International Journal of Combinatorial Optimization Problems and Informatics, 16(3), May-Aug 2025, 563-577. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2025.v16i3.871

www.editada.org

EDNN-PCXR: Enhanced Pediatric Chest X-Ray Classification using Fine-Tuned Deep Neural Networks

S. S. Thakur[1], Soma Bandyopadhyay[1], Mahika Thakur[1], Anwesha Laha[1], Aditya Singh[1], Aditya Kumar Singh[1], Shyam Sunder Singh[1], Srajan Mishra[1]
[1] MCKV Institute of Engineering, Liluah, Howrah, West Bengal, India
subroto_thakur@yahoo.com, somabanmuk@yahoo.co.in, {thakur.mahika02, anweshalaha2002, adityajeet0510, 18adityasingh03, shyamsundersinghofficial, srajanmishra2003} @gmail.com.

Abstract. In today's world, the rapid progress of artificial intelligence (AI) and machine learning (ML) presents remarkable opportunities for developing innovative solutions to tackle various challenges within the healthcare sector. Deep learning (DL) has become a powerful tool in healthcare, transforming patient care and improving clinical support. It is increasingly utilized to identify critical features in medical images that go beyond what the human eye can naturally detect. Chest X-ray images are a widely used medical tool for detecting various health conditions. This covers pneumonia, lung cancer, and other issues such as tissue damage and bone fractures. Regardless of experience, for radiologists, accurately identifying diseases from X-ray images can be a strenuous task. Diagnosing pneumonia, a viral lung infection, is especially difficult because its symptoms closely resemble those of other pulmonary diseases. This similarity reduces the accuracy of current diagnostic methods. The vast amount of information contained in X-ray images has created an increasing demand for computerized support systems. This paper compares various computer-aided pneumonia identification methods, incorporating different deep learning approaches to streamline diagnosis using images of chest X-rays. In this study, seven types of deep convolutional neural networks have been applied to a dataset containing 5,856 chest X-ray images of normal and pneumonia cases. It has been observed that VGG-16, VGG-19, and ResNet-50 effectively classify images of Chest X-ray into normal and pneumonia affected cases. Among these architectures, VGG-16 performs the best with an accuracy of 91%, followed by VGG-19 at 90.38% and ResNet-50 at 89.94%. The results surpass those of the advanced techniques mentioned in the literature.

Keywords: Pneumonia detection, CNN, VGG16, VGG19, ResNet-50, Mobile-Net, NasNet-Mobile, DenseNet etc.

Article Info
Received February 09, 2025
Accepted March 21, 2025

1 Introduction

Pneumonia is a lung infection that causes inflammation and fluid buildup in the air sacs, mainly caused by infectious agents like bacteria and viruses. Pneumonia presents with a range of symptoms, including a productive or dry cough, chest pain, fever. Its effects can vary widely from person to person. It can cause pericardial effusion, a condition in which fluid builds up in the chest and leads to breathing difficulties. Several risk factors can increase the likelihood of developing pneumonia. These include conditions like cystic fibrosis, asthma, diabetes, chronic obstructive pulmonary disease (COPD), and heart failure. Additionally, factors such as a history of smoking, weakened cough reflexes from a stroke, and a compromised immune system also contribute to the risk (Musher & Thorner, 2014 and Rudan, 2008). Infection by bacteria or viruses in the alveoli prompts the body's immune system to respond, causing immune cells to accumulate and inflammatory mediators to be released. Consequently, the alveoli accumulate pus and cellular debris, impairing gas exchange and leading to respiratory difficulty. In more severe cases, the infection may extend to surrounding lung tissues, resulting in consolidation, abscesses, or pleural effusion. The severity and progression of pneumonia are determined by the type of pathogen and the individual's overall health. Tragically, pneumonia stands out as one of the top killers of children under the age of five, responsible for approximately 1.4 million deaths each year. This accounts for 18% of global child mortality in this age group. Although pneumonia affects children globally, this condition is particularly widespread in South Asia and sub-Saharan Africa. Pneumonia is categorized as either infectious or non-infectious, depending on its underlying cause. Infectious pneumonia is further divided into subtypes, including bacterial, viral, mycoplasma, and chlamydial pneumonia. Not all pneumonia cases are linked to infections, certain types such as immune-associated pneumonia and aspiration pneumonia, are non-infectious in nature. Pneumonia is categorized by its infection source, such as community-acquired pneumonia (CAP), hospital acquired (HAP), and ventilator associated (VAP) types. Among these, CAP is the most common (Ukwuoma et al., 2023). However, pneumonia can be prevented through basic interventions and treated effectively with affordable and simple medical care (Rello and Diaz, 2003). Timely and appropriate medical intervention is also essential in determining the outcome of the condition.

Early detection is crucial for ensuring proper treatment and enhancing survival rates. Computed tomography (CT), X-rays and magnetic resonance imaging (MRI) are frequently used for diagnosing lung conditions. These radiological methods help in assessing lung conditions effectively. For the early-stage pneumonia detection the most widely used method is Chest X-ray imaging. Figure 1 displays an X-ray of a normal lung, an X-ray of a lung affected by bacterial pneumonia, and an X-ray of a lung affected by viral pneumonia. The chest X-ray on the left panel shows clear lungs and there is no abnormal opacity present. Bacterial pneumonia as shown in the middle panel, typically appears as a localized lobar consolidation. This typically appears in the right upper lobe, highlighted by white arrows. In contrast, viral pneumonia, shown in the right panel, exhibits a more dispersed "interstitial" pattern affecting both lungs (Kermany et al., 2018). X-ray scans have been used for years to examine sensitive areas such a head, chest and bones. However, recognizing pneumonia in chest X-ray images can be a complex process. It can sometimes be confused with other diseases or share symptoms with frequently occurring ailments. These uncertainties have resulted in a significant amount of subjective judgment when diagnosing pneumonia. As a result, there are often differences in diagnosis among radiologists (Ayan & Unver, 2019). Even for experienced doctors, diagnosing pneumonia is a complex task, as the images often share similar features with other diseases like lung cancer. Traditional methods are time and energy consuming, making it difficult to diagnose pneumonia through a standardized process. Therefore, an automated system is necessary for pneumonia detection.

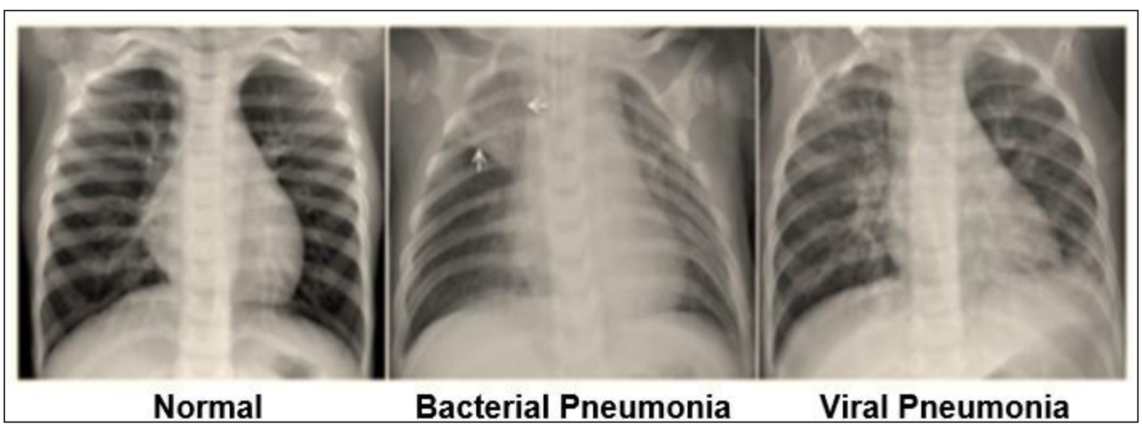


Fig. 1. Normal and pneumonia affected X-Ray Plate

Deep learning (DL) in healthcare offers remarkable innovation by analyzing large data sets quickly and accurately. Unlike rule-based algorithms, machine learning improves with increased data exposure. Depending on the experience gained, this allows ML algorithms to learn and make predictions. DL is a section of artificial neural networks (ANN), inspired by the biological nervous system. It consists of interconnected neurons spread across input layer, hidden and output layers. The connections between neurons are the weights assigned. To optimize the model's performance and its accuracy, these weights are adjusted during training (Pillai, 2022). X-ray images are 2-dimensional (D) representations of a 3D body. Convolutional neural networks (CNNs) demonstrate superior performance when handling 2-dimensional images. Manual assessment of X-rays by radiologists can be time consuming and prone to errors, while computer algorithms quickly process large datasets, improving analysis speed and accuracy. These algorithms also detect subtle patterns beyond human perception, enhancing pneumonia detection sensitivity.

