Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A novel predictive model decision system for the detection and classification of congenital heart disease and respiratory diseases from chest X-ray images based on IoT and cloud

Madhumita Pal
Government College of Engineering
Subhankit prusti
Odisha University of Technology and Research
Rajiv lochan Bal
Orissa School of Mining EngineeringDepartment of Mining EngineeringDepartment of Mining Engineering
Sujata Dash
Sujata238dash@gmail.com
Nagaland University
Ganapati Panda
C.V. Raman Global University
Zhen Wang
Westlake University School of Medicine
Haipeng Liu
Coventry University

Article

Keywords: Congenital heart disease, accuracy, precision, recall, F1-score, Vision transformer, IoT

Posted Date: May 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-7597193/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A novel predictive model decision system for the detection and classification of congenital heart disease and respiratory diseases from chest X-ray images based on IoT and cloud

Madhumita Pal[1], Subhankit prusti[2], Rajiv lochan Bal[3], Sujata Dash[4*], Ganapati Panda[5], Zhen Wang[6*], Haipeng Liu[7]

[1]Department of Electrical Engineering, Government College of Engineering, Keonjhar, Odisha, India,758001, https://orcid.org/0000-0003-0680-3747, madhumitapal@gcekjr.ac.in

[2]Odisha University of Technology and Research, Bhubaneswar, Odisha, India. subhankitprusti2000@gmail.com

[3]Department of Mining Engineering, Orissa School of Mining Engineering, Keonjhar, Odisha, India,758001. rajiv.osme2017@gmail.com

[4]Department of Information Technology, School of Engineering and Technology, Nagaland University, Meriam, Kohima Campus, India. Sujata238dash@gmail.com

[5]Department of Electronics and Communication Engineering, C.V. Raman Global University, Bidyanagar, Mahura, Janla, Bhubaneswar, Odisha 752054, India. ganapati.panda@gmail.com

[6]Hangzhou First People's Hospital, Westlake University School of Medicine, Hangzhou, China. wangzhen@hospital.westlake.edu.cn

[7]Centre for Intelligent Healthcare, Coventry University, Coventry, UK. haipeng.liu@coventry.ac.uk

Corresponding author:

Zhen Wang (wangzhen@hospital.westlake.edu.cn);

Sujata Dash (sujata@nagalanduniversity.ac.in)

Abstract

Congenital heart diseases (CHDs), including Ventricular Septal Defect (VSD), Atrial Septal Defect (ASD), and Patent Ductus Arteriosus (PDA), are among the most prevalent birth-related anomalies in children, often requiring early diagnosis to prevent severe complications. Similarly, respiratory diseases such as tuberculosis, pneumonia, and COVID-19 continue to pose significant public health challenges globally. Early and precise identification of these conditions through chest X-ray analysis plays a crucial role in supporting accurate diagnosis and guiding treatment strategies. This research presents an automated deep learning framework designed to classify congenital heart diseases (CHDs) and respiratory disorders by utilizing two distinct chest X-ray datasets. To capture and learn discriminative image features, advanced models, including Vision Transformer (ViT), ConvMixer, and VGG-19, were applied. The models were trained and evaluated on labeled datasets containing paediatric chest X-rays for CHD detection and mixed adult/paediatric chest X-rays for respiratory disease classification. The performance of each model was assessed using metrics such as accuracy(acc), precision(pr), recall(re), F1-score, Area under the curve (AUC), and Precision Recall(P-R) curves. Experimental results demonstrate that the Vision Transformer outperforms traditional convolution-based models, achieving superior classification accuracy across all disease categories. Based on our findings, the ViT model outperformed the other two models across both datasets. For Dataset-1, the ViT model achieved an average acc of 0.86**,** pr of 0.83, re of 0.71, an F1-

score of 0.76, and an AUC of 0.80. For Dataset-2, the model achieved an acc of 0.93**, a** pr of 0.89, a re of 0.88, an F1-score of 0.87, and an AUC of 0.89. This work emphasizes the effectiveness of transformer-based models in the field of healthcare imaging analysis tasks and their applicability in real-time healthcare environments through integration with IoT and cloud platforms.

Keywords: Congenital heart disease, accuracy, precision, recall, F1-score, Vision transformer, IoT

1. Introduction

CHD constitutes the most common form of structural cardiac anomalies present at birth, affecting approximately 1.1 % of neonates worldwide and up to 1 % of annual births in the United States alone [1]. Among these, VSD, ASD, and PDA represent three of the most prevalent subtypes, frequently leading to hemodynamic disturbances, heart failure, and long-term morbidity if not identified and managed early [2]. In parallel, respiratory pathologies such as pneumonia, tuberculosis, and coronavirus disease 2019 (COVID-19) continue to impose substantial global health burdens, particularly in low-resource and remote settings where access to expert radiological interpretation is limited [3], [4]. Chest radiography remains the diagnostic mainstay for initial screening of both CHDs (via evaluation of cardiac silhouette and pulmonary vasculature) and respiratory diseases (through assessment of parenchymal opacities, cavitations, and consolidation patterns). However, manual analysis of X-ray images is labor-intensive, subject to inter-observer variability, and often constrained by the availability of specialized clinicians.

Contemporary improvements in DL have demonstrated remarkable success in automating medical image interpretation, achieving expert-level performance in tasks such as diabetic retinopathy grading, pulmonary nodule detection, and arrhythmia classification from electrocardiograms [5]–[7]. Traditional convolutional neural networks (CNNs) like VGG-19 have delivered high accuracy in chest-X-ray classification, yet they may struggle to capture long-range dependencies and global context within medical images [8]. More recently, hybrid and transformer-based architectures—including ConvMixer and Vision Transformer (ViT)—have emerged as powerful alternatives, leveraging multi-head self-attention and patch-based embeddings to model spatial relationships more effectively [9], [10].

Despite these successes in generic image domains, their application to the simultaneous detection of both CHDs and diverse respiratory conditions from chest X-rays remains underexplored. This study introduces an integrated deep learning framework for the automatic identification of VSD, ASD, PDA, pneumonia, tuberculosis, and COVID-19 from two separate chest X-ray datasets. To ensure fairness, the performance of three advanced architectures—ViT, ConvMixer, and VGG-19—is compared under uniform training and evaluation settings. The models are evaluated using multiple metrics, including acc, pr, re, F1-score, ROC curves, and precision-recall analysis. In addition, a deployment strategy is proposed that incorporates Internet of Things (IoT) devices and cloud infrastructure to facilitate real-time applications.

The objectives of this work are as follows:

1. To conduct a comprehensive study on various types of congenital heart diseases and respiratory illnesses.
2. To implement DL models for the detection of congenital heart disease and respiratory conditions using two distinct chest X-ray image datasets.
3. To apply and evaluate the performance of ViT, ConvMixer, and VGG-19 architectures for the classification of multiple types of congenital and respiratory diseases from chest radiographs.
4. To assess the effectiveness of the deep learning models using performance metrics such as acc, pr, re, F1-score, AUC, and P-R curve.

5. To explore the integration of Internet of Things (IoT) and cloud computing technologies for the deployment of deep learning models in a real-time clinical environment.

The structure of this paper is as follows: Section 2 presents a review of existing studies on deep learning approaches for chest X-ray analysis. Section 3 details the datasets, pre-processing techniques, and model architectures employed. Section 4 explains the methods adopted for detecting congenital heart disease (CHD) and respiratory disorders. Section 5 provides an overview of the deep learning models, while Section 6 discusses the experimental results. The performance assessment of the proposed DL model has been compared with prior work in Section 7. Finally, Section 8 concludes the work and highlights potential directions for future research

2. Literature Survey

Several researchers have employed ML and DL techniques for the classification and diagnosis of cardiovascular diseases using phonocardiogram (PCG) and electrocardiogram (ECG) signals. Chen et al. [11] utilized a deep neural network (DNN) model for PCG signal classification and achieved an acc of 93%, sensitivity(sens) of 98%, and specificity(sf) of 88.5%. Dominguez-Morales et al. [12] employed a convolutional neural network (CNN) using the PhysioNet database and reported an accuracy of 81.5%, sensitivity of 84.5%, and specificity of 78.5% for heart disease detection. Ghaemmaghami et al. [13] achieved an acc of 95.5%, sf of 97%, and sens of 89.3% using a DNN-based approach. Sujadevi et al. [14] applied recurrent neural network (RNN), long short-term memory (LSTM), and gated recurrent unit (GRU)models on the PhysioNet dataset, obtaining an acc of 85.5%, sf of 86%, and sens of 83.9%. Chen et al. [15] diagnosed extrasystole sound signals using a DNN, achieving an acc of 93%, sens of 94.5%, and sf of 98.33%. Bozkurt et al. [16] used a CNN model on the UoC-Murmur and PhysioNet-2016 databases, reporting an F1-score of 96% for heart abnormality detection via PCG signals. Messner et al. [17] achieved an acc of 95.8%, sens of 99.2%, and sf of 83.2% using the PhysioNet database. Wu et al. [18] implemented a CNN for heart disease prediction and achieved an acc of 89.81% on the PhysioNet dataset. Abduh et al. [19] used a DNN model to classify heart sound signals, reporting an acc of 95.5%. Alafif et al. [20] applied a 2D-CNN combined with transfer learning on the PASCAL database and achieved an accuracy of 89.5%. Deng et al. [21] reported an acc of 98% for PCG signal classification using a hybrid CNN-RNN architecture. Nguyen et al. [22] obtained an acc of 99.67% using a CNN+LSTM model on the PhysioNet 2016 dataset. Yang et al. [23] achieved an acc of 98.74% and an AUC of 0.99 using the AudioSet database for heart sound classification. Acharya et al. [24] used a CNN on the MIT-BIH ECG dataset and reported an acc of 94% and an AUC of 89.07%. Krishnan et al. [25] implemented RNN, GRU, and LSTM models on the Cleveland Heart Disease dataset and achieved an acc of 98% and an F1-score of 91%. Mehmood et al. [26] reported an acc of 97% using a CNN for cardiovascular disease prediction. Pal et al. [27] employed a random forest model and achieved an acc of 86.9%, sens of 90.6%, and sf of 82.7% for heart disease classification.

From this literature review, it is evident that a wide range of ML and DL models have been employed for the detection and classification of cardiovascular diseases. However, there is limited research focusing on the detection of **congenital heart disease (CHD) and respiratory diseases** using advanced models such as **Vision Transformer (ViT)**, **ConvMixer**, and **VGG-19**, particularly from chest X-ray imaging data. This highlights a significant opportunity for further exploration in this domain.

Compared with existing works, the contributions of this work are as follows:
1. A comparative study of transformer-based (ViT), hybrid (ConvMixer), and classical (VGG-19) DL models for multi-disease classification on chest X-rays.
2. Comprehensive performance evaluation across congenital and respiratory disease cohorts, highlighting strengths and limitations of each architecture.

3. A proof-of-concept IoT/cloud integration pipeline for on-device image acquisition, model inference, and centralized result dissemination.

