# CS633 Final Project: Improving Generalization of Deep AUC Maximization for Medical Image Classification

Due date: May 08, 2023 (11:59pm)

## 1 Introduction

Deep AUC Maximization (DAM) is a new paradigm for learning a deep neural network by maximizing the AUC score of the model on a dataset. It has broad range of applications [15], including training ChatGPT [8]. After a series of research on non-convex optimization from our group [9, 13, 7, 2], Yuan et al. [16] demonstrated great success of DAM on multiple medical image classification tasks, e.g., the 1st place at Stanford CheXpert competition [5]. This technique has been adopted by many projects and achieved great success for solving different machine learning problems [4, 12, 10, 1, 3, 6] and achieved great success in solving real-world problems, e.g., the 1st Place at MIT AICures Challenge [11].

## 2 Project Goal: Improving the Generalization of DAM

Although DAM has achieved great success on large-scale datasets, it might easily overfit on small training data. For example, one a small dataset BreastMnist with 546 training samples, simply optimizing AUCM loss by our LibAUC library (lr = 0.01, margin = 1.0, epoch_decay = 0.003, weight_decay = 0.0001, batch_size=128), we only get 0.888 testing AUC score while optimizing the cross-entropy (CE) loss can get 0.901 AUC score. However, the AUC on the training data by DAM is much better than optimizing the CE loss. This indicates the DAM can overfit to small training data in terms of AUC score. This project aims to improve the generalization ability of DAM for medical image classification tasks.

## 3 Requirements

You need to download the LibAUC library (https://libauc.org), which has implemented a series of algorithms for optimizing AUROC, AUPRC, partial AUC, ranking measures, and other contrastive losses. You are asked to conduct experiments on 7 medical image classification tasks from MedMNIST website (`https://medmnist.com`), namely BreastMNIST, PneumoniaMNIST, ChestMNIST, NoduleMNIST3D, AdrenalMNIST3D, VesselMNIST3D, SynapseMNIST3D. Among these seven datasets, ChestMNIST is a multi-label classification tasks, and others are binary classification tasks. For ChestMNIST, each label is considered as a binary classification problems, and you need to report the averaged performance on all labels. Except for ChestMNIST, other datasets are relatively small. Your goal is to improve the benchmark performance reported in the MedMNIST paper [14]. For fair comparison, you need to use the same network structure as in the

MedMNIST paper. While we noticed that MedMNIST paper has tried multiple network structures. It is up to you to choose the network structure.

## 4 Grading

Your technical report and code will be the basis for grading. If you don't submit a report, you will receive a score of zero.

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.

This is a comprehensive project, and we expect to see your efforts in multiple directions. If you choose to focus on one area of improvement, we want to see that you have explored that direction to a deep level. A good project will demonstrate that you have some innovative ideas for improving the generalization of LibAUC.

## 5 Software

The code package on CANVAS is a demo to load the MedMNIST data and do a standard training and evaluation. For a more detailed and adavanced tutorial of loss functions for AUC maximizing loss such as the AUROC loss and Partial AUROC loss, please refer to https://libauc.org. In particular, coding instruction is under its "Example" sections. Instruction on how to use High Performance Research Computing (HPRC) Resource was shared on CANVAS/module/Student Resources. You can either run LibAUC on TAMU HPRC or Google Colab.

## 6 Teaming

We suggest you to find a partner to form a team of two persons and divide the work evenly among the team members. Teams with more than 2 persons are not allowed. It is OK if you choose to work on your own. But remember that there are multiple datasets you need to explore.

## 7 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 standard regularization (weight decay) and epoch-wise regularization (epoch decay). What other approaches are useful?

- Distributional shift between training/validation and testing datasets. Do training data, validation data and testing data have the same imbalance ratio (proportion of positive examples)? If not, then hyper-parameter tuning according to the validation data might not yield a good model on testing data. However, it is prohibited to use the testing data to do the hyper-parameter tuning and model selection. You can consider how to construct your own validation data (from training + validation) to do the model selection.

- 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.

- Data-centric approaches. This is a new paradigm that focuses the construction of a good training dataset. You can think about how to construct a good training data from the provided training and validation data.

- 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 since the baseline for optimizing CE loss would also benefit from transfer learning. However, co-training using all data with consistent data format is allowed. You need to submit the model and we will evaluate your model on the testing dataset.

## 8 Incentive Program

The full credit of the project is 100 points. We also give some incentive points according to

- Top 3 teams get 50 addtional points, and a certificate mentioning the 1st, 2nd, 3rd places.

- Top 5 teams get 30 additional points.

- Top 10 teams get 20 additional points.

To participate the incentive program, you need to use fixed network structure. For BreastMNIST, PneumoniaMNIST, ChestMNIST, you need to use ResNet-18 (28), and for NoduleMNIST3D, AdrenalMNIST3D, VesselMNIST3D, SynapseMNIST3D, you need to use Reset-18 + 3D.

## References


[1] Chih cheng Hsieh. Multimodal-xai-medical-diagnosis-system, 2022.

[2] Zhishuai Guo, Zhuoning Yuan, Yan Yan, and Tianbao Yang. Fast objective and duality gap convergence for non-convex strongly-concave min-max problems. *Journal of Machine Learning Research*, 2023.

[3] Siyuan He, PENGCHENG XI, Ashkan Ebadi, Stéphane Tremblay, and Alexander Wong. Performance or trust? why not both. deep auc maximization with self-supervised learning for covid-19 chest x-ray classifications. *Journal of Computational Vision and Imaging Systems*, 7(1):37–39, Apr. 2022.

[4] Mengda Huang, Yang Liu, Xiang Ao, Kuan Li, Jianfeng Chi, Jinghua Feng, Hao Yang, and Qing He. Auc-oriented graph neural network for fraud detection. In *Proceedings of the ACM Web Conference 2022*, WWW '22, page 1311–1321, New York, NY, USA, 2022. Association for Computing Machinery.

[5] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 590–597, 2019.

[6] Tharun Kumar. Rsna brain tumor classification, 2021.

[7] 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.

[8] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback, 2022.

[9] Hassan Rafique, Mingrui Liu, Qihang Lin, and Tianbao Yang. Non-convex min-max optimization: Provable algorithms and applications in machine learning. *Optimization Methods and Software*, 2020.

[10] Tencent Youtu Research. Heterogeneous interpolation on graph, 2021.

[11] 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.

[12] Lanning Wei, Huan Zhao, Quanming Yao, and Zhiqiang He. Pooling architecture search for graph classification. In *CIKM*, 2021.

[13] Yan Yan, Yi Xu, Qihang Lin, Wei Liu, and Tianbao Yang. Optimal epoch stochastic gradient descent ascent methods for min-max optimization. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, 2020.

[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), jan 2023.

[15] Tianbao Yang and Yiming Ying. Auc maximization in the era of big data and ai: A survey. *ACM Computing Surveys*, 2022.

[16] Zhuoning Yuan, Yan Yan, Milan Sonka, and Tianbao Yang. Robust deep AUC maximization: A new surrogate loss and empirical studies on medical image classification. In *Interntional Conference on Computer Vision*, volume abs/2012.03173, 2020.