Deep Learning in Olfactory Learning

A thesis submitted in partial fulfillment of the requirements for Honors in the Department of Computer Science at Brown University

Sean Yu

Ritambhara Singh, Ph.D *(Advisor)*

Alexander Fleischmann, Ph.D *(Second Reader)*

BROWN

April 18, 2025

Acknowledgements

My deepest gratitude goes to Professor Singh for her careful and attentive guidance throughout my time in the Singh Lab. She welcomed me without hesitation and taught me so much, and none of this project or what I've learned about deep learning and research would have been possible without her support and mentorship. Tuan, my Ph.D student mentor in the Singh Lab, has also been instrumental in showing and teaching me what research looks like, helping me debug, interpret and ideate research ideas. In the Fleischmann lab, I'm so grateful for Professor Fleischmann, Keeley Baker, and Tuan Pham (neuro) for their support in teaching me the biology of olfactory learning and showing me what productive interdisciplinary research looks like. Finally, I have been deeply blessed by my friends and family for their unwavering and constant support throughout my time doing research at Brown. Without a single one of these people above, this thesis would not have been possible.

Deep Learning in Olfactory Learning

Sean Yu
Ritambhara Singh, Ph.D
Alexander Fleischmann, Ph.D

Abstract
Given the recent "AI Boom" and advances in single-cell RNA-sequencing (scRNA-seq), applications of AI, specifically deep learning (DL), to scRNA-seq data have become increasingly popular. Even with this growing popularity, however, there exist gaps in this scRNA-seq/DL intersection, two of which this study focuses on. (1) Most applications of DL to scRNA-seq data seek to impute, cluster, annotate, or clean scRNA-seq datasets, and DL is rarely used to directly uncover biological insights. (2) Secondly, most current models fail to allow for primary objectives (e.g. predicting experimental conditions, cell cycle stage, etc) to be built upon secondary objectives like cell-type embedding. This limits our ability to use DL in more constrained circumstances where, for instance, cell-type-based analysis is required. As such, this study builds a multi-label model that can predict experimental condition exposure with 77% and 88% accuracy in two samples of mouse piriform cortex neurons while simultaneously maintaining an accurate cell-type embedding, evidenced by distinct UMAP clustering and positive silhouette scoring. The model is then used to perform SHaply Additive exPlanation (SHAP) analysis to understand influential features in the model's decision making, which in theory translate to genes that are biologically significant to the process of olfactory learning. SHAP analysis yielded the top 40 most influential genes in the model's decision making, but significant validation is needed to elucidate the uncovered genes' biological relation to olfactory learning. All in all, this study shows that a multi-label model can be used to effectively retain cell-type embeddings during scRNA-seq/DL analysis, and that SHAP analysis applied to such models may yield biologically significant findings.

Contents

1 Introduction

The sense of smell is a powerful gateway through which we understand the world around us. Walking in and around this campus, I can smell the fresh cut grass of the main green, the fragrant stir-fries of the wok station at Andrews Dining Hall, and the pungent stench of the CIT bathrooms. But how does my brain know what these smells are and how to associate them with the main green, food, or the bathroom? How does the brain learn and change, specifically the piriform cortex, when exposed to a variety of olfactory stimulants? In this study, I approach this question through a computational lens, applying deep learning to shed more light on the biological pathways that facilitate olfactory learning.

1.1 Background

1.1.1 Olfaction

In mammals, olfaction is unique among the senses in that it bypasses the thalamus, achieving processing by different means.[12;27] When scent molecules are inhaled through the nose, they're received by the olfactory epithelium, which then carries signals to the olfactory bulb. From the olfactory bulb, signals are then passed into the piriform cortex, which then passes signals to higher-order processing units.[7] Within the piriform cortex, a great deal of processing takes place, and it has even been implicated in studies related to memory and learning.[21;13] Specifically, it has been found that odors' molecular features and quality (i.e., the "perceptual character" of a smell) are encoded in the anterior and posterior regions of the piriform cortex.[10]

The mammalian piriform cortex is primarily broken down into three layers[36] that contain different types of neurons: pyramidal cells, semilunar cells, and inhibitory cells.[7] While the function of each layer may not be entirely discrete or unique, layer 1 is generally responsible for carrying signals from the olfactory bulb into layer 2.[31] In layer 2 and 3, these signals are processed and encoded by a combination of excitatory and inhibitory cells.[31] The signals from layer 2 are then bilaterally projected on to various parts of the olfactory tract (olfactory bulb, piriform cortex), and the signals from layer 3 are projected to regions of the amygdala[25].

There are two main excitatory cells in layer 2: pyramidal cells and semilunar cells. Pyramidal cells form large interconnected networks that fire in conjunction with each other, associationally and recurrently projecting on to each other.[13] These networks of pyramidal cells can reach deep into layer 3, where further associational projection occurs.[13] Semilunar cells, on the other hand, are also equally excitatory in layer 2 but differ in connectivity in that they receive strong afferent input from the olfactory bulb but do not participate in associational or recurrent circuits.[31] As such, while both pyramidal and semilunar cells play excitatory roles in the piriform cortex, they serve in distinct roles in signal processing and learning.[31]

In addition to these two types of cells, inhibitory cells in the piriform cortex can be found across all layers.[33] In response to pyramidal cell excitation, inhibitory cells inhibit pyramidal cell firing to temper and avoid runaway excitation in response to odor.[9] Together, the activation of excitatory cells and their inhibition by inhibitory cells shape odor representation in the piriform cortex.[9]

Overall, depending on where a cell lies in the layers of the piriform cortex, its role varies, which leads to further variability in the cell's plasticity. Shallower (layer 1a/1b) cells tend to be less plastic and act more as transduc-

ers of signals, whereas deeper cells (layer 2/3) tend to be more plastic and influenced by different scents.[7] More plastic cells develop different gene expression patterns with learning, and the converse is true for less plastic cells.[26]

In approaching olfactory learning from a broader biological perspective, I aim to understand how olfactory experiences influence gene expression at the cell-type level, leading to functional changes in specific cell-types that ultimately drive behavior. Additionally, I investigate how behavioral states reciprocally impact cellular function and gene regulation.

1.1.2 Single-cell RNA Sequencing (scRNA-seq)

(Bulk) RNA sequencing (RNA-seq) is a method that enables high-throughput sequencing of the RNA transcriptome of a population of cells.[17] Since its inception,[35] it's become a standard method that's been used to make advances in everything from agriculture to cancer biology.[34] The typical RNA-seq procedure involves collecting a sample of cells, which then undergo RNA isolation as a whole sample, destroying any granularity in the data beyond the sample level (i.e., cell-types cannot be distinguished from a traditional RNA-seq analysis).[4] This RNA isolate is then sequenced and aligned, quantifying the expression of mRNA within the cell sample.[4]

Single Cell RNA-sequencing is a newer type of RNA-seq that follows a similar procedure to measure gene expression, but scRNA-seq preserves the single-cell granularity within a potentially diverse group of cells within a sample.[14] As a result, scRNA-seq can be used to distinguish cell-types within heterogeneous samples,[8] which opens the door for numerous other routes of downstream analysis, including cell-type-specific analyses and more computationally intensive analyses.

