

A Appendix for Fairness of Neural Collapse

A.1 Empirical Neural Collapse Metrics

$$\text{NC1} := \frac{1}{K} \text{tr} \left\{ \Sigma_W \Sigma_B^\dagger \right\}$$

where $\Sigma_W = \frac{1}{n} \sum_{k=1}^K \sum_{i=1}^{n_k} ((h_{i,k} - \mu_k)(h_{i,k} - \mu_k)^T)$ is the within class covariance and $\Sigma_B = \frac{1}{K} \sum_{k=1}^K ((\mu_k - \mu_G)(\mu_k - \mu_G)^T)$ is the between class covariance, \dagger denotes the pseudo-inverse. Optimal value: NC1 = 0.

$$\text{NC2} = \left| 1 - \frac{s_1}{s_2} \right|$$

where s_1 and s_2 are the singular values of the matrix defined as $[\mu_0, \mu_1]$ in the case of binary classification. Optimal value: NC2 = 0.

$$\text{NC3} = \left\| \frac{W^T}{\|W\|_F} - \frac{M}{\|M\|_F} \right\|_F$$

where W is the weights matrix of the linear classifier and M is the matrix defined as $M = [\mu_0 - \mu_G, \mu_1 - \mu_G]$ in the binary classification case. Optimal value: NC3 = 0.

$$\text{NC4} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(\text{argmax}_k \langle h_i, w_k \rangle \neq \text{argmin}_k \|h_i - \mu_k\|)$$

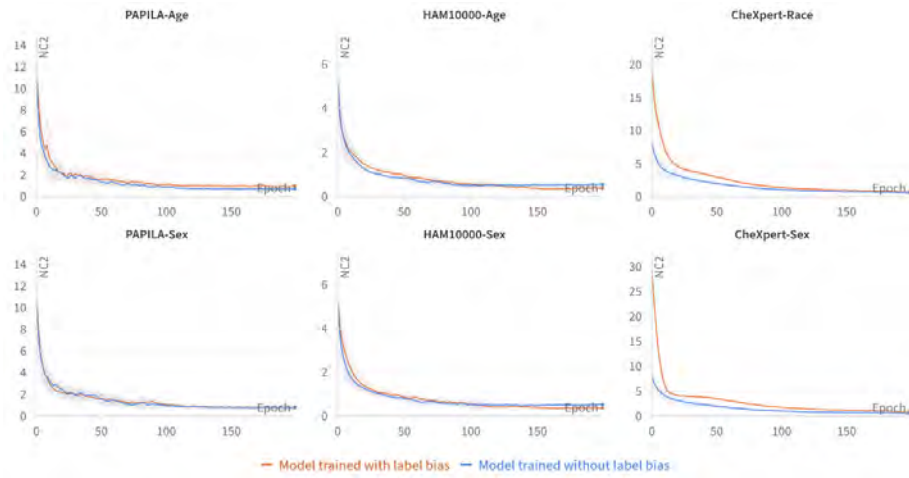
where $\mathbb{I} : \{True, False\} \rightarrow \{1, 0\}$ is the indicator function. Optimal value: NC4 = 0.

A.2 Implementation details

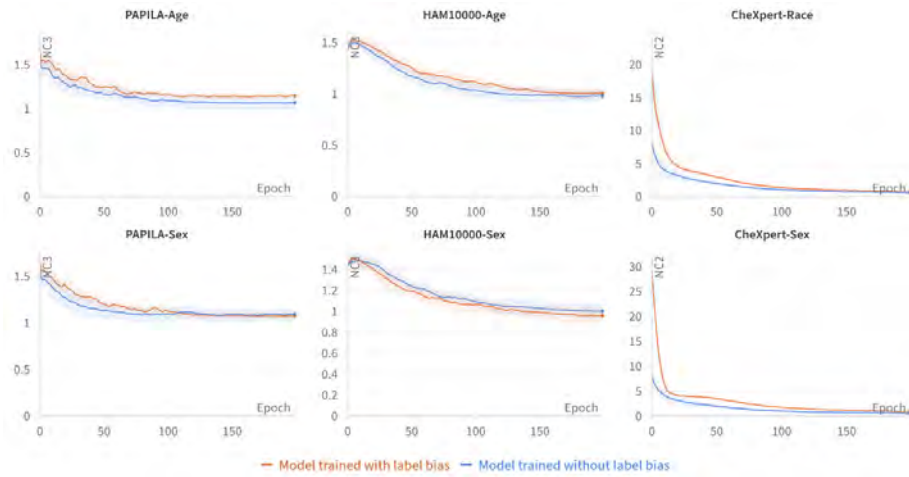
Table 1: Hyperparameters using in the training for all experiments.

Parameter	Value
Architecture	ResNet34
Optimizer	Adam [$\beta_1 = 0.9$, $\beta_2 = 0.999$]
Learning rate	Initial LR = 5×10^{-4} , Minimum LR = 5×10^{-7} .
LR scheduler	ReduceOnPlateau [monitor = train loss, patience=20, factor=0.1]
Early stopping	[monitor = val loss, patience=10, max epochs=200]
Batch size	256 (32 for PAPILA)
Train time	PAPILA=4h, HAM10000=5h, CheXpert=8h (per model)
GPU	NVIDIA A40 48GB

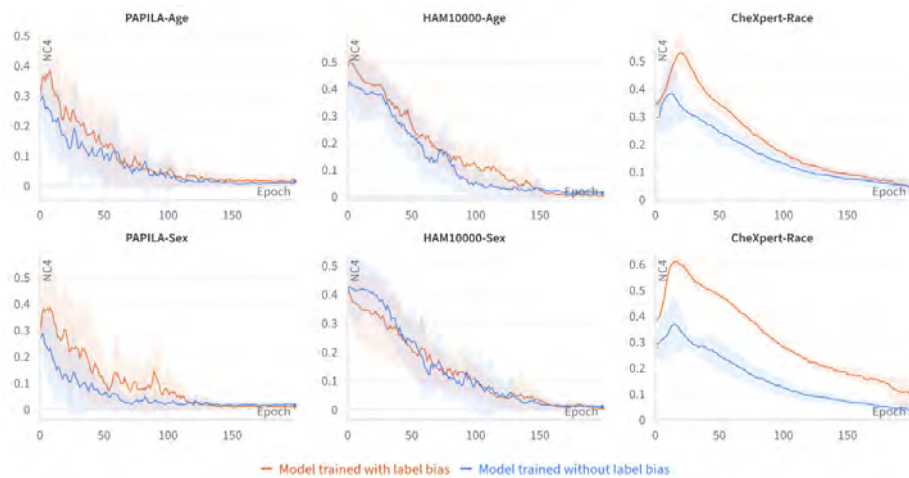
A.3 Monitoring NC2, NC3 and NC4 during Training



(a) NC2 metric per epoch for each dataset-attribute combination.



(b) NC3 metric per epoch for each dataset-attribute combination.



(c) NC4 metric per epoch for each dataset-attribute combination.