A HULP: Human-in-the-Loop for Prognosis (Appendix)

Table A1. Distribution of custom validation split (64 patients) from the ChAImeleon [4] lung cancer dataset for the imputation experiments. Survival range is the recorded time in months. "X" represents unknown or missing data. Count is the count of patients with the same clinical information.

Count	Gender	Smoking Status	T-stage	N-stage	M-stage	Surv. Range (months)
10	Female	Smoker	X	X	X	4.43 to 52.37
5	Female	Smoker	X	X	cM0	8.60 to 61.00
4	Female	Ex-smoker	X	X	X	4.27 to 58.27
3	Female	X	cT4	cN2	cM1	7.40 to 28.97
3	Female	Non-smoker	X	X	X	2.33 to 24.50
2	Female	X	cT3	cN2	cM1	6.50 to 28.60
2	Female	Ex-smoker	X	X	cM1	10.40 to 26.47
1	Female	Smoker	X	X	cM1	0.83 to 0.83
1	Female	Ex-smoker	X	X	cM0	36.17 to 36.17
9	Male	Ex-smoker	X	X	X	2.27 to 57.10
7	Male	Smoker	X	X	X	2.57 to 56.53
6	Male	Non-smoker	X	X	X	2.83 to 23.80
4	Male	Smoker	X	X	cM1	5.63 to 20.27
4	Male	Ex-smoker	X	X	cM1	2.43 to 16.90
3	Male	X	X	X	X	9.07 to 34.50

Algorithm 1 Pseudo-code for Human-in-the-Loop for Prognosis (HuLP) Model

Given a set of clinical features P; m number of unique concepts in each clinical feature; input image x.

Let $M = \sum_{j=1}^{|P|} m_j$.

Encoder

$y = \xi(x)$ ▷ $x \in \mathbb{R}^{H \times W \times D \times C}, y \in \mathbb{R}^{(K \times M)}$

for i in range(M) **do**

  Intervention

  $c_i = \alpha_i(y)$ ▷ $c \in \mathbb{R}^K, K\%2 = 0$

  $p_i = \sigma(c_i)$ ▷ $p \in \mathbb{R}^1, p \in [0, 1]$

  During inference, and during training with a probability of 0.25, p_i is replaced with the ground-truth label ▷ $p \in \mathbb{Z}^1, p \in [0, 1]$

  $[c_i^+, c_i^-] = c_i$ ▷ $c^+, c^- \in \mathbb{R}^{K/2}$

  $c_{Fi} = p_i c_i^+ + (1 - p_i) c_i^-$ ▷ $c_F \in \mathbb{R}^{K/2}$

  Classifier

  $q_i = \beta_i(c_{Fi})$ ▷ $q \in \mathbb{R}^1$

end for

Prognosticator

$w = [c_{F1}, c_{F2}, ..., c_{FM}]$

$y = \gamma(w)$ ▷ $y \in \mathbb{R}^n$