Explainable AI in Pathology - Concept Based Explainability for Mitotic Figure Detection in Whole Slide Images

Adam Tegart DCU BSc in Data Science

Abstract

- Artificial Intelligence (AI) advancements are allowing the performance of models to reach new heights, though increasing complexity reduces their explainability.
- These "black-box" models bring with them ethical implications when the autonomy and livelihood of individuals is at stake due to lack of interpretability, such as in the health sector.
- This project aims to discover if automated approaches can be used to bootstrap concept-based explainability approaches for use in mitotic figure detection, eliminating the need for an expert to curate a set of concepts manually.
- The approach used in this paper builds on previous work that creates segments from images and clusters those that are visually similar. These clusters act as potential visual concepts, which can then be tested using statistical methods.
- The automated approach outlined in this paper creates a good baseline set of concepts but requires an iterative approach to refine these concepts for complex use cases that typically require an expert.

Introduction

- There is a shortage of pathologists globally with only 0.8% of medical doctors being pathologists and a workload that is increasing[1].
- Humans are also prone to error and there is the possibility for inter-observer variability between pathologists in relation to the counts of mitotic figures.
- The increased workload and possibility of varying results between pathologists motivates the use of AI, but adequate XAI techniques must allow the detections to be understood by pathologists.
- This will allow pathologists to ensure the model is detecting based on the correct features and will allow them to be more efficient.
- Transparency is a key element to the success of AI as a tool in healthcare as decisions can carry the weight of individuals' lives and every reasonable precaution should be taken to ensure these models don't result in harm.

Dataset

- the Mitosis Domain Generalisation (MIDOG) challenge dataset[2] was used, consisting of 405 whole slide images covering 6 different types of tissue between human and canine sourced from various scanners.
- Image snippets/tiles containing the annotations were extracted to remove redundant data, lowering the storage requirement by 90%.
- A tiling approach which allowed for overlap was used for sectioning the tiles to avoid annotations being cut-off if they occurred on the boundary of tiles.

Methodology

- A Faster R-CNN[3] model with a ResNet50[4] backbone was trained using a 90:10 train and validation split.

