Prototypical Clustering Networks for Dermatological Disease Diagnosis

Viraj Prabhu[1], Anitha Kannan[2], Murali Ravuri[2], Manish Chablani[2], David Sontag[3], Xavier Amatriain[2]

[1] Georgia Tech [2] Curai [3] MIT

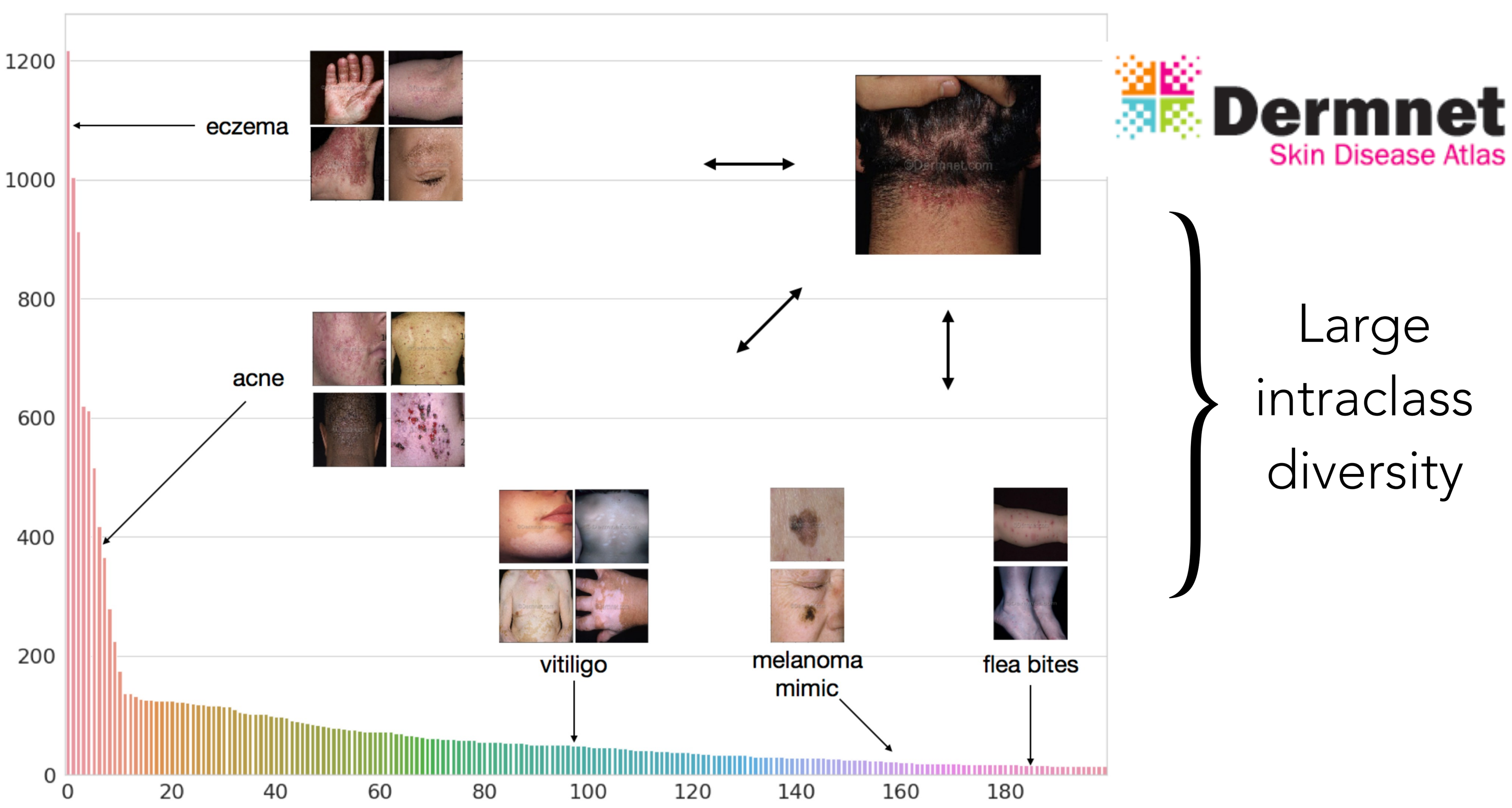
1 Motivation

Setting: Teledermatology with doctor-in-the-loop

Dermatological Diagnosis is challenging
- Over 1000 skin conditions exist[1]
- PCP's typically only trained to diagnose most common conditions

