Confidence and Dispersity Speak: Characterizing Prediction Matrix for Unsupervised Accuracy Estimation

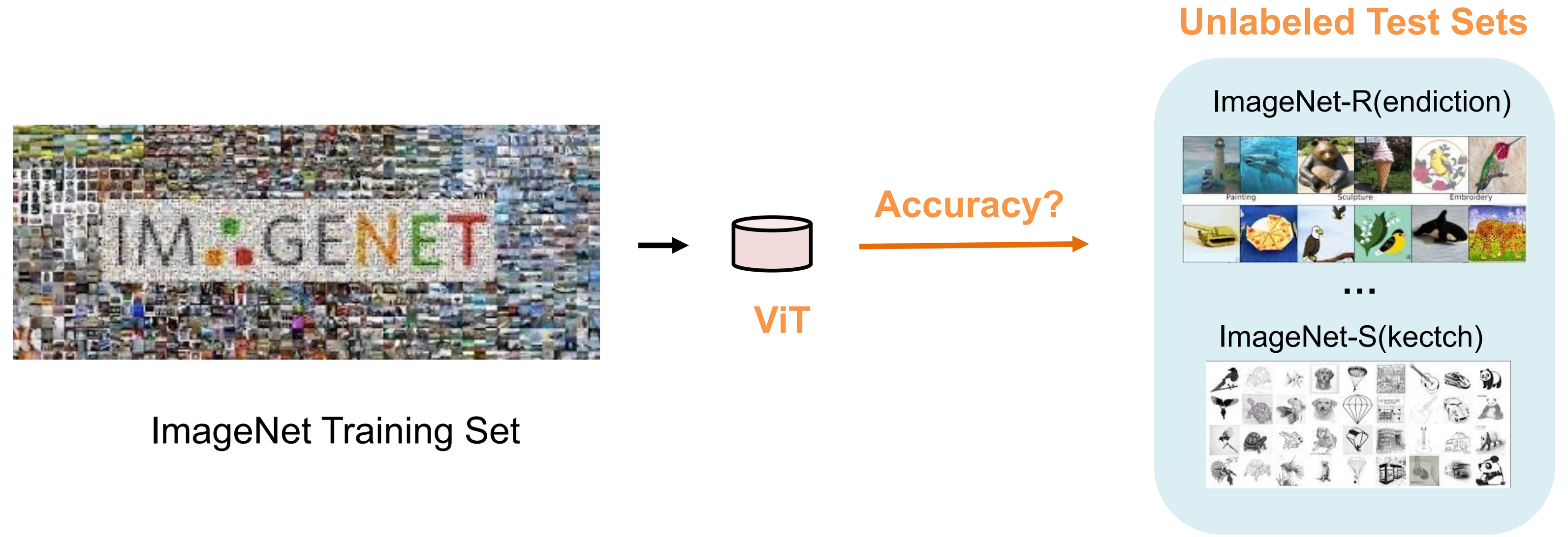
Australian National University
NEC
NEC Laboratories America

Weijian Deng[1] Yumin Suh[2] Stephen Gould[1] Liang Zheng[1]
[1]Australian National University [2]NEC Labs America

