Spatial Coordinates as a Cell Language: A Multi-Sentence Framework for Imaging Mass Cytometry Analysis

Chi-Jane Chen[*,1], **Yuhang Chen**[*,1], **Sukwon Yun**[*,1],
Natalie Stanley[†,1,2,3], **Tianlong Chen**[†,1]
[1]Department of Computer Science, The University of North Carolina at Chapel Hill,
[2]Computational Medicine Program, The University of North Carolina at Chapel Hill,
[3]Department of Genetics, The University of North Carolina at Chapel Hill

Abstract

Image mass cytometry (IMC) enables high-dimensional spatial profiling by combining mass cytometry's analytical power with spatial distributions of cell phenotypes. Recent studies leverage large language models (LLMs) to extract cell states by translating gene or protein expression into biological context. However, existing single-cell LLMs face two major challenges: (1) **Integration of spatial information**: hey struggle to generalize spatial coordinates and effectively encode spatial context as text, and (2) **Treating each cell independently**: they overlook cell-cell interactions, limiting their ability to capture biological relationships. To address these limitations, we propose `Spatial2Sentence`, a novel framework that integrates single-cell expression and spatial information into natural language using a multi-sentence approach. `Spatial2Sentence` constructs expression similarity and distance matrices, pairing spatially adjacent and expressionally similar cells as positive pairs while using distant and dissimilar cells as negatives. These multi-sentence representations enable LLMs to learn cellular interactions in both expression and spatial contexts. Equipped with multi-task learning, `Spatial2Sentence` outperforms existing single-cell LLMs on preprocessed IMC datasets, improving cell-type classification by `5.98%` and clinical status prediction by `4.18%` on the diabetes dataset while enhancing interpretability. The source code can be found here: `https://github.com/UNITES-Lab/Spatial2Sentence`

1 Introduction

Single-cell technologies, such as flow and mass cytometry (CyTOF) and single-cell RNA sequencing, have revolutionized our ability to analyze cellular heterogeneity in blood and tissue samples (Bendall et al., 2012; Brodin et al., 2019; Jagadeesh et al., 2022). These techniques provide high-resolution insights into the human immune system, enabling targeted therapeutic strategies for disease treatment and prevention (Reece et al., 2016). CyTOF, for instance, identifies immune cells based on protein expression (Bendall et al., 2012). However, traditional suspension-based proteomic approaches lack spatial context, limiting our understanding of cell-cell interactions and tissue organization.

Recent advancements, such as imaging mass cytometry (IMC) and multiplexed ion beam imaging (MIBI), overcome this limitation by integrating mass cytometry with spatial profiling (Kakade et al., 2021; Shaaban et al., 2024). These next-generation single-cell proteomics technologies enable high-dimensional immune profiling while preserving spatial relationships (Giesen et al., 2014; Nair et al., 2015). Such insights are crucial for characterizing the tumor microenvironment, where spatial interactions between immune and tumor cells influence prognosis (Keren et al., 2018). By capturing cell types, states, and interactions, spatial proteomics enhances our understanding of the immune system's complexity and disease mechanisms (Hartmann and Bendall, 2020).

Inspired by the advances in natural language processing (NLP), researchers have started conceptualizing cellular information as "words" and "sentences," enabling deep learning models to interpret cellular behavior within a context-dependent framework (Fang et al., 2024). A recent innovative approach applies large language model to extract cell (gene) state by translating cell (gene) expression information into biological context (Levine et al., 2023; Chen and Zou, 2024). By leveraging vast single-cell datasets, those models can generalize across gene expression dynamics, phenotype, and disease status, paving the way for new discoveries in immune system and therapeutic development. Despite their effectiveness in various downstream

[*]Equal contribution. [†] Co-corresponding authors. Correspondence: `{natalies,tianlong}@cs.unc.edu`

Findings of the Association for Computational Linguistics: ACL 2025, pages 13241–13252

tasks, current approaches face two major bottlenecks:

Integration of Spatial Information. With rapid advancements in spatial transcriptomics and multiplexed imaging, capturing spatial context has become increasingly important (Rao et al., 2021; Tian et al., 2023; Yun et al., 2024). Cells interact within their microenvironment, influencing differentiation, immune response, and disease progression (Marx, 2021). However, current models rely solely on expression matrices (e.g., gene or protein) and overlook spatial organization, leading to suboptimal representations. Spatial context is essential for modeling intercellular interactions, uncovering cell-type-specific behaviors, and improving biological relevance. Ignoring it risks missing key regulatory mechanisms that shape cellular function within complex tissues.

Treating each cell independently. Current single-cell LLM approaches (Levine et al., 2023; Cui et al., 2024) treat each cell as an independent entity, overlooking broader cellular interactions across similar or distinct groups of cells. Learning the similarities and distinctions between cells within the same or different functional groups (*e.g.,* cell types or tissue niches) can enrich the biological context captured in cell sentences (Keren et al., 2018; Hartmann and Bendall, 2020). Without this information, models may fail to capture key regulatory relationships or misinterpret cellular function within a tissue-specific context.

