A framework for river connectivity classification using temporal image processing and attention based neural networks

Timothy James Becker*, Derin Gezgin, JunYi He Wu
Department of Computer Science
Connecticut College

Mary Becker
Water Monitoring and Assessment
Connecticut Department of Energy and Environmental Protection

*corresponding author

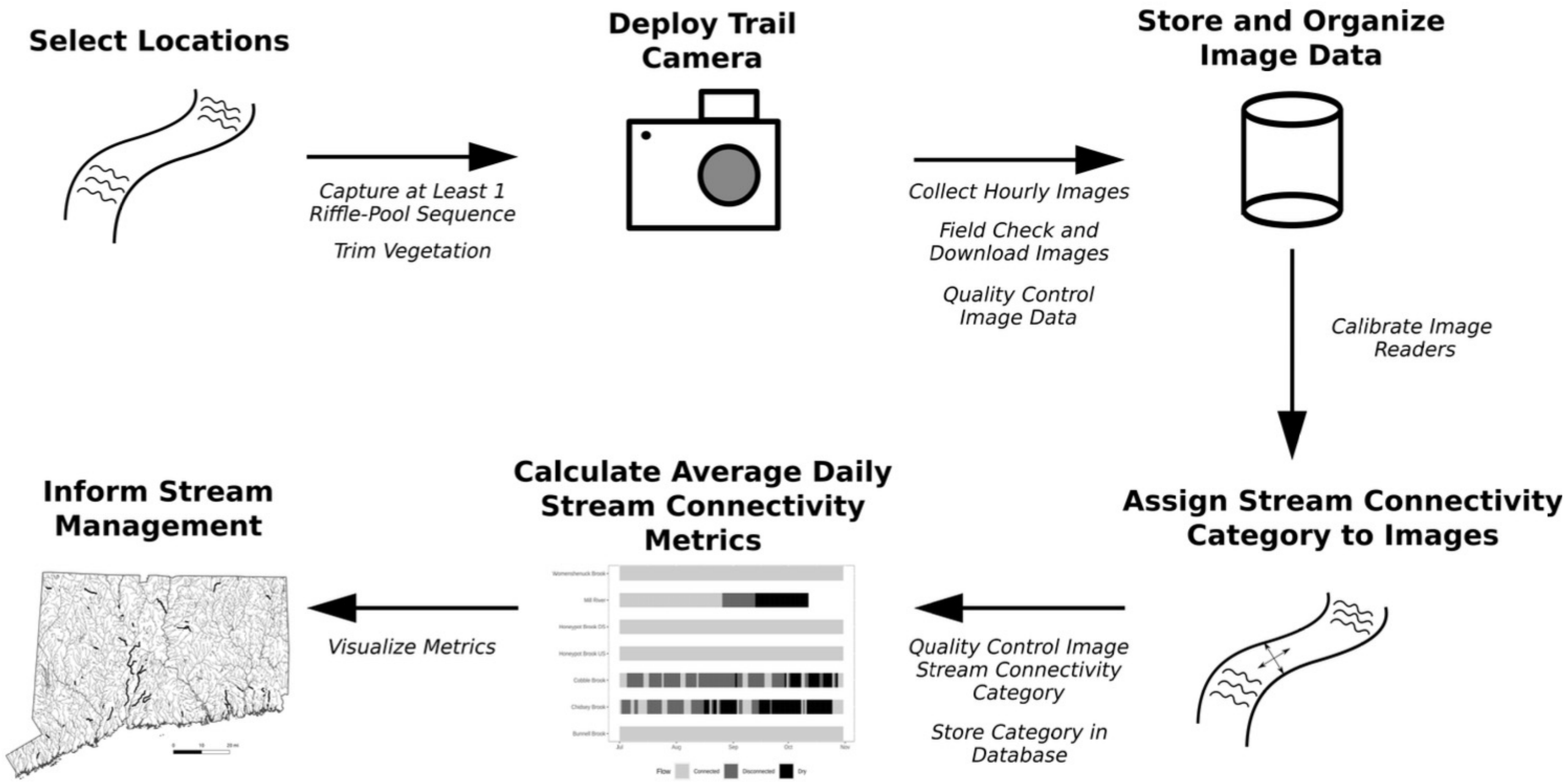
COMPASS2025 (Toronto, CA) 7/25/25
8th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies

Introduction: stream connection background

I

- **stream connectivity best informs management**
- **hourly trail camera capture is:**
  - cost effective
  - deploys into sensitive areas (headwaters)
  - informative (once QCed and labeled)
  - issues (staff time labeling could be improved)

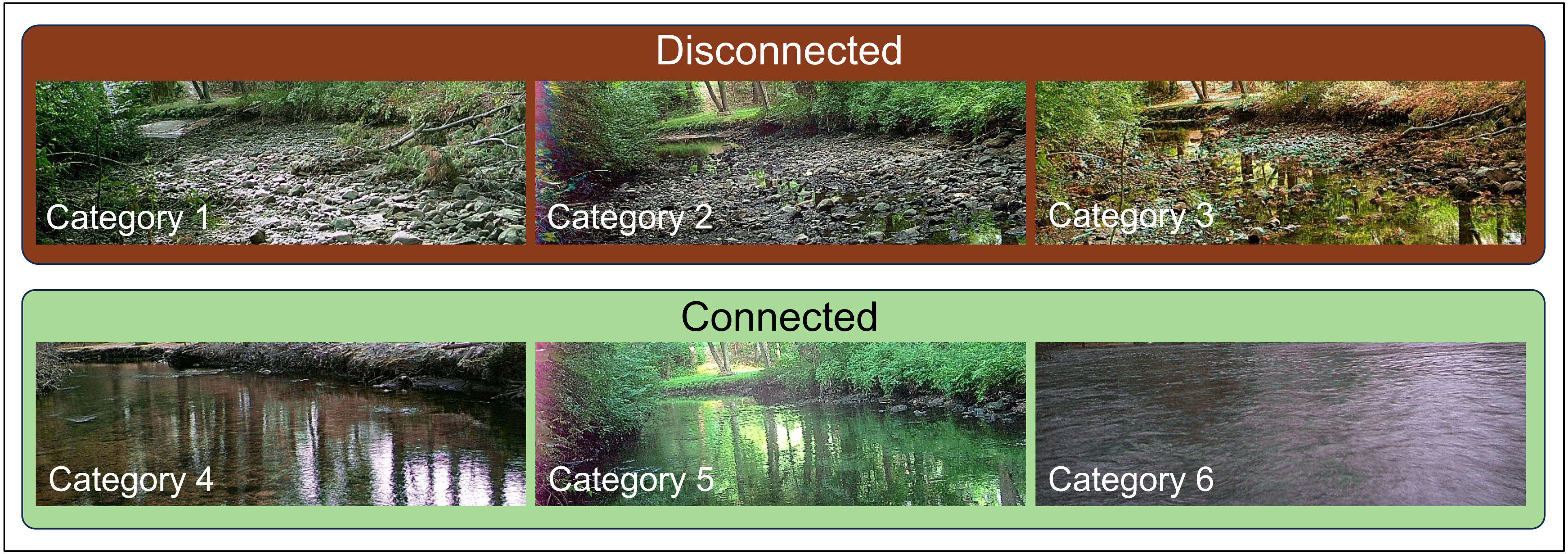
Introduction: stream connection monitoring



Introduction: stream connection states

I

- same site in all six connection states



Bellucci, CJ, Becker, ME, Czarnowski, M, Fitting, C. A novel method to evaluate stream connectivity using trail cameras. River Res Applic. 2020; 36: 1504– 1514.