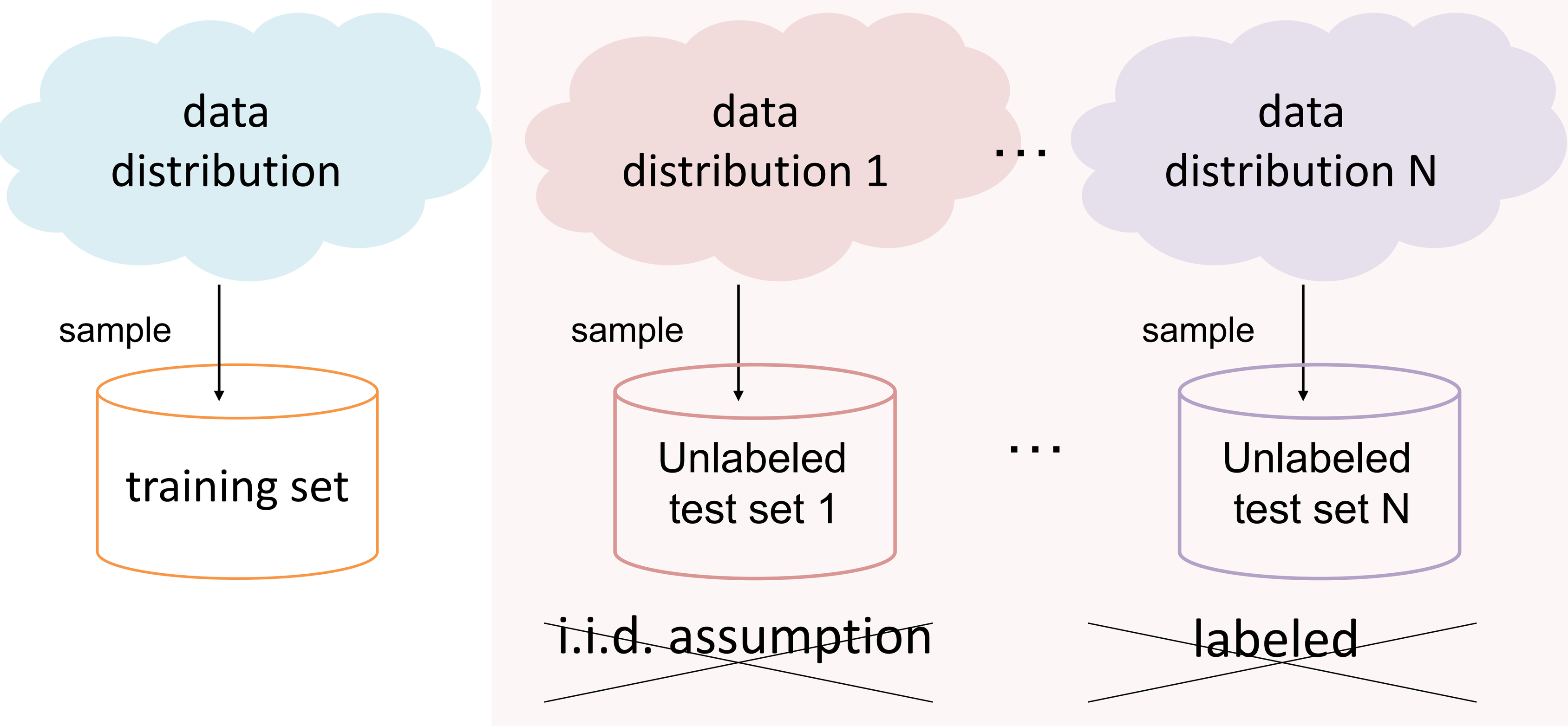
ICML 40 Years International Conference On Machine Learning

Unsupervised Accuracy Estimation

- **Definition**: given a trained model, the goal is to estimate its accuracy on various test datasets **without labels**



Real-world evaluation: 1) the distributions of test sets are often *different* from that of training set (*no i.i.d*); 2) test labels are *unavailable* or *expensive to obtain*.

