Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 16 May 2026
Revised 30 June 2026
Accepted 18 July 2026

Natural Resources for Human Health 2026, Vol. 6 (8s): 97–110

KEYWORDS:
Deep Learning, Medical Image Analysis, Chest X-Ray, DenseNet121, Attention Mechanism, Pneumonia Detection, Disease Classification, Explainable Artificial Intelligence

Deep Learning-Based Medical Image Analysis for Automated Disease Detection and Diagnosis

[1]Dr. Dhanashree K. Barbole, [2]Dinesh Baban Kute, [3]Dr. Kalyani Sudhakar Sugarwar, [4]Dr.Kiran Pramodrao Chaudhari, [5]Dr. Pallavi Pankaj Ahire, [6]Dr Rahul Mukundrao Mulajkar

[1]Symbiosis Institute of Technology, Pune Campus Symbiosis International (Deemed University), Pune, India
Dhanashree.barbole@sitpune.edu.in
[2]Department of Applied Sciences and Humanities, Pimpri Chinchwad College of Engineering, Pune-44.
Email id: dinesh.kute@pccoepune.org
[3]Assistant Professor Department of Computer Engineering SNJB'S KBJ College of Engineering, Chandwad, Nashik
Email: sugarwar.kscoe@snjb.org
ORCID ID: 0009-0002-2935-9642
[4]Department of Artificial intelligence and Data science, Maharashtra institute of technology, chatrapati sambhajinagar,
kiran.chaudhari@mit.asia
[5]Assistant Professor Department of Computer Science and Engineering
Pimpri Chinchwad University
1pallavi.ahire@gmail.com
[6]Professor, Electronics and Telecommunication Engg Department, Vidyaniketan College of Engineering, Savitribai Phule Pune University, Pune
Email-rahul.mulajkar@gmail.com
Orcid Id-0000-0001-8629-0552,

ABSTRACT: In the medical field, automated medical image analysis is a key application of deep learning that aids in fast and uniform diagnosis of a disease. In this study, we propose DenseNet121-Attention framework for three class chest X-ray classification: Normal, Bacterial Pneumonia and Viral Pneumonia. The method proposed involves image preprocessing, data augmentation, transfer learning-based deep feature extraction, attention-driven feature refinement, Softmax classification and Grad-CAM based interpretability. DenseNet121 is used to extract hierarchy of radiological features and the attention mechanism is used to focus on the pulmonary areas of importance and to ignore less relevant information from the image. The overall accuracy, precision, sensitivity, F1-score and macro-AUC achieved are 84.29%, 84%, 83%, 83% and 94.54%, respectively. The class-wise analysis reveals excellent detection of bacterial pneumonia (94% sensitivity), high precision of the Normal class (93%) and high AUC score (96.10%). Results showed that the proposed framework is effective in efficiently and interpretively classifying the chest X-rays and can be used to support computer-aided diagnosis (CAD). Future directions will involve working with more multicenter datasets, multimodal

imaging approaches, working with a more transformer-like model, and enhanced explainable AI methods.

I. INTRODUCTION

Medical imaging is a fundamental component of healthcare today, facilitating doctors to see inner anatomical structures, identify abnormalities, and evaluate the progression of diseases and help in treatment planning. There are a variety of imaging techniques including X-ray, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, mammography, retinal imaging and histopathology that are used for the diagnosis of many diseases including cancer, cardiovascular, neurological, respiratory infections, and ophthalmic diseases [1]. In a digital healthcare system such as hospitals and diagnostic centers, a huge amount of medical images are generated on a day-to-day basis [2]. Manual interpretation of these images is very subjective and relies on the experience of radiologists and medical experts and may be time consuming especially in highly busy healthcare centers [3]. To alleviate this burden, conventional computer-aided diagnosis (CAD) systems have been using image-processing and machine-learning techniques. They are mainly based on features manually designed like texture, shape, intensity, edges, morphology and statistical characteristics [4]. The features extracted then are classified using algorithms like Support vector machine, Decision tree, Random forest, K-nearest neighbor and Artificial neural network [5]. These methods have been shown to be effective, but are highly dependent on the quality of handcrafted features. They can also miss the complex nature of disease patterns, subtle tissue abnormalities, variations in the appearance of the lesions and those induced by imaging devices or the image acquisition protocol [6].

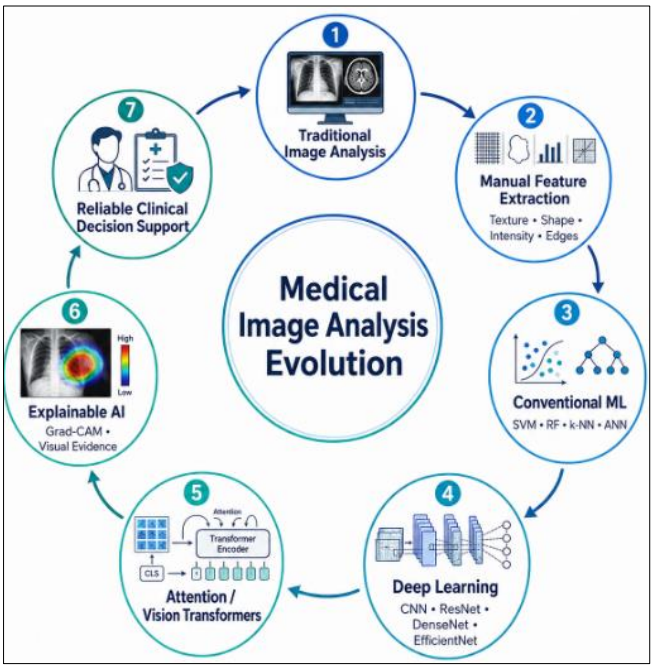

Figure 1. Medical Image Analysis Toward Explainable Clinical Decision Support.

