A New Semi-Supervised Learning Benchmark for Classifying Views and Diagnosing Aortic Stenosis from Echocardiograms

Zhe Huang[*], Gary Long[1], Benjamin Wessler[2], and Michael C. Hughes[*]

[*] Dept. of Computer Science, Tufts Univ.
[1] CVAI Solutions
[2] Division of Cardiology, Tufts Med. Center

Tufts Medical Echocardiogram Dataset (TMED): **https://TMED.cs.tufts.edu**

Clinical Motivation

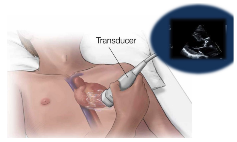
Aortic stenosis (AS) is a common cardiac valve condition, best detected using **echocardiograms** (ultrasound images of the heart).

Many patients are missed by current practice
- Up to 66% of symptomatic AS patients may not be referred for care

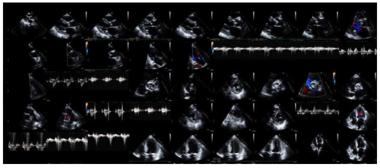
Improved early detection of AS sorely needed
- If undetected, severe AS is often fatal (higher mortality than some metastatic cancers)
- With timely detection, severe AS is treatable with low mortality

Automating preliminary screening of echocardiograms for AS *via machine learning* may improve detection (and thus improve outcomes).

Background: Echocardiogram Workflow



Handheld transducer is used to capture different views of the heart's anatomy. There are dozens of standard view types.



One study yields ~100 images of diverse view and quality. Images are not labeled with view type or diagnosis.

Challenge: Lack of labeled data

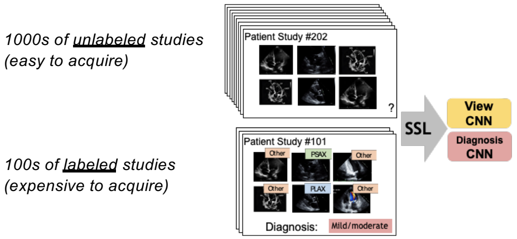
Most classifiers require **large training sets of labeled images** to be successful

Echocardiogram imagery is easy to collect from existing records
However, **labels are difficult and expensive to acquire**
- View and diagnostic labels not recorded when imagery is captured
- Require post-hoc annotation by clinical experts

Recent SSL methods show promise on standard vision tasks (e.g. CIFAR-10)
But use class-balanced data and artificially forget labels to make unlabeled set

Can SSL methods handle an uncurated unlabeled set of real medical images?

Solution: Semi-supervised Learning (SSL) 2



Number of Unlabeled Patients	Images	Method	View Task Balanced acc. on test set
0	0	Basic WRN	81.97
380	~41k	Pseudo-Label (Lee et al. '13)	84.23
380	~41k	VAT (Miyato et al. '18)	87.31
380	~41k	Augment-Only MixMatch	88.75
380	~41k	MixMatch (Berthelot et al. '19)	**91.11**

Takeaways:
- Modern SSL can use a large uncurated unlabeled set to boost performance over using only the modest-size labeled set.
- Among several methods, MixMatch is particularly effective.

Contributions

1. New open-access dataset: TMED
2. Analysis of recent SSL classifiers
   - What works on medical images?
3. Methods for coherent patient diagnosis from many images

Future Work

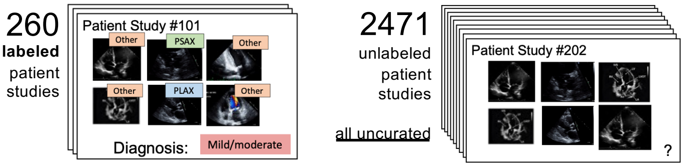
Data represents only 1 site. External validation needed.

Try multi-task SSL

Go beyond AS: more diagnoses and detailed measurements

Open-Access Dataset Release: TMED 1

Authentic benchmark for vision methods that learn from limited labeled data



Two classification tasks relevant to automatic diagnosis of aortic stenosis

1. Classify the view type of each *image*: PSAX / PLAX / Other

2. Diagnose AS severity of each *patient*: None / Mild/moderate / Severe

Existing public echo datasets (EchoNet or CAMUS) are great, but not suitable for AS diagnosis.

Challenge: Predict diagnosis from many images

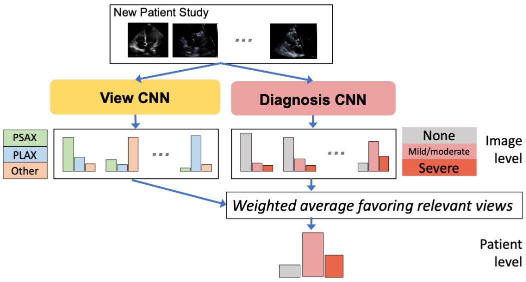
Most classifiers are designed to take in only one image and predict its class.

One echocardiogram study of one patient produces **~100 diverse images.**
- Most images show views that are irrelevant to the AS diagnosis task
- Only some view types are relevant (e.g. PLAX and PSAX show the aortic valve)
- Labels identifying which images are relevant are not available

Clinicians can take in many uncurated images, identify which ones are relevant views, and aggregate information from relevant images to make a diagnosis.

Can we automate diagnosis from many images?

Solution: Prioritize Relevant Views 3



Aggregation across images	Diagnosis Task Balanced acc. on test set
Simple average	81.77
Prioritize relevant view	**90.11**

Takeaways:
- Using view and diagnosis classifiers together can improve diagnosis.
- Manually curating relevant views is not necessary.