Comparative Detection of Carpal Tunnel Syndrome in Muscle Ultrasound Images leveraging Heterogeneous Datasets

Jungsub Sim, Seunghyun Kim, Sungche Lee, Joonshik Yoon, Seungjun Baek
wishmaster@korea.ac.kr

Department of Computer Science, Korea University, Seoul, Republic of Korea
Department of Physical Medicine & Rehabilitation, Korea University, Seoul, Republic of Korea

Background

- Carpal Tunnel Syndrome (CTS) is a disease caused by compression of the median nerve through the carpal tunnel in the wrist.

- Nerve Conduction study/Electromyography (NCS/EMG) is the gold standard for diagnosing CTS and monitoring a patient's condition, but as EMG requires the insertion of a needle, which can injure patient's muscle, Ultrasonography (US) is emerging as an alternative diagnostic tool.

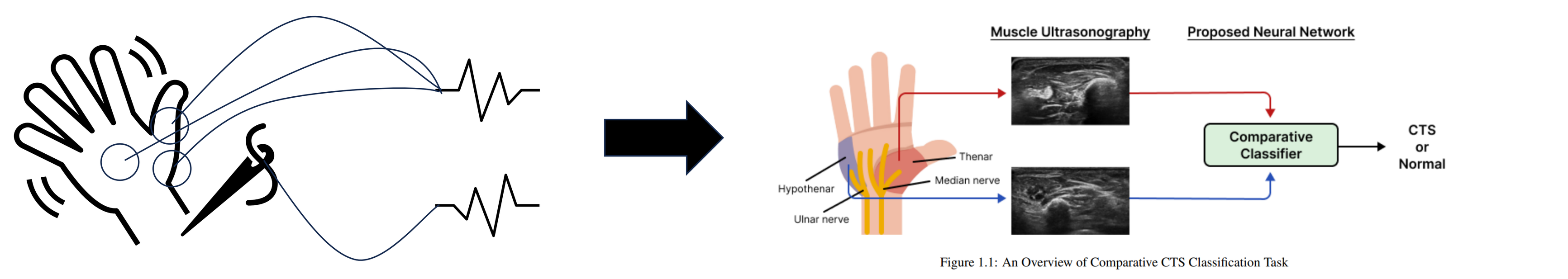
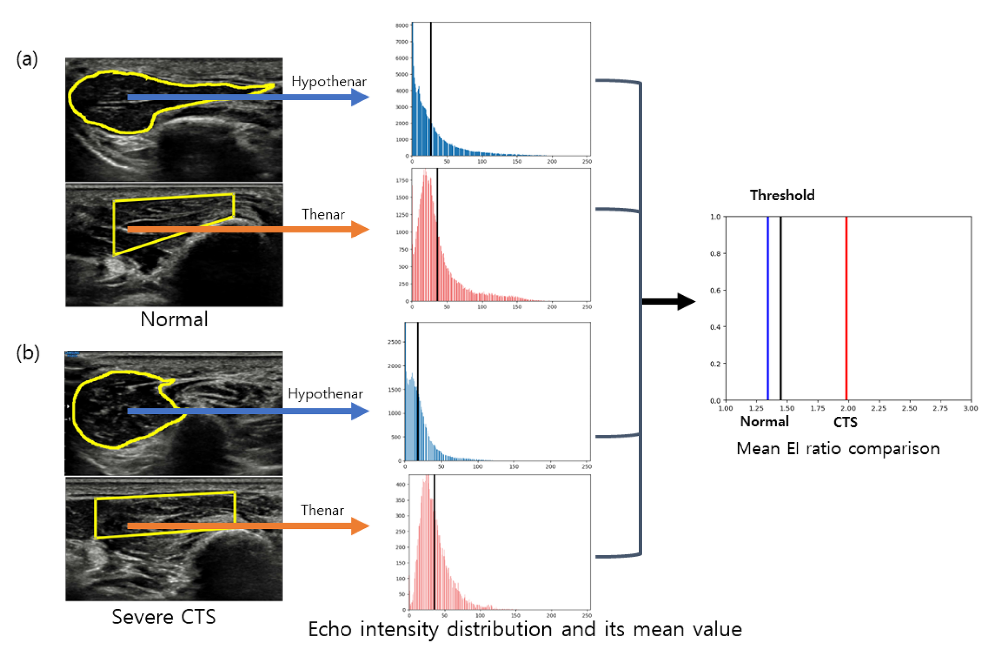