This research aims to evaluate the effectiveness of CNN, including VGG-16, VGG-19, ResNet-50, MobileNet, NasNet-Mobile, and DenseNet. The purpose is to diagnose pneumonia in young children aged 1 to 5 years through chest X-ray imaging. The goal of this research is to assess the accuracy of deep learning models in categorizing images as either Pneumonia or Normal. The structure of this paper is arranged as follows: Section 1 offers an overview of the study and its context. Section 2 delves into previous studies relevant to this research. Section 3 details the datasets and the proposed approach. The result and evaluation criteria have been explored in Section 4. Finally, Section 5 presents a summary of the study while outlining possible research directions.

2 Literature Survey

Artificial intelligence and deep learning have increasingly become focal points of interest and widespread discussion in recent years. (Dixit, 2018). Among these, deep learning has stood out as an essential subset of machine learning, modeled after the intricate processes of the human brain. The healthcare industry, characterized by vast amounts of data and a pressing demand for efficiency and accuracy, provides an optimal setting for leveraging machine learning(ML) methods such as DL (Krumholz). The scan line technique is utilized to optimize X-ray images of the chest to exclude unrelated body regions and prevent diagnostic inaccuracies, as described in (Hermann, 2014). Deep learning-driven approaches for identifying chest pathologies were also discussed (Bar et al., 2015). A methodology using CNN was proposed for prostate segmentation in MRI volumes (Milletari, 2026), while a sophisticated deep neural network model was presented for skin cancer classification with dermatologist-level expertise (Esteva et al., 2017). CheXNet is presented as a deep convolutional neural network comprising 121 layers. ChestX-ray14, an extensive public dataset with more than 100,000 frontal chest X-ray images covering 14 diseases, serves as the training foundation. They compare CheXNet with radiologists using the Receiver Operating Characteristic (ROC) curve, which plots the model's sensitivity against 1 – specificity (Rajpurkar, 2017). A fast-growing technology that's making waves in recent research, deep learning stands out for its efficiency in processing and analyzing vast data (Al-Antari et al., 2018 and Al-Masni et al., 2018). The effectiveness of customized CNNs was examined to recognize pneumonia and distinguish between bacterial and viral types in pediatric chest X-ray images (Kermany et al., 2018). A study to expand a dataset by generating synthetic X-ray samples of the chest using GANs, which enable the exploration of the fundamental structure within diagnostic images, making it possible to generate accurate, high-definition samples (Madani et al, 2018). Transfer learning (TL) was utilized to train a neural network with significantly less data than traditional methods require. This approach also enhanced diagnostic transparency and interpretability by emphasizing the regions identified by the neural network (Kermany et al., 2018). Pneumonia detection was performed using a CNN model (Abiyev & Ma'aita, 2018). Modern strides in deep learning (DL) have surpassed performance by humans in several practices. DL can be used to predict treatment results, exemplified in Chevalier studies and cancer therapies. Promising results in categorizing thoracic illnesses using X-ray images are shown by DL algorithms and labeled data.

Traditionally, human professionals have developed and tested deep neural network (DNN) models through a time-consuming and resource-intensive trial-and-error process (Stephen et al., 2019). For pneumonia detection, a region-based CNN was applied alongside image augmentation for pulmonary image segmentation (Sirazitdinov et al., 2019). A pioneering multi-scale heterogeneous 3D convolutional neural network (MSH-CNN), grounded in chest computed tomography (CT) images, was proposed (Xiao et al., 2019). The hierarchical CNN framework, along with a unique sin-loss function design, was employed to enhance pneumonia detection (Xu et al., 2018). The evaluation focused on four deep learning models where two pre-trained models namely MobileNetV2 and ResNet152V2 was used along with a CNN model, and an LSTM model (Elshennawy & Ibrahim, 2020). Each model was evaluated with a range of parameters, applying traditional classification evaluation metrics. Each model underwent evaluation under different parameters, employing standard classification evaluation metrics. Pneumonia detection was carried out using the ResNet152 model, employing TL techniques. It achieved a notable recognition rate of 97.4% without any adjustments or feature preprocessing (Talo, 2019). A study (Varshni et al., 2019) investigated detection of pneumonia using multiple CNN-based models for feature extraction via TL employing various classifiers as predictive tools. It was found that the integration of CNN model along with supervised classifiers plays a crucial role in assessing images of chest X-ray, particularly for identifying pneumonic conditions. Feature extraction with DenseNet-169 and prediction using SVM produced the best results.

Segmentation of lung regions was performed on chest X-rays, and eight statistical features were extracted for classification using multi-layer perceptron (MLP), random forest, sequential minimal optimization (SMO), classification via regression, and logistic regression. The MLP classifier achieved 95.39% accuracy on 412 images (Chandra & Verma, 2020). In (Chen et al., 2020) the model was developed and validated using the dataset from the DLAI3 Kaggle Challenge (Jonathan, 2021), consisting of 50 CXR images (25 Covid-19 positive and 25 negative). The dataset was split, with 80% allocated for training and the remaining 20% reserved for validation. A total of 250 images distributed across different file formats were used for testing, which included 7 positive cases. The proposed model utilizes VGG16 framework, enhanced with pre-trained weights from ImageNet. Among the models evaluated VGG16 model outperformed the others. After 10 epochs, the VGG16 model reached 98% accuracy on the test data. Additionally, VGG16 demonstrated high accuracy in identifying the positive class, showing superior performance. However, they trained the model with very limited number of training set. In (Mabrouk et al., 2022) ensemble method was proposed that integrates forecasts from several CNN models for enhanced classification performance. Transfer learning (TL) and fine-tuning were utilized, and the framework was optimized with batch normalization and dropout layers. Three well-established pre-trained models namely DenseNet169, MobileNetV2, and Vision Transformer were utilized, initially trained on the ImageNet database and later fine-tuned on a X ray images of chest dataset. During experimentation, the model's extracted features were fused, resulting in superior performance compared to top-tier methods. The proposed ensemble learning approach achieved a testing accuracy of 93.91% and an F1-Score of 93.88%. In (Pugliesi, 2019) VGG-16 framework attained an accuracy of 74.9% alongside a loss of 48.8% during a testing. In contrast, the ResNet-50 framework outperformed it, achieving an accuracy of 88.9% with a lower loss of 28.9% during testing. The analysis reveals the potential of DL models for pediatric pneumonia detection, with ResNet-50 showing a marked advantage over VGG-16.

3 Proposed Methodology

This research focuses on utilizing deep learning for automated pneumonia diagnosis. Its goal is to assess and compare the effectiveness of a standard CNN model with varied transfer learning approaches on chest X-ray datasets. The dataset is split into 3 separate folders: train, test, and validation. Each of these folders include subdirectories for the two image categories: Pneumonia and Normal. It consists of 5,856 JPEG images of chest X-ray distributed across two categories. X-ray images of Chest were taken from the anterior-posterior view were sourced from Mendeley data. They were collected from past cases of pediatric patients aged 1 to 5 years at Guangzhou Women and Children's Medical Center, Guangzhou (Kermany, 2018). The X-ray scans were carried out as a regular part of the patients' treatment. To maintain accuracy and reliability in the analysis, a rigorous quality control process was implemented, eliminating any low-quality or unreadable scans (Kermany, 2018). Two experienced physicians reviewed and confirmed the diagnoses of the selected images to ensure their appropriateness for enhancing the AI model. To mitigate scoring errors, a third expert conducted an additional evaluation. This ensured the accuracy of the assessment set. The dataset is structured into three sets: a training set with 5,216 images, a testing set contains 624 images, and a validation set with 16 images. Regarding the distribution of cases, a total of 4,273 images are categorized as pneumonia cases, while 1,583 images represent normal cases.

The pneumonia dataset was prepared for training using ImageGenerator, which was utilized to rescale and augment the images, enhancing the model's robustness by introducing diverse image variations. During training, data augmentation techniques were applied, including pixel value rescaling (normalized to a range of 0 to 1 by dividing by 255), random shearing, zooming, and horizontal flipping. These transformations help to reduce overfitting by exposing the model to varied image versions, promoting the learning of generalizable features rather than memorizing specific details. The validation and test images were rescaled to normalize pixel values, with no additional augmentation applied, preserving their original structure for unbiased evaluation. Images from their respective directories were utilized for training, validation, and testing, configured with specified target dimensions and batch size. To meet the requirement of the task, this classification was done, and the images were categorized into two groups namely Normal and Pneumonia. This study evaluates a baseline CNN for detecting pneumonic conditions from X-ray images of chest. After that, transfer learning (TL) is applied using DL models such as VGG-16, VGG-19, ResNet-50, MobileNet, NasNet-Mobile, and DenseNet etc. This approach boosts the model's performance.

