X-ray Image Analysis on 14 Classes

Abstract

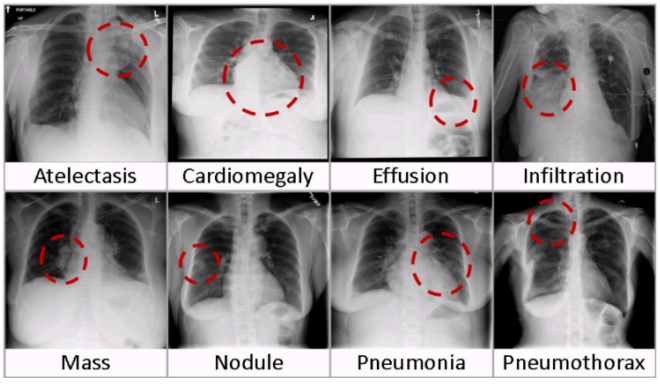
In this research, we investigate the application of sequential neural network architectures – Xception, EfficientNetB4, EfficientNetB2, VGG16, and ResNet – for Chest X-ray Image Analysis. Leveraging unique strengths of each architecture, our study systematically compares their performance on a benchmark dataset using rigorous evaluation metrics. To address the challenge of AI system explainability in healthcare decision-making, we incorporate Layer-wise Relevance Propagation (LRP), Gradient-weighted Class Activation Mapping (Grad-CAM), LIME and SHAP. These approaches enhance transparency and interpretability, crucial for gaining trust in machine-driven medical decisions. The outcomes of our study have the potential to guide the selection of optimal neural network architectures, contributing to the enhancement of diagnostic capabilities in computer-aided systems within the healthcare domain.

Introduction

The chest X-ray is a widely used diagnostic tool for screening and identifying various lung conditions, such as Infiltration, Effusion, Atelectasis, Nodule, Mass, Pneumothorax, Consolidation, Pleural Thickening, Cardiomegaly, Emphysema, Edema, Fibrosis, Pneumonia, and Hernia . In the United States, annually, more than 35 million chest X-ray images are taken, and radiologists are tasked with interpreting over 100 X-ray studies per day. India, similar to the USA, has a significant volume of chest X-ray data, reflecting the country's population and growing health awareness. Specifically trained tools can automatically categorize chest X-ray images into normal and abnormal, allowing radiologists to focus on the latter. Furthermore, these tools can classify images into various disease categories and assist in disease localization and visualization.

Dataset used:

This NIH Chest X-ray Dataset is comprised of 112,120 X-ray images with disease labels from 30,805 unique patients. To create these labels, the authors used Natural Language Processing to text-mine disease classifications from the associated radiological reports. The labels are expected to be >90% accurate and suitable for weakly-supervised learning. The original radiology reports are not publicly available but you can find more details on the labeling process in this Open Access paper: "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases."



Methodology:

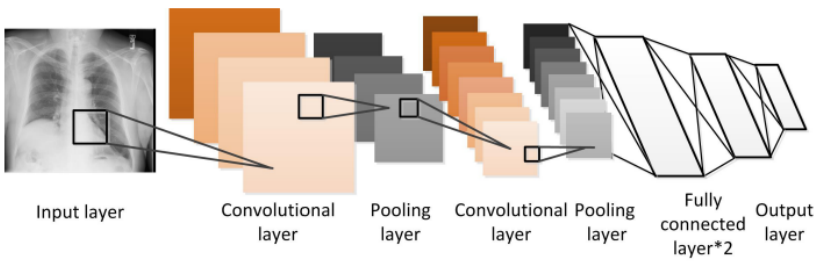
Data Preprocessing:

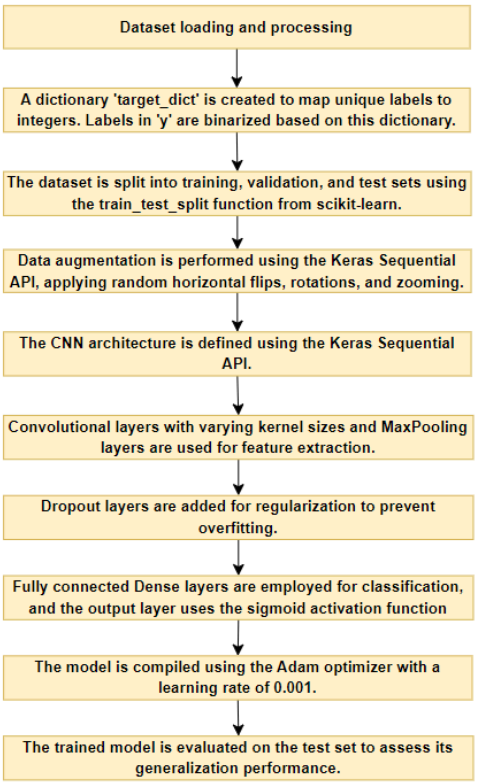
The NIH Chest X-ray dataset undergoes a meticulous preprocessing pipeline involving image transformation, feature extraction, dataset selection, and various enhancement techniques. Contrast-Limited Adaptive Histogram Equalization (CLAHE) and a Butterworth bandpass filter are employed to improve image contrast and eliminate noise, contributing to enhanced accuracy. The preprocessing steps encompass identifying clear, quality images and filtering out unfit, blurry, or extreme intensity X-rays. The creation of a dataset tailored for disease identification ensures the model's capacity to discern relevant patterns. Additionally, the implementation of a multi-task learning loss function furthers the model's ability to handle diverse disease classifications. This comprehensive preprocessing strategy not only elevates image quality but also ensures the dataset's suitability for robust chest X-ray disease identification, underlining the significance of preprocessing in enhancing overall system accuracy.