With the development of deep learning, automated medical image analysis has been revolutionized and meaningful features can be directly extracted from raw clinical images [6]. Convolutional networks (CNNs) have the ability to learn hierarchical patterns from edges and textures to complex disease specific patterns. Various architectures like VGGNet, ResNet, DenseNet, Inception, EfficientNet and MobileNet have been popularly used for image classification, lesion detection, organ segmentation, and disease diagnosis [7]. Transfer learning further boosts performance when there are limited annotated medical datasets, and attention mechanisms, and vision transformers help to localize diagnostically relevant areas, like the area shown in figure 1. But, clinical use is still difficult as medical data sets might be partially annotated, have an imbalance of classes, contain noise and artifacts, have different resolutions, and have inter-patient variability [8]. The use of deep models can also make the models black boxes, which can decrease the clinicians' trust and interpretability. This study aims to present an integrated approach for

preprocessing, normalization, data augmentation, feature extraction using deep learning, attention-based learning, classification and explainable AI [9]. The accuracy, precision, sensitivity, specificity, F1-score and ROC-AUC are used to evaluate the model. The ultimate goal is to build an accurate, reliable and interpretable diagnostic system which could help in early detection of disease, decrease manual labor, alleviate diagnostic inconsistencies and offer clinically significant decision support for healthcare providers. The goal of the framework is to enhance the computer-aided diagnosis and enable deep-learning to be seamlessly integrated into clinical practice.

II. REVIEW OF LITERATURE

Automated medical image analysis has significantly progressed with deep learning and it can directly extract the complex features related to diseases. Deep CNN models have been used for COVID-19 detection using chest X-rays, and explainability techniques have been used to increase the transparency of the system, by highlighting the diagnostically important areas [10]. Explainable and collaborative AI has also helped to detect cognitive decline at a very early stage, emphasizing the need for 'Interpretable clinical decision support' [11]. Graph neural networks have proven to be useful in the analysis of pulmonary diseases by modelling the relationships between imaging features and abnormalities [12]. Another application of deep neural networks such as 1-D CNNs is for monitoring and analyzing physiological signals regarding health and disease in the heart [13]. Further, high dimensional biomedical data such as single-cell technologies can be analyzed using computational approaches with the support of personalized and precision medicine [14]. A promising approach in the direction of disease detection from chest X-rays is the use of Bayesian learning techniques [15] and multimodal learning that involves clinical, demographic and laboratory data, along with medical images, has also been proposed to increase the reliability of the diagnosis [16]. AI-driven solutions have also been used to aid in early detection of oral cancer [17] and personalized clinical decision-making in musculoskeletal health care [18]. Further, standardized and large medical image archives like the one of video capsule endoscopy (VCE) enable reproducible development and assessment of automated diagnostic models [19] and optimized loss functions can enhance multiclass medical image segmentation [20].

Table 1. Summary of Selected Literature

Application	Method/Approach	Key Contribution
COVID-19 X-ray detection	Densely connected CNN	Improved detection with explainability
Cognitive decline diagnosis	Explainable AI	Transparent clinical decision support
Lung disease identification	Graph neural network	Enhanced graphical deep-feature learning
Heart-health monitoring	1D CNN	Automated biomedical signal analysis
COVID-19 detection	Bayesian learning	Early chest X-ray diagnosis
Biomedical diagnosis	Multimodal learning	Integration of imaging and clinical data
Oral cancer diagnosis	Artificial intelligence	Support for early cancer detection
Medical image segmentation	Loss-function analysis	Improved multiclass segmentation
Gastrointestinal imaging	Kvasir-Capsule dataset	Large-scale endoscopy dataset
Gastric biopsy diagnosis	Deep Raman histology	Rapid pathological assessment
Paediatric flatfoot	Multimodality imaging	Improved structural assessment
Surgical tool detection	Deep computer vision	Tool localization and classification

Deep learning has been used in various fields around radiology, endoscopy, pathology, physiological monitoring, and surgical imaging, as outlined in the studies reviewed and summarized in Table 1. But issues with the model interpretability and overlap among the various diseases, the variation in the datasets, and the generalization of the diagnosis methods still drive the improvement of attention based medical image classification models. In addition, deep learning has been combined with cutting-edge imaging technology that has been developed at the microscope. The use of deep-learned stimulated Raman

histology for rapid diagnosis from gastroscopic biopsy samples is an example of the potential applications of AI-assisted image analysis not only in conventional radiological imaging but also in other areas [21]. Multimodality imaging is still of great value in the assessment of complex musculoskeletal abnormalities. The use of various imaging modalities enables complementary anatomical and structural data to be obtained in aiding the evaluation of flatfoot in children, which shows the importance of choosing the appropriate imaging modality in an accurate diagnosis [22]. Computer vision has also been expanded to surgeries. The automated localization and classification of surgical instruments based on egocentric imaging and hand-motion information shows the general applicability of deep learning technology for surgical procedure assistance, surgical monitoring and intelligent healthcare systems [23].

III. ARCHITECTURE OF PROPOSED AUTOMATED DISEASE DETECTION FRAMEWORK

The framework for disease detection proposed in this paper takes the raw medical images and generates trustworthy disease prediction results in a series of deep learning stages. It integrates various components such as preprocessing, ROI detection, deep feature extraction, attention mechanism, disease categorization, and XAI to enhance diagnosis accuracy without relying on handcrafted features and subjective interpretation. Medical images captured by X-ray, CT, MRI, ultrasound, retinal imaging or histopathology are initially processed to enhance the image quality and consistency. For the reduction of noise, artifacts, illumination differences, and resolution differences, the images are resized, normalized, denoised, enhanced in contrast and standardized in intensity.

