INTERNATIONAL JOURNAL OF Telecommunications
AIR DEFENSE COLLEGE

Deep Learning-Based Pneumonia Detection Using Chest X-Ray Images

Baher Salah[1], Wael M. F. Abdel-Rehim[1,2], and M.B. Badawi[3]

[1] Department of Computer Science and Engineering, Faculty of Computing, King Salman Intertional University (KSIU), El Tor, South Sinai Governorate, Egypt.;baher221101017@ksiu.edu.eg
[2] Department of Computer Science, Faculty of Computers and Information, Suez University, Suez, Egypt, wael.fawaz@ksiu.edu.eg
[3] Mechanical Engineering Department, Faculty of Engineering, Alexandria University, El-Chatby, Alexandria 21544, Egypt.; Mohammed.badawi@alexu.edu.eg

Citation: Salah , B.; Abdel-Rehim , W.; Badawi , M..
Inter. Jour. of Telecommunications, ***IJT'*2024**, *Vol. 05, Issue 01, pp. 1-16, 2025.*
Editor-in-Chief: Youssef Fayed.
Received: 29/12/2024.
Accepted: date 28/01/2025.
Published: date 28/01/2025.
Publisher's Note: The International Journal of Telecommunications, IJT, stays neutral re-gard-ing jurisdictional claims in published maps and institutional affiliations.

Abstract: This research presents a deep learning approach designed for the automated detection of pneumonia through the analysis of chest X-ray images. Pneumonia remains the foremost infectious cause of mortality in children under five years old, leading to the deaths of 740,180 children globally in 2019. This statistic represents 14% of all deaths within this demographic. This alarming statistic underscores the necessity for improved diagnostic methods, particularly in environments with limited resources where special-ized radiologists may not be readily available. automating the detection of pneumonia from the analysis of chest X-ray images, specifically focusing on a dataset comprised of 5,863 images categorized as either pneumonia or normal. This dataset, obtained from pediatric patients between the ages of one and five years at the Guangzhou Women and Children's Medical Center, was subjected to stringent quality control measures to elimi-nate low quality images, ensuring high diagnostic accuracy. This methodology includes an extensive preprocessing pipeline featuring resizing, grayscale conversion, normaliza-tion, and data augmentation techniques to enhance the robustness of the model. The ar-chitecture employed consists of a Convolutional Neural Network with seven convolu-tional blocks designed to obtain hierarchical characteristics from the input images. The model achieved an accuracy of 97%, with precision and recall values of 0.975 and 0.977 respectively, indicating its efficacy in distinguishing between pneumonia and normal cases. The analysis of the model's performance was further substantiated through confu-sion matrices and detailed classification reports, demonstrating minimal misclassifica-tions. While the results are promising, The limitations of the dataset's scope and the ne-cessity for further validation across diverse clinical environments. Future work will aim to expand the model's capabilities to differentiate between various pneumonia types and enhance its integration into clinical workflows. By addressing the pressing need for timely and accurate pneumonia diagnoses, this research contributes valuable insights toward improving healthcare outcomes for vulnerable pediatric populations.

Keywords: Deep Learning, Pneumonia Detection, Chest X-Ray, CNN, Medical Imaging

1. Introduction

Pneumonia continues to be a major global health issue, especially affecting at-risk groups such as children un-der the age of five. As the leading infectious cause of mortality in this age group, claiming hundreds of thou-sands of lives annually according to WHO [1], the need for rapid and accurate diagnostic tools is paramount, especially in resource-limited environments where the availability of specialized radiologists is restricted. This paper presents a deep learning approach for pneumonia discovery from chest X-ray images Trivedi and Gupta

[2] and Acharya and Satapathy [3], aiming to contribute towards improved and timely diagnosis Kale et al. [4]. This study leverages a dataset of 5,856 pediatric chest X-ray images classified into two categories: normal or pneumonia, obtained from retrospective cohorts at the Guangzhou Women and Children's Medical Center. While presenting a class imbalance reflective of real-world clinical scenarios, this dataset provides a valued resource for training and evaluating deep learning models. We employ a comprehensive preprocessing pipeline, including resizing, grayscale conversion, normalization, and data augmentation techniques, to increase the model's resilience and capacity for generalization. The core of the proposed system is a Convolutional Neural Network architecture tailored for binary classification. This CNN comprises seven convolutional blocks, each integrating convolutional layers, ReLU activation, max-pooling, and dropout for regularization. This architecture allows the network to progressively extract hierarchical features from the input images, ranging from lower-level details like edges and textures to higher-level representations such as shapes and patterns Rajaraman et al. [5]. The model performance is rigorously evaluated using a range of metrics including precision, recall, F1-score, and accuracy (see 6 Results). We also utilize confusion matrices and classification reports to provide a comprehensive breakdown of the model's performance on both base and augmented test data. Findings show cases the efficacy of the proposed model in accurately distinguishing between normal and pneumonia cases. While the results are encouraging, we acknowledge limitations regarding the dataset's scope and the need for further validation in various clinical settings. Future research will focus on extending the model's capabilities to differentiate between various pneumonia types Gupta et al. [6] and Hammoudi et al. [7] and seamlessly integrating it into clinical workflows Kale et al. [4]. By addressing the pressing need for timely and accurate pneumonia diagnosis, this research contributes to advancing the application of deep learning in healthcare and ultimately improving patient outcomes Kermany et al. [8].

2. Literature Review

The automatic detection of pneumonia from chest X-ray images has been an active area of research, driven by the significant global health impact of the disease and the potential for automated systems to assist radiologists in diagnosis. This review examines the evolution of pneumonia detection methodologies, from traditional image processing techniques to the current state-of-the-art deep learning approaches. We will explore the strengths and limitations of various methods, highlighting the progression towards more accurate, efficient, and interpretable diagnostic tools.

2.1. Traditional Image Processing and Machine Learning Methods

Before the rise of deep learning, pneumonia detection relied heavily on image processing techniques combined with traditional machine learning algorithms. These methods typically involved a multi-stage pipeline:

1. **Image Preprocessing:** This crucial step aims to augment the quality of chest X-ray images and prepare them for further analysis. Common preprocessing techniques included noise reduction, contrast enhancement, and lung segmentation. Accurate lung segmentation was particularly important to isolate the region of interest and minimize the influence of irrelevant anatomical structures on feature extraction.