3.1 Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) have transformed deep learning, particularly in image recognition. Their advanced architectural design enables them to extract intricate features from images, surpassing traditional neural networks (O'Shea, 2015 and Aghdam & Heravi, 2017). A CNN generally consists of three primary layers i.e. the input layer, hidden layer, and the output layer (Wu, 2017). Raw images are initially fed into the input layer. They are then passed to the hidden layer, where feature extraction takes place. The hidden layer consists of three key components i.e. convolution layers, pooling layers, and fully connected (FC) layers. The foundation of a CNN model is its ability to extract features, a process primarily handled by the convolution layer (Lavin & Gray, 2016). The convolution layer analyzes raw images using a sliding window technique, having a filter of fixed-size. This allows it to automatically detect and extract relevant features from the images. This capability to learn and detect essential patterns makes CNNs highly effective in image recognition. The CNN architecture is shown in Figure 2.

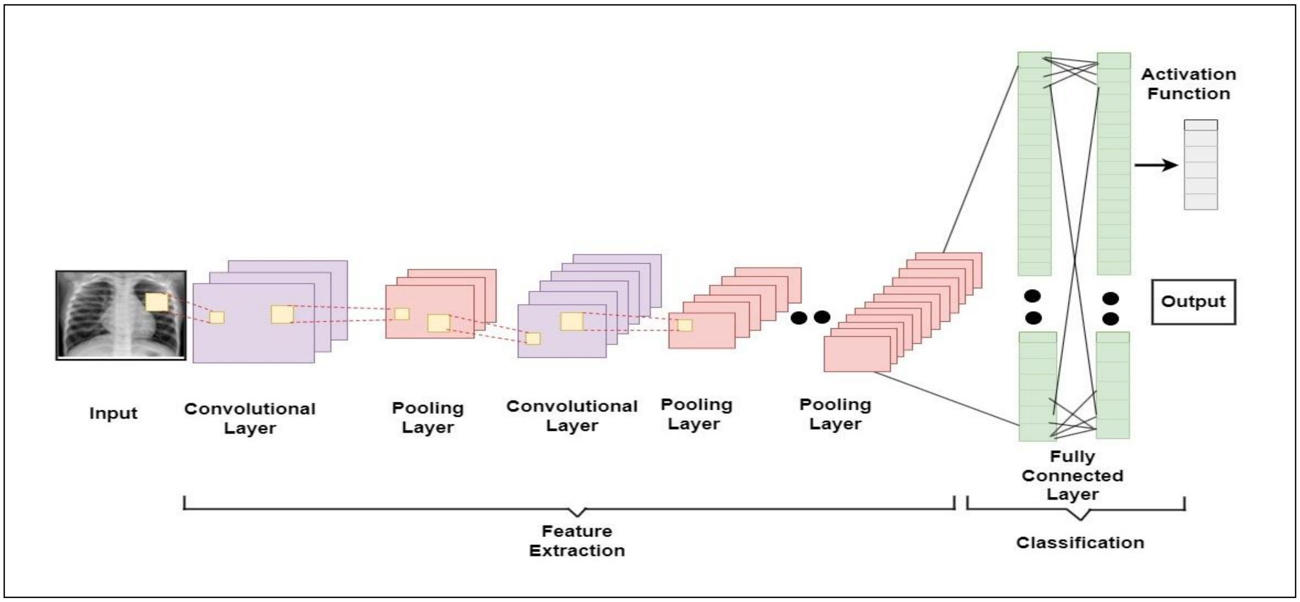


Fig. 2. CNN architecture.

The primary objective of convolutional layer's is to extract features from images. It contains multiple convolution kernels, with the layer's output being calculated by applying these kernels to the input data as shown in equation 1.

$$x_j^l = f\left(\sum_{i \in M_j} x_j^{l-1} * k_{ij}^l + b_j^l\right) \qquad (1)$$

In this context, x_j^{l-1} represents the feature map which is output from previous layer, while x_j^l is the resultant output of ith channel in the jth convolutional layer. The activation function is denoted as $f(.)$. M_j is the subset of input feature maps, k_{ij}^l is the convolutional kernel, and b_j^l is the corresponding weight for this kernel.

After the convolutional layer, the pooling layer helps in reducing dimensionality, making the model more efficient. While preserving essential features, the pooling layer minimizes the dimensionality of feature maps from the convolution layer. This enhances computational efficiency and improves pattern recognition. The FC layer represents the final stage of the model, integrating extracted features for classification (Yamashita et al., 2018). This layer fully connects and flattens the processed feature maps for classification, enabling the CNN to accurately categorize images based on extracted features. This enables CNNs to be highly effective for tasks like object detection and image classification (Bouvrie, 2006).

3.2 VGG16

Visual Geometry Group 16 (VGG16), a prominent CNN framework as illustrated in Figure 3. It is recognized for its outstanding performance in computer vision tasks, particularly those related to image analysis (Rezende, 2018). The creators of VGG16 carefully examined existing networks and made notable improvements by adopting a compact (3 × 3) convolution filter architecture, which surpassed earlier cutting-edge model designs. The "16" in VGG16 refers to its depth, consisting of 16 layers with trainable weights. With its extensive structure, VGG16 contains a staggering 138 million parameters, making it one of the largest and most powerful neural networks in computer vision. VGG-16 processes fixed-size 244×244 pixel RGB images, with each pixel's RGB value pre-processed by subtracting its mean before inputting into the network.

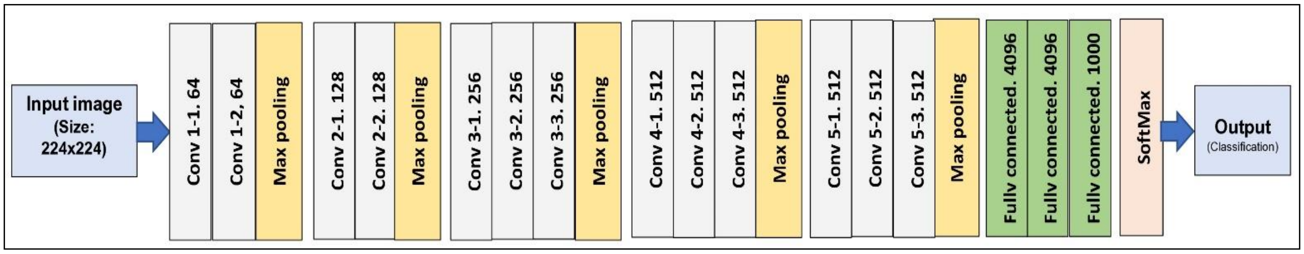


Fig. 3. VGG16 architecture.

After preprocessing, images pass through convolutional layers with 3x3 filters, and in some cases, 1x1 filters for linear transformations followed by non-linear activation (Qassim et al., 2017). The convolution operation uses a default stride of 1 for accurate feature extraction. Following several convolutional layers, five max-pooling layers help shrink the size of the data, making processing faster and more efficient (Xie et al., 2027).

3.3 VGG19

Visual Geometry Group 19 (VGG-19) is a deep neural network (DNN) consisting of 19 weighted layers, that include 16 convolutional layers and 3 FC layers. Figure 4 illustrates the structure of its architecture.

