML/DL based models. In Section 3, we will define our proposed framework; and in Sections 4,5 and 6, we will explore the results of the proposed framework, compare it to other implementations in the field, and appropriately derive conclusions, and future work on our model, whilst also exploring the limitations of the model.

2. Related Work

The development and application of several lung imaging modalities stand out as essential strategies for lung disease diagnosis as we begin the literature study. In the field of lung disease

detection research, the main emphasis is on finding abnormalities and examining their features to determine the disease that a person is suffering from. The idea of dividing an input image into discrete "segments" or regions is fundamental to segmentation modeling. This segmentation method is a fundamental part of the investigation since it is so effective at identifying different things in the image.

2.1.Traditional Models

Traditionally, various techniques have been created to segment input lung scans for disease detection. These techniques leveraged the use of mathematical systems to delineate the structures of interest within a given input image. Understanding the different techniques employed is essential for contextualizing the advancements made towards the deployment of ML algorithms in this domain. There are many statistical models that exist for image segmentation. One such model is described in [10] that explains the use of a hierarchical Bayesian approach towards lung CT image segmentation. It uses the Dirichlet process (DP) to determine the structure and the hierarchical Bayesian paradigm to jointly segment the images. DP is defined in Equation 1,

$$(\boldsymbol{G(A_1)}, \dots, \boldsymbol{G(A_r)}) \sim \boldsymbol{Dir(\alpha_0 G_0(A_1)}, \dots \boldsymbol{\alpha_0 G_0(A_r)}) \tag{1}$$

where α_0,G_0, ($G(A_1)$, …, $G(A_r)$) denote the scaling factor, base measure, and random vector over a finite partition (A_1, … ,A_r) in a measurable space Θ respectively. DP involves the random distribution of elements across random distributions. Due to the non-parametric nature of DP, it can automatically determine the anatomy of a given CT scan, and thus it doesn't require the definition of manual Region of Interest (ROI) and sample seed sets. The overall generative model is defined by the system in Equation 2,

$$\begin{pmatrix} \boldsymbol{x_{ji}|\theta_{ji} \sim F(\theta_{ji})} \\ \boldsymbol{\theta_{ji}|G_j \sim G_j} \\ \boldsymbol{G_i|\alpha_0, G_0 \sim DP(\alpha_0, g_0)} \\ \boldsymbol{G_0|\gamma, H \sim DP(\gamma, H)} \end{pmatrix} \tag{2}$$

where x_{ji} denotes the feature for the i-th image's j-th pixel. Another approach is the fuzzy c-means algorithm (FCM), which segments many pixels in an image by iterating over two essential conditions until a solution is found. The FCM aims to compute the center of the clusters and identify the class membership matrix c × n, where c denotes the number of clusters and n denotes the number of samples. Another such algorithm is the Fuzzy Possibilistic C-Means (FPCM) algorithm, which was first proposed in [11]. It is defined by Equation 3,

$$\boldsymbol{\min\{J_{m,n}(U, T, V; X)\}} = \boldsymbol{\sum_{i=1} \sum_{k=1} (u_{ik}^m + t_{ik}^n) D_{ikA}^2} \tag{3}$$

$$\text{subject to, } \begin{pmatrix} \boldsymbol{m>1, n>1, 0 \le u_{ik}, t_{ik} \le 1} \\ \boldsymbol{D_{ikA} = \|x_k - v_i\|A} \\ \boldsymbol{\sum_{i=1}^{c} u_{ik} = 1 \forall k, i.e., U \in M_{f\,cn}} \\ \boldsymbol{\sum_{k=1}^{n} u_{ik} = 1 \forall i, i.e., T^t \in M_{f\,nc}} \end{pmatrix}$$

where, U, T, V, c, and n denote the membership matrix, the possibilistic matrix, the resultant cluster centers, the number of the cluster, and the data point respectively. It uses a possibilistic membership within the function to illustrate the degree of absolute belonging for any given point in the lung image within a particular cluster. Although accurate, the high mathematical

complexity means that creating higher number of segments within the lungs leads to extremely complex computation.

2.2. Deep Learning Models

The advent of the modern segmentation algorithm based on clustering in [15] led to a surge in ML-based segmentation models for lung images. Table 2 compares papers on lung image segmentation, detailing ML algorithms, metrics, and objectives. In [12], the U-net architecture is suggested for lung CT image segmentation, achieving a Dice-Coefficient of 0.9502, calculated using equation (4),

$$\boldsymbol{DSC} = \mathbf{2} * \frac{||\boldsymbol{S} \cap \boldsymbol{T}||}{||\boldsymbol{S} \oplus \boldsymbol{T}||} \tag{4}$$

where T is the lung parenchyma area obtained using manual segmentation, and S is the lung parenchyma area produced using segmentation based on the model's output. In the study [13], an evaluation of four lung x-ray segmentation models, including Fully Convolutional Networks (FCN), SegNet, U-Net, and U-Net++, reveals U-Net++ as the most effective for tuberculosis identification. U-Net++ outperforms other models across various metrics, such as loss, dice coefficient, specificity, mean_iou, sensitivity, recall, precision, and accuracy, achieving scores of (-0.9796, 0.9796, 0.9932, 0.9598, 0.9753, 0.9838, 0.9685, 0.9874). Clustering methods, including k-means, k-median, particle swarm optimization (PSO), inertia-weighted PSO (IWPSO), and guaranteed convergence PSO (GCPSO), are assessed for detecting lung cancer through CT scan segmentation, with GCPSO proving more accurate at 95.3%.

In [15], the study introduces Encoder-Decoder Convolutional Neural Networks (ED-CNN) for accurate lung segmentation, achieving an average Dice-Coefficient of 0.962. Another model, CA-UNet, is presented in [16] for lung nodule segmentation, combining classic UNet with convolution and attention fusion. Compared to other models, CA-UNet demonstrates superior performance, reflected in high dice-coefficient, precision, and recall scores (0.8986, 0.8907, and 0.9204, respectively). ResBCDU, presented in [17], utilizes ResNet-34 and Bidirectional Convolutional Long Short-Term Memory Recurrent Neural Networks (BiConvLSTM) for lung CT image segmentation, surpassing other models in precision, recall, f1, accuracy, and dice coefficient index scores. Finally, [18] introduces a segmentation network called NASNet-Large for lung images, achieving commendable IoU and DICE Coefficient Index scores, outperforming SegNet, UNet, DeepLab, and other researchers using CRF-LSTM [31]. The IoU metric was computed using Equation 5[18],

$$\boldsymbol{IoU} = \frac{|\boldsymbol{A} \cap \boldsymbol{B}|}{|\boldsymbol{A} \cup \boldsymbol{B}|} \tag{5}$$

where A and B denote the ground truth mask and prediction mask respectively. In their work [19], the authors introduce the DEHA-Net framework for lung nodule segmentation, employing a dual encoder setup within a hard attention network incorporating an adaptive region of interest (ROI) algorithm. The framework achieves DICE coefficient Index, Sensitivity, and Positive Predictive Value (PPV) scores of (87.91%, 90.84%, 89.56%), calculated using equations (6) and (7),

$$\boldsymbol{SEN} = \frac{\boldsymbol{Y'} \cap \boldsymbol{Y}}{\boldsymbol{Y}} \tag{6}$$

$$\boldsymbol{PPV} = \frac{\boldsymbol{Y'} \cap \boldsymbol{Y}}{\boldsymbol{Y'}} \tag{7}$$

where SEN and PPV represent sensitivity and positive predictive value, respectively. Another proposed model [20] modifies the U-Net segmentation network for accurate segmentation in COVID-19 Lung CT-scan datasets. The alterations include adjustments in convolution kernel numbers, a switch from RMSprop to Adaptive Moment Estimation (Adam) for optimization, employment of early stopping-based implicit regularization with weight decay regularization, and a cosine annealing function for the learning rate scheduler as described in Equation (8).

$$\boldsymbol{\eta_t = \eta^i_{min} + \frac{1}{2}(\eta^i_{max} - \eta^i_{min})\left(1 + \cos\frac{T_{cur}}{T_i}\pi\right)} \quad (8)$$

where, η denotes the learning rate, with η^i_{min} and η^i_{max} being the learning rate range, and T_{cur} represents the number of epochs since the models restart and it also incorporates a binary cross-entropy and dice loss function, described in Equation 9.

$$\boldsymbol{L(Y,\hat{Y}) = -\frac{1}{N}\sum_{b=1}^{N}\left(\frac{1}{2}\, Y_b \cdot \log \hat{Y}_b + \frac{2\cdot Y_b \cdot \hat{Y}_b}{Y_b + \hat{Y}_b}\right)} \quad (9)$$

where, $\hat{\mathbf{Y}}_\mathbf{b}$ and Y_b, represent prediction and ground truth values of the b^{th} image input into the model. N represents the batch size. This model achieves superior accuracy, dice, sensitivity, and specificity scores (0.9928, 0.8724, 0.8642, 0.9969) compared to models like Inf-Net, nnU-Net, and R2-UNet. Additionally, the utilization of a Mask-R-CNN model with a K-means kernel for automatic lung segmentation on lung CT images, as proposed in [21], achieves a remarkable dice coefficient score of 0.97, outperforming competing implementations such as Optimal Path Snake (OPS), Vector Field Convolution (VFC), Crisp Adaptive (CRAD), System to Detect and Quantify Pulmonary Emphysema (SISDEP), and Gradient Vector Flow (GVF), with dice coefficient scores of 0.93, 0.84, 0.94, 0.93, and 0.82, respectively.