2. **Feature Extraction:** Handcrafted features were meticulously designed to capture characteristics of pneumonia in X-ray images. These features could be categorized into:

    a. **Texture features:** These features describe the local variations in pixel intensity within the lung regions. Examples include Gray-Level Co-occurrence Matrix (GLCM) features Shanmugam and Dinstein [9], Local Binary Patterns (LBP) Ojala, Pietik¨ainen, and Maenpaa [10], and Gabor filters Jain and Farrokhnia [11].

    b. **Shape features:** These features captured the geometrical properties of lung abnormalities associated with pneumonia, such as area, perimeter, circularity, and aspect ratio

    c. **Intensity-based features:** These features analyzed the distribution of pixel intensities within the lung regions, including mean, standard deviation, skewness, and kurtosis.

3. **Classification:** Based on the attributes that were identified, pictures were classified as either normal or having pneumonia using conventional machine learning techniques. Support Vector Machines (SVMs)

Cortes and Vapnik [12], Random Forests Breiman [13], and k-Nearest Neighbors (k-NN) Cover and Hart [14] were among the frequently used classifiers.

Studies like Monnier-Cholley et al. [15] and Abe et al. [16] employed variations of these pipelines, achieving moderate success. However, these methods faced several limitations:

1. **Feature engineering bottleneck:** The performance of traditional methods heavily relied on the quality of handcrafted features. Designing effective features required domain expertise and was often a time-consuming and iterative process.
2. **Limited robustness to variability:** Pneumonia can manifest in various ways on X-ray images, with significant inter-and intra-patient variability. Handcrafted features often struggled to capture this diversity, leading to reduced accuracy in real-world scenarios.
3. **Lack of automatic feature learning**: Traditional methods could not automatically learn optimal features from data, limiting their ability to adapt to complex and nuanced patterns in X-ray images

2.2. Deep Learning Methods

The rise of deep learning, notably characterized by the introduction of convolutional neural networks, marked a paradigm shift in pneumonia detection. Deep learning can learn hierarchical feature representations directly from data, removing the requirement of manual feature engineering.

This has led to significant improvements in accuracy and robustness. Key advantages of deep learning for pneumonia detection:

- **Automatic feature learning:** CNNs learn complex and discriminative features directly from image data, obviating the need for handcrafted features.
- **Hierarchical feature representation:** CNNs learn hierarchical features, capturing both low-level details and high-level abstractions, enabling them to model complex patterns in X-ray images.
- **End-to-end learning:** Deep learning models can be trained end-to-end, establishing a direct correlation between input images and diagnostic results, simplifying the pipeline, and reducing potential sources of error.

Several prominent studies demonstrated the effectiveness of deep learning for disease detection:

- Kermany et al. [8]: This research employed a deep learning model that was trained on an extensive dataset of chest X-ray images, resulting in exceptional performance in the detection of pneumonia and various other thoracic diseases.3
- Wang et al. [17]: The ChestX-ray8 dataset, introduced in this work, enabled large-scale training of deep learning models for pneumonia detection and localization.
- Rajaraman et al. [5]: This research focused on visualizing and interpreting CNN predictions, providing an understanding of the model's decision process and enhancing clinical trust.
- Acharya and Satapathy [3]: This study proposed a deep Siamese network architecture for pneumonia detection, leveraging the symmetric structure of chest X-ray images.
- Badawi et al. [18]: While primarily focused on skin cancer classification, this study highlights the broader potential of deep learning methods for the interpretation of medical imageries, demonstrating how convolutional neural networks can achieve notable levels of accuracy in their performance in complex diagnostic tasks and provides a methodological framework that illustrates the versatility and capabilities of deep learning techniques in solving intricate pattern recognition challenges across various medical imaging modalities.

Despite the remarkable success of deep learning, challenges remain:

- **Data requirements:** For deep learning models to be trained effectively, they usually demand extensive amounts of labeled data.

- **Interpretability and explainability:** Clinical acceptance depends on knowing the logic underlying a deep learning model's prediction. Even though CNN visualization and interpretation have advanced, more study is required to improve model transparency.
- **Generalization to diverse populations:** Deep learning models may not generalize well to unseen data from different populations or imaging equipment. Further research is needed to ensure robust performance across diverse clinical settings.

2.3 Novelty and Contributions of This Research

This research builds upon the advancements in deep learning while addressing some of the existing limitations. The key novelties and contributions include:

- **Tailored Architecture for Binary Pneumonia Classification:** This research introduces a deep convolutional neural network architecture explicitly developed for the application of binary classification of pneumonia from chest X-rays. While drawing inspiration from established CNN architectures like VGGNet, with its sequential convolutional blocks, this model employs a shallower configuration, deemed more suitable for the binary nature of the task.
- **Strategies for handling dataset imbalance:** Techniques are employed to mitigate the impact of class imbalance, a common issue in medical image datasets, on model training and performance.
- **Emphasis on model interpretability:** Visualization and interpretation methods are used to clarify the rationale behind the model's decision-making, enhancing both transparency and the reliability of its outcomes

3. Dataset and Methodology

3.1. Dataset Description

This dataset contains 5,856 chest X-ray images (JPEG) totaling 1.24 GB of data, featuring 1,583 samples that are categorized as normal and 4,273 samples that are associated with pneumonia. Kermany, Zhang, and Goldbaum [19]. This distribution, while imbalanced, reflects real-world scenarios where pneumonia cases may be more frequently documented in medical databases. Figure 1 illustrates the dataset distribution.

