



BLAST IDENTIFICATION IN CHILDHOOD ACUTE MYELOID LEUKAEMIA USING WGAN LATENT SPACE EMBEDDINGS

Licandro R.*^{1,2}, Schlegl T.*¹, Reiter M.², Diem M.², Dworzak M.^{3,4}, Schumich A.⁴, Langs G.¹, Kampel M.²

(1) Department of Biomedical Imaging and Image-guided Therapy, Computational Imaging Research Lab, Medical University of Vienna

(2) Institute of Visual Computing and Human-Centered Technology, Computer Vision Lab, TU Wien

(3) Children's Cancer Research Institute and St. Anna Children's Hospital, Medical University of Vienna

(4) LabDia Labordiagnostik GmbH

Abstract

Acute Myeloid Leukaemia is a rare type of childhood blood cancer and is monitored using flow cytometry. During treatment, the quantification of remaining cancer cells (Minimal Residual Disease (MRD)) is particularly important to determine therapy response and to guide treatment. We propose a Wasserstein Generative Adversarial Network (WGAN) latent space embedding strategy, trained in a semi-supervised fashion for the classification of cancer cells. Results suggest that our approach outperforms semi-supervised learned PCA embeddings and fully connected neural networks.

Introduction

Acute Leukaemia

- Blood cancer affects blood progenitor cells
- Acute Myeloid and Acute Lymphoblastic Leukaemia (AML, ALL)
- Proliferation of undifferentiated cells

Flow CytoMetry (FCM)

- Reliable assessment of Minimal Residual Disease (MRD)
- Detection of leukaemia specific immunophenotypes
- Laser-based biophysical technology

Methodology

Wasserstein Generative Adversarial Network (WGAN)

- WGAN z embeddings as feature for cancer cell identification
- Semi-supervised learning
- Minimization of Wasserstein distance to stabilize training procedure

Network Architecture

- Fully connected network for generator and discriminator
- Generator: Mapping from low (3) to high dimensional space (9)
 - Activation function: Rectified Linear Unit (ReLU)
 - *tanh*-function (on top of linear output)
- Discriminator: Mapping from high to low dimensional space
 - Activation function: Leaky ReLU
 - Linear output layer
- 4 hidden layers (128 neurons), 1 output layer

WGAN Training

- Unsupervised learning of a low dimensional latent space (parametrization of generator and discriminator)
- Parametrization freeze
- Mapping procedure from data to embedding space with minimum residual
- Supervised Neural Network classifier (NN) training

Alternative Approaches Evaluated

- Unsupervised PCA + NN
- Fully connected Neural Network (FNN)