Table 2. A comparison of Lung Segmentation models on publicly available dataset

Paper	Algorithm(s) Implemented	Dataset Used	Metrics Computed
Skourt et al, 2018 [12]	U-Net	LIDC-IDR	Dice-Coefficient index
Gite et al, 2022 [13]	FCN, SegNet, UNet, UNet++	NIH, JRS	loss, dice coefficient, specificity, mean_iou, sensitivity, recall, precision, accuracy
Senthil et al, 2019 [14]	K-Means, K-Median, PSO, IWPSO, GCPSO	LungCT-Diagnosis	accuracy
Kalinovsky et al, 2016 [15]	ED-CNN	JRST	Dice-Coefficient
Wang et al, 2023 [16]	CA-UNet, UNet, DenseUNet, UNet ++, CoLe-CNN, Trans-UNet, Swin-UNet,	LIDC-IDRI	dice-coefficient index, precision and recall scores
Jalali et al, 2021 [17]	ResBCDU, U-Net, RU-Net, ResNet34-UNet, BCDUNet	LIDC-IDRI	precision, recall, f1, accuracy and dice coefficient index scores

Zhang et al,2023 [18]	NASNet, SegNet, UNet DeepLab	RSNA Pneumonia	IoU and DICE Coefficient Index scores
Usman et al, 2023 [19]	DEHA-Net	LIDC-IDRI	DICE coefficient Index, Sensitivity and Positive Predictive Value (PPV) scores
Upadhyay et al,2023 [20]	Modified U-Net, Inf-Net, nnU-Net, R2-UNet	Covid CT	accuracy, dice, sensitivity, and specificity scores
Hu et al, 2020 [21]	Mask-R-CNN, OPS, VFC, CRAD, SISDEP, GVF	Lung CT Dataset	dice coefficient scores
Proposed Implementation	U-Net, FCN, SegNet, GCPSO, Mask-R-CNN, and U-Net++	NIH, CheXpert, Covid CT, Lung-PET-CT-Dx	Accuracy, Sensitivity, Specificity, F1 Score, Intersection over Union (IoU), Hausdorff Distance, Dice Coefficient, Dice Loss, and Matthews Correlation Coefficient (MCC)

2.3.Autoencoder Based Models

The use of advanced ML and DL techniques has completely transformed medical image analysis, especially for lung disease segmentation and classification. In the following section, we provide a comprehensive analysis of the current studies with a focus on the role of autoencoders in this very important area. Autoencoders are unsupervised learning techniques that have attracted a great deal of interest and attention because they can learn to represent the input data effectively. In the case of lung disease segmentation and classification, autoencoders have been used as an integral part of the framework. Autoencoders have been widely used as feature extractors and their efficacy has been proven [22] using the compressed representation given by Equation 10. The use of such feature extractors can help to learn very complex pattern and allow to discover subtle irregularities in the lung images, thus increasing the segmentation accuracy. Another characteristic of autoencoders is the ability to use unsupervised learning that has been particularly effective in applications with only a limited number of labelled data. The unsupervised learning can help with the generalisation of the segmentation models by training them on unlabelled images, and this characteristic is very useful when there is a limited number of medical images that can be annotated.

$$\boldsymbol{h} = \boldsymbol{f_\theta} = \boldsymbol{\sigma}(\boldsymbol{Wx} + \boldsymbol{b}) \tag{10}$$

where the variable h represents the encoded representation, $\boldsymbol{f_\theta}$ represents the whole encoding function, σ represents the activation function, W represents the weight matrix, x represents the input, and b represents the bias term. A lung disease segmentation model, which seamlessly incorporates autoencoders, was proposed in a groundbreaking study [23]. The autoencoder served as a feature extractor as shown in Equation 11, allowing the model to derive meaningful representations from a wide range of lung images. The results demonstrated superior segmentation accuracy in comparison to conventional CNN-based methods.

$$\boldsymbol{L} = \boldsymbol{\Sigma}_{\mathrm{i}}\, \boldsymbol{D}(\boldsymbol{F}(\boldsymbol{X}_{\mathrm{i}}), \widehat{\boldsymbol{X}_{\boldsymbol{i}}}) \tag{11}$$

Where L represents the loss function that needs to be minimized during training, N is the total number of samples in the dataset, $\boldsymbol{X}_{\mathrm{i}}$ is the input data, F is the segmentation model, $\widehat{\boldsymbol{X}_{\boldsymbol{i}}}$ is the reconstructed output of the segmentation model for the input and D is a dissimilarity or

reconstruction loss function, measuring the difference between the input and the reconstructed output.

Autoencoders, particularly explored for classifying lung disorders, have shown promise in unsupervised learning [24]. Despite their potential, challenges persist, including interpretability, dataset heterogeneity, and relevance to real-world clinical scenarios [25]. Variational autoencoders (VAEs) outperform traditional autoencoders but face challenges like training difficulty and collapse mode [26]. While autoencoders significantly impact lung disease segmentation and classification, a unified framework for lung image quality evaluation and disease diagnosis is lacking. To address this gap, we propose GAME, integrating attention-enhanced generative autoencoders into ensemble and unsupervised learning. This approach tackles issues like low-quality labeled data and mild symptoms, enhancing robustness and accuracy in lung image analysis. Data scarcity and diverse architectural frameworks pose challenges to modern models. While pipeline architectures address the latter, the lack of reliable data impedes model deployment in real-world scenarios. Our primary goal is to develop an automated method capable of analyzing and quantifying any lung image dataset, providing insights into model effectiveness in diverse circumstances.

3. Proposed Methodology

The proposed methodology aims to address current limitations and enhance the precision and robustness of disease identification. The approach focuses on improving lung image quality assessment by integrating attention-enhanced autoencoders into an ensemble and unsupervised learning paradigm. Leveraging attention mechanisms highlights salient features in complex lung images, providing a nuanced understanding of pathophysiology. The integration of multiple models allows a comprehensive analysis, and unsupervised learning enhances generalization using massive unlabelled datasets. This sophisticated methodology promises a more accurate and autonomous system for image quality evaluation and disease diagnosis, advancing lung image analysis. Figure 2 illustrates the higher-level description of the framework's three essential layers.

- Phase 1: GAME - This critical initial stage enhances the effectiveness of subsequent autoencoder processing. Standardizing input photos to a chosen resolution reduces variability and dimensions. The attention mechanism in GAME focuses on salient regions, improving feature extraction. A data loader efficiently gathers images into batches for further processes.
- Phase 2: Classification and Unsupervised ROI Generation - Models for classifiers and unsupervised ROI generators are applied to the pre-processed dataset. Evaluating the dataset's suitability for ROI creation and classification is the goal, forming a crucial basis for subsequent decision-making.
- Phase 3: Image Segmentation Ensemble - The preprocessed dataset undergoes evaluation in an image segmentation ensemble model. Performance indicators generate a comprehensive report, determining optimal use cases for each model and providing insights for further investigation and framework enhancement.

3.1. Datasets Used

Various datasets were chosen to create a fair analysis and to suitably benchmark all different data representations available in the field for lung image segmentation. This shall aid in the development of a robust pipeline for dataset evaluation. In this section, we will discuss the datasets that were utilized in this model. Each dataset was split into training and test sets with an 80%-20% split ratio. Before anything else, we need to look at the NIH Chest-XRay dataset. It is made up of 112,120 frontal view X-ray images of 30,805 patients, and each image can

have one of 14 illness labels. These disease labels were identified from the appropriate reports using data mining techniques [5]. Every label included in the dataset is listed in Table 3.

Table 3: NIH Chest-X-Ray Data Description

Label	**Information**
Atelectasis	Patient's lung has either collapsed or closed
Cardiomegaly	Patient has an abnormal enlargement of the heart
Effusion	Patient's lung has excess fluid around it
Infiltration	Presence of Infiltration
Mass	Presence of Mass
Nodule	Presence of Nodule
Pneumonia	Patient is diagnosed with Pneumonia
Pneumothorax	Presence of Pneumothorax
Consolidation	Patient's lungs are partially filled with liquid
Edema	Patient has excessive fluid in lung tissues.
Emphysema	Patient is diagnosed with Emphysema
Fibrosis	Patient is diagnosed with Fibrosis
Pleural Thickening	Patient's lungs may have pleural thickening
Hernia	Patient is diagnosed with hernia.
No Finding	Patient's diagnosis is normal

ChexPert is a comprehensive collection of chest X-rays that are used for the diagnosis of lung diseases. It is equipped with uncertainty labels as well as standard evaluation sets that are labeled by radiologists. In total, there are 224,316 chest radiographs from 65,240 patients that have been tagged for the presence of 14 common chest radiographic observations with this collection. The functions of each label in the dataset are outlined in Table 4, along with the distribution of those labels.