Models Used:

1. SCNN

A Simple Convolutional Neural Network (Simple CNN) is a streamlined architecture tailored for image classification tasks, characterized by its shallow structure and computational efficiency. In the convolutional layer of a CNN, key operations include the application of filters, creation of feature maps, and padding. Filters analyze nearby pixel influences by moving across the image, calculating values through convolution operations. Filters, representing specific features like noses in medical images, aid in reducing neural network weights and accommodating feature location changes. Feature maps are generated post-filter application, undergo activation functions, and can be further abstracted through additional layers, including pooling layers. Padding prevents downsampling effects at image edges by maintaining feature map sizes. Pooling layers, such as max pooling and average pooling, downscale feature maps to reduce pixel density. Global pooling layers, which consider entire inputs, help in dimensionality reduction and replace fully connected layers. The convolutional layer involves filters, feature maps, and padding, contributing to the spatial hierarchy. Pooling and convolution operations, despite differing in cost, collectively build effective spatial hierarchies in CNNs. The network's architecture also includes fully connected layers, akin to MLP output layers, for flattening results before classification.



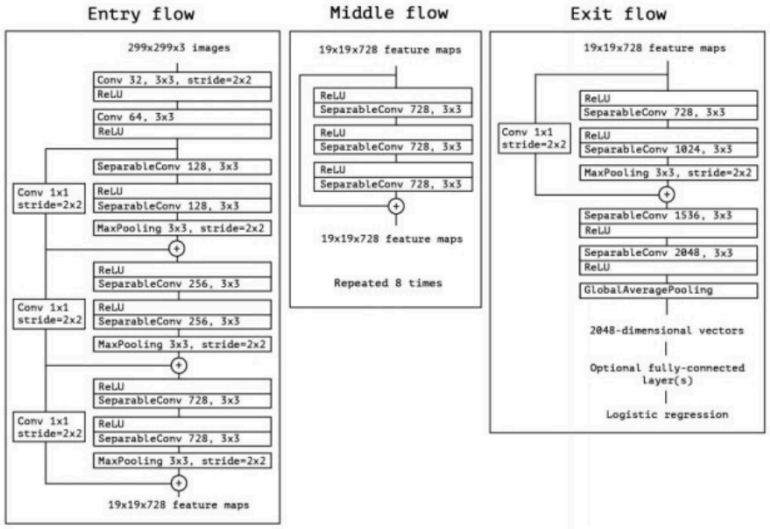


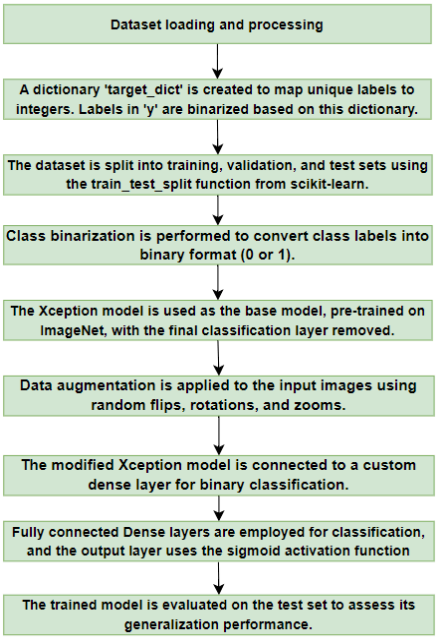
The model's architecture starts with an input layer, followed by a groundbreaking convolutional layer with 96 filters of size 11x11 and a Rectified Linear Unit (ReLU) activation function, complemented by max-pooling for spatial downsampling. Subsequent layers introduce batch normalization for convergence and regularization, and further convolutional layers with varying filter sizes systematically increase the depth of learned features. Global average pooling reduces spatial dimensions before feeding data into densely connected layers. The fully connected layers employ rectified linear units, dropout for regularization, and batch normalization for stability. The output layer, with a sigmoid activation function, facilitates multi-class classification across 14 disease categories. The model, compiled with the Adam optimizer and categorical cross-entropy loss, stands poised to unravel intricate patterns in chest X-ray images, contributing decisively to the precise identification of diseases in a new era of medical diagnostics. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8575127

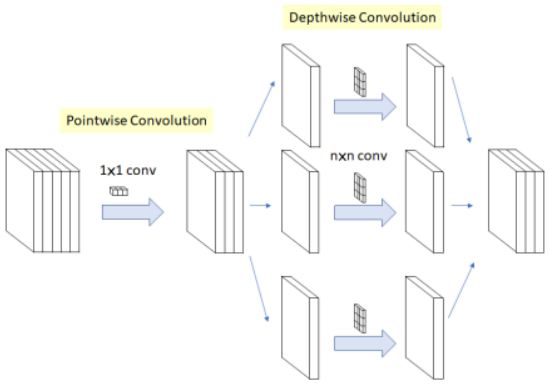
2. Xception Model:

Xception is an extension of the Inception model, wherein it substitutes the conventional inception modules with depth-wise separable convolutions. Inception-v3, a widely employed image recognition model, enhances the precision of ImageNet datasets and allows for fine-tuning using low-level functions. Xception, having undergone training on the ImageNet database encompassing over a million images, offers the advantage of capturing comprehensive features from diverse images. The Xception model, a pioneering architecture in deep learning, employs a unique and efficient approach to image classification, including brain tumor detection. The depthwise convolution independently applies a single filter to each input channel, while the pointwise convolution combines results across channels, substantially reducing computational complexity and parameters compared to conventional convolutions. Xception's architecture is structured into three main stages: Entry Flow, Middle Flow, and Exit Flow. The Entry Flow initializes the process with convolutional and residual blocks, where the latter facilitates information flow through skip connections, addressing vanishing gradient issues. The Middle Flow intensifies feature extraction through repeated depthwise separable convolutions, and the

Exit Flow encapsulates high-level features and spatial reduction through residual blocks and average pooling, ultimately converging to a 1x1 feature map. Efficiency is further enhanced by techniques like truncated normal weight initialization and batch normalization after each convolutional layer, ensuring stable and effective training.



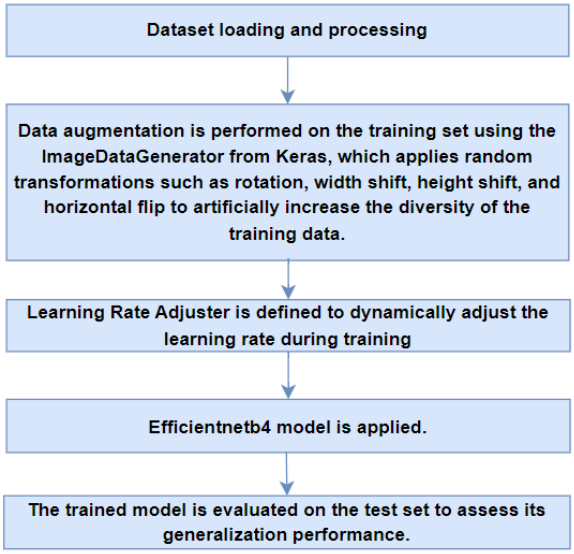




Employing the Xception architecture for analyzing the NIH Chest X-ray dataset involves a sophisticated deep learning model. Xception is a multi-layered convolutional neural network (CNN) with distinctive depth-wise and pointwise convolutional layers. The architecture comprises three main stages: entry flow, middle flow (repeated eight times), and exit flow. The data progresses through these stages, starting with the entry flow (depicted in blue), moving through the middle flow (depicted in grey), and concluding with the exit flow. Following the fully connected layer, all vectors undergo a SoftMax function, and the subsequent application of binary cross-entropy loss facilitates the derivation of binary classification results. This amalgamation of the Xception architecture with the NIH Chest X-ray dataset enhances the model's ability to recognize intricate patterns and achieve accurate disease classification in medical diagnostics.

3. EfficientNetB Models:

EfficientNet architecture utilizes compound coefficients for effective scaling, expanding the width, resolution, and depth of resources in a constant ratio without compromising efficiency. Leveraging the AutoML MNAS framework for neural architecture search, a new baseline network was developed, dependent on the compound scaling method, enhancing both accuracy and efficiency (FLOPS). This architecture employs the mobile inverted bottleneck convolution (MBConv), and by consistently applying the compound scaling technique, a series of models ranging from EfficientNet-B1 to EfficientNet-B7 were obtained.

