Deep AUC Maximization (DAM)

Tianbao Yang

Department of Computer Science
The University of Iowa

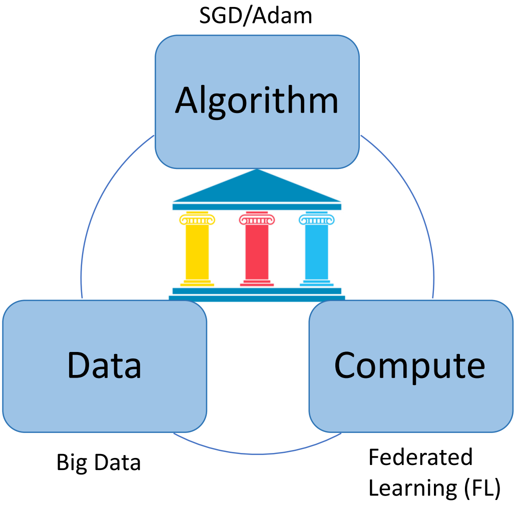
Outline

The AI Revolution

Deep Learning

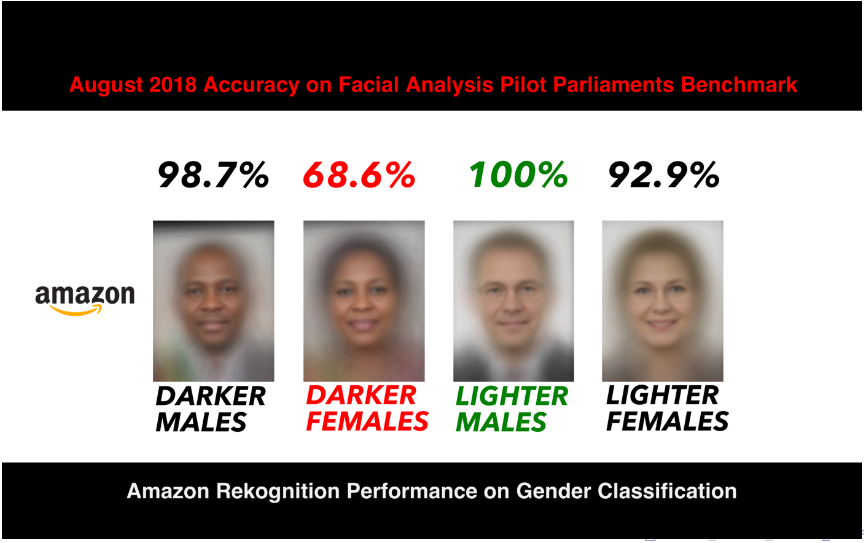
- AI beats human on Image Recognition (2015)
- AlphaGo beats human champion (2017)
- AI beats radiologists on interpreting X-ray images (2019)
- AlphaFold solves Protein Folding (2020)
- ...

Three Pillars of Deep Learning

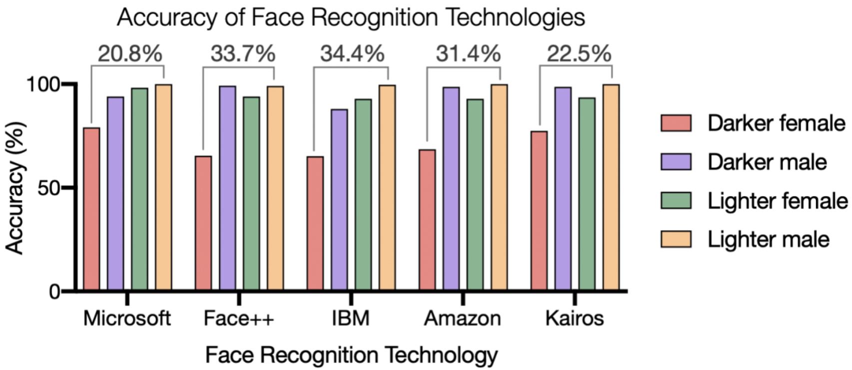


Challenges for Accelerating AI **Democratization**

Face Recognition



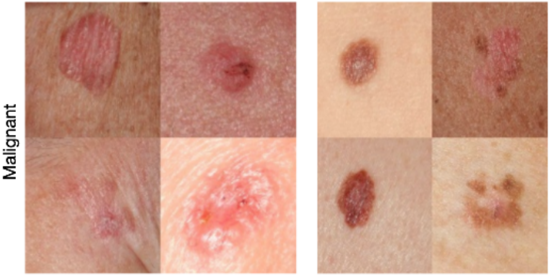
Challenges for Accelerating AI **Democratization**