Table 4: CheXpert Data Description

Label	**Information**
No Finding	Patients' diagnosis is normal
Enlarged Cardio mediastinum	Patient has an abnormally enlarged cardio mediastinum
Cardiomegaly	Patient has an abnormal enlargement of the heart
Lung Opacity	Patient has opaque lung area
Lung Lesion	Patient has an abnormal lesion in the lung
Edema	Patient has excessive fluid in lung tissues
Consolidation	Patient's lungs are partially filled with liquid
Pneumonia	Patient is diagnosed with Pneumonia
Atelectasis	Patient's lung has either collapsed or closed
Pneumothorax	Presence of air or gas in the cavity surrounding the Patient's lungs
Pleural Effusion	Patient's lung has excess fluid around it.
Pleural Other	Other abnormality in the pleural space of the Patient's chest
Fracture	Patient has a fracture
Support Devices	Patient has support devices

Another dataset that is investigated in this research is the chest x-ray dataset, which is used for the categorization of respiratory diseases. This is intended to be used for the purpose of distinguishing COVID-19 from other respiratory disorders such as pneumonia and

tuberculosis. It is made up of 32687 frontal view X-ray images of 30,805 patients, and each image can contain one of four disease labels. These disease labels were derived from the appropriate reports using data mining techniques. An overview of the dataset and its labels is provided in Table 5, which may be found here.

Table 5: Chest X-Ray Data Description

Label	Information
COVID-19	Patient diagnosed with COVID-19
Lung-Opacity	Patient has an opaque lung area
Normal	Patient's lungs are normal
Viral Pneumonia	Patient is diagnosed with Pneumonia
Tuberculosis	Patient is diagnosed with tuberculosis

A database that is based on CT scan images, Curated Chest CT. A database that is employed for the detection of COVID-19 disease through the utilization of lung-CT scan imaging is referred to as reference source not found. It is produced by combining CT scans from seven independent datasets into a single piece of data. There are 112,120 frontal view X-ray images of 30,805 patients that make up this collection. Each image has the potential to have one of three illness diagnoses, all of which have been validated by radiologists. An overview of the labels that are included in the dataset is provided in Table 6.

Table 6: Curated Chest CT Data Description

Label	Information
COVID-19	Patient diagnosed with COVID-19
Normal	Patient is normal
CAP	Patient diagnosed with Community Acquired Pneumonia (CAP)

The Lung-PET-CT-Dx is presented by the Cancer Imaging Archive, involves the use of lung CT images and lung PET scans to create a large-scale dataset for the identification of lung cancer. There are a total of 112,120 frontal view X-ray images of 30,805 individuals included in this collection. Each image can be assigned one of four illness classifications. There were five academic thoracic radiologists who were accountable for the annotation that indicated the location of each tumor. An overview of the labels that are included in the dataset is provided in Table 7.

Table 7: Lung-PET-CT-Dx Data Description

Label	Information
Adenocarcinoma	Patient has cancer in the lung's glandular cells.
Small Cell Carcinoma	Patient's lungs have small, cancerous cells.
Large Cell Carcinoma	Patient's lungs have large, abnormal cancerous cells
Squamous Cell Carcinoma	Patient's lungs have cancerous squamous epithelial cells

3.2. Model Pipeline

Figure 3 pipeline integrates proposed cutting-edge methods for enhanced efficiency and accuracy in lung image analysis. Comprising three stages, Stage 1 initiates with meticulous preprocessing, standardizing photos to a set resolution and introducing a pivotal GAME for advanced representation learning and selective attention. In Stage 2, preprocessed data undergoes an unsupervised ROI generator and classifier, evaluating dataset suitability for classification tasks and precise ROI definition. Stage 3 utilizes an ensemble model for image

segmentation, producing a detailed report on optimal model use cases. Leveraging autoencoders, attention mechanisms, and ensemble learning, this pipeline signifies a comprehensive advancement in lung image processing for illness diagnosis and image quality assessment.

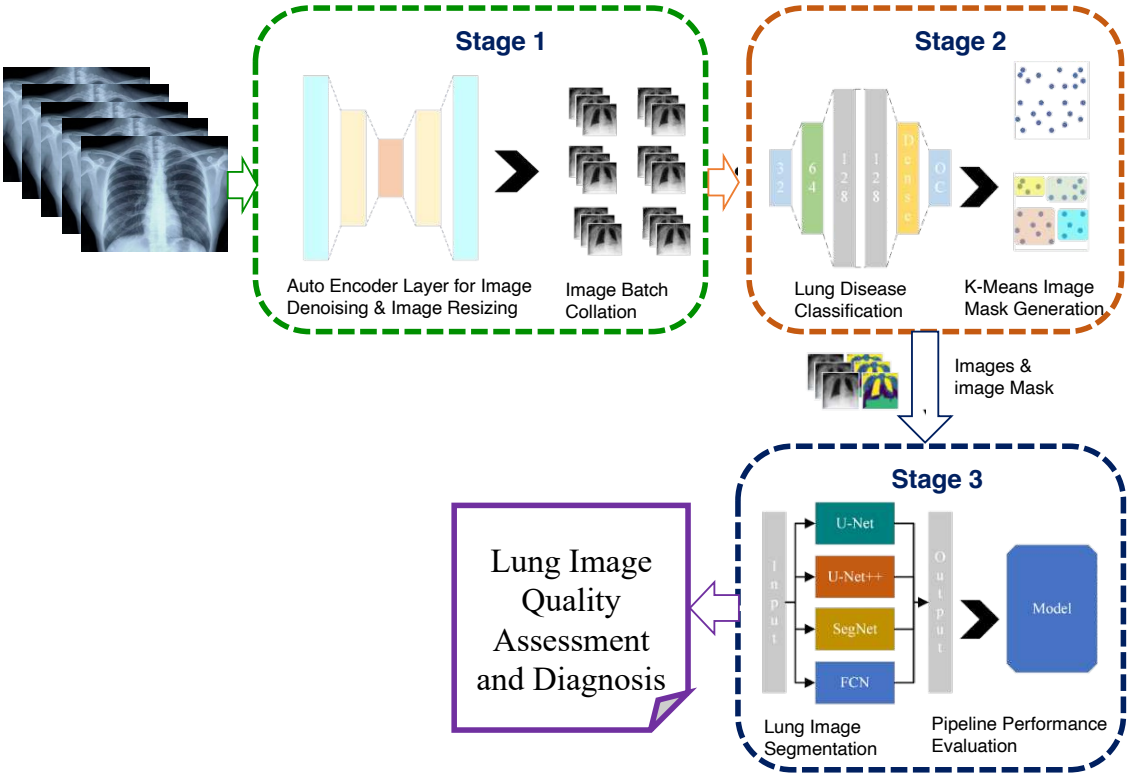

Fig. 2. Overview of Proposed Architecture

3.2.1. Stage 1: Generative Autoencoders with Attention Mechanisms

As the first phase of our research, Stage 1 manages a sophisticated preprocessing model that combines enhanced interpretability from a GAME with dimensionality reduction. By instructing the autoencoder to concentrate on significant regions of lung images, this critical step not only addresses the requirement for image standardization but also enhances the preprocessing pipeline's adaptive encoding capabilities. Every lung image is meticulously converted to a uniform format to start the standardization process.

The following is the formula for this standardization (Equation 12).

$$X_{std} = \frac{(X - \mu)}{\sigma} \tag{12}$$

Where X_{std} denotes Standardized image, X is Original image, μ is a Mean and σ is a Standard deviation. The distinguishing feature of Stage 1 innovation is the integration of a GAME. By deviating from traditional autoencoders, the encoding process gains dynamism. The embedded attention mechanism within dynamically allocates weights during encoding, allowing the autoencoder to identify and prioritize prominent regions within the input images. This is the definition of the encoding function (Equation 13 & 14). Since crucial features are frequently localized in the field of medical imaging, this flexibility is especially helpful.

$$a = \text{Softmax}\left(W_{att} * \tanh\left(W_g * x_i + b_g\right)\right) \tag{13}$$

$$h_i = a_i * x_i \tag{14}$$

Where a is the Attention weights, $\boldsymbol{W_{att}}$ is the attention weight parameters, tanh is the hyperbolic tangent activation function, $\boldsymbol{W_g}$ is the parameters for the transformation of input x, $\boldsymbol{b_g}$ is bias term for the transformation of input x, $\boldsymbol{att}$ enhanced encoded representation, $\boldsymbol{h_i}$ is the *i*-th element of the attention-enhanced encoded representation, $\boldsymbol{a_i}$ is the i-th element of the attention weights vector and $\boldsymbol{x_i}$ is the i-th element of the input image vector.