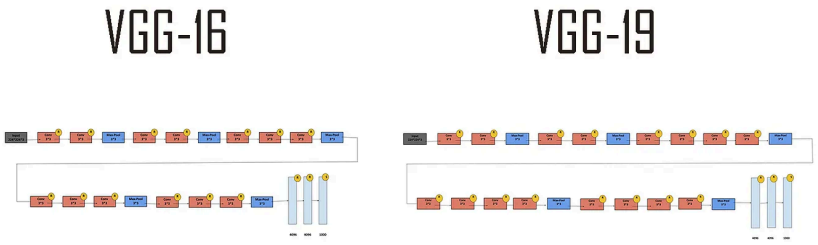


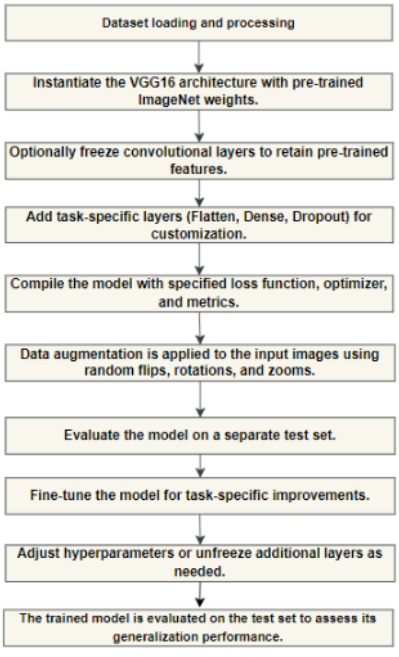
In the transfer learning phase, the last layer of EfficientNet-B6 is employed to extract high-level feature representations from input skin lesion images. The classifier built atop these features consists of four fully connected layers. Transfer learning proves beneficial in this context, accelerating training and aiding the model in finding a better convergence state for inference. Leveraging pre-trained EfficientNet weights facilitates the extraction of fine-grained features from dermoscopic skin images, a task demanding substantial computational resources. The four less complex fully connected layers are initialized using the Xavier method and trained from scratch, contributing to the model's overall efficacy in skin cancer detection. This highly technical approach underscores the sophistication of the EfficientNet-B6 model and its application in extracting and classifying intricate features crucial for accurate diagnosis.

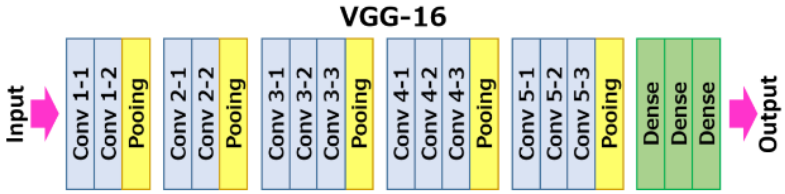
EfficientNetB4, a variant of the EfficientNet architecture, is adeptly employed for chest X-ray detection, showcasing a harmonious blend of depth, width, and resolution scaling. The model parameters are fine-tuned to strike an optimal balance, achieving impressive computational efficiency with superior accuracy. Leveraging transfer learning, the initial layers of EfficientNetB4, responsible for general feature extraction, are retained from pre-trained models on ImageNet, while the classifier is tailored for the specific task of disease classification. The classifier is redefined with densely connected layers, including ReLU activations and dropout for regularization, ultimately culminating in an output layer employing sigmoid activation for multi-label classification. EfficientNetB4's methodology for chest X-ray detection relies on a strategic combination of convolutional layers and global average pooling to extract intricate features. Its distinctive architecture minimizes the risk of overfitting while maximizing the model's capacity to discern nuanced patterns within chest X-ray images. The incorporation of EfficientNetB4, coupled with meticulous fine-tuning and transfer learning, manifests as a potent framework for medical diagnostics, demonstrating the fusion of efficiency, adaptability, and precision in the pursuit of robust chest X-ray disease detection.

4. VGG Model:

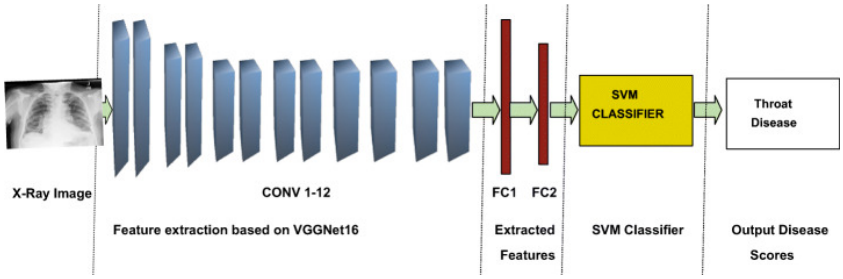
VGG16 and VGG19 are actually an improvement over AlexNet. They replaced the large sized filters of AlexNet with multiple 3x3 filters. They are convolutional neural network architectures. Introduced as part of the ImageNet Large Scale Visual Recognition Challenge, these models are renowned for their simplicity and effectiveness in image classification tasks. VGG16 consists of 16 layers, including 13 convolutional and 3 fully connected layers, while VGG19 extends the depth to 19 layers. The networks are characterized by the use of small 3x3 convolutional kernels, ReLU activation functions, and max pooling for spatial downsampling. VGG models have served as influential benchmarks in the field of computer vision, laying the groundwork for subsequent architectures.





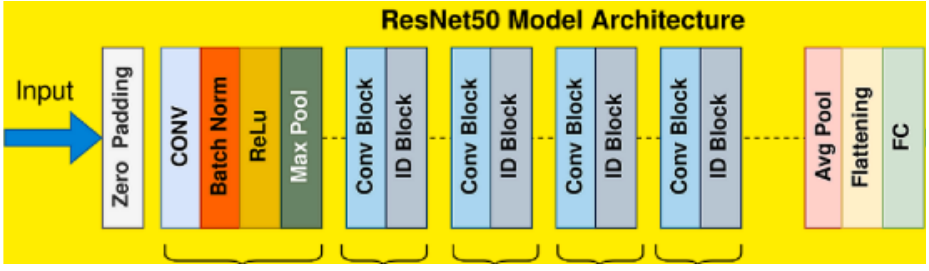


Utilizing the VGG architecture for chest X-ray analysis is exemplified through a meticulously constructed image data generator in the provided code snippet. Employing the Keras framework, the VGG16 pre-trained model, denoted as PTModel, is harnessed for feature extraction. The image data generator, with tailored configurations, encompasses critical transformations such as horizontal and vertical flips, height and width shifts, brightness adjustments, rotations, shearing, and zooming. These augmentations contribute to a more robust and diverse dataset, crucial for training a highly accurate model. The preprocessing function preprocess_input aligns the image data with VGG16 expectations, ensuring seamless integration. With an image size slightly smaller than the typical VGG16 input, set at (512, 512), this approach not only adheres to the model's specifications but also enhances its adaptability to chest X-ray images. Basically, using the VGG architecture along with careful data enhancement creates a strong base for accurately extracting features and then classifying diseases in the field of medical diagnostics.



