Deep Learning Based Pulmonary Edema Detection: Performance Evaluation of Optimizers

Indushree Shetty [a*][0009-0007-6446-4583], **Prerna Agrawal** [b][0000-0003-097-639X], **Savita Gandhi** [c][0000-0002-6581-2814]

[a] Research Scholar, SDRI, GLS University, Gujarat, India
[b] Assistant Professor, FCAIT-PG, GLS University, Gujarat, India
[c] Dean, FCAIT-PG, GLS University, Gujarat, India

*Corresponding Email: shettyindu.05@gmail.com

Abstract

Pulmonary edema is a life-threatening condition that is caused by fluid accumulation in the lungs, remains a diagnostic challenge, particularly in resource limited healthcare settings. This research presents a deep learning-based approach for automated detection of pulmonary edema using frontal chest X-rays from the NIH Chest X-ray14 dataset. The study proposes a methodology for pulmonary edema detection using the custom 7-layer convolutional neural network (CNN) in combination with different deep learning and hybrid algorithms. A dataset of total 4,606 images was used to train, validate and evaluate the algorithms. A total of 30 different deep learning and hybrid algorithms were experimented on different thresholds of 0.3, 0.4 and 0.5. Total 6 algorithms on threshold 0.5 were selected for further experimentation using the proposed methodology namely DenseNet121, ANN, DenseNet121+VGG16, DenseNet121+VGG19 and DenseNet121+NasNet. To enhance model performance and generalization, various preprocessing techniques, hyperparameters and optimizers including SGD, RMSProp, Adam, Nadam, AdaDelta and AdaGrad were applied on threshold 0.5. The DenseNet121+NasNet model demonstrated an excellent performance at 0.5 threshold, achieving 94.55% accuracy, 99.27% precision, 89.77% recall, an F1-score of 0.9428, and an AUC of 0.9948. The results of the proposed methodology and other existing methodologies are also compared using AUC. The comparison clearly shows that our proposed methodology results are excellent with highest AUC of 0.9948. The findings clearly show that our proposed methodology is an accurate and reliable solution to develop a scalable diagnostic tool for pulmonary edema, supporting medical professionals in underprivileged regions with a shortage of skilled radiologists.

Keywords: *Lung Diseases; Chest X-rays; Edema; CNN; ANN; DenseNet; VGG16; VGG19; NasNet; RMSProp*

1. INTRODUCTION

Pulmonary edema occurs on accumulation excess fluid in the lungs air sacs (alveoli), impairing the exchange of oxygen and carbon dioxide. This condition is most frequently triggered by left-sided heart failure, wherein the left ventricle fails to pump blood effectively, causing increased pressure in the pulmonary veins and subsequent fluid leakage into the lungs. Other causes include acute respiratory distress syndrome (ARDS), kidney failure, severe infections, and high-altitude exposure. Clinically, pulmonary edema manifests as shortness of breath, persistent coughing, and hypoxia. If left untreated, it can progress rapidly to respiratory failure and death. Diagnosis typically involves a combination of clinical examination, laboratory tests, and imaging, with chest X-rays being the primary imaging modality in most healthcare settings. Characteristic radiographic features include bilateral infiltrates, vascular congestion, perihilar haziness, and horizontal lines suggesting interstitial fluid accumulation. However, accurately interpreting these signs especially in early stages or in resource-limited settings with few trained radiologists remains a considerable challenge. [1, 2]

Pulmonary edema continues to burden healthcare systems worldwide, particularly among aging populations where comorbidities like heart failure are prevalent. Given its potential severity and the need for timely intervention, the development of efficient and automated diagnostic support tools is critical. This research investigates the application of deep learning models and various hyperparameter configurations to detect pulmonary edema from chest X-ray images. The aim is to create scalable, reliable, and accessible tools that can aid clinicians, especially in under-resourced areas, in identifying pulmonary edema swiftly and accurately. [1]

The global burden of pulmonary edema is staggering, driven largely by its association with heart failure. As per the global index 2023, 60-70 million people are living with heart failure from which 75-83% people are affected by pulmonary edema. Figure 1 shows the highest pulmonary edema cases in most affected 5 countries. [1, 2]

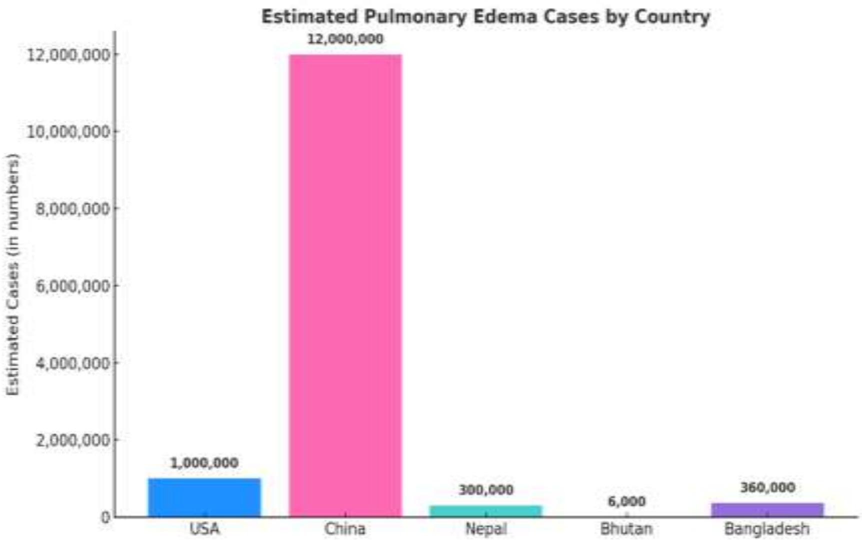

Fig. 1. Total number of pulmonary edema cases in top 5 countries

This study's main contribution is the development, testing, and evaluation of an effective deep learning-based system for the automated identification of pulmonary edema from chest radiography images. The Figure 2 depicts how by utilizing the Chest X-ray14 dataset, our study methodically investigates the performance of several neural network architectures, including trained models like DenseNet121 and baseline models like Artificial Neural Networks (ANNs). Furthermore, this study suggests and assesses new hybrid model configurations namely DenseNet121+ANN, DenseNet121+NasNet, DenseNet121+VGG16 and DenseNet121+VGG19 with addition of 7 convolutional layer in each of them designed to improve feature extraction and classification accuracy for the identification of pulmonary edema. Comparing seven optimization algorithms (SGD, Adam, RMSprop, Adagrad, Adadelta, Nadam, and Adamax) across each model architecture to determine the best optimizer model pairings is a unique feature of the research. In order to achieve generalizability, the models are trained and validated using pre-processed image data using standardization techniques such as pixel rescaling, image resizing to 150*150 dimensions, and validation splitting. A collection of relevant evaluation metrics, namely test accuracy, precision, recall, F1-score, AUC (Area Under the ROC Curve), confusion matrix parameters (True Positive, True Negative, False Positive, False Negative), are used to thoroughly evaluate each model's performance. The study offers a scalable and deep learning architecture that can help with the quick and accurate detection of pulmonary edema in chest X-rays through this multi-model, multi-optimizer inquiry. The results highlight how hybrid CNN architectures can enhance diagnostic procedures in clinical and resource-constrained settings, with customized optimization.