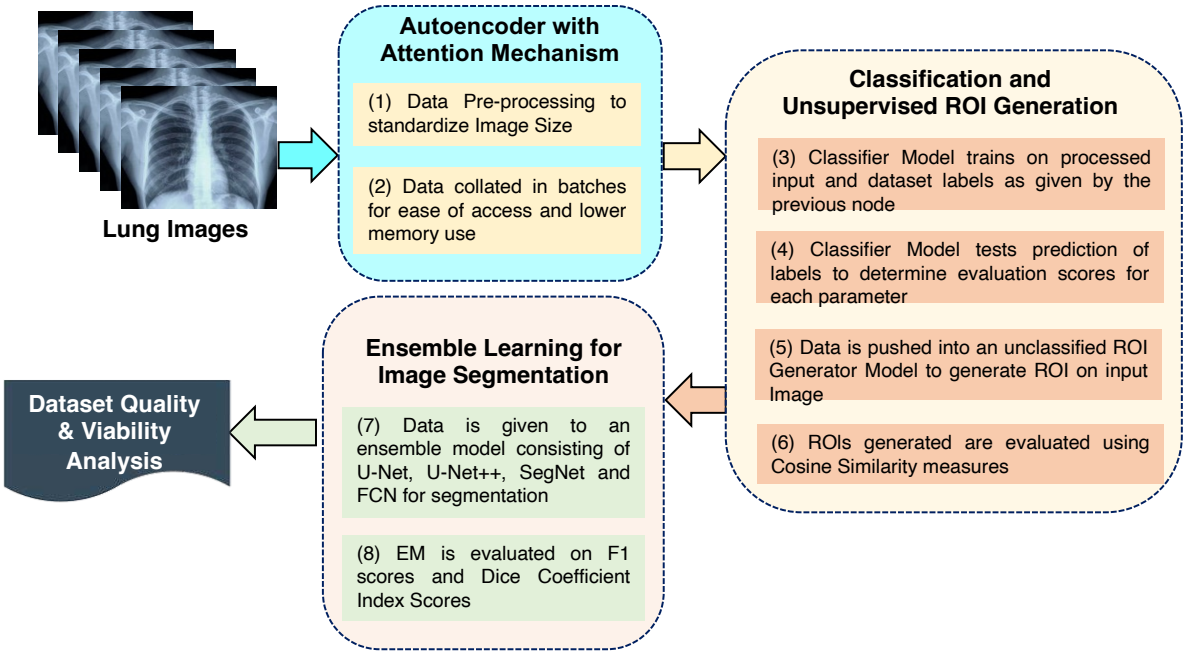

Fig. 3. Workflow of the Proposed Model

Algorithm I GAME

Step 1: Load_image Function

```
function load_image(image_file: str) → Tensor
    image ← read_and_decode(image_file)
    return image
end load_image
```

Step 2: Encoder Function

```
function encoder(labels: List[str]) → Array
    label_binarizer ← create_label_binarizer()
    ohe_labels ← label_binarizer.fit_transform(labels)
    return ohe_labels
end encoder
```

Step 3: Normalize_image Function

```
function normalize_image(img_array: Array) → Array
    norm_img ← []
    for img in img_array do
        img_np ← convert_to_numpy(img)
        scal ← max(img_np) - min(img_np)
        sub ← min(img_np)
        norm_img_array ← (img_np - sub) / scal
        norm_img.append(norm_img_array)
    end for
    return norm_img
```

```
end normalize_image
Step 4: Resize_and_normalize_image Function
function resize_and_normalize_image(image: Tensor, target_size: dims) → Tensor
    attention_enhanced_image ← autoencoder_with_attention_preprocessing(image)
    resized_image ← resize_to_target(attention_enhanced_image, target_size)
    normalized_resized_image ← normalize_image(resized_image)
    return normalized_resized_image
end resize_and_normalize_image
Step 5: Generator Function
function generator(images: List[str], labels: List[str], batch_size: int, target_size: dims) →
Tuple(Array, Array)
    L ← length(images)
    while True do
        batch_start ← 0
        batch_end ← batch_size
        while batch_start < L do
            limit ← min(batch_end, L)
            paths ← images[batch_start:limit]
            X ← []
            for path in paths do
                img ← load_image(path)
                processed_img ← resize_and_normalize_image(img, target_size)
                X.append(processed_img)
            end for
            Y ← encoder(labels[batch_start:limit])
            yield (X, Y)
            batch_start += batch_size
            batch_end += batch_size
        end while
    end while
end generator
Step 6: Autoencoder_with_attention_preprocessing Function
function autoencoder_with_attention_preprocessing(image: Tensor) → Tensor
    attention_enhanced_image ← autoencoder_attention_process(image)
    return attention_enhanced_image
end autoencoder_with_attention_preprocessing
```

The processed images are effectively arranged into batches using a dataloader after attention-enhanced encoding. In order to optimize computational resources and guarantee a smooth transition of pre-processed images into pipeline stages that follow, this step is essential. The core of Stage 1 is captured by this method, which coordinates the incorporation of attention-enhanced autoencoder preprocessing. The method converts unprocessed lung images into a dataloader that is ready for the next steps in our ground-breaking lung image analysis pipeline. To sum up, Stage 1 establishes a complex equilibrium between standardization and adaptive encoding, laying the groundwork for a sophisticated, effective, and cutting-edge lung image analysis research project. Algorithm 1 represents the data processing using GAME.

3.2.2. Stage 2: Classification and Unsupervised ROI Model

The input datasets are first loaded into a CNN based classifier model, which is constructed based on the number of classes to predict for that input dataset. CNN classifier is used for

benchmarking input lung scans is due to their lower computational complexity in processing medical image datasets, especially for the disease classification task [35] . Furthermore, such models remain advantageous for this application due to their robust performance with relatively smaller datasets [36]. Our proposed implementation is visualized in Figure 4. This step involves classifying the input dataset using a model intended to classify lung images according to predetermined classes. Using previously processed photos, the classification model forecasts the probability of each class. Multiclass classification frequently makes use of the SoftMax activation function. One way to view the output is as probability assigned to every class as shown in Equation 15.

$$P\left(\frac{y_i}{X}\right) = \frac{e^{X_i}}{\sum_{j=1}^{N} e^{X_j}} \tag{15}$$

where $P\left(\frac{y_i}{X}\right)$is the predicted probability of class, X_i is the logit (raw score) for class i and N is the total number of classes. An unsupervised model for generating Regions of Interest is simultaneously applied to the input dataset. This method may detect important regions inside the lung images without the need for labelled training data. ROIs designate areas that, because of possible anomalies or distinctive features, demand closer inspection. The procedure of creating a ROI from a pre-processed lung image is described in Equation 16.

$$ROI = f_{ROI}(X_i) = \text{Morph}(\text{Threshold}(\text{Segment}(\text{Extract}(\text{Image})))) \tag{16}$$

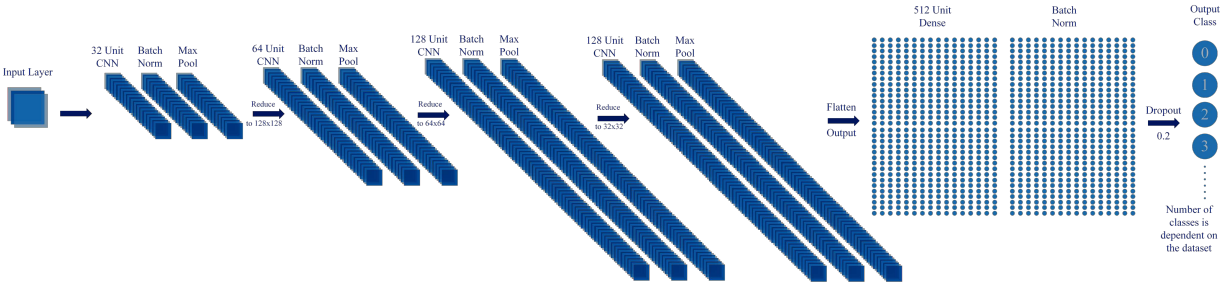

Fig. 4. Implemented Classifier Architecture

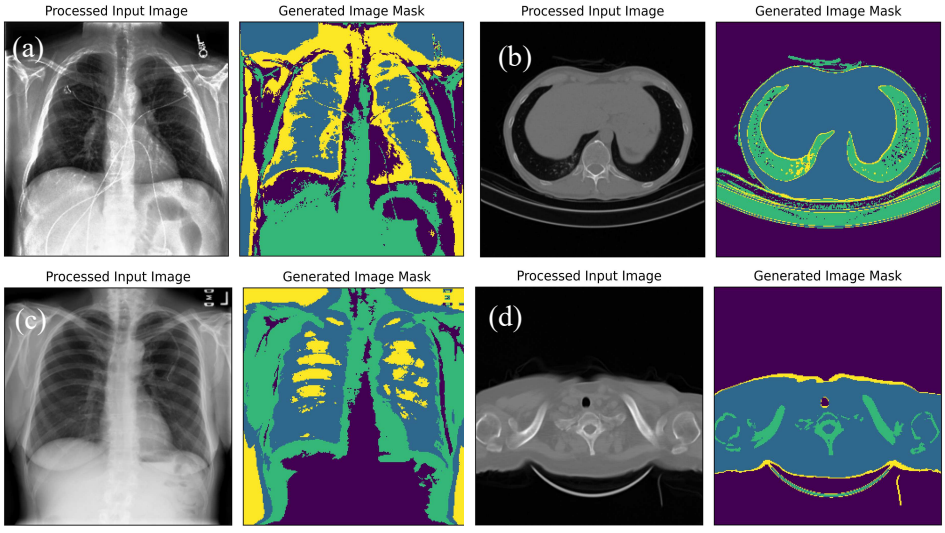

Fig. 5. Generated Regions of Interest Masks for input images from the (a) Chexpert dataset, (b) Covidct dataset, (c) Chest-x-ray dataset

The function f_ROI autonomously identifies and extracts relevant regions in an image, targeting anomalies or distinctive features without requiring labelled training data. Employing diverse image processing techniques and unique unsupervised ROI algorithms, f_ROI aims to

pinpoint regions of interest. The subsequent Extract function focuses on acquiring pertinent features from the pre-processed lung image, such as edges, textures, and intensity changes. Pixel grouping through clustering or segmentation processes, like region-growing or k-means, identifies coherent areas. Intensity thresholds highlight abnormal pixel intensities, with specific thresholds applied to define ROI pixels. To enhance ROI accuracy, morphological operations like erosion and dilation refine the discovered region's shape and boundaries. This ROI model, utilizing a k-means clustering algorithm with four clusters, automatically generates image segmentation masks, as illustrated in Figure 5, while Algorithm 2 elucidates the functioning of the classification and unsupervised ROI Model.

