PURDUE UNIVERSITY

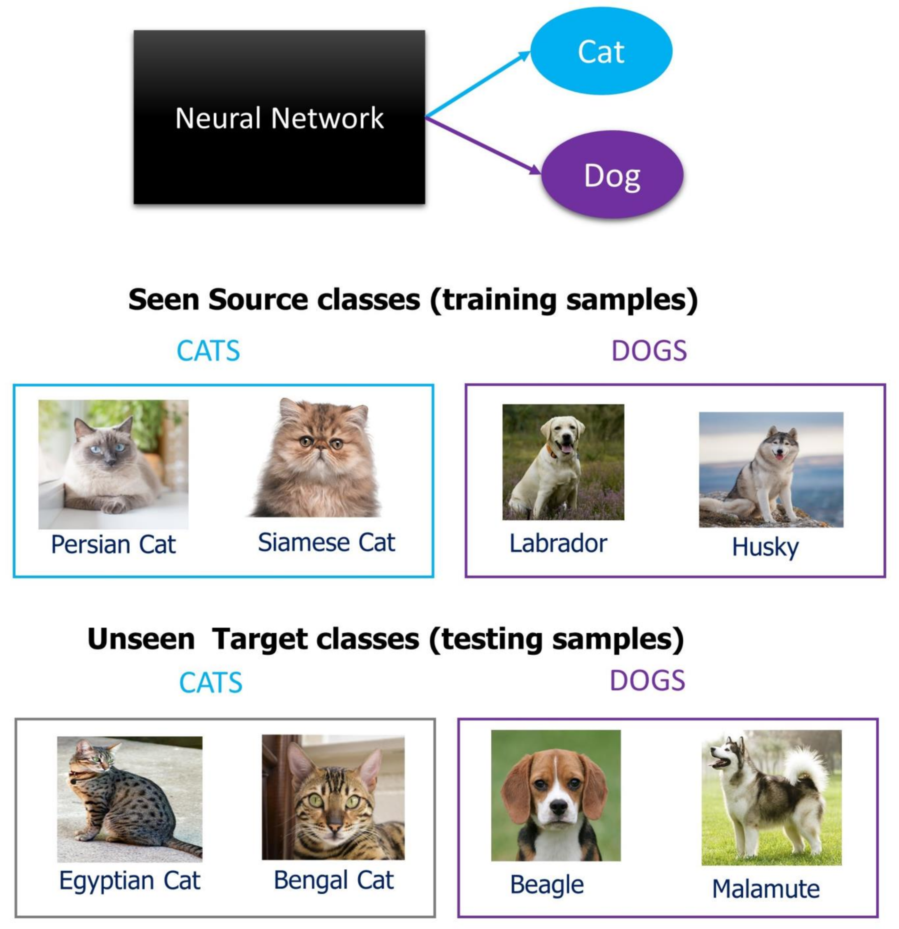
C-BRIC

# Encoding Hierarchical Information in Neural Networks Helps in Subpopulation Shift

Amitangshu Mukherjee*, Isha Garg*, Kaushik Roy

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

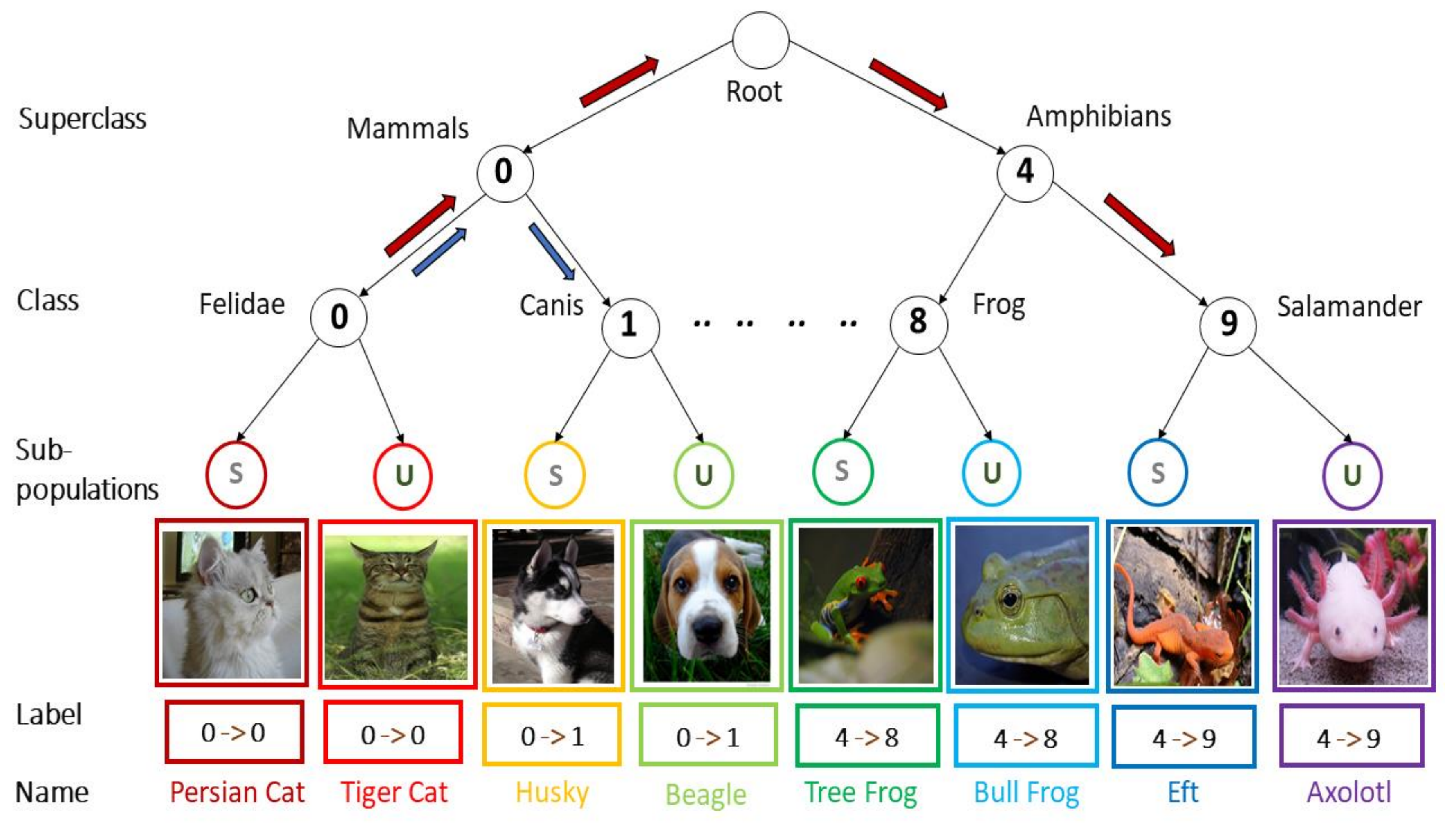
## 1. Subpopulation Shift and Proposed Method




- All subpopulations (such as different breeds of dogs) are not represented in the training dataset
- The learned model is unable to generalize to newer subpopulations (breeds) not seen during training [1]
- Subpopulation Shift implicitly involves the notion of hierarchy. We propose to explicitly learn representations that are trained with hierarchical labels
- Conditional Training allows for heads at different levels to collaborate
- Hierarchical Trees enable quantification of the impact of mispredictions via graph traversal length between true and incorrectly predicted labels