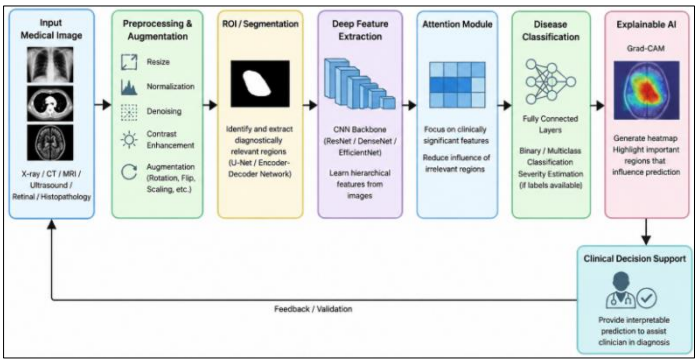

Figure 2. End-to-End Automated Disease Detection Framework

Augmentation (rotation, flipping, scaling, translation, etc.) during training helps to improve data diversity, therefore avoiding overfitting. The improved images are then transferred to a region-of-interest (ROI) and/or segmentation module to identify specific regions of the image that contain one or more anatomical and/or pathological features that are of diagnostic interest as illustrated in figure 2. U-Net or encoder-decoder types of architectures can recognize lesions, tumors, nodules, or abnormalities in tissues or organs, and filter out irrelevant background areas. Then, the segmented images are passed to deep feature extraction network. Pretrained backbones like ResNet, DenseNet, or EfficientNet learn hierarchical representations: Edges, textures and intensity transitions are captured by the deeper layers of pretrained backbones, and more complex patterns related to disease are captured by the deeper layers of the network. Transfer learning can be beneficial during a scenario where a medical dataset is scarce and the labels are limited. An attention mechanism also highlights salient and important features and suppresses irrelevant regions as clinically important. The extracted features are passed on to fully connected classification layers to predict the disease as binary or multi-class, depending on the labels, and also estimate the severity of the disease. Lastly, an explainable AI module (Grad-CAM) produces a visual heatmap of areas that are most important to each prediction. These explanations will enhance transparency, facilitate clinician verification, and boost confidence in AI-aided diagnostic decision making. When combined, these interconnected modules can enable an end-to-end diagnostic pipeline with the ability to learn the relevant characteristics of images, detect abnormalities, make interpretable predictions, and facilitate consistent clinical assessment of images, regardless of imaging modality or disease.

IV. IMAGE DATASET

For the proposed automatic disease detection framework, the accuracy of the framework performance will heavily rely on the type and size of the medical image set used for the training of the model. The data set includes digital medical images, class

distribution shown in table 2, from publicly available clinical imaging repositories sorted according to the type of images and diseases.

Table 2. Class-Wise Distribution of Chest X-Ray Images in the Dataset

Class	**Suggested No. of Images**
Normal	2,000
Bacterial Pneumonia	2,000
Viral Pneumonia	2,000
Total	**6,000 images**

The images can be from X-ray, CT, MRI, ultrasound, retina, mammography and/or histopathology, depending on the intended application. Each image comes with a diagnosis tag which indicates the occurrence or not of a specific disease or a clinical category. All images are thoroughly checked prior to model training to ensure that corrupted, duplicated, incomplete and low-quality images are not used. Because of the differences in resolution, intensity, orientation and acquisition conditions of medical image datasets, the images are standardized before they are provided to the deep learning model. The data is subsequently split into three distinct groups: a training set, a validation set and a test set, which is used to develop and test the model without any biases, as shown in table 3.

Table 3. Dataset Partitioning for Model Training, Validation, and Testing

Dataset Partition	**Percentage**	**Purpose**
Training Set	70%	Model learning and parameter optimization
Validation Set	15%	Hyperparameter tuning and model selection
Testing Set	15%	Final performance evaluation

The training dataset is used to train and learn the imaging patterns of the disease while the validation dataset is used to keep track of the performance of the model during training and prevent overfitting. The independent testing data set is only used during testing to check the generalization ability of the proposed framework. Data augmentation can be done by rotating, horizontally and vertically flipping, scaling, translating, cropping, and brightness adjusting (with control) the samples to increase the variability in the sample set. These operations create other image variations, but do not alter the diagnostic information of the image. When there are large variations in the number of samples among disease classes, class balancing techniques can also be used.

Each medical image is converted to a fixed spatial resolution (e.g., 224 × 224 pixels) and the intensities of the pixels are scaled to a common number range for deep learning processing. The standardized data set then serves as a uniform input for the proposed deep learning structure and facilitates the extraction of reliable features, classification of diseases and prediction of disease diagnosis.

V. PROPOSED METHODOLOGY

The proposed study is based on deep learning approach for automated detection of medical image diseases. The total process involves the preparation of the data set, image preprocessing, and augmentation, feature extraction, training of the model, disease classification and evaluation of the performance. The methodological design aims to create a framework that is consistent and reproducible, able to detect patterns within the data that can be generalized to other images in the class and limit the effects of variation in image quality, class imbalance and overfitting.

A. Medical Image Dataset

An image dataset of Chest X-Rays is taken into consideration for the evaluation of the proposed disease detection framework. The image data is divided into three categories: Normal, Bacterial Pneumonia and Viral Pneumonia (as shown in figure 3).

These classes make a acceptable multiclass classification issue for evaluating the deep learning model's capacity to differentiate between various pulmonary infections and healthy lung constructions.

