TEXAS A&M UNIVERSITY
SAN ANTONIO

Multi-Scale Self-Supervised Consistency Training for Trustworthy Medical Imaging Classification

Bonian Han[1], Cristian Moran[2], Gongbo Liang[2]

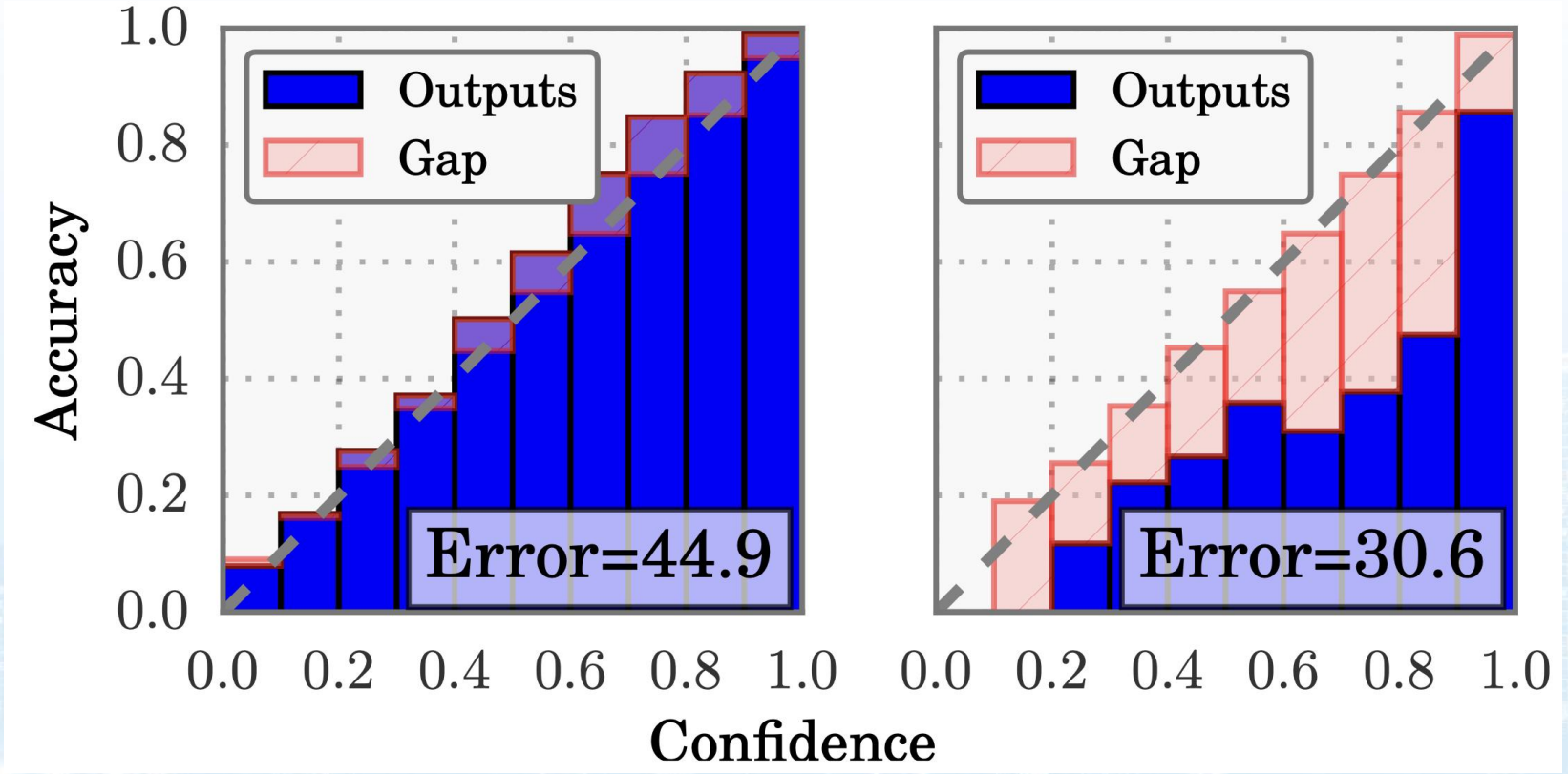
[1]Hangzhou Dianzi University, Hangzhou, China

[2] Texas A&M University-San Antonio, San Antonio, TX, USA

Neural network-powered CADs are exciting! But...

Can we trust them in medical practice?

A&M-SA



Modern Neural Networks are often poorly calibrated!