(Buolamwini & Gebru, 2018. Gender Shades)

AI for Medical Image Classification

Dermatologist-level classification of skin cancer



Esteva et al. (Nature, 2017), reported AUC>0.91

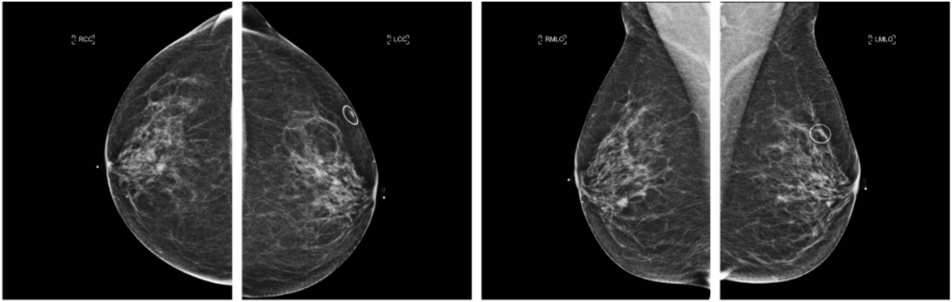
AI for Medical Image Classification

Radiologist-level Interpretation of X-ray images

Irvin, et al. (AAAI, 2019), reported AUC>0.90

AI for Medical Image Classification

Radiologist-level Screening of Breast Cancer



Wu, et al. (IEEE T. Medical Imaging, 2020), reported AUC=0.895

Keys to "Success" for Medical AI

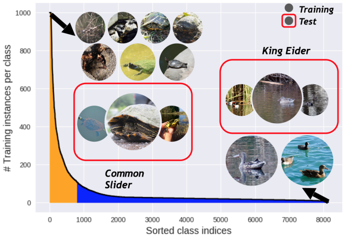
- Large-scale Datasets (100,000+ $\sim$ 1,000,000 images)
- Domain-specific techniques (e.g., network structures)

But Performance for Under-represented Classes could be Much Worse

Data Imbalance

is very common in real world

- Rare Disease Identification (e.g, Takotsubo)
- Terrorist Identification
- Credit Card Fraud Detection
- ...



picture courtesy: Jamal et al. 2020.

would cause

- dramatic performance drop
- unfairness, ethical issues

DL with Imbalanced Data Faces New Challenges

Performance Metrics of Imbalanced Data

- Accuracy
  - not suitable for imbalanced data

- Area under the Curve (AUC)
  - area under ROC curve (AUROC)
  - area under Precision-Recall curve (AUPRC)
  - widely used for evaluating the performance

How to Optimize AUC for Deep Learning?

Performance Metrics of Imbalanced Data

- Accuracy
  - not suitable for imbalanced data

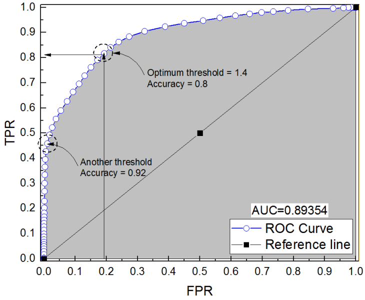
- Area under the Curve (AUC)
  - area under ROC curve (AUROC)
  - area under Precision-Recall curve (AUPRC)
  - widely used for evaluating the performance

How to Optimize AUC for Deep Learning?

Outline

AUROC

Area under ROC Curve



AUC Max. is more Difficult Accuracy Max.

