Researching the cell – amyloid plaque relationship in Alzheimer's disease

Dimitar Smenovski[1]

Supervisor(s): Marcel Reinders[1]**, Roy Lardenoije**[1]**, Timo Verlaan**[1]**, Gerard Bouland**[1]

[1]**EEMCS, Delft University of Technology, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 22, 2025

Name of the student: Dimitar Smenovski
Final project course: CSE3000 Research Project
Thesis committee: Marcel Reinders, Roy Lardenoije, Timo Verlaan, Gerard Bouland, Ricardo Marroquim

Abstract

Single-cell RNA sequencing (scRNAseq) is a measuring technique of gene expressions in single cells that has allowed researchers to tackle Alzheimer's disease (AD) in many ways. Single-cell data has been joined with machine learning to classify brain cells as affected by AD. However, not much is known regarding the usage of such classification models in a spatial setting. This paper analyzes how models trained on scRNAseq data can be used to find AD properties of single cells when measuring them with spatially resolved transcriptomics. With that we study the hypothesis that cells labeled as affected by the disease should appear closer to amyloid plaques, than those that are unaffected. To find out if this holds, three models are used to classify single cells spatially and their predictions are analyzed. Two single-cell datasets are used for training, each giving a drastically different classification outcome. The models do not come to a consensus on the hypothesis' validity either, as the analysis finds no significant correlation between the variables.

1 Introduction

Alzheimer's disease (AD) is a neuro-degenerative disorder that affects people's memory, thinking and communication skills [1]. This loss of cognitive abilities, called dementia, varies largely depending on the stage of the disease people are in, as it progresses over a long time period. The most widely affected group are the elderly, for instance in America the majority of cases are people aged 65 years or older [1]. Several features characteristic to the disease are known, such as build up of amyloid plaques and tau tangles inside the brain [1]. However, despite scientist's understanding of these features, the exact causes of the disease are not known, making it difficult to diagnose patients in early stages, offer effective treatments or develop a cure.

Knowing individual cells' gene expression helps scientists uncover more details about their heterogeneous nature and how they are affected by the disease [2]. These expressions can be measured using a modern technique known as single-cell RNA sequencing (scRNAseq). Despite its advantages, scRNAseq has one caveat which is that it does not capture the information regarding cells' location in the tissue, as the sequencing process requires to isolate them. This limits the possibility to study the disease in more detail [3].

To circumvent this issue, scientists have put forth another method for measuring cells' gene expression - spatially resolved transcriptomics. Its value comes from the ability to again measure expressions at single-cell resolution, while preserving the spatial information regarding a cell, such as its coordinates inside the tissue [3]. The downside of this approach is that it can only measure a subset of the whole transcriptome, unlike scRNAseq which does so in its entirety [4].

Fortunately, researches have already taken notice of this. Several methods have been developed that integrate spatial transcriptomic and scRNAseq datasets, including SpaGE [4] and Tangram [5]. These tools can predict the expression of genes not found in the spatial data, but which are present in the scRNAseq data. For this to work, they require that the datasets have a set of shared genes to be used as reference for the expected expression values.

With the rising adoption of machine learning in all fields of science, some researchers have delved into how such systems can be used to determine the extent to which single cells are affected by AD. For example, cell-level classification models do this by projecting sample metadata onto individual cells [2]. However, this results in all cells having the same disease status, which does not correspond with the notion that different cells are affected at different degrees by the disease [2]. To avoid this, sample-level classification models, such as scAGG, label cells individually based on gene expressions, before aggregating them to determine the entire sample's disease status [2]. Nonetheless, the aforementioned studies do not consider how their models can be used in a spatial setting.

A study which employs spatial transcriptomics has identified several effects that plaques have on the transcription of cells within a 100 micron diameter [6]. Such effects are the expression of plaque-induced genes in astrocytes and microglia, as well as AD-associated genes in oligodendrocyte cells aptly named OLIGs [6]. Another study finds that microglial cells in close proximity of plaques have increased Ca^{2+} activity, which is a response that is thought to be associated with inflammation and phagocytosis [7].

Based on these findings, we put forward the hypothesis, H1, that cells affected by AD should be located closer to plaques than unaffected healthy cells. When combined with spatial transcriptomics, gene expression based cell-level classification makes it possible to study how strongly affected cells are located relative to AD pathology.

The main aim of this study is to analyze the validity of H1 by examining the relationship between cells' spatial profile and their gene expression. Additionally, it provides a method for labeling spatial transcriptomic data using a classification model pre-trained on whole-transcriptome scRNAseq data, and determines the effectiveness of different models on this task.

2 Methodology

2.1 Datasets

This study makes use of three datasets from AD donors: ROSMAP [8][9], SEA-AD [10] [11] and Xenium single cell spatial transcriptomics.

- ROSMAP consists of gene expression data with samples taken from the dorsolateral prefrontal cortex (DLPFC) of donors. The pathological severity differs across donors, although a direct measure is not present in the data. The full data consists of 725109 observations (obs) with over 19000 expressed genes from 450 donors. The observations are further divided into cell types, where microglia is 86612 obs, astrocytes is 228925 obs and oligodendrocytes is 409572 obs.

- SEA-AD, specifically the 10x single nucleus RNAseq data, also stores the gene expressions of cells in the PFC, but here only the nucleus is sequenced rather than the entire cell. The pathological severity of each donor is denoted in the data with four stages - No AD, Low, Intermediate and High. For the purposes of this study, 9 donor objects were combined - 3 with No AD, 3 with Low AD, and 3 with High AD. Their IDs are respectively H19.33.004, H20.33.002, H20.33.036, H20.33.001, H20.33.032, H21.33.001, H21.33.003, H21.33.032, H21.33.045 in order of severity. The full SEA-AD data contains 34319 obs and 36601 genes. Per cell type this is 9674 astrocyte cells, 5536 microglia and 19109 oligodendrocytes.
- Xenium is a spatial transcriptomics dataset, which contains both gene expression data and information regarding each cell' location in the tissue, such as coordinates and distance to plaque in microns. However, it is not whole-transcriptome, unlike the other two datasets, as only 266 genes are measured. Furthermore, it comes from only 1 donor who had severe AD and Cerebral Amyloid Angiopathy (CAA). The total number of observations is 58132, consisting of 45138 astrocytes, 2160 microglia and 10834 oligodendrocytes.

All three datasets measure multiple cell types - from neurons to glial cells. However, for this study only astrocytes, microglia and oligodendrocytes are considered. This choice reflects literature that has identified activation of such cells around amyloid plaques [6][7].

2.2 Pre-processing

All of the pre-processing and later computations are handled with random state set to 56. This ensures that the results remain the same when reproduced, given the same data.

Initially, the ROSMAP dataset only contains raw counts. However, as this counts data is highly sparse, it cannot be used directly for classification. Instead it must be filtered, normalized and standardized. Although SEA-AD and Xenium do have scaled variants of their gene expression counts, they are also preprocessed from the raw counts in order to ensure all datasets are in the same final format. For SEA-AD the raw counts are stored in layer 'UMIs'. Most of the preprocessing follows the standard SCANPY approach [12].

