



The 65th ASH Annual Meeting Abstracts

ORAL ABSTRACTS

803. EMERGING TOOLS, TECHNIQUES AND ARTIFICIAL INTELLIGENCE IN HEMATOLOGY

Synthetic Histopathological Images Generation with Artificial Intelligence to Accelerate Research and Improve Clinical Outcomes in Hematology

Gianluca Asti, MSc¹, Saverio D'Amico, MSc¹, Nico Curti, PhD^{2,3}, Gianluca Carlini, PhD^{2,3}, Elisabetta Sauta, PhD¹, Nicolas Riccardo Derus, PhD⁴, Daniele Dall'Olio, PhD⁴, Claudia Sala, PhD⁴, Lorenzo Dall'Olio, PhD^{3,2}, Luca Lanino, MD⁵, Giulia Maggioni, MD¹, Alessia Campagna, MD¹, Marta Ubezio, MD¹, Antonio Russo, MD¹, Gabriele Todisco, MD¹, Cristina Astrid Tentori, MD¹, Pierandrea Morandini, MEng¹, Marilena Bicchieri, PhD¹, Maria Chiara Grondelli, BSc¹, Matteo Zampini, PhD¹, Victor Savevski, MEng¹, Armando Santoro, MD^{6,7}, Shahram Kordasti, MDPHD⁸, Valeria Santini, MD⁹, Anne Sophie Kubasch, MD¹⁰, Uwe Platzbecker, MD¹¹, Maria Diez-Campelo, MD PhD¹², Pierre Fenaux, MD PhD¹³, Lin-Pierre Zhao, MD¹³, Amer M. Zeidan, MBBS, MHS¹⁴, Torsten Haferlach, MD PhD¹⁵, Gastone Castellani, PhD⁴, Matteo Giovanni Della Porta, MD^{1,16}

¹ Humanitas Clinical and Research Center, IRCCS, Rozzano, Italy

² Department of Physics and Astronomy, University of Bologna, Bologna, Italy

³ Data Science and Bioinformatics Laboratory, IRCCS Institute of Neurological Sciences of Bologna, Bologna, Italy

⁴ University of Bologna, Bologna, Italy

⁵ Humanitas Clinical and Research Center, IRCCS, Rozzano, Italy

⁶ Humanitas Cancer Center, Rozzano, Italy

⁷ Humanitas University, Pieve Emanuele and IRCCS Humanitas Research Hospital- Humanitas Cancer Center Rozzano, Milan, Italy

⁸ King's College London, London, United Kingdom

⁹ MDS Unit, DMSC, AOU Careggi, University of Florence, Firenze, Italy

¹⁰ Department of Hematology, Cellular Therapy, Hemostaseology and Infectious Diseases, University Medical Center Leipzig, Leipzig, Germany

¹¹ Medical Clinic and Policlinic 1, Hematology and Cellular Therapy, University Hospital Leipzig, Leipzig, Germany

¹² Department of Hematology, Salamanca-IBSAL University Hospital, Salamanca, Spain

¹³ Saint Louis hospital APHP, Paris, France

¹⁴ Section of Hematology, Department of Internal Medicine, Yale School of Medicine and Yale Cancer Center, New Haven, CT

¹⁵ MLL Munich Leukemia Laboratory, Munich, Germany

¹⁶ Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Italy

Background: Hematological malignancies are rare and complex diseases and as a consequence, multimodal data (ranging from clinical and genomic information to images) are required to improve diagnosis, prognosis and personalized treatments. However, collecting all these layers of information is challenging, in particular when collecting cytological and histological images from the bone marrow (BM) reproducing disease morphologic features. Synthetic data generation by Artificial Intelligence (AI) can circumvent these issues by generating images conditioned from textual inputs (i.e. reports from pathologists), which are widely available and contain many useful clinical information. This technology can enrich data with synthetic images, thus boosting translational research and improving the performances of precision medicine strategies based on multimodal information.

Aims: This project was conducted by GenoMed4all and Synthema EU consortia, with the aim to: 1) Apply generative models to real-world dataset with histological images of patients with myeloid neoplasms (MN). 2) Develop a Synthetic Images Validation Framework (SIVF) to evaluate the utility and fidelity of generated images. 3) Verify the capability of synthetic images to accelerate research and to improve clinical models.