Driven by these motivations, we propose `Spatial2Sentence`, a novel single-cell LLM framework that integrates spatial information into language using a multi-sentence approach to capture cellular interactions. Specifically, we construct an expression similarity matrix and a distance matrix from the expression matrix and spatial coordinates, respectively. For each cell, we identify the most similar and adjacent cells as positive pairs, while dissimilar and distant cells serve as negative pairs. These structured pairs are then used to prompt an LLM, enabling it to capture cell-cell interactions effectively. Using our newly preprocessed IMC dataset, which containing protein expression matrices, spatial coordinates, and designated cell-type annotations for diabetes and brain tumor samples, `Spatial2Sentence` achieves state-of-the-art performance, improving cell-type classification by 5.98% and clinical status prediction by 4.18% compared to recent single-cell LLM approaches. Furthermore, our approach provides new insights into interpretability by identifying which cell types and protein markers are most crucial for distinguishing clinical states, offering a deeper understanding of their roles in disease progression.

In summary, our main contributions are summarized as follows:

- We highlight the limitations of existing single-cell LLMs, particularly their lack of spatial information integration and their limitation in capturing contextual information from neighboring cells, which can be important for improving annotation accuracy in spatial omics data.
- We propose a novel single-cell LLM framework, `Spatial2Sentence` that integrates spatial information as language and introduces multi-sentence contrastive prompting, enabling LLMs to capture cell interactions using positive and negative pairs based on both expression and spatial proximity.
- We preprocess and transform two IMC datasets into cell × protein feature matrices, providing spatial coordinates and cell-type annotations for each cell.
- Our method achieves state-of-the-art performance, surpassing previous models by 5.98% in cell-type classification and 4.18% in clinical status prediction on the Diabetes dataset.

2 Related Works

Cells as language. Transformer-based models have made significant advances in natural language processing due to their exceptional parallel processing capabilities and highly adaptable attention mechanisms. Building on these successes, researchers have begun applying transformer architectures to the modeling of single-cell data (Lan et al., 2024). scGPT is a deep learning-based approach designed for cell identification, particularly in the context of single-cell RNA sequencing (scRNA-seq) (Cui et al., 2024). The model has demonstrated strong performance in various downstream tasks, including multi-batch integration, multiomic synthesis, cell-type classification, genetic perturbation prediction, and gene network inference. Cell2Sentence (C2S) is another pre-trained model, fine-tuned from GPT-2, specifically designed to process textual sequences containing

gene names (Levine et al., 2023). The model generates new cell-level textual representations, and conversely, it can transform these textual sequences back into corresponding gene expression vectors. By converting cell text sequences back into gene expression profiles, C2S ranks genes based on their expression levels and preserves key information from the original data in the majority of cases. However, both scGPT and C2S have not yet been utilized for spatial datasets, such as those obtained from imaging mass cytometry (IMC) (Walsh and Quail, 2023). Additionally, if applied to spatial datasets without accounting for spatial location of individual cells, scGPT would overlook the geometric context inherent to imaging data, thereby forfeiting one of the key advantages provided by spatial modalities. Such existing approaches typically analyze cell data individually, failing to account for the potential interactions and relationships that may exist between pairs of cells.

CyTOF. Several methodologies have started leveraging CyTOF for cell annotation to study disease status or disease pathology (Milosevic, 2023). Take two examples of approaches that have been developed for performing cell annotation. Automated cell-type discovery and classification (ACDC) is an algorithm that can use in CyTOF or high-dimension mass cytometry like IMC to classify cell and new cell types by combining profile matching and semi-supervised learning (Lee et al., 2017). Linear discriminant analysis (LDA) also another automatic classifier for cell classification enables the analysis of large CyTOF datasets without requiring prior biological knowledge of marker expression patterns across different cell types (Abdelaal et al., 2018). While these techniques enable automated annotate cells, they process each cell independently and disregard spatial information. Our study utilizes IMC data, which captures spatial information for each cell across all samples. Therefore, incorporating spatial information is crucial for a more comprehensive analysis. The integration of spatial characterization with cell expression remains largely unexplored, and effectively incorporating spatial information remains a challenge.

3 Methodology

`Spatial2Sentence` consists of three stages: expression similarity and distance matrix generation, contrastive multi-sentence learning, and multi-sentence prompting. In Section 3.1, we define the preliminaries and notations used in this paper. In Section 3.2, we describe the generation of expression similarity and distance matrices, enabling each cell to identify its similar counterparts based on expression and spatial information. In Section 3.3, we introduce the multi-sentence technique, designed to capture interactions between cells through contrastive prompting with positive and negative pairs derived from the similarity and distance matrices. Figure 1 provides a detailed illustration of the overall framework of `Spatial2Sentence`.