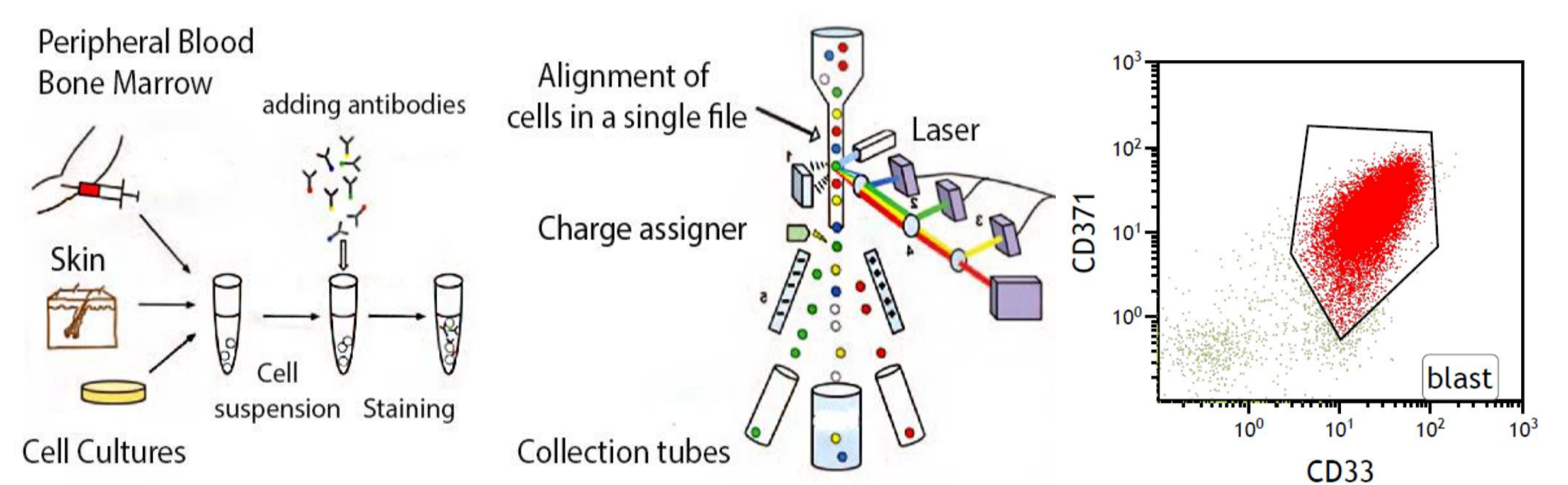
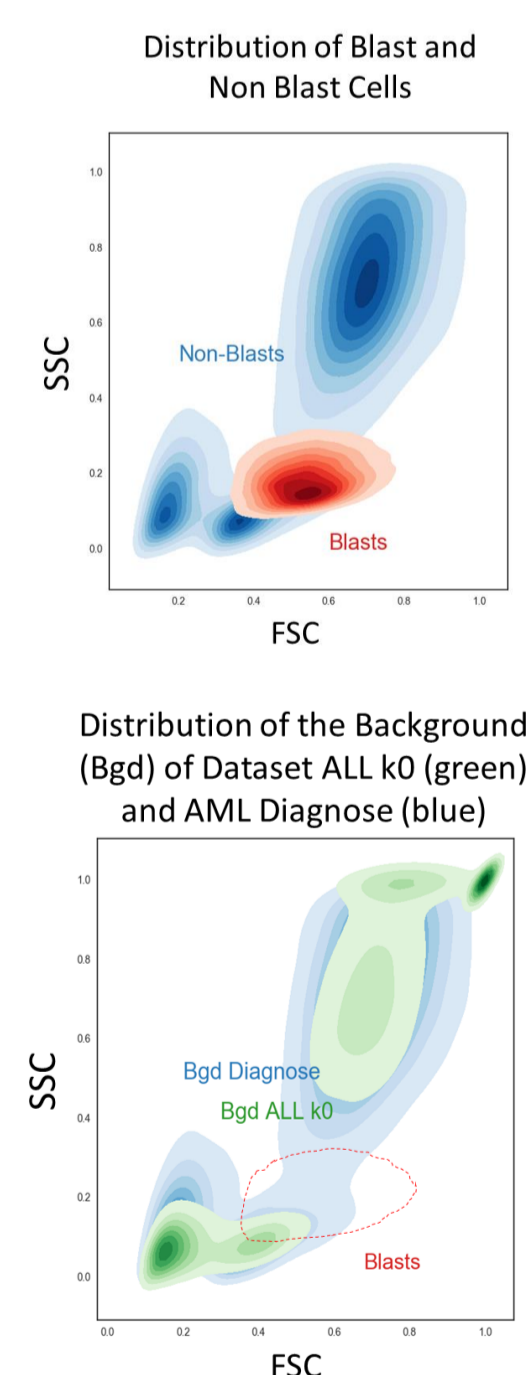
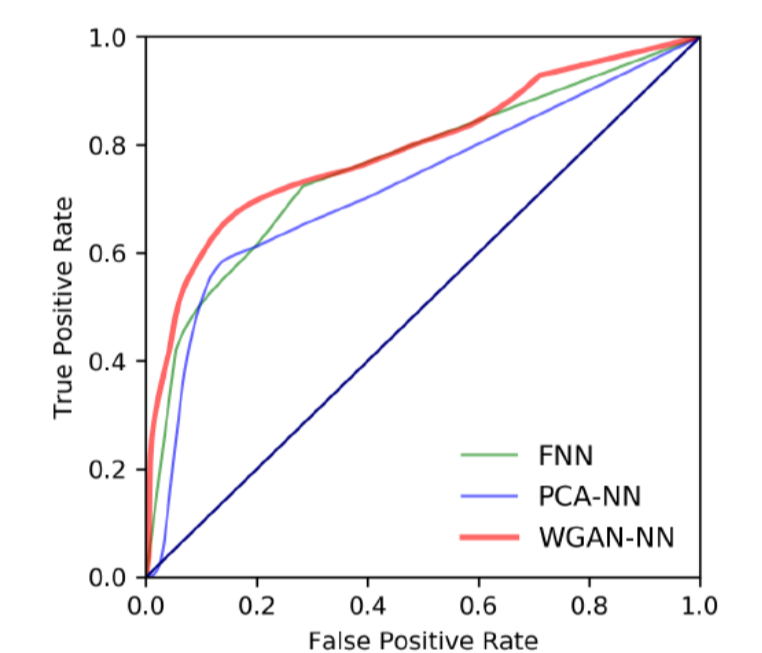


