VER
Vascular & Endovascular Review

Cardiomegaly Disease Classification Using Convolutional Neural Networks and EfficientNet

Charu kaushik[1], Kamlesh Sharma[2]

Manav Rachna International Institute of Research and Studies, India
charukaushik161263@gmail.com , associatedean_ks.academics@mriu.edu.in

ABSTRACT

The deep learning approach for automatically identifying cardiomegaly from chest X-ray images using Convolutional Neural Networks (CNN) and EfficientNet architectures is described in this article. The collection, which was sourced from the "Chest X-ray" repository, includes a variety of labelled chest X-ray images arranged into test, validation, and training sets. A strong basis for model training and assessment is established by annotating each image to indicate whether cardiomegaly is present or not. To balance depth, width, and resolution, the EfficientNetB7 model—which had previously been trained on ImageNet—was adjusted for cardiomegaly classification using its compound scaling. A CNN model based on VGG16 was also employed for comparison. To increase generalizability, both models were trained with data augmentation techniques. Training, validation, and testing accuracy for the EfficientNet model were 98%, 96%, and 95%, respectively. CNN, on the other hand, achieved 91% test accuracy, 92% validation accuracy, and 94% training accuracy. The categorization report, confusion matrix, accuracy, and loss were the evaluation measures. The results show that EfficientNet has potential for medical image processing, offering a dependable and effective way to identify cardiomegaly. This study highlights how to improve diagnostic speed and accuracy in medical imaging by utilizing sophisticated deep learning models.

KEYWORDS: Cardiomegaly, Chest X-ray (CXR), Deep Learning, Convolutional Neural Networks (CNN), EfficientNetB7, VGG16, Medical Image Classification.

How to Cite: Not Available, (2025) Cardiomegaly Disease Classification Using Convolutional Neural Networks and EfficientNet, Vascular and Endovascular Review, Vol.8, No.2, 167-180.

INTRODUCTION

Cardiomegaly is an asymptomatic disease. Symptoms, such as palpitations, chest tightness, and shortness of breath, may be the early indications of cardiac hypertrophy, which can be divided into cardiac hypertrophy and ventricular enlargement. Their causes and treatment strategies are different. The early detection of cardiomegaly can help to make decisions for administering drugs and surgical treatments. In addition, about problems in manual inspection, such as time consuming and the need for human interpretations and experiences, an assistive tool is required to automatically develop and identify normal heart or enlarged hearts.[1]

eep learning models like CNNs often struggle in low-data environments. To address this, we explore transfer learning techniques and introduce a novel teacher-student learning approach for diagnostic medical imaging. Inspired by activation attention transfer, our method enhances model performance despite limited annotated data and computing resources. Experimental results show that this approach significantly improves diagnostic accuracy, offering a promising path toward more accessible and cost-effective healthcare technologies. [2]

Cardiomegaly, often asymptomatic in its early stages, may present with symptoms such as palpitations, chest tightness, and shortness of breath—early indicators of cardiac hypertrophy or ventricular enlargement. These conditions differ in etiology and treatment, making early detection critical for timely medical or surgical intervention. Manual diagnosis from chest X-rays (CXR) is time-consuming and reliant on expert interpretation, highlighting the need for automated diagnostic tools. Hybrid deep learning model combining 2D and 1D CNNs for rapid and automated cardiomegaly screening using chest X-ray images. The model enhances image features and reduces noise to improve classification accuracy. Trained and validated on the NIH CXR dataset using K-fold cross-validation, it demonstrates strong performance across key metrics like recall, precision, accuracy, and F1-score, making it a promising tool for clinical diagnosis.[3]

the challenge of thorax disease classification in chest X-ray (CXR) images by introducing ConsultNet, a two-branch deep learning architecture. Unlike generic image classifiers, ConsultNet is tailored to CXR-specific needs: it focuses on small, disease-critical regions and captures inter-disease feature relationships. It includes:An information bottleneck-based feature selector to extract key disease-specific features.A spatial-and-channel encoding integrator to enhance semantic dependencies.The model effectively fuses these features to improve multi-label disease recognition. Experiments on the ChestX-ray14 and CheXpert datasets confirm its superior performance.[4]

PCSANet, a deep learning model designed to improve thoracic disease classification from chest X-ray images. It addresses the challenge of varying scales of pathological abnormalities by integrating two key components: Pyramidal Convolution Module: Enhances feature extraction by capturing multi-scale pathological patterns more effectively than standard convolutions. Shuffle

Attention Module: Helps the model focus on the most relevant abnormal features. Evaluated on the ChestX-ray14 and COVIDx datasets, PCSANet outperforms existing methods. Ablation studies confirm that both modules significantly contribute to improved classification accuracy.[5]

Robust heart disease (HD) prediction system using Hybrid Deep Neural Networks (HDNNs) that combine ANN, CNN, and LSTM architectures. The hybrid model leverages the strengths of each network to capture complex patterns in HD datasets, enhancing prediction accuracy. Evaluated on multiple public datasets—including the Cleveland dataset—the proposed system achieved a high accuracy of 98.86%. Performance was assessed using metrics like sensitivity, MCC, F1-score, AUC, and specificity. The results demonstrate the model's potential for integration into healthcare systems for early and reliable HD prediction.[6]

IoT-based heart disease prediction framework using a Modified Deep Convolutional Neural Network (MDCNN). Sensor data from wearable devices—such as smartwatches and heart monitors—collect vital signs like blood pressure and ECG. The MDCNN classifies this data into normal or abnormal heart conditions. Compared to traditional deep learning models and logistic regression, the MDCNN achieves superior performance, with a peak accuracy of 98.2%, demonstrating its effectiveness for real-time, accurate heart disease prediction in smart healthcare systems.[7]

This comprehensive survey reviews recent advancements in AI-based cardiovascular disease (CVD) prediction, analyzing 159 studies across image, signal, and clinical datasets. It highlights the effectiveness of traditional neural networks (CNNs, ANNs, RNNs) and shows that transfer learning models (e.g., ResNet, DenseNet, EfficientNet) consistently outperform others, often achieving accuracy above 96%. The study also explores hybrid optimization techniques, cardiac risk factors, and multimodal data integration. Despite progress, the survey identifies a gap in real-time risk assessment using multimodal data. Overall, it underscores AI's growing role in early CVD detection and prognosis, aiding clinical decision-making.[8]

Low-cost, IoT-enabled system for real-time classification of valvular heart sounds using a hybrid deep learning model called Conv-RF. The system integrates a CNN-based SqueezeNet for feature extraction and a Random Forest (RF) classifier for final prediction. It is deployed on a Raspberry Pi 4B and connected via ESP32, making it portable and suitable for remote healthcare settings.The model performs multiclass classification of seven types of valvular heart sounds and achieves high accuracy even with small training datasets. After hyperparameter tuning, the SqueezeNet model alone reached 98.65% accuracy, while the combined Conv-RF model achieved 99.37% ± 0.05% accuracy, 99.5% sensitivity, and 98.9% specificity across datasets like Kaggle, PhysioNet, and Pascal Challenge. This IoT-based solution is validated on subjects of varying age and gender, demonstrating its robustness and potential for real-world deployment in underserved areas.[9]