1.1.3 Deep Learning (DL)

DL is the backbone of the recent "AI Boom" and is responsible for technology like ChatGPT and Alpha Fold. As a subset of machine learning, DL is particularly effective in analyzing and processing large amounts of data. DL is built on the premise of trying to (loosely) mimic the processing capability of neurons that exist in our brains, such that DL models can be collectively referred to as neural networks (NN). NNs are fundamentally built upon the premise that large collections of nodes exist in layers (analogous to neurons), which each feed forward to the next layer, relaying a signal that has been modified in some calculated manner. After being forwarded through a number of layers, the signal that started out as a raw input vector and has fully propagated through the network can be used as a classification or regression signal. Due to its wide applicability from language to images to even molecules, DL has become an increasingly popular tool for synthesizing biological data like magnetic resonance imaging scans, RNA-seq data,[6] and even scRNA-seq data.[8]

Along with NNs' ability to synthesize large datasets, the problem of interpretability also arises: on what basis is an NN making its decisions? How is the model making sense of the data? NNs are also known to be black-box models: models that we cannot reason about without specific interpretation methods, due to the complex patterns they learn and the millions of parameters they're built on. This has driven the development of model interpretation algorithms like LIME, SmoothGrad and SHaply Additive exPlanations (SHAP).[19] For the purposes of this study, I utilize SHAP, which is a model-agnostic algorithm based in game-theory that calculates the importance of each feature for a given prediction.[22]

1.2 Related Works

Within the realm of current research being done on the piriform cortex, the majority of existing literature relies on wet lab techniques and in-vivo animal models.[5;37;18] Not much has been done with deep learning, especially. Outside of the realm of piriform cortex and olfactory learning research, however, there are a couple DL-related works that I draw inspiration from:[24]

Deep generative modeling for single-cell transcriptomics:[20]
In this paper, Lopez et al. construct an NN that is able to impute single-cell sequencing data on the order of hundreds of thousands of single-cell samples. They achieve this by building out a variational autoencoder that first encodes the raw data and the batch id as a low-dimensional vector of random variables. They then feed this vector into a decoder that then batch-corrects, de-noises, or provides a reconstruction of the data, depending on the use case.

The main finding I seek to expand on in this study is the inclusion of batch id in encoding cells. While I don't seek to decode the results of our model as Lopez et al. do in their model, I do wish to incorporate cell-type information into my model, which in theory functions similarly to batch id. Specifically, I seek to recreate the inclusion of a one-hot encoding of cell-type in my raw data, emulating the strategy Lopez et al. employ.

CAMML: Multi-Label Immune Cell-Typing and Stemness Analysis for Single-Cell RNA-sequencing:[30]
Schiebout and Frost developed a multi-label classification model to predict cell-type given scRNA data taken from immune cell samples. In particular, they posit that immune cell types exist as more of a continuous spectrum as opposed to a set of discrete cell-types, and this makes single-label classification models incomplete in predicting cell-type within this context. To address this, they create a multi-label classifier that can take in a second cell-type label for each data point to provide the model with additional phenotypic context. This additional phenotypic context allows the model to learn more from each sample than a traditional single-label model would enable, hence leading to a more nuanced understanding of the data by the model.

This work is related to my project in that including a second label as a means to add phenotypic context to inform the model's training is something I seek to replicate. If I were to adapt this approach to my project, my application would be slightly different in that my two labels wouldn't exist in the same domain(i.e., both labels are not cell-types). Instead, I would use a second label to ensure that the model learns cell-type information in addition to gene expression patterns that result from different condition exposures. Given the cell-type-specific nature of the analysis in this project, the model must retain cell-type information in addition to the primary condition information in order for any interpretive analysis to be biologically meaningful.

Comprehensive SHAP Values and Single-Cell Sequencing Technology Reveal Key Cell Clusters in Bovine Skeletal Muscle:[11]
Guo et al. utilize Random Forest models trained on bovine and porcine skeletal muscle single-cell transcriptome data to make predictions about 14 possible cell-types. They then employ SHAP analysis to extract the top genes on a cell-type basis, which they then further analyze to identify critical genes in muscle growth and development across cell-types.

This workflow is very similar to the pipeline I hope to build out in my own study, only I hope to apply my pipeline to a mouse scRNA dataset and use NNs as opposed to random forest models. Moreover, Guo et al. do not seem to take into account cell-type when architecting their models, which is something I wish to do in this study.

1.3 Motivation

Upon surveying the current literature, my proposed pipeline and application is novel. In the piriform cortex literature space, there exists a lack of research that utilizes computational methods, specifically machine and deep learning-based methods, to analyze and produce new insights. And from a computational biology perspective, I seek to fill two gaps. (1) Most DL pipelines that are applied in the context of scRNA-seq datasets seek to impute, annotate, or clean data[24]as opposed to using DL to directly uncover biological insights, especially via model interpretation. (2) The second gap is that not many current models unify both cell-type information and a separate predictive category as a means of making cell-type-informed predictions.

As such, I fill these gaps by constructing a multi-label model that intakes scRNA data, sourced from mouse piriform cortex neurons, and makes predictions about cell-type and olfactory experience (i.e. what scent a mouse has learned). I then seek to use this model and interpretive methods like SHAP to uncover insights into genes that might play a significant role in how olfactory experiences translate to functional and behavioral changes in a cell-type specific manner.

2 Methods

2.1 Dataset Curation

In collaboration with the Fleischmann Lab in the Department of Neuroscience, I received an scRNA-seq dataset comprising 103,987 samples, each with expression data for 17,089 genes. These samples were harvested from mouse (mus musculus) piriform cortices after in-vivo experiments wherein the mice were exposed to either no olfactory stimulants (bored), sucrose (positive valence), quinine (negative valence), or a naive stimulant (no valence). Mice that were exposed to any of sucrose, quinine, or naive were exposed 5 times to promote learning before a final exposure, after which cells were harvested either 30 minutes post-exposure or 90-minutes post-exposure (figure 1). In total, the conditions resulting from these experiments were as follows: bored (no exposure, no learning), naive30, naive90, sucrose30, sucrose90, quinine30, quinine90.

After the dataset was harvested, the Fleischmann Lab then carried out further analysis to harmonize, cluster and annotate the dataset. The cell-types included in the dataset are as follows: IN_CGE-NPY, IN_CGE-VIP, IN_MGE-Pvalb, IN_MGE-SST, Pyrlayer2a/b, Pyrlayer3, SL1, SL2, Vglut2. Cell-types that begin with 'IN' are considered to be inhibitory-type cells, SL refers to semilunar, and cells that begin with Pyrlayer refer to pyramidal cells. Further studies were done to identify immediate early genes (IEGs), which are most likely to be at the peak of their expression 30-minutes post-exposure, and late response genes (LRGs), which are most likely to be at the peak of their expression 90-minutes post-exposure, within both the dataset and current literature.

2.2 Preprocessing

In this project, the primary goal was to study olfactory *learning*, so I focused on the naive, quinine and sucrose groups. As such, I took a subset of all naive, sucrose, and quinine-conditioned cells in the dataset, yielding a total of $\sim 61,000$ effective samples. From here, given that the original feature space of the dataset is $\sim$ 17000 genes large, I sought to reduce this space to simplify the training process and also enable clearer interpretation of the models downstream. As such, from this

