



# Abstract: Cytologic Scoring of Equine Exercise-induced Pulmonary Hemorrhage

## Performance of Human Experts and a Deep Learning

Christof A. Bertram<sup>1</sup>, Christian Marzahl<sup>2</sup>, Alexander Bartel<sup>3</sup>, Jason Stayt<sup>4</sup>,  
Federico Bonsembiante<sup>5</sup>, Janet Beeler-Marfisi<sup>6</sup>, Ann K. Barton<sup>3</sup>, Ginevra Brocca<sup>5</sup>,  
Maria E. Gelain<sup>5</sup>, Agnes Gläsel<sup>7</sup>, Kelly du Preez<sup>8</sup>, Kristina Weiler<sup>7</sup>,  
Christiane Weissenbacher-Lang<sup>1</sup>, Katharina Breininger<sup>2</sup>, Marc Aubreville<sup>9</sup>,  
Andreas Maier<sup>2</sup>, Robert Klopff/eisch<sup>3</sup>, Jenny Hill<sup>4</sup>

<sup>1</sup>University of Veterinary Medicine Vienna, Vienna, Austria

<sup>2</sup>Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

<sup>3</sup>Freie Universität Berlin, Berlin, Germany

<sup>4</sup>Novavet Diagnostics, Bayswater, Western Australia

<sup>5</sup>University of Padova, Legnaro, Italy

<sup>6</sup>University of Guelph, Guelph, Ontario, Canada

<sup>7</sup>Justus-Liebig-Universität Giessen, Giessen, Germany

<sup>8</sup>University of Pretoria, Pretoria, South Africa

<sup>9</sup>Technische Hochschule Ingolstadt, Ingolstadt, Germany

christof.bertram@vetmeduni.ac.at

Exercise-induced pulmonary hemorrhage (EIPH) is a common respiratory condition in race horses with negative implications on performance. The gold standard diagnostic method is cytology of bronchoalveolar lavage fluid using the time-consuming total hemosiderin score (THS). For the routine THS, 300 alveolar macrophages are classified into 5 grades based on the amount of intracellular hemosiderin pigment (degradation product of heme iron of red blood cells). Besides the high time investment, there is notable inter-rater variability in assigning hemosiderin grades. Thus automated image analysis is of high interest to improve this diagnostic test. In this study [1] we validated a deep learning-based algorithm (RetinaNet) in 52 whole slide images (WSI) against the performance of 10 experts (each graded 300 cells per case) and a ground truth with labels for all macrophage in the WSI (range: 596 - 8954 macrophages). Compared to the ground truth reference, the algorithm had a diagnostic accuracy of 92.3%, while the 10 experts had an accuracy of 75.5% (range: 63.4 - 92.3%). Automated analysis of a single WSI took on average 1:37 minutes, while experts required an average of 14 minutes for 300 cells. In conclusion, the deep learning-based algorithm has a high diagnostic accuracy and is, therefore, a promising tool to reduce expert labor and to facilitate the routine use of the THS.

## References

1. Bertram CA, Marzahl C, Bartel A, Stayt J, Bonsembiante F, Beeler-Marfisi J et al. Cytologic scoring of equine exercise-induced pulmonary hemorrhage: performance of human experts and a deep learning-based algorithm. Vet Pathol. 2023;60.