**Pneumonia[[1]](#footnote-1)**

Pneumonia is an infection of the respiratory system that primarily impacts the lungs. This condition arises when harmful microorganisms infiltrate the lung tissue, leading to the destruction of the pulmonary alveoli, which are responsible for oxygen absorption. As a result, these areas become filled with inflammatory fluid, impairing their functionality. Pneumonia can be triggered by a variety of pathogens, including bacteria, viruses, and parasites.

Common symptoms associated with pneumonia include coughing, fever, chills, difficulty breathing, and chest discomfort. The intensity of pneumonia can range significantly, from mild cases to those that pose a serious threat to life. In more severe instances, pneumonia may result in complications such as respiratory failure, sepsis, and potentially death. Certain groups, including infants, young children, the elderly, and individuals with compromised immune systems, are at an increased risk of developing pneumonia.

Annually, pneumonia impacts approximately 450 million individuals worldwide, and leads to around 4 million fatalities. The advent of antibiotics and vaccines in the 20th century has significantly enhanced survival rates. However, pneumonia continues to be a primary cause of mortality in developing nations, as well as among the elderly, infants, and those with chronic health conditions.

**Chest X-Rays and Pneumonia Detection in Radiology[[2]](#footnote-2)**

Chest X-rays are one of the most used imaging techniques in radiology, especially for diagnosing conditions affecting the lungs and thoracic region. Interpreting chest X-rays can be challenging due to the complex anatomy and various structures that must be evaluated systematically. A typical chest X-ray includes both a posterior-anterior (PA) and a lateral view, allowing radiologists to assess the lungs, soft tissues, bones, and the mediastinum. In the case of pneumonia, chest X-rays often reveal opacities in the lungs, which may present as either patchy or confluent areas depending on the extent of infection. Pneumonia typically appears as a consolidation of lung tissue, with visible air bronchograms indicating infection within the airspaces.

In radiological practice, detecting pneumonia is crucial as it can lead to complications if untreated. Radiologists evaluate signs such as the silhouette sign, which helps localize lesions, and air bronchograms, which suggest alveolar disease.

**Active Learning in Machine Learning[[3]](#footnote-3)**

Active learning is a subfield of machine learning and artificial intelligence where the learning algorithm selects the data from which it learns, effectively allowing it to be "curious." The main idea is that by choosing its training data, the algorithm can perform better with less labeled data. This is particularly valuable because many supervised learning systems require large amounts of labeled instances to perform well, and obtaining these labels can be costly, time-consuming, or difficult.

Active learning addresses this issue by asking an oracle, often a human annotator, to label only the most informative unlabeled instances. This approach aims to achieve high accuracy while minimizing the number of labeled instances required, making it an effective solution in situations where data is abundant, but labeled examples are scarce or expensive to acquire.

**Active Learning State-of-the-Art Solutions in X-Ray Imaging**

Active learning has emerged as a promising approach for addressing the challenges of data scarcity and annotation costs in medical image analysis, particularly in the domain of X-Ray imaging. By strategically selecting the most informative images for labeling, active learning algorithms can significantly improve the performance of machine learning models while minimizing the annotation effort required.

According to Biswas et al. (2023)[[4]](#footnote-4), the current main key aspects of active learning technologies used in X-ray imaging are:

* Selective Sampling: Active learning algorithms can identify which unlabeled X-ray images would provide the most value if labeled. This is often done by assessing the uncertainty of the model's predictions.
* Iterative Learning Process: The active learning process is iterative. The model is trained on a small set of labeled data, then it queries additional samples from a larger pool of unlabeled data. After obtaining labels for these samples, the model is retrained, and the cycle continues until a desired performance level is achieved.
* Types of Active Learning: There are several strategies within active learning, including Stream-based Selective Sampling, which means that the model queries one instance at a time and receives feedback before proceeding to the next; Pool-based Sampling, meaning that the model queries multiple instances from a pool of unlabeled data at once; and Membership Query Synthesis which means that the model generates synthetic instances that it is uncertain about and queries for their labels.
* Integration with Federated Learning: Active learning can be combined with federated learning, a technique that allows multiple clients to collaboratively train a shared model without sharing their raw data, allowing multiple healthcare organizations to collaboratively train models on private data without sharing sensitive patient information. This enhances the model's ability to learn from diverse datasets while maintaining confidentiality.

In recent years, several active learning techniques have been applied to X-ray image classification tasks with promising results. Those include the "GOAL" strategy, proposed by Nguyen et al. (2021), focusing on selecting the samples for tagging not only based on uncertainty but also on informativeness, derived by clustering.[[5]](#footnote-5) Hao et al. (2021) proposed another data-efficient methodology that can achieve the same level of accuracy as a non-active-learning model using significantly fewer images and labels, based on CNN unsupervised representation learning, followed by a Gaussian Process (GP) classifier that conducts the active learning pipeline over the learned representations.[[6]](#footnote-6)

While these methods have demonstrated effectiveness in various X-ray image classification tasks, there is still ongoing research to develop even more efficient and robust active learning strategies. Recent advancements in deep learning and generative models have opened new possibilities for active learning, such as using generative models to synthesize new, informative images for labeling. In this project, we explore the application of active learning to the domain of pediatric radiology, specifically focusing on the classification of chest X-rays into "sick" (indicating pneumonia) or "not sick." Our goal is to develop an active learning pipeline that can effectively identify the most informative images for annotation, thereby improving the performance of a pneumonia classification model while minimizing the annotation effort required.

1. https://en.wikipedia.org/wiki/Pneumonia [↑](#footnote-ref-1)
2. https://www.glowm.com/atlas-page/atlasid/chestXray.html [↑](#footnote-ref-2)
3. Settles, B. (2009). Active learning literature survey. [↑](#footnote-ref-3)
4. Biswas, A., Nasim, M. a. A., Ali, M. S., Hossain, I., Ullah, M. A., & Talukder, S. (2023). *Active learning on medical image*. arXiv.org. https://arxiv.org/abs/2306.01827 [↑](#footnote-ref-4)
5. Nguyen, C., Huynh, M. T., Tran, M. Q., Nguyen, N. H., Jain, M., Vo, T. D., ... & Truong, S. Q. H. (2021, August). GOAL: gist-set online active learning for efficient chest X-ray image annotation. In Medical Imaging with Deep Learning (pp. 545-553). PMLR.‏ [↑](#footnote-ref-5)
6. Hao, H., Didari, S., Woo, J. O., Moon, H., & Bangert, P. (2021). Highly efficient representation and active learning framework for imbalanced data and its application to covid-19 x-ray classification.‏ [↑](#footnote-ref-6)