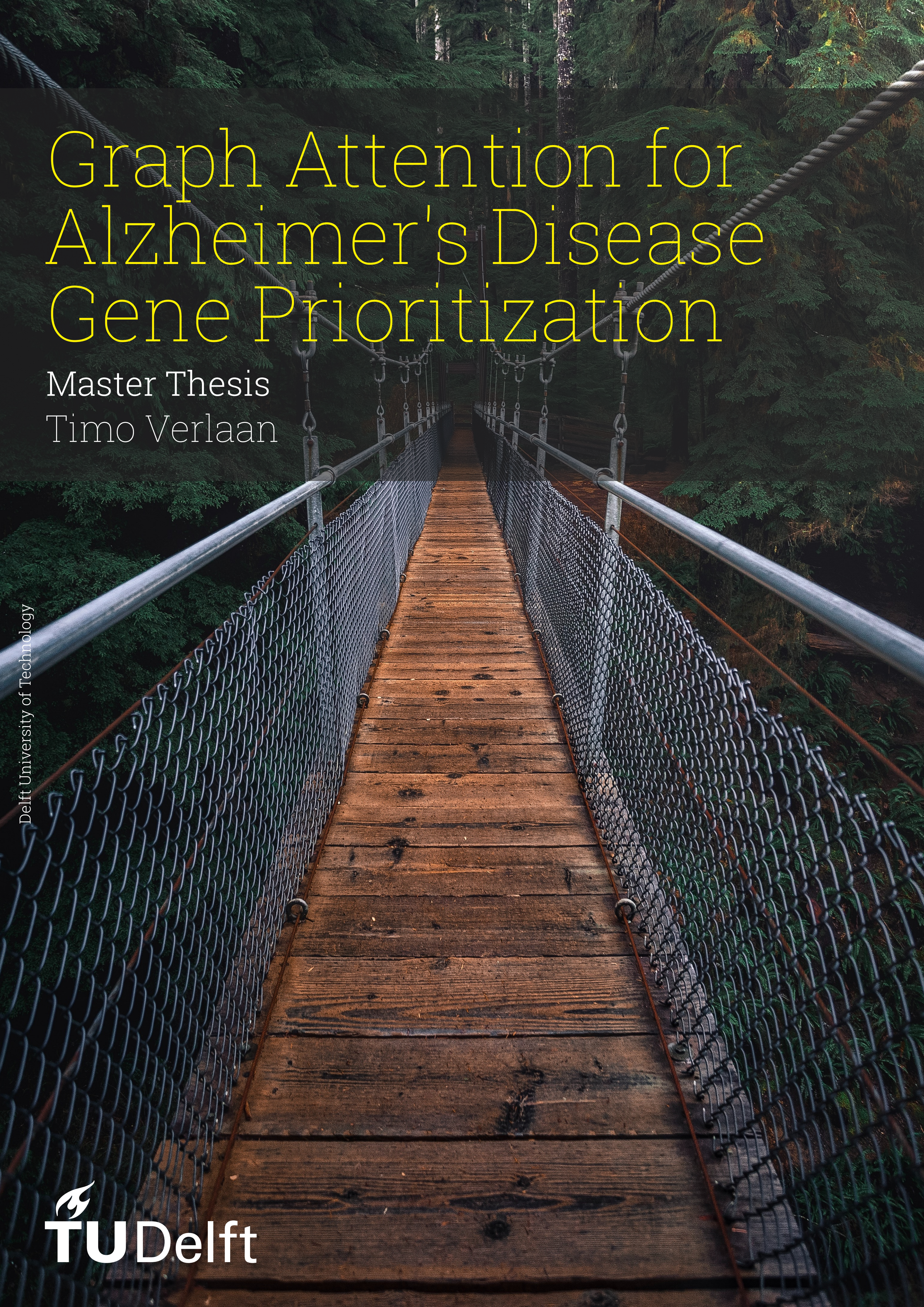


Graph Attention for Alzheimer's Disease Gene Prioritization

Master Thesis

Timo Verlaan

Delft University of Technology

Graph Attention for Alzheimer's Disease Gene Prioritization

by

Timo Verlaan

Student Name	Student Number
Timo Verlaan	4687108

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Monday July 10th, 2023 at 15:00.

Thesis Committee:	Prof. Dr. Ir. M. J. T. Reinders	TU Delft, Supervisor
	Dr. M. Khosla	TU Delft
	Ir. G. Bouland	TU Delft
Project Duration:	September, 2022 - July, 2023	
Faculty:	Faculty of Electrical Engineering, Mathematics & Computer Science, Delft	

TU Delft

Graph Attention for Alzheimer's Disease Gene Prioritization

Timo Verlaan

July 4, 2023

Abstract

Identifying key genes in Alzheimer's Disease (AD) is important in increasing understanding about its pathogenesis, and discovering potential therapeutic targets. Recent advances in single-cell RNA sequencing (scRNAseq) technology have provided unprecedented opportunities to study the molecular mechanisms underlying AD at the cellular level. In this study, we have trained a Graph Attention Network (GAT) to predict disease status of single cells from the SEA-AD scRNAseq dataset. Furthermore, we propose a method for interpreting the learned attention weights of the GAT to score genes on their importance for the prediction, treating gene prioritization as a feature importance problem. We have identified several genes associated with AD, including RBFOX1, NRG1, NRG3, GPC6, HNRNPC and CSMD1. We also found significant gene set enrichment in several terms related to AD and dementia, warranting future research about the presented top genes.

Introduction

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder that primarily affects the elderly, causing symptoms such as memory loss, cognitive decline, and behavioral changes. The disease progression is typically slow, spanning several years, and eventually leads to the loss of independent function, and ultimately death [1].

AD is increasingly becoming a public health concern due to the aging global population. Advances in modern medicine have increased life expectancy, resulting in a larger proportion of elderly individuals who are at risk of developing AD. Current estimates suggest that around 50 million people worldwide are living with dementia, with AD accounting for the majority of these cases. This number is expected to triple by 2050. The disease has a profound economic impact, with costs associated with healthcare and caregiving. The global worldwide cost of dementia, of which AD constitutes the majority, is estimated to be a trillion US dollars per year [2] [3].

The disease involves the accumulation of amyloid-beta plaques and tau tangles in the brain, leading to neuronal death [4], but many details of the pathogenesis of AD remain elusive. Specifically, discovery of the genes involved, and their interactions, could aid in understanding disease mechanisms and finding therapeutic targets [5]. To this end, genome-wide association studies (GWAS) are performed, where we look for single-nucleotide polymorphisms (SNPs) that are associated with the disease. It is through such studies, that we now know about variants of genes, including APOE4, APP, PSEN1, and PSEN2, that are associated with increased AD risk [6]. However, in many other cases it is difficult to match the loci found in GWAS directly to a gene because the majority are found in non-coding parts of the genome. Knowing the affected genes is paramount in studying the effects of the variant, and their role in the disease. Furthermore, the consideration of only the genome does not unveil the cascade of changes in expression and interaction between genes, that can provide more insight in the disease.

To this end, gene expression is measured through RNA sequencing (RNAseq), where expression levels of individual genes are measured. Recent advances in technology now even allow for the measurement of gene expression at the individual cell level, through single-cell RNA sequencing (scRNAseq). This was not possible in traditional "bulk" RNAseq, where expression needs to be averaged over many cells to get a sufficiently accurate estimate of expression level, thus masking potentially important variability between individual cells [7]. Capturing this heterogeneity in the cellular landscape enables researchers to study individual cell types and states, and their roles in disease [8] [9] [10].

Analysis of the vast amounts of data produced by scRNAseq experiments requires dealing with new challenges, including high dimensionality, high levels of noise, and sparsity [11]. In bulk RNAseq analysis, differential expression was applied to find genes that have altered expression levels in diseased cells [12] [13]. But in single-cell analyses, it lacks precision in capturing cell type-specific effects or lowly expressed genes that might still be relevant, and may yield false positives [14]. Even the calculation of differential expression between clusters of cells [15] [16] [17] is limited because over-clustering yields false positives and under-clustering causes us to miss genes. These limitations call for more advanced analysis methods that can capture highly heterogeneous and sometimes lowly expressed signals, while scaling well to the ever-increasing dimensionality of scRNAseq datasets.

An approach that is being applied increasingly in many fields, including bioinformatics, is to represent the data in graphs, rather than in a tabular form. This allows us to integrate information from neighboring samples, that are connected in the graph if they have a high similarity. In scRNAseq data specif-

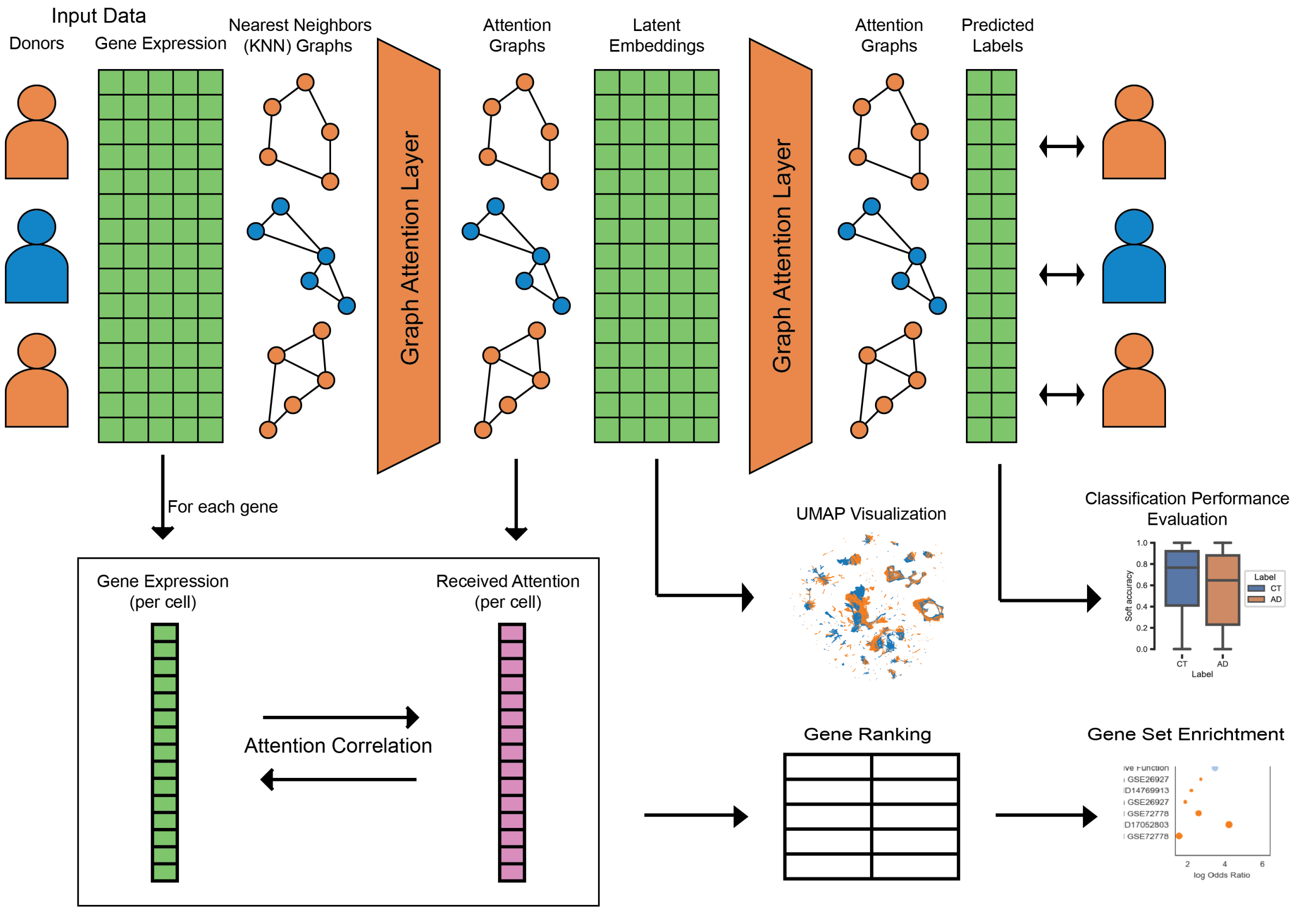


Figure 1: A visual schematic of the model architecture and proposed gene prioritization method, and performed analyses. A first Graph Attention Network (GAT) layer constructs embeddings for each node, that include learned representations of neighbors. These embeddings can be used directly, to visualize the latent space, or be passed through yet another GAT layer, to assign a label. Attention weights from the first GAT layer are used to rank genes using the proposed attention correlation technique.

ically, this means we can integrate the gene expression from neighboring cells, in addition to a cell's own expression levels. This helps estimate missing values, thus combatting the sparsity issue inherent to scRNAseq data, or more precisely estimate the expression level of a particular gene, thereby reducing noise. Furthermore, this allows us to use complementary signals from neighboring cells in the graph, enhancing the overall data interpretation. This is particularly relevant as cells with highly similar gene expression are likely to be physically close to each other in the brain. Considering these cells simultaneously might allow us to pick up on some communication between cells, which would be missed when considering cells in isolation.

Graph Neural Networks (GNNs) [18][19] are a class of deep learning models that generalize many of the techniques in the deep learning fields to graph data. Similarly to (convolutional) neural networks, layers with learnable weights are stacked, to find embeddings of cells that integrate information from neighboring cells, optimized for a specific task, like disease or cell-type classification. Additionally, GNNs are scalable, making them useful for processing vast amounts of graph data, and allowing them to discover more complex non-linear patterns, by increasing the number of model weights and layers. This property directly addresses the limitations of existing models, that have trouble discovering the highly heterogeneous expression profiles present in single cell data. A Graph Attention Network (GAT) [20] is a specific subtype, that assigns weights to neighboring nodes, based on how important they are in the downstream prediction task. These weights are referred to as attention scores, and have the added benefit of providing insight in the model's reasoning, that is often not available in similar models [21].

