AI for Health: Learning from Imperfect Data

Le Peng[1], Yash Travadi[2], Ju Sun[1]

[1]Department of Computer Science and Engineering, University of Minnesota

[2]School of Statistics, University of Minnesota

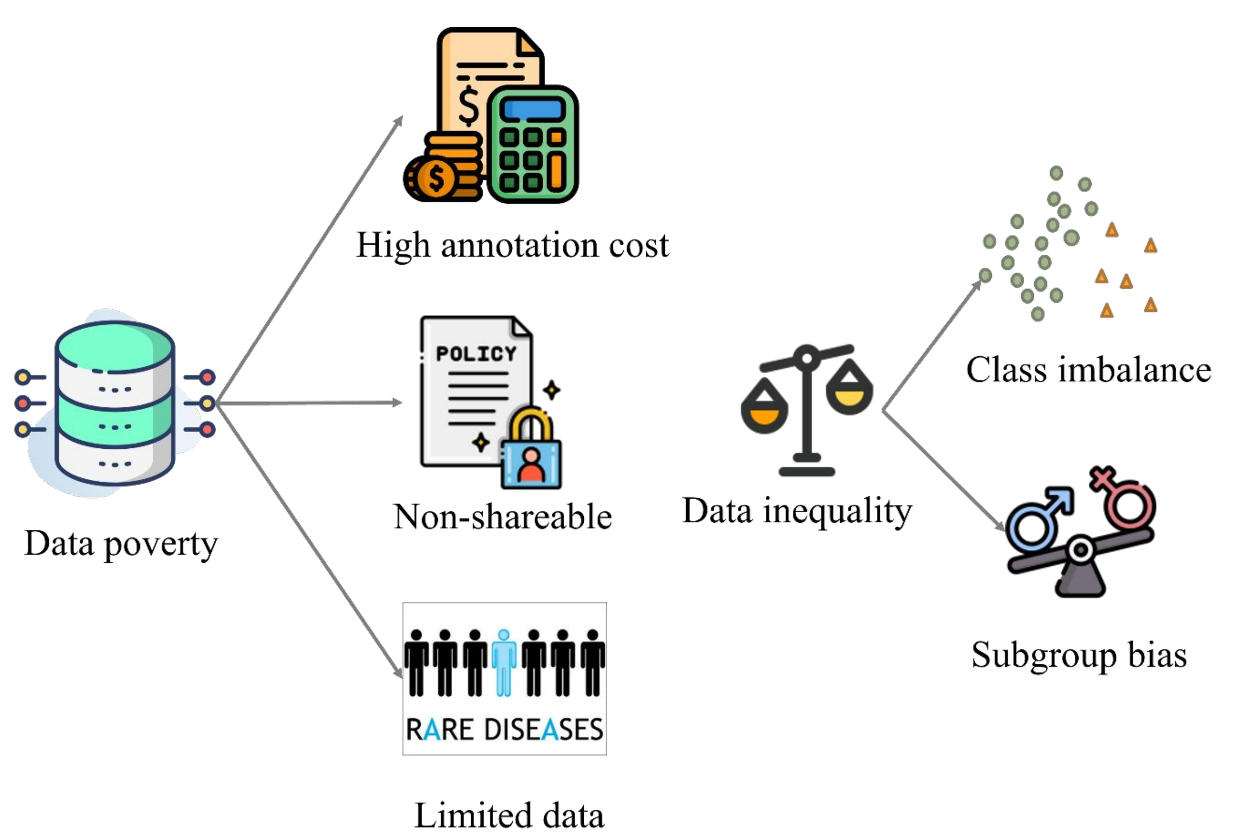
GLOVEX

