# Calibration of the Neural Network PiCAP (Protein Interaction of CArbohydrates Predictor)

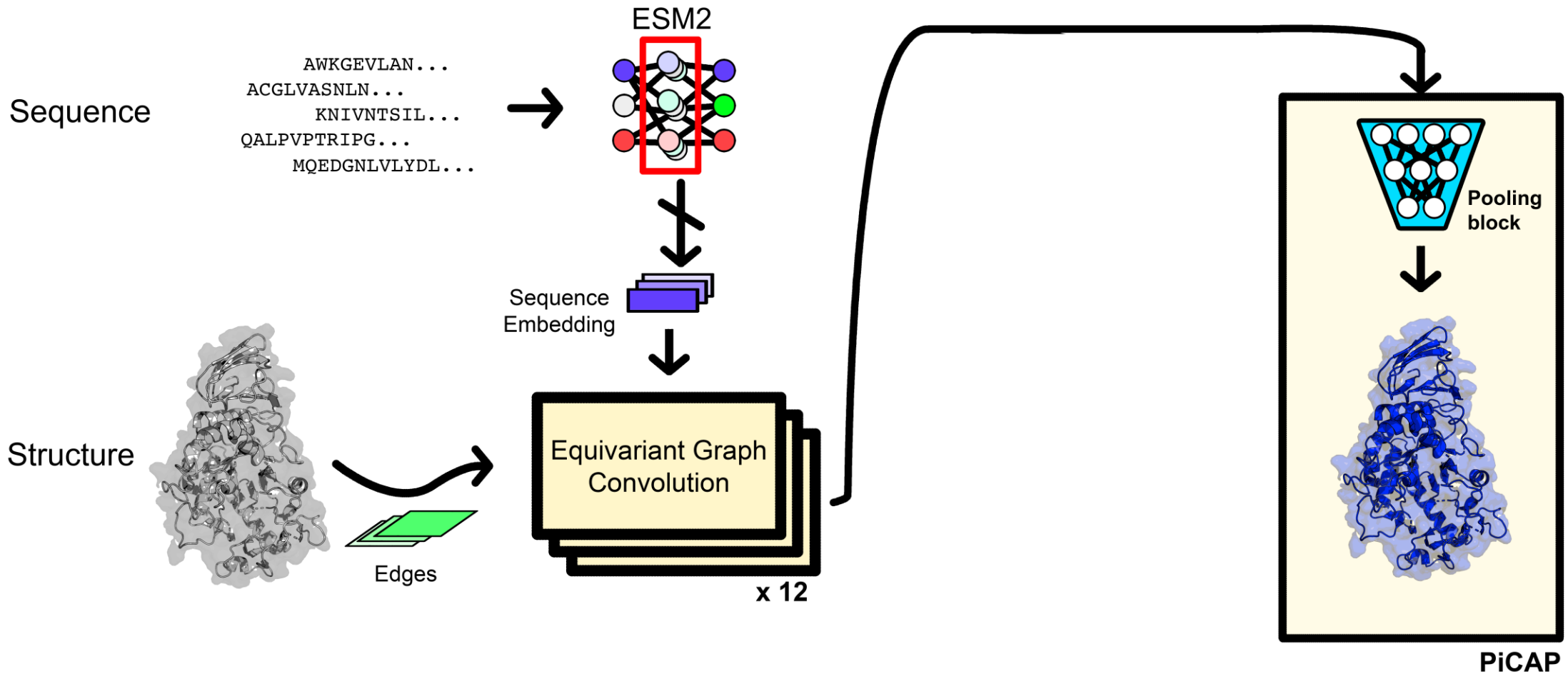
**Sho Takeshita**

Gray Lab

3.14.2025

JOHNS HOPKINS
UNIVERSITY

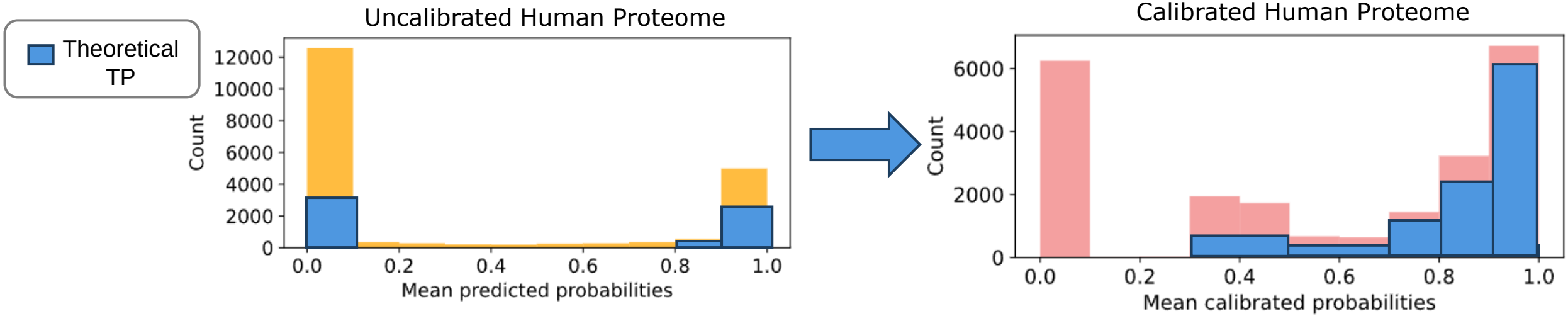
# Sam developed an algorithm to predict whether proteins bind to carbohydrates or not




Courtesy of SW Canner

# Calibration is a technique that smooths binary classifications to a scalar distribution

- PBind does not reflect true positive rate
- If there are 100 predictions of .8 predicted PBind, 80 of the predictions should be true
- Reduces difference between TP rate and predicted PBind



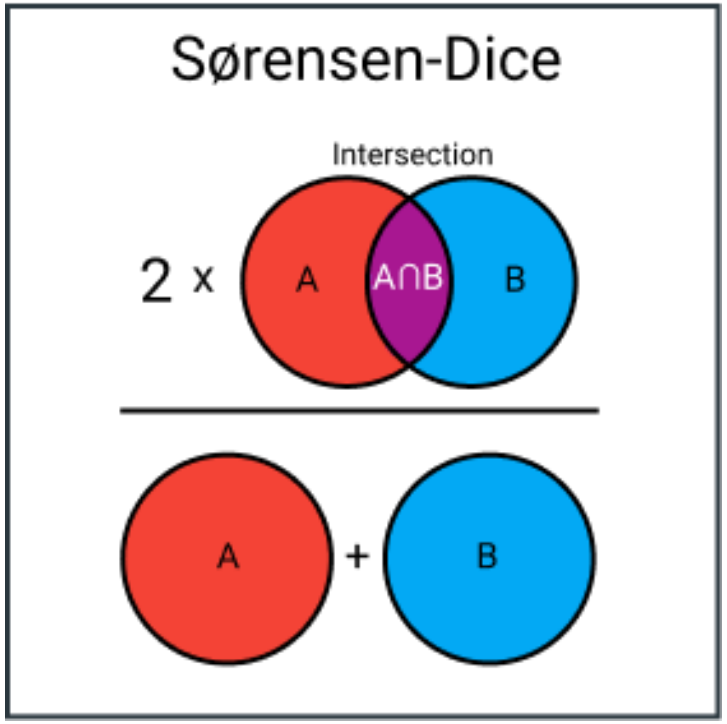
# Types of Calibration

- Regularization methods
- Post-hoc Methods
- Uncertainty Estimation

# Regularization Using Dice++

$$\mathcal{L}_{Dice} = 1 - \frac{2TP}{2TP + FP + FN} \rightarrow \mathcal{L}_{Dice++} = 1 - \frac{2TP}{2TP + (FP)^{\gamma} + (FN)^{\gamma}}$$

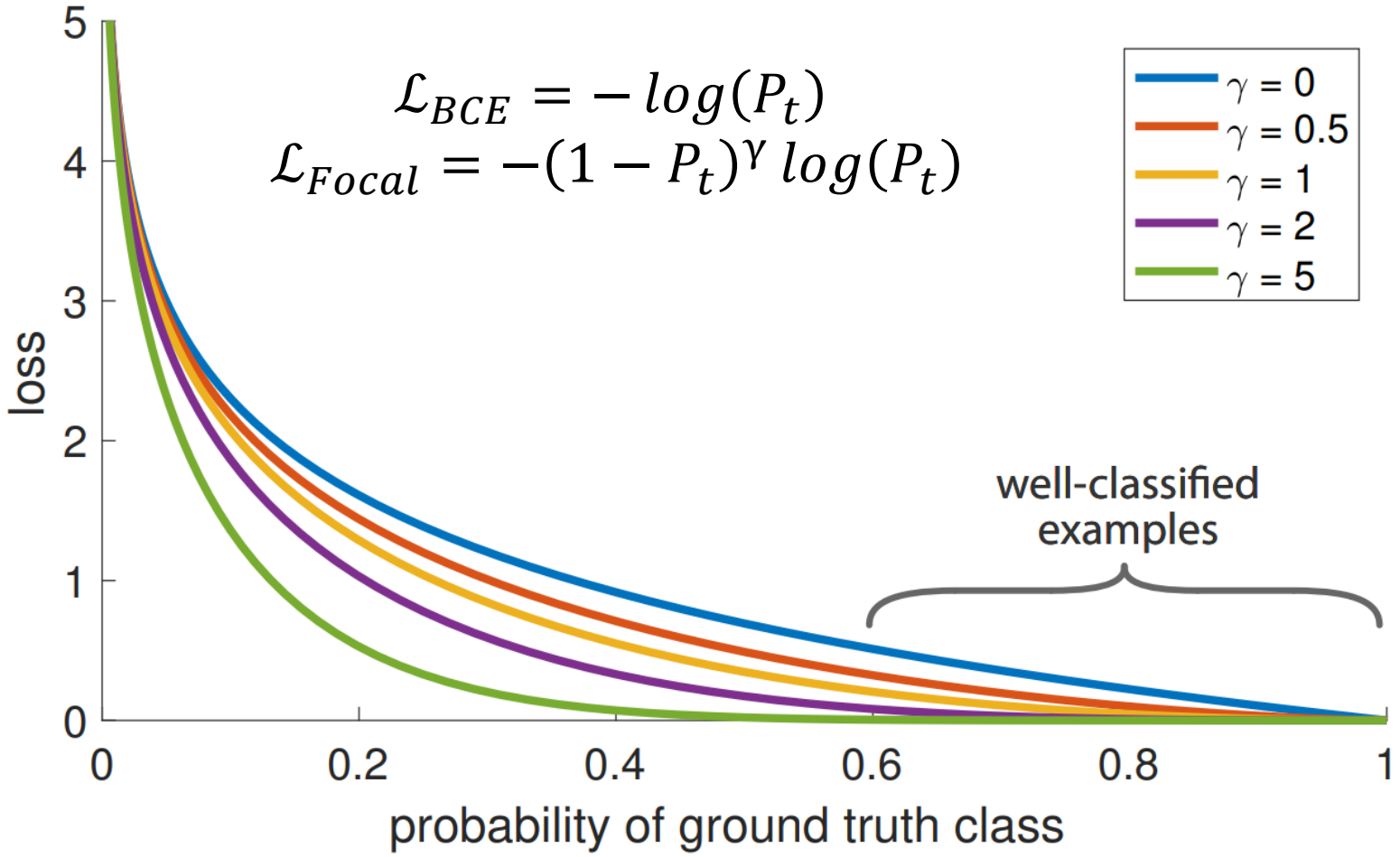
- Emphasizes greater loss when FP and FN are greater by adding $\gamma$ hyperparameter