In this work, we harness the interpretability of these attention scores to find genes that are involved in Alzheimer's disease. We propose to train a GAT model on scRNAseq data, to predict disease status of each cell, which in this case corresponds to the clinical diagnosis of the donor the cell belonged to. We then consider the resulting attention scores, and propose a novel attention correlation technique, that allows us to derive the importance of each gene in the classification task. Because we train the model on a prediction task, rather than gene priorities directly, we do not require existing functional annotations of genes, or known disease associated genes, which many existing methods do [22][23][24]. Resulting highly ranking genes can be compared with genes known to be involved in AD, and previously undiscovered genes may warrant further research.

Neither the use of GNNs in the context of single-cell data, nor the interpretation of attention weights, are novel techniques. But to our knowledge, this is the first attempt to address the challenges in disease gene prioritization through interpretation of the attention weights, especially in the context of AD. As such, this work underscores the potential of integrating advanced deep learning techniques, like GATs, with cutting-edge scRNAseq biological data, to unravel the complexities of human diseases such as AD.

Results

In the following sections, we present the achieved results. Before we do so, however, we briefly discuss the dataset we used, and the proposed method, to provide context to the results. A thorough explanation of the method can be found in the final Method section. The results are first analysed in terms of disease classification accuracy, which reveals biological patterns that characterize changes in the transcriptomic landscape due to AD. Next, we present the genes identified by our proposed method, which are compared to knowledge from existing literature in the discussion. In the discussion, we also recommend directions for future work.

Dataset

In this study, we use the SEA-AD [25] dataset, which includes single cell RNAseq data of 84 brain donors, from various demographics and stages of AD, providing a comprehensive view of the disease and its progression. After pre-processing (see Methods), 49 donors remained, totaling $744,503$ cells. Of these remaining donors, 21 were diagnosed with AD, and 28 are age-matched controls. The AD and control donors comprised $426,761$ cells (57%) and $317,742$ (43%) respectively. Overall, the average cell count per donor is $15,194$, with a standard deviation of $2,587$. The dataset also provides celltype annotations, including 18 neuronal and 6 non-neuronal celltypes. Each donor was also assigned a made-up name, to allow easy referral to them in the results. Names start with A for donors from the AD class, and with C for the CT class.

We also consider the most important demographical, pathological, and genetic features of the donors in the dataset (Figure 2). First, we note how the classes of Alzheimer's disease (AD) and the control group (CT) are reasonably age-, sex- and race-matched, although the donors labeled as 'White' in the dataset are significantly over-represented. In terms of pathology, AD individuals are more represented in higher Thal- and Braak stages, which represent the progression of amyloid-beta plaques and neurofibrillary tangles (intracellular accumulations of tau protein) respectively. This is to be expected, as both are hallmark symptoms of AD [4]. Similarly, higher neuropathological change is found in AD individuals. The LATE-stage and primary location of found Lewy bodies do not show significant imbalance between disease classes, but in both, a significant portion of donors is not assigned a class altogether, limiting the interpretability of these variables. Finally, we consider the APOE4 status. Individuals with a positive APOE4 status (that is, they have at least one copy of the APOE E4 allele) are more represented in the AD class, which is to be expected because this is associated with an increased risk of getting AD.

Graph Attention Network (GAT) for Alzheimer's Disease Prediction

In this study, we used Graph Attention Networks (GATs) to predict the disease states of single cells. The input consists of a KNN graph per individual, where each node represents a single cell from that patient. Nodes are connected to their k-nearest neighbors based on the smallest Euclidean distance in the input feature space. In this study, we set $k = 30$.

GATs, a subtype of Graph Neural Networks (GNNs), incorporate an attention mechanism, which enables the model to assign varying degrees of importance to connected nodes in the graph. This property allows GATs to compute adaptive weights for neighboring nodes, in a way that can be transferred to unseen graphs. These attention weights determine the contribution of each neighbor's features when calculating the latent representation of a given node. These attention weights also enable downstream interpretation, which is valuable for our goal of identifying the most important genes or features associated with the disease.

In Figure 1, we show the proposed model architecture, with its two GAT layers. The first layer uses eight parallel attention heads, which allows different parts of the embeddings to use different attention weights for the neighbors. The concatenation of these parts from each of the attention heads gives us a single embedding per cell, that includes the learned representation of the target node and their direct neighbors, optimized for separating the input classes. The second layer is designed for the soft label assignment, and has two output features and a single attention head, such that it aligns with the two disease classes (CT and AD) that we are predicting. The model is trained using a categorical cross-entropy loss function, and its performance is evaluated through leave-one-out cross validation, where every donor is held out once, and evaluated on a model trained on the remainder of the dataset.

The method is described in more detail in the Method section.

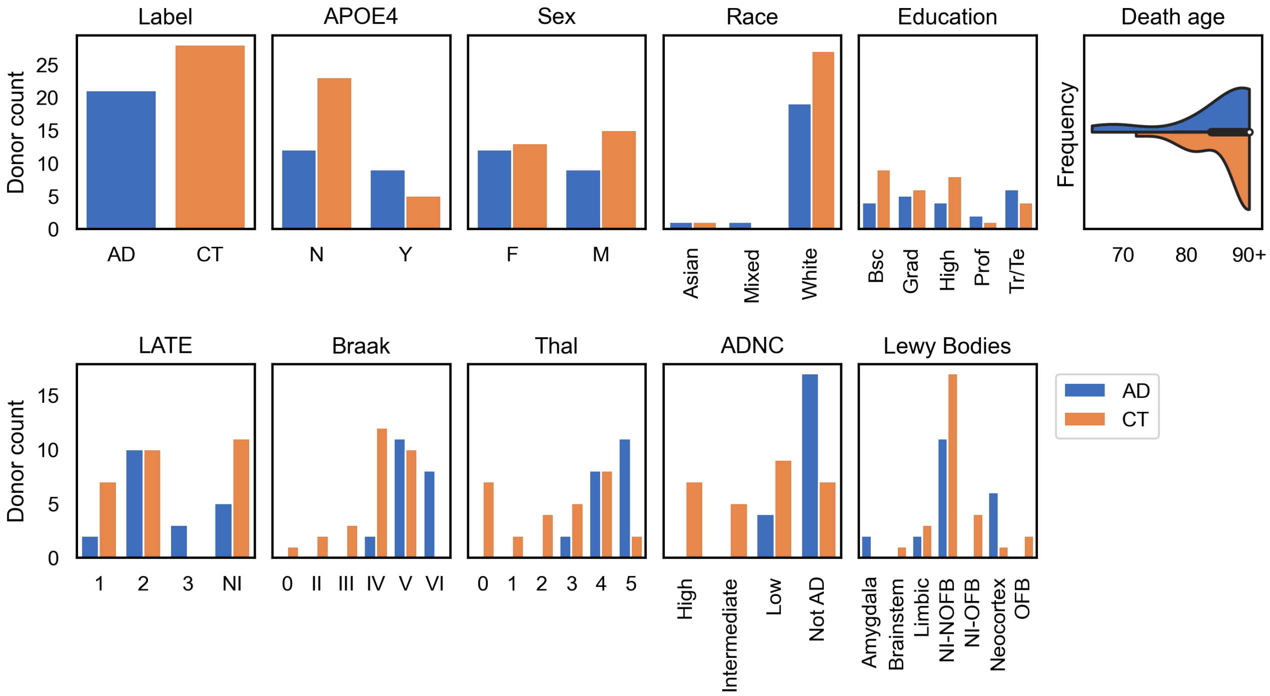


Figure 2: Demographics of the individuals in the SEA-AD dataset, summarizing the extent of diversity in the dataset. Each sub-figure compares the amount of diseased (AD) and control group (CT) individuals belonging to a certain class over eleven categories. The first four figures describe the core demographics of the dataset (disease class, sex, race and the highest level of education completed), followed by APOE4 status, which denotes whether an individual has at least one copy of the APOE E4 allele. In the top right, we show the distribution of ages at death between the classes. On the second row, several pathological features are presented, each characterizing progression of the disease, and overall neurological degeneration (ADNC). In the LATE stage and primary lewy bodies location, metadata is not available for some donors, which are marked “NI” (for not investigated).

Embedded space shows increased separability between disease classes

Using UMAP visualizations, we compare the input space and the embedded space that is obtained from the hidden dimension of our model (see Methods). We show that the disease classes are more separable in the latent space, and illustrate that the achieved separation is more likely to be due to biological meaningful patterns, rather than donor-specific differences.

The embeddings produced by our model separate healthy and diseased cells better than the input space (Figure 3). This is further exemplified by a LISI [26] score of 1.177 for our latent embeddings and a LISI score of 1.707 for the input space, which means that neighbors in the latent space are much more likely to be of the same class. Although there is no perfect separation between healthy and diseased cells, the increased separability does suggest that the GAT model has learned discriminating features between the disease states.

Furthermore, we find that the produced embeddings retain separability between cell types (Figure 3). This distinction between the celltypes is already present in the input space, but it’s important that this remains the case in the latent space because retention of the heterogeneity between celltypes and states is essential for downstream interpretability. In our embeddings, cell types are generally adjacent in clusters, indicating that the prediction task the embeddings are optimized for relies on biologically meaningful patterns because otherwise these would likely not have been so apparent in the latent space. We also note, however, that for some celltypes the clusters are more spread out, which is likely because the model recognized the heterogeneity within these cell types. This is somewhat apparent in each of the cell types, but especially so in the glutamatergic (excitatory) neurons, astrocytes and oligodendrocytes.

Moreover, we observe that within the clusters that correspond to the different celltypes, there is some separation between the different donors (Figure 3. This separation is not present at all in the input space, which is likely due to the batch correction that was applied to the original dataset [25]. We note that some degree of separation between donors in the latent space is necessary to achieve separation between their corresponding disease classes. It is also apparent that in many cell types, the donors are arranged on a spectrum that corresponds to the direction in which the disease class varies. This indicates our model recognizes various degrees of AD characteristics in these donors, which is consistent with the fact that AD patients can exhibit various degrees of neurodegenera-

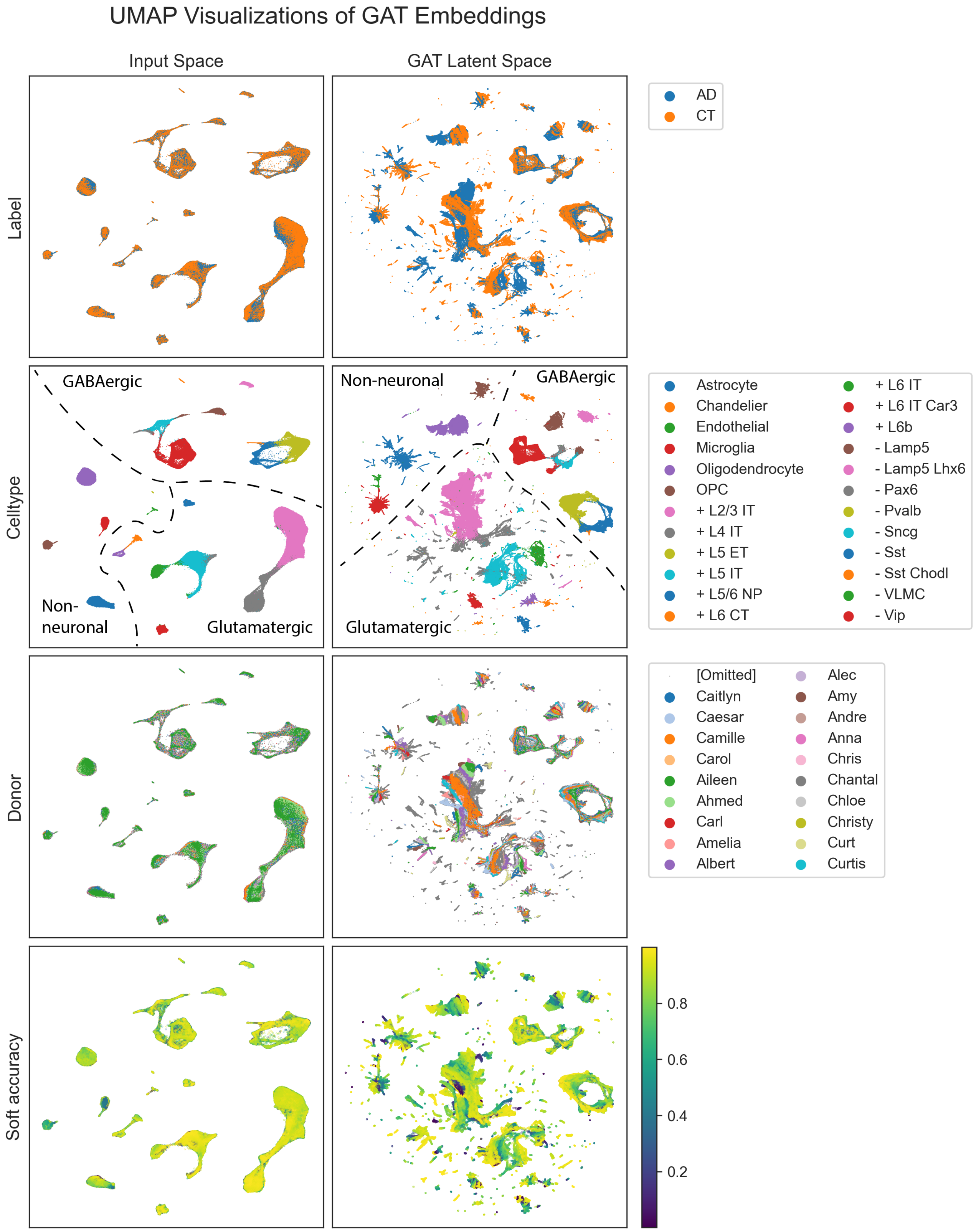


Figure 3: UMAP visualizations of the latent space (right) compared to the input space (left). In the top row, each cell is colored according to the disease state of the donor it belongs to, in which we distinguish between AD (Alzheimer's Disease) and CT (Control group). In the second row, cells are colored according to cell type, where glutamatergic (excitatory) neurons are marked "+", and GABAergic (inhibitory) neurons "-". The remaining cell types are non-neuronal. The third row colors cells according to the donor they belong to, where only 19 random donors are shown to improve comprehensibility. In the bottom row, cells are colored according to soft accuracy, which is defined as (1 − absolute error) and corresponds to how well each cell was classified.

tion.