Example 1		Example 2		Example 3	
Prediction	Ground Truth	Prediction	Ground Truth	Prediction	Ground Truth
0.9	1	0.9	1	0.9	1
0.8	1	**0.41**(↓)	1	**0.41**(↓)	1
0.7	1	0.7	1	**0.40**(↓)	1
0.6	0	0.6	0	**0.49**(↓)	0
0.6	0	**0.49**(↓)	0	**0.48**(↓)	0
0.47	0	0.47	0	0.47	0
0.47	0	0.47	0	0.47	0
⋮	⋮	⋮	⋮	⋮	⋮
0.1	0	0.1	0	0.1	0
Acc=0.92		Acc=0.92 (—)		Acc=0.92 (—)	
AUC=1.00		AUC=**0.89** (↓)		AUC=**0.78** (↓)	

AUC Surrogate Loss

$$\text{True-AUC}(h) = \Pr(h(\mathbf{x}) \geq h(\mathbf{x}')|y = 1, y' = -1)$$

- h: prediction model (e.g., deep neural network)
- $\mathbf{x}, \mathbf{x}'$ random data

$$\text{True-AUC}(h) = \mathrm{E}[\mathbb{I}(h(\mathbf{x}) - h(\mathbf{x}') \geq 0)|y = 1, y' = -1]$$

$$\min_h \text{AUC-Surrogate}(h) = \frac{1}{n_+}\frac{1}{n_-}\sum_{\mathbf{x}_i \in \mathcal{D}_+}\sum_{\mathbf{x}_j \in \mathcal{D}_-} \ell(h(\mathbf{x}_i) - h(\mathbf{x}_j))$$

AUC Surrogate Loss

$$\text{True-AUC}(h) = \Pr(h(\mathbf{x}) \geq h(\mathbf{x}')|y = 1, y' = -1)$$

- h: prediction model (e.g., deep neural network)
- $\mathbf{x}, \mathbf{x}'$ random data

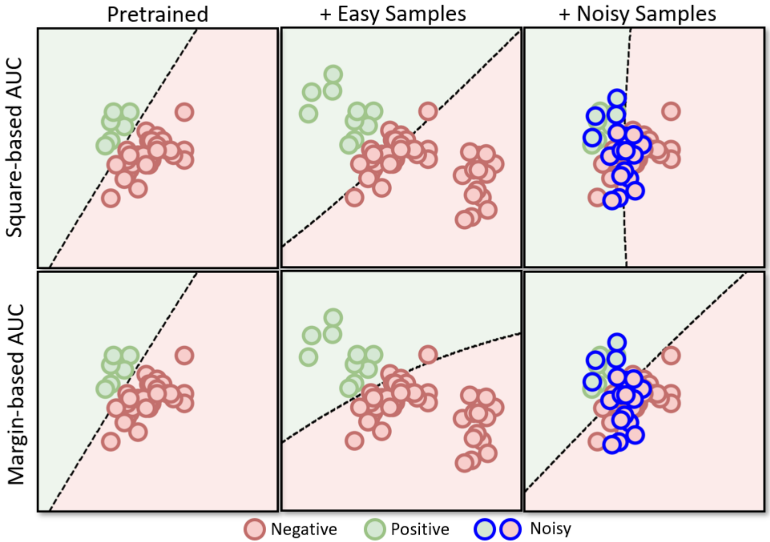
$$\text{True-AUC}(h) = \mathrm{E}[\mathbb{I}(h(\mathbf{x}) - h(\mathbf{x}') \geq 0)|y = 1, y' = -1]$$

$$\min_h \text{AUC-Surrogate}(h) = \frac{1}{n_+}\frac{1}{n_-}\sum_{\mathbf{x}_i \in \mathcal{D}_+}\sum_{\mathbf{x}_j \in \mathcal{D}_-} \ell(h(\mathbf{x}_i) - h(\mathbf{x}_j))$$

Challenges of Optimizing AUROC

- Scalability: scale up $> 10^6$ examples
- Robustness: robust to noise in the data
- Theoretical Guarantee: Yes, we are doing Science!

