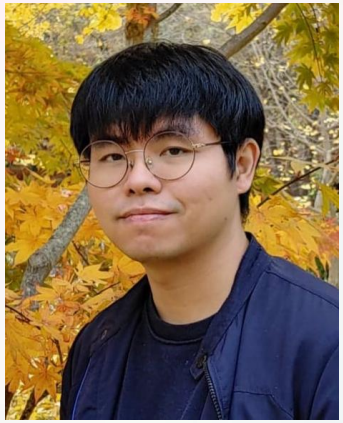
UBC Robotics & Control Laboratory

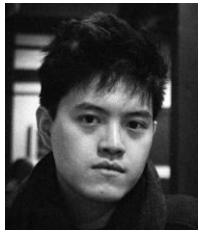
ICML International Conference On Machine Learning

DPE

Diverse Prototypical Ensembles Improve Robustness to Subpopulation Shift

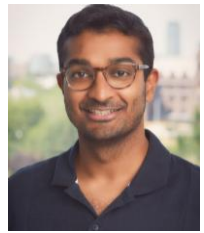


Minh To



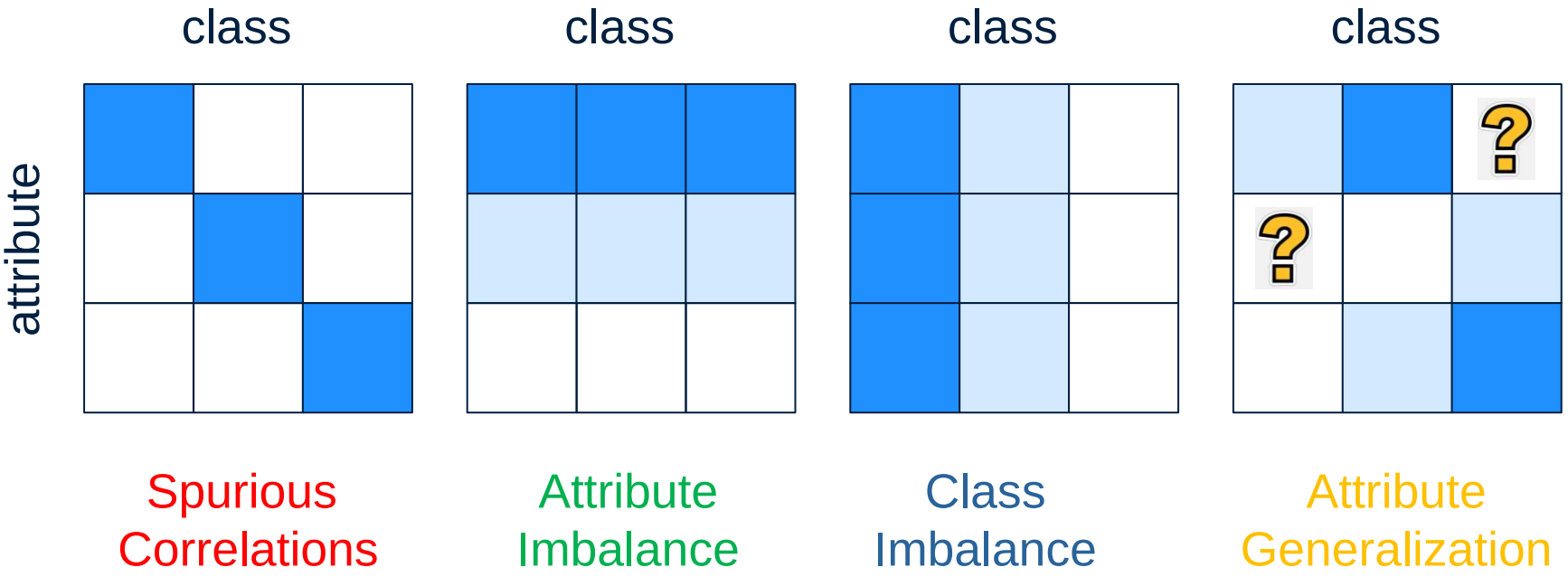




UBC · Queen's University · CIFAR · Vector Institute