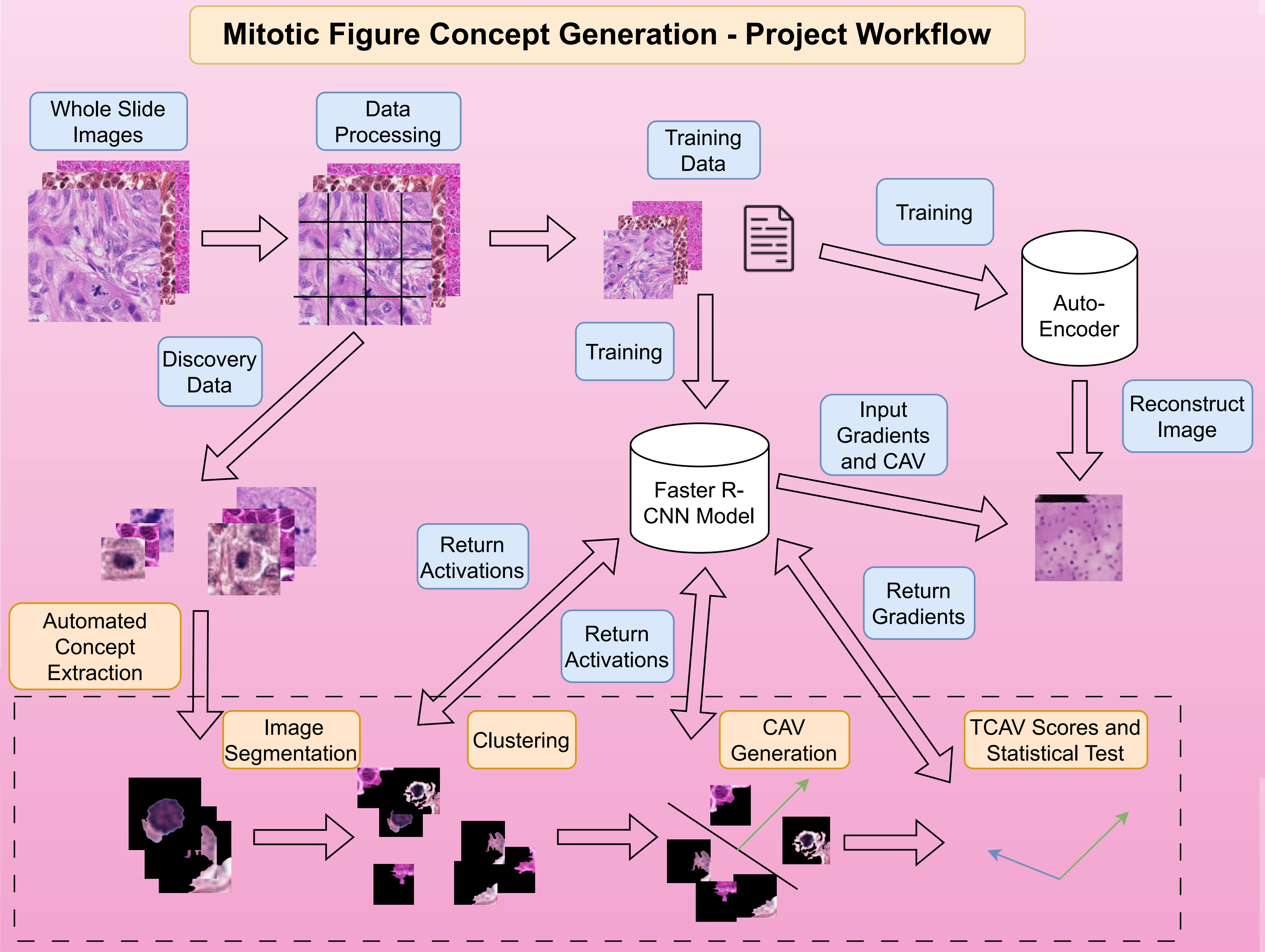

Fig. 1. A figure depicting the project workflow.

- **Image Segmentation:** The annotations are segmented into patches.
- **Segment Clustering:** The segments are passed through the model and the activations extracted from the bottlenecks. These activations give us a latent representation which we cluster to find visually similar segments.
- **Random Sampling:** A random concept and random samples are generated from the segments. These are needed for training the concept activation vectors (CAVs)[5] and allow for a two-sided t-test.
- **CAV Generation:** A linear model is trained to separate the concept images from the random samples. The vector orthogonal to the hyperplane is taken as the CAV.
- **TCAV Scores:** The dot product of the gradients for a class in each bottleneck layer is taken with the CAV. This tells us if the CAV contributes to increasing the probability, i.e., moving away from the gradient.
- **Gifsplanation[6]:** A ResNet50 based autoencoder was trained to allow the tiles to be reconstructed. The CAVs and gradients were added to the activations passing through to alter the image. This would allow for a visual aid to understand the effect of the CAV.

Results and Discussion

- The primary results of the project are the returned concepts and both their TCAV scores and p-values from the two-sided t-test. These are plotted below.

