Deep Learning for Thyroid Nodule Detection and Classification
Multi-Stage Automated Framework

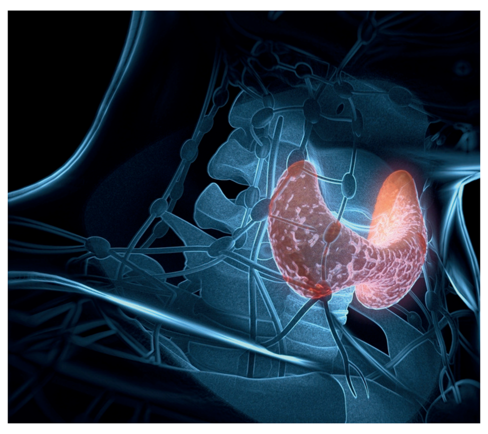
Zhou Zihan

Supervisor: Prof. Kenneth K.Y. Wong

GROUP
24022

Introduction

- Thyroid cancer is the most common endocrine malignancy
- Affects 15 per 100,000 individuals annually in the US.
- Ranks as the 7th most common cancer in women.
- Ultrasound is widely used in diagnosis.



Problem Statement

Current diagnosis procesure still have many problems

Time-Consuming and Mentally Demanding
Radiologists must evaluate a large number of images daily, which requires sustained focus and effort.
Subjectivity and variability
The manual interpretation of ultrasound images is subjective, often causing variability and uncertainty in diagnoses.
Lack of Automation
Current methods lack reliable, automated tools to improve diagnostic accuracy and reduce inefficiencies in clinical workflows.

TI-RADS

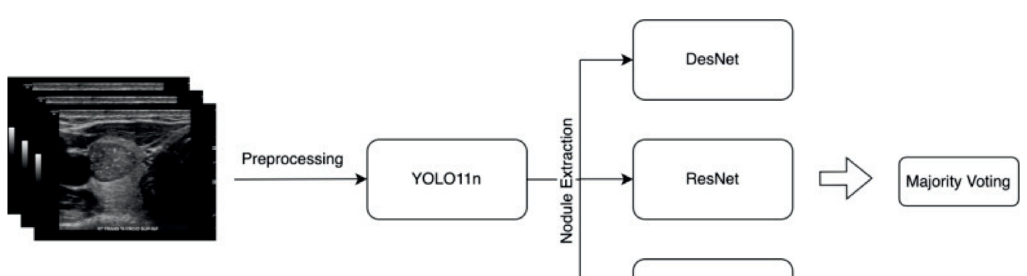
Thyroid Imaging Reporting and Data System (TI-RADS)
A classification method that categorizes thyroid nodules into risk levels based on features such as size, echogenicity, margin, shape, and echogenic foci. It guides clinicians in deciding whether a biopsy or other further evaluation is needed.

Score	ACR TI-RADS Category	Malignancy Rate (%)
0	TR I	Benign
2	TR 2	Not suspicious
3	TR 3	Mildly suspicious
4-6	TR 4	Moderately suspicious
>7	TR 5	Highly suspicious

Objective

This project aims to develop a deep learning-based system that:

1. Locates thyroid nodules using object detection techniques.
2. Classifies the malignant risk of nodules based on features extracted from the nodule region.
3. Assigns a standardized TI-RADS level, where a higher level indicates a greater risk of malignancy.



Dataset

All the training process is based on **Stanford AIMI Thyroid Ultrasound Cine-clip Dataset**:
192 Nodule Cases
Each Nodule represented by 100 Consecutive Ultrasound Frames
Forming a total of 18,000 Ultrasound Frames dataset

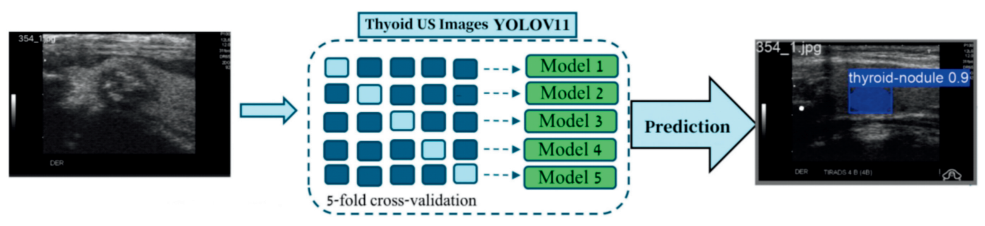
Characteristic	Stanford AIMI Dataset	
	Benign	Malignant
Age (y)	56.8 ± 15.2	48.3 ± 14.1
Sex - Female	144 (82.3%)	15 (88.2%)
Sex - Male	31 (17.7%)	2 (11.8%)
TI-RADS Level		
1	1 (0.6%)	0 (0.0%)
2	10 (5.7%)	0 (0.0%)
3	52 (29.7%)	0 (0.0%)
4	78 (44.6%)	5 (29.4%)
5	34 (19.4%)	12 (70.6%)
Total	175	17