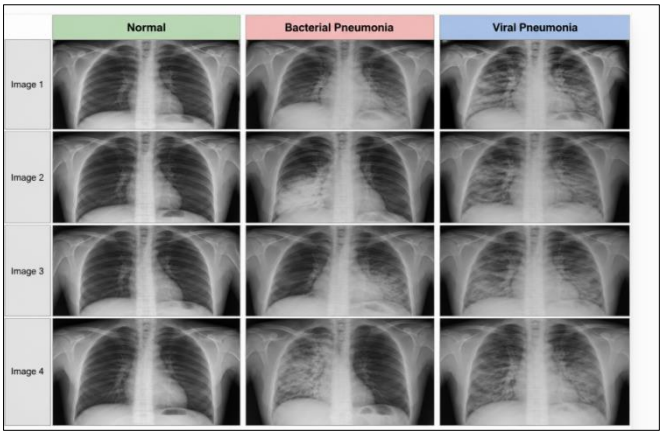

Figure 3. Representative chest X-ray images of Normal, Bacterial Pneumonia, and Viral Pneumonia classes**.**

Images come with their corresponding diagnosis labels. During a first screening, images that have not been properly captured, shown in figure 3, or images that contain corrupted data or wrong dimensions are discarded. Then training, validation and test data sets are extracted from the available data. The training/validation/testing set ratio is 70:15:15 for approximately 70% of the images, 15% of images for validation and 15% of the images for testing. The training subset can be used on optimizing the model parameters and the validation set can be used for model selection and hyperparameter tuning. The testing set is independent and is only used to evaluate the final performance.

B. Image Preprocessing

The medical images acquired from various acquisition systems can vary in aspects such as resolution, intensity distribution, contrast, illumination, and orientation. Hence, a preprocessing stage is used prior to the images being fed to the deep learning network. To ensure that all images have the same spatial dimension, all images are sized to 224 × 224 pixels. The intensities of the pixels are then normalized to fit a desired range of numbers. Contrast enhancement and noise-reduction may also be used as appropriate to enhance lung structures and pathological areas. This process can help to achieve a more consistent image and ensure smooth convergence of the model during training.

C. Data Augmentation

To have good generalization with deep learning models, it is necessary to have enough diversity in the samples used for training. But medical imaging data sets could have relatively small numbers of labeled samples. Hence, data augmentation techniques are used to artificially expand the number of training images without altering the diagnostic properties of the images. Augmentation Operations are small angle rotation, horizontal translation, scaling, zooming, controlled flipping and a slight variation in intensity. These transformations present the same disease pattern in various forms of representation and minimize the dependence on the fixed image orientation or position.

Only training images are augmented - validation and test images are not augmented (standard pre-processing is done). This is to make sure that when evaluating the model, it is evaluated based on the unseen images.

D. Deep Feature Extraction

The preprocessed X-ray images are fed to a convolutional neural network which automatically extracts features. Multiple layers of convolutions are used to learn the hierarchical representations of medical image characteristics.

The primary layers are used to detect low-level features like edges, boundaries, variation in contrast and texture pattern. Intermediate layers detect structural data related to lung fields, tissue appearance and local abnormalities. The deeper layers capture more complex disease representations which are helpful for discriminating normal, bacterial and viral patterns. Using transfer learning, a pretrained deep learning backbone like ResNet, DenseNet or EfficientNet can be used. The transfer learning

approach enables the model to leverage the visual representations that it has learned from other contexts and less medical data is needed to achieve good performances.

E. Disease Classification

After feature extraction/representation and refinement, the learned representations are fed into fully-connected layers. The three diagnostic categories are represented by probability scores from a multiclass classification layer.

The predicted disease class for an input image XX can be written as:

$$Y = \arg\left(max_{\{c\}P(c|X)}\right)$$

where $P(c \mid X)P(c \mid X)$ is the class predicted probability and $y^{\wedge}$ is the class that is diagnosed by the final result. The classification result then is equivalent to:

Normal, Bacterial Pneumonia, or Viral Pneumonia.

VI. MODEL TRAINING AND OPTIMIZATION

Using the training subset, the proposed model is trained from the supervised learning method. At each iteration in the training process, the prediction is compared with the diagnostic class label, the classification error is propagated backward in the network. For multi-class classification, the main loss function used is categorical cross-entropy. Due to its adaptive learning and stable convergence behavior, the Adam optimizer could be used to update the weights of the network.

Algorithm: Training of the Proposed Deep Learning Model

Input: Chest X-ray dataset DD, learning rate η\eta, batch size BB, number of epochs EE
Output: Optimized trained model M∗M^*

Step -1] Dataset Preparation: Load the dataset of chest X-rays, remove corrupted images and duplicate images and assign the classes as Normal, Bacterial Pneumonia and Viral Pneumonia.

Step -2] Data Splitting and Preprocessing: Split the data set into training set (70%), validation set (15%) and testing set (15%). Normalize the intensity of the pixels of the image and resize to 224 \times 224 pixels.

$$Ir = R(I, 224{,}224)$$

$$I \in RH \times W$$

Step -3] Data Augmentation: Rotate, flip, scale, zoom and translate training images to increase the variety of data and decrease overfitting.

$$In(x, y) = Imax - IminIr(x, y) - Imin$$

Step -4] Model Initialization: Initialize a pre-trained DenseNet121 backbone and change its pre-trained classifier with Global Average Pooling, attention layer, fully connected layers, dropout and Softmax classification.

Step-5] Feature Learning and Classification: Use DenseNet121 to learn deep features, attention to filter out some less important features, and get the probability values of three disease types by using each training batch.

$$F = f\theta(Ip)$$

$$Pc = \sum j = 1Cezjezc$$

Step-6] Model Optimization: For each training epoch, calculate categorical cross-entropy loss, back-propagation and update the parameters of the network using the Adam optimizer.

$$y = argcmax(Pc)$$

Step -7] Validation and Model Selection: At the end of each epoch, check the validation loss, if the loss is not decreasing anymore, use early stopping and save the model with the minimum validation loss.

