Expert-annotated datasets for deep-learning based classification methods for morphologic Leukaemia diagnostics

Christian Matek [1,2]   CHRISTIAN.MATEK@UK-ERLANGEN.DE

Carsten Marr[2]   CARSTEN.MARR@HELMHOLTZ-MUNICH.DE

[1] *Institute of Pathology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany*

[2] *Institute of AI for Health, Helmholtz Munich, Neuherberg, Germany.*

Editors: Under Review for MIDL 2024

Abstract

In recent years, image analysis and classification methods have become increasingly popular for segmentation and classification tasks in pathology imaging, specifically for the evaluation of tissues sections and single cells.

Hematological cytomorphology, with single cell-classification at its base and lacking some of the complexities of tissue architecture, has been at the forefront of this development. Recent additions of large-scale datasets to the public domain open up new possibilities for developing and benchmarking algorithms for morphologic diagnostics, prognostication and subgrouping.

Keywords: Microscopy imaging, deep learning, single-cell data, multiple-instance learning.

1. Introduction

Over the past 150 years, morphologic examination of blood and bone marrow cells using light microscopy has formed the basis of leukemia diagnostics. Presently, high-resolution microscopy of peripheral blood and bone marrow samples still represents an important step in the diagnostic workup of patients suffering from hematologic malignancies (Means , ed.). Used in conjunction with more recently developed immunologic and genetic tests in modern routine diagnostics, cytomorphology still stands out for its accessibilty, comparatively low cost and wide availability.

A notable problem of morphologic evaluation in current routine practice is its heavy use of trained human staff. Firstly, readling cytology specimens without preselection of diagnostically relevant areas can be a tedious and time consuming task. Secondly, many diagnostic settings worldwide suffer from a lack of trained examiners experienced in cytomorphology. Finally, the intrinsically subjective nature of morphologic classificaiton can lead to significant inter- and intra-rater variabiliy of results (Fuentes-Arderiu and Dot-Bach; Font et al.). Hence implementing robust and widely applicable diagnostic decision support systems for cytomorphologic examination corresponds to a relevant clinical need. Recent advances in deep learning-based image classification offer promising methods for developing such tools. As for many data-driven methods in medical diagnostics, progress in algorithm development can be limited by the lack of publicly available high-quality datasets. Here we present a set of recently published curated reference datasets containing microscopic images of blood and bone marrow cells, as well as a set of algorithms developed based on these datasets which answer diagnostically and therapeutically relevant questions.

2. Datasets

Recently, a number of datasets on hematological cytomorphology have been published for both peripheral blood and bone marrow. Specifically, datasets related to work from our group were anonymized in concordance with the relevant ethics votes, and deposited in The Cancer Imaging Archive (TCIA) (Clark et al.). Relevant dataset parameters are summarized in Tab. 4, and sample images from two datasets are shown in Fig. 1. Refs. (Matek et al., c) and (Hehr et al., b) cover single-leukocyte images from peripheral blood, while ref. (Matek et al., a) contains microscopic images of bone marrow smears, which represents the diagnostic gold standard. All three image databases are curated, and ref. (Hehr et al., b) additionally contains clinical patient information. While the peripheral blood datasets focus on acute myeloid leukemia (AML), ref. (Matek et al., a) contains bone marrow samples from a wide range of hematologic diseases detailed in ref. (Matek et al., b). While expert-annotated datasets have been published before for peripheral blood (Acevedo et al.) and - more rarely - for bone marrow samples (Choi et al.), these to the authors' knowledge represent the largest publicly available databases today, providing a benchmark and resource for future algorithm development.

3. Classification approaches

Expert annotations accompanying the datasets described in this short paper allow training and evaluating algorithms for both single-leukocyte classification and patient-level subtyping. Specifically, we used the ResNeXt architecture in order to train a classifier for leukocytes in AML which attains human-level performance for blast cell recognition (Matek et al., d). For differention of leukocytes relevant in the bone marrow, a classifier can be trained that shows similar performance (Matek et al., b). Finally, a classifier can be trained to predict the genetic subtypes of AML from a bag of leukocyte images using attention-based multiple instance learning (Hehr et al., a). Predictions are amenable to explainability methods, both for single and multiple instance classifiers (Matek et al., b; Sadafi et al., a). All single-leukocyte classifiers obtain a reduced accuracy for minority classes, reflecting the importance of class imbalance. Variable imbalance between different domains can be accounted for using cross-domain feature extraction (Salehi et al.) or dedicated domain-generalization algorithms (Umer et al.). Additionally, trained algorithms can be sensitive towards biased patient age and biological sex distributions (Sadafi et al., b).

4. Conclusions and outlook

Comprehensive datasets of malignant and non-malignant leukocyte classes allow harnessing the power of current deep learning-based image classification algorithms to train classifiers that achieve a high diagnostic accuracy in single-cell classification of many relevant cell classes in both peripheral blood and bone marrow. For some specific classification tasks, human-level performance is achieved. Major challenges in routine implementation of these methods are a significant class imbalance between individual cell types, an effect due to the biological distribution of relevant leukocyte types, as well as domain-specific staining and cell-type distribution differences. These challenges can be met by utilizing dedicated

domain-generalization methods, as well as increasing the size and diversity of datasets used as training data.