Figure: Flow cytometry measurement procedure and annotated data example

Results

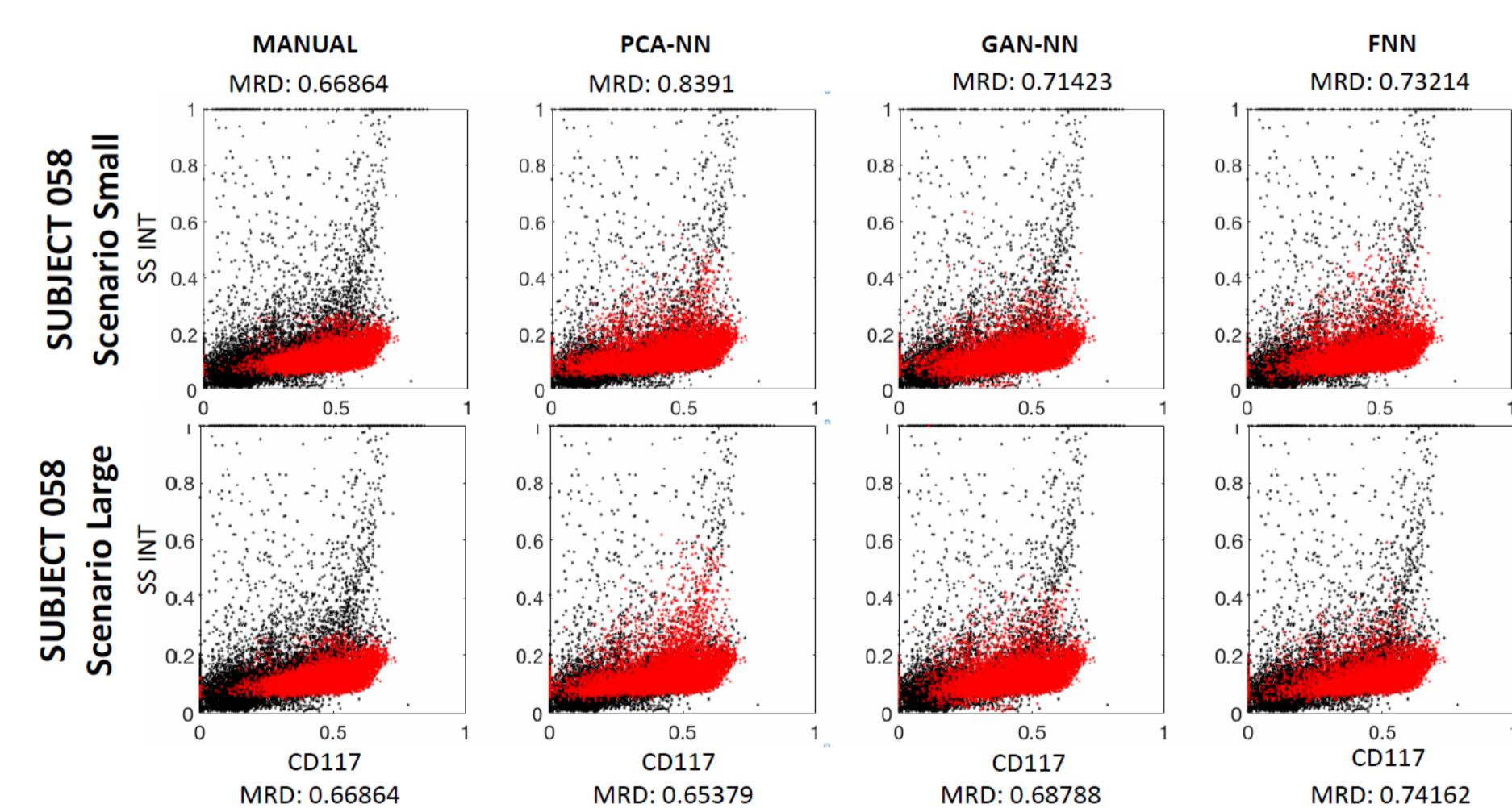
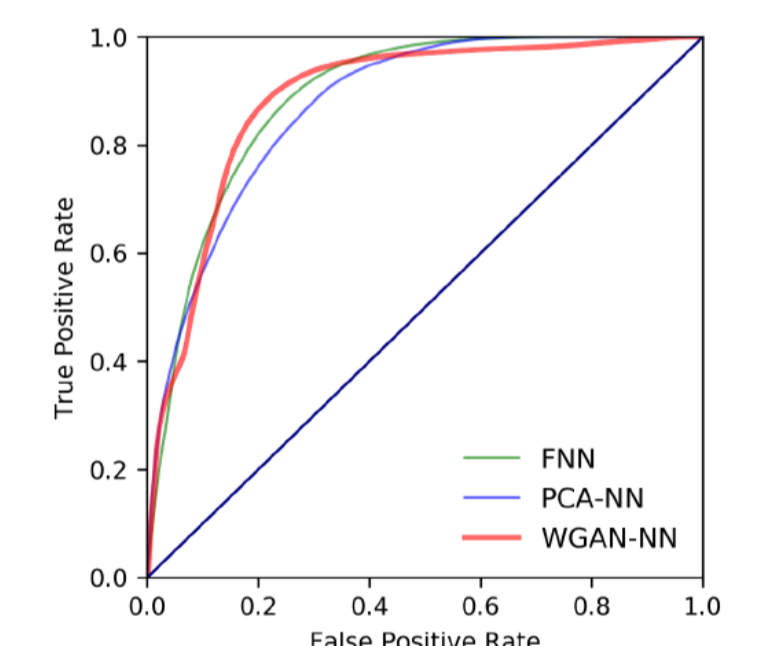
Scenario Small Dataset