Overall, the fact that the latent space is organized per cell type first, and by patient only within these structures, supports the notion that the achieved separation is more likely due to disease-related patterns, rather than donor- or batch-specific differences.

GAT Outperforms Baseline on AD Prediction Task

We further evaluate the model by comparing it with a baseline model. Specifically, we evaluate the ability to correctly classify cells from entirely unseen donors, based on the disease status of their donor (either CT or AD). The baseline is a two-layer neural network with an equally sized hidden dimension of 1024 (see Methods). Both models are evaluated using leave-one-out crossvalidation (LOOCV), where each donor is evaluated on a model trained on the data of all other donors.

The GAT model outperforms the baseline on all metrics (Table 1). Since they have an equally size hidden dimension, this is likely to be attributed to the GAT's attention mechanism, and ability to integrate information for neighboring cells.

Moreover, we note that (in both models) the recall and precision are very similar. This means that the difference in accuracy between the two disease classes can be attributed to the class imbalance. Misclassifications made in either class can be partially accounted for by considering the nature of the data. First, cells belonging to AD patients are not necessarily all affected by the disease, so some might be harder to classify as such. Conversely, because the dataset is age-matched, many individuals from the CT group will exhibit neuropathological degeneration due to aging, possibly making it hard to distinguish from degeneration that is specifically due to AD. This may account for some of the misclassification among the CT class. Finally, we note that cells are labelled according to their donor's clinical diagnosis, which does not account for diseased individuals that have not been diagnosed before they died. This limitation is explored further in the discussion.

Higher classification accuracy in specific cell types suggests increased sensitivity to AD

Analysis of classification performance compared between cell types indicates that in some cell types, the cells are more difficult to label according to disease state (Figure 4). In many of these cell types, this coincides with a low cell count for that type, which we consider a likely cause of the decreased performance. Some cell types, however, are frequently misclassified despite their sufficiently large representation. Cell types where we attribute poor classification performance to under-representation include the Sst Chodl, Pax6, and Chandelier GABAergic neuron types, as

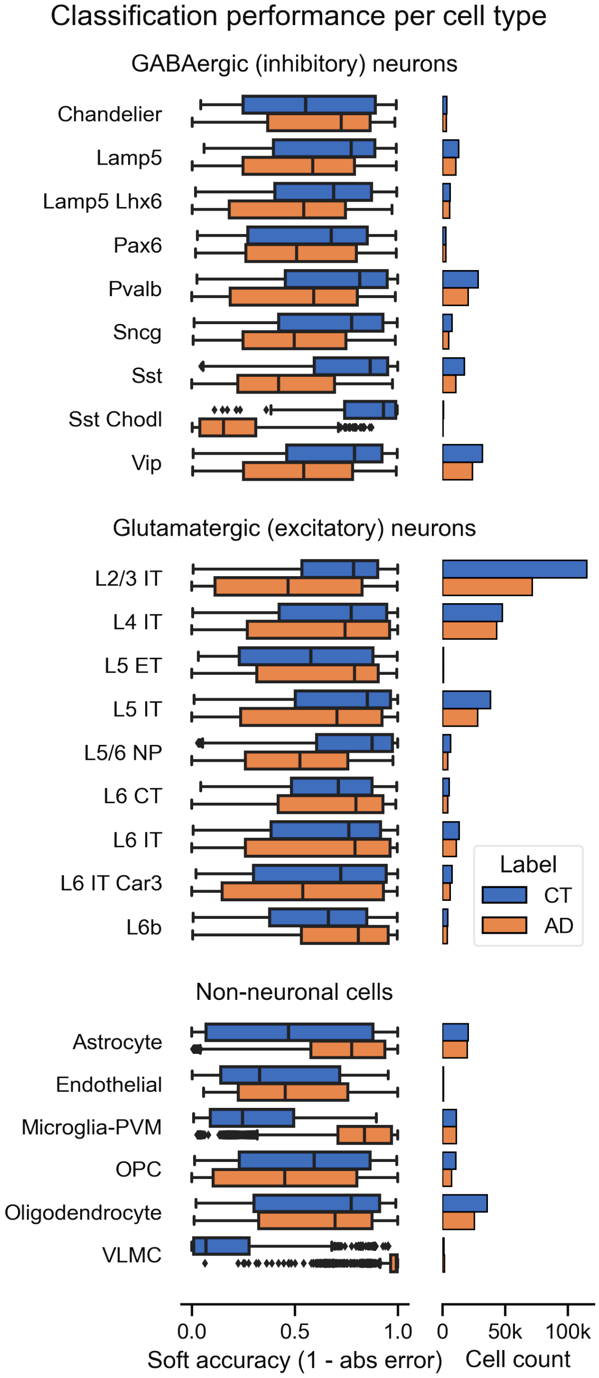


Figure 4: Classification performance compared per cell type. We compare train and test performance between AD and CT cells for each cell type. Cells are grouped by super type and the number of cells per type is shown. Performance is quantified using "soft accuracy", which is defined for each cell as (1 − absolute error) and corresponds to how well each cell was classified.

	LOOCV Accuracy			Precision	Recall	F1	AUC
	Full	CT	AD				
Baseline	0.619	0.690	0.523	0.558	0.523	0.540	0.607
GAT	**0.658**	**0.701**	**0.600**	**0.599**	**0.600**	**0.599**	**0.651**

Table 1: Comparative performance of the proposed GAT model, and the baseline linear neural network. Both models are evaluated through leave-one-out crossvalidation (see Methods). In all metrics, a higher score is better. Accuracy is calculated as the fraction of correctly predicted cells in a given set, without accounting for the class imbalance. It is also calculated for the set of CT and AD cells separately. Precision denotes the fraction of cells that were correctly identified as having AD out of all cells classified as AD, whereas recall signifies the fraction of cells accurately predicted as AD out of all actual AD cells. The F1 score is the harmonic mean of the precision and recall, which expresses the performance on an imbalanced dataset better than the accuracy. AUC denotes the area under the ROC curve, which can be found in supplementary figure S1.

well as the L5 ET glutamatergic neurons. From the non-neuronal cell types, we expect the lower performance for VLMC and endothelial cells to be due to under-representation.

Furthermore, we note how in most cell types AD cells are misclassified more often. This typically also aligns with the class imbalance with that cell type, with the exception of the microglia, which are significantly more often classified as AD. This is remarkable, because we found earlier that microglia are quite visibly separated in the latent space (Figure 3). A possible cause could be that even microglial cells in healthy individuals exhibit gene expression profiles similar to those of AD cells. Overall, it should be noted though, that the asymmetry in misclassifications does not necessarily mean anything more than that those cells cannot be distinguished, and an arbitrary decision boundary has to be drawn somewhere.

On the other hand, for cell types that are classified correctly more often overall, we can argue that they are therefore more easily distinguishable. Such is the case for astrocytes and oligodendrocytes, which are both classified more accurately than other non-neuronal cells. Conversely, L2/3 IT glutamatergic (excitatory) neurons, which are from the outer-most layer present in the dataset, are misclassified relatively often if we take into account how well they are represented in the dataset. This indicates that inner-most layers of the brain are affected more distinguishably in AD, compared to the outer layers. Among the GABAergic (inhibitory) neurons, there is very little difference in classification performance.

Finally, we note that overall, glutamatergic neurons (acc=0.68) are classified moderately better than GABAergic neurons (acc=0.65). Although the difference is small, this could indicate glutamatergic (excitatory) neurons are affected more in the disease.

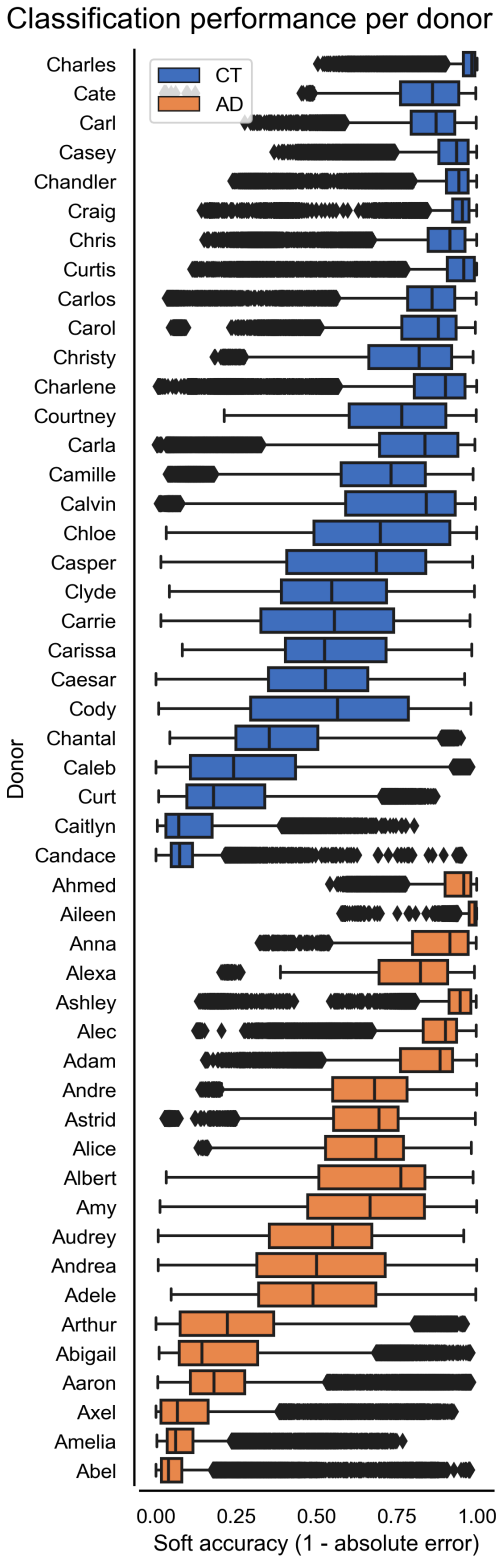
Strong association between neuropathological degeneration and classification performance

We also consider the classification accuracy per individual donor, from the LOOCV iteration where they were held out (Figure 5). Primarily, we note how the performance differs quite drastically between donors within both classes, raising the question of what is the cause behind this. It is not likely this variation is caused by differences in cell counts between donors because they do not vary as much as per cell type. All donor cell counts are between $10,000$ and $20,000$, which is a direct result of our input processing.

If we consider the metadata available, we can account for much of the variability in accuracy between individuals. First, we consider the age at death and how it compares to the donors accuracy. In both classes, the majority of donors under the age of 90 is classified below-average (60% in AD, and 67% in CT). Similarly, if we compare the relative performance between donors in each of the Braak and Thal stages, we find that AD donors in a higher stage are classified better (Figure 6). This indicates that classification is based on features related to the extent of neurodegeneration. In CT donors, however, this relationship is less apparent. Furthermore, we see an even stronger correlation in the CERAD [27] and CAA [28] scores and the classification accuracy of AD donors (Figure 6), where AD donors with a CAA score of "Mild", are classified wrong almost exclusively. There is also a slight relationship apparent between misclassification of CT donors and higher, CERAD/CAA scores, albeit less pronounced.

Finally, we consider the misclassification of CT donors in particular. Although some of them can be attributed to their relatively high CERAD/CAA scores, the performance throughout the difference stages of degeneration is more-or-less stable. If we consider the actually misclassified donors, though, of which there are only five (named Chantal, Caleb, Curt, Caitlyn and Candace), we see that two of them have a high overall ADNC, and two others medium. The remaining misclassified CT donor (Caitlyn) still has a Braak stage of IV, which is well above average for the CT class.

Overall, we consider the relation between degeneration and classification accuracy to be strong, where AD donors with low degeneration are likely to be confused for healthy individuals that may have a similar extent of neurodegeneration due to aging.



Acc	Sex	Death	LATE	Braak	Thal	ADNC	APOE4
1.000	M	90+	2	IV	4	Med.	N
0.996	F	90+	1	IV	3	Med.	N
0.993	M	90+	2	V	4	High	N
0.992	M	81	1	V	0	Not AD	N
0.970	M	90+	NI	II	0	Not AD	N
0.967	M	89	1	III	3	Med.	Y
0.964	M	90+	2	IV	2	Low	N
0.957	M	90+	2	V	2	Med.	N
0.957	M	90+	1	V	5	High	N
0.946	F	82	NI	IV	3	Med.	N
0.940	F	90+	2	V	4	High	N
0.923	F	90+	2	IV	2	Low	N
0.893	F	90+	NI	IV	0	Not AD	N
0.831	F	90+	NI	III	1	Low	N
0.804	F	90+	2	IV	0	Not AD	N
0.781	M	90+	2	V	4	Med.	N
0.738	F	90+	2	V	5	High	N
0.682	M	78	NI	0	0	Not AD	N
0.569	M	90+	1	IV	0	Not AD	N
0.554	F	87	NI	III	4	Med.	Y
0.549	F	90+	NI	V	4	High	Y
0.538	M	82	2	IV	2	Low	N
0.536	M	72	NI	II	1	Low	N
0.255	F	90+	NI	V	4	High	Y
0.198	M	90+	NI	IV	3	Med.	N
0.123	M	83	1	V	4	High	Y
0.025	F	80	NI	IV	0	Not AD	N
0.006	F	90+	1	IV	3	Med.	N
1.000	M	81	2	VI	5	High	Y
1.000	F	81	2	VI	5	High	Y
0.995	F	90+	NI	V	4	High	N
0.976	F	90+	2	V	4	High	N
0.974	F	65	NI	VI	5	High	Y
0.960	M	90+	2	VI	5	High	N
0.953	M	69		VI	4	High	N
0.838	M	90+	3	V	3	Med.	N
0.791	F	90+	NI	VI	5	High	N
0.782	F	90+	2	V	4	High	N
0.753	M	90+	NI	IV	4	Med.	N
0.721	F	70	2	VI	5	High	Y
0.590	F	90+	2	V	5	High	Y
0.500	F	86	2	V	4	High	N
0.480	F	90+	NI	IV	4	Med.	N
0.140	M	90+	1	V	5	High	Y
0.102	F	90+	3	V	5	High	Y
0.086	M	86	1	V	5	High	Y
0.053	M	84	2	V	5	High	Y
0.035	F	87	2	VI	4	High	N
0.029	M	88	3	V	3	Med.	N

Figure 5: Classification performance per individual donor, compared between AD and CT donors. For the boxplots, performance is quantified using "soft accuracy", which is defined for each cell as $(1 - \text{absolute error})$. In the table on the right, we highlight several metadata columns for each of the donors, where accuracy is the rate of correctly classified cells, "Death" refers to the age at death, LATE, Thal and Braak stages characterize progression of neuropathological degeneration, and ADNC denotes the overall AD neuropathological change. The right-most APOE4 column denotes whether the donors has at least one APOE E4 allele.