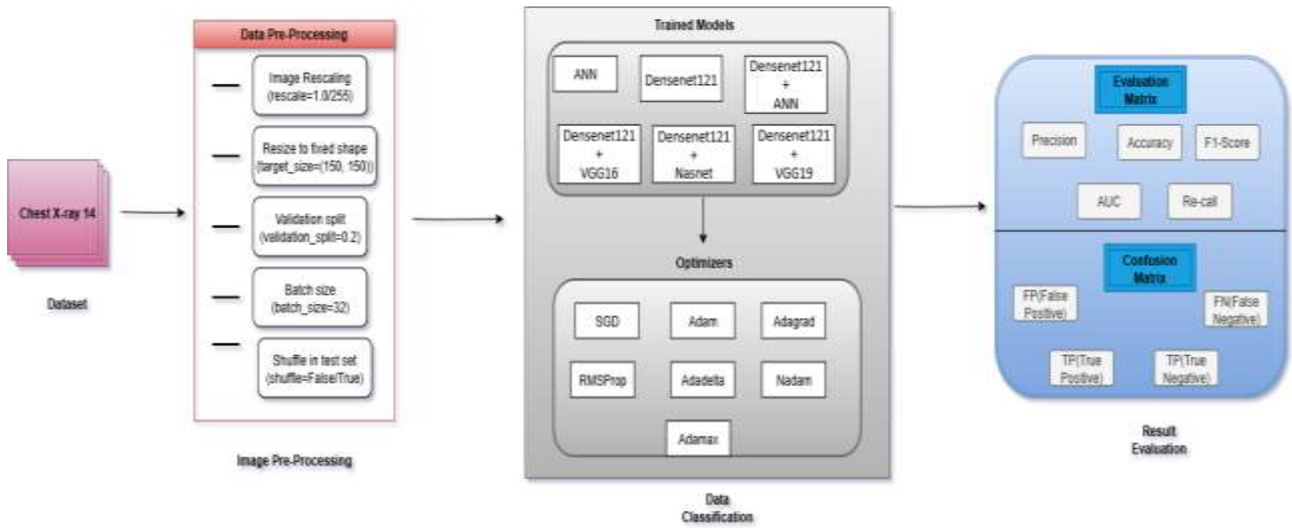

Fig. 2. Overview of the approach of edema classification

The paper is divided into 6 sections. Literature Review is presented in Section 2, whereas Section 3 contains the dataset used for experimentation. Section 4 explains the Experimentation Methodology used for conducting experiments. Section 5 discusses the results of the experiments conducted. Section 6 concludes with the conclusion and future Scope.

2. LITERATURE REVIEW

Recent advances in deep learning have greatly improved the identification of lung diseases, such as pulmonary edema, using chest X-ray (CXR) images. These studies use a wide range of datasets, from publicly available repositories to proprietary hospital collections, and a variety of convolutional neural network (CNN) architectures, including ANN, NasNet, DenseNet, ResNet, and other hybrid models, to detect conditions such as Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural Thickening, Hernia etc. Reported indicators of performance, such as accuracy, precision, recall, F1-Score and AUC, show significant gains, with some models obtaining accuracies more than 90%. The variety of methodologies including transfer learning, data augmentation, and ensemble techniques demonstrate the strength of deep learning in resolving issues such as class imbalance and picture variability, leading the path for more accurate and scalable diagnostic tools in medical imaging.

Table 1. Comparison of Deep Learning Models for Lung Disease Detection from Chest X-rays

Paper Title	**Dataset**	**Data Used**	**Total Images/data**	**Diseases Covered**	**Models Used**	**AUC**
M. Irtaza et al. [3]	NIH ChestX-ray14 Dataset	CXR images	112,120 images (30,805 patients)	Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural	MobileNetV1, MobileNetV2, VGG16, InceptionV3, Xception, DenseNet-121, EfficientNet, ResNet50	Atelectasis: 0.75, Cardiomegaly: 0.73, Effusion: 0.87, Infiltration: 0.81, Mass: 0.89, Nodule: 0.72, Pneumonia: 0.69, Pneumothorax: 0.88, Consolidation: 0.85, Edema: 0.76, Emphysema: 0.78,

				Thickening, Hernia		Fibrosis: 0.82, PleuralThickening: 0.77, Hernia: 0.84
F. J. M. Shamrat et al. [4]	NIH ChestX-ray14 dataset	CXR images	112,120 images	Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Pulmonary Fibrosis, Pleural Thickening, Hernia, No Finding	InceptionV3, AlexNet, DenseNet121, VGG19, MobileNetV2	Atelectasis:0.936, Cardiomegaly:0.855, Effusion:0.931, Infiltration:0.912, Mass:0.883, Nodule:0.962, Pneumonia:0.904, Pneumothorax:0.967, Consolidation:0.889, Edema:0.958, Emphysema:0.935, Pulmonary Fibrosis:0.891, Pleural Thickening:0.872, Hernia:0.975
T.K.K. Ho and J. Gwak [5]	ChestX-ray14	Chest X-ray radiographs	112,120 images	Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Pulmonary Fibrosis, Pleural Thickening, Hernia, No Finding	deep features (DenseNet-121)	Atelectasis: 0.795, Cardiomegaly: 0.887, Effusion: 0.875, Infiltration: 0.703, Mass: 0.835, Nodule: 0.716, Pneumonia: 0.742, Pneumothorax: 0.863, Consolidation: 0.786, Edema: 0.892, Emphysema: 0.875, PulmonaryFibrosis: 0.756, Pleural Thickening: 0.774, Hernia: 0.836
X. Gong et al. [6]	ChestX-ray14	Chest X-ray radiographs	112,120 images	Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Pulmonary Fibrosis, Pleural	Deformable Gabor Feature Network (DGFN) with DenseNet121, DGFN with AlexNet and ResNet18	Atelectasis 0.817, Cardiomegaly: 0.928, Effusion: 0.875, Infiltration: 0.745, Mass: 0.8803, Nodule: 0.786, Pneumonia: 0.779, Pneumothorax: 0.893, Consolidation: 0.809, Edema: 0.892, Emphysema: 0.939, PulmonaryFibrosis:

				Thickening, Hernia, No Finding		0.817, Pleural Thickening: 0.814, Hernia: 0.921,
T.K.K. Ho and J. Gwak [7]	ChestX-ray14	Chest X-ray radiographs	112,120 images	Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Pulmonary Fibrosis, Pleural Thickening, Hernia, No Finding	DenseNet-121 (base), ResNet-152, ResNet-50, ResNet-32, MobileNet-v1/v2, VGG-19 with various KD strategies	Atelectasis: 0.797, Cardiomegaly: 0.896, Effusion: 0.882, Infiltration: 0.705, Mass: 0.844, Nodule: 0.752, Pneumonia: 0.763, Pneumothorax: 0.878, Consolidation: 0.798, Edema: 0.870, Emphysema: 0.981, PulmonaryFibrosis: 0.803, Pleural Thickening: 0.799, Hernia: 0.876,
Baltruschat et al. [8]	ChestX-ray14 [8]	chest X-ray images + metadata (age, gender, view position)	112,120 images	Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Pulmonary Fibrosis, Pleural Thickening, Hernia, No Finding	ResNet-50 (OTS, fine-tuned), ResNet-38, ResNet-101, ResNet-50-large-meta	Atelectasis: 0.747, Cardiomegaly: 0.865, Effusion: 0.818, Infiltration: 0.686, Mass: 0.796, Nodule0.738, Pneumonia: 0.694, Pneumothorax: 0.839, Consolidation: 0.734, Edema: 0.828, Emphysema: 0.868, PulmonaryFibrosis: 0.778, PleuralThickening: 0.739, Hernia: 0.855
S. Gündel et al. [9]	ChestX-ray14 & PLCO	chest X-ray images	112,120 images	Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Pulmonary	DenseNet (DNetLoc) & DNet and DNetLoc (with spatial labels)	Atelectasis: 0.826, Cardiomegaly: 0.911, Effusion: 0.885, Infiltration: 0.716, Mass: 0.854, Nodule: 0.774, Pneumonia: 0.765, Pneumothorax: 0.872, Consolidation: 0.806, Edema: 0.892, Emphysema: 0.925,

				Fibrosis, Pleural Thickening, Hernia, No Finding		PulmonaryFibrosis0.820, Pleural Thickening: 0.785, Hernia: 0.941

As per Table 2, the Advanced deep learning techniques have significantly improved the automated classification of lung diseases using chest X-ray (CXR) images. The NIH Chest X-ray14 dataset, having more than 112,000 images with 14 major abnormalities, is a key component of these efforts. Using models like MobileNetV1/V2, VGG16, InceptionV3, Xception, DenseNet121, EfficientNet, and ResNet50, M. Irtaza et al. [3] were able to obtain a F1 score of 0.552, specification of 0.983 for edema, binary accuracy of 0.93, recall of 0.55, precision of 0.56, and AUC of 0.76. F. J. M. Shamrat et al. [4] reported a result for edema with AUC of 0.958 accuracy of 0.9956, precision of 0.9522, recall of 0.9736, specificity of 0.9962, F1-Score of 0.9628 using InceptionV3, AlexNet, DenseNet121, VGG19, and MobileNetV2. T.K.K. Ho and J. Gwak [5] obtained an AUC of 0.892, accuracy of 0.8097 F1-Score of 0.8638, recall of 0.8799 and precision of 0.8489 for edema utilizing deep features collected using DenseNet-121. Similarly, X. Gong et al. [6] combined a Deformable Gabor Feature Network (DGFN) with DenseNet121, AlexNet, and ResNet18 to obtain an AUC of 0.892 for the diseases called edema. With DGFN DenseNet121 achieved the highest AUC of 0.870 for edema, with an overall average AUC of 0.8501 across 14 diseases, however, F1-score, precision, and recall were not provided. T.K.K. Ho and J. Gwak [7] used a combination of DenseNet121, ResNet variants, MobileNet, and VGG19 under various knowledge distillation approaches and achieved AUC of 0.870 for edema. The author, Baltruschat et al. [8] improved ResNet designs, achieving an AUC of 0.828 with 0.9895 accuracy, 0.700 precision, 0.85 recall, and 0.76 F1Score. After adding spatial labels to DenseNet variants (DNet and DNetLoc), S. Gündel et al. [9] achieved accuracy of 0.98, precision of 0.70, recall of 0.82, F1-Score of 0.76, and AUC of 0.892 for edema. All of these results point to a steady upward trend in edema detection performance, especially with improved feature learning.

Our work focuses on pulmonary edema detection it attempts to provide a customized solution for accurate and dependable binary classification of edema in chest X-rays by developing a specialized 7-layer CNN architecture using various deep learning & hybrid models and contrasting various optimizers meeting a crucial need in medical diagnostics. The next section discusses the dataset used for conducting experiments on Edema Detection.

3. DATASET USED

The ChestX-ray14 [10] dataset, a well-known and reliable collection assembled by the NIH Clinical Centre, is used in this investigation. One of the largest publicly accessible datasets of its type, this dataset consists of more than 112,000 frontal-view chest X-ray scans from 30,805 distinct patients. A useful tool for medical imaging studies, each image has labels for 14 common thoracic disorders, such as Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Pulmonary Fibrosis, Pleural Thickening, Hernia, Normal. In order to differentiate pulmonary edema from typical instances, we concentrated on binary classification for this study. For our experiments, we used a total of 4606 images, 606 images are used for testing, consisting of 303 edema and 303 normal images. Total 4000 training images were 2000 consists of edema and 2000 normal images. The table 2 shows total no. of edema and normal images. The dataset was divided into training & validation in the ratio of 80:20. The next section discusses the experimentation methodology used for conducting experiments.

Table 2. Pulmonary edema Dataset

Dataset	Type	Edema Images	Normal Images
Chest X-ray 14 [10]	Chest X-ray images	2303	2303

4. EXPERIMENTATION METHODOLOGY

This section describes the experiments conducted using Deep Learning models in the detection of pulmonary edema. All experiment were carried on using 64-bit Windows and 12th Gen Intel® Core™ i7-12650H CPU running at 2.30 GHz, 16 GB of RAM. TensorFlow was used for implementation. Standard libraries including NumPy, Pandas, Seaborn, and Matplotlib were utilized for data processing and visualization. In this research, 30 different algorithms have been evaluated completely to determine the best successful models for detecting edema in chest X-ray images. All the models were tested for different threshold values of 0.3, 0.4, 0.5. All the models were evaluated with respect to different evaluation parameters like AUC, precision, accuracy, F1-Score, and recall. After all the experimentations 6 best algorithms were chosen for further experimentation. These algorithms are ANN, DenseNet121, DenseNet121+ANN, DenseNet121+NasNet, DenseNet121+VGG16, DenseNet121+VGG19.