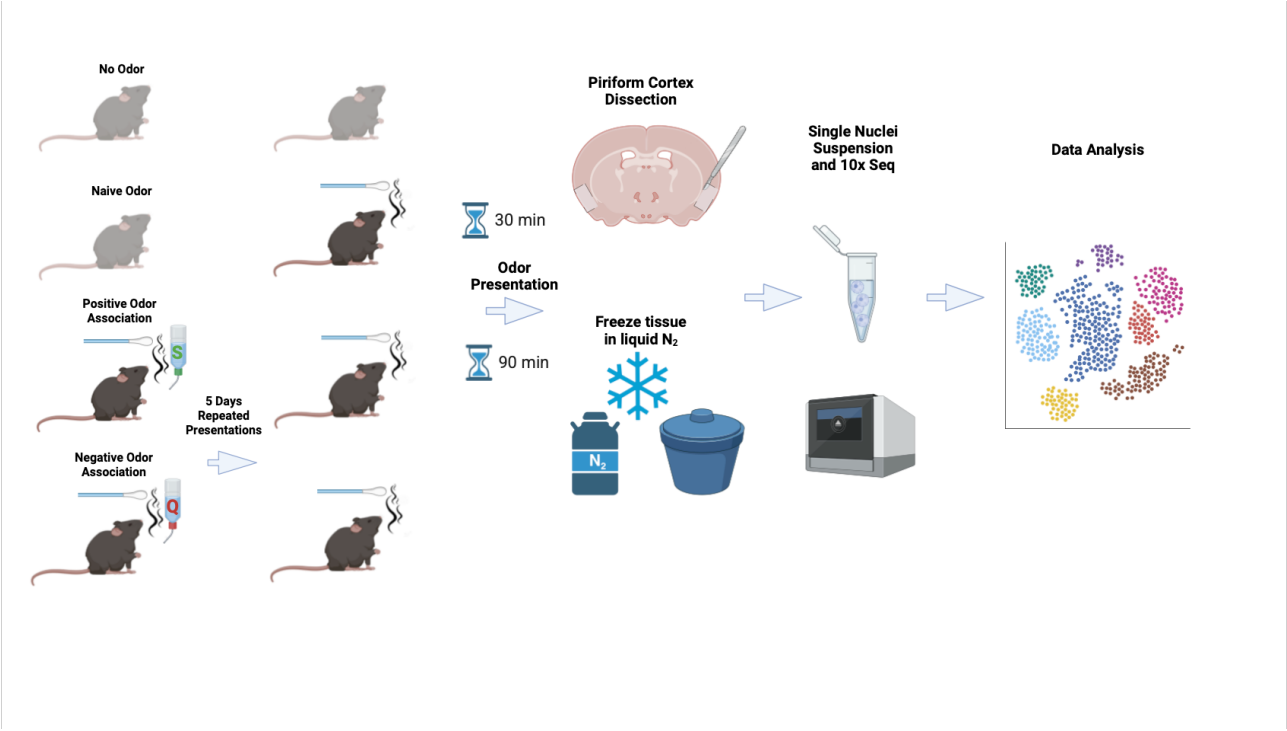


Figure 1: this graphic visually shows the in-vivo procedure done to facilitate learning in the mice neuron cells harvested. Figure made by Keeley Baker using BioRender.

subset of all naive, sucrose and quinine cell samples, I generated the top 3000 most variable genes, excluding mitochondrial genes and genes that do not have a canonical name, and reduced the feature space to these 3000 genes. For the labels of this dataset, I designated the condition (either naive, sucrose, or quinine) as the label. In choosing condition as a label, I sought to mimic the mice' learning process in having the model predict the condition cells were exposed to. From here, the dataset was then split into a training-test split of 80/20, stratified across the conditions. I saved all data (training and test sets) as pickle files to be used in later analyses.

2.2.1 Feature Selection

In designing this experiment, I separated the 30 and 90 minute models, since the patterns learned from the 30 minute data might interfere with patterns learned from the 90 minute data and vice-versa. This potential interference would likely lead to inaccurate or ineffective interpretation results. As such, in order to compare the 30 and 90 minute models effectively, as well as their results, it was important to have them share the same feature space. To ensure that I was building my model on impactful features for both 30 and 90 minute models, I conducted an overlap analysis across the top 3000 variable genes from just the 30 minute naive, quinine and sucrose cells (n/q/s); the top 3000 variable genes from just the 90 minute n/q/s cells; and the top 3000 variable genes from the set of set of all 30 and 90 minute n/q/s (figure 2). From this study, I found that generating the top 3000 variable genes from the collection of all 30 and 90 minute n/q/s cells was sufficient in creating a feature space that was at least 2/3 highly variable across both the 30 and 90 minute time intervals.

2.3 Models

The basic architecture of all models in this project is a multilayer perceptron (MLP). MLPs are the most basic form of neural network, where the basic building block is a dense

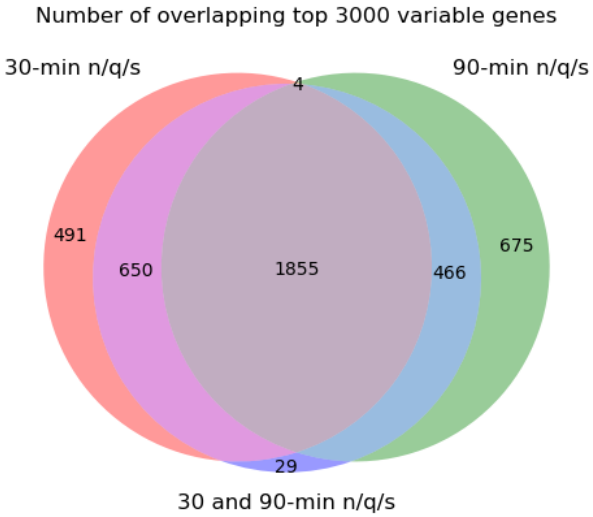


Figure 2: 30-minute naive, quinine and sucrose (n/q/s) samples were subsetted out and the top 3000 most variable genes were generated. The same was done with the 90-minute (n/q/s) samples and the subset of both 30 minute and 90-minute n/q/s samples. This venn diagram was then generated to show how many genes overlap between each subset. 1855 genes are shared between all three groups, and there are only 29 genes that are in the 30 and 90-min n/q/s subset that are not shared with any of the other subsets, making it the ideal candidate for our feature space.

layer that comprises a number of nodes that each take in an input and feed it into every node in the next layer. All models were trained using Tensorflow 2.18.0[1], ADAM optimizer and a sparse categorical entropy loss function[15].

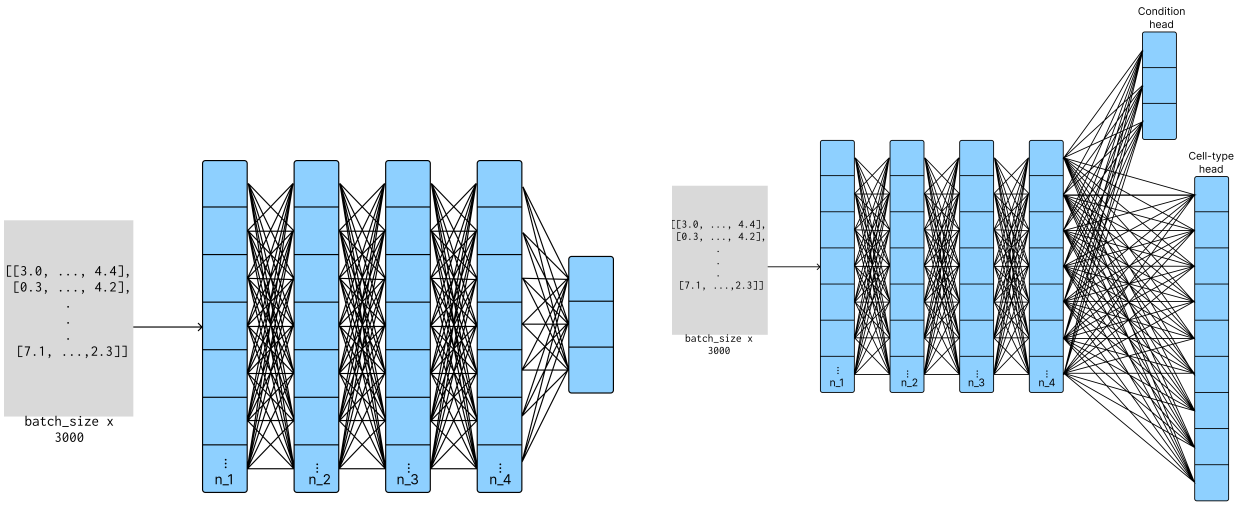
2.3.1 Cell-type oblivious model

As a baseline, I first created 30min and 90min MLPs that were oblivious to any explicit cell-type information, outside of what was already implicitly encoded into the raw data. I constructed both models as 4-layer perceptrons. Each layer in each model comprised a dense layer followed by a rectified linear unit layer (ReLU)[2]. These dense layers were then followed by a final dense layer with 3 units, representing the three conditions, followed by a softmax[28] activation function. In training these models, I fed in the preprocessed data as above, and trained for 10 epochs at a learning rate of 0.0003.

