International Journal of Enhanced Research in Medicines & Dental Care (IJERMDC),
ISSN: 2349-1590, Vol. 12 Issue 10, October 2025, Impact Factor: 8.325

A Study on Deep Learning Techniques for Phemonia Detection

G. Arunalatha

Assistant Professor, Department of Computer Science and Engineering, Perunthalaivar Kamarajar Institute of Engineering and Technology (PKIET), Karaikal, Puducherry.

ABSTRACT

Pneumonia is inflammation and fluid in your lungs caused by a bacterial, viral or fungal infection. It makes it difficult to breathe and can cause a fever and cough with yellow, green or bloody mucus. The flu, COVID-19 and pneumococcal disease are common causes of pneumonia. The severity of the disease can range from mild to life-threatening, especially for vulnerable populations such as infants, the elderly, and individuals with weakened immune systems. Treatment depends on the cause and severity of pneumonia. In this paper, the various deep learning techniques used to detect phemonia disease is discussed.

Keywords: Deep learrning, CNN, Phemonia

INTRODUCTION

Pneumonia is a common but potentially serious respiratory infection that affects the lungs. It occurs when the air sacs (alveoli) in one or both lungs become inflamed and filled with fluid or pus, making it difficult to breathe. Pneumonia can be caused by a variety of microorganisms, including bacteria, viruses, and fungi. The severity of the disease can range from mild to life-threatening, especially for vulnerable populations such as infants, the elderly, and individuals with weakened immune systems. Prompt diagnosis and treatment are essential to prevent complications. With proper medical care, most people recover fully, though preventive measures like vaccination and hygiene practices play a key role in reducing the risk of infection.

Pneumonia can be spread in several ways. The viruses and bacteria that are commonly found in a child's nose or throat can infect the lungs if they are inhaled. They may also spread via air-borne droplets from a cough or sneeze. In addition, pneumonia may spread through blood, especially during and shortly after birth. The presenting features of viral and bacterial pneumonia are similar. However, the symptoms of viral pneumonia may be more numerous than the symptoms of bacterial pneumonia. In children under 5 years of age who have cough and/or difficult breathing, with or without fever, pneumonia is diagnosed by the presence of either fast breathing or lower chest wall indrawing where their chest moves in or retracts during inhalation (in a healthy person, the chest expands during inhalation). Wheezing is more common in viral infections.

Pneumonia killed 740 180 children under the age of 5 in 2019, accounting for 14% of all deaths of children under 5 years old but 22% of all deaths in children aged 1 to 5 years. Pneumonia affects children and families everywhere, but deaths are highest in southern Asia and sub-Saharan Africa. Children can be protected from pneumonia, it can be prevented with simple interventions, and it can be treated with low-cost, low-tech medication and care. While most healthy children can fight the infection with their natural defences, children whose immune systems are compromised are at higher risk of developing pneumonia. A child's immune system may be weakened by malnutrition or undernourishment, especially in infants who are not exclusively breastfed. Pre-existing illnesses, such as symptomatic HIV infections and measles, also increase a child's risk of contracting pneumonia.

LITERATURE SURVEY

An approach for dentifying and categorizing pneumonia in chest radiographs is designed. This deep learning object detection and is based on CoupleNet, a fully convolutional network that combines global and local features for object detection. Our method achieves robustness by utilizing a unique ensembling algorithm that mixes bounding boxes from many models and by implementing significant modifications to the training process. We tested our detection algorithm on a dataset of 3000 chest radiographs as part of the 2018 RSNA Pneumonia Challenge; our solution was recognized as a winning submission in a competition that attracted more than 1400 competitors from all over the world.

CheXNe t[2], a 121-layer convolutional neural network is designed to detect pneumonia from chest X-ray images at a level exceeding practicing radiologists. The model was trained on the ChestX-ray14 dataset, which contains over 100,000 chest X-ray images labeled with 14 different thoracic diseases. The authors compared the performance of CheXNet to that of four practicing radiologists on a test set of 420 images, and found that CheXNet achieved a significantly higher F1 score than the average radiologist performance. The authors also extended CheXNet to detect all 14 diseases in the ChestX-ray14 dataset.