4.1 Image Preprocessing

The Figure 3 displays the image preprocessing using Chest X-ray dataset as an essential part for deep learning models as it improves training effectiveness. All of the images in this study were normalized by rescaling the pixel values using a factor of 1/255 from the conventional [0, 255] range to [0, 1]. In order to balance computational speed with the preservation of critical anatomical details required for precise edema identification, images were also scaled to a uniform 150×150-pixel resolution. The training and validation sets were divided in the ration of 80:20. In order to maintain the alignment, the test set of 606 images was processed using the same preprocessing pipeline but loaded independently without shuffling. With image preprocessing, each image was consistently prepared, allowing for accurate and effective model evaluations.

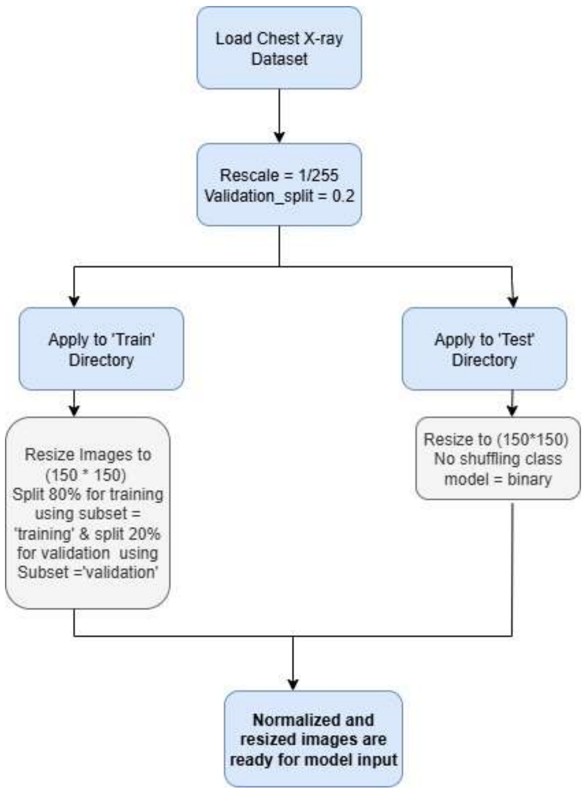

Fig. 3. Image preprocessing steps for pulmonary edema

4.2 Algorithms and Hyperparameters used

Different Deep Learning algorithm such as Artificial Neural Network (ANN), DenseNet121, DenseNet121+ANN, DenseNet+NasNet, DenseNet121+VGG16 and DenseNet121+VGG19 with different hyperparameters were used for detecting pulmonary edema based on a custom-designed 7-layer Convolutional Neural Network (CNN) that has been carefully developed to perform binary classification of chest X-ray images into NORMAL and EDEMA categories, addressing a critical diagnostic challenge in respiratory medicine. Table 3 exhibits the different hyperparameters used for all the algorithms in Edema detection. In order to standardize the data distribution and improve model convergence by reducing the effect of fluctuating image brightness, the preprocessing pipeline starts with the normalization of pixel values.

Table 3: Description for the hyperparameters used

Parameters	Values
Rescale	1/255 – Normalizes pixel values from [0, 255] to [0, 1]
Target_Size	(150, 150) – All images resized to 150×150 pixels
Batch_Size	32 – Number of images per batch
Class_Mode	'binary' – Binary classification: NORMAL vs EDEMA
Optimizers Used	SGD, Adam, Adagrad, RMSProp, Adadelta, Nadam, Adamax
Output Layer	Function: Sigmoid, Layer: Dense
Activation Function	Relu
Convolutional Layers	7
Epochs	5
Threshold	0.5

A batch size of 32 is intentionally selected to balance memory usage and training efficiency, enabling the model to handle large datasets without charging computational resources. Utilizing ReLU activation functions in each of its 7 convolutional layers, the model's design introduces non-linearity to capture complex spatial patterns and enables effective feature extraction, which is essential for differentiating minute pathological alterations in lung tissue. Global average pooling, which aggregates spatial information into a compact feature representation, comes after these layers to reduce dimensionality and prevent overfitting. This feature representation is then fed into a series of fully connected layers, each of which is equipped with 512, 256, and 128 neurons. Each of these layers is strengthened with a 0.5 dropout rate to improve regularization and lessen an excessive dependence on particular features. The final output layer of the classification process has a sigmoid activation, which generates probability scores that are thresholded at 0.5 to assign binary labels. The model is methodically trained using 7 different optimizers SGD, Adam, Adagrad, RMSProp, Adadelta, Nadam, and Adamax each applied to the same architecture in order to examine the effects of optimization strategies on convergence behaviour and classification performance. Training is carried out over 5 epochs, which is a sufficient amount of time to assess initial learning trends while minimizing computational cost. These metrics together offer a thorough understanding of the model's diagnostic capability. In addition to facilitating reliable and understandable edema classification, this methodical approach which combines careful preprocessing with a variety of optimization techniques has potential for practical implementation, especially in settings with limited resources where prompt diagnosis can have a substantial impact on patient outcomes.

4.3 Proposed Methodology

This section discusses the proposed methodology used for detection of pulmonary combining various optimizers for systematic evaluation with deep learning. To guarantee uniformity in input dimensions, all chest X-ray images are first reduced to a fixed resolution of 150*150 pixels and normalized to a pixel intensity range of [0, 1]. After that, the dataset is split into three subsets: a separate, unshuffled test set for objective assessment, 20% for validation, and 80% for training. 7 optimizers are used to explore how optimization strategies impact model performance: SGD, Adam, Adagrad, RMSProp, Adadelta, Nadam, and Adamax. A customized CNN approach is used for each optimizer. It consists of three fully connected dense layers (512 → 256 → 128), a Flatten layer, and 7 convolutional layers with ReLU activation and MaxPooling, all of which are regularized with a dropout rate of 0.5. The final output layer performs binary classification (Normal vs. Edema) using a sigmoid activation function. The training and validation sets are used to train the model for 5 epochs, with accuracy and binary crossentropy loss serving as the assessment metrics. Using a threshold of 0.5, predictions on the test set are transformed into binary class labels. The precision, recall, F1-score, acuraccy and the Area Under the ROC Curve (AUC) confusion matrix (TP, TN, FP, FN), are among the measures used to evaluate the model's performance. For every optimizer, the trained model and confusion matrix are saved, and the evaluation findings are combined into a thorough report that is saved in a given file. The proposed algorithm guarantees the reliable, accurate and effective results in edema detection. The next sections discuss the results, and findings of different experimentation conducted.