1. 30min model architecture (figure 3a): Dense(1024)→ReLU→Dense(512)→ ReLU→Dense(640)→ReLU→ Dense(256)→ReLU→Dense(3)→Softmax
2. 90min model architecture (figure 3a): Dense(1024)→ReLU→Dense(512)→ ReLU→Dense(896)→ReLU→ Dense(128)→ReLU→Dense(3)→Softmax

2.3.2 18-model cohort

To instill cell-type information into my models, I then opted to create a cohort of 18 4-layer MLPs for each cell type in each timepoint. This resulted in models for the following subgroups: 30minIN_CGE-NPY, 30minIN_CGE-VIP, 30minIN_MGE-Pvalb, 30minIN_MGE-SST, 30minPyr-layer2a/b, 30minPyrlayer3, 30minSL1, 30minSL2, 30minVglut2, 90minIN_CGE-NPY, 90minIN_CGE-VIP, 90minIN_MGE-Pvalb, 90minIN_MGE-SST, 90minPyr-layer2a/b, 90minPyrlayer3, 90minSL1, 90minSL2, and 90minVglut2. For each of these models, I took the preprocessed data prior to splitting into training and test datasets and segmented out each cell-type and timepoint to create the same time/cell-type combinations listed above. For each sub-



(a) Single-label model **(b)** Double-label model

Figure 3: Visual representation of the single-label and double-label models. n_1 is a dimension between 1024 and 2048, n_2 and n_3 are dimensions between 512 and 1024, and n_4 is a dimension between 128 and 256.

set, I then performed an 80/20 train-test split and trained. Each model comprised four layers in the same configuration as the baseline model with varying numbers of output units in each Dense layer (figure 3a). In training this model, I trained for 10 epochs at a learning rate of 0.0003 as well.

2.3.3 Multi-label model

To reconcile the baseline model and 18-model cohort, I then forced the model to predict both condition and cell-type. During preprocessing, in addition to the condition as a label, I then added cell-type as an additional label. When architecting my model, I used the same 4-layer perceptron architecture as in the base model, except when defining the terminal classification layers, I added both a condition head that comprised a Dense layer with 3 units feeding into a final softmax function in addition to a cell-type head that comprised a Dense layer with 9 units feeding into a final softmax function (figure 3b). I then also designed a total loss function for this model, weighting the loss from the condition head with a weight of 0.9 and the loss from the cell-type head with a weight of 0.1. In training this model, I trained for 10 epochs at a learning rate of 0.0003.

2.4 Fine-tuning

Each model was then further fine-tuned using the KerasTuner package[29], and each model was fine-tuned over 10 trials. For each model trained, the number of output units in each layer in the network was fine-tuned. The number of output units in first dense layer (applicable to all models in the project) was tuned to a value between 1024 and 2048 with the tuner stepping in intervals of 512, the second dense layer was tuned to a value between 512 and 1024 with intervals of 128, the third dense layer was tuned to a value between 512 and 1024 with steps of 128, and the fourth dense layer was tuned to a value between 128 and 256 with steps of 128. Other fine-tuned variables included the batch size, which was at minimum 32 and at maximum 256 with tuning intervals of 32. For all hypertuning experiments, in the base model and the 18-model cohort, validation accuracy was used as the tuning metric, and in the multi-label model, total loss was used as the tuning metric.

2.5 SHAP Analysis

All SHAP analysis was performed using version 0.46.0 of the SHAP library[23]. I used the SHAP library's DeepExplainer object to explain the model, and I instantiated this object with a background context of $max(1000, len(X_{train}))$ randomly selected data points from the training set. I then randomly selected $max(300, len(X_{test}))$ test samples from the test set to have the explainer produce explanations for. After generating the SHAP values, they were then saved in pickle files to be used in visualization and further experiments.

3 Methods

3.1 Multi-label model can accurately predict conditions and cell-types

I measured the performance of each of the base, 18-cohort, and multi-label models using testing accuracy and compared them against the base MLP classifier predicting condition only. Overall, the baseline model performed the best on the task of purely predicting conditions with 80% and 88% accuracy in the 30 and 90-minute time points, respectively. Comparatively, the multi-label model performed almost as well as the base model with 30 and 90-minute condition accuracies of 77% and 88%. On the task of predicting cell-type, the multi-label models also performed extremely well at 88% and 91% in the 30 and 90-minute timepoints. Lastly, the 18-model cohort performed with varying accuracy across cell-types, ranging from 60% to 88%. In the 18-model cohort, there is a 0.5090 Pearson correlation between accuracy and sample training-set size. Overall, there is a general trend in that models trained on the 90-minute data are more accurate (Table 1).

3.2 UMAP shows simultaneous cell-type and condition embeddings being learned

I visualized the UMAP of the output of the fourth dense layer (penultimate layer, right before classification occurs) of both the base model and multi-label model to see how well the model had learned both cell-type and condition information. I then also calculated the silhouette scores for both UMAPs; the cell-type silhouette score was based solely on the nine cell-type clusters. For condition, I calculated the silhouette score based on the three condition clusters in addition to a silhouette score calculated on a collection of 18 unique clusters associated with each possible (cell-type, condition) tuple. In choosing to calculate the silhouette scores in this manner, I noticed that a given cell-type cluster in the cell-type UMAP would be broken up into three clusters in the analogous condition UMAP, with each cluster representing one of the three conditions within the same cell-type. As a result, I reasoned that each (cell-type, condition) tuple exists as a discrete cluster in the condition UMAP (Table 2).

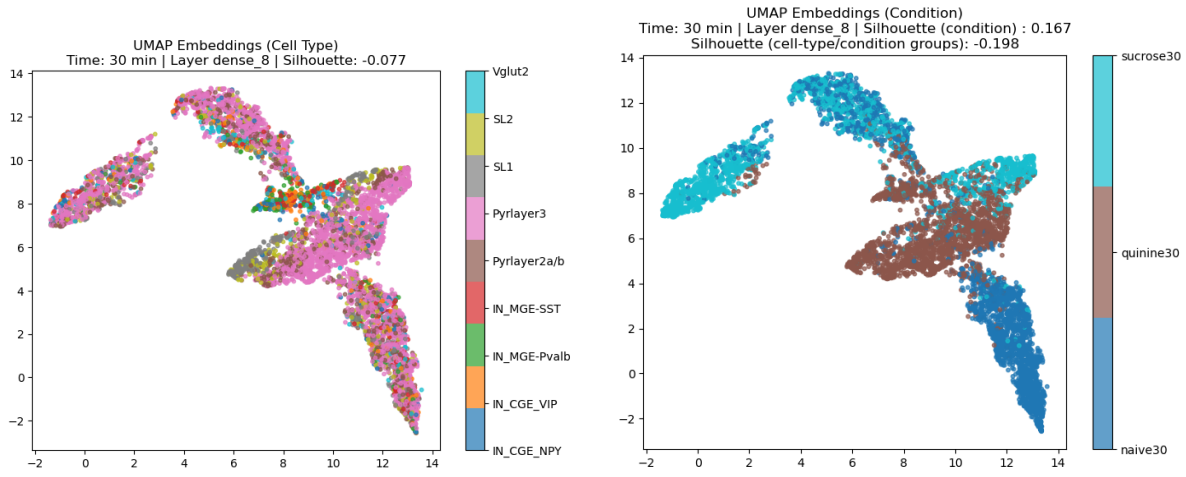
As expected, the base model without any explicit cell-type influence entirely lost any understanding of a cell-type embedding in the UMAP representation. This is evident by the approximate -0.1 silhouette score for both 30-minute and 90-minute base model cell-type UMAPs. Conversely, the condition UMAP is highly clustered in both the 30-minute and 90-minute base models, with silhouette scores over 0.150 in both models.

In the multi-label models, I observed that there is a fine balance between the silhouette scores for the cell-type and condition UMAPs. In both 30 and 90-minute multi-label cell-type UMAPs, there are visually observable discrete

clusters, despite the modest ~ 0.05 silhouette score, which indicate that the model is able to maintain an accurate representation of cell-type information while simultaneously learning a strong condition embedding, evident by high condition silhouette scores and condition accuracy. Finally, the cell-type/condition silhouette scores also remain positive, indicating that the cells of the same (cell-type, condition) tuple exist in loosely defined clusters in the condition UMAPs.