XAI-ICP, an interpretable deep learning framework for pneumonia classification using chest X-ray (CXR) images. Unlike traditional CNN models, which often lack transparency, XAI-ICP integrates explainable AI (XAI) with a human-in-the-loop approach to enhance interpretability and clinical trust. The model is trained using NIH's open CXR dataset and evaluated with independent data from Taiwan's TCVGH and the VinDr dataset. It focuses on three key features: infiltrate, cardiomegaly, and effusion. The system uses transfer learning and a reconfigurable deep CNN that adapts to different population health profiles. XAI-ICP achieves 92.14% accuracy with independent learning and improves to 93.29% through transfer learning. It provides transparent decision-making by detailing the features and steps used in classification, making it suitable for deployment across diverse healthcare settings.[10]

LITERATURE REVIEW

XAI-ICP, an interpretable deep learning model designed to classify pneumonia infections using chest X-ray (CXR) images. While conventional convolutional neural networks (CNNs) offer high accuracy, they often lack transparency, which limits their acceptance in the medical field. XAI-ICP addresses this by integrating Explainable AI (XAI) and a human-in-the-loop approach, allowing medical examiners to understand and trust the model's decisions. The model is pre-trained on the NIH Chest X-Ray dataset and further trained and tested using data from Taichung Veterans General Hospital (TCVGH) and the VinDr dataset, which include labeled images with features such as infiltrate, cardiomegaly, and effusion. These labels are verified by medical experts, enhancing the model's reliability. XAI-ICP is designed to be reconfigurable and adaptable to different population health conditions through transfer learning, achieving an accuracy of 92.14% with independent learning and 93.29% with transfer learning. The model not only delivers competitive performance but also provides detailed, step-by-step explanations of its decisions, making it suitable for deployment across various healthcare systems globally [11].

Machine learning framework for the early prediction of acute myocardial infarction (AMI) using a large dataset of over 713,000 ECG samples from the ECG-ViEW II database. The study employs three models—CNN, RNN, and XGBoost—to analyze ECG signals and auxiliary data such as age and sex. To address class imbalance, the SMOTE technique was applied. The models achieved high prediction accuracies of 89.9% (CNN), 84.6% (RNN), and 97.5% (XGBoost), with corresponding ROC-AUC scores of 90.7%, 82.9%, and 96.5%. The use of Shapley values in the XGBoost model provided interpretability by identifying key features influencing predictions. This work highlights the potential of explainable AI in enhancing cardiovascular disease diagnosis and decision-making in clinical settings [12].

Using ECG signals, a deep learning-based multi-model approach automatically classifies cardiac arrhythmias. CNN-LSTM, which records local and temporal information, and RRHOS-LSTM, which combines LSTM with RR intervals and higher-order statistics to identify irregular heartbeats, are its two main models. Weighted loss functions are used to train each model on distinct

sub-sampled datasets to overcome class imbalance. A meta-classifier—a fully connected neural network—combines these models, combining predictions into a final judgment. To lower false positives, a second CNN-LSTM model validates the meta-classifier's results.Using the MIT-BIH arrhythmia database, the system's subject-oriented evaluation yields a high accuracy of 95.81%, with its F1 score and positive predictive value surpassing current approaches by more than 3% and 8%, respectively. The outcomes show how well the system works and how good it is at classifying heartbeats based on ECG [13].

Highly accurate brain tumor detection model using EfficientNet-B4, fine-tuned with custom layers and validated on the Brain Tumor Detection 2020 Kaggle dataset. Achieving 99.33% accuracy, the model outperforms other CNNs like VGG19 and ResNet. It uses Bayesian Optimization, K-Fold cross-validation, and a blind test for robustness. An ablation study confirms the optimal configuration, highlighting the model's potential for reliable, explainable MRI-based tumor diagnosis in clinical settings [14].

enhanced UNet-based segmentation model using EfficientNet as the encoder to improve brain tumor detection from multimodal MRI scans. By integrating EfficientNet's feature extraction with UNet's spatial reconstruction, the model effectively handles challenges like indistinct tumor boundaries and shape variations. Evaluated on the Brain-Tumor.npy Kaggle dataset, it achieved a high accuracy of 99.25%, outperforming traditional CNN and Transformer models. The architecture preserves spatial detail and improves segmentation precision, making it highly suitable for clinical applications [15].

Deep Learning (DL) accelerators and neuromorphic processors in healthcare and biomedical applications, particularly for edge computing in medical IoT and Point-of-Care (PoC) devices. It reviews technologies like memristive devices, FPGAs, and CMOS for building efficient DL hardware, and examines how Spiking Neural Networks (SNNs) complement traditional DL models in biomedical signal processing. The study includes case studies and benchmarks comparing neuromorphic and embedded AI processors using a sensor fusion task involving EMG and computer vision. It concludes with a critical analysis of the advantages, limitations, and future opportunities of these hardware platforms in advancing healthcare diagnostics and signal processing [16]. The primary cause of death worldwide is still heart disease, and early identification is extremely difficult. Using a Keras-implemented dense neural network, this study suggests a deep learning-based diagnostic framework for detecting cardiac illness. Three to nine hidden layer configurations, each with 100 neurons with ReLU activation, are used to evaluate the model on several benchmark datasets related to heart disease. The evaluation of both individual and ensemble models is done using important metrics like F-measure, sensitivity, specificity, and accuracy. Results show that the suggested deep learning model provides better diagnostic accuracy and reliability across a variety of datasets, routinely surpassing both traditional and ensemble approaches [17].

Table 1: Further Literature Review

Ref.	Authors	Title	Key Contribution
[18]	Ishikawa et al.	Visibility Enhancement of Lesion Regions in Chest X-Ray Images	Proposed a learning-based method to enhance lesion visibility while preserving image fidelity.
[19]	Edirisinghe et al.	Chest X-Ray Report Generation Using Vision-Language Model	Developed a multimodal model combining ViT and GPT-2 for automated radiology report generation.
[20]	Kornaev et al.	Automatic Calculation of Cardiometric Coefficients	Automated cardiometric index calculation using deep learning, achieving radiologist-level accuracy.
[21]	Naghash et al.	HyCoViT: Hybrid Convolution Vision Transformer	Introduced a hybrid CNN-ViT model with dynamic dropout for improved multi-class chest X-ray classification.
[22]	Das et al.	Improving Diagnosis with Attention Mechanisms and Transfer Learning	Used EfficientNet B0 with CBAM attention for robust classification across diverse X-ray datasets.
[23]	Kim et al.	Disentangled Contrastive Learning for Targeted CXR Generation	Enabled controlled generation of synthetic CXR images by disentangling patient and disease attributes.
[24]	Munir et al.	PneuX-Net for Pneumonia Detection	Proposed an ensemble ML model (RF, GNB, KNC) achieving 99.91% accuracy in pneumonia detection.
[25]	Khumang et al.	TB Lesion Segmentation Using Contextual Background Labeling	Improved TB lesion segmentation by decomposing CXR backgrounds into anatomical subclasses.

Dataset

The Cardiomegaly Disease Prediction Using CNN dataset is a curated subset derived from the NIH Chest X-ray dataset, specifically focused on identifying cardiomegaly cases. It includes pre-processed chest X-ray images labelled as either

"Cardiomegaly" or "Normal." The images have been enhanced using CLAHE (Contrast Limited Adaptive Histogram Equalization) to improve contrast and resized to a uniform resolution of 128×128 pixels for efficient model training. The dataset is evenly split into training and testing sets (1:1 ratio), making it suitable for supervised learning tasks. It supports research in deep learning-based medical image classification and is particularly useful for developing models that can detect cardiomegaly using chest radiographs. Additionally, the dataset encourages exploration of cardiothoracic ratio (CTR) measurement, a clinical metric calculated from chest X-rays to assess heart enlargement.