Methods: We implemented Stable Diffusion (SD) generative model fine-tuned on hematological data to generate Hematoxylin and Eosin (H&E) images of MN patients. We implemented a domain specific language model (HematoBERT) to encode textual input as condition for the generation process. Use cases were Myelodysplastic Syndrome (MDS), Acute Myeloid Leukemia

(AML) and Myeloproliferative Neoplasm (MPN) patients, with available BM biopsies and their reports from pathologists, genomic and clinical data. We applied SIVF to evaluate distributions of morphological features extracted from real and synthetic images.

Clinical validation was performed on disease classification and survival probability prediction, using real and synthetic images features (experimental setting is reported in Figure 1).

Results: We trained SD model on 200 patients with available BM biopsies and associated reports. We first performed SIVF to compare extracted morphological features (geometrical, color and texture features of cells nuclei) from synthetic and real images of 55 patients never seen by the model. Results proved that features distributions and correlations in both datasets were comparable. Similar results were obtained performing SIVF on each single patient data.

We verified if synthetic data augmentation could improve performances on MN classification (i.e. models able to correctly assign a single patient to a specific clinical entity according to the 2022 WHO classification criteria). We implemented three XGBOOST models to classify patients' disease. Classifiers were trained and validated on morphological extracted features of images from a real set of patients (n=614), a synthetic group (n=396) and a mixed dataset (n=1010). Data augmentation improved classification performance by 10% (F1 Score) when we tested it on the three different validation sets.

Finally, demographics, clinical features, genomics (cytogenetics and gene mutations) were included as covariates together with morphological features extracted from BM biopsies in L1 penalized Cox's proportional hazards models, considering Overall Survival as primary endpoint. Models were fitted on two different cohorts of real patients (n=182, n=294). Then we added 112 synthetic patients to both sets and refitted the models. We observed an improvement in performances of >10% (C-Index) for both cases (Figure 2), with morphological features (such as "major axis" of nuclei) being selected among the best predictors.

All these results confirmed that data augmentation through synthetic data is a viable approach and can significantly improve the models capability to efficiently capture clinical outcomes at individual patient level.

Conclusion: AI generated images preserve properties of real-world images, replicating cells morphological features relevant to identify hematological diseases and their clinical status. This approach based on widely available textual data allows effective data augmentation and effortless data sharing, thus accelerating and improving precision medicine research in hematology.