Methods: accuracy on a new unseen site

- **partition data to keep sites unseen in model training**

II

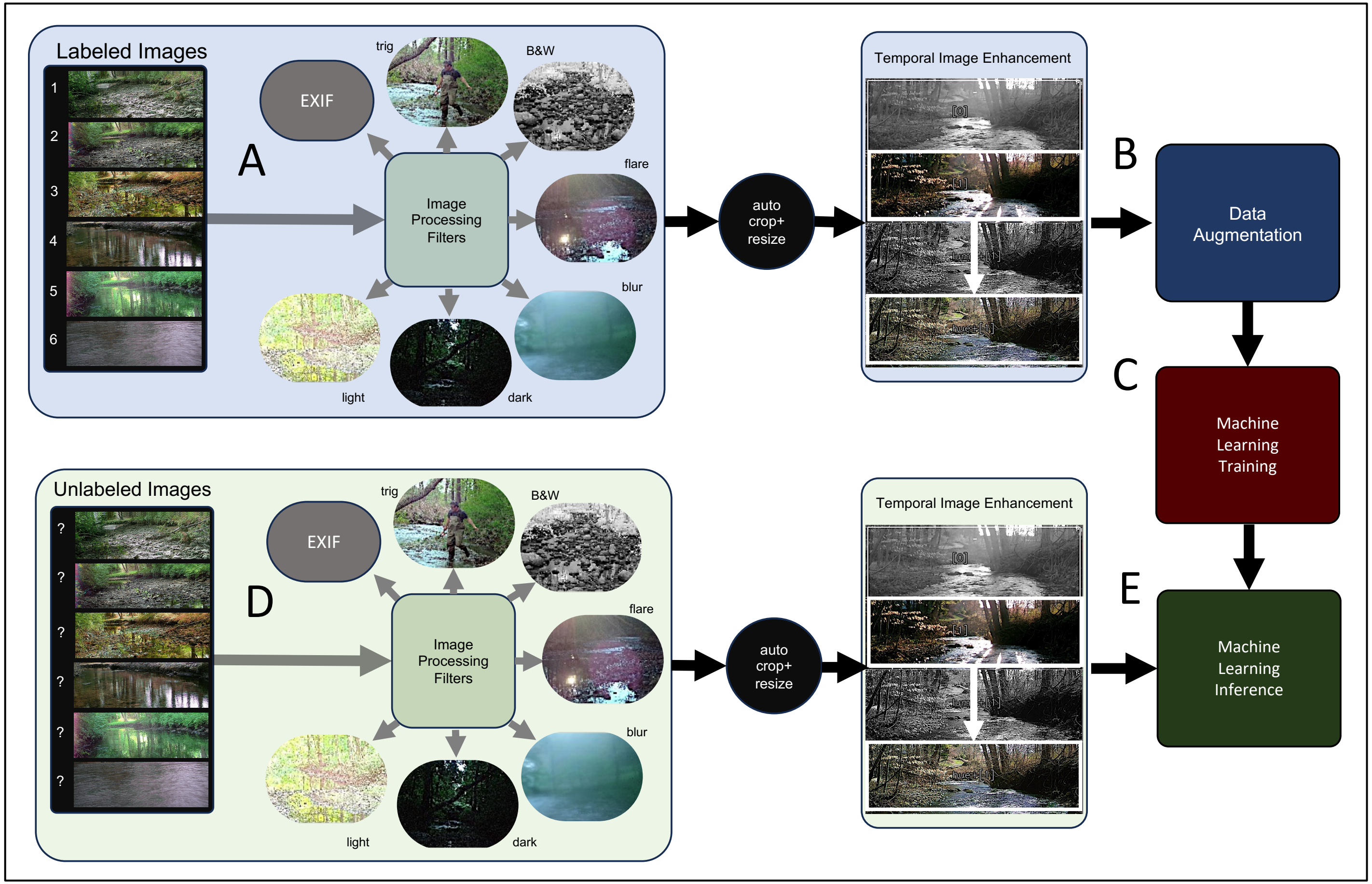
$$\{[X_{s_1}^{i1},\dots,X_{s_1}^{n_1}],\ ,\dots\ [X_{s_m}^{i1},\dots,X_{s_m}^{n_m}]\} \qquad \{[Y_{s_1}^{i1},\dots,Y_{s_1}^{n_1}],\ ,\dots\ [Y_{s_m}^{i1},\dots,Y_{s_m}^{n_m}]\}$$

$$\underset{y\in L}{\mathrm{P}(\mathbb{Y}=\mathrm{y})} = \left[\frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} b\left(Y_{s_i}^{n_j}, y\right)}{\sum_{i=1}^{m}\sum_{j=1}^{n_i} 1}\right] : b(y,l) = \begin{cases} 0, & y = l \\ 1, & else \end{cases} \quad (1)$$

$$H(s_i,l) = \sum_{j=1}^{n_i} b\left(Y_{s_i}^{n_j}, l\right) \quad (2)$$

$$\delta(u,t) = \left|\sum_{l=1}^{|L|}\left(\sum_{i\in u} H(S_i,l) - \sum_{j\in t} H(S_j,l)\right)\right| \quad (3)$$

