Automated Detection of Pneumonia from Chest X-Ray Images Using Machine Learning

Ranganathan Srinivasan

Abstract

Pneumonia remains a major global health concern, and early diagnosis from chest X-ray images is critical for effective treatment. Here we evaluate machine learning methods for automated pneumonia detection using chest radiographs. Three classifiers: VGG16, Support Vector Machine (SVM), and XGBoost were implemented and compared in terms of accuracy, precision, recall, and F1-score. VGG16 achieved the highest overall accuracy (78.3%) with balanced performance across classes, while SVM demonstrated very high sensitivity for pneumonia (recall = 0.99) but lower specificity for normal cases. XGBoost provided intermediate performance, favoring pneumonia detection but misclassifying some normal cases. To further interpret the learned feature space, Uniform Manifold Approximation and Projection (UMAP) was applied, revealing distinct but overlapping clusters of pneumonia and control samples. These results highlight both the potential and limitations of machine learning for pneumonia detection, emphasizing that while models can assist in screening, challenges in specificity and overlapping feature representations remain. Future work should focus on larger datasets, integration of clinical information, and improved interpretability to enhance real-world applicability.

Introduction

Pneumonia is a leading cause of morbidity and mortality worldwide, affecting millions of people each year across all age groups. According to the World Health Organization, pneumonia is responsible for approximately 15% of all deaths in children under five, and it remains a major health burden among elderly and immunocompromised populations[5, 3]. Early and accurate diagnosis is crucial, as timely treatment can significantly reduce complications and improve patient outcomes.

Chest radiography (X-ray imaging) is the most commonly used diagnostic tool for pneumonia due to its wide availability, low cost, and ability to reveal lung abnormalities. However, interpretation of chest X-rays is challenging even for experienced radiologists, as visual differences between healthy lungs, pneumonia, and other thoracic conditions can be subtle and overlapping. These challenges often lead to inter-observer variability and diagnostic delays, particularly in resource-limited settings where expert radiologists may not be available.

In recent years, machine learning (ML) and deep learning techniques have shown promising results in automating disease detection from medical images[1]. For pneumonia detection, ML models can learn discriminative patterns from large collections of chest X-ray images, potentially assisting radiologists in screening and diagnosis. At the same time, dimensionality reduction and visualization methods such as Uniform Manifold Approximation and Projection (UMAP)[2] can provide insight into the latent structure of the data, offering a way to interpret and explore the separability of pneumonia and non-pneumonia cases in reduced-dimensional spaces.

Despite the progress, several challenges remain. Many existing studies focus primarily on deep neural networks trained on large datasets, which may not always be practical in smaller-scale clinical applications. Furthermore, there is a growing need for interpretable methods that not only achieve high classification accuracy but also provide visual understanding of how cases cluster in feature space.

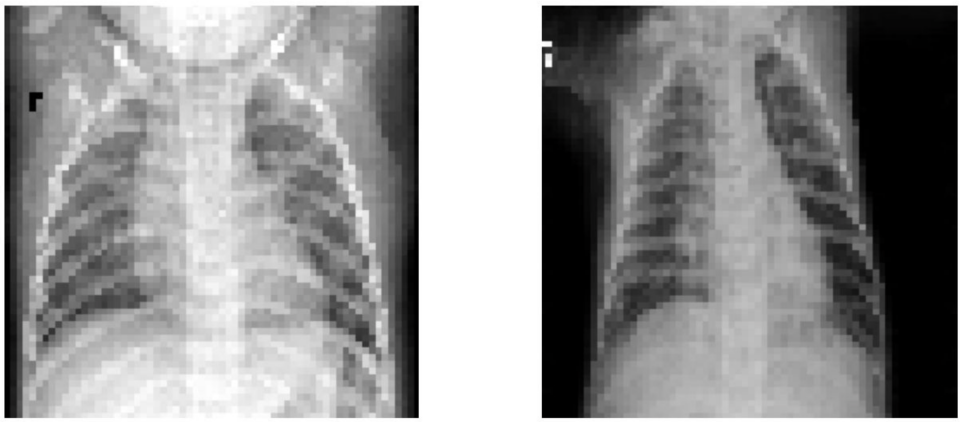


Figure 1: Sample chest X-ray images illustrating the difference between a normal lung (left) and a pneumonia-affected lung (right). The pneumonia X-ray shows increased opacities consistent with infection.

