Contextualized Transfer Learning

Transforming Heterogeneity into Predictive Power with Generative Latent Structures in Resource-Limited Settings

Siddharth Nirgudkar

Mentor: Dr. Ben Lengerich

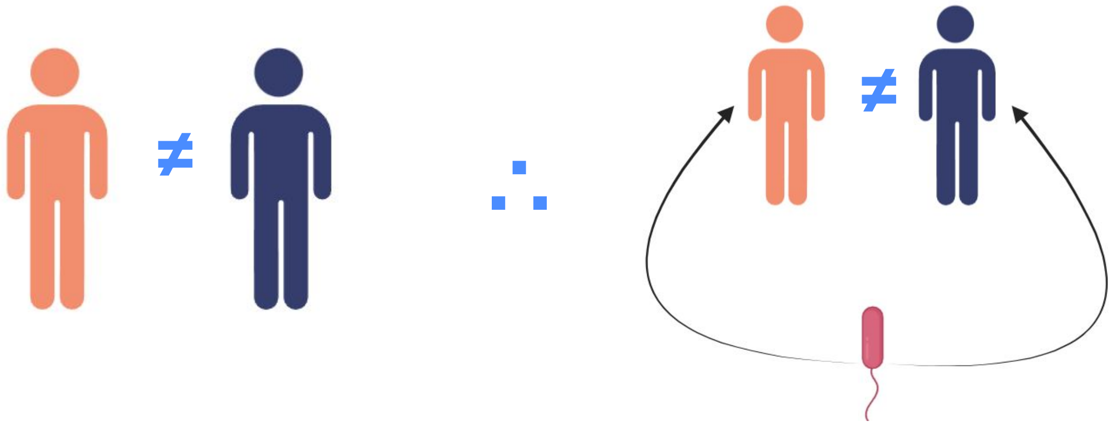
Acton-Boxborough Regional High School

MIT PRIMES October Conference: 10/12/2024

01

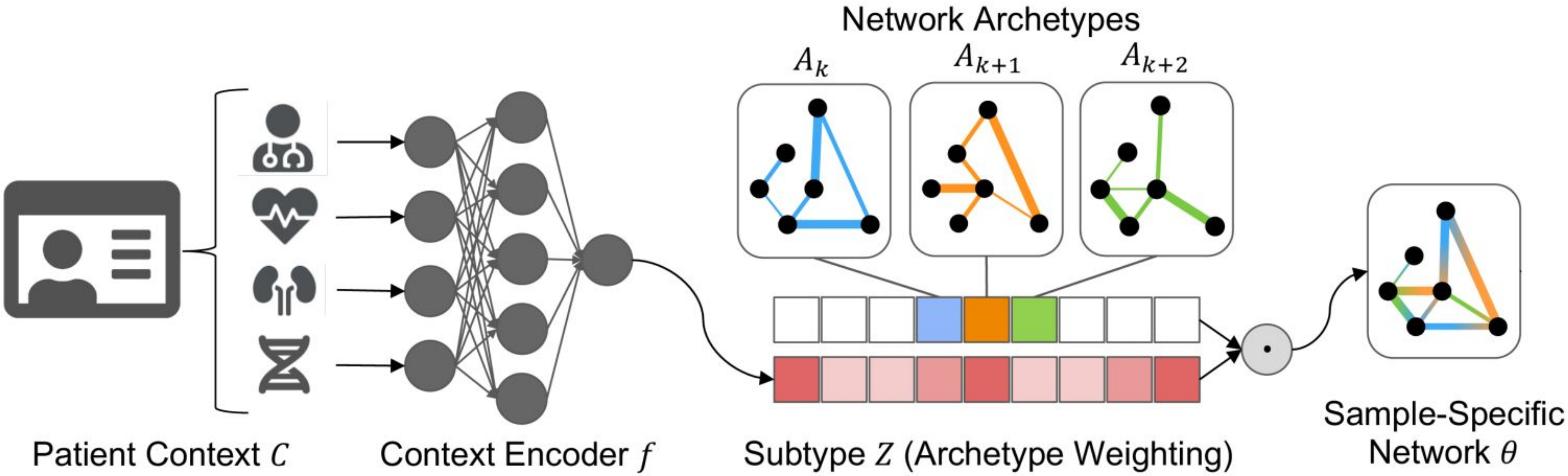
Preliminaries

Introduction to Personalized Medicine



Patient Specific
Understanding

The Response: Contextualized Learning



Ellington Et. Al 2023 https://www.biorxiv.org/content/10.1101/2023.12.01.569658v1.full.pdf

