

# A Calibrated approach to Semi-Supervised Learning For Medical Image Classification

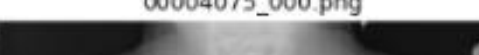
**Shreya Singh | [singh.shreya1@northeastern.edu](mailto:singh.shreya1@northeastern.edu)**

## Opportunity

## Introduction :

- Deep learning models achieve strong performance for radiology image classification, but their practical application is bottlenecked by the need for large labeled training datasets.
- Semi-supervised learning (SSL) approaches use small labeled datasets alongside larger unlabeled datasets for machine learning tasks and offer potential for reducing labeling cost.
- Confidence calibration – the problem of predicting probability estimates representative of the true correctness likelihood – is also important, especially when dealing with medical dataset.

### Research Question:

- Can state of the art methods for Semi-supervised Learning for image Classification like Fix-Match and Mean-Teacher be applied to radiology images ?
  - Will these methods achieve results that are confident enough to be used by the medical fraternity?
  - Are these models stable and calibrated?
- 

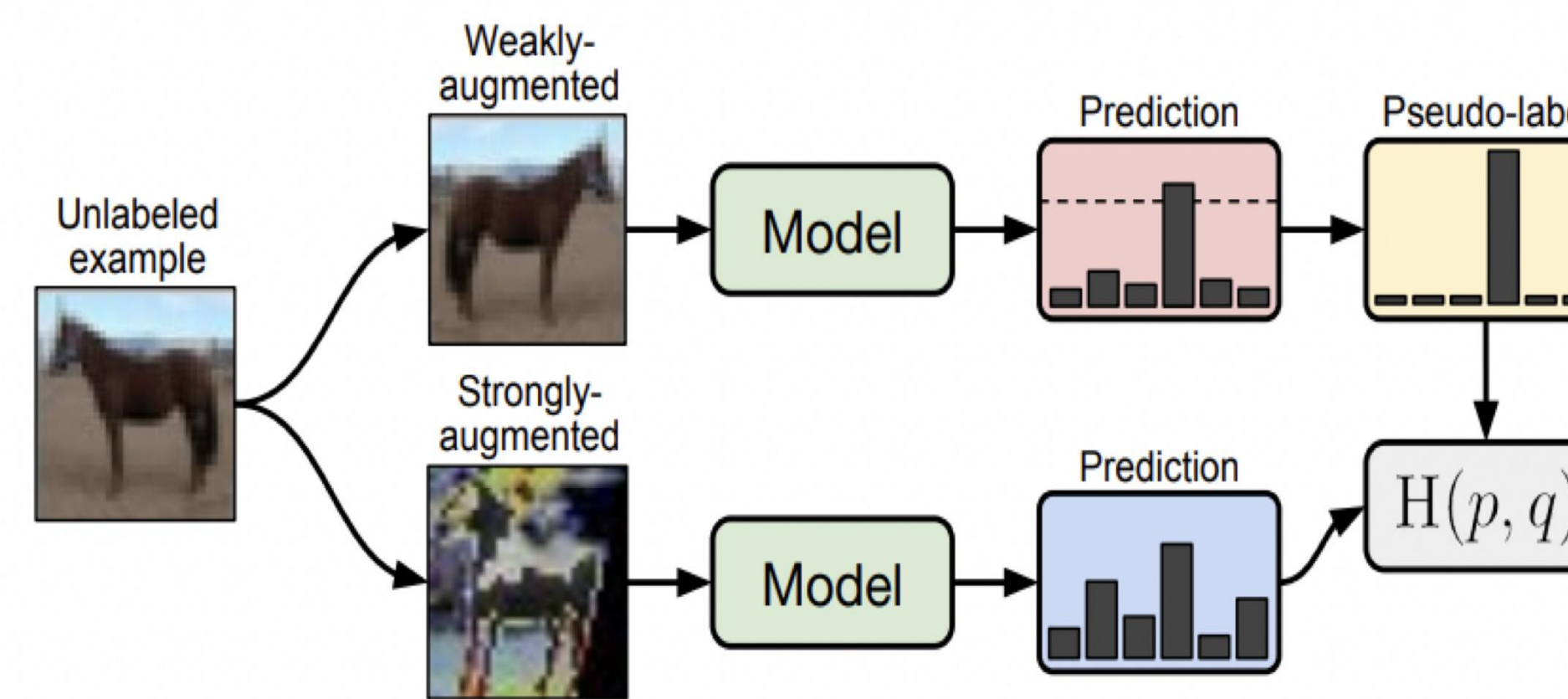
### Short Answer:

- Even on a very small dataset, both Fix-Match and Mean Teacher perform well.
- Temperature scaling improves ECE on both Methods.

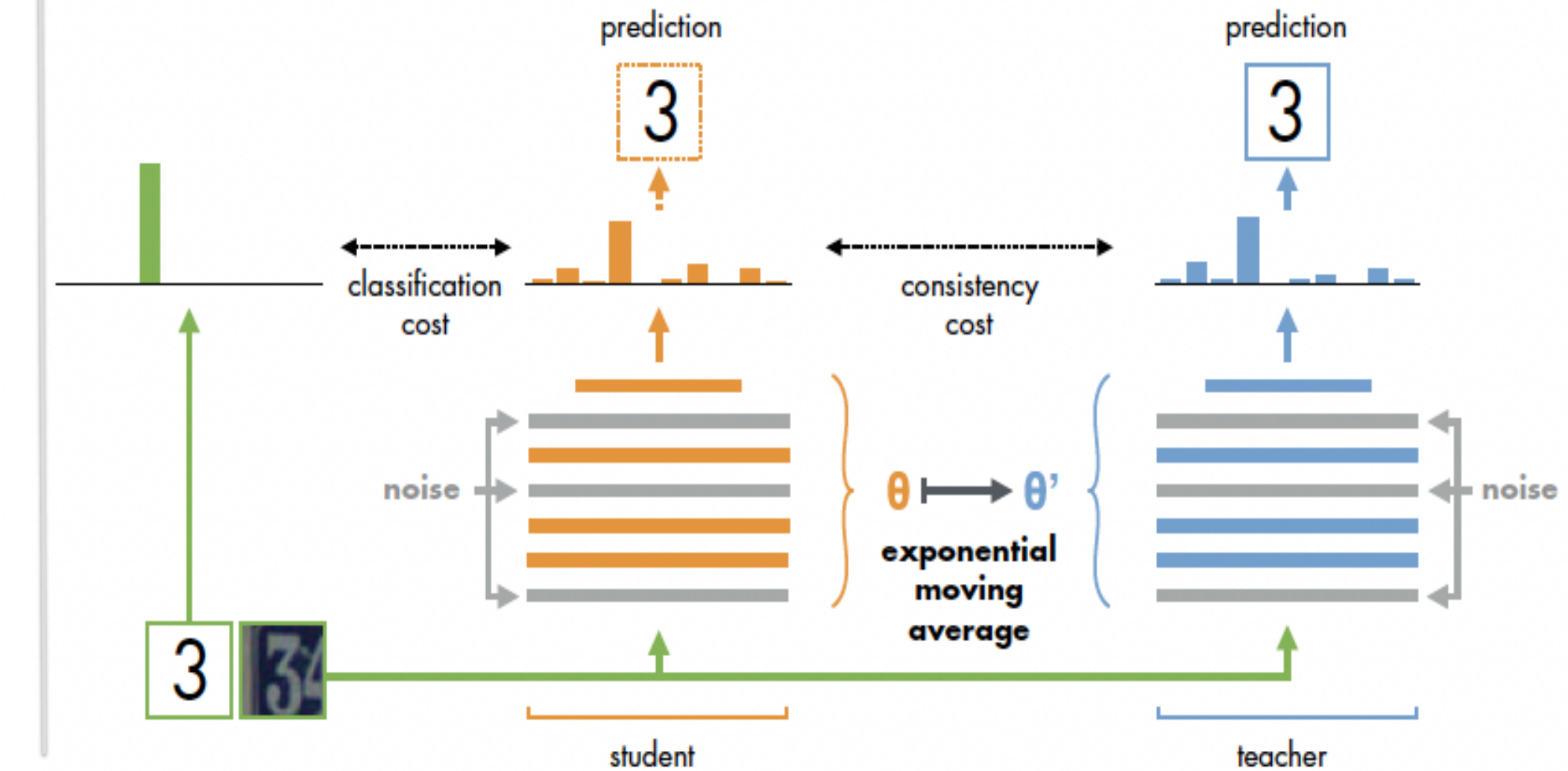


## Methods

### FixMatch:

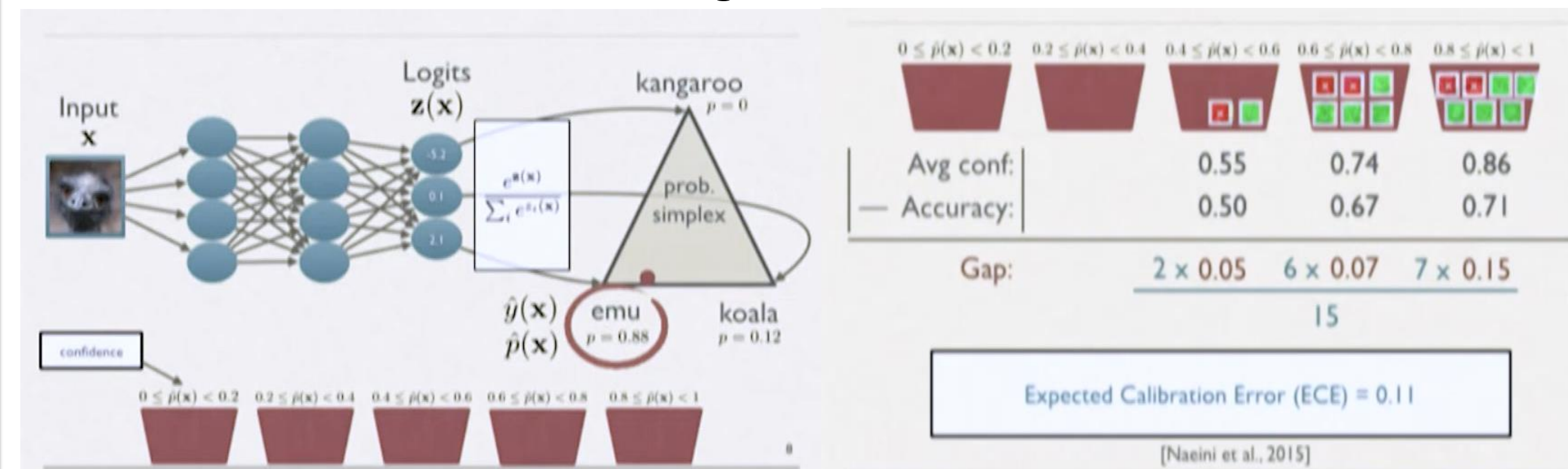


**Mean Teacher :**

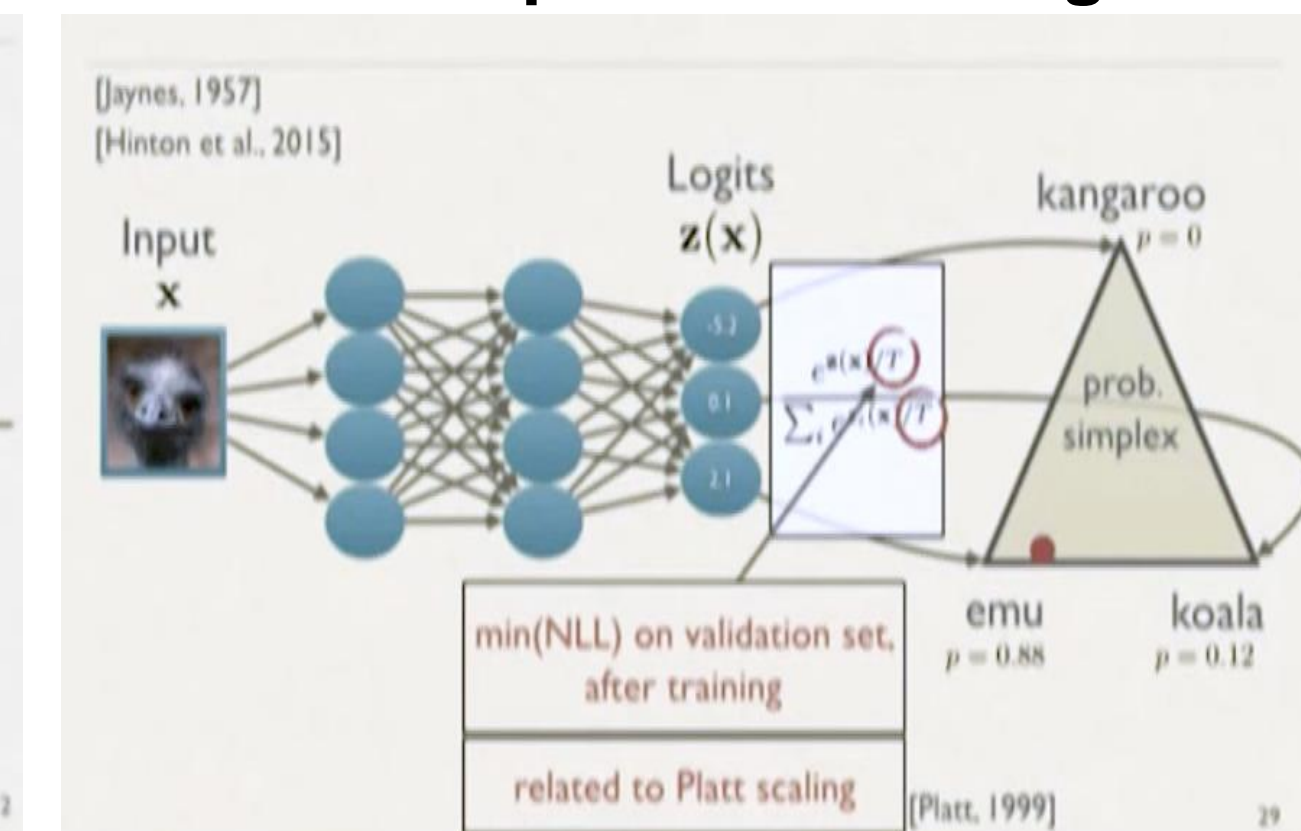


## Calibration:

## Measuring Calibration Error



## Temperature Scaling



## Results

**Dataset used :**

- Chest X-ray scans from NIHCC
- There are 14 classes of diseases in these scans, some scans contain possibility of multiple diseases.
- For the first set of experiments, only unique image per class was used, 65 in training and 25 in validation set. Labelled images were 50 per class.

Method	Top-5 Accuracy	ECE before calibration	ECE after Temp Scaling
FixMatch	36.86	0.091	0.041
Mean-Teacher	43.75	0.098	0.056

## Impact and Discussion

- Obtaining medical scans is often far cheaper than paying an expert doctor to label every image.
- FixMatch can begin to bridge the gap between low-label semi-supervised learning and few-shot learning or clustering.
- Mean Teacher scales well to state-of-the-art architectures and large image sizes.
- Using these semi-supervised techniques with temperature scaling leads to lesser expected Calibration error.
- Modern neural networks tend to be mis calibrated, with factors related to model capacity, batch normalization and weight-decay.
- Future experiments on larger non-unique balanced and unbalanced dataset.
- Comparing and exploring graph-based algorithms like No-Teacher for semi-supervised learning.

**Key References :**  
<https://arxiv.org/pdf/1705.02315.pdf>  
<https://arxiv.org/pdf/2001.07685.pdf>  
<https://arxiv.org/abs/1703.01780>  
<https://arxiv.org/pdf/1706.04599.pdf>

Acknowledgements :  
Prof . Hongyang Zhang  
Dongyue Li