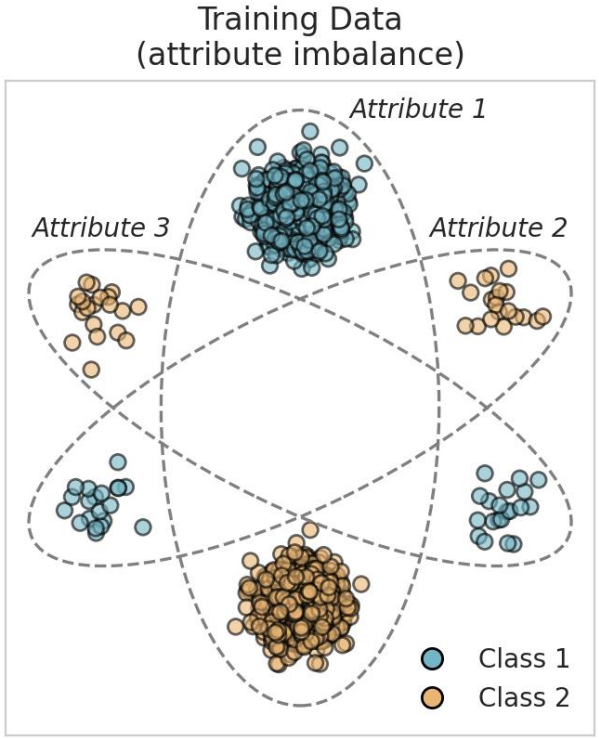
Subpopulation Shifts

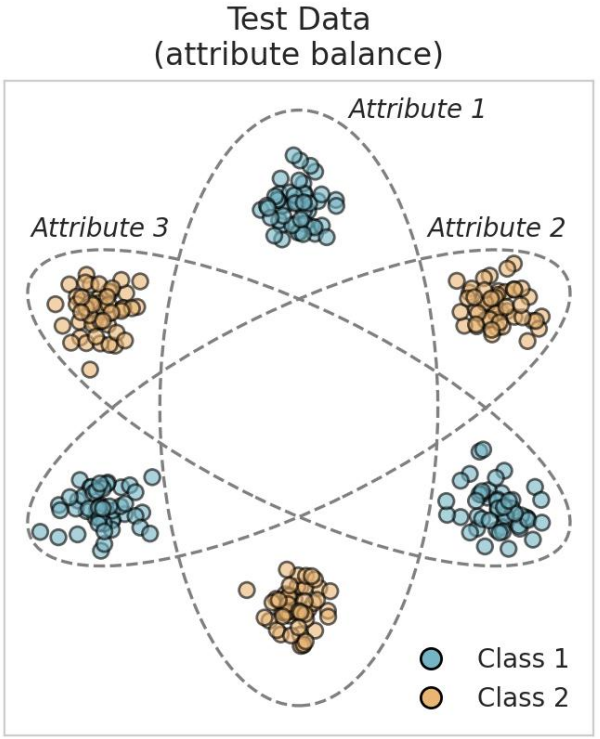
Subpopulation shift = distribution mismatch between training and test groups.



* *Slide adapted from* ***Change is Hard: A Closer Look at Subpopulation Shift (Yang et al.)***

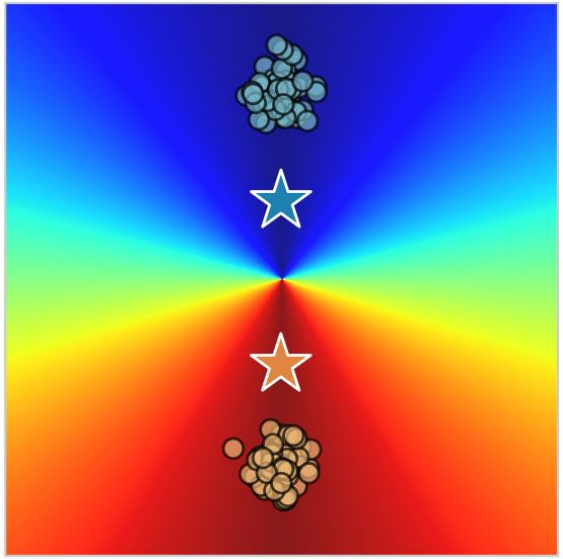
Motivation


Training Data
(attribute imbalance)
Attribute 1
Attribute 3
Attribute 2
Class 1
Class 2