Promising avenue for AI-assisted diagnosis

However, off-the-shelf classifiers fail on derm datasets



Large intraclass diversity

Long Tail

[1] ICD-10

2 Setup

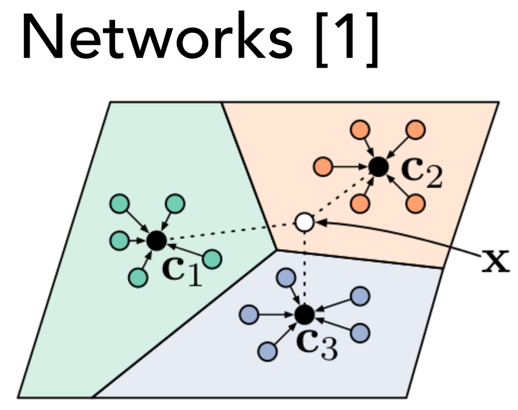
Goal #1: Learn robust representations for tail
- Given very few examples
- Be resistant to overfitting

Natural fit for few-shot learning

Train base classifier on head classes and deploy
- Diagnose new conditions from few shots
- Learning without forgetting

Dermnet dataset (23k images, 600+ classes)
Top 150: Base
150-200: Novel

For few shot learning, we employ Prototypical Networks [1]
- Learn embedding 'prototype' for each class
- $p_\phi(y=k|x) \propto exp(-d(f_\phi(x), c_k))$
- Learn end to end with episodic training



Goal #2: Model diverse multimodal classes
- Prototypical Networks make unimodal assumption
- To model diverse classes, learn mixture of prototypes per class

Our Approach: **Prototypical Clustering Networks**

- @epoch: Cluster class embeddings, centroids are 'prototypes'
- @episode:
  - Infer cluster responsibilities using support set
  - 'Online' prototype update with memory trade-off
  - Infer query set responsibilities, and compute loss

$$p_\phi(y=k|x) \propto exp(\Sigma_z - q(z|k,x) d(f_\phi(x), c_{z,k}))$$

(Mixture of prototypes weighted by posteriors over within-class assignments)

3 Results

Metrics: Mean class accuracy (mca), balanced recall@k

Baselines:
- Finetune on $K_{base+novel}$ with CE loss (FT-CE) (needs retraining)
- FT-*NN: Finetune on K_{base}, at test perform NN on $K_{base+novel}$
- Prototypical Network (PN) [1]