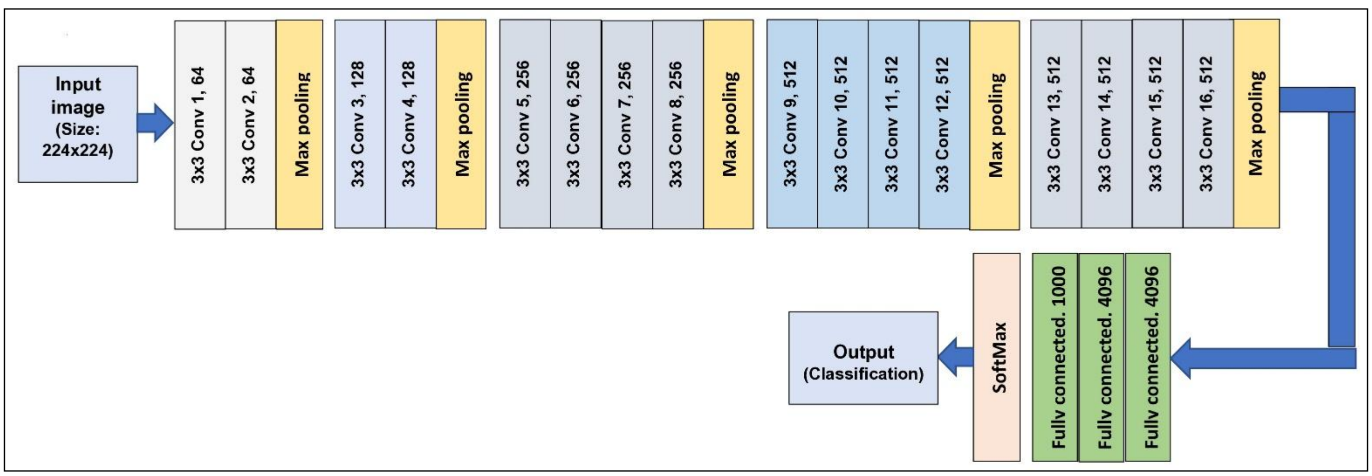


Fig. 4. VGG19 architecture.

It's simple yet deep architecture showcased that enhancing network depth can greatly boost performance in image recognition tasks. While the deeper connectivity of the VGG architecture enhances model performance, it may encounter the "Vanishing Gradient" issue due to the increased network depth. The training time for VGG is quite extensive (Khan et al., 2021). Despite this, the architecture has delivered excellent results in straightforward application in classification and object recognition, thereby a go-to choice for deep learning applications.

3.4 Resnet-50

ResNet, abbreviated as Residual Network, is a unique type of CNN highly regarded for its innovative approach to managing deep network structures. It incorporates the principle of residual learning facilitates efficient training of exceptionally deep neural models while addressing the problem of gradient disappearance (Akiba et al., 2017). This issue previously limited deep networks by obstructing gradient propagation during backpropagation. The solution is the use of shortcut connections, also called skip or identity mappings. ResNet uses shortcut connections to bypass certain layers, forming residual blocks that retain original information while enabling the network to learn residual features. This design converts a traditional network into a residual network, forming the basis for various ResNet variants with different layer counts. ResNet-50, with its 50 layers, is one of the most commonly used variants.

ResNet-50 consists of 48 convolutional layers, along with one MaxPool layer and one average pooling layer. Figure 5 illustrates the architecture of ResNet-50.

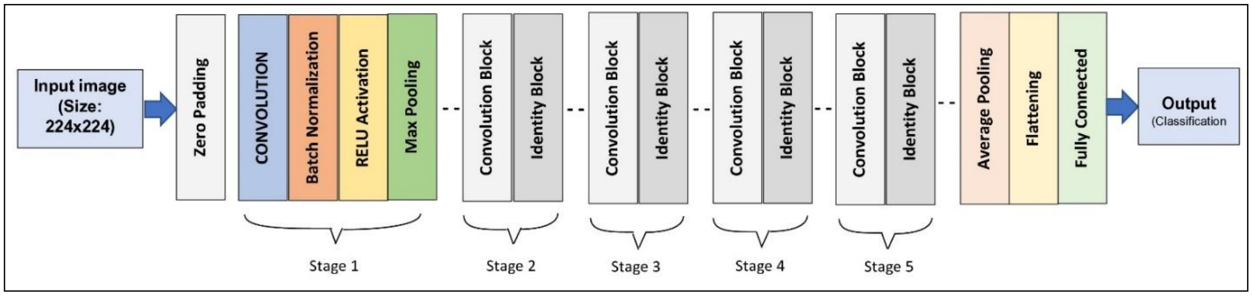


Fig. 5. Resnet-50 architecture.

It showcases the effectiveness of stacking multiple residual blocks to build a strong and expressive CNN. With its deep architecture, ResNet-50 is highly effective at learning intricate structures and attributes extracted from data, rendering it highly suitable for demanding computer vision applications like image recognition, identification of objects, and segmentation (Sankupellay & Konovalov, 2018, Mikami et al., 2018). The ResNet-50 architecture has established itself as a standard model for assessing reliability of different DL models in image analysis, demonstrating its adaptability and dependability. The ResNet series began with the ResNet-34 architecture, featuring 34 layers. It introduced shortcut connections, which enhanced CNN depth, improved information flow during training, reduced vanishing gradients, and facilitated better optimization, paving the way for advancements in residual learning.

3.5 MobileNet

MobileNet, introduced by Google in 2017, is a compact CNN designed for enhanced performance and speed with minimal computational requirements. This makes it well-suited for mobile and embedded systems with limited processing power. The architecture utilizes Depth-wise Separable Convolution, which divides conventional convolutions into two stages: a depth-wise convolution for filtering individually for each input channel and a convolution at the point level for combining these filtered channels. This architecture greatly minimizes the count of parameters and computational steps, resulting in faster and compact model. Its versatile architecture, shown in Figure 6, is predominantly used for tasks like object detection, classification, face analysis, and localization.

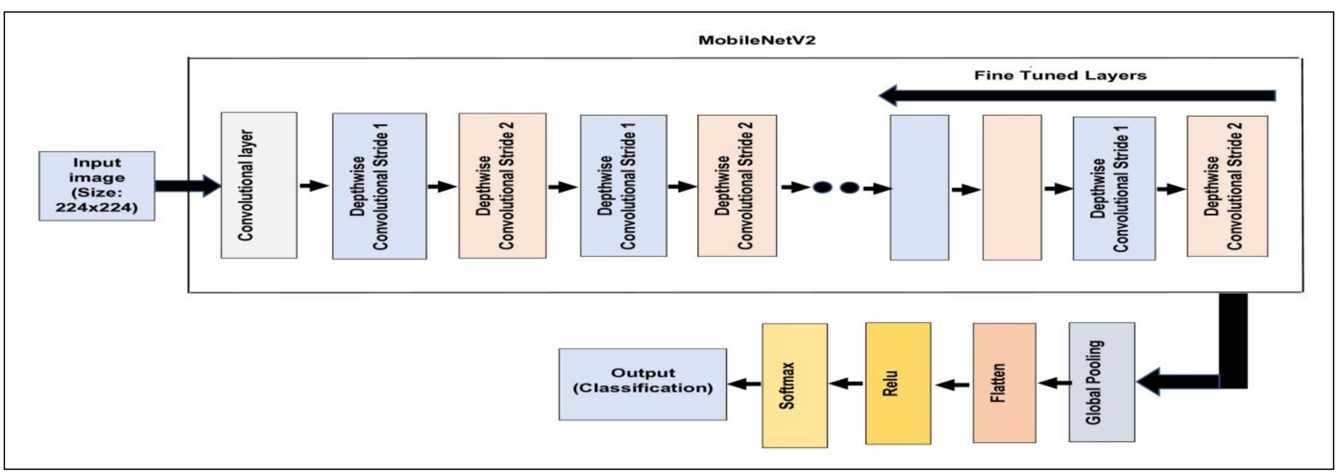


Fig. 6. MobileNet architecture.

Additionally, MobileNet (Mabrouk, 2022) incorporates adjustable hyperparameters—α (width multiplier) and ρ (resolution multiplier), which enable users to modify the model's size and computational demands, making it adaptable to different resource limitations. MobileNet has three versions: V1, V2, and V3, each enhancing accuracy and efficiency. It is supported by major ML frameworks and offers pre-trained weights, making it ideal for practical computer vision applications.

3.6 NasNet-Mobile

NasNet-Mobile is a pre-configured neural network built for use in mobile and low-power environments. It is a compact version of the NasNet (Neural Architecture Search Network) models, offering a practical equilibrium between computational operational efficiency and prediction reliability. This model has been trained on a subset of the ImageNet dataset to enhance its performance across a diverse array of tasks in image classification. The stem cell layer, made up of a Conv2D layer with kernel of size 3x3 and 32 filters, reduces the spatial dimensions of the input image to enable efficient feature extraction and conserve computational resources. The architecture includes six Normal Cells, which maintain the spatial extent of the feature maps. Each cell incorporates depth wise separable convolutions (3x3), inverted residual blocks (1x1 followed by 3x3), and squeeze-and-excitation blocks, enabling the model to capture complex features efficiently with reduced computational overhead. The NasNet-50 architecture is depicted in Figure 7.