Guo, Chuan, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. "On calibration of modern neural networks." In International conference on machine learning, pp. 1321-1330. PMLR, 2017.

A&M-SA

Miscalibration = Estimate Uncertainty Wrong

- E.g.,
  - For any binary classification tasks, given 100 predictions with an average confidence of 0.95, we would expect around 95 correct predictions.

$$\mathbb{P}(\hat{\boldsymbol{y}} = \boldsymbol{y} | \hat{\boldsymbol{p}} = \boldsymbol{p}) \neq \boldsymbol{P}$$

In reality, a model with 0.95 confidence often has less accuracy than 95%.

A&M-SA

$$\mathbb{P}(\hat{y} = y | \hat{p} = p) \neq P$$

In reality, a model with 0.95 confidence often has less accuracy than 95%.

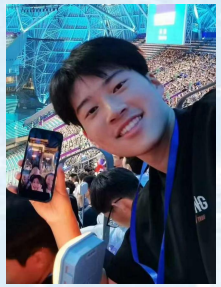
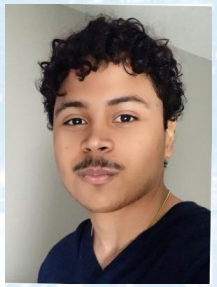
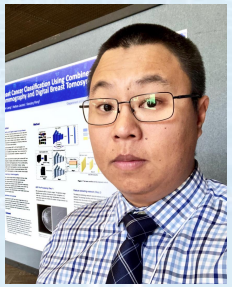
Miscalibrated models provide...

Misleading information that might lead to catastrophic consequence

A&M-SA

Multi-Scale Self-Supervised Consistency Training for Trustworthy Medical Imaging Classification

Bonian Han[1], Cristian Moran[2], Gongbo Liang[2]