In this work, we present a machine learning-based framework for pneumonia detection from chest X-ray images. We evaluate classification performance across multiple algorithms and employ UMAP to visualize the latent feature space, highlighting the separability between pneumonia and normal cases. Our experiments demonstrate both the effectiveness of machine learning methods for automated pneumonia detection and the value of UMAP as a tool for interpretability and data exploration.

Methods

Dataset

We utilized a publicly available chest X-ray dataset containing images of pneumonia and normal cases (see Figure 1). The dataset was curated to include frontal chest radiographs labeled by clinical experts. Images were divided into two main classes: pneumonia and normal. To ensure balanced evaluation, the dataset was split into training, validation, and test sets with stratification by class.

Machine Learning Classifiers

We employed three different machine learning classifiers to evaluate the effectiveness of automated pneumonia detection:

1. **VGG16:**[4] We employed the VGG16 deep convolutional neural network as a feature extractor and end-to-end classifier. Pretrained weights on ImageNet were used for initialization, followed by fine-tuning on the pneumonia dataset. The final fully connected layers were replaced with task-specific layers to classify chest X-ray images into pneumonia or normal categories. Training was performed using the Adam optimizer with categorical cross-entropy loss.

2. **Support Vector Machine (SVM):**[6] A supervised learning algorithm that finds an optimal hyperplane to separate classes in feature space. For this study, features extracted from

VGG16 were used as input to the SVM classifier.

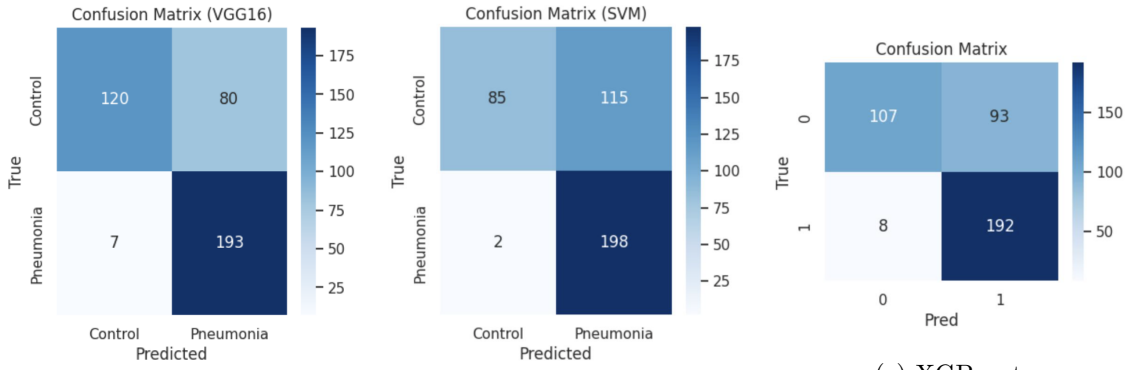
3. **Extreme Gradient Boosting (XGBoost):** A gradient-boosted decision tree model designed for efficient and scalable learning. Similar to the SVM approach, deep features extracted from VGG16 were used as input to train the XGBoost classifier.

Dimensionality Reduction and Visualization

To gain insights into the learned representations, we employed Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction. UMAP projects high-dimensional features into a two-dimensional space while preserving local and global data structure. The resulting visualizations allow us to observe the clustering of pneumonia and normal cases, providing an interpretable view of classifier performance.