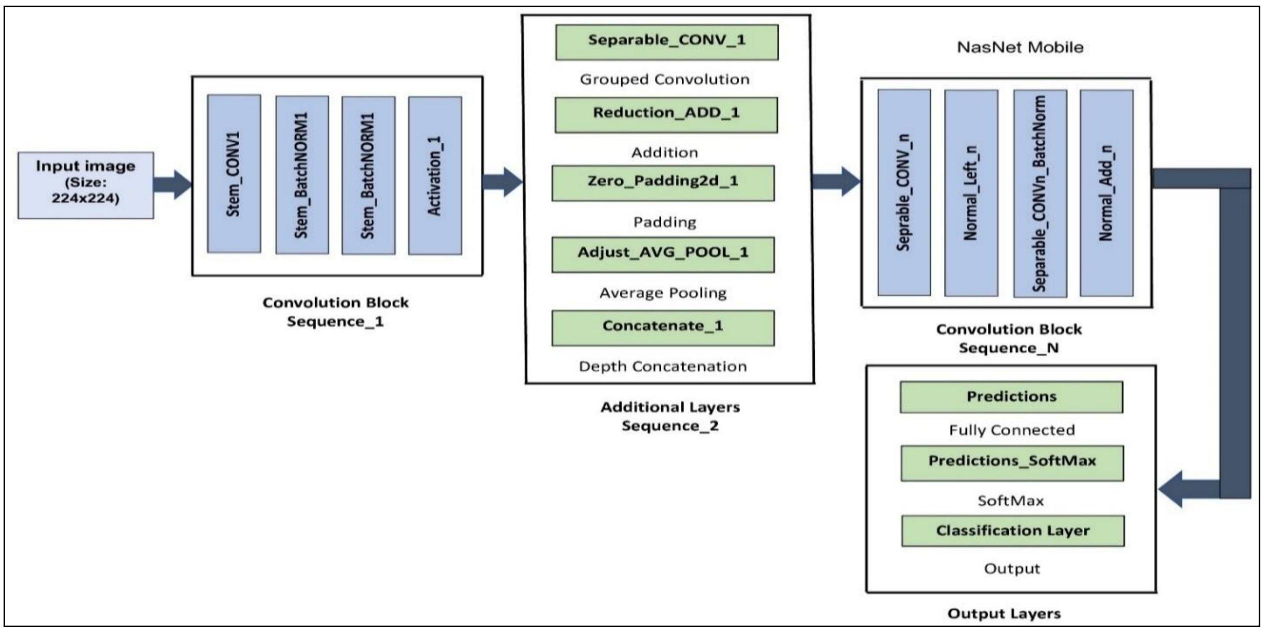


Fig. 7. NasNet-Mobile architecture.

The model includes two Reduction Cells, which use depth wise separable convolutions (3x3) and inverted residual blocks to halve the spatial dimensions and double the feature channels, effectively condensing essential patterns for deeper layers. The top layers consist of a Global Average Pooling (GAP) layer that condenses each feature map into a single value, subsequently, a Dropout layer with a 0.5 rate is applied to reduce overfitting, and a Dense layer with one neuron and sigmoid activation to produce binary classification outputs (Naskinova, 2023).

3.7 DenseNet

DenseNet architecture was engineered to fix the vanishing gradient issue in deep learning models ensuring better information flow between input and output layers (Pillai, 2022). In a neural network with N layers, a standard architecture features N connections, whereas a DenseNet model has N(N+1)/2 connections, greatly improving the inter-layer connectivity. Figure 8 shows the DenseNet architecture.

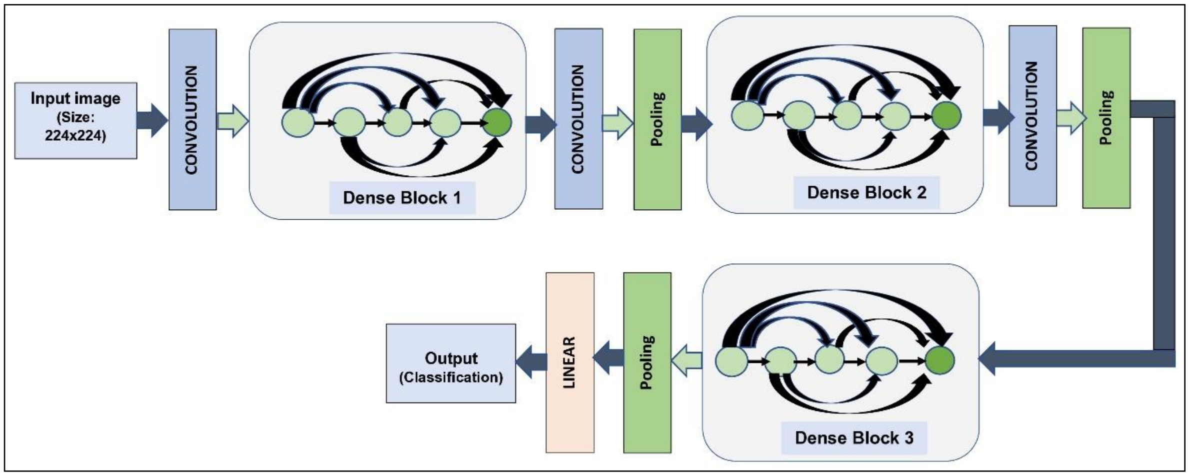


Fig. 8. DenseNet architecture.

As the network depth increases, the number of connections becomes unsustainable, with each layer receiving inputs from all preceding layers. For instance, in a network with ten layers, the 10th layer uses feature maps from all nine previous layers. If every layer generates 128 feature maps, it leads to a rapid increase in the number of feature maps. To address this, dense blocks are introduced, containing a predefined number of layers. The output from a dense block is passed through a transition layer, which applies 1x1 convolution followed by max pooling to reduce feature map size. DenseNet improves information flow by connecting each layer to all preceding layers, enabling enhanced learning compared to traditional architectures.

4 Results

The study focusses on the crucial role of CNN and DL techniques in Pneumonia detection. Table 1 shows the evaluation metrics of different DNN architectures. It has been observed that among all deep learning models, the top three performers in terms of accuracy are VGG-16, VGG-19 and ResNet-50, as evaluated on the chest X-ray image dataset. Using CNN, especially VGG-16, VGG-19 and ResNet-50, our study investigated their potential in classifying images of chest X-ray into pneumonia and normal conditions accurately. In case of VGG-16 and VGG-19, testing accuracies of 91.02% and 90.38% were observed, whereas in case of ResNet-50, testing accuracy of 89.94% was noticed.

Table 1. Evaluation metrics of different deep neural network architecture

Model	Class (False/True)	Precision	Recall	F1-Score	Accuracy in percent - % (average)
Baseline CNN	0	0.62	0.78	0.65	
	1	0.63	0.74	0.66	**88.89**
VGG-16	0	0.91	0.85	0.88	
	1	0.91	0.95	0.93	**91.02**
VGG-19	0	0.89	0.83	0.86	
	1	0.88	0.94	0.92	**90.38**
ResNet-50	0	0.93	0.76	0.83	
	1	0.87	0.96	0.91	**89.94**
MobileNet	0	0.61	0.75	0.63	
	1	0.62	0.74	0.64	**89.42**
NasNet-Mobile	0	0.82	0.87	0.86	
	1	0.84	0.89	0.87	**84.13**
DenseNet	0	0.62	0.76	0.67	
	1	0.60	0.72	0.65	**85.58**

To assess the reliability of our classification approach, we analysed key metrics such as precision, recall, F1-score, and accuracy, as outlined in Table 1. Accuracy measures overall correct predictions, but in imbalanced datasets, it may not reflect the model's ability to distinguish classes equally. In medical image classification, precision and recall are more informative. Precision shows the accuracy of positive predictions, while recall measures the percentage of actual positives correctly identified. Additionally, it is it's important to consider metrics like F1-Score rather than just accuracy, as it helps ensure precise identification of both diseased and non-diseased individuals. In the pneumonia detection task, the performance of the proposed technique is typically assessed by calculating well-known performance metrics, namely Confusion Matrix (CM). In binary classification, the terms true positive (TP), true negative (TN), false positive (FP), false negative (FN) are defined based on the correctness of predictions. The CM for VGG-16 and VGG-19 is shown Figure 9 and Figure 10.