Methods: framework overview



A - preprocessing

B - augmentation

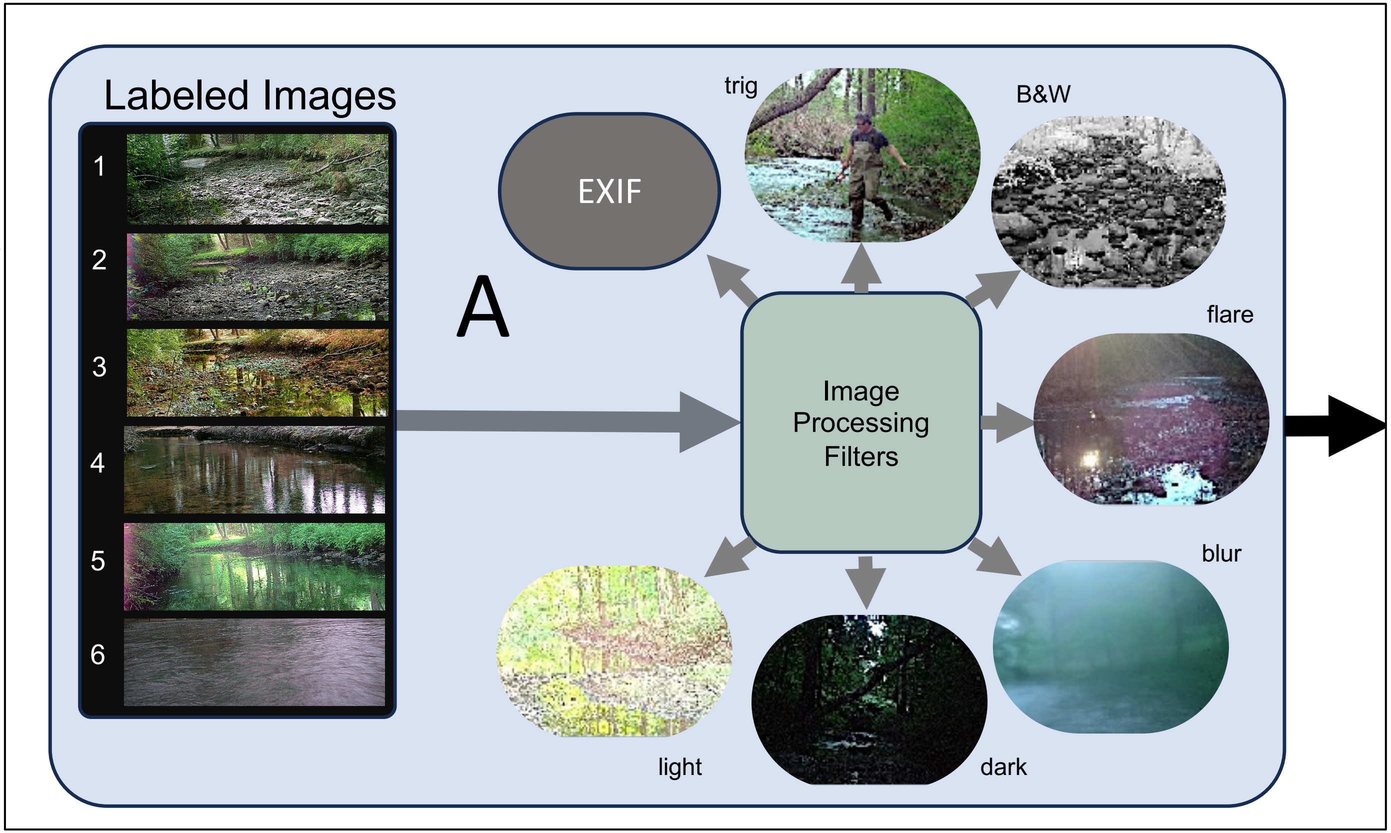
C - training

D - preprocessing

E - inference

Methods A: preprocessing

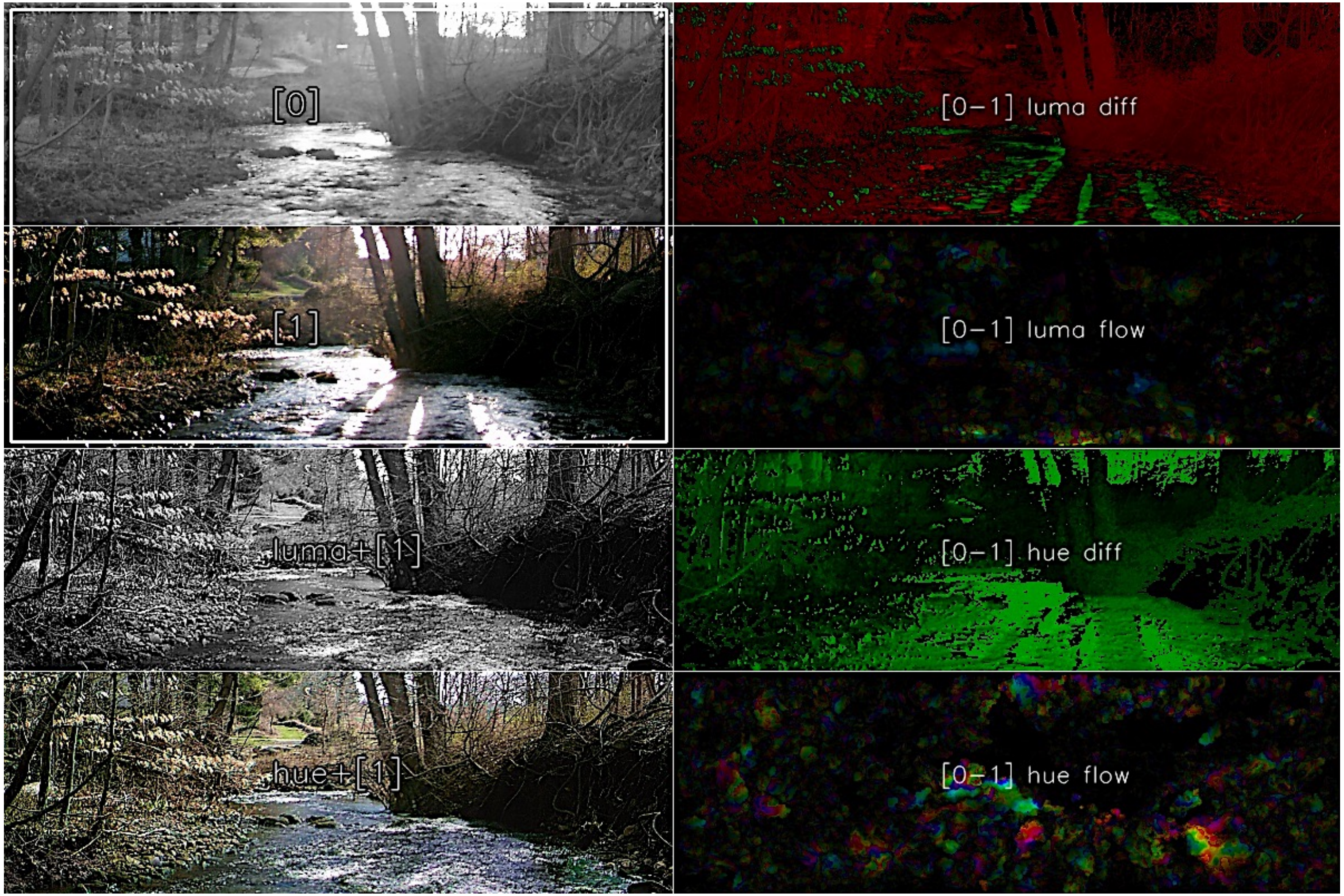
II



- **5 quality filters:**
  - B&W
  - Lens Flare
  - Blur
  - Dark
  - Light
- **2 metadata filters:**
  - Triggered
  - EXIF issues

Methods A: temporal enhancement

II



- Sorted by time
- Composited
- Luma channel
- Hue channel
- Homography check

Methods: dataset 2018-2020 CT DEEP labeled

II

Bellucci, CJ, Becker, ME, Czarnowski, M, Fitting, C. A novel method to evaluate stream connectivity using trail cameras. River Res Applic. 2020; 36: 1504– 1514.

