Enhancing Pneumonia Detection Using Deep Learning

Amal Alshahrani[1], Batul Mrakkan[1], Hala Algethami[1], Hanan Alhaj[1], Jana Hariri[1], Waseelah Alamoudi[1]

Faculty of Computers, Umm Al-Qura University, Makkah, Saudi Arabia

Abstract
Pneumonia represents a significant public health challenge. This condition, caused by bacteria, viruses, or fungi, leads to the soreness of the air sacs in one or both lungs, resulting in health issues ranging from mild to life-threatening. The severity of pneumonia needs timely and accurate diagnosis for effective treatment, especially for some cases such as young children, the elderly, and individuals with weakened immune systems or pre-existing health conditions. Neural networks, particularly in the field of medical imaging analysis, have improved the way healthcare professionals diagnose and treat conditions. Neural networks are inspired by the human brain's structure and function, enabling them to learn from vast amounts of data and identify complex patterns that may not be immediately apparent to human observers. We aim to introduce an approach employing advanced deep learning technologies and algorithms, specifically designed to improve the accuracy and efficiency of pneumonia detection, to help healthcare specialists, and to save as many innocent lives as possible.
Keywords:
Deep Learning, Pneumonia, Artificial Intelligence, Classification

1. Introduction

Pneumonia [1], as a global health concern, causes millions of deaths annually, particularly among young children and the elderly. Early diagnosis is crucial for effective treatment and improving survival rates. Traditional diagnostic methods such as chest X-rays are widely used but often require skilled radiologists, whose availability can be limited in resource-constrained areas. In response to these challenges, recent advances in artificial intelligence, particularly deep learning, have shown great promise in automating pneumonia detection [4], which paves the way to optimize the results of automated detection using artificial neural networks and deep learning.

Deep learning models [2], particularly convolutional neural networks (CNNs) [3], have demonstrated exceptional performance in image classification tasks. These models can be trained on large datasets of chest X-rays to automatically identify patterns indicative of pneumonia. However, applying these models requires careful consideration of data quality, annotation accuracy, and model architecture to guarantee that the results are as optimized as possible.

This paper explores the employment and evaluation of various deep-learning models for pneumonia detection. We investigate different architectures, training strategies, and tools to optimize pneumonia detection results. Our research aims to provide a comprehensive comparison of different classification models used in such operations, We will focus on models like YoloV8, YoloV5, and VGG16, while implementing these models in different settings to improve the results as much as possible.

By leveraging publicly available datasets and standardized evaluation metrics, we establish a benchmark for future research and demonstrate the potential of artificial intelligence to transform pneumonia diagnosis. The findings of this study could help guide the development of more accurate, accessible, and efficient diagnostic tools, ultimately improving pneumonia detection outcomes globally.

2. Related Work

The rapid advancements in deep learning have catalyzed significant progress in medical imaging, particularly in disease detection and diagnosis. In recent years, numerous studies have explored the application of convolutional neural networks (CNNs) for detecting pneumonia from chest X-rays, motivated by the potential to automate the diagnostic process and reduce the burden on radiologists. Early approaches leveraged relatively simple architectures, while more recent work has focused on sophisticated models and transfer learning to achieve higher accuracy. This section provides an overview of existing methodologies, their strengths and limitations, and how they have shaped the development of automated pneumonia detection.

Alaa M et al. [4] conducted a study named Diagnosis of Pneumonia Using Deep Learning, where they used convolutional neural networks (CNNs) to diagnose pneumonia from chest X-ray images. Focusing on deep learning, the research outlines the development of a CNN

Manuscript received May 5, 2025
Manuscript revised May 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.5.22

model designed to accurately identify pneumonia cases in X-ray scans. After training the CNN model to distinguish between normal and pneumonia-affected lungs. The model showed good results, and the authors are optimistic about having more accurate diagnoses in the health sector using technology in the future.

Abhay Chopde et al. [5] proposed a study about detecting pneumonia in patients using chest X-ray images based on a Conventional Neural Network (CNN). At the core of this study is the implementation of transfer learning models, such as VGG16, pre-trained on an extensive dataset that contains 5206 images, to classify chest X-ray images into normal or pneumonic categories. In addition, they utilized data augmentation to enlarge the data and prevent overfitting. The accuracy of the VGG16 was 96%.

C. Leracitano et al, [6] responded to the urgent need for efficient COVID-19 pneumonia detection tools. Furthermore, They proposed a novel methodology integrating fuzzy logic with deep learning, detailed in their study, "A fuzzy-enhanced deep learning approach for early detection of COVID-19 pneumonia from portable chest X-ray images." Published in Neurocomputing, this work presents an innovative approach to enhance the interpretability and robustness of deep learning models. The study utilizes convolutional neural networks (CNNs) as the backbone for deep learning and incorporates fuzzy logic to improve model interpretability. By combining fuzzy logic's human-like reasoning with deep learning's computational power, the study aims to overcome the challenges associated with pneumonia detection from chest X-ray images, offering a promising solution to improve patient outcomes in the context of the ongoing pandemic. The results demonstrate that the proposed approach achieved an accuracy of 92.5%, a significant improvement in early detection of COVID-19 pneumonia compared to traditional deep learning methods, highlighting the potential of interdisciplinary approaches in addressing complex healthcare challenges.

Rohit K. et al. [7] conducted research on detecting pneumonia in chest X-ray images using a combination of learning models. They utilized transfer learning to address the availability of data. The study involved an ensemble of three network models. GoogLeNet, ResNet 18, and DenseNet 121. The proposed approach was evaluated using the Kermany dataset,t achieving an accuracy rate of 98.81% and a sensitivity rate of 98.80%. Similarly, evaluation with the RSNA dataset resulted in an accuracy rate of 86.85% and a sensitivity rate of 87.02%.

