Identifying Trusted And Ambiguous Regions In Neural Network Predictions: High-Fidelity AI For Image Pathology

Kenneth Wenger[1,2], Katayoun Hossein Abadi[1], Damian Fozard[1], Kayvan Tirdad[2], Alex Dela Cruz[2], Alireza Sadeghian[2]

[1]Dept. of Advanced Artificial Intelligence & Cognition, Squint AI Inc, Waterloo, Ontario, Canada
[2]Department of Computer Science, Faculty of Science, Toronto Metropolitan University, Toronto Ontario, Canada

Summary

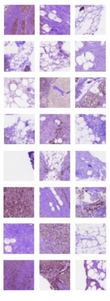
- AI in medicine can benefit from human-in-the-loop[1].
- Squint pipelines optimize the work of human pathologists by identifying when AI predictions can't be trusted, and when the assessment of human professionals is required[2].

Introduction

- Recent years have seen a rise in deep learning (neural network) algorithms being applied to image pathology and cancer grading research[3,4].
- Still, wide adoption of AI in the medical domain remains low[5].
- High-confidence mistakes are common and differ from the types of mistakes humans make[6].
- In this work we show how to use eXplainable AI (XAI) to improve the performance of cancer grading algorithms by combining AI models and human pathologists.

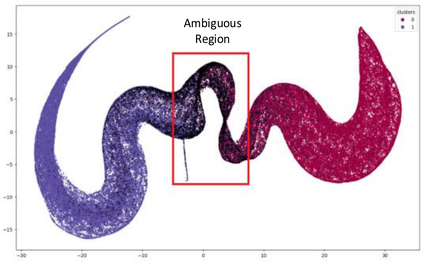
Data Set

- Breast cancer dataset consisting of biopsy scans of breast tissue[7].
- The dataset is divided into 126,056 training images and 15,758 testing images.
- **Training Data:** 62,901 positive samples, and 63,155 negative samples.
- **Testing Data:** 7,947 positive samples, and 7,811 negative samples.

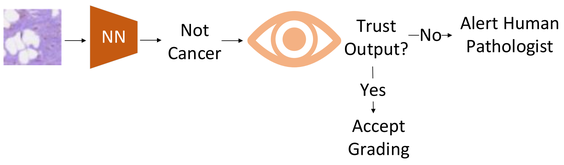


Methods

- We trained a cancer grading CNN to state-of-the-art performance of 89.37% accuracy on binary classification of cancer vs healthy tissue.
- Using XAI clustering algorithms t-SNE[8] and PacMap[9] we generate a knowledge map for the CNN that identifies Trusted and Ambiguous regions in the model's decision surface.



- When CNN decisions originate in Trusted regions the decision is accepted, otherwise it is rejected, and a pathologist is alerted for further analysis.
- The Squint framework is described by the following diagram:



Results

- Our framework identified 65% of CNN decisions as trusted and flagged 35% of decisions as requiring a human pathologist's input.
- Human involvement was only required for 5,515 cases out of 15,758.
- Human effort reduced by two thirds.
- Our Squinting framework produced **75% fewer mistakes** than the state-of-the-art model.
- Our framework **automatically identified 1,257 mistakes** (out of 1675 total mistakes) and alerted a human pathologist**.**
- In 455 of the identified mistakes, the CNN was highly confident (> 80% confidence) of its prediction.

Conclusion

- Squinting pipelines constitute a paradigm shift in AI deployment.
- Our work shows that AI decisions can be monitored to know when they can be trusted, and when to involve a human pathologist.
- Our proposed framework produced 1257 fewer mistakes than the state-of-the-art model.
- XAI algorithms facilitate deploying AI models in medical use cases.

References

- Onur, A et al., 2020, 10.2196/15154
- Wenger, K et al., 2023, 10.1016/j.mlwa.2023.100491
- Habibi, K, et al., 2022, 10.1016/j.mlwa.2022.100387
- Wenger, K, et al., 2022, 10.1016/j.mlwa.2022.100347
- Bauranov, A, et al., 2021, 10.1016/j.paerosci.2021.100726
- Nguyen, A, et al., 2015, 10.1109/CVPR.2015.7298640
- Mooney, P, 2023, https://www.kaggle.com/datasets/paultimothymooney/breast-histopathology-images/code
- Van der Maaten, et al., 2008, "Visualizing Data using t-SNE"
- Wang, Y, et al., 2021, 10.5555/3546258.3546459

Kenneth Wenger
ken.wenger@squintinc.com