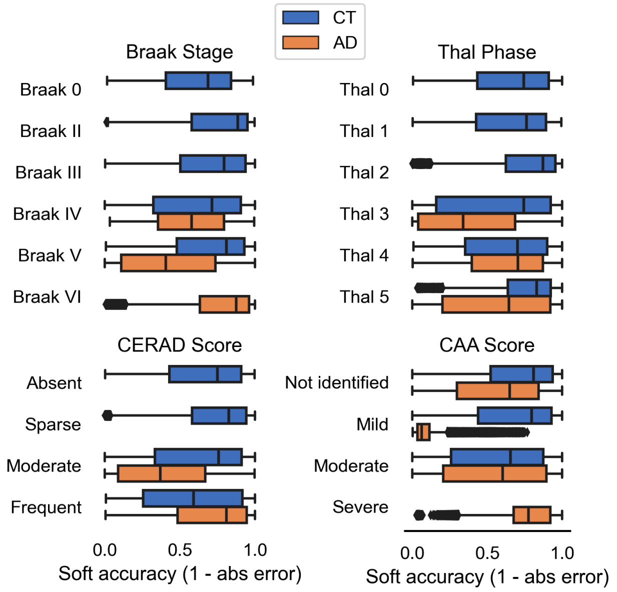


Figure 6: Classification performance is compared between different stages of neuropathology. Each of the Thal and Braak stages, as well as the CERAD and CAA scores, characterize progression of neuropathological degeneration and AD cognitive symptoms.

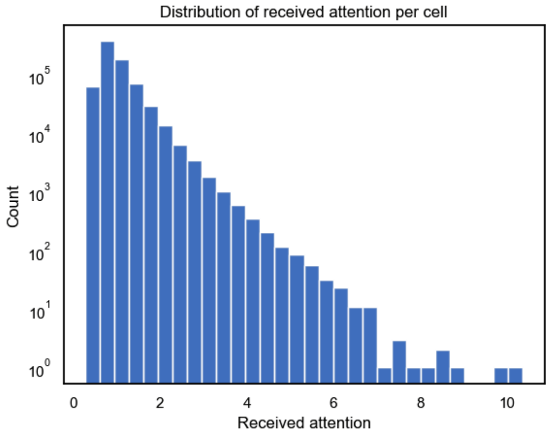


Figure 7: The distribution of total received attention per cell, that is calculated for each cell by summing the attention paid by each neighbor (see Methods).

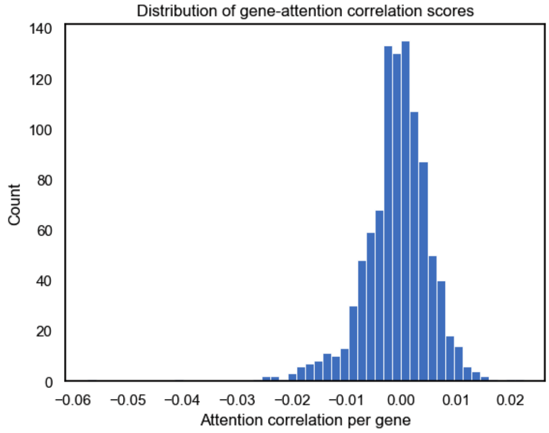


Figure 8: Distribution of attention correlation scores that are found for each of the genes ($\mu = 0$, $\sigma = 0.006$).

Attention correlation allows for prioritization of genes

Through interpretation of the attention scores that are assigned by the GAT model to each neighbor of each cell, genes can be ranked according to importance in disease classification (see Methods). Aggregation of summed received attention per cell yields a single score per cell, which is highly variable between cells and can be interpreted as their importance according to their neighbors. Because attention scores are softmaxed in the GAT model, the average received attention is 1. Notably, some cells receive up to 10x more attention than this average (Figure 7).

We use the received attention to rank genes, according to their importance in the classification of AD. To achieve this, we rank genes according to the correlation between their expression in each cell, and this cell's total received attention. Genes that have a high correlation with received attention are likely to have played a role during the classification of the cell. This indicates that the gene may be related to Alzheimer's disease. Note that this analysis is not limited to the genes the GAT was trained on, because correlation can be calculated with any gene in the original SEA-AD dataset.

Analysis of the distribution of gene-attention correlation scores shows that they follow a normal distribution (Figure 8). For several of the most accurately classified cell types, we present the top 20 genes (Table 2). These include several genes that have been implicated in Alzheimer's disease (NRG1 [29], NRG3 [30], CSMD1 [31], GPC6 [32], RBFOX1 [33], HNRNPC [34]).

Moreover, we performed gene set enrichment analysis (GSEA) on the found genes, to find disease and pathway gene sets in which the reported genes are overrepresented (Figure 9). The terms that are most enriched are not directly AD terms, but many can be related to AD, which further supports the relevance of the genes we found. This includes the volumes of the lingual gyrus and cerebellum ($p = 0.011$ and $p = 0.012$ respectively) which are affected in neurodegeneration [35][36][37][38] and Crohn's disease ($p = 0.0008$) which is associated with a more than doubling of AD risk [39]. The remaining terms are studied in the discussion. Finally, we consider the enrichment of several terms directly related to AD or dementia (Figure 9). For example, we found significant enrichment in terms corresponding to: AD with visuospatial impairment ($p = 0.022$), dementia in Parkinson's disease ($p = 0.043$), and AD in APOE e4+ carriers ($p = 0.0003$). Altogether, the significant enrichment in the terms discussed, further underline the importance of the found genes in Alzheimer's disease.

	Oligodendrocyte		Astrocyte		Microglia		Excitatory Neurons		Inhibitory Neurons	
Rank	**Gene**	**Corr**	**Gene**	**Corr**	**Gene**	**Corr**	**Gene**	**Corr**	**Gene**	**Corr**
1	MIR646HG	-0,099	MALAT1	0,103	PCDH9	-0,129	IDS	0,064	ATP1B1	0,067
2	GPC5	-0,092	KCNIP4	-0,072	KCNIP4	-0,120	YWHAZ	0,060	IDS	0,065
3	SDK1	-0,088	MIR646HG	-0,072	LRP1B	-0,110	MIR646HG	-0,058	MIR646HG	-0,065
4	KCNIP4	-0,087	MYO6	0,066	CSMD1	-0,108	AUXG01000058,1	0,057	CLASP2	0,064
5	RBFOX1	-0,085	HNRNPC	0,065	NRG3	-0,108	GLS	0,057	SERINC1	0,063
6	SGCZ	-0,079	WDFY3	0,064	KCTD12	0,107	YWHAG	0,056	PPP6R3	0,063
7	KIAA1217	-0,076	DDR2	0,064	RBFOX1	-0,106	ACSL3	0,056	GPR155	0,062
8	CSMD1	-0,074	LRRFIP2	0,063	SYT1	-0,104	PLEKHA1	0,055	GSK3B	0,061
9	SYT1	-0,073	RALGAPA1	0,063	CADM2	-0,102	NPTN	0,055	MARCH6	0,061
10	NRG1	-0,073	PDCD6IP	0,062	CNTNAP2	-0,100	PRNP	0,055	OIP5-AS1	0,060
11	NRG3	-0,07	SOX6	0,062	ADGRB3	-0,098	ARHGEF12	0,055	DYRK1A	0,059
12	DLGAP1	-0,068	DDX17	0,062	PABPC1	0,097	PRKACB	0,054	MAP4K3	0,058
13	BAZ2B	0,068	CSMD1	-0,061	GPC5	-0,095	PJA2	0,054	ANKRD17	0,058
14	AC109466,1	-0,067	LIFR	0,061	PTPRD	-0,093	GSK3B	0,054	RBFOX2	0,058
15	GPC6	-0,066	DCAF6	0,061	OPCML	-0,092	KIF3A	0,054	NCOA1	0,058
16	CLSTN2	-0,066	KIF21A	0,061	COP1	0,092	ZDHHC21	0,054	GLS	0,058
17	NRXN1	-0,066	SFPQ	0,061	SGCZ	-0,089	WASF1	0,053	NCKAP1	0,057
18	PCDH11X	-0,066	PAPOLA	0,061	DLGAP1	-0,089	ANKRD17	0,053	PNMA2	0,056
19	AL589740,1	-0,065	ZDHHC21	0,06	WAC	0,088	NCKAP1	0,053	FRRS1L	0,056
20	DPYD	0,064	LINC01578	0,059	BIRC6	0,088	SERINC1	0,053	YWHAG	0,056

Table 2: Top 20 genes produced through the proposed attention correlation technique. Analyses were performed for the three most well classified cell types, as well as all excitatory and inhibitory neurons. For each gene, the attention correlation score is presented. Corrected p-values for all genes are 0, as derived from 10^7 iterations of permutation testing (see Methods).

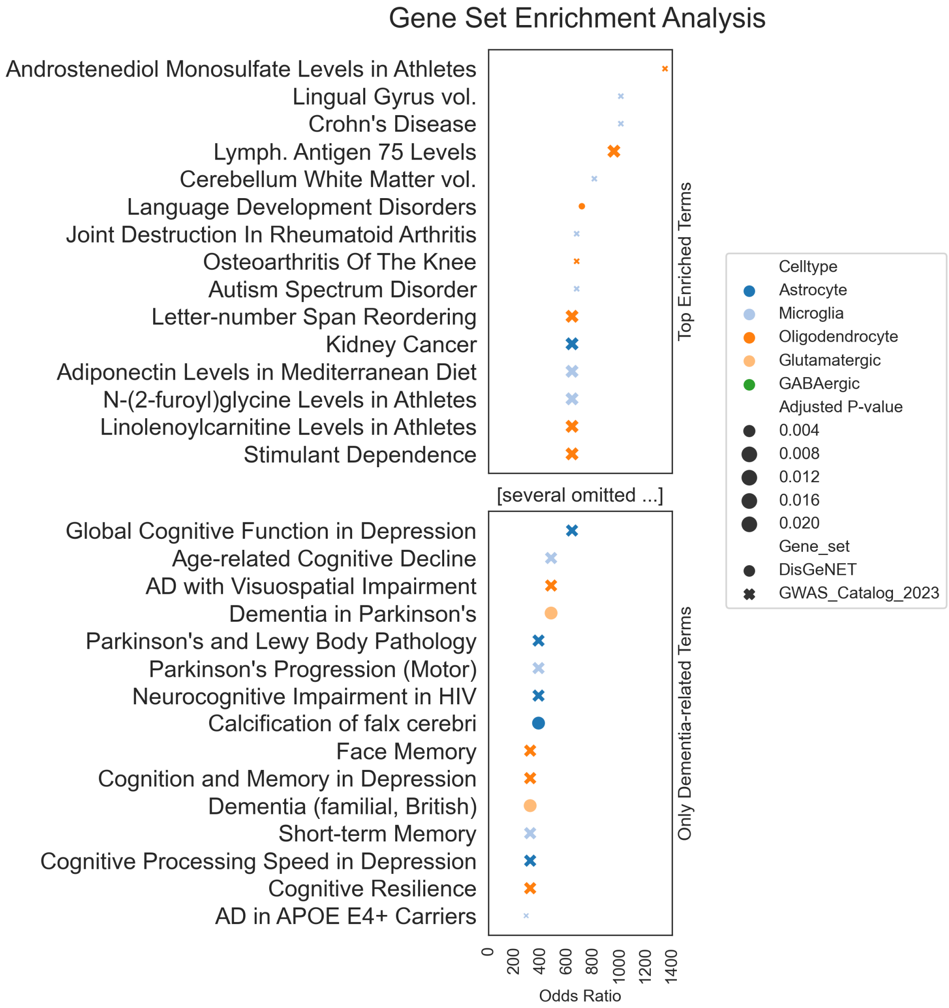


Figure 9: Gene set enrichment results of the top 20 genes that were found for each cell type. The overall top 15 terms are shown (top), as well as the 15 highest enriched terms that are explicitly related to AD (bottom).

Discussion

We introduced a graph attention network (GAT) model that can predict disease status of single cells from Alzheimer's disease (AD) and control group (CT) donors, represented in graphs. This model significantly outperforms the baseline model, and has the added benefit of providing attention scores, that are assigned to neighboring cells in the graphs, which denote their importance in the classification process and provide insight in the decision-making process that is not as apparent in alternative models. We then used these attention scores to characterize the gene expression of cells that were most important during classification of their neighbors, allowing us to rank genes based on their connection with Alzheimer's disease. We extensively verified cell classification results to show that the learned embeddings used to make the predictions are biologically meaningful and not batch- or patient-specific. Finally, we evaluated the performance of the proposed gene prioritization technique by performing enrichment analysis, and considering the differential expression of the top-ranked genes. In the remainder of this section, we further discuss some results presented in the previous chapter, the limitations of the proposed work and performed analyses, and how they might be overcome in future work.

Embedded Space

We have shown an increased separability between the disease classes in the embedded space, compared to the input space. This was achieved though dimensionality reduction using UMAP, and coloring each cell according to their label, and visually confirming a decrease in overlap between classes. Coloring cells according to cell type or the donor it belonged to, shows that there is still a clear separation between cell types, whereas donors still overlap. This supports the idea that it is likely increased class separation is achieved through use of biologically meaningful data, rather than batch- or patient-specific effects. Additionally, we quantified the separability using the LISI score [26], which provides a less subjective way to compare the separability in both spaces.

We do have to note, though, that UMAP visualizations might not give a fully accurate view of the class separability. Because dimensions are drastically reduced in a non-linear fashion, much of the variability in the data is lost, so a single UMAP visualization can never fully characterize the entire space spanned by the learned representations. They do, however, show that there is some extent of separation, that was not visible in the input space at all. Additional analyses of the latent space, for example by exploring the latent embeddings produced by each of the separate attention heads, may provide additional insight.