Model: ResNet v2 (50 layers), pretrained on ImageNet

Train: Base classifier trained on K_{base}=150 classes

Eval: 5-shot $K_{base+novel}$=200-way evaluation

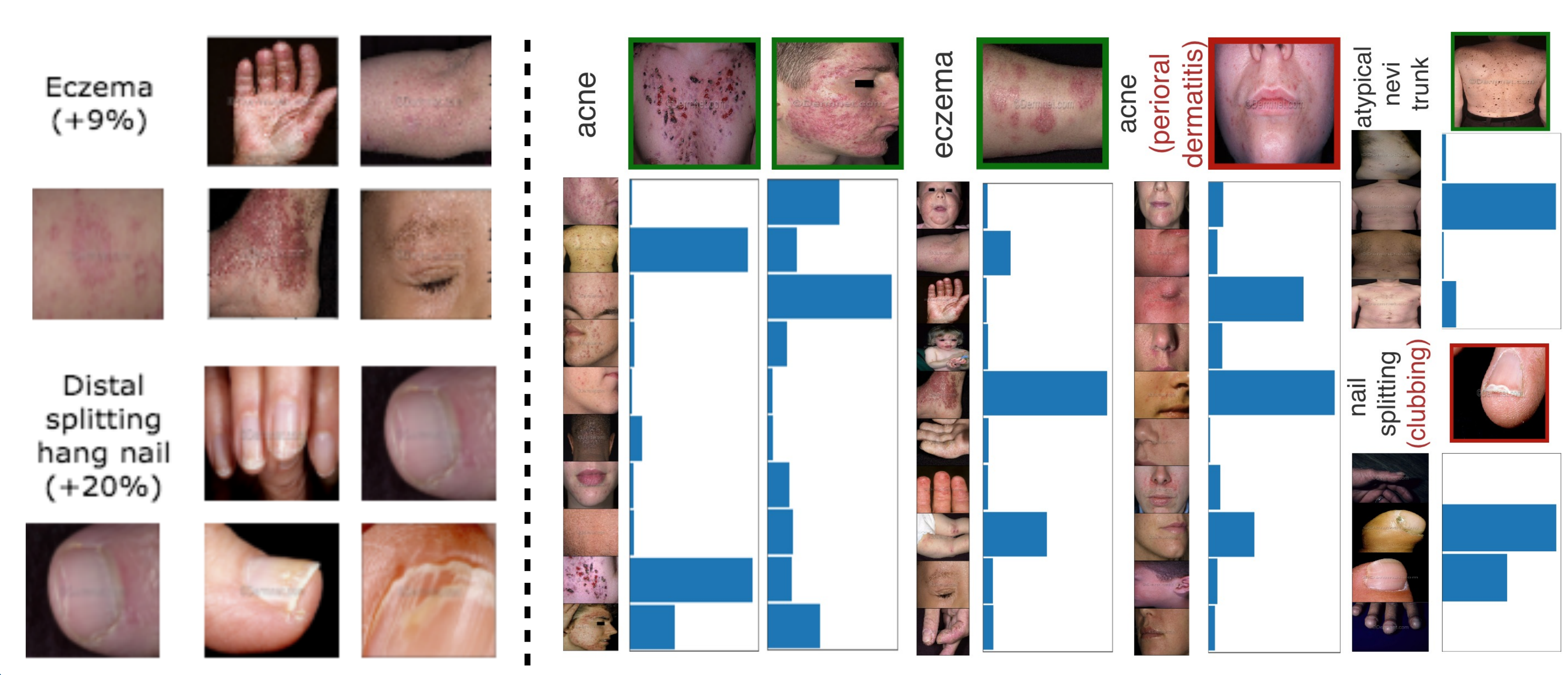
	mca (%)			balanced recall (%)	
	base+novel	base	novel	r@5	r@10
FT-*NN	46.2	55.3	18.8	-	-
FT-CE	47.8	55.8	24.0	65.4	73.1
PN	43.9	48.7	29.6	66.5	75.3
PCN	47.8	53.7	30.0	70.7	79.1

PCN performs strongly across the board

- FT-*NN, FT-CE learn good representations for K_{base}, but are unable to generalize to K_{novel} from few shots
- PN fails to model diverse K_{base} classes, but generalizes to K_{novel}
- PCN closes the gap on K_{base}, and maintains K_{novel} performance

PCN vs PN head to head
- Consistent gains across shot and tail length
- Strong gains (+4% mca) over PN with post-hoc clustering

Sanity check: Generalization to standard *mini*ImageNet low shot benchmark, PCN achieves near SoTA accuracy



4 Summary

- Few shot learning is promising avenue for data starved applications with open-world assumptions
- Prototypical Networks are a good choice, with 1) strong implicit and explicit regularization 2) privacy preservation
- To model diverse classes, relaxing modeling constraints helps

Future Work

- Incorporating multiple modalities (symptoms, text, etc.)
- Uncertainty guided data acquisition (active few shot learning)
- Domain adaptation to real-world (patient-clicked) images
- Semi-supervised few-shot learning

[1] Snell, Jake, Kevin Swersky, and Richard Zemel. "Prototypical networks for few-shot learning." *Advances in Neural Information Processing Systems*. 2017.