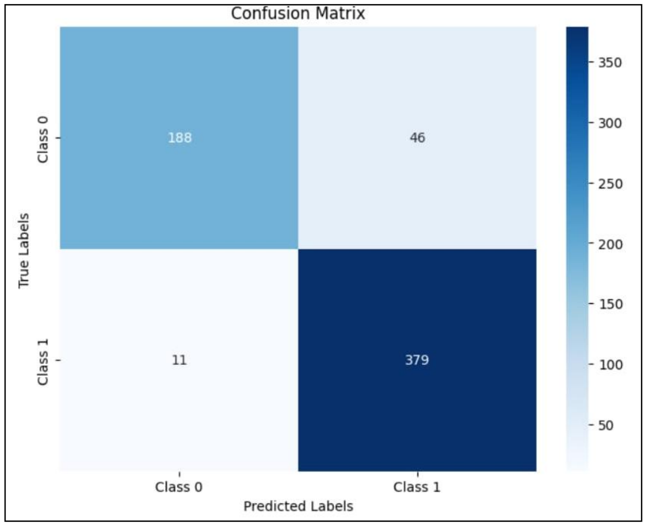


Fig. 9 Confusion Matrix using VGG-16

In case of VGG-16. it can be observed that out of 624 test X-ray images, 379 images have been truly predicted as pneumonia affected images, whereas188 images have been truly predicted as normal images. Similarly, for VGG-19, 379 images have been accurately classified as pneumonia affected images, whereas 186 images have been correctly classified as normal images. Using the above two VGG-16, VGG-19 architecture only 11 images has been predicted as normal images, though they were pneumonia affected.

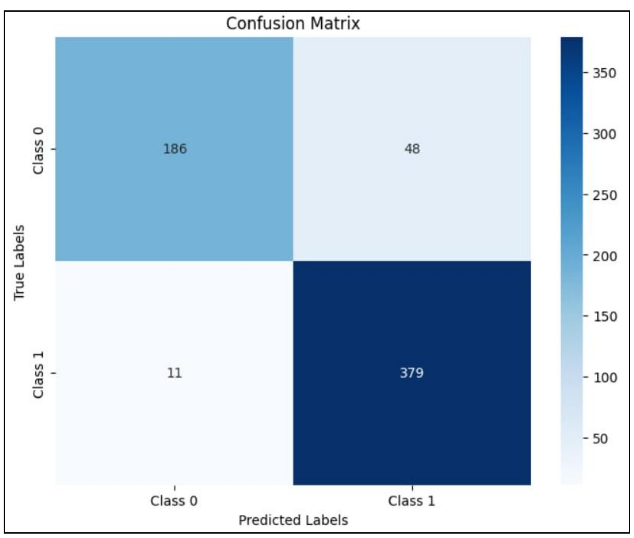


Fig. 10 Confusion Matrix using VGG-19

The Confusion Matrix (CM) for ResNet-50 is shown in Figure 11.

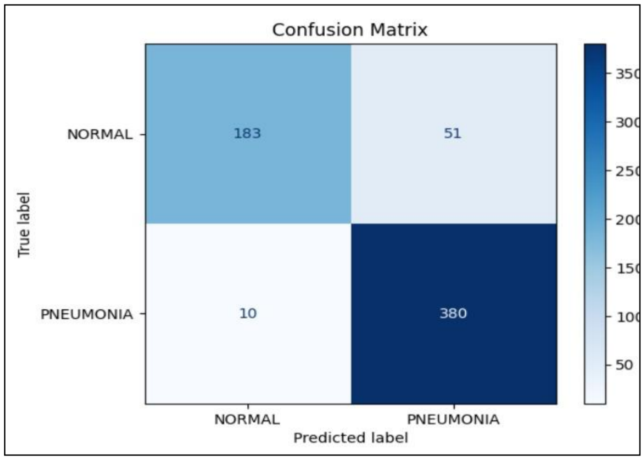


Fig. 11 Confusion Matrix using ResNet-50

Using the ResNet-50 architecture, it can be observed that 380 images were correctly identified as pneumonia affected, whereas 183 images were correctly identified as normal images. Additionally, it was observed that 10 images have been predicted as normal images, though they were pneumonia affected, whereas 51cases of pneumonia affected patients were wrongly predicted as normal patients.

The VGG-16 used for the research work has a total of 21,137,729 parameters. Out of these total parameters 6,423,041 are trainable parameters and 14,714,688 are non-trainable parameters. While plotting the Training vs. Validation Accuracy graph and Training vs. Validation Loss graph for VGG-16 , VGG-19 and ResNet-50, the model has been run with 30 epochs. Figure 12(a) shows the Training vs. Validation Accuracy graph, Figure 12(b) displays the Training vs. Validation Loss graph using VGG-16. Whereas Figure 13(a) shows the Training vs. Validation Accuracy graph, Figure 13(b) displays the Training vs. Validation Loss graph using VGG-19.

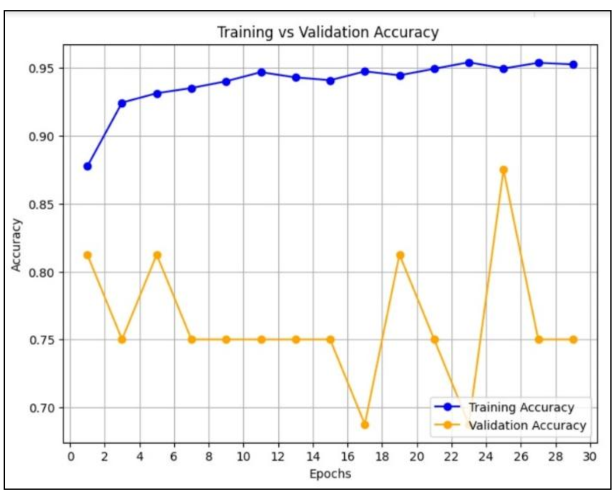


Fig. 12(a). Training vs. Validation Accuracy using VGG-16

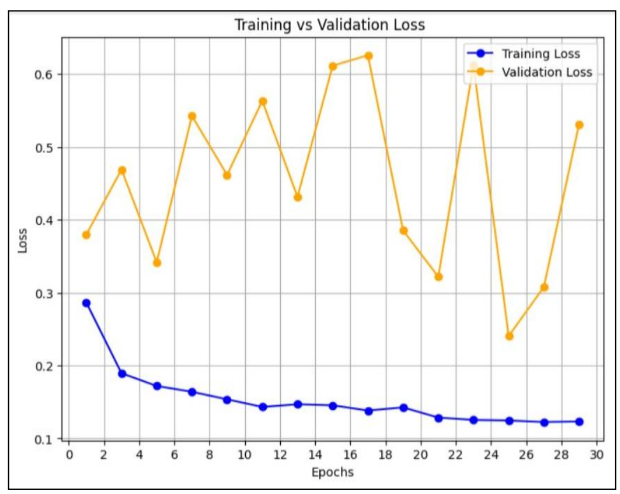
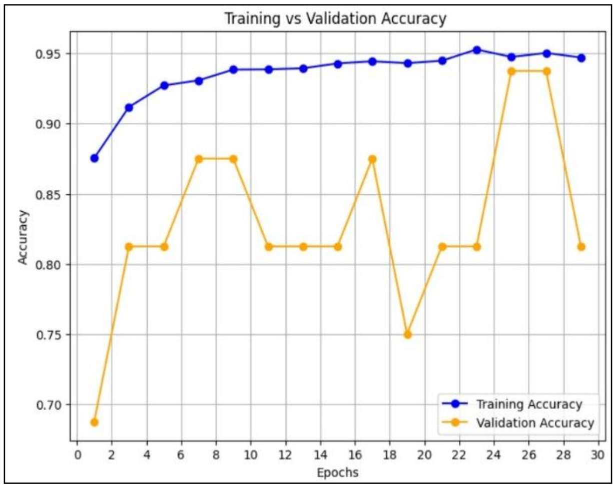


Fig. 12(b). Training vs. Validation Loss using VGG-16

It can be observed from Figure 12(a) and Figure 12(b) that for VGG-16 model, the training accuracy obtained was 0.9413 and training loss achieved is 0.1666. Another key aspect of this model is that it achieved a validation accuracy of 0.7613 and a validation loss of 0.4593.