We Proposed a More Robust Approach



AUC Maximization: Zero-Sum Game Problem

Consider

$$\min_{\mathbf{w}} \max_{\alpha} f(\mathbf{w}, \alpha) = \mathrm{E}_{\mathbf{z}}[f(\mathbf{w}, \alpha, \mathbf{z})]$$

Stochastic Gradient Descent Ascent (SGDA)

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t \nabla_w f(\mathbf{w}_t, \alpha_t, \mathbf{z}_t), \quad \alpha_{t+1} = \alpha_t + \eta_t \nabla_\alpha f(\mathbf{w}_t, \alpha_t, \mathbf{z}_t)$$

Our Contributions

- First Proof of Convergence for Deep Learning
- Optimal Complexity Results

Summary of Our Theoretical Results

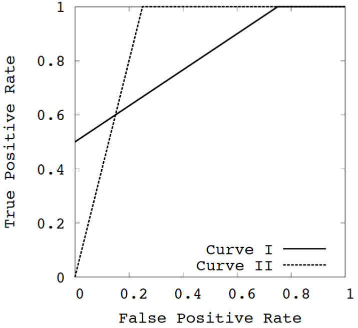
Table: Blue are our results. Red indicate optimal results. SC: strongly concave, PL: Polyak-Łojasiewicz condition. OGDA: optimistic gradient descent ascent.

Work	Conditions	Batch Size	$\mathcal{A}$	Sample Complexity
Rafique et al.'18	Concave	O(1)	SGDA	$O(\frac{1}{\epsilon^6})$
Rafique et al.'18	SC	O(1)	SGDA	$O(\frac{1}{\epsilon^4} + \frac{n}{\epsilon^2})$
Yan et al.'20	SC	$O(1)$	SGDA	$O(\frac{1}{\epsilon^4})$
Liu et al.'20	SC, PL	$O(1)$	SGDA AdaGrad	$O(\frac{1}{\mu^2\epsilon})$
Guo et al.'20	SC, PL	$O(1)$	OGDA STORM	$O(\frac{1}{\mu\epsilon})$
Lin et al.'19	Concave	$O(1)$	SGDA	$O(1/\epsilon^8)$
Lin et al.'19	SC	$O(1/\epsilon^2)$	SGDA	$O(1/\epsilon^4)$

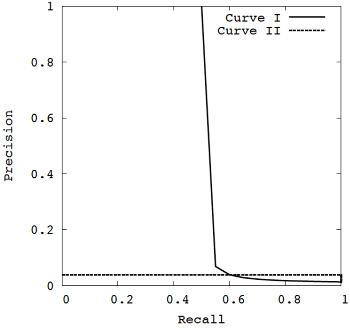
where ϵ is the accuracy level

AUPRC Maximization

Maximizing AUROC does not maximize AUPRC



(a) Comparing AUC-ROC for two algorithms



(b) Comparing AUC-PR for two algorithms

(picture courtesy: Davis&Goadrich, ICML'04)

Highly Imbalanced Data

AUROC vs AUPRC

Kaggle Melanoma Classification:

#	△pub	Team Name	Notebook	Team Members	Score	Entries	Last
1	▲880	All Data Are Ext			0.9490	116	1y
2	▲55	aloe			0.9485	61	1y
3	▲262	Deloitte Analytics Spain			0.9484	118	1y
4	▲210	Atagi Yuya			0.9476	23	1y
5	▲723	Wenlu			0.9475	19	1y
6	▲155	<^.^>			0.9468	168	1y
7	▲502	James Sebastian			0.9466	75	1y
8	▲218	Charlie			0.9463	58	1y
9	▲243	Rai			0.9462	90	1y
10	▲263	thakurudit			0.9461	67	1y
11	▲21	DSRGN			0.9459	387	1y

Our AUROC Maximization: 0.9438 (33/3314), But AUPRC is 0.19

AUPRC Maximization is even more Challenging

Mathematically Complex

$$\text{AUPRC} = \int_{-\infty}^{\infty} \Pr(Y = 1 | f(X) \geq c) d \Pr(f(X) \leq c | Y = 1),$$

Challenges of Optimizing AUPRC

- Much more Complex mathematical form
- Scalability: scale up $> 10^6$ examples.
- Theoretical Guarantee: Yes, we are doing Science!

