

Abstract: A completely annotated whole slide image dataset of canine breast cancer to aid human breast cancer research

Marc Aubreville^{1,2}, Christof A. Bertram³, Taryn A. Donovan⁴,
Christian Marzahl², Andreas Maier², Robert Klopffleisch³

¹Technische Hochschule Ingolstadt, Ingolstadt, Germany

²Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg,
Germany

³Institute of Veterinary Pathology, Freie Universität Berlin, Germany

⁴Department of Anatomic Pathology, Animal Medical Center, New York, USA
`marc.aubreville@thi.de`

Canine mammary carcinoma (CMC) has been used as a model to investigate the tumorigenesis of human breast cancer and the same histological grading scheme is commonly used to estimate patient outcome for both. One key component of this grading scheme is the density of cells undergoing cell division (mitotic figures, MF). Current publicly available datasets on human breast cancer only provide annotations for small subsets of whole slide images (WSIs).

We present a novel dataset[1] of 21 WSIs of CMC completely annotated for MF. For this, a pathologist screened all WSIs for potential MF and structures with a similar appearance. A second expert blindly assigned labels, and for non-matching labels, a third expert assigned the final labels. Additionally, we used machine learning to identify previously undetected MF. Finally, we performed representation learning and two-dimensional projection to further increase the consistency of the annotations.

Our dataset consists of 13,907 MF and 36,379 hard negatives. We achieved a mean F1-score of 0.791 on the test set. Testing our algorithms without any further adaptation on a human breast cancer dataset (AMIDA13) yielded a mean F1-score of 0.635. The F1-score increased to 0.696 when using threshold optimization and model selection, and to 0.733 using transfer learning, both on the human tissue training set.

References

1. Aubreville M, Bertram CA, Donovan TA, et al. A completely annotated whole slide image dataset of canine breast cancer to aid human breast cancer research. *Scientific Data*. 2020;7(417).