Contextualized Transfer Learning: Transforming Heterogeneity into Predictive Power with Generative Latent Structures in Resource-Limited Settings

Problem

- Artificial intelligence has proliferated across every sector, but currently models are not interoperable
- Current medical models only have local insights to work with, which is detrimental for patient outcomes in resource constrained settings
- No current method for sharing insights across diverse contexts/ tasks

Current Limitations

- Traditional transfer learning only works if the two tasks are similar, e.g. deep net layers → specific image detection
- Heterogenous transfer learning creates a high rank subspace and performs calculations there before mapping down → lack of interpretability, which is essential for medical taks

Solution - Contextualized Transfer Learning

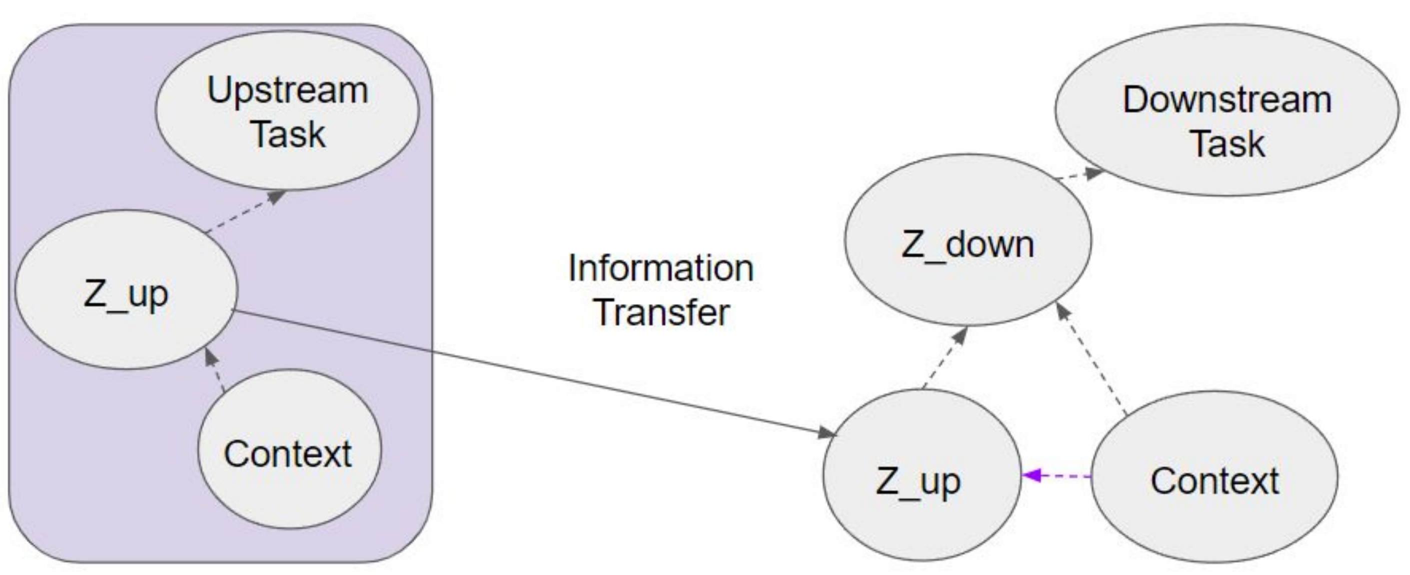

Fig 1. We hypothesize that there is a shared latent space (z) across observations, tasks, and settings. Formally, $p(x, y, c) \sim z$ with $y, x \perp c \mid z$. This allows for the following decomposition: $p(y \mid x, c) = \int_z dZ\, p(y \mid x, z) p(z \mid c)$. **Made by Sid Nirgudkar, Google Slides, 2024*

