





# Customized Hybrid Deep Learning Approach for Symptom-Driven Pneumonia Diagnosis

Nouf Alshawoosh, Rahaf Alerwi, Maha Alsayed, Samah Alhilale

**Supervised By:** Dr.Tanveer Hussain

# Abstract

This report presents the development of a hybrid deep learning model aimed at enhancing pneumonia detection in medical diagnostics by integrating image analysis and natural language processing (NLP). Leveraging a VGG16-based U-Net model for analyzing chest X-rays (CXRs) and a fine-tuned BioBERT model for symptom classification, the system offers a robust diagnostic tool.

The VGG16-Unet model achieved an Intersection over Union (IoU) of 0.88 and an accuracy of 95% on the validation set. The integrated approach demonstrated strong potential in improving diagnostic accuracy and reliability, combining the strengths of image segmentation and NLP for comprehensive medical diagnosis.

# Table of Contents

Abstract ..... 2

Introduction ..... 3

Related Work ..... 5

Methodology ..... 6

Results ..... 10

Discussion ..... 11

Challenges ..... 12

Future Work ..... 13

Conclusion ..... 14

References ..... 15

# Introduction

Pneumonia remains one of the most pressing global health challenges, responsible for over 2.5 million deaths annually[1], with children and the elderly being particularly vulnerable. Despite advances in medical imaging, the traditional method of diagnosing pneumonia—primarily through the manual interpretation of chest X-rays—continues to be a time-consuming and error-prone process.

The complexity of interpreting chest X-rays, combined with the high volume of cases that clinicians must assess, often leads to delays in diagnosis and treatment, potentially worsening patient outcomes. Given the limitations of current diagnostic practices, there is a critical need for innovative solutions that can enhance both the accuracy and efficiency of pneumonia detection.

This project seeks to address this challenge by developing a deep learning-based model that integrates image analysis with natural language processing techniques. By leveraging the strengths of these two domains, the model is designed to analyze chest X-rays alongside patient symptoms, providing a more comprehensive and accurate diagnostic tool. The main objective is to significantly improve the speed and reliability of pneumonia diagnosis, thereby enabling timely and effective treatment interventions and reducing the burden of this life-threatening disease on vulnerable populations.

# Related Work

Previous studies have explored various approaches to pneumonia detection, primarily focusing on either image analysis or text processing. Linghua Wu et al. (2024)[2] developed an anchor-free model for pneumonia detection, achieving an average precision of 51.5%. Anand Nayyar et al. (2020)[3] introduced a Mask-RCNN model with a 0.155 IoU score, demonstrating improvements over methods like YOLOv3 and ResNet.

This project distinguishes itself by integrating both image and text data into a single diagnostic model, leveraging the strengths of both modalities to improve diagnostic accuracy and reliability.

# Methodology

The proposed Pneumonia Diagnosis Assistant integrates deep learning for image analysis and natural language processing (NLP) to assess the likelihood of pneumonia in patients. The methodology is divided into two main components:

**1. Image Processing Pipeline:**

A U-Net model with a VGG16 encoder is utilized to process chest X-ray images. The VGG16 encoder extracts deep features from the input X-ray, while the U-Net decoder upscales these features to generate a segmentation map. This map highlights potential regions of interest within the lungs that may indicate pneumonia.

**2. Text Processing Pipeline:**

The BioBERT model, pre-trained on biomedical literature, processes the clinical symptoms provided by the user. It analyzes the text input to correlate the symptoms with known medical knowledge about pneumonia.

The outputs from both pipelines are integrated within the Pneumonia Diagnosis Assistant. This integration enables the system to provide a comprehensive diagnostic recommendation based on the analysis of both the chest X-ray findings and the patient's symptoms.

The figure below illustrates this methodology.

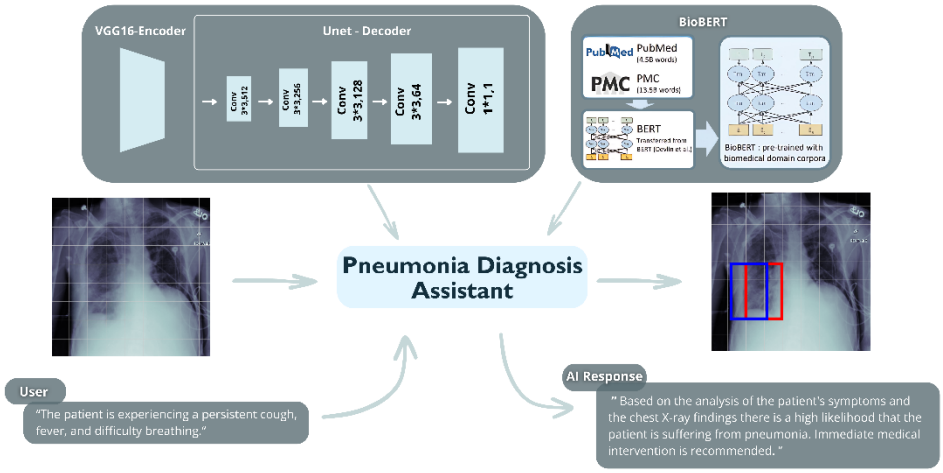



*Figure 1 Pneumonia Diagnosis Assistant Model*

This dual approach ensures a robust and accurate assessment, aiding healthcare professionals in making informed decisions regarding the patient's care.