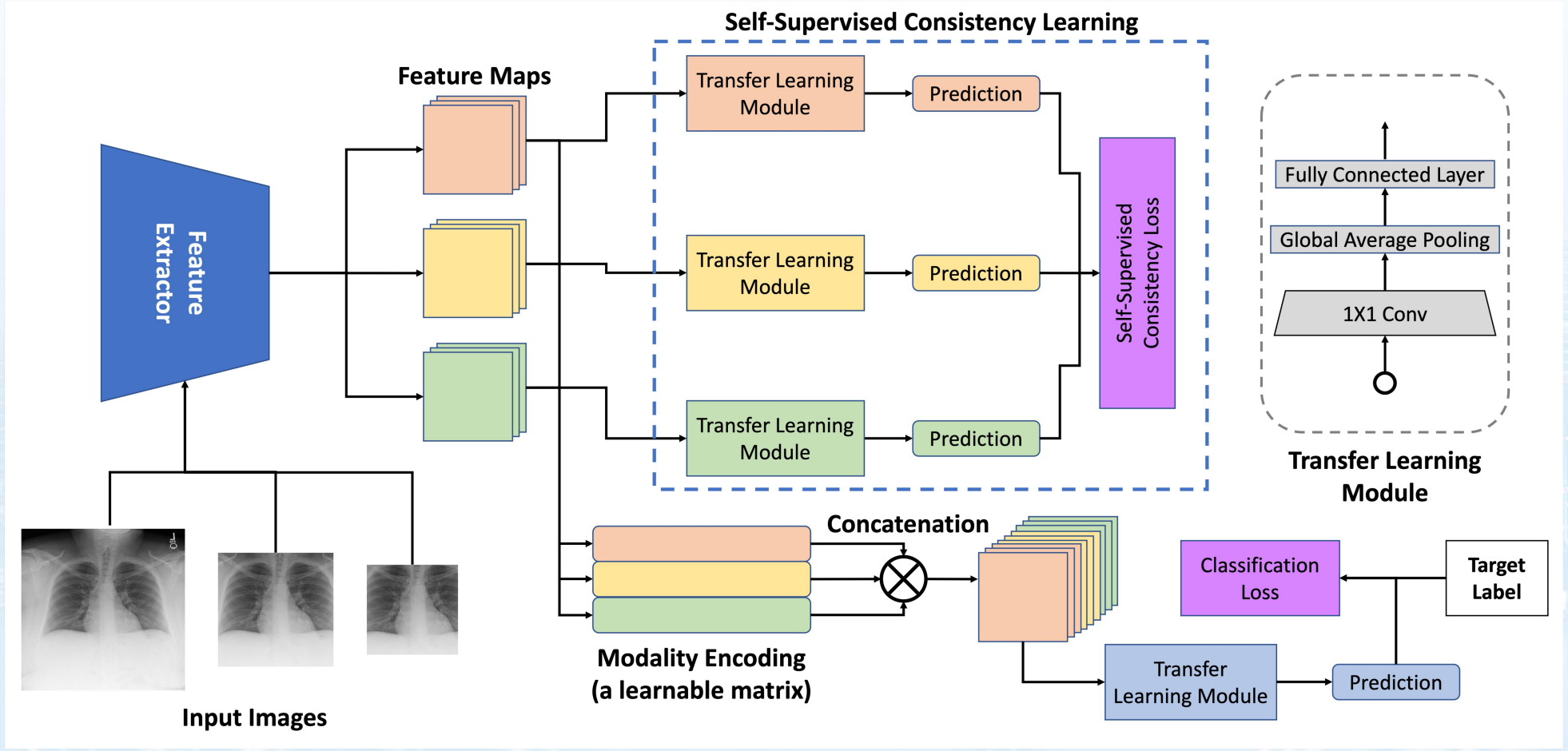




[1]Hangzhou Dianzi University, Hangzhou, China

[2] Texas A&M University-San Antonio, San Antonio, TX, USA

A&M-SA

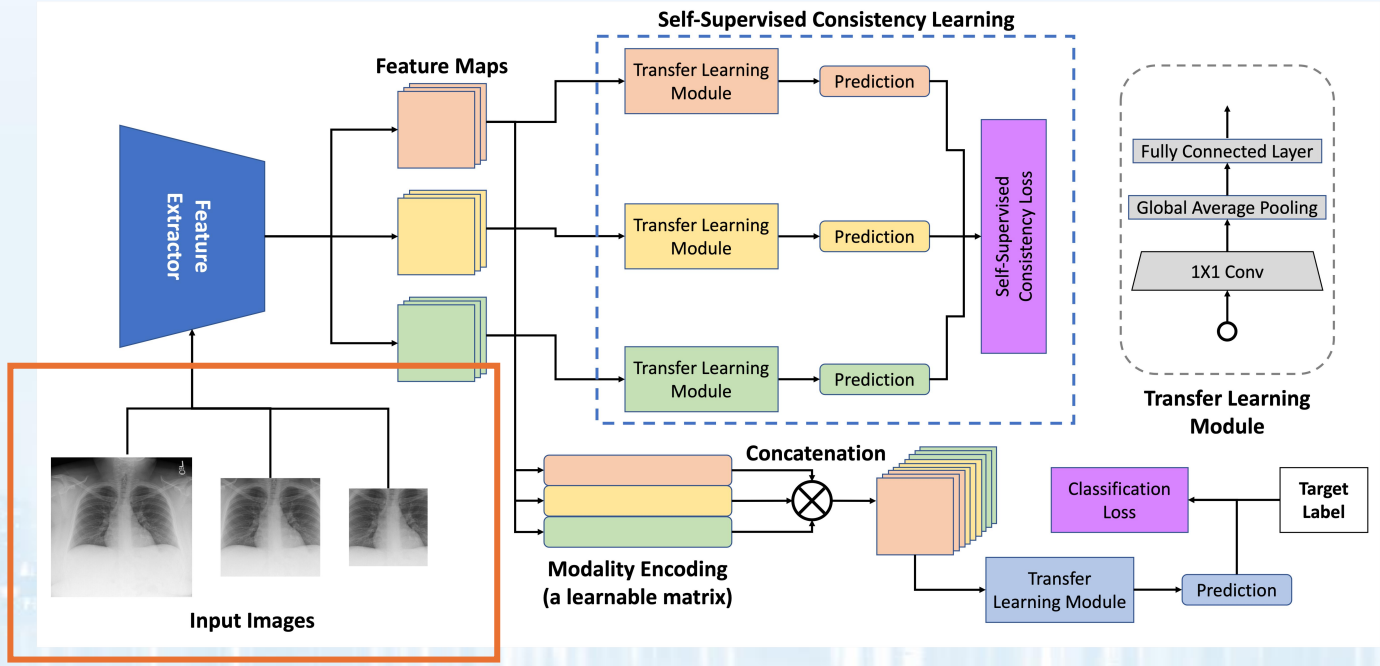
Proposed Architecture



A&M-SA