Our Method: SOAP

$$\max_{h} \frac{1}{n_+} \sum_{\mathbf{x}_i \in \mathcal{D}_+} \frac{\mathrm{rank}(\mathbf{x}_i, \mathcal{D}_+; h)}{\mathrm{rank}(\mathbf{x}_i, \mathcal{D}; h)},$$

- $h(\mathbf{x})$: prediction network
- $\mathcal{D} = \{(\mathbf{x}_i, y_i), i = 1, \ldots, n\}$, $\mathcal{D}_+$ is the positive set
- Our Contributions: First Practical and Provable Algorithm

Outline

CheXpert Competition: Classifying X-ray Images

The 1st Place

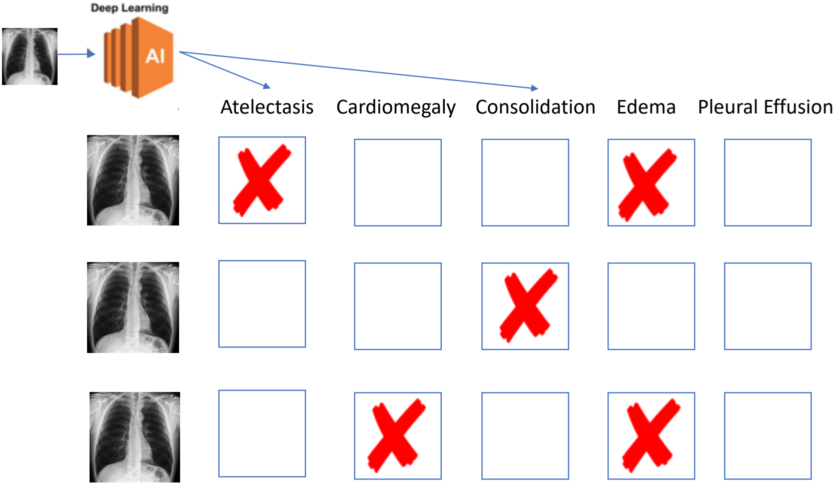
Stanford ML Group (Andrew Ng)
150+ teams worldwide

Leaderboard

Will your model perform as well as radiologists in detecting different pathologies in chest X-rays?

Rank	Date	Model	AUC	Num Rads Below Curve
1	Aug 31, 2020	DeepAUC-v1 *ensemble*	0.930	2.8

CheXpert Competition: Classifying X-ray Images



CheXpert Competition: Classifying X-ray Images

Data Set

- 224,316 chest X-rays images of 65,240 patients
- Only 5 selected diseases for evaluation: Atelectasis, Cardiomegaly, Consolidation, Edema, Pleural Effusion
- optimize CNNs

Results: 2%+ AUC improvement of DAM over standard DL

Model	AUROC	NRBC	Rank
Stanford Baseline (Irvin et al, AAAI'19)	0.9065	1.8	85
Hierarchical Learning (Pham et al. 2020)	0.9299	2.6	2
Ours (Yuan et al, 2020)	**0.9305**	**2.8**	**1**

CheXpert Competition: Classifying X-ray Images

Data Set

- 224,316 chest X-rays images of 65,240 patients
- Only 5 selected diseases for evaluation: Atelectasis, Cardiomegaly, Consolidation, Edema, Pleural Effusion
- optimize CNNs

Results: 2%+ AUC improvement of DAM over standard DL

Model	AUROC	NRBC	Rank
Stanford Baseline (Irvin et al, AAAI'19)	0.9065	1.8	85
Hierarchical Learning (Pham et al. 2020)	0.9299	2.6	2
Ours (Yuan et al, 2020)	**0.9305**	**2.8**	**1**

Kaggle Melanoma Classification Competition

Kaggle Competition

- May 27, 2020 - August 10, 2020
- 33,126 training images, with only 584 malignant melanoma samples

>2% AUC improvement of DAM over standard DL

Results in AUROC:

- Top 1% rank (ranked 33 out of 3314 teams)
- Ensemble: our (0.9438, 10 models) vs winner (0.9490, 18 models)
- Single Model: our (0.9423) vs winner (0.9167)
- Post-competition: DAM + standard DL gives 0.9503.