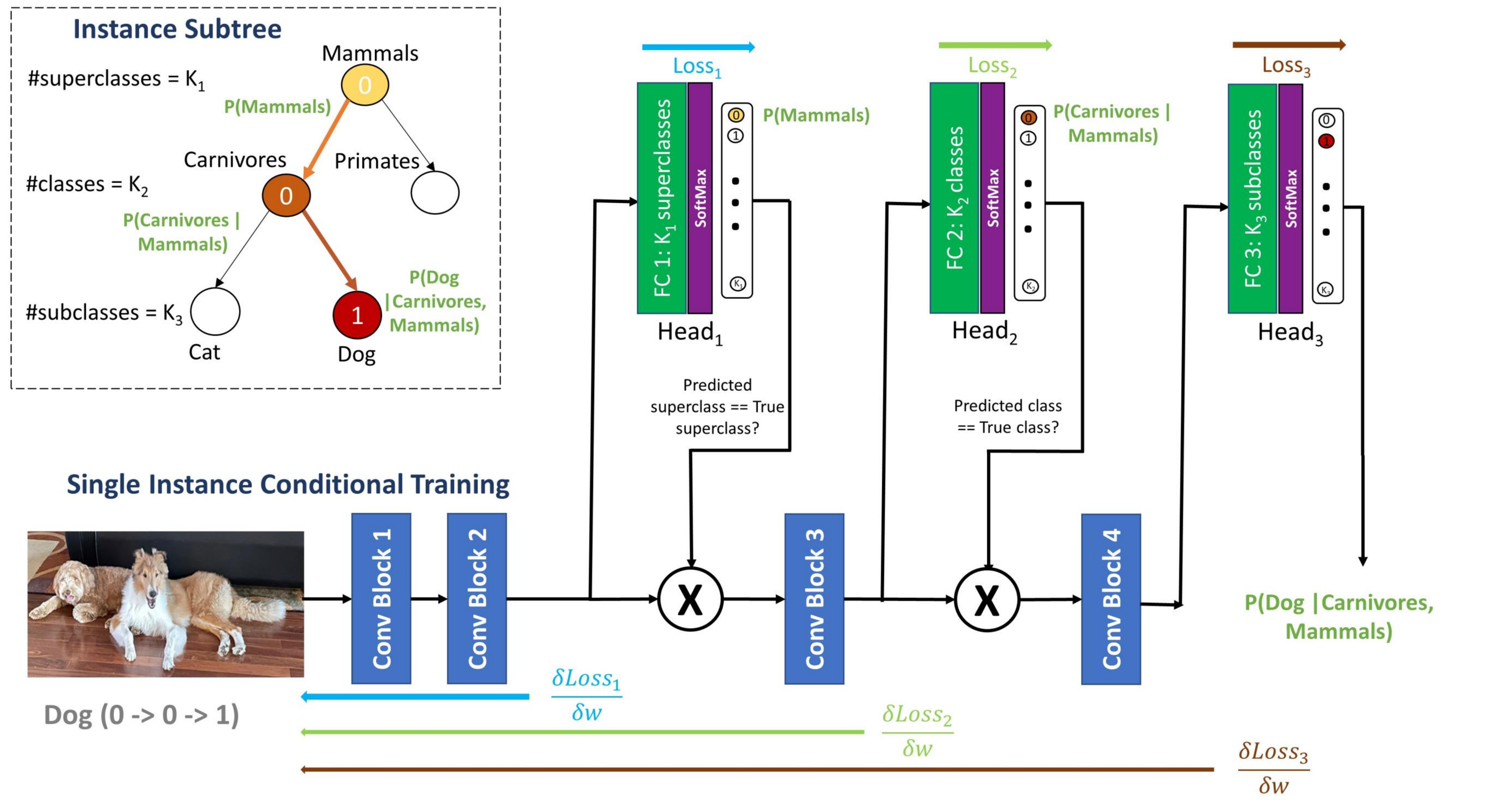
Subpopulation shift is the problem of not generalizing to unseen subpopulations of classes

