Received: 2025-10-28
Accepted: 2026-01-11
Published Online: 2026-01-20
DOI: 10.70454/JRICST.2026.30103

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

AI-based Diagnostic System for Chest X-rays: A Multi-Labeled Classification Approach using Deep Learning

Deepak Kumar
Professor & Dean, College of Science and Technology Surajmal University, Kichha Uttrakhand, India.
Email: deepakraj@smu.ac.in
*Corresponding Author. deepakraj@smu.ac.in

Abstract

Although chest X-rays (CXRs) are still a vital diagnostic tool for detecting thoracic disease, their interpretation can be challenging because of their multilevel findings and contradictory visual patterns. As a result, we examine how well deep convolutional neural networks (CNNs) with transfer learning perform automated multi-label classification of CXRs. Extensive preprocessing and augmentation techniques were used to address class imbalance and normalise image quality using the CheXpert dataset. Under consistent experimental conditions, several CNN architectures, including CustomNet, DenseNet121, ResNet50, InceptionV3, and VGG16, were trained and evaluated. With an AUROC of 0.78 and an accuracy of 87% on test data, DenseNet121 performs significantly better than all other models, according to a comparative analysis of AUROC and accuracy. Additional evaluation by disease category on an individual basis showed excellent performance for pleural effusion (AUROC 0.93) and lung opacity (AUROC 0.91). These results indicate the promise of Dense Net-based architectures to deliver accurate, automated diagnostic assistance to clinical radiology. The work emphasizes the utility of transfer learning in enhancing generalization with sparse labeled data and offers pragmatic guidance to model choice in the analysis of medical images.

Keywords: Deep learning, Multi-label classification, Chest X-ray, Transfer learning, DenseNet, Medical image analysis

1.0 Introduction

There are two subspecialties of radiology in medicine: diagnostic radiology and interventional radiology [1]. In diagnostic radiology, abnormalities and diseases are identified by analyzing medical images. The most used test that is correctly and quickly interpreted to prevent potentially fatal diseases is chest X-ray radiography. The difficulty comes when radiologists are required to interpret these images, and their abilities are restricted by time, experience, and the need to hire a certified radiologist. To automate and produce accurate radiology reporting, the healthcare industry turned to deep learning algorithms [2].

These neural networks consist of layers of interconnected nodes (artificial neurons) that process and transform input data into meaningful output. Deep learning is characterized by the depth of these neural networks [3]. The strength of such a connection is determined by the weights of the variable or features that associate inputs with outputs.

Possible Advantages for Medical Fields:

These are just some of the many advantages AI is said to offer the medical field-

- **More affordable treatment:** Automation makes diagnosis quicker and more accurate. Physicians can then prescribe the best courses of action or step in early to avoid sickness and necessitating costlier care.
- **More secure options:** A lower risk of complications from patients receiving ineffective or incorrect treatment is associated with more accurate diagnosis.
- **More patients received care:** Labs can run more tests when the time it takes to finish a diagnostic analysis is reduced. More patients will be covered in less time as a result of this.
- **Handling the "Physician Shortage" on a global scale:** Many countries are concerned about the growing gap between the demand and supply of physicians. Global shortages of doctors, nurses, and other health workers are estimated by the WHO. Because of the scarcity of medical schools and their limited capacity, the shortage is frequently worse in developing countries [4]. These countries also have a far higher proportion of rural and isolated areas, which exacerbates the problem, like poor transportation. Similar to an unfinished jigsaw puzzle, the world's emerging need for more advanced medical personnel will require time and money to train, which will make meeting demand unlikely. Deep learning systems combined with automation offer a comparatively faster, more scalable, and more affordable solution to salvage the situation [5].

Given the world's population expansion, we will need alternative AI-based medical workers to assist us accomplish the sustainable development objective of delivering "Affordable, Accurate, and Adequate Healthcare for All". After considering the trends and current advancements in deep learning, we chose to experiment with many variants of convolutional neural networks for this project.

We accepted this discovery and used the power of the Dense Convolutional Network (DenseNet), which connects each layer to the next in a feed-forward fashion. We also tested with alternative architectures to compare their relative performances using standard performance matrices such as AUROC and Accuracy.