3. Dataset Description

In this study, two chest X-ray image datasets were utilized for simulation purposes. The first dataset (D-1) was employed for the detection of CHD, while the second dataset (D-2) was used to detect various respiratory diseases, including pneumonia, COVID-19, and tuberculosis. Both datasets were obtained from the Kaggle platform [28], [29]. Dataset D-1 contains 828 paediatric chest X-ray images in DICOM format, divided into four categories: normal (208), VSD (210), ASD (194), and PDA (216). Of these, 662 images were allocated for training and 166 for testing the deep learning models, with a summary provided in Table 1. Dataset D-2 comprises chest X-rays related to tuberculosis, pneumonia, COVID-19, and normal cases, including 6,326 training samples, 38 validation samples, and 771 test samples, which were utilized to develop and assess the models for respiratory disease classification

The class levels mentioned for different chest X-ray images are as follows:
ASD:0, Normal:1, PDA:2, VSD:3, pneumonia:0, tuberculosis:1, normal:2, covid-19:3

Table 1. Details of the dataset-1

Types of defects found in the chest X-ray image	Male	Female	Number of patients	Median age
ASD	83	111	194	6.7
VSD	90	120	210	6.5
PDA	93	123	216	6.8
Normal control group	89	119	208	7.1
Total.	355	473	828	473

4. Methodology Framework: IoT & Cloud for Disease Detection

For the simulation work, we utilized the congenital heart disease dataset and the chest X-ray image dataset from the publicly available Kaggle data source [28] [29]. The image dataset was then pre-processed using various data augmentation techniques, including normalization, zooming, and random flipping. After that ViT model is implemented on the pre-processed data. Eighty percent of the data is used for training, and 20 percent is reserved for testing the model. After that, the ViT model predicts the X-ray image as ASD, Normal, PDA, and VSD. The details are given in Fig.1.

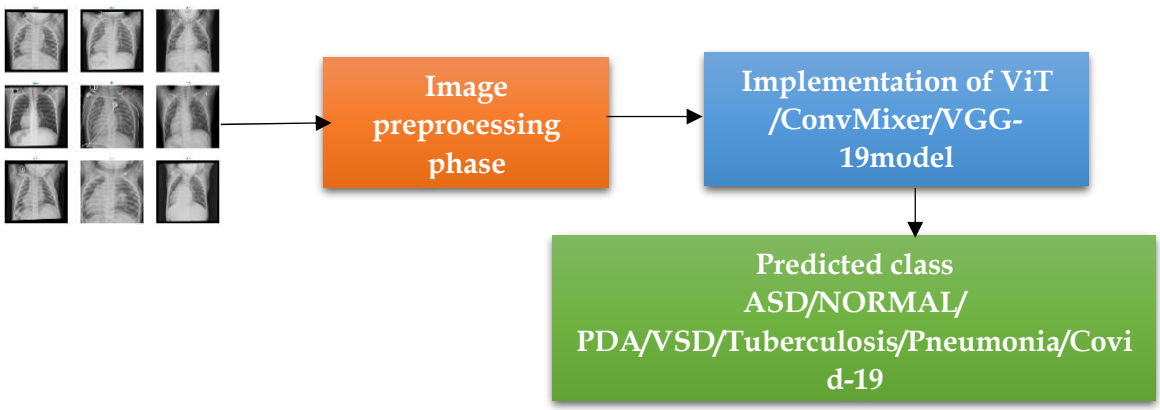


Figure 1. Block diagram for the detection of heart diseases from chest x-ray image

As shown in Fig.2, detection of heart diseases has been done using IoT and cloud computing technology.

IoT-enabled X-ray Imaging:

Step 1: Smart X-ray machines capture chest X-ray images.

Step 2: IoT sensors monitor the imaging environment (e.g., temperature, machine calibration).

Data Transmission to Edge Devices:

Step 3: X-ray images and environmental data are transmitted to edge devices for pre-processing.

Edge Computing:

Step 4: Edge devices pre-process the images (e.g., noise reduction, contrast enhancement).

Step 5: Pre-processed images are temporarily stored and metadata is generated.

Cloud Integration and Storage:

Step 6: Pre-processed images and metadata are uploaded to cloud storage (e.g., AWS S3).

Step 7: Metadata of the image used for labeling the class.

Model Deployment on AWS Sage Maker:

Step 8: Pre-processed images are fed to the deployed model on AWS Sage Maker for classification.

Step 9: The model performs classification and returns results.

Result Analysis and Alert System:

Step 10: Classification results are analyzed.

Step 11: Alerts are generated and sent to healthcare providers if abnormalities are detected.

Data Visualization and Access:

Step 12: Results and data are visualized on dashboards accessible to clinicians.

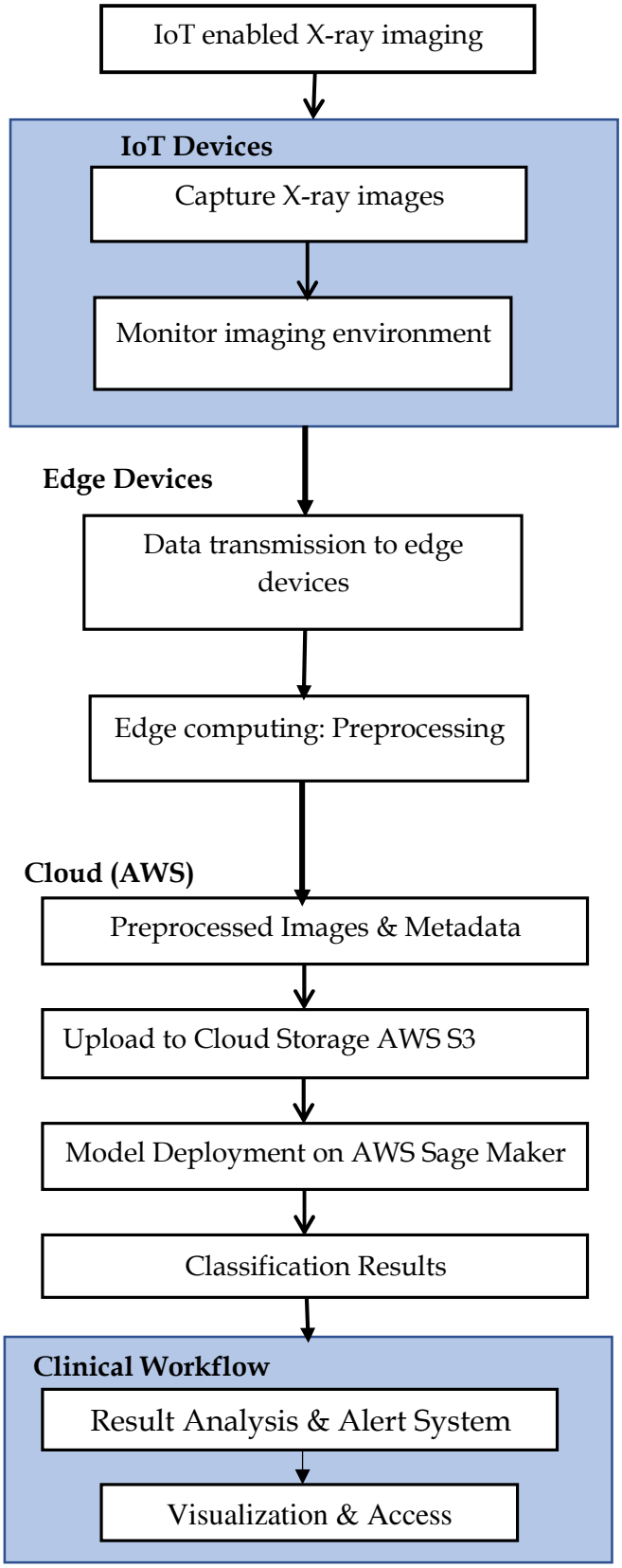


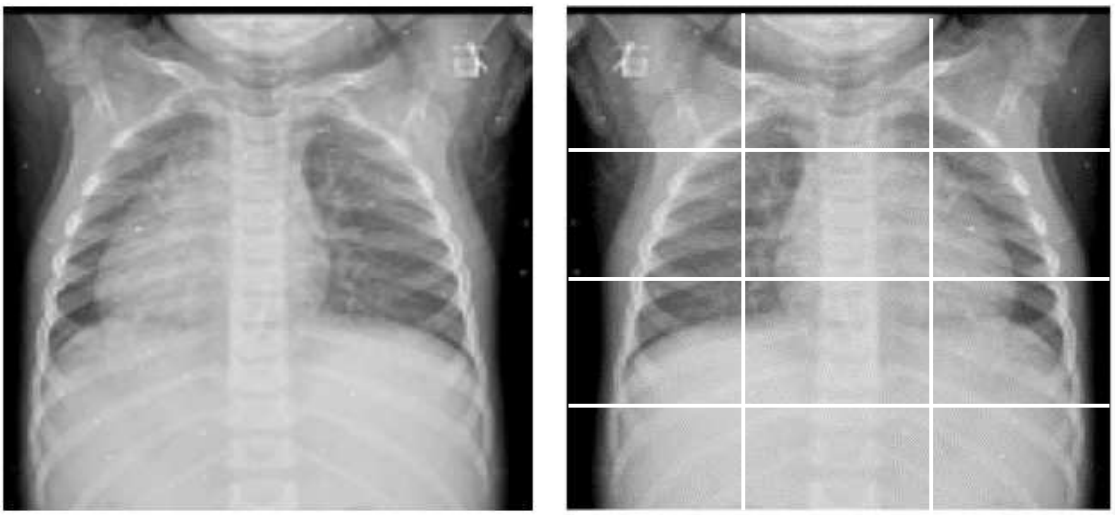
Figure 2. Flow diagram for CHD detection using IoT technology

5. Deep Learning Models

Three DL models are used in the study to classify congenital heart disease and respiratory diseases: ViT, ConvMixer, and VGG-19. Details of the DL models are given below.

5.1 Vision transformer model:

The **Vision Transformer (ViT)** architecture is composed of multiple transformer blocks, each containing two primary sub-layers: a multi-head self-attention layer and a feed-forward neural network layer. The self-attention mechanism [30] calculates attention weights for each image patch, enabling the model to focus on relevant parts of the image. The feed-forward layer then applies a non-linear transformation to the outputs of the self-attention mechanism, enhancing feature representation. A key component of ViT is the patch embedding layer, which divides the input image into fixed-size, non-overlapping patches. Each patch is then flattened and projected into a high-dimensional vector space to form a sequence of embeddings. These embeddings are further enriched by adding positional encodings to retain spatial relationships among the patches. The final output of the transformer encoder is passed through a classification head to generate the class prediction. The self-attention mechanism in ViT enables global context modelling, allowing the network to integrate information across the entire image. Additionally, ViT learns spatial relationships between patches by encoding their relative positions during training, effectively reconstructing the image's structure from its patch-based representation. The attention maps generated by the Vision Transformer (ViT) represent the relative importance assigned to different regions of the input image within the model's internal feature representations. As illustrated in Fig. 3a, the input image is initially partitioned into non-overlapping patches (Fig. 3b). These patches are then flattened and passed into the transformer encoder, as depicted in Fig. 4. The attention maps are derived from the self-attention mechanism, which computes attention weights between each patch (or token) and all other patches in the image. These weights are used to generate a weighted aggregation of contextual information across the image, enabling the model to capture long-range dependencies and spatial relationships.



