

A institution  
(Source domain)

Knowledge:  
“Similarity features”  
“Missing patterns”

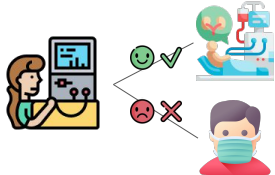


Domain shift  
(Table 2)

B, C, D institutions  
(Target domain)

## Knowledge-Guided Data Augmentation

### Similarity-guided Mixup



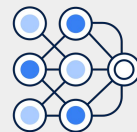
$$S(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^2}{2\sigma^2}\right)$$

$$P(x_j|x_i) = \frac{S(x_i, x_j)}{\sum_{k \neq i, x_k \in \mathcal{D}} S(x_i, x_k)}$$

### Group-based Masking

$$m_j = \begin{cases} b_k & \text{if } j \in \mathcal{G}_k \text{ for some group } k \\ \sim \text{Bernoulli}(1 - p_{\text{ind}}) & \text{if } j \text{ is not in any group} \end{cases}$$

## Model Training



## Evaluation and Analysis

### Recall

Feature space analysis  
Center-wise performance  
...

B	C	D	Mean ± Std
0.4074	0.0333	0.2258	0.2222 ± 0.0958
0.3210	0.0667	0.1613	0.1830 ± 0.0958
0.4074	0.1667	0.2581	0.2774 ± 0.2170
0.4444	0.1667	0.4839	0.3650 ± 0.0958
0.7284	0.5000	0.4516	0.5600 ± 0.1635
0.4198	0.2000	0.7097	0.4431 ± 0.2170
0.7407	0.5333	0.4516	0.5752 ± 0.1209
0.8765	0.5000	0.8065	0.7277 ± 0.1209
0.4321	0.2000	0.7097	0.4473 ± 0.0958
<b>0.9259</b>	<b>0.5667</b>	<b>0.8710</b>	<b>0.7879 ± 0.1473</b>