For model training, the ROSMAP and SEA-AD data are filtered on cells and genes. Cells with less than 100 expressed genes and genes expressed in less than 3 cells are left out. Moreover, only those cells which have less than 5% mitochondrial genes expressed are kept [12][13]. Next, the most highly variable 1000 genes are selected using the 'seurat_v3' flavor which requires raw counts. The resulting data is then normalized, so that all counts per cell add up to the same number equal to the median of the total counts before normalization, logarithmized and scaled to unit variance and zero mean.

For imputation, the datasets, including the spatial transcriptomic data, are normalized and logarithmized. After this, a clustering is performed on the data using leidenalg [14], as this is required by Tangram in order to map the clusters onto the spatial data [15]. Additionally, the ROSMAP data undergoes a downsampling procedure by randomly selecting a quarter of the olygodendrocyte cells and half of the astrocyte cells, reducing each to about 100k cells. This step is done, because the unfiltered observations for these two cell types are in the order of hundreds of thousands, which greatly slows down the imputation and increases RAM usage.

Before classification, the imputed spatial data transcriptome is subset to the same 1000 genes used in training. It also undergoes the same procedure as the model training data by filtering before selecting the top genes, then normalizing, logarithmizing and scaling after the genes have been subset. This ensures the spatial data is in the same format as the model training data.

The ROSMAP data is labeled using the "ROSMAP_clinical.csv" metadata file according to the approach described in [16]:

- AD (109 samples): cogdx = 4, braaksc $\geq$ 4, ceradsc $\leq$ 2
- CT (61 samples): cogdx = 1, braaksc $\leq$ 3, ceradsc $\geq$ 3
- Other (279 samples): all other samples.

The metrics used are cogdx (cognitive diagnosis) - a clinical assessment of cognitive impairment in patients, braaksc - a measure of the neurofibrillary pathological severity [17][18] and ceradsc - a measure of the density of amyloid-beta protein neuritic plaques [19]. The outcome gives twice as many AD cells as CT for all cell types.

The SEA-AD data already contains a column regarding donors' pathological severity as mentioned previously. To create the appropriate labels, the 'No AD' and 'Low' severities are marked as CT, while the 'Intermediate' and 'High' are marked as AD. The total outcome is 25413 CT and 8906 AD cells, nearly a 3 times difference. When considering just oligodendrocytes, the CT cells are roughly 4.2 times more, while for the other two cell types this is only around 2 times.

2.3 Approach

The general approach is to train AD classification models from scRNAseq data and use them to classify the cells in a spatially resolved transcriptomic dataset as healthy (CT) or diseased (AD). The single-cell training data is labeled based on donor-level pathological severity that is then transferred to the cell-level by the associated donor ID.

ML models

Classification is performed with three models. Those being a linear classifier (LC), a multilayer perceptron classifier (MLP), and an MLP classifier with dropout (MLP+Dropout). For the purpose of model training, the data target labels are converted to numerical form where CT cells label is set to 0 and AD cells is set to 1. The models architecture is as follows:

- Linear classifier - one linear layer with 1000 input and 2 output dimensions.
- MLP classifier - 3 layers, two Linear and one ReLU in between, with the linear inputs and outputs being (1000, 512) and (512, 2) respectively.
- MLP classifier with dropout - 4 layers, the same as the general MLP model, with the addition of a dropout layer at 0.3 dropout rate after the ReLU layer.

Model training
All models use the ADAM optimizer with a learning rate of 0.0001. Similarly, all share the same cross-entropy loss function. The losses are summed per batch, rather than averaged. No regularization parameter is added.

The models are trained two times. Once on the ROSMAP data and once on the SEA-AD data. Training occurs in batches of size 128 with shuffling of the data. This is done on 100% of the data for 21 epochs. The resulting model weights of both runs are saved for classification.

The training quality of the models is measured using accuracy. The training accuracy is calculated at each epoch as the average of the accuracies per batch. To calculate the latter, the output of each batch is matched with the training labels and the number of correct predictions is divided by the batch size. The resulting accuracies per epoch are visualized in a line plot, which consists of three distinctly colored and shaped lines for each model.

Model evaluation
The models' ability to classify unseen data is evaluated after each training epoch by again calculating accuracy. The same loss function is used as for training. The evaluation is done twice for each model training run. When the models are trained on ROSMAP, they are evaluated on SEA-AD and vice versa.

The accuracy is calculated in the same manner as during training. It is again the average of the accuracies per batch, where each batch is of size 128 and shuffled. All of the data is used for evaluation.

Applying models to spatial transcriptomic data
After the models are trained, they are used to classify the spatial data. However, this is not directly possible as Xenium's gene set is only 266 transcripts, while the training data uses 1000. Despite being able to reduce the dimensionality of both to an equal number, not including all relevant genes in the process is likely to lead to poor results. To deal with this, first the missing genes in the spatial data are imputed from the two single-cell datasets using Tangram [5]. The imputation occurs in three steps: finding the common genes using 'tg.pp_adatas', creating a matrix based on these shared genes that maps single-cell to spatial data entries using 'tg.map_cells_to_space', and based on this mapping projecting the single-cell gene expressions to the spatial data with 'tg.project_genes'. For Tangram to work reliably, the single-cell and spatial datasets' gene sets must intersect. The imputation is done using the whole reference dataset transcriptome, rather than only the 1000 gene subsets. This is because the total shared genes are higher before subsetting - all 266 Xenium genes are present in SEA-AD, and all but one in ROSMAP. Before the actual imputation, the spatial and reference datasets are preprocessed, without excluding any shared genes, as described in the previous subsection. The mapping from cells to space is done for 5 epochs at cell level (mode="cells") with density_prior="rna_count_based" on CPU [15]. The number of epochs was chosen to be the highest number which is able to be run on the cluster. Even for 5 epochs, the full ROSMAP data requires over 300GB of RAM, so anything higher than 5 was not possible.

The imputation is validated by a leave-one-gene-out procedure for 10 cluster marker genes of the spatial data. The genes are determined by a differential expression test using the wilcoxon method. For each marker gene, the spatial data gene expressions are imputed for 5 epochs by leaving this marker gene out. After imputation, the predicted expression is compared to the true measurement using Spearman's rank correlation coefficient. This validation is performed only for the SEA-AD imputation, due to limitations.

Once the spatial data is imputed and processed, it can be classified using the pre-trained models. When classifying, there are two configurations: one is to impute the spatial data using SEA-AD as reference and to label it using the model trained on ROSMAP. The other is to impute using ROSMAP and classify using the SEA-AD model. This is done to prevent information that was used for training from leaking into the spatial data. The model outputs two raw values per class for each data point. To map the output to a list of predicted labels, an argmax function is used, which takes the higher value and reduces the inner list to this value's index. The share of the resulting labels is visualized using a bar chart. Furthermore, the raw output is converted to a list of probability values, representing the likelihood of a cell being AD, using the softmax function. These probabilities are then plotted on a histogram with bin size 15. The histogram could be used to check whether the output follows a random uniform distribution or a U-shaped distribution where the two extremes are predicted more frequently than mid-range values.

