CSCE 421 Final Project: AUC Maximization by Deep Learning for Imbalanced Medical Image Classification

Due date: Monday, May 06, 2024 (11:59pm). No Extension.

1 Introduction

In past decades, machine learning, especially deep learning methods have revolutionized many domains such as machine vision and natural language processing. In the realm of medical image analysis, deep learning has demonstrated promising capabilities such as in classification of skin lesions [3, 8], interpretation of chest radiographs [1, 7], and breast cancer screening [2, 11, 12]. Some works have already achieved expert-level performance in different tasks [1, 11, 9].

A common challenge in many medical imaging tasks is the imbalance in data availability, where the dataset predominantly consists of negative instances (images without any indication of disease) and a much smaller proportion of positive instances (images showing signs of disease). This imbalance renders accuracy as a unreliable metric, leading to the widespread preference for the AUC (Area Under the ROC Curve) metric as a more informative measure of performance.

In this project, you are required to train deep leaning models to maximize AUC for imbalanced medical imaging tasks.

2 Setup

Data You are asked to conduct experiments on two medical image classification tasks from MedMNIST [14] website (`https://medmnist.com`), namely BreastMNIST, PneumoniaMNIST.

Deep Neural Network ResNet18 [5] is mandatory, but experimenting with alternative network architectures is encouraged for potential enhancements.

Software The code in this package is a demo to load the MedMNIST data and do a standard training and evaluation. You need to install MedMNIST using "pip install medmnist"[1]. You need to install LibAUC library using "pip install -U libauc"[2].

Instruction on how to use High Performance Research Computing (HPRC) Resource can be found at this link[3]. You can either run on TAMU HPRC or Google Colab.

[1] `https://github.com/MedMNIST/MedMNIST`

[2] `https://github.com/Optimization-AI/LibAUC`

[3] `https://docs.google.com/document/d/1fbYU9mCFOoYa8ROk2szh2MtjoJvLptOAJsTLoXAQ7vo/edit?usp=sharing`

3 Requirements

Your project should have the following blocks, where the credits for each block is highlighted.

1. (**10%**) Train a ResNet18 with Cross Entropy Loss (referred to as CE in below).
2. (**20%**) Use the AUCM Loss in **LibAUC**[4] to train a ResNet18.

Background: Our group has developed a series of research on deep AUC maximization [15, 10, 4] for medical image classification and achieved success on some challenging tasks, such as the 1st place in CheXpert competition for chest X-ray image classification [7] and the 1st Place at MIT AICures Challenge [13].

3. (**70%**) Explore techniques to improve the performance based on training the AUCM Loss of LibAUC. Feel free to try anything. The next section will provide some possible techniques.

4 Examples of things you can work with

Below is a list of things that may help you start brainstorming your own solutions.

- Data Augmentations. Data augmentation is always a good strategy for improving the generalization of deep learning. The literature has proposed many data augmentations techniques. Is there a good data augmentation strategy working for AUC maximization?
- Control Overfitting. In the past, we have tried different approaches for controlling the overfitting, including L2 and L1 regularization. Would they work here and what other approaches are useful?
- Does optimizer matter? You could tune step size (aka learning rate) and batch size in the optimizer. You can also try other optimizers such as Momentum method or the Adam optimizer.
- Choice of neural network structures. You can try differe nerual networks such as ResNet50 (similar structure as ResNet18 but has more layers) [5], or DenseNet121 [6].
- Transfer learning is *out of scope*. While pretraining the network on a large external dataset (e.g., ImageNet) would improve the performance on a small data, however, this is out of scope of this project.

5 Submission and Grading

5.1 What to submit

You need to submit all of the following

- A pdf report, without which you will get 0.
- Training code
- Saved checkpoints of the model, i.e., saved trained models

[4]`https://libauc.org`

- Code to test the performance
- Instructions to run the training and testing code. Put this in a file named instructions.txt. It can have just a few lines.

5.2 Report

The report should be no less than 3 pages in one column style. No upper limit. It should be structured as an academic paper with the following components:

- Abstract
- Introduction
- Method
- Experiments

5.3 Grading Criteria

Your report's clarity will contribute 40% to your overall grade, and the technical soundness of your approach and code will account for 60%. While achieving a high testing AUC score is desirable, it is more important to demonstrate your efforts. Please report on the performance of different versions of your methods and what you have tried for each version, including why it worked or didn't work.

6 Teaming

We suggest you to have a team of two or three members and divide the work evenly among the team members. Teams with more than three persons are not allowed. It is okay if you choose to work on your own.

References

[1] Diego Ardila, Atilla P Kiraly, Sujeeth Bharadwaj, Bokyung Choi, Joshua J Reicher, Lily Peng, Daniel Tse, Mozziyar Etemadi, Wenxing Ye, Greg Corrado, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine*, 25(6):954–961, 2019.

[2] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22):2199–2210, 2017.

[3] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.

[4] Zhishuai Guo, Mingrui Liu, Zhuoning Yuan, Li Shen, Wei Liu, and Tianbao Yang. Communication-efficient distributed stochastic AUC maximization with deep neural networks. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 3864–3874. PMLR, 2020.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[6] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017.

[7] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn L. Ball, Katie S. Shpanskaya, Jayne Seekins, David A. Mong, Safwan S. Halabi, Jesse K. Sandberg, Ricky Jones, David B. Larson, Curtis P. Langlotz, Bhavik N. Patel, Matthew P. Lungren, and Andrew Y. Ng. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI)*, pages 590–597, 2019.

[8] Yuexiang Li and Linlin Shen. Skin lesion analysis towards melanoma detection using deep learning network. *Sensors*, 18(2):556, 2018.

[9] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[10] Mingrui Liu, Zhuoning Yuan, Yiming Ying, and Tianbao Yang. Stochastic AUC maximization with deep neural networks. In *8th International Conference on Learning Representations (ICLR)*, 2020.

[11] Scott Mayer McKinney, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, Hutan Ashrafian, Trevor Back, Mary Chesus, Greg S Corrado, Ara Darzi, et al. International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788):89–94, 2020.

[12] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H Beck. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*, 2016.

[13] Zhengyang Wang, Meng Liu, Youzhi Luo, Zhao Xu, Yaochen Xie, Limei Wang, Lei Cai, Qi Qi, Zhuoning Yuan, Tianbao Yang, and Shuiwang Ji. Advanced graph and sequence neural networks for molecular property prediction and drug discovery, 2021.

[14] Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data*, 10(1):41, 2023.

[15] Zhuoning Yuan, Yan Yan, Milan Sonka, and Tianbao Yang. Large-scale robust deep auc maximization: A new surrogate loss and empirical studies on medical image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.