Multimodal Graph Convolutional Network for Cortical Layer Classification in Human DLPFC Spatial Transcriptomics

Tomás Agustín González Orlando
Senior Software Engineer and Independent AI Researcher
Cambridge, United Kingdom
taomasgonzalez@gmail.com

***Abstract*—Accurate annotation of cortical layers is fundamental to the study of human brain organisation yet remains manual and time-consuming in high-resolution spatial transcriptomics datasets. We present a multimodal graph convolutional network (GCN) that integrates gene expression profiles and histological context to automatically classify spots in the dorsolateral prefrontal cortex (DLPFC) into cortical layers. Using the publicly–available 10x Genomics DLPFC Visium dataset comprising 12 tissue sections ($33\,685$ spots), we represent each spot as a node with 51 features: 55 principal components of log-normalised gene expression and a single colour feature that summarises local Haematoxylin & Eosin (H&E) stain intensity. Edges are weighted by the Euclidean distance in the $(x, y,$ colour$)$ space, embedding both spatial proximity and morphology. A 3-layer GCN (208-102-40 hidden units) trained with inverse frequency class weights and label smoothing reaches 85% test accuracy and a 0.85 macro F_1 across eight layer classes.**

***Index Terms*—Spatial transcriptomics, graph neural networks, multimodal learning, cortical layering, histology, DLPFC**

I. Introduction

The dorsolateral prefrontal cortex (DLPFC) is crucial for cognition and exhibits a six-layered architecture with an additional white-matter compartment. Recent spatial transcriptomic technologies such as 10x Genomics Visium provide genome-wide expression with precise spatial coordinates, enabling computational layer annotation [1]. Graph convolutional networks (GCNs) excel at modelling neighbourhood dependencies [2], making them natural candidates for combining molecular and spatial cues. Recent advances in deep learning have demonstrated the effectiveness of integrating spatial transcriptomics with histological imaging data [4].

We propose a *lightweight* alternative that augments gene expression with a single, standardised colour feature per spot and constructs an edge-weighted graph capturing both spatial distance and stain similarity. Our open-source PyTorch implementation (https://github.com/taomasgonzalez/GNNCellClassification) is accompanied by a reproducible pipeline orchestrated with `dvc` and tracked with MLflow.

II. Dataset

The study uses the publicly-released DLPFC dataset consisting of 12 fresh-frozen human sections (IDs 151507-151676). After quality control, we retain $33\,685$ spots and $15\,876$ high-variance genes. Ground-truth layer labels provided by the Allen Institute were converted to eight classes (layers I-VI, white matter, unknown).

The training set contains $31\,731$ spots with the following class distribution:

TABLE I
Training data distribution by cortical layer class.

Class	Layer	Count (%)
0	Layer I	8,631 (27.20%)
1	Layer II	2,541 (8.01%)
2	Layer III	7,977 (25.14%)
3	Layer IV	3,494 (11.01%)
4	Layer V	4,121 (12.99%)
5	Layer VI	2,831 (8.92%)
6	White Matter	2,037 (6.42%)
7	Unknown	99 (0.31%)

The dataset exhibits significant class imbalance, with Layer I and Layer III being the most abundant classes, while the Unknown class represents only 0.31% of the training samples.

III. Methods

A. Pre-processing

Raw unique molecular identifier (UMI) counts were normalised with `scanpy` using total count scaling followed by natural logarithm transformation $\ln(1 + x)$. UMI counts can vary dramatically across genes and cells - highly expressed genes may have counts in the hundreds while lowly expressed genes have counts of 0, 1 or 2. The $\ln(1 + x)$ transformation compresses this wide dynamic range while preserving the relative relationships between expression levels, making the data more suitable for neural network training.

Dimensionality reduction via PCA is essential because the original dataset contains around $33\,500$ gene ids per spot, which would create computational bottlenecks and overfitting in neural networks. PCA captures the most informative variance in the gene expression space while reducing the feature dimension from $33\,500$ to 55, enabling efficient training while preserving the biological signal relevant for layer classification. The top 55 principal components (PCs) were computed on the training set and applied to validation and test splits.

B. Histological feature

For each spot, we extract a 96×96 region in the aligned H&E image and compute the mean RGB vector. Channel variances are used to form a weighted grey-scale value, which is subsequently z-normalised across the slide, yielding the scalar colour feature.

C. Graph Construction

Nodes correspond to spots. The adjacency matrix $A \in \mathbb{R}^{N\times N}$ is defined as the pairwise Euclidean distance in a three-dimensional space $[x, y, \alpha\, c]$ with spatial pixel coordinates (x, y) and the colour feature c. The scale factor α is set to the maximum spatial standard deviation to isotropically balance modalities. Edges with zero distance are pruned, resulting in a sparse weighted graph saved as a NumPy matrix per patient.