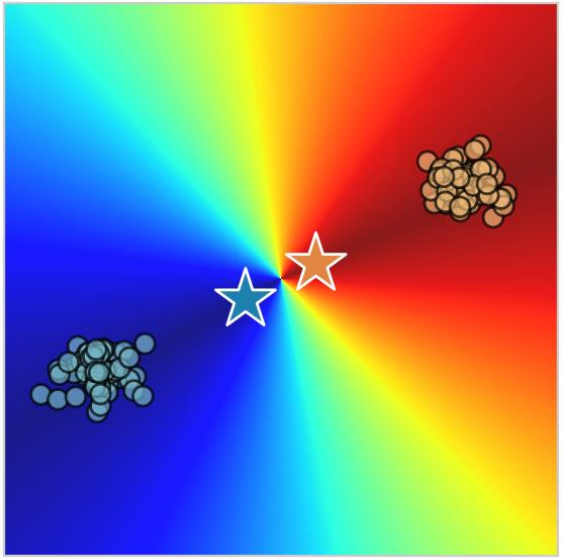

Test Data
(attribute balance)
Attribute 1
Attribute 3
Attribute 2
Class 1
Class 2

Motivation

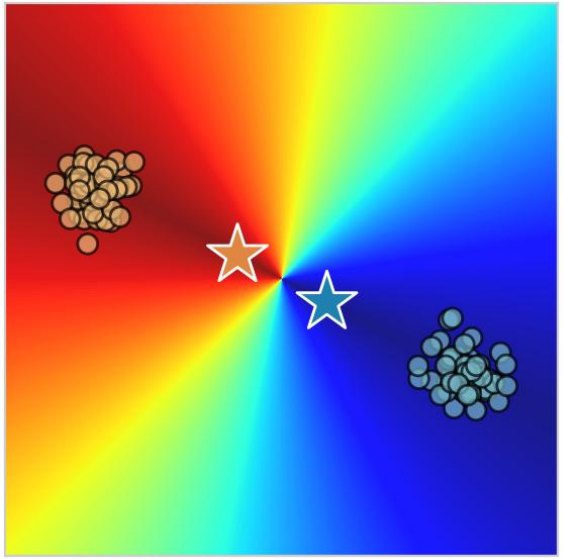
Distance to Class 1 Prototype
(Test set - Attribute 1)



Distance to Class 1 Prototype
(Test set - Attribute 2)