Disclosures Santoro: *Incyte*: Consultancy; *Celgene (BMS)*: Speakers Bureau; *BMS*: Membership on an entity's Board of Directors or advisory committees, Speakers Bureau; *Sanofi*: Consultancy; *Amgen*: Speakers Bureau; *AstraZeneca*: Speakers Bureau; *Servier*: Membership on an entity's Board of Directors or advisory committees, Speakers Bureau; *Gilead*: Membership on an entity's Board of Directors or advisory committees, Speakers Bureau; *Pfizer*: Membership on an entity's Board of Directors or advisory committees, Speakers Bureau; *Eisai*: Membership on an entity's Board of Directors or advisory committees, Speakers Bureau; *Bayer*: Membership on an entity's Board of Directors or advisory committees, Speakers Bureau; *Merck MSD*: Membership on an entity's Board of Directors or advisory committees, Speakers Bureau; *Takeda*: Speakers Bureau; *Roche*: Speakers Bureau; *Abbvie*: Speakers Bureau; *Eli Lilly*: Speakers Bureau; *Sandoz*: Speakers Bureau; *Novartis*: Speakers Bureau; *Arqule*: Other. **Kordasti:** *Beckman Coulter*: Honoraria; *Novartis*: Honoraria, Membership on an entity's Board of Directors or advisory committees; *MorphoSys*: Research Funding. **Santini:** *BMS*, *Abbvie*, *Geron*, *Gilead*, *CTI*, *Otsuka*, *servier*, *janssen*, *Syros*: Membership on an entity's Board of Directors or advisory committees. **Platzbecker:** *Silence Therapeutics*: Consultancy, Honoraria, Research Funding; *Takeda*: Consultancy, Honoraria, Research Funding; *Bristol Myers Squibb*: Consultancy, Honoraria, Membership on an entity's Board of Directors or advisory committees, Other: travel support; medical writing support, Research Funding; *Servier*: Consultancy, Honoraria, Research Funding; *Janssen Biotech*: Consultancy, Research Funding; *Syros*: Consultancy, Honoraria, Research Funding; *Merck*: Research Funding; *Curis*: Consultancy, Research Funding; *Geron*: Consultancy, Research Funding; *Roche*: Research Funding; *BeiGene*: Research Funding; *BMS*: Research Funding; *MDS Foundation*: Membership on an entity's Board of Directors or advisory committees; *AbbVie*: Consultancy; *Novartis*: Consultancy, Honoraria, Research Funding; *Celgene*: Honoraria; *Jazz*: Consultancy, Honoraria, Research Funding; *FibroGen*: Research Funding; *Amgen*: Consultancy, Research Funding. **Diez-Campelo:** *Novartis*: Consultancy, Honoraria, Membership on an entity's Board of Directors or advisory committees; *Gilead Sciences*: Other: Travel expense reimbursement; *BMS/Celgene*: Consultancy, Honoraria, Membership on an entity's Board of Directors or advisory committees, Other: Advisory board fees; *GSK*: Consultancy, Membership on an entity's Board of Directors or advisory committees. **Fenau:** *AbbVie*: Consultancy, Honoraria, Research Funding; *Jazz*: Consultancy, Honoraria, Research Funding; *Janssen*: Consultancy, Honoraria, Research Funding; *French MDS Group*: Honoraria; *Novartis*: Consultancy, Honoraria, Research Funding; *Bristol Myers Squibb*: Consultancy, Honoraria, Research Funding. **Zeidan:** *Novartis*: Consultancy, Honoraria; *Boehringer-Ingelheim*: Consultancy, Honoraria; *Incyte*: Consultancy, Honoraria; *Agios*: Consultancy, Honoraria; *Servier*: Consultancy, Honoraria; *Seattle Genetics*: Consultancy, Honoraria; *Amgen*: Consultancy, Honoraria; *Janssen*: Consultancy, Honoraria; *Genentech*: Consultancy, Honoraria; *Zentalis*: Consultancy, Honoraria; *Astex*: Research Funding; *Shattuck Labs*: Research Funding; *Syros*: Consultancy, Honoraria; *Lox Oncology*: Consultancy, Honoraria; *ALX Oncology*: Consultancy, Honoraria; *Orum*: Consultancy, Honoraria; *Notable*: Consultancy, Honoraria; *BioCryst*: Consultancy, Honoraria; *Takeda*: Consultancy, Honoraria; *Ionis*: Consultancy, Honoraria; *BeyondSpring*: Consultancy, Honoraria; *Otsuka*: Consultancy, Honoraria; *Epizyme*: Consultancy, Honoraria; *Syndax*: Consultancy, Honoraria; *Gilead*: Consultancy, Honoraria; *Kura*: Consultancy, Honoraria; *Chiesi*: Consultancy, Honoraria; *Mendus*: Consultancy, Honoraria; *Tyme*: Consultancy, Honoraria; *Schrödinger*: Consultancy, Honoraria; *Regeneron*: Consultancy, Honoraria; *Foran*: Consultancy, Research Funding; *Taiho*: Consultancy, Honoraria; *Geron*: Consultancy, Honoraria; *Astellas*: Consultancy, Honoraria; *Daiichi Sankyo*: Consultancy, Honoraria; *Jazz*: Consultancy, Honoraria; *Celgene/BMS*: Consultancy, Honoraria; *Pfizer*: Consul-

tancy, Honoraria; *AbbVie*: Consultancy, Honoraria. **Haferlach**: MLL Munich Leukemia Laboratory: Current Employment, Other: Equity Ownership. **Della Porta**: Bristol Myers Squibb: Honoraria, Membership on an entity's Board of Directors or advisory committees.