Methodology	Precision	Recall	f-score	Specificity	AUC
PCA-NN	0.5471	0.7690	0.5829	0.8151	0.7952
FNN	0.6037	0.7543	0.6164	0.8594	0.8336
WGAN-NN	0.5502	0.9049	0.6483	0.7452	0.8592



Scenario Large Dataset

Methodology	Precision	Recall	f-score	Specificity	AUC
PCA-NN	0.5986	0.9213	0.6893	0.7916	0.9008
FNN	0.6526	0.9250	0.7305	0.7933	0.8892
WGAN-NN	0.6320	0.8942	0.7147	0.8364	0.9139



References

Licandro R.* and Schlegl T.*, Reiter M., Diem M., Dworzak M., Schumich A., Langs G., Kampel M., "WGAN Latent Space Embeddings for Blast Identification in Childhood Acute Myeloid Leukaemia", *24th International Conference on Pattern Recognition (ICPR) 2018*, Beijing, August 2018.

Licandro R., Reiter M., Diem M., Dworzak M., Schumich A., Kampel M., "Application of Machine Learning for Automatic MRD Assessment in Paediatric Acute Myeloid Leukaemia", in *7th International Conference on Pattern Recognition Application and Methods (ICPRAM) 2018*. IEEE, Madeira, January 2018.

*) contributed equally to this work

Conclusion

- Novel semi-supervised learning approach based on a combination of a WGAN latent space embedding and a simple fully connected neural network.
- Rareness of the diseases: creation of small datasets simulating the fact.
- WGAN outperforms PCA-NN and FNN.
- Future work: Focus on obtaining machine and country independent data representations.