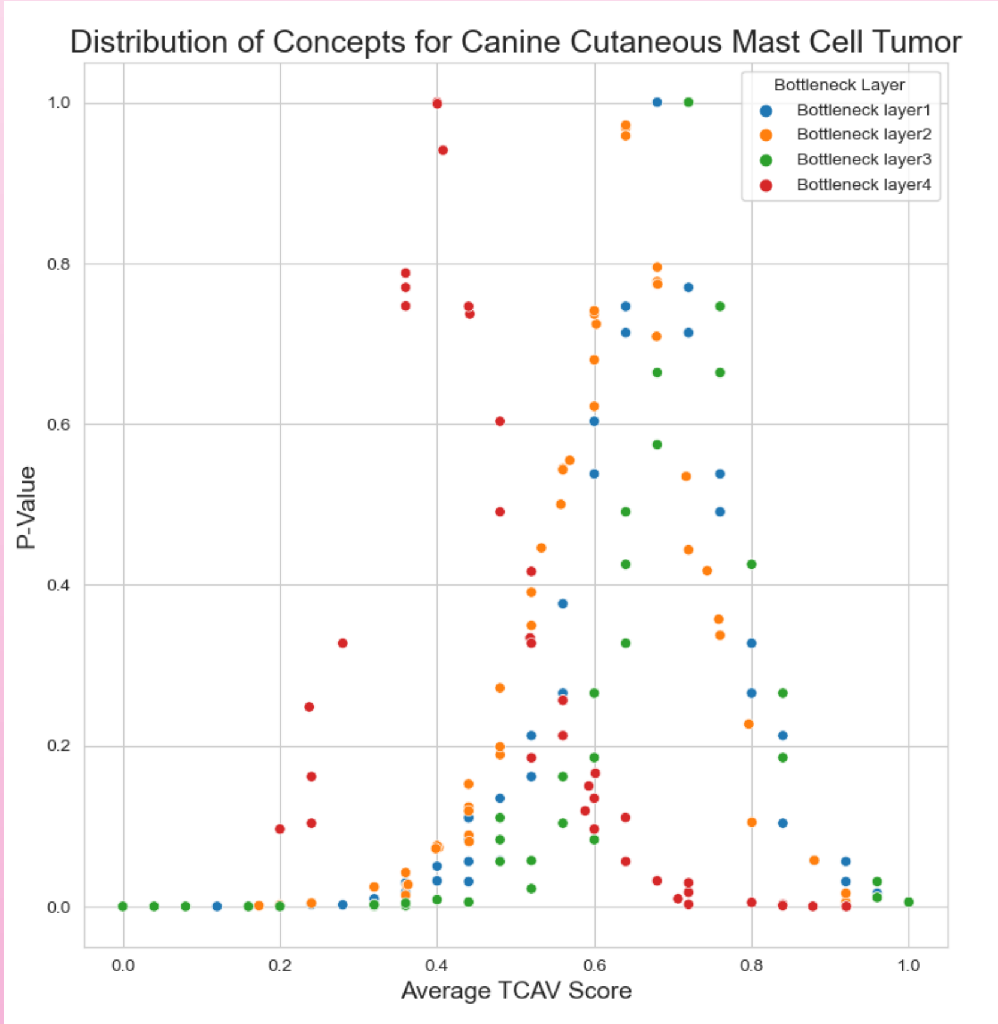

Fig. 2. Scatterplot of cutaneous mast scores and p-values.

