



## The 65th ASH Annual Meeting Abstracts

## POSTER ABSTRACTS

## 803. EMERGING TOOLS, TECHNIQUES AND ARTIFICIAL INTELLIGENCE IN HEMATOLOGY

**Inter-Facility Evaluation of Machine Learning Based Deep Hematopoietic Cell Classifiers for Cytological Evaluation of Hematopoiesis**

Julian Thull<sup>1</sup>, Martina Crysandt, MD<sup>2,3</sup>, Barbara Klinkhammer<sup>4</sup>, Peter Boor<sup>5,6</sup>, Tim H. Brümmendorf<sup>7,8</sup>, Volkmar Schulz<sup>1,9</sup>, Barbara Klinkhammer, MD<sup>10</sup>, Dorit Merhof<sup>11</sup>, Edgar Jost, MD<sup>12,13</sup>

<sup>1</sup>Institute for Experimental Molecular Imaging, RWTH Aachen, Aachen, Germany

<sup>2</sup>University Hospital RWTH Aachen, Aachen, DEU

<sup>3</sup>Department of Hematology, Oncology, Hemostaseology and Stem Cell Transplantation, Medical Faculty, RWTH Aachen University, Aachen, Germany

<sup>4</sup>Department of Pathology, University Hospital Aachen, Aachen, Germany

<sup>5</sup>Center for Integrated Oncology Aachen Bonn Cologne Düsseldorf (CIO ABCD), Aachen, Germany

<sup>6</sup>Institute of Pathology, RWTH Aachen University, Faculty of Medicine, Aachen, Germany

<sup>7</sup>Hematology/Oncology, Universitätsklinikum RWTH Aachen, Aachen, Germany

<sup>8</sup>Center for Integrated Oncology - Aachen Bonn Cologne Düsseldorf (CIO-ABCD), Aachen, Germany

<sup>9</sup>Fraunhofer Institute for Digital Medicine MEVIS, Bremen, Germany

<sup>10</sup>Department of Pathology, University Hospital Aachen, Aachen, Germany

<sup>11</sup>Institute of Image Analysis and Computer Vision, University of Regensburg, Regensburg, Germany

<sup>12</sup>Department of Hematology, Oncology, Hemostaseology and Stem Cell Transplantation, University Hospital Aachen, Aachen, Germany

<sup>13</sup>Center for Integrated Oncology CIO ABCD Aachen, Bonn Cologne, Düsseldorf, Aachen, Germany

**Introduction:** Cytology, histology, flow cytometry, cytogenetics and molecular genetics are the cornerstones of diagnosis in hematology. Cytological evaluation of blood and bone marrow is performed by trained experts. However, this task is time consuming and subject to inter- and intra-rater variability. Several studies proposed processing pipelines for peripheral blood and bone marrow smears. Deep learning approaches by neural networks have been evaluated for automated classification of blood and bone marrow cells. Nevertheless, issues concerning the evaluation methodology with respect to real-world applicability remain, e.g., accuracy of designation of cell types, differences in staining and imaging protocols, dataset splitting, number of cell types as well as hierarchy and selection of regions of interest.

**Methods:** We present a slide and facility cross-validated evaluation of machine-learning based automated hematopoietic cell classifiers of bone marrow smears. The analysis is based on the DenseNet121 and ResNet152 architectures of our own database collected at University Hospital RWTH Aachen consisting of 11,899 labeled objects from 24 different classes collected from 6 different patients. In addition, we analyzed 3 publicly available bone marrow and blood cell datasets for staining, color variability and other quality parameters. Coupled with cross-validation we employed model ensembling, augmentation, normalization, and pseudo-labeling to train domain-adapted models for each dataset. Our evaluations are contrasted by an inter-rater analysis performed on a subset of our labeled granulocytic cell line data.

**Results:** A review of performance scores from related works in the field revealed that Convolutional Neural Network (CNN) architectures were used in all analyses, except for Krappe et al. where a tree of polynomial classifiers was employed. Only one publication used slide cross validation (SCV) and the number of cell classes applied was variable as well as the number of cells annotated by experts and used for analysis. Different numbers of cell classes were defined. Accuracy of cell classification was also variable, and F-scores were used to improve interpretation of results also for small cell populations. Fig 1 shows Facility cross-validation results for baseline evaluation (pink) and new model training considering brightness, contrast, saturation, and hue (gold) with DenseNet and ResNet CNN for granulopoiesis divided into 5 classes of cells. Dark bars indicate bagging performance. X-axes indicate the evaluation data set. Green shaded regions denote the lower bound for optimal performance, red shaded regions show the upper bound for worst performance. The variations in smear preparation, fixation, and embedding as well as the usage of different staining protocols and scanning devices lead to biases in data collected from different sites. Accuracies of in 13-, 8- and 5-class cross-validation setups were obtained when using different amounts of available data each.

Incorporating confusion tolerance fields (one-class difference in granulocytic precursors) yields accuracies of , whilst classifier ensembling increases performance scores similarly well to training on larger combined datasets. The interrater accuracy of experts was 96 % when the one-class difference was applied as confusion tolerance.

Conclusion: Deep neural network models as suggested here are most sensitive to color augmentations in the domain of brightness, contrast, hue, and saturation. If configured properly, respective augmentations increase performance scores significantly. However, if not accounted for, the models experience severe performance drops when evaluated on slide data from other sites. Enlarging the corpus of training data via pseudo-labeling improves results only slightly, whilst normalization approaches are outperformed by color augmentation strategies. Even though, our results show inferior accuracy by deep learning approaches compared to medical experts.

Deep neural network approaches will become the platform for automated cytological analysis of blood and bone marrow smears in the future. However, harmonization of preanalytical and technical aspects will be necessary to improve accuracy of cell classification and further extension for recognition of specific morphological abnormalities e.g., dysplasia, will be necessary.

**Disclosures Crysandt:** Pfizer: Other: Travel grant. **Brümmendorf:** Janssen: Consultancy, Honoraria, Speakers Bureau; Merck: Consultancy, Honoraria, Speakers Bureau; Novartis: Consultancy, Honoraria, Patents & Royalties, Research Funding, Speakers Bureau; Pfizer: Consultancy, Honoraria, Research Funding, Speakers Bureau; Gilead: Consultancy, Speakers Bureau. **Just:** JAZZ: Honoraria; BMS Celgene: Honoraria; Amgen: Honoraria; Medac: Other: travel costs.

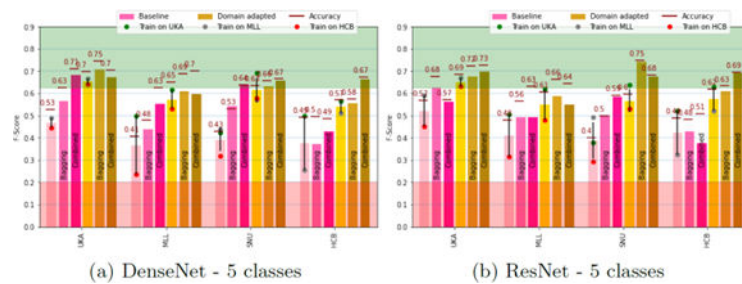


Figure 1

<https://doi.org/10.1182/blood-2023-184399>