Analyzing spatial transcriptomic data distribution
To illustrate the difference in how the two class labels, AD and CT, are distributed along the distance to plaque metric, a split violin plot with quartile interior is drawn where 'x' is the distance and 'hue' is the label set with order 1, 0 for AD, CT. For a quantitative evaluation, the Mann-Whitney U test is applied [20]. This is a nonparametric statistical test which measures the difference between two distributions and whether one is stochastically greater than the other. In the case of H1, it tests whether the AD label distribution is stochastically lower than the CT label distribution. The test returns a U statistic and a p-value. The former is the number of times an element from one distribution appears closer to a plaque than an element of the other. The latter represents how confident the test is that one distribution is stochastically greater or lower than the other, with values less than or equal to 0.05 implying strong significance.

Visual analysis of spatial transcriptomic data
Four maps of all cells in the tissue are created; three colored using the predicted AD probabilities of cells and the one using the distance to plaque in microns. The AD probability maps are three as there is one for each model. The cells (dots) are enlarged to make them more clearly visible. Furthermore, their alpha value is set to 0.5 to give them some transparency as they can overlap when enlarged. By comparing the AD probability maps with the distance to plaque map, the expectation is to see a negative correlation, such that cells with higher probability of being AD have a lower distance to plaques. This correlation is quantitatively measured using Spearman's rho. The test outputs a correlation coefficient in

range -1 to 1, where the two end points represent strong negative correlation and strong positive correlation respectively, with 0 being no correlation. It also returns a p-value, which hints that the observed correlation has not occurred purely by chance when less than or equal to 0.05.

Implementation details

The implementation is written in Python. Used frameworks include anndata 0.10.9 [21], matplotlib 3.9.4 [22], numpy 2.0.2 [23], pandas 2.2.3 [24], scanpy 1.10.3 [12], scipy 1.13.1 [25], seaborn 0.13.2 [26], tangram-sc 1.0.4 [5], torch 2.7.0 [27], leidenalg 0.10.9 [14] and their required dependencies. A full list of the requirements can be found in Appendix A.

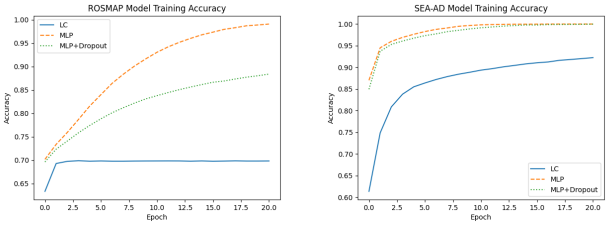
The reason that Tangram (tangram-sc) [5] was used for imputation and not another algorithm, such as SpaGE [4], is that it happened to be the fastest one on the machines used for the experiments in this paper. Nonetheless, any imputation method should work for reproducing the results, as long as it is designed for integrating scRNAseq and spatial transcriptomic datasets.

As the ROSMAP data comes in individual h5ad files for each cell type, they are combined into one using anndata's concat along axis 0. However, this is only done when building the models. When doing the imputation, the data is imputed per cell type, as the entire ROSMAP data is too large to apply at once. However, for SEA-AD the imputation happens in one go using all three cell types.

3 Results

Three models - Linear Classification (LC), Multilayer Perceptron (MLP) and MLP + Dropout are trained on scRNAseq datasets ROSMAP and SEA-AD. Their performance is measured using accuracy. These models are applied to spatial transcriptomic data, of which the missing gene expressions are imputed. The imputation is validated using a leave-one-gene-out procedure. The results of the models' classification on the spatial data are analyzed qualitatively using violin plots of the distance to plaque distribution of labels and maps of cells' location within the tissue, as well as quantitatively using Mann-Whitney U Test and Spearman's rho statistics.

Model training performance



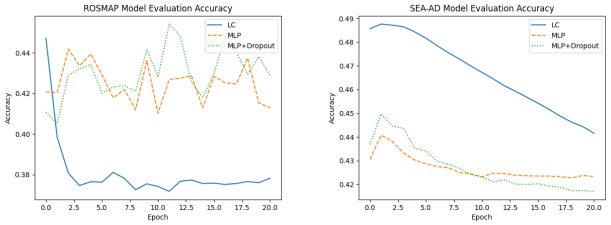
(a) (b)

Figure 1: Training accuracy across epochs of the three models trained on (a) ROSMAP and (b) SEA-AD

The training accuracies per epoch for the three models, LC, MLP and MLP+Dropout, on the ROSMAP and SEA-AD datasets, and the evaluation accuracies of the models' predictions for each training epoch on the dataset they are not being trained on, are illustrated in Figures 1a, 1b and Figures 2a, 2b respectively. During training the most performant model is MLP, reaching nearly 100% accuracy on both datasets. Following it are MLP+Dropout in second place and LC in last. The exact accuracies are presented in Table 1.

	LC	MLP	MLP+Dropout
ROSMAP	0.6984	0.9909	0.8840
SEA-AD	0.9225	1.0000	0.9997

Table 1: Last epoch training accuracy of each model by dataset



(a) (b)

Figure 2: Evaluation accuracy across epochs of the three models (a) trained on ROSMAP, but evaluated on SEA-AD and (b) trained on SEA-AD, but evaluated on ROSMAP

The testing accuracy is quite a lot lower compared to training, as pointed out in Table 2. The two MLP models perform equally well on both datasets. However, when evaluated on SEA-AD they appear stuck, showing no signs of improvement as training progresses. During evaluation on ROSMAP, the trend is even negative, with accuracy appearing to decrease as the models train. The situation with the LC model is even worse. With SEA-AD, its accuracy quickly drops below 38% in the first 3 epochs, flattening out around that value without any increase later on. With ROSMAP there is also a sharp decline after the first few epochs. Despite this, the LC model still performs better until the 20th epoch during evaluation on ROSMAP, compared to the other two. However, this is unlikely to remain the case if more epochs were added as it declines much faster than the other models.

training / evaluation	LC	MLP	MLP+Dropout
ROSMAP / SEA-AD	0.3783	0.4130	0.4288
SEA-AD / ROSMAP	0.4416	0.4233	0.4171

Table 2: Last epoch evaluation accuracy of each model by training dataset (left side) and evaluation dataset (right side)

Validation of gene expression imputation

The validation of SEA-AD imputed data for 10 marker genes is illustrated in Figure 3. There appears to be moderate positive correlation between the true and predicted gene expressions. This implies that the imputation works well, though it is not perfect.

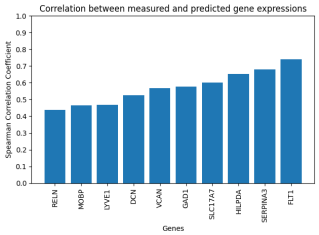
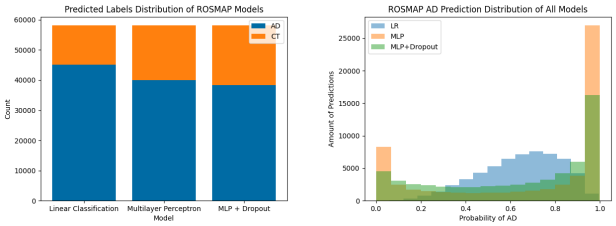


Figure 3: Spearman rank correlation coefficient between measured and predicted using SEA-AD marker genes expression

Classification output for Xenium data

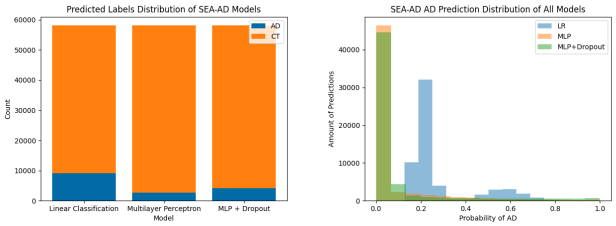
The results of the classification on the Xenium data using the models trained on ROSMAP and SEA-AD are presented in Figures 4a, 4b and 5a, 5b respectively.