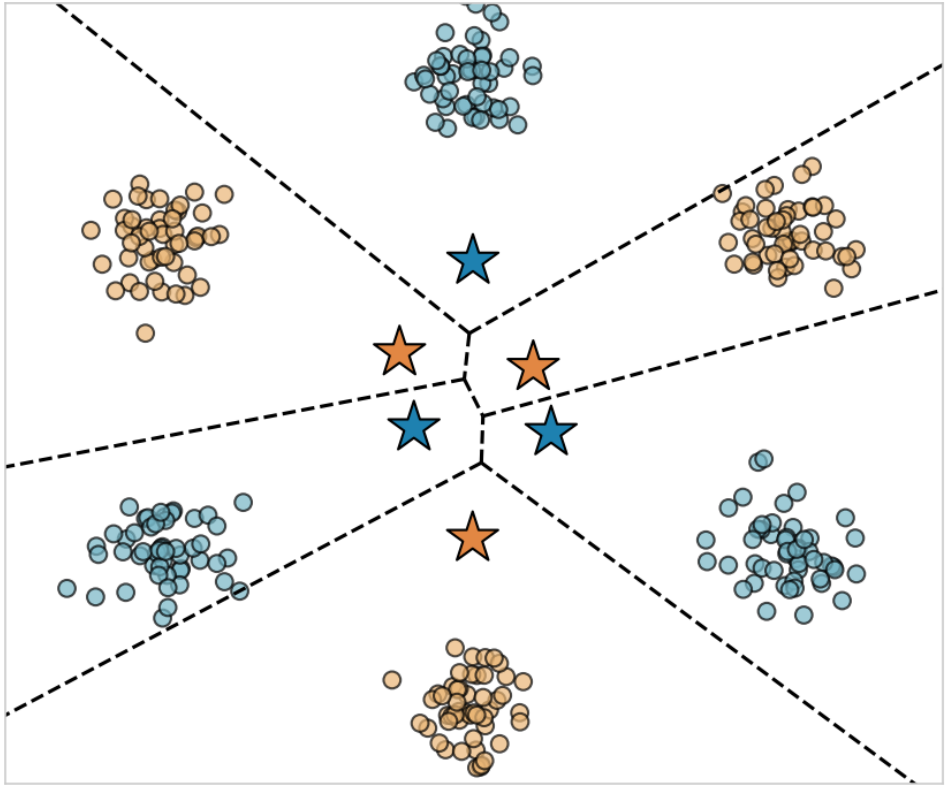


Distance to Class 1 Prototype
(Test set - Attribute 3)