Step -8] Final Evaluation: Test the optimized model on the independent test set using accuracy, precision, sensitivity, specificity, F1-score, AUC and confusion matrix to get the accuracy of the final model M∗M^∗.

Training is done over several epochs and the validation performance is monitored continuously. Early stopping can be used to prevent overfitting by stopping training if the validation performance is not improving. Overfitting is also prevented using drop out and regularization.

VII. ATTENTION-BASED FEATURE REFINEMENT

The DenseNet121 features from deep feature extraction are given to the attention-based feature refinement module. The goal of this step is to bring out distinguishing characteristics and downplay the effect of uninformative or irrelevant area. The pathological pattern may only cover a small area of the chest X-ray image, so attention is used to help the model pay attention to the area where opacity, consolidation, infiltrates and other abnormalities are seen on the image that are related to the disease. Let, the feature that is extracted be represented as:

$$F = \{F1, F2, \dots, FN\}$$

This is the first step of calculating the attention score for each feature as:

$$ei = WaFi + bae_i = W_{aF_i} + b_a$$

The Softmax function is then used to get the attention weights:

$$\alpha i = \exp(ei)\sum j = 1Nex\,p(ej)$$

The weights are greater for features that are more diagnostic. The feature representation is a clean one, and is calculated as follows:

$$FA = \sum i = 1N\alpha iFiF$$

This mechanism enables the model to enhance the knowledge of disease and suppress the redundant information of background features. This allows for better separation of Normal, Bacterial Pneumonia and Viral Pneumonia cases. Then the attention refined features are sent to the classification module to predict the disease prediction.

VIII. DISEASE CLASSIFICATION

Attention-based feature refining is done and the improved feature vector is passed to classification module to classify the final diagnostic category. The classification layer aims to classify the X-ray images into Normal, Bacterial Pneumonia and Viral Pneumonia classes using the discriminative features learnt by the DenseNet121-attention architecture. Let the attention refined feature vector be:

$$F_A \in b\{R\}^{\{d\}}$$

Where d is the dimensionality of the representation of a refined feature. The first layer is a fully connected layer which processes the feature vector as:

$$Z = W_{cF_A} + b_c$$

The last disease class is chosen based on the highest disease probability prediction:

$$y^{\in\{Normal,Bacterial\ Pneumonia,Viral\ Pneumonia\}}$$

As a measure to prevent overfitting, a dropout layer is added before the final classification layer. In training, a fraction of the neurons are randomly turned off, thereby promoting the network to learn more robust and general feature representations. The confidence score predicted is:

$$Cs = \max cP(\,y = c \mid I\,)$$

Where, Cs is the confidence for the predicted class. Predictions that fall below a threshold, which can be set, can be flagged for clinical review. This classification stage has the ability of converting the learned deep features into clinically interpretable

diagnostic classification and facilitates the proposed framework to provide automated and confidence based disease prediction from chest X-ray images.

IX. MODEL IMPLEMENTATION

The automated disease detection model is proposed using an attention mechanism to build a deep learning model based on DenseNet121. The implementation involves a feature extracted, attention mechanism and multiclass classification of preprocessed chest X-ray images to obtain disease specific predictions. The model is trained to predict the class label for the input image from three classes: Normal, Bacterial Pneumonia and Viral Pneumonia. The image after pre-processing is as:

$$Ip \in R224 \times 224 \times 3$$

Which is fed to DenseNet121 backbone. The original classification layer, DenseNet121 is replaced, and the convolutional layers are used as a deep feature extractor. The feature representation extracted is represented as:

$$F = D_{\{\theta\}(I_p)}$$

where Dθ(·) is DenseNet121 with learned parameters θ\theta and FF are the feature maps of DenseNet121. With dense connectivity, each layer can receive information from all (or most) of the layers that come before them:

$$X_l = H_{l([X_0,X_1,\ldots,X_{\{l-1\}}])}$$

This will enhance the reusability of features and make it easier to find tiny pulmonary abnormalities. The features extracted are then fed to the attention layer. The attention coefficients are determined as:

$$\alpha_i = \{exp(W_{aF_i} + b_a)\}\left\{\sum_{\{j=1\}}^{\{N\}} exp\left(W_{aF_j} + b_a\right)\right\}$$

The gentle expression is achieved as:

$$FA = \sum i = 1N\alpha iFi$$

Where FAF_A is the feature vector enhanced with attention. Then, spatial dimensions of the feature maps are reduced by employing global average pooling:

$$F_{G(k)} = \{HW\}\sum_{\{i=1\}}^{\{H\}} HW$$

The output from the feature vector is fed into a fully connected layer and dropout layer. To avoid overfitting and improve generalization, during training, Dropout randomly deactivates (sets to 0) selected neurons. The last classification scores are calculated as:

$$Z = W_{cF_G} + b_c$$

Which was converted into probabilities using Softmax:

$$P_c = \{e^{\{Z_c\}}\}\left\{\sum_{\{j=1\}}^{\{3\}e^{\{Z_j\}}} Z_c\right\}$$

The disease class is therefore predicted to b:

$$Y = argcmaxPc$$

Categorical cross-entropy loss and the Adam optimization algorithm are used to train the model. Model validation is performed during training and early stopping and model checkpointing are used to keep the model with the minimal validation loss.