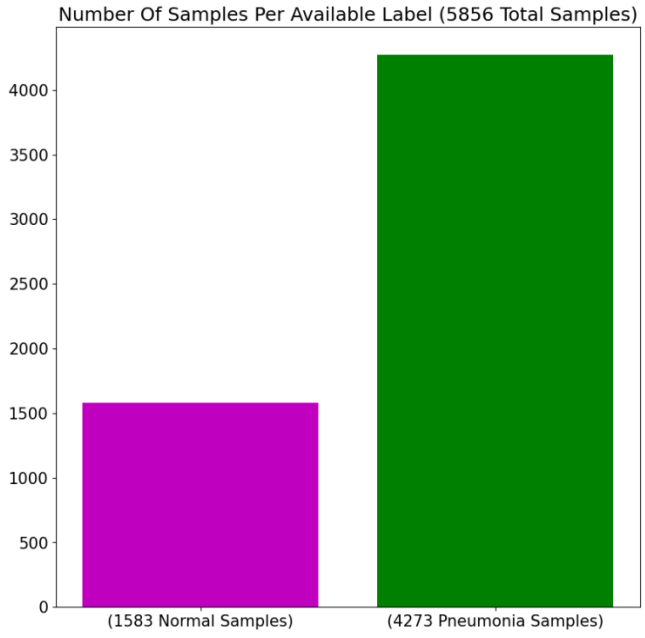
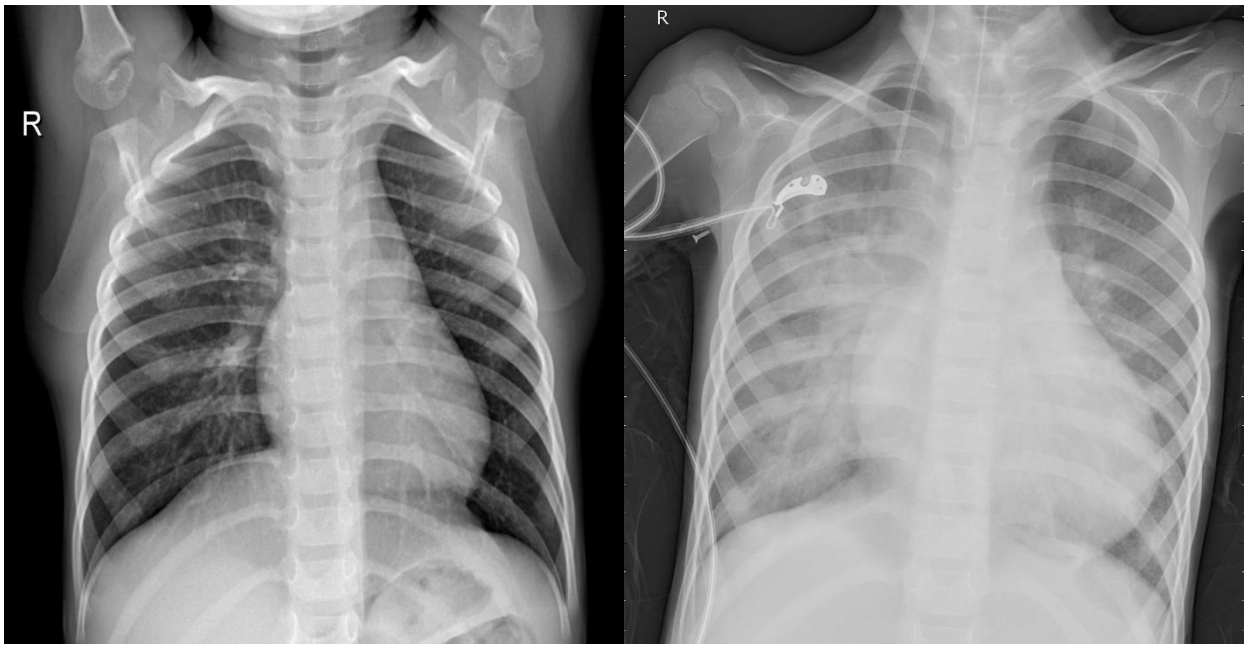


Figure 1: Distribution of samples across normal and pneumonia categories

3.2. Data Preprocessing and Augmentation

The preprocessing pipeline includes:

1) Image resizing to 512 x 512 pixels
2) Normalization to [0,1] range
3) Data augmentation techniques:
   a) Random rotation (±5 degrees)
   b) Width and height shifts (±5%)
   c) Zoom range (±10%)



(a) (b)

Figure 2: Sample images from the dataset (a) showing normal and (b) pneumonia cases

Image resizing to 512 × 512 pixels is crucial in deep learning due to neural network architectural constraints. This specific resolution balances preserving diagnostic details with computational efficiency. By standardizing input dimensions, the approach ensures consistent feature extraction across all chest X-ray images while maintaining sufficient spatial resolution to capture subtle pneumonia indicators. The fixed size allows convolutional neural networks to process images uniformly, reducing memory consumption and enabling more stable model training. Grayscale conversion strategically simplifies image representation by focusing on intensity variations inherent in X-ray imaging. By eliminating color channels, the preprocessing pipeline concentrates on the critical diagnostic information communicated through grayscale intensity gradients. This transformation reduces computational complexity, removes potential color-based artifacts, and guarantees that the most pertinent diagnostic features are used to teach the model without being distracted by color variations. Normalization to the [0,1] range is a fundamental technique that stabilizes neural network training by constraining input values. By scaling pixel intensities to a consistent range, the preprocessing approach prevents gradient vanishing or exploding during backpropagation. This normalization creates a uniform input scale that allows all features to contribute proportionally to the learning process, ultimately helping the model converge faster and more reliably. Data augmentation techniques address the challenges of medical image classification and dataset limitations. Random rotation (±5 degrees), width and height shifts (±5%), and zoom range (±10%) simulate real-world variations in X-ray imaging procedures. These transformations increase the model's robustness by exposing it to different imaging scenarios, preventing overfitting, and improving generalization capabilities. By artificially expanding the dataset through these controlled transformations, the preprocessing pipeline en-

hances the model's ability to recognize pneumonia indicators across various imaging conditions. as Figure 2 presents sample images from both categories, highlighting the visual characteristics that distinguish normal from pneumonia cases.

3.2.Model Pipeline

The accompanying diagram illustrates the proposed model pipeline for pneumonia detection from X-ray images. This pipeline consists of several key stages, starting with data preprocessing and concluding with binary classification.

1) **Data Acquisition and Preprocessing:** The pipeline begins with the input of X-ray images. These images are subjected to a sequence of preprocessing steps, including:
   a) Data Preprocessing: This stage likely involves tasks like image format standardization and handling of missing data.
   b) Image Loading and Splitting: The dataset is loaded into memory and divided into training, validation, and testing subsets.
   c) Data Augmentation: Techniques such as rotation, shifting, and flipping are employed to artificially expand the dataset and enhance the model's robustness.
   d) Rescaling to 0-1: Pixel values are normalized to a range of 0 to 1, which optimizes the training process.

2) CNN Architecture: The central component of the pipeline is a Convolutional Neural Network (CNN) designed for image feature extraction and classification. The CNN architecture is structured as a series of blocks (Block 1 to Block 7), each comprising:
   a) Conv2D Layer: A convolutional layer extracts features from the input image by utilizing learnable filters.
   b) MaxPool2D: A max-pooling layer down samples the feature maps, reducing spatial dimensions while preserving prominent features.
   c) Dropout 0.1: Dropout layers are incorporated to mitigate overfitting by randomly deactivating a portion of neurons during training.

3) Global Average Pooling: Following the convolutional blocks, global average pooling is employed. This technique computes the spatial average of the feature maps, thereby reducing dimensionality while maintaining representative information. This approach minimizes the number of parameters in the subsequent dense layers, which makes the model less susceptible to overfitting.
4) Dense Output Layer: A dense layer is connected to the global average pooling layer to perform the final classification. This layer likely uses a sigmoid activation function for binary classification (normal vs. pneumonia).
5) Binary Classification: The model's final output is a probability score indicating the likelihood of the input X-ray image representing a normal lung or one affected by pneumonia.