(a) (b)

Figure 3. (a) Original image, (b) Implementation of the patch encoding layer.

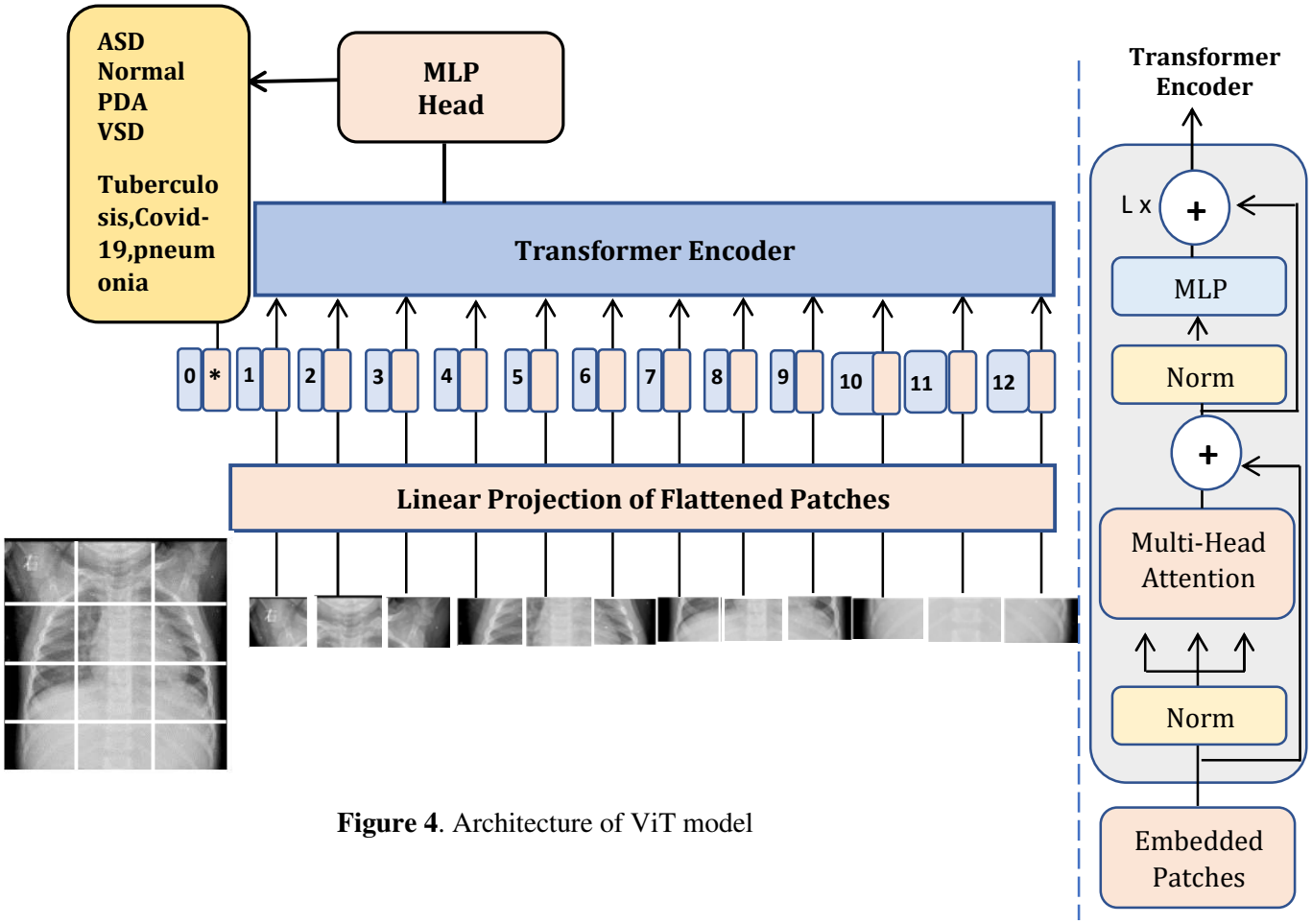


Figure 4. Architecture of ViT model

The parameters used for the implementation of the ViT model are as follows:
(i) Number of classes=4
(ii) Input image shape= (128*128*3)
(iii) Learning rate(η)=0.001
(iv) Weight decay=0.0001
(v) Batch size=16
(vi) Number of epochs=100
(v) Image size=72
(vi) Transformer layers=8
(vii) MLP head units= [2048,2024]

5.2 ConvMixer model:

The architecture of the convmixer model is based on convolution rather than a transformer. It uses image patches as input and performs depth-wise convolution followed by pointwise convolution for classifying the input image. The performance of this model can be enhanced by using a large kernel size. It splits the image into patches rather than pixels. It separates the spatial and channel dimensions. It maintains equal size and resolution throughout the network. It uses batch normalization rather than layer normalization. It takes input as a patch embedding layer, which consists of a convolutional layer of equal size to the patch, with C input channels and h output channels, followed by an activation function and a batch normalization layer (Equation 1).

$Y_0 = \mathrm{BN}\,(\sigma\ \mathrm{Conv}_o \rightarrow h\ (M, stride = s, kernel\ size = k))$.....(1)

It performs two types of convolution such as depth and point-wise convolution on the input image. Depth-wise convolution is used to mix the spatial dimension of the channel and point-wise convolution is used (1*1) kernel to iterate over every point in the image. pointwise convolution is also used to increase the number of filters in the image [30][32].

5.3 VGG-19

The VGG-19 model is a deep convolutional neural network architecture that was proposed by the Visual Geometry Group at the University of Oxford. It is an extension of the VGG family and is known for its simplicity, uniform architecture, and strong performance in image classification tasks. VGG-19 consists of 19 layers, including 16 convolutional layers and 3 fully connected layers, along with max-pooling and ReLU activation functions applied throughout the network [33]. The model uses small 3×3 convolutional filters and a fixed stride and padding, which helps in capturing fine details in images. One of the key advantages of VGG-19 is its deep hierarchical structure, which enables it to learn complex and high-level features from input images. Despite its relatively large number of parameters and high computational requirements, VGG-19 has been widely adopted in computer vision tasks due to its effectiveness and ease of implementation. In medical imaging applications, such as chest X-ray analysis, VGG-19 is often used as a backbone network or a feature extractor to classify various diseases with high accuracy.

Parameters used for the VGG-19 model

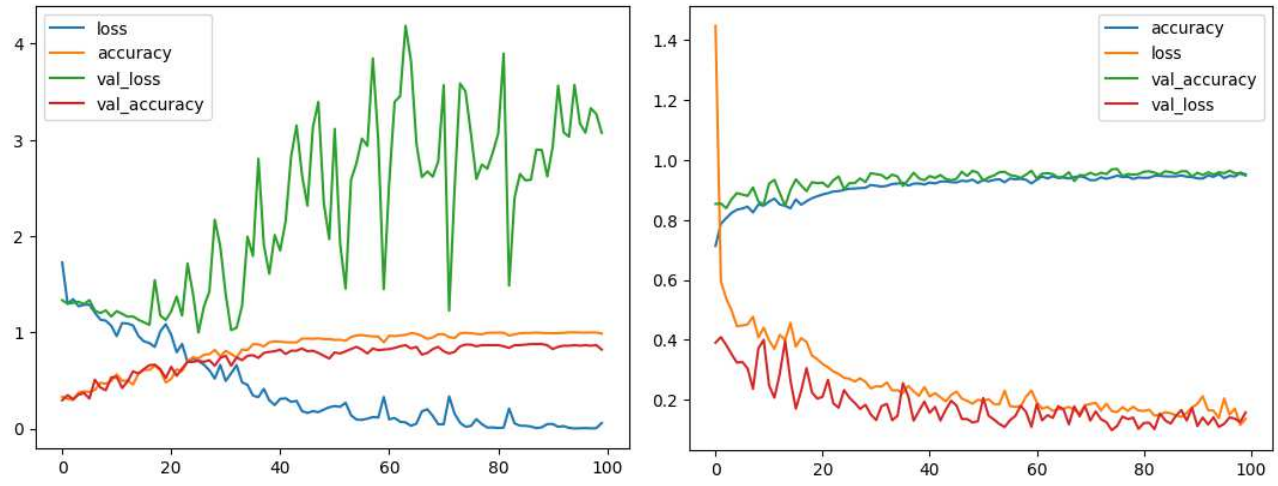
Batch Size=8, optimizer=Adam, Activation=Relu, Softmax, Loss=sparse categorical loss function, Epoch=100, Dropout=0.5.

6. Results

In this section, the performance of three DL models has been evaluated in terms of accuracy, precision, recall, F1-score, AUC, and P-R curve.

6.1 Performance evaluation of the ViT model

The training and validation accuracy and loss curve of the ViT model are shown in Fig.5.



(a) (b)

Figure 5. Visualization of the training and validation accuracy and loss curves of the ViT model for (a) dataset D-1, (b) D-2.

According to Fig. 5(a), the model achieved a training accuracy of **98.70%** with a loss of **0.05**, whereas the testing phase resulted in an accuracy of **82.08%** and a loss of **3.07**. The entire training process was completed in **10.7 seconds**. In contrast, when the Vision Transformer (ViT) was applied (Fig. 5b), the training phase yielded an accuracy of **95.26%** with a loss of **0.1229**, while the validation results showed **94.79%** accuracy and a loss of **0.1581**. The training duration for the ViT model was considerably longer, taking **38 minutes and 4 seconds**. The plotted accuracy and loss curves across

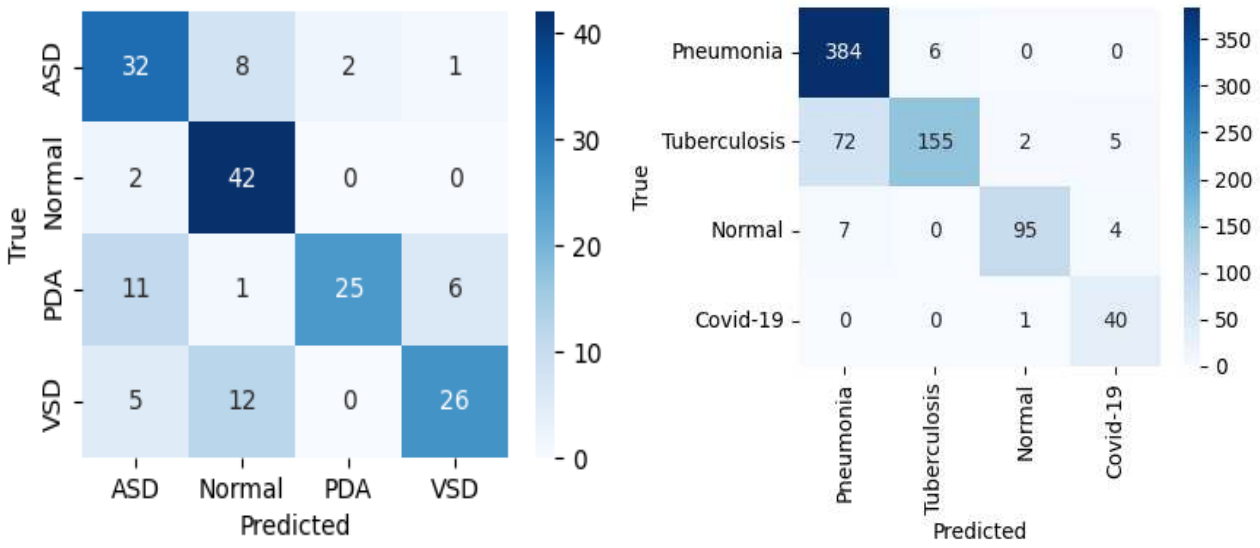
100 epochs clearly illustrate the model's learning dynamics in chest X-ray image classification. From the plot, the training accuracy (blue line) shows a steady increase, starting around 0.75 and converging near 0.95. Similarly, the validation accuracy (green line) closely tracks the training accuracy, maintaining a value between 0.90 and 0.95 across epochs. The close alignment of both curves suggests that the model generalizes well to unseen data and does not suffer from overfitting. The training loss (orange line) shows a rapid decrease in the initial epochs—from a high of over 1.4 down to below 0.2—which is typical behavior as the model adjusts its weights to minimize classification error. The validation loss (red line) follows a similar trend, decreasing smoothly and stabilizing at a low value. This consistent decline without divergence between training and validation loss further supports the conclusion that the model is well-tuned and not overfitting.