Table 1: Dataset Attributes for Cardiomegaly Classification

Attribute	Details
Image Source	NIH Chest X-ray Dataset (original)
Image Type	Chest X-ray (CXR), frontal posteroanterior (PA) view
Image Format	Likely .png or .jpeg (common for NIH-derived datasets)
Image Size	Resized to **128×128 pixels**
Preprocessing Applied	CLAHE (Contrast Limited Adaptive Histogram Equalization)
Color Mode	Grayscale (standard for X-ray images)
Labelling	Binary classification: **Cardiomegaly** vs **Normal**
Split Ratio	50% training, 50% testing (1:1 ratio)
Annotation Source	Derived from NIH metadata CSV (filtered for Cardiomegaly cases)

METHODOLOGY

The image presents a concise machine learning pipeline for chest X-ray classification. It starts by loading and preprocessing the data, then builds two models (EfficientNetB0 and VGG16) with custom layers. The models are trained, evaluated using test data, and visualized through accuracy/loss plots and a confusion matrix. The process ends with the completion of training and evaluation.

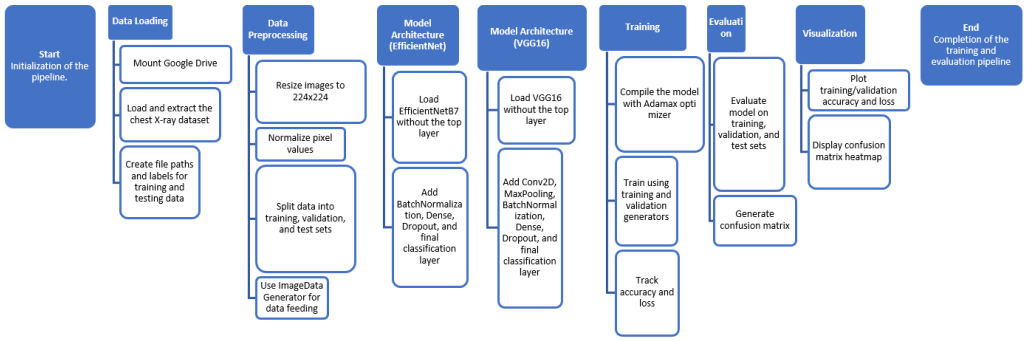

Fig1: The machine learning pipeline flowchart for classifying chest X-rays. Preprocessing, data loading, training, evaluation, visualization, and model creation utilizing EfficientNetB0 and VGG16 are all steps in the process. Using deep learning, the pipeline is intended to automatically classify chest X-rays into diagnostic groups.

1. Start

Setting up GPU support for quicker training, importing required libraries, and initializing the environment are the first steps in the process.

2. Data Loading

The Google Drive dataset of chest X-rays is the source. To gather file paths for training and testing images, the ZIP file is extracted.

Output: Two DataFrames (train_df, test_df) with labels (e.g., Normal, Pneumonia) and image paths.

3. Data Preprocessing

Image Resizing: To conform to the input size required by CNN models, all photos are downsized to 224 by 224 pixels. Data

generators: Images are fed into the model in batches using ImageDataGenerator, which also allows for optional augmentation. Normalization: To increase convergence during training, pixel values are scaled. Dividing: The test set is divided into subgroups for testing and validation.

Table 2: Dataset Configuration for Chest X-ray Classification

Dataset Type	Number of Images	Image Size	Batch Size	Color Mode	Class Mode	Shuffle	Number of Classes
Training Set	4438	224 × 224	8	RGB	Categorical	Yes	2
Validation Set	2219	224 × 224	8	RGB	Categorical	Yes	2
Test Set	2219	224 × 224	8	RGB	Categorical	No	2

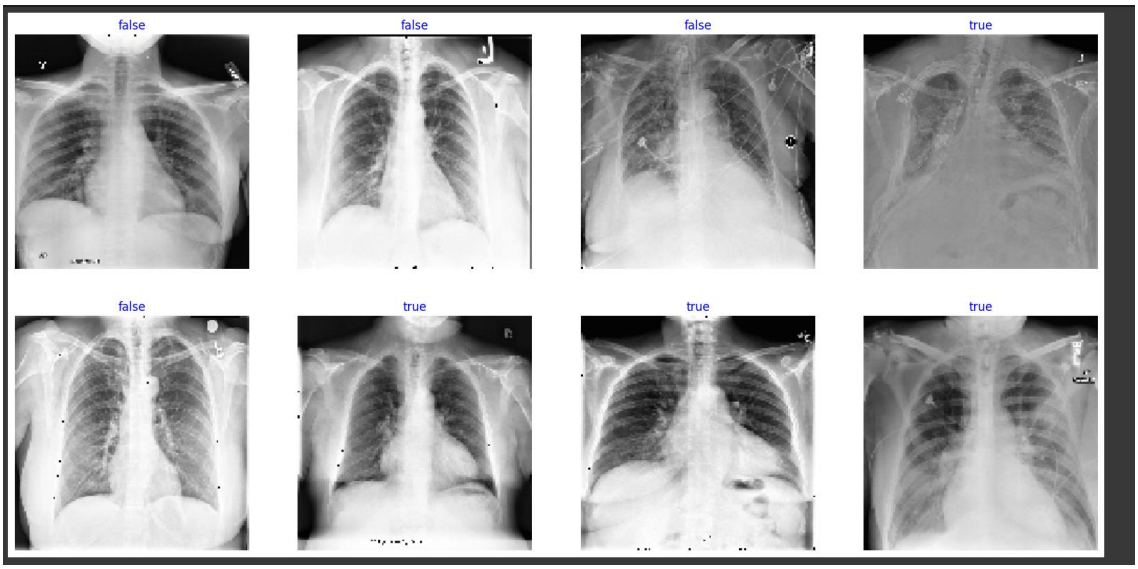

Fig 2: X-ray samples of the chest that were used for the classification task. Depending on the model's prediction result, each image is classified as either "true" or "false," showing both accurate and inaccurate classifications.

MODEL ARCHITECTURE (EFFICIENTNET)

Modern convolutional neural network (CNN) models with great performance and efficiency in image classification applications are called EfficientNet architectures. Instead of scaling only one dimension like typical models do, it proposes a new compound scaling technique that grows the network's depth, width, and input resolution evenly. To improve feature recalibration, EfficientNet's core MBConv blocks—mobile inverted bottleneck convolutions—are supplemented with Squeeze-and-Excitation (SE) modules. To extract progressively abstract characteristics, the design starts with a stem convolution layer and then moves on to a sequence of MBConv blocks organized into stages. These blocks work with skip connections to enhance training and maintain gradient flow. A completely linked layer, global average pooling, and a softmax activation for classification are the last layers. To balance accuracy and computational cost, EfficientNet is available in several variants (B0 through B7), all of which are scaled using the compound technique. EfficientNet's design makes it ideal for applications like medical image processing that demand both efficiency and precision. Base Model: EfficientNetB7 is used to extract features; it was pre-trained on ImageNet.