Algorithm II Classification and Unsupervised ROI Model

Inputs: X - Input image dataset
Outputs: Y - Predicted labels

Step 1: Define CNN Classification Model

Define the CNN architecture with convolutional layers, pooling layers, batch normalization , and fully connected layers
Initialize model weights and biases.
Forward pass through the CNN:

function classifier(input_shape: shape, num_classes: int) → Model:

$conv_{32} = \sigma(conv_{d=32}X^{32}; w^d = 32, \mathrm{dc})$
$X^{32}_{\mathrm{normalized}} = \mathrm{BatchNorm}(X^{32})$
$\Gamma_{s1} = maxpool\ (d_{32}X^{32}_{\mathrm{normalized}})$
$conv_{64} = \sigma(conv_{d=64}X^{64}; w^d = 64, \mathrm{dc})$
$X^{64}_{\mathrm{normalized}} = \mathrm{BatchNorm}(X^{64})$
$\Gamma_{s2} = maxpool\ (d_{64}X^{64}_{\mathrm{normalized}})$
$conv_{128} = \sigma(conv_{d=128}X^{128}; w^d = 128, \mathrm{dc})$
$X^{128}_{\mathrm{normalized}} = \mathrm{BatchNorm}(X^{128})$
$\Gamma_{s3} = maxpool\ \left(d_{128}X^{128}_{\mathrm{normalized}}\right)$
$X_{flattened} = Flatten(X_{pool4})$
$FC_{512} = \sigma(dense_{d=512}X^{512}; w^d = 512, \mathrm{dc})$
model ← Model(inputs ← inputs, outputs ← outputs)
return model

end classifier

Step 2: Unsupervised Clustering

class Cluster:
procedure initiate_class(cluster, no_units: integer):
cluster.k_model ← KMeans(n_clusters ← no_units)
end procedure initiate_class
function train_predict_labels(cluster, dataset: Array) → Array:
cluster.k_model.fit(dataset)
return cluster.k_model.predict(dataset)
end function train_predict_labels
function model_return(cluster) → KMeans:
return cluster.k_model
end function model_return
end class Cluster

Step 3: Classification Followed by Clustering

function classify_and_cluster(images: Array, num_classes: int, no_units: int) → Array:
classification_results ← classifier(images)

```
    c ← Cluster(no_units ← no_units)
    cluster_labels ← c.train_predict_labels(classification_results)
    return cluster_labels
end function classify_and_cluster
Step 4: Perform Classification Followed by Clustering
hybrid_results ← classify_and_cluster(images, num_classes, no_units)
```

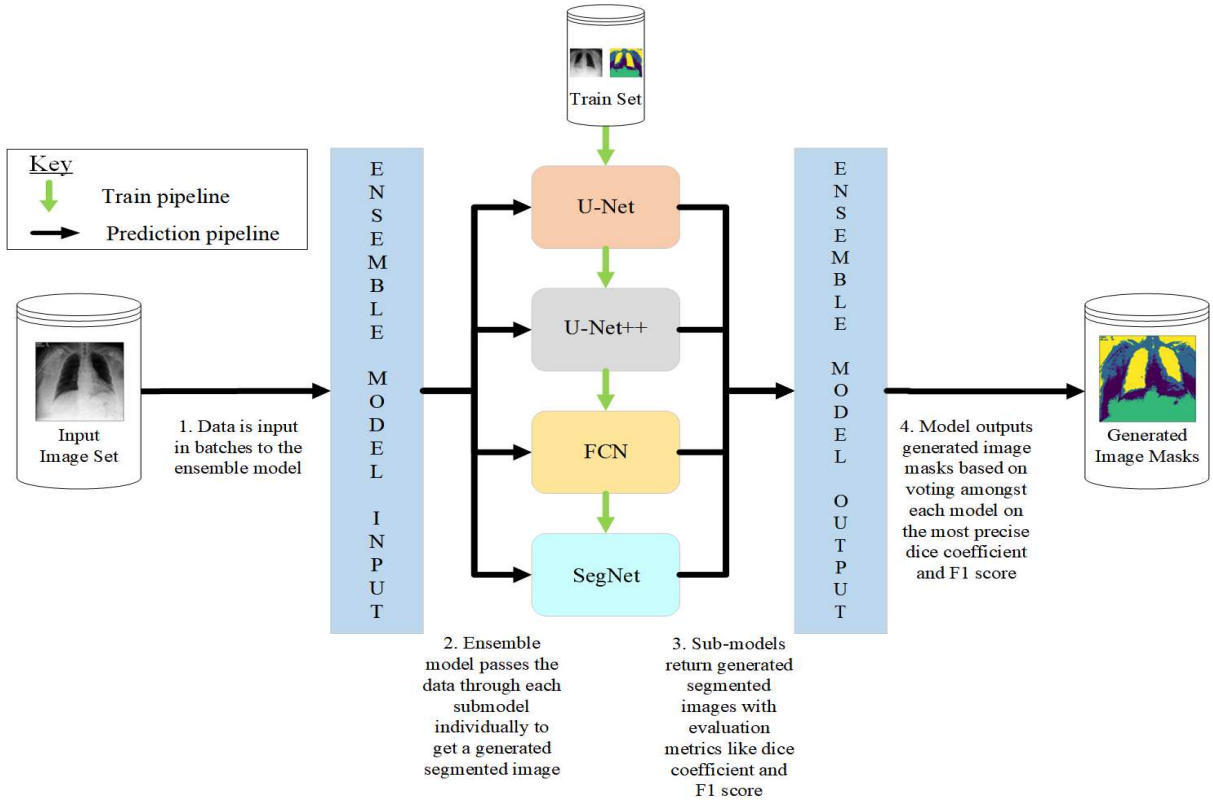

Fig. 6. Proposed Ensemble Model Architecture

3.2.3. Stage 3: Ensemble Architecture

By combining various segmentation methods, such as U-Net, FCN, SegNet, GCPSO, Mask-R-CNN, and U-Net++, a potent ensemble model specifically designed for lung disease segmentation is created. Through independent training on datasets, each algorithm learns to recognize distinct patterns and traits linked to lung disorders. With this customized training, the group can better tackle the different obstacles that come with analyzing lung images by utilizing the unique capabilities of each algorithm. This hybrid technique combines the output of each algorithm to produce a robust and flexible segmentation model, is responsible for the effectiveness of ensemble learning. This ensemble is evaluated using samples from multiple datasets, providing a comprehensive view of its performance in different scenarios. The quality of segmentation is quantified using metrics such as recall, accuracy, precision, and Dice coefficient. Apart from improving accuracy, the ensemble facilitates detailed investigation of algorithmic contributions and emphasizes the relative benefits and drawbacks of each. The result of this repeated evaluation and improvement procedure is an ensemble model that performs very well in lung disease segmentation, providing a promising new avenue for the development of medical image analysis. After receiving individualized training, every model in the ensemble is put through a rigorous evaluation process.

The models are given five training epochs to fine-tune their parameters based on the characteristics of the sampled images. After this training phase, the predictions made by each model are combined using an average method. Through the utilisation of each model's unique insights and abilities, this prediction pooling ensures a comprehensive and well-balanced ensemble output. The averaging process increases the resilience of the ensemble by lessening the impact of any outliers and raising the precision of the final segmented mask predictions. In essence, this approach leverages the features of many segmentation algorithms while emphasizing a collaborative and synergistic approach. The ensemble's ability to integrate several model predictions into a coherent output highlights its flexibility and suitability for difficult medical image analysis tasks, particularly in the field of lung disease segmentation. The ensemble architecture of the suggested method is shown in Figure 6. In the final stage, a report will be produced after the ensemble model results are considered. The viability of the input dataset is indicated in this report.