Model	Condition Accuracy	Cell-type Accuracy
Base model 30-min	0.8032	NA
Base model 90-min	0.8824	NA
IN_CGE-NPY 30-min	0.7071	NA
IN_CGE-NPY 90-min	0.7387	NA
IN_CGE-VIP 30-min	0.6919	NA
IN_CGE-VIP 90-min	0.6997	NA
IN_MGE-Pvalb 30-min	0.7082	NA
IN_MGE-Pvalb 90-min	0.7463	NA
IN_MGE-SST 30-min	0.6083	NA
IN_MGE-SST 90-min	0.6771	NA
Pyrlayer2a/b 30-min	0.6463	NA
Pyrlayer2a/b 90-min	0.7867	NA
Pyrlayer3 30-min	0.7586	NA
Pyrlayer3 90-min	0.8839	NA
SL1 30-min	0.7070	NA
SL1 90-min	0.7860	NA
SL2 30-min	0.6497	NA
SL2 90-min	0.7667	NA
Vglut2 30-min	0.6078	NA
Vglut2 90-min	0.7500	NA
Multi-label 30-minutes	0.7728	0.8805
Multi-label 90-minutes	0.8786	0.9130

Table 1: Condition and cell-type accuracy for all evaluated models.



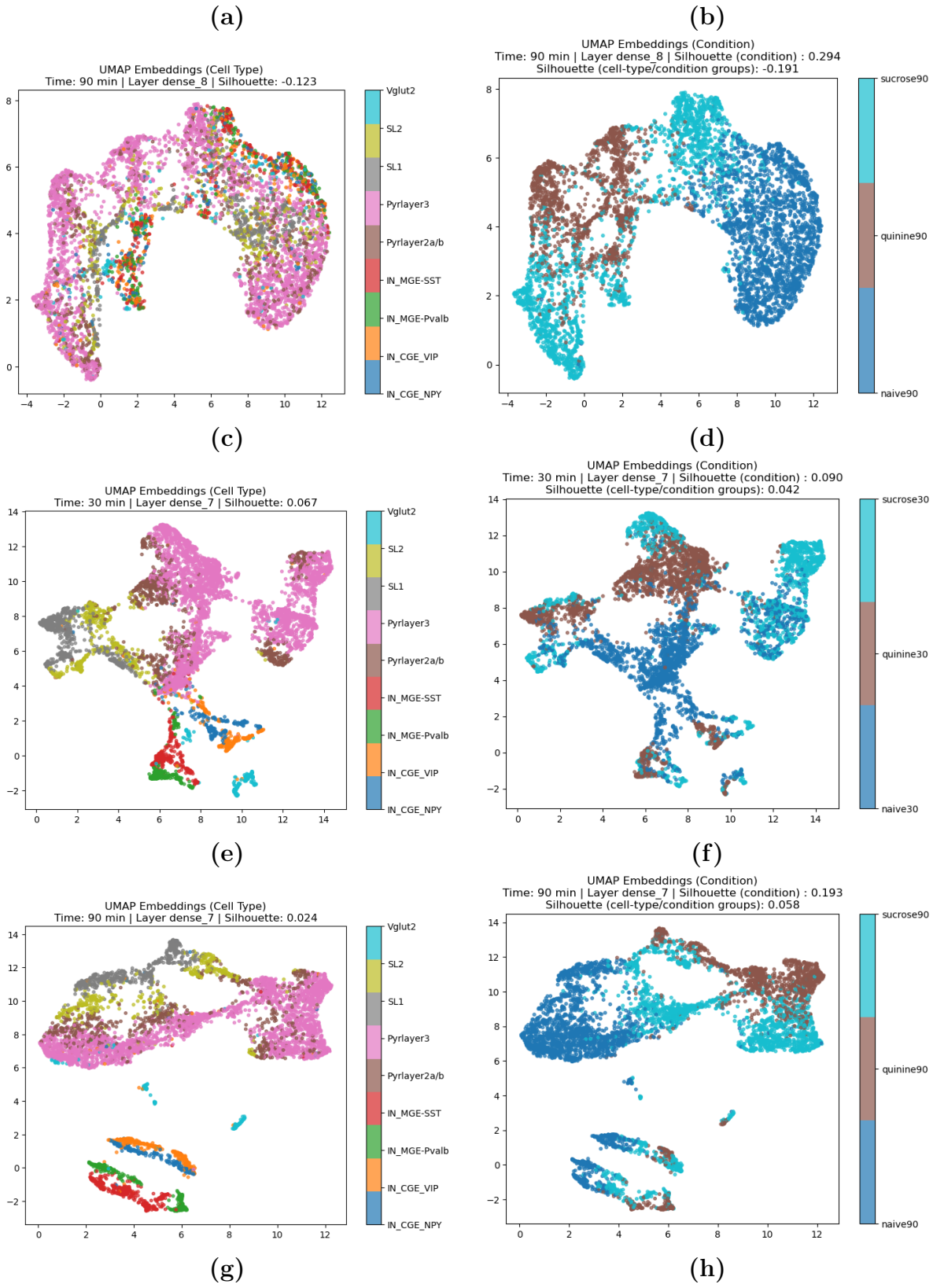
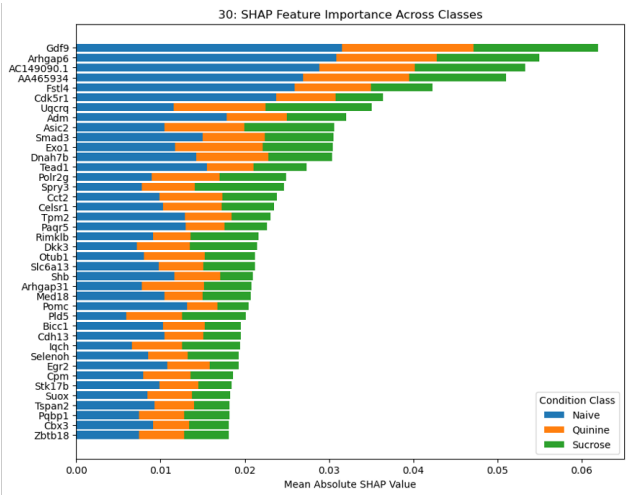
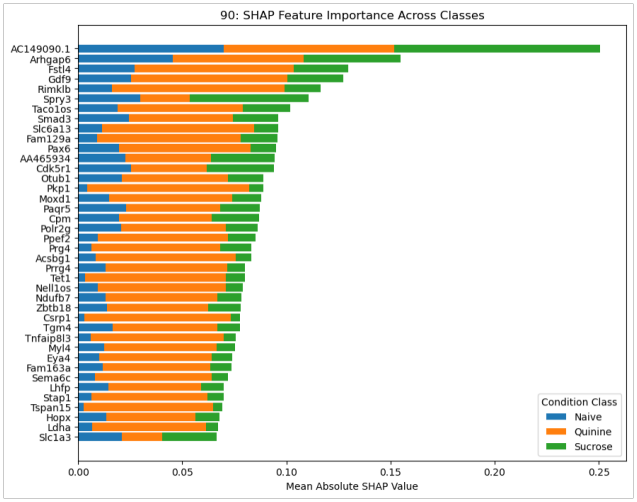


Table 2: UMAP visualizations of the penultimate Dense layer in all 30 and 90-min base and multi-label models before data is fed into the classification heads. Cell-type UMAPs are in the left column, and Condition UMAPs are in the right column. (a) and (b) are from the 30-minute base model, (c) and (d) are from the 90-minute base model, (e) and (f) are from the 30-minute multi-label model, and (g) and (h) are from the 90-minute multi-label model



(a) The top 40 most influential features in the 30-min multi-label model.



(b) The top 40 most influential features in the 90-min multi-label model.

Figure 4: SHAP Plots show the amount of influence each feature has in the models' decision making.

3.3 SHAP Analysis identifies potential genes of interest in the models' decision-making

After training my multi-label model, I performed SHAP analysis on both 30-minute and 90-minute models to generate the top 40 most influential genes across all cell-types in the model's decision making when predicting condition (figure 4). Across the top 40 genes from both 30 and 90-minute models, there is only an overlap of 15 out of 40 genes, suggesting that there is a distinction between what the model is learning at both time points. I then performed gene ontology (GO) profiling on the two lists provided and received no significant results[16].