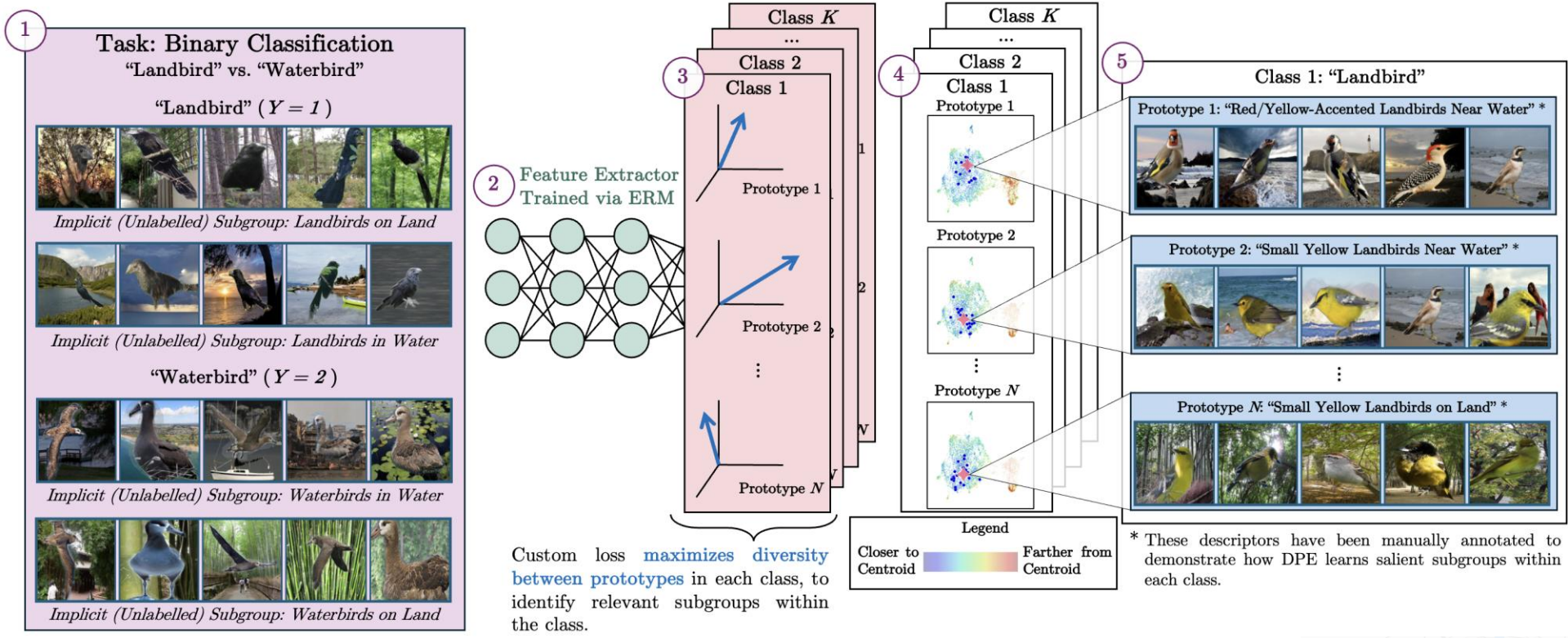


Motivation

Voronoi Diagram

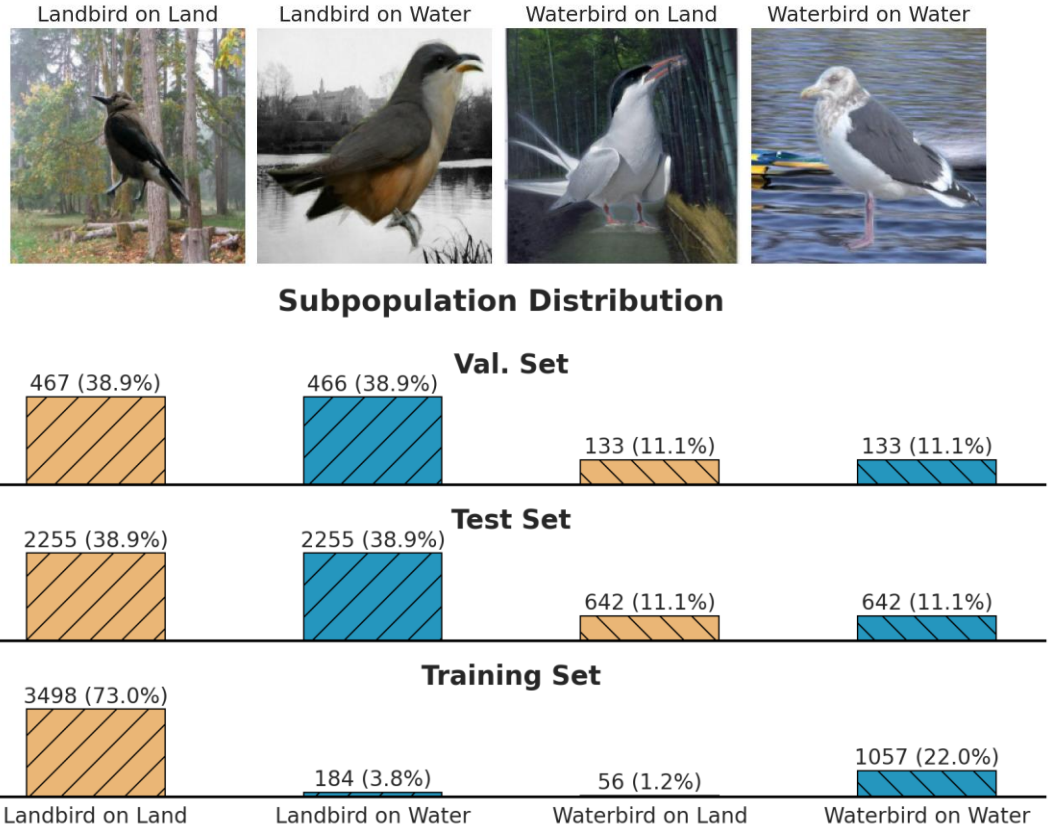


Diversified Prototypical Ensemble



Datasets

- Waterbirds – Standard Benchmark Dataset
- Spurious correlation between the bird types and the background
- And other 8 Real-world datasets



Worst-Group Accuracy (With Subgroup Annotations)