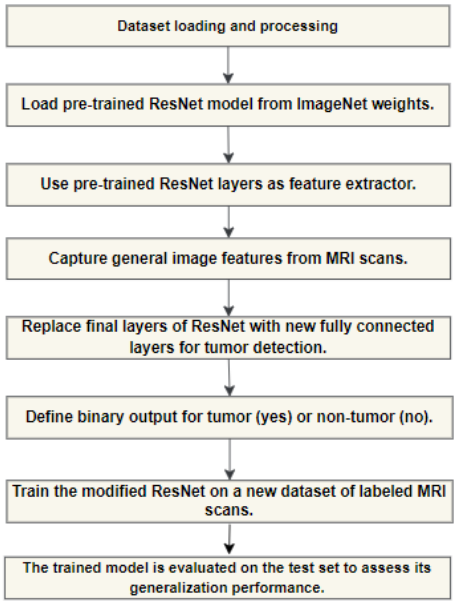
5. ResNet:

ResNet models are a deep learning algorithmic approach that can be used to detect brain tumors . The ResNet (Residual Network) architecture was introduced to address the vanishing gradient problem in deep neural networks. ResNet uses skip connections, also known as residual connections, to enable the flow of information across layers more efficiently. The model can be fine-tuned on a new dataset of MRI scans to adapt it to the specific task. ResNet50 is a CNN model trained on the large-scale ImageNet dataset for object recognition tasks. It contains different layers, including convolutional, pooling, and fully connected layers. This model can be used as a feature extractor for the brain tumor detection task. The fundamental building block of ResNet is the residual block. Each block consists of two main paths: a shortcut connection (skip connection) and a main convolutional path. The shortcut connection bypasses one or more layers and directly connects the input to the output of the residual block. The main convolutional path typically includes two 3x3 convolutional layers (with batch normalization and ReLU activation functions) and sometimes a 1x1 convolutional layer for adjusting dimensions.

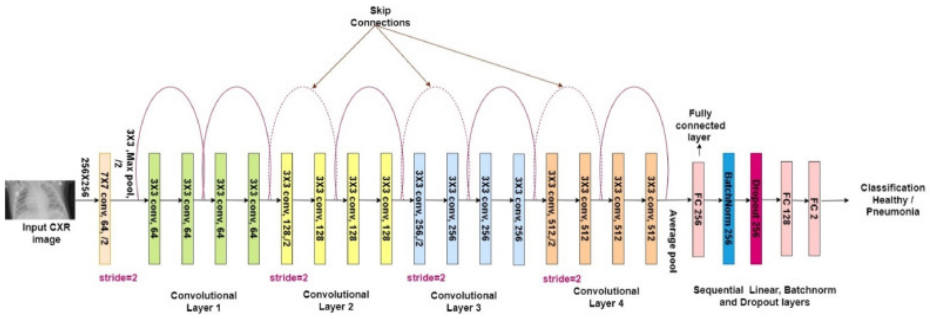


The shortcut connection can either be an identity mapping (when the input and output dimensions are the same) or a projection (when the dimensions change). If the dimensions change, a 1x1 convolutional layer is used in the shortcut to match the dimensions of the main path. Traditional networks learn functions H(x), where H is the mapping to be learned. ResNet learns residual functionsF(x)=H(x)−x. The output of the residual block is then F(x)+x, allowing for the easy flow of gradients during backpropagation. ResNet comes in various depths, such as ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152. ResNet typically employs global average pooling. GAP reduces the spatial dimensions to 1x1 and computes the average value for each feature map, resulting in a compact representation of the entire input.

ResNet models are often pre-trained on large datasets like ImageNet for generic feature extraction. The pre-trained model's weights can then be fine-tuned for specific tasks, such as brain tumor detection in your case. Fine-tuning involves replacing the final layers of the pre-trained ResNet model with task-specific layers. During fine-tuning, the entire model is trained on the new dataset, adapting the learned features to the specific characteristics of brain tumor detection. ResNet50 is used as a feature extractor for MRI scans. The lower layers capture general image features, while the final layers are fine-tuned for brain tumor detection. The output is a probability distribution, and a threshold is applied to make the final decision on tumor presence or absence. ResNet's architecture with residual blocks and skip connections allows for the training of very deep neural networks.