Dataset	Individuals	Leukocytes	Resolution	Reference
AML-Cytomorphology_LMU	200	18,365	0.07072 $\mu m/px$	(Matek et al., c)
AML-Cytomorphology_MLL_Helmholtz	189	81,214	0.1725 $\mu m/px$	(Hehr et al., b)
Bone-Marrow-Cytomorphology_MHF	945	170,000	0.086 $\mu m/px$	(Matek et al., a)

Table 1: Overview of three reference datasets for leukemia cytomorphology deposited in The Cancer Imaging Archive (TCIA).
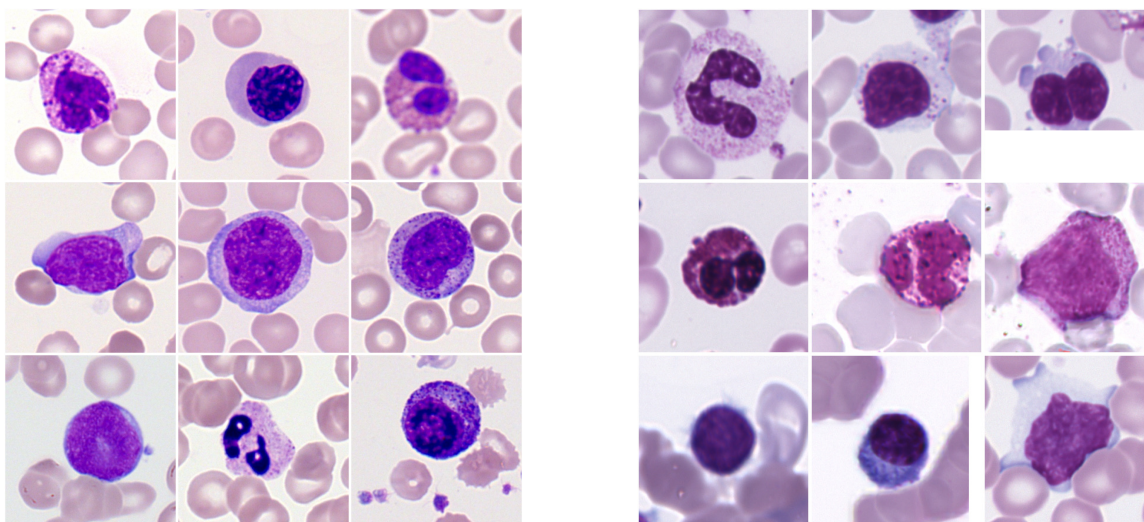

Figure 1: Sample images of leukocytes from peripheral blood smears (left, (Matek et al., c)) and bone marrow (right, (Matek et al., a))

Acknowledgments

References

Andrea Acevedo, Anna Merino, Santiago Alférez, Ángel Molina, Laura Boldú, and José Rodellar. A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. 30:105474. ISSN 23523409. doi: 10.1016/j.dib.2020.105474. URL https://linkinghub.elsevier.com/retrieve/pii/S2352340920303681.

Jin Woo Choi, Yunseo Ku, Byeong Wook Yoo, Jung-Ah Kim, Dong Soon Lee, Young Jun Chai, Hyoun-Joong Kong, and Hee Chan Kim. White blood cell differential count of maturation stages in bone marrow smear using dual-stage convolutional neural networks. 12(12):e0189259. ISSN 1932-6203. doi: 10.1371/journal.pone.0189259. URL https://dx.plos.org/10.1371/journal.pone.0189259.

Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, Lawrence Tarbox, and Fred Prior. The cancer imaging archive (TCIA): Maintaining and operating a public information repository. 26(6):1045–1057. ISSN 0897-1889, 1618-727X. doi: 10.1007/s10278-013-9622-7. URL http://link.springer.com/10.1007/s10278-013-9622-7.

Patricia Font, Javier Loscertales, Carlos Soto, Pilar Ricard, Carolina Muñoz Novas, Estela Martín-Clavero, Montserrat López-Rubio, Luis Garcia-Alonso, Marta Callejas, Alfredo Bermejo, Celina Benavente, Mónica Ballesteros, Teresa Cedena, María Calbacho, Raquel Urbina, Jesús Villarrubia, Santiago Gil, José María Bellón, José Luis Diez-Martin, and Ana Villegas. Interobserver variance in myelodysplastic syndromes with less than 5 % bone marrow blasts: unilineage vs. multilineage dysplasia and reproducibility of the threshold of 2 % blasts. 94(4):565–573. ISSN 0939-5555, 1432-0584. doi: 10.1007/s00277-014-2252-4. URL http://link.springer.com/10.1007/s00277-014-2252-4.

Xavier Fuentes-Arderiu and Dolors Dot-Bach. Measurement uncertainty in manual differential leukocyte counting. 47(1). ISSN 1437-4331, 1434-6621. doi: 10.1515/CCLM.2009.014. URL https://www.degruyter.com/document/doi/10.1515/CCLM.2009.014/html.