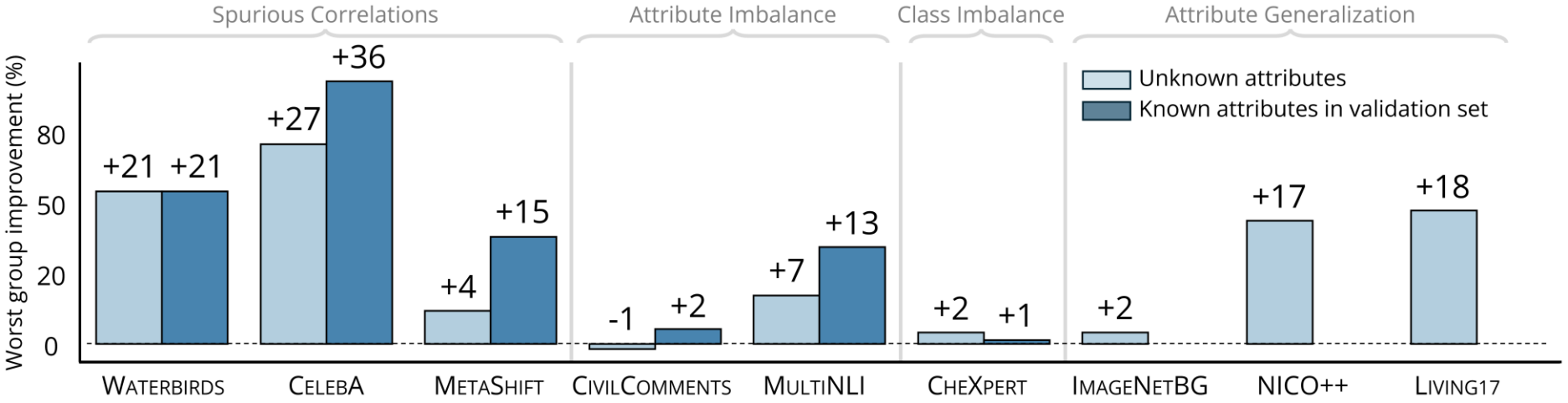
Algorithm	Group Info (Train/Val)	Waterbirds	CelebA	CivilComments	MultiNLI	MetaShift	CheXpert
ERM	✗/✗	69.1 ±4.7	57.6 ±0.8	63.2 ±1.2	**66.4** ±2.3	82.1 ±0.8	41.7 ±3.4
CRT	✗/✓	76.3 ±0.8	70.4 ±0.4	**68.5** ±0.0	65.4 ±0.1	83.1 ±0.0	**74.0** ±0.2
ReWeightCRT	✗/✓	76.3 ±0.2	71.1 ±0.5	68.2 ±0.4	65.3 ±0.1	**85.1** ±0.4	73.9 ±0.2
DFR	✗/✓✓	**89.0** ±0.2	**86.3** ±0.3	66.5 ±0.2	63.8 ±0.0	81.5 ±0.0	**75.4** ±0.6
ERM + DPE	✗/✓✓	**91.0** ±0.4	**87.7** ±0.6	**71.5** ±0.6	**74.8** ±0.3	**87.9** ±0.7	–
ERM*	✗/✗	77.9 ±3.0	66.5 ±2.6	69.4 ±1.2	66.5 ±0.7	80.0 ±0.0	75.6 ±0.4
Group DRO	✓/✓	91.4 ±1.1	88.9 ±2.3	70.0 ±2.0	**77.7** ±1.4	–	–
RWG	✓/✓	87.6 ±1.6	84.3 ±1.8	**72.0** ±1.9	69.6 ±1.0	–	–
JTT	✗/✓	86.7	81.1	69.3	72.6	–	–
CnC	✗/✓	88.5 ±0.3	88.8 ±0.9	68.9 ±2.1	–	–	–
SSA	✗/✓✓	89.0 ±0.6	89.8 ±1.3	69.9 ±2.0	76.6 ±0.7	–	–
DFR*	✗/✓✓	92.9 ±0.2	88.3 ±1.1	70.1 ±0.8	74.7 ±0.7	–	–
GAP (Last Layer)	✗/✓✓	93.2 ±0.2	**90.2** ±0.3	–	74.3 ±0.2	–	–
GAP (All Layer)	✗/✓✓	**93.8** ±0.1	**90.2** ±0.3	–	**77.8** ±0.6	–	–
ERM* + DPE	✗/✓✓	**94.1** ±0.4	**90.3** ±0.7	**70.8** ±0.8	75.3 ±0.5	**91.7** ±1.3	**76.0** ±0.3