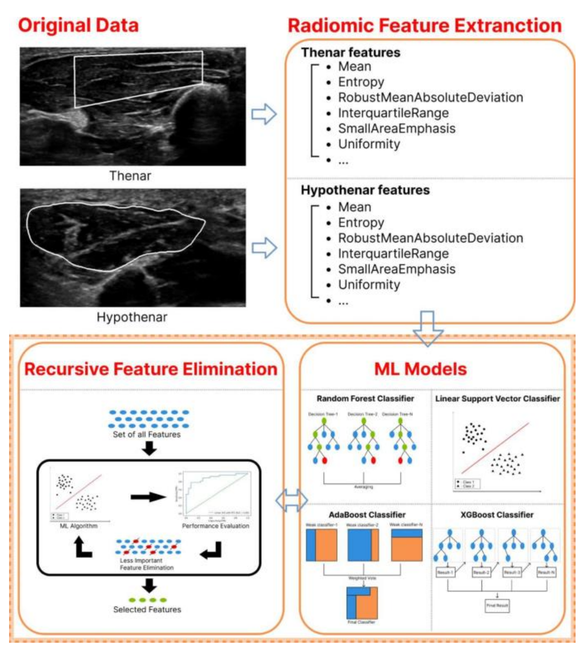
Figure 1.1: An Overview of Comparative CTS Classification Task

- When CTS occurs, the median nerve is denervated and the muscle fibers of the thenar muscle, which is connected to the median nerve, are replaced by fibrous tissue, while the hypothenar muscle which is connected to the ulnar nerve, remains unaffected.

- Changes in muscle tissues are reflected in US images and two methods were formerly employed to diagnose CTS on US images:

1. Comparing the echo intensity (EI) ratio of the thenar muscle and the hypothenar muscle.
2. Utilizing various machine learning classifiers to diagnose CTS based on features extracted from US images using an open source medical feature extraction library.





Introduction

- Unlike EI ratio-based methods and machine learning models, deep learning models do not require the specification of a seperate region of interest (ROI) or features to be extracted for image classification, which is convenient.

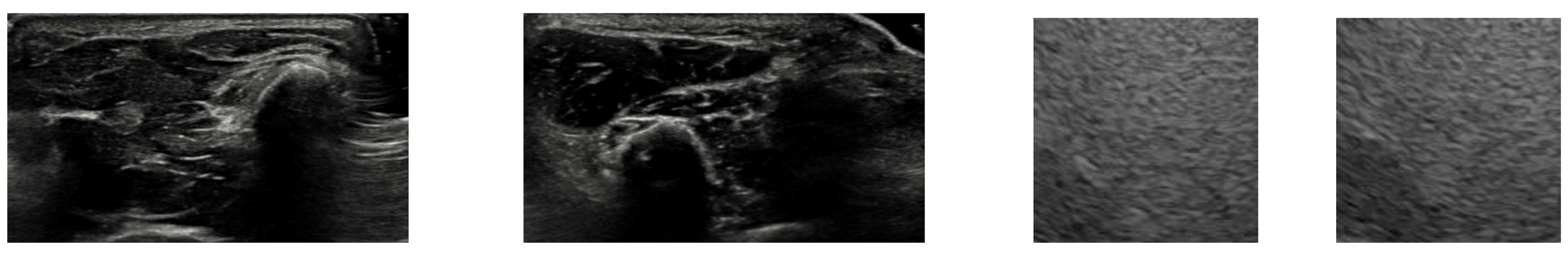
- As deep learning models extract features that are deemed important through learning, they can discover other significant features that may have been overlooked by the practitioner, and are more adaptive to the task compared to machine learning methods that use predefined features for analysis.

- However, for the advantages of deep learning models to be realized, a significant amount of data is required for training, which is challenging for medical data.

-To overcome this challenge, we attempted to train deep learning models with combined datasets of different domains: the CTS dataset and the Fatty Liver dataset. Initially, our goal was to merge the two datasets to obtain a sufficient amount of data and create a general medical diagnosis model that is not limited to specific domain. However, through various experiments, we found that training both datasets together in a single model yielded higher diagnostic performance than training on the CTS dataset alone.

Materials

- Each CTS data consists of a pair of thenar and hypothenar area images. Similarly, the Fatty Liver data consists of a pair of liver and kidney area images.



Thenar **Hypothenar** **Liver** **Kidney**

- Thenar area in CTS dataset and Liver area in Fatty Liver dataset is a lesion area affected by occurrence of specific symptom, and used as diagnosing symptoms. On the other hand, hypothenar area and kidney area is used to compensate individual patient's features because it's not affected by symptoms.