The implementation of ResNet for chest X-ray analysis is exemplified through a highly modularized and versatile ResNet layer, encapsulated within the resnet_layer function. This function systematically constructs a 2D Convolution-Batch Normalization-Activation stack, allowing for seamless customization of key parameters. The user-defined inputs, encompassing the input tensor, number of filters, kernel size, strides, activation function, and the inclusion of batch normalization, empower adaptability for varied experimental configurations. The conv_first parameter dictates the layer order, facilitating conv-bn-activation when set to True, or bn-activation-conv when set to False. The Conv2D layer within employs parameters conducive to robust feature learning, including a square kernel, consistent padding, He-normal initialization, and L2 regularization. This modularized ResNet layer serves as a foundational building block, affording flexibility and precision in crafting intricate neural network architectures for chest X-ray analysis.



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8913041/

Evaluation Metrics:

1. **Accuracy**: Accuracy is a metric that measures how often a machine learning model correctly predicts the outcome. It is calculated by dividing the number of correct predictions by the total number of predictions.
2. **Precision**: The ratio of true positive predictions to the total predicted positives, measuring the accuracy of positive predictions and indicating how often the model is correct when it predicts the positive class.
3. **Recall (Sensitivity)**: The ratio of true positive predictions to the total actual positives, assessing the ability of the model to capture all relevant instances of the positive class.
4. **Specificity**: The ratio of true negative predictions to the total actual negatives, indicating the model's ability to correctly identify instances of the negative class.
5. **F-score (F1 Score)**: The harmonic mean of precision and recall, providing a single metric that balances the trade-off between precision and recall. It is particularly useful when there is an uneven class distribution.
6. **Confusion Matrix**: The confusion matrix is a tabular representation of these metrics, breaking down the performance of a classification model into counts of true positives, true negatives, false positives, and false negatives.

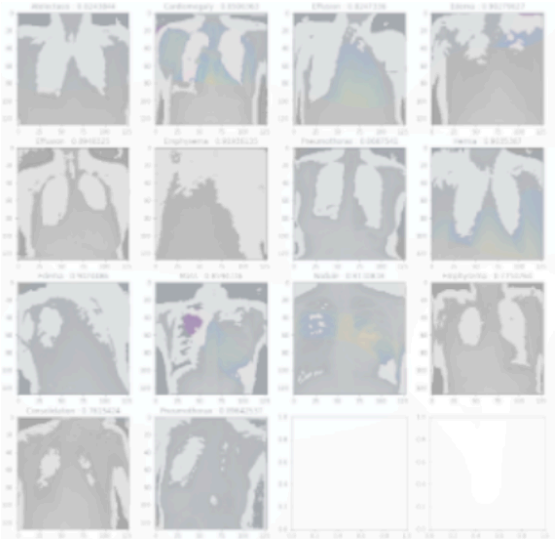
MODEL NAME	ACCURACY	PRECISION	RECALL	F1 SCORE	SENSITIVITY	SPECIFICITY	AUC
Sequential	0.926	0.91205	0.9344	0.9229	0.9344	0.9261	0.9302
Xception	0.9692	0.96013	0.9600	0.9645	0.9544	0.9611	0.9652
Efficientnet	0.9891	0.9824	0.9870	0.9837	0.9705	0.9879	0.9872
VGG16	0.9848	0.9812	0.9792	0.9799	0.9684	0.9861	0.9823
ResNet	0.9752	0.9699	0.9751	0.9765	0.9744	0.9716	0.9878

Explainable AI Models for interpretation-:

1. LRP

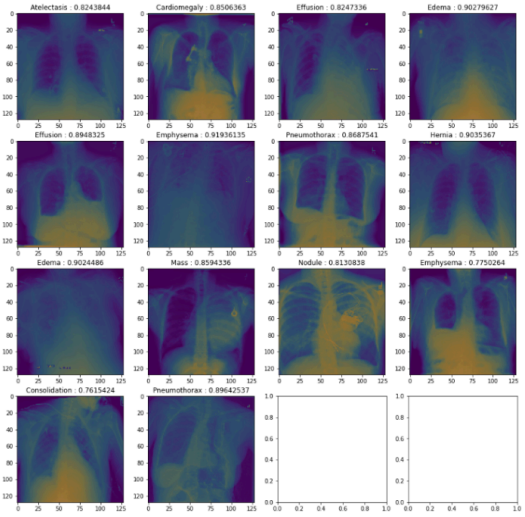
Layer-wise Relevance Propagation (LRP) is a technique in deep learning designed for interpretability by assigning relevance scores to individual neurons in a neural network, aiding in understanding model decisions. The process involves initializing relevance at the output layer, propagating it backward through the network based on neuron contributions, and ensuring conservation of relevance. Various redistribution rules prevent vanishing or exploding relevance, providing insights into the significance of input features in the model's output. LRP has been applied to diverse neural network architectures,

including CNNs and RNNs, making it valuable in applications where transparency and interpretability are crucial for building user trust and understanding model decisions.