(a) (b)

Figure 4: All Cell Type Spatial Data ROSMAP Models Distribution of (a) Predicted Label and (b) Probability of AD

The ROSMAP-trained models, classifying the SEA-AD imputed spatial data, label the majority of cells as AD. Similarly, the histogram shows two spikes in predictions around the values 0 and 1, representing CT and AD respectively, with the spike at 1 for AD being much higher. However, this is only the case for the two MLP models. Despite their resemblance, the general MLP model makes a much clearer distinction between the two class labels. In contrast, the LR model predictions are much more condensed with the mean appearing around 0.7, which explains the high number of predicted AD labels.



(a) (b)

Figure 5: All Cell Type Spatial Data SEA-AD Models Distribution of (a) Predicted Label and (b) Probability of AD

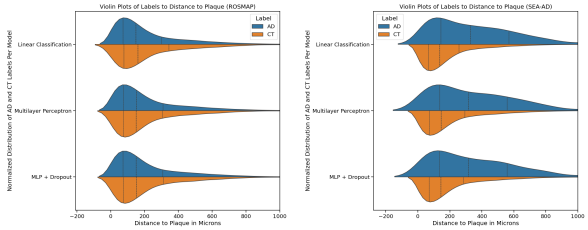
On the other hand, the models trained with SEA-AD data, when classifying the full spatial data that is imputed using ROSMAP, label the majority of cells as CT instead. This split is also noticed in the histogram, which shows a large spike at 0 for the MLP models, and one smaller at 0.2 for LR. Again, the general MLP model's predictions are more skewed toward 0 probability than the one with dropout.

Analysis of predicted AD and CT labels distribution across distance to plaque metric

Figures 6a and 6b both depict three violin plots, one per model, of the distance to plaque metric distribution by label, where the orange half is CT and the blue half is AD. The plots also contain dashed vertical lines representing the quartiles of the distributions, where the middle one is the median. The distributions are normalized, as to have equal heights rather than being disproportionate.

The first figure, 6a, shows the distance-label distributions for the SEA-AD imputed spatial data classified by the ROSMAP-trained model. The violin plots for each of the three models are quite similar, and more importantly so are the AD and CT halves of each distribution. Some small deviations in the means can be seen for LC and MLP+Dropout, where the AD distribution median is slightly closer to 0 than that of the CT labels.

The violin plots of the SEA-AD trained models also appear to be quite similar. However, in this case the two halves of each violin appear to be much more different. All models' AD labels distribution is heavily shifted to the right compared to the upper half. Moreover, even the median of the AD distribution is positioned higher along the axis than the other distribution's 3rd quartile. Another notable difference is the LC model's CT labels distribution being skewed slightly more toward 0. This is also evident by the lower median line, which appears before the first AD quartile, as opposed to the other two models where the CT median comes after it.



(a) (b)

Figure 6: Distance to Plaque Distribution of Labels Using (a) ROSMAP Models and (b) SEA-AD Models

The above observations are quantitatively measured using Mann-Whitney U tests for the ROSMAP-trained models and SEA-AD trained models on full data. The results are presented in Tables 3 and 4 respectively.

The U tests on ROSMAP-trained models identify a statistically significant result for the LC model with p-value of near 0, thus favoring the alternative hypothesis H1. On the other hand, the two MLP models do not cross the 0.05 confidence threshold, although MLP + Dropout is close.

Model	U statistic	p-value
Linear Classification	277427877.0	0.0
Multilayer Perceptron	364146700.0	0.6498
MLP + Dropout	377508247.0	0.0926

Table 3: Mann-Whitney U Test for ROSMAP-trained models

Regarding the SEA-AD trained models, all U tests show that the inverse of H1 holds, as the p-value is equal to 1. This implies that the AD labels are distributed significantly higher on the distance to plaque axis relative to the CT labels.

Model	U statistic	p-value
Linear Classification	313531076.5	1.0
Multilayer Perceptron	99405932.5	1.0
MLP + Dropout	154318028.5	1.0

Table 4: Mann-Whitney U Test for SEA-AD trained models

Analysis of correlation between AD probability and distance to plaque metric

Following are several maps of cells' location within the sample tissue, starting with Figures 7a to 7d. The first image depicts where to expect more cells that are labeled as AD and where as CT. The maps of cells colored based on ROSMAP-trained models' predictions, i.e. 7b to 7d, show a seemingly random assignment of probabilities with overwhelmingly red-colored AD cells and a few blue CT dots spread uniformly throughout the tissue.

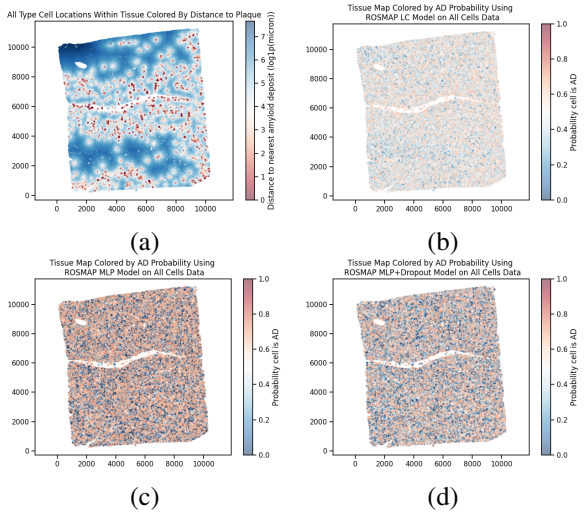


Figure 7: Maps of All Cells Inside Tissue Colored by (a) Distance to Plaque and (b) Probability of AD Based on LC, (c) MLP and (d) MLP+Dropout ROSMAP Models

The tissue maps colored based on SEA-AD trained models predictions for all cell types shown in Figure 8 have majority blue CT cells with a few bright regions where the probability of AD is higher. It is most notable that these bright regions coincide with the areas that are farthest from plaques. According to the reference map in Figure 7a, these regions should be darker compared to the surroundings instead.

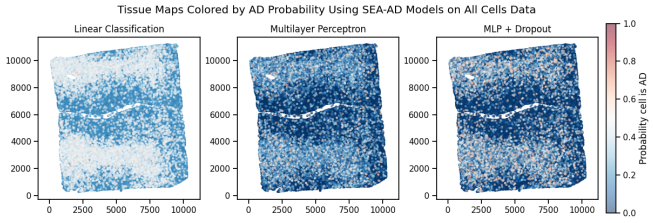


Figure 8: Maps of All Cells Inside the Tissue Colored by Probability of AD for SEA-AD Trained Models

The SEA-AD models' predictions are also analyzed on the spatial data filtered to include only oligodendrocyte cells. From the first map, 9a, it can be seen that most cells appear in the areas which are far from plaques. In Figures 9b to 9d these cells are also colored in blue, which means the SEA-AD model accurately predicts AD for this cell type, even if by chance. Nonetheless, some orange to red dots can still be seen in the tissue, with more of them being visible in the darker MLP model maps.