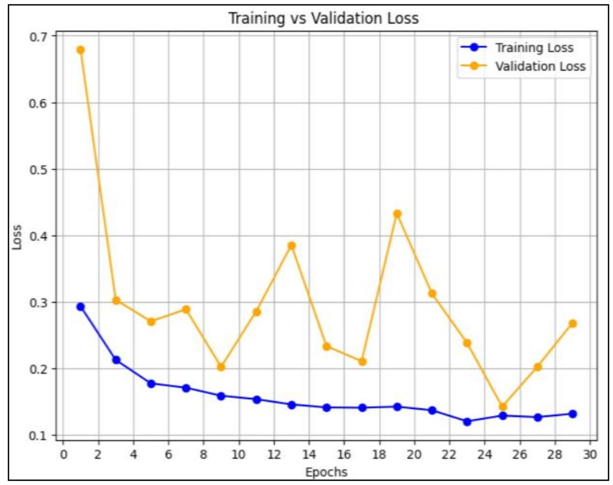
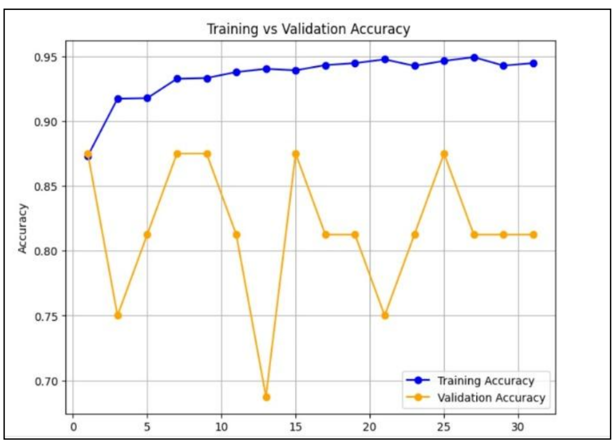


Fig. 13(a). Training vs. Validation Accuracy using VGG-19

Fig. 13(b). Training vs. Validation Loss using VGG-19

It can be observed from Figure 13(a) and Figure 13(b) that for VGG-19 model, the training accuracy obtained was 0.9326 and training loss achieved is 0.1716. It's worth mentioning that the same model achieved a validation accuracy of 0.8276 and a validation loss of 0.2996.

Figure 14(a) shows the Training vs. Validation Accuracy graph, and Figure 14(b) show the Training vs. Validation Loss graph using ResNet-50.



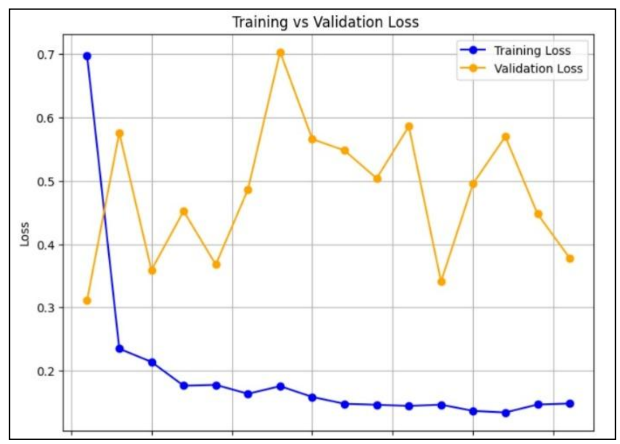


Fig. 14(a). Training vs. Validation Accuracy using ResNet-50

Fig. 14(b). Training vs. Validation Loss using ResNet-50

It can be observed from Figure 14(a) and Figure 14(b) that for ResNet-50 model, the training accuracy obtained was 0.9333 and training loss achieved is 0.2026. Another key point is that for the same model, the validation accuracy obtained was 0.815, and the validation loss achieved was 0.484 respectively.

Table 2 shows the Performance comparison of DL models (Pugliesi, 2019 for VGG-16), (Pugliesi, 2019 for ResNet-50) and the proposed DL work with VGG-16, VGG-19 and ResNet-50 as the backbone network. In this work, an average accuracy of 91.02% was achieved for VGG-16, an accuracy of 90.38% is obtained for VGG-19, and an accuracy of 89.74% was achieved for ResNet-50 models. The results clearly indicate that our approach for the VGG-16 deep learning model outperforms the work of Pugliesi (2019), in terms of accuracy. Additionally, our ResNet-50 deep learning model also demonstrates superior accuracy compare to Pugliesi (2019) work.

Table 2. Comparative analysis with the work (Pugliesi, 2019) for VGG-16, Resnet-50 and our proposed DL models.

Parameters Measured	DL Models – Classification (Pugliesi, 2019) VGG-16	DL Models – Classification (Pugliesi, 2019) ResNet-50	Proposed (Backbone N/W: VGG-16)	Proposed (Backbone N/W: VGG-19)	Proposed (Backbone N/W: ResNet-50)
Precision	NA	NA	**0.91***	0.88	0.87
Recall	NA	NA	0.95	0.94	**0.96***
F1- score	NA	NA	**0.93***	0.92	0.91
Accuracy in percent - % (average)	**74.9 %**	**88.9 %**	**91.02 %**	**90.38 %**	**89.74 %**

NA – Not available, * - best performing values are indicated in bold. Accuracy – values are indicated in bold.

It can be observed from Table 2, that the precision achieved is 0.91 for VGG-16 DL model, which is higher than the precision of VGG-19 and ResNet-50 models. Additionally, it is noteworthy that the ResNet-50 model achieved a recall of 0.96, which is higher than that of VGG-16 and VGG-19 models. Similarly, F1-score achieved is 0.93 for VGG-16 model, which is higher than the F1-score of VGG-19 and ResNet-50 models.

In our work, we utilize pre-trained DL models, like VGG-16, VGG-19, ResNet-50, MobileNet, NasNet Mobile, and DenseNet. All code is executed on a Kaggle notebook with a P100 GPU, ensuring efficient training and evaluation of the models. The code leverages essential libraries such as TensorFlow, Keras, Sklearn, and Seaborn for model training, evaluation, and visualization. To enhance model generalization and prevent overfitting, we apply data augmentation to the training dataset using ImageDataGenerator. To enhance the dataset, we apply various augmentation techniques, including scaling pixel values to the [0,1] range, rotating images randomly up to 20 degrees, shifting them horizontally and vertically up to 20%, applying shear transformations nearly 20%, adjusting zoom levels up to 20%, and flipping them horizontally. Additionally, we use fill_mode to smoothly fill any gaps created during these transformations. For the validation dataset, only rescaling is applied to maintain consistency and ensure fair model evaluation. The key parameters used in our model training for enhanced transparency and reproducibility are the Learning rate which is set at 0.0001, Batch size is 32, Optimizer used is Adam, Regularization technique i.e. Dropout rate = 0.5, Loss function used is Binary Cross-Entropy, Activation functions is ReLU for the dense layer with 256 units, and Sigmoid for the output layer, suitable for binary classification. This method helps transfer learning models adapt better, making them more accurate when handling new data.

5 Conclusions

The study highlights the significant impact of medical imaging and deep learning in diagnosing pediatric pneumonia. The research utilizes standardization and various pre-processing techniques to improve data quality, thoroughly analysing their influence on the effectiveness of the CNN model. Exploring alternative pre-processing methods could lead to varying outcomes and conducting a sensitivity analysis on different techniques would further reinforce the study's conclusions. The research exclusively assesses the effectiveness of CNN frameworks on the dataset utilized for training and testing. Our study explored the effectiveness of CNN, specifically VGG-16, VGG-19, and ResNet-50, in classification of X-ray images of the chest into pneumonic and normal condition. For VGG-16 and VGG-19, testing accuracies of 91.02% and 90.38% were achieved, with corresponding testing losses of 25.22% and 28.90%, respectively. These models outperformed ResNet-50, which recorded an accuracy of 89.94% and a loss of 35.42% during testing. Although the study compares VGG-16, VGG-19, and ResNet-50, it does not evaluate their performance against other pneumonia detection approaches, such as traditional machine learning algorithms or radiologist assessments. Comparing these models with other approaches would provide a broader perspective on the advantages of deep learning frameworks for this specific implementation.

The research does not consider the potential impact of false positives on clinical decision-making. False positives could result in unnecessary follow-up tests or treatments, potentially straining healthcare resources and affecting patient well-being. A thorough analysis of false positives and their implications would enhance the study's relevance and applicability in clinical practice. Our research contributes to growing evidence supporting the use of DL models for pediatric pneumonia detection. The promising results achieved with the VGG-16, VGG-19, and ResNet-50 models highlight their potential in the medical imaging field, offering significant improvements in patient care. Accurate and efficient pneumonia diagnosis in pediatric patients is essential for timely treatment and improved health outcomes. As technology progresses, integrating DL architectures into clinical practice could result in more effective disease detection and improved resource allocation within healthcare settings.

In the future, we plan to develop a web-based application to make pneumonia detection more accessible and efficient for healthcare professionals. This platform will enable users to upload images of chest X-ray and receive instant insights powered by deep learning models. By developing an interface and ensure reliable performance, we hope that the gap will be reduced between medical practice and research. In the future, we also aim to expand the system's capabilities to detect other respiratory conditions, making it a more comprehensive tool for early diagnosis and better patient care.