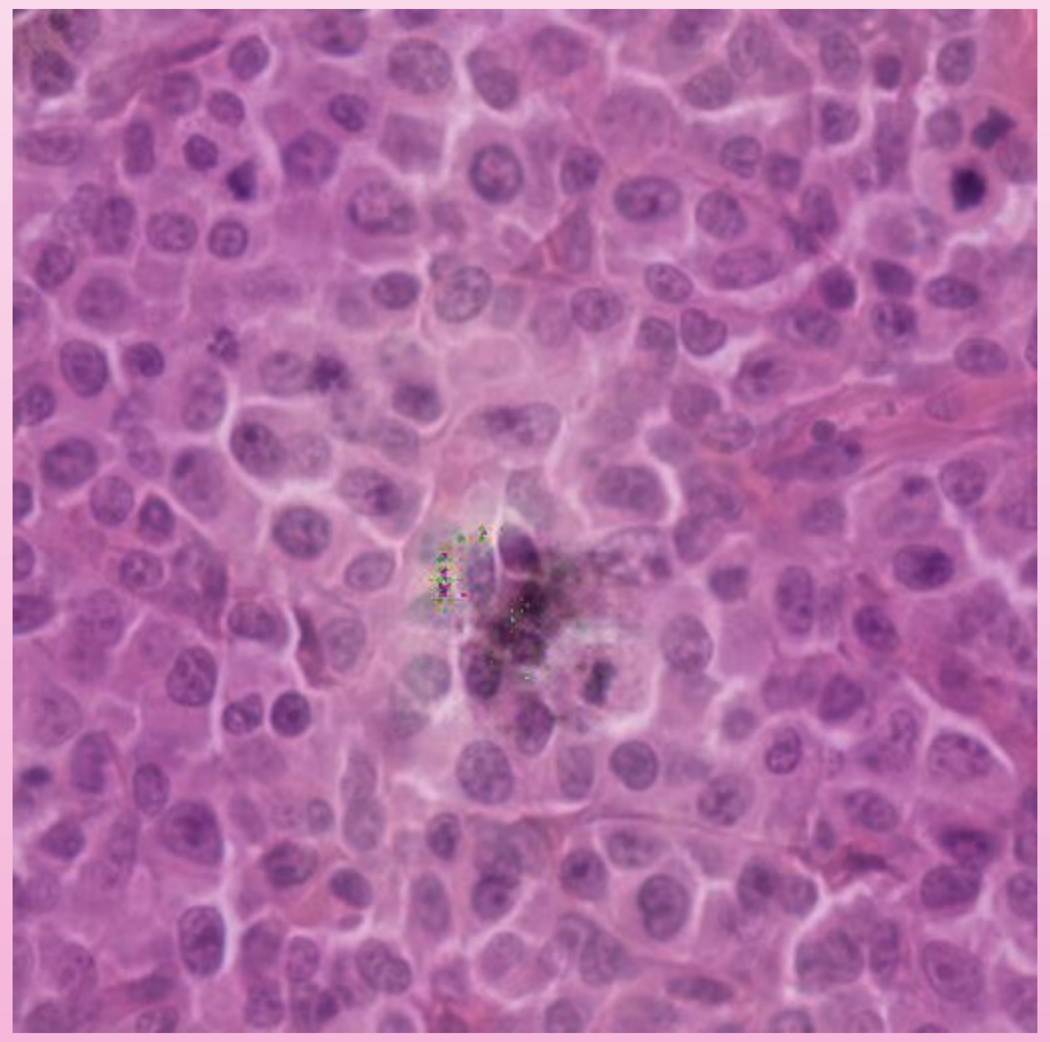

Fig. 3. Reconstructed images with a large portion of the gradient (0.048) added to the activations.

- We can see from the plots that the values with an average TCAV score are correlated with the random concept, shown by their high p-value. This means that mildly influential concepts were found to be correlated with the random concept. The highly influential concepts are statistically different from the random concept and so are meaningful and not influential just by sheer coincidence.
- The Gifsplanation output shows that the changes can be localized, though the autoencoder is not sufficient to allow for the CAV to be seen in a meaningful way.

Conclusion

- The automated approach to finding concepts allowed for a good baseline implemented of concept-based explainability. An iterative approach that samples images from clusters would allow concepts to be refined.
- There were shortcomings regarding both overlooking the option to jointly train both models and my inability to effectively analyse the results due to my lack of expertise in the area.

References

[1] A. Bychkov et al., "Constant demand, patchy supply," The Pathologist, Feb. 2023. [Online]. Available: https : / / thepathologist .com/outside-the-lab/constant-demand-patchy-supply.
[2] M. Aubreville,et al., "Mitosis domain generalization challenge 2022," Mar. 2022. doi:10.5281/zenodo.6362337.
[3] S. Ren, wt al., "Faster r-cnn: Towards real-time object detection with region proposal networks," Advances in neural information processing systems, vol. 28, 2015.
[4] K. He, et al., "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[5] B. Kim, et al., "Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav)," in International conference on machine learning, PMLR, 2018, pp. 2668–2677.
[6] J. P. Cohen, et al., "Gifsplanation via latent shift: A simple autoencoder approach to counterfactual generation for chest x-rays," in Medical Imaging with Deep Learning, PMLR, 2021, pp. 74–104.