Evaluation Metrics

The performance of all classifiers was evaluated using standard metrics, including accuracy, precision, recall, and F1-score. Additionally, confusion matrices were generated to assess class-wise performance. UMAP visualizations were used qualitatively to analyze the separability of pneumonia and normal cases in feature space.



(a) VGG16 (b) SVM (c) XGBoost

Figure 2: Confusion matrices for the three classifiers: (a) VGG16, (b) SVM, and (c) XGBoost.

Results

Classification Performance

We evaluated three classifiers: VGG16, Support Vector Machine (SVM), and XGBoost on the chest X-ray dataset. The performance was assessed using accuracy, precision, recall, and F1-score, along with confusion matrices for detailed class-wise analysis.

VGG16: The VGG16 model achieved a test accuracy of 78.3%. It demonstrated high sensitivity for pneumonia cases, with a recall of 0.96 and precision of 0.71, resulting in an F1-score of 0.82. For control cases, the precision was high (0.94) but recall was comparatively lower (0.60), indicating that some normal cases were misclassified as pneumonia. The confusion matrix (Figure 2) shows 193 true positives and 120 true negatives.

SVM: The SVM classifier obtained a test accuracy of 70.8%. It exhibited strong performance in detecting pneumonia (recall = 0.99, F1-score = 0.77), but performance on control cases was limited (recall = 0.42, F1-score = 0.59). This indicates a bias toward predicting pneumonia, as reflected in the confusion matrix (Figure 2), where 198 pneumonia cases were correctly classified but 115 control cases were misclassified as pneumonia.

XGBoost: The XGBoost model achieved a test accuracy of 74.8%. Similar to VGG16 and SVM, it favored pneumonia detection with a recall of 0.96 and an F1-score of 0.79. For control cases, precision was high (0.93), but recall was relatively low (0.54). The confusion matrix (Figure 2) shows that 192 pneumonia cases were correctly classified, but 93 control cases were misclassified as pneumonia.

UMAP Visualization

To further investigate the separability of pneumonia and control cases, we applied Uniform Manifold Approximation and Projection (UMAP) to the feature space. As shown in Figure 3, UMAP revealed two partially overlapping clusters corresponding to pneumonia and control samples. Although distinct groupings are visible, some degree of overlap remains, which may explain misclassifications observed across all models.

All three classifiers achieved comparable performance, with VGG16 showing the best balance between precision and recall. SVM demonstrated the highest recall for pneumonia but struggled with specificity, while XGBoost provided a middle ground between the two. UMAP visualization confirmed that the pneumonia and control cases are separable to a large extent but not perfectly distinct, highlighting the inherent difficulty of the task.

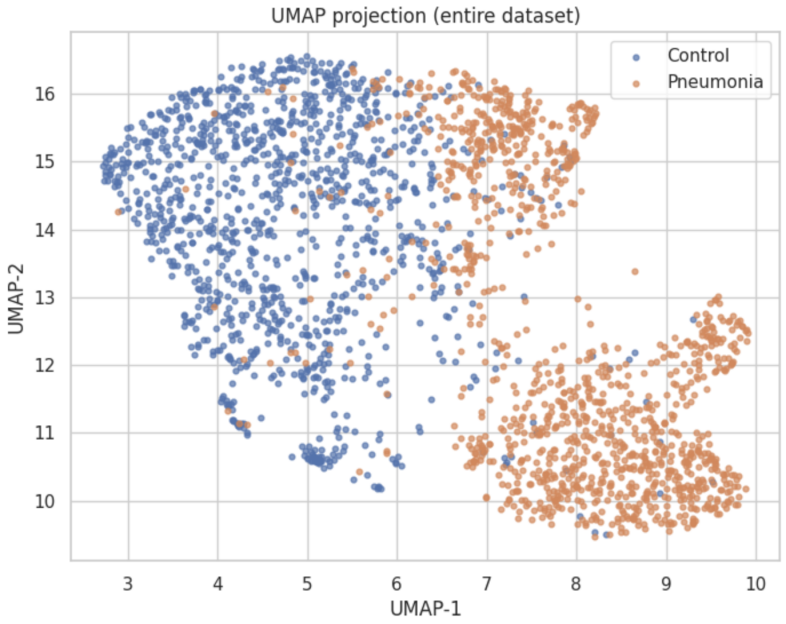


Figure 3: UMAP projection of the dataset showing clustering of pneumonia and control cases.

