Using Artificial Intelligence To Guide Proper Probe Positioning on Ultrasound Videos of the Abdominal Wall

Brian Yuen[1], L. Colby Mangum[2], Laura Brattain[1]

[1]College of Medicine, University of Central Florida;
[2]READY Lab, Institute of Exercise Physiology & Rehabilitation Science, University of Central Florida

UCF READY LAB
REhabilitation, Athletic assessment, & DYnamic imaging

UCF College of Medicine
UNIVERSITY OF CENTRAL FLORIDA

Motivation

- Low back pain (LBP) is one of the move prevalent musculoskeletal (MSK)-related diagnoses in military population.
- Core muscles are typically measured with MRI and fine-wire EMG. These are expensive and time-consuming. MSK ultrasound (US) imaging provides a reliable and cost-effective alternative, but is heavily operator dependent.
- AI offers a way to standardize quality by guiding positioning.
- EfficientNet is well established for accurately segmenting and identifying medical images across varied contexts.

Hypothesis: *Deep learning can accurately classify ultrasound images which are malpositioned, and therefore can potentially be used to help guide proper probe placement.*

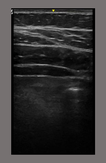
Goal: *Evaluate the efficacy of EfficientVit for the classification of ultrasound videoframes based on probe positioning.*

Materials

Dataset

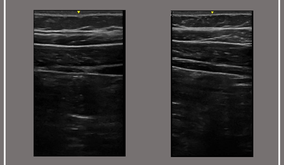
10 ultrasound videos were taken from 13 subjects. The videos captured common positioning mistakes:

- **Panning** too far medially or laterally
- **Tilting** up, down, left, right
- **Rotating** counterclockwise or clockwise



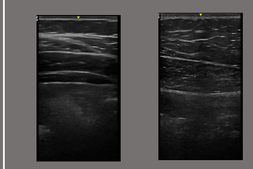
Ideal Position

Rotation



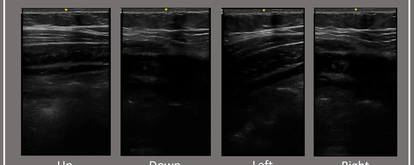
Counterclockwise | Clockwise

Pan



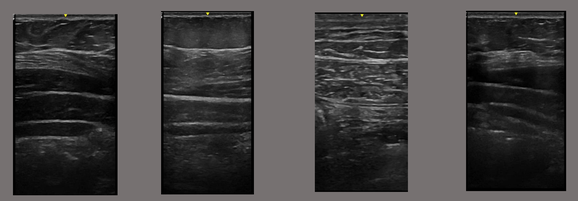
Lateral | Medial

Tilt



Up | Down | Left | Right

Frame-by-Frame Labeling



Pan: “Ideal”
Rotation: “Ideal”
Tilt: "Ideal"

Pan: “Ideal”
Rotation: “Clockwise"
Tilt: "Ideal"

Pan: "Medial“
Rotation: "Ideal"
Tilt: "Ideal"

Pan: “Ideal“
Rotation: "Ideal"
Tilt: “Up"

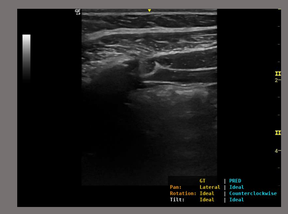
Methods

1. Segmentation + Classification

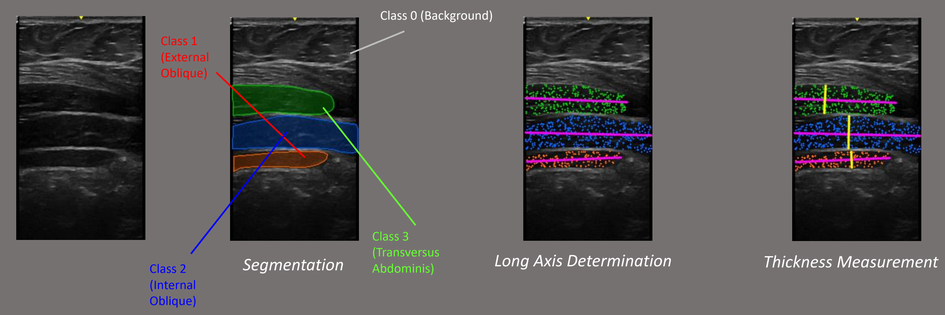
- **EfficientVit:** hybrid of convolution and Vision Transformer for fast, low-latency inference.
  - ImageNet Backbone
  - Batch Size: 32
  - Learning Rate: 3e-4 (default)
  - Decay: 1e-4 (default)