3.1 Preliminary Definitions

Let $\mathbf{B} = \{b_1, b_2, \ldots, b_M\} \in \mathbb{R}^{1\times M}$ represent the names of the proteins, where each element b_k corresponds to the name of the k-th protein. Let $V \in \mathbb{R}^{N\times M}$ represent the protein expression profiles, where N is the number of cells and M is the number of proteins. Each entry $v_{i,j}$ in the matrix V corresponds to the expression level of protein j in cell i, and the spatial positions of the cells are denoted by $C \in \mathbb{R}^{N\times 2}$, where each row $\{x_i, y_i\}$ represents the 2D spatial coordinates of cell i.

Following C2S (Levine et al., 2023), we convert the expression data for each cell into a linguistic sentence. Specifically, each cell's protein expression profile $\mathbf{v}_i = \{v_{i,1}, v_{i,2}, \ldots, v_{i,M}\}$ is transformed into a sentence by rank-ordering the proteins according to their expression levels. This is based on the hypothesis that the rank of protein expression reflect the property of cell (*i.e.*, certain cell types may exhibit higher expression levels of specific proteins). More formally, we convert the expression matrix V into a cell sentence by ordering the proteins in decreasing order of expression:

$$r_{i,k} = \text{Rank}(v_{i,k}, \mathbf{v}_i) \tag{1}$$

where $\text{Rank}(v_{i,k}, \mathbf{v}_i)$ denotes the descending rank of $v_{i,k}$ within $\mathbf{v}_i$. The resulting cell protein sentence is then given by:

$$S_i = \{b_{r_{i,1}}, b_{r_{i,2}}, \ldots, b_{r_{i,M}}\} \tag{2}$$

This transformation represents a cell's high-dimensional expression data as a token sequence, making it more accessible for LLMs.

3.2 Expression Similarity & Distance Matrix

To effectively incorporate both expression values and spatial information into the multi-sentence prompt design, we first compute two types of pairwise similarity measures: the cosine similarity

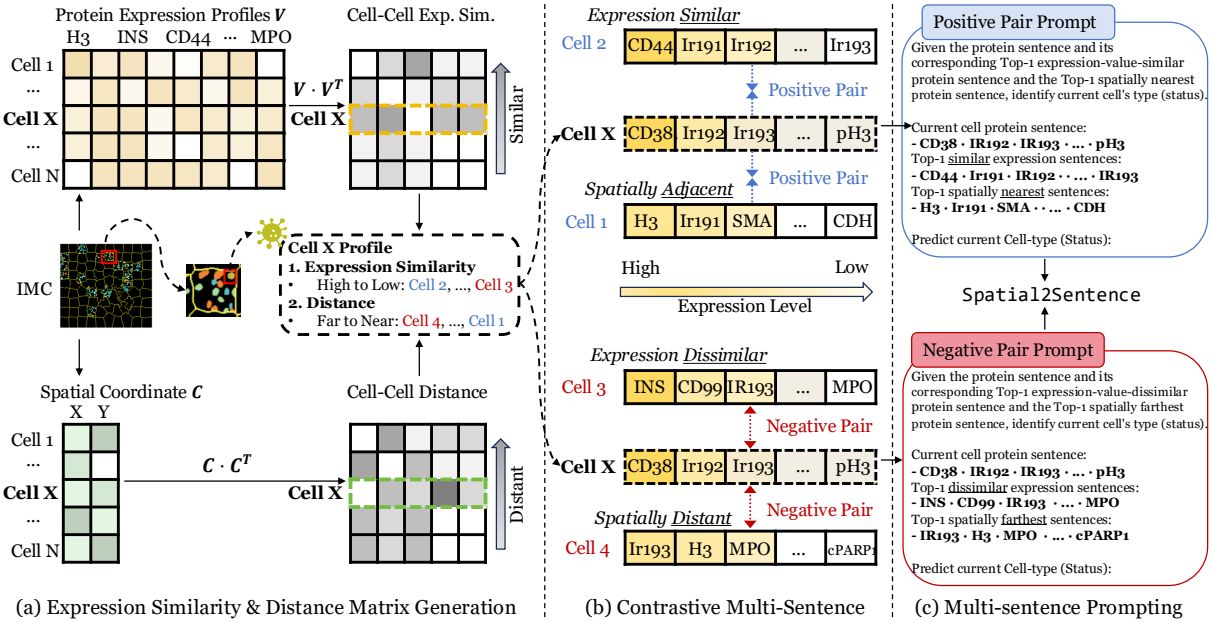

Figure 1: Overall framework of Spatial2Sentence. (a) Given IMC data, we integrate both the protein expression matrix and the spatial coordinate information simultaneously. The protein expression matrix (with cells as rows and proteins as columns) is transposed (i.e., V^T) and multiplied to compute cosine similarity. Similarly