Resource Scarcity in Biomedical Contexts

Data Availability

- Hospitals only have a small, local amount of information to make triage decisions
    - Covid
    - ICU Availability
    - Treatment Priorities

Biological Availability

- Very hard to take samples from organs
- Some organs (like the brain) can only be observed once - postmortem

General Problem

This seemingly medical problem turned into a…

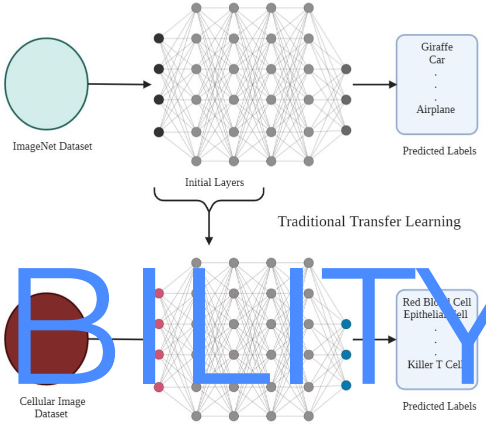
Generative Modeling | **Contextualized Learning**

“ How do we share information across disparate tasks?”

… general information sharing problem

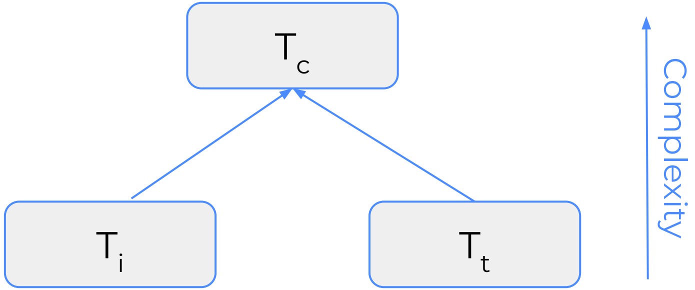
Current Transfer Learning

- **Traditional Transfer Learning**
  - Assumes $T_i \approx T_t$
  - Directly shares information from M_i to M_t
  - Fails when $T_i \neq T_t$



INTERPRETABILITY!

- **Heterogeneous Transfer Learning**
  - Does not require similar tasks
  - Maps T_i , $T_t \rightarrow T_c$
  - Computations and information transfer performed in higher complexity T_c space
  - Maps back down to T_t