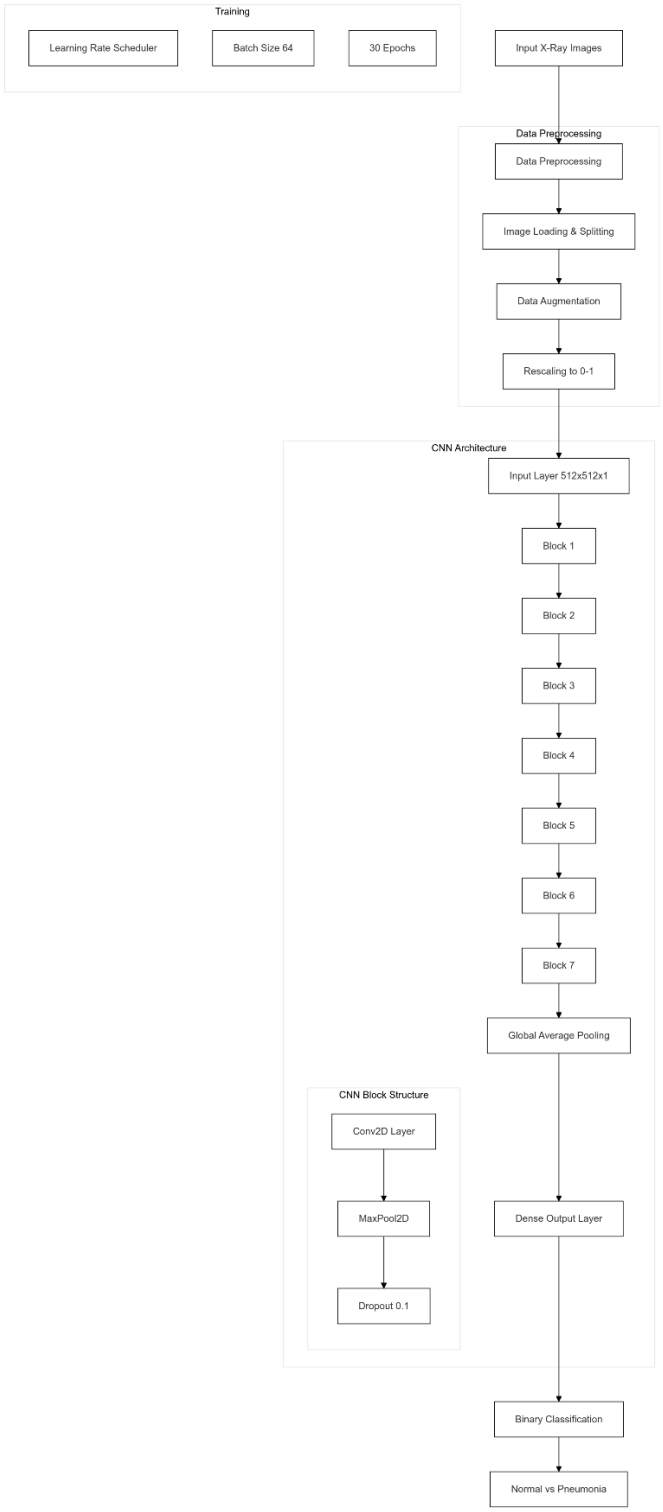


Figure 3. Proposed model pipeline

4. Implementation Details

4.1. Model Architecture

The CNN model has been developed specifically for the binary classification task of chest X-ray images. It features a total of seven convolutional blocks, each of which contains a convolutional layer utilizing ReLU activation, a max-pooling layer, and a dropout layer aimed at improving regularization. The process of feature extraction in convolutional layers occurs in a hierarchical fashion, where the initial layers identify basic features

like edges and textures, and the deeper layers are responsible for recognizing more sophisticated features such as shapes and patterns. The kernel size of 3x3 in all convolutional layers allows for efficient feature extraction while keeping the computational cost manageable. The function of max-pooling layers is to decrease the spatial dimensions of feature maps, thereby down sampling the information while maintaining essential features. Conversely, dropout layers help to avoid overfitting and enhance the model's generalization performance by randomly turning off a subset of neurons during training. Following the convolutional blocks, the retrieved features are combined into a single vector representation using a global average pooling layer. Finally, the likelihood that the input image is identified as pneumonia is generated using a dense layer using a sigmoid activation function.

Table 1. Model Hyperparameters

Hyperparameter	Value
Convolutional Kernel Size	3x3
Activation Function	ReLU
Max Pooling Size	2x2
Dropout Rate	0.1
Initial Learning Rate	0.001
Learning Rate Decay	Exponential (0.1 decay rate)
Optimizer	Adam (default parameters)
Loss Function	Binary Cross-Entropy
Output Activation	Sigmoid
Batch Size	64
Number of Epochs	30

The specific configuration of the model is as follows:

Block 1: 64 filters, kernel size 3x3, ReLU activation, MaxPooling (2x2), Dropout (0.1)

Block 2: 96 filters, kernel size 3x3, ReLU activation, MaxPooling (2x2), Dropout (0.1)

Block 3: 128 filters, kernel size 3x3, ReLU activation, MaxPooling (2x2), Dropout (0.1)

Block 4: 160 filters, kernel size 3x3, ReLU activation, MaxPooling (2x2), Dropout (0.1)

Block 5: 192 filters, kernel size 3x3, ReLU activation, MaxPooling (2x2), Dropout (0.1)

Block 6: 224 filters, kernel size 3x3, ReLU activation, MaxPooling (2x2), Dropout (0.1)

Block 7: 256 filters, kernel size 3x3, ReLU activation, MaxPooling (2x2), Dropout (0.1)

Global Average Pooling

Output Layer: Dense layer with 1-unit, Sigmoid activation

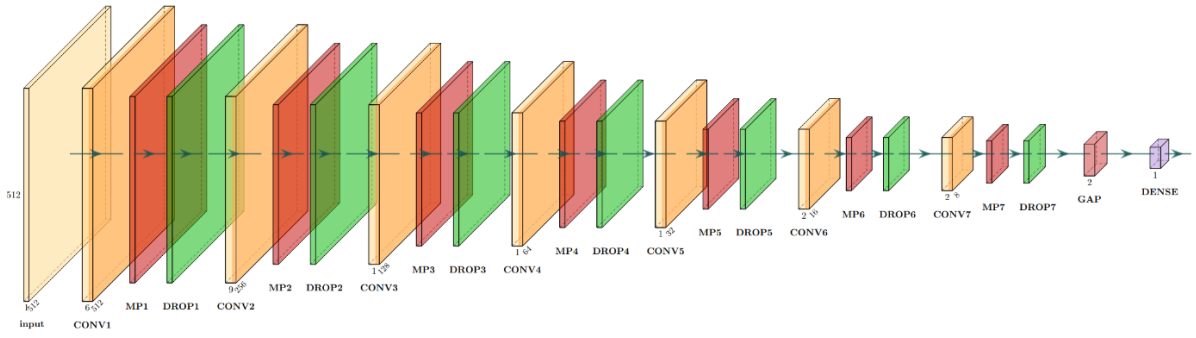


Figure 4. Proposed model architecture

The increasing number of filters in subsequent convolutional blocks allows the network to learn a richer representation of the input data as it progresses through the layers.