A deep learning-based model [3] is designed for differentiating between normal and severe cases of pneumonia using chest X-ray images. The model uses eight pre-trained deep learning architectures - MobileNet, ResNet50, ResNet152V2, DenseNet121, DenseNet201, Xception, VGG16, and EfficientNet. These models were evaluated on two datasets containing 5,856 and 112,120 chest X-ray images. The best performance was achieved by the MobileNet model, with accuracy scores of 94.23% and 93.75% on the two datasets. The paper also analyzes the impact of different hyperparameters like batch size, number of epochs, and optimizers on the model's performance.

An automated system for pneumonia detection [4] using chest X-ray images, employing an ensemble of three CNN models:is developed. GoogLeNet, ResNet-18, and DenseNet-121, enhanced by a novel weighted average ensemble technique. The system achieved impressive accuracy rates of 98.81% on the Kermany dataset and 86.85% on the RSNA dataset, surpassing existing state-of-the-art methods. This reliable computer-aided diagnosis (CAD) system is significant for early pneumonia diagnosis, which is crucial for improving treatment outcomes, and the code is available on GitHub.

An advanced technique [5] for identifying pneumonia using the Vision Transformer (ViT) architecture on a publicly accessible chest X-ray dataset that is accessible on Kaggle. The suggested approach uses the ViT model, which combines transformer architecture and self-attention mechanisms, to extract global context and spatial relationships from chest X-ray pictures. Our tests using the suggested Vision Transformer-based framework show that it can detect pneumonia from chest X-rays with a higher accuracy of 97.61%, sensitivity of 95%, and specificity of 98%. For processing images with varying resolutions, understanding spatial relationships, and collecting global context, the ViT model is better. The framework outperforms convolutional neural network (CNN) based architectures, demonstrating its effectiveness as a reliable pneumonia diagnosis solution.

Traditional methods of detection depend on the interpretation of chest X-rays by expert radiologists, which can vary in reliability. A deep learning model [6] that employs single-shot detectors and multi-task learning for more accurate pneumonia detection, achieving high accuracy in the RSNA Pneumonia Detection Challenge is designed. The model was trained on a balanced dataset of 26,684 X-ray images, utilizing techniques such as heavy augmentations, dropout for regularization, and ensemble methods to enhance performance.

A deep learning method for automatic pneumonia detection utilizing transfer learning [7] to enhance accuracy. It preprocesses chest X-ray images through U-Net segmentation and classifies pneumonia into normal and abnormal categories using pre-trained models like ResNet50, InceptionV3, and InceptionResNetV2. The study evaluates the impact of optimizers Adam and Stochastic Gradient Descent (SGD) with batch sizes of 16 and 32. Comparative analysis shows that the ResNet50 model achieved superior performance with 93.06% accuracy, 88.97% precision, 96.78% recall, and a 92.71% F1-score, surpassing other CNN models including DenseNet-169+SVM and VGG16.

A pneumonia identification model utilizing chest X-ray images has been developed based on an upscaled ResNet50 [8] architecture. Data augmentation techniques were employed to enhance the limited dataset, and transfer learning was incorporated during training. The model aids radiologists in clinical decision-making and was rigorously evaluated for overfitting and generalization errors. It achieved a test accuracy of 98.14% and an AUC score of 99.71 using data from the Guangzhou Women and Children’s Medical Center pneumonia dataset.

Convolutional neural networks [9], or CNNs, have drawn a lot of interest in the classification of diseases. Additionally, pre-trained CNN models' features from extensive datasets are quite helpful for image classification applications. In this study, we evaluate how well pre-trained CNN models perform as feature extractors, which are then used by several classifiers to distinguish between abnormal and normal chest X-rays. We find the best CNN model for the job analytically. The statistical findings show that the use of pretrained CNN models in conjunction with supervised classifier algorithms can be highly advantageous when examining chest X-ray pictures, particularly for the purpose of detecting pneumonia.

