

Introduction

Radiologists may miss up to 30% of the lung nodules on Chest X-ray (CXR). The purpose of this study was to compare the efficacy of Gray Scale Inversion (GSI) to improve the detection of pulmonary nodules on CXRs between radiologists and Artificial Intelligence (AI).

Methods

This study was approved by the Institutional Review Board (IRB) of Thomas Jefferson University Hospital. Pulmonary nodules detected on CXR with both posteroanterior (PA) and lateral views from August 2016 to February 2019 were retrospectively collected from institution PACS. A positive dataset (n=60) consisting of CT-confirmed benign and malignant pulmonary nodules and a negative dataset (n=62) of normal 2-view CXR from the same time period were entered into this study. GSI was accomplished through the included PACS algorithm without additional postprocessing by the readers. For AI study, 20% of cases (n=24, 12 positive and 12 negative) were randomly selected for testing and the remaining 80% cases (n=98) were mirrored to augment the dataset to reach the minimum number of images required for training the model. Four separate models for grayscale (GS) PA, GS lateral, GSI PA, and GSI lateral with 196 images in each model were trained using the Google AutoML algorithm. Three experienced cardiothoracic radiologists read the GS and GSI images with a minimum of two weeks interval in between. The mean of two radiologists' performance in the testing set was compared to AI across all four views using a paired t-test.

Results

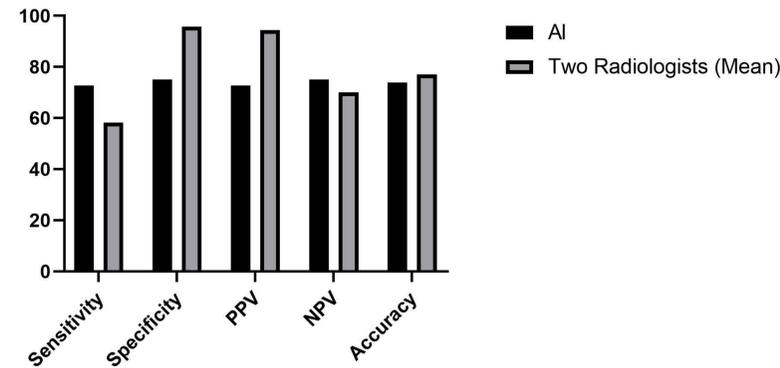
Among all four models, highest accuracy (78.2%) and Negative Predictive Value (76.9%) were achieved by the GSI lateral model. GS lateral model had the highest specificity (88.8%) and Positive Predictive Value (87.5%). GS frontal and GSI lateral model equally showed the highest sensitivity (72.7%). GSI did not significantly improve performance for the detection of pulmonary nodules in either human readers or the AI models. The paired t-test was not significantly different between performance of AI and the mean of two radiologists across all four models in the testing set.

AI Internal Validation Report

Model	Precision (PPV)	Recall (Sensitivity)
Grayscale PA	85.71%	85.71%
Grayscale Lateral	90.48%	90.48%
GSI PA	90.48%	90.48%
GSI Lateral	100%	95.24%

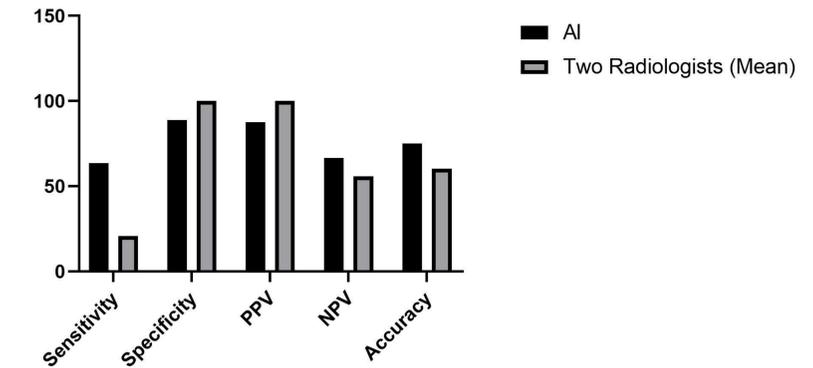
AI and Radiologist Performance in PA and Lateral views of GS and GSI

GS-PA



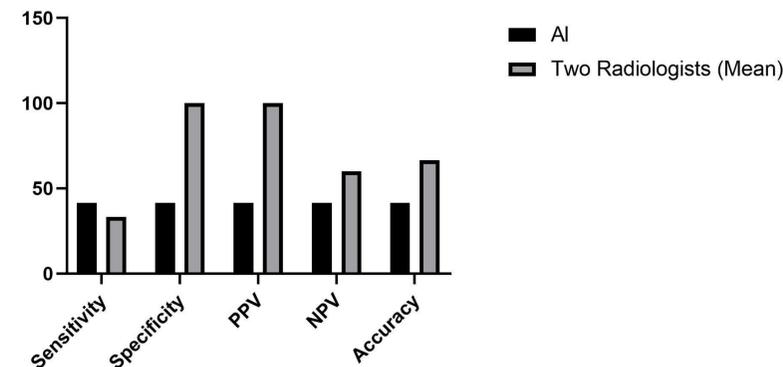
AI Versus human readers performance in GS PA view

GS-Lateral



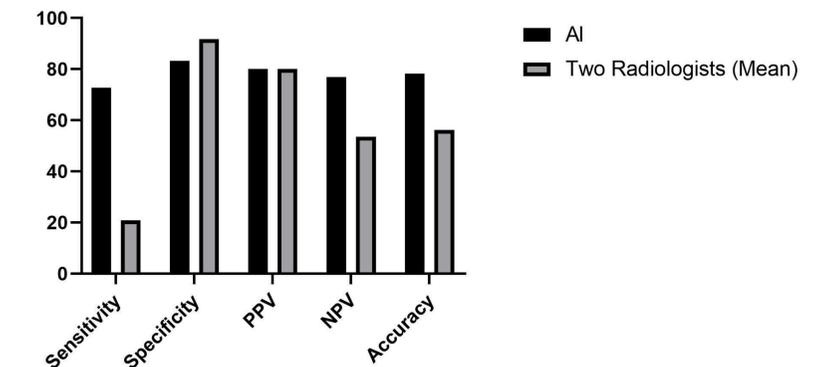
AI Versus human readers performance in GS Lateral view

GSI-PA



AI Versus human readers performance in GSI PA view

GSI-Lateral



AI Versus human readers performance in GSI Lateral view

Conclusion

Improving diagnostic performance on CXRs has the advantages of reducing CT follow-ups, and thus downstream costs and radiation exposure. Therefore, continued investigation in postprocessing techniques may improve performance among both radiologists and AI.