M. Grootendorst, 2021, Medium.com

# Regularization Using Focal loss

Puts more focus on hard, misclassified data





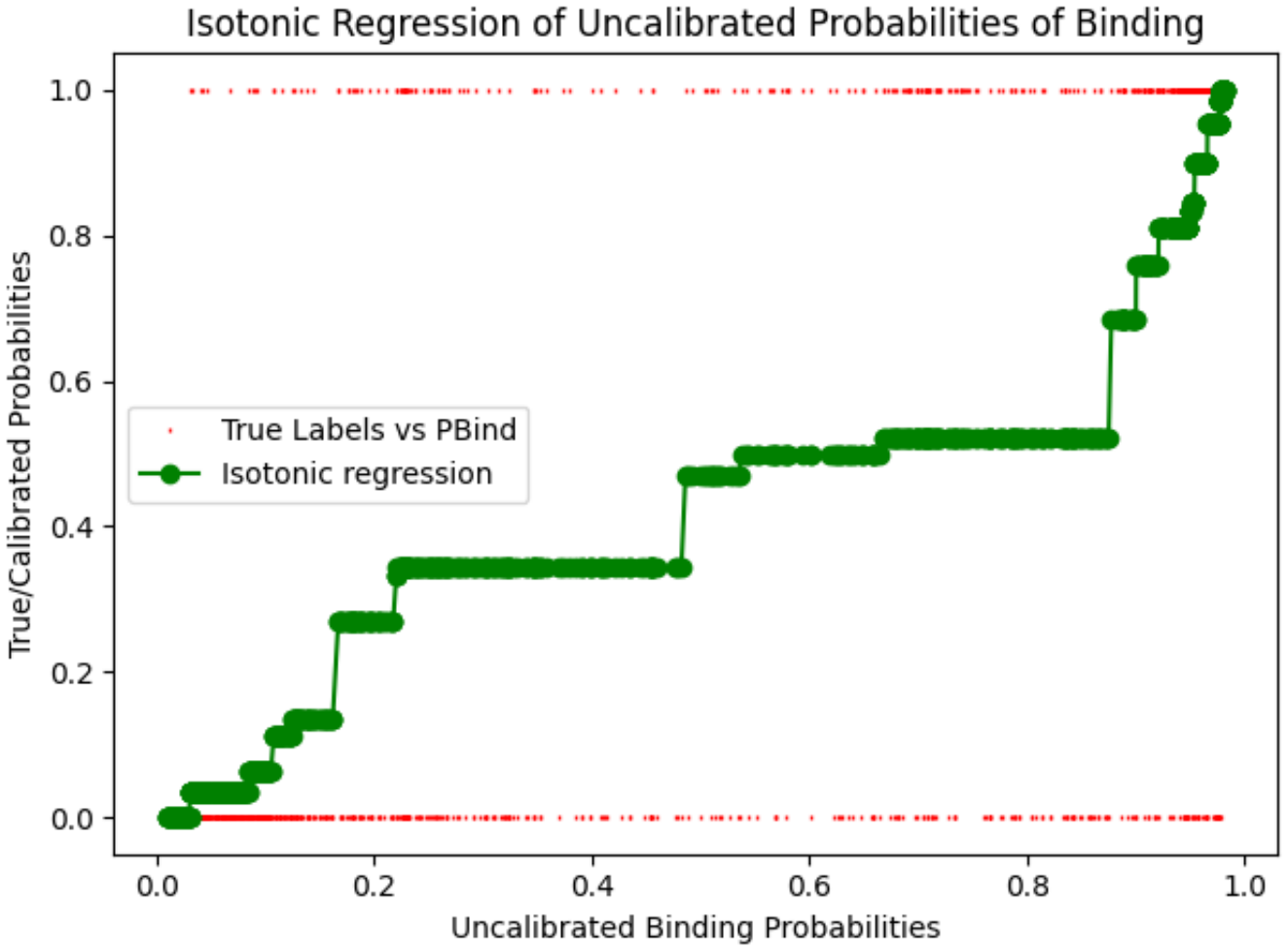

T. Lin, et al., 2018, arXiv

# Post-hoc Calibration with Isotonic Regression

Fits monotonically increasing function to True Probabilities vs. Raw Predictions on secondary training set

$$\min \sum_{i=1}^{n} w_i (\hat{y}_i - y_i)^2$$

$$\hat{y}_i \leq \hat{y}_j \text{ for all } (i, j) \in E$$

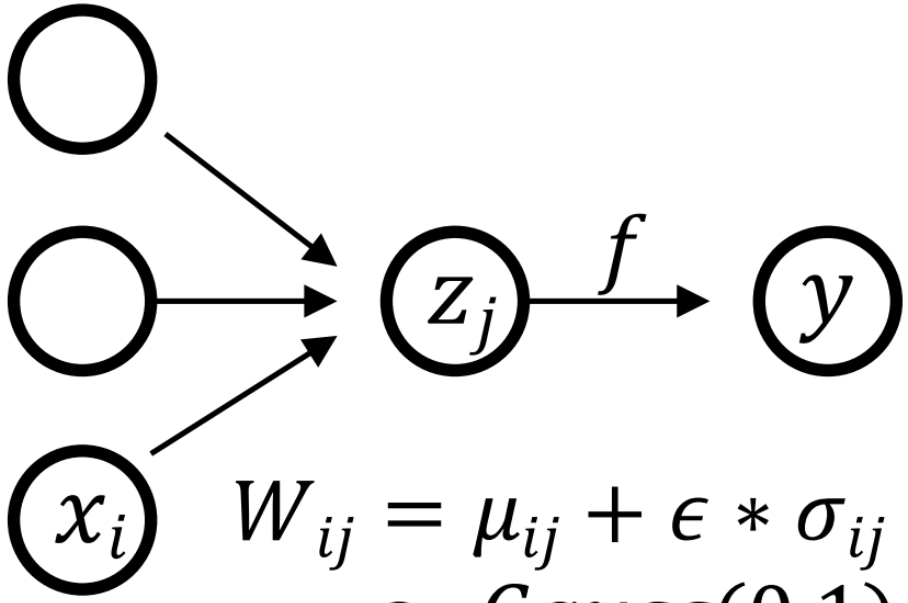





Isotonic Regression, 2025, Wikipedia

# Uncertainty Estimation with Bayesian Neural Network (BNN)

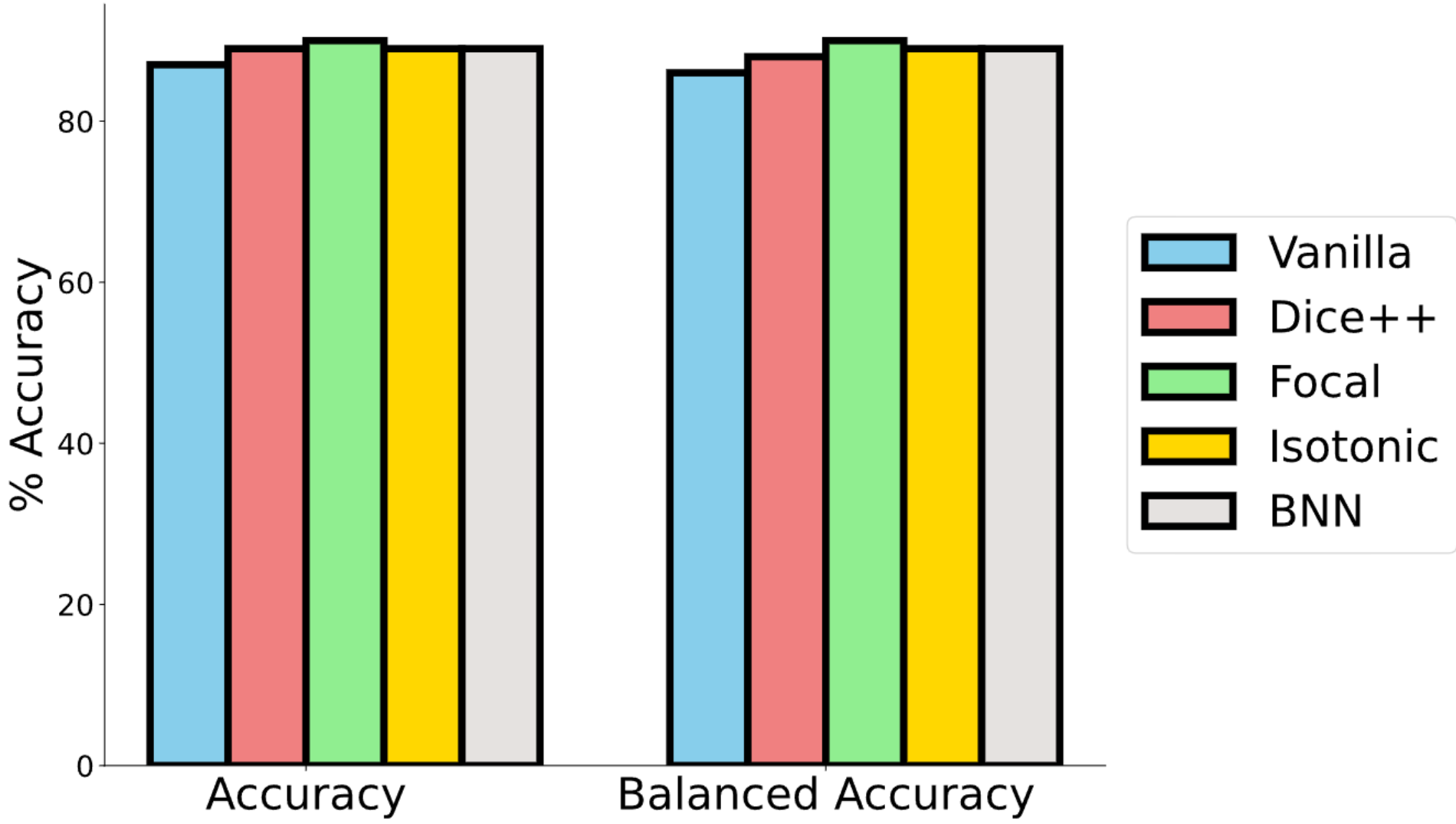
- Learns a gaussian distribution for weights instead of 1 scalar value
- During inference samples from distribution
- Can quantify uncertainty