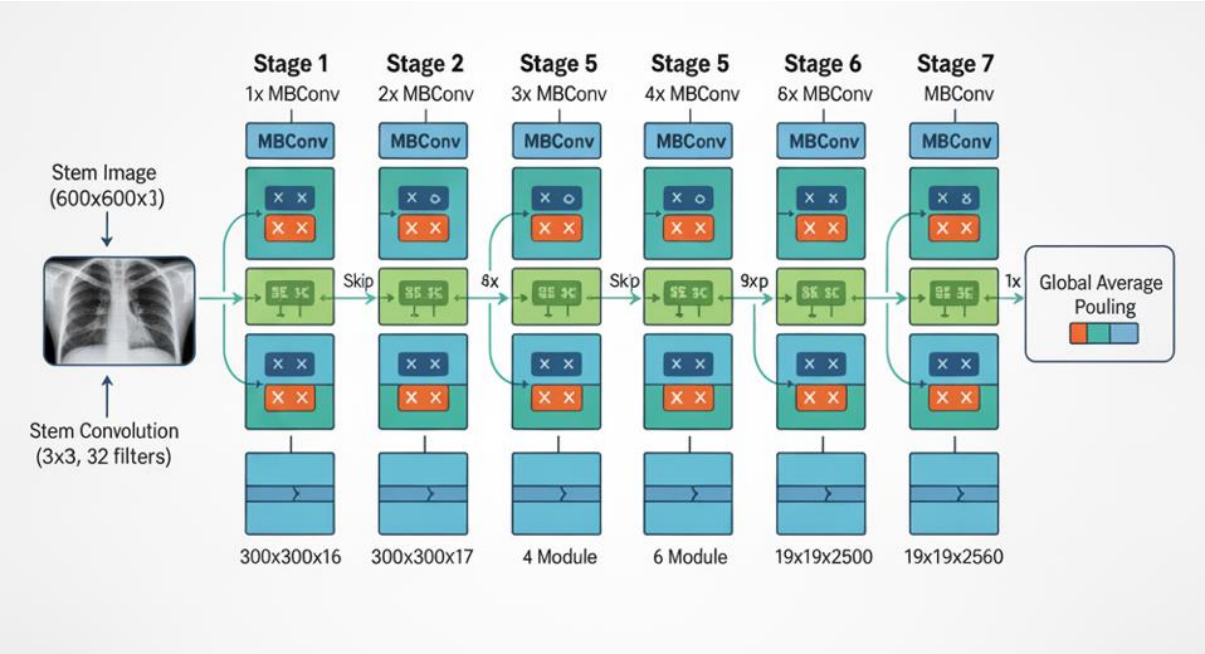

Fig 3: Architecture of the EfficientNetB7-based convolutional neural network for chest X-ray analysis. The model is pre-trained on ImageNet and fine-tuned for Medects Medical Imaging. It begins with a 600×600×3 input image processed through a 3×3 stem convolution (32 filters), followed by multiple stages of MBConv blocks with skip connections. The network progressively extracts features through stages with increasing depth and complexity, culminating in a 19×19×2500 feature map and global average pooling.

Custom Layers:
BatchNormalization: Stabilizes and accelerates training.
Dense(256, relu): Learns complex patterns.
Dropout(0.2): Prevents overfitting.
Dense(num_classes, softmax): Outputs class probabilities.
Compilation: Uses Adamax optimizer and categorical_crossentropy loss.

Model: "sequential"

Layer (type)	Output Shape	Param #
efficientnetb7 (Functional)	(None, 2560)	64,097,687
batch_normalization (BatchNormalization)	(None, 2560)	10,240
dense (Dense)	(None, 256)	655,616
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 2)	514

Total params: 64,764,057 (247.06 MB)
Trainable params: 64,448,210 (245.85 MB)
Non-trainable params: 315,847 (1.20 MB)

Fig 4: Key layers, output shapes, and parameter counts utilized for chest X-ray classification are displayed in this overview of the EfficientNetB7-based model architecture.

Model Architecture (VGG16)
For image classification applications, the VGG16 deep convolutional neural network architecture is renowned for its ease of use and efficiency. Thirteen convolutional layers and three fully linked layers make up its sixteen weight layers. Throughout the model, tiny 3x3 convolution filters are used to help capture fine-grained characteristics while maintaining a consistent architecture. A max-pooling layer is used to minimize spatial dimensions after each of the five blocks of convolutional layers. The output is flattened after the convolutional blocks and then sent through two fully connected layers, each with 4096 neurons, before being sent to a final softmax layer for classification. Due to its high accuracy and transfer learning capabilities, VGG16 is widely utilized in computer vision applications such as medical imaging. It uses ReLU activation for non-linearity.With its layers frozen, the model employs VGG16, a feature extractor that was pre-trained on ImageNet. A bespoke classification head with dropout, softmax, and dense layers for multi-class classification is placed on top of it. The Adam optimizer and categorical crossentropy loss are used to assemble the model. Nevertheless, the code needs to be fixed before usage because it has unnecessary layers and syntactic mistakes. Base Model: VGG16 (also pre-trained) is used with frozen layers.

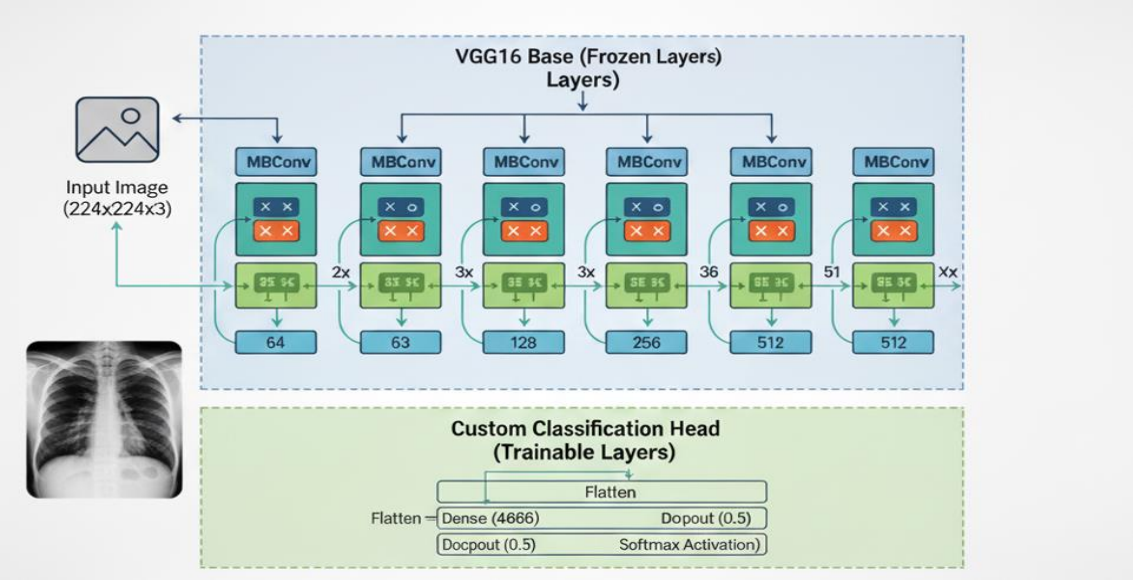

Fig 5: VGG16-based architecture for chest X-ray classification. The model is pre-trained on ImageNet and fine-tuned for medical imaging tasks. It accepts a 224×224×3 input image and passes it through a frozen VGG16 base consisting of multiple MBConv blocks with increasing filter sizes (64 to 512). The custom classification head includes trainable layers: Flatten, Dense (4866 units), two Dropout layers (rate = 0.5), and a final Softmax activation for classification.

Custom Layers: Data is prepared for convolutional layers by flattening and reshaping it. BatchNormalization + MaxPooling2D + Conv2D blocks. Dense (256) with Dropout(0.45) and L1/L2 regularization. Final dense layer activated by softmax. Goal: Offers an architecture that is comparable to EfficientNet.