We will explore the major and popular transfer learning techniques and see their performances on our Chest X-rays data. Over the past, transfer learning had been applied to improve the efficiency of CNNs or other neural network architectures and performed positively [6]. They have several benefits that enhance the applicability and performance of deep learning A.I. models in various domains. By using knowledge from pre-training, models are set to better utilize limited labeled data for different tasks. Pre-trained models serve as a strong starting point, reducing the time and resources needed for training. Transfer learning can handle changes in data distribution, making it useful in diverse real-world scenarios. Overall, it plays a crucial role in improving the performance and efficiency of machine learning models in a wide variety of applications and domains [7].

Received: 2025-10-28
Accepted: 2026-01-11
Published Online: 2026-01-20
DOI: 10.70454/JRICST.2026.30103

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

2.0 Literature Review

There has been a notable public release of large radiology image datasets in the recent past. Utilizing these datasets has been essential to maximizing group efforts for the creation and evaluation of machine learning models. [2], [3], [4], and [5]. The NIH Clinical Center released more than 100,000 chest X-ray images to the scientific community in 2017. Of these, 108,948 frontal-view X-ray images were part of 32,717 unique cases. Many studies have made use of this dataset, including Wang et al.'s [6] work, which showed how to use a unified weakly supervised multi-label image classification to detect and spatially locate common thoracic diseases in their paper.

Subsequent paragraphs, however, are indented. Techniques based on deep learning have been created to categorize X-ray pictures of the chest and pinpoint potential diseases. ROC statistics and rank correlation can be used to compare the effectiveness of several deep learning models in the categorization of chest X-rays [7]. A number of deep learning architectures, such as an expanded ResNet-50 architecture and a network that integrates non-image data to take use of the great spatial resolution of X-ray data, have been studied for the classification of chest X-rays . Class activation maps are useful for comprehending the classification procedure and for thoroughly examining the influence of non-image features on the classification of chest X-rays For the categorization of chest The application of numerous ResNet depths, including ResNet-38 and ResNet-101, to X-rays can also be studied. For chest X-ray classification, studies have shown that the X-ray-specific ResNet-38, which incorporates non-image data, produces the best overall results [7].

Furthermore, ResNet-50 has been recognised as a potent network architecture for classifying chest X-rays that can be applied to training from scratch, fine-tuning, or transfer learning. Deep learning methods for chest X-ray classification have gained popularity due to the availability of labelled X-ray image archives. Several well-designed CNN architectures, including VGG, GoogleNet, ResNet, and DenseNet, have been used to train deep learning models for CXR analysis [7][8].

Using 112,120 frontal chest X-rays from 30,805 patients, the ChestX-ray14 dataset has been used to assess deep learning-based methods for multi-label disease classification. It's a multilabel classification challenge because the dataset includes photos with numerous abnormalities and images clear of disease. Three open-source datasets were also utilized in the study: the ImageNet, ChestX-ray, and CheXpert datasets [7][8]. A thorough examination of the various transfer learning attributes in medical image analysis is necessary to optimize the effectiveness of transfer learning for CXR image classification. The study's findings might help develop best practices for the effective utilization of various data sets to help alleviate the lack of training data and improve deep learning models' performance in the medical domain [8].

In recent years, there are challenges in chest radiography, where a long-tailed distribution of clinical findings poses difficulties for standard deep learning methods biased towards more common classes. Proposing some effective techniques, the paper [9] employs EfficientNetV2 and ConvNeXt as primary architectures, incorporating image size influence in architectural decisions. To counter dataset imbalance and the multi-label nature of chest X- ray detections, the paper utilizes many augmentations, including mosaic augmentation, and modifies label acquisition

Received: 2025-10-28
Accepted: 2026-01-11
Published Online: 2026-01-20
DOI: 10.70454/JRICST.2026.30103

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

method, and performance is enhanced through ensemble strategies. Overall, the paper provides aextensive review of methods addressing the long-tailed imbalance in chest X-ray datasets using advanced techniques and architectures. DenseNet has outperformed other models that we used for this paper later on. A chest radiograph will usually reveal a limited number of common findings and a considerably greater number of rare findings. An experienced radiologist may learn the visual presentation of rare illnesses by looking at a few representative examples, whereas an algorithm requires many more. Teaching a machine to learn from a "long-tailed" distribution is difficult since typical methods build models that are biassed towards the common classes and perform poorly on the unusual. According to [11], the benchmark consists of two chest X-ray datasets for 19- and 20-way thorax illness classification, with classes having as many as 53,000 and as few as 7 labelled training pictures. This new benchmark is used to evaluate both standard and cutting-edge long-tailed learning approaches, allowing us to determine which features of these methods are most advantageous for long-tailed medical picture categorization.