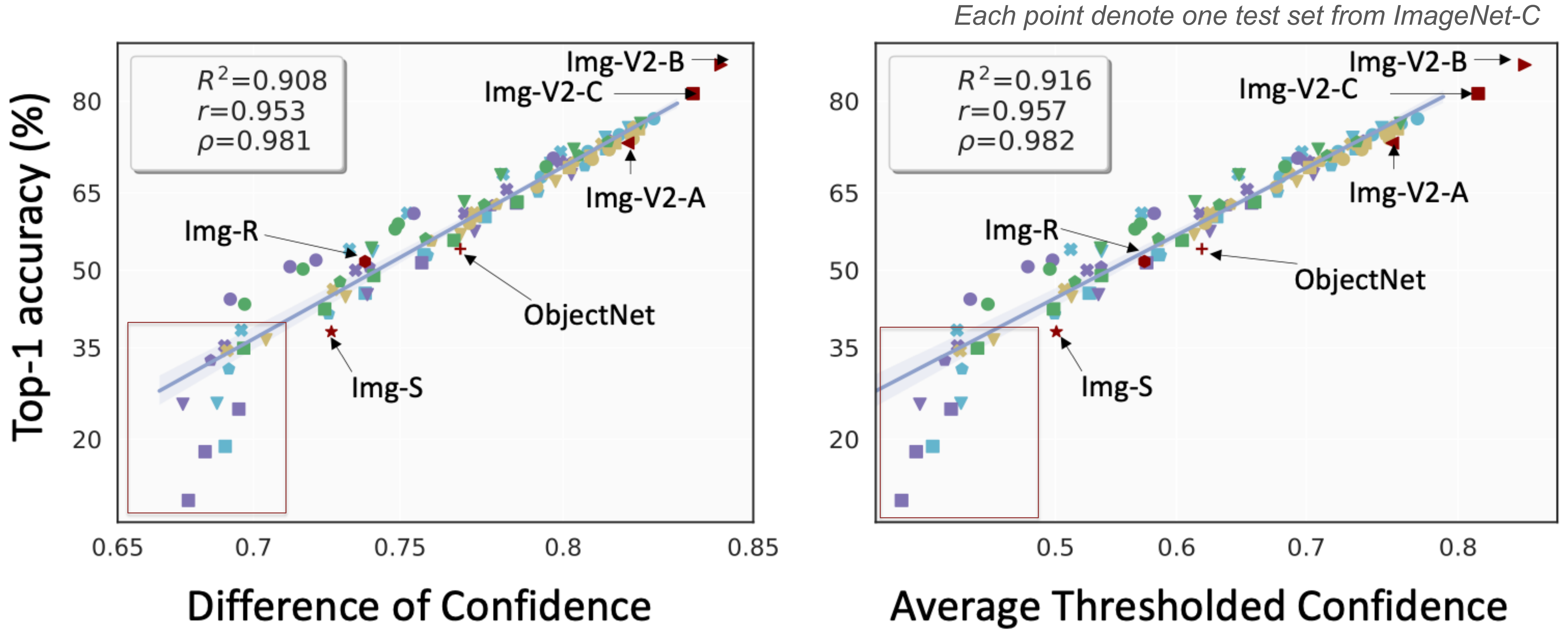


In-distribution accuracy may only be a weak predictor of performance on out-of-distribution data;
Evaluation without labels and under distribution shifts

Prediction Confidence

- **Confidence** reflects whether the individual prediction is certain
  Existing methods (*e.g.*, DoC and ATC) explore such information

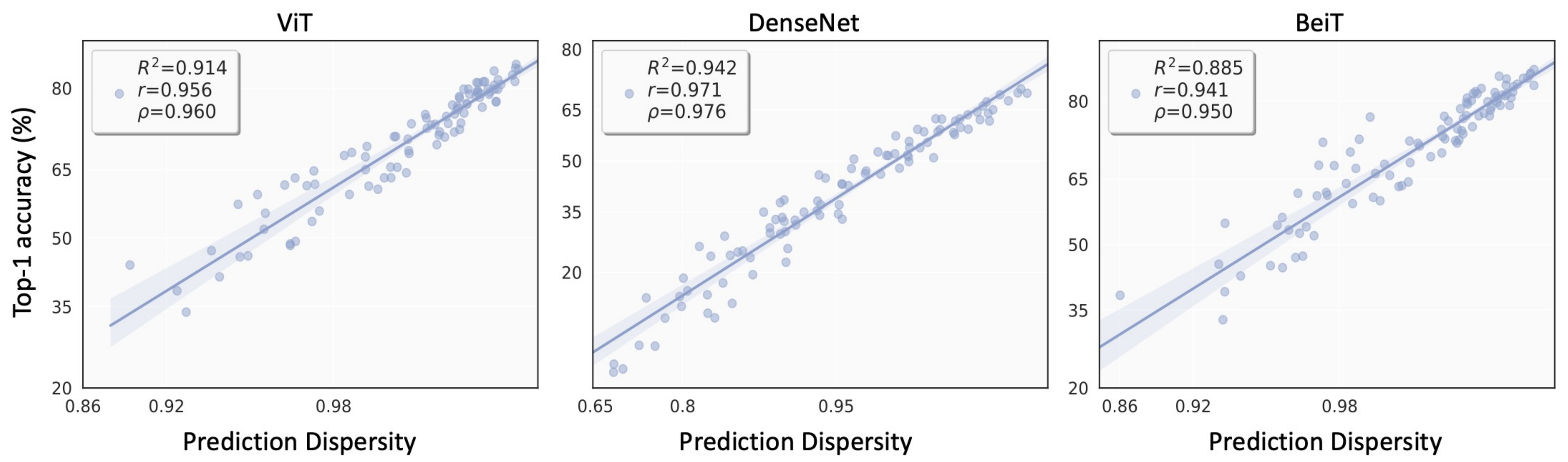
Each point denote one test set from ImageNet-C



- **Confidence may be a weak indicator**
  Predication score-based methods **cannot well capture** the test sets in the **low-accuracy region** (bottom-left area of the above correlation figure)

Prediction Dispersity

- **Dispersity** indicates how the predictions are distributed across all categories
  Prediction Dispersity Score: we first calculate the histogram of the number of the predicted class and then use entropy to measure **the degree of balance**