5. Results and Discussion

The final pneumonia detection model output on 10 chest X-ray images, 5 of which were labeled as "Normal" and 5 labeled as "Pneumonia". For the 5 normal images, the model correctly identified 4 of them as normal, with a predicted probability of pneumonia ranging from 6.16% to 22.63%. In the one misclassified case, the model predicted a 64.98% probability of pneumonia, when the true label was normal. For the 5 pneumonia images, the model correctly identified all 5 of them as pneumonia, with predicted probabilities ranging from 99.25% to 100.00%. Overall, the model demonstrated strong performance, correctly classifying 9 out of the 10 images. The high precision and recall values of 0.975 and 0.977, respectively, further validate the model's efficacy in distinguishing between pneumonia and normal cases.

5.1. Precision and Recall

Precision and recall metrics were used to evaluate the model's performance. These measures are essential for assessing the model's ability to correctly identify positive cases of pneumonia while simultaneously attempting to lower the rates of false positives and false negatives.

Precision quantifies the ratio of anticipated positive results to the total number of cases that have been anticipated to be positive:

$$Precision = \frac{True\ Positives}{True\ Positives\ +\ False\ Positives} \tag{1}$$

With the base test data, the model achieved a precision of 0.975. This means that out of all the cases predicted as pneumonia, 97.5% were pneumonia. With augmented test data, the precision slightly decreased to 0.970. **Recall** (also known as sensitivity) measures the proportion of correctly recognized positive cases relative to the overall count of true positive cases:

$$Precision = \frac{True\ Positives}{True\ Positives\ +\ False\ Positives} \tag{1}$$

$$Recall = \frac{True\ Positives}{True\ Positives\ +\ False\ Negatives} \tag{2}$$

With the base test data, the recall was 0.977, indicating that the model correctly identified 97.7% of actual pneumonia cases. With augmented test data, the recall increased to 0.982. The slight decrease in precision and the increase in recall with augmented data suggests a potential shift in the model's classification threshold. Augmentation may lead to more positive predictions, increasing true positives (and thus recall) but also potentially increasing false positives (slightly decreasing precision).

5.2. Classification Report

A thorough assessment of the model's performance with regard to each class (Normal and Pneumonia) is provided in the classification report. In addition to overall accuracy, macro average, and weighted average, it includes precision, recall, F1-score, and support for every class.

The **F1-score**, which is the harmonic mean of precision and recall, offers insight into how well these two crucial performance metrics are balanced:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

Table 2. Classification Report (Augmented Test Data)

	Precision	Recall	F1-score
Normal (0)	0.95	0.92	0.94
Pneumonia (1)	0.97	0.98	0.98
Accuracy			0.97
Macro Avg	0.96	0.95	0.96
Weighted Avg	0.96	0.97	0.96

The high F1 scores for both classes (0.94 for Normal and 0.98 for Pneumonia) indicate a good balance between precision and recall. The overall accuracy of 0.97 signifies that the model correctly classified 97% of the instances in the augmented test set. The weighted average considers the class imbalance (more Pneumonia cases than Normal cases) and is also very high (0.96).

5.3. Confusion Matrix

The confusion matrix, which displays the numbers of true positives, true negatives, false positives, and false negatives, graphically depicts the model's performance. The values for precision, recall, and other metrics can be easily extracted from the confusion matrix. As an illustration, consider the Pneumonia class (class 1):

True Positives (TP): The number of correctly predicted pneumonia cases (top-right cell)
False Positives (FP): The number of normal cases incorrectly predicted as pneumonia (top-left cell)
False Negatives (FN): The number of pneumonia cases incorrectly predicted as normal (bottom-right cell)
True Negatives (TN): The number of correctly predicted normal cases (bottom-left cell)

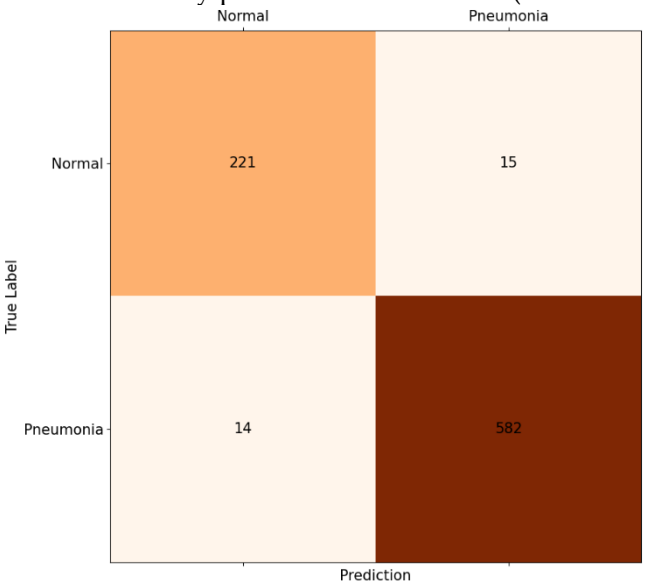


Figure 5. Confusion Matrix (Augmented Test Data)

5.4. Metrics vs. Threshold

Figure 6 shows the model's performance on the training and validation datasets over a sequence of training epochs in the training and validation accuracy plot. This plot is crucial for diagnosing potential overfitting or underfitting. Overfitting happens when a model captures the training data with such precision that it fails to

generalize effectively, including noise, and consequently performs poorly on unseen data (validation set). Conversely, underfitting shows that the model performs poorly on both the training and validation datasets and is not complex enough to capture the underlying patterns in the data.