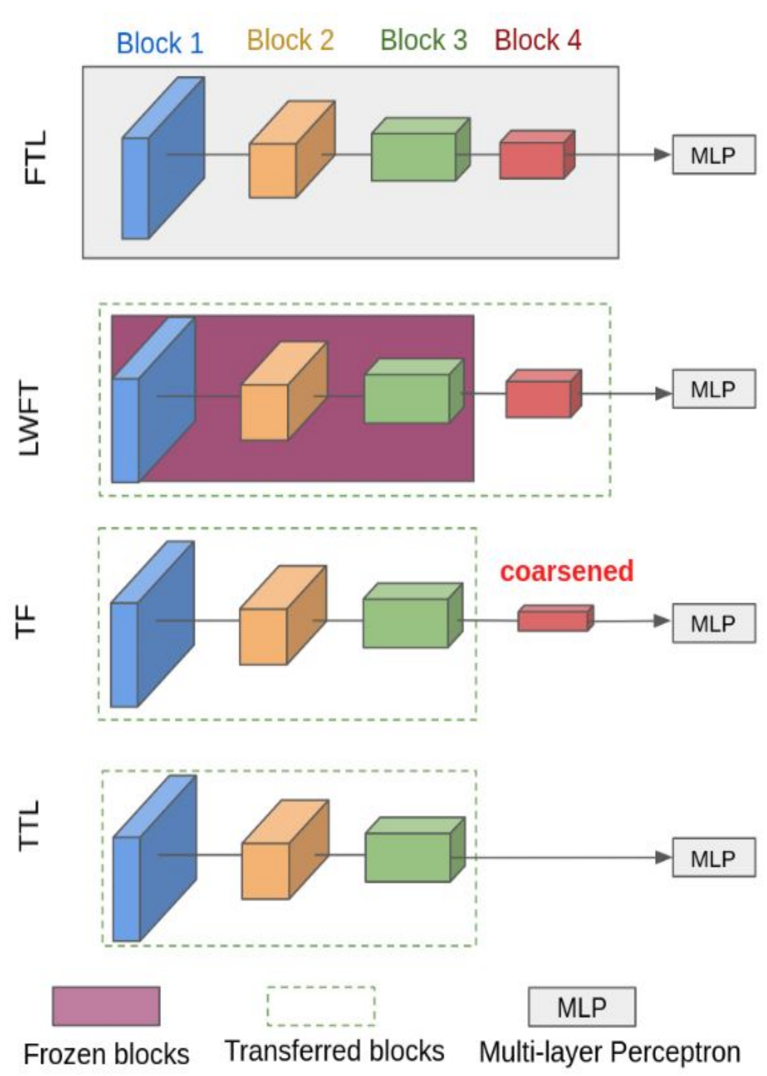
Group of **L**earning, **O**ptimization, **V**ision, healthcar**E** , and **X**

Objective and Background

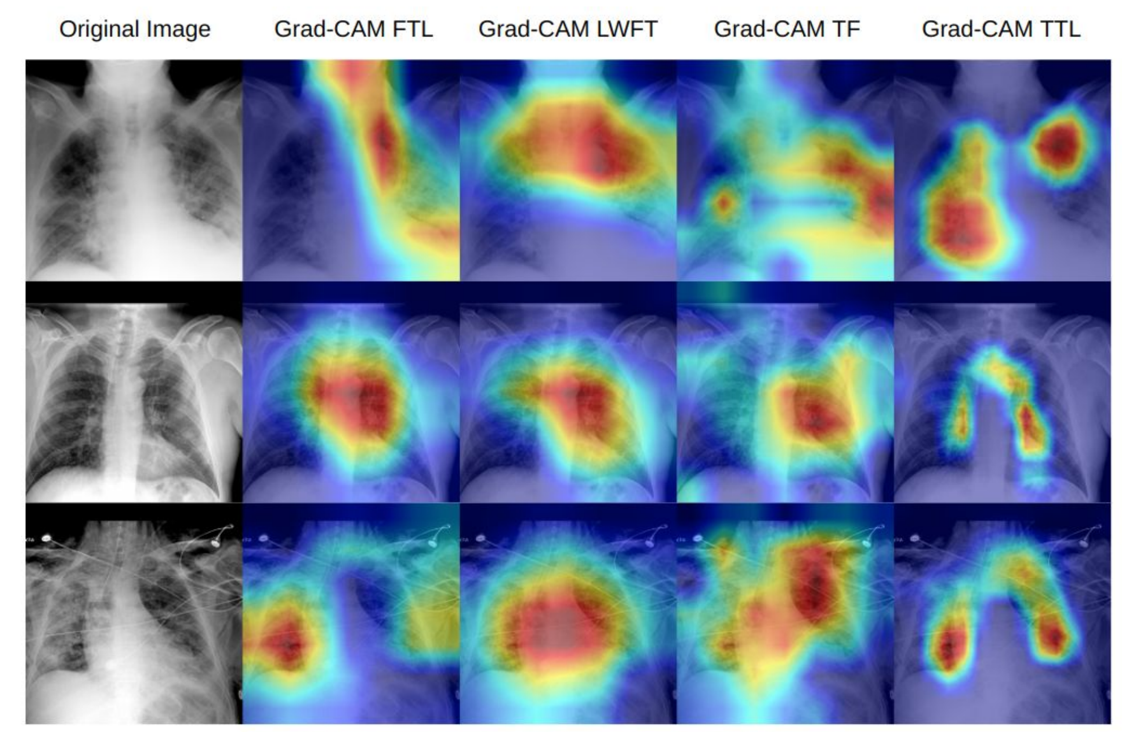
- Deep learning (DL) are data hungry
- Popular DL tools designed on well-curated data