Mohammad F.et al. [8] explore how deep learning methods can be utilized to detect pneumonia in chest X-ray images. They introduced a revised ResNet50 model known as compound-scaled ResNet50 for pneumonia identification. To boost the training dataset size, they applied data augmentation techniques and utilized transfer learning during model training. The final model achieved a test accuracy of 98.14% and an AUC score of 99.71 on the pneumonia dataset from the Guangzhou Women and Children Medical Center.

Luka Racic et al. [9] proposed a study about detecting pneumonia in patients using chest X-ray images based on a Conventional Neural Network (CNN). The study comprised 5856 chest X-ray images categorized into training, validation, and testing sets. Additionally, pre-processing techniques such as data augmentation and normalization are detailed to enhance the quality and balance of the dataset. They achieved nearly 90% accuracy with the possibility of overfitting due to the size of the dataset.

Kong et al. [10] Recent advancements in deep convolutional neural network (CNN) models have greatly improved pneumonia diagnosis, a leading cause of child mortality worldwide. This study proposes a novel approach combining the Xception CNN with LSTM technology for automatic pneumonia detection in X-ray images. By integrating these frameworks, the model achieves a remarkable 96% accuracy rate, surpassing existing methods. This advancement promises increased reliability in childhood pneumonia classification, aiding clinicians in timely and accurate diagnoses.

Tatiana G et al. [11], in their paper, Deep Learning for Automatic Pneumonia Detection explored the use of deep learning for pneumonia detection from chest X-ray images in their study. By employing single-shot detectors and deep convolutional neural networks with augmentations and multi-task learning, they developed a model to identify pneumonia-related lung opacities. Their model, which combines an SSD RetinaNet with a SE-ResNext101 encoder pre-trained on ImageNet, showed good performance in detecting pneumonia. This study indicates the potential of deep learning methods to enhance diagnostic processes for pneumonia.

Zhenjia Yue et al. [12] conducted a study named Comparison and Validation of Deep Learning Models for the Diagnosis of Pneumonia, which highlights the importance of early detection and treatment of pneumonia. It discusses the limitations of traditional X-ray diagnosis and the potential of deep learning, particularly convolutional neural networks (CNNs), in improving pneumonia detection using chest X-ray images. It also emphasizes the need for improvements in migration learning models and the development of algorithms capable of distinguishing between different lung diseases.

Garima Verma et al. [13] focused on using deep learning, specifically Convolutional Neural Networks (CNNs), for pneumonia classification in chest X-ray images. They highlight the growing interest in deep learning and the availability of large datasets. The proposed CNN model aims to improve pneumonia diagnosis by accurately identifying the disease from X-ray images. Data augmentation techniques and Python programming are used to enhance classification accuracy. The study suggests that deep learning methods outperform traditional approaches to handling complex tasks.

In summary, the body of existing work in pneumonia detection through deep learning has significantly advanced the field, demonstrating the potential for automated diagnosis. Researchers have employed a range of deep learning models, from basic CNNs to more intricate architectures, along with data augmentation and transfer learning techniques.

While many approaches have shown promising results, challenges remain, including the need for diverse and comprehensive datasets, improved interpretability, and consistent performance in clinical settings. This paper builds upon these findings by evaluating a range of deep learning models and comparing their effectiveness on a publicly available dataset, aiming to identify the most suitable approaches for practical implementation in healthcare environments.

3. Materials and Methods

By comparing and evaluating different tools for classifying and detecting normal and pneumonia-affected lungs using a publicly available dataset, this section will provide a detailed explanation of the implementation and evaluation methodology.

3.1 Dataset

The "Chest X-Ray Images (Pneumonia)" [14] dataset comprises 5,863 images across two categories: normal and pneumonia, organized into three folders (train, test, and val) with subfolders for each category. The images, in JPEG format, originate from retrospective cohorts of pediatric patients aged one to five years from the Guangzhou Women and Children's Medical Center, Guangzhou.

All images were screened for quality control, with diagnoses graded by two expert physicians and a third expert checking the evaluation set to ensure accuracy. The dataset aims to facilitate the automated detection and classification of human diseases from medical images, specifically pneumonia through the use of advanced imaging technology. It acknowledges the data source as a Mendeley dataset and is licensed under CC BY 4.0. This dataset has been extensively used for learning, research, and application purposes, demonstrating its significant contribution to the medical and scientific communities in enhancing the understanding and diagnosis of pneumonia through imaging.

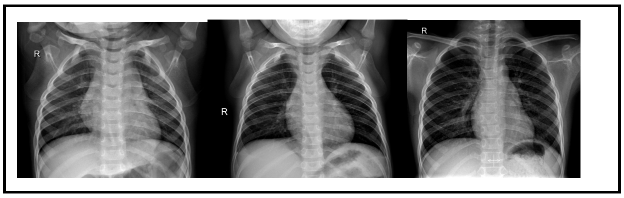

Figure 1: Normal lungs

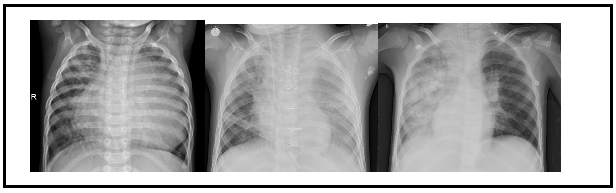

Figure 2: Pneumonia lungs

Figures 1 and 2 show samples of the images used in our training process, for both normal and pneumonia-affected lungs.

3.2 Methodology

This section explains the stages of building a model that predicts the diagnosis of chest X-ray images, which are in the following order: data acquisition, model development, and model evaluation. Figure 3 illustrates the methodology.

