

Predicting treatment of rheumatoid arthritis with LIVI

Prediction with a classifier on top of the LIVI model

Esther Wit: e.f.n.wit@student.tudelft.nl

1 Introduction

Developments in **single-cell RNA sequencing (scRNA-seq)** have enabled large scale research into how genomic variations such as single-nucleotide polymorphisms (SNPs) affect the phenotype, allowing for better understanding of diseases such as **Rheumatoid Arthritis (RA)**.

The LIVI Model (Vagiaki et al., 2026)

LIVI (Latent Interaction Variational Inference) is a novel **Variational Autoencoder (VAE)** for modeling scRNA-seq gene expression. Its key innovation is multiple latent spaces, capturing **both cell and donor space**.

This research applies LIVI to scRNA-seq data from **82 synovial tissue samples** from RA and osteoarthritis (OA) patients. This project researches whether simple supervised **classifiers** (MLP/SVM/Random forest) can recover biological and clinical labels from the LIVI latent spaces.

Main research question

Do the latent spaces of the LIVI model capture the major cell type, sub-cell type and treatment succes/failure?

2 Data

The dataset (table 1) consists of synovial tissue samples (n=85) of individuals with RA (n=72) or OA (n = 10). On 3 individuals repeat sampling has been performed. In total, scRNA-seq on these samples resulted in n = 314 011 cells.

Table: Population characteristics of the dataset of Zhang et al.

	n or Mean (SD)
Diagnosis	
Rheumatoid arthritis (RA)	72
Osteoarthritis (OA)	10
Demographics	
Age (years)	57.9 (14.8)
Sex – men / women	22 / 60
Treatment group	
Naive	28
Methotrexate failure	28
TNF failure	16
Osteoarthritis control	10
Repeat	3

3 Model

LIVI separates latent spaces: C for cell-factors, D for donor factors and V for global donor-variation. DxC is a cell-donor combination.

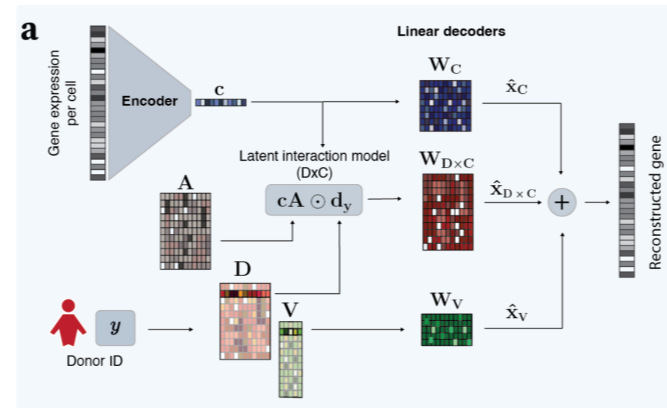
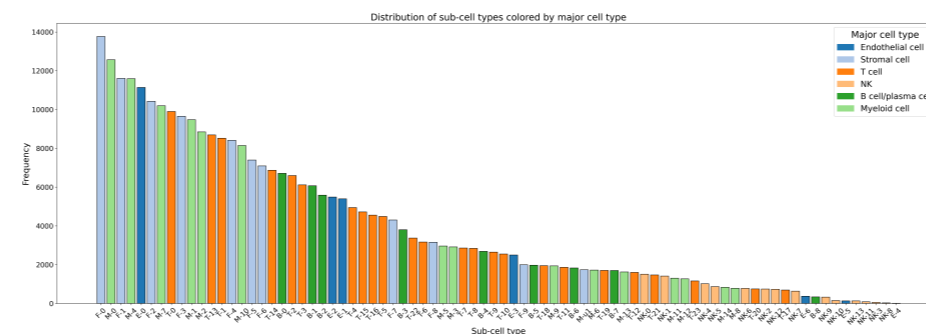


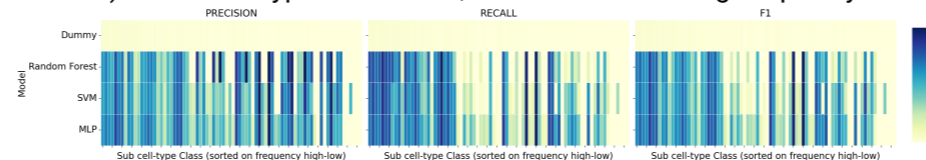
Figure: LIVI encodes gene-expression (scRNA-seq) in C,D and V latent space and can be decoded using decoders WC, WDXC, WV. The $D \times C$ latent space (latent interaction model) maps the donor level of the respective patient to the cell level using assignment matrix A, resulting in a row for each cell with donor factor levels based on the donor id and cell expression. Figure from Vagiaki et al.