The pneumonia diagnosis assistant is composed of two models:

## 1. VGG16-Unet Model Architecture

**RSNA Pneumonia Dataset:** The RSNA Pneumonia dataset [4] is a collection of chest radiograph images used to train algorithms for detecting and localizing pneumonia in medical images. This dataset includes expert annotations to assist in identifying visual signals of pneumonia. It was part of a competition organized by the Radiological Society of North America (RSNA) aimed at leveraging AI to improve pneumonia detection.

**VGG16 Model Initialization:** The VGG16 model, a widely recognized convolutional neural network pre-trained on the ImageNet dataset, serves as the base of the architecture. The top classification layer is excluded (include_top=False), leaving only the convolutional layers for feature extraction.

**Input Configuration:** The input to the model is defined as (imgHeight, imgWidth, 3), where imgHeight and imgWidth represent the dimensions of the input images, and 3 corresponds to the RGB color channels.

**Frozen Layers:** All layers of the VGG16 model are set to non-trainable (layer. trainable = False) to retain the pre-trained weights, thereby preserving the model's learned features during training.

**Encoder (Feature Extraction):** extracts feature maps at various levels of abstraction from specific convolutional layers of the VGG16 model:

- **Block 1:** Outputs from block1_conv2 capture low-level features.
- **Block 2:** Outputs from block2_conv2 capture slightly higher-level features.
- **Block 3:** Outputs from block3_conv3 capture mid-level features.
- **Block 4:** Outputs from block4_conv3 capture more abstract features.
- **Block 5:** Outputs from block5_conv3 capture the most abstract features.
- **Block 6:** Outputs from block5_pool mark the end of the VGG16 encoder, serving as the bottleneck layer.

**Decoder (Upsampling and Reconstruction):**

- **Transposed Convolutions:** The decoder upsamples the feature maps back to the original image size using transposed convolution layers (deconvolutions).
- **Skip Connections:** The model employs skip connections by concatenating feature maps from corresponding encoder layers with the upsampled feature maps, aiding in the recovery of spatial information lost during downsampling.
- **Block 6 to Block 5:** A Conv2DTranspose layer upsamples the output from Block 6. and the upsampled output is concatenated with the output from Block 5. Also, three Conv2D layers refine the features, followed by batch normalization and dropout for regularization.
- **Block 5 to Block 4:** Similar steps are performed but with fewer filters (256).
- **Block 4 to Block 3:** The same process continues with further upsampling, concatenation, convolution, and regularization.
- **Block 3 to Block 2:** The output is upsampled to align with the output from Block 2.
- Block 2 to Block 1: Final upsampling and concatenation with low-level features from Block 1.
- **Output Layer:** A Conv2D layer with a kernel size of 1 and a sigmoid activation function is used as the output layer. This produces a binary segmentation map where each pixel represents the probability of belonging to a specific class (e.g., object vs. background).

## 2. BioBERT

**Dataset Preparation:** The BioBERT model, a specialized language representation model for biomedical text, was fine-tuned using a symptom-based disease labeling dataset sourced from Kaggle [5]. This dataset included 1,200 English sentences describing symptoms related to 24 diseases, including pneumonia.

**Pneumonia-Specific Data:** For this project, the focus was on the 50 sentences related to pneumonia. An additional 50 synthetic entries representing healthy symptoms were generated using ChatGPT, ensuring a balanced dataset.

**Text Augmentation:** Techniques like synonym replacement were employed to diversify the dataset and improve the model's robustness.

**Training Process:**

| Training Configuration | Details |
|---|---|
| Number of Epochs | 5 epochs (to balance underfitting and overfitting) |
| Batch Size | 8 (optimized for GPU memory usage and stable gradient updates) |
| Learning Rate and Scheduler | Initial learning rate of 0.01 with a warmup phase of 500 steps; a learning rate scheduler was used to gradually reduce the learning rate for improved convergence |
| Weight Decay | 0.01 (applied to regularize the model and prevent overfitting) |
| Training Loss | 0.62 (demonstrating the model's ability to learn the associations between symptoms and disease labels) |
| Validation Loss | 0.49 (indicating good generalization performance) |

*Figure 2 Table of training configuration and performance details*

The outputs from the fine-tuned BioBERT model, which classified symptoms as either pneumonia-related or not, were integrated with the VGG16-Unet model's analysis of chest X-rays. This combined approach allowed the system to provide comprehensive diagnostic suggestions by leveraging both textual and visual data, thereby improving diagnostic accuracy and reliability.

# Results

The VGG16-Unet model demonstrated robust performance in detecting pneumonia from chest X-rays, achieving an Intersection over Union (IoU) of 0.88 and an accuracy of 95% on the validation set. These metrics highlight the model's ability to accurately segment and identify pneumonia-affected regions in medical images, ensuring reliable visual diagnostics.