Multi-Scale Input Generation

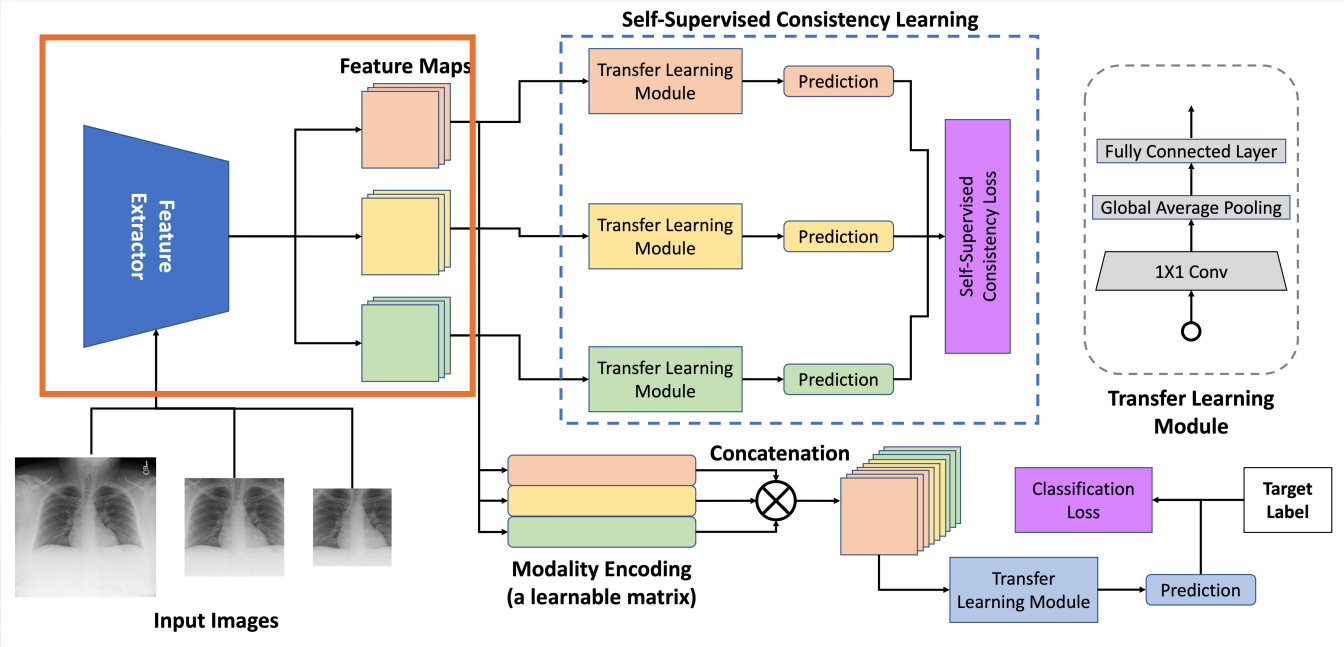
- Generates *k* sub-views of an input image by applying random crops.
- The *k+1* views, including the original one, are then fed into the shared feature extractor.



A&M-SA