## 2. Taxonomical Hierarchy and Catastrophic Coefficient



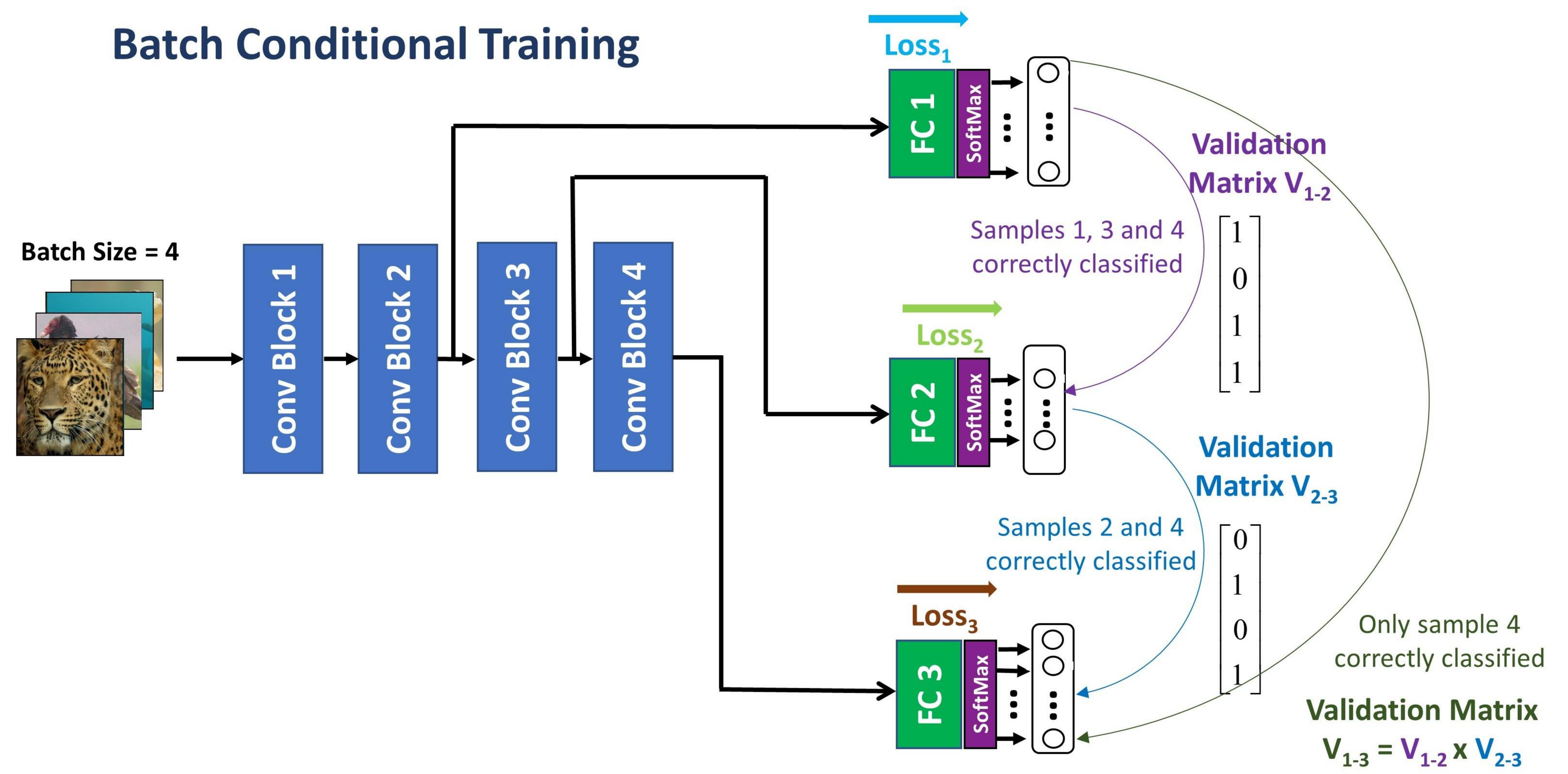

Mispredicting a Cat as a Dog has an impact of 2, vs mispredicting it as a Salamander has an impact of 4

## 3. Enabling Collaboration Between Levels




Level specific heads pass only correctly classified instances for learning to the next levels

## 4. Conditional Training Through Validity Matrices




Conditional-$Loss_1$ = $Loss_1$

Conditional-$Loss_2$ = $Loss_2$ x $V_{1-2}$

Conditional-$Loss_3$ = $Loss_3$ x $V_{1-3}$ = $Loss_3$ x $V_{2-3}$ x $V_{1-2}$

Validation Matrix

$V_{1-3} = V_{1-2} \times V_{2-3}$

$$\begin{bmatrix}1\\0\\1\\1\end{bmatrix} * \begin{bmatrix}0\\1\\0\\1\end{bmatrix} = \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

$V_{l1-l2}$ represent all the samples that are correctly classified between all the levels between l1 and l2

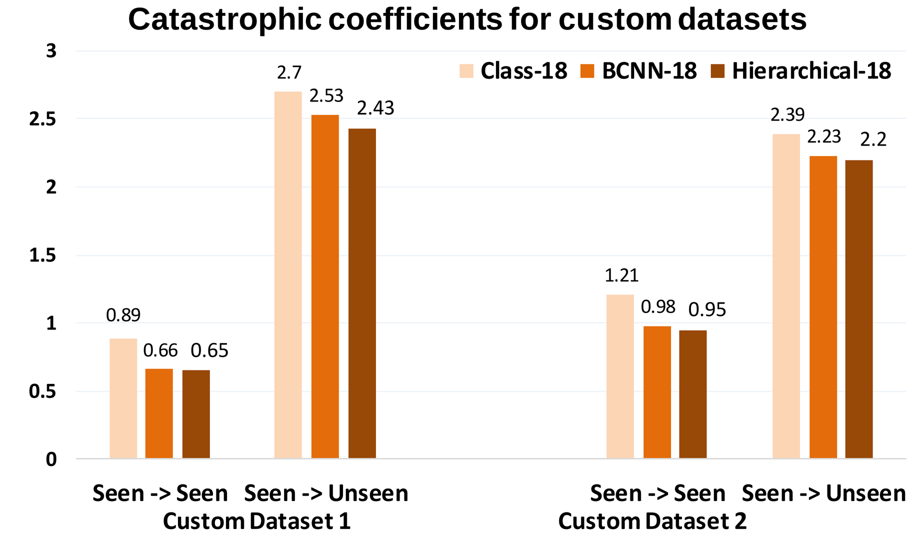
## 5. Results

### Custom Datasets 1 and 2

Accuracy results for custom datasets (%)