But, medical data face





TTL vs others

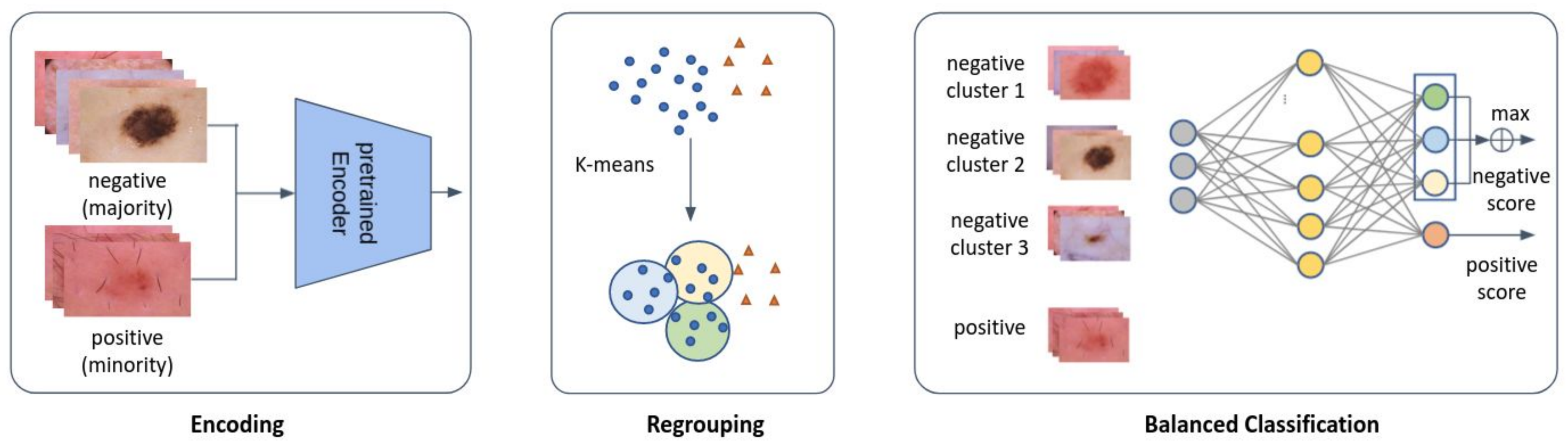


Method	AUROC↑	AUPRC↑	Params(M)↓	MACs(G)↓	CPU(ms)↓	GPU(ms)↓
FTL	0.849 ± 0.001	0.857 ± 0.003	23.5	4.12	79.6	3.59
(l)1-7 TF-1	0.856 ± 0.011	0.863 ± 0.012	12.9	3.56	67.0	3.55
LWFT-1	0.848 ± 0.002	0.861 ± 0.004	23.5	4.12	76.9	3.59
TTL-1	0.851 ± 0.002	0.860 ± 0.002	8.55	3.31	59.7	3.19
TF-2	0.856 ± 0.011	0.863 ± 0.012	12.9	3.56	72.7	3.56
LWFT-2	0.853 ± 0.005	0.861 ± 0.001	23.5	4.12	79.7	3.56
TTL-2 (ours)	**0.861 ± 0.013**	**0.871 ± 0.008**	**6.31**	**2.87**	**53.1**	**2.97**

Improved performance with lower costs!

