Out-of-Distribution Detection for Dermoscopic Image Classification

Mohammadreza Mohseni[1, 2] *Jordan Yap*[2] *William Yolland*[2] *Majid Razmara*[2] *M. Stella Atkins*[1, 2, 3]
[1]*School of Computing Science, Simon Fraser University* [2]*MetaOptima Technology Inc*
[3]*Department of Dermatology and Skin Science, University of British Columbia*

Introduction

- Medical image diagnosis can be achieved by deep neural networks, provided there is enough varied training data for each disease class.
- Well-instructed primary care physicians (PCPs) are able to consistently capture dermoscopic images from a similar domain as training images, rendering classification of dermoscopic lesions feasible. However, training on every possible class of skin diseases is not possible, because some diseases are extremely rare.
- A hitherto unknown disease class not encountered during training will inevitably be misclassified, even if predicted with low probability.
- In this work, we aim to efficiently identify novel skin disease images while mitigating the performance loss introduced by multi-class classification.

Data

In order to train, validate, and evaluate our algorithms, we used publicly available dermoscopic images from the ISIC challenges.

Class Name	Train	Dev	Test
BCC	2807	299	217
BKL	2397	259	147
DF	187	28	24
MEL	4720	193	193
NV	15609	1230	1229
SCC	496	79	53
VASC	184	40	29
AK	0	428	439

We treated Actinic Keratosis as novel disease in our experiments. All other diseases were considered as common diseases.

Method

In order to make the neural network able to accurately classify in-distribution classes and detect OOD samples at the same time, we use a BinaryHeads network multi-label classifier with conflict resolution at the inference time. In the case that an image is not associated with any class, our network predicts that image as OOD.
Our conflict resolution strategy uses per-class-thresholds to mitigate the unfairness in imbalanced setting.

Out-of-Distribution Detection

Novel disease detection can be viewed as an out-of-distribution (OOD) detection problem. However, we observed that samples from more numerous diseases are less prone to be misclassified as OOD compared to less numerous diseases.

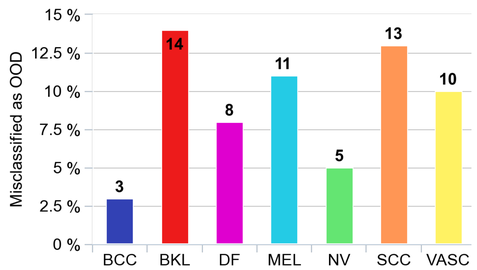

Figure 1: percentage of samples from each in-distribution class mispredicted as OOD using the baseline approach[1]

Evaluation Framework

We evaluate our work using two types of evaluations called internal evaluation and external evaluation. Internal evaluation is concerned with the performance of the same system after it is equipped with the ability to detect OOD. External evaluation is concerned with comparing performance of our algorithm against other OOD detection algorithms. In each scenario, we also show how the presence of differing amounts of OOD in the dataset affects performance of the system as this can be never known in real world scenarios.

Internal Evaluation

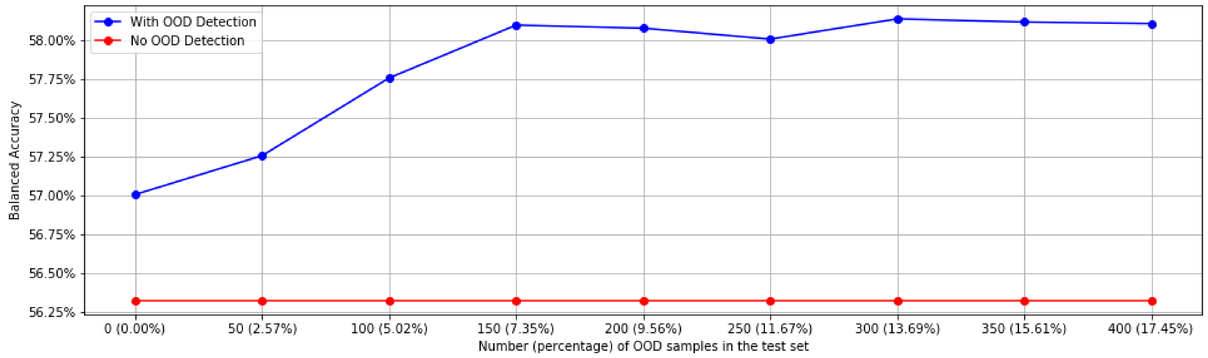

Figure 2: Comparing a BinaryHeads model with and without OOD detection capability

Figure 2 shows that our algorithm in addition to detecting OOD samples is able to boost the balanced accuracy of the classifier. We believe this is happening as a result of using per-disease thresholds.

External Evaluation

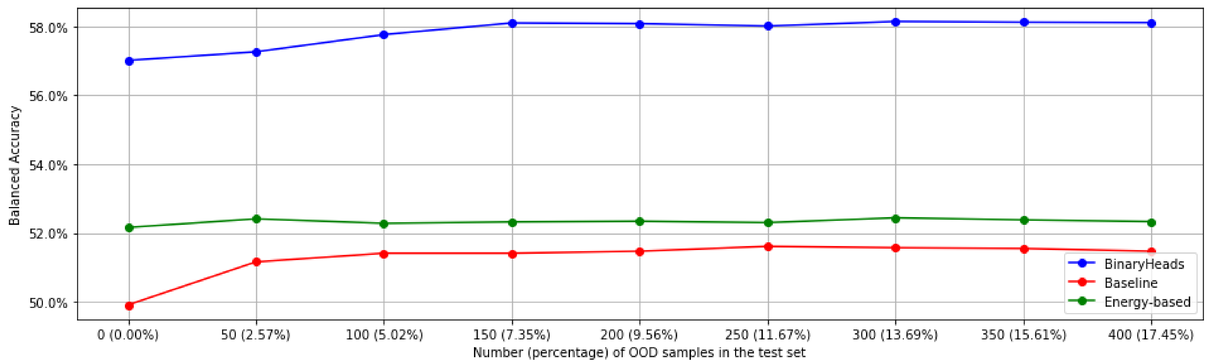

Figure 3: Comparing balanced accuracy of our OOD detection algorithm against baseline[1] and energy-based[2] OOD detection algorithms

Figure 3 shows that by combining BinaryHeads network with per-class thresholds, our algorithm is able to achieve higher balanced accuracy compared to previous work.

Conclusion

In this work:

- We empirically show that current OOD detection algorithms act unfairly when the in-distribution classes are imbalanced, by favouring the most numerous diseases in the training set.
- We developed a simple method to train and validate neural networks only on in-distribution dermoscopic skin disease images, which equips them with the additional ability to detect novel diseases from dermoscopic images at the test time.
- We introduce a method to investigate the effectiveness of OOD detection methods based on presence of varying amounts of OOD data, which may arise in real-world settings.

References

[1] D. Hendrycks and K. Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *ICLR*, 2018.

[2] Weitang Liu, Xiaoyun Wang, John D Owens, and Yixuan Li. Energy-based out-of-distribution detection. *arXiv preprint arXiv:2010.03759*, 2020.