Conclusion

This paper explored the use of machine learning classifiers and dimensionality reduction techniques for pneumonia detection from chest X-ray images. Three classifiers VGG16, SVM, and XGBoost

were compared in terms of accuracy, precision, recall, and F1-score, while UMAP was employed to visualize feature separability between pneumonia and control cases.

Across all experiments, the models consistently showed stronger performance in detecting pneumonia cases compared to normal cases. Both VGG16 and XGBoost achieved balanced overall performance, with VGG16 reaching the highest accuracy of 78.% and providing a good trade-off between precision and recall. XGBoost achieved slightly lower accuracy (74.8%) but showed comparable recall for pneumonia. The SVM classifier yielded the highest recall for pneumonia (0.99), indicating an ability to capture nearly all pneumonia cases, but this came at the cost of misclassifying many normal cases, reflected in its low recall for controls (0.42). These results highlight a common trade-off in medical image classification: maximizing sensitivity to avoid missed disease cases may increase false positives, which has implications for clinical deployment.

The UMAP projection provided additional insight into the structure of the data. Pneumonia and control cases largely clustered into distinct regions, yet noticeable overlap was present. This overlap corresponds to cases that the models misclassified, suggesting that some X-ray images contain subtle or ambiguous features that are challenging even for automated methods to distinguish. This finding reinforces the clinical reality that pneumonia diagnosis can be complex and often requires complementary information beyond imaging, such as patient history and laboratory results.

Taken together, the results demonstrate the potential of machine learning models to support radiologists in pneumonia screening. VGG16, in particular, showed strong balanced performance, while SVM's high sensitivity may be useful in triaging cases when false negatives must be minimized. However, the imperfect separability observed in UMAP and the reduced specificity of some classifiers underline the need for careful integration of such models into clinical workflows. Rather than replacing expert interpretation, these models may serve best as decision-support tools to flag suspicious cases for further review.

So, this work confirms that machine learning methods, coupled with visualization tools such as UMAP, can provide valuable assistance in pneumonia detection from chest radiographs. While the current models achieve promising results, future work should focus on improving specificity, incorporating larger and more diverse datasets, and exploring hybrid approaches that combine imaging features with clinical data. Such directions may enhance both the accuracy and interpretability of automated pneumonia detection systems, ultimately supporting more reliable and timely patient care.

References

[1] Gregory F Cooper et al. "An evaluation of machine-learning methods for predicting pneumonia mortality". In: *Artificial intelligence in medicine* 9.2 (1997), pp. 107–138.

[2] Leland McInnes, John Healy, and James Melville. "Umap: Uniform manifold approximation and projection for dimension reduction". In: *arXiv preprint arXiv:1802.03426* (2018).

[3] Andrew Prayle, Maria Atkinson, and Alan Smyth. "Pneumonia in the developed world". In: *Paediatric respiratory reviews* 12.1 (2011), pp. 60–69.

[4] Hussam Qassim, Abhishek Verma, and David Feinzimer. "Compressed residual-VGG16 CNN model for big data places image recognition". In: *2018 IEEE 8th annual computing and communication workshop and conference (CCWC)*. IEEE. 2018, pp. 169–175.

[5] Olli Ruuskanen et al. "Viral pneumonia". In: *The Lancet* 377.9773 (2011), pp. 1264–1275.

[6] Shihong Yue, Ping Li, and Peiyi Hao. "SVM classification: Its contents and challenges". In: *Applied Mathematics-A Journal of Chinese Universities* 18.3 (2003), pp. 332–342.