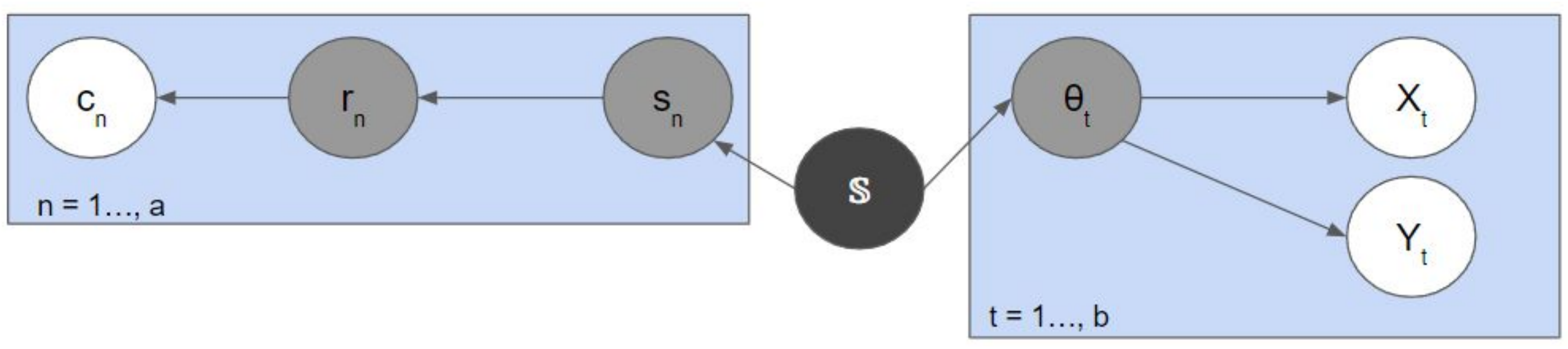

Fig 2. Graphical solution: We seek to estimate **S**, a subset of Z that is shared across context for the current task, and information from upstream models $c_1 \ldots c_n$. **Made by Sid Nirgudkar, Google Slides, 2024*