Matthias Hehr, Ario Sadafi, Christian Matek, Peter Lienemann, Christian Pohlkamp, Torsten Haferlach, Karsten Spiekermann, and Carsten Marr. Explainable AI identifies diagnostic cells of genetic AML subtypes. 2(3):e0000187, a. ISSN 2767-3170. doi: 10.1371/journal.pdig.0000187. URL https://dx.plos.org/10.1371/journal.pdig.0000187.

Matthias Hehr, Ario Sadafi, Christian Matek, Peter Lienemann, Christian Pohlkamp, Torsten Haferlach, Karsten Spiekermann, and Carsten Marr. A morphological dataset of white blood cells from patients with four different genetic AML entities and non-malignant controls (AML-cytomorphology_mll_helmholtz), b. URL https://www.cancerimagingarchive.net/collection/aml-cytomorphology_mll_helmholtz/.

Christian Matek, Sebastian Krappe, Christian Münzenmayer, Torsten Haferlach, and Carsten Marr. An expert-annotated dataset of bone marrow cytology in hematologic malignancies, a. URL https://www.cancerimagingarchive.net/collection/bone-marrow-cytomorphology_mll_helmholtz_fraunhofer/.

Christian Matek, Sebastian Krappe, Christian Münzenmayer, Torsten Haferlach, and Carsten Marr. Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set. 138(20):1917–1927, b. ISSN 0006-4971, 1528-0020. doi: 10.1182/blood.2020010568. URL https://ashpublications.org/blood/article/138/20/1917/477932/Highly-accurate-differentiation-of-bone-marrow.

Christian Matek, Simone Schwarz, Carsten Marr, and Karsten Spiekermann. A single-cell morphological dataset of leukocytes from AML patients and non-malignant controls, c. URL https://www.cancerimagingarchive.net/collection/aml-cytomorphology_lmu/.

Christian Matek, Simone Schwarz, Karsten Spiekermann, and Carsten Marr. Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. 1(11):538–544, d. ISSN 2522-5839. doi: 10.1038/s42256-019-0101-9. URL https://www.nature.com/articles/s42256-019-0101-9.

Robert T. Means (ed.). Wintrobe's clinical hematology.

Ario Sadafi, Oleksandra Adonkina, Ashkan Khakzar, Peter Lienemann, Rudolf Matthias Hehr, Daniel Rueckert, Nassir Navab, and Carsten Marr. Pixel-level explanation of multiple instance learning models in biomedical single cell images. In Alejandro Frangi, Marleen De Bruijne, Demian Wassermann, and Nassir Navab, editors, *Information Processing in Medical Imaging*, volume 13939, pages 170–182. Springer Nature Switzerland, a. ISBN 978-3-031-34047-5 978-3-031-34048-2. doi: 10.1007/978-3-031-34048-2_14. URL https://link.springer.com/10.1007/978-3-031-34048-2_14. Series Title: Lecture Notes in Computer Science.

Ario Sadafi, Matthias Hehr, Nassir Navab, and Carsten Marr. A study of age and sex bias in multiple instance learning based classification of acute myeloid leukemia subtypes. In Stefan Wesarg, Esther Puyol Antón, John S. H. Baxter, Marius Erdt, Klaus Drechsler, Cristina Oyarzun Laura, Moti Freiman, Yufei Chen, Islem Rekik, Roy Eagleson, Aasa Feragen, Andrew P. King, Veronika Cheplygina, Melani Ganz-Benjaminsen, Enzo Ferrante, Ben Glocker, Daniel Moyer, and Eikel Petersen, editors, *Clinical Image-Based Procedures, Fairness of AI in Medical Imaging, and Ethical and Philosophical Issues in Medical Imaging*, volume 14242, pages 256–265. Springer Nature Switzerland, b. ISBN 978-3-031-45248-2 978-3-031-45249-9. doi: 10.1007/978-3-031-45249-9_25. URL https://link.springer.com/10.1007/978-3-031-45249-9_25. Series Title: Lecture Notes in Computer Science.

Raheleh Salehi, Ario Sadafi, Armin Gruber, Peter Lienemann, Nassir Navab, Shadi Albarqouni, and Carsten Marr. Unsupervised cross-domain feature extraction for single blood cell image classification. In Linwei Wang, Qi Dou, P. Thomas Fletcher, Stefanie Speidel, and Shuo Li, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, volume 13433, pages 739–748. Springer Nature Switzerland. ISBN 978-3-031-16436-1 978-3-031-16437-8. doi: 10.1007/978-3-031-16437-8_71. URL https://link.springer.com/10.1007/978-3-031-16437-8_71. Series Title: Lecture Notes in Computer Science.

Rao Muhammad Umer, Armin Gruber, Sayedali Shetab Boushehri, Christian Metak, and Carsten Marr. Imbalanced domain generalization for robust single cell classification in hematological cytomorphology. doi: 10.48550/ARXIV.2303.07771. URL https://arxiv.org/abs/2303.07771. Publisher: [object Object] Version Number: 3.