Classifyable Cell Types

We compared classification performance between different cell types. Poor classifications of several cell types were mostly likely due to low cell counts, which means there were too few cells to capture their underlying transcriptional profiles and how they might be affected in AD. Of the remaining cell types, some had substantially higher classification accuracy than others.

In particular, we reported astrocytes and oligodendrocytes to be classified accurately most consistently. If we confer with existing literature, we find ample evidence incriminating both in Alzheimer's disease. Astrocytes are known to become activated in an immunoresponse against β-amyloid (Aβ) [40], a major component of amyloid plaques, and they are linked with neuroinflammation in AD [41] [42]. Similarly, vulnerability of oligodendrocytes is known to lead to myelin breakdown, further progressing the disease [43] [44] [45].

Furthermore, we noted that overall classification performance among glutamatergic (excitatory) neurons was better. This can partially be attributed to their higher cell counts, as discussed in the results. If we compare this result with existing literature, however, excitatory neurons are indeed known to be affected strongly in AD. The over-abundance of Aβ in AD brains is believed to dysregulate glutamate receptors, causing overactivation of excitatory neurotransmission [46], which causes many of the clinical manifestations of AD such as memory loss [47].

We also noted that classification performance seems to be better in the inner-most cortical layers. This is could be because both neurogeneration due to aging, as well as due to AD progress from the outside in [48]. However, in natural aging this is not likely to progress all the way to the inner-most layers of the cortex (V-VI), which is also reflected in the Braak stages of our donors: None of the CT donors have Braak stage VI, where pTau has progressed to the 6th and inner-most layer [49] [50]. As such, the inner layers of the cortex in AD patients are more likely to show more degeneration than in age-matched controls, and therefore they are easier to classify accurately.

Finally, we note that a limiting factor in truly characterizing the altered transcriptional profiles in different celltypes is the low cell count for many cells. Certain cell types are significantly underrepresented. Consequentially, the learned representation likely prioritizes more frequent cell types, which have a higher payoff in the loss function. This can partially be aided by stratifying per cell type, or using weighed sampling during training, but preferably more data would be used, possibly from other studies.

Variable Classification Performance Between Donors

We considered the classification performance per individual donor, and found that most of the variability can be accounted for by considering the metadata. Specifically, the extent of neurodegeneration, which is characterized by the Braak stage among others, is found to affect the classification accuracy a lot. AD donors in the highest Braak stages are classified accurately, as are CT donors in the lowest Braak stages. Donors of both classes in intermediate Braak stages, however, are seemingly harder to classify reliable. This is likely due to similarities between the donors that make it hard to distinguish between them. This also supports the claim made earlier, that the model relies on biologically meaningful patterns to make predictions. We also found that donors under the age of 90 are harder to classify than donors of age 90 and above. This may be because degeneration has not progressed as much in these relatively younger donors, and their pathology might be more similary to changes due to general aging.

Furthermore, we should consider the labels used when training and evaluating the model. Each cell was labelled according to the clinical diagnosis of the donor they belonged to. Alzheimer's disease is diagnosed by physicians through several tests, including neuroimaging and cognitive exams that evaluate cognitive function. It is, however, not impossible that some donors have been wrongfully diagnosed with AD. Especially since AD is known to have a high number of comorbidities, such as Parkinson and other types of dementia, which can have very similar symptoms [51][52][53][54]. Control group donors, on the other hand, are also not guaranteed to not have AD. The fact that they do not have a clinical diagnosis only means that they did not have severe enough symptoms to go to a physician. However, it is known that Alzheimer's disease can progress quite far before symptoms arise, due to neuroplasticity [55] [56], so it's not impossible some CT donors actually have Alzheimer's disease.

Reported Genes

Through the proposed attention correlation method, we found genes that play an important role in the classification of AD in single cell data. We performed this analysis for several different celltypes, and for each a ranking of top genes was presented. Many of the found genes are associated with AD in literature, but it is beyond the scope of this work to investigate the role in AD for each of the presented genes. This requires a dedicated literature study where each of the genes found is thoroughly investivated. To evaluate the correctness and relevance of the genes we presented however, we provide a brief overview of some of the top genes in found in the oligodendrocytes and astrocytes, several of which were also found in the remaining cell types.

Starting with the genes found in oligodendrocyte cells, we note that GPC6 is involved in the recruitment of glutamate receptors, and together with RBFOX1, it is associated with Aβ load in the brain [32]. RBFOX1 is also associated with global cognitive decline during life [33]. SDK1 is related to amygdala cell death [57], which plays a big role in AD. Neuregulin (NRG) 1 is also associated with formation of plaques [29] and is in particular believed to be responsible for schizophrenia-like symptoms in AD with psychosis [58] [59]. Similarly, NRG3 is associated with schizophrenia symptoms [60], as well as an increased risk of getting AD, as well as the age at onset [30].

For the astrocytes, the top ranked gene was MALAT1, which is known to inhibit neuroinflammation, and plays an important role in protection against Aβ-induced toxicity [61] [62] [63]. The second highest ranked gene is KCNIP4, which interacts with presenilins [64] [65], that are an important cause of excess Aβ formation [66]. We also found HNRNPC, which is known to promote APP transcription [34], which plays a key role in AD pathogenesis, due to its cleavage product Aβ [67]. Consequentially, HNRNPC indirectly incluences the amount of Aβ that builds up in the brain. The CSMD1 gene is in the top 20 genes of astrocytes, as well as oligodendrocytes and microglia. Mutations in this gene result in inflammation in the central nervous system [31] that leads to familial Parkinson's disease. Several variants are also associated with an increased predisposition to getting AD [68] [69] [70] [71].

Enriched Pathways

To further verify the found gene sets in a more hollistic fashion, we also performed gene set enrichment analysis, where we found significant enrichment in many AD-related gene sets. This further shows that the set of genes we found are significantly related to AD. We also found several gene sets, that are not explicitly AD sets, with even higher odds ratios and p-values. For many of them, however, there are clear connections to be found in existing literature.

Two of the top terms are lingual gyrus, and cerebellum white matter volume, both of which are known to be affected in AD [35][36][37][38]. We also reported enrichment in Crohn's disease, which is known to more than double AD risk, and accelerate onset by around 7 years compared to individuals without the disease [39][72][73]. Furthermore, we found several terms that are known to be highly related to aging, including rheumatoid arthritis [74][75], osteoarthritis [76], and kidney cancer [77]. This may indicate that some of the aging-related processes that lie at the core of AD, are shared with these diseases. Moreover, we found enrichment in autism spectrum disorder, which is also known to have significant overlap with AD in genetics and disease mechanics [78] [79].

We also found three terms related to affected metabolite levels in professional athletes. The top term of the three concerns the level of androstenedione, which is found to be involved with protective mechanisms that mitigate development of AD [80], and is shown to have increased levels in subjects with AD [81]. Additionally, studies have shown that former elite athletes in general, have poorer brain health and increased AD risk [82] [83].

Finally, we consider the term labeled "Adiponectin Levels in Mediterranean Diet". This is remarkable because, not only is adiponectin, a hormone involved in several metabolic processes, is known to be involved in AD [84] [85], but mediterranean diets are also shown to decrease AD risk [86] [87], improve memory, and lead to less amyloid and pTau pathology [88]. This could indicate that the decreased AD risk due to mediterranean diet adherence is due to its effect on adiponectin levels. Further research in these, and aforementioned pathways, can contribute to a better understanding of AD.

Recommendations

Aside from the performed enrichment analyses, and comparing the found genes with existing literature, the options for evaluating the results are limited. This is because there is no ground truth set of genes known to be associated with AD, that we can easily compare it with, as is the case in some other diseases. One option that we have not explored in this work is attempting to retrain the classifier on the most important genes found through attention correlation. If prediction performance remains similar, this indicates these genes were indeed important in the prediction. An alternative way of validating the feature selection capabilities of attention correlation, rather than the found genes directly, could be to test it on a synthetic dataset, where we control how important each of the input features are. This way, the method can also be compared to other feature importance techniques like SHAP or permutation importance.

A limitation of the technique for associating the attention paid to each cell with expression of the input genes, is that it only considers one graph attention layer at a time. In practice, GAT layers are often part of multi-layered models, as is the case in our proposed model too. In this work, we disregarded the attention scores assigned by the second layer, but these might contain valuable information. Similarly, it is likely that better results can be achieved if the proposed attention correlation technique can be extended to interpret multiple attention heads. In the current implementation, attention scores from different heads are aggregated by taking their average, but in this step too, valuable information may be lost. An alternative could be to do the attention correlation calculations for each head or layer separately, and then combine the individual rankings into a joint ranking through a technique like majority voting. Finally, gene prioritization performance could be improved by training the interpreted GAT model on more genes to start with. This will likely improve the classification performance, and with it, potentially the resulting gene rankings.

Finally, the results of the proposed gene prioritization may also improve through the integration of additional data types. For example, we note the availability of chromatin accessibility (ATACseq) data in the SEA-AD dataset [25]. Integration of this data could not only improve classification performance, but the proposed attention correlation method can be extended to assign priorities to genes described by the ATACseq data as well. Similarly, the use of spatial transcriptomics data can help unveil patterns of gene expression that are limited to specific parts of the brain.

References

[1] P. Scheltens, B. De Strooper, M. Kivipelto, *et al.*, "Alzheimer's disease," *The Lancet*, vol. 397, no. 10284, pp. 1577–1590, 10284 Apr. 2021, ISSN: 01406736. DOI: 10.1016/S0140-6736(20)32205-4.

[2] "Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 17, no. 3, pp. 327–406, 3 Mar. 2021, ISSN: 1552-5260, 1552-5279. DOI: 10.1002/alz.12328.

[3] C. Patterson, "World Alzheimer Report 2018," 2018.

[4] A. Hung, Y. Liang, T. C. Chow, *et al.*, "Mutated tau, amyloid and neuroinflammation in Alzheimer disease—A brief review," *Progress in Histochemistry and Cytochemistry*, vol. 51, no. 1, pp. 1–8, 1 May 2016, ISSN: 00796336. DOI: 10.1016/j.proghi.2016.01.001.

[5] N. Gill, S. Singh, and T. C. Aseri, "Computational Disease Gene Prioritization: An Appraisal," *Journal of Computational Biology*, vol. 21, no. 6, pp. 456–465, 6 Jun. 2014, ISSN: 1066-5277, 1557-8666. DOI: 10.1089/cmb.2013.0158.

[6] L. M. Bekris, C.-E. Yu, T. D. Bird, and D. W. Tsuang, "Review Article: Genetics of Alzheimer Disease," *Journal of Geriatric Psychiatry and Neurology*, vol. 23, no. 4, pp. 213–227, 4 Dec. 2010, ISSN: 0891-9887, 1552-5708. DOI: 10.1177/0891988710383571.

[7] Y. Wang and N. E. Navin, "Advances and Applications of Single-Cell Sequencing Technologies," *Molecular Cell*, vol. 58, no. 4, pp. 598–609, 4 May 2015, ISSN: 10972765. DOI: 10.1016/j.molcel.2015.05.005.

[8] T. Nawy, "Single-cell sequencing," *Nature Methods*, vol. 11, no. 1, pp. 18–18, 1 Jan. 2014, ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.2771.

[9] S. S. Potter, "Single-cell RNA sequencing for the study of development, physiology and disease," *Nature Reviews Nephrology*, vol. 14, no. 8, pp. 479–492, 8 Aug. 2018, ISSN: 1759-5061, 1759-507X. DOI: 10.1038/s41581-018-0021-7.

[10] F. Buettner, K. N. Natarajan, F. P. Casale, *et al.*, "Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells," *Nature Biotechnology*, vol. 33, no. 2, pp. 155–160, 2 Feb. 2015, ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt.3102.

[11] D. Lähnemann, J. Köster, E. Szczurek, *et al.*, "Eleven grand challenges in single-cell data science," *Genome Biology*, vol. 21, no. 1, p. 31, 1 Dec. 2020, ISSN: 1474-760X. DOI: 10.1186/s13059-020-1926-6.

[12] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biology*, vol. 15, no. 12, p. 550, 12 Dec. 2014, ISSN: 1474-760X. DOI: 10.1186/s13059-014-0550-8.

[13] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR : A Bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics*, vol. 26, no. 1, pp. 139–140, 1 Jan. 1, 2010, ISSN: 1367-4811, 1367-4803. DOI: 10.1093/bioinformatics/btp616.

[14] Y. Li, X. Ge, F. Peng, W. Li, and J. J. Li, "Exaggerated false positives by popular differential expression methods when analyzing human population samples," *Genome Biology*, vol. 23, no. 1, p. 79, 1 Dec. 2022, ISSN: 1474-760X. DOI: 10.1186/s13059-022-02648-4.

[15] R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, and A. Regev, "Spatial reconstruction of single-cell gene expression data," *Nature Biotechnology*, vol. 33, no. 5, pp. 495–502, 5 May 2015, ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt.3192.

[16] 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, 1 Dec. 2018, ISSN: 1474-760X. DOI: 10.1186/s13059-017-1382-0.

[17] V. Y. Kiselev, K. Kirschner, M. T. Schaub, *et al.*, "SC3: Consensus clustering of single-cell RNA-seq data," *Nature Methods*, vol. 14, no. 5, pp. 483–486, 5 May 2017, ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.4236.

[18] M. Gori, G. Monfardini, and F. Scarselli, "A new model for learning in graph domains," in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, vol. 2, Montreal, Que., Canada: IEEE, 2005, pp. 729–734, ISBN: 978-0-7803-9048-5. DOI: 10.1109/IJCNN.2005.1555942.

[19] F. Scarselli, M. Gori, Ah Chung Tsoi, M. Hagenbuchner, and G. Monfardini, "The Graph Neural Network Model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 1 Jan. 2009, ISSN: 1045-9227, 1941-0093. DOI: 10.1109/TNN.2008.2005605.

[20] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," version 3, 2017. DOI: 10.48550/ARXIV.1710.10903.

[21] R. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "GNNExplainer: Generating Explanations for Graph Neural Networks," version 4, 2019. DOI: 10.48550/ARXIV.1903.03894.