Worst-Group Accuracy (Without Subgroup Annotations)

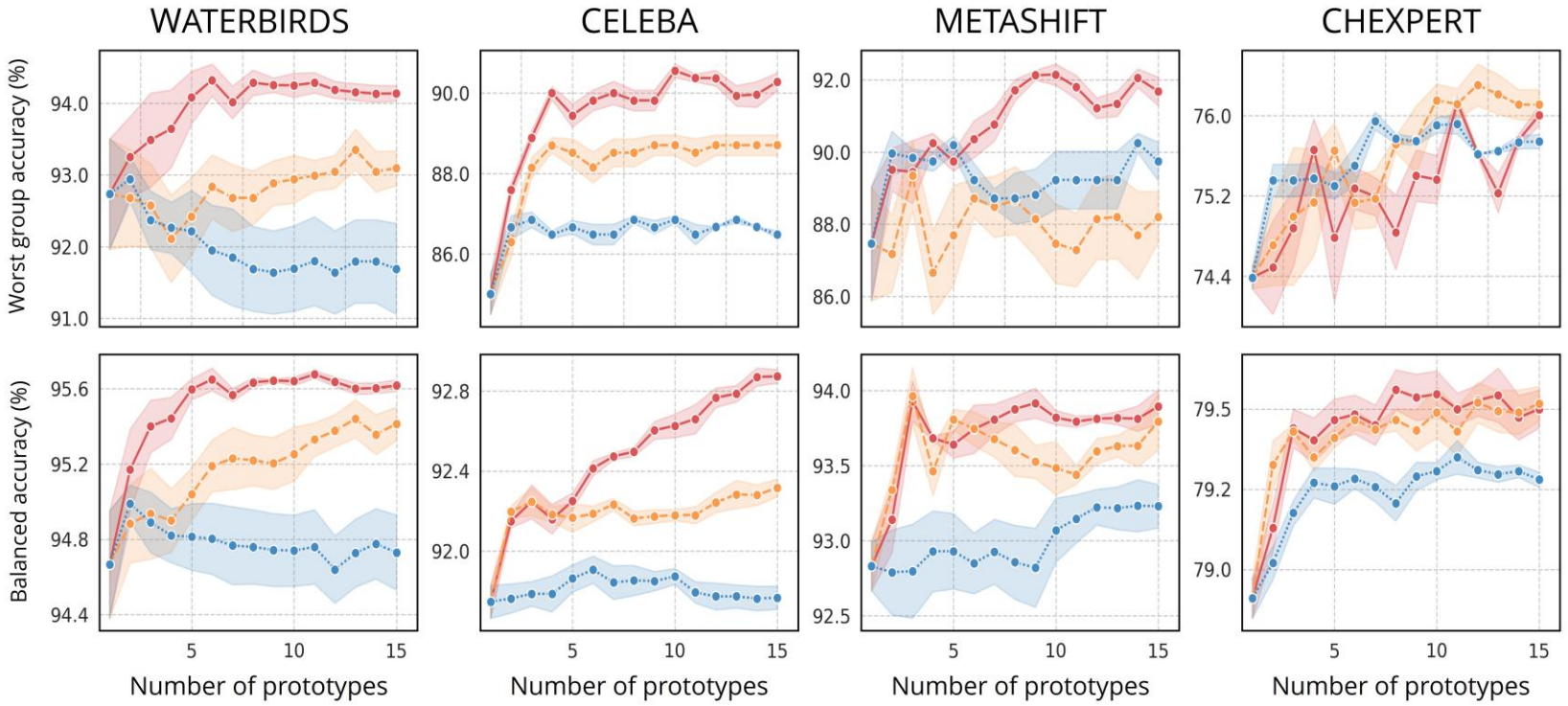
The top section presents SubpopBench-style baselines using an ERM backbone. The bottom section includes recent state-of-the-art methods and our method using a stronger ERM* backbone.