The confusion matrix shown in Fig. 6 represents the classification outcomes of the applied machine learning model. It highlights four essential parameters: True Positive (TP.), where both the actual and predicted outputs are positive; False Positive (FP.), when the true label is negative but the prediction is positive; True Negative (TN.), where both prediction and actual class are negative; and False Negative (FN.), when the true label is positive but is incorrectly predicted as negative. By leveraging deep learning, the model is able to extract meaningful and discriminative features from chest radiographs to support accurate classification. The overall performance, measured in terms of acc, pr, re, and F1-score for congenital heart disease (CHD) and respiratory illness detection, is presented in Table 2. For dataset D-1, the Vision Transformer (ViT) model achieved a maximum acc of 0.82, pr of 0.76, re of 0.73, and an F1-score of 0.72. For dataset D-2, the model attained average values acc of 93.5%, 88.84% precision, 87.42% recall, and 86.89% F1-score.

Table 2. Performance evaluation of the ViT model for D-1& D-2

Performance parameter	ASD	Normal	PDA	VSD	Pneumonia	Tuberculosis	Normal	Covid-19
Acc= $(\frac{TP\prime+TN\prime}{TP\prime+TN\prime+FP\prime+FN\prime})$	0.83	0.87	0.88	0.86	0.88	0.89	0.98	0.99
pr = $(\frac{TP\prime}{TP\prime+FP\prime})$	0.84	0.68	0.93	0.88	0.83	0.96	0.97	0.82
re= $(\frac{TP\prime}{TP\prime+FN\prime})$	0.74	0.95	0.58	0.60	0.98	0.66	0.90	0.98
F1-score= $(\frac{2*P*Re}{P+Re})$	0.69	0.83	0.71	0.81	0.90	0.78	0.93	0.89

*Pr=precision
*Re=Recall



(a) (b)

Figure 6. Confusion matrix of ViT model (a) D-1, (b) D-2.

As shown in Fig.6(a)we have obtained the TP, TN, FP, and FN values for ASD classification are 32,112,18 &11respectively. Similarly, the TP, TN, FP & FN values for the normal group are 42,108,21 &2, for PDA classification are 25,128,2 &18, and for the VSD group are 26,123,27&17 respectively. For Pneumonia classification the obtained confusion matrix(Fig.6(b)) parameter values are TP = 384 (Correctly predicted Pneumonia),FN = 6 (Tuberculosis predicted as Pneumonia),FP = 72 + 7 + 0 = 79 (Other classes predicted as Pneumonia),TN = Total other correct predictions = 155 + 2 + 5 + 95 + 4 + 1 + 40 = 302.Similarly for Tuberculosis the parameter values obtained from the confusion matrix are TP = 155, FN = 72 (Tuberculosis misclassified as Pneumonia) + 2 (as Normal) + 5 (as Covid-19) = 79, FP = 6 (Pneumonia predicted as Tuberculosis) + 0 (Normal) + 0 (Covid-19) = 6, TN = Remaining correct predictions = 384 + 95 + 40 = 519, For Normal chest TP = 95, FN = 7 (as Pneumonia) + 0 (as Tuberculosis) + 4 (as Covid-19) = 11, FP = 2 (Tuberculosis predicted as Normal) + 1 (Covid-19 predicted as Normal) = 3, TN = 384 + 6 + 155 + 5 + 40 = 59.For Covid-19 the obtained confusion matrix parameter vales are TP = 40,FN = 0 + 0 + 1 (Normal misclassified as Covid-19) = 1,FP = 0 (Pneumonia) + 5 (Tuberculosis) + 4 (Normal) = 9,TN = 384 + 6 + 72 + 155 + 7 + 0 + 95 = 719.