4. Results and Discussion

4.1. GAME Performance

The proposed lung image quality assessment method was notably enhanced by integrating attention mechanisms into generative autoencoders (GAME). This improvement skillfully highlighted salient features in medical images, resulting in enhanced clarity and preservation of anatomical details. Quantitative measures, including structural similarity index (SSI) and signal-to-noise ratio (SNR), demonstrated substantial improvements in image quality. Sample images from NIH Chest X-ray and cheXpert data, showcased in Figure 7, depict the impact of GAME on image quality before and after application. Table 8 presents SNR and SSI values, indicating increased image quality after employing GAME in the data pre-processing step. Noteworthy improvements in structural similarity index and signal-to-noise ratio were observed across datasets, exemplified by a 4.6% SNR increase and a 7% improvement in SSI for NIH Chest X-ray. This pattern persisted in various datasets, illustrating the efficacy of the proposed method in enhancing lung image quality.

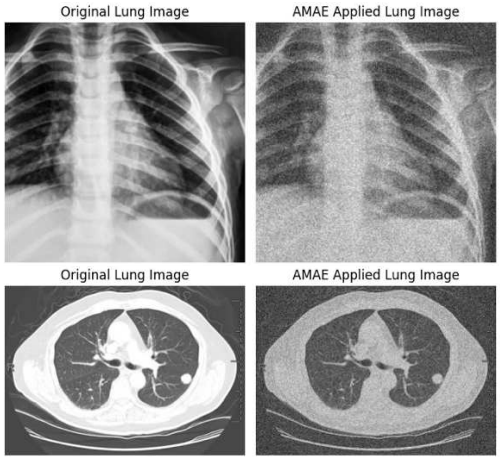

Fig. 7. GAME in NIH Chest X-Ray and cheXpert datasets

Table 8: Comparison of GAME Performance (Four Datasets)

Dataset	Before SNR	After SNR	Before SSI	After SSI
NIH Chest Xray	20.7	25.3	0.84	0.91
CheXpert	18.6	27.8	0.77	0.86
Curated Chest CT	22.6	28.2	0.91	0.97
Lung-PET-CT-Dx	19.9	24.5	0.86	0.95

4.2. Classification and Unsupervised ROI Performance

Our CNN-based classification step underwent rigorous testing on four distinct datasets: NIH Chest Xray, Curated Chest CT, Lung-PET-CT-Dx, and CheXpert. The evaluation showcased the model's effectiveness in recognizing and categorizing various lung conditions. Notably, on the NIH Chest Xray dataset, the model achieved an F1 Score of 0.954, Accuracy of 0.957, Precision of 0.956, and Recall of 0.957, with an AUC-ROC value of 0.754. The CheXpert dataset demonstrated high accuracy, resulting in an F1 Score of 0.951, Accuracy of 0.952, Precision of 0.953, and Recall of 0.952, accompanied by an AUC-ROC value of 0.874. Remarkable performance on the Curated Chest CT dataset yielded an F1 Score of 0.960, Accuracy of 0.960, Precision of 0.962, and Recall of 0.960, with an AUC-ROC score of 0.958. Similarly, the Lung-PET-CT-Dx dataset showed proficiency, achieving an F1 Score of 0.960, Accuracy of 0.961, Precision of 0.960, and Recall of 0.961, supported by an AUC-ROC value of 0.862. AUC curves for NIH and CheXpert datasets are illustrated in Figures 8 and 9, providing unique AUC values for each class. The comprehensive analysis presented in Table 9 affirms the classification model's accuracy and reliability in recognizing intricate patterns indicative of lung diseases, facilitated by performance metrics for precise diagnosis and categorization.

Table 9: Comparison of Classification Performance (Four Datasets)

Dataset	F1	Accuracy	Precision	Recall	AUC- ROC
NIH Chest Xray	0.954	0.957	0.956	0.957	0.754
CheXpert	0.951	0.952	0.953	0.952	0.874
Curated Chest CT	0.960	0.960	0.962	0.960	0.958
Lung-PET-CT-Dx	0.960	0.961	0.960	0.961	0.862

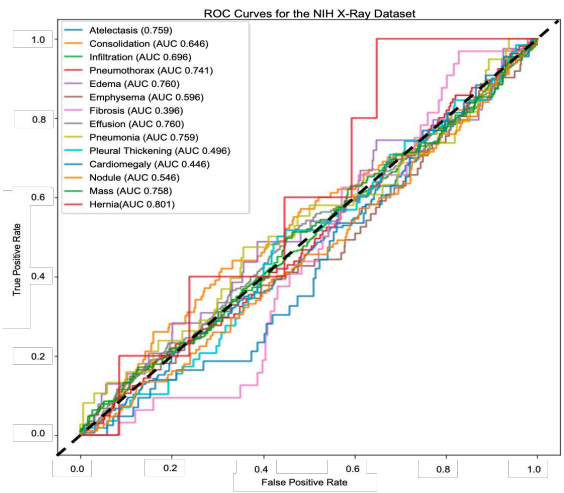

Fig. 8. ROC Curve for the NIH X-ray Dataset

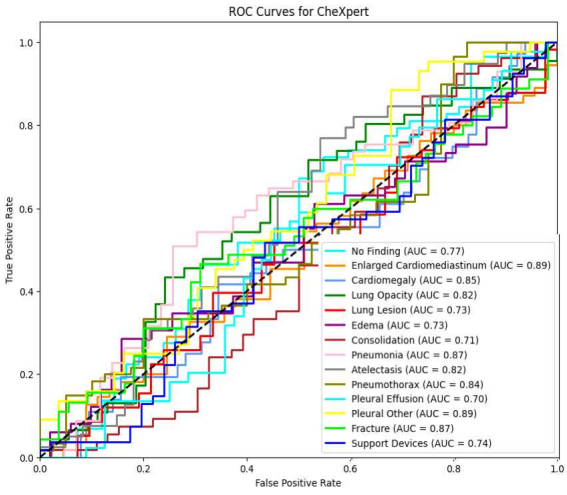

Fig. 9. ROC Curve for the CheXpert Dataset

The proposed method's efficacy was simultaneously assessed using the Unsupervised ROI model, focusing on its capability to identify relevant Regions of Interest (ROI) through Cosine Similarity across datasets. Notably, the Unsupervised ROI model exhibited remarkable efficacy on the NIH Chest Xray dataset, achieving a Cosine Similarity of 0.981, highlighting its ability to recognize and emphasize pertinent regions in lung scans. Similarly, the model performed well on the CheXpert dataset, achieving a Cosine Similarity of 0.988, highlighting its competence in locating significant areas of interest in lung images. Exceptional performance was observed on the Curated Chest CT dataset, with a Cosine Similarity of 0.915, indicating the model's capability to identify and extract relevant regions from complex chest CT images.

Although slightly lower, the Unsupervised ROI model demonstrated notable effectiveness on the Lung-PET-CT-Dx dataset, with a Cosine Similarity of 0.715. Table 10 summarizes these Cosine Similarity values, affirming the model's ability to recognize, highlight, and capture significant regions of interest in lung images across diverse medical imaging datasets, highlighting its precision and focus on image analysis.

Table 10: Comparison of Unsupervised ROI Performance (Four Datasets)

Dataset	Cosine Similarity
NIH	0.981
ChexPert	0.988
Curated Covid CT	0.915
Lung-PET-CT-Dx	0.715

Table 11: Comparison of Ensemble Model Performance (Four Datasets)

	Ensemble Model	
Dataset	F1 Score	Dice Coefficient Index
NIH Chest XRay	0.842	0.852
ChexPert	0.857	0.865
Curated Chest CT	0.766	0.794
Lung-PET-CT-Dx	0.701	0.820

The Ensemble Model, incorporating advanced techniques like U-Net, FCN, SegNet, GCPSO, Mask-R-CNN, and U-Net++, underwent thorough testing on four datasets. It exhibited precision and accuracy in sickness diagnosis and segmentation, achieving an F1 Score of 0.842 and a Dice Coefficient Index of 0.852 on the NIH Chest XRay dataset. On the CheXpert dataset, it demonstrated proficiency with an F1 Score of 0.857 and a Dice Coefficient Index of 0.865. Remarkable performance on the Curated Chest CT dataset was evident with an F1 Score of 0.766 and a Dice Coefficient Index of 0.794. Despite challenges in PET-CT images, the model identified disease-related regions effectively on the Lung-PET-CT-Dx dataset, achieving an F1 Score of 0.701 and a Dice Coefficient Index of 0.820. These results, detailed in Table 11, underscore the Ensemble Model's adaptability and effectiveness across various medical imaging modalities, establishing it as a reliable tool for comprehensive lung imaging analysis.

4.3. Ensemble Model Performance

The ensemble algorithms' performance, focusing on both training and testing accuracy, was evaluated across four datasets: NIH Chest XRay, CheXpert, Curated Chest CT, and Lung-PET-CT-Dx. Trained for a brief five epochs, the model showcased rapid learning and accurate predictions with minimal CPU resources. Figure 10 visually represents the training and testing accuracy patterns for each dataset over the training epochs, offering insights into the models' adaptability to dataset features and their generalization in a limited timeframe. This emphasis on a short training phase underscores the goal of optimizing computing efficiency while maintaining robust learning. The graphical depiction serves as a valuable tool for assessing the model's swift learning and prediction accuracy across diverse medical imaging datasets, providing a concise overview of the training process.