Kaggle Melanoma Classification Competition

Kaggle Competition

- May 27, 2020 - August 10, 2020
- 33,126 training images, with only 584 malignant melanoma samples

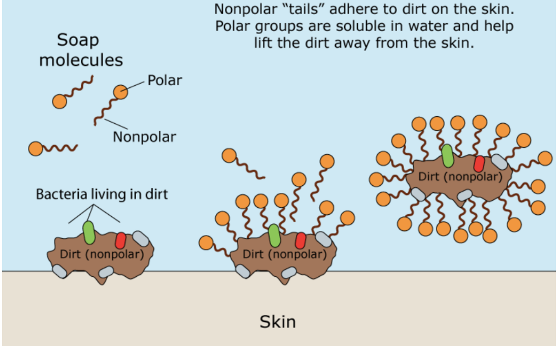
$>$2% AUC improvement of DAM over standard DL

Results in AUROC:

- Top 1% rank (ranked 33 out of 3314 teams)
- Ensemble: our (0.9438, 10 models) vs winner (0.9490, 18 models)
- Single Model: our (0.9423) vs winner (0.9167)
- Post-competition: DAM + standard DL gives 0.9503.

Molecules Property Prediction for Drug Discovery

Drug Discovery by predicting Antibacterial properties of molecules



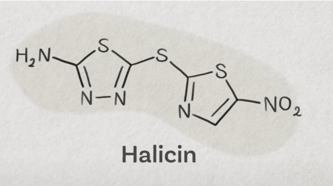
Molecules Property Prediction for Drug Discovery

Traditional Approach for Drug Discovery

- Expensive + Long Cycle

Machine Learning Approach for Drug Discovery

- Data-based for Molecules Properties Prediction
- Efficient & Fast (millions of molecules)



Stokes et al. 2020. Cell.

MIT AICures Challenge: 1st Place

Fighting Secondary Effects of Covid by predicting antibacterial properties

With DAM, > 5% AUPRC improvement and >2% AUROC improvement

- Collaboration with TAMU
- Optimize GNN
- The Original Result (without using DAM): AUPRC: 0.677

Rank	Model	Author	Submissions	10-fold CV ROC-AUC	10-fold CV PRC-AUC	Test ROC-AUC	Test PRC-AUC
1		DIVE@TAMU	11			0.957	0.729
2	MolecularG	AIDrug@PA	9			0.7	0.725
3		AGL Team	20			0.675	0.702
4		phucdoitoan@Fujitsu	14	0.898 +/- 0.113	0.508 +/- 0.253	0.867	0.694
5	GB	BI	6			0.698	0.67
6	Chemprop ++	AICures@MIT	4			0.877	0.662
7		Mingjun Liu	3			0.72	0.657
8	Pre-trained OGB-GIN (ensemble)	Weihua Hu@Stanford	2	0.905 +/- 0.133	0.494 +/- 0.333	0.837	0.651

MIT AICures Challenge: 1st Place

Fighting Secondary Effects of Covid by predicting antibacterial properties

With DAM, > 5% AUPRC improvement and >2% AUROC improvement

- Collaboration with TAMU
- Optimize GNN
- The Original Result (without using DAM): AUPRC: 0.677

Rank	Model	Author	Submissions	10-fold CV ROC-AUC	10-fold CV PRC-AUC	Test ROC-AUC	Test PRC-AUC
1		DIVE@TAMU	11			0.957	0.729
2	MolecularG	AIDrug@PA	9			0.7	0.725
3		AGL Team	20			0.675	0.702
4		phucdoitoan@Fujitsu	14	0.898 +/- 0.113	0.508 +/- 0.253	0.867	0.694
5	GB	BI	6			0.698	0.67
6	Chemprop ++	AICures@MIT	4			0.877	0.662
7		Mingjun Liu	3			0.72	0.657
8	Pre-trained OGB-GIN (ensemble)	Weihua Hu@Stanford	2	0.905 +/- 0.133	0.494 +/- 0.333	0.837	0.651

