Introduction

Pneumonia, a prevalent respiratory infection, poses a significant global health challenge, affecting individuals of all ages. The timely and accurate detection of pneumonia is critical for effective treatment and the prevention of severe complications. Moreover, distinguishing between viral and bacterial pneumonia is essential, as it aids healthcare professionals in tailoring treatment plans to better address the specific cause of the illness.

Non-invasive imaging techniques, particularly X-ray imaging, have become indispensable tools in the diagnosis of pneumonia. However, the manual interpretation of X-rays can be time-consuming and subject to human error, especially in resource-limited settings. To overcome these challenges, the integration of Machine Learning (ML) and Deep Learning (DL) techniques into medical diagnostics has gained considerable attention. These advanced computational methods not only reduce the workload of medical practitioners but also enhance diagnostic accuracy and efficiency.

In recent years, the availability of large-scale medical image datasets and advancements in deep learning have facilitated the adoption of sophisticated models, such as Convolutional Neural Networks (CNNs) and Transformers, for pneumonia detection. Among these, the ResNet model and Vision Transformer (ViT) have shown significant promise in medical image analysis.

In this study, we aim to improve pneumonia detection and classification by leveraging both ResNet and Vision Transformer models. Specifically, we explore various training strategies for the ResNet model, including training from scratch, transfer learning, and fine-tuning the last layer. Additionally, we investigate the use of a self-supervised Transformer model, DINO ViT-B/16, fine-tuned with a linear classifier for pneumonia classification.

Furthermore, we employ Extreme Learning Machines (ELM), a fast and efficient method for training linear classifiers, in conjunction with Principal Component Analysis (PCA) based on Singular Value Decomposition (SVD) for feature extraction and elimination. This approach provides a rapid alternative to traditional gradient-based methods, achieving successful generalization with training speeds comparable to one epoch of fine-tuning.