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

Framework

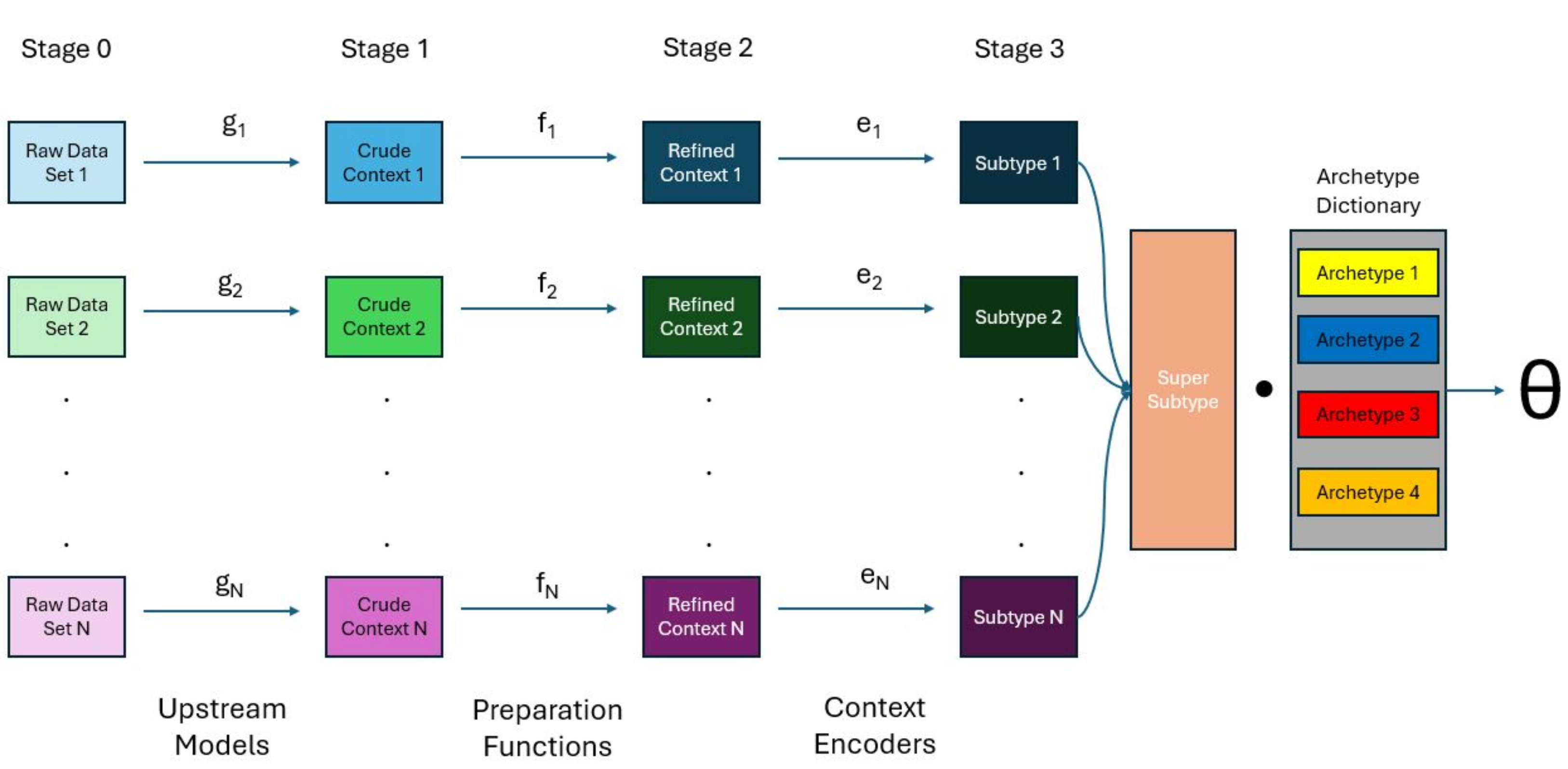

Fig. 3 Deep learning architecture for Contextualized Transfer Learning. **Made by Sid Nirgudkar, Google Slides, 2024*

Stage 0: Raw data is collected and used in upstream models g_i, which are inaccessible. Each g function generates a unique context modality.

Stage 1: Crude context modalities are refined into simpler (order 1) representations called *refined context* via shared transformation functions F. These functions are linear for order 1 modalities and multi-layer perceptrons for higher orders.

Stage 2: Shared Neural Generative Additive Models process the refined context to produce a *subtype*, a vector that weights archetypes to create personalized models.

Stage 3: Subtypes are linearly combined into a *super subtype* to balance information streams. The dot product with an archetype dictionary (containing K extrema sample models) yields a sample-specific model.

Stage 4: Patient observations are passed through this model to generate predictions

Results

	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**

Table 1. CTL outperforms other modes of Contextualized Learning due to transfer of information from upstream models. **Made by Sid Nirgudkar, Latex, 2024*