Despite potential limitations that I will discuss below, my model consistently identifies AC149090.1 as a top 5 important gene in the 30-minute model and the most important gene in the 90-minute model. Though significantly more research is needed to understand what the role of AC149090.1 in the piriform cortex is, in literature, it has been implicated in biological clocks and aging[3]. Another example of a potentially significant gene is Fstl4, which was identified as a top 5 influential gene in both 30 and 90-minute multi-label models. In literature, Fstl4's mouse ortholog, Spig1 has been implicated as a regulator of brain-derived neurotrophic factor-driven long term potentiation and extinction of inhibitory avoidance memory[32].

4 Discussion

In this study, I developed a multi-label classifier to successfully combine the task of cell-type prediction and condition prediction into a single model that performs with high accuracy. I then used UMAP to visualize the embedding space of the model before predictions were made, and UMAP showed distinct embeddings on both a cell-type and condition level. This allowed me to capture condition embeddings in a cell-type specific manner. I then utilized this model to perform SHAP analysis to reveal potentially biologically significant genes that might play a role in how behavior changes with learning.

4.1 Biological Significance and Applications

These findings are significant in two ways. (1) By creating a model that predicts both cell-type and condition, I show that a multi-label model is an effective way to ensure that cell-type embeddings are maintained throughout training, even when a single-label model traditionally loses such embeddings. Given the rise of scRNA-seq experiments, this type of model paradigm, combining cell-type with an additional task, might be increasingly useful for cell-type-specific analysis. Furthermore, through my model's accuracy in comparison to a standard single-label base model, I show that my model can perform well while still maintaining cell-type embeddings in its training.

(2) Secondly, by using such a model to generate SHAP values and further investigate the top most influential genes in the model's decision making, I have found genes that are potentially, and still unknown to be, significant in olfactory learning. Though performing GO profiling on the gene lists generated yielded no results, I speculate that the top features that the model is using to make decisions may not be very interrelated, resulting in a lack of GO hits. Furthermore, there are many non-canonical genes like AC149090.1 and AA465934 that the model seems to be dependent on that are not even found in the GO profiler, which may require further literature search to reason about. Overall, while I only present an extremely preliminary literature search on two influential genes in my results, many of the genes identified in the SHAP analysis need further vetting to understand their relation to the piriform cortex and olfactory learning. SHAP analysis, therefore, has the potential to be useful in finding previously non-canonical genes in many different applications, especially olfactory learning.

4.2 Limitations

There are a couple limitations with this study: namely dataset balance and SHAP analysis methodology.

4.2.1 Dataset Balance

With regards to dataset balance, while the dataset is equally balanced across the 30 and 90 minute time points, the balance between each cell-type is quite skewed. In the dataset, there are significantly more pyramidal 3 cells than any other cell-type in both the 30 and 90-minute time points (supplementary figure 1 shows this disparity in the training set). Especially when training the 18-model cohort, this presented many difficulties in creating equally reliable models to be used in further SHAP analysis. The fact that there is a ~ 0.5 positive Pearson correlation between training set size and model accuracy in the 18-model cohort is concerning. Furthermore, the cell-type disparity could also cause present issues with the multi-label model as well. Especially

when distinguishing between pyramidal 3 and pyramidal 2a/b cells, the model might be inclined to predict a cell as pyramidal 3, simply because of this imbalance, skewing the embedding learned.

4.2.2 SHAP Analysis Methodology

While it's general practice to randomly select a small random sample, in my case 1000 samples, from the training data to serve as the background data for the SHAP explainer, in addition to a further select number of random test samples to have the explainer explain, this has led to some instability in my SHAP results. Upon generating SHAP values across multiple trials with the same trained model, the model produced different results every couple trials. I speculate that this is likely a combination of the dataset being skewed by cell-type, so it may be the case that data needs to be sampled in a more stratified or uniform manner in future SHAP experiments. As a result of this instability, however, some of my findings that are built on my SHAP analysis pipeline have been moved to my supplementary figures section. All SHAP analysis results that I have included in my results section have proven to be stable across multiple trials of SHAP analysis.

In supplementary figure 2, I created an overlap matrix of the top 40 genes in each time and cell-type, along with a dendrogram (supplementary figure 3) that shows how related two cell-types might be with one another. In generating this figure, I first initiated my SHAP explainer with a random 1000 cells from my testing set. I then fed in a maximum of 300 cells of each cell type to generate my SHAP values, and then extracted the top 40 most influential genes from the SHAP values. From there, I took the intersection between each possible (time, cell-type) tuple combination and plotted their rate of overlap. To generate supplementary figure 3, I calculated the distance/dissimilarity by performing 1 - rate of overlap, and then I plotted the distances in a dendrogram.

Had my SHAP analysis been stable, I would have expected to have seen higher overlap between the 30SL1/2 and 90SL1/2 groups, based on their expected lower plasticity as semilunar cells. I would have also expected to have seen the 30SL1 group be more related to 30SL2, but instead the 30SL1 group was found to be most related to 90Pyrlayer3 in the dendrogram, which is unexpected and may point to instability in the SHAP results. These are just a few of the findings that I found to be unstable and counter to my hypothesis.

4.3 Future Work

At its core, my project is aimed at providing insights into genes that might be of interest in the context of olfactory learning. After reevaluating my SHAP analysis methodology, the next step after discovering that certain genes might be implicated in olfactory learning would be to conduct wet-lab studies, either in-vitro or in-vivo, to empirically study their behavior and function in live models.

References

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: A system for large-scale machine learning, 2016.

[2] Abien Fred Agarap. Deep learning using rectified linear units (ReLU). 2018.

[3] Matthew T Buckley, Eric D Sun, Benson M George, Ling Liu, Nicholas Schaum, Lucy Xu, Jaime M Reyes, Margaret A Goodell, Irving L Weissman, Tony Wyss-Coray, Thomas A Rando, and Anne Brunet. Cell-type-specific aging clocks to quantify aging and rejuvenation in neurogenic regions of the brain. *Nature Aging*, 3(1):121–137, January 2023.

[4] A. Chatterjee, A. Ahn, E. J. Rodger, P. A. Stockwell, and M. R. Eccles. A guide for designing and analyzing rna-seq data. In N. Raghavachari and N. Garcia-Reyero, editors, *Gene Expression Analysis*, volume 1783 of *Methods in Molecular Biology*, pages 39–59. Humana Press, New York, NY, 2018.

[5] Heming Cheng, Yi Wang, Junzi Chen, and Zhong Chen. The piriform cortex in epilepsy: What we learn from the kindling model. *Exp. Neurol.*, 324(113137):113137, February 2020.

[6] Travers Ching, Daniel S Himmelstein, Brett K Beaulieu-Jones, Alexandr A Kalinin, Brian T Do, Gregory P Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M Hoffman, Wei Xie, Gail L Rosen, Benjamin J Lengerich, Johnny Israeli, Jack Lanchantin, Stephen Woloszynek, Anne E Carpenter, Avanti Shrikumar, Jinbo Xu, Evan M Cofer, Christopher A Lavender, Srinivas C Turaga, Amr M Alexandari, Zhiyong Lu, David J Harris, Dave DeCaprio, Yanjun Qi, Anshul Kundaje, Yifan Peng, Laura K Wiley, Marwin H S Segler, Simina M Boca, S Joshua Swamidass, Austin Huang, Anthony Gitter, and Casey S Greene. Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface*, 15(141):20170387, April 2018.

[7] Ronald L Davis. Olfactory learning. *Neuron*, 44(1):31–48, September 2004.

[8] Nafiseh Erfanian, A Ali Heydari, Adib Miraki Feriz, Pablo Iañez, Afshin Derakhshani, Mohammad Ghasemigol, Mohsen Farahpour, Seyyed Mohammad Razavi, Saeed Nasseri, Hossein Safarpour, and Amirhossein Sahebkar. Deep learning applications in single-cell genomics and transcriptomics data analysis. *Biomed. Pharmacother.*, 165(115077):115077, September 2023.