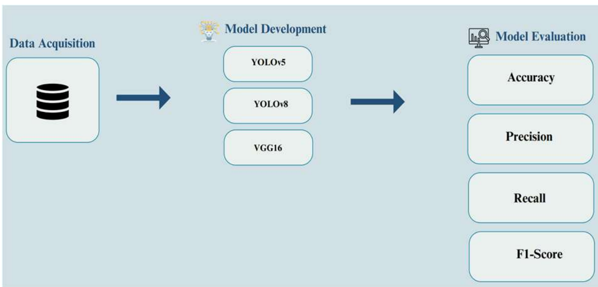

Figure 3: Methodology

We collected our data from the Kaggle website, which comprised approximately 6,000 chest X-ray images. These images were meticulously classified into two distinct categories: normal and pneumonia (infected). The substantial size of this dataset played a critical role in the training phase, as it provided a diverse range of images and clinical cases that enhanced our analysis process, classification accuracy, and ultimately, the ability to predict diagnoses correctly.

To determine the most effective model for our needs, we experimented with two machine learning models to evaluate each's performance in handling medical imaging data. The models we tested included well-known architectures such as YOlO and VGG16. These were implemented with various training epochs to find the best results for classifying.

A. Yolo

YOLO is a popular object detection algorithm in computer vision. It's used to identify objects within images or video frames. Unlike traditional object detection algorithms, YOLO processes the entire image in a single pass, making it faster and more efficient.

The main advantage of YOLO is speed and accuracy. By dividing the image into a grid and predicting bounding boxes and class probabilities for each grid cell, YOLO can efficiently detect multiple objects in real-time. This makes it useful in various applications such as autonomous vehicles, surveillance systems, robotics, and image analysis. YOLO has undergone several iterations and improvements over time, with each version refining its accuracy and speed.

1) YOLOv8

YOLO8 maintains its hallmark feature of processing images swiftly, making it suitable for real-time applications such as autonomous driving, surveillance, and augmented reality. This iteration likely further refines object localization, recognition, and tracking capabilities, cementing YOLO's status as a cornerstone in the field of computer vision. Using YOLOv8, we expanded our experimentation to include this model with varying numbers of training epochs, which showed very promising results.

2) YOLOv5

YOLO5 streamlines the object detection process by dividing images into a grid and predicting bounding boxes and class probabilities directly. This approach enables real-time detection, making it ideal for applications where speed is crucial, such as autonomous vehicles, surveillance systems, and robotics. Which we have also experimented with. to compare the results with newer versions like YOLOv8 and detect which of these models classifies pneumonia-affected lungs better and has the best accuracy and results.

B. VGG16

VGG16 [15] is a deep convolutional neural network known for its simplicity and effectiveness. With 16 layers, it's adept at capturing intricate image features, making it a popular choice for tasks like image classification and object recognition. Despite its computational demands, its impressive performance on benchmarks like ImageNet has solidified its status as a foundational model in computer vision.

Our final set of experiments involved training our model on VGG16 for a relatively short span of 5 epochs. This approach was aimed at assessing the model's quick adaptability and initial performance metrics without extensive training. The optimal outcome was 98.9% accuracy.

C. Evaluation

During the evaluation phase, we applied a series of robust metrics to assess each model's accuracy, sensitivity, and specificity of each model. This allowed us to compare their performance comprehensively in diagnosing pneumonia from normal X-ray images. The evaluation involved not only quantitative metrics but also qualitative analysis to ensure that the models provided reliable and consistent predictions.

3.3 Results

In our investigation, we evaluated various deep-learning models for pneumonia detection using publicly available datasets of chest X-ray images. The primary objective was to assess the performance of these models in accurately identifying pneumonia cases from radiographic images.

A. VGG16

Hyperparameters	Value for Vgg16
image size	224x224

epochs	5
batch size	32
learning rate	0.001
optimizer	SGD
momentum	0.9
weight decay	0.0005

Table 1: VGG16 Parameters

We have applied and evaluated VGG16 on a larger number of epochs, but the performance did not improve. The results we had using 5 epochs gave the best results, as specified in Table 1.

- **VGG16:**

```
history = model.fit(x = train_batches,
          steps_per_epoch = len(train_batches),
          epochs = 5,
          verbose = 2
        )
```

Figure 4: VGG16 training

Training the model on five epochs and the rest of the default settings of VGG16 as specified in Figure 4, the model achieved a high accuracy of 98.9%, as shown in Figure 6. This high accuracy is a result of the model effectively learning and capturing the important features from the training data, leading to accurate predictions.

Additionally, Figure 5 shows the loss result of the model, which initially decreases and then fluctuates. The initial decrease in loss indicates that the model is learning and improving its performance, while the fluctuations are due to the learning rate adjustments and the model's adaptation to the data. Overall, the decreasing trend in loss and the high accuracy suggest that the model is optimizing well.

The loss decreases initially, reflecting the model's learning process as it optimizes its weights to minimize errors. The fluctuations in the loss may be due to the learning rate or the model's ongoing adjustments to the data. Despite these fluctuations, the overall trend of decreasing loss is a positive sign that the model is improving its performance with each epoch.

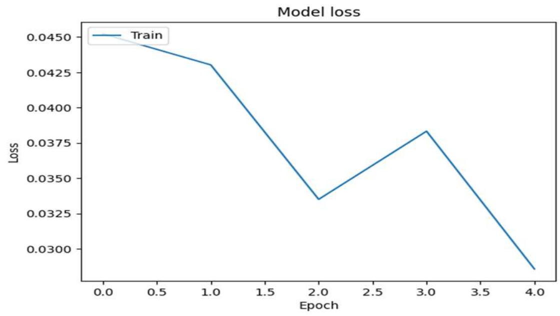

Figure 5: VGG16 loss

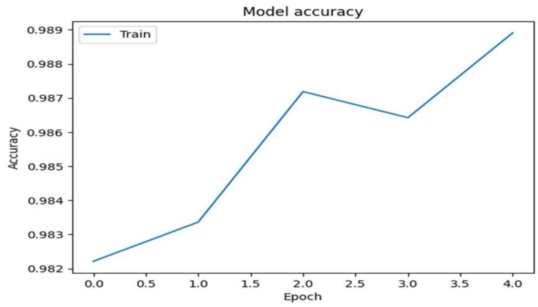

Figure 6: VGG16 Accuracy

B. YOLO

Hyperparameters	Value for YoloV5	Value for YoloV8
image size	64	128
epochs	20	20
batch size	16	32
learning rate	0.000714	0.001
optimizer	AdamW	Adam
momentum	0.9	0.9
weight decay	0.0005	5e-05

Table 2: Yolo Parameters

We trained two YOLO versions while changing the number of epochs, and in Table 2, the best settings that gave the best results are specified.