An efficient algorithm[10] for locating the locations of lung opacities was pre-trained on the ImageNet dataset and was based on the single-shot detector RetinaNet with Se-ResNext101 encoders. The model's accuracy was raised by implementing a number of changes. To generalize the model, the ensemble of four folds and many checkpoints was unified, the global classification output was specifically included to the model, and extensive augmentations were applied to the data. Ablation research have demonstrated how the suggested methods increase the accuracy of the model.

1. Deep Learning Models:

i) CNN:

Convolutional neural network (CNN/ConvNet) is a type of deep neural network that is most typically used to evaluate visual data. Instead of relying exclusively on matrix multiplications like standard neural networks do, the CNN architecture employs a unique approach known as convolution. Convolutional networks use a method known as convolution, which mixes two functions to demonstrate how one influences the structure of another.A convolution neural network consists of several hidden layers that aid in the extraction of information from images. CNNs use linear algebra principles, specifically convolution operations, to extract features and find patterns in images. Although CNNs are mostly used to process pictures, they can also be programmed to process audio and other signal data. CNNs have a number of layers, each of which identifies a different aspect of an input image. A CNN can include dozens, hundreds, or even thousands of layers, depending on the complexity of the task at hand, with each layer building on the outputs of preceding layers to recognize precise patterns. The procedure begins with sliding a filter designed to detect specific features over the input image, which is known as a convolution operation, hence the name convolutional neural network. This technique generates a feature map that highlights the existence of the detected features in the image. This feature map is then used as an input for the following layer, allowing a CNN to gradually construct a hierarchical representation of the image.

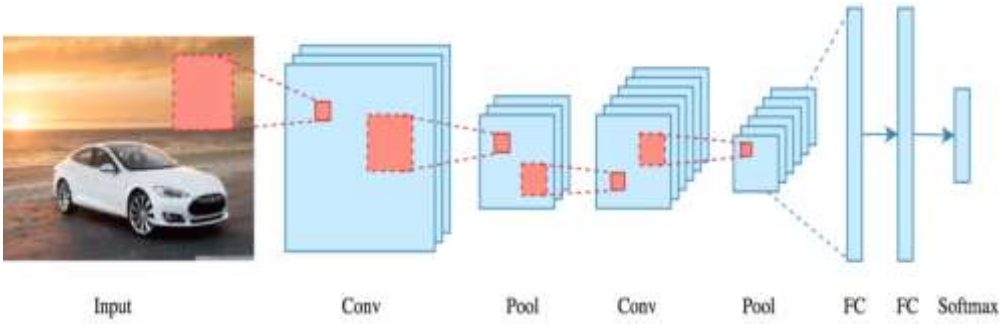

Fig. 1 Convolution Neural Network

Initial filters often recognize basic features like lines or simple textures. The filters in subsequent layers are more complicated, combining previously recognized fundamental elements to recognize more complex patterns. For example, once an early layer recognizes the presence of edges, a deeper layer may use that information to begin distinguishing forms. Between these layers, the network works to minimize the spatial dimensions of the feature maps (height and breadth) in order to enhance efficiency and accuracy. In the last layers of a CNN, the model makes a final conclusion, such as classifying an object in an image, based on the results of the preceding levels

ii) Couplenet:

A framework that links protein sequences and structures to generate insightful protein representations. CoupleNet integrates various levels of features in proteins, encompassing residues and their locations for sequences, along with geometric representations for three-dimensional configurations. We develop two types of graphs to model the derived sequential features and structural geometries, ensuring comprehensiveness on these graphs, respectively, and execute convolution on nodes and edges concurrently to derive enhanced embeddings. The ImageNet pre-trained ResNet-101 serves as the backbone, with the final average pooling layer and the fully connected layer omitted. Subsequently, each proposal branches into two distinct paths: the local FCN and global FCN. Ultimately, the outputs from the global and local FCNs are merged to produce the final object score.

