# Supplementary Material for ACPL: Anti-curriculum Pseudo-labelling for Semi-supervised Medical Image Classification

Fengbei Liu[1*] Yu Tian[1*] Yuanhong Chen[1] Yuyuan Liu[1]
Vasileios Belagiannis[2] Gustavo Carneiro[1]
[1] Australian Institute for Machine Learning, University of Adelaide
[2] Universität Ulm, Germany

Table 1. Ablation study of the number of information content sets in Eq.2 (2, 3, 4 sets) with model training performance (in terms of mean AUC testing set results) and number of training stages with 2% and 20% labelled set on Chest X-ray14 [2].

| Number of Inform. Cont. Sets in Eq.2 | 2 | 3 | 4 |
|---|---|---|---|
| Number of Training Stages | 5 | 5 | 9 |
| 2% | 71.28 | 74.44 | 74.37 |
| 20% | 79.56 | 81.51 | 81.60 |

## 1. Additional Ablation study

**The Number of Information Content Sets in Eq.2** is studied in Table 1, which shows the model training performance (in terms of mean AUC testing set results) and number of training stages using 2% and 20% labelled set on Chest X-ray14 [2]. The default setting used in the paper is to have three information content sets, namely low, medium, high. As shown in Table 1, the selection of only two sets produces the worst results because the pseudo-labelled set becomes less informative and imbalanced. The selection of four sets produces similar results as with three sets. However, with this additional set, the number of new pseudo labelled samples are greatly reduced for every training stage, forcing the number of training stages to grow. Hence, by selecting three sets we reach a good balance between training time and accuracy.

## 2. Data Distribution

In Figure 1, we show the data distribution of all classes of Chest X-ray14 (plus the class 'No Findings') [2]. Notice that the selection of high information content samples (blue) creates a more balanced distribution compared with the selection of low information content (yellow) or the original data distribution (green).

*First two authors contributed equally to this work.

## 3. Visualization of Classification Results

Figure 2 shows examples of pseudo-labels produced by our density mixup for both Chest Xray-14 [2] (top) and ISIC2018 [1] (bottom) datasets.

## References

[1] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9, 2018. 1, 2

[2] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017. 1, 2

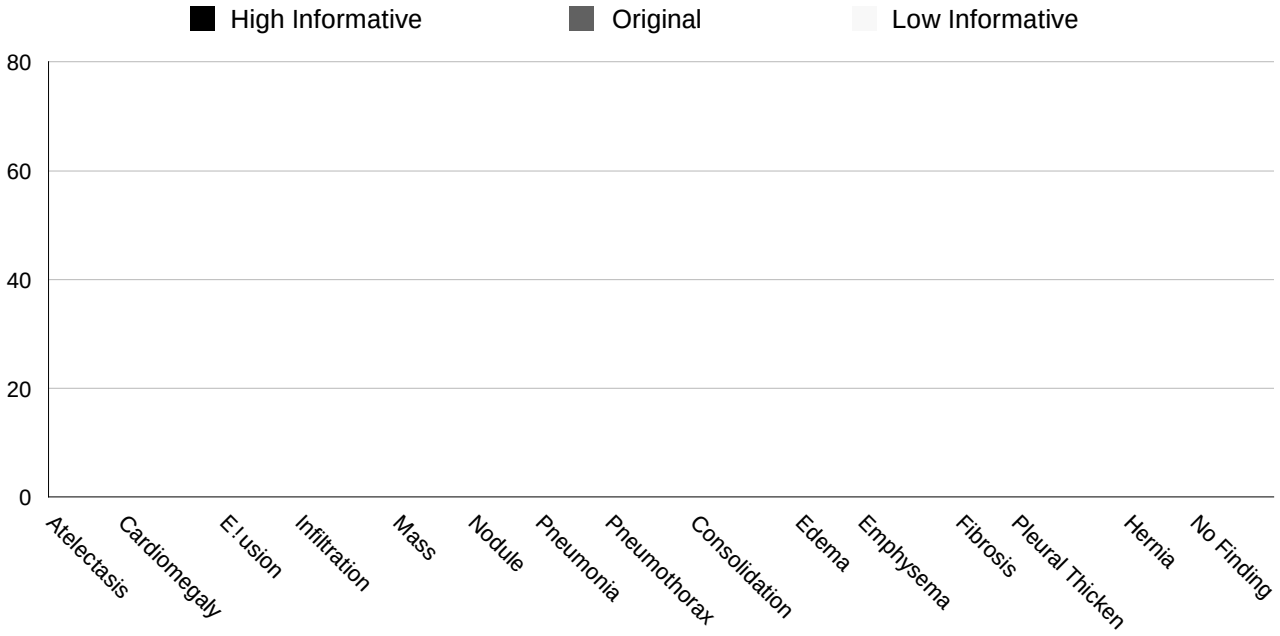


Figure 1. Histogram of label distribution in percentage of all 14 classes from Chest X-ray14 plus the class 'No Finding'. Blue for high information content subset and yellow for low information content subset. Green is the original data distribution.

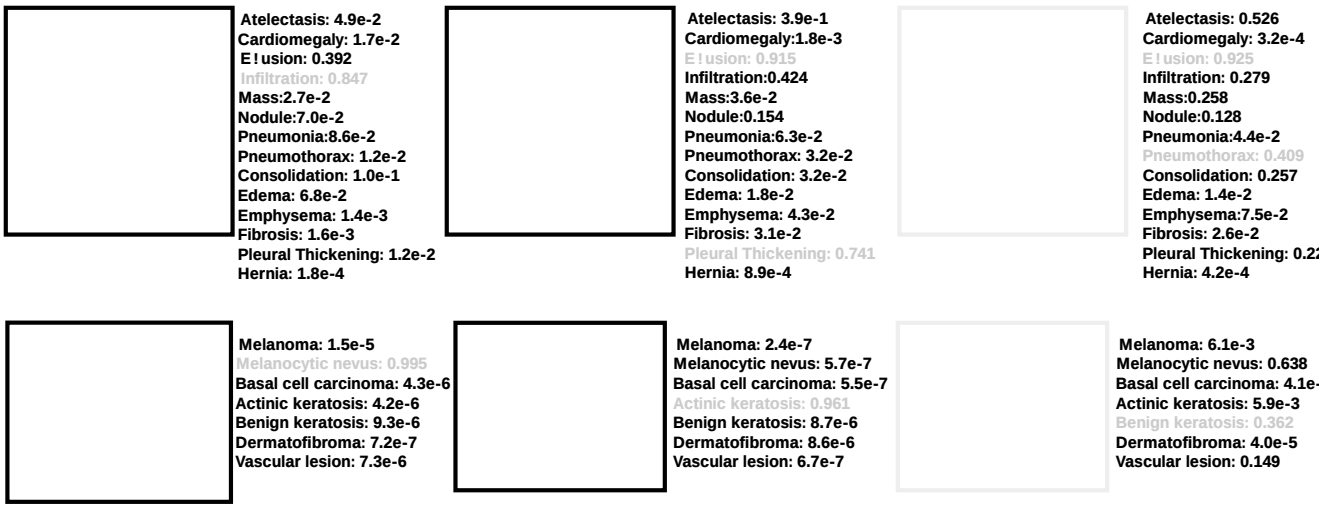


Figure 2. Pseudo-labelling of high-information content unlabelled samples estimated with the **density mixup** prediction for Chest Xray-14 [2] (top) and ISIC2018 [1] (bottom) datasets. Green border denotes accurate prediction and red border represents inaccurate prediction. Classes with red color represent the ground truth.