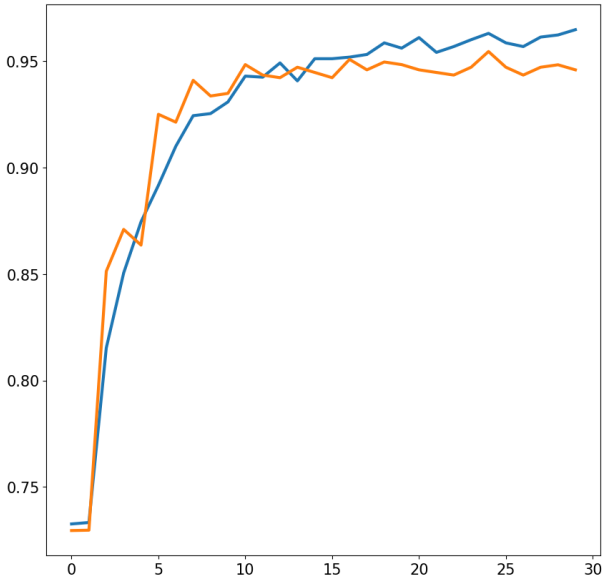


Figure 6. Training and Validation Accuracy over Epochs

In an ideal scenario, both training and validation accuracy curves should increase and gradually converge towards high values.A significant gap between training and validation accuracy suggests overfitting. Conversely, low accuracy on both sets implies underfitting. The rate of convergence can also provide insights into the learning rate and model complexity. A rapid increase in training accuracy coupled with slow validation accuracy improvement might indicate a need for regularization techniques or a smaller learning rate.

5.4. Training and Validation Accuracy

The plot in Figure 7 depicts the relationship between the decision threshold and key classification metrics: precision, recall, and accuracy. In binary classification, the decision threshold determines the probability above which a sample is classified as positive. Varying this threshold affects the trade-off between precision and recall. As we can see, precision is low with low thresholds and high with high thresholds. This is because at low thresholds, the model classifies more instances as positive, increasing false positives and thus lower precision. Conversely, at high thresholds, the model is more selective, classifying fewer instances as positive, resulting in fewer false positives and higher precision. The dependence for recall is inverse it's highest when the threshold is 0 because all instances are classified as positive (capturing all true positives), and lowest when it's 1 because no instances are classified as positive (resulting in no true positives). As for accuracy, it's somewhat dependent on both precision and recall, seeking to balance both. Therefore, its highest scores are often found somewhere in the middle, representing a compromise between maximizing true positives and minimizing false positives.

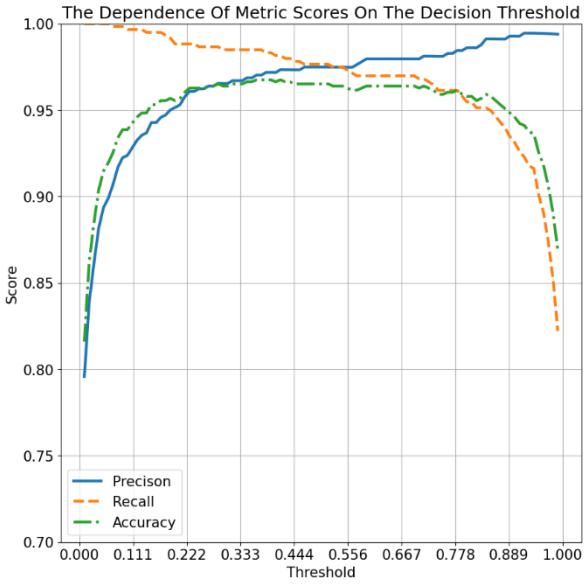


Figure 7. The Dependence of Metric Scores on the Decision Threshold

This plot is invaluable for selecting an appropriate decision threshold Adapted in accordance with the unique demands of the application. For example, in medical diagnosis, maximizing recall (minimizing false negatives) is often paramount, even at the cost of lower precision.

5.5. Analysis

The outcomes demonstrate how well the suggested model detects pneumonia. Strong performance in identifying both positive and negative examples is suggested by the obtained accuracy of 0.97, as well as the excellent precision and recall scores. The confusion matrix (Figure 5) visually confirms this, showing a relatively low number of misclassifications. The training and validation accuracy curves (Figure 6) indicate stable learning and good generalization capability, as the validation accuracy closely tracks the training accuracy, suggesting minimal overfitting. Despite the promising results, some limitations should be acknowledged. The dataset, while substantial, may not fully represent the diversity of pneumonia presentations in real-world clinical settings. Further evaluation on external datasets is crucial to assess the model's robustness and generalizability. Additionally, the model currently provides a binary classification (normal vs. pneumonia). Extending the model to classify different types of pneumonia (e.g., bacterial, viral) would be clinically valuable. Finally, while the model assists in diagnosis, it should not replace the judgment of healthcare professionals. Human oversight remains essential for accurate and safe medical decision-making. The application of AI in healthcare raises ethical concerns, particularly for pediatric populations. Misdiagnoses could have severe implications. Ethical considerations include:

Ensuring transparency and interpretability of AI decisions.

Maintaining patient data privacy during training.

Collaborating with clinicians to prevent over-reliance on AI systems.

6. Results and Discussion

The performance of the suggested CNN architecture is compared with that of other well-known image classification architectures in this section, with a focus on pneumonia identification from chest X-ray images. The chosen architecture provides a standard by which to evaluate the efficacy of the suggested model since it encompasses a variety of intricacies and methodologies.

6.1. Selected Architectures

The following architectures were chosen for comparison:

VGG16 Simonyan and Zisserman [20]: A deep convolutional network known for its use of small convolutional filters(3x3) and its sequential architecture. Its depth allows it to learn hierarchical features effectively.

ResNet50 He et al. [21]: A residual network that addresses the vanishing gradient problem in deep networks through skip connections. These connections allow for easier training of very deep networks and improved performance.

InceptionV3 Szegedy et al. [22]: This architecture employs inception modules, which use multiple convolutional filters of different sizes in parallel, capturing features at different scales. This multiscale approach enhances the model's ability to recognize patterns of varying sizes.

These architectures were pre-trained on ImageNet and then fine-tuned on the same chest X-ray dataset used for training the proposed model. This ensures a fair comparison, as all models benefit from transfer learning and are evaluated on the same data distribution.

