Risk Stratifying Thyroid Nodules with Deep Learning

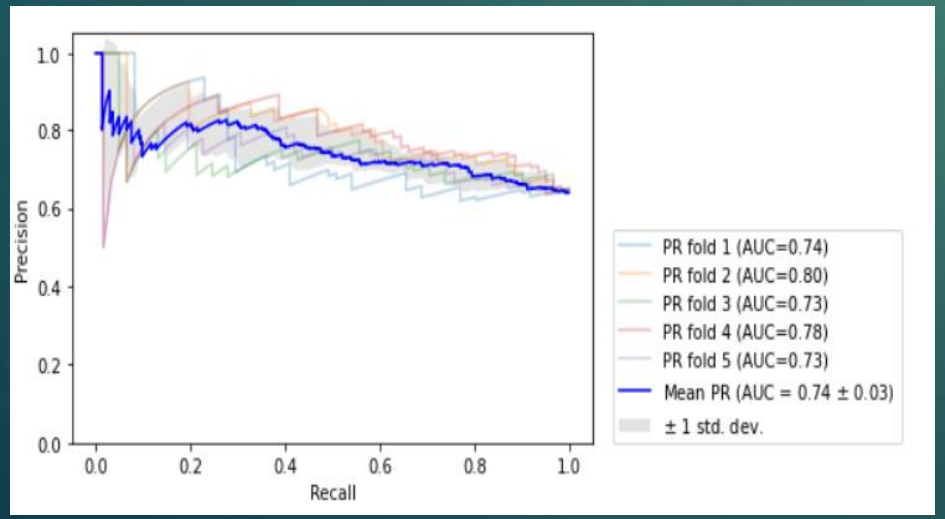
Author: George Zhou, BS

Purpose:

The ACR Thyroid Imaging Reporting & Data System (TI-RADS) was created to help risk stratify thyroid nodules seen on ultrasound for chance of malignancy to help guide downstream management. While introducing a level of standardization, some of the main drawbacks that persist are the inter-observer discrepancies and the dependence on the operator skill level. To address these limitations, this study aims to develop a deep learning model to help risk stratify thyroid nodules.

Results

On binary classification of thyroid ultrasounds into low- or high-risk groups, the ViT and the CNN was able to achieve an area under the precision-recall curve (PR AUC) of 0.74 ± 0.03 and 0.47 ± 0.06 respectively. The PR curve from 5-fold cross-validation is shown below.



Methods/Materials

The dataset includes 480 thyroid ultrasound images (DOI: 10.1117/12.2073532). Each ultrasound image was assigned a TI-RADS score by a radiologist. The dataset was divided into two groups: low-risk (TI-RADS score of 1 to 4a) and high-risk (TI-RADS score of 4b to 5) . The final dataset included 192 high-risk and 233 low-risk thyroid ultrasounds. Two deep learning models were studied: a convolutional neural network (CNN) and a vision transformer (ViT). The models were trained using the Adam optimizer over the binary cross-entropy loss function. The models were evaluated using 5-fold cross validation.

Conclusions

The superior performance of the ViT over the CNN, which previously defined state of the art performance in various computer vision tasks, can be attributed to the ViT's self-attention mechanism. Self-attention allows the ViT to learn long range dependencies and aggregate global information in early layers.

Overall, our preliminary results show that deep learning has the potential to risk-stratify thyroid nodules seen on ultrasound in accordance to the ACR TI-RADS grading system. With advances in deep learning, it may be worthwhile to further study if augmenting the TI-RADS grading system with predictions from deep learning models can improve risk-stratification.