

Explaining Model Predictions When Users Aren't Sure: XAI for scRNA-seq Cell-Type Classification

Josh Shell*
University of Cincinnati

Ashley Kuenzi-Davis†
Cincinnati Childrens HMC

Jun Bai‡
University of Cincinnati

Jillian Aurisano§
University of Cincinnati

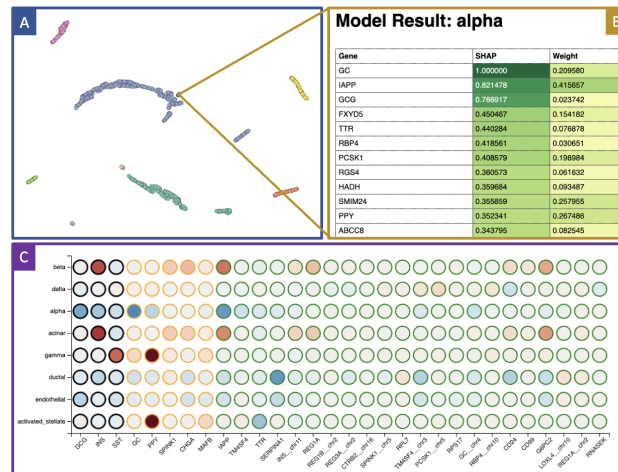


Figure 1: Overview of model outputs and corresponding explainability mechanisms for scRNA-seq cell-type classification: (A) UMAP projection of predicted cell types shown in color, allowing users to explore cell clustering, (B) Explanation panel showing top genes contributing to the selected prediction including SHAP values and learned model weights, (C) Dot plot heatmap showing expression of top genes across predicted cell types, enabling comparison between the model's top genes vs the user's. Dot color indicates importance; outlines represent the model's top genes (green), user-selected genes (orange), and areas of overlap (black).

ABSTRACT

Visualization techniques that support Explainable AI (XAI) are becoming increasingly important in scientific domains. Many XAI approaches assume the user has a fixed understanding of the correct answer and needs help understanding how the model arrived at its conclusion. However this assumption may not hold in complex domains like biology, where even expert knowledge is uncertain or incomplete. In these situations, XAI is not just a matter of verifying correctness but supporting scientific reasoning under uncertainty. This poster explores visualization techniques to support user reasoning in cell-type classification using scRNA-seq data. In a collaboration between biomedical and computer science researchers we iteratively developed a prototype system to understand how users reconcile model outputs when the ground-truth is unknown or ambiguous. To help users interpret model behavior, we incorporated familiar visualizations such as UMAP and dot plots, alongside XAI metrics like SHAP values and model weights. We present preliminary insights and discuss how these features resonate with domain experts. This work lays the foundation for a broader study examining how experts reason with AI model explanations where the basis for a decision is uncertain.

Index Terms: Explainable AI, Biological Data Visualization

*e-mail: ShellJJ@mail.uc.edu

†e-mail: Ashley.Kuenzi@cchmc.org

‡e-mail: baiju@ucmail.uc.edu

§e-mail: aurisajm@ucmail.uc.edu

1 INTRODUCTION

As artificial intelligence (AI) becomes increasingly integrated into scientific workflows, ensuring that model outputs are not only accurate, but trustworthy and explainable, is critical. Explainable AI (XAI) seeks to make model behavior interpretable to human users [5]. However, many XAI methods assume the user already knows the “correct” answer and simply needs help understanding how the model arrived at that answer. For example, if an image classifier highlights stripes to justify its decision to label an image as a zebra rather than a horse, this makes sense to a user who already knows zebras have stripes. But what happens when the user doesn't know the answer? This poster explores classification scenarios where both the model's behavior and the user's expectations may be incomplete or evolving. We focus on the task of cell-type classification using single-cell RNA sequencing (scRNA-seq) data, where cell-types are often unknown or ambiguous. In these settings, explanation is not about verifying a known answer but rather about supporting reasoning under uncertainty.

We present a preliminary prototype visualization application, developed through an iterative, user-centered process with domain experts. We intend to use this prototype as a design probe to examine how domain expert users reason about model predictions when the correct answer is ambiguous or unknown. This will support future development of XAI techniques.

2 BACKGROUND

2.1 What is scRNA-seq and why is it hard to explain?

Single cell RNA sequencing (scRNA-seq) measures gene expression, which approximates gene activity at the level of individual cells. It produces large, sparse datasets with tens of thousands of gene features. A common task is to classify or annotate cells by

cell type (i.e., alpha, beta, delta) based on these gene expression profiles to be used in downstream biomedical research.