Shared Feature Extractor

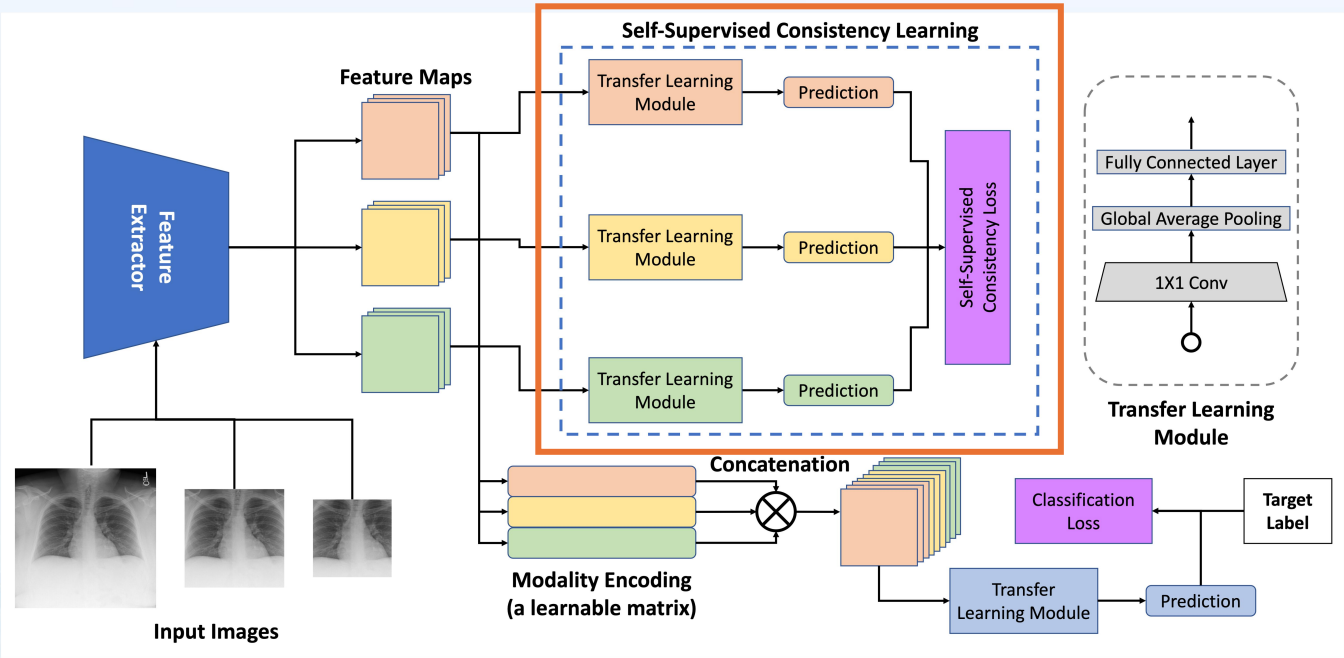
- A feature extractor is shared between all the *k+1* views.
- The *k+1* feature maps are fed to the Self-Supervised Consistency Learning module and the Main Predicting Head.