2. Training and Prediction with EfficientVit

- *EfficientVit was trained on a patient-stratified 80/10/10 split.*
- *Predictions for all three categories were made.*
- *Agreeance was mapped on a bar graph*



3. Measuring Accuracy and Width



Generation of Muscle thickness:

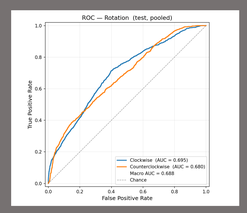
- Long Axis calculated as the regression line through the abdominal muscle points. The thickness was then determined by finding the tallest orthogonal line near the center of the muscle. Conversion ratios were taken from the source files.
- DICE Scores were also calculated as a measure of accuracy and specificity.

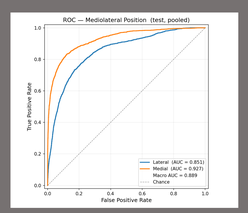
Results

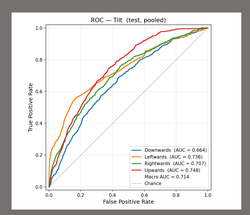
Performance Metrics

Probe Position	Accuracy	F1	AUC
Mediolateral Position (Overall)	0.94	0.83	0.88
Lateral	0.96	0.80	0.85
Medial	0.98	0.91	0.93
Rotation (Overall)	0.93	0.75	0.69
Clockwise	0.96	0.65	0.70
Counterclockwise	0.94	0.96	0.68
Tilt (Overall)	0.91	0.71	0.71
Downwards	0.97	0.59	0.66
Leftwards	0.98	0.68	0.73
Rightwards	0.98	0.66	0.70
Upwards	0.97	0.68	0.75

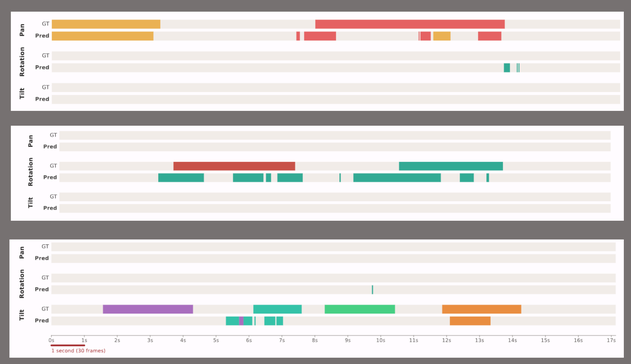
Fig 1. Performance metrics for EfficientVit Classification depicted through Accuracy, F1, and AUC. Overall the model demonstrated high accuracy across all position classifications. F1 Score was comparatively lower.







Classification Distribution across Three Videos



Yellow = Medial
Light Red = Lateral
Dark Red = Clockwise
Teal = Counterclockwise
Purple = Up
Dark Green = Down
Light Green = Right
Orange = Left
Grey = Ideal

Fig 2. The above diagram is a visual representation of the EfficientVit predictions for each frame for three different videos. The Odd-numbered rows are the ground truth labels, and the even-numbered rows are the predictions. Overlapping colors represents agreeance.

- *Due to a large portion of Ideal frames, the accuracy of the model remained high despite the comparatively lower number of correct frame predictions.*

Discussion

- *EfficientVit demonstrated a high level of accuracy across the 130 videos. However, in both Rotation and Tilt, the F1 scores comparatively underperformed. This is likely due to the prevalence of Ideally labeled frames compared to mistaken ones.*
- *To improve classification of frames depicting improper position, possible methods to explore in the future to improve detection involve increasing the proportion of these frames, either by artificially making them appear more often in the training data or by recording improper frames for longer.*
- *The method of frame level classifications is nevertheless unique and can be applied to many situations.*

Summary

EfficientVit demonstrated that it is capable of depicting accurate results, but more work needs to be done in order to determine the effect of using comparatively larger number of ideal frames compared to those depicting ultrasound mispositioning.

References

Xie, Y., & Liao, Y. (2023, April 7). *Efficient-VIT: A light-weight classification model based on CNN and VIT | Proceedings of the 2023 6th International Conference on Image and Graphics Processing*. Association for Computing Machinery, Digitla Library. https://dl.acm.org/doi/abs/10.1145/3582649.3582676

Zhou, T., Porikli, F., Crandall, D. J., Van Gool, L., & Wang, W. (2023). A survey on Deep learning technique for video segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(6), 7099–7122. https://doi.org/10.1109/tpami.2022.3225573