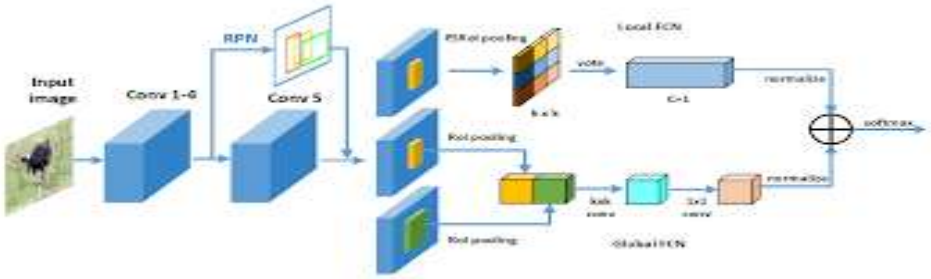

Fig:2 CoupleNet Architecture

iii) ChexNet:

CheXNet is a sophisticated 121-layer convolutional neural network (CNN) trained on the ChestX-ray14 dataset, which comprises over 100,000 frontal-view X-ray images featuring 14 distinct diseases. The task of detecting pneumonia is

framed as a binary classification issue, where the input consists of a frontal-view chest X-ray image X, and the output yields a binary label y, signifying the absence or presence of pneumonia. A 121-layer DenseNet pre-trained on ImageNet is utilized for this purpose. Prior to feeding the images into the network, they undergo downscaling to a dimension of 224×224 and are normalized using the mean and standard deviation from ImageNet. To enhance the dataset, random horizontal flipping is employed as a form of data augmentation.

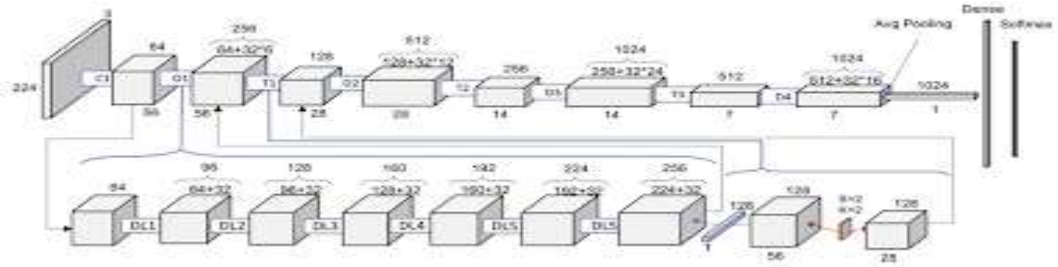

Fig:3 CheXNet

iv) ResNet

ResNet, also known as Residual Network, represents a significant breakthrough in the training of deep convolutional neural networks (CNNs) within the field of computer vision. A key feature of ResNet is the concept of skip connections, which are alternatively referred to as shortcuts. In a conventional CNN, the output from each layer is directed to the subsequent layer without any alternative pathways. As the CNN architecture deepens, the model encounters increasing challenges in learning. This difficulty arises during the backpropagation phase, where the gradients intended for updating the model's weights may either diminish or become ineffective, leading to poorer performance in deeper networks compared to their shallower counterparts.

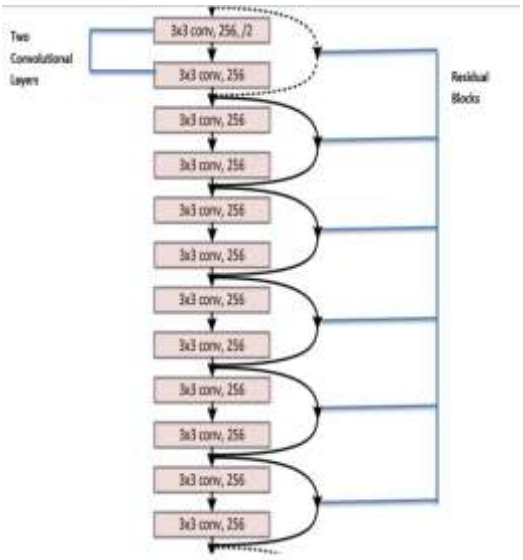

Fig: 4 ResNet

ResNet mitigates this learning challenge by permitting the input from a layer or a set of layers to be combined with the output from that layer. A skip connection effectively creates an alternative pathway for the gradients to propagate, promoting stable training even in networks with numerous layers. Consequently, ResNet emphasizes the learning of residual functions; in other words, each layer is tasked with learning the difference, or residual, between the input and the target output, rather than mastering the entire transformation from input to output. If a particular layer proves to be unhelpful, the model can simply adjust the weights of that layer to a value near zero, allowing the input to be seamlessly routed through the shortcut. This design enables the model to maintain both efficiency and effectiveness in its learning process, even as it grows deeper.