As we all know, deep neural networks are difficult to train. This study [12] introduces a residual learning paradigm for facilitating the training of networks that are substantially deeper than those previously used. This explicitly rewrites the layers as learning residual functions with respect to the layer inputs rather than learning unreferenced functions. It presents thorough factual information suggesting that these residual networks are easier to optimize and can benefit from more depth.

3.0 Methodology

Our main focus is on developing a CNNbased model to predict the models. The approach used in this paper involves the utilization Convolutional Neural Network(CNN) architectures for disease prediction using chest X-ray data. To make sure the data was appropriate for training and assessment, we first implemented preprocessing procedures. We investigate several cutting-edge CNN architectures, including ResNet and DenseNet, with the primary goal of contrasting how well they predict diseases. These were chosen because prior research had demonstrated their effectiveness in medical image analysis tasks. A meticulously planned experimental setup that includes crucial hyper parameters like learning rate and batch size is then used to train, validate, and test the models in order to guarantee reliable performance evaluation. Performance is used to evaluate the models' efficacy. Our research attempts to shed light on the relative advantages and disadvantages of various CNN architectures in relation to disease prediction using chest X-rays.

3.1 Data Collection and Exploration

For training our model, we used a version of the ChexPert chest X-ray dataset. The CheXpert dataset is a famous benchmark dataset in medical imaging, especially for interpreting chest radiographs. Irvin et al. introduced it in their publication, "CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison" (2019). The dataset includes chest radiographs (X-ray images), as well as radiologist interpretations and labelling for 14 common thoracic diseases such as pneumonia, pleural effusion, and cardiomegaly. Each X-ray radiograph in the dataset is accompanied by a feature vector containing the patient ID, study number, X-ray view type, and a collection of fourteen expert-labelled observations. We clean the data to omit the age and sex features from the equation because we are evaluating solelyon the image.

Received: 2025-10-28
Accepted: 2026-01-11
Published Online: 2026-01-20
DOI: 10.70454/JRICST.2026.30103

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

3.2 Data Pre-processing and Augmentation

The effectiveness of CNN models for analyzing X-rays depends very much on how we prepare data. As we initially checked the dataset, we noticed large variations in how the X-ray pictures looked, like their resolution and brightness were totally off. To address this, we used a procedure to standardize the pixel values in the images, making comparisons much simpler. We continued by resizing the images uniformly to a 224x224 pixel square in accordance with the specifications of the chosen CNN design. Since the model's strength is crucial, we added various techniques to the data to increase its diversity. This entailed chaotically rotating, flipping, and shifting the images at random. controlled rotation between -10 and -10 degrees, made possible by expansion to enhance the dataset while keeping crucial diagnostic characteristics. This makes it possible for the model to learn from a greater range of X-ray variations.

We know that every grouched fair mix of the cases; with chest illnesses, so our model wouldn't be biased by having a lot of one kind of data, hopefully. All of these steps in preparing the data make a good foundation for using the model to identify illnesses from X-rays, tackling challenges like variances in the data and having only a restricted amount of examples to learn from.

The transformational pipeline was designed in such a way that there is a balanced data augmentation with the preservation of clinically significant details. Extreme alterations such a scrapping were avoided, there by, ensuring the retention of important clinical features within the images. Sample images from the transformed dataset are as below:

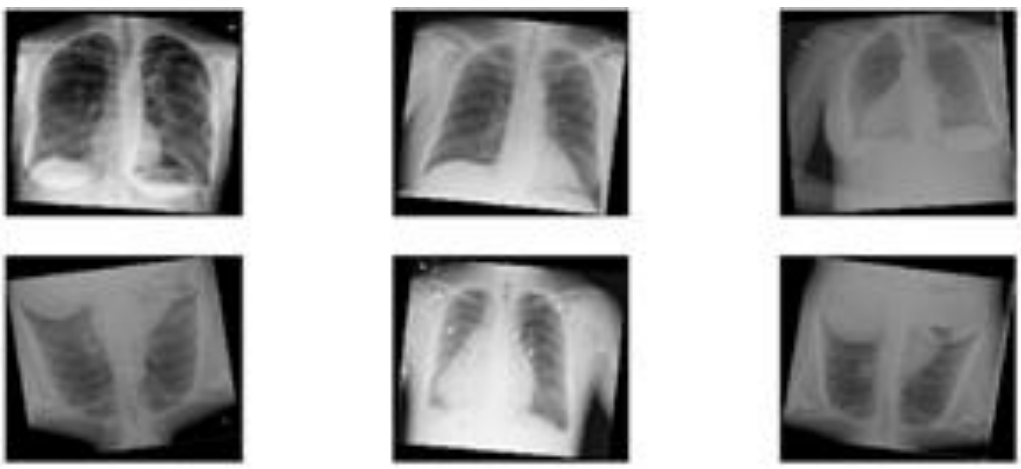

Fig.1. Snapshotofthe X-ray images in the data set after augmentation

3.3Modelling

Our research brought us to the point that CNN architecture excels in the multi-class, multi-label classification of image datasets. This is due to a decrease in the number of parameters. This reduction doesn't compromise crucial features that are necessary for accurate predictions. Consequently, wedeled into several CNN-based models, which will be explored in more detail later. Each of these models won't through testing using a batch size of 96 for a maximum of 40

Received: 2025-10-28
Accepted: 2026-01-11
Published Online: 2026-01-20
DOI: 10.70454/JRICST.2026.30103

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

epochs. Binary cross-entropy was used as the loss function. Adam optimizer with an starting learning rate of 0.001, and an adaptive learning rate adjustment: multiplying it by 10 was used when the validation loss levels off after an epoch.

3.4 Custom Net

We started by developing a basic custom CNN model. The model was built from scratch, starting with random settings and adjusting them. The input then goes through 4 layers, and each layer does specific tasks like identifying patterns and simplifying information. In each layer, max pooling is done thus reducing the size and activation step using ReLU. In Figure2, you can see the details for each layer. After going through these layers, the information heads to a fully connected part that uses a sigmoid function. This function helps turn the raw output into probabilities, telling us how likely it is that a chest X-ray has disease we're looking for. If the probability is found greater than 0.50, we see it as a positive finding.

CustomNet

Modules	Parameters
ConvLayer1.0.weight	216
ConvLayer1.0.bias	8
ConvLayer1.1.weight	1152
ConvLayer1.1.bias	16
ConvLayer2.0.weight	12800
ConvLayer2.0.bias	32
ConvLayer2.1.weight	9216
ConvLayer2.1.bias	32
ConvLayer3.0.weight	18432
ConvLayer3.0.bias	64
ConvLayer3.1.weight	102400
ConvLayer3.1.bias	64
ConvLayer4.0.weight	204800
ConvLayer4.0.bias	128
ConvLayer4.1.weight	147456
ConvLayer4.1.bias	128
Lin1.0.weight	7168
Lin1.0.bias	14

Total Trainable Params: 504126

Fig2. DetailsoftheCustomNetmodel

3.5 Dense Net

Dense Net, Densely Connected Convolutional Network, is a type of neural network architecture known for its dense connectivity pattern, where each layer is connected to other layer in a densely packed way. Dense blocks are made up of many units. Each unit produces a fixed-size feature vector after packing two convolutions that are each followed by Batch Normalization and ReLU activations. The amount of new information that the layers permit to pass through is controlled by a parameter known as the growth rate.

Conversely, transition layers are really basic parts made to down sample the features that are moving through the network. Each transition layer has three layers: a 1x1 convolution, a 2x2 average pooling, and a Batch Normalization layer. For this paper, DenseNet121 was trained with initial weights from a pre-trained network on ImageNet data [13].

Received: 2025-10-28
Accepted: 2026-01-11
Published Online: 2026-01-20
DOI: 10.70454/JRICST.2026.30103