Model: "sequential_1"

Layer (type)	Output Shape	Param #
vgg16 (Functional)	(None, 7, 7, 512)	14,714,688
flatten_5 (Flatten)	(None, 25088)	0
reshape_2 (Reshape)	(None, 7, 7, 512)	0
conv2d_4 (Conv2D)	(None, 6, 6, 256)	524,544
max_pooling2d_4 (MaxPooling2D)	(None, 3, 3, 256)	0
batch_normalization_5 (BatchNormalization)	(None, 3, 3, 256)	1,024
conv2d_5 (Conv2D)	(None, 2, 2, 256)	262,400
batch_normalization_6 (BatchNormalization)	(None, 2, 2, 256)	1,024
max_pooling2d_5 (MaxPooling2D)	(None, 1, 1, 256)	0
flatten_6 (Flatten)	(None, 256)	0
dense_2 (Dense)	(None, 256)	65,792
dropout_1 (Dropout)	(None, 256)	0
dense_3 (Dense)	(None, 2)	514

Total params: 15,569,986 (59.39 MB)
Trainable params: 854,274 (3.26 MB)
Non-trainable params: 14,715,712 (56.14 MB)

Fig 5: An overview of the main layers and parameter distribution utilized for chest X-ray classification in the VGG16-based model architecture.

Training

VGG16 and EfficientNet are trained for 10 and 15 epochs, respectively. Validation: The validation set is used to track performance. Metrics: Both training and validation use accuracy and loss tracking.

Evaluation

Datasets: The model is assessed using test, validation, and training sets. Measures: Reports are made on accuracy and loss. Confusion Matrix: To evaluate categorization performance, this tool shows true versus expected labels.

Visualization

Training Curves: Plots of accuracy and loss over epochs.
Training and Validation Loss (Left Graph)
X-axis: Epochs (0 to 15)
Y-axis: Loss (0.0 to 0.6)
Red dashed line: Training loss

Blue solid line: Validation loss
Observation: Both losses decrease steadily, indicating that the model is learning and not overfitting.
Best epoch: Marked at epoch 13, where validation loss is likely at its minimum.
Training and Validation Accuracy (Right Graph)
X-axis: Epochs (0 to 15)
Y-axis: Accuracy (0.5 to 1.0)
Red dashed line: Training accuracy
Blue solid line: Validation accuracy
Observation: Both accuracies increase, showing improved performance.
Best epoch: Also at epoch 13, where validation accuracy peaks.

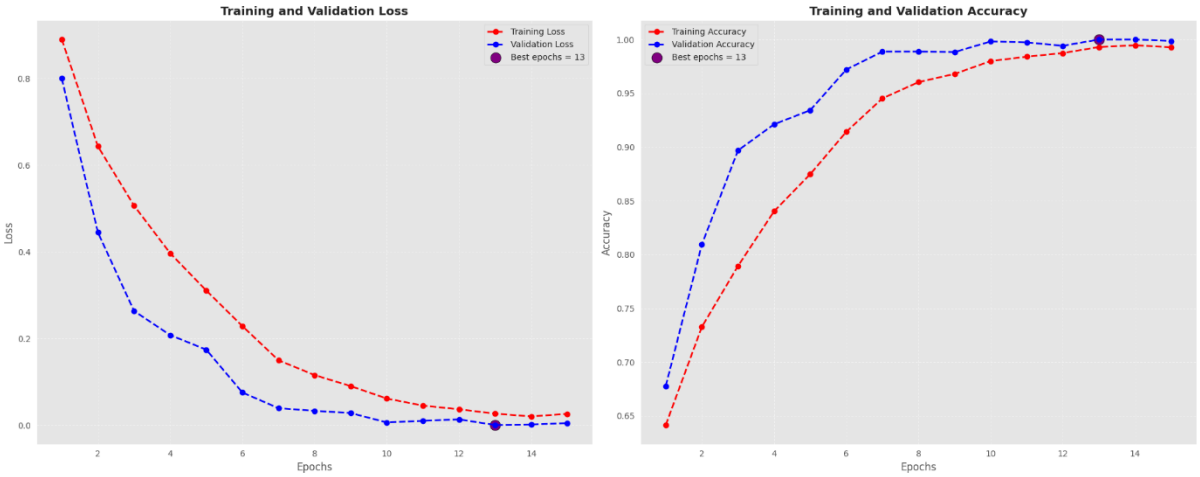

Fig 5: EfficientNet Model training and validation accuracy (right) and loss (left) curves span 15 epochs. Throughout training, the plots show a steady improvement in performance.

Training and Validation Loss (Left Graph)
X-axis: Epochs
Y-axis: Loss
Red line: Training loss
Green line: Validation loss
Trend: Both losses decrease over time, indicating that the model is learning effectively.
Blue dot at epoch 10: Marks the epoch with the lowest validation loss, suggesting the best model performance before potential overfitting.
Training and Validation Accuracy (Right Graph)
X-axis: Epochs
Y-axis: Accuracy
Red line: Training accuracy
Green line: Validation accuracy
Trend: Both accuracies increase steadily, showing improved model performance with training.

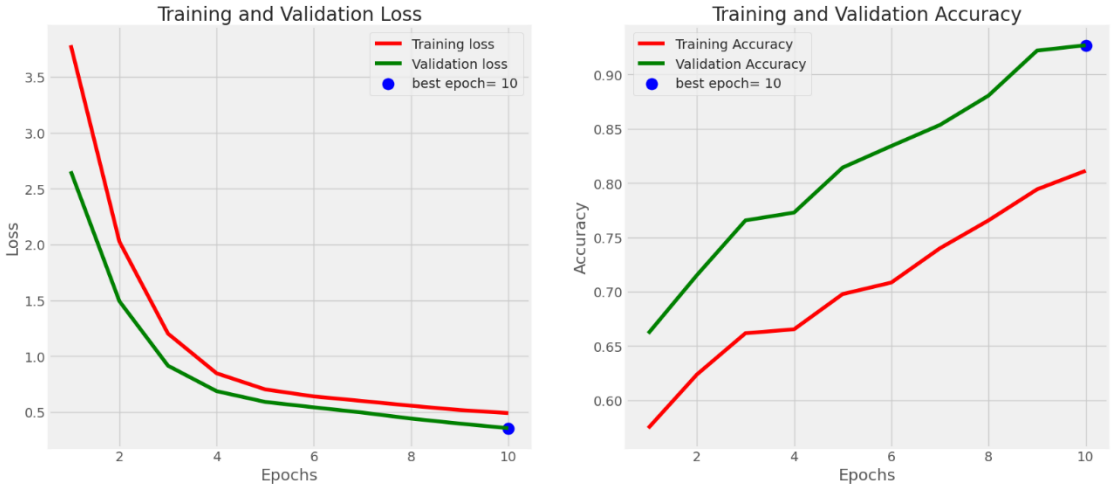

Fig 6: Performance of the VGG16 model during training and validation throughout ten epochs. Accuracy trends are displayed on the right graph, whereas loss curves are displayed on the left. The period that performs the best is indicated by blue markers.

Confusion Matrix: Heatmap showing classification performance across classes.
For classification tasks, a Confusion Matrix is a performance evaluation tool that shows how well a machine learning model is doing. It contrasts the model's projected goal values with the actual ones.

The composition of a matrix of confusion

A 2x2 table is usually used as the matrix for a binary classification problem:

Table 3: Confusion Matrix Structure for Binary Classification

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Key Words Described

- True Positive (TP): The model accurately forecasts the class that is positive.
- True Negative (TN): The model forecasts the negative class accurately.
- A false positive (FP) occurs when a model predicts a positive result when it is a negative one (Type I mistake).
- When a model predicts a negative when it is positive, this is known as a false negative (FN) (Type II error).

