Improved Multimodal Fusion for Small Datasets with Auxiliary Supervision

Gregory Holste[1,2], Douwe Van der Wal[2], Hans Pinckaers[2], Rikiya Yamashita[2], Akinori Mitani[2], Andre Esteva[2]

[1] The University of Texas at Austin, [2] Artera AI

MOTIVATION

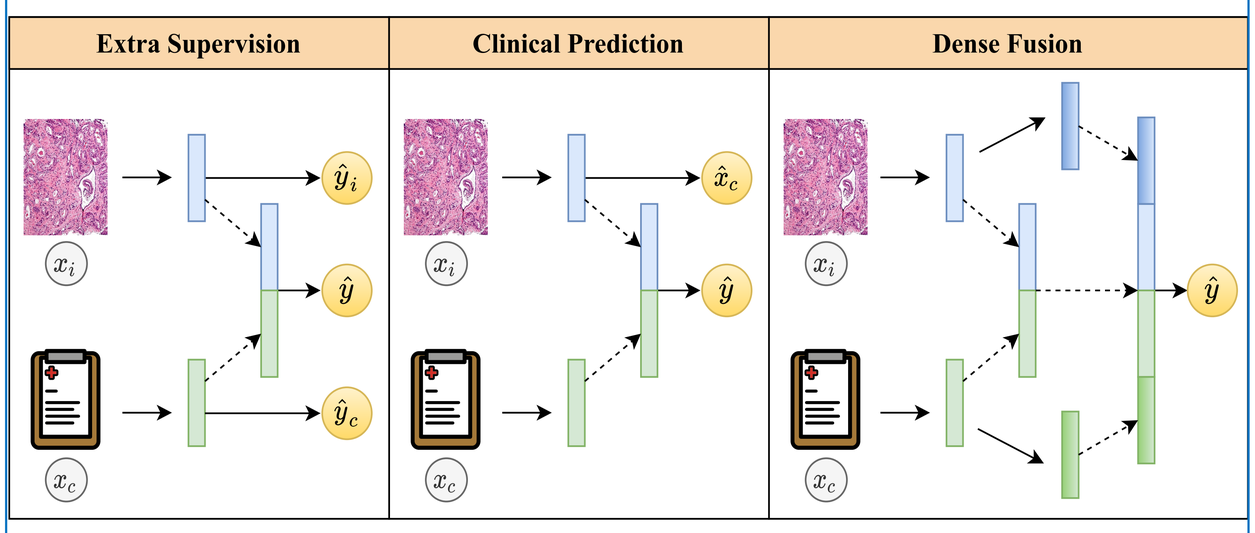
- Background:
  - Prostate cancer is a common, deadly disease
  - Diagnosis is **multimodal**: *histopathology imaging* + structured *clinical risk factors*
- Problem: Existing methods for fusing histopathology imaging + non-image data are extremely expressive (ex: Kronecker fusion)
  - Likely to overfit small/low-dimensional data
- Question: Can we develop a *parameter-efficient* method to learn from multimodal medical data?

DATASET & TASK

- 4,581 patients from five phase III clinical trials w/ paired histopathology imaging + clinical data
  - Kx128-dim "bag" of pre-extracted image features
  - 6-dim vector of numeric clinical features
    - Age, PSA, T-stage, Gleason scores
- Goal: Predict prostate cancer distant metastasis (DM)

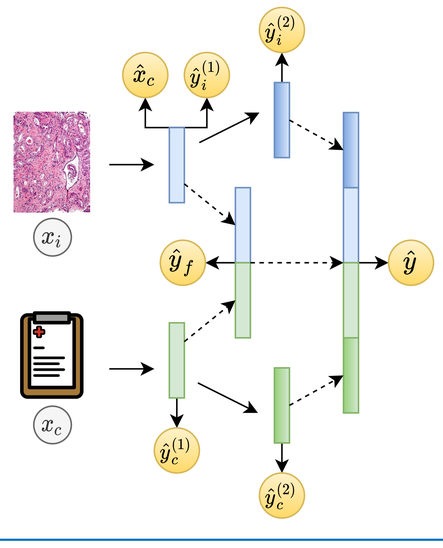
METHODS

- Solution: Use **auxiliary sources of supervision**



- **Extra supervision**: generate additional image-only and clinical-only outcome predictions
  - Minimize sum of cross-entropy (CE) losses
- **Clinical prediction**: use image-only features to predict/regress associate non-image inputs
  - Minimize simple MSE loss
- **Dense fusion**: encourage dense interaction of image-only, clinical-only, and fused features

- These approaches are all *complementary*
  - Ex: a combination of all 3 methods can be optimized with 6 CE losses + 1 MSE loss



EXPERIMENTAL SETUP

- Train baseline *late joint fusion* approach that concatenates image and non-image features
- Exp. #1: Observe effect of auxiliary supervision on concatenation vs. Kronecker fusion
- Exp. #2: Isolate which auxiliary supervision methods improve upon the baseline (ablation)
- Use 5-fold cross-validation (CV) with identical base architecture and hyperparameters
- Evaluate with **mean AUC** across CV folds

RESULTS

Fusion Operation	Auxiliary Supervision	# Params	AUC
Concatenation		17.1K	0.781 ± 0.024
Kronecker		66.3K	0.770 ± 0.018
Concatenation	✓	43.1K	$\mathbf{0.792 \pm 0.014}$
Kronecker	✓	207.1K	0.781 ± 0.013

Experiment #1:

- Auxiliary supervision improves performance
- Kronecker fusion increases parameter count by 4-5x and decreases performance

Extra Supervision	Clinical Prediction	Dense Fusion	AUC
			0.781 ± 0.024
✓			0.778 ± 0.021
	✓		0.789 ± 0.013
		✓	0.776 ± 0.025
✓	✓		0.787 ± 0.015
✓		✓	0.785 ± 0.016
	✓	✓	$\underline{0.790 \pm 0.012}$
✓	✓	✓	$\mathbf{0.792 \pm 0.014}$

Experiment #2:

- Combination of all three methods works best
  - Gain in performance is not purely additive
- Clinical prediction is the single most impactful
  - Highest mean AUCs and lowest std AUCs

ACKNOWLEDGMENTS

This project was supported by U10CA180822 (NRG Oncology SDMC), U10CA180868 (NRG Oncology Operations), and U24CA196067 (NRG Specimen Bank) from the National Cancer Institute.

Special thanks to the entire AI team at Artera (https://artera.ai/)!

PAPER