[9] Kevin M Franks, Marco J Russo, Dara L Sosulski, Abigail A Mulligan, Steven A Siegelbaum, and Richard Axel. Recurrent circuitry dynamically shapes the activation of piriform cortex. *Neuron*, 72(1):49–56, October 2011.

[10] Jay A. Gottfried, Joel S. Winston, and Raymond J. Dolan. Dissociable codes of odor quality and odorant structure in human piriform cortex. *Neuron*, 49(3):467–479, 2006.

[11] Y. Guo, F. Ma, P. Li, L. Guo, Z. Liu, C. Huo, C. Shi, L. Zhu, M. Gu, R. Na, and W. Zhang. Comprehensive shap values and single-cell sequencing technology reveal key cell clusters in bovine skeletal muscle. *International Journal of Molecular Sciences*, 26(5):2054, 2025.

[12] Norimitsu Suzuki John M. Bekkers. Neurons and circuits for odor processing in the piriform cortex. *Trends in Neurosciences*, 36:429–438, 7 2013.

[13] David M. Johnson, Kevin R. Illig, Mark Behan, and Lewis B. Haberly. New features of connectivity in piriform cortex visualized by intracellular injection of pyramidal cells suggest that "primary" olfactory cortex functions like "association" cortex in other sensory systems. *Journal of Neuroscience*, 20(18):6974–6982, September 2000.

[14] Y. Kashima, Y. Sakamoto, K. Kaneko, et al. Single-cell sequencing techniques from individual to multiomics analyses. *Exp Mol Med*, 52(9):1419–1427, 2020.

[15] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. 2014.

[16] Liis Kolberg, Uku Raudvere, Ivan Kuzmin, Priit Adler, Jaak Vilo, and Hedi Peterson. g:profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.*, 51(W1):W207–W212, July 2023.

[17] K. R. Kukurba and S. B. Montgomery. Rna sequencing and analysis. *Cold Spring Harbor Protocols*, 2015(11):951–969, April 2015.

[18] Amit Kumar, Edi Barkai, and Jackie Schiller. Plasticity of olfactory bulb inputs mediated by dendritic nmda-spikes in rodent piriform cortex. *eLife*, 10:e70383, 2021.

[19] X. Li, H. Xiong, X. Li, L. Zhang, D. Zhang, and C. Chen. Interpretable deep learning: interpretation, interpretability, trustworthiness, and beyond. *Knowledge and Information Systems*, 64(12):3197–3234, 2022.

[20] R. Lopez, J. Regier, M.B. Cole, et al. Deep generative modeling for single-cell transcriptomics. *Nat Methods*, 15:1053–1058, Nov 2018.

[21] Wolfgang Löscher and Ulrich Ebert. The role of the piriform cortex in kindling. *Progress in Neurobiology*, 50(5–6):427–481, 1996.

[22] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions. 2017.

[23] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[24] Brendel M, Su C, Bai Z, Zhang H, Elemento O, and Wang F. Application of deep learning on single-cell rna sequencing data analysis: A review. *Genomics Proteomics Bioinformatics*, 20(5):814–835, Oct 2022.

[25] Fernando Martínez-García, Amparo Novejarque, Nuria Gutiérrez-Castellanos, and Enrique Lanuza. Piriform cortex and amygdala. In Charles Watson and George Paxinos, editors, *The Mouse Nervous System*, pages 140–172. Academic Press, 2012.

[26] Keiichiro Minatohara, Mika Akiyoshi, and Hiroyuki Okuno. Role of immediate-early genes in synaptic plasticity and neuronal ensembles underlying the memory trace. *Front. Mol. Neurosci.*, 8:78, 2015.

[27] K.R. Neville and L.B. Haberly. Olfactory cortex. In G.M. Shepherd, editor, *The Synaptic Organization of the Brain*, pages 415–454. Oxford University Press, 5 edition, 2004.

[28] Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. Activation functions: Comparison of trends in practice and research for deep learning. 2018.

[29] Tom O'Malley, Elie Bursztein, James Long, François Chollet, Haifeng Jin, Luca Invernizzi, et al. Kerastuner. https://github.com/keras-team/keras-tuner, 2019.

[30] C. Schiebout and H.R. Frost. Camml: Multi-label immune cell-typing and stemness analysis for single-cell rna-sequencing. In *Pac Symp Biocomput*, volume 27, pages 199–210, 2022.

[31] N. Suzuki and J. M. Bekkers. Neural coding by two classes of principal cells in the mouse piriform cortex. *Journal of Neuroscience*, 26(46):11938–11947, November 2006.

[32] Ryoko Suzuki, Akihiro Fujikawa, Yukio Komatsu, Kazuya Kuboyama, Naomi Tanga, and Masaharu Noda. Enhanced extinction of aversive memories in mice lacking sparc-related protein containing immunoglobulin domains 1 (spig1/fstl4). *Neurobiology of Learning and Memory*, 152:61–70, 2018.

[33] D. N. Vaughan and G. D. Jackson. The piriform cortex and human focal epilepsy. *Frontiers in Neurology*, 5:259, December 2014.

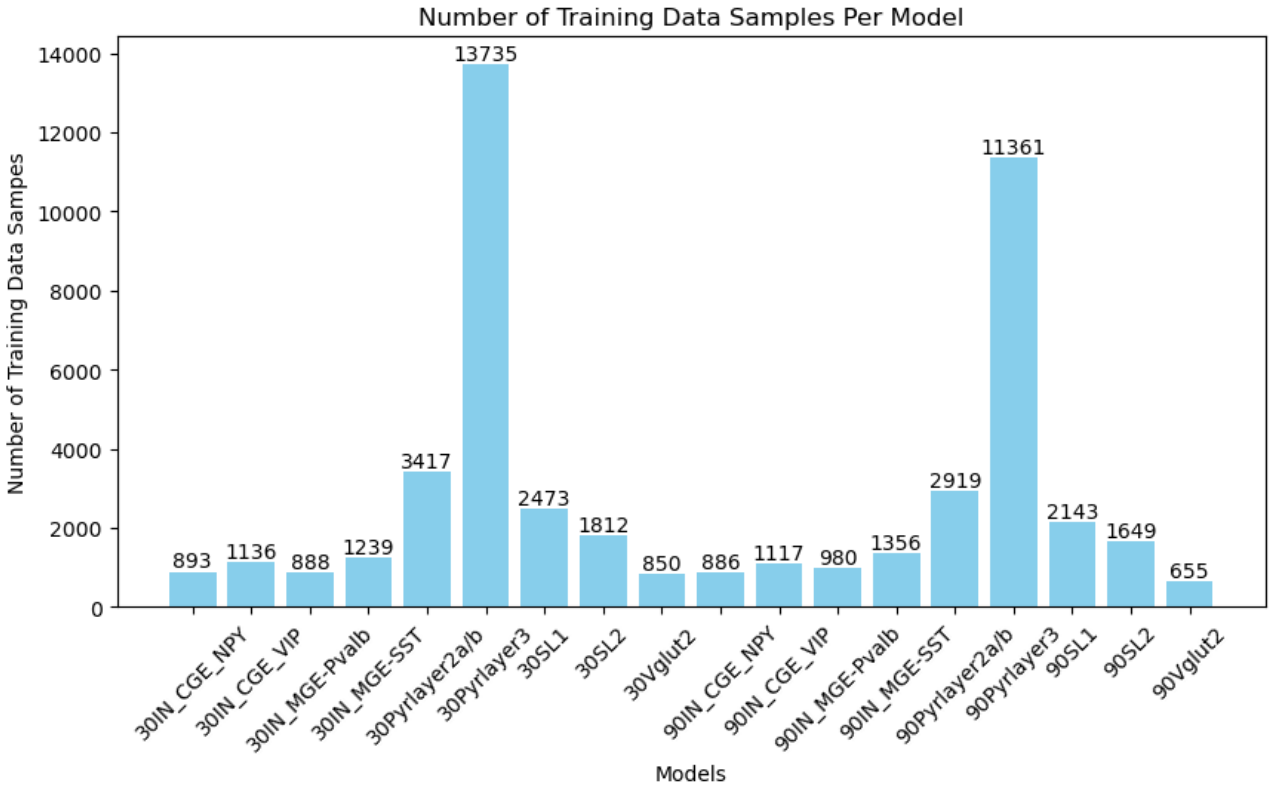
[34] Irina Vlasova-St. Louis. Introductory chapter: Applications of RNA-seq diagnostics in biology and medicine. In *Applications of RNA-Seq in Biology and Medicine*. IntechOpen, October 2021.