v) Vision transformer

Transformer is a deep learning model that adopts the self-attention mechanism, differentially weighting the significance of each part of the input data. Transformers are increasingly the model of choice for NLP problems, replacing RNN models such as long short-term memory (LSTM). Vision transformer (ViT) is a type of neural network that can be used for image classification and other computer vision tasks. It marks an interesting evolution of various methods that deal with sequential data.

Vision transformers first divide the image into a sequence of patches. Each patch is then represented as a vector. The vectors for each patch are then fed into a transformer encoder. The transformer encoder is a stack of self-attention layers. Self-attention is a mechanism that allows the model to learn long-range dependencies between the patches. This is important for image classification, as it allows the model to learn how the different parts of an image contribute to its overall label. The output of the transformer encoder is a sequence of vectors. These vectors represent the features of the image.

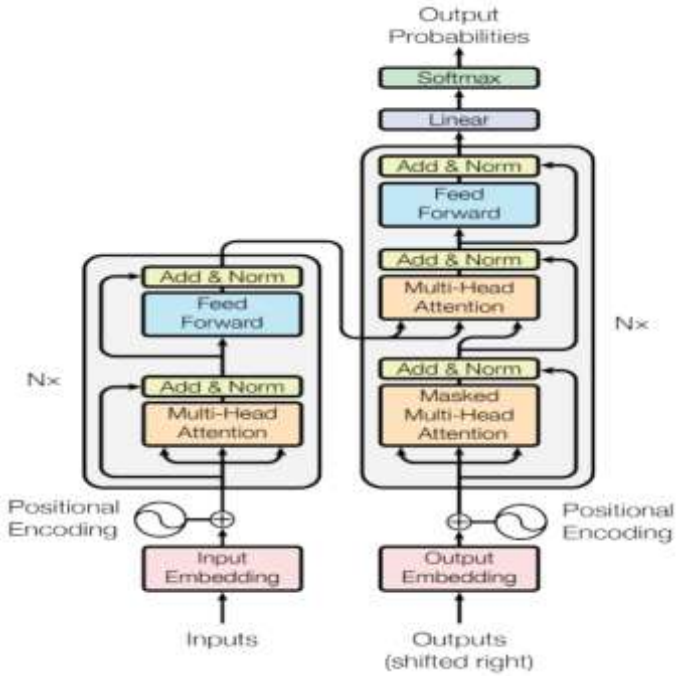

Fig: 5 ResNet

The features are then used to classify the image. Vision Transformer (ViT) achieves remarkable results compared to CNNs while obtaining substantially fewer computational resources for pre-training. In comparison to CNNs, Vision Transformer (ViT) shows a generally weaker inductive bias resulting in increased reliance on model regularization or data augmentation (AugReg) when training on smaller datasets.The ViT is a visual model based on the architecture of a transformer originally designed for text-based tasks. The ViT model represents an input image as a series of image patches, like the series of word embeddings used when using transformers to text, and directly predicts class labels for the image. ViT exhibits an extraordinary performance when trained on enough data, breaking the performance of a similar SOTA CNN with 4x fewer computational resources.

vi) MobileNet

Google created a lightweight deep convolutional neural network called MobileNet. It is designed to work well with computer vision tasks, especially on small devices that have limited processing power, like smartphones. MobileNet uses a special technique called depthwise separable convolution, which is different from regular convolutional networks. This technique splits the standard convolution process into two parts. The first part is depthwise convolution, where each input channel is processed separately. Then, in the second part, called pointwise convolution, the results from each channel are combined. This method keeps the model's accuracy high while making it run faster and use less memory.