[22] J. Chen, E. E. Bardes, B. J. Aronow, and A. G. Jegga, “ToppGene Suite for gene list enrichment analysis and candidate gene prioritization,” *Nucleic Acids Research*, vol. 37, no. Web Server, W305–W311, Web Server Jul. 1, 2009, ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkp427.

[23] S. Aerts, D. Lambrechts, S. Maity, *et al.*, “Gene prioritization through genomic data fusion,” *Nature Biotechnology*, vol. 24, no. 5, pp. 537–544, 5 May 1, 2006, ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt1203.

[24] A. Altabaa, D. Huang, C. Byles-Ho, H. Khatib, F. Sosa, and T. Hu, “geneDRAGNN: Gene Disease Prioritization using Graph Neural Networks,” in *2022 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, Ottawa, ON, Canada: IEEE, Aug. 15, 2022, pp. 1–10, ISBN: 978-1-66548-462-6. DOI: 10.1109/CIBCB55180.2022.9863043.

[25] M. I. Gabitto, K. J. Travaglini, V. M. Rachleff, *et al.* “Integrated multimodal cell atlas of Alzheimer’s disease.” (May 9, 2023), [Online]. Available: https://www.biorxiv.org/content/10.1101/2023.05.08.539485v1 (visited on 05/22/2023), preprint.

[26] I. Korsunsky, N. Millard, J. Fan, *et al.*, “Fast, sensitive and accurate integration of single-cell data with Harmony,” *Nature Methods*, vol. 16, no. 12, pp. 1289–1296, Dec. 2019, ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-019-0619-0.

[27] H. C. Rossetti, C. Munro Cullum, L. S. Hynan, and L. H. Lacritz, “The CERAD Neuropsychologic Battery Total Score and the Progression of Alzheimer Disease,” *Alzheimer Disease & Associated Disorders*, vol. 24, no. 2, pp. 138–142, Apr. 2010, ISSN: 0893-0341. DOI: 10.1097/WAD.0b013e3181b76415.

[28] P. Yilmaz, M. A. Ikram, M. K. Ikram, *et al.*, “Application of an Imaging-Based Sum Score for Cerebral Amyloid Angiopathy to the General Population: Risk of Major Neurological Diseases and Mortality,” *Frontiers in Neurology*, vol. 10, p. 1276, Dec. 6, 2019, ISSN: 1664-2295. DOI: 10.3389/fneur.2019.01276.

[29] A. R. Chaudhury, K. M. Gerecke, J. M. Wyss, D. G. Morgan, M. N. Gordon, and S. L. Carroll, “Neuregulin-1 and ErbB4 Immunoreactivity Is Associated with Neuritic Plaques in Alzheimer Disease Brain and in a Transgenic Model of Alzheimer Disease,” *Journal of Neuropathology & Experimental Neurology*, vol. 62, no. 1, pp. 42–54, Jan. 2003, ISSN: 0022-3069, 1554-6578. DOI: 10.1093/jnen/62.1.42.

[30] K.-S. Wang, N. Xu, L. Wang, *et al.*, “NRG3 gene is associated with the risk and age at onset of Alzheimer disease,” *Journal of Neural Transmission*, vol. 121, no. 2, pp. 183–192, Feb. 2014, ISSN: 0300-9564, 1435-1463. DOI: 10.1007/s00702-013-1091-0.

[31] M. Patel, “CSMD1 gene mutations can lead to familial Parkinson disease,” *Nature Reviews Neurology*, vol. 13, no. 11, pp. 641–641, Nov. 2017, ISSN: 1759-4758, 1759-4766. DOI: 10.1038/nrneurol.2017.132.

[32] B. W. Kunkle, M. Schmidt, H.-U. Klein, *et al.*, “Novel Alzheimer Disease Risk Loci and Pathways in African American Individuals Using the African Genome Resources Panel: A Meta-analysis,” *JAMA Neurology*, vol. 78, no. 1, p. 102, Jan. 1, 2021, ISSN: 2168-6149. DOI: 10.1001/jamaneurol.2020.3536.

[33] N. S. Raghavan, L. Dumitrescu, E. Mormino, *et al.*, “Association Between Common Variants in *RBFOX1* , an RNA-Binding Protein, and Brain Amyloidosis in Early and Preclinical Alzheimer Disease,” *JAMA Neurology*, vol. 77, no. 10, p. 1288, Oct. 1, 2020, ISSN: 2168-6149. DOI: 10.1001/jamaneurol.2020.1760.

[34] E. K. Lee, H. H. Kim, Y. Kuwano, *et al.*, “hnRNP C promotes APP translation by competing with FMRP for APP mRNA recruitment to P bodies,” *Nature Structural & Molecular Biology*, vol. 17, no. 6, pp. 732–739, Jun. 2010, ISSN: 1545-9993, 1545-9985. DOI: 10.1038/nsmb.1815.

[35] X. Liu, W. Chen, H. Hou, *et al.*, “Decreased functional connectivity between the dorsal anterior cingulate cortex and lingual gyrus in Alzheimer’s disease patients with depression,” *Behavioural Brain Research*, vol. 326, pp. 132–138, May 2017, ISSN: 01664328. DOI: 10.1016/j.bbr.2017.01.037.

[36] S. Köhler, S. Black, M. Sinden, *et al.*, “Memory impairments associated with hippocampal versus parahippocampal-gyrus atrophy: An MR volumetry study in Alzheimer’s disease,” *Neuropsychologia*, vol. 36, no. 9, pp. 901–914, Sep. 1998, ISSN: 00283932. DOI: 10.1016/S0028-3932(98)00017-7.

[37] S. Xie, J. X. Xiao, G. L. Gong, *et al.*, “Voxel-based detection of white matter abnormalities in mild Alzheimer disease,” *Neurology*, vol. 66, no. 12, pp. 1845–1849, Jun. 27, 2006, ISSN: 0028-3878, 1526-632X. DOI: 10.1212/01.wnl.0000219625.77625.aa.

[38] N. Hirono, H. Kitagaki, H. Kazui, M. Hashimoto, and E. Mori, “Impact of White Matter Changes on Clinical Manifestation of Alzheimer’s Disease: A Quantitative Study,” *Stroke*, vol. 31, no. 9, pp. 2182–2188, Sep. 2000, ISSN: 0039-2499, 1524-4628. DOI: 10.1161/01.STR.31.9.2182.

[39] B. Zhang, H. E. Wang, Y.-M. Bai, *et al.*, “Inflammatory bowel disease is associated with higher dementia risk: A nationwide longitudinal study,” *Gut*, vol. 70, no. 1, pp. 85–91, Jan. 2021, ISSN: 0017-5749, 1468-3288. DOI: 10.1136/gutjnl-2020-320789.

[40] R. G. Nagele, M. R. D’Andrea, H. Lee, V. Venkataraman, and H.-Y. Wang, “Astrocytes accumulate Aβ42 and give rise to astrocytic amyloid plaques in Alzheimer disease brains,” *Brain Research*, vol. 971, no. 2, pp. 197–209, May 2003, ISSN: 00068993. DOI: 10.1016/S0006-8993(03)02361-8.

[41] A. Verkhratsky, M. Olabarria, H. N. Noristani, C.-Y. Yeh, and J. J. Rodriguez, “Astrocytes in Alzheimer’s disease,” *Neurotherapeutics*, vol. 7, no. 4, pp. 399–412, Oct. 2010, ISSN: 1933-7213, 1878-7479. DOI: 10.1016/j.nurt.2010.05.017.

[42] R. Medeiros and F. M. LaFerla, “Astrocytes: Conductors of the Alzheimer disease neuroinflammatory symphony,” *Experimental Neurology*, vol. 239, pp. 133–138, Jan. 2013, ISSN: 00144886. DOI: 10.1016/j.expneurol.2012.10.007.

[43] Z. Cai and M. Xiao, “Oligodendrocytes and Alzheimer’s disease,” *International Journal of Neuroscience*, vol. 126, no. 2, pp. 97–104, Feb. 2016, ISSN: 0020-7454, 1543-5245. DOI: 10.3109/00207454.2015.1025778.

[44] X. Zhan, B. Stamova, and F. R. Sharp, “Lipopolysaccharide Associates with Amyloid Plaques, Neurons and Oligodendrocytes in Alzheimer’s Disease Brain: A Review,” *Frontiers in Aging Neuroscience*, vol. 10, p. 42, Feb. 22, 2018, ISSN: 1663-4365. DOI: 10.3389/fnagi.2018.00042.

[45] J. S. Sadick, M. R. O’Dea, P. Hasel, T. Dykstra, A. Faustin, and S. A. Liddelow, “Astrocytes and oligodendrocytes undergo subtype-specific transcriptional changes in Alzheimer’s disease,” *Neuron*, vol. 110, no. 11, 1788–1805.e10, Jun. 2022, ISSN: 08966273. DOI: 10.1016/j.neuron.2022.03.008.

[46] A. C. Paula-Lima, J. Brito-Moreira, and S. T. Ferreira, “Deregulation of excitatory neurotransmission underlying synapse failure in Alzheimer’s disease,” *Journal of Neurochemistry*, vol. 126, no. 2, pp. 191–202, Jul. 2013, ISSN: 00223042. DOI: 10.1111/jnc.12304.

[47] J. Greenamyre and A. B. Young, “Excitatory amino acids and Alzheimer’s disease,” *Neurobiology of Aging*, vol. 10, no. 5, pp. 593–602, Sep. 1989, ISSN: 01974580. DOI: 10.1016/0197-4580(89)90143-7.

[48] R. R. Romito-DiGiacomo, H. Menegay, S. A. Cicero, and K. Herrup, “Effects of Alzheimer’s Disease on Different Cortical Layers: The Role of Intrinsic Differences in Aβ Susceptibility,” *The Journal of Neuroscience*, vol. 27, no. 32, pp. 8496–8504, Aug. 8, 2007, ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.1008-07.2007.

[49] H. Braak and E. Braak, “Neuropathological stageing of Alzheimer-related changes,” *Acta Neuropathologica*, vol. 82, no. 4, pp. 239–259, Sep. 1991, ISSN: 0001-6322, 1432-0533. DOI: 10.1007/BF00308809.

[50] H. Braak, E. Braak, and J. Bohl, “Staging of Alzheimer-Related Cortical Destruction,” *European Neurology*, vol. 33, no. 6, pp. 403–408, 1993, ISSN: 0014-3022, 1421-9913. DOI: 10.1159/000116984.

[51] P. M. Doraiswamy, J. Leon, J. L. Cummings, D. Marin, and P. J. Neumann, “Prevalence and Impact of Medical Comorbidity in Alzheimer’s Disease,” *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 57, no. 3, pp. M173–M177, Mar. 1, 2002, ISSN: 1079-5006, 1758-535X. DOI: 10.1093/gerona/57.3.M173.

[52] A. Duthie, D. Chew, and R. L. Soiza, “Non-psychiatric comorbidity associated with Alzheimer’s disease,” *QJM*, vol. 104, no. 11, pp. 913–920, Nov. 1, 2011, ISSN: 1460-2725, 1460-2393. DOI: 10.1093/qjmed/hcr118.

[53] R. Matej, A. Tesar, and R. Rusina, “Alzheimer’s disease and other neurodegenerative dementias in comorbidity: A clinical and neuropathological overview,” *Clinical Biochemistry*, vol. 73, pp. 26–31, Nov. 2019, ISSN: 00099120. DOI: 10.1016/j.clinbiochem.2019.08.005.

[54] M. L. Haaksma, L. R. Vilela, A. Marengoni, *et al.*, “Comorbidity and progression of late onset Alzheimer’s disease: A systematic review,” *PLOS ONE*, vol. 12, no. 5, S. D. Ginsberg, Ed., e0177044, May 4, 2017, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0177044.

[55] B. Teter and J. W. Ashford, “Neuroplasticity in Alzheimer’s disease,” *Journal of Neuroscience Research*, vol. 70, no. 3, pp. 402–437, Nov. 1, 2002, ISSN: 0360-4012, 1097-4547. DOI: 10.1002/jnr.10441.

[56] G. Koch and D. Spampinato, “Alzheimer disease and neuroplasticity,” in *Handbook of Clinical Neurology*, vol. 184, Elsevier, 2022, pp. 473–479, ISBN: 978-0-12-819410-2. DOI: 10.1016/B978-0-12-819410-2.00027-8.

[57] X. Yao, J. Yan, K. Liu, *et al.*, “Tissue-specific network-based genome wide study of amygdala imaging phenotypes to identify functional interaction modules,” *Bioinformatics*, vol. 33, no. 20, pp. 3250–3257, Oct. 15, 2017, ISSN: 1367-4803,

1367-4811. DOI: 10.1093/bioinformatics/btx344.

[58] R. C. P. Go, R. T. Perry, H. Wiener, *et al.*, "Neuregulin-1 polymorphism in late onset Alzheimer's disease families with psychoses," *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, vol. 139B, no. 1, pp. 28–32, Nov. 5, 2005, ISSN: 1552-4841, 1552-485X. DOI: 10.1002/ajmg.b.30219.

[59] B. Vincent and S. Maitra, "BACE1-dependent metabolism of neuregulin 1: Bridging the gap in explaining the occurrence of schizophrenia-like symptoms in Alzheimer's disease with psychosis?" *Ageing Research Reviews*, vol. 89, p. 101 988, Aug. 2023, ISSN: 15681637. DOI: 10.1016/j.arr.2023.101988.

[60] W.-T. Kao, Y. Wang, J. E. Kleinman, *et al.*, "Common genetic variation in Neuregulin 3 (*NRG3*) influences risk for schizophrenia and impacts *NRG3* expression in human brain," *Proceedings of the National Academy of Sciences*, vol. 107, no. 35, pp. 15 619–15 624, Aug. 31, 2010, ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1005410107.

[61] P. Ma, Y. Li, W. Zhang, *et al.*, "Long Non-coding RNA MALAT1 Inhibits Neuron Apoptosis and Neuroinflammation While Stimulates Neurite Outgrowth and Its Correlation With MiR-125b Mediates PTGS2, CDK5 and FOXQ1 in Alzheimer's Disease," *Current Alzheimer Research*, vol. 16, no. 7, pp. 596–612, Sep. 4, 2019, ISSN: 15672050. DOI: 10.2174/1567205016666190725130134.