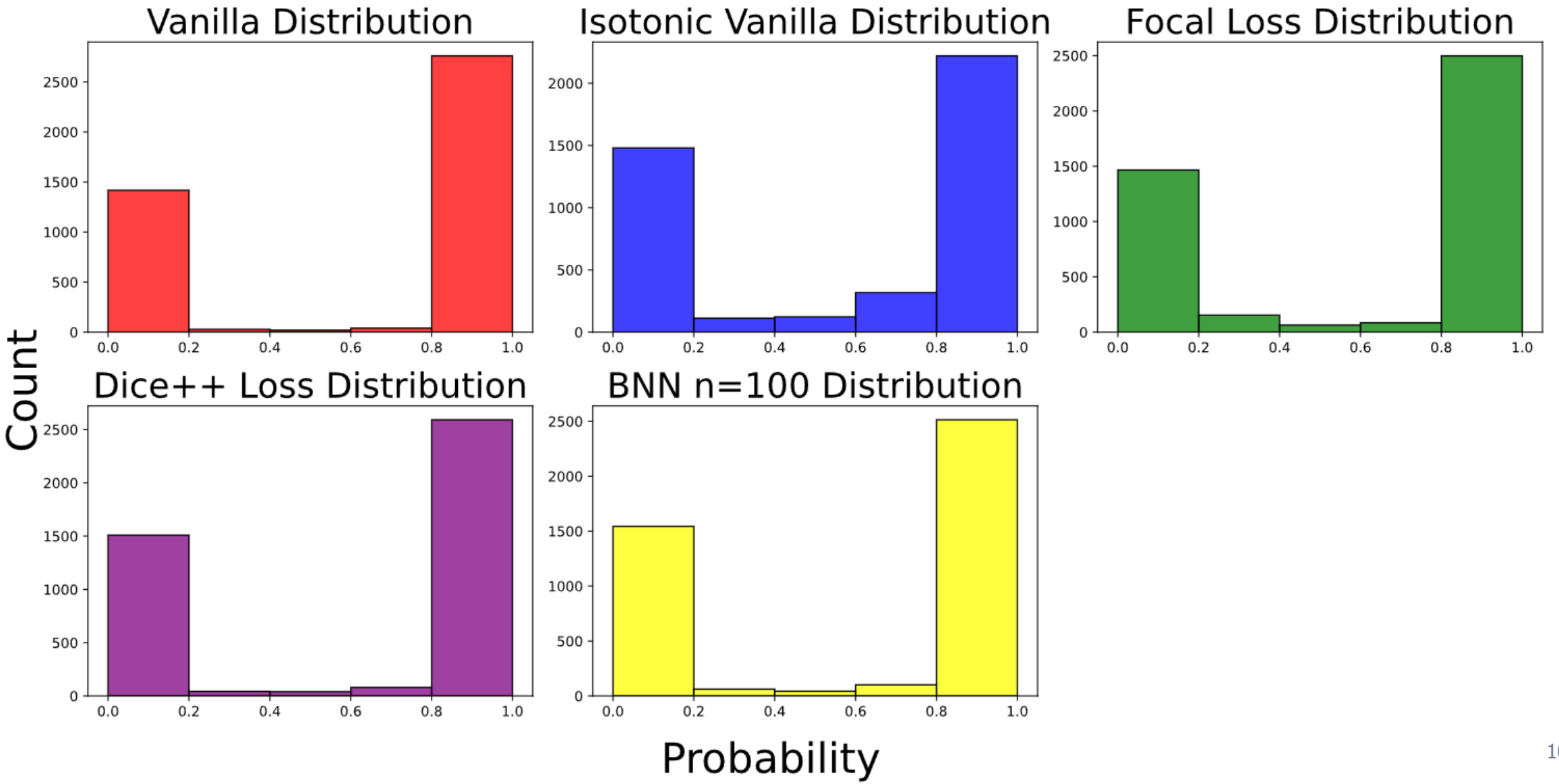



C. Wang, 2024, arXiv

# Accuracies


% Accuracy
0
20
40
60
80
Accuracy
Balanced Accuracy
Vanilla
Dice++
Focal
Isotonic
BNN

# Calibration shifts the distributions



# Metrics

$$\mathrm{acc}(b_m) = \frac{1}{|b_m|} \sum_{i \in B_m} \mathbf{1}(\hat{y}_i = y_i),\ \mathrm{conf}(b_m) = \frac{1}{|b_m|} \sum_{i \in B_m} p_i.$$

## Maximum Calibration Error

$$\mathrm{MCE} = \max_{m \in \{1,\ldots,M\}} |\mathrm{acc}(b_m) - \mathrm{conf}(b_m)|$$

## Expected Calibration Error

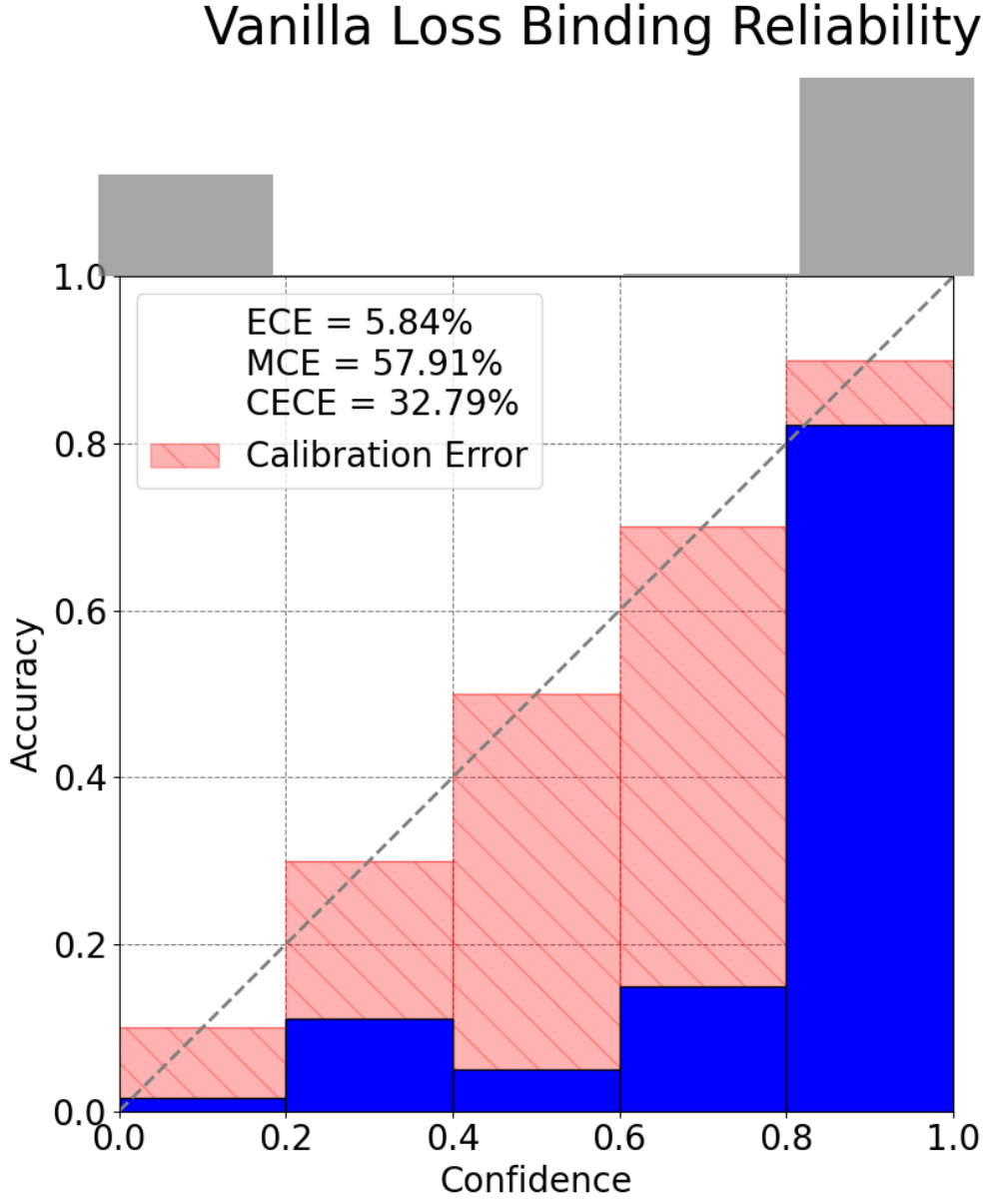
$$\mathrm{ECE} = \frac{1}{N} \sum_{m=1}^{M} |b_m|\, |\mathrm{acc}(b_m) - \mathrm{conf}(b_m)|$$