References

Abiyev, R. H., & Ma'aita H, M. K. S. (2018). Deep convolutional neural networks for chest diseases detection. *Journal of Healthcare Engineering, 2018*(1), 4168538.

Aghdam, H. H., & Heravi, E. J. (2017). *Guide to convolutional neural networks*. Springer.

Akiba, T., Suzuki, S., & Fukuda, K. (2017). Extremely large minibatch SGD: Training ResNet-50 on ImageNet in 15 minutes. *arXiv preprint* arXiv:1711.04325.

Al-Antari, M. A., Al-Masni, M. A., Choi, M. T., Han, S. M., & Kim, T. S. (2018). A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *International Journal of Medical Informatics, 117*, 44–54.

Al-Masni, M. A., Al-Antari, M. A., Park, J. M., Gi, G., Kim, T. Y., Rivera, P., … & Kim, T. S. (2018). Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Computer Methods and Programs in Biomedicine, 157*, 85–94.

Ayan, E., & Ünver, H. M. (2019, abril). Diagnosis of pneumonia from chest X-ray images using deep learning. En *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)* (pp. 1–5). IEEE.

Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., & Greenspan, H. (2015, abril). Chest pathology detection using deep learning with non-medical training. En *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)* (pp. 294–297). IEEE.

Bouvrie, J. (2006). Notes on convolutional neural networks.

Chan, J. (2021). COVID-19 chest X-ray challenge. *Kaggle*. https://kaggle.com/competitions/csc532

Chandra, T. B., & Verma, K. (2020). Pneumonia detection on chest X-ray using machine learning paradigm. En *Proceedings of 3rd International Conference on Computer Vision and Image Processing: CVIP 2018, Vol. 1* (pp. 21–33). Springer Singapore.

Chen, A., Jaegerman, J., Matic, D., Inayatali, H., Charoenkitkarn, N., & Chan, J. (2020, noviembre). Detecting COVID-19 in chest X-rays using transfer learning with VGG16. En *CSBio'20: Proceedings of the Eleventh International Conference on Computational Systems-Biology and Bioinformatics* (pp. 93–96).

Dixit, R. R. (2018). Factors influencing healthtech literacy: An empirical analysis of socioeconomic, demographic, technological, and health-related variables. *Applied Research in Artificial Intelligence and Cloud Computing, 1*(1), 23–37.

Elshennawy, N. M., & Ibrahim, D. M. (2020). Deep-pneumonia framework using deep learning models based on chest X-ray images. *Diagnostics, 10*(9), 649.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115–118.

Hermann, S. (2014). Evaluation of scan-line optimization for 3D medical image registration. En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3073–3080).

Kermany, D. (2018). Labeled optical coherence tomography (OCT) and chest X-ray images for classification [Dataset]. *Mendeley Data*.

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., … & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell, 172*(5), 1122–1131.

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., … & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell, 172*(5), 1122–1131.

Khan, M. A., Rajinikanth, V., Satapathy, S. C., Taniar, D., Mohanty, J. R., Tariq, U., & Damaševičius, R. (2021). VGG19 network assisted joint segmentation and classification of lung nodules in CT images. *Diagnostics, 11*(12), 2208.

Krumholz, H. M. (2014). Big data and new knowledge in medicine: The thinking, training, and tools needed for a learning health system. *Health Affairs, 33*(7), 1163–1170.

Lavin, A., & Gray, S. (2016). Fast algorithms for convolutional neural networks. En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4013–4021).

Mabrouk, A., Diaz Redondo, R. P., Dahou, A., Abd Elaziz, M., & Kayed, M. (2022). Pneumonia detection on chest X-ray images using ensemble of deep convolutional neural networks. *Applied Sciences, 12*(13), 6448.

Madani, A., Moradi, M., Karargyris, A., & Syeda-Mahmood, T. (2018, marzo). Chest X-ray generation and data augmentation for cardiovascular abnormality classification. En *Medical Imaging 2018: Image Processing (Vol. 10574, pp. 415–420)*. SPIE.

Mikami, H., Suganuma, H., Tanaka, Y., & Kageyama, Y. (2018). Massively distributed SGD: ImageNet/ResNet-50 training in a flash. *arXiv preprint* arXiv:1811.05233.

Milletari, F., Navab, N., & Ahmadi, S. A. (2016, octubre). V-net: Fully convolutional neural networks for volumetric medical image segmentation. En *2016 Fourth International Conference on 3D Vision (3DV)* (pp. 565–571). IEEE.

Musher, D. M., & Thorner, A. R. (2014). Community-acquired pneumonia. *New England Journal of Medicine, 371*(17), 1619–1628.

Naskinova, I. (2023). Transfer learning with NASNet-Mobile for pneumonia X-ray classification. *Asian-European Journal of Mathematics, 16*(01), 2250240.

O'Shea, K. (2015). An introduction to convolutional neural networks. *arXiv preprint* arXiv:1511.08458.

Pillai, A. S. (2022). Multi-label chest X-ray classification via deep learning. *arXiv preprint* arXiv:2211.14929.

Pugliesi, R. A. (2019). Deep learning models for classification of pediatric chest X-ray images using VGG-16 and ResNet-50. *Sage Science Review of Applied Machine Learning, 2*(2), 37–47.

Qassim, H., Feinzimer, D., & Verma, A. (2017). Residual squeeze VGG16. *arXiv preprint* arXiv:1705.03004.

Rajaraman, S., Candemir, S., Kim, I., Thoma, G., & Antani, S. (2018). Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. *Applied Sciences, 8*(10), 1715.

Rajpurkar, P. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv preprint* arXiv:1711.5225.

Rello, J., & Diaz, E. (2003). Pneumonia in the intensive care unit. *Critical Care Medicine, 31*(10), 2544–2551.

Rezende, E., Ruppert, G., Carvalho, T., Theophilo, A., Ramos, F., & Geus, P. D. (2018). Malicious software classification using VGG16 deep neural network's bottleneck features. En *Information Technology-New Generations: 15th International Conference on Information Technology* (pp. 51–59). Springer International Publishing.

Rudan, I., Boschi-Pinto, C., Biloglav, Z., Mulholland, K., & Campbell, H. (2008). Epidemiology and etiology of childhood pneumonia. *Bulletin of the World Health Organization, 86*, 408–416B.

Sankupellay, M., & Konovalov, D. (2018, noviembre). Bird call recognition using deep convolutional neural network, ResNet-50. En *Proceedings of Acoustics* (Vol. 7, No. 2018, pp. 1–8).

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint* arXiv:1409.1556.

Sirazitdinov, I., Kholiavchenko, M., Mustafaev, T., Yixuan, Y., Kuleev, R., & Ibragimov, B. (2019). Deep neural network ensemble for pneumonia localization from a large-scale chest X-ray database. *Computers & Electrical Engineering, 78*, 388–399.

Stephen, O., Sain, M., Maduh, U. J., & Jeong, D. U. (2019). An efficient deep learning approach to pneumonia classification in healthcare. *Journal of Healthcare Engineering, 2019*(1), 4180949.

Talo, M. (2019, abril). Pneumonia detection from radiography images using convolutional neural networks. En *2019 27th Signal Processing and Communications Applications Conference (SIU)* (pp. 1–4). IEEE.

Ukwuoma, C. C., Qin, Z., Heyat, M. B. B., Akhtar, F., Bamisile, O., Muaad, A. Y., … & Al-Antari, M. A. (2023). A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images. *Journal of Advanced Research, 48*, 191–211.

Varshni, D., Thakral, K., Agarwal, L., Nijhawan, R., & Mittal, A. (2019, febrero). Pneumonia detection using CNN based feature extraction. En *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)* (pp. 1–7). IEEE.

Wu, J. (2017). Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology, Nanjing University, China, 5*(23), 495.

Xiao, Z., Du, N., Geng, L., Zhang, F., Wu, J., & Liu, Y. (2019). Multi-scale heterogeneous 3D CNN for false-positive reduction in pulmonary nodule detection, based on chest CT images. *Applied Sciences, 9*(16), 3261.

Xie, X., Han, X., Liao, Q., & Shi, G. (2017, junio). Visualization and pruning of SSD with the base network VGG16. En *Proceedings of the 2017 International Conference on Deep Learning Technologies* (pp. 90–94).

Xu, S., Wu, H., & Bie, R. (2018). CXNet-m1: Anomaly detection on chest X-rays with image-based deep learning. *IEEE Access, 7*, 446.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into Imaging, 9*, 611–629.