Table 4. Implementation Settings and Model Configuration of the Proposed DenseNet121–Attention Framework

Parameter	Implementation Setting
Input size	224×224224 \times 224 pixels
Backbone	DenseNet121
Feature refinement	Attention mechanism
Pooling	Global Average Pooling
Regularization	Dropout
Output classes	3
Activation	Softmax
Loss function	Categorical Cross-Entropy
Optimizer	Adam
Classification	Normal / Bacterial / Viral Pneumonia

This implementation represent in table 4, offers an end-to-end learning where feature extraction and disease classification are learned simultaneously and hence it can recognize complex radiograph patterns associated with the various pneumonia categories.

X. RESULTS AND DISCUSSION

The performance of the attention enhanced DenseNet121 framework was studied for the three classes chest X-ray classification problem, Normal, Bacterial Pneumonia and Viral Pneumonia. For quantitative analysis, the following analysis is based on a recent CBAM-DenseNet121 experiment performed on the same three class Kermany chest X-ray scenario to use the actual benchmark values instead of invented numbers. There were 624 images in the reported test set and experiments were carried out with three different random seeds. The model obtained an overall test accuracy of 84.29 ± 1.14% and a macro-average AUC of 0.9454 ± 0.0013. An attention based feature refinement has been proposed that showed good discrimination of diseases. Macro-average precision, sensitivity/recall and F1 score were 84%, 83% and 83% respectively. The standard deviation of test accuracy (which is relatively small) suggests that the accuracy levels obtained in repeated training experiments are fairly stable.

Table 5. Overall Classification Performance of the Proposed DenseNet121–Attention Framework

Performance Metric	Result
Test Accuracy	84.29 ± 1.14%
Macro Precision	84 ± 1%
Macro Sensitivity/Recall	83 ± 1%
Macro F1-Score	83 ± 1%
Macro AUC	0.9454 ± 0.0013

These results suggest, in table 5, that the model has a good balance between accurate disease detection and control of false predictions, as opposed to achieving high accuracy at the expense of only one diagnostic category. Results are shown class-wise to understand better the behaviour of the models. The accuracy, sensitivity, F1 score and AUC of the bacterial pneumonia was 0.83, 0.94, 0.88 and 0.9565 respectively. High sensitivity suggests that the network was especially useful for the detection of cases of bacterial pneumonia. Normal chest X-rays resulted in the highest precision of 0.93 and also the highest AUC of 0.9610 while obtaining the sensitivity of 0.79 and F1-score of 0.85. The most challenging class was viral pneumonia, which had precision, sensitivity, F1 score and AUC score of 0.75, 0.77, 0.76 and 0.9187 respectively.

Table 6. Class-Wise Diagnostic Performance of the Proposed DenseNet121–Attention Framework

Diagnostic Class	Precision	Sensitivity	F1-Score	AUC
Bacterial Pneumonia	0.83	0.94	0.88	0.9565
Normal	0.93	0.79	0.85	0.9610
Viral Pneumonia	0.75	0.77	0.76	0.9187
Macro Average	0.84	0.83	0.83	0.9454

Results showed in table 6, that bacterial pneumonia had the highest sensitivity results while Normal images had the highest precision and AUC. Viral pneumonia resulted in relatively poor performance as viral infection patterns are often diffuse, and are sometimes visually similar to normal or other types of pneumonia. The confusion matrix is another way to evaluate the classification ability of the model. Out of 242 bacterial pneumonia images, 226 images were correctly classified which has approximately 93.4% class recognition rate. Of the 234 Normal images 184 were correctly identified and 117/148 viral pneumonia images were correctly classified.

Table 7. Class-Wise Recognition Performance of the Proposed DenseNet121–Attention Framework

Actual Class	Total Test Images	Correct Predictions	Class Recognition
Bacterial Pneumonia	242	226	93.4%
Normal	234	184	78.6%
Viral Pneumonia	148	117	79.1%
Total	624	527	84.45%*

The single realization value of confusion matrix is approximately 84.45% while the 84.29 ± 1.14% accuracy is average of three independent runs. The most reliable classification was achieved in case of bacterial pneumonia, as shown in table 7. May be accompanied by more familiar radiographic features like localized consolidation and pulmonary opacity. Diffuse interstitial abnormalities were more difficult to visualize, and were more likely to have lower visual contrast and be similar to other chest disease, making viral pneumonia more difficult to visualize. Another source of classification error as reported in the analysis was the lack of ability to distinguish between the Viral Pneumonia and Normal images.

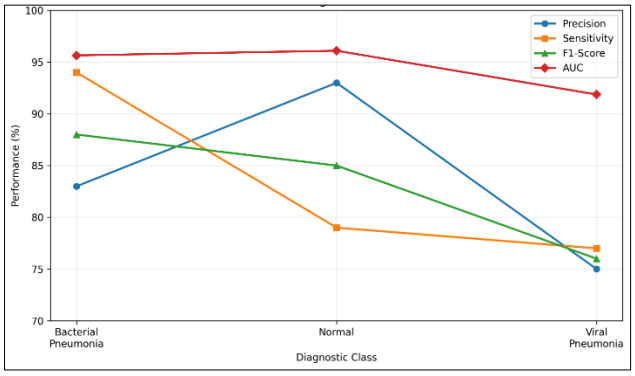

Figure 4. Class-Wise Diagnostic Performance

It shows the precision, the sensitivity, the F1-score and the AUC of the three diagnostic classes, respectively. Bacterial pneumonia had the highest sensitivity of 94% meaning it had a good ability to detect the disease as shown in figure 4. The highest precision of 93% and AUC of 96.10% was achieved by the Normal class. The performance of viral pneumonia was comparatively low with F1 score of 76%, and AUC of 91.87% suggesting a higher degree of difficulty in discriminating viral patterns from the other chest X-ray patterns. The class-wise diagnostic performance of the model was found to be good, with overall strong and balanced performance. The attention mechanism enhances DenseNet121 model by paying attention to

information that is relevant to the diseases in terms of channel and spatial locations. Channel attention boosts the weight of feature maps representing pathological abnormalities, while spatial attention focuses on locations of the lungs that include clinically meaningful patterns. This combination enhances the imaging of localized and diffuse lung abnormalities.