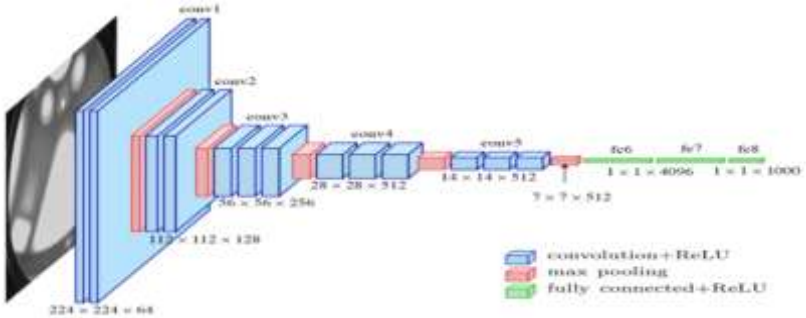

Fig: 6 MobileNet

MobileNet has two important settings, the width multiplier and the resolution multiplier, which let users choose between speed and accuracy. These settings help adjust the model to fit different needs. Since its creation, MobileNet has gone through several improvements. The first version was MobileNetV1. MobileNetV2 came next and made the network more efficient by using inverted residuals and linear bottlenecks. Then MobileNetV3 was released, which further improved performance by using new methods like squeeze-and-excitation blocks and neural architecture search. Because of its high accuracy, small size, and fast performance, MobileNet is often used in real-time applications such as object detection, face recognition, and medical image analysis. This makes it a great choice for many tasks.

vii) GoogLeNet

GoogLeNet was designed to overcome the limitations of earlier convolutional neural network architectures by introducing a new type of building block called the inception module. This concept is significant because it allows the network to process data in parallel across multiple different scales, making it more efficient at capturing a wide range of features.Each inception module is made up of several convolutional layers, each using a different number of filters.

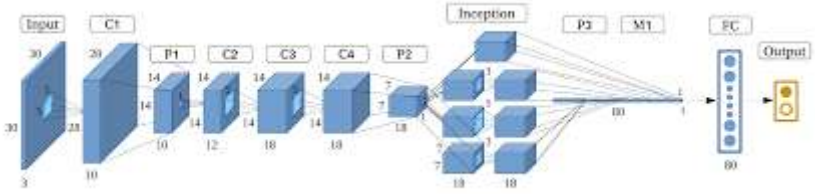

Fig: 7 GoogLeNet

These layers work at the same time, handling the input data at various levels of detail. The outputs from these layers are combined together, and then passed through a pooling layer, which helps to reduce the amount of data while maintaining important features. Over time, different versions of the inception module were developed, and these variations were incorporated into the overall structure of the network. These versions include different combinations of layers and filter sizes, which helped improve the model's performance and adaptability.In GoogLeNet, the entire network is built using a total of nine inception blocks arranged in three main groups. Between each group, a max-pooling layer is included, which helps to reduce the size of the data, making it easier to process. At the end of the network, a global average pooling layer is used to produce the final output.The initial part of the network, known as the stem, is similar to the structures found in earlier networks like AlexNet and LeNet, which helps to provide a solid foundation for the more complex layers that follow.

2. Datasets

i) Kaggele-Chest X-ray images(Phemonia)

There are 5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal). Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children’s Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients’ routine clinical care.

Fig: 8 Kaggele-Chest X-ray images (Phemonia)

ii) RSNA Pneumonia Detection Challenge (2018)

There are 30,000 frontal view chest radiographs from the 112,000-image public National Institutes of Health (NIH) CXR8 dataset