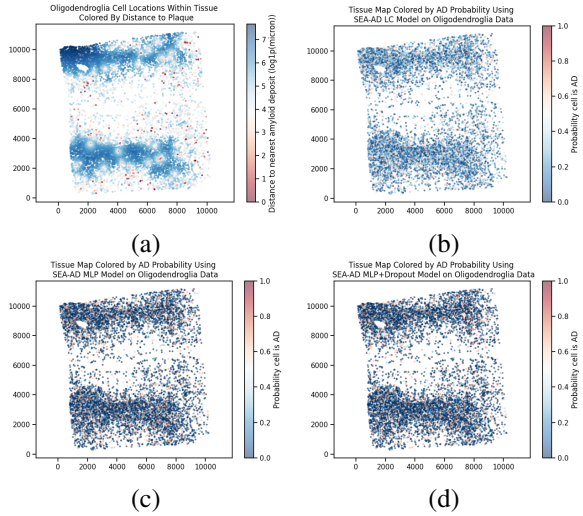


Figure 9: Maps of Oligodendrocyte Cells Inside Tissue Colored by (a) Distance to Plaque and (b) Probability of AD Based on LC, (c) MLP and (d) MLP+Dropout SEA-AD Models

The correlation between a cell's proximity to a plaque and the probability of it being AD, visualized in the tissue maps, is quantitatively analyzed using Spearman's rho. The results are given in Tables 5, 6 and 7.

The test on the three ROSMAP-trained models finds a coefficient very close to 0 for all. This implies little to no correlation between the AD probabilities predicted by the models and cells' distance to pathology.

Model	correlation coefficient	p-value
Linear Regression	-0.0480	0.0000
Multilayer Perceptron	0.0005	0.5506
MLP + Dropout	-0.0063	0.0638

Table 5: Spearman's rho Between Distance to Plaque Metric and Probability of AD for All Cell Types from ROSMAP-trained models

The correlation coefficients for the SEA-AD trained models are a lot higher. However, the correlation is positive, meaning the cells farther from plaques are more likely to have AD. This is the opposite of the proposed hypothesis, which expects a negative correlation. This interpretation is also supported by the high p-values.

Model	correlation coefficient	p-value
Linear Regression	0.2021	1.0
Multilayer Perceptron	0.1956	1.0
MLP + Dropout	0.1718	1.0

Table 6: Spearman's rho Between Distance to Plaque Metric and Probability of AD for All Cell Types from SEA-AD Trained Models

The correlation test of SEA-AD models predictions for oligodendrocyte cells only is more similar to the ROSMAP models results, than the one for all cell types. Despite what was seen in the Figure 9 subplots, where a low likelihood of AD coincided with a long distance from plaques, the correlation coefficients are still close to 0. This is likely caused by the models predicting low AD too often, thus including also cells close to plaques, which counteracts the negatively correlated predictions.

Model	correlation coefficient	p-value
Linear Regression	-0.0197	0.0200
Multilayer Perceptron	-0.0148	0.0615
MLP + Dropout	-0.0157	0.0502

Table 7: Spearman's rho Between Distance to Plaque Metric and Probability of AD for Oligodendrocytes Cells from SEA-AD Trained Models

4 Discussion

Models overfit during training and underperform during evaluation

The results in the previous section of the models' accuracy illustrate their inability to generalize to unseen data. This is evidenced by the lack of increase in accuracy throughout evaluation epochs. As all models score below 50% prediction accuracy, they work no better than a random classifier.

The drops in accuracy when evaluating the models trained with SEA-AD data on ROSMAP data are likely related to the datasets' composition. SEA-AD contains predominately CT cells, while ROSMAP has majority AD cells. Due to overfitting, the SEA-AD models end up labeling a high number of ROSMAP cells as CT, with this number increasing in each epoch, resulting in a decreasing evaluation accuracy.

In contrast, the training accuracy is high for all three models. This, in combination with the evaluation results, implies that the models simply memorize the input dada, rather than learn underlying patterns and features that define it. Further increasing the training epochs is unlikely to improve performance, as all evaluation accuracies are either flat, fluctuating around the same value, or even decreasing.

Nevertheless, some insight can be gained on which model is better suited for classification. Out of the three models, MLP performs better than LC when evaluated on SEA-AD data, and better than MLP + Dropout for ROSMAP data. Even though adding dropout improves accuracy for SEA-AD data, the increase is marginal and does not justify its use. Furthermore, while LC achieves higher accuracy than MLP on ROSMAP, this is unlikely to hold up if the number of epochs increases as the LC model's accuracy follows a sharp downward trend during each evaluation step. This suggests that the best model to use is MLP without dropout.

Classification of spatial data differs significantly depending on the datasets used for training

Changing the training data seems to heavily affect the end result of the spatial data classification. As was postulated in the previous discussion point, the SEA-AD trained models do predict majority CT, whereas ROSMAP-trained models predict AD cases. With that in mind, the reason for the discrepancy in classification output can also be attributed to the models overfitting.

Analysis on predicted AD and CT label distributions across distance to plaque metric does not conclusively reject null hypothesis

The analysis results indicate that AD cells do not appear closer to plaques than CT cells on average. The only model for which this does occur is the LC model when trained on ROSMAP data. On its own, this result is not enough to reject the null hypothesis in favor of the alternative H1.

When trained on SEA-AD data, the models predict AD in cells significantly farther from plaques than cells it labels as CT. This contradicts H1, which is highly unlikely to be true considering the literature on cells' response to plaques. Rather, it could be that the SEA-AD models have learned to predict oligodendrocytes, instead of AD, as these cells' location in the tissue coincides with areas farthest from plaques. As noted in the methodology, SEA-AD does contain a great amount of oligodendrocytes compared to astrocytes and microglia. However, this is also the case for ROSMAP, but its models do not exhibit the same defect.

Analysis finds almost no negative correlation between AD probability and distance to plaque

The models trained on ROSMAP data have not identified any spatial pattern in their predictions, as the maps of cells' AD probability resemble random noise, more so than the map of cells' distance to a plaque. Although the LC and MLP+Dropout correlation coefficients are on the negative side, as desired, they both approach 0, indicating complete lack of correlation.

On the other hand, the SEA-AD trained model predictions for all cell types show a clear spatial pattern. However, the

high AD probabilities in the regions far from plaques hint that this is more a coincidence and an error than actual spatial pattern identification by the models. This contradiction is unlikely to stem from mistakenly using the probability of CT instead of AD, as this would have reflected in the ROSMAP results. The cause could be related to higher number of oligodendrocyte cells in the SEA-AD data. It might also be that the spatial data imputation on ROSMAP causes the gene expression of the spatial oligodendrocyte cells to be representative of AD. This would explain why the areas with such cells are predicted to be closer to AD than CT by SEA-AD models.