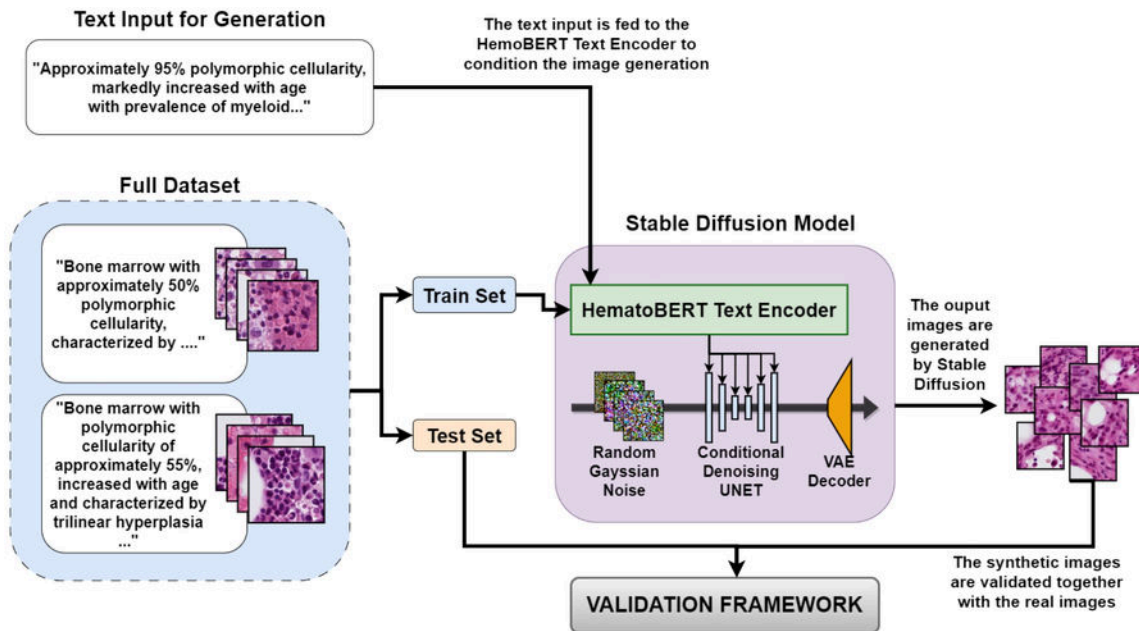


Figure 1: Overview of the SD fine-tuning process and the clinical and statistical validation process. The Validation Framework consisted of 3 steps: Validation through SVF to verify data similarities through Distributions, Correlations and PCA; Validation through data augmentation of XGBOOST multi-classification model; Validation through data augmentation of Cox’s proportional hazards model.

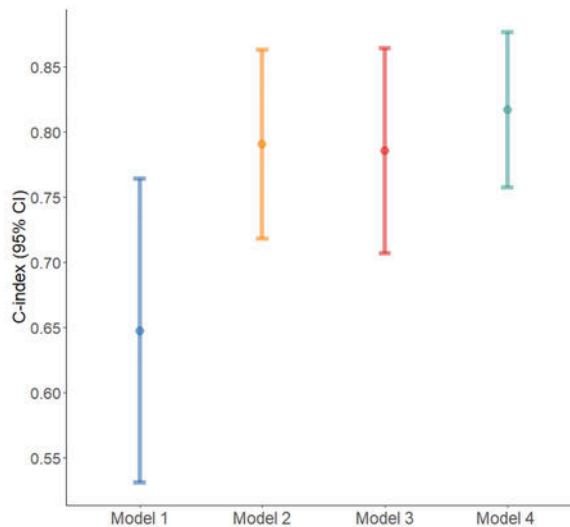


Figure 2: Cox’s proportional hazards model to predict individual probability of overall survival in patients affected with myeloid neoplasms. Demographics, clinical features, genomics (cytogenetics and gene mutations) were included as covariates together with morphological features extracted from bone marrow biopsies. The first model was trained only on 182 real patients and model 2 was trained on 294 patients. Model 3 was trained on 182 real patients and 112 synthetic ones. Model 4 was trained on 294 real patients and 112 synthetic ones.

Figure 1

<https://doi.org/10.1182/blood-2023-187521>

Downloaded from http://ashpublications.net/blood/article-pdf/142/Supplement_1/902/2202807/blood-5506-main.pdf by guest on 22 May 2024