- **YOLOv8:**

Trained on the first try using 40 epochs and the other settings set to default.

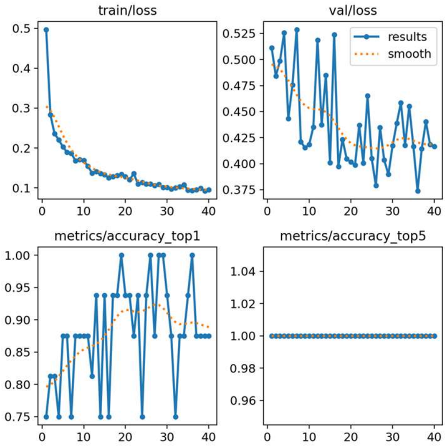

Figure 7: YOLOv8 results

ep	train.	metrics/a	metrics/a	val/	lr/p	lr/p	lr/pg2
1	0.49711	0.75	1	0.51142	0.000237	0.000237	0.000237
2	0.28377	0.8125	1	0.4843	0.000464	0.000464	0.000464
3	0.23576	0.8125	1	0.4985	0.000678	0.000678	0.000678
4	0.22065	0.75	1	0.52557	0.000661	0.000661	0.000661
5	0.20271	0.875	1	0.44323	0.000643	0.000643	0.000643
6	0.18948	0.875	1	0.47563	0.000626	0.000626	0.000626
7	0.18669	0.75	1	0.52888	0.000608	0.000608	0.000608
8	0.16849	0.875	1	0.42099	0.00059	0.00059	0.00059
9	0.1707	0.875	1	0.41534	0.000573	0.000573	0.000573
10	0.16879	0.875	1	0.41832	0.000555	0.000555	0.000555
11	0.15418	0.875	1	0.43518	0.000537	0.000537	0.000537
12	0.13712	0.8125	1	0.51872	0.00052	0.00052	0.00052
13	0.14039	0.9375	1	0.43737	0.000502	0.000502	0.000502
14	0.13601	0.75	1	0.48472	0.000484	0.000484	0.000484
15	0.13281	0.9375	1	0.4011	0.000467	0.000467	0.000467
16	0.12542	0.75	1	0.52374	0.000449	0.000449	0.000449
17	0.12788	0.9375	1	0.39738	0.000431	0.000431	0.000431
18	0.13105	0.9375	1	0.42328	0.000414	0.000414	0.000414
19	0.13384	1	1	0.40455	0.000396	0.000396	0.000396
20	0.1279	0.9375	1	0.40152	0.000378	0.000378	0.000378
21	0.12069	0.9375	1	0.39877	0.000361	0.000361	0.000361
22	0.1357	0.875	1	0.43689	0.000343	0.000343	0.000343
23	0.10959	0.9375	1	0.40056	0.000325	0.000325	0.000325
24	0.11355	0.75	1	0.46503	0.000308	0.000308	0.000308
25	0.11032	0.9375	1	0.40512	0.00029	0.00029	0.00029
26	0.11016	1	1	0.37896	0.000272	0.000272	0.000272
27	0.10605	0.875	1	0.43465	0.000255	0.000255	0.000255
28	0.1091	1	1	0.4034	0.000237	0.000237	0.000237
29	0.10074	1	1	0.38969	0.000219	0.000219	0.000219
30	0.10083	0.9375	1	0.41712	0.000202	0.000202	0.000202
31	0.09725	0.875	1	0.43876	0.000184	0.000184	0.000184
32	0.09978	0.75	1	0.45852	0.000166	0.000166	0.000166
33	0.10267	0.875	1	0.41768	0.000149	0.000149	0.000149
34	0.10785	0.875	1	0.45511	0.000131	0.000131	0.000131
35	0.09429	0.9375	1	0.4166	0.000113	0.000113	0.000113
36	0.093	1	1	0.37378	9.54	9.54	9.5498e-05
37	0.09573	0.875	1	0.41451	7.78	7.78	7.7826e-05
38	0.09875	0.875	1	0.4402	6.01	6.01	6.0154e-05
39	0.09223	0.875	1	0.41839	4.24	4.24	4.2483e-05
40	0.09526	0.875	1	0.41648	2.48	2.48	2.4812e-05

Figure 8: YOLOv8 train details

As provided in Figures 7 and 8, it is shown that using 40 epochs, the model was overfitted and gave a 1 in metrics/accuracy, which means 100% accuracy in the training results, and that means that the model is overfitted.

- **YOLOv8:**

  Trained on the second try using 20 epochs, and the other settings being set as default.