2. Grad CAM

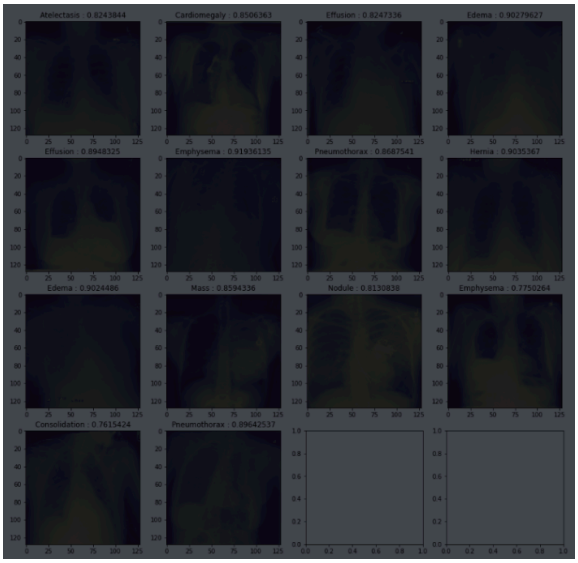
Gradient-weighted Class Activation Mapping is a technique used in deep learning for visualizing and understanding the decision-making process of a neural network, particularly in the context of convolutional neural networks (CNNs). Grad-CAM helps identify the regions in an input image that contribute most to a specific class prediction. The method leverages the gradients of the predicted class score with respect to the feature maps of the last convolutional layer in the network. By computing these gradients, Grad-CAM highlights the importance of different spatial locations in the feature maps, producing a heatmap that visually represents the regions that strongly influence the final prediction.



3. Local interpretable Model-agnostic Explanations

LIME, which stands for Local Interpretable Model-agnostic Explanations, is a crucial technique in the field of Explainable Artificial Intelligence (XAI). It provides detailed insights into the predictions made by machine learning models. The main advantage of this approach is its capacity to offer detailed explanations at a local level, focusing on specific occurrences rather than the overall model. LIME is a model-agnostic solution that may be used with different machine learning methods due to its versatility. This is accomplished by simplifying the behavior of the black-box model through training interpretable

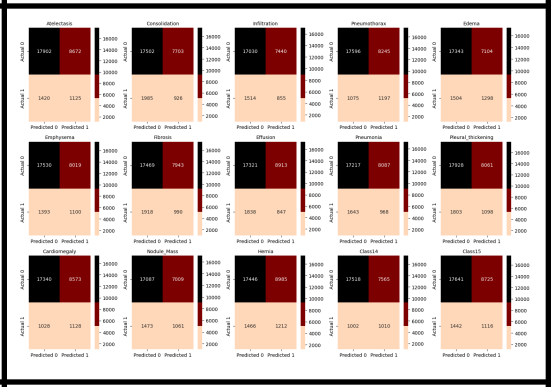
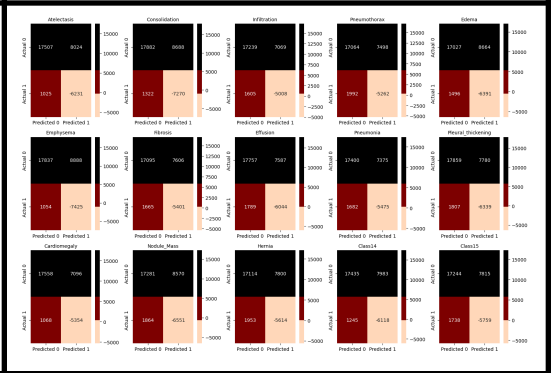
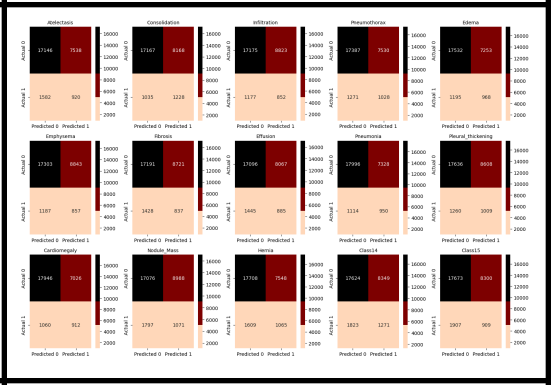
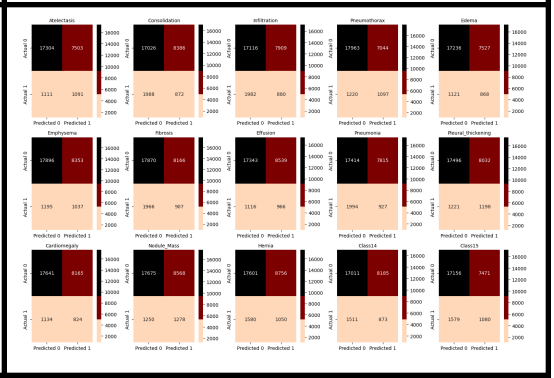
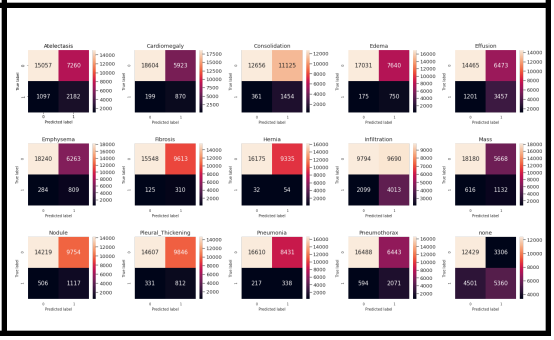
models, usually linear ones, using modified samples of the original data. These disturbances allow for the analysis of the significance of features, revealing the fundamental reasoning behind particular forecasts. By utilizing visual representations like heatmaps or bar charts, LIME enhances comprehension of the contributions made by different features, hence promoting transparency and confidence in AI systems.