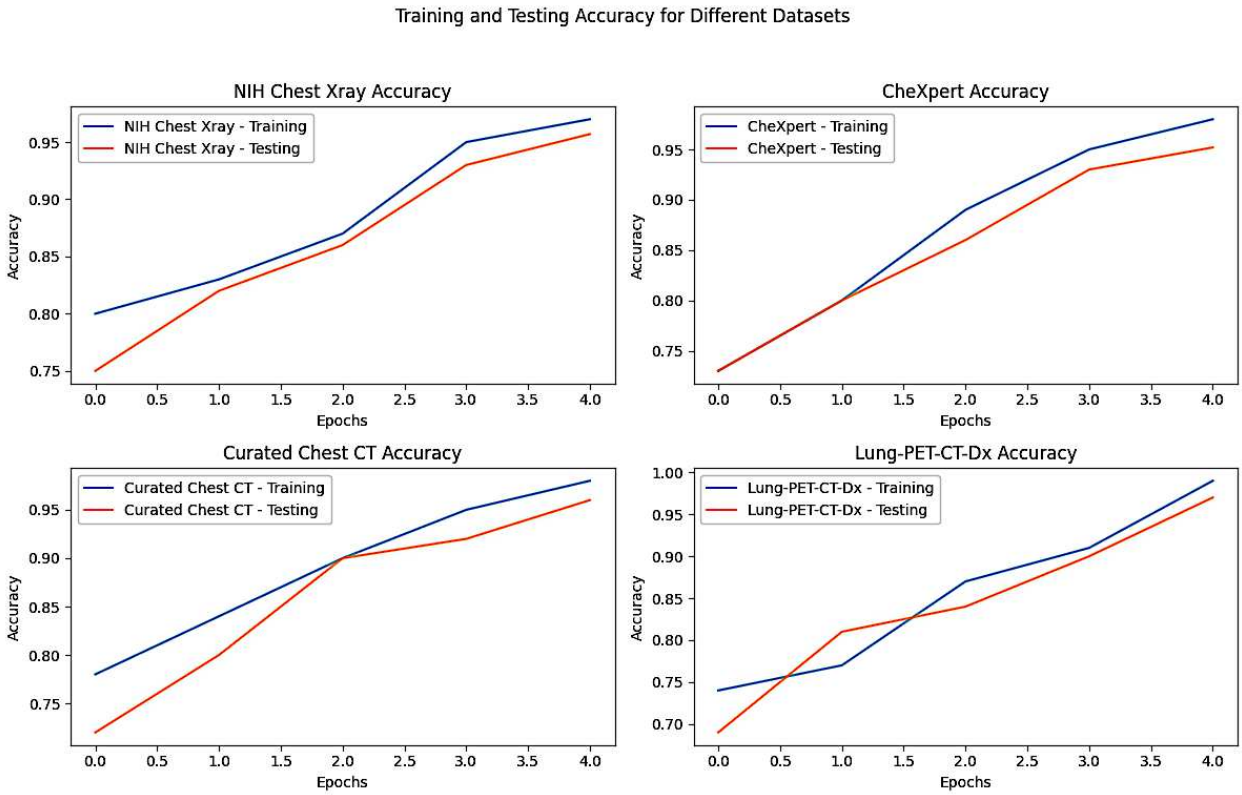

Fig. 10. Training and Testing Accuracy for 5 Epochs (Four Datasets)

4.4. Comparative Analysis

Using the NIH Chest X-ray, CheXpert, Curated Chest CT, and Lung-PET-CT-Dx datasets, among other datasets, the suggested technique demonstrated impressive resilience. The group of models demonstrated consistently strong performance on datasets containing a range of imaging modalities and acquisition strategies after integrating GAME with ensemble learning. The remarkable ability of the suggested method to generalize across a range of datasets demonstrates both its adaptability and possible therapeutic value. The proposed model outperformed the most sophisticated methods currently available for lung image analysis, according to a comparative performance analysis that measured its accuracy, sensitivity, specificity, F1 score, intersection over union (IoU), Hausdorff distance, dice coefficient, dice loss, and Matthews Correlation Coefficient (MCC). The proposed method is a promising tool for automated lung image quality assessment and lung disease diagnosis, utilizing attention-guided methods in combination with a robust and adaptable ensemble. The results obtained, as shown in Table 12, indicate that the suggested approach has the potential to be a reliable and flexible solution in the field of medical image analysis. This is evidenced by its ability to produce high-quality findings across different datasets.

Table 12: Comparative Analysis with State-of-the-Art Techniques (Two Datasets)

Dataset	**Metric**	**Proposed Approach**	**Nastnetlarge-net-Post [25]**	**ConvNeXt-ImageNet-21K [26]**
NIH X-ray	Accuracy	0.96	0.92	0.84
	Sensitivity	0.96	0.88	0.82
	Specificity	0.96	0.94	0.82
	F1 Score	0.95	0.90	0.26 (mean)

	IoU	0.88	0.87	0.83
	Hausdorff Distance	3.21	5.61	6.89
	Dice Coefficient	0.85	0.92	0.73
	Dice Loss	0.15	0.08	0.27
	MCC	0.87	0.82	0.74
	Metric	**Proposed Approach**	**CheXpertShifa-NET [27]**	**CheXpert Architecture [28]**
	Accuracy	0.95	0.89	0.93
	Sensitivity	0.95	0.84	0.91
	Specificity	0.95	0.91	0.95
cheXpert	F1 Score	0.95	0.92	0.92
	IoU	0.86	0.85	0.85
	Hausdorff Distance	3.79	6.45	4.89
	Dice Coefficient	0.82	0.86	0.89
	Dice Loss	0.18	0.14	0.11
	MCC	0.84	0.79	0.85

Table 12 provides a thorough comparison between the proposed method and two existing approaches, Nastnetlarge-net-Post and ConvNeXt-ImageNet-21K, using various performance metrics on the NIH X-ray dataset. The proposed strategy demonstrates superior specificity, sensitivity, and accuracy compared to the other methodologies. Notably, it excels in maintaining a balance between precision and recall, evident in the F1 Score and Dice Coefficient metrics. Evaluating the cheXpert dataset, the developed technique exhibits impressive performance metrics, showcasing its effectiveness in disease identification and classification. Accurate segmentation is highlighted by higher IoU and Dice Coefficient values, confirming the model's precision in identifying and delineating diseased areas. In summary, the comparative study validates the recommended method as a reliable and versatile solution for medical image analysis, outperforming other approaches across various datasets.

5. Future Works & Limitations

Further development of our work could be explored by the implementation of explainable ai (xAI) based layers atop across each stage of the pipeline, that could provide a reasonable, trustworthy and logical understanding of the model's decision [32]. Additionally, integrating feature attribution methods or saliency maps could help highlight regions that contribute significantly to the predicted outcome. Another avenue that could be explored is the implementation of advanced computer vision models such as vision transformers (ViT), which are known for being able to capture long range dependencies and fine-grained details within images [33], within the ensemble model layer.

5.1.Limitations

Although our model has performed admirably in identifying suitable datasets for lung image segmentation and classification, certain limitations can be observed in its implementation. One such limitation is the lack of transparency due to the use of black-box deep learning-based algorithms that do not provide reasoning for the decisions made during the prediction of the

output [32]. Another limitation is the heavy use of computational resources, such as memory, during model training, which may require future implementations to either devise hardware-accelerated solutions or utilize algorithms that are less resource-intensive. Apart from this, the evaluation of the images on the classification task is done solely using a CNN-based classifier. Future work could explore the implementation of alternate classifier models, such as vision transformers. Furthermore, it should be noted that data sources used to train the model may have an inherent bias in the collection process due to various factors, such as underrepresentation of a specific population demographic for lung X-rays [34] or variations in imaging protocols across studies. This unknown bias may slightly impact the future quality assessment of new images due to differences in the model's expectations.

6. Conclusion

This research paper proposed a novel framework for lung image quality analysis and diagnosis. The proposed approach involved integrating Attention Mechanisms in Generative Autoencoders in ensemble and unsupervised learning, showed notable improvements in lung disease diagnosis and image quality assessment. Across a range of datasets, the ensemble model—which comprised U-Net, FCN, SegNet, GCPSO, Mask-R-CNN, and U-Net++—performed admirably. These datasets comprised Lung-PET-CT-Dx, Curated Chest CT, CheXpert, and NIH Chest Xray. Not only did the GAME help to detect illnesses and improve image quality, but they also greatly increased the model's ability to identify key components. The comparison analysis revealed that the suggested method outperformed the current methods in terms of accuracy, sensitivity, specificity, F1 Score, IoU, Hausdorff Distance, Dice Coefficient, Dice Loss, and MCC. The ensemble model's adaptability and effectiveness were immediately obvious, and it continuously yielded positive results for a variety of metrics and datasets. The investigation additionally showcased the ensemble model's versatility in handling diverse medical imaging modalities, indicating its potential for application in an expanded array of clinical contexts. The approach described here offers a strong and flexible tool for automated diagnosis of lung disorders, in addition to providing important insights into disease patterns and improving patient treatment. The proposed method not only advances the field of lung image analysis but also sets a benchmark for future medical image processing and machine learning research. Future work could explore the implementation of xAI atop the model layers, and ViT based algorithms within the ensemble model.