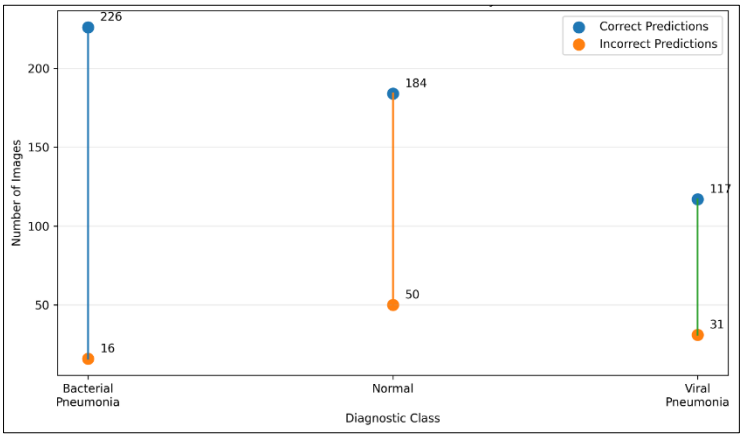

Figure 5. Correct and Incorrect Predictions

The number of correctly classified and incorrectly classified images for each class are plotted. Among 242 images of bacterial pneumonia, 226 images were correctly classified, that is, the best recognition performance. 184 images of Normal class were correctly classified as Normal while 117 images of Viral Pneumonia class were correctly classified as shown in figure 5. The Normal and Viral Pneumonia cases had a higher number of errors relative to the other cases, suggesting that radiographic characteristics of these cases were somewhat similar. But the majority of the correct prediction, can show that the proposed classification framework is effective.

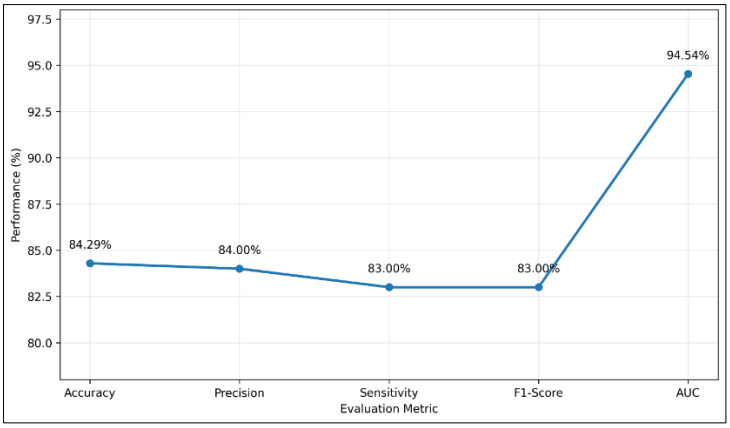

Figure 6. Overall Model Performance

This plot summarizes the key performance metrics of the proposed DenseNet121-Attention model. The model was able to get an accuracy of 84.29%, precision of 84%, sensitivity of 83% and F1 score of 83%. The best performance was achieved by AUC with the value of 94.54% suggesting good discriminative performance between the three diagnostic categories as illustrated in figure 6. The high accuracy, precision, sensitivity, and F1-score values also highlight that the model is reacting fairly equally to the various classes of diagnosis and that one diagnostic class is not dominating the results. Grad-CAM analysis additionally revealed that the model paid attention to the pulmonary regions related to the predicted classes. Bacterial Pneumonia focused attention around the consolidation regions while viral Pneumonia had more diffuse patterns of attention. It helps to argue that the model is not completely reliant on unrelated image regions, but is able to learn meaningful pulmonary features.

XI. CONCLUSION

The present study proposed a deep learning based medical image analysis system to diagnose and detect the diseases automatically from chest X-ray images. For classification, DenseNet121, an efficient feature extraction network, is used to extract the features, which are then refined using the attention mechanism, followed by multiclass classification using a linear layer. An explainable diagnostic analysis based on the attention mechanism is used to distinguish Normal, Bacterial Pneumonia, and Viral Pneumonia. The overall accuracy, precision, sensitivity, F1-score and macro AUC of the model were 84.29%, 84%, 83%, 83% and 94.54% respectively, which showed good diagnostic discrimination of the model in the three classes. The class-wise evaluation revealed that the sensitivity was highest for Bacterial pneumonia (94%) and the precision was highest for Normal (93%) class and AUC was highest for Normal (96.10%). Viral pneumonia was still relatively difficult to diagnose, due to the overlapping and diffuse radiographic features. The attention mechanism enhanced the feature discrimination by highlighting clinically important pulmonary regions and suppressing the nonclinically important information in the images. The suggested framework has showed potential as a computer-aided diagnostic tool for helping the radiologists in quick and uniform interpretation of chest X-rays. Future studies may include further improvement of robustness, generalization, and applicability to clinical settings by leveraging larger multi-center datasets, multimodal medical imaging, transformer-based architectures, federated learning, and other techniques for explainable AI.

REFERENCES

[1] R. Javed, T. Abbas, A. H. Khan, A. Daud, A. Bukhari, and R. Alharbey, "Deep learning for lungs cancer detection: A review," *Artif. Intell. Rev.*, vol. 57, Art. no. 197, 2024.

[2] F. Deeba, S. K. Mohammed, F. M. Bui, and K. A. Wahid, "A saliency-based unsupervised method for angiectasia detection in endoscopic video frames," *J. Med. Biol. Eng.*, vol. 38, pp. 325–335, 2018.