A&M-SA

Self-Supervised Consistency Learning

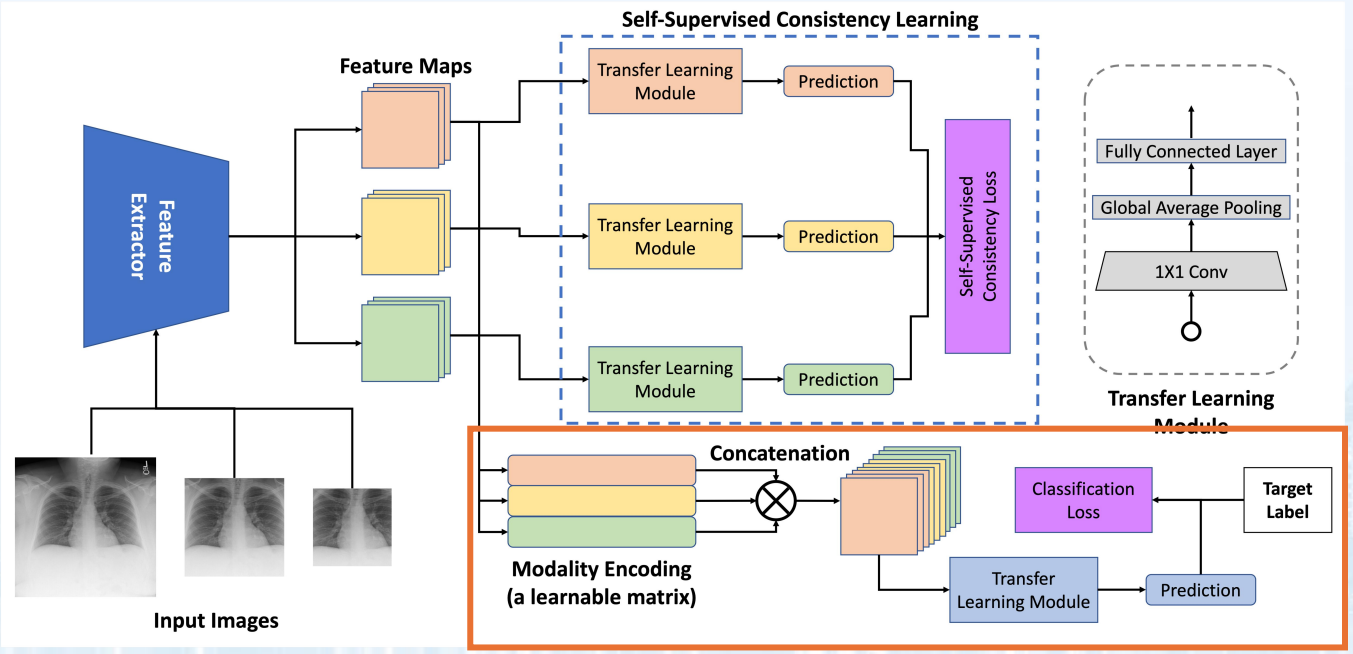
- Contains *k+1* auxiliary prediction heads, one for each view.
- KL divergence is used as the loss to measure the consistency of the prediction of every two auxiliary heads



A&M-SA

Main Predicting Head w/ Modality Encoding

- Modality Encoding
  - A learnable (k+1) x 8
  - Provide information about the views in the multi-scale input set
  - Concatenated with the corresponding feature maps
- The concatenated feature maps are used as input for the final prediction.