4 C latent space for sub-cell types

Major cell-types are easily distinguished, opening up the possibility to classify sub-cell types from the C latent space.



a) Plot of cell type distribution, sorted on decreasing frequency



b) Heatmap of sub-cell type classifier on the C latent space

Figure: Distribution of the cell types (a), colored by major cell type. Heatmap of sub-cell type classifier on the C latent space with the metrics precision, recall and F1 score (b). Classes are sorted descending on frequency on the x-axis (order equal to Figure a). Larger classes (left) have better precision, recall and F1 scores. Smaller classes (right) have worse recall, indicating that these labels are predicted relatively less compared to larger classes.

5 DxC for treatment success/failure

DxC latent space captures both donor-and cell variation. The DxC space is used to predict the treatment **failure/success** per cell based on the corresponding donor factors.

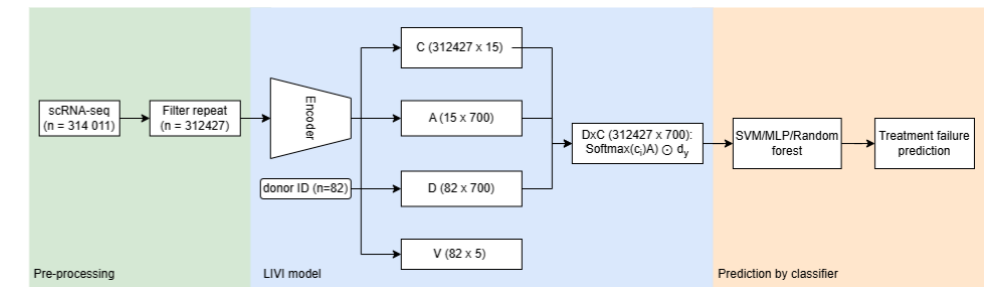


Figure: Study design of prediction treatment failure/success based on the DxC latent space. Repeat sampling is filtered out of the data.

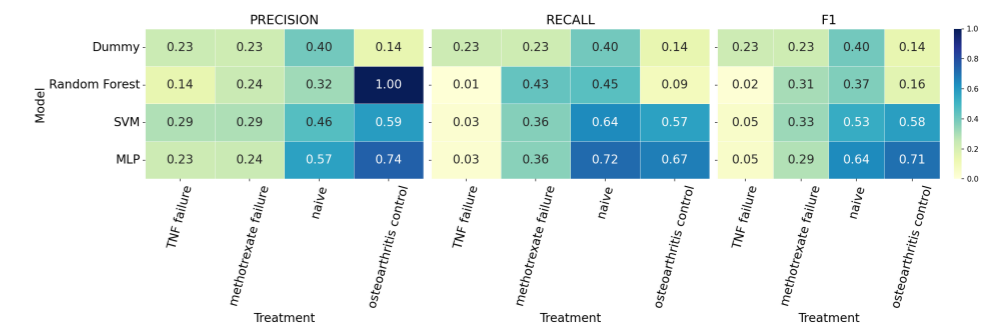


Figure: Heatmap of the precision, recall and F1 scores per classifier for each class on the test data. Precision is high for OA for all classifiers. Random Forest is outperformed by SVM and MLP on the other categories. Difference between SVM and MLP can be seen in methotrexate failure.

6 Conclusion

Major cell types are clearly distinguishable from the C latent space, sub-cell types are partially captured in the C space, and treatment response is partially captured in the DxC space, but not fully separable by a simple classifier. These findings indicate that **biologically and clinically** relevant information is **preserved** within the **LIVI latent spaces**.

7 Future Work

- Gradient boosting as alternative Random Forest
- Evaluate number of layers MLP as a hyperparameter
- Research D latent space if more donors available