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

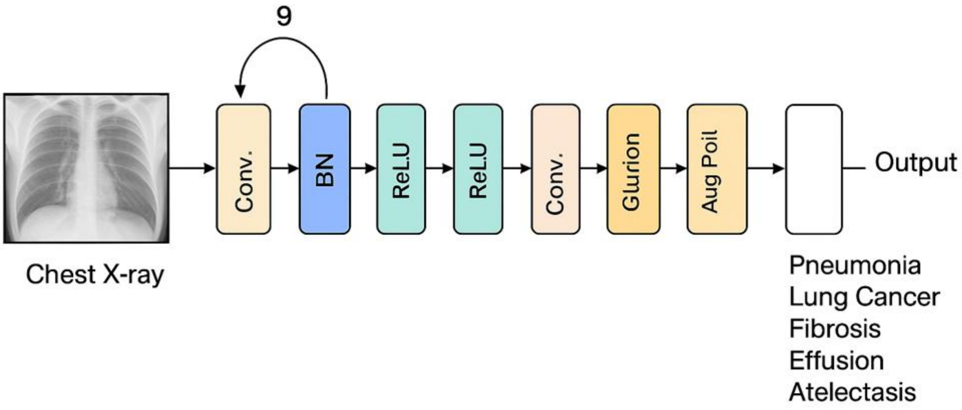

Fig3. DenseNet architecture

4.0 Results & Analysis

ResNet50

A deep residual network, often known as deep ResNet, is a type of specialised neural network that helps with more complex deep learning tasks by allowing for better results with deeper designs. Its effectiveness in training deep networks has received increased attention in recent years. ResNet-50, a 50-layer deep convolutional neural network, was pre-trained using more than a million images from the ImageNet dataset. As a result, it can classify photos into 1000 different object categories and has developed detailed feature representations for a wide range of images. One issue that experts regularly raise with deep networks composed of hundreds of layers is that accuracy can get saturated and degraded. The problem of vanishing gradient worries researchers equally. We freeze pre-trained network initial weights for the first 6 layers used for feature extraction and only the weights of the last layer are adapted to re-train one or more layers with samples from the X-ray dataset [12]

Inception_V3

The network has a depth of 48 layers, incorporating several enhancements such as label smoothing, factorized convolutions, and the inclusion of an auxiliary classifier to propagate label information throughout the network. Pre-trained weights obtained from the ImageNet dataset are used to initialize the model. To preserve the learned low-level features, the weights of the first eight layers are frozen during the initial training phase [13].

Vgg16

The VGG network, a deeply neural network, was pre trained on an extensive dataset comprising more than a million images from the ImageNet database. This network comes with a total of 41 layers, involving 16 layers with changeable weights, 13 convolutional layers, and 3 fully connected layers. An adverse feature of the VGG16 Neural Network is its overwhelming number of trainable parameters, surpassing 134 million!!! To tackle this problem, we made the choice to freeze the initial 6 layers, thus restricting the trainable parameters to 57,000 .

Received: 2025-10-28
Accepted: 2026-01-11
Published Online: 2026-01-20
DOI: 10.70454/JRICST.2026.30103

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

Table1. Trainable Parameters

S.No	Model Name	Total Parameters	Trainable Parameters
1	CustomNet	504126	504126
2	DensNet121	6968206	6968206
3	ResNet50	23772110	264078
4	Inception	21814254	28686
5	Vgg16	134317902	57358

Model Training

We divided the training data into two sets: 80 percent training and 20 percent validation data, then trained the model for a maximum of 40 epochs. Most of the models performed best at 20 to 25 epochs. Table 2 shows that the DenseNet model achieved the highest training AUROC of 78 and the highest training accuracy of 87%.

Table2. Models with performance metrics

S.No	Model Name	AUROC	Accuracy
1	CustomNet	0.740989	0.862637
2	DensNet121	0.779881	0.866486
3	ResNet50	0.716902	0.854088
4	Inception	0.650021	0.846562
5	Vgg16	0.67147	0.848543

We used unseen test data to predict the multi classifications labels and hence evaluate and compare performances of various models. DenseNet121 performed the best overall based on test metrics. With a ROC score of 0.78 and accuracy of 87%, this model performed well. The accuracy and ROC values of the other models varied from 83% to 86% and 0.69 to 0.75 respectively

Table3. Models with performance metrics on test set

S.No	Model Name	AUROC	Accuracy
1	CustomNet	0.75376	0.863197
2	DenseNet121	0.783894	0.866112
3	ResNet50	0.730819	0.845245
4	Inception	0.69211	0.829782
5	Vgg16	0.716072	0.838025

AUROC (Area the Receiver-Operatingistic) values for different labels were being assessed to check the performance of each model. In the context of individual, Dense net 121showcaseda

Received: 2025-10-28
Accepted: 2026-01-11
Published Online: 2026-01-20
DOI: 10.70454/JRICST.2026.30103

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

better performance, as illustrated in the **Figure.** The model managed to achieve AUROC scores that ranged from 0.82 right upto 0.93 for the five labels that were included in the Stanford ML group study[0]. Additionally, quite notably, an extraordinary performance was observed for a select few labels such as Pleural Other(AUROC: 0.97) and Lung Opacity(AUROC: 0.91). Some labels showed average performance like Enlarged Cardio mediastinum, showcased AUROC of 0.49.