[62] L. Li, Y. Xu, M. Zhao, and Z. Gao, "Neuroprotective roles of long non-coding RNA MALAT1 in Alzheimer's disease with the involvement of the microRNA-30b/CNR1 network and the following PI3K/AKT activation," *Experimental and Molecular Pathology*, vol. 117, p. 104 545, Dec. 2020, ISSN: 00144800. DOI: 10.1016/j.yexmp.2020.104545.

[63] K. Chanda, N. R. Jana, and D. Mukhopadhyay, "Long non-coding RNA MALAT1 protects against Aβ1–42 induced toxicity by regulating the expression of receptor tyrosine kinase EPHA2 via quenching miR-200a/26a/26b in Alzheimer's disease," *Life Sciences*, vol. 302, p. 120 652, Aug. 2022, ISSN: 00243205. DOI: 10.1016/j.lfs.2022.120652.

[64] H. Kitagawa, W. J. Ray, H. Glantschnig, *et al.*, "A Regulatory Circuit Mediating Convergence between Nurr1 Transcriptional Regulation and Wnt Signaling," *Molecular and Cellular Biology*, vol. 27, no. 21, pp. 7486–7496, Nov. 1, 2007, ISSN: 1098-5549. DOI: 10.1128/MCB.00409-07.

[65] Y. Morohashi, N. Hatano, S. Ohya, *et al.*, "Molecular Cloning and Characterization of CALP/KChIP4, a Novel EF-hand Protein Interacting with Presenilin 2 and Voltage-gated Potassium Channel Subunit Kv4," *Journal of Biological Chemistry*, vol. 277, no. 17, pp. 14 965–14 975, Apr. 2002, ISSN: 00219258. DOI: 10.1074/jbc.M200897200.

[66] M. Newman, F. Musgrave, and M. Lardelli, "Alzheimer disease: Amyloidogenesis, the presenilins and animal models," *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, vol. 1772, no. 3, pp. 285–297, Mar. 2007, ISSN: 09254439. DOI: 10.1016/j.bbadis.2006.12.001.

[67] Y.-w. Zhang, R. Thompson, H. Zhang, and H. Xu, "APP processing in Alzheimer's disease," *Molecular Brain*, vol. 4, no. 1, p. 3, 2011, ISSN: 1756-6606. DOI: 10.1186/1756-6606-4-3.

[68] S. Swaminathan, S. Kim, L. Shen, *et al.*, "Genomic Copy Number Analysis in Alzheimer's Disease and Mild Cognitive Impairment: An ADNI Study," *International Journal of Alzheimer's Disease*, vol. 2011, pp. 1–10, 2011, ISSN: 2090-0252. DOI: 10.4061/2011/729478.

[69] V. A. Stepanov, A. V. Bocharova, A. V. Marusin, N. G. Zhukova, V. M. Alifirova, and I. A. Zhukova, "Replicative association analysis of genetic markers of cognitive traits with Alzheimer's disease in the Russian population," *Molecular Biology*, vol. 48, no. 6, pp. 835–844, Nov. 2014, ISSN: 0026-8933, 1608-3245. DOI: 10.1134/S0026893314060168.

[70] V. Stepanov, A. Marusin, K. Vagaitseva, A. Bocharova, and O. Makeeva, "Genetic Variants in CSMD1 Gene Are Associated with Cognitive Performance in Normal Elderly Population," *Genetics Research International*, vol. 2017, pp. 1–5, Dec. 12, 2017, ISSN: 2090-3154, 2090-3162. DOI: 10.1155/2017/6293826.

[71] T. Porter, V. L. Villemagne, G. Savage, *et al.*, "Cognitive gene risk profile for the prediction of cognitive decline in presymptomatic Alzheimer's disease," *Personalized Medicine in Psychiatry*, vol. 7–8, pp. 14–20, Mar. 2018, ISSN: 24681717. DOI: 10.1016/j.pmip.2018.03.001.

[72] M. Aggarwal, M. Alkhayyat, M. Abou Saleh, *et al.*, "Alzheimer Disease Occurs More Frequently In Patients With Inflammatory Bowel Disease: Insight From a Nationwide Study," *Journal of Clinical Gastroenterology*, vol. 57, no. 5, pp. 501–507, May 2023, ISSN: 1539-2031. DOI: 10.1097/MCG.0000000000001714.

[73] N. Liu, Y. Wang, L. He, J. Sun, X. Wang, and H. Li, "Inflammatory bowel disease and risk of dementia: An updated meta-analysis," *Frontiers in Aging Neuroscience*, vol. 14, p. 962 681, Oct. 5, 2022, ISSN: 1663-4365. DOI: 10.3389/fnagi.2022.962681.

[74] G. E. Wright, J. C. Parker, K. L. Smarr, J. C. Johnson, J. E. Hewett, and S. E. Walker, "Age, depressive symptoms, and rheumatoid arthritis," *Arthritis & Rheumatism*, vol. 41, no. 2, pp. 298–305, 1998, ISSN: 1529-0131. DOI: 10.1002/1529-0131(199802)41:2<298::AID-ART14>3.0.CO;2-G.

[75] D. van Schaardenburg, J. M. Hazes, A. de Boer, A. H. Zwinderman, K. A. Meijers, and F. C. Breedveld, "Outcome of rheumatoid arthritis in relation to age and rheumatoid factor at diagnosis," *The Journal of rheumatology*, vol. 20, no. 1, pp. 45–52, Jan. 1, 1993, ISSN: 0315-162X.

[76] A. Shane Anderson and R. F. Loeser, "Why is osteoarthritis an age-related disease?" *Best Practice & Research Clinical Rheumatology*, vol. 24, no. 1, pp. 15–26, Feb. 2010, ISSN: 15216942. DOI: 10.1016/j.berh.2009.08.006.

[77] G. Scelo and T. L. Larose, "Epidemiology and Risk Factors for Kidney Cancer," *Journal of Clinical Oncology*, vol. 36, no. 36, pp. 3574–3581, Dec. 20, 2018, ISSN: 0732-183X, 1527-7755. DOI: 10.1200/JCO.2018.79.1905.

[78] S. A. Khan, S. A. Khan, A. Narendra, *et al.*, "Alzheimer's Disease and Autistic Spectrum Disorder: Is there any Association?" *CNS & Neurological Disorders - Drug Targets*, vol. 15, no. 4, pp. 390–402, Mar. 28, 2016, ISSN: 18715273. DOI: 10.2174/1871527315666160321104303.

[79] M. S. Nadeem, S. Hosawi, S. Alshehri, *et al.*, "Symptomatic, Genetic, and Mechanistic Overlaps between Autism and Alzheimer's Disease," *Biomolecules*, vol. 11, no. 11, p. 1635, Nov. 4, 2021, ISSN: 2218-273X. DOI: 10.3390/biom11111635.

[80] M. Vaňková, M. Velíková, D. Vejražková, *et al.*, "The Role of Steroidomics in the Diagnosis of Alzheimer's Disease and Type 2 Diabetes Mellitus," *International Journal of Molecular Sciences*, vol. 24, no. 10, p. 8575, May 10, 2023, ISSN: 1422-0067. DOI: 10.3390/ijms24108575.

[81] C. J. Cunningham, M. Sinnott, A. Denihan, *et al.*, "Endogenous Sex Hormone Levels in Postmenopausal Women with Alzheimer's Disease," *The Journal of Clinical Endocrinology & Metabolism*, vol. 86, no. 3, pp. 1099–1103, Mar. 2001, ISSN: 0021-972X, 1945-7197. DOI: 10.1210/jcem.86.3.7289.

[82] G. D. Batty, P. Frank, U. M. Kujala, S. J. Sarna, C. A. Valencia-Hernández, and J. Kaprio, "Dementia and Alzheimer's disease in former contact sports participants: Population-based cohort study, systematic review, and meta-analysis," Epidemiology, preprint, May 25, 2022. DOI: 10.1101/2022.05.24.22275500.

[83] M. Yamazaki, Q. De Larochelambert, G. Sauliere, J.-F. Toussaint, and J. Antero, "Heads-Up: Risk-Specific Neurodegenerative Mortality and Years-Saved Analysis on the US Olympian Cohort," *Frontiers in Physiology*, vol. 12, p. 705 616, Sep. 9, 2021, ISSN: 1664-042X. DOI: 10.3389/fphys.2021.705616.

[84] J. Y. Kim, S. Barua, Y. J. Jeong, and J. E. Lee, "Adiponectin: The Potential Regulator and Therapeutic Target of Obesity and Alzheimer's Disease," *International Journal of Molecular Sciences*, vol. 21, no. 17, p. 6419, Sep. 3, 2020, ISSN: 1422-0067. DOI: 10.3390/ijms21176419.

[85] M. Waragai, G. Ho, Y. Takamatsu, *et al.*, "Adiponectin Paradox in Alzheimer's Disease; Relevance to Amyloidogenic Evolvability?" *Frontiers in Endocrinology*, vol. 11, p. 108, Mar. 4, 2020, ISSN: 1664-2392. DOI: 10.3389/fendo.2020.00108.

[86] P. Agarwal, S. E. Leurgans, S. Agrawal, *et al.*, "Association of Mediterranean-DASH Intervention for Neurodegenerative Delay and Mediterranean Diets With Alzheimer Disease Pathology," *Neurology*, vol. 100, no. 22, e2259–e2268, May 30, 2023, ISSN: 0028-3878, 1526-632X. DOI: 10.1212/WNL.0000000000207176.

[87] O. M. Shannon, J. M. Ranson, S. Gregory, *et al.*, "Mediterranean diet adherence is associated with lower dementia risk, independent of genetic predisposition: Findings from the UK Biobank prospective cohort study," *BMC Medicine*, vol. 21, no. 1, p. 81, Mar. 14, 2023, ISSN: 1741-7015. DOI: 10.1186/s12916-023-02772-3.

[88] T. Ballarini, D. Melo Van Lent, J. Brunner, *et al.*, "Mediterranean Diet, Alzheimer Disease Biomarkers, and Brain Atrophy in Old Age," *Neurology*, vol. 96, no. 24, e2920–e2932, Jun. 15, 2021, ISSN: 0028-3878, 1526-632X. DOI: 10.1212/WNL.0000000000012067.

Methods

Data pre-processing

The first pre-processing step involves the removal of all data from donors whose diagnoses are unknown. This ensures the accuracy and relevance of our analysis, as a precise diagnosis provides a basis for understanding the cell-specific variations associated with AD.

Next, we remove donors with under 10,000 cells (which is one standard deviation under the mean), and we downsample the cells of donors with a cell count above 20,000 (which is one standard deviation above the mean). As a result, cell counts differ at most a factor 2 between donors (Figure 10). To motivate this step, we note that the expected value of similarity between a cell and its i'th closest neighbor is proportional to the number of cells samples from the donor. Consequently, if donors' cell counts vary too much, the similarity between their cells and their neighbors will also vary a lot. This procedure aims to stabilize the similarity between neighbors.

The remaining cells are labeled according to the clinical diagnosis of the donor they are derived from. Cells from donors with diagnoses labeled as "AD Likely" and "AD Probable" are also considered AD donors because without them there would be too few AD samples.

To limit computational load and focus on the most dynamic features in the data, we train classification models only on the 1000 most highly variable genes. These are selected using the standard procedure in Seurat [1], where genes are first standardized, to calculate the normalized variance. This makes the method less prone to selecting just the highly expressed genes, which would naturally have a higher variance.

Finally, we construct k-Nearest Neighbor (KNN) graphs with $k = 30$, for each donor separately. Each graph consists of nodes representing cells, which are connected to the k cells that have the highest similarity (i.e., the lowest Euclidean distance in the feature space). Besides the reduction in computational overhead, we argue that representing donors in separate graphs also makes more sense from a biological perspective. Cells that are connected in a donor-specific graph are likely to also have been located physically close to each other in the donor's brain. Therefore, the integration of information from neighboring cells in the KNN graph resembles the communication between the cells.

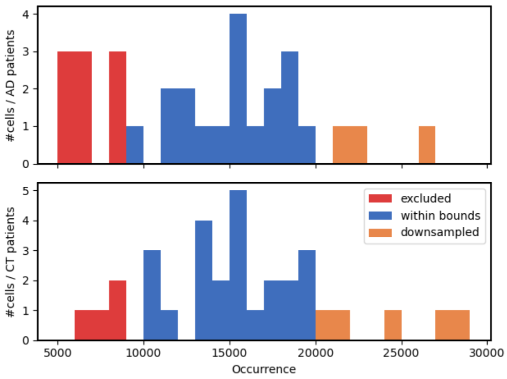


Figure 10: Cell count distribution per patient, split between the two label classes (AD in the top row, CT in the bottom row). Colors show the donors that have been excluded or downsampled, because of their too low or high cell counts.

Graph Attention Network (GAT)

Disease predictions for the cells are made using a two-layer Graph Attention Network (GAT) [2], with a total hidden dimension size of 1024. The full embeddings in this hidden dimension are the concatenation of the embeddings from each of the 8 attention head, with a size of 128 each. In each attention head, the latent representation h_i for a node i is calculated as:

$$h_i = \sigma\left(\sum_{j \in N_k(i)} \alpha_{i \to j} W x_j\right)$$

Here, $N_k(i)$ denotes the k-neighborhood of i, x_j is the feature vector of neighbor j, and W is a learnable linear transformation. $\alpha_{i \to j}$ denotes the softmaxed attention that distinguishes GATs from other GNNs. The attention value is derived from feature vectors of nodes i and j, and allows the model to learn unique weightings for neighbors, that are optimized for the downstream prediction task. The attention $\alpha_{i \to j}$ that is paid to j when calculating the embedding of i is calculated following the definition of GATv2 [3], where different weights are applied to i and j, to allow for "dynamic" attention:

$$\alpha_{i \to j} = softmax_j \left(a \cdot \text{LeakyReLU}\left(W' \left[x_i || x_j\right]\right)\right)$$

where a and W' are learnable transformations, and the $||$ operator denotes the concatenation of vectors. Note that by softmaxing over all neighbors j, the attention values sum to 1, allowing us to directly take the weighted average over the neighbors. Finally, the results of the above calculations from each of the parallel attention heads are concatenated, to construct the latent representation of size 1024.