Proposed Algorithm for Edema Detection

Step 1. BEGIN

Step 2. Normalize all input images to range [0, 1]

Step 3. Resize images to 150×150 pixels

Step 4. Split dataset into:

a. Training set (80%)
b. Validation set (20%)
c. Test set (separate, unshuffled)

Step 5. Define list of optimizers:

OPTIMIZERS ← {SGD, Adam, Adagrad, RMSProp, Adadelta, Nadam, Adamax}

Step 6. FOR each optimizer OPT in OPTIMIZERS DO

Step 7. Initialize model with:

- 7 convolutional layers (ReLU + MaxPooling)
- Flatten layer
- Dense layers: 512 → 256 → 128 (each with Dropout=0.5)
- Output layer with sigmoid activation

Step 8. Compile model with:

- Optimizer = OPT
- Loss = Binary Crossentropy
- Metric = Accuracy

Step 9. Train model using training and validation sets for 5 epochs

Step 10. Predict on test set to get probabilities PROB

Step 11. Convert PROB to class labels PRED using threshold = 0.5

Step 12. Compute metrics:

- Confusion Matrix → TN, FP, FN, TP
- Precision, Recall, F1-score
- ROC curve → AUC

Step 13. Save trained model as “edema_model”

Step 14. Plot and save confusion matrix for OPT

Step 15. Store all metrics in RESULTS list

Step 16. END FOR

Step 17. Save RESULTS to “Edema_Result.csv”

Step 18. END

5. RESULTS, FINDING & DISCUSSION

This section discusses the evaluation results experiments. Optimizers like Adam, Nadam, and RMSProp consistently performed better than the rest among the evaluated configurations in terms of accuracy and AUC, suggesting improved generalization and convergence during training & validation. Interestingly, the RMSProp optimizer produced the best test accuracy & AUC, proving that it is capable of lowering loss while maintaining adaptive learning rates. SGD and Adagrad, on the other hand, showed relatively lower recall values despite being consistent. The confusion matrix showed that models assessed at the conventional threshold of 0.5 produced a lower false positives and false negatives, whereas assessments at lower thresholds (0.3 and 0.4) increased false positives and false negatives. The F1-score and AUC were found to be trustworthy measures of the model's robustness in every trial high-performing model’s AUCs continuously exceeded 0.85, confirming their potent selective powers. Overall, the findings confirm that the suggested model can be a dependable tool for detecting pulmonary edema in clinical situations, especially when adjusted with adaptive optimizers. It shows a good balance between computational efficiency and diagnostic accuracy.

5.1 Evaluation Criteria

The evaluation criteria are intended to offer an accurate assessment of the model's capacity to correctly identify pulmonary edema from chest X-ray images. The evaluation parameters on which the models are evaluated are accuracy, AUC, precision, recall & F1-Score with confusion matrix values such as False Positive (FP), True Positive (TP), False Negative (FN) & True Negative (TN).

5.2 Evaluation Results

With the three classification probability thresholds (0.3, 0.4, and 0.5) top 10 best performance results are shown in Table 4, 5, 6 in experimentation to see how well our deep learning models performed in identifying pulmonary edema from chest X-ray images. These thresholds gave us a better understanding of how each model functions at various levels of decision confidence. To provide an in-depth overview of model performance, the assessment concentrated on a number of metrics, including test accuracy, precision, recall, F1-score, Area Under the ROC Curve (AUC), and confusion matrix components (True Positives [TP], True Negatives [TN], False Positives [FP], and False Negatives [FN]). The efficacy of results was shown better at threshold 0.5.

Table 4. Result for threshold 0.3 Edema

Sr. No.	Algorithm Used	Accuracy	Precision	Recall	F1-Score	AUC	TN	TP	FN	FP
1	EfficientNetB0+ DenseNet	0.9290	0.9167	0.9439	0.9301	0.9821	277	286	17	26
2	RondomForest	0.9109	0.8903	0.9373	0.9132	0.9708	268	284	19	35
3	DenseNet121	0.9290	0.9431	0.9307	0.9369	0.9400	286	282	21	17
4	DenseNet169	0.9109	0.8978	0.9274	0.9123	0.9721	271	281	22	32
5	DenseNet121+AlexNet	0.9488	0.9823	0.9142	0.9470	0.9862	298	277	26	5
6	DenseNet121+NasNet	0.9505	0.9928	0.9076	0.9483	0.9902	301	275	28	2
7	DenseNet121+CNN	0.9307	0.9611	0.8977	0.9283	0.9858	292	272	31	11

8	SVM	0.9175	0.9600	0.8713	0.9135	0.9828	292	264	39	11
9	Efficient+NasNet	0.9043	0.9329	0.8713	0.9010	0.9702	284	264	39	19
10	DenseNet121+VGG19	0.9092	0.9593	0.8548	0.9040	0.9642	292	259	44	11

Table 5. Result for threshold 0.4 Edema

Sr. No.	Algorithm Used	Accuracy	Precision	Recall	F1-Score	AUC	TN	TP	FN	FP
1	ANN	0.8762	0.8081	0.9868	0.8886	0.884	232	299	4	71
2	DenseNet121+NasNet	0.9604	0.9861	0.934	0.9593	0.9901	299	283	20	4
3	DenseNet121	0.9158	0.9574	0.8911	0.9231	0.9300	291	270	33	12
4	DenseNet121+AlexNet	0.9323	0.9712	0.8911	0.9294	0.9797	295	270	33	8
5	Efficient+NasNet	0.9076	0.9364	0.8746	0.9044	0.9718	285	265	38	18
6	DenseNet121+VGG19	0.9224	0.9848	0.8581	0.9171	0.9753	299	260	43	4
7	DenseNet169	0.8894	0.9338	0.8383	0.8835	0.9539	285	254	49	18
8	SVM	0.9059	0.9695	0.8383	0.8991	0.9828	295	254	49	8
9	DenseNet121+ANN	0.9059	0.9767	0.8317	0.8984	0.9671	297	252	51	6
10	CNN+DenseNe121+NasNet	0.9142	1	0.8284	0.9061	0.9769	303	251	52	0