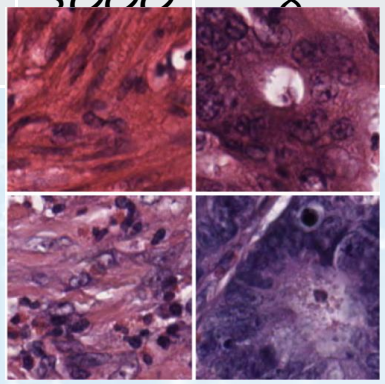
A&M-SA

Experiment Setup

- Dataset

Name	Imaging Modality	# of Images	# of Classes
Mendeley	Chest X-Ray	5856	2
Kather	Histologic	5000	8

Mendeley

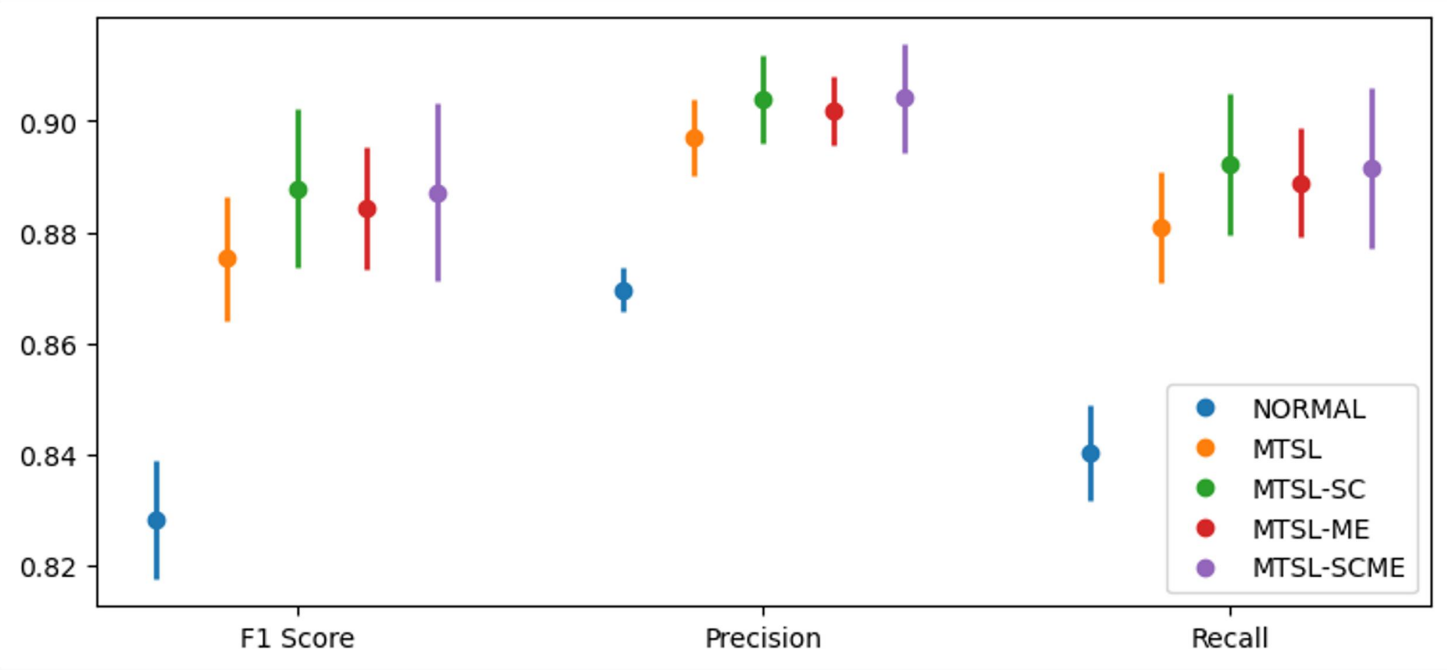


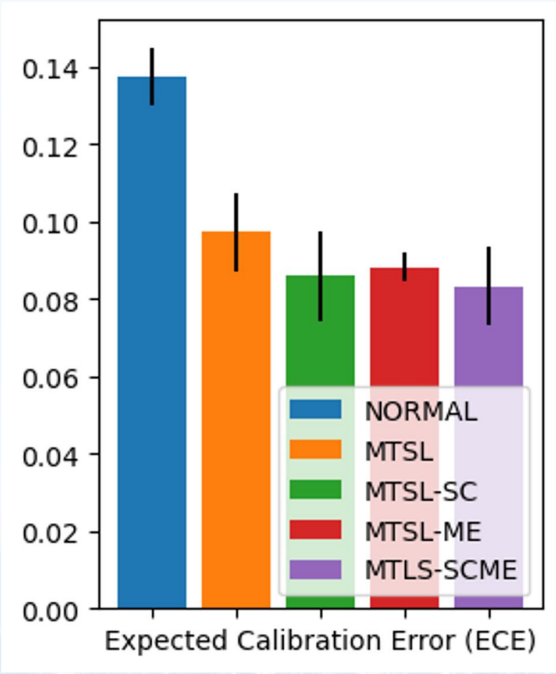
Kather

- Architecture
  - ResNet-50
- Evaluation Metrics
  - ECE – Expected Calibration Error
    - The most common metric for calibration
  - F1 Score
  - Precision
  - Recall