In contrast, the same SEA-AD models' predictions for oligodendrocyte-only spatial data appear much more accurate in the areas far from plaques, labeling cells predominantly as CT. Unfortunately, this contradicts the previous claim about imputation significantly biasing cells' gene expressions toward AD, making it less plausible. Nonetheless, the respective correlation coefficients are very close to 0, implying a high classification error rate and models that merely predict the majority of cells as CT off of memory.

Limitations

As the spatial data is so limited, it cannot be deemed sufficient enough to critically assess the alternative hypothesis' validity based on the observed results. At best it can provide support for one of the two outcomes, but for a definitive conclusion further experiments using data from more patients must be carried out.

Another limiting factor in the experimentation process is the number of epochs used for imputation. As the datasets are so large and the mapping matrix that tangram creates has size equal to the multiple of the single-cell observations by the spatial observations the amount of memory and time required to compute the imputation are in the order of 100s of GB and several hours of runtime. This is only possible to do on DAIC, but it comes at the risk of timeout or out-of-memory errors. This further exacerbates the process of imputing missing gene expressions. As such doing it for more than 5 epochs becomes infeasible in the time-frame of the project.

In addition, the models' architecture is quite simple for the same reason as the imputation epochs. Although the current one is able to run on a local machine, more complex networks would require usage of DAIC as well. If more powerful models are used, it might lead to more confident and reliable results that shed new light on the hypothesis.

5 Conclusions

This study provides insights into the hypothesis that cells should be labeled as AD when closer to plaques. This is the expectation, as previous research has found amyloid plaques to have significant impact on neighboring cells, particularly in the range below 200 microns from a plaque. Alongside this, it proposes a method for classifying single cells as AD using spatial transcriptomic data. Furthermore, it contributes by assessing the suitability of different pre-trained models for AD classification.

The analysis on the classified spatial data is highly contradictory. Some models find a moderate, but positive correlation between cells' distance to plaques and their probability of AD, while other models do not identify any correlation. Therefore, these results are not enough to determine the validity of H1. Irrespective of this, MLP without dropout is found to be the preferred model to use for classification, based on training and evaluation accuracy.

Further Research

Researchers who want to build upon this study can do so in several ways. One of the main bottlenecks of this study is the number of imputation epochs. Due to the previously mentioned limitations, only 5 epochs were able to be run. Increasing this amount could improve the quality of the imputed missing gene expressions and in turn provide more conclusive results.

Another problem with the proposed approach is the model training. As it stands the models quickly overfit, and likely do not learn the correct features, but ones that are correlated but easier to learn, like whether a cell is of type oligodendrocyte or not. In order to resolve this, one could dive deeper into methods such as regularization and hyper-parameter optimization.

In addition to the above proposal, aspiring researchers could be more rigorous when preparing the datasets used for training. This is to avoid the issue where the model learns to predict one class, or worse one cell type, purely because it constitutes the majority of the data. As the data in this study has not been so carefully prepared, it occurs, for example, that the SEA-AD data has more CT labeled points than AD, while the opposite is true for ROSMAP. A solution to these issues would be to ensure all datasets contain equal proportions of cell types and classes.

6 Responsible Research

The main ethical consideration when conducting this research is the kind of data being used. For this project, that data comes from human donors who have agreed to have samples taken from their brains postmortem for studying [8]. Ensuring that the donors agree to this procedure is crucial for preserving the ethical integrity of this research.

Most of the data comes from online sources. For ROSMAP, this is the Synapse platfrom. This platform requires users to be granted permission in order to access datasets consisting of human data. This is done by submitting a Data Use Certificate (DUC). For this project the DUC, accepted on 4/12/2025, can be found at https://adknowledgeportal.synapse.org/Data%20Access by searching for "tverlaan". In order to ensure the confidentiality of the data, it has not been redistributed through third-party applications or other means and after completion of the study it is wiped from the author's personal computer.

The Xenium spatial transcriptomic data is not publicly available, as it has not yet been published by the TU Delft. In order to use it all students had to sign a data usage agreement before the start of the project. This agreement again enforces that all data be wiped from local machines and no redistribution occur.

The DUC for ROSMAP also requests access to SEA-AD on Synapse's platform. However, the files on

the platform are in raw format. Luckily, a processed version of the data in h5ad format is available online at https://sea-ad-single-cell-profiling.s3.amazonaws.com/index.html#PFC/RNAseq/. The data is under the license of the Allen Institute (https://alleninstitute.org/terms-of-use/). As with the other datasets, any locally stored data is wiped and any redistribution is strictly avoided.

Another important ethical consideration is the presence of bias in the training data. ROSMAP participants are in part people from the catholic community, in part people from retirement homes [8]. This setting captures a large portion of the elderly population in the United States, which makes it a fairly reliable data source for classification models of Alzheimer's disease.

The same can be said for SEA-AD, which contains data from many different individuals across the US. However, in this paper only a few semi-randomly selected individuals are used for the experiments, due to the immense size of the data. This could lead to unintentional biases that hinder the integrity of the trained models. Despite this, such bias is unlikely to significantly affect the results of the experiments, as what matters more is the disease status of the patients, rather than their diversity.

On the other hand, the Xenium data used for classification comes from only one patient. This makes it largely unrepresentative of the general population. While this does not distort the results, it does mean that any conclusion on the hypothesis' validity is not directly applicable to the general population. However, it can still serve as the basis for further research.

For reproduction of the results, one can make use of the project code in Github (https://github.com/dsmenovski/Research-Project-2425). Setup requires downloading the list of dependencies that are given in Appendix A. Detailed implementation details are also present in the last methodology subsection. The project files include scripts which can be run locally or on a server with arguments for path to input and output files. When combined, the scripts make up the full experimental pipeline used for generating the results. The order in which to run them is written in the README on Github. Anyone who wishes to recreate the project from scratch can follow the approach outlined in the methodology. As mentioned in the pre-processing subsection, the results are generated using random seed '56' to ensure reproducibility. Links to the single-cell data are also provided in the README. As the spatial data is not publicly available, there is no link to it. To request access to it, contact the Pattern Recognition & Bioinformatics group at TU Delft. It is suggested to execute the imputation script on an HPC cluster, such as DAIC, as it is quite computationally intensive, especially for a high number of epochs. For users with different file systems who want to run the experiments on their machines, the scripts should also enable them to do so, as they require custom input/output paths to be entered for saving and loading the data. However, in order to use different datasets than the ones in this study, they will have to write their own scripts for label extraction following the example of the code in the repository, as the exact method differs slightly from dataset to dataset and cannot be generalized easily. Moreover, the names of the AnnData columns accessed in the code will have to be adapted if using different datasets.

Acknowledgments

Research reported in this work was partially or completely facilitated by computational resources and support of the Delft AI Cluster (DAIC) at TU Delft (RRID: SCR_025091), but remains the sole responsibility of the authors, not the DAIC team [28].

The results published here are in whole or in part based on data obtained from the AD Knowledge Portal (https://adknowledgeportal.org).

ROSMAP: Study data were generated from postmortem brain tissue provided by the Religious Orders Study and Rush Memory and Aging Project (ROSMAP) cohort at Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. This work was funded by NIH grants U01AG061356 (De Jager/Bennett), RF1AG057473 (De Jager/Bennett), and U01AG046152 (De Jager/Bennett) as part of the AMP-AD consortium, as well as NIH grants R01AG066831 (Menon) and U01AG072572 (De Jager/St George-Hyslop).