Severity	#hands	#data
Normal	70	560
Mild	21	129
Moderate	34	187
Severe	27	171
Total	152	1047

Table 2. Number of hands and data by severity

Status	#patients	#data
Normal	17	229
Abnormal	38	378
Total	55	537

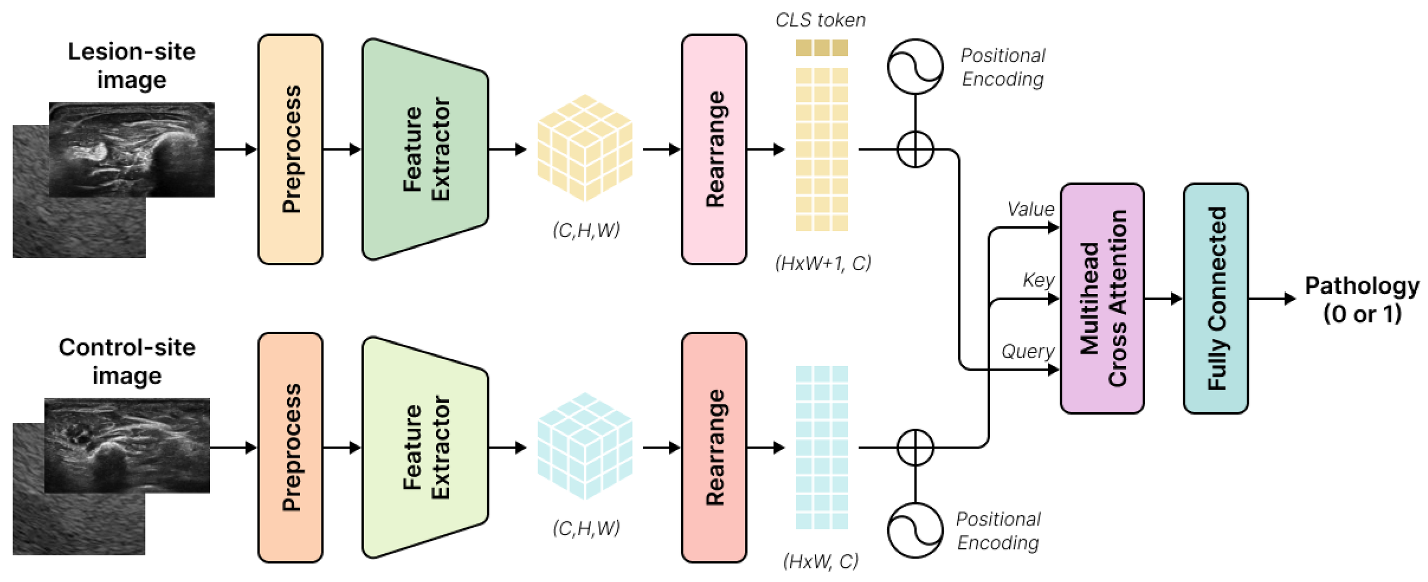
Table 1. Number of patients and data in Fatty Liver dataset

- The CTS dataset consists of 1047 data from 152 hands and Fatty Liver dataset consists of 537 data from 55 patients

Method

- The proposed model compares the images of lesion-site (thenar area) and control-site (hypothenar area) using deep neural networks

- It consists of two preprocessors, two feature extractors and a fusion module using multi-head attention. One may choose various models for feature extractor, e.g., ViT, Swin Transformer or ResNet



- The model is jointly trained using heterogeneous datasets: CTS and Fatty Liver datasets. In the inference phase, the Fatty Liver dataset is not employed and the diagnostic performance is evaluated on the CTS dataset.

Result & Discussion

- Overall, deep learning (DL) models outperformed machine learning (ML) baselines. DL models without comparative classification consist of a feature extractor and a linear classifier, among which the Swin transformer (SwintT) gave the best accuracy.

- Finally, the comparative network with the ResNet feature extractor trained with heterogeneous datasets (CTS + Fatty Liver) achieved the highest accuracy.

Method	Accuracy
ML (AdaBoost)	0.710 ± 0.002
ML (Random Forest)	0.719 ± 0.002
ML (XGBoost)	0.724 ± 0.001
DL (SwinT)	0.737 ± 0.026
Proposed (Comparative, SwinT)	0.749 ± 0.023
Proposed (Comparative, ResNet, 2 datasets)	$\mathbf{0.771 \pm 0.017}$