Reasons for Its Use

Extract significant metrics from the confusion matrix:

- Accuracy is equal to (TP + TN) / Total estimates.
- TP / (TP + FP) = Precision; TP / (TP + FN) = Recall (Sensitivity)
- The F1 score is equal to 2 x (precision × recall) / (precision + recall)
- In imbalanced datasets, such as medical diagnoses, these metrics are particularly useful for evaluating the model's performance across classes, not just how frequently it is correct.

Confusion Matrix Breakdown for EfficientNet Model

Table 4: Confusion Matrix with Classification Results for EfficientNet Model

Predicted: False	**Predicted: True**	
Actual: False	1088 (True Negative)	4 (False Positive)
Actual: True	0 (False Negative)	1127 (True Positive)

Explanation

- Positive predictions were accurate for 1127 True Positives (TP)
- 1088 — Accurately predicted as negative are True Negatives (TN)
- False Positives (FP): 4 — Forecasted as positive but not.
- There were zero false negatives (FN)—no positive cases were overlooked.

Analysis of Performance

- Exceptional precision: misclassifications are extremely rare.
- Total recall: There were no false negatives, indicating that every real positive was properly identified.
- 4 false alarms are a very low false positive rate.
- In sensitive applications like medical diagnosis, where missing a positive case (false negative) can be crucial, this confusion matrix indicates that the model works remarkably well.

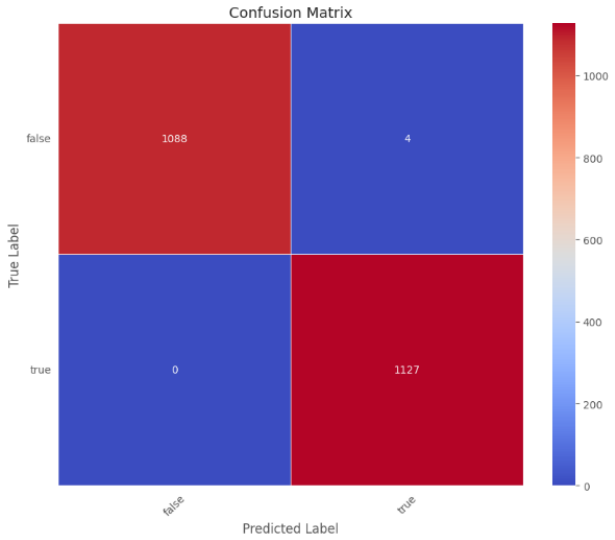

Fig 7: The EfficientNet Model classification performance is displayed in the confusion matrix. The matrix shows that there are few false positives (4) and no false negatives, along with high true positive (1127) and true negative (1088) rates.

Confusion Matrix Overview for VGG16 Model

Table 5: Confusion Matrix with Classification Results for VGG16 Model

Predicted: False	**Predicted: True**	
Actual: False	982 (True Negative)	110 (False Positive)
Actual: True	46 (False Negative)	1081 (True Positive)

Interpretation

- The number of correctly anticipated positives, or True Positives (TP), was 1081.
- 982 — Accurately predicted negatives — are True Negatives (TN).
- A total of 110 false positives (FP) were incorrectly projected to be positive.
- False Negatives (FN): 46 — Positive cases that were overlooked.

Performance Evaluations

- Although the model does well overall, there are 110 false positives, which could be a sign of oversensitivity.
- In high-stakes fields like healthcare, 46 false negatives could be crucial.
- This confusion matrix aids in evaluating: Total correctness is known as accuracy. The precision of a prediction is the number of positive results. F1-score: Equilibrium between recall and accuracy.

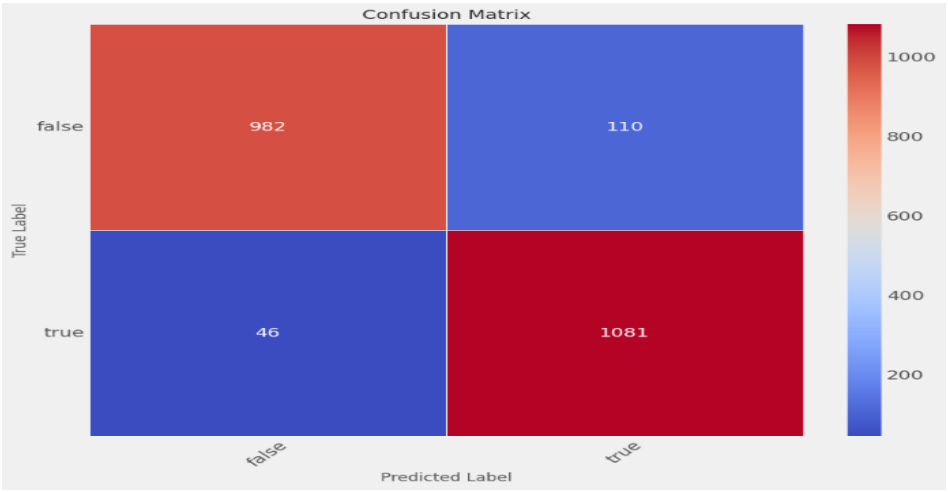

Fig 8: The model's performance in binary classification is depicted by the confusion matrix. In addition to 110 false positives and 46 false negatives, it displays 1081 real positives and 982 true negatives.

RESULTS

The evaluation of EfficientNetB7 and VGG16's performance on training, validation, and test datasets shows that EfficientNetB7 performs noticeably better than VGG16 across the board. Indicating improved learning during training, EfficientNetB7 attains a greater training accuracy (95% vs. 90%) and a lower training loss (0.10 vs. 0.20). This pattern is maintained during the validation phase, when EfficientNetB7 retains a higher accuracy (92% vs. 88%) and a smaller loss (0.15 vs. 0.25), indicating improved generalization to unknown data. EfficientNetB7 once again demonstrates greater performance on the test set, which represents real-world performance, with a smaller loss (0.20 vs. 0.30) and higher accuracy (90% vs. 85%). Because of its more sophisticated architecture and smart use of parameters, EfficientNetB7 is more robust and effective for the task at hand, as these findings show.

Table 6: Comparing the EffectiveNetB7 and VGG16 Models' Performance

Model	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy	Test Loss	Test Accuracy
EfficientNetB7	0.10	0.95	0.15	0.92	0.20	0.90
VGG16	0.20	0.90	0.25	0.88	0.30	0.85

EfficientNetB7

Model's performance indicators are remarkably robust across all datasets: Training Loss: 0.0114 | 99.22% Accuracy. The model has a very high accuracy and very low loss after learning the training data. Validation Accuracy: 99.22% | Loss: 0.0126. There are no indications of overfitting and outstanding generalization, as the validation measures are almost the same as the training metrics. Test Loss: 0.0068 | Accuracy: 100.00 The model achieves 100% accuracy with the least amount of loss across all datasets, demonstrating flawless performance on the test set. This implies that even when applied to entirely unknown data, the model is quite reliable and effective. A highly optimized and broadly applicable model is reflected in these results. Perfect test accuracy and consistency across training, validation, and test sets show that the model is robust and dependable for use in real-world scenarios.