Table4. AUROC values of different models

Model Name Label	CustomNet	DenseNet121	Inception	ResNet50	Vgg16
Atelectasis	0.78474	0.824351	0.750568	0.75349	0.77857
Cardiomegaly	0.757884	0.834515	0.745836	0.72431	0.70712
Consolidation	0.841248	0.897332	0.672396	0.82662	0.81713
Edema	0.882775	0.883598	0.811875	0.83857	0.78895
Enlarged Cardiomediastinum	0.538789	0.485358	0.512073	0.56881	0.52095
Lung Lesion	0.04721	0.227468	0.085837	0.09871	0.28755
Lung Opacity	0.864051	0.911229	0.782775	0.82393	0.82136
No Finding	0.865602	0.850161	0.843582	0.85419	0.87218
Pleural Effusion	0.865135	0.925552	0.780409	0.83368	0.83332
Pleural Other	0.935622	0.969957	0.849785	0.88412	0.65665
Pneumonia	0.586836	0.682522	0.373341	0.60564	0.4812
Pneumothorax	0.63219	0.68031	0.681416	0.61117	0.83075
Support Devices	0.854588	0.882258	0.705276	0.74892	0.71565

Received: 2025-10-28
Accepted: 2026-01-11
Published Online: 2026-01-20
DOI: 10.70454/JRICST.2026.30103

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

Densenet121 gave the best accuracy for individual labels. Accuracy values for various labels. Fracture, Pneumothorax Pneumonia, Lung Lesion and Pleural Other achieved more than 95% accuracy.

Table5. Accuracies of different models

Model	**CustomNet**	**DenseNet121**	**Inception**	**ResNet50**	**Vgg16**
Atelectasis	0.700856	0.752137	0.65812	0.675214	0.65812
Cardiomegaly	0.24359	0.713675	0.709402	0.709402	0.7094
Consolidation	0.868974	0.858974	0.858974	0.858974	0.85897
Edema	0.863248	0.863248	0.803419	0.811966	0.81624
Enlarged Cardio mediastinum	0.534188	0.534188	0.534188	0.534188	0.53419
Fracture	1	1	1	1	1
Lung Lesion	0.995726	0.991453	0.995726	0.995726	0.99573
Lung Opacity	0.739816	0.782051	0.628205	0.75641	0.73932
No Finding	0.863248	0.807692	0.837607	0.858974	0.83761
Pleural Effusion	0.833333	0.871795	0.782051	0.773504	0.77778
Pleural Other	0.996726	0.995726	0.995726	0.995726	0.99573
Pneumonia	0.965812	0.965812	0.965812	0.965812	0.96581
Pneumothorax	0.965812	0.961538	0.965812	0.965812	0.96581
Support Devices	0.777778	0.769231	0.615385	0.679487	0.64103

DenseNet121 has performed relatively well as compared to the other models that we used for our work. However, to determine which model is definitely the best depends on cases. Multiple factors such as the overall size of your enormous dataset, available plentiful computational resources, and the specific, unique characteristics of the significant problem you are currently working on should be considered. It's quite often a very good and recommended practice to experiment with multiple various architectures and then to choose the one that significantly performs best through validation, testing on your specific, particular task. Additionally, pre-trained models on extremely large datasets (e.g.,ImageNet) can potentially provide a slightly good starting point and might possibly be fine-tuned for the specific task.

We contrasted our DenseNet121-based model's performance with that of other recent studies that used comparable datasets. Rajpurkar et al. (2018) used the CheXNet model (DenseNet121 fine-tuned on 224×224 CXRs) and obtained an average AUROC of 0.78 on the CheXpert dataset. Using transfer learning on the ChestX-ray14 dataset, Baltruschat et al. (2019) reported AUROC values ranging from 0.76 to 0.83 for different thoracic disease labels. Similarly, Nguyen-Mau et al. (2023) combined the EfficientNetV2 and ConvNeXt architectures with advanced augmentation and ensemble learning to achieve an AUROC of 0.81. Despite utilizing a simpler training pipeline and fewer computational resources, our DenseNet121 model demonstrated comparable diagnostic strength, achieving an overall AUROC of 0.78 and accuracy of 87%. This suggests that our framework continues to function well when clinical integration is used.