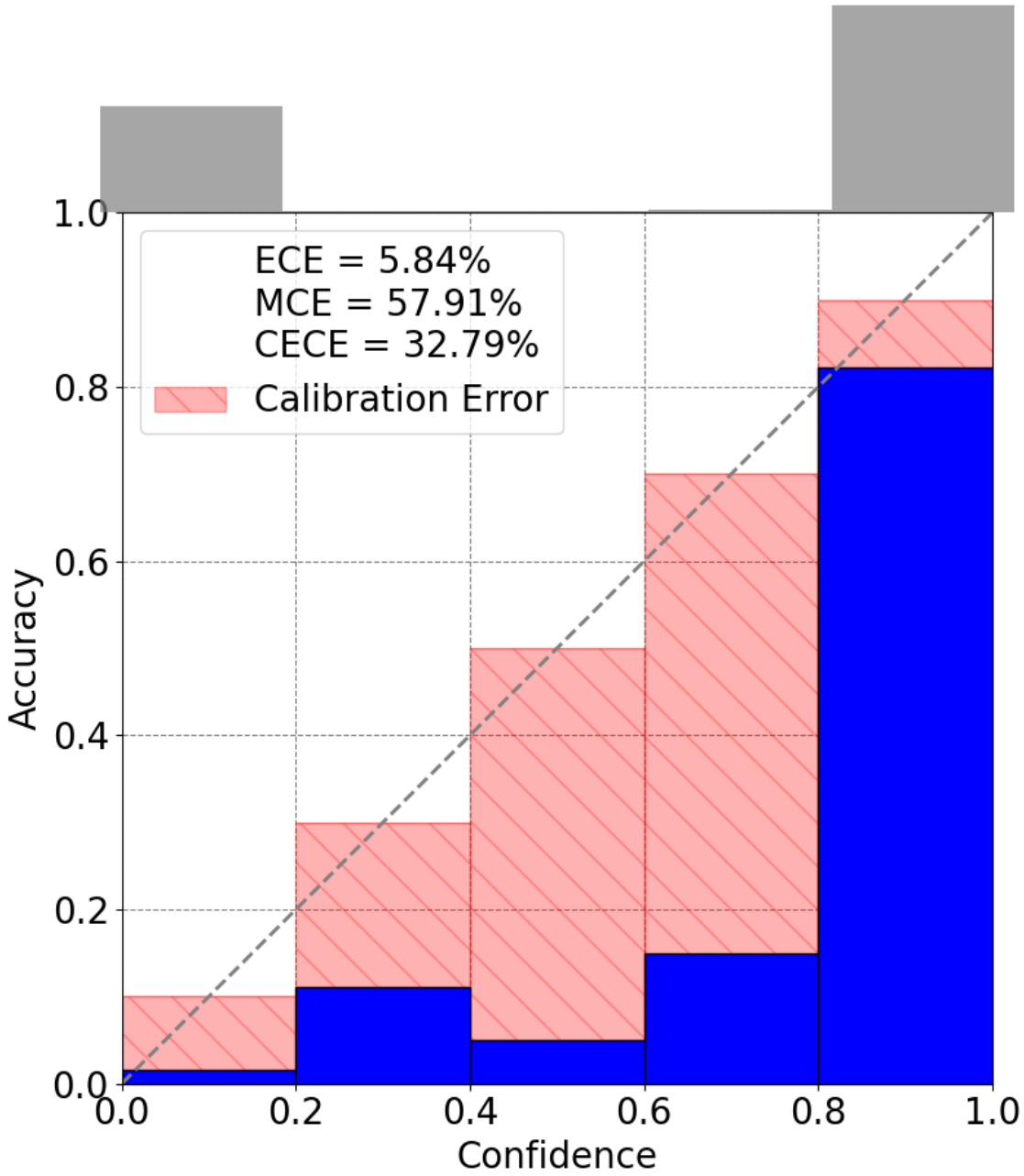
## Classwise Expected Calibration Error

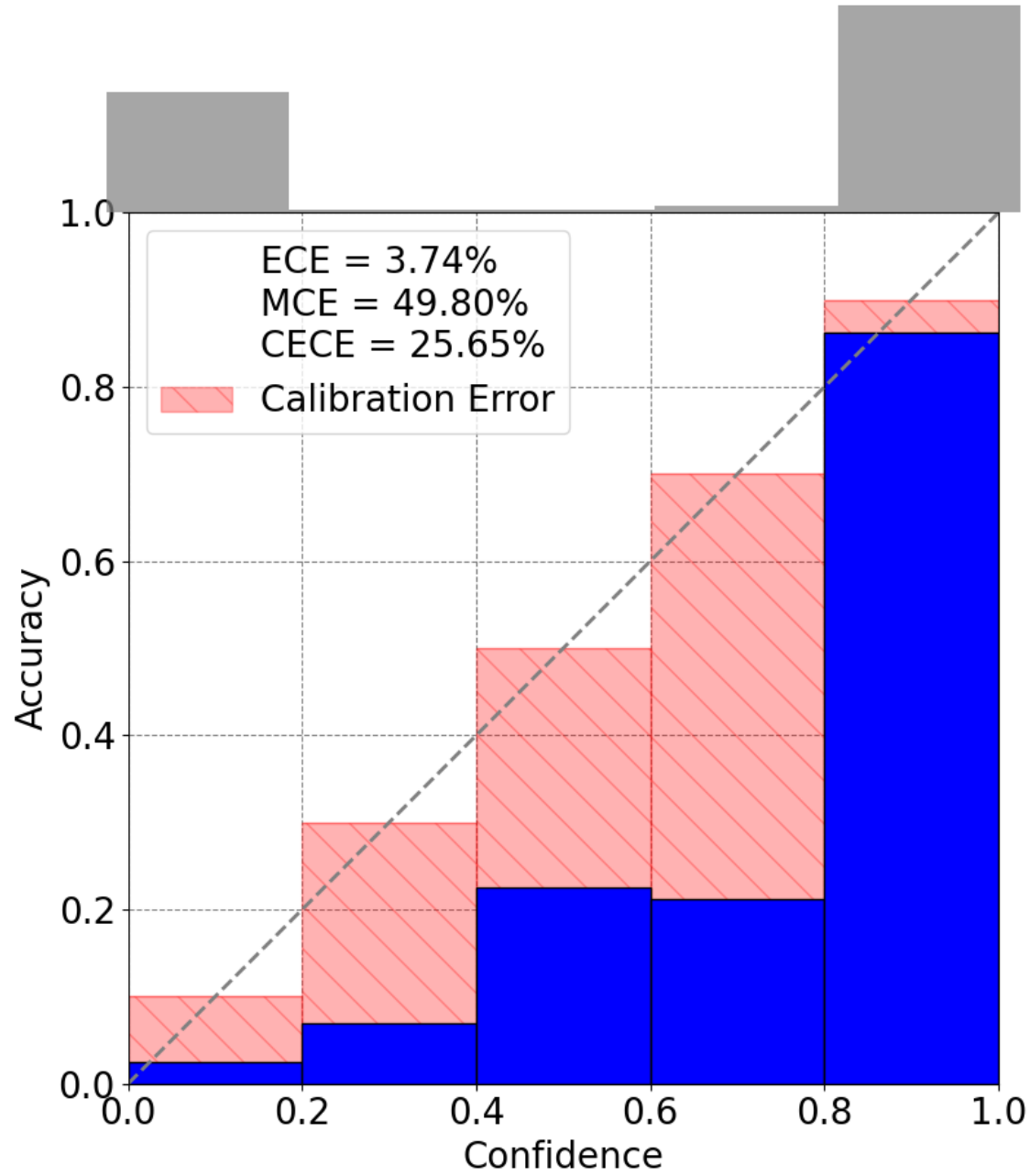
$$\mathrm{CECE} = \sum_{m=1}^{M} \sum_{c=1}^{K} \frac{|b_{m,c}|}{NK} |\mathrm{acc}_c(b_{m,c}) - \mathrm{conf}_c(b_{m,c})|$$

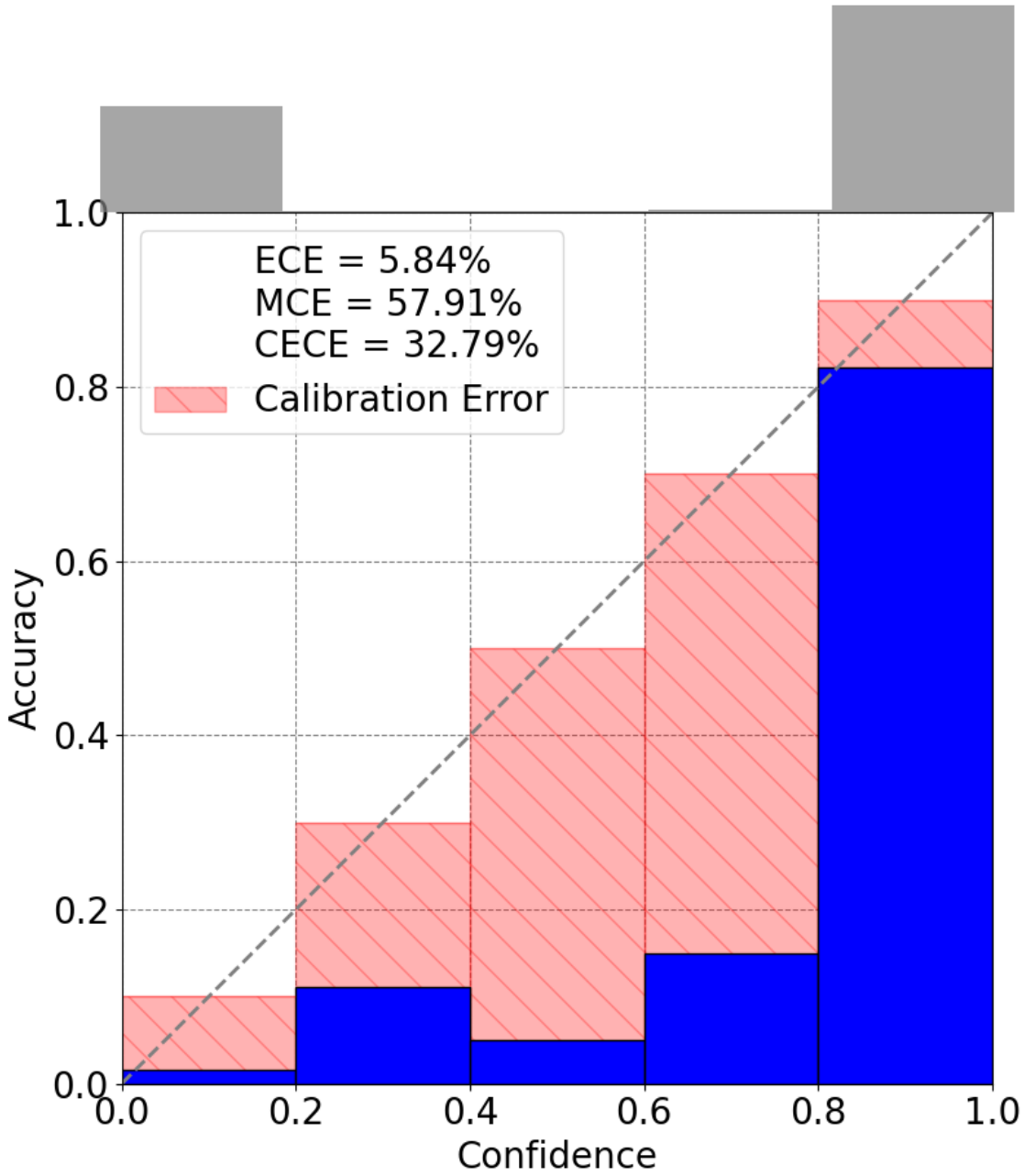
# Reliability Diagram

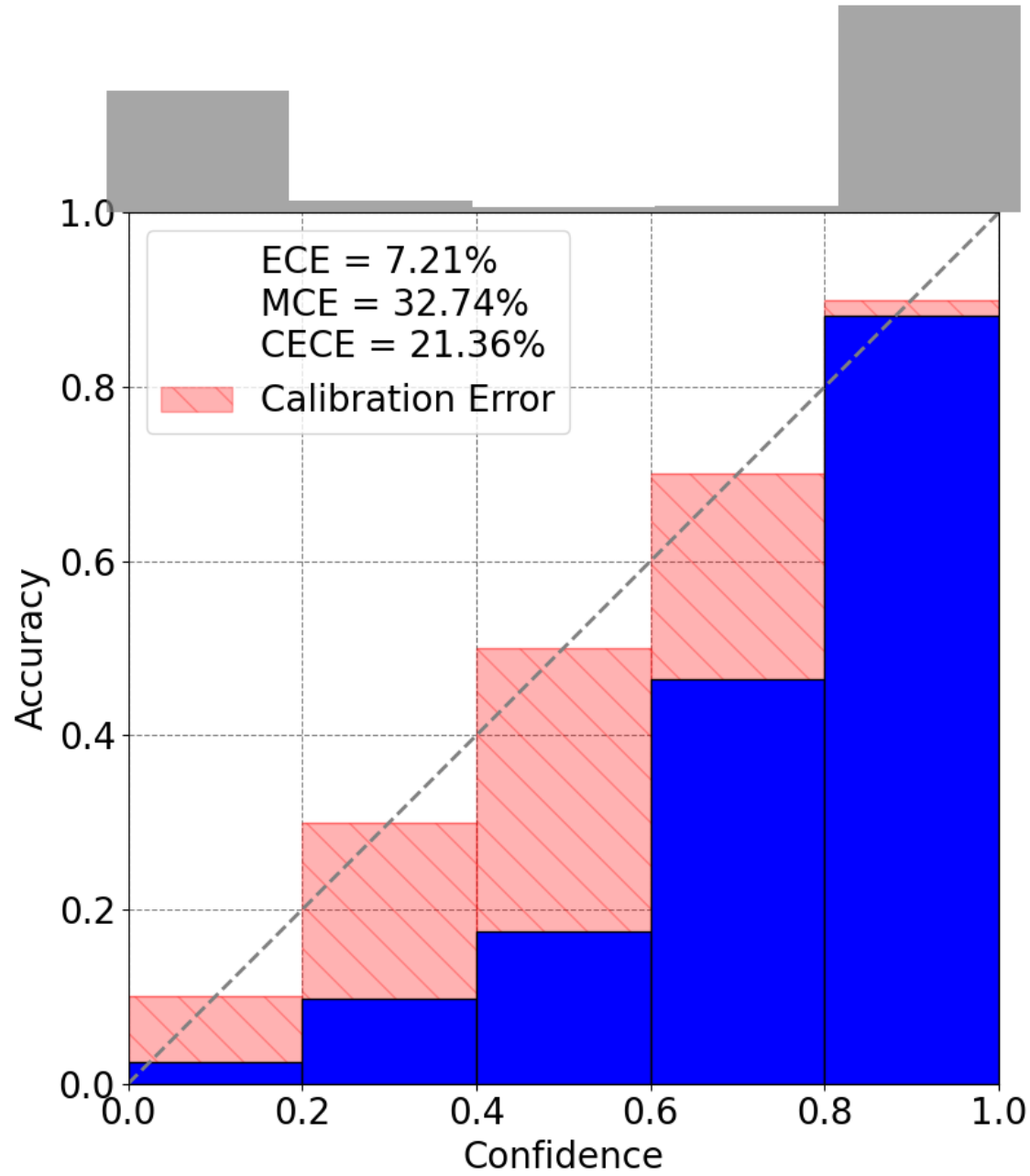
- Gives a sense of how calibrated the model is
- Dashed line represents perfect calibration
- Red bars represent gap to perfect calibration