Cell-type classification involves dimensionality reduction using methods such as Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP [8]), followed by clustering and comparison to annotated reference datasets. Marker genes, or genes known to be associated with specific cell types, are commonly used for validation, but may be inconsistent across species, tissue types, or samples [6]. scRNA-seq data is sensitive to “batch effects” which are systematic variations introduced by lab protocols or sampling conditions which can further complicate interpretation [4].

2.2 The role of AI and Viz in classification

As noted above, biologists commonly use visualization techniques such as plots of UMAP projections, dot plots, and heatmaps to explore scRNA-seq data [2, 9, 13, 12]. While these methods are effective for examining gene expression, these tools are not designed to explain model decisions. They help visualize the data, not the reasoning behind the model’s classification. The use of AI in scRNA-seq is becoming more common, yet explainability remains underexplored. Related to our work, Polyphony presents a framework for interactive transfer learning for cell-type annotation with scRNA-seq data [3]. Our work focuses on automated classification and supporting explainability for model decisions.

3 PROCESS AND IDENTIFIED XAI CHALLENGE

This work is a collaboration between three computer science researchers and two biomedical researchers who are experts in scRNA-seq data. Over the course of five months we met bi-weekly to understand cell-type classification workflows and challenges. We follow Munzner’s nested design model [10] and incorporate principles from User-Centered XAI framework from Wang et al. [14].

From initial discussions with domain collaborators multiple XAI challenges for our problem. Because ground truth is often unavailable and classifications are based on partial or uncertain knowledge, explaining model behavior becomes especially difficult. Users must reason not only about what the model did, but whether that decision fits their evolving and incomplete understanding of cell identity. Visualizing feature importance is a common technique in XAI. However, a unique challenge for this task is even domain experts are not familiar with all features (genes) in the input data, so they often encounter unfamiliar genes in scRNA-seq analysis and must refer to external resources to learn more about them. This issue highlights a need for XAI systems that expand beyond output explanations and incorporate external biological knowledge.

4 PROTOTYPE OVERVIEW AND DESIGN PROBE

To understand how users reason about model predictions under uncertainty, we developed an interactive prototype. This prototype is not a finalized tool but a vehicle to elicit user feedback on visualization techniques and learn about their reasoning methods in a planned study. The features of the prototype were iteratively developed in collaboration with domain experts and is embedded with visualizations and XAI techniques familiar to target users and appropriate for scRNA-seq analysis.

The prototype is a web-application, developed using d3 [1], (Figure 1) that visualizes the outputs of a three-layer deep learning classifier trained to label human pancreatic cells into cell types, using training data used by Muraro et al. [11]. The underlying data consists of 14,000 cells and over 20,000 features, reduced to a subset using the top 200 features and normalization. The model predictions are shown using a UMAP projection (Figure 1A) where each dot represents a classified cell. When a user clicks on a cell, (Figure 1B) a side panel table which reveals two complementary XAI

metrics: 1) SHAP values [7] showing instance-level feature importance for that particular classification and 2) Model weights to give the user context into the global feature importance across the training data. A diverging color scale is used to help the user identify places where SHAP values are high, but model weights are low, and vice versa. This table supports the task of comparing local and global reasoning.

To support class-level interpretability we included a dot plot-style heatmap of all eight cell-types found in the dataset (Figure 1C). Circle marks are organized into a grid, with the y-axis listing cell-types and the x-axis listing genes. These genes are initially the top-ranked genes by the model per cell-type, based on their SHAP values. A fill color is used to encode the importance of a gene within a given cell type using a diverging color scale (blue:low, red:high). The user is able to select genes of interest, such as marker genes, and add them to the dot-plot. The circles’ outlines are colored to encode membership in one of three categories: green for top-ranked genes by the model, orange as user-selected genes, and black for overlap between the user selected and model’s top 30. With this view, we aim to help the user compare their expectations against the model’s reasoning.