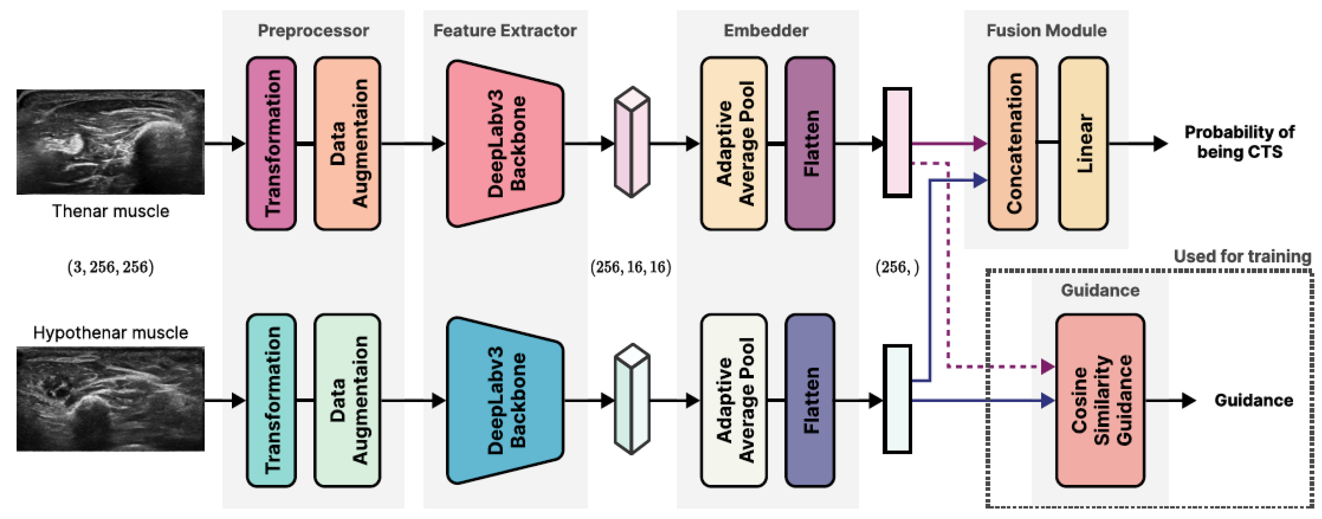
Table 1. Detection Performance

- The result is attributable to the synergetic effect of learning common features of ultrasound images, which outweighs the possible confusion caused by the domain shift across heterogeneous datasets.

Conclusion & Future works

- In this work we found that a comparative framework to detect CTS from US images led to more accurate classification, and further improvement was achieved by leveraging a fatty liver dataset.

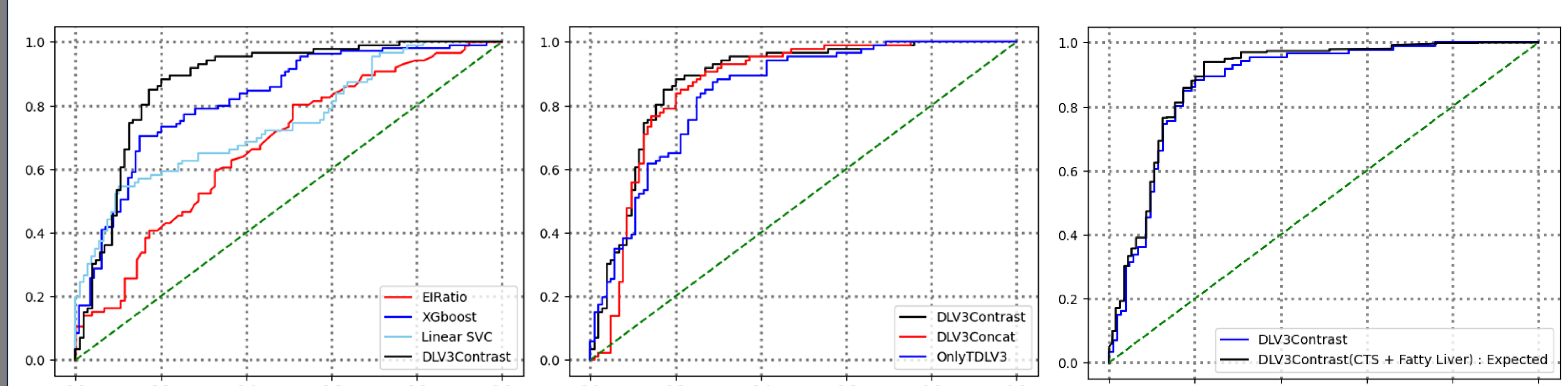
- Since the discovery of inherent features from medical data is challenging, in the present work, we implemented a comparative guidance based on the cosine similarity between two feature vectors from lesion area and control area.



$$G(\mathbf{x}_{t}, \mathbf{x}_{ht}) = \begin{cases} 1 - \cos(\mathbf{x}_{t}, \mathbf{x}_{ht}), & \text{for normal people,} \\ \max(\mathbf{0}, \cos(\mathbf{x}_{t}, \mathbf{x}_{ht}) - \alpha), & \text{for abnormal patients.} \end{cases}$$

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda G, \quad \text{where } \lambda > 0$$

- Based on the notion that the echo intensity difference between lesion area and control area is small in normal people and large in patients, it penalizes if the feature vectors are not similar in normal group and similar in patients.



- By combining this comparative guidance module with Deeplabv3 based feature extractor, we implemented our DLV3Contrast model and this model outperformed all the other baseline models on the CTS dataset.

- In future work, we can jointly train the CTS dataset and the Fatty Liver dataset together on our DLV3Contrast model and see if the Fatty Liver dataset as a catalyst also works well for Comparative guidance.