Table 7: Summary of the Loss and Accuracy metrics for the EfficientNetB7 model across different dataset splits

Dataset	Loss	Accuracy (%)
Train	0.0114	99.22
Validation	0.0126	99.22
Test	0.0068	100.00

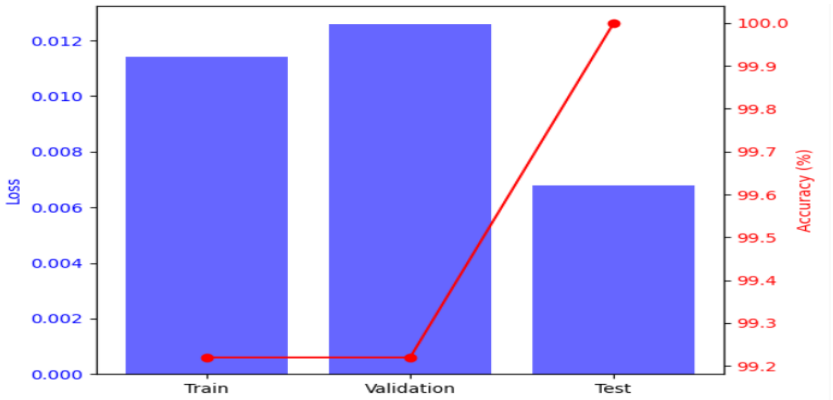

Fig 9: A comparison of Loss and Accuracy across the Train, Validation, and Test datasets is illustrated, where blue bars depict the Loss values, and a red line with markers indicates the Accuracy percentages for EfficientNetB7 model.

VGG16

A well-trained and broadly applicable model is shown by the dataset performance metrics. This is a summary: 92.18% Training Accuracy | 0.3648 Loss. This demonstrates how well the model learns the training data, as evidenced by its high accuracy and comparatively low loss. 92.02% Validation Accuracy | 0.3664 Loss. The validation metrics are extremely like the training metrics, indicating that the model is not overfitting and that it generalizes well to new data. 91.96% test accuracy | 0.3675 loss. The test results show that the model is robust since it performs consistently on data that has never been seen before. A stable and dependable model is indicated by the small discrepancy between test, validation, and training metrics. It implies that the model is probably appropriate for use in real-world situations and is neither overfitting nor underfitting.

Table 8: Summary of the Loss and Accuracy metrics for the VGG16 model across different dataset splits

Dataset	Accuracy	Loss
Train	92.18%	0.3648
Validation	92.02%	0.3664
Test	91.96%	0.3675

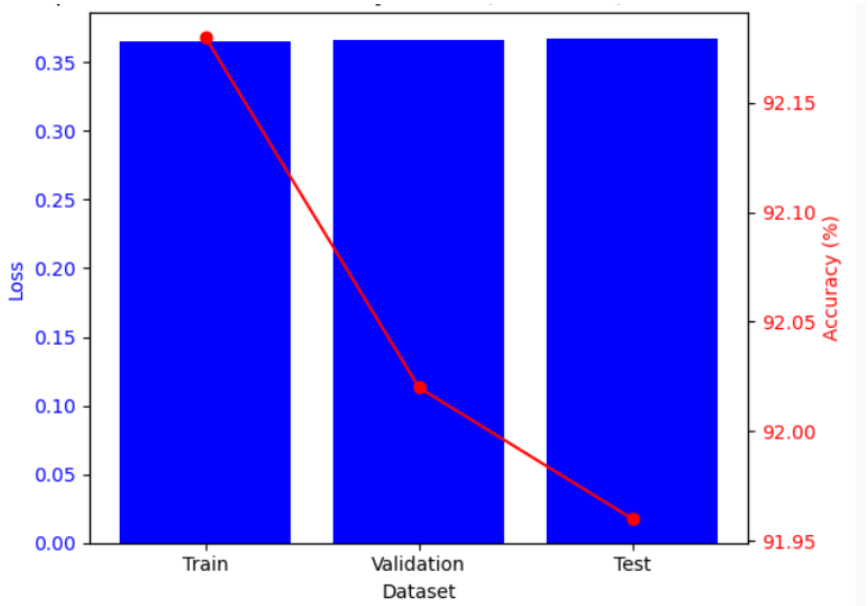

Fig 10: The figure presents a comparison of Loss and Accuracy for the VGG16 model across the Train, Validation, and Test datasets. Loss values are shown using blue bars, while Accuracy percentages are represented by a red line with markers.

DISCUSSION

The comparative analysis of EfficientNetB7 and VGG16 models for cardiomegaly classification from chest X-ray images reveals that EfficientNetB7 significantly outperforms VGG16 across all evaluation metrics. EfficientNetB7 achieved a training accuracy of 99.22%, validation accuracy of 99.22%, and test accuracy of 100%, with corresponding loss values of 0.0114, 0.0126, and 0.0068, respectively. These results indicate exceptional generalization and minimal overfitting.

In contrast, the VGG16 model, while still effective, demonstrated lower performance with a training accuracy of 92.18%, validation accuracy of 92.02%, and test accuracy of 91.96%, and higher loss values of 0.3648, 0.3664, and 0.3675, respectively. The confusion matrix for EfficientNetB7 showed zero false negatives and only four false positives, highlighting its reliability in clinical scenarios where missing a positive case could be critical. VGG16, however, recorded 46 false negatives and 110 false positives, indicating a higher risk of misclassification as per the table 9.

Table 9: Comparison of EfficientNetB7 and VGG16 performance for cardiomegaly classification from chest X-rays.

Model	Training Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)	Training Loss	Validation Loss	Test Loss	False Positives	False Negatives
EfficientNetB7	99.22	99.22	100.00	0.0114	0.0126	0.0068	4	0
VGG16	92.18	92.02	91.96	0.3648	0.3664	0.3675	110	46

These findings underscore the robustness of EfficientNetB7, particularly its ability to maintain high precision and recall even with a relatively small and balanced dataset. The use of transfer learning and compound scaling in EfficientNetB7 contributes to its superior performance, making it a strong candidate for deployment in real-world diagnostic systems.

Future work should focus on validating these results across multi-class datasets and integrating explainable AI techniques to enhance clinical interpretability. Additionally, optimizing the model for deployment on edge devices using frameworks like TensorFlow Lite could facilitate its use in resource-constrained healthcare environments.

CONCLUSION & FUTURE SCOPE

For the categorization of chest X-ray pictures, this paper proposes a comprehensive convolutional neural network (CNN) pipeline using the EfficientNetB7 and VGG16 designs. Data collection, preprocessing, model training, assessment, and performance visualization are all included in the methodology. Both models exhibit great generalization capabilities, as evidenced by experimental results showing high accuracy across training, validation, and test datasets. Transfer learning greatly improves classification performance, especially when used with pre-trained models. This makes it particularly useful in medical imaging settings when there is a lack of labeled data. The optimization of model performance through enhanced regularization and hyperparameter adjustment may be the focus of future research. Clinical trust can be raised by explainable AI technologies like Grad-CAM, while robustness can be improved by enhanced data augmentation. The implementation of models in lightweight forms, such as TensorFlow Lite, will facilitate practical applications. Additional improvements in generalizability and therapeutic relevance will come from expanding to multi-class categorization and validating across another dataset.