A&M-SA

Result -- Mendeley

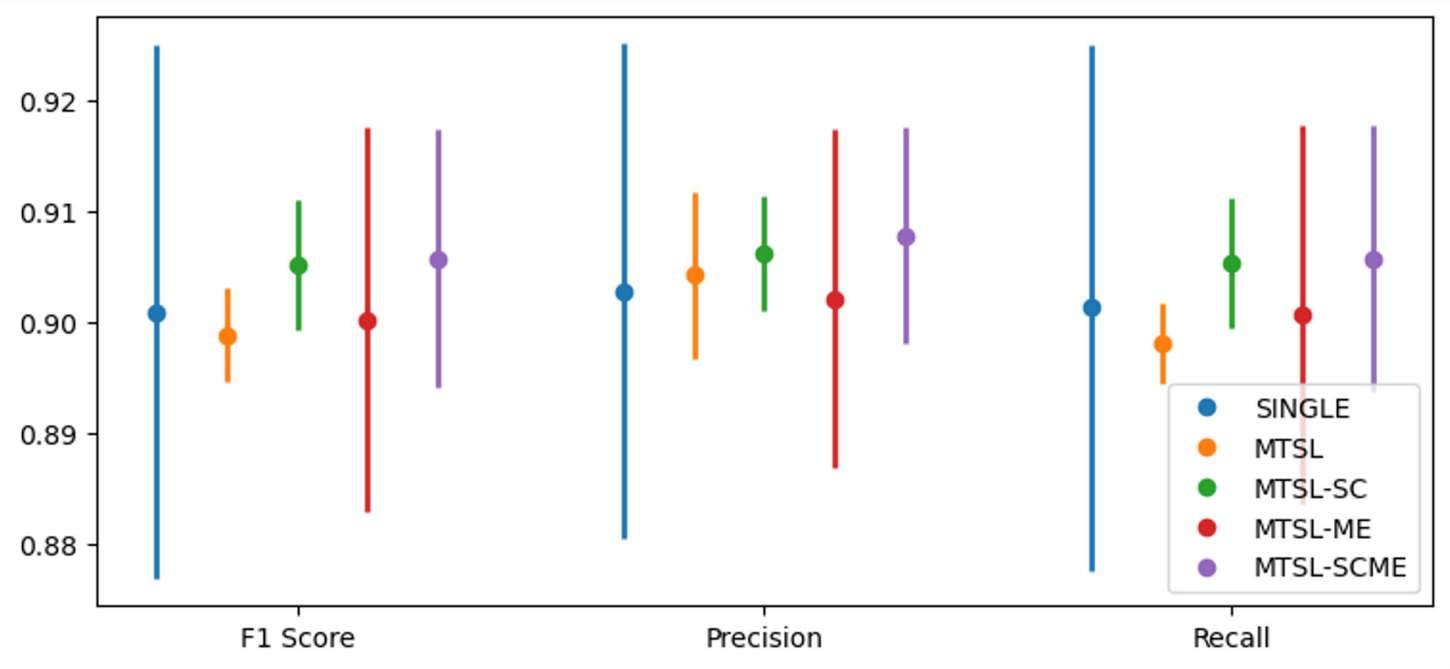


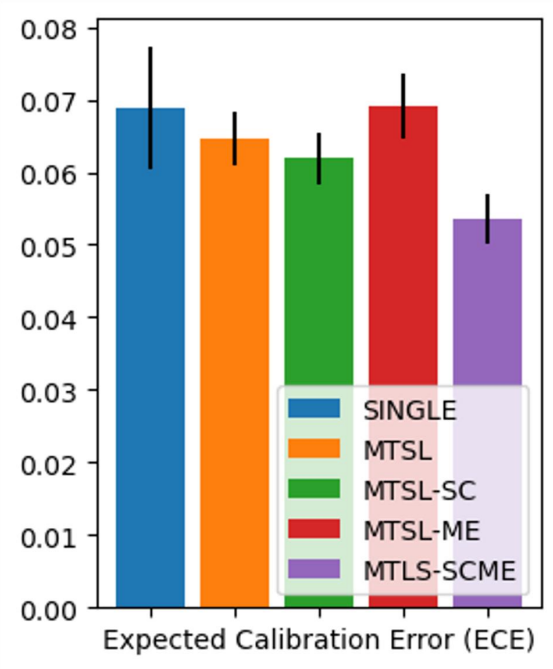


SINGLE: The ResNet-50 model with Image-Net pre-training
MTSL: The ResNet-50 model with **multi-scale input**
MTSL-SC: The ResNet-50 model with **multi-scale input & self-supervised consistency module**
MTSL-ME: The ResNet-50 model with **multi-scale input & modality encoding**
MTSL-SCME: The proposed model

A&M-SA

Result -- Kather

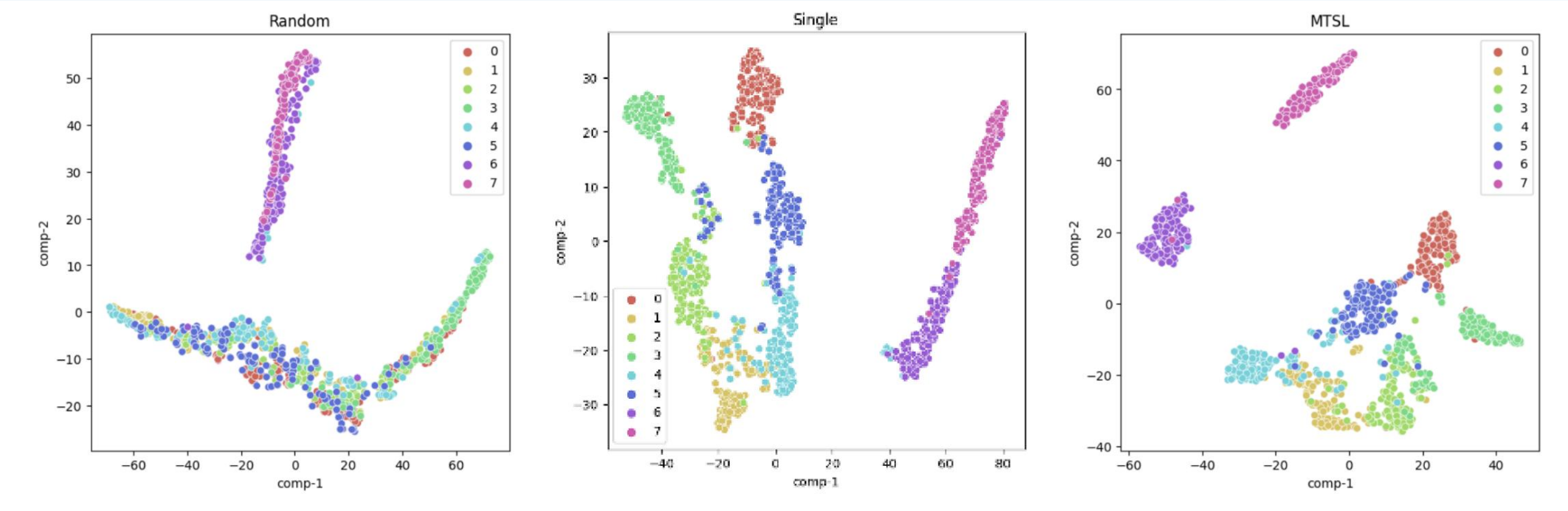




SINGLE: The ResNet-50 model with Image-Net pre-training
MTSL: The ResNet-50 model with **multi-scale input**
MTSL-SC: The ResNet-50 model with **multi-scale input & self-supervised consistency module**
MTSL-ME: The ResNet-50 model with **multi-scale input & modality encoding**
MTSL-SCME: The proposed model

A&M-SA

T-SNE for Feature Space Visualization (Kather)



Random Model

Image-Net Pre-Trained ResNet-50 Model

ResNet-50 with Multi-Scale Input

A&M-SA

Conclusion

TEXAS A&M UNIVERSITY
SAN ANTONIO

Thank you!

Questions?