| Model | Seen -> Seen (1) | Seen -> Unseen (1) | Seen -> Seen (2) | Seen -> Unseen (2) |
|---|---|---|---|---|
| Baseline-18 | 84.8 | 49.53 | 77.7 | 55.47 |
| BCNN-18 | 87.84 | 51.64 | 81.8 | 58.53 |
| Hierarchical-18 | **88.13** | **53.92** | **82.31** | **59.28** |

Catastrophic coefficients for custom datasets

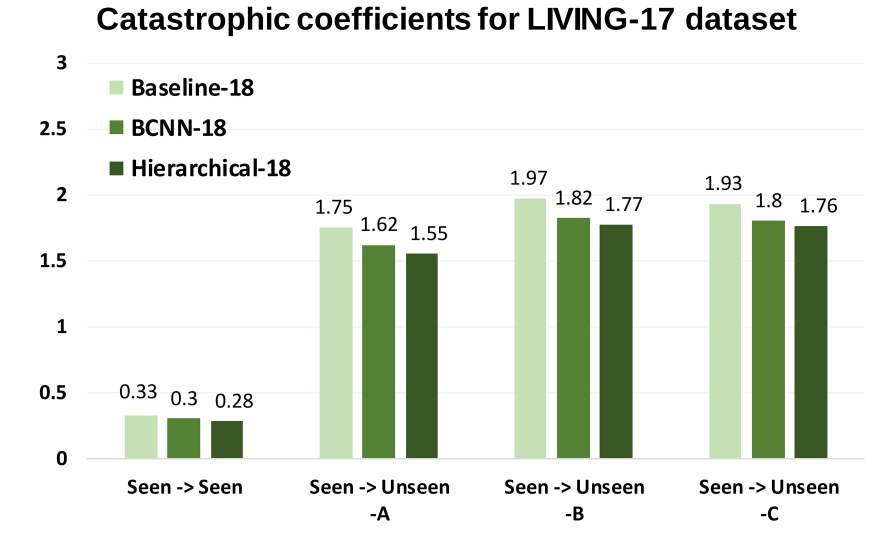



### LIVING-17 –A, -B and –C Dataset

Accuracy results for LIVING-17 dataset (%)

| Model | Seen -> Seen | Seen -> Unseen –A | Seen -> Unseen –B | Seen -> Unseen –C |
|---|---|---|---|---|
| Baseline-18 | 92.3 | 57.02 | 53.54 | 53.04 |
| BCNN-18 | 93.02 | 58.8 | 55.66 | 55.1 |
| Hierarchical-18 | 93.17 | 60.53 | 56.6 | 55.62 |

Catastrophic coefficients for LIVING-17 dataset




- Baseline-18 : Resnet-18 model trained in a traditional supervised manner.
- BCNN-18 : Modified Resnet-18 trained via Branch Training Strategy
- Hierarchical-18 : Modified Resnet-18 trained conditionally via the approach explained
  [All models are trained till the level of classes on 5 random seeds with mean reported]

## 6. Summary

- The problem of subpopulation shift is mitigated by a novel conditional hierarchical training framework
- The implicit notion of hierarchy in subpopulation shift problem framing is made explicit by encoding the hierarchical information in class labels
- We train modified versions of standard networks on the hierarchical labels while enabling conditional label propagation to encourage collaboration between different levels of hierarchy
- We show that our method outperforms standard baselines by up to 3% in terms of accuracy and 11% in terms of catastrophic co-efficient on the LIVING-17 subpopulation shift benchmark
- Scan the QR code for more results outlined in our paper on ArXiV

This work was supported in part by the Center for Brain Inspired Computing (C-BRIC), one of the six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA, by the Semiconductor Research Corporation GRC grant, the National Science Foundation, Intel Corporation, the DoD Vannevar Bush Fellowship, and by the U.S. Army Research Laboratory

[1] S. Santurkar, D. Tsipras, & A. Madry (2021). BREEDS: Benchmarks for Subpopulation Shift. In *International Conference on Learning Representations*.