02

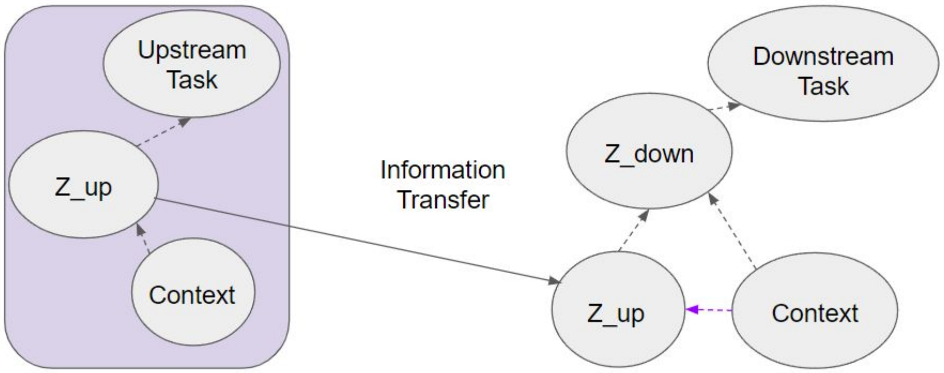
Rationale

A New Perspective: Generative Modeling

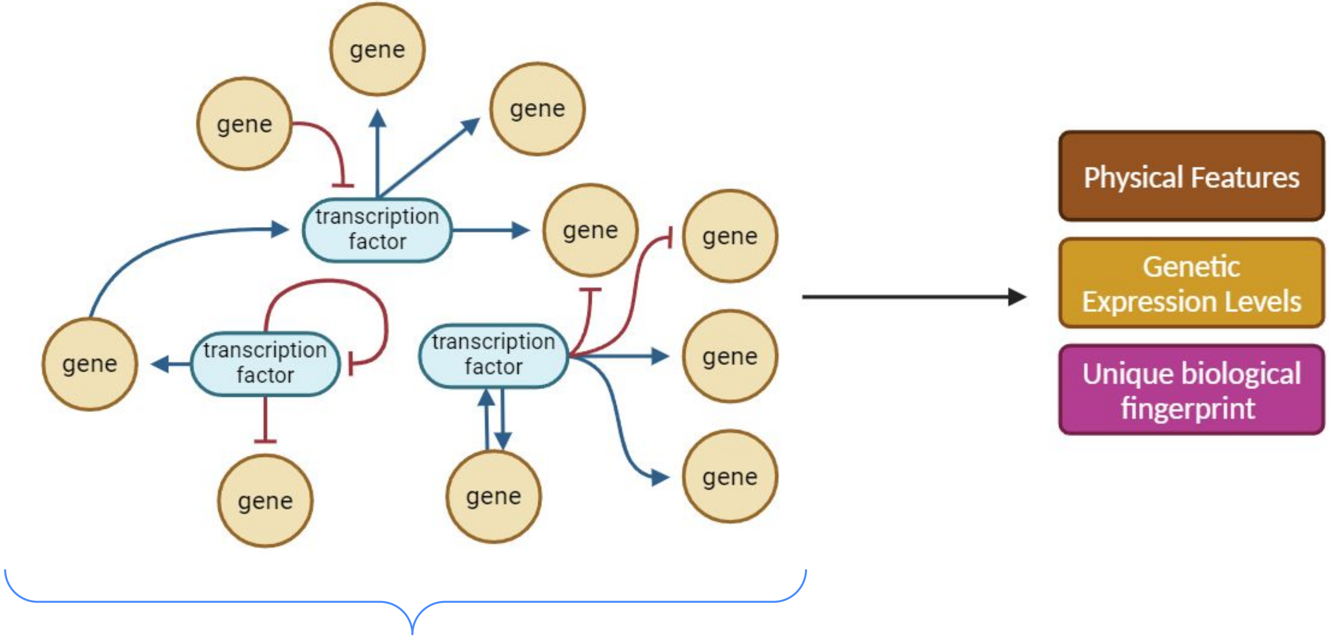
- Estimating join distributions $p(x, y)$ instead of conditional distributions $p(y \mid x)$ allows models to uncover latent structures within data that are consistent across disparate tasks