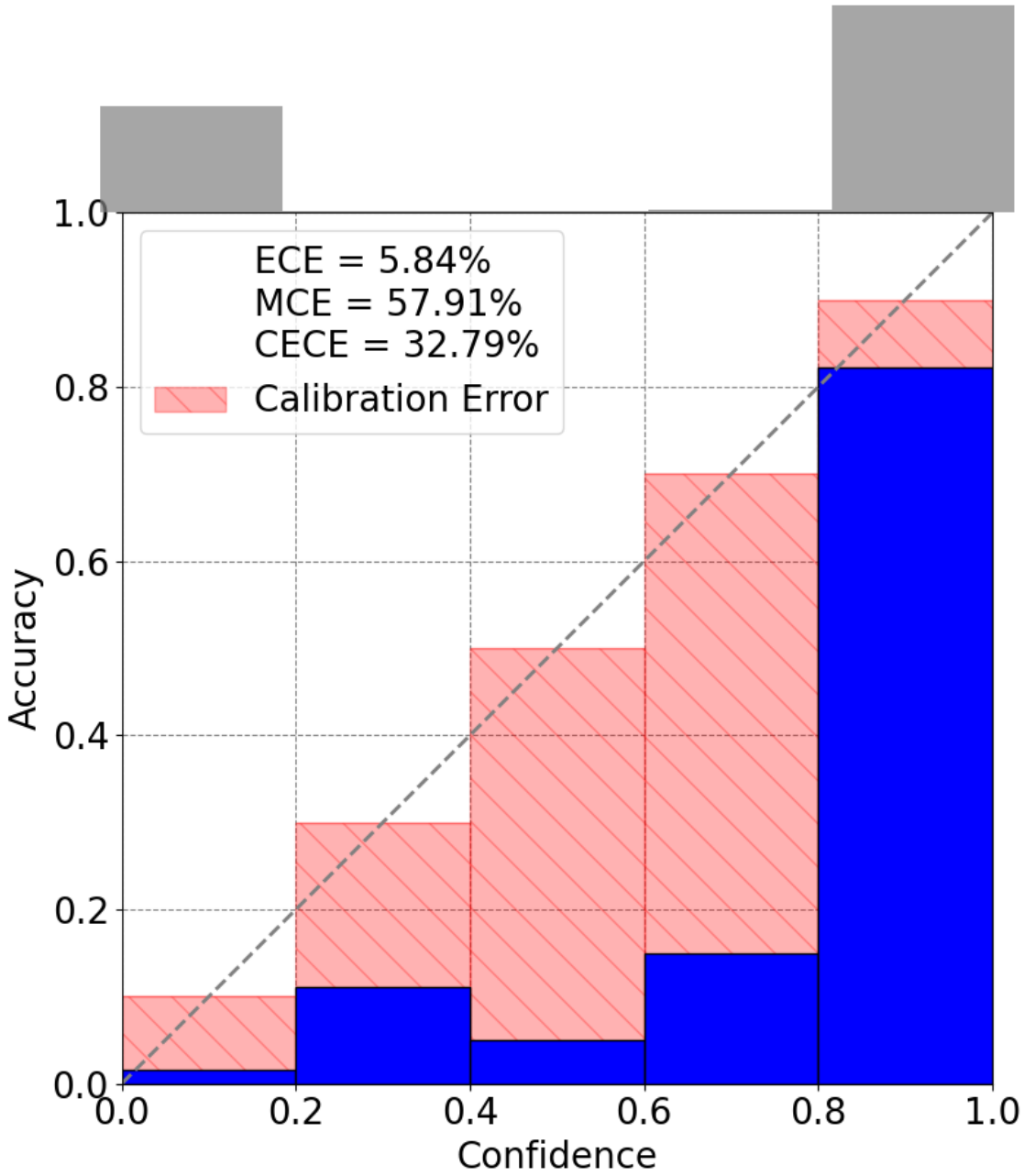


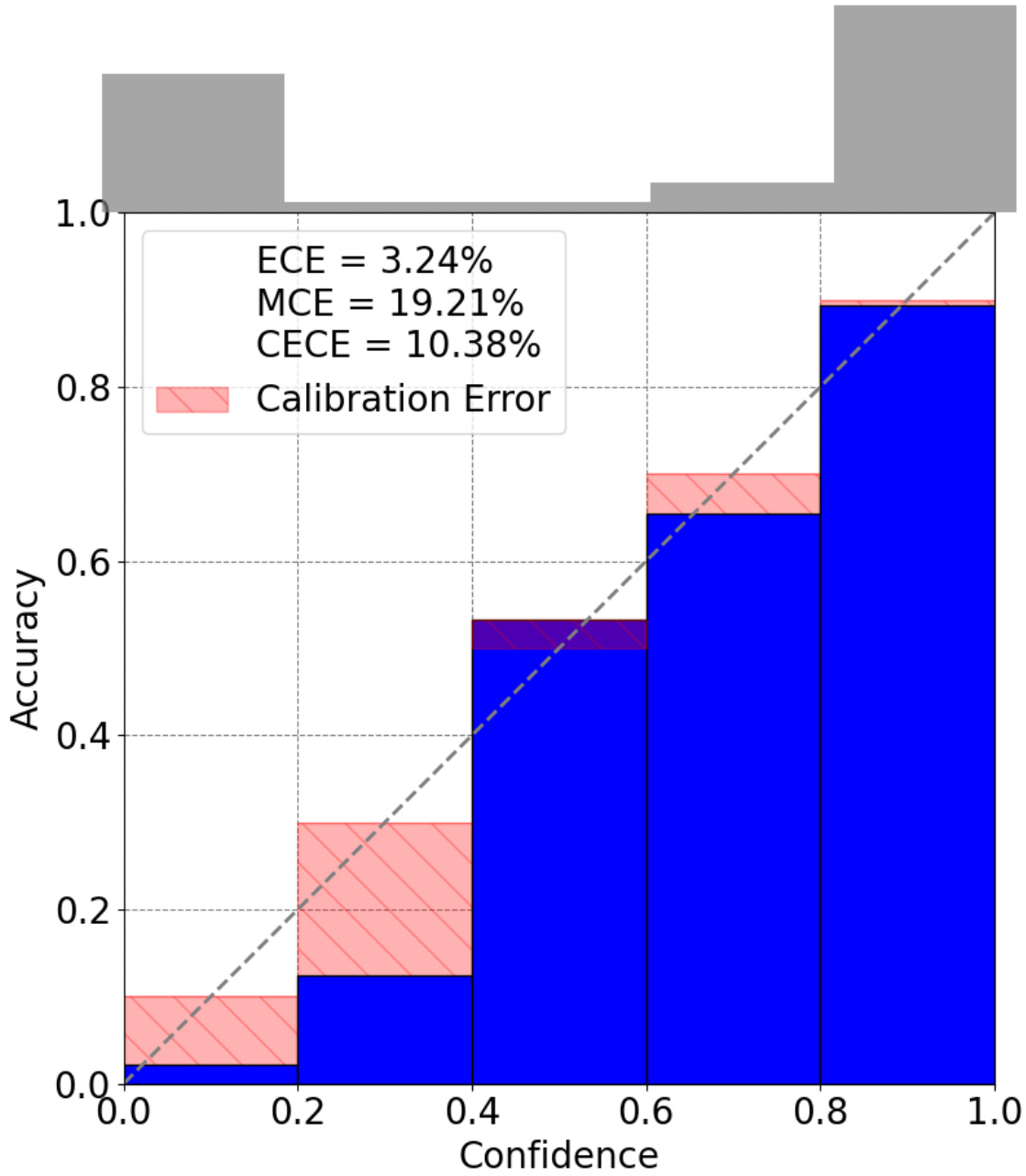

Vanilla Loss Binding Reliability
ECE = 5.84%
MCE = 57.91%
CECE = 32.79%
Calibration Error
Accuracy
Confidence
0.0
0.2
0.4
0.6
0.8
1.0