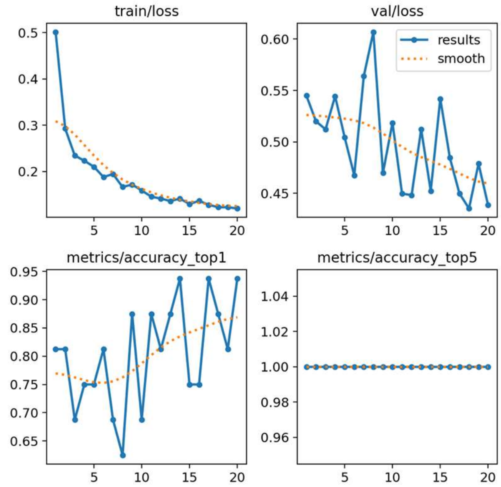

Figure 9: YOLOv8 results

	train/loss	metrics/accuracy_to	metrics/accuracy_to	val/loss	lr/pg0	lr/pg1	lr/pg2
1	0.50134	0.8125	1	0.54492	0.00023727	0.00023727	0.00023727
2	0.29241	0.8125	1	0.52002	0.00045174	0.00045174	0.00045174
3	0.23429	0.6875	1	0.51221	0.00064266	0.00064266	0.00064266
4	0.223	0.75	1	0.54443	0.00060797	0.00060797	0.00060797
5	0.21006	0.75	1	0.50439	0.00057263	0.00057263	0.00057263
6	0.18792	0.8125	1	0.46753	0.00053728	0.00053728	0.00053728
7	0.19429	0.6875	1	0.56396	0.00050194	0.00050194	0.00050194
8	0.16634	0.625	1	0.60693	0.0004666	0.0004666	0.0004666
9	0.17167	0.875	1	0.46973	0.00043126	0.00043126	0.00043126
10	0.1588	0.6875	1	0.51855	0.00039591	0.00039591	0.00039591
11	0.14542	0.875	1	0.44971	0.00036057	0.00036057	0.00036057
12	0.1413	0.8125	1	0.44824	0.00032523	0.00032523	0.00032523
13	0.1359	0.875	1	0.51221	0.00028988	0.00028988	0.00028988
14	0.14123	0.9375	1	0.45239	0.00025454	0.00025454	0.00025454
15	0.12953	0.75	1	0.54199	0.0002192	0.0002192	0.0002192
16	0.13645	0.75	1	0.48462	0.00018385	0.00018385	0.00018385
17	0.12785	0.9375	1	0.44995	0.00014851	0.00014851	0.00014851
18	0.12278	0.875	1	0.4353	0.00011317	0.00011317	0.00011317
19	0.12244	0.8125	1	0.47876	7.78E-05	7.78E-05	7.78E-05
20	0.11998	0.9375	1	0.43848	4.25E-05	4.25E-05	4.25E-05

Figure 10: YOLOv8 train details

The best accuracy in classification using YOLOv8 with 20 epochs was 93.75%, which is considered good accuracy, but it is not better than VGG16, as provided in Figures 9 and 10.

- **YOLOv8:**

  Trained in the last try using 10 epochs and the other default settings.

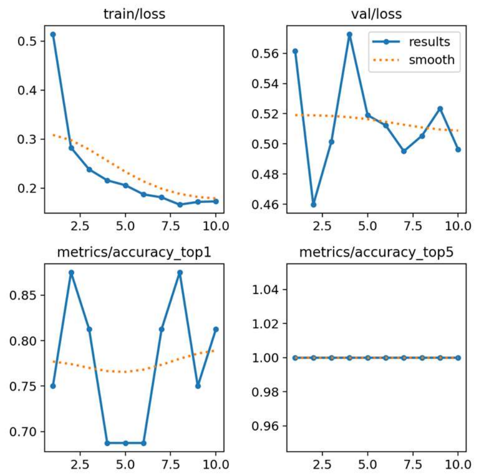

Figure 11: YOLOv8 results

	train/loss	metrics/accuracy_to	metrics/accuracy_to	val/loss	lr/pg0	lr/pg1	lr/pg2
1	0.51434	0.75	1	0.56152	0.00023727	0.00023727	0.00023727
2	0.28245	0.875	1	0.45972	0.00042822	0.00042822	0.00042822
3	0.23814	0.8125	1	0.50146	0.00057204	0.00057204	0.00057204
4	0.2158	0.6875	1	0.57275	0.00050194	0.00050194	0.00050194
5	0.20586	0.6875	1	0.51904	0.00043126	0.00043126	0.00043126
6	0.18719	0.6875	1	0.51221	0.00036057	0.00036057	0.00036057
7	0.18106	0.8125	1	0.49512	0.00028988	0.00028988	0.00028988
8	0.16621	0.875	1	0.50537	0.0002192	0.0002192	0.0002192
9	0.17178	0.75	1	0.52344	0.00014851	0.00014851	0.00014851
10	0.17285	0.8125	1	0.49634	7.78E-05	7.78E-05	7.78E-05

Figure 12: YOLOv8 train details

In Figures 11 and 12, they show that using 10 epochs with YOLOv8 the best accuracy was 87.5%, which is also a good result, but using more epochs as 20 gave a better result.

In conclusion, using a large number of epochs as 40 overfitted the model, and using a smaller number of epochs as 10, gave less accuracy than when training the model on 20 epochs, which was the best number of epochs.

- **YOLOv5:**
  Trained using 20 epochs, and the other settings set as default.

ep	train/	test/	metrics/a	metrics/a	lr/0
0	0.51912	0.8957	0.625	1	0.000951
1	0.47267	0.59924	0.69712	1	0.000901
2	0.46424	0.47061	0.85737	1	0.000852
3	0.44187	0.49475	0.79487	1	0.000802
4	0.42825	0.58131	0.76603	1	0.000753
5	0.41741	0.40059	0.8766	1	0.000703
6	0.39675	0.37072	0.90224	1	0.000654
7	0.38647	0.37854	0.88942	1	0.000604
8	0.37629	0.41497	0.875	1	0.000555
9	0.36645	0.43752	0.86058	1	0.000505
10	0.36079	0.44677	0.84936	1	0.000456
11	0.36199	0.3807	0.89263	1	0.000406
12	0.34577	0.37985	0.90064	1	0.000357
13	0.34865	0.39195	0.88141	1	0.000307
14	0.33625	0.40912	0.87821	1	0.000258
15	0.33831	0.3381	0.92949	1	0.000208
16	0.33671	0.32664	0.93429	1	0.000159
17	0.3324	0.33632	0.91827	1	0.000109
18	0.32699	0.31515	0.9375	1	5.95E-05
19	0.31256	0.33534	0.92628	1	1.00E-05