- **Core Idea:** $p(x, y, c) \sim z \text{ with } y, x \perp c|z$

- **Probability Decomposition:**
  $p(y|x, c) = \int_z dZ p(y|x, z) p(z|c)$



A look at z through a biological lens

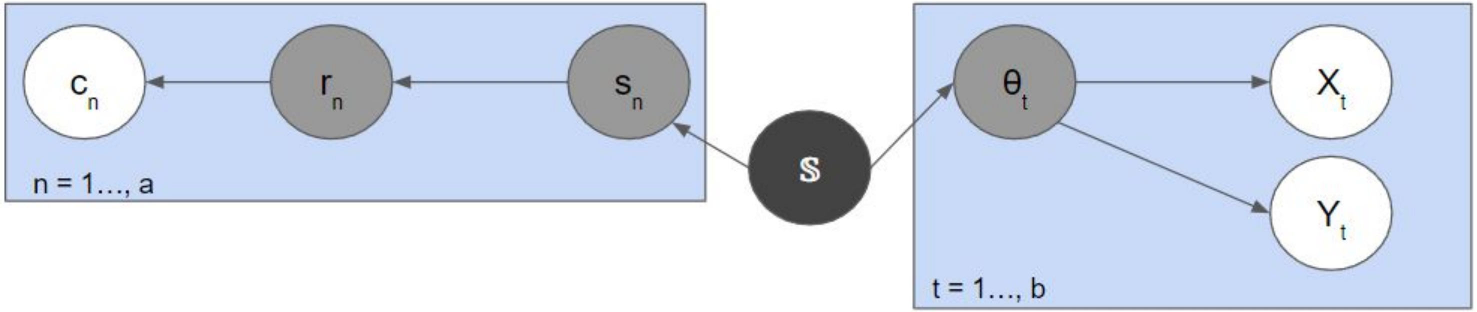


Finite Set of Shared Regulators

03

Implementation

Core Idea



Sample Specific Model

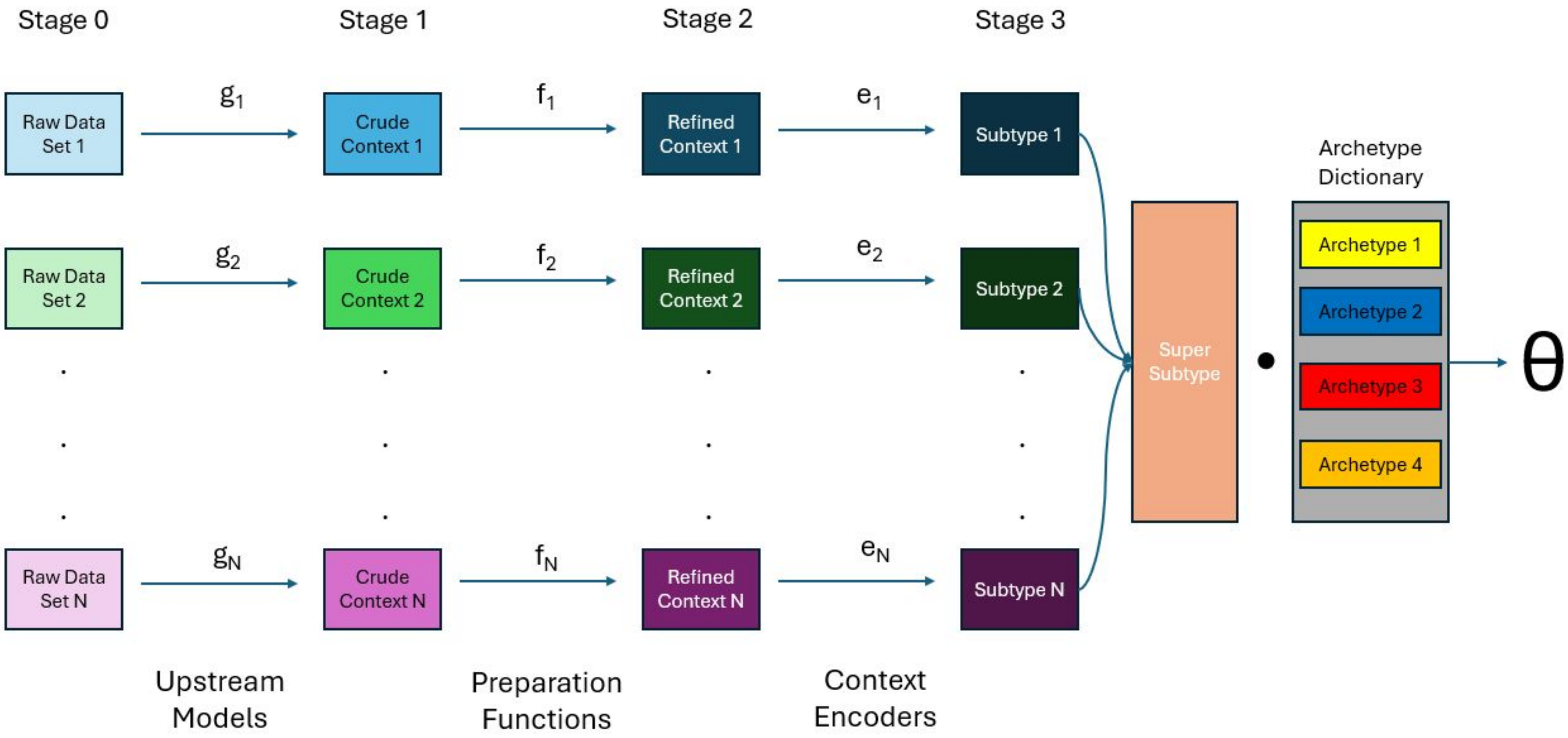
Model Generator

Super Subtype

Refined Context

$$P(\mathbf{Y} \mid X, \mathbb{C}) = \int_\theta P(\mathbf{Y} \mid X, \theta) \cdot P(\theta \mid X, \mathbb{S}) \cdot P(\mathbb{S} \mid \mathbf{R}) \cdot P(\mathbf{R} \mid \mathbb{C}) d\theta$$