5.0 Conclusion

This work proposes an extensive assessment of deep convolutional neural networks (CNNs) using transfer learning for multi-label classification of chest X-rays automatically. The experiments performed using the CheXpert dataset confirm the fact that DenseNet121 far surpasses other state-of-the-art models, such as CustomNet, ResNet50, InceptionV3, and VGG16. DenseNet121 attained the highest diagnostic performance with an AUROC of 0.78 and accuracy of 87%, particularly in detecting key thoracic diseases like pleural effusion (AUROC 0.93) and lung opacity (AUROC 0.91). These results validate the excellent generalization ability of DenseNet models in learning discriminative features from intricate radiographic textures.

The problems of class imbalance and a lack of labelled medical data were effectively resolved by combining transfer learning with advanced data augmentation techniques, producing classification results that were more accurate and trustworthy. In addition to increasing diagnostic precision, the framework presented here allows for scalable implementation in actual healthcare systems with a shortage of radiologists.

Despite the positive outcomes, more work is required to improve clinical validation, domain adaptability, and model interpretability. In order to bring transparency and reliability to medical decision-making, future research must take into account explainable AI (XAI) techniques, uncertainty estimation, and attention-based mechanisms. Clinical workflows can also be accelerated by expanding this framework to real-time edge or cloud-based diagnostic systems.

Author Contributions

Authors contributed completely in all phases.

Funding
No funding was provided for this study.

Conflicts of Interest

Received: 2025-10-28
Accepted: 2026-01-11
Published Online: 2026-01-20
DOI: 10.70454/JRICST.2026.30103

Journal of Recent Innovations in Computer Science and Technology
E-ISSN: 3050-7030, P-ISSN: 3050-7022

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Data Set

CheXpert — A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison (Irvin et al., 2019).

References

[1] Imaging and Radiology: MedlinePlus Medical Encyclopedia. Available: https://medlineplus.gov/ency/article/007451.ht

[2] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund et al., "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," Proc. AAAI Conf. Artif. Intell., vol. 33, pp. 590–597, 2019.

[3] A. Nair, U. Nair, P. Negi, P. Shahane, and P. Mahajan, "Detection of diseases on chest X-ray using deep learning," Cikitusi J. Multidiscip. Res., vol. 6, no. 5, pp. 298–305, 2019.

[4] N. Wall, M. Palanisamy, and J. Santerre, "Automated pleural effusion detection on X-rays," SMU Data Sci. Rev., vol. 2, no. 2, p. 15, 2019.

[5] A. Madani, M. Moradi, A. Karargyris, and T. Syeda-Mahmood, "Chest X-ray generation and data augmentation for cardiovascular abnormality classification," in Proc. Med. Imag. 2018: Image Process., vol. 10574, pp. 415–420, 2018.\

[6] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 2097–2106.

[7] I. M. Baltruschat, H. Nickisch, M. Grass, T. Knopp, and A. Saalbach, "Comparison of deep learning approaches for multi-label chest X-ray classification," Sci. Rep., vol. 9, no. 1, p. 6381, 2019.

[8] "Deep transfer learning for the multilabel classification of chest X-ray images," Diagnostics, vol. 12, no. 6, p. 1457, 2022.

[9] T. H. Nguyen-Mau, T. L. Huynh, T. D. Le, H. D. Nguyen, and M. T. Tran, "Advanced augmentation and ensemble approaches for classifying long-tailed multi-label chest X-rays," 2023.

[10] L. Van Der Maaten and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 4700–4708.

[11] G. Holste, S. Wang, Z. Jiang, T. C. Shen, G. Shih, R. M. Summers, et al., "Long-tailed classification of thorax diseases on chest X-ray: A new benchmark study," in MICCAI Workshop on Data Augmentation, Labelling, and Imperfections, Cham: Springer Nature Switzerland, 2022, pp. 22–32.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 2818–2826.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. 25th Int. Conf. Neural Inf. Process. Syst. (NIPS), vol. 1, pp. 1097–110