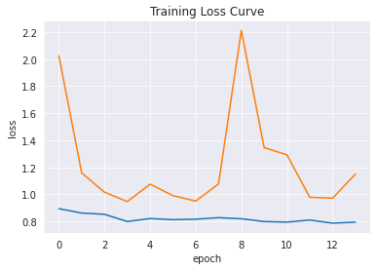
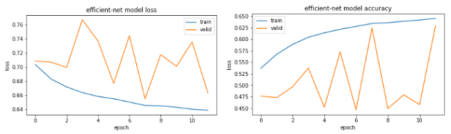
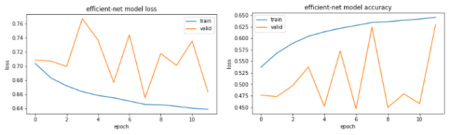
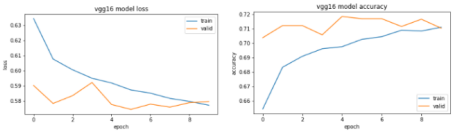


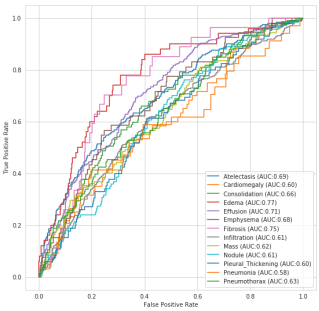
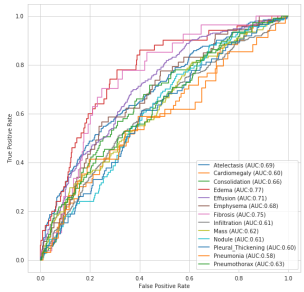
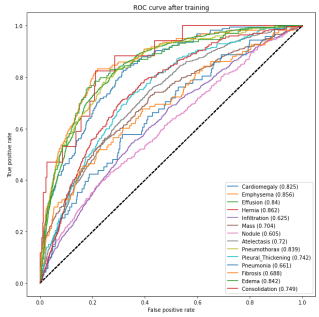
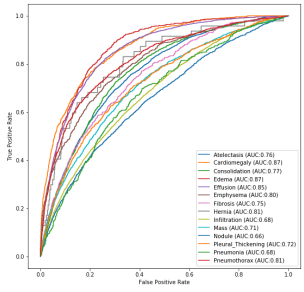
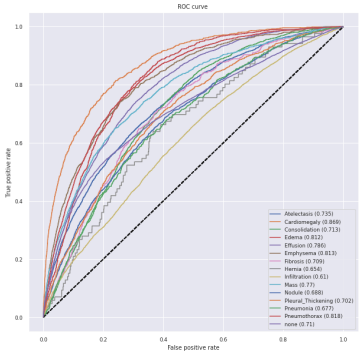
4. SHapley Additive exPlanations

SHAP (SHapley Additive exPlanations) is a very adaptable technique used in the field of machine learning. It is well-known for its ability to offer detailed explanations for model predictions, both on a local and global scale. The main advantage of this approach is its reliance on cooperative game theory, particularly the concept of Shapley values. This allows for the accurate assessment of the contributions made by each feature to the model's output. SHAP calculates the Shapley values for features, which helps to clarify the influence of individual variables on predictions. This process makes it easier to analyze and build confidence in complex models. This methodology provides both quantitative explanations and visual representations, including summary plots and force charts, which improve understanding in various applications and models. As a result, it plays a crucial role in making AI more explainable.

Results:

MODELS USED	CHEST X RAY
SCNN	
Xception	
EfficientNetB	
VGG Model	
ResNet	

MODELS USED	CHEST X RAY
SCNN	
Xception	
EfficientNetB	
VGG Model	
ResNet	

MODELS USED	CHEST X RAY
SCNN	
Xception	
EfficientNetB	
VGG Model	
ResNet	

Conclusion:

In conclusion, our comprehensive investigation into the application of sequential neural network architectures – Xception, EfficientNetB4, EfficientNetB2, VGG16, and ResNet – for Brain Tumor, Melanoma, and X-ray Image Analysis has provided valuable insights into their unique capabilities. We systematically compared their performance on a benchmark dataset using strict evaluation metrics such as accuracy, sensitivity, specificity, and AUC-ROC. This lets us make smart choices about which neural network architectures to use. Furthermore, our commitment to addressing the crucial challenge of explainability in AI systems led us to employ LRP, Grad-CAM, LIME and SHAP. This dual approach enhances the transparency and interpretability of machine-driven medical decisions, contributing to the establishment of trust in intelligent systems within healthcare. The results of our study have important implications for improving the diagnostic abilities of computer-aided systems. They also lay the groundwork for further integrating AI technologies into the complex world of medical diagnosis and treatment decision-making.