Table 6. Result for threshold 0.5 Edema

Sr. No.	Algorithm Used	Accuracy	Precision	Recall	F1-Score	AUC	TN	TP	FN	FP
1	**DenseNet121+VGG16**	**0.9191**	**0.8825**	**0.9670**	**0.9228**	**0.9830**	**264**	**293**	**10**	**39**
2	**ANN**	**0.9323**	**0.9119**	**0.9571**	**0.934**	**0.9406**	**275**	**290**	**13**	**28**
3	**DenseNet121+ANN**	**0.9224**	**0.9384**	**0.9043**	**0.9210**	**0.9718**	**285**	**274**	**29**	**18**
4	**DenseNet121+VGG19**	**0.9158**	**0.9846**	**0.8449**	**0.9094**	**0.9738**	**299**	**256**	**47**	**4**
5	**DenseNet121**	**0.9125**	**0.9883**	**0.8350**	**0.9052**	**0.9100**	**300**	**253**	**50**	**3**
6	**DenseNet121+NasNet**	**0.9175**	**1**	**0.8350**	**0.9101**	**0.9883**	**303**	**253**	**50**	**0**
7	Efficient+NasNet	0.8960	0.9580	0.8284	0.8885	0.8960	292	251	52	11
8	DenseNet121+AlexNet	0.8944	0.9878	0.7987	0.8832	0.9694	300	242	61	3
9	EfficientNetB0+ DenseNet	0.8894	0.9836	0.7921	0.8775	0.9735	299	240	63	4
10	CNN+DenseNe121+NasNet	0.8944	1	0.7888	0.8819	0.9624	303	239	64	0

5.3 Proposed Methodology Results

To enhance the performance of pulmonary edema detection models, we implemented a series of optimization techniques, and fine-tuning the hyperparameters. The 0.5 threshold consistently delivered the most robust outcomes, striking an optimal balance between high precision and strong recall, making it ideal for clinical applications where both accurate detection and minimal false positives are crucial. The algorithms experimented using proposed methodology with optimizers were SGD, Adam, Adagrad, RMSProp, Adadelta, Nadam, Adamax. Table 7 to 12 shows the evaluation results with their optimizers. Models were paired with the RMSprop optimizer at 0.5 threshold achieved notable improvements, with higher precision, F1-score, Recall, reflecting enhanced prediction confidence and reliability. Consequently, we selected the 0.5 threshold as the optimal decision boundary for final reporting, as it best supports real-world diagnostic needs by ensuring reliable edema detection while reducing unnecessary false alarms, demonstrating a robust model.

Table 7. Optimizer-wise Edema Results for ANN Model using 0.5 threshold

Edema_ANN Convolutional Layer 7										
Sr. No.	**Optimizer**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC**	**TN**	**FP**	**FN**	**TP**
1	SGD	0.8383	0.9362	0.7261	0.8178	0.9245	288	15	83	220
2	Adam	0.9010	0.9841	0.8152	0.8917	0.9750	299	4	56	247
3	Adagrad	0.8069	0.9650	0.6370	0.7674	0.9268	296	7	110	193
4	**RMSProp**	**0.9224**	**0.9444**	**0.8977**	**0.9205**	**0.9824**	**287**	**16**	**31**	**272**
5	Adadelta	0.5000	0.5000	1.0000	0.6667	0.7032	0	303	0	303
6	Nadam	0.8383	0.9904	0.6832	0.8086	0.9763	301	2	96	207
7	Adamax	0.8317	0.9763	0.6799	0.8016	0.9577	298	5	97	206

Table 8. Optimizer-wise Edema Results for DenseNet121Model using 0.5 threshold

Edema_DenseNet121 + Convolution layer 7										
Sr. No.	**Optimizer**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC**	**TN**	**FP**	**FN**	**TP**
1	SGD	0.8927	1.0000	0.7855	0.8799	0.9812	303	0	65	238
2	Adam	0.9208	1.0000	0.8416	0.9140	0.9892	303	0	48	255
3	Adagrad	0.8762	0.9634	0.7822	0.8634	0.9656	294	9	66	237
4	**RMSProp**	**0.8762**	**0.8497**	**0.9142**	**0.8808**	**0.9583**	**254**	**49**	**26**	**277**
5	Adadelta	0.4983	0.4933	0.1221	0.1958	0.5029	265	38	266	37
6	Nadam	0.8944	0.9918	0.7954	0.8828	0.9256	301	2	62	241
7	Adamax	0.8696	1.0000	0.7393	0.8501	0.9751	303	0	79	224

Table 9. Optimizer-wise Edema Results for DenseNet121+ANN Model using 0.5 threshold

Edema_DenseNet121+ANN + Convolution layer										
Sr. No.	**Optimizer**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC**	**TN**	**FP**	**FN**	**TP**
1	SGD	0.8663	0.9587	0.7657	0.8514	0.9550	293	10	71	232
2	Adam	0.9125	0.9808	0.8416	0.9059	0.9724	298	5	48	255
3	Adagrad	0.7921	0.8766	0.6799	0.7658	0.9014	274	29	97	206
4	RMSProp	0.9076	0.9805	0.8317	0.9000	0.9813	298	5	51	252
5	Adadelta	0.5990	0.8191	0.2541	0.3879	0.7203	286	17	226	77
6	**Nadam**	**0.9158**	**0.9632**	**0.8647**	**0.9113**	**0.9683**	**293**	**10**	**41**	**262**
7	Adamax	0.9092	0.9429	0.8713	0.9057	0.9685	287	16	39	264

Table 10. Optimizer-wise Edema Results for DenseNet121+NasNet Model using 0.5 threshold

Edema_Densenet121+Nasnet + Convolution layer 7										
Sr. No.	**Optimizer**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC**	**TN**	**FP**	**FN**	**TP**
1	SGD	0.9307	0.9925	0.8680	0.9261	0.9901	301	2	40	263
2	Adam	0.8333	1.0000	0.6667	0.8000	0.9905	303	0	101	202
3	Adagrad	0.8498	0.9609	0.7294	0.8293	0.9486	294	9	82	221
4	**RMSProp**	**0.9455**	**0.9927**	**0.8977**	**0.9428**	**0.9948**	**301**	**2**	**31**	**272**
5	Adadelta	0.5330	0.5239	0.7228	0.6075	0.5620	104	199	84	219
6	Nadam	0.8663	1.0000	0.7327	0.8457	0.9786	303	0	81	222
7	Adamax	0.9224	0.9812	0.8614	0.9174	0.9796	298	5	42	261

Table 11. Optimizer-wise Edema Results for DenseNet121+VGG16 Model using 0.5 threshold