~98K filtered 600x200 RGB images processed using the SRIP framework (9.7GB compressed): https://zenodo.org/records/14681118

Methods C/E: ViT and DCNN

II

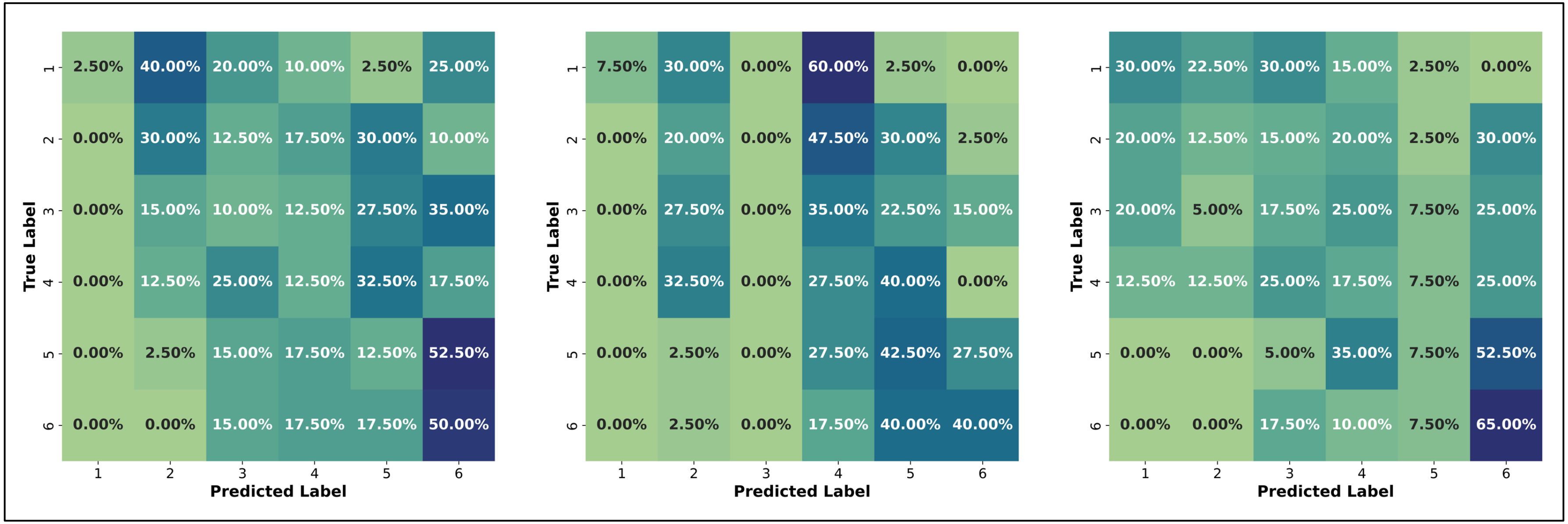
- **Multi-head attention-based vision transformer ViT model***
  - easy to setup and train
  - finds gradient patterns via patches

- **A Deep Convolutional (DCNN) RESNET style***
  - less overhead than ViT
  - finds gradient patterns using convolutional layers
  - requires more training