CRediT authorship contribution statement

Elakkiya R: Methodology, Software, Writing-complete draft, Writing - review &editing. **Harshiv Chandra:** Methodology, Writing – initial draft, **Nick Pears:** Supervision, Methodology, Validation, **Subramaniyaswamy V:** Data Preparation, Writing -review &editing **Ketan Kotecha:** Review & editing, Inspection.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was funded by The Royal Society, London under International Exchange Scheme (Grant Number: IES\R3\223017.) We would like to express our deepest gratitude for the BITS Pilani Dubai Campus and University of York for the continued support.

Data availability

The data that has been used is opensource.

References

[1] Schittny, Johannes C. "Development of the lung." *Cell and tissue research* 367 (2017): 427-444.
[2] Van Melle, William. "MYCIN: a knowledge-based consultation program for infectious disease diagnosis." International journal of man-machine studies 10, no. 3 (1978): 313-322.
[3] Kukar, M. "Estimating the reliability of classifications and cost-sensitive combining of different machine learning methods." PhD diss., Ph. D. thesis, Faculty of Computer and Information Science, Univ. of Ljubljana, Ljubljana, Slovenia, 2001.
[4] De Wever, W., J. Coolen, and J. A. Verschakelen. "Imaging techniques in lung cancer." *Breathe* 7, no. 4 (2011): 338-346.
[5] Wang, Xiaosong, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers. "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2097-2106. 2017.
[6] Kinsinger LS, Anderson C, Kim J, et al. Implementation of Lung Cancer Screening in the Veterans Health Administration. JAMA Intern Med. 2017;177(3):399–406. doi:10.1001/jamainternmed.2016.9022
[7] Campbell-Washburn, Adrienne E., Rajiv Ramasawmy, Matthew C. Restivo, Ipshita Bhattacharya, Burcu Basar, Daniel A. Herzka, Michael S. Hansen et al. "Opportunities in interventional and diagnostic imaging by using high-performance low-field-strength MRI." Radiology 293, no. 2 (2019): 384-393.
[8] Bar, Stéphane, Céline Yee, Daniel Lichtenstein, Magali Sellier, Florent Leviel, Osama Abou Arab, Julien Marc, Matthieu Miclo, Hervé Dupont, and Emmanuel Lorne. "Assessment of fluid unresponsiveness guided by lung ultrasound in abdominal surgery: a prospective cohort study." Scientific Reports 12, no. 1 (2022): 1350.
[9] Vansteenkiste, Johan F., and Sigrid S. Stroobants. "PET scan in lung cancer: current recommendations and innovation." Journal of Thoracic Oncology 1, no. 1 (2006): 71-73.
[10] Cheng, Wenjun, Luyao Ma, Tiejun Yang, Jiali Liang, and Yan Zhang. "Joint lung CT image segmentation: A hierarchical Bayesian approach." PloS one 11, no. 9 (2016): e0162211.
[11] Pal, Nikhil R., Kuhu Pal, and James C. Bezdek. "A mixed c-means clustering model." In Proceedings of 6th international fuzzy systems conference, vol. 1, pp. 11-21. IEEE, 1997.
[12] Skourt, Brahim Ait, Abdelhamid El Hassani, and Aicha Majda. "Lung CT image segmentation using deep neural networks." Procedia Computer Science 127 (2018): 109-113.
[13] Gite, Shilpa, Abhinav Mishra, and Ketan Kotecha. "Enhanced lung image segmentation using deep learning." Neural Computing and Applications (2022): 1-15.
[14] Senthil Kumar, K., K. Venkatalakshmi, and K. Karthikeyan. "Lung cancer detection using image segmentation by means of various evolutionary algorithms." Computational and mathematical methods in medicine 2019 (2019).
[15] Kalinovsky, Alexander, and Vassili Kovalev. "Lung image segmentation using deep learning methods and convolutional neural networks." In XIII Int. Conf. on Pattern Recognition and Information Processing Belarus State University (2016): 3-5.
[16] Wang, Tong, Fubin Wu, Haoran Lu, and Shengzhou Xu. "CA-UNet: Convolution and attention fusion for lung nodule segmentation." International Journal of Imaging Systems and Technology (2023).
[17] Jalali, Yeganeh, Mansoor Fateh, Mohsen Rezvani, Vahid Abolghasemi, and Mohammad Hossein Anisi. "ResBCDU-Net: a deep learning framework for lung CT image segmentation." Sensors 21, no. 1 (2021): 268.

[18] Zhang, Youshan. "Lung segmentation with NASNet-Large-Decoder Net." arXiv preprint arXiv:2303.10315 (2023).
[19] Usman, Muhammad, and Yeong-Gil Shin. "DEHA-Net: A Dual-Encoder-Based Hard Attention Network with an Adaptive ROI Mechanism for Lung Nodule Segmentation." Sensors 23, no. 4 (2023): 1989.
[20] Upadhyay, Ashwini Kumar, and Ashish Kumar Bhandari. "Semi-Supervised Modified-UNet for Lung Infection Image Segmentation." IEEE Transactions on Radiation and Plasma Medical Sciences (2023).
[21] Hu, Qinhua, Luis Fabricio de F. Souza, Gabriel Bandeira Holanda, Shara SA Alves, Francisco Hercules dos S. Silva, Tao Han, and Pedro P. Reboucas Filho. "An effective approach for CT lung segmentation using mask region-based convolutional neural networks." Artificial intelligence in medicine 103 (2020): 101792.
[22] Souza, Johnatan Carvalho, João Otávio Bandeira Diniz, Jonnison Lima Ferreira, Giovanni Lucca França Da Silva, Aristofanes Correa Silva, and Anselmo Cardoso de Paiva. "An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks." Computer methods and programs in biomedicine 177 (2019): 285-296.
[23] Ghosal SS, Sarkar I, El Hallaoui I. Lung Nodule Classification Using Convolutional Autoencoder and Clustering Augmented Learning Method (CALM). InHSDM@ WSDM 2020 Feb 5 (pp. 19-26).
[24] Yadav P, Menon N, Ravi V, Vishvanathan S. Lung-GANs: unsupervised representation learning for lung disease classification using chest CT and X-ray images. IEEE Transactions on Engineering Management. 2021 Aug 30.
[25] Rashid N, Hossain MA, Ali M, Sukanya MI, Mahmud T, Fattah SA. AutoCovNet: Unsupervised feature learning using autoencoder and feature merging for detection of COVID-19 from chest X-ray images. biocybernetics and biomedical engineering. 2021 Oct 1;41(4):1685-701.
[26] Saldanha J, Chakraborty S, Patil S, Kotecha K, Kumar S, Nayyar A. Data augmentation using Variational Autoencoders for improvement of respiratory disease classification. Plos one. 2022 Aug 12;17(8):0266467.
[27] Zhang Y. Lung segmentation with NASNet-Large-Decoder Net. arXiv preprint arXiv:2303.10315. 2023 Mar 18
[28] Jeong, Jaehyup, Bosoung Jeoun, Yeonju Park, and Bohyung Han. "An Optimized Ensemble Framework for Multi-Label Classification on Long-Tailed Chest X-ray Data." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2739-2746. 2023.
[29] Bajwa, Nabit, Kedar Bajwa, Muhammad Faique Shakeel, Atif Rana, Kashif Haqqi, and Suleiman Khan. "A generalized deep learning model for multi-disease Chest X-Ray diagnostics." In International Work-Conference on Artificial Neural Networks, pp. 541-552. Cham: Springer Nature Switzerland, 2023.
[30] Wollek A, Haitzer P, Sedlmeyr T, Hyska S, Rueckel J, Sabel B, Ingrisch M, Lasser T. Automated Labeling of German Chest X-Ray Radiology Reports using Deep Learning. arXiv preprint arXiv:2306.05997. 2023 Jun 9.
[31] Yang Wang, Pandi Vijayakumar, Brij B.Gupta, Wadee Alhalabi, Audithan Sivaraman, “An Improved Entity Recognition Approach to Cyber-Social Knowledge Provision of Intellectual Property Using a CRF-LSTM Model”, Pattern Recognition Letters, Elsevier, Vol.163, pp.145-151, November 2022.
[32] Xu, Feiyu, Hans Uszkoreit, Yangzhou Du, Wei Fan, Dongyan Zhao, and Jun Zhu. "Explainable AI: A brief survey on history, research areas, approaches and challenges." In *Natural language processing and Chinese computing: 8th cCF international*

conference, NLPCC 2019, dunhuang, China, October 9–14, 2019, proceedings, part II 8, pp. 563-574. Springer International Publishing, 2019.

[33] Dosovitskiy, Alexey. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).

[34] Seyyed-Kalantari, Laleh, Haoran Zhang, Matthew BA McDermott, Irene Y. Chen, and Marzyeh Ghassemi. "Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations." *Nature medicine* 27, no. 12 (2021): 2176-2182.

[35] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234-241. Springer International Publishing, 2015.

[36] Vaswani, A. "Attention is all you need." *Advances in Neural Information Processing Systems* (2017).