## Supplementary material for Multi-class Classification from Multiple Unlabeled Datasets with Partial Risk Regularization

Yuting Tang Nan Lu Tianyi Zhang The University of Tokyo/RIKEN, Japan.

Masashi Sugiyama RIKEN/The University of Tokyo, Japan. TANG@MS.K.U-TOKYO.AC.JP LU@MS.K.U-TOKYO.AC.JP ZHANG@MS.K.U-TOKYO.AC.JP

SUGI@K.U-TOKYO.AC.JP

Editors: Emtiyaz Khan and Mehmet Gönen

## 1. Experiments on large-scale datasets

We further tested our method on Kuzushiji-49 dataset<sup>1</sup>, which has 49 classes (28x28 grayscale, 270,912 images) of Hiragana characters, and SVHN dataset<sup>2</sup>, which has 10 classes (32x32 RGB, 531,131 images) of printed digits cropped from house number plate photos. A 5-layer fully-connected network is used for the Kuzushiji-49 dataset, and ResNet-20 is used for the SVHN dataset. We ran each experiment five times, and we trained the model for 100 epochs on all datasets. Other experimental setups followed Section 5.1 in the paper. The experiments on the symmetric class-prior matrix are reported in Table 1.

Table 1: Experimental results on large-scale datasets. Means (standard deviations) of the classification error (Err) and the drop ( $\Delta_E$ ) over five trials in percentage. The best and comparable methods based on the paired t-test at the significance level 5% are highlighted in boldface.

Dataset	a,b	Biased		Prop		Prop-CR		Unbiased		U-correct		U-flood		U-PRR	
		Err	$\Delta_E$	Err	$\Delta_E$	Err	$\Delta_E$	Err	$\Delta_E$	Err	$\Delta_E$	Err	$\Delta_E$	Err	$\Delta_E$
Kuzushiji-49	0.51, 0.01	54.59 (0.52)	30.11 (0.46)	$  45.76 \\ (0.89)  $	1.05 (0.84)	$\begin{vmatrix} 38.90 \\ (0.80) \end{vmatrix}$	$\begin{array}{c} 0.12 \\ (0.17) \end{array}$	53.27 (0.45)	19.67 (0.54)	$  \begin{array}{c} 41.33 \\ (1.67) \end{array}  $	6.91 (1.51)	$  \begin{array}{c} 38.02\\ (0.87) \end{array}  $	2.66 (0.95)	$30.15 \ (0.50)$	1.55 (0.66)
SVHN	0.5, 0.05	$ \begin{array}{c c} 12.79 \\ (2.40) \end{array} $	3.47 (2.12)	$\begin{vmatrix} 27.75\\(2.37) \end{vmatrix}$	$0.28 \\ (0.44)$	22.58 (1.62)	$\begin{array}{c} 0.01 \\ (0.01) \end{array}$	88.22 (6.45)	72.71 (6.53)	$ \begin{array}{c c} 10.92 \\ (0.34) \end{array} $	$\begin{array}{c} 0.61 \\ (0.19) \end{array}$	$  13.52 \\ (0.79)  $	$1.28 \\ (0.90)$	$\begin{array}{ }9.27\\(0.37)\end{array}$	$\begin{array}{c} 0.26 \\ (0.35) \end{array}$

The experimental results show the proposed U-PRR method outperforms baselines and successfully mitigates the overfitting of the Unbiased method.

<sup>1.</sup> http://codh.rois.ac.jp/kmnist/

<sup>2.</sup> http://ufldl.stanford.edu/housenumbers/