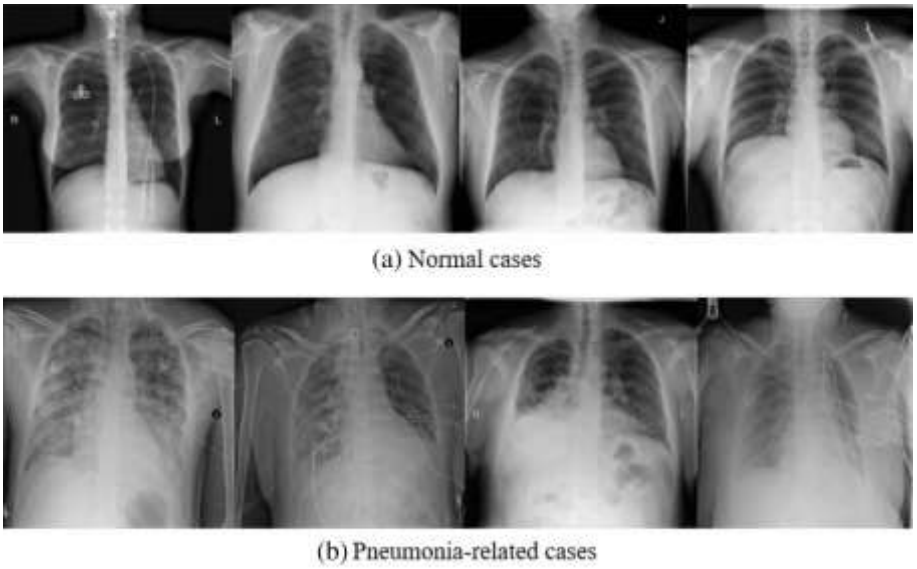

Fig: 9 RSNA Pneumonia Detection Challenge (2018)

CONCLUSION

Pneumonia is an infection in your lungs caused by bacteria, viruses or fungi. Pneumonia causes your lung tissue to swell (inflammation) and can cause fluid or pus in your lungs. Bacterial pneumonia is usually more severe than viral pneumonia, which often resolves on its own. Pneumonia is the single largest infectious cause of death in children worldwide. The various deep learning methods to detect phemonia disesease is presented in this paper. More research needs to be done on the different pathogens causing pneumonia and the ways they are transmitted, as this is of critical importance for treatment and prevention.

REFERENCES

[1]. Team, The DeepRadiology. "Pneumonia detection in chest radiographs." arXiv preprint arXiv:1811.08939 (2018).
[2]. Rajpurkar, Pranav & Irvin, Jeremy & Zhu, Kaylie & Yang, Brandon & Mehta, Hershel & Duan, Tony & Ding, Daisy & Bagul, Aarti & Langlotz, Curtis & Shpanskaya, Katie & Lungren, Matthew & Ng, Andrew. (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. 10.48550/arXiv.1711.05225.
[3]. Reshan, Mana Saleh Al, Kanwarpartap Singh Gill, Vatsala Anand, Sheifali Gupta, Hani Alshahrani, Adel Sulaiman, and Asadullah Shaikh. 2023. "Detection of Pneumonia from Chest X-ray Images Utilizing MobileNet Model" Healthcare 11, no. 11: 1561. https://doi.org/10.3390/healthcare11111561
[4]. Kundu R, Das R, Geem ZW, Han GT, Sarkar R. Pneumonia detection in chest X-ray images using an ensemble of deep learning models. PLoS One. 2021 Sep 7;16(9):e0256630. doi: 10.1371/journal.pone.0256630. PMID: 34492046; PMCID: PMC8423280.
[5]. Singh, S., Kumar, M., Kumar, A. et al. Efficient pneumonia detection using Vision Transformers on chest X-rays. Sci Rep 14, 2487 (2024). https://doi.org/10.1038/s41598-024-52703-2
[6]. Gabruseva, Tatiana, Dmytro Poplavskiy, and Alexandr Kalinin. "Deep learning for automatic pneumonia detection." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020.
[7]. Manickam, Adhiyaman & Jiang, Jianmin & Zhou, Yu & Sagar, Abhinav & Soundrapandiyan, Rajkumar & Samuel, R.. (2021). Automated pneumonia detection on chest X-ray images: A deep learning approach with different optimizers and transfer learning architectures. Measurement. 184. 109953. 10.1016/j.measurement.2021.109953.
[8]. Hashmi, Mohammad Farukh & Katiyar, Satyarth & Hashmi, Abdul Wahab & Keskar, Avinash. (2021). Pneumonia detection in chest X-ray images using compound scaled deep learning model. Automatika. 62. 397-406. 10.1080/00051144.2021.1973297.
[9]. Varshni, Dimpy, et al. "Pneumonia detection using CNN based feature extraction." 2019 IEEE international conference on electrical, computer and communication technologies (ICECCT). IEEE, 2019.