*using python keras 3 API

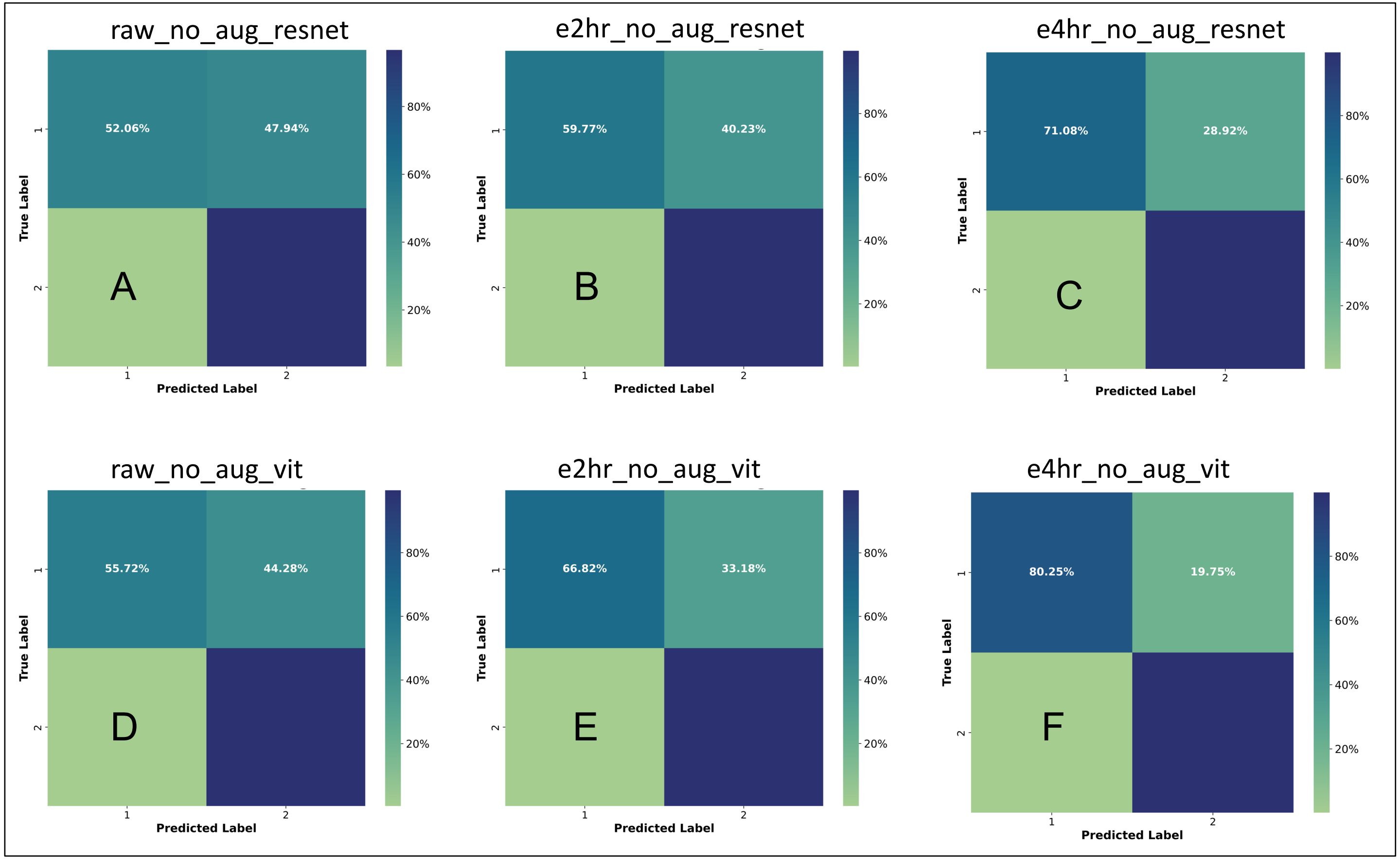
Results: Limitations (using ViT)

