Chest X-ray Classifier vs Clinical Diagnosis Performance Version 1

This summary compares the performance of an AI model trained on chest X-rays with published clinical benchmarks. It highlights areas where the AI aligns with or falls short of typical radiologist-level sensitivity.

|  |  |  |  |
| --- | --- | --- | --- |
| Condition | Model Performance | Clinical Benchmark | Remarks |
| NORMAL | 0.991 | 95-99% | Excellent (matches expert-level review) |
| BACTERIAL PNEUMONIA | 0.866 | 80-90% | Strong, close to clinical diagnosis |
| COVID 19 | 0.855 | 70-90% | Comparable to early COVID imaging studies |
| PNEUMOTHORAX | 0.718 | 85-95% | Needs improvement (misses some cases) |
| VIRAL PNEUMONIA | 0.676 | 70-85% | Slightly below range |
| LUNG NODULES | 0.669 | 65-85% | Acceptable, on the lower end |
| MASS | 0.633 | 60-80% | Within clinical variation |
| EMPHYSEMA | 0.506 | 50-75% | Borderline (often underdiagnosed in X-rays) |
| PLEURAL THICKENING | 0.372 | 40-70% | Significantly underperforming |

We are actively working to enhance this model's accuracy and reliability. Upcoming updates will include additional training data, better handling of non-chest X-ray inputs, and expanded support for edge-case diagnoses. Our goal is to build a more robust and clinically aligned system through iterative validation and model refinement.

## Sources & References

Bairwa, H., & `Jangid, R. (2024). *Pneumonia Detection from Chest X-Rays Using the Chexnet Deep Learning Algorithm*. https://doi.org/10.20944/preprints202407.0104.v1

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). Chestx-Ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3462–3471. https://doi.org/10.1109/cvpr.2017.369

Johnson, A.E.W., Pollard, T.J., Berkowitz, S.J. et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Sci Data 6, 317 (2019). https://doi.org/10.1038/s41597-019-0322-0