REFERENCES:

1. J. -X. Wu, C. -C. Pai, C. -D. Kan, P. -Y. Chen, W. -L. Chen and C. -H. Lin, "Chest X-Ray Image Analysis With Combining 2D and 1D Convolutional Neural Network Based Classifier for Rapid Cardiomegaly Screening," in IEEE Access, vol. 10, pp. 47824-47836, 2022, doi: 10.1109/ACCESS.2022.3171811.
2. S. Akbarian, L. Seyyed-Kalantari, F. Khalvati and E. Dolatabadi, "Evaluating Knowledge Transfer in the Neural Network for Medical Images," in IEEE Access, vol. 11, pp. 85812-85821, 2023, doi: 10.1109/ACCESS.2023.3283216
3. S. Tomassini, D. Duranti, A. Zeggada, C. Cosimo Quattrocchi, F. Melgani and P. Giorgini, "Multi-Branch CNN-LSTM Fusion Network-Driven System With BERT Semantic Evaluator for Radiology Reporting in Emergency Head CTs," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 13, pp. 61-74, 2025, doi: 10.1109/JTEHM.2025.3535676.
4. Q. Guan, Y. Huang, Y. Luo, P. Liu, M. Xu and Y. Yang, "Discriminative Feature Learning for Thorax Disease Classification in Chest X-ray Images," in IEEE Transactions on Image Processing, vol. 30, pp. 2476-2487, 2021, doi: 10.1109/TIP.2021.3052711.
5. K. Chen, X. Wang and S. Zhang, "Thorax Disease Classification Based on Pyramidal Convolution Shuffle Attention Neural Network," in IEEE Access, vol. 10, pp. 85571-85581, 2022, doi: 10.1109/ACCESS.2022.3198958.
6. M. S. A. Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani and A. Shaikh, "A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks," in IEEE Access, vol. 11, pp. 121574-121591, 2023, doi: 10.1109/ACCESS.2023.3328909.
7. M. A. Khan, "An IoT Framework for Heart Disease Prediction Based on MDCNN Classifier," in IEEE Access, vol. 8, pp. 34717-34727, 2020, doi: 10.1109/ACCESS.2020.2974687.
8. G. Sunilkumar and P. Kumaresan, "Deep Learning and Transfer Learning in Cardiology: A Review of Cardiovascular Disease Prediction Models," in IEEE Access, vol. 12, pp. 193365-193386, 2024, doi: 10.1109/ACCESS.2024.3514093.
9. T. Sinha Roy, J. K. Roy and N. Mandal, "Conv-Random Forest-Based IoT: A Deep Learning Model Based on CNN and Random Forest for Classification and Analysis of Valvular Heart Diseases," in IEEE Open Journal of Instrumentation and Measurement, vol. 2, pp. 1-17, 2023, Art no. 2500717, doi: 10.1109/OJIM.2023.3320765.
10. A. Mendoza, S. Tume, K. Puri, S. Acosta and J. R. Cavallaro, "Clinical Features and Physiological Signals Fusion Network for Mechanical Circulatory Support Need Prediction in Pediatric Cardiac Intensive Care Unit," in IEEE Journal of Biomedical and Health Informatics, vol. 29, no. 2, pp. 783-791, Feb. 2025, doi: 10.1109/JBHI.2024.3510217.
11. R. -K. Sheu, M. S. Pardeshi, K. -C. Pai, L. -C. Chen, C. -L. Wu and W. -C. Chen, "Interpretable Classification of Pneumonia Infection Using eXplainable AI (XAI-ICP)," in IEEE Access, vol. 11, pp. 28896-28919, 2023, doi: 10.1109/ACCESS.2023.3255403.
12. L. Ibrahim, M. Mesinovic, K. -W. Yang and M. A. Eid, "Explainable Prediction of Acute Myocardial Infarction Using Machine Learning and Shapley Values," in IEEE Access, vol. 8, pp. 210410-210417, 2020, doi: 10.1109/ACCESS.2020.3040166.
13. E. Essa and X. Xie, "An Ensemble of Deep Learning-Based Multi-Model for ECG Heartbeats Arrhythmia Classification," in IEEE Access, vol. 9, pp. 103452-103464, 2021, doi: 10.1109/ACCESS.2021.3098986.
14. R. Preetha, M. J. P. Priyadarsini and J. S. Nisha, "Automated Brain Tumor Detection From Magnetic Resonance Images Using Fine-Tuned EfficientNet-B4 Convolutional Neural Network," in IEEE Access, vol. 12, pp. 112181-112195, 2024, doi: 10.1109/ACCESS.2024.3442979.
15. P. Kumar Tiwary, P. Johri, A. Katiyar and M. K. Chhipa, "Deep Learning-Based MRI Brain Tumor Segmentation With EfficientNet-Enhanced UNet," in IEEE Access, vol. 13, pp. 54920-54937, 2025, doi: 10.1109/ACCESS.2025.3554405.
16. M. R. Azghadi et al., "Hardware Implementation of Deep Network Accelerators Towards Healthcare and Biomedical Applications," in IEEE Transactions on Biomedical Circuits and Systems, vol. 14, no. 6, pp. 1138-1159, Dec. 2020, doi: 10.1109/TBCAS.2020.3036081.

17. A. K. Jameil and H. Al-Raweshidy, "Efficient CNN Architecture on FPGA Using High Level Module for Healthcare Devices," in IEEE Access, vol. 10, pp. 60486-60495, 2022, doi: 10.1109/ACCESS.2022.3180829.
18. R. Ishikawa, T. Yuzawa, T. Fukiage, M. Kagesawa, T. Watsuji and T. Oishi, "Visibility Enhancement of Lesion Regions in Chest X-Ray Images With Image Fidelity Preservation," in IEEE Access, vol. 13, pp. 11080-11094, 2025, doi: 10.1109/ACCESS.2025.3528489.
19. D. Edirisinghe, W. Nimalsiri, M. Hennayake, D. Meedeniya and G. Lim, "Chest X-Ray Report Generation Using Abnormality Guided Vision Language Model," in IEEE Access, vol. 13, pp. 157651-157673, 2025, doi: 10.1109/ACCESS.2025.3606961.
20. A. Kornaev et al., "Automatic Calculation of Cardiometric Coefficients on Chest X-Ray Images," in IEEE Access, vol. 13, pp. 10702-10712, 2025, doi: 10.1109/ACCESS.2024.3524116.
21. O. A. Naghash, N. Ling and X. Li, "HyCoViT: Hybrid Convolution Vision Transformer With Dynamic Dropout for Enhanced Medical Chest X-Ray Classification," in IEEE Access, vol. 13, pp. 112623-112641, 2025, doi: 10.1109/ACCESS.2025.3584065.
22. I. Das et al., "Improving Medical X-Ray Imaging Diagnosis with Attention Mechanisms and Robust Transfer Learning Techniques," in IEEE Access, vol. 13, pp. 159002-159027, 2025, doi: 10.1109/ACCESS.2025.3607639.
23. E. Kim, S. Lee and K. Mu Lee, "Disentangled Contrastive Learning from Synthetic Matching Pairs for Targeted Chest X-Ray Generation," in IEEE Access, vol. 13, pp. 15453-15468, 2025, doi: 10.1109/ACCESS.2025.3531366.
24. K. Munir, M. Usama Tanveer, H. J. Alyamani, A. Bermak and A. Ur Rehman, "PneuX-Net: An Enhanced Feature Extraction and Transformation Approach for Pneumonia Detection in X-Ray Images," in IEEE Access, vol. 13, pp. 84024-84037, 2025, doi: 10.1109/ACCESS.2025.3568885.
25. S. Khumang, S. Kansomkeat, W. Tanomkiat and S. Intajag, "Tuberculosis Lesion Segmentation Improvement in X-Ray Images Using Contextual Background Label," in IEEE Access, vol. 13, pp. 36611-36625, 2025, doi: 10.1109/ACCESS.2025. 3532631.