III

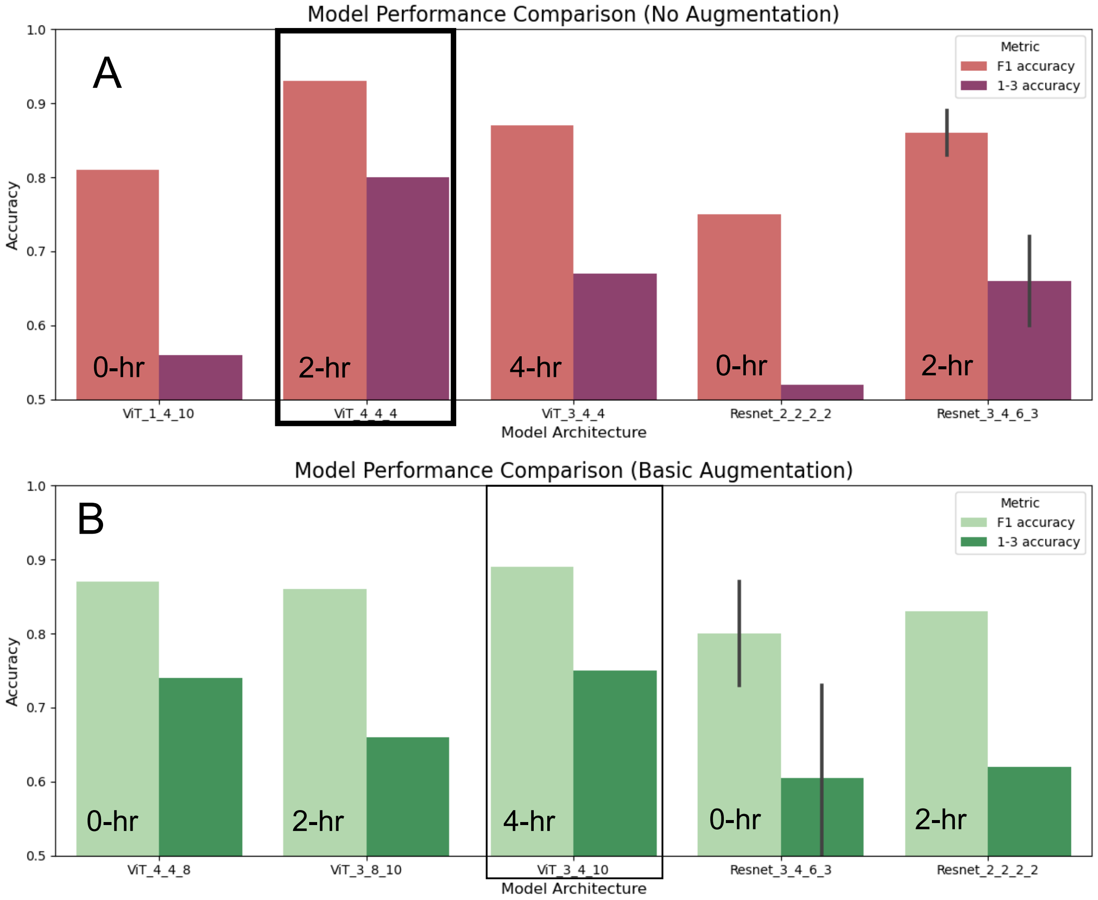


- Confusion between 3 and 4
- Class 1 is difficult to detect
- Classification of all six is very difficult!

Results: ViT with TE and augmentation



Results: enhancement and augmentation



III

- ViT was best
- 2-hours of TE was best
- Augmentation didn’t help

Results: generative augmentation

III

- L1 looks very similar
- L2 and L3 needs more work to be convincing
- Augmentation didn't help (L2 and L3)

Discussion: ViT and TE are ~90% accurate

- **ViT with TE performs best for (feature learning)**

- **Generative Augmentation had issues with L2 and L3**
  - traditional aug works for CNN models
  - generative aug would allow smaller training sets

- **Future direction are to use low-level eng features**
  - multiple segmentation model ensemble

IV

Conclusion: good overall framework is needed

- **We have approached human labeling for CT DEEP data**
  - good data collection practice
  - good data QC and human labeling (>100K images)
  - use of preprocessing filters and TE are important
  - ViT works well on high quality data
  - lower accuracy is expected in different environments

V

Acknowledgments

- Connecticut College SSRI program (A4000 ada GPU)
- Department of Computer Science (A4000 ada GPU)
- AI@Conn program (RAM and SSD)
- Conn Student Travel Funding
- Conn Staff Jim O'Connor, Danielle Fontaine
- CT DEEP Water Planning Staff, Melissa Czarnowski, Chris Bellucci

COMPASS25: https://doi.org/10.1145/3715335.3735455

preprint: https://doi.org/10.48550/arXiv.2502.00474

https://informatics.digital.conncoll.edu

Connecticut College