D. Split Strategy

Patients are split by donor: eight for training (66%), two for validation (16%) and two for held-out testing (16%) (with a seed of 42) to prevent spot-level leakage.

E. Model Architecture

We adopt a 3-layer GCN implemented in PyTorch Geometric:

$$\begin{aligned}\text{GCN}(51, 208) &\to \text{ReLU} \to \text{Dropout}(0.25)\\ \text{GCN}(208, 102) &\to \text{ReLU} \to \text{Dropout}(0.25)\\ \text{GCN}(102, 40) &\to \text{ReLU} \to \text{Linear}(40, 8).\end{aligned}$$

Weights are optimised with Adam (η=10^{-3}, weight decay 10^{-5}) and a cosine annealing scheduler with a period of 50 epochs (T_max = 400/8). Training continues for a maximum of 400 epochs but is terminated early if validation loss does not improve for 40 consecutive epochs. The model actually trained for 255 epochs before early stopping was triggered.

F. Loss Function

To mitigate severe class imbalance (Layer I accounts for 27.20% of spots while the Unknown class represents only 0.31%), we apply a weighted cross-entropy loss with inverse class frequencies and 0.2% label smoothing.

IV. Results

A. Quantitative Performance

Table II reports validation and test metrics at the epoch with highest validation accuracy. The model trained for 255 epochs before early stopping was triggered due to no improvement in validation loss for 40 consecutive epochs.

TABLE II
PERFORMANCE COMPARISON. THE GENE-ONLY BASELINE USES IDENTICAL ARCHITECTURE WITHOUT THE COLOUR FEATURE AND WITH PURELY SPATIAL EDGE WEIGHTS.

Model	Val Acc (%)	Val Macro F_1	Test Acc (%)	Macro F_1
GNN	**80.7**	**0.72**	**85.7**	**0.71**

Note on the Unknown Class: The unknown class (class 7) represents spots that could not be definitively classified by domain experts during the original annotation process. This class has extremely low occurrence in the dataset and, as expected, achieves 0% accuracy and F_1 score of 0.000. This is not a failure of the model but reflects the inherent ambiguity in the ground truth labels. The macro F_1 score is affected by the poor performance on this class, which contributes to the overall macro F_1 score despite its low frequency. When excluding the unknown class, the macro F_1 score has a value of 0.82 and is better reflecting the model's performance on the well-defined cortical.

Per-class metrics on the test set are shown in Tables III and IV. The unknown class remains challenging, whereas Layer-I (class 0) reaches 92.5% accuracy.

TABLE III
PER-CLASS ACCURACY ON THE HELD-OUT TEST SET.

Class	Layer	Accuracy (%)
0	Layer I	92.5
1	Layer II	80.8
2	Layer III	78.7
3	Layer IV	86.9
4	Layer V	79.0
5	Layer VI	87.0
6	White Matter	85.3
7	Unknown	0.0

TABLE IV
PER-CLASS F_1 SCORES ON THE HELD-OUT TEST SET.

Class	Layer	F_1 Score
0	Layer I	0.953
1	Layer II	0.700
2	Layer III	0.852
3	Layer IV	0.689
4	Layer V	0.815
5	Layer VI	0.791
6	White Matter	0.921
7	Unknown	0.000

B. Ablation Study

We ablate key components on the validation set:

- **No colour feature**: TBD accuracy.
- **Unweighted loss**: TBD accuracy.
- **No cosine schedule**: TBD accuracy.

V. Conclusion

We introduce a lightweight multimodal GCN that efficiently combines gene expression and histological staining to classify cortical layers in human DLPFC spatial transcriptomics. The proposed integration of a single colour-based morphological feature yields a significant accuracy boost at negligible computational cost, making the method attractive for large-scale atlasing studies. Future work will explore attention-based message passing and inclusion of learned image embeddings.

ACKNOWLEDGEMENTS

We thank the Allen Institute for Brain Science for releasing the dataset and the open-source community for developing PyTorch Geometric and MLflow.

REFERENCES

[1] L. A. Huuki-Myers *et al.*, "A data-driven single-cell and spatial transcriptomic map of the human prefrontal cortex," *AAAS*, 2022.

[2] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," in *Proc. ICLR*, 2017.

[3] J. Liu *et al.*, "SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains," *Nature Methods*, 2022.

[4] J. Luo, J. Fu, Z. Lu, and J. Tu, "Deep learning in integrating spatial transcriptomics with other modalities," *Briefings in Bioinformatics*, vol. 26, no. 1, pp. bbae719, 2025.