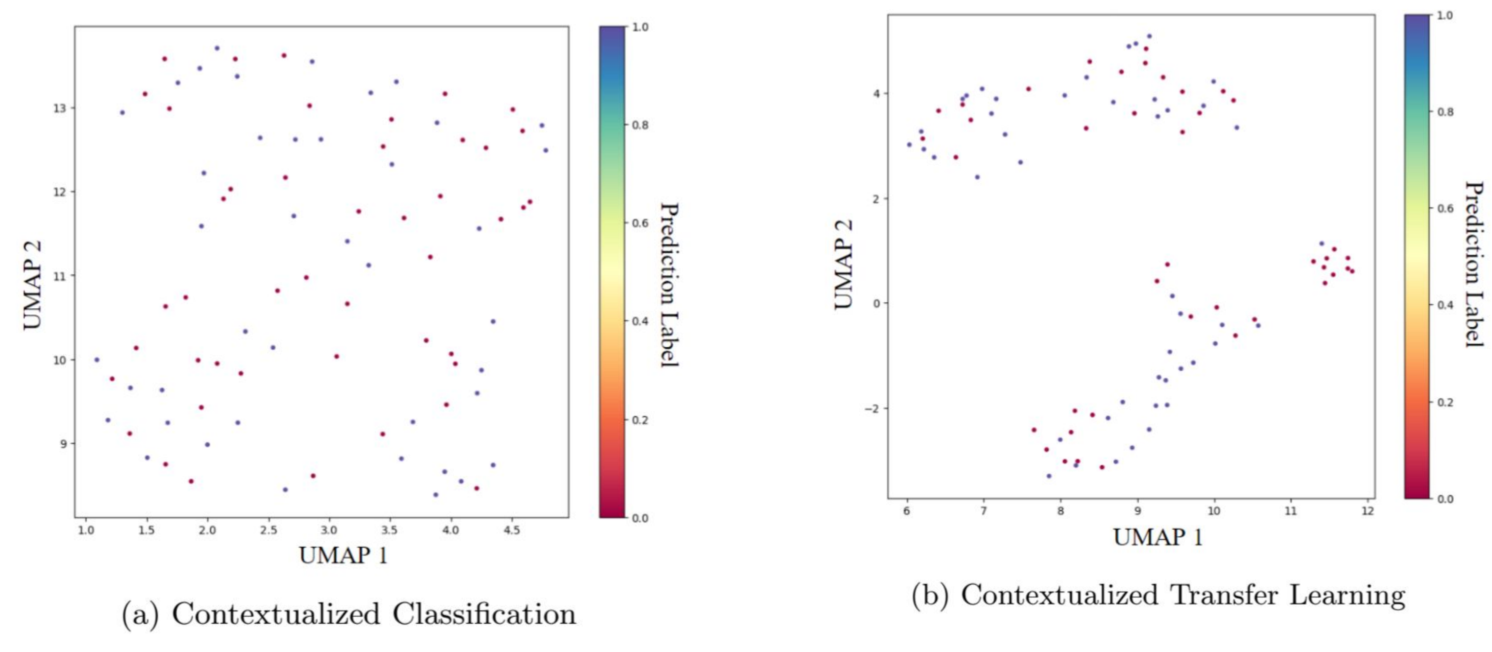

Fig 4. CTL captures underlying patterns that other models cannot. **Made by Sid Nirgudkar, Python, 2024*

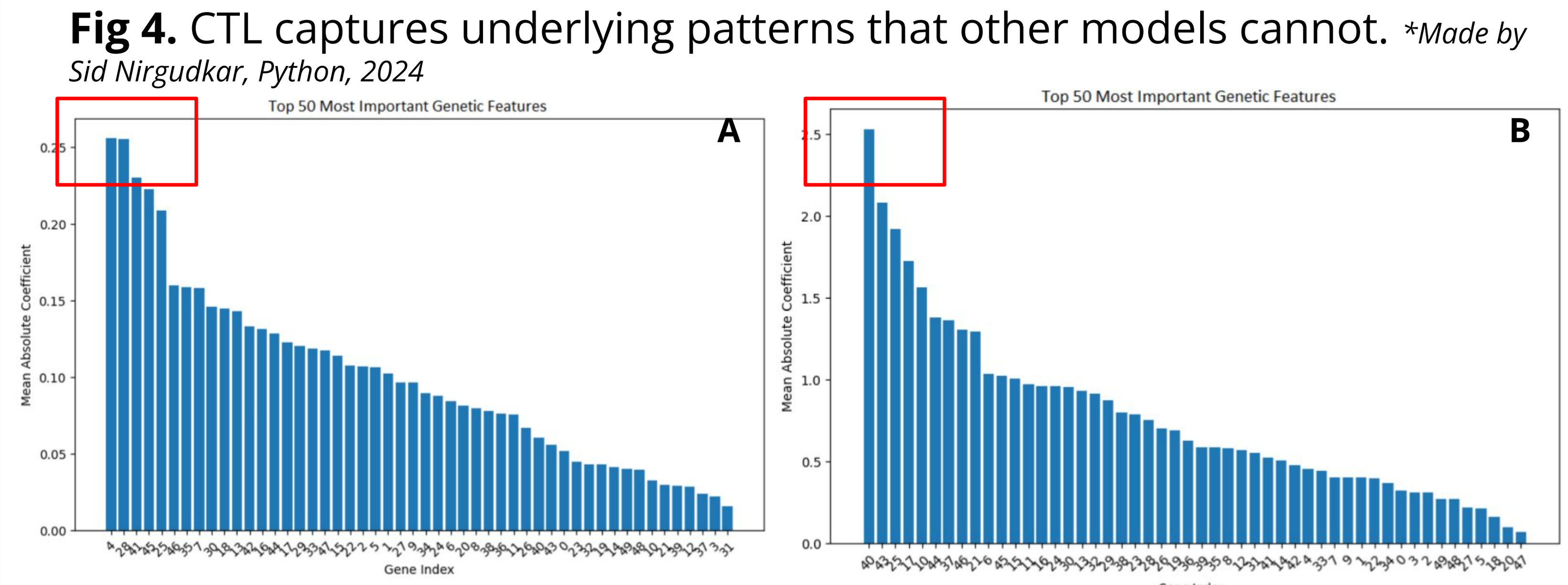

Fig 5. CTL is decisive when predicting **(B)** as compared to other contextualized models **(A)**. This improves interpretability significantly - a key necessity in biomedicine. **Made by Sid Nirgudkar, Python, 2024*

Contextualized Transfer Learning (CTL) leverages stable data distributions to enhance predictions without extensive retraining, as seen in Alzheimer's prediction using shared predictors. It improves accuracy in low-resource settings by integrating upstream information and contextual data for personalized medicine. However, its effectiveness depends on shared features in upstream models, and added context modalities can create non-convex solution spaces, reducing efficiency.