



Contribution ID: 32

Type: Talk

DinoBloom: A Foundation Model for Generalizable Cell Embeddings in Hematology

Tuesday 14 May 2024 12:00 (15 minutes)

In hematology, computational models offer significant potential to improve diagnostic accuracy, streamline workflows, and reduce the tedious work of analyzing single cells in peripheral blood or bone marrow smears. However, clinical adoption of computational models has been hampered by the lack of generalization due to large batch effects, small dataset sizes, and poor performance in transfer learning from natural images. To address these challenges, we introduce DinoBloom, the first foundation model for single cell images in hematology, utilizing a tailored DINOv2 pipeline. Our model is built upon an extensive collection of 13 diverse, publicly available datasets of peripheral blood and bone marrow smears, the most substantial open-source cohort in hematology so far, comprising over 380,000 white blood cell images. To assess its generalization capability, we evaluate it on an external dataset with a challenging domain shift. We show that our model outperforms existing medical and non-medical vision models in (i) linear probing and k-nearest neighbor evaluations for cell-type classification on blood and bone marrow smears and (ii) weakly supervised multiple instance learning for acute myeloid leukemia subtyping by a large margin. A family of four DinoBloom models (small, base, large, and giant) can be adapted for a wide range of downstream applications, be a strong baseline for classification problems, and facilitate the assessment of batch effects in new datasets.

Primary author: KOCH, Valentin Korbinian (HMGU)

Presenter: KOCH, Valentin Korbinian (HMGU)

Session Classification: Thematic Session: Image Analysis - part II

Track Classification: Image Analysis (focus on representation learning and foundation models): Thematic focus: Image Analysis