Methodology

To design a two-stage pipeline that accurately detects thyroid nodules and classifies their cancer risk levels.

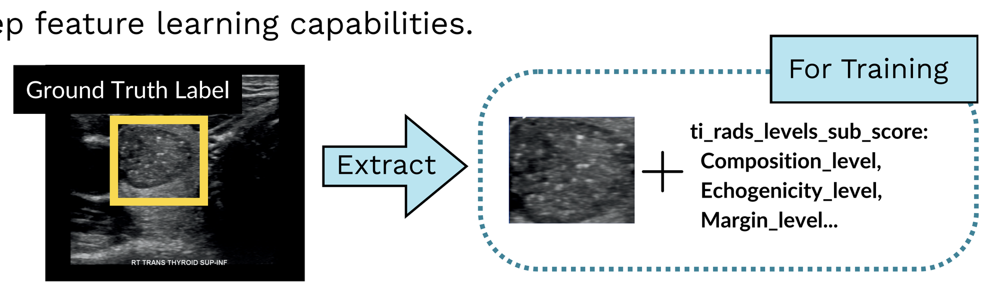
Data Preprocessing
Data Augmentation: Applied horizontal flipping to expand the dataset while preserving diagnostic features.
Cross-Validation: Used 5-fold cross-validation to ensure robust and reliable results.
ROI Extraction: For classification model training, Focused on thyroid nodule regions to eliminate irrelevant background.
Resizing: Standardized image dimensions to match model input requirements.

Object Detection Training
The lastest version of You Only Look Once (YOLO) model YOLO11 was selected for thyroid nodule detection due to its proven track record in medical imaging applications.



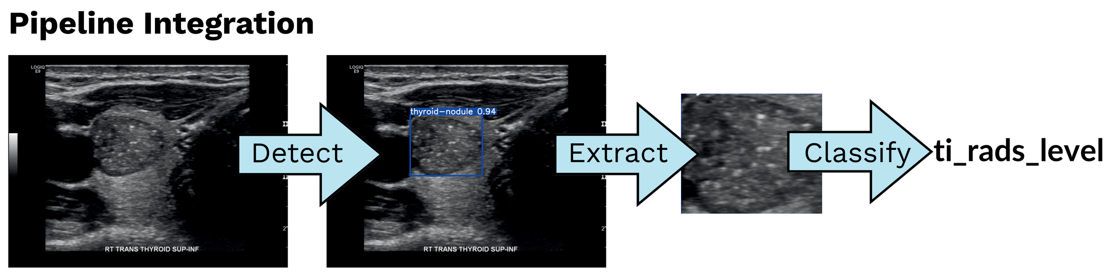
During training, the model processed thyroid ultrasound images from the Stanford AIMI Dataset using YOLO-formatted annotations (normalized center_x, center_y, width, height) and systematically evaluated key hyperparameters such as learning rate, batch size, and scheduler.

Classification Model Training
Several convolutional neural network architectures are evaluated, including VGG16, VGG19, and ResNet50, ResNet101, leveraging their deep feature learning capabilities.



The model takes ground truth thyroid nodules as input and predicts sub-scores for features. These sub-scores are summed to derive the final TI-RADS level.

