**Naive Prediction of Cancer Outcomes using Machine Learning**

Matt Bixley, Mik Black  
University of Otago  
[matt.bixley@otago.ac.nz](mailto:matt.bixley@otago.ac.nz)  
[mik.black@otago.ac.nz](mailto:mik.black@otago.ac.nz)

### ABSTRACT / INTRODUCTION

*Please include any of the following points, but not limited to:*

* *Background/Context*
* *Related work*
* *Method*
* *Results*
* *Conclusion*
* *Acknowledgments*

Prediction of 5 year cancer outcomes from histology images has been undertaken using machine learning (ML), artificial intelligence (AI) and deep learning techniques has been undertaken by various groups, with success, (INSERT PAPER REF) for a number of different cancers. Eg Prostate, Breast and Colorectal cancers. A key outcome in this approach is the easy translation of technology to allow pathologist to access the applications in their workflow.

With regard to tumor detection in prostate tissues, Litjens et al. ([65](https://www.frontiersin.org/articles/10.3389/fmed.2019.00185/full#B65)) used a convolutional auto-encoder for tumor detection in H&E stained biopsy specimens. Substantial gains in efficiency were possible by using CNNs to exclude tumor-negative slides from further human analysis; showing the potential to reduce the workload for pathologists. **Bulten et al. (**[**66**](https://www.frontiersin.org/articles/10.3389/fmed.2019.00185/full#B66)**) developed an algorithm for automated segmentation of epithelial tissue in prostatectomy slides using CNN. The generated segmentation can be used to highlight regions of interest for pathologists and to improve cancer annotations.**

For virtually every patient with colorectal cancer (CRC), hematoxylin–eosin (HE)–stained tissue slides are available. These images contain quantitative information, which is not routinely used to objectively extract prognostic biomarkers. In the present study, we investigated whether deep convolutional neural networks (CNNs) can extract prognosticators directly from these widely available images.

We hand-delineated single-tissue regions in 86 CRC tissue slides, yielding more than 100,000 HE image patches, and used these to train a CNN by transfer learning, reaching a nine-class accuracy of >94% in an independent data set of 7,180 images from 25 CRC patients. With this tool, we performed automated tissue decomposition of representative

multitissue HE images from 862 HE slides in 500 stage I–IV CRC patients in the The Cancer Genome Atlas (TCGA) cohort, a large international multicenter collection of CRC tissue.

A key requirement for technology translation is the need to embed AI within diagnostic workflow—to ensure that the pathologist can easily access AI applications for diagnostics.

References

Colorectal Cancer  
<https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002730>

Breast Cancer<https://link.springer.com/article/10.1007/s10549-019-05281-1>

# Machine learning for tissue diagnostics in oncology: brave new world <https://www.nature.com/articles/s41416-019-0535-1>

### ABOUT THE AUTHOR(S)

* Name
* Bio