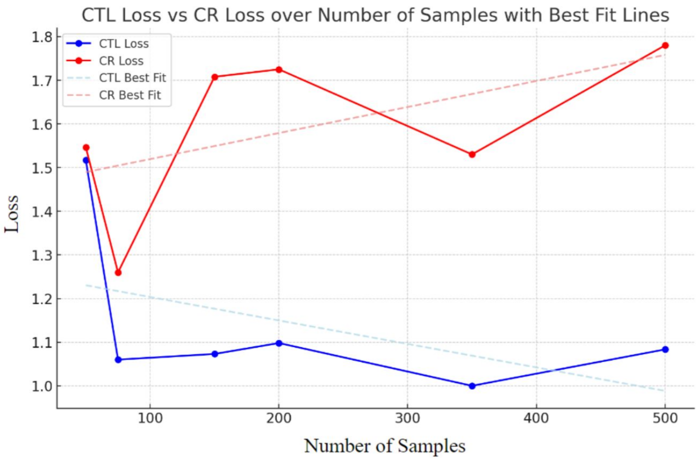
Mechanism for Information Transfer

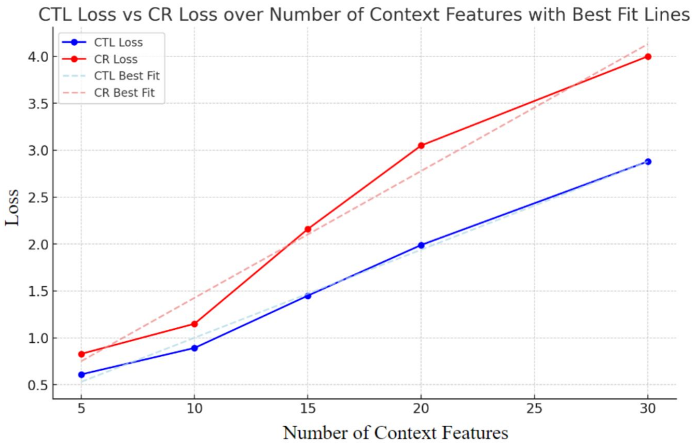


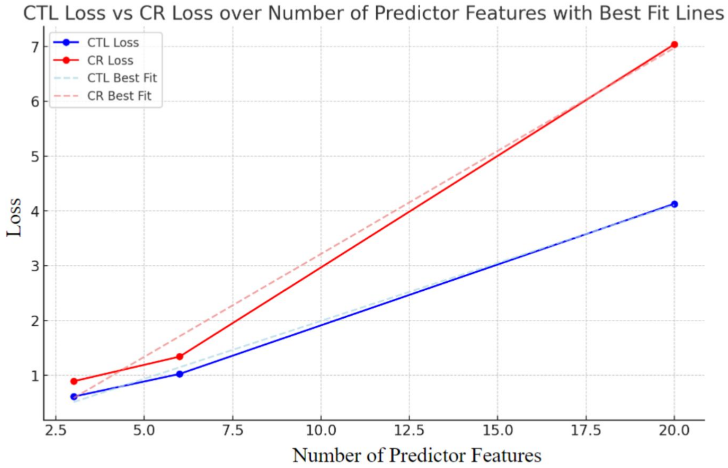
04

Experiments

Synthetic Data

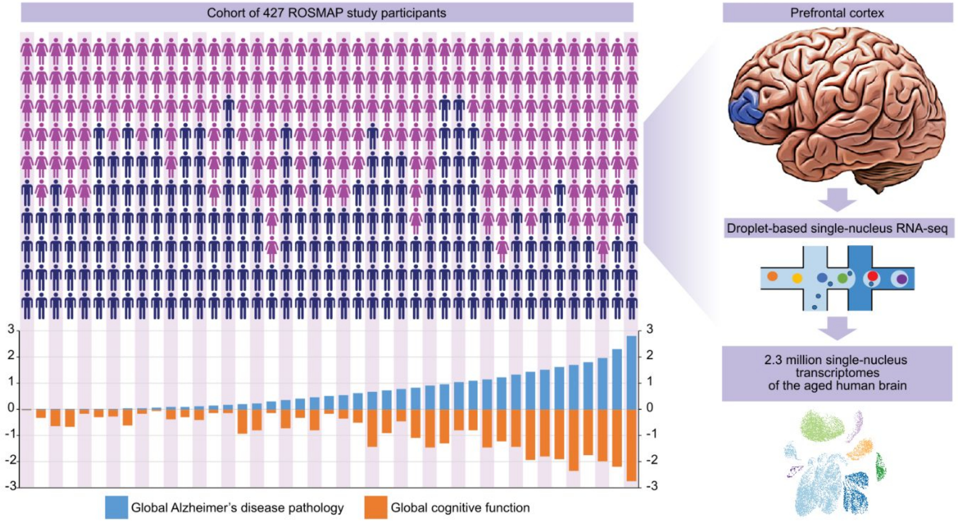






Rosmap Alzheimer's Dataset

- Alzheimer's disease (AD) is a progressive neurodegenerative disorder
- Highly heterogeneous making prediction and diagnosis challenging
- ROSMAP dataset contains extensive genomic data
- We seek to use CTL to make better predictions of alzheimer and gain a better understanding of underlying drivers