SEA-AD: Study data were generated from postmortem brain tissue obtained from the University of Washington BioRepository and Integrated Neuropathology (BRaIN) laboratory and Precision Neuropathology Core, which is supported by the NIH grants for the UW Alzheimer's Disease Research Center (P50AG005136 and P30AG066509) and the Adult Changes in Thought Study (U01AG006781 and U19AG066567). This study is supported by NIA grant U19AG060909.

Xenium: This work was supported by NIH common fund project number 1U54EY032442-01, NIH T32 DK101003, NIH project number 3OT2OD033759-01S1 subaward number 1090719-473495, and National Institute on Aging (NIA) R01AG078803.

A Python Library Requirements

```
anndata==0.10.9
anyio==4.9.0
argon2-cffi==23.1.0
argon2-cffi-bindings==21.2.0
array_api_compat==1.11.2
arrow==1.3.0
asttokens==3.0.0
async-lru==2.0.5
attrs==25.3.0
babel==2.17.0
beautifulsoup4==4.13.4
bleach==6.2.0
certifi==2025.4.26
cffi==1.17.1
charset-normalizer==3.4.2
colorama==0.4.6
comm==0.2.2
contourpy==1.3.0
cycler==0.12.1
debugpy==1.8.14
decorator==5.2.1
defusedxml==0.7.1
exceptiongroup==1.2.2
```

```
executing==2.2.0
fastjsonschema==2.21.1
filelock==3.18.0
fonttools==4.57.0
fqdn==1.5.1
fsspec==2025.3.2
get-annotations==0.1.2
h11==0.16.0
h5py==3.13.0
httpcore==1.0.9
httpx==0.28.1
idna==3.10
igraph==0.11.8
importlib_metadata==8.7.0
importlib_resources==6.5.2
ipykernel==6.29.5
ipython==8.18.1
isoduration==20.11.0
jedi==0.19.2
Jinja2==3.1.6
joblib==1.5.0
json5==0.12.0
jsonpointer==3.0.0
jsonschema==4.23.0
jsonschema-specifications==2025.4.1
jupyter-events==0.12.0
jupyter-lsp==2.2.5
jupyter_client==8.6.3
jupyter_core==5.7.2
jupyter_server==2.15.0
jupyter_server_terminals==0.5.3
jupyterlab==4.4.2
jupyterlab_pygments==0.3.0
jupyterlab_server==2.27.3
kiwisolver==1.4.7
legacy-api-wrap==1.4.1
leidenalg==0.10.2
llvmlite==0.43.0
MarkupSafe==3.0.2
matplotlib==3.9.4
matplotlib-inline==0.1.7
mistune==3.1.3
mpmath==1.3.0
natsort==8.4.0
nbclient==0.10.2
nbconvert==7.16.6
nbformat==5.10.4
nest-asyncio==1.6.0
networkx==3.2.1
notebook==7.4.2
notebook_shim==0.2.4
numba==0.60.0
numpy==1.24.4
overrides==7.7.0
packaging==25.0
pandas==2.2.3
pandocfilters==1.5.1
parso==0.8.4
patsy==1.0.1
pillow==11.2.1
platformdirs==4.3.8
prometheus_client==0.21.1
prompt_toolkit==3.0.51
psutil==7.0.0
pure_eval==0.2.3
pycparser==2.22
Pygments==2.19.1
pynndescent==0.5.13
pyparsing==3.2.3
python-dateutil==2.9.0.post0
python-json-logger==3.3.0
pytz==2025.2
pywin32==310
pywinpty==2.0.15
PyYAML==6.0.2
pyzmq==26.4.0
referencing==0.36.2
requests==2.32.3
rfc3339-validator==0.1.4
rfc3986-validator==0.1.1
rpds-py==0.24.0
scanpy==1.10.3
scikit-learn==1.6.1
scikit-misc==0.3.1
scipy==1.13.1
seaborn==0.13.2
Send2Trash==1.8.3
session_info==1.0.1
six==1.17.0
sniffio==1.3.1
soupsieve==2.7
stack-data==0.6.3
statsmodels==0.14.4
stdlib-list==0.11.1
sympy==1.14.0
tangram-sc==1.0.4
terminado==0.18.1
texttable==1.7.0
threadpoolctl==3.6.0
tinycss2==1.4.0
tomli==2.2.1
torch==2.7.0
tornado==6.4.2
tqdm==4.67.1
traitlets==5.14.3
types-python-dateutil==2.9.0.20241206
typing_extensions==4.13.2
tzdata==2025.2
umap-learn==0.5.7
uri-template==1.3.0
urllib3==2.4.0
wcwidth==0.2.13
webcolors==24.11.1
webencodings==0.5.1
websocket-client==1.8.0
zipp==3.21.0
```

References

[1] National Institute on Aging. "Alzheimer's disease fact sheet." Accessed: 2025-04-30, National Institutes of Health. (Apr. 2023), [Online]. Available: https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet.

[2] T. Verlaan, G. Bouland, A. Mahfouz, and M. Reinders, "scAGG: Sample-level embedding and classification of Alzheimer's disease from single-nucleus data," *bioRxiv*, p. 2025.01.28.635240, Jan. 2025. DOI: 10.1101/2025.01.28.635240. [Online]. Available: http:

//biorxiv.org/content/early/2025/01/30/2025.01.28.635240.abstract.

[3] C. G. Williams, H. J. Lee, T. Asatsuma, R. Vento-Tormo, and A. Haque, "An introduction to spatial transcriptomics for biomedical research," *Genome Medicine*, vol. 14, no. 1, p. 68, Jun. 27, 2022, ISSN: 1756-994X. DOI: 10.1186/s13073-022-01075-1. [Online]. Available: https://doi.org/10.1186/s13073-022-01075-1.

[4] T. Abdelaal, S. Mourragui, A. Mahfouz, and M. J. T. Reinders, "SpaGE: Spatial gene enhancement using scRNA-seq," *Nucleic Acids Research*, vol. 48, no. 18, e107–e107, Oct. 9, 2020, ISSN: 0305-1048. DOI: 10.1093/nar/gkaa740. [Online]. Available: https://doi.org/10.1093/nar/gkaa740 (visited on 05/20/2025).

[5] T. Biancalani, G. Scalia, L. Buffoni, *et al.*, "Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram," *Nature Methods*, vol. 18, no. 11, pp. 1352–1362, Nov. 1, 2021, ISSN: 1548-7105. DOI: 10.1038/s41592-021-01264-7. [Online]. Available: https://doi.org/10.1038/s41592-021-01264-7.

[6] W.-T. Chen, A. Lu, K. Craessaerts, *et al.*, "Spatial transcriptomics and in situ sequencing to study alzheimer's disease," *Cell*, vol. 182, no. 4, 976–991.e19, Aug. 20, 2020, ISSN: 0092-8674. DOI: 10.1016/j.cell.2020.06.038. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0092867420308151.