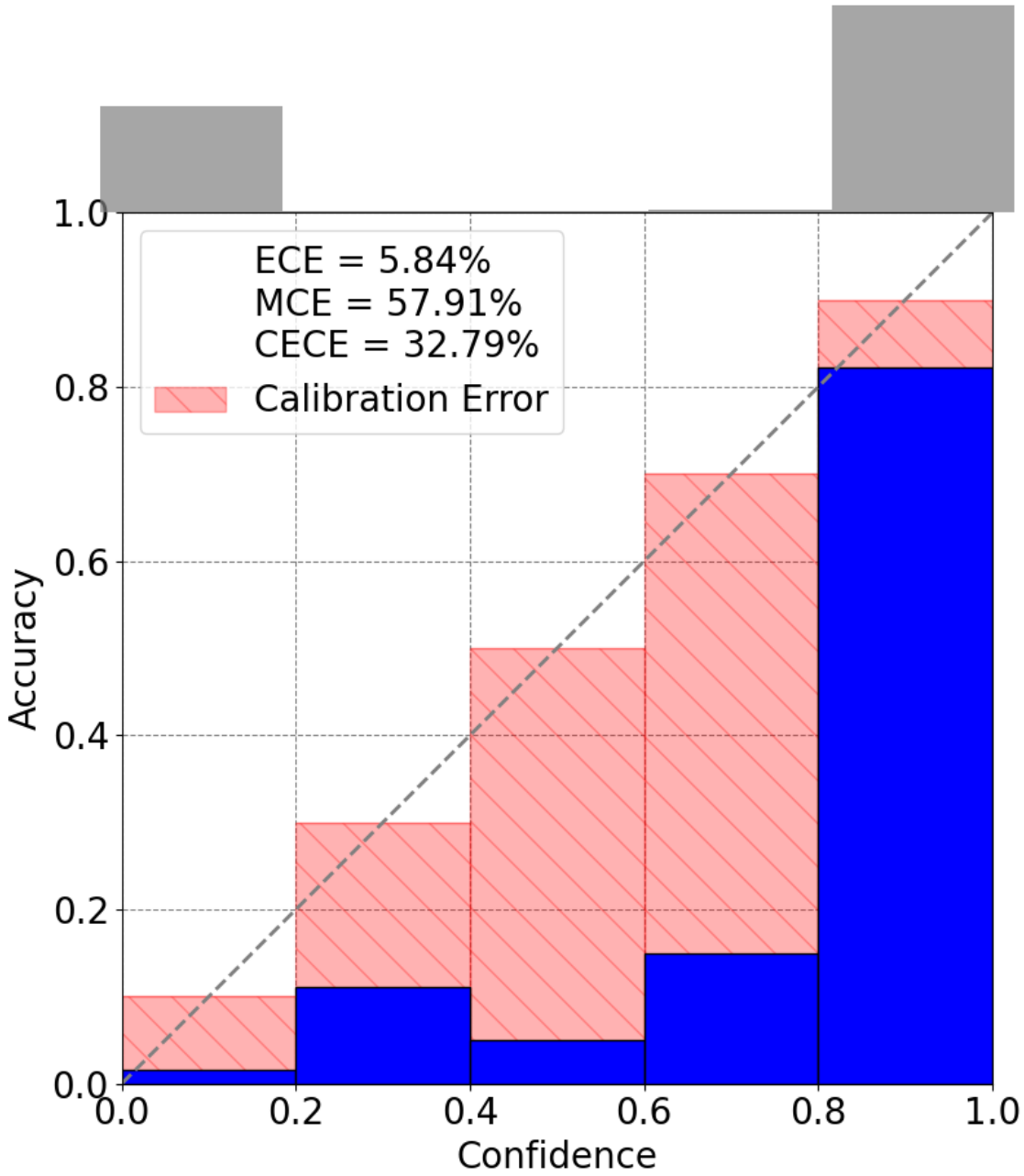

Dice++ Loss Binding Reliability
ECE = 3.74%
MCE = 49.80%
CECE = 25.65%
Calibration Error
Accuracy
Confidence
0.0
0.2
0.4
0.6
0.8
1.0

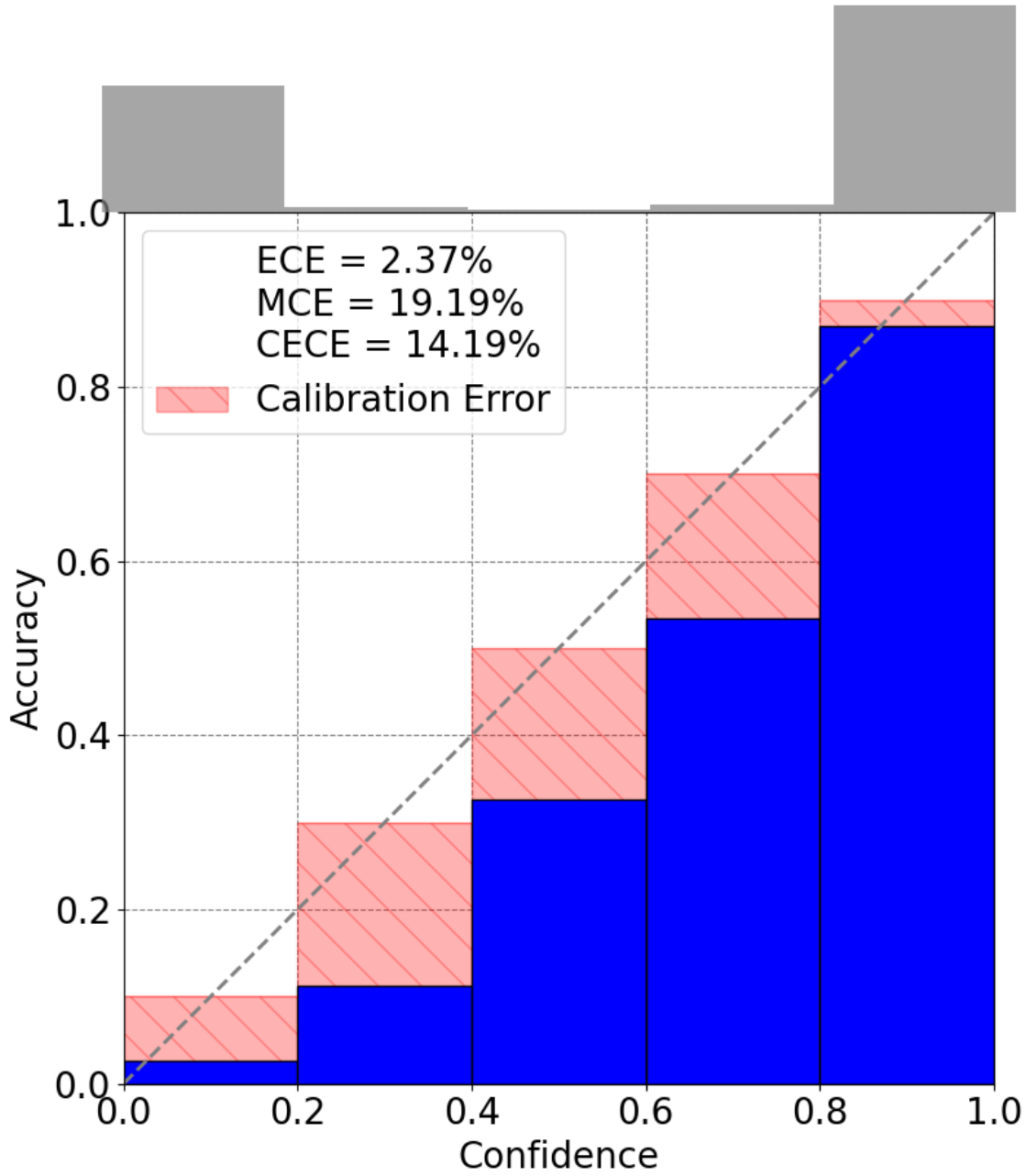

Vanilla Loss Binding Reliability
ECE = 5.84%
MCE = 57.91%
CECE = 32.79%
Calibration Error
Accuracy
Confidence
0.0
0.2
0.4
0.6
0.8
1.0


Focal Loss Binding Reliability
ECE = 7.21%
MCE = 32.74%
CECE = 21.36%
Calibration Error
Accuracy
Confidence
0.0
0.2
0.4
0.6
0.8
1.0

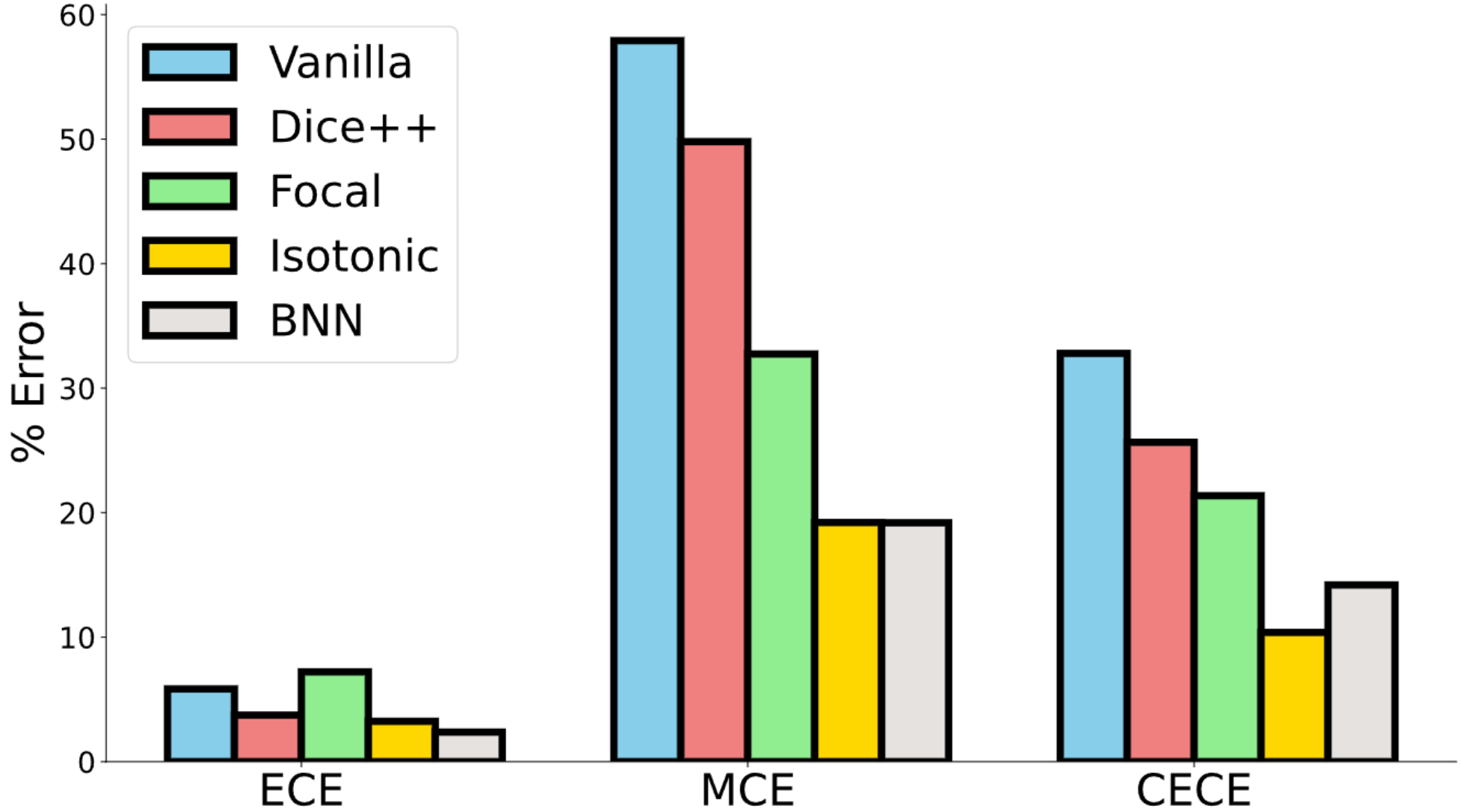

Vanilla Loss Binding Reliability
ECE = 5.84%
MCE = 57.91%
CECE = 32.79%
Calibration Error
Accuracy
Confidence
0.0
0.2
0.4
0.6
0.8
1.0


Isotonic Vanilla Binding Reliability
ECE = 3.24%
MCE = 19.21%
CECE = 10.38%
Calibration Error
Accuracy
Confidence
0.0
0.2
0.4
0.6
0.8
1.0