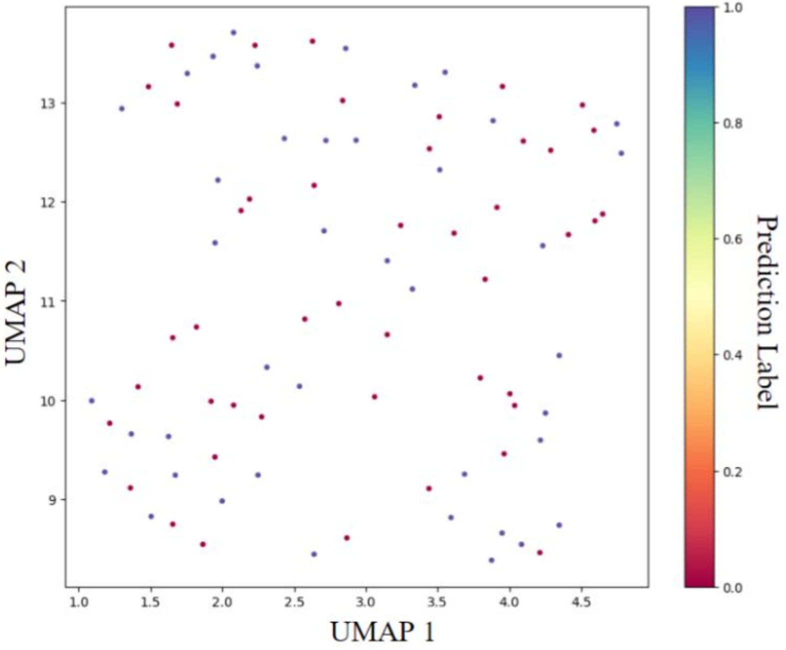


Kellis Lab https://compbio.mit.edu/

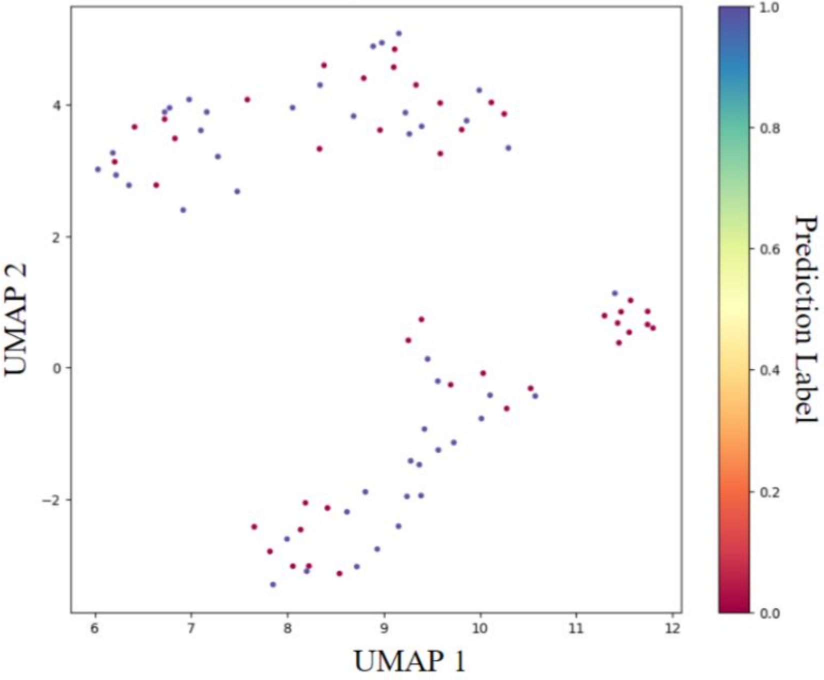
Performance

	Classification		Regression
	Correct Classifications	Incorrect Classifications	Mean Squared Error Loss
Population	30	56	0.4531
Contextualized	47	39	0.3652
CTL (ours)	**56**	**30**	**0.2817**