As opposed to most implementations of GAT, we do not include self-loops: $i \notin N_k(j)$. This is because that would also divert much of the attention to nodes themselves, which is usually not problematic if we are merely interested in maximizing classification performance. By not including the self-connecting edges, however, we force the model to make predictions using neighboring nodes, improving the interpretability of the attention scores, which better suits the main objective in this work.

To train the model, we use a categorical cross-entropy loss function, that minimizes the difference between the soft label assignment predicted by the model and the actual class labels. Model weights are optimized using the Adam optimizer [4], with a learning rate of 0.0001, and to combat overfitting we use a weight decay of 0.00005, and a dropout rate of 0.75. These hyper parameters (including hidden dimension size and number of heads) have been tuned only roughly, because thorough optimization would require averaging over too many runs to get a stable prediction of the performance.

Mini-batches that the model is trained upon are constructed using neighbor sampling [5], where 32 seed nodes are sampled from the different training graphs. Sampling is done in a weighted fashion, to correct for the For each seed node, 15 random neighbors are added to the batch graph, and for each neighbor, 15 additional randomly selected (1-hop) neighbors are added to the batch graph. As a result, each batch has at most $32 \times 15 \times 15 = 7200$ nodes, though in practice this number is smaller, because many neighbors are shared. We provide the model with only 200 of such batches every training epoch, to ensure it sees roughly the same amount of data as the baseline model every epoch. Overall, this batching technique reduces computational overhead, and allows us to use class-balanced batches, which stabilizes the optimization process. In addition, this random sampling of neighborhoods acts as a regularization technique, similar to node-wise dropout, enhancing generalizability of the model.

Baseline Model (2-Layer MLP)

To evaluate classification performance, we compare several metrics against a baseline model. We use a two-layer neural network, with an equally sized hidden dimension of 1024. This allows for a fair comparison with our model, that is presented the same input samples and features. The main difference is therefore the attention mechanism employed by the GAT that enables it to integrate information from neighboring cells.

The same loss function and optimizer are used for the baseline, with slightly different hyperparameters. The learning rate remains 0.0001, but a larger weight decay of 0.005 was required to keep the model from overfitting. A dropout rate of 0.5 proved best in tuning experiments, and the batchsize was set to 2000, which is roughly the same size as the average number of nodes in the batch graphs used to train the GAT model.

Model Performance Evaluation

To quantify the models' abilities to generalize to unseen data, we employ leave-one-out crossvalidation (LOOCV). Where the model is trained from scratch for each individual donor, holding out just that donor, and training on the rest of the dataset. This approach was preferred over regular hold-out crossvalidation, because this would have further reduced the already limited number of donor samples, which makes assessing generalizability even harder.

Classification performance of both models is quantified using several metrics. We consider the accuracy, which denotes the ratio of correctly classified samples in the evaluation set. In the comparison with the baseline, we also report precision, recall and the F1-score (Table 1).

In the analyses where we explore the biological patterns that underlie the differences in predictive power between different cell types and donors, we also use what we refer to as the "soft accuracy" (Figures 4 and 5). For each individual cell, this is defined as the distance between the predicted soft-label and the target hard label. This measure provides insight in the certainty of classifications per individual cell that is not available with traditional "hard accuracy".

UMAP Visualization

We further assess the embeddings generated by the model by visualizing them using UMAP [6], which reduces the dimensionality to two dimensions, to allow for 2D visualization. The UMAP representations of the embeddings from the GAT model are then compared with UMAPs calculated for the input feature space, to compare the different representations (Figure 3). The UMAP algorithm aims to preserve the global structure of the data, while locally optimizing the visual representation of distances between samples in the original space. For both the input space and the GAT's latent space visualizations, first do an initial dimensionality reduction where we select the first 50 principal components using PCA. The settings of UMAP that we used are all left to the defaults, except n_neighbors = 100.

Furthermore, we note that in order to produce the embeddings used for the UMAPs, we separately train a new GAT model on the entire dataset. This circumvents the bias introduced by having to use the embeddings from one of the LOOCV iterations. Note that the performance can be expected to be accurately represented by the LOOCV results, because the training set is nearly the same (only one donor is added, compared to each LOOCV iteration), and we use the same hyper parameters.

Attention Correlation

From the GAT model trained on the full dataset, attention weights can be retreived, that assign a weight to each neighbor, for each cell. These weights sum to 1 over all neighbors of each cell. To be able to interpret these, we first aggregate the received attention of each cell, by summing over the received attention from all neighbors. On average the received attention is 1, but some cells receive as much as 10× more attention (Figure 7. We do this for the attention weights from each of the eight parallel attention heads, and then take mean over the heads for each cell, to end up with a score $att_{recv}(i)$ for each cell i, given it's neighborhood $N(i)$:

$$att_{recv}(i) = \frac{1}{n_{heads}} \sum_{h \in n_{heads}} \sum_{i \in N(i)} \alpha_{j \to i}$$

Next, we use the received attention per cell, to assign scores to each of the genes in the dataset. We do this through a measure we refer to as "attention correlation". This attention correlation scores genes according to the correlation between the expression of that gene in each cell, and the attention those cells received. Given the received attention att_{recv} defined as above, the attention correlation $att_{corr}(g)$ for gene g is calculated as:

$$att_{corr}(g) = \rho\left(\begin{pmatrix} att_{recv}(0) \\ att_{recv}(1) \\ \vdots \\ att_{recv}(n) \end{pmatrix}, \begin{pmatrix} X_{0,g} \\ X_{1,g} \\ \vdots \\ X_{n,g} \end{pmatrix}\right)$$

Here, $X_{i,g}$ denotes the (log-normalized) expression of gene g in cell i, and ρ denotes the Pearson correlation coefficient. A higher attention correlation indicates the gene was used more extensively in generating the embeddings, and during the prediction task the model is trained on. Note that attention correlation can also be calculated for genes that were excluded before training the GAT model (in our case, because we only selected the most highly variable genes).

Top ranking genes were selected for each major cell type in the dataset. Considering the attention correlation on this scale allows us to find patterns in gene expression that are not available when considering it in it's entirety.

To evaluate the statistical significance of the found top genes (Table 2), we perform permutation testing. Each gene's expression is randomly permuted 10^7 times, and for each permutation the attention correlation is calculated again. Scores were deemed significant if the amount of times they are higher than the random permutation scores exceeds the confidence threshold. This threshold is derived using $p = 0.05$, and correcting it for multiple testing (Bonferroni correction), with the total amount of genes $n_{genes} = 36,601$. As such, the resulting confidence boundary is $B = 0.05/n_{genes} = 1.366 \times 10^{-6}$.

Finally, we perfomed enrichment analysis on the top 20 genes of each of the considered cell types. For this we used Enrichr [7], and considered the databases GWAS Catalog (2023) [8] and DisGeNET [9] [10] [11]. The top 15 terms over all celltypes are reported in the results, as well as the top 15 terms that are explicitly related to Alzheimer's disease.

Methods References

[1] T. Stuart, A. Butler, P. Hoffman, *et al.*, "Comprehensive Integration of Single-Cell Data," *Cell*, vol. 177, no. 7, 1888–1902.e21, Jun. 2019, ISSN: 00928674. DOI: 10.1016/j.cell.2019.05.031.

[2] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," version 3, 2017. DOI: 10.48550/ARXIV.1710.10903.

[3] S. Brody, U. Alon, and E. Yahav, "How Attentive are Graph Attention Networks?" Version 3, 2021. DOI: 10.48550/ARXIV.2105.14491.

[4] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," version 9, 2014. DOI: 10.48550/ARXIV.1412.6980.

[5] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive Representation Learning on Large Graphs," version 4, 2017. DOI: 10.48550/ARXIV.1706.02216.

[6] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," version 3, 2018. DOI: 10.48550/ARXIV.1802.03426.

[7] E. Y. Chen, C. M. Tan, Y. Kou, *et al.*, "Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool," *BMC Bioinformatics*, vol. 14, no. 1, p. 128, Dec. 2013, ISSN: 1471-2105. DOI: 10.1186/1471-2105-14-128.

[8] E. Sollis, A. Mosaku, A. Abid, *et al.*, "The NHGRI-EBI GWAS Catalog: Knowledgebase and deposition resource," *Nucleic Acids Research*, vol. 51, no. D1, pp. D977–D985, Jan. 6, 2023, ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkac1010.

[9] J. Pinero, N. Queralt-Rosinach, A. Bravo, *et al.*, "DisGeNET: A discovery platform for the dynamical exploration of human diseases and their genes," *Database*, vol. 2015, no. 0, bav028–bav028, Apr. 15, 2015, ISSN: 1758-0463. DOI: 10.1093/database/bav028.

[10] J. Piñero, À. Bravo, N. Queralt-Rosinach, *et al.*, "DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants," *Nucleic Acids Research*, vol. 45, no. D1, pp. D833–D839, Jan. 4, 2017, ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkw943.

[11] J. Piñero, J. M. Ramírez-Anguita, J. Saüch-Pitarch, *et al.*, "The DisGeNET knowledge platform for disease genomics: 2019 update," *Nucleic Acids Research*, gkz1021, Nov. 4, 2019, ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkz1021.

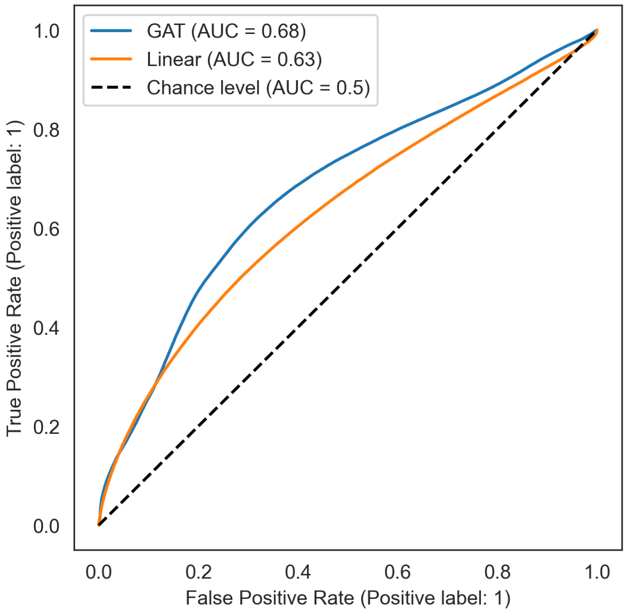


Figure S1: Receiver operating characteristic (ROC) curve, showing the trade-off between the true positive rate (TPR) and false positive rate (FPR). The AUC score (between 0 and 1) is calculated as the area under this curve, where a higher AUC corresponds to true positive requiring less false positives.

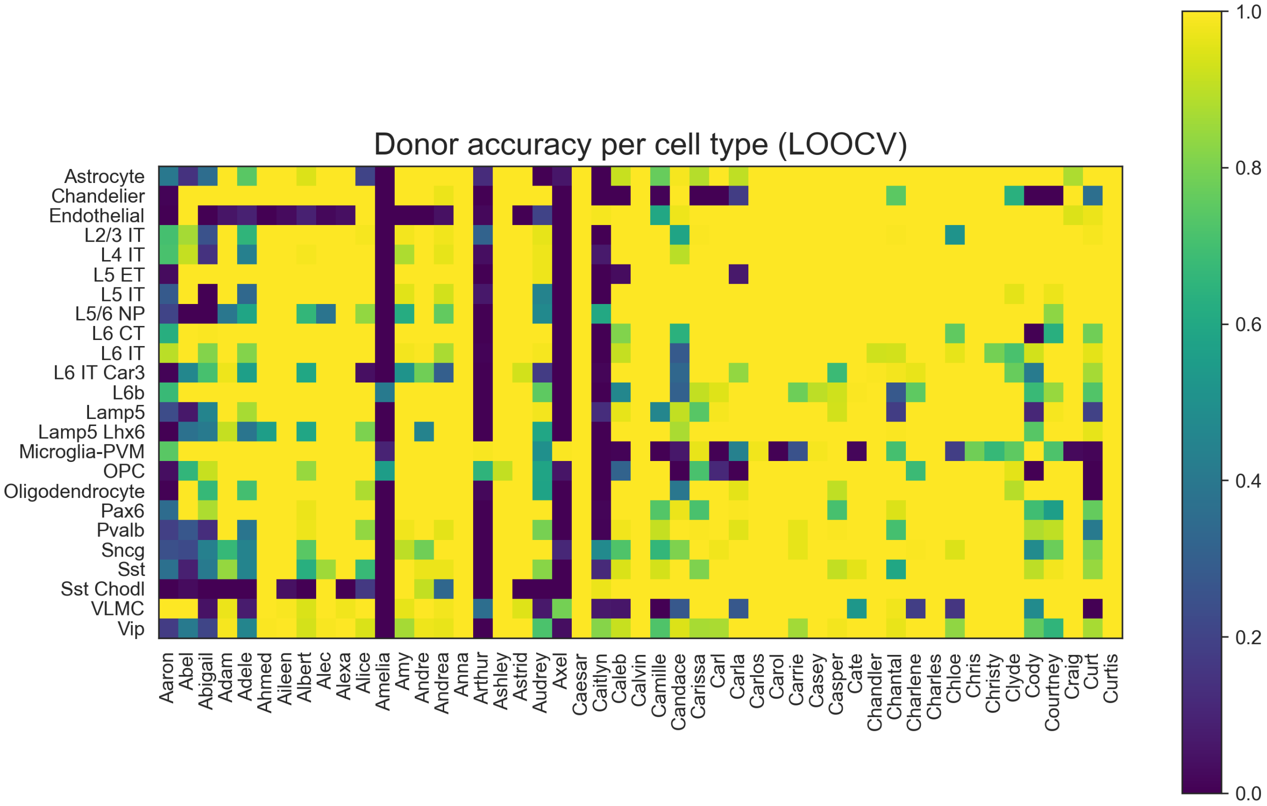


Figure S2: Donor accuracy per cell type. Donor's whose name starts with an 'A' are AD donors and starting with a 'C' are CT donors. Accuracy denotes the fraction of cells correctly classified in that donor, for that celltype.