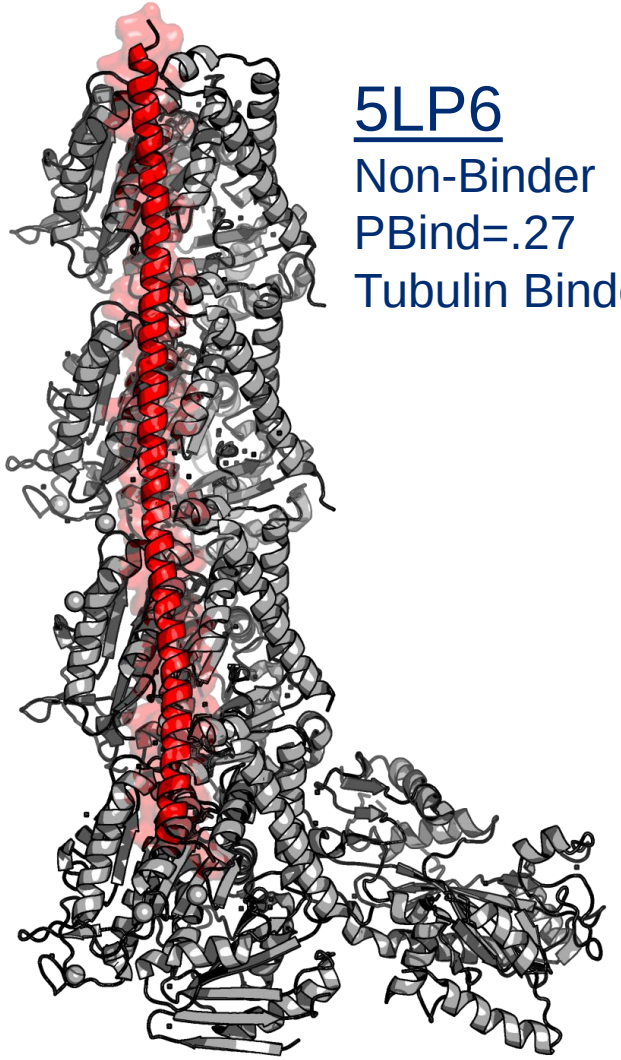

Vanilla Loss Binding Reliability
ECE = 5.84%
MCE = 57.91%
CECE = 32.79%
Calibration Error
Accuracy
Confidence
0.0
0.2
0.4
0.6
0.8
1.0


BNN n=100 Binding Reliability
ECE = 2.37%
MCE = 19.19%
CECE = 14.19%
Calibration Error
Accuracy
Confidence
0.0
0.2
0.4
0.6
0.8
1.0

# Comparison of Metrics



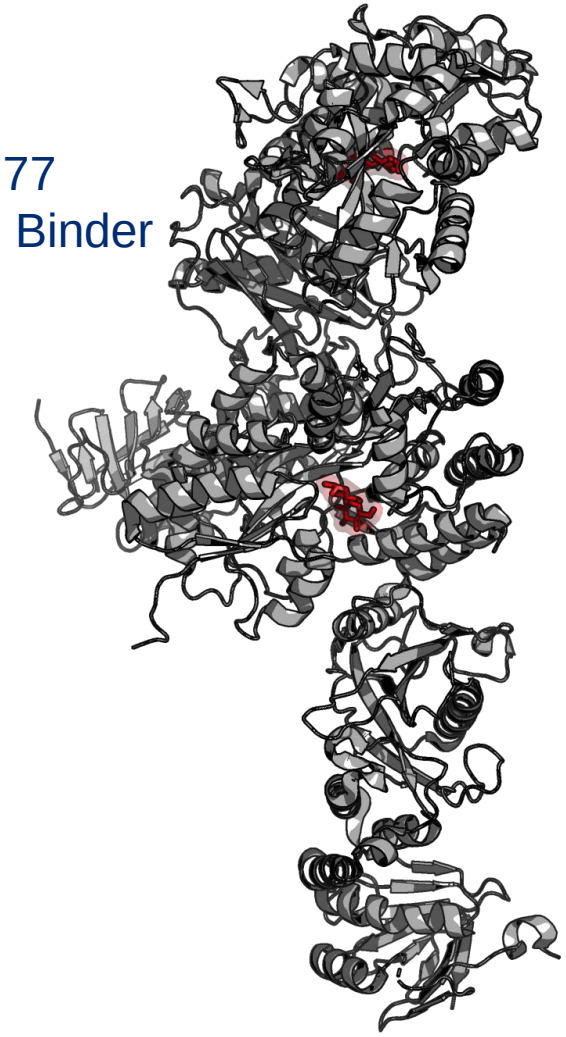
# Some Isotonic Predictions



5LP6
Non-Binder
PBind=.27
Tubulin Binder

6ZJW
Binder
PBind=.50
β-Galactosidase

4BL8
Binder
PBind=.77
Maltose Binder



# Thoughts and Conclusions

- Later in the training cycle, the more dramatic the effects are on model calibration
- Isotonic regression may be the way to go
- Still needs improvement

# Acknowledgements


- Samuel W Canner
- Dr. Jeffrey J Gray
- Gray Lab

# Thank you for listening

# Evaluating Protein-Carb Docking

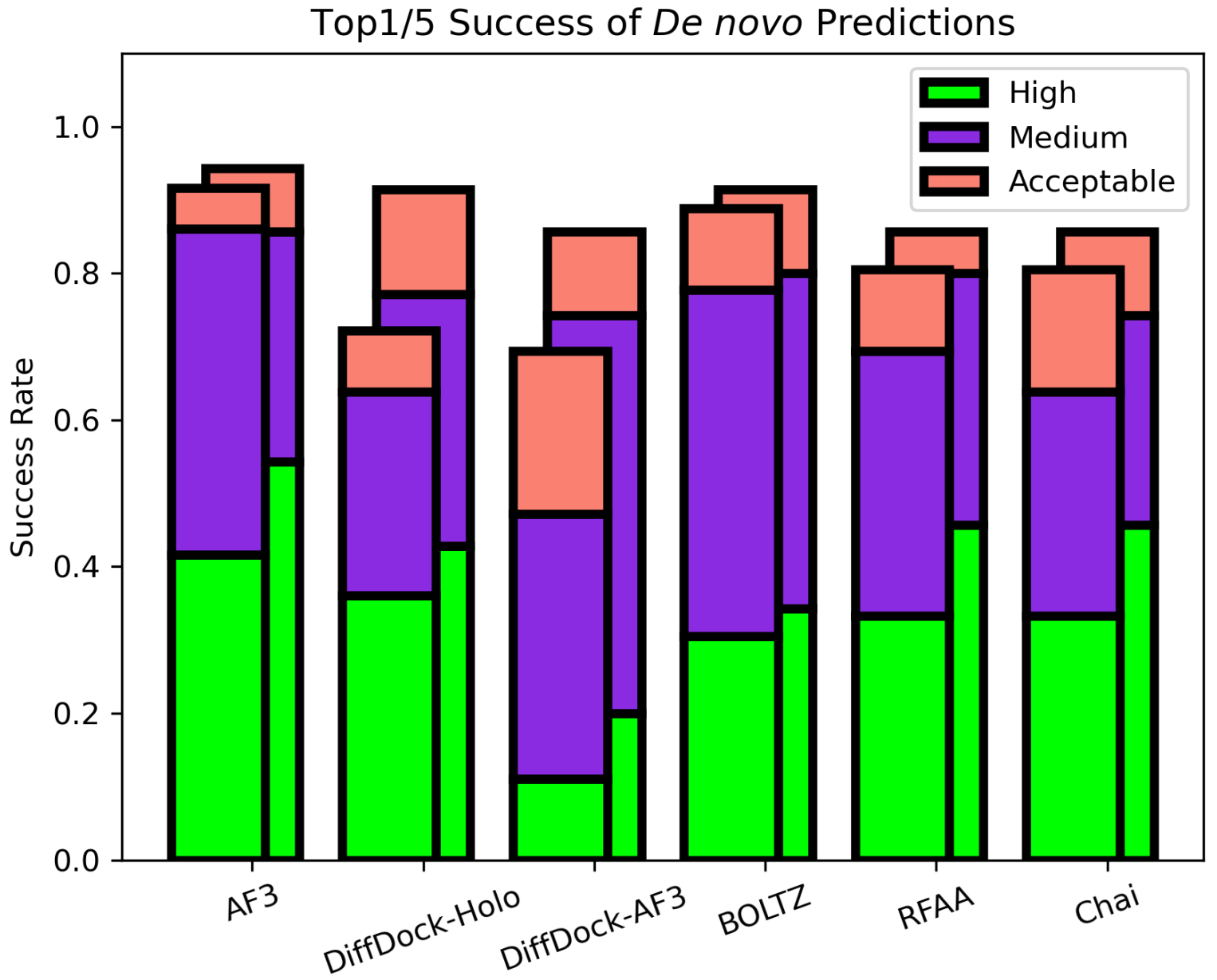
JOHNS HOPKINS
UNIVERSITY

# Models

Google DeepMind

Chai Discovery

# Results

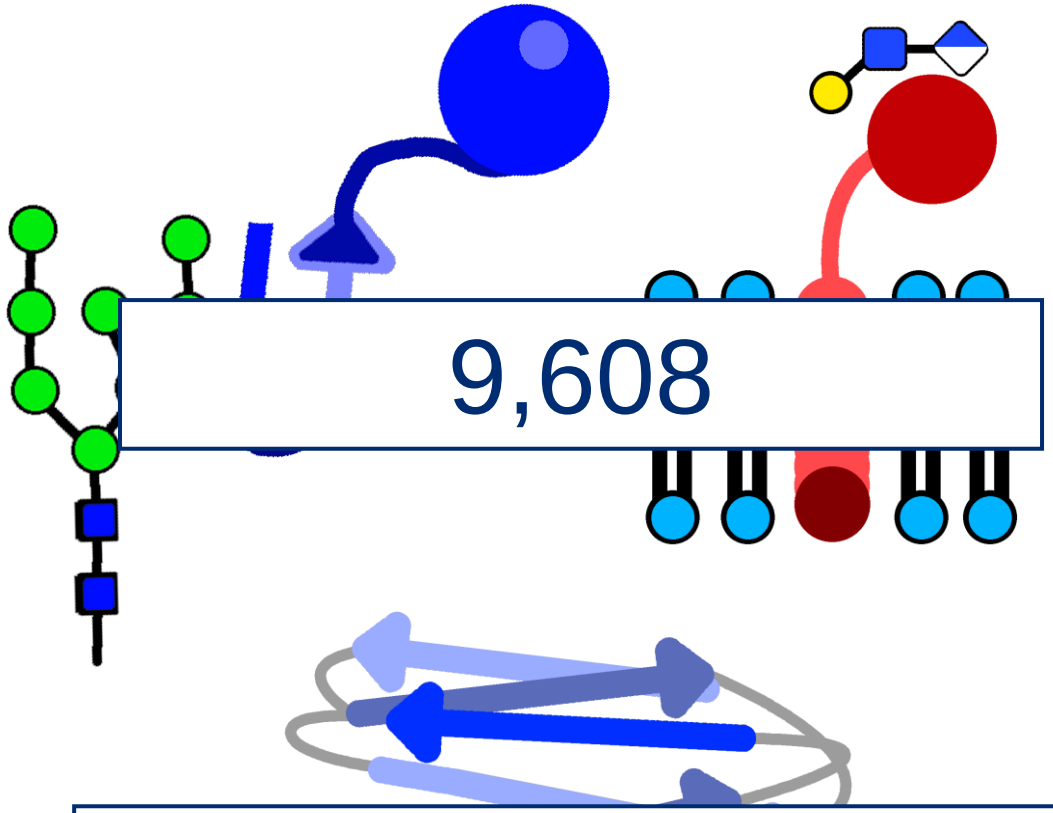



N=35 protein-carbohydrate complexes after cutoff date

# Supplemental Slides

# Nonbinder and binder of CArbohydrate Protein interactions dataset (NoCAP) characterizes proteins based on biological intuition