In parallel, the fine-tuned BioBERT model, enhanced with a Named Entity Recognition (NER) system, effectively classified patient symptoms as either pneumonia-related or not. The NER model accurately extracted relevant medical entities and symptoms from patient records, which were then processed by BioBERT. This combination allowed for precise and context-aware symptom analysis, demonstrating good generalization across different cases.

When integrated, the VGG16-Unet and BioBERT models provided a powerful diagnostic tool that leveraged both visual and textual data. The integrated model not only confirmed the presence of pneumonia through detailed image analysis but also contextualized these findings with patient symptoms and medical history. This dual-modality approach resulted in a comprehensive diagnostic suggestion, significantly improving the accuracy and reliability of pneumonia detection. The integrated system, therefore, offers a promising solution for enhancing early diagnosis and treatment, ultimately contributing to better patient outcomes.

# Discussion

The hybrid model's success underscores the potential of combining deep learning techniques from different domains to enhance medical diagnostics. The integration of NLP with image analysis allows for a more accurate diagnosis, capturing both visual indicators and patient-reported symptoms.

Compared to previous models that focused solely on image analysis or text processing, the integrated approach provides a more holistic view of the patient's condition, improving diagnostic accuracy and reliability.

This project's findings have significant implications, particularly in enhancing diagnostic tools for complex medical conditions like pneumonia.

# Challenges

## Technical Challenges:

**Dataset Balancing for BioBERT Fine-Tuning:** Addressing the imbalance in the symptom-based dataset required generating synthetic data and using augmentation techniques to ensure the model didn't become biased towards more common symptoms.

**Seamless Integration of Image and Text Models:** Integrating the VGG16-Unet and BioBERT models required careful design to ensure they communicated effectively, synchronized outputs, and provided a coherent diagnostic result.

**Hyperparameter Optimization:** Optimizing the training process for both models involved fine-tuning multiple hyperparameters, such as learning rates, batch sizes, and regularization techniques, to balance accuracy and generalizability.

## Research Limitation:

**Limited Dataset Diversity:** The datasets used were not fully representative of the variety of imaging conditions and symptom presentations, which could affect the model's accuracy in diverse real-world scenarios.

**Dependence on Publicly Available Datasets:** The reliance on publicly available datasets may limit the generalizability of the model's performance across different patient populations and clinical settings.

**Need for Broader Validation:** Future research should include more diverse datasets from multiple clinical sources to validate the model's performance across different demographics and medical conditions.

# Future Work

**Model Expansion:**

Future research could focus on expanding the model's capabilities to detect multiple lung-related conditions, thereby improving its versatility in medical diagnostics.

**Fine-Tuning BioBERT:**

Further fine-tuning of the BioBERT model using larger, more diverse symptom datasets could enhance its accuracy in symptom extraction and classification, enabling more precise integration with imaging data.

**Data Augmentation:**

Implementing more advanced data augmentation techniques could further improve the model's robustness and generalizability.

# Conclusion

This project successfully developed a hybrid deep learning model that integrates image analysis and text processing to detect pneumonia, achieving high accuracy and reliability. The model's strong performance highlights the potential of combining different deep learning techniques to enhance diagnostic tools for healthcare professionals.

The integrated approach not only improves pneumonia diagnosis but also lays the groundwork for future advancements in AI-driven medical diagnostics. Expanding this model to other conditions could significantly impact healthcare, offering faster and more accurate diagnostic tools for a wide range of medical issues. Continued research and development in integrating multi-modal data for medical diagnostics are recommended to further enhance the accuracy and reliability of AI-driven tools in healthcare.

# References

[1] Cohen, M., Levine, S. M., & Zar, H. J. (2022). World Lung Day: impact of "the big 5 lung diseases" in the context of COVID-19. *American journal of physiology. Lung cellular and molecular physiology*, 323(3), L338–L340. https://doi.org/10.1152/ajplung.00261.2022

[2] Wu, L., Zhang, J., Wang, Y., Ding, R., Cao, Y., Liu, G., Liufu, C., Xie, B., Kang, S., Liu, R., Li, W., & Guan, F. (2024). Pneumonia detection based on RSNA dataset and anchor-free deep learning detector. *Scientific Reports*, *14*(1), 1929. https://doi.org/10.1038/s41598-024-52156-7

[3] Nayyar, A., Jain, R., & Upadhyay, Y. (2020). *Object detection based approach for Automatic detection of Pneumonia*. https://doi.org/10.1109/icabcd49160.2020.Year).

[4] Kaggle. (n.d.). *RSNA Pneumonia detection challenge*. Kaggle. Retrieved August 31, 2024, from https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/overview/acknowledgements

[5] Kaggle. (n.d.). *Symptom-based disease labeling dataset*. Kaggle. Retrieved August 31, 2024, from https://www.kaggle.com/datasets/krish0202/symptom-based-disease-labeling-dataset