5 PRELIMINARY FINDINGS

With our domain collaborators, we have captured initial insights into how experts might understand model outputs for this data, as well as feedback on the visualization techniques in our prototype:

- **Agreement and Disagreement of Gene Importance:** When the model identified marker genes, our collaborators found this validating. However, when expected genes were missing or unfamiliar genes were emphasized they felt uncertain. Our domain collaborators explained that they trusted the model’s classification, but were not sure how to treat its reliance on these particular genes in its decision. These situations reemphasized the need for XAI systems to explore disagreement between the model and human expectations.
- **External Knowledge Integration:** Users requested links to external datasources (e.g., PubMed) to understand what is known about these unfamiliar genes. External knowledge might be useful in contextualizing and verifying the model’s inference of biological data. As stated above, seeing that the model used unexpected or unfamiliar genes in its decisions may lead to uncertainty in the domain expert users.
- **Interpretability of Metrics:** Our collaborators interpreted the comparison on SHAP (local) and model weights (global) as follows: a high SHAP and low weight was seen as similar to a marker gene- a feature that specifically marks as cell as a member of a specific type. This interpretation of the data mimics their existing processes for identifying cell types.
- **Visualization Familiarity:** Our collaborators recognized and understood the UMAP, table and dot plot views, noting how these are commonly used in their workflow.

These initial results suggest potential to develop techniques that utilize large language models for integrating external knowledge.

6 PLANNED STUDY AND FUTURE WORK

The initial sessions represent the pilot phase of this work, highlighting how users rationalize model predictions in various forms of uncertainty. We plan to conduct a study using the prototype as a design probe to examine how domain experts reason about model outputs and incorporate external knowledge. We will recruit biologists with scRNA-seq experience, who will interact with the tool, explore model-selected genes, and explain what those selections suggest. Resources such as PubMed will be available to support this reasoning. We will analyze session transcripts to identify common reasoning strategies and use those findings to guide the next prototype iteration.

REFERENCES

- [1] M. Bostock. D3.js - data-driven documents. <https://d3js.org>, 2012. Accessed: 2025-06-30. 2
- [2] B. Cakir, M. Prete, N. Huang, S. Van Dongen, P. Pir, and V. Y. Kiselev. Comparison of visualization tools for single-cell rnaseq data. *NAR Genomics and Bioinformatics*, 2(3):lqaa052, 2020. 2
- [3] F. Cheng, M. S. Keller, H. Qu, N. Gehlenborg, and Q. Wang. Polyphony: An interactive transfer learning framework for single-cell data analysis. *IEEE transactions on visualization and computer graphics*, 29(1):591–601, 2022. 2
- [4] L. Haghverdi, A. T. Lun, M. D. Morgan, and J. C. Marion. Batch effects in single-cell rna-sequencing data are corrected by matching mutual nearest neighbors. *Nature biotechnology*, 36(5):421–427, 2018. 2
- [5] S. R. Hong, J. Hullman, and E. Bertini. Human factors in model interpretability: Industry practices, challenges, and needs. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–26, 2020. 1
- [6] M. D. Luecken and F. J. Theis. Current best practices in single-cell rna-seq analysis: a tutorial. *Molecular systems biology*, 15(6):e8746, 2019. 2
- [7] W. E. Marcilio and D. M. Eler. From explanations to feature selection: assessing shap values as feature selection mechanism. In *2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 340–347. Ieee, 2020. 2
- [8] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018. 2
- [9] S. Mohseni, N. Zarei, and E. D. Ragan. A multidisciplinary survey and framework for design and evaluation of explainable ai systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 11(3-4):1–45, 2021. 2
- [10] T. Munzner. A nested model for visualization design and validation. *IEEE transactions on visualization and computer graphics*, 15(6):921–928, 2009. 2
- [11] M. J. Muraro, G. Dharmadhikari, D. Grün, N. Groen, T. Dielen, E. Jansen, L. Van Gurp, M. A. Engelse, F. Carlotti, E. J. De Koning, et al. A single-cell transcriptome atlas of the human pancreas. *Cell systems*, 3(4):385–394, 2016. 2
- [12] C. C. S. Program, S. Abdulla, B. Aevermann, P. Assis, S. Badajoz, S. M. Bell, E. Bezzi, B. Cakir, J. Chaffer, S. Chambers, et al. Cz cellxgene discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. *Nucleic Acids Research*, 53(D1):D886–D900, 2025. 2
- [13] M. L. Speir, A. Bhaduri, N. S. Markov, P. Moreno, T. J. Nowakowski, I. Papatheodorou, A. A. Pollen, B. J. Raney, L. Seninge, W. J. Kent, et al. Ucs cell browser: visualize your single-cell data. *Bioinformatics*, 37(23):4578–4580, 2021. 2
- [14] D. Wang, Q. Yang, A. Abdul, and B. Y. Lim. Designing theory-driven user-centric explainable ai. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–15, 2019. 2