Imbalance Learning

Imbalance learning via regrouping



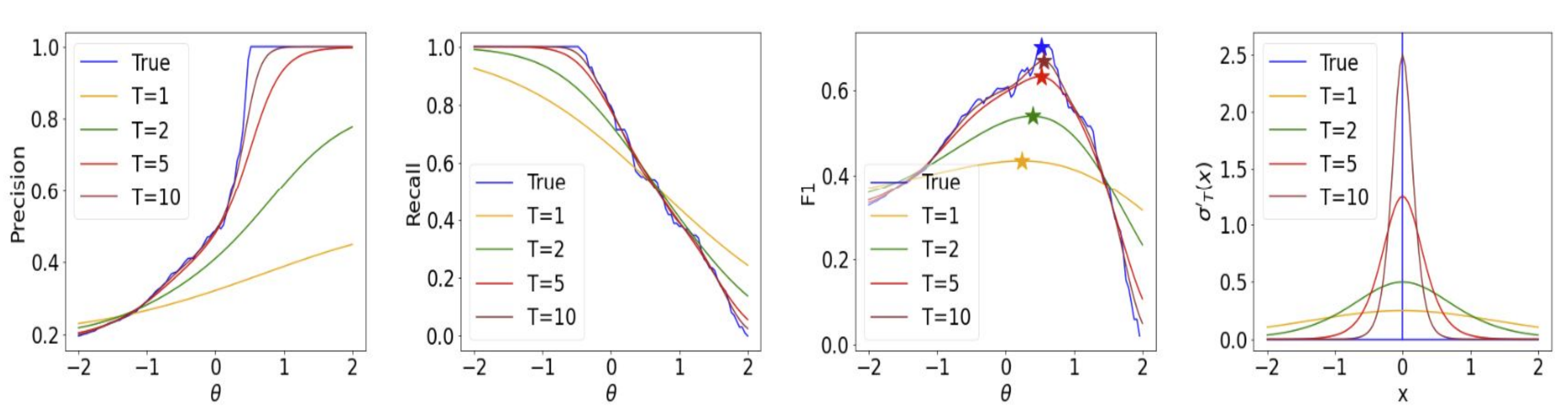
Direct metric optimization

$$
\begin{aligned}
\textbf{FPOR:} &\quad \max_{\boldsymbol{\theta},t}\ \mathrm{recall}(f_{\boldsymbol{\theta}},t) \quad \mathrm{s.t.}\ \mathrm{precision}(f_{\boldsymbol{\theta}},t)\geq\alpha,\\
\textbf{FROP:} &\quad \max_{\boldsymbol{\theta},t}\ \mathrm{precision}(f_{\boldsymbol{\theta}},t) \quad \mathrm{s.t.}\ \mathrm{recall}(f_{\boldsymbol{\theta}},t)\geq\alpha,\\
\textbf{OFBS:} &\quad \max_{\boldsymbol{\theta},t}\ F_\beta(f_{\boldsymbol{\theta}},t)\\
\textbf{OAP:} &\quad \max_{\boldsymbol{\theta}}\ \mathrm{AP}(f_{\boldsymbol{\theta}}).
\end{aligned}
$$

Example: reformulation for FPOR

$$
\begin{aligned}
\max_{\boldsymbol{\theta},\boldsymbol{s},t}\quad & \frac{1}{n_+}\sum_{i\in\mathcal{P}} s_i\\
\text{s.t.}\quad & \sum_{i\in\mathcal{P}} -(1-\alpha)s_i + \sum_{i\in\mathcal{N}} \alpha s_i \leq 0,\\
& \max(s_i + f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - t - 1, 0) -\\
&\quad \max(-s_i, f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - t) \geq 0 \quad \forall i\in\mathcal{N},\\
& \max(s_i + f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - t - 1, 0) -\\
&\quad \max(-s_i, f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - t) \leq 0 \quad \forall i\in\mathcal{P}
\end{aligned}
$$

Relaxation to indicator functions is problematic



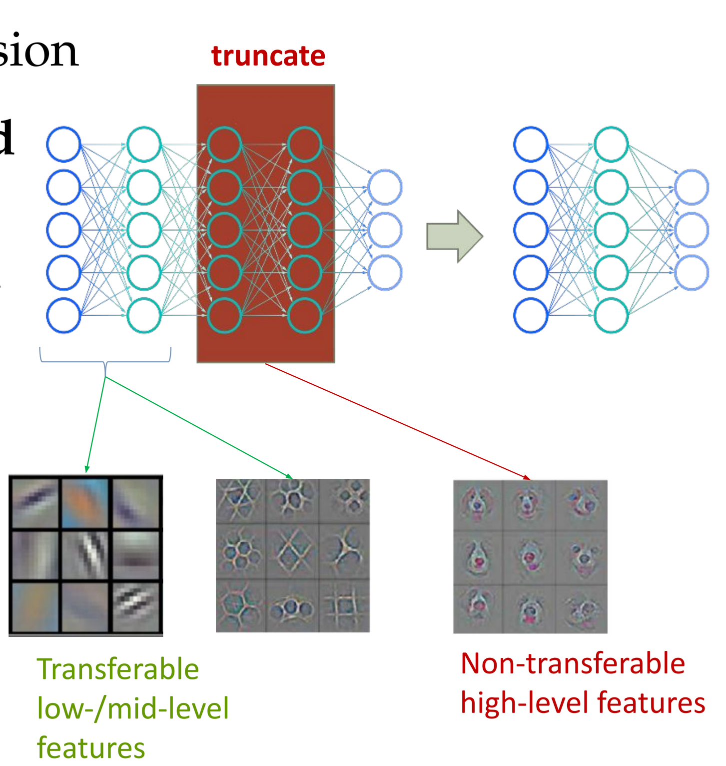
Transfer Learning

Truncated Transfer Learning (TTL)