Figure 13: YOLOv8 train details

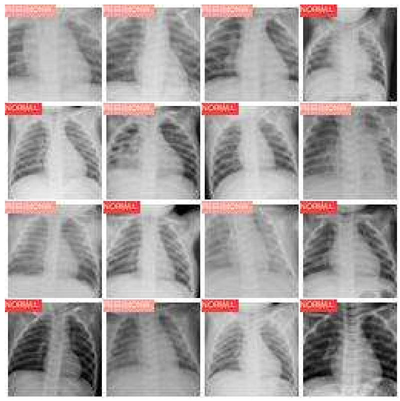

Figure 14: YOLOv5 classify

We finally wanted to see if older versions of YOLO could give better results. As provided in Figures 13 and 14, we used 20 epochs as it gave the best result in YOLOv8, and the results showed that YOLOv5 had an accuracy of 93.4%, which is less than YOLO8 with 20 epochs and VGG16.

C. Evaluation

A confusion matrix is a fundamental tool in machine learning and statistical classification that provides a detailed breakdown of how a classification model performs. It is a tabular representation that contrasts the actual and predicted classifications made by a model, allowing for a nuanced understanding of the model's accuracy, precision, recall, and overall effectiveness.

Precision is a measure of the accuracy of the positive predictions made by the model. It is defined as the ratio of true positive results to the total number of positive predictions (both true positives and false positives). Precision is particularly important in scenarios where the cost of false positives is high, as it indicates how many of the predicted positive cases are positive.

Recall, also known as sensitivity or true positive rate, measures the model's ability to identify all relevant positive cases within a dataset. It is the ratio of true positive results to the total number of actual positives (true positives and false negatives). Recall is crucial in medical diagnostics, where missing a positive case (false negative) can have serious implications.

The Figures [15, 16, 17] below show the confusion matrices for the best results for each, YOLOv8, YOLOv5, and VGG16.

- YOLOv8

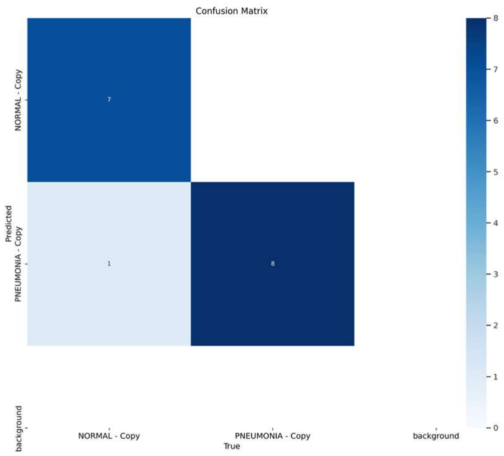

Figure 15: YOLOv8 20 epoch

- YOLOv5

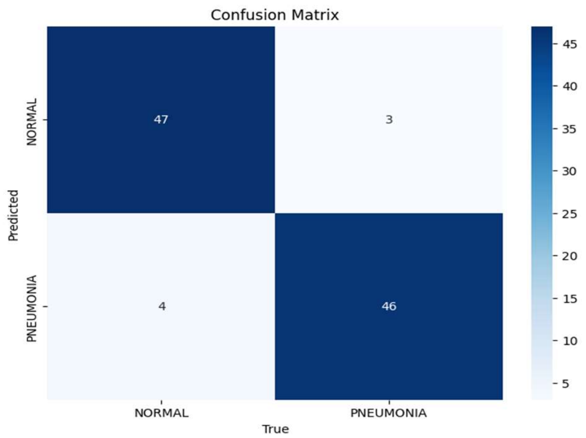

Figure 16: YOLOv5 20 epoch

- VGG16

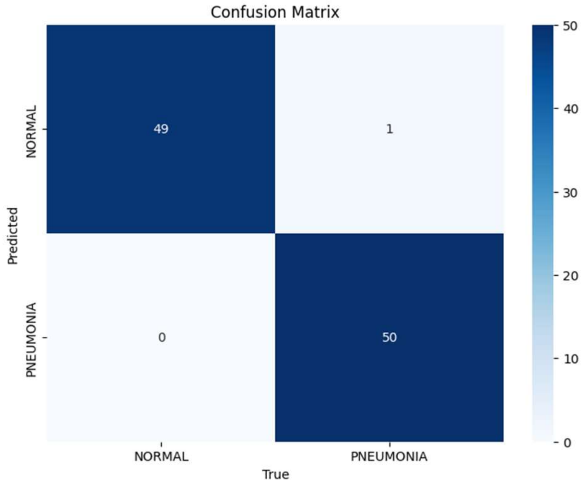

Figure 17: VGG16 5 epoch

The confusion matrices for the best results of each model are shown in Figures 15, 16, and 17. These matrices provide a detailed breakdown of the model's performance, including true positives, false positives, true negatives, and false negatives.

- VGG16: Accuracy: 98.9%, Precision: 99.02%, Recall: 99%,
- YOLOv8: Accuracy: 93.75% Precision: 88.2%, Recall: 100%
- YOLOv5: Accuracy: 93.4%, Precision: 93.02%, Recall: 93%,

VGG16 has the best overall performance across all metrics, with the highest accuracy, precision, recall, and F1-score. YOLOv8 is the second-best, performing better than YOLOv5, but not as well as VGG16. YOLOv5 has the lowest performance among the three models.

In conclusion, our evaluation demonstrates the effectiveness of VGG16 for pneumonia detection with the highest accuracy, precision, recall, and F1-score. These metrics highlight the robustness and reliability of VGG16 in identifying pneumonia cases from chest X-ray images, making it a valuable tool for medical image analysis.

D. Compare

Finally, we can see that the results, as simplified in Figure 18, showed that VGG16 helped in classifying whether the lung images were infected with pneumonia or not with an accuracy rate of 98.9%.

