

## Appendix A. Comparison of different margin settings

To measure the impact of different margin settings, we train the model 10 iterations on a histopathological image. 1000 responses are simulated in the first iteration, and 1500 responses are simulated in each following one. Answered triplets are doubled by the proposed enhancement algorithm before feeding them into the model. Dendrogram purity is applied to measure the performance. We applied different margin values with constant and dynamic (Yin et al., 2020) settings. Results are reported in Table 2 and Fig 12.

It could be seen that there is no much difference in performance under constant and dynamic margin settings when margin value is small, but a noticeable performance drops under dynamic 0.8 margin. That is because the extra variables introduced by dynamic margin may destabilize the model, and a large margin value could enlarge this impact. Therefore, in this paper, we choose a constant margin with value 0.2.

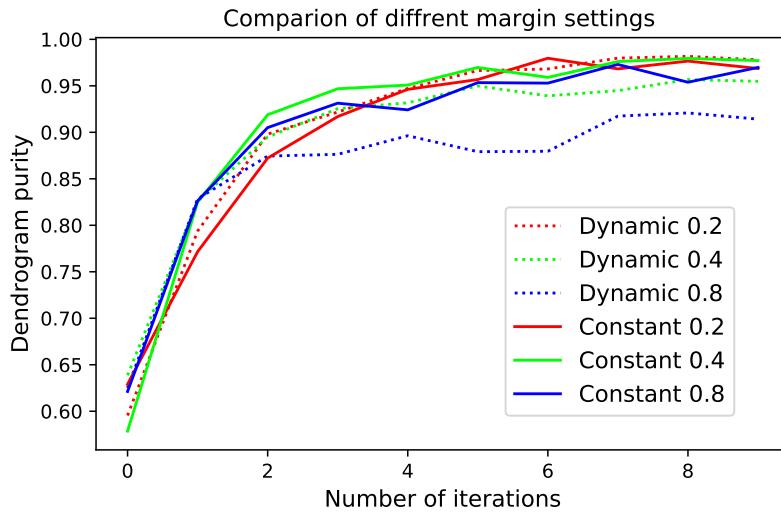


Figure 12: Training curve with different margin settings on histopathological image

Table 2: Dendrogram purity of different margin settings on histopathological image (%)

Iterations	Margin value		
	Constant/Dynamic	0. 2	0. 4
0	62.90/59.54	57.86/63.87	62.12/62.51
1	77.16/79.37	82.48/82.61	82.63/82.85
2	87.22/89.81	91.89/89.50	90.51/87.44
3	91.69/92.18	94.68/92.55	93.13/87.63
4	94.62/94.72	95.08/93.17	92.42/89.63
5	95.68/96.64	96.97/94.98	95.34/87.92
6	97.97/96.81	95.91/93.92	95.29/87.96
7	96.82/97.98	97.61/94.48	97.30/91.74
8	97.66/98.17	97.95/95.67	95.39/92.08
9	96.86/97.80	97.71/95.49	96.98/91.42