□ **Compact models**
- Up to 75% compression

□ **Fast inference speed**
- Up to 25% speedup

□ **Great compatibility**
- Working with major DL models



Visit our project

- Low-level: general, task-invariant
- High-level: specific, task-dependent

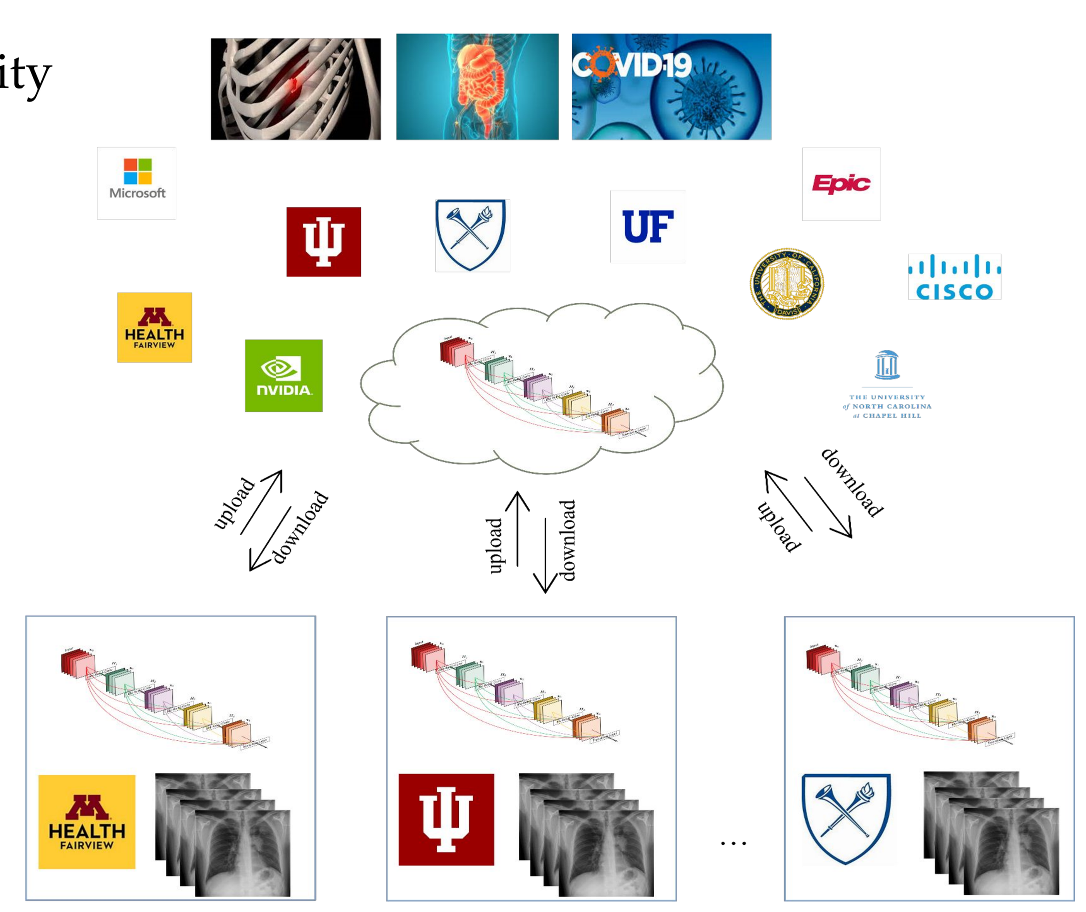
Federated Learning

- Mitigate the data poverty & inequality
- Respect data privacy

Our work covers:

□ Computer Vision
- COVID-19 detection
- Rib fracture detection

□ Natural Language Processing
- Medical information extraction (e.g., NER and RE))



Our federation highlighted in NVIDIA white paper on federated learning

References

[1] Sun, J., **Peng, L.**, et al., (2022) Performance of a chest radiograph ai diagnostic tool for COVID-19: a prospective observational study. Radiology: Artificial Intelligence, 4(4),.

[2] **Peng, L.**, et al.. (2021). Rethink Transfer Learning in Medical Image Classification. submitted to MedIA.

[3] **Peng, L.**, et al. (2022). Evaluation of Federated Learning Variations for COVID-19 diagnosis using Chest Radiographs from 42 US and European hospitals. Accepted to JAMIA.

[4] **Peng, L.**, et al., (2022) Imbalanced Data Classification using Regrouping. In preparation for JMLR.

[5] **Peng, L.**, et al., (2022) An empirical study on imbalanced classification. In preparation for JMLR.