

Abstract

Objectives

To evaluate the efficacy of combining predictive artificial intelligence and image similarity model to risk stratify thyroid nodules, using a retrospective study.

Methods

Two datasets were used to determine the efficacy of the algorithm. One was the publicly available Stanford dataset consisting of ultrasound images of 192 nodules between April 2017 to May 2018 and the second one was from a private practice setting consisting of 118 thyroid nodule images from 2018-2023. All the nodules had definitive diagnosis either by biopsy or by surgery. The software was used to predict the diagnosis and TI-RADS score.

Results

In the Stanford dataset, the AI algorithm predicted malignancies with a sensitivity of 1.0 and a specificity of 0.55. The PPV was 0.18 and the NPV was 1.0. The AUCROC was 0.78. The AI algorithm did not miss any cases of cancer. TI-RADS based clinical recommendation had a polychoric correlation of 0.67. In the private dataset, the AI algorithm predicted malignancies with a sensitivity of 0.91 and a specificity of 0.95. The PPV was 0.8 and NPV was 0.98. AUCROC was 0.93 and accuracy was 0.94. TI-RADS based clinical recommendation had a polychoric correlation of 0.94 for this dataset.

Conclusion

The AI model demonstrated high negative predictive value with a potential for 60% reduction in the need for biopsy. This could reduce the burden on patients and healthcare costs.