Edema_DenseNet121+VGG16 + Convolution layer 7										
Sr. No.	**Optimizer**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC**	**TN**	**FP**	**FN**	**TP**
1	SGD	0.9059	0.9960	0.8152	0.8966	0.9791	302	1	56	247
2	Adam	0.9092	0.9806	0.8350	0.9020	0.9828	298	5	50	253
3	Adagrad	0.8003	0.8922	0.6832	0.7738	0.8792	278	25	96	207
4	RMSProp	0.8597	0.9910	0.7261	0.8381	0.9563	301	2	83	220
5	Adadelta	0.4109	0.4405	0.6601	0.5284	0.3833	49	254	103	200
6	**Nadam**	**0.9340**	**0.9782**	**0.8878**	**0.9308**	**0.9532**	**297**	**6**	**34**	**269**
7	Adamax	0.8894	0.9876	0.7888	0.8771	0.9705	300	3	64	239

Table 12. Optimizer-wise Edema Results for DenseNet121+VGG19 Model using 0.5 threshold

Edema_DenseNet121+VGG19 + Convolution layer 7										
	Optimizer	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC**	**TN**	**FP**	**FN**	**TP**
1	SGD	0.8944	0.9959	0.7921	0.8824	0.9791	302	1	63	240
2	**Adam**	**0.9125**	**0.9032**	**0.9241**	**0.9135**	**0.9611**	**273**	**30**	**23**	**280**
3	Adagrad	0.8812	0.9425	0.8119	0.8723	0.9608	288	15	57	246
4	RMSProp	0.8977	0.9547	0.8350	0.8908	0.9625	291	12	50	253
5	Adadelta	0.5149	0.5158	0.4851	0.5000	0.5000	165	138	156	147
6	Nadam	0.8894	0.9758	0.7987	0.8784	0.9408	297	6	61	242
7	Adamax	0.9010	1.0000	0.8020	0.8901	0.9507	303	0	60	243

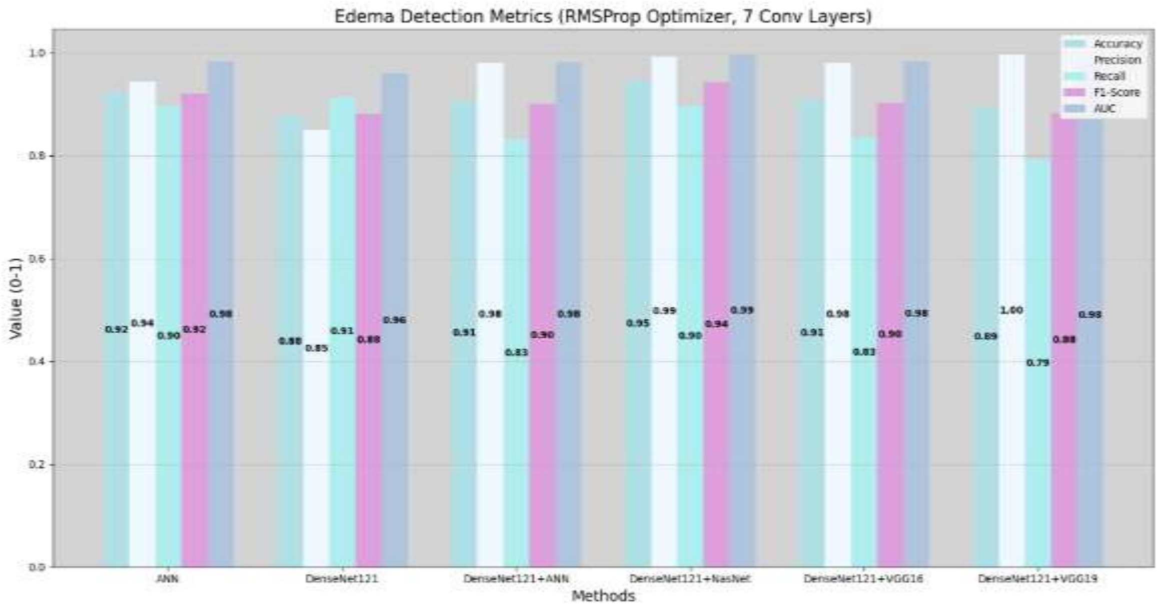

Fig 4. Comparison of proposed methodology with Best Optimizers

Figure 4 shows how well different deep learning models performed in detecting pulmonary edema using the RMSProp optimizer with 7 convolutional layers. It compares six models: Artificial Neural Network (ANN), DenseNet121, DenseNet121+ANN, DenseNet+NasNet, DenseNet121+VGG16 and DenseNet121+VGG19. Among these, the DenseNet121+NasNet model gave the best results, with the highest accuracy (0.94), AUC (0.99), precision (0.99), recall (0.89), and F1-score (0.94). This means it was very good at correctly identifying both positive and negative cases.

The ROC (Receiver Operating Characteristic) curve is a key visual representation of the diagnostic performance of various deep learning and hybrid models in classifying pulmonary edema from chest X-ray images, providing insight into their ability to balance sensitivity and specificity. The True Positive Rate (Recall) and False Positive Rate (FPR) are carefully plotted for six different models in this graph: Artificial Neural Network (ANN), DenseNet121, DenseNet121+ANN, DenseNet+NasNet, DenseNet121+VGG16 and DenseNet121+VGG19. The Figure 5 shows the ROC curve for all 5 algorithms. Each model is displayed as a distinct point on the curve, with its position relative to the diagonal line representing the performance of a random classifier clearly indicating its discriminative power. The DenseNet121+NasNet hybrid is the best of these, with a remarkable recall of 0.99. This suggests an ideal trade-off that improves sensitivity while preserving a respectable level of specificity, making it especially useful in healthcare settings where early detection is crucial.