Algorithm	Waterbirds	CelebA	CivilComments	MultiNLI	MetaShift	CheXpert	ImageNetBG	NICO++	Living17
ERM	69.1 ±4.7	57.6 ±0.8	63.2 ±1.2	**66.4** ±2.3	82.1 ±0.8	41.7 ±3.4	76.8 ±0.9	35.0 ±4.1	48.0 ±1.5
CRT	76.3 ±0.8	69.6 ±0.7	**67.8** ±0.3	65.4 ±0.2	83.1 ±0.0	74.6 ±0.4	**78.2** ±0.5	33.3 ±0.0	–
ReWeightCRT	76.3 ±0.2	70.7 ±0.6	64.7 ±0.2	65.2 ±0.2	**85.1** ±0.4	**75.1** ±0.2	77.5 ±0.7	33.3 ±0.0	–
DFR	**89.0** ±0.2	**73.7** ±0.8	64.4 ±0.1	63.8 ±0.0	81.4 ±0.1	**75.8** ±0.3	74.4 ±1.8	**38.0** ±3.8	–
ERM + DPE	**91.0** ±0.5	**81.9** ±0.2	**69.9** ±0.9	**69.3** ±0.8	**84.1** ±1.5	–	**87.9** ±0.6	**50.0** ±0.0	**54.0** ±4.0
ERM*	77.9 ±3.0	66.5 ±2.6	**69.4** ±1.2	66.5 ±0.7	80.0 ±0.0	75.6 ±0.4	86.4 ±0.8	33.3 ±0.0	53.3 ±0.9
RWY	86.1 ±0.7	**82.9** ±2.2	67.5 ±0.6	68.0 ±1.9	–	–	–	–	–
AFR	**90.4** ±1.1	82.0 ±0.5	68.7 ±0.6	**73.4** ±0.6	–	–	–	–	–
ERM* + DPE	**94.1** ±0.2	**84.6** ±0.8	**68.9** ±0.6	**70.9** ±0.8	**83.6** ±0.9	**76.8** ±0.1	**88.1** ±0.7	**50.0** ±0.0	**63.0** ±1.7

Improvement over ERM Training



Effect of Diversification Strategy



Prototypes

Subgroups

Alignment

“Diving and surface-swimming waterbirds in open aquatic environments”
This prototype captures ecologically coherent waterbirds — mostly **grebes, loons, cormorants, and gulls** — all adapted to **diving or floating** on water. The visual consistency in **plumage contrast, aquatic setting, and pose** makes this a prototypical representation of the waterbird class.

“Large-bodied waterbirds in extended flight or lift-off poses”
Prototype 2 groups waterbirds with **broad wingspans**, often shown **in flight or preparing for takeoff**, including albatrosses, gulls, and terns. While backgrounds vary (bamboo, docks, ponds), the shared latent theme centers on **aerodynamic posture and silhouette geometry**, forming a distinct subpopulation that reflects a **kinetic visual mode** of aquatic birds.

“Compact-bodied diving birds with rounded silhouettes in terrestrial or ambient-light settings”
This prototype groups **grebes, auklets, guillemots**, and terns — birds that share a **morphologically compact, rounded form** with low profiles and dark plumage. Despite their aquatic nature, they're often shown in **mismatched land-based or bamboo-heavy scenes**, suggesting a **visually-coherent but ecologically-confounded** latent cluster.

“Standing gulls in upright posture with clean visual separation”
This prototype forms a highly coherent group of **California, Glaucous-winged, Heermann’s, and Ring-billed Gulls**, unified by their **gray-and-white plumage**, **yellow bills**, and **erect standing poses**. It tolerates a wide range of **backgrounds**, but emphasizes **pose uniformity and size**. A single **Hooded Merganser outlier** appears, likely due to **visual mimicry in shape** rather than ecological or taxonomic alignment.

THANK YOU FOR YOUR ATTENTION