

Performance Benchmarks: Performance Evaluation

Comparing multimodal models vs. single-modal benchmarks and evaluation metrics (accuracy, AUC, F1)

Accuracy

Overall accuracy
Ratio of correct predictions

AUC-ROC

Classification performance
Sensitivity-specificity balance

F1-Score

Precision-recall
Harmonic mean

Sensitivity

Sensitivity
True Positive Rate

Specificity

Specificity
True Negative Rate

Dice/IoU

Segmentation accuracy
Region overlap measurement

Pneumonia Diagnosis Performance Comparison Example

Model	Accuracy	AUC	F1
X-ray Only	84.2%	0.87	0.82
Clinical Only	79.5%	0.83	0.78
X-ray + Clinical (Multimodal)	91.3%	0.94	0.90