Analyzing Sample Specific Models

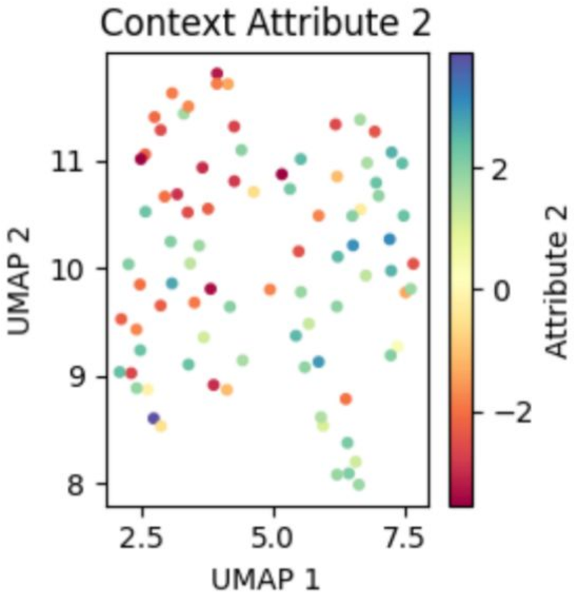


(a) Contextualized Classification

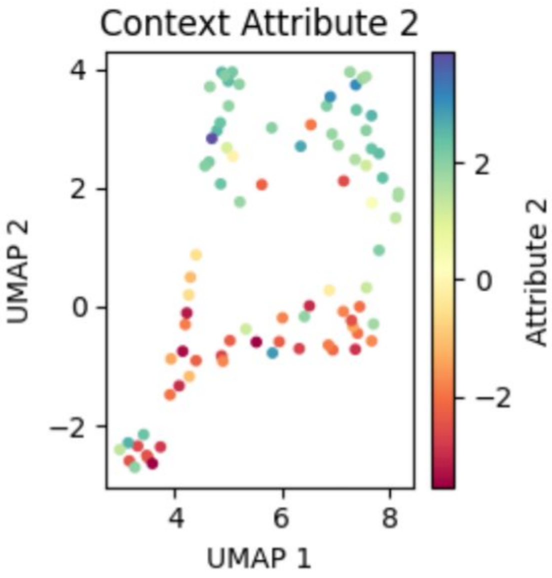


(b) Contextualized Transfer Learning

Elucidating Importance of Context

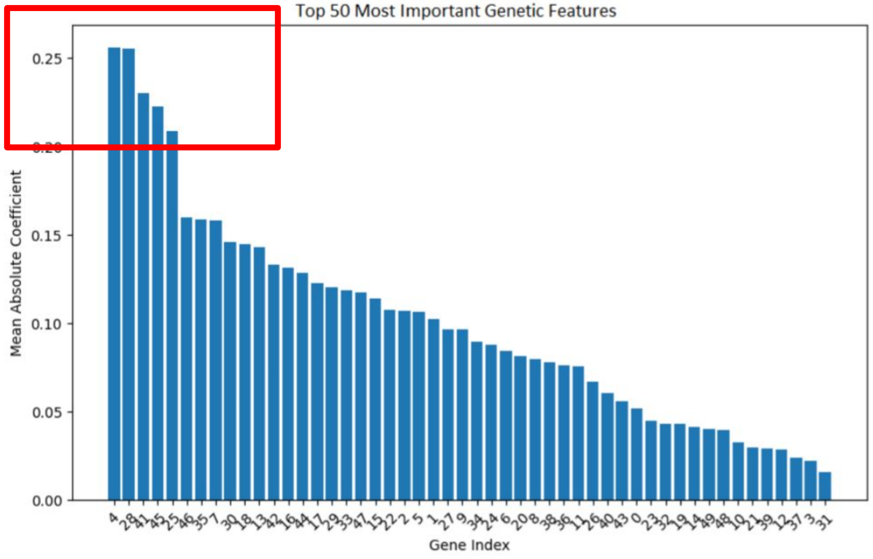


Contextualized Learning

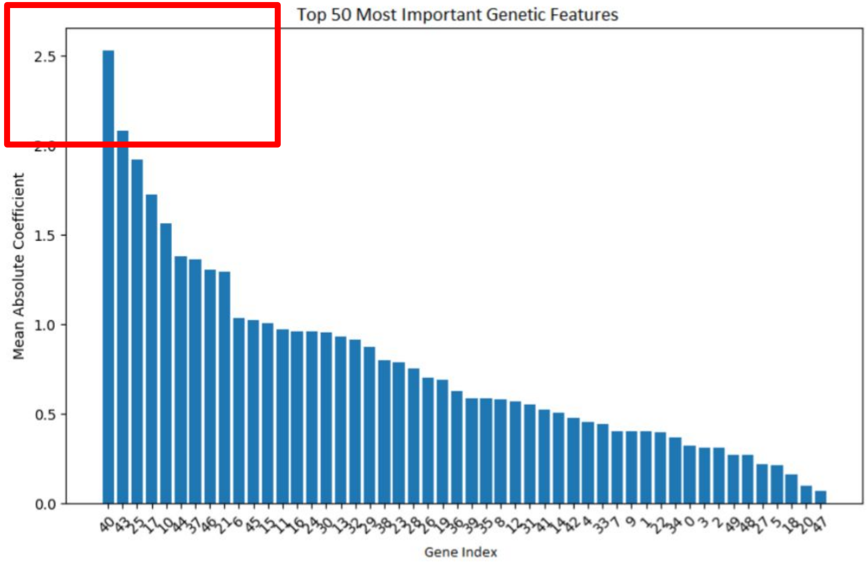


Contextualized Transfer Learning

Genetic Importance for Sample Specific Models



Contextualized Learning



Contextualized Transfer Learning

05

Future Work

- Formalize proof for Convex Hull Optimization for Archetype Dictionary and implement it

Acknowledgements

I would like to sincerely thank the following people:

- Prof. Lengerich (my mentor)
- Prof. Kellis (PI of the Lab)
- Dr. Gerovitch, Prof. Devadas, and rest of MIT Primes team for giving me this opportunity
- My parents

Questions?

References

[1] C. N. Ellington *et al.*, “Contextualized Networks Reveal Heterogeneous Transcriptomic Regulation in Tumors at Sample-Specific Resolution,” *bioRxiv (Cold Spring Harbor Laboratory)*, Dec. 2023, doi: https://doi.org/10.1101/2023.12.01.569658.

[2] B. Lengerich, 1, C. Ellington, A. Rubbi, M. Kellis, and E. Xing, “Contextualized Machine Learning,” 2023. Available: https://arxiv.org/pdf/2310.11340

[3] B. Lengerich, C. Ellington, B. Aragam, E. P. Xing, and M. Kellis, “NOTMAD: Estimating Bayesian Networks with Sample-Specific Structures and Parameters,” *arXiv:2111.01104 [cs, stat]*, Nov. 2021, Available: https://arxiv.org/abs/2111.01104

[4] B. Lengerich, B. Aragam, and E. P. Xing, “Learning Sample-Specific Models with Low-Rank Personalized Regression,” *arXiv.org*, 2019. https://arxiv.org/abs/1910.06939 (accessed Oct. 10, 2024).