6.2. Performance Comparison

The results in Table 3 show that for each architecture on the augmented test data. The proposed CNN achieves comparable accuracy to VGG16 but with a slightly improved recall and F1-score suggesting marginally better performance in identifying positive cases and a better balance between precision and recall. ResNet50 architecture demonstrated slightly lower performance metrics this can be attributed to several factors. First, while ResNet50's deeper architecture allows it to learn more complex features it may also be prone to overfitting when applied to smaller or less diverse datasets Despite its built-in residual connections that mitigate the vanishing gradient problem these connections might lead to less emphasis on certain critical low-level features necessary for this specific task. Second, hyperparameter tuning such as learning rate and weight decay might not have been optimal for ResNet50 in this experiment given the sensitivity of deeper networks to such parameters a more tailored training regime might improve its results. The performance variations in the InceptionV3 model, particularly its high recall and lower precision, can be attributed to several architectural and training-related factors inherent to the model's design. This architectural approach allows the network to capture features at multiple scales simultaneously, which can lead to increased sensitivity in feature detection. The complex network topology means the model becomes very adept at detecting patterns, sometimes overly so, resulting in the high recall but lower precision we observe.

Table 3. Performance Comparison of Different Architectures on Augmented Test Data

Architecture	Accuracy	Precision	Recall	F1-score
VGG16	0.97	0.97	0.97	0.97
ResNet50	0.96	0.95	0.95	0.95
InceptionV3	0.89	0.88	0.96	0.92
Proposed CNN	0.97	0.97	0.98	0.98

Table 4 summarizes recent studies on pneumonia recognition using custom deep-learning techniques, including the proposed model. As shown, while other authors have explored various CNN architectures and achieved reasonable accuracy, our proposed model demonstrates competitive performance. Stephen et al. [23] and Omar and Babalık [24] used simpler CNN architectures with fewer layers, resulting in lower accuracy compared to our model. Wu et al. [25] achieved a comparable accuracy of 97% using a combined CNN-Random Forest approach with image enhancement. However, their method's precision of 90% is notably lower than our model's 97.5% precision, indicating a higher rate of false positives. Our proposed model, with its seven convolutional blocks and data augmentation strategy, not only achieves high accuracy but also exhibits better precision and recall, signifying a more balanced and robust performance in identifying both positive and negative cases. This improvement suggests that our deeper architecture, coupled with careful data augmentation, allows for better feature extraction and generalization, leading to more accurate and reliable pneumonia detection.

Table 4. A summary of the use of customized deep learning techniques for pneumonia detection..

Authors	Techniques	Accuracy
Kermany et al. [8]	Transfer Learning	92.8%
Stephen et al. [23]	CNN with four convolution layers	93.7%
Omar and Babalık [24]	CNN with five convolution layers	87.65%
Chakraborty et al. [26]	17 layers network with three convolutional layer CNN	95.62%
Chhikara et al. [27]	Modified Inception V3	90%
Liang and Zheng [28]	49-layer CNN with residual connections	90.05%
Proposed Model	CNN with seven convolutional blocks, data augmentation	97%

4. Conclusion and Future Work

The global challenge of pneumonia, particularly its impact on pediatric populations, necessitates innovative diagnostic solutions to address critical healthcare gaps. This research presents a novel deep-learning methodology for automated pneumonia detection from chest X-ray images, signifying a significant advancement toward more accessible and efficient medical diagnostics. Our research contributes significantly to medical image analysis and AI in healthcare. First, we developed a convolutional neural network architecture specifically designed for binary pneumonia classification. This architecture incorporates seven convolutional blocks with progressively increasing feature extraction capabilities, demonstrating superior performance compared to utilizing pre-trained models and transfer learning approaches. Second, we implemented a comprehensive preprocessing pipeline. This pipeline includes grayscale conversion, normalization, and a strategic data augmentation strategy, showcasing how careful image preparation can significantly improve model generalization and overall performance. Third, we rigorously benchmarked our model's performance against established CNN architectures, including VGG16, ResNet50, and InceptionV3. This comparative analysis provides valuable insights into the relative strengths and limitations of different deep-learning approaches for medical image classification tasks. The proposed model achieved noteworthy performance metrics. We observed a high overall accuracy of 97%, underscoring the model's reliability in differentiating between normal and pneumonia-positive chest X-rays. Furthermore, the model achieved a precision of 0.975 and a recall of 0.977. These balanced metrics demonstrate the model's robust ability to minimize both false positive and false negative classifications, crucial for clinical applications. The minimal variation in performance between the base and augmented test datasets further indicates the model's inherent stability and potential for generalization to unseen data. Beyond these technical achievements, this research has broader implications for global healthcare. The automated detection mechanism offers a potential solution for regions with limited access to specialized radiological expertise, particularly in low-resource settings where pediatric pneumonia mortality rates remain alarmingly high. The model's architecture, while achieving high performance, maintains computational efficiency, making it potentially scalable for widespread clinical deployment and integration into existing healthcare workflows. This work also exemplifies the transformative potential of interdisciplinary research, bridging computer science, medical imaging, and healthcare innovation. While the current results are promising, we acknowledge limitations, particularly regarding the dataset's limited diversity, which poses challenges for generalization to broader populations. Future research will address these limitations and explore several key avenues: expanding the model to include multi-class classification in order to distinguish between bacterial, viral, and fungal pneumonia types; conducting comprehensive testing and validation across more diverse demographic and geographic datasets to assess true generalizability; developing more advanced visualization techniques to explain the model's decision-making process and build clinical trust; and collaborating with healthcare institutions to develop seamless, clinically validated integration strategies for real-world deployment.