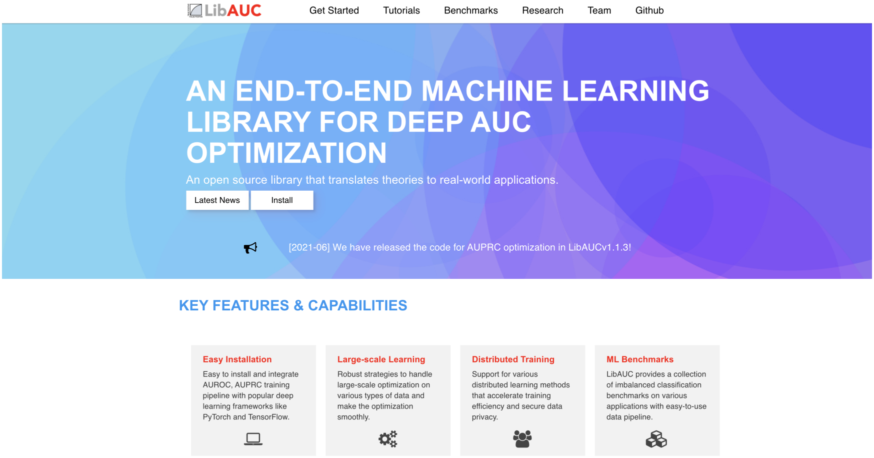
Outline

Conclusions

Our Achivements

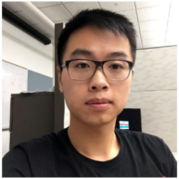
- A new learning paradigm for DL with imbalanced data
- Provable and Practical Stochastic Algorithms
- For AUROC and AUPRC Maximization
- The 1st Place at Stanford CheXpert Competition
- The 1st Place at MIT AICures Challenge

LibAUC: www.libauc.org



Acknowledgements: Students

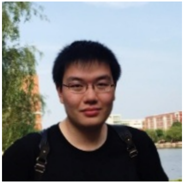
Current and Former PhD Students and Postdoc:



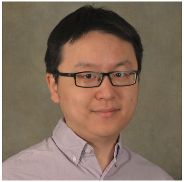
Zhuoning Yuan

Zhishuai Guo

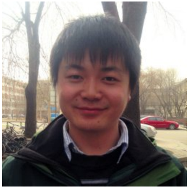
Qi Qi



Mingrui Liu
(AP, GMU)



Yi Xu
(Alibaba)



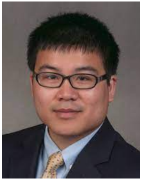
Yan Yan
(AP, WSU)

Acknowledgements: Collaborators

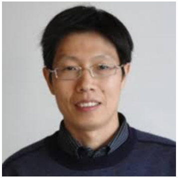
Milan Sonka (UIowa)

Nitesh Chawla (ND)

Hassan Rafique (UIndy)



Qihang Lin (UIowa)



Yiming Ying (UAlbany)

Shuiwang Ji (TAMU)

Acknowledgements: Funding Support

Funding Support from NSF Career Award

References

This talk include some results from the following Papers:

1. *Non-Convex Min-Max Optimization: Provable Algorithms and Applications in Machine Learning.* Optimization Methods and Software, 2020 (2018).
2. *Stochastic AUC Maximization with Deep Neural Networks.* ICLR'20.
3. *Communication-Efficient Distributed Stochastic AUC Maximization with Deep Neural Networks.* ICML'20.
4. *Optimal Epoch Stochastic Gradient Descent Ascent Methods for Min-Max Optimization.* NeurIPS'20.
5. *Federated Deep AUC Maximization for Heterogeneous Data with a Constant Communication Complexity.* ICML'21.
6. *Fast Objective and Duality Gap Convergence for Non-convex Strongly-concave Min-max Problems.* arXiv, 2020.
7. *Robust Deep AUC Maximization: A New Surrogate Loss and Empirical Studies on Medical Image Classification.* arXiv, 2020.
8. *Stochastic Optimization of Areas Under Precision-Recall Curves with Provable Convergence.* arXiv, 2021.

THANK YOU!

QUESTIONS?

Collaborations are more than Welcome!