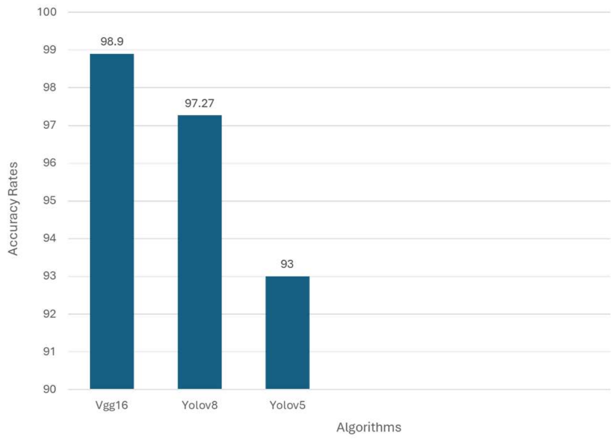

Figure 18: Accuracy Rates of the Classification Algorithms

- VGG16 emerged as the top performer in terms of both accuracy and recall, achieving an impressive accuracy and recall rate of 98.9% in classifying pneumonia cases.
- Additionally, YOLOv8 and YOLOv5 showed promising results, with YOLOv8 achieving an accuracy of 97.27% and YOLOv5 achieving an accuracy of 93%. These results demonstrate their capability to handle complex image data with high precision.
- The ease of use and accessibility of VGG16 make it a valuable resource for healthcare professionals and researchers alike.

4. Discussion

Throughout our investigation, we have examined the performance of various deep learning models in pneumonia detection, leveraging publicly available datasets of chest X-ray images. Our findings have shed light on the potential of these models to revolutionize diagnostic practices in healthcare, offering a glimpse into a future where AI-driven technologies play a central role in disease detection and patient care.

In Table 3, we can see the simplified results and see that VGG16 gave the best results compared with the two versions of YOLO.

4.1 Findings

In this section, we present the findings of our study, which involved the evaluation of various deep-learning tools and models for pneumonia detection using publicly available datasets of chest X-rays. The aim was to assess the performance of these tools and models in accurately identifying pneumonia cases from radiographic images.

Algorithms	Epoch	Best Accuracy
Yolov5	20	93.4%
Yolov8	10	87.5%
	20	93.75%
	40	overfitting
Vgg16	5	98.9%

Table 3: Accuracy Rates of the Classification Algorithms

Our analysis revealed that VGG16, used for creating machine learning models, emerged as the top performer in terms of both accuracy and recall. Remarkably, VGG16 achieved an impressive accuracy and recall rate of 98.9% in classifying whether an individual had pneumonia or not based on the chest X-ray images.

The exceptional performance of VGG16 underscores its effectiveness for medical image analysis, particularly in the context of pneumonia detection. Its high accuracy rate indicates its robustness in distinguishing between normal and pneumonia-afflicted chest X-ray images, thereby facilitating prompt and accurate diagnosis.

Furthermore, the ease of use and accessibility of VGG16 make it a valuable resource for healthcare professionals and researchers alike. Its intuitive interface and simplified workflow enable users to quickly develop and deploy machine learning models for medical image analysis.

4.2 Implication

The promising results of our study hold significant implications for the fields of medical imaging and pneumonia diagnosis, suggesting that deep learning models have the potential to revolutionize current diagnostic practices. By demonstrating the efficacy of these models in accurately detecting pneumonia from chest X-ray images, our findings underscore the transformative impact of artificial intelligence in healthcare.

The deployment of deep learning models as reliable, automated tools for pneumonia diagnosis could significantly enhance clinical practice, particularly in regions with limited access to expert radiologists. By leveraging the capabilities of these models, healthcare professionals can expedite the diagnostic process, leading to timely interventions and improved patient outcomes.

Moreover, our study highlights the value of data augmentation and transfer learning techniques in enhancing the performance of deep learning models for medical image analysis. Through the augmentation of training data and the transfer of knowledge from pre-trained models, we observed notable improvements in model accuracy and robustness. This emphasizes the importance of leveraging advanced techniques to optimize model performance and reliability.

The integration of deep learning models into existing hospital systems holds the potential to standardize diagnosis practices and improve overall healthcare delivery. By incorporating these models into routine clinical workflows, healthcare institutions can streamline diagnostic processes, reduce variability in interpretations, and ensure consistent quality of care across diverse patient populations.

Furthermore, the adoption of deep learning-based diagnostic tools has the potential to alleviate the workload of healthcare professionals, enabling them to focus on more complex tasks and patient care. By automating routine diagnostic procedures, these tools can significantly reduce diagnosis times and increase the efficiency of healthcare delivery.

In conclusion, our study highlights the transformative potential of deep learning models in pneumonia diagnosis and underscores the importance of integrating these technologies into clinical practice. Through continued research and innovation, we can harness the power of artificial intelligence to improve healthcare outcomes, enhance diagnostic accuracy, and increase accessibility to high-quality medical imaging analysis.

4.3 Limitations and Future Work

While our study has yielded promising findings regarding the potential of deep learning models in pneumonia diagnosis, it is important to acknowledge several limitations that may impact the generalizability and applicability of our results.

First and foremost, the models developed in this study were trained and validated using publicly available datasets of chest X-ray images. While these datasets provide valuable resources for research purposes, they may not fully capture the diversity and complexity of clinical cases encountered in real-world practice settings. The limited variability in these datasets may affect the generalizability of our findings to broader patient populations, particularly those with unique demographic characteristics or comorbidities.

Second, the binary classification task employed in our study (pneumonia vs. normal) oversimplifies the complexities of real-world diagnostic scenarios. In clinical practice, distinguishing between different etiologies of pneumonia (e.g., bacterial, viral, or fungal) is essential for guiding appropriate treatment decisions. The lack of granularity in our classification approach may limit the clinical utility of the models, particularly in scenarios where precise differentiation between pneumonia subtypes is necessary.