[35] Z. Wang, M. Gerstein, and M. Snyder. Rna-seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63, January 2009.

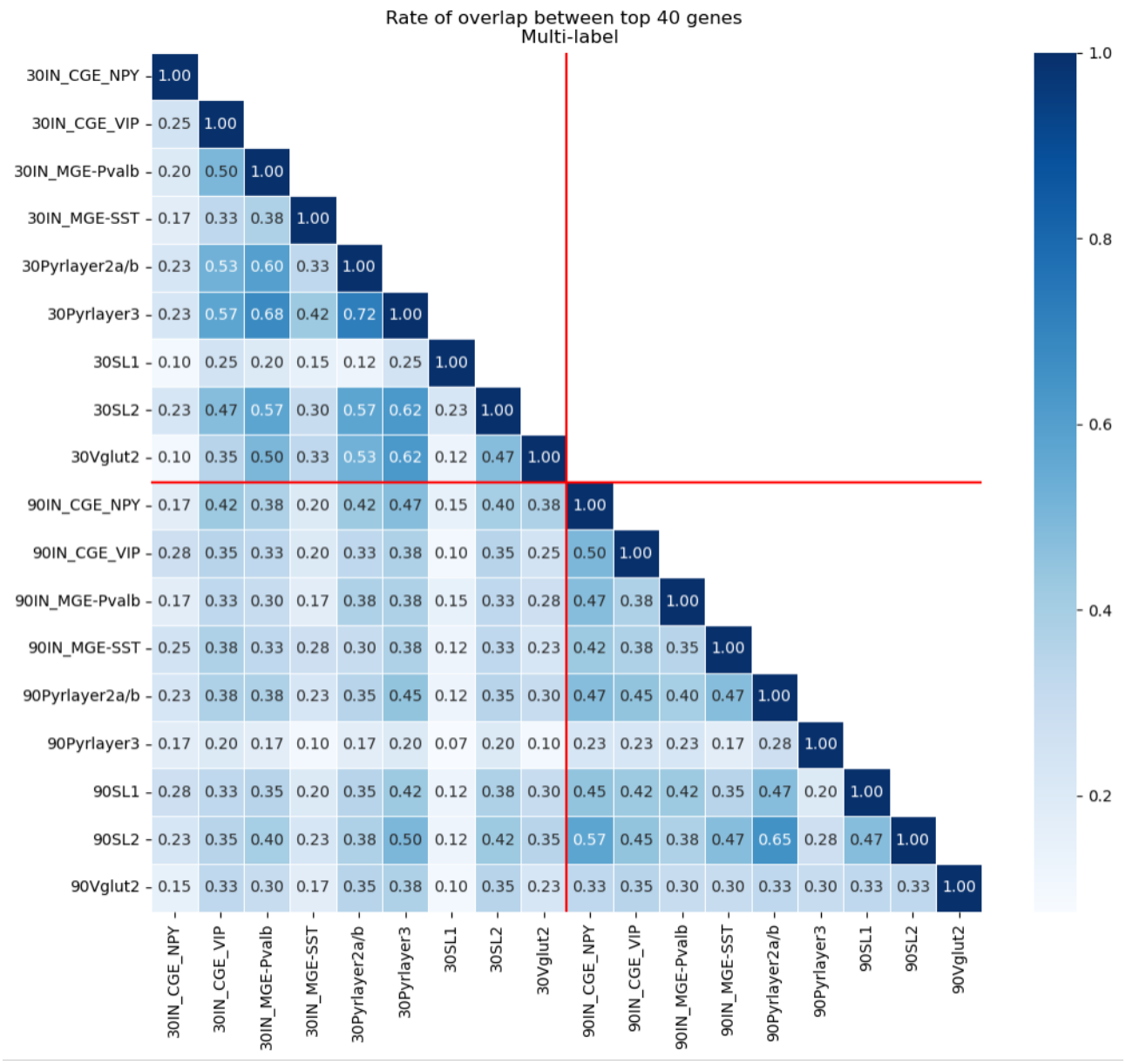
[36] D. A. Wilson. Receptive fields in the rat piriform cortex. *Chemical Senses*, 26(5):577–584, 2001.

[37] Zhang Z, Collins DC, and Maier JX. Network dynamics in the developing piriform cortex of unanesthetized rats. *Cereb Cortex*, 31(2):1334–1346, 2021.

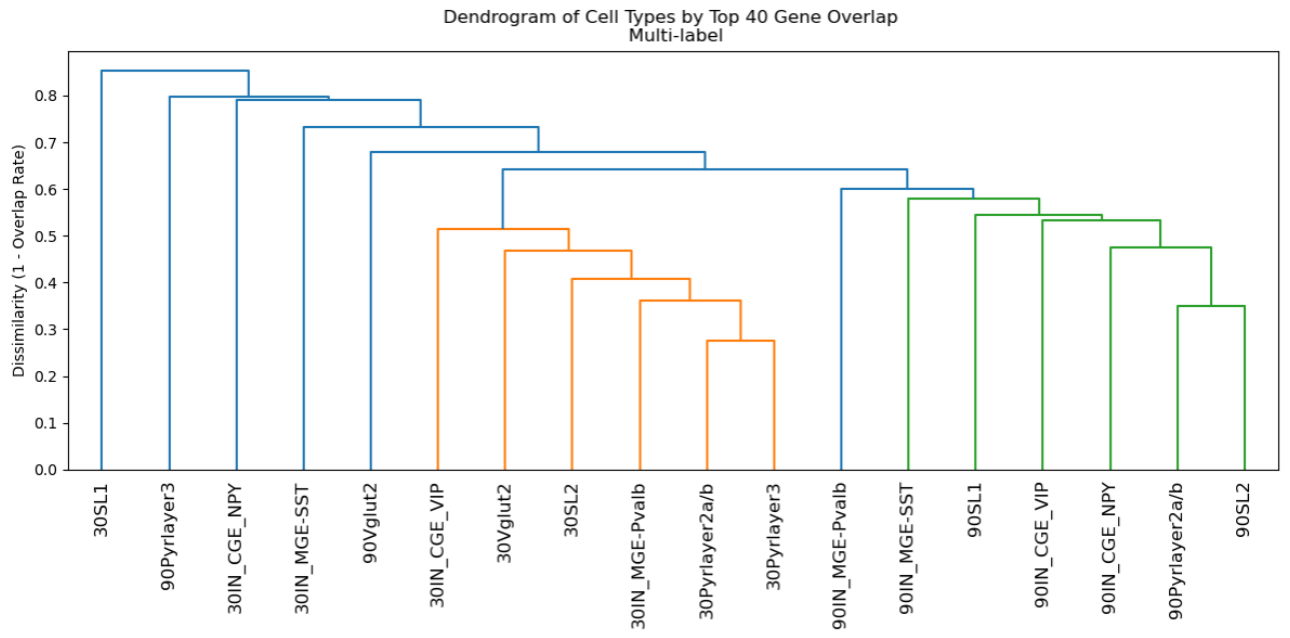
5 Appendix



Supplementary Figure 1. This histogram shows the distribution of cells across the cell-types within the training data. Training and testing data were split in a stratified manner.



Supplementary Figure 2. An overlap matrix that shows the rate of overlap between the top 40 genes of each cell-type after SHAP Analysis was done on the multi-label model.



Supplementary Figure 3. A dendrogram that shows how related cell-types are based off on the overlap matrix