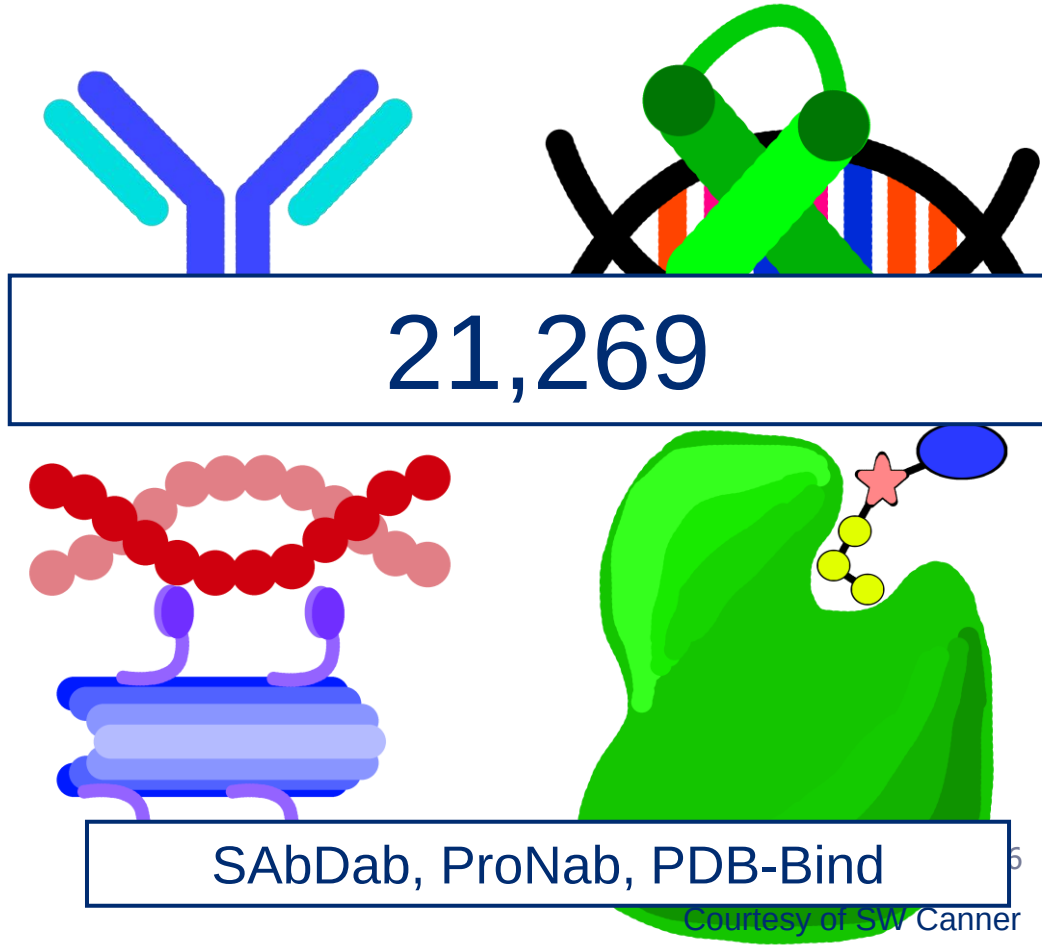
## Binders



9,608

PDB subsets: UniLectin, CAPSIF, DIONYSUS

## Non-binders



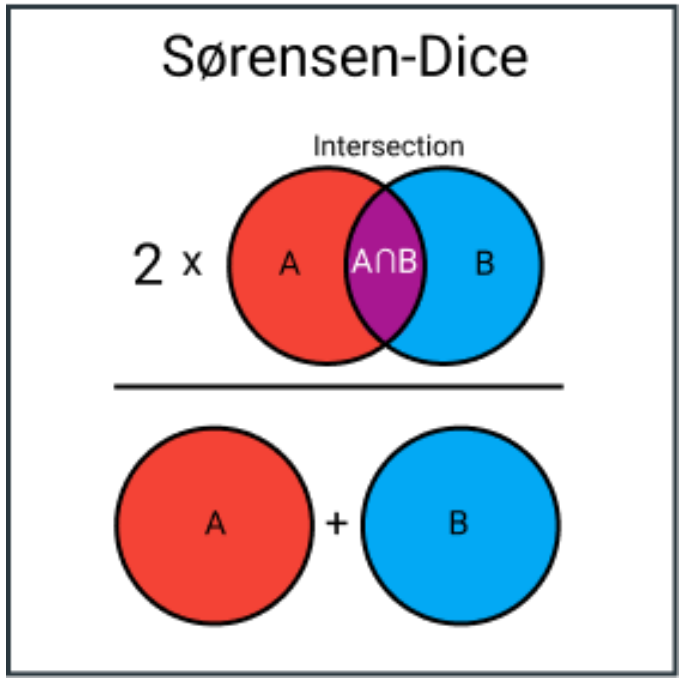
21,269

SAbDab, ProNab, PDB-Bind

Courtesy of SW Canner

# Vanilla PiCAP

- Pre-training: Dice loss
- Fine-tuning: Binary Cross Entropy (BCE) loss



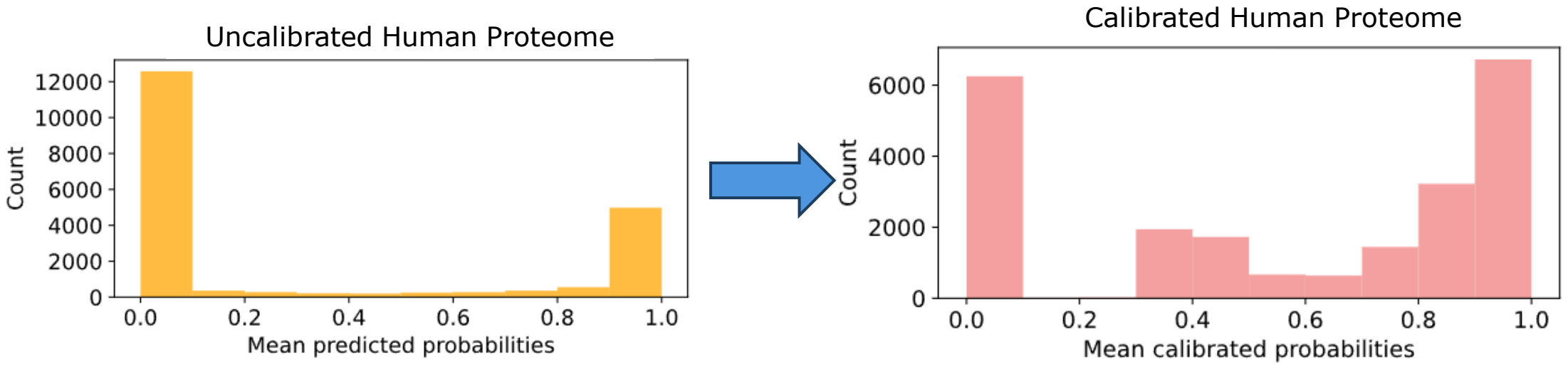

$$\mathcal{L}_{Dice} = 1 - \frac{2TP}{2TP + FP + FN}$$

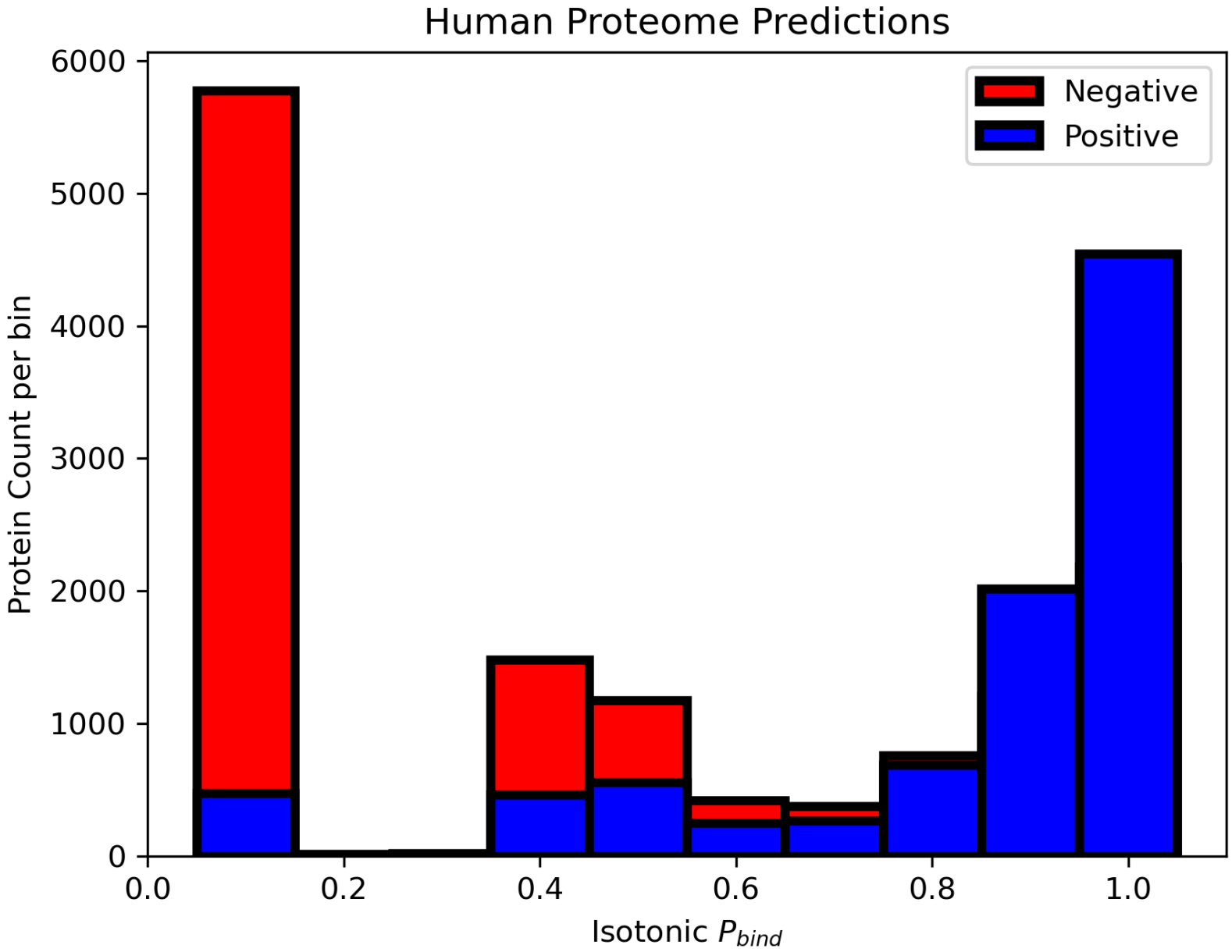
$$\mathcal{L}_{BCE} = -log(P_t)$$

$P_t$ = probability of predicting the ground truth class

# Calibration is a technique that smooths binary classifications to a scalar distribution

- Predictions reflect confidence
- If there are 100 predictions of .8 probability, 80 of the predictions should be true
- Prevents over-fitting and smooths distribution




Human Proteome Predictions
Negative
Positive
Protein Count per bin
6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
Isotonic $P_{bind}$