Furthermore, while our study demonstrates the potential of deep learning models to streamline the diagnostic process, it is essential to recognize that these models are not intended to replace the expertise of healthcare professionals. Rather, they should be viewed as decision-support tools that augment clinical judgment and facilitate more efficient diagnostic workflows. Continued collaboration between AI researchers and healthcare practitioners is essential to ensuring the responsible and ethical integration of these technologies into clinical practice.

Finally, it is important to acknowledge that the healthcare sector is dynamic and constantly evolving, with ongoing advancements in medical imaging technology, diagnostic techniques, and treatment modalities. As such, there is always room for improvement in optimizing the performance and utility of AI-based diagnostic tools. Future research efforts should focus on addressing these limitations, including the development of more diverse and representative datasets, the refinement of classification algorithms to account for multiple pneumonia subtypes, and the validation of model performance in real-world clinical settings.

In conclusion, while our study has provided valuable insights into the potential of deep learning models for pneumonia diagnosis, it is essential to interpret our findings within the context of these limitations. By addressing these challenges and continuing to advance the field of AI in healthcare, we can unlock the full potential of these technologies to improve patient care and outcomes.

5. Conclusion

In conclusion, our paper demonstrates the potential of artificial neural networks, specifically deep learning, to enhance pneumonia detection. Through thorough evaluation of various models like VGG16, YOLOv5, and YOLOv8, we establish benchmarks for future research.

VGG16 emerged as a top performer, highlighting its efficacy in classifying pneumonia cases from chest X-ray images. While our findings hold promise, addressing limitations such as dataset diversity and real-world implementation challenges is crucial. Overall, our work contributes to the advancement of AI-driven healthcare, aiming to improve patient outcomes and save lives.

The promising results of our study have significant implications for the fields of medical imaging and pneumonia diagnosis. Deep learning models have the potential to revolutionize current diagnostic practices, offering automated and accurate detection of pneumonia from chest X-ray images. Integration of these models into clinical practice could expedite diagnostic processes, improve patient outcomes, and standardize diagnosis practices across diverse patient populations.

Acknowledgment

We extend our sincerest gratitude to all those who have contributed to the completion of this paper, making it possible for us to delve into the fascinating realm of artificial neural networks.

We extend our appreciation to the academic staff and resources at Umm Al-Qura University, whose unwavering commitment to excellence has provided us with the necessary tools and infrastructure to pursue our academic endeavors.

References

[1] National Heart, Lung, and Blood Institute, "Pneumonia," National Institutes of Health, [Online]. Available: https://www.nhlbi.nih.gov/health/pneumonia. [Accessed: 10-May-2024].

[2] IBM, "Convolutional Neural Networks," [Online]. Available: https://www.ibm.com/topics/convolutional-neural-networks. [Accessed: 10-May-2024].

[3] IBM, "Deep Learning," [Online]. Available: https://www.ibm.com/topics/deep-learning#:~:text=for%20AI%20updates-,What%20is%20deep%20learning%3F,AI)%20in%20our%20lives%20today. [Accessed: 10-May-2024].

[4] A.M.A. Barhoom and S.S. Abu Naser, "Diagnosis of Pneumonia Using Deep Learning," International Journal of Academic Engineering Research (IJAER), vol. 6, no. 2, pp. 48-68, Feb. 2022.

[5] A. Chopde, F. Aga, and A. Kumar, "Detection of Pneumonia in Patients Using Chest X-Ray Images Based on Convolutional Neural Networks," International Journal of Medical and Biomedical Studies, vol. 6, no. 2, pp. 18-24, Feb. 2022.

[6] C. Ieracitano et al., 'A fuzzy-enhanced deep learning approach for early detection of Covid-19 pneumonia from portable chest X-ray images', Neurocomputing, vol. 481, pp. 202–215,Apr. 2022, doi: 10.1016/j.neucom.2022.01.055.

[7] L. Racic, T. Popovic, S. Cakic, and S. Sandi, "Pneumonia Detection Using Deep Learning Based on Convolutional Neural Network," Journal of Information Technology, vol. 25, no. 2, Feb. 2021.

[8] R. Kundu, R. Das, Z.W. Geem, G.-T. Han, and R. Sarkar, "Pneumonia detection in chest X-ray images using an ensemble of deep learning models," PLoS ONE, vol. 16, no. 9, Sep. 2021, Art. no. e0256630.

[9] M.F. Hashmi, S. Katiyar, A.W. Hashmi, and A.G. Keskar, "Pneumonia detection in chest X-ray images using compound scaled deep learning model," Automatika, vol. 62, no. 3-4, pp. 397-406, Sep. 2021.

[10] L. Kong and J. Cheng, "Based on improved deep convolutional neural network model pneumonia image classification," PLOS ONE, vol. 16, no. 11, Nov. 2021, Art. no. e0258804. doi: 10.1371/journal.pone.0258804.

[11] T. Gabruseva, D. Poplavskiy, and A. Kalinin, "Deep Learning for Automatic Pneumonia Detection," CVPRW 2020, Conference on Computer Vision and Pattern Recognition Workshops, Jun. 2020.

[12] Z. Yue, L. Ma, and R. Zhang, "Comparison and Validation of Deep Learning Models for the Diagnosis of Pneumonia," Computational Intelligence and Neuroscience, vol. 2020, Art. no. 8876798, Sep. 2020.

[13] G. Verma and S. Prakash, "Pneumonia Classification using Deep Learning in Healthcare," International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 4, Feb. 2020.

[14] T. Dincer, "Labeled Chest X-ray Images," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/tolgadincer/labeled-chest-xray-images. [Accessed: 25-Mar-2024].

[15] Datagen, "VGG16," [Online]. Available: https://datagen.tech/guides/computer-vision/vgg16/. [Accessed: 13-May-2024].