[7] P. Izquierdo, R. B. Jolivet, D. Attwell, and C. Madry, "Amyloid plaques and normal ageing have differential effects on microglial ca2+ activity in the mouse brain," *Pflügers Archiv - European Journal of Physiology*, vol. 476, no. 2, pp. 257–270, Feb. 1, 2024, ISSN: 1432-2013. DOI: 10.1007/s00424-023-02871-3. [Online]. Available: https://doi.org/10.1007/s00424-023-02871-3.

[8] D. A. Bennett, A. S. Buchman, P. A. Boyle, L. L. Barnes, R. S. Wilson, and J. A. Schneider, "Religious orders study and rush memory and aging project," *Journal of Alzheimer's Disease*, vol. 64, S161–S189, s1 Jun. 12, 2018, Publisher: SAGE Publications, ISSN: 1387-2877. DOI: 10.3233/JAD-179939. [Online]. Available: https://journals.sagepub.com/action/showAbstract (visited on 05/20/2025).

[9] S. Mostafavi, C. Gaiteri, S. E. Sullivan, *et al.*, "A molecular network of the aging human brain provides insights into the pathology and cognitive decline of alzheimer's disease," *Nature Neuroscience*, vol. 21, no. 6, pp. 811–819, Jun. 1, 2018, ISSN: 1546-1726. DOI: 10.1038/s41593-018-0154-9. [Online]. Available: https://doi.org/10.1038/s41593-018-0154-9.

[10] Allen Institute for Brain Science, University of Washington Alzheimer's Disease Research Center, and Kaiser Permanente Washington Health Research Institute, *Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) – 10x single nucleus RNAseq*, Dataset. AD Knowledge Portal, 2022. [Online]. Available: https://adknowledgeportal.synapse.org/Explore/Studies/DetailsPage/StudyDetails?Study=syn26223298.

[11] M. I. Gabitto, K. J. Travaglini, V. M. Rachleff, *et al.*, "Integrated multimodal cell atlas of alzheimer's disease," *Nature Neuroscience*, vol. 27, no. 12, pp. 2366–2383, Dec. 1, 2024, ISSN: 1546-1726. DOI: 10.1038/s41593-024-01774-5. [Online]. Available: https://doi.org/10.1038/s41593-024-01774-5.

[12] F. A. Wolf, P. Angerer, and F. J. Theis, "SCANPY: Large-scale single-cell gene expression data analysis," *Genome Biology*, vol. 19, no. 1, p. 15, Feb. 6, 2018, ISSN: 1474-760X. DOI: 10.1186/s13059-017-1382-0. [Online]. Available: https://doi.org/10.1186/s13059-017-1382-0.

[13] Scanpy development team. "Preprocessing and clustering – scanpy." Accessed: 2025-06-19. (Jun. 2025), [Online]. Available: https://scanpy.readthedocs.io/en/stable/tutorials/basics/clustering.html.

[14] V. A. Traag, L. Waltman, and N. J. van Eck, "From louvain to leiden: Guaranteeing well-connected communities," *Scientific Reports*, vol. 9, no. 1, p. 5233, Mar. 26, 2019, ISSN: 2045-2322. DOI: 10.1038/s41598-019-41695-z. [Online]. Available: https://doi.org/10.1038/s41598-019-41695-z.

[15] L. Heumos, A. C. Schaar, C. Lance, *et al.*, "Best practices for single-cell analysis across modalities," *Nature Reviews Genetics*, vol. 24, no. 8, pp. 550–572, Aug. 1, 2023, ISSN: 1471-0064. DOI: 10.1038/s41576-023-00586-w. [Online]. Available: https://doi.org/10.1038/s41576-023-00586-w.

[16] Q. Wang, K. Chen, Y. Su, E. M. Reiman, J. T. Dudley, and B. Readhead, "Deep learning-based brain transcriptomic signatures associated with the neuropathological and clinical severity of Alzheimer's disease," *Brain Communications*, vol. 4, no. 1, fcab293, Feb. 2022, ISSN: 2632-1297. DOI: 10.1093/braincomms/fcab293. [Online]. Available: https://doi.org/10.1093/braincomms/fcab293.

[17] H. Braak and E. Braak, "Neuropathological stageing of alzheimer-related changes," *Acta Neuropathologica*, vol. 82, no. 4, pp. 239–259, Sep. 1, 1991, ISSN: 1432-0533. DOI: 10.1007/BF00308809. [Online]. Available: https://doi.org/10.1007/BF00308809.

[18] H. Braak, I. Alafuzoff, T. Arzberger, H. Kretzschmar, and K. Del Tredici, "Staging of alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry," *Acta Neuropathologica*, vol. 112, no. 4, pp. 389–404, Oct. 1, 2006, ISSN: 1432-0533. DOI: 10.1007/s00401-006-0127-z. [Online]. Available: https://doi.org/10.1007/s00401-006-0127-z.

[19] S. S. Mirra, A. Heyman, D. McKeel, *et al.*, "The consortium to establish a registry for alzheimer's disease (CERAD)," *Neurology*, vol. 41, no. 4, pp. 479–479, Apr. 1, 1991, Publisher: Wolters Kluwer. DOI: 10.1212/WNL.41.4.479. [Online]. Available: https://doi.org/10.1212/WNL.41.4.479 (visited on 06/19/2025).

[20] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, Mar. 1, 1947. DOI: 10.1214/aoms/1177730491. [Online]. Available: https://doi.org/10.1214/aoms/1177730491.

[21] I. Virshup, S. Rybakov, F. J. Theis, P. Angerer, and F. A. Wolf, "Anndata: Access and store annotated data matrices," *Journal of Open Source Software*, vol. 9, no. 101, p. 4371, 2024. DOI: 10.21105/joss.04371. [Online]. Available: https://doi.org/10.21105/joss.04371.

[22] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, Jun. 2007, ISSN: 1558-366X. DOI: 10.1109/MCSE.2007.55.

[23] C. R. Harris, K. J. Millman, S. J. van der Walt, *et al.*, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 1, 2020, ISSN: 1476-4687. DOI: 10.1038/s41586-020-2649-2. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2.

[24] The pandas development team, *Pandas-dev/pandas: Pandas*, version v2.2.3, Sep. 2024. DOI: 10.5281/zenodo.13819579. [Online]. Available: https://doi.org/10.5281/zenodo.13819579.

[25] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, "SciPy 1.0: Fundamental algorithms for scientific computing in python," *Nature Methods*, vol. 17, no. 3, pp. 261–272, Mar. 1, 2020, ISSN: 1548-7105. DOI: 10.1038/s41592-019-0686-2. [Online]. Available: https://doi.org/10.1038/s41592-019-0686-2.

[26] M. L. Waskom, "Seaborn: Statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021. DOI: 10.21105/joss.03021. [Online]. Available: https://doi.org/10.21105/joss.03021.

[27] J. Ansel, E. Yang, H. He, *et al.*, "PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation," in *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)*, ACM, Apr. 2024. DOI: 10.1145/3620665.3640366. [Online]. Available: https://docs.pytorch.org/assets/pytorch2-2.pdf.

[28] Delft AI Cluster (DAIC), *The delft ai cluster (daic), rrid:scr_025091*, 2024. DOI: 10.4233/rrid:scr_025091. [Online]. Available: https://doc.daic.tudelft.nl/.