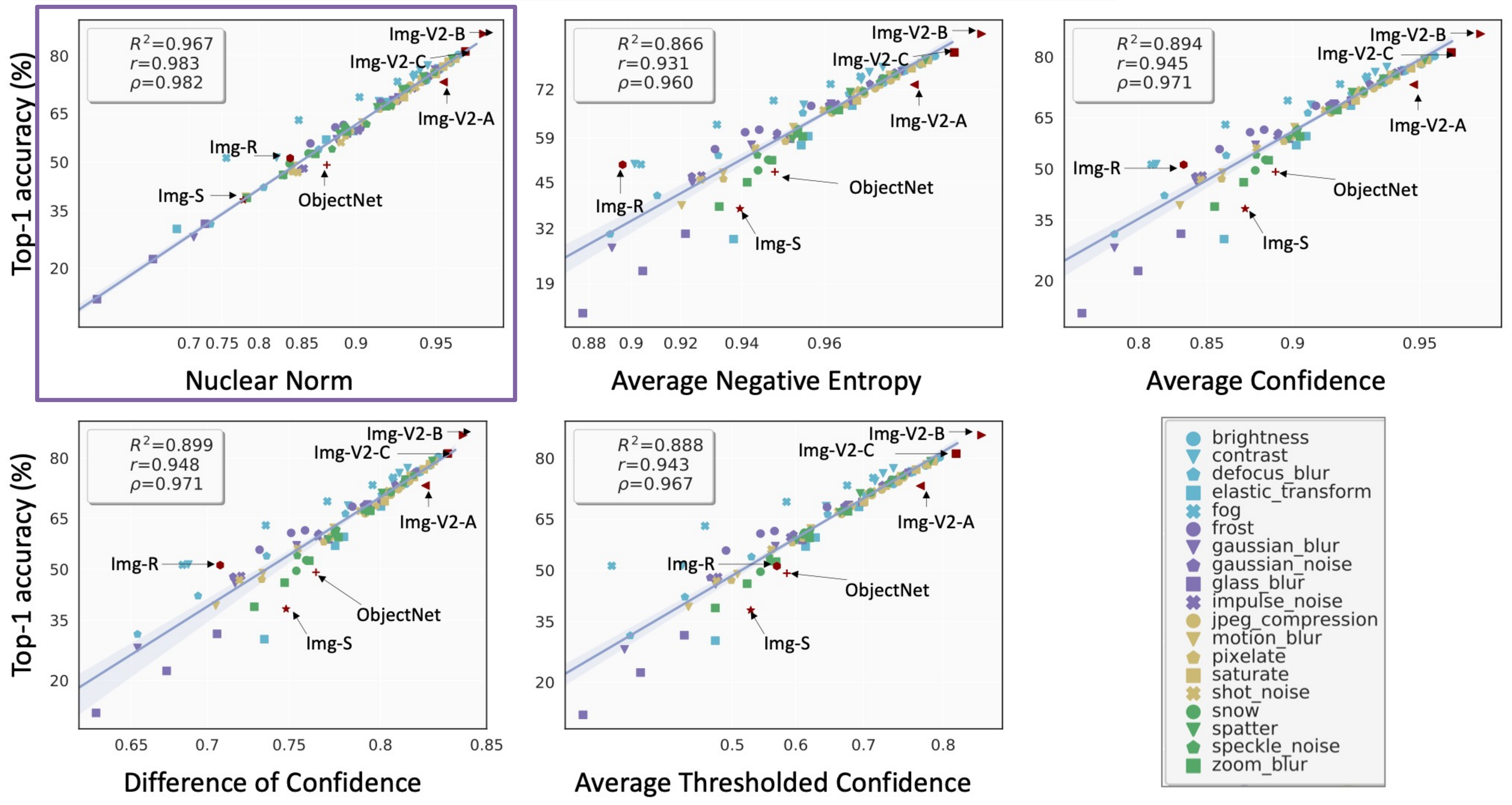


Nuclear Norm

- **Nuclear norm** is effective in characterizing both confidence and dispersity

Prediction Matrix $P \in \mathbb{R}^{N_t \times K}$ (N_t test samples, and K classes)

Nuclear Norm: the sum of singular values of prediction matrix



Nuclear norm exhibits the highest correlation strength with OOD accuracy

Potential Direction

1) Other methods are stable under class imbalance;
2) Nuclear Norm **is resistant to moderate class imbalance;**
3) Nuclear Norm **is less effective under severe class imbalance.**

If we have **prior knowledge** about the imbalanced class distribution, we can expect class predictions to follow it rather than a uniform one