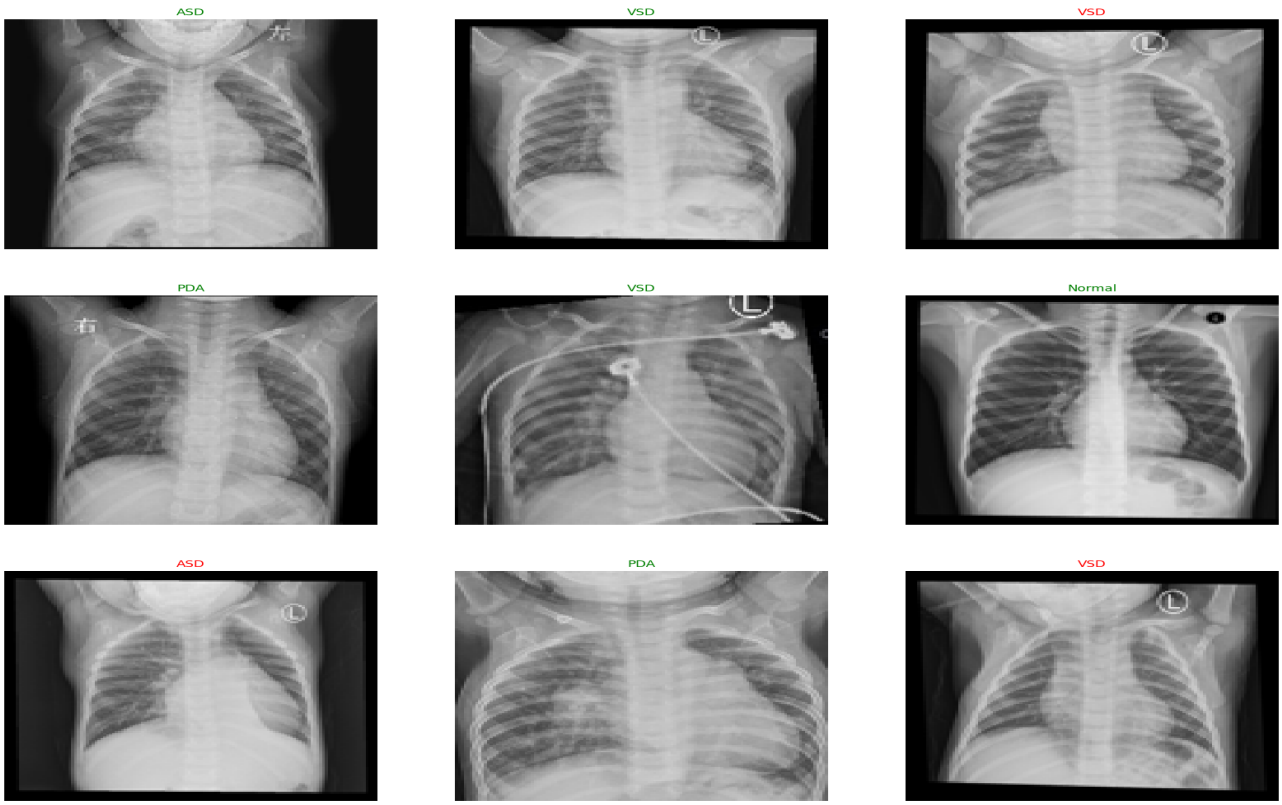


Figure 7. Output of ViT model for dataset-1

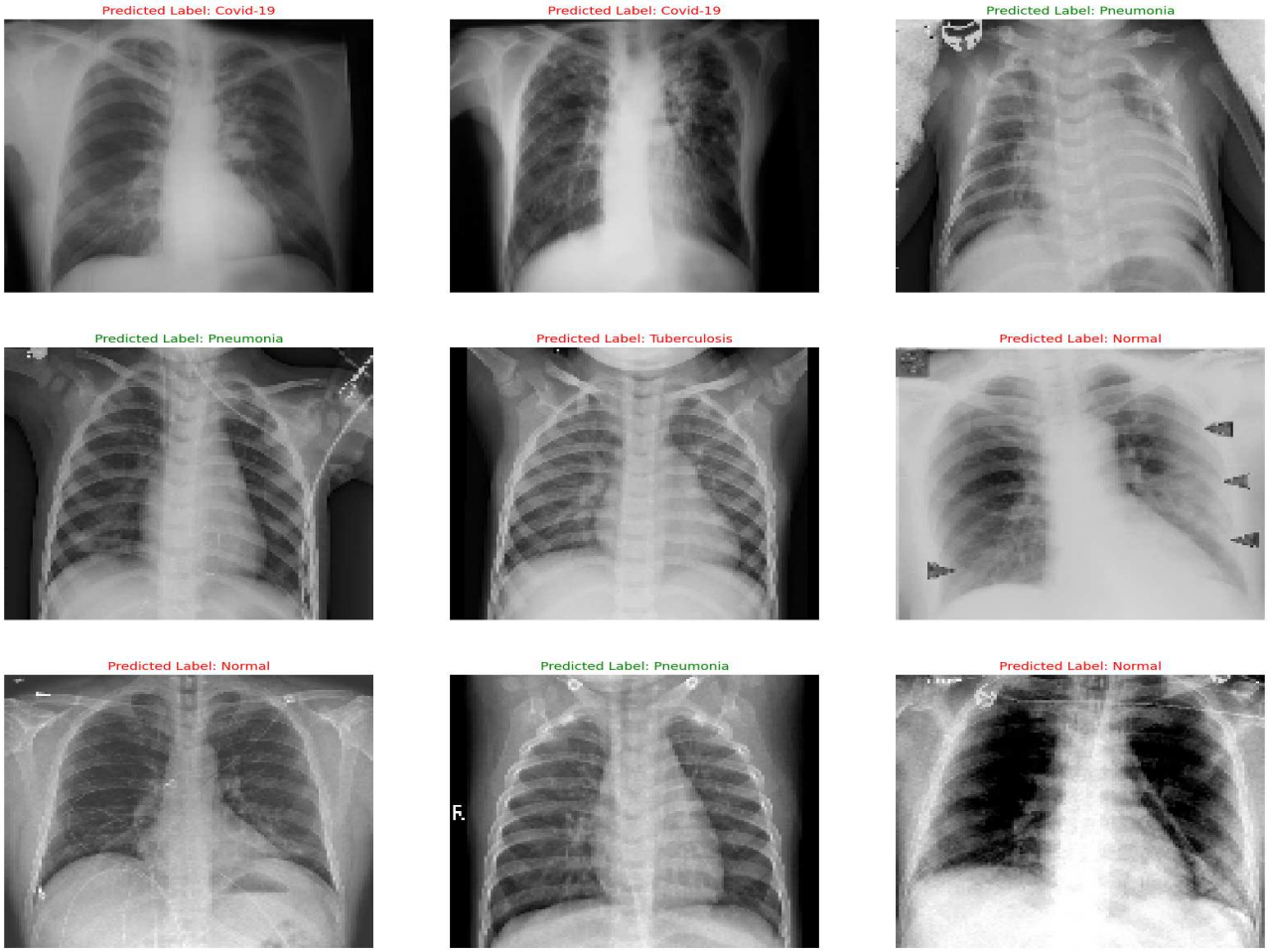
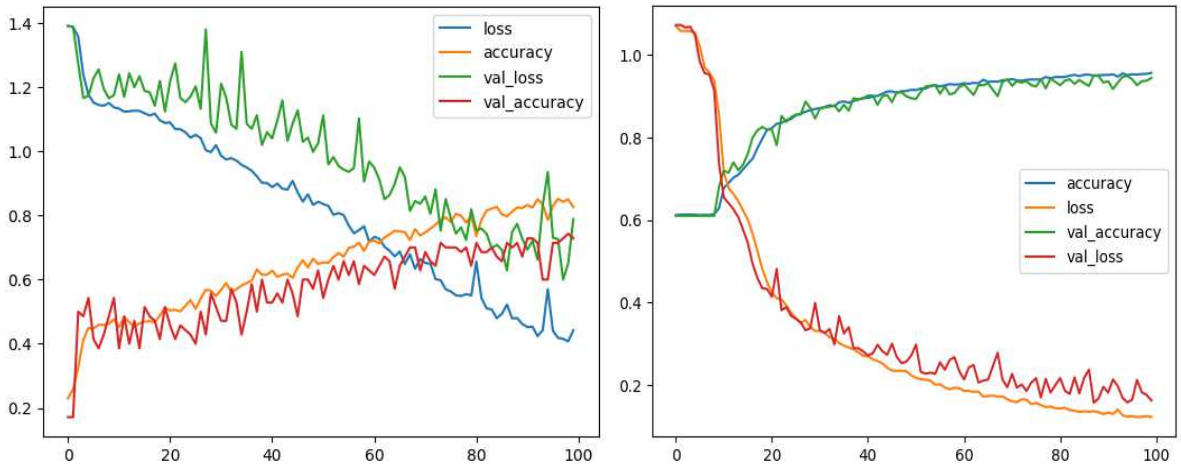


Figure 8. Output of the ViT model for dataset-2

If a patient suffers from ASD, the x-ray image shows a large right atrium with a ventricle as shown in Fig.7. The PDA patient's x-ray image shows a prominent left atrium with a prominent left ventricle and an increase in the marking of the pulmonary vessels. VSD shows cardiomegaly and increased vascular marking. The Fig.8. illustrates the visual predictions made by the ViT model on chest X-ray images for the classification of four conditions: COVID-19, Pneumonia, Tuberculosis, and Normal. Each image is accompanied by a predicted label, highlighted in green for correct predictions and red for incorrect ones.

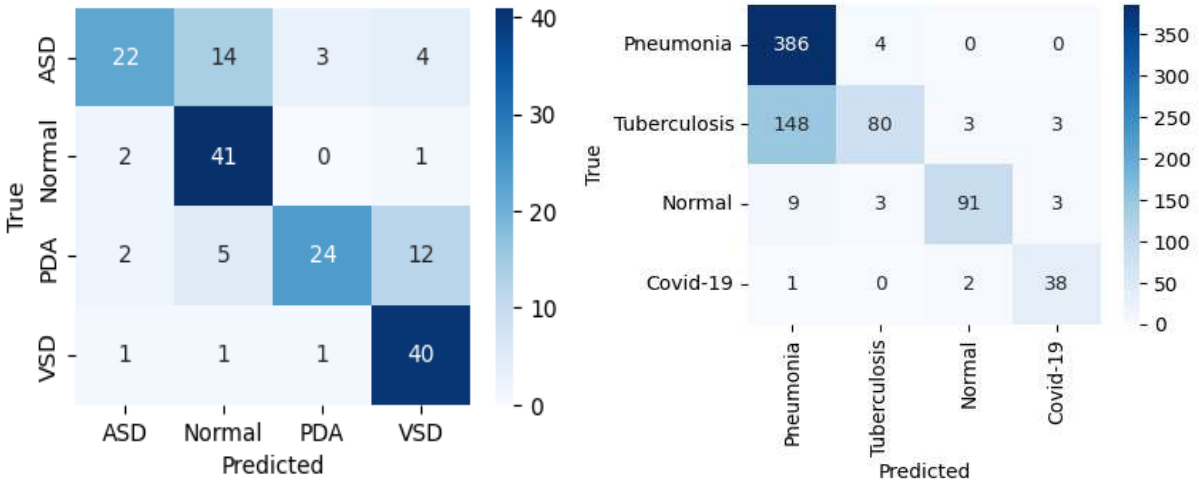
6.2 Performance evaluation of the Convmixer model



(a) **(b)**

Figure 9. Accuracy and curve comparison of the convmixer model for (a)D-1 (b) D-2

From the accuracy curve shown in Fig.9a. we have obtained training loss and accuracy of 0.44 and 82.61 percent, validation loss and accuracy of 0.77 and 73.41 percent. The total run time is 23 minutes and 22 seconds. The plotted training curve (Fig.9b) represents the model's training and validation accuracy and loss over 100 epochs. The graph includes four key metrics: Training Accuracy (blue line): Remains flat at around 0.60 across all epochs. Training Loss (orange line): Shows a gradual decrease from approximately 1.08 to just below 1.00. Validation Accuracy (green line): Also remains constant at around 0.60. Validation Loss (red line): Decreases steadily from around 1.08 to 0.95



(a) **(b)**

Figure 10. Confusion matrix of Convmixer model for (a) dataset-1, (b)dataset-2

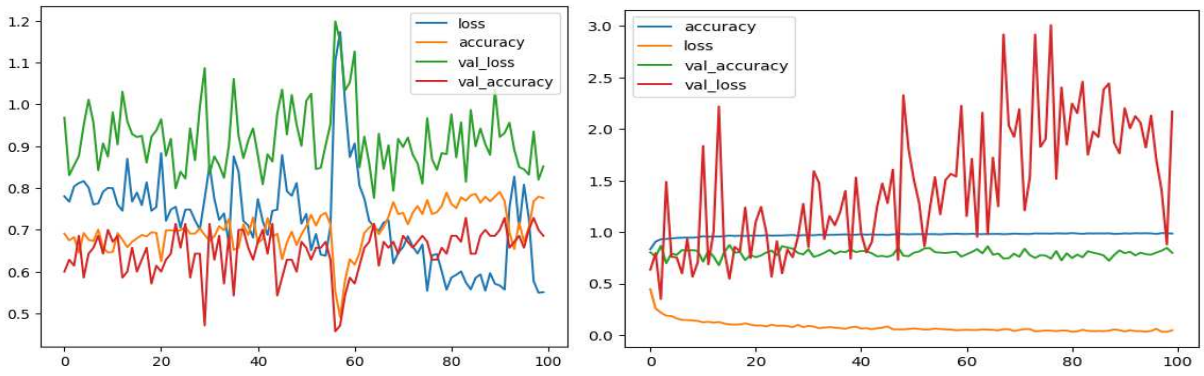
From the confusion matrix (Fig.10a) we have obtained the TP values for ASD, Normal, PDA, and VSD classification are 22,41,24, and 40, respectively. Similarly, the TN values for ASD, Normal, PDA, and VSD classifications are 125,83,126,113, respectively. The FN values for ASD, Normal, PDA, and VSD classification are 21,3,19, and 3, respectively. Similarly, the FP values for the classification of ASD, Normal, PDA, and VSD are 5,20,4 and 17, respectively. The confusion matrix shown in Fig.10b illustrates the classification performance of a deep learning model on four classes of respiratory diseases: Pneumonia, Tuberculosis, Normal, and COVID-19. Pneumonia cases are classified with high accuracy: out of 390 true Pneumonia samples, 386 were correctly classified, while only 4 were misclassified as Tuberculosis. This indicates the model's strong ability to detect Pneumonia. Tuberculosis shows noticeable misclassification: only 80 out of 234 cases were correctly identified, whereas 148 were wrongly predicted as Pneumonia, suggesting that the model struggles to distinguish between Tuberculosis and Pneumonia, possibly due to similar radiographic patterns. For the Normal class, 91 out of 106 samples were correctly identified. A small number were misclassified as Pneumonia (9), Tuberculosis (3), and COVID-19 (3), showing relatively good performance. In the COVID-19 class, 38 out of 41 samples were correctly predicted, with only 3 being misclassified into other categories (2 as Normal and 1 as Pneumonia), indicating the model is effective at distinguishing COVID-19 from other conditions. The average acc, pr, re, and F1-score values for the ConvMixer model are 0.73,0.78,0.72, and 0.7, respectively, for D-1 and 0.89,0.86,0.78 and 0.78 respectively (Table 3).

Table 3. Performance metrics comparison of the ConvMixer model for D-1 & D-2.