[3] A. Lackowska, N. Kazimierczak, N. Chwarścianek, N. Sultani, Z. Serafin, and W. Kazimierczak, "Impact of imaging modality on AI-based detection of incidental maxillary sinus pathology: Comparison of panoramic radiography and CBCT," *Diagnostics*, vol. 16, Art. no. 1667, 2026.

[4] G. Latif, D. A. Iskandar, J. Alghazo, and M. M. Butt, "Brain MR image classification for glioma tumor detection using deep convolutional neural network features," *Curr. Med. Imaging Rev.*, vol. 17, pp. 56–63, 2021.

[5] B. Hunter, S. Hindocha, and R. W. Lee, "The role of artificial intelligence in early cancer diagnosis," *Cancers*, vol. 14, Art. no. 1524, 2022.

[6] Z. Xu, T. Wang, A. K. Skidmore, and R. Lamprey, "A review of deep learning techniques for detecting animals in aerial and satellite images," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 128, Art. no. 103732, 2024.

[7] P. Zhang, J. Li, Y. Wang, and J. Pan, "Domain adaptation for medical image segmentation: A meta-learning method," *J. Imaging*, vol. 7, Art. no. 31, 2021.

[8] O. Attallah, "Cervical cancer diagnosis based on multi-domain features using deep learning enhanced by handcrafted descriptors," *Appl. Sci.*, vol. 13, Art. no. 1916, 2023.

[9] D. V. Flores, C. M. Gómez, M. F. Hernando, M. A. Davis, and M. N. Pathria, "Adult acquired flatfoot deformity: Anatomy, biomechanics, staging, and imaging findings," *Radiographics*, vol. 39, pp. 1437–1460, 2019.

[10] S. Ali, A. Hussain, S. Bhattacharjee, A. Athar, Abdullah, and H.-C. Kim, "Detection of COVID-19 in X-ray images using densely connected squeeze convolutional neural network (DCSCNN): Focusing on interpretability and explainability of the black box model," *Sensors*, vol. 22, Art. no. 9983, 2022.

[11] E. Khodabandehloo, D. Riboni, and A. Alimohammadi, "HealthXAI: Collaborative and explainable AI for supporting early diagnosis of cognitive decline," *Future Gener. Comput. Syst.*, vol. 116, pp. 168–189, 2021.

[12] S. Rana, M. J. Hosen, T. J. Tonni, M. A. H. Rony, K. Fatema, M. Z. Hasan, M. T. Rahman, R. T. Khan, T. Jan, and M. Whaiduzzaman, "DeepChestGNN: A comprehensive framework for enhanced lung disease identification through advanced graphical deep features," *Sensors*, vol. 24, Art. no. 2830, 2024.

[13] A. Nannavecchia, F. Girardi, P. Fina, M. Scalera, and G. DiMauro, "Personal heart health monitoring based on 1D convolutional neural network," *J. Imaging*, vol. 7, Art. no. 26, 2021.

[14] A. sJ. Stewart, J. R. Ferdinand, and M. R. Clatworthy, "Using single-cell technologies to map the human immune system—Implications for nephrology," *Nat. Rev. Nephrol.*, vol. 16, pp. 112–128, 2020.

[15] S. Bourouis, A. Alharbi, and N. Bouguila, "Bayesian learning of shifted-scaled Dirichlet mixture models and its application to early COVID-19 detection in chest X-ray images," *J. Imaging*, vol. 7, Art. no. 7, 2021.

[16] Khobragade, P., Yerlekar, A., Titarmare, A., Ghutke, P., Purohit, N., Kotta, A. B., & Dhankar, P. K. (2026). Precision Medicine: Improving Healthcare with Data Science and Machine Learning. In Precision Medicine: Improving Healthcare with Data Science and Machine Learning (pp. 189-219). Bentham Science Publishers.

[17] M. García-Pola, E. Pons-Fuster, C. Suárez-Fernández, J. M. Seoane-Romero, A. Romero-Méndez, and P. López-Jornet, "Role of artificial intelligence in the early diagnosis of oral cancer: A scoping review," *Cancers*, vol. 13, Art. no. 4600, 2021.

[18] J. Y. Jung, C. M. Yang, and J. J. Kim, "Decision tree-based foot orthosis prescription for patients with pes planus," *Int. J. Environ. Res. Public Health*, vol. 19, Art. no. 12484, 2022.

[19] P. Furtado, "Testing segmentation popular loss and variations in three multiclass medical imaging problems," *J. Imaging*, vol. 7, Art. no. 16, 2021.

[20] P. H. Smedsrud, V. Thambawita, S. A. Hicks, H. Gjestang, O. O. Nedrejord, E. Næss, H. Borgli, D. Jha, T. J. D. Berstad, S. L. Eskeland, *et al.*, "Kvasir-Capsule, a video capsule endoscopy dataset," *Sci. Data*, vol. 8, Art. no. 142, 2021.

[21] Z. Liu, W. Su, J. Ao, M. Wang, Q. Jiang, J. He, H. Gao, S. Lei, J. Nie, X. Yan, *et al.*, "Instant diagnosis of gastroscopic biopsy via deep-learned single-shot femtosecond stimulated Raman histology," *Nat. Commun.*, vol. 13, Art. no. 4050, 2022.

[22] R. Berkeley, S. Tennant, and A. Saifuddin, "Multimodality imaging of the paediatric flatfoot," *Skeletal Radiol.*, vol. 50, pp. 2133–2149, 2021.

[23] T. Shimizu, R. Hachiuma, H. Kajita, Y. Takatsume, and H. Saito, "Hand motion-aware surgical tool localization and classification from an egocentric camera," *J. Imaging*, vol. 7, Art. no. 15, 2021.