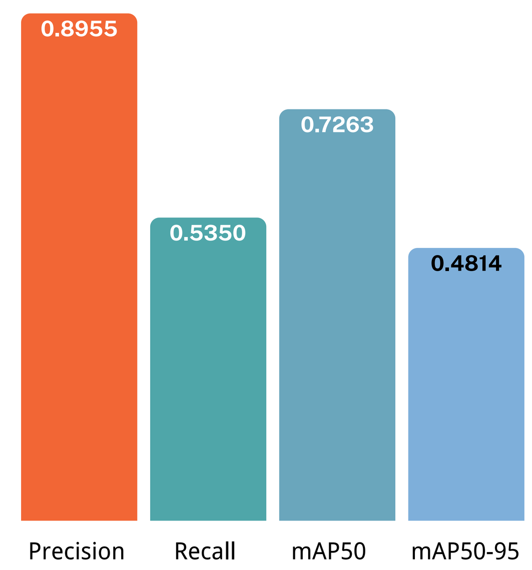
Pipeline Integration



The system processes ultrasound examination data, either as a video or a sequence of consecutive images. A YOLO-based detection model identifies thyroid nodules in each frame, generating bounding boxes that define regions of interest (ROIs). These ROIs are extracted and passed to a classification model for analysis.

Results

Detection Model Performance (YOLO11) on the reserved test set

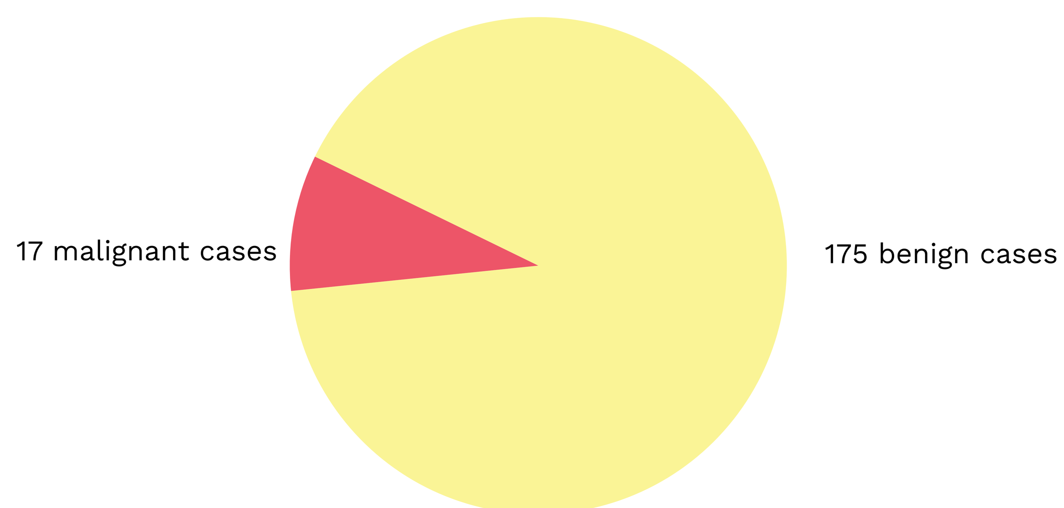


A histogram for the YOLO11 model on the Stanford AIMI dataset shows high precision (~0.90), moderate recall (~0.54), strong mAP50 (~0.73), and mAP50-95 (~0.48), with fitness averaging 0.51, indicating balanced performance.

The model is still in the development stage, and is expected to take the ground truth nodule regions as input and output their corresponding TI-RADS levels, providing an automated risk assessment for each identified nodule. Based on preliminary testing with VGG16, VGG19, and ResNet50 architectures on a small subset of the data, the classification model is expected to achieve accuracy rates between 65-75% for TI-RADS level prediction.

Challanges

Class Imbalance
75 benign vs. 17 malignant cases may bias the model.
Solution: Use loss function adjustment and plan to collect more data.
Preservation of Diagnostic Features
Preserving diagnostic features limits augmentation techniques.
Solution: Use simple augmentations like horizontal flipping to expand data.
Generalization to Real-World Data
Ensuring robust performance on real-world clinical data.
Solution: Fine-tune the model with diverse datasets.
Time Limit
Limited project timeline for model training and evaluation.
Solution: Prioritize critical tasks and adopt iterative improvements.



Conclusion

This study designed an automated thyroid nodule analysis pipeline that integrated thyroid nodule detection and cancer risk assessment classification together to assist doctors in making informed decisions. The latest investigation used the Stanford AIMI dataset, and implemented image preprocessing, data augmentation, YOLO-based detection model development with five-fold cross-validation for evaluation, and trained a basic framework for the cancer risk classification model.

Future research could prioritize data collection and distribution balancing. A comprehensive data collection strategy across multiple institutions might help establish a more diverse and representative dataset.