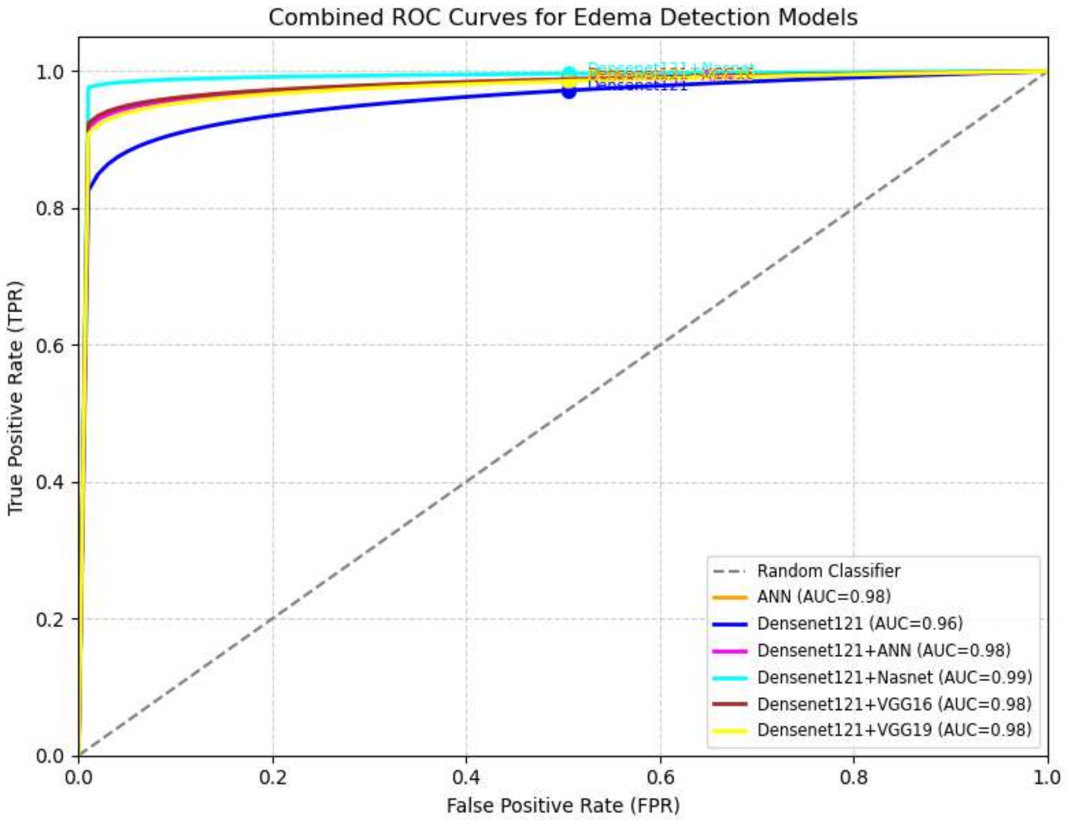

Fig 5. ROC Curve for the classification of Edema

5.4 Comparison Between Proposed Methodology & Existing Methodology

The hybrid DenseNet121 + NasNet model, which produced excellent precision, recall, accuracy, F1-score, and AUC, making it a suitable option for identifying pulmonary edema in chest X-rays. Because of its capacity to find a balance between recall and precision, it detects the majority of edema cases while minimizing false positives, which is essential for practical clinical use. Our proposed methodology results are compared with the existing models of edema detection in terms of AUC. Our model outshines with the 0.99 AUC value as compared to all other existing models. Table 12 shows the comparison results between our proposed methodology and existing models. The next section discusses the conclusion and future scope.

Table 12. Comparison between existing models and proposed methodology

Other Approaches/Models	AUC
M. Irtaza et al. [3]	0.76
F. J. M. Shamrat et al. [4]	0.95
T.K.K. Ho and J. Gwak [5]	0.89
X. Gong et al. [6]	0.89
T.K.K. Ho and J. Gwak [7]	0.87
Baltruschat et al. [8]	0.82
S. Gündel et al. [9]	0.89
Proposed Methodology	**0.99**

6. CONCLUSION AND FUTURE SCOPE

This research successfully proves the use of deep learning and hybrid models to enable the diagnostic identification of pulmonary edema in chest X-rays. Using 30 deep learning and hybrid models with different thresholds (0.3, 0.4, 0.5), the top performing 6 models were selected namely DenseNet121, ANN, DenseNet121+VGG16, DenseNet121+VGG19 and DenseNet121+NasNet. These selected models were used further for advanced optimization with different hyperparameters and optimizers namely SGD, RMSProp, Adam, Nadam, AdaDelta and AdaGrad. The DenseNet121 + NasNet with RMSProp optimizer model gave excellent results at the 0.5 threshold with 94.55% accuracy, 99.27% precision, 89.77% recall, and F1-score of 0.9428 and AUC 0.9948, outperforming all the current comparative approaches. These results provide evidence that the proposed methodology represents a reliable, scalable and accurate diagnostic support available for use in low-resource healthcare contexts where access to expert radiologists is sparse. Future work can focus on a larger dataset and additional multi-modal clinical data. A real-time implementation for edema detection on affordable or portable devices to improve accessibility in underserved areas can also be a focus in future. This research work lays a strong foundation for the development of AI-assisted, point of care diagnostic systems that can improve pulmonary care around the world.

References

[1] Degefu N, Jambo A, Nigusse S, Dechasa M, Gashaw T, Getachew M. The Burden and Contributing Factors of Cardiogenic Pulmonary edema Among Acute Heart Failure Patients Admitted to Tertiary Hospital, Eastern Ethiopia. Open Access Emerg Med. 2023; 15:405-414 https://doi.org/10.2147/OAEM.S436352

[2] T. Yan et al., “Burden, Trends, and Inequalities of Heart Failure Globally, 1990 to 2019: A Secondary Analysis Based on the Global Burden of Disease 2019 Study,” Journal of the American Heart Association, vol. 12, no. 6, Mar. 2023, doi: https://doi.org/10.1161/jaha.122.027852.

[3] M. Irtaza, A. Ali, M. Gulzar and A. Wali, "Multi-Label Classification of Lung Diseases Using Deep Learning," in IEEE Access, vol. 12, pp. 124062-124080, 2024, doi: 10.1109/ACCESS.2024.3454537.

[4] F. J. M. Shamrat, S. Azam, A. Karim, K. Ahmed, F. M. Bui, and F. De Boer, “High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images,” Computers in Biology and Medicine, vol. 155, p. 106646, Mar. 2023, doi: 10.1016/j.compbiomed.2023.106646.

[5] T.K.K. Ho, J. Gwak, Multiple feature integration for classification of thoracic disease in chest radiography, Appl. Sci. 9 (19) (2019) 4130.

[6] X. Gong, X. Xia, W. Zhu, B. Zhang, D. Doermann, L.A. Zhuo, Deformable gabor feature networks for biomedical image classification, in: Proceedings of the IEEE/ CVF Winter Conference on Applications of Computer Vision, 2021, pp. 4004–4012.

[7] T.K.K. Ho, J. Gwak, Utilizing knowledge distillation in deep learning for classification of chest X-ray abnormalities, IEEE Access 8 (2020) 160749–160761.

[8] M. Ivo, Baltruschat, hannes nickisch, michael grass, tobias knopp, and axel saalbach. Comparison of deep learning approaches for multi-label chest x-ray classification, Sci. Rep. 9 (6381) (2019).

[9] S. Guendel, S. Grbic, B. Georgescu, S. Liu, A. Maier, D. Comaniciu, learning to recognize abnormalities in chest X-rays with location-aware dense networks, in: Proc. Iberoamer. Congr. Pattern Recognit., Springer, Cham, Switzerland, Nov. 2018, 757765.

[10] “NIH Chest X-rays,” www.kaggle.com. https://www.kaggle.com/datasets/nih-chest-xrays/