References

1. WHO. Pneumonia in children. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/pneumonia, 20 November 2024.
2. Megha Trivedi and Abhishek Gupta. "A lightweight deep learning architecture for the automatic detection of pneumonia using chest X-ray images". In: Multimedia Tools and Applications, Vol. 81, 2022, pp. 5515–5536. ISSN: 1573-7721. DOI: 10.1007/s11042-021-11807-x
3. Anuja Acharya and Rajalakshmi Satapathy. "A Deep Learning Based Approach towards the Automatic Diagnosis of Pneumonia from Chest Radio-Graphs". In: Biomedical and Pharmacology Journal, Vol. 13, issue 1, Mar. 2020, pp. 449–455. DOI: 10.13005/bpj/1905
4. Shiv Kale, Jigisha Patil, Akshada Kshirsagar, and Varsha Bendre. "Early Lungs Tuberculosis Detection Using Deep Learning". In: Springer Nature Singapore, Vol. 1, Jan. 2022, pp. 287–294. ISBN: 978-981-16-6308-6. DOI: 10.1007/978-981- 16-6309-329
5. Sivaramakrishnan Rajaraman, Sema Candemir, Incheol Kim, George Thoma, and Sameer Antani. "Visualization and Interpretation of Convolutional Neural Network Predictions in Detecting Pneumonia in Pediatric Chest Radiographs". In: Applied Sciences, Vol. 8, Issue 10, 2018. ISSN: 2076-3417. DOI: 10.3390/app8101715. URL: https://www.mdpi.com/2076-3417/8/10/1715
6. Rajeev Gupta, Yatendra Sahu, Nilesh Kunhare, Abhishek Gupta, and Deo Prakash. "Deep Learning-based Mathematical Model for Feature Extraction to Detect Corona Virus Disease using Chest X-Ray Images". In: International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 29, Issue 6, Nov. 2021, pp. 921-947. DOI: 10.1142/S0218488521500410
7. Karim Hammoudi, Halim Benhabiles, Mahmoud Melkemi, Fadi Dornaika, Ignacio Arganda-Carreras, Dominique Collard, and Arnaud Scherpereel. "Deep Learning on Chest X-ray Images to Detect and Evaluate Pneumonia Cases at the Era of COVID-19". In: Journal of medical systems, Vol. 45, Article No. 7, 2021, p. 75. DOI: 10.1007/s10916-021-01745-4
8. Daniel S. Kermany et al. "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning". In: Cell, Vol. 172, Issue 5, 2018, PP. 1122–1131. DOI: 10.1016/j.cell.2018.02.010
9. Kalaivani Shanmugam and Itshak Dinstein. "Haralick RM, Shanmuga K, Dinstein ITextural features for image classification. IEEE Trans Syst Man Cybern 3: 610-621". In: Systems, Man and Cybernetics, IEEE Transactions, Vol. SMC-3, Issue 6, Dec. 1973, pp. 610 –621. DOI: 10.1109/TSMC.1973.4309314
10. Timo Ojala, Matti Pietik¨ainen, and T. Maenpaa. "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns". In: Pattern Analysis and Machine Intelligence, IEEE Transactions, Vol. 24, Issue 7, Aug. 2002, pp. 971–987. DOI: 10.1109/TPAMI.2002.1017623
11. Anil K. Jain and Farshid Farrokhnia. "Unsupervised texture segmentation using Gabor filters". In: Pattern recognition, Vol. 24, Issue 12, 1991, pp. 1167–1186. DOI: 10.1016/0031-3203(91)90143-S
12. Corinna Cortes and Vladimir Vapnik. "Support-vector networks". In: Machine learning, Vol. 20, Issue 3, 1995, pp. 273–297. DOI:10.1007/BF00994018
13. Leo Breiman. "Random forests". In: Machine learning, Vol. 45, Issue 1, 2001, pp. 5–32. DOI: 10.1023/A:1010950718922
14. Thomas Cover and Peter Hart. "Nearest neighbor pattern classification". In: IEEE transactions on information theory, Vol. 13, Issue 1, 1967, pp. 21–27. DOI: 10.1109/TIT.1967.1053964
15. Laurence Monnier-Cholley, Heber MacMahon, Shigehiko Katsuragawa, Junichi Morishita, Takayuki Ishida, and KunioDoi. "Computer-aided diagnosis for detection of interstitial opacities on chest radiographs". In: American journal of roentgenology, Vol. 171, Issue 6, 1998, pp. 1651–1656. DOI: 10.2214/ajr.171.6.9843307
16. Hiroyuki Abe, Heber Macmahon, Junji Shiraishi, Qiang Li, Roger Engelmann, and Kunio Doi. "Computer-aided diagnosis in chest radiology". In: Seminars in ultrasound, CT and MR 25, Vol. 25, Issue 5, Nov. 2004, pp. 432–437. DOI: 10.1053/j.sult.2004.02.004
17. Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald Summers. "ChestX-ray8:Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of CommonThorax Diseases". In Proceedings of the IEEE conference on computer vision and pattern recognition, 21-26 July 2017, Honolulu, HI, USA, pp. 2097–2106. DOI: 10.48550/arXiv.1705.02315
18. M.B. Badawi, Rania Elgohary, Mostafa Tarek, Mohamed EzzAlRegal, Abdulrahman Ahmed, Ahmed Samir, and Nour Ehab. "Skin Cancer Classification and Segmentation Using Deep Learning". In: International Journal of Telecommunications, Vol. 4, Issue 1, Feb. 2024, pp. 1–23. DOI: 10.21608/ijt.2024.280957.1045
19. Daniel Kermany, Kang Zhang, and Michael Goldbaum. "Labeled Optical Coherence Tomography (OCT) and ChestX-Ray Images for Classification". In: Mendeley Data Version 2, 2018. DOI: 10.17632/rscbjbr9sj.2
20. Karen Simonyan and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". In:arXiv (2015). DOI: 10.48550/arXiv.1409.1556
21. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition". In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 27-30 June 2016, Las Vegas, NV, USA, pp. 770–778. DOI: 10.1109/CVPR.2016.90

22. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. "Rethinking the InceptionArchitecture for Computer Vision". In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),), 27-30 June 2016, Las Vegas, NV, USA, pp. 2818–2826. DOI: 10.1109/CVPR.2016.308
23. Okeke Stephen, Mangal Sain, Uchenna Joseph Maduh, and Do-Un Jeong. "An efficient deep learning approach topneumonia classification in healthcare". In: Journal of healthcare engineering, Vol. 1, 2019, p. 4180949.
24. H Omar and A Babalık. "Detection of pneumonia from X-ray images using convolutional neural network". In: International Conference on Data Science, Machine Learning and Statistics - 2019 (DMS-2019), 26-29 June 2019, Faculty of Economics and Administrative Sciences, Van Yuzuncu Yil University, 65080 Van, Turkey, pp 183-185.
25. Huaiguang Wu, Pengjie Xie, Huiyi Zhang, Daiyi Li, and Ming Cheng. "Predict pneumonia with chest X-ray images basedon convolutional deep neural learning networks". In: Journal of Intelligent & Fuzzy Systems, Vol. 39, Issue 3, 2020, pp. 2893–2907.
26. Sabyasachi Chakraborty, Satyabrata Aich, Jong Seong Sim, and Hee-Cheol Kim. "Detection of pneumonia from chestx-rays using a convolutional neural network architecture". In: International conference on future information & commu-nication engineering, Vol. 11, Issue 1, 2019, pp. 98–102.
27. Prateek Chhikara, Prabhjot Singh, Prakhar Gupta, and Tarunpreet Bhatia. "Deep convolutional neural network with transfer learning for detecting pneumonia on chest X-rays". In: Advances in Bioinformatics, Multimedia, and Electronics Circuits and Signals: *Proceedings of GUCON 2019* (pp. 155-168). Springer Singapore.
28. Gaobo Liang and Lixin Zheng. "A transfer learning method with deep residual network for pediatric pneumonia diagnosis". In: Computer methods and programs in biomedicine Vol. 187, 2020, p. 104964.