Types of class	**Types of Datasets**	**Acc**	**Pr**	**Re**	**F1-score**
ASD	D-1	0.81	0.81	0.72	0.65
Normal		0.93	0.67	0.93	0.79
PDA		0.55	0.86	0.58	0.68
VSD		0.93	0.79	0.68	0.68
Pneumonia	D-2	0.97	0.71	0.99	0.83
Tuberculosis		0.87	0.92	0.34	0.50
Normal		0.77	0.95	0.86	0.90
Covid-19		0.97	0.86	0.93	0.89

6.3 Performance evaluation of the VGG-19 model

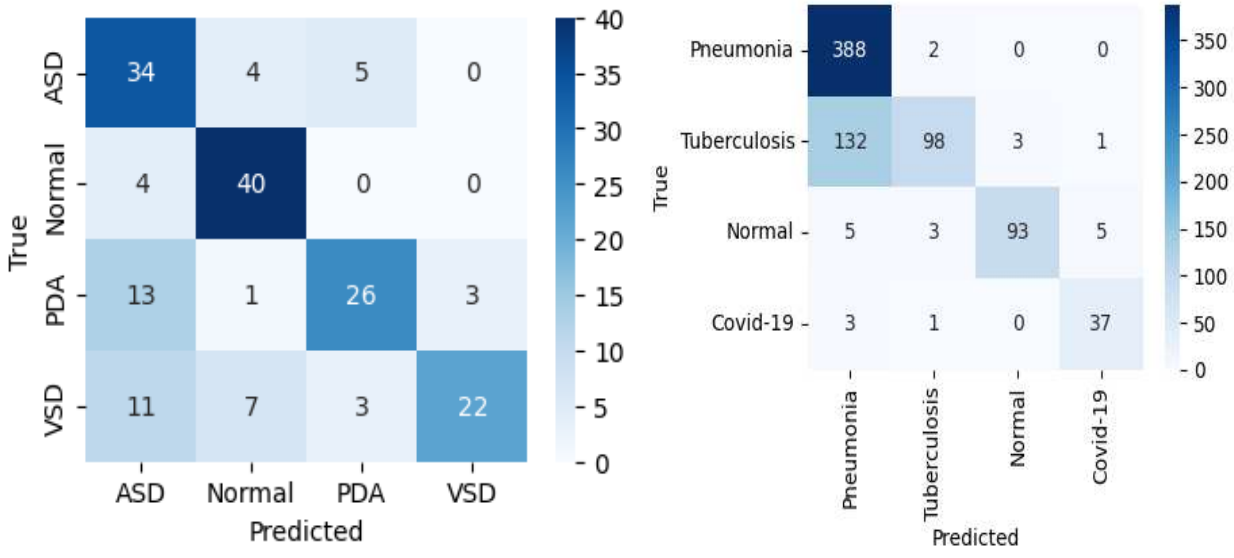
The VGG-19 model, a deep convolutional neural network with 19 weight layers, has been employed for the detection of congenital heart disease (CHD) as well as respiratory illnesses due to its strong feature extraction capability. By leveraging its deep architecture, the model can capture fine-grained spatial patterns from chest X-ray images, which are essential for identifying structural abnormalities associated with CHD and pathological variations linked to respiratory conditions. The performance of the VGG-19 model was measured in terms of acc, pr, re, F1-score, AUC, and P-R curve.



(a) **(b)**

Figure 11. Accuracy and loss curve comparison of the VGG-19 model for (a)D-1, (b)D-2

Using the VGG-19 model, we achieved a training loss of **0.55** with an accuracy of **77%**, while the validation loss reached **0.80** and the validation accuracy was approximately **72.25%**,for D-1 as illustrated in Fig. 11(a).The CPU times: user 2min 56s, sys: 11.4 s, total: 3min 7s, Wall time: 3min 22s.The training and validation curves displayed in Fig.11b illustrate the learning dynamics of a deep learning model applied to chest X-ray image classification over 100 epochs. The metrics plotted include training accuracy and loss (accuracy, loss) as well as validation accuracy and loss (val_accuracy, val_loss). The training loss (orange line) shows a consistent downward trend, stabilizing at a low value, which indicates that the model is learning effectively on the training data. The training accuracy (blue line) steadily increases and flattens near 1.0, showing that the model performs very well on the training set. The validation accuracy (green line) shows modest improvement initially but stabilizes below the training accuracy, indicating a gap between training and validation performance. The validation loss (red line) exhibits significant fluctuation and an increasing trend over time, suggesting overfitting**.**



(a) **(b)**

Figure 12. Confusion matrix of the VGG-19 model for (a)D-1, (b)D-2

The confusion matrix of the VGG-19 model has been shown in Fig.12a.The TP values for ASD, normal, PDA, and VSD are 34,40,26,22, respectively. The TN values for the classification of ASD, normal, PDA, and VSD are 102,117,122 and 127, respectively. The FN values for the classification

of ASD, normal, PDA, and VSD are 9,4,17, and 21, respectively. Similarly, the FP values for the classification of ASD, normal, PDA, and VSD are 28,12,8 and 3, respectively. The confusion matrix presented in Fig.12b offers a detailed insight into the model's classification performance across four classes: Pneumonia, Tuberculosis, Normal, and COVID-19. The diagonal elements represent the correctly classified cases, while the off-diagonal elements highlight misclassifications. From the confusion matrix, we have observed VGG-19 model obtained 388 instances of Pneumonia that were correctly classified. Only two pneumonia cases were incorrectly predicted as Tuberculosis, and none were misclassified as Normal or COVID-19. The VGG-19 model correctly classified 98 cases, while 132 instances were misclassified as Pneumonia, 3 as Normal, and 1 as Covid-19. For the normal case, 93 was correctly identified. 5 cases misclassified as Pneumonia and 5 as Covid-19, with 3 misclassified as Tuberculosis. For COVID-19, 37 cases were correctly classified. Three cases were misclassified as Pneumonia and one as tuberculosis. The model performs fairly well for Covid-19 detection but may benefit from more training data or improved feature extraction to distinguish it more robustly from Pneumonia.

Table 4 provides a comprehensive summary of the performance of the VGG-19 model across multiple medical conditions based on chest X-ray images, including cardiac defects (ASD, PDA, VSD), respiratory diseases (Pneumonia, Tuberculosis, Covid-19), and Normal cases. Each class is evaluated based on acc, pr, re, and F1-score**.** All three cardiac anomalies demonstrate a relatively modest accuracy of 0.71. ASD (Atrial Septal Defect) has the lowest precision (0.55) and recall (0.51) among the three, suggesting the model struggles to correctly identify and distinguish ASD cases. PDA (Patent Ductus Arteriosus) demonstrates higher precision (0.76) but a low recall (0.56), indicating it predicts fewer false positives but misses many true cases.VSD (Ventricular Septal Defect) achieves a more balanced profile with pr (0.70) and re (0.67), resulting in a consistent F1-score (0.65) for dataset 1. Pneumonia exhibits excellent re (0.96), indicating that nearly all positive cases are correctly identified. The pr(0.73) is also reasonable, and the F1-score of 0.82 confirms strong classification ability. Tuberculosis, on the other hand, shows a striking imbalance: while it has a very high pr (0.94), the re is low (0.42). This suggests that while the model rarely mislabels non-TB cases as TB, it often misses actual TB cases, a concern in clinical settings. COVID-19 classification is strong, with both pr (0.85) and re (0.89) being high. The F1-score (0.87) suggests a reliable balance between sensitivity and specificity. The normal class appears twice in the table, which could reflect evaluations on different datasets or model iterations. The first entry shows moderate performance with precision (0.65) and recall (0.91). The second entry (with accuracy of 0.87) demonstrates excellent pr (0.97) and re (0.88), leading to a high F1-score (0.90). This suggests the model is highly effective in distinguishing healthy individuals in that particular instance.

Table 4. Performance comparison of the VGG-19 model for D-1 and D-2

Types of class	**Types of Datasets**	**acc**	**pr**	**re**	**F1-score**
ASD	D-1	.79	.55	.51	.63
Normal		.90	.65	.91	.78
PDA		.55	.76	.56	.67
VSD		.51	.70	.67	.65
Pneumonia	D-2	.96	.73	.98	.82
Tuberculosis		.41	.94	.42	.51
Normal		.87	.97	.88	.90
Covid-19		.90	.85	.89	.87

The comparative evaluation of three DL models—ViT, ConvMixer, and VGG-19—across two distinct datasets as given in Table 5 highlights their ability to detect both congenital heart diseases (Dataset-1) and respiratory conditions (Dataset-2) using Area Under the Curve (AUC), 95%

Confidence Intervals (CI), and Precision-Recall (P-R) values. The ViT model achieved average acc, pr, re, and F1-score values of 0.86, 0.83, 0.71, and 0.76, respectively, on Dataset-1 (D-1), and 0.935, 0.895, 0.88, and 0.875, on Dataset-2 (D-2). Using the ConvMixer model, the average values for acc, pr, re, and F1-score were 0.73, 0.78, 0.7275, and 0.70 for D-1, and 0.77, 0.86, 0.78, and 0.78 for D-2. For the VGG-19 model, the corresponding metrics for D-1 were 0.68, 0.66, 0.66, and 0.68, while for D-2, the model achieved 0.79, 0.87, 0.79, and 0.77, respectively.

6.4.AUC and P-R value comparison of the DL model for D-1 and D-2

Table 5 illustrates the comparison of AUC and P-R value for three DL models using two datasets.

6.4.1 Congenital heart disease classification (Dataset-1: ASD, Normal, PDA, and VSD)

ViT outperforms the other models with AUCs consistently above 0.78 for all CHD classes, particularly achieving 0.90 for Normal, and 0.80 for ASD, suggesting robust class separation. The P-R values are moderately strong, with VSD reaching 0.67 and ASD at 0.54, indicating fair precision-recall trade-offs. ConvMixer shows inconsistent performance. While it achieves a good AUC of 0.84 for Normal, its performance deteriorates for ASD (0.67) and VSD (0.65). The lower confidence bounds for PDA (0.411) and VSD (0.814, likely a typo needing correction) indicate significant prediction uncertainty.VGG-19 maintains balanced performance with AUCs ranging from 0.74 to 0.91, indicating that it is generally stable across CHD types. The highest P-R value (0.72) is observed for Normal, but is relatively low for ASD (0.49) and PDA (0.56), suggesting limited precision in those classes.

6.4.2 Respiratory disease detection (Dataset-2: pneumonia, tuberculosis, normal, and Covid-19)

ViT excels in detecting COVID-19 with an AUC of 0.94, supported by a narrow CI (0.979–0.995), indicating high model confidence. It also performs well on Pneumonia (AUC: 0.87) and Normal (0.87). Tuberculosis has an AUC of 0.79, but a CI of 0.868–0.912. VGG-19 demonstrates mixed results. While it performs impressively for Covid-19 (AUC: 0.96, P-R: 0.80) and Normal (AUC: 0.93, P-R: 0.83), it significantly underperforms for Tuberculosis (AUC: 0.66, CI: 0.284–0.405), indicating weak model reliability for this class. Convmixer provides consistently high AUCs across all classes, including Covid-19 (0.95) and Normal (0.95). The model's ability to detect Pneumonia (AUC: 0.83, P-R: 0.73) is also strong. However, like VGG-19, it struggles with Tuberculosis (AUC: 0.68, CI: 0.355–0.482), suggesting this class is intrinsically harder to classify, possibly due to overlapping radiographic features with other lung conditions. ViT demonstrates the best overall AUC performance in both datasets, particularly excelling in COVID-19 and Normal detection.VGG-19 lags in CHD classification but offers competitive results in respiratory diseases, especially Covid-19 detection. Convmixer shows balanced performance, proving effective in both CHD and respiratory domains with high AUC and relatively stable precision-recall scores.

Table 5. AUC and Precision-Recall (P-R) values of deep learning models with 95% confidence intervals for two types of chest X-ray datasets.

Dataset-1						Dataset-2					
Model name	Types of Class	AUC	95% CI lower	95% CI upper	P-R value	Model name	Types of Class	AUC	95% CI lower	95% CI upper	P-R value
ViT	ASD (0)	0.80	0.777	0.888	0.54	ViT	Pneumonia (0),	0.87	0.867	0.911	0.81
	Normal (1)	0.90	0.816	0.918	0.64		Tuberculosis (1),	0.79	0.868	0.912	0.67
	PDA (2)	0.78	0.837	0.932	0.59		Normal (2)	0.87	0.973	0.991	0.70
	VSD (3)	0.78	0.810	0.913	0.67		Covid-19(3)	0.94	0.979	0.995	0.50
VGG-19	ASD (0)	0.67	0.368	0.654	0.54	VGG-19	Pneumonia (0),	0.79	0.974	0.996	0.71
	Normal (1)	0.84	0.818	0.977	0.65		Tuberculosis (1),	0.66	0.284	0.405	0.51
	PDA (2)	0.74	0.411	0.696	0.64		Normal (2)	0.93	0.780	0.912	0.83
	VSD (3)	0.65	0.814	0.976	0.57		Covid-19(3)	0.96	0.806	0.975	0.80
ConvMixer	ASD (0)	0.79	0.648	0.886	0.59	ConvMixer	Pneumonia (0),	0.83	0.987	1.000	0.73
	Normal (1)	0.91	0.788	0.964	0.72		Tuberculosis (1),	0.68	0.355	0.482	0.57
	PDA (2)	0.77	0.456	0.736	0.66		Normal (2)	0.95	0.815	0.939	0.87
	VSD (3)	0.74	0.868	0.954	0.67		Covid-19(3)	0.95	0.812	0.992	0.78

***CI=confidence interval**

7. Discussion

This section represents the performance comparison of the DL models with existing literature. Table 6 presents a comparative evaluation of the proposed ViT-based approach with existing deep learning and machine learning models applied for disease detection. Earlier studies employing conventional architectures such as ResNet-18, CNN, and VGG-16 for congenital heart disease (CHD) and respiratory illnesses reported accuracies ranging from 57.9% to 91.6%, depending on the dataset size and disease complexity. For instance, the ResNet-18 model applied to ECG data for CHD detection achieved 73.9%, while machine learning–based discharge prediction using ultrasound images yielded only 57.9%, reflecting the limited capacity of traditional models to capture complex patterns in heterogeneous clinical data. Similarly, CNN and VGG-16 approaches for pneumonia and COVID-19 classification attained accuracies in the range of 87.2%–87.5%, demonstrating effectiveness but with room for improvement. The proposed Vision Transformer (ViT) approach demonstrates superior adaptability across both CHD and respiratory disease detection tasks. For CHD classification, it achieved 86% accuracy, outperforming several existing models that rely heavily on handcrafted feature extraction or conventional CNN backbones. Furthermore, when applied to multi-disease

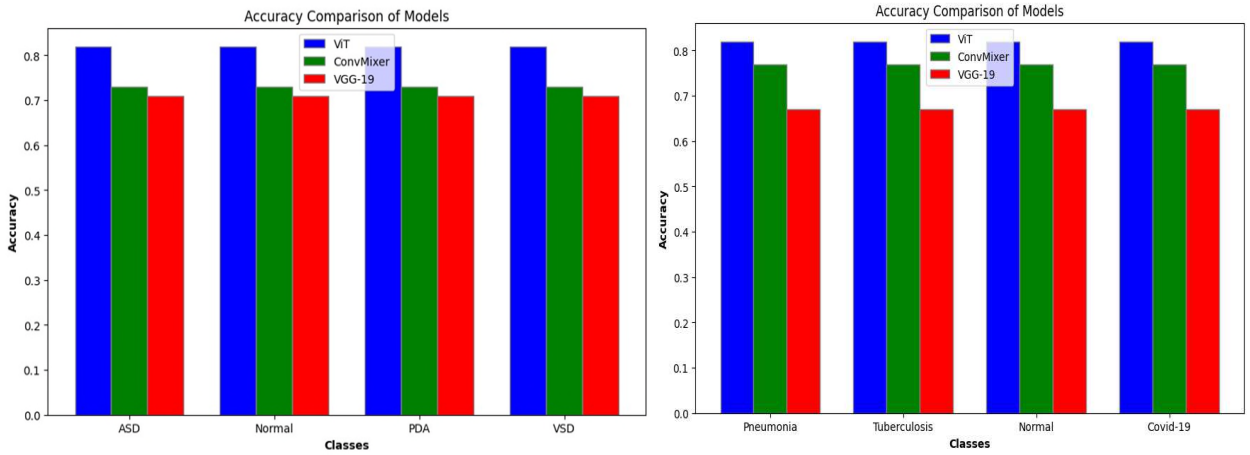
detection (pneumonia, tuberculosis, and COVID-19), the ViT model achieved 93% accuracy, which surpasses or remains competitive with the highest reported accuracies from ensemble and federated learning–based approaches. This improvement can be attributed to the ViT architecture's ability to capture global dependencies through attention mechanisms, enabling more robust learning even from complex and diverse medical datasets.Overall, the results indicate that the proposed ViT model not only enhances detection performance in congenital heart disease but also generalizes well to other thoracic diseases, thereby highlighting its potential as a unified framework for computer-aided diagnosis across multiple medical imaging modalities.

Table 6. Comparative performance analysis of the proposed model and prior studies

Exiting literature	Model name	Types of disease for detection using DL model	Class	Total no. of image samples	Accuracy	precision	Recall	F1-score	Validation type
[36]	ResNet-18	CHD	3	1035 ECG samples	73.9	73.8	73.6	75.8	5-fold cross-validation on an 80/20 split (no external test set)
[37]	Machine Learning-Based Discharge Prediction	CHD	3	6720 ultrasound images	57.9	59.78	40.22	-	Not specified
[38]	LSTM+DenseNet121	CHD	2	13 hypoplastic left heart syndrome ultrasound video images and 9 healthy images	77.3	-	-	-	Hold out test set
[39]	Deep learning model	ASD	2	300 samples of children patient	84.95	-	-	-	5-fold cross validation
[40]	Pretrained DL model	TB	2	Normal image=420, tuberculosis image=400	76.05	-	-	-	-
[41]	VGG-16	Pneumonia & Covid-19	4	Pneumonia (B, V) =657, normal image=310, Covid-19=284	87.49	-	-	-	Hold out

[42]	ResNet 18	Pneumonia,TB & Covid-19	5	COVID-19 image = 435, TB = 434, Normal image439, Pneumonia(B, V)=878	91.60	-	-	-	Five fold cross validation
[43]	CNN	Pneumonia & Covid-19	3	(Pneumonia = 500, Covid-19=130) Normal = 500	87.20	89.98	-	87.73	Five fold cross validation
[44]	Ensemble model (ResNet 50& Xception)	Pneumonia & Covid-19	3	Covid19 = 180, Pneumonia = 6054, Normal = 8851	91.40	35.27	80.53	-	Hold out test
[45]	Federated learning with a neural network	Pneumonia & Covid-19	3	Pneumonia = 4273, Healthy = 1583,COVID-19=2358	89.9	-	-	-	-
Proposed approach	ViT	CHD	4	normal (208), VSD (210), ASD (194), and PDA (216)	86	0.83	0.71	0.76	Test
Proposed approach	ViT	Pneumonia,TB & Covid-19		6,326 training samples, 38 validation samples, and 771 test samples	93	0.89	0.88	0.87	Test

7.1. Overall comparison in performance: potential of the ViT model



(a) **(b)**

Figure 13. Accuracy comparison plot of DL models for(a) D-1& (b) D-2

As illustrated in Fig.13, the ViT model achieves the highest prediction accuracy across all types of respiratory diseases, followed by ConvMixer and VGG-19 for both datasets.The Vision Transformer (ViT) model frequently surpasses ConvMixer and VGG-19 in performance because it effectively captures long-range dependencies and global contextual information in images. Its self-attention mechanism enables the model to assign varying importance to different regions within an image, thereby enhancing feature representation. ViT contains self-attention mechanisms, allowing it to weigh the importance of different image regions when making predictions, which can be particularly beneficial for complex tasks and larger datasets [34]. Whereas, ViT is not suitable for small datasets in the diagnosis of breast lesions [35]. The application of ViT in IoT-based large imaging datasets deserves further investigation.

7.2 Limitations and future directions

This study has some limitations. First, the sample size is small considering the number of different diseases, which may affect the performance of some models like ViT. Secondly, we only included three novel deep learning models, while some advanced architectures were not investigated. Third, the data were from open-source datasets, where the data have been pre-screened. In real-world applications, the IoT-based real-time images may vary in quality, where a structured screening and preprocessing workflow is needed. In the future, more advanced deep learning architectures can be tested on big data derived from real-world IoT devices to validate the results and explore the best implementation in clinical applications.

8. Conclusion

In this study, we presented a deep learning-based framework for the automated classification of congenital heart diseases (CHDs) and respiratory illnesses using chest X-ray images. By leveraging advanced architectures such as Vision Transformer (ViT), ConvMixer, and VGG-19, we effectively demonstrated the potential of deep learning in detecting both paediatric and adult thoracic conditions. Additionally, the article proposes a practical framework for real-world chest X-ray image classification by integrating IoT and cloud computing technologies. Model performance was evaluated based on key metrics such as acc, pr, re, F1-score, AUC, and the Precision-Recall (P-R) curve. Among the evaluated models, the ViT consistently outperformed its convolution-based counterparts, achieving superior accuracy, precision, recall, F1-score, and AUC across two diverse datasets. These findings highlight the efficacy of transformer-based models in extracting high-level discriminative features from medical images, making them well-suited for clinical decision support systems. The proposed deep learning framework not only facilitates early screening of congenital heart disease and respiratory diseases in resource-limited healthcare settings but also aids researchers in understanding the variations in chest X-ray features associated with different types of congenital heart conditions. Early detection through such models could significantly help in reducing complications associated with congenital heart disease and respiratory diseases. For future work, we aim to enhance the performance of the ViT model by addressing class imbalance and incorporating larger datasets. Moreover, congenital heart disease detection can be explored using other deep learning and machine learning models, such as Synaptic Neural Networks, Autoencoders, Deep Belief Networks, ResNet-50, meta ensemble learning method to achieve even greater accuracy. Future research can be directed toward the following enhancements: Incorporating larger and more diverse datasets—including images from different age groups, imaging devices, and geographical regions—can improve model robustness and generalizability. Integrating additional diagnostic modalities such as ECG, CT scans, or clinical metadata with chest X-ray images may enhance prediction performance, particularly for complex cases. Implementing explainable AI (XAI) techniques will help build clinician trust by providing visual or textual justifications for predictions. Collaborating with

healthcare professionals for prospective clinical validation and feedback will be essential to transition the model from research to practical use.

Availability of Data and Materials

The datasets for the current study are available from:
1.https://www.kaggle.com/competitions/congenital-heart-disease.
2.https://www.kaggle.com/datasets/newra008/cxr-data-for-multiclass-classification

Declaration on the use of AI and AI-assisted tools in manuscript preparation:
No AI-based or automated writing tools were employed in the preparation of this paper.

Consent for Publication:
Not applicable.

Funding Statement:
This research was conducted without any external financial support.

Authors' contributions:

Conceptualization, Investigation, Writing – original draft: Madhumita Pal.

Software-Subhankit prusti

Validation, Writing – review & editing: Ganapati Panda.

Writing – review & editing: Rajiv Lochan Bal.

Methodology, Writing- original draft: Haipeng Liu

Investigation and methodology-Zhen Wang

Writing – original draft, writing-review & editing: Sujata Dash.

All authors have agreed and read the published version of the manuscript.

Acknowledgement: Thanks to the Government College of Engineering, Keonjhar for providing me with the platform to work.

References

[1] R. van der Linde, E. E. Konings, M. A. Slager, J. W. Witsenburg, W. A. Helbing, J. J. Takkenberg, and B. J. Roos-Hesselink, "Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis," *J. Am. Coll. Cardiol.*, vol. 58, no. 21, pp. 2241–2247, Nov. 2011.

[2] A. Marelli, A. Ionescu-Ittu, K. M. Mackie, M. Guo, C. Dendukuri, and L. J. Kaouache, "Lifetime prevalence of congenital heart disease in the general population from 2000 to 2010," *Circulation*, vol. 130, no. 9, pp. 749–756, Aug. 2014.

[3] World Health Organization, "Global tuberculosis report 2023," Geneva: WHO, 2023. [Online]. Available: https://www.who.int/tb/publications/global_report/en/

[4] R. Wu, A. Wang, and L. Wang, “Challenges of radiology services in rural and low-resource settings during COVID-19,” *Lancet Glob. Health*, vol. 9, no. 3, pp. e244–e245, Mar. 2021.

[5] M. Pal et al., “Risk prediction of cardiovascular disease using machine learning classifiers,” *Open Med.*, vol. 17, no. 1, pp. 1100–1113, 2022.

[6] D. Ardila et al., “End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography,” *Nat. Med.*, vol. 25, no. 6, pp. 954–961, Jun. 2019.

[7] A. Hannun et al., “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” *Nat. Med.*, vol. 25, no. 1, pp. 65–69, Jan. 2019.

[8] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[9] I. Trockman and J. B. Kolter, “Patches are all you need?,” *arXiv preprint arXiv:2201.09792*, 2022.

[10] A. Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2021.

[11] T.-E. Chen et al., “S1 and S2 heart sound recognition using deep neural networks,” *IEEE Trans. Biomed. Eng.*, vol. 64, no. 2, pp. 372–380, 2016, doi: 10.1109/TBME.2016.2574613.

[12] J. P. Dominguez-Morales, A. F. Jimenez-Fernandez, M. J. Dominguez-Morales, and G. Jimenez-Moreno, “Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 1, pp. 24–34, 2017, doi: 10.1109/TBCAS.2017.2751441.

[13] H. Ghaemmaghami et al., “Automatic segmentation and classification of cardiac cycles using deep learning and a wireless electronic stethoscope,” in *Proc. IEEE Life Sci. Conf. (LSC)*, 2017, pp. 9–12, doi: 10.1109/LSC.2017.8268124.

[14] V. Sujadevi, K. Soman, R. Vinayakumar, and A. P. Sankar, “Deep models for phonocardiography (PCG) classification,” in *Proc. Int. Conf. Intell. Commun. Comput. Tech. (ICCT)*, 2017, pp. 160–164, doi: 10.1109/INTELCCT.2017.8324014.

[15] L. Chen, J. Ren, Y. Hao, and X. Hu, “The diagnosis for the extrasystole heart sound signals based on the deep learning,” *J. Med. Imag. Health Inform.*, vol. 8, no. 5, pp. 959–968, 2018, doi: 10.1166/jmihi.2018.2449.

[16] B. Bozkurt, I. Germanakis, and Y. Stylianou, “A study of time-frequency features for CNN-based automatic heart sound classification for pathology detection,” *Comput. Biol. Med.*, vol. 100, pp. 132–143, 2018, doi: 10.1016/j.compbiomed.2018.07.001.

[17] E. Messner, M. Zöhrer, and F. Pernkopf, “Heart sound segmentation—An event detection approach using deep recurrent neural networks,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1964–1974, 2018, doi: 10.1109/TBME.2017.2762738.

[18] J. M.-T. Wu et al., “Applying an ensemble convolutional neural network with Savitzky–Golay filter to construct a phonocardiogram prediction model,” *Appl. Soft Comput.*, vol. 78, pp. 29–40, 2019, doi: 10.1016/j.asoc.2019.01.011.

[19] Z. Abduh, E. A. Nehary, M. A. Wahed, and Y. M. Kadah, "Classification of heart sounds using fractional Fourier transform-based mel-frequency spectral coefficients and stacked autoencoder deep neural network," *J. Med. Imag. Health Inform.*, vol. 9, no. 1, pp. 1–8, 2019, doi: 10.1166/jmihi.2019.2597.

[20] T. Alafif et al., "Normal and abnormal heart rate recognition using transfer learning," in *Proc. 12th Int. Conf. Knowl. Syst. Eng. (KSE)*, 2020, pp. 225–230, doi: 10.1109/KSE50997.2020.9287761.

[21] M. Deng et al., "Heart sound classification is based on improved MFCC features and convolutional recurrent neural networks," *Neural Netw.*, vol. 130, pp. 22–32, 2020, doi: 10.1016/j.neunet.2020.06.010.

[22] M. T. Nguyen, W. W. Lin, and J. H. Huang, "Heart sound classification using deep learning techniques based on log-mel spectrogram," *Circ. Syst. Signal Process.*, vol. 42, no. 1, pp. 344–360, 2023, doi: 10.1007/s00034-022-02066-9.

[23] D. Yang et al., "Assisting heart valve diseases diagnosis via transformer-based classification of heart sound signals," *Electronics*, vol. 12, no. 10, p. 2221, 2023, doi: 10.3390/electronics12102221.

[24] U. R. Acharya et al., "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, 2017, doi: 10.1016/j.compbiomed.2017.08.022.

[25] S. Krishnan, P. Magalingam, and R. Ibrahim, "Hybrid deep learning model using recurrent neural network and gated recurrent unit for heart disease prediction," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 6, pp. 5303–5311, 2021, doi: 10.11591/ijece.v11i6.pp5303-5311.

[26] A. Mehmood et al., "Prediction of heart disease using deep convolutional neural networks," *Arab. J. Sci. Eng.*, vol. 46, no. 4, pp. 3409–3422, 2021, doi: 10.1007/s13369-020-05159-1.

[27] M. Pal and S. Parija, "Prediction of heart diseases using random forest," *J. Phys.: Conf. Ser.*, vol. 1964, p. 042005, 2021, doi: 10.1088/1742-6596/1964/4/042005.

[28] Kaggle, "Congenital heart disease," [Online]. Available: https://www.kaggle.com/competitions/congenital-heart-disease [Accessed: Apr. 20, 2025].

[29] Kaggle, "CXR data for multiclass classification," [Online]. Available: https://www.kaggle.com/datasets/newra008/cxr-data-for-multiclass-classification

[30] M. Pal et al., "Ensemble approach of deep learning models for binary and multiclass classification of histopathological images for breast cancer," *Pathol. Res. Pract.*, vol. 263, p. 155644, 2024, doi: 10.1016/j.prp.2023.155644.

[31] A. Vaswani et al., "Attention is all you need," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017.

[32] Keras, "ConvMixer for image classification," [Online]. Available: https://keras.io/examples/vision/convmixer/

[33] M. Pal, S. Parija, G. Panda, S. Mishra, R. K. Mohapatra, and K. Dhama, "COVID-19 Prognosis from Chest X-ray Images by using Deep Learning Approaches: A Next Generation Diagnostic Tool," *J. Pure Appl. Microbiol.*, vol. 17, no. 2, 2023.

[34] Padmavathi, V., & Ganesan, K. (2025). Metaheuristic optimizers integrated with vision transformer model for severity detection and classification via multimodal COVID-19 images. *Scientific Reports*, *15*(1), 13941.

[35] Wang, L., Fang, S., Chen, X., Pan, C., & Meng, M. (2025). Comparative analysis of convolutional neural networks and vision transformers in identifying benign and malignant breast lesions. *Medicine*, *104*(23), e42683.

[36] Y. S. Lee, H. T. Chung, J. J. Lin, M. S. Hwang, H. C. Liu, H. M. Hsu, Y. T. Chang, and S. J. Peng, "Prediction of significant congenital heart disease in infants and children using continuous wavelet transform and deep convolutional neural network with 12-lead electrocardiogram," *BMC Pediatrics*, vol. 25, no. 1, p. 324, Apr. 2025, doi: 10.1186/s12887-025-05628-2.

[37] A. A. AlZubi and A. Alkhanifer, "Study on the prediction of congenital cardiac abnormalities using various machine learning models," *Journal of King Saud University – Science*, vol. 36, no. 11, p. 103555, 2024, doi: 10.1016/j.jksus.2024.103555.

[38] T. Rahman, M. K. A. A. Al-Ruweidi, M. S. I. Sumon, R. Y. Kamal, M. E. H. Chowdhury, and H. C. Yalcin, "Deep learning technique for congenital heart disease detection using stacking-based CNN-LSTM models from fetal echocardiogram: A pilot study," *IEEE Access*, vol. 11, pp. 110375–110390, 2023, doi: 10.1109/ACCESS.2023.3316719.

[39]Liu Y, Huang Q, Han X, Liang T, Zhang Z, Lu X, Dong B, Yuan J, Wang Y, Hu M, Wang J, Stefanidis A, Su J, Chen J, Li Q, Zhang Y. *Atrial Septal Defect Detection in Children Based on Ultrasound Video Using Multiple Instances Learning***.** Journal of Imaging Informatics in Medicine. 2024. DOI: 10.1007/s10278-024-00987**-1**.

[40] D. Das, K. Santosh, and U. Pal, "Cross-population train/test deep learning model: Abnormality screening in chest x-rays," in *Proc. 2020 IEEE 33rd Int. Symp. Computer-Based Medical Systems (CBMS)*, 2020, pp. 514–519, doi: 10.1109/CBMS49503.2020.00101.

[41] C. Sitaula and M. B. Hossain, "Attention-based VGG-16 model for COVID-19 chest x-ray image classification," *Applied Intelligence*, vol. 51, pp. 2850–2863, 2021, doi: 10.1007/s10489-020-02055-x.

[42] A. H. Al-Timemy, R. N. Khushaba, Z. M. Mosa, and J. Escudero, "An efficient mixture of deep and machine learning models for COVID-19 and tuberculosis detection using x-ray images in resource limited settings," in *Artificial Intelligence for COVID-19*, 2021, pp. 77–100, doi: 10.1016/B978-0-323-85780-2.00014-8.

[43] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with x-ray images," *Computers in Biology and Medicine*, vol. 121, p. 103792, 2020, doi: 10.1016/j.compbiomed.2020.103792.

[44] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest x-ray images based on the concatenation of Xception and ResNet50v2," *Informatics in Medicine Unlocked*, vol. 19, p. 100360, 2020, doi: 10.1016/j.imu.2020.100360.

[45] A. Makkar and K. Santosh, "Securefed: Federated learning empowered medical imaging technique to analyze lung abnormalities in chest x-rays," *International Journal of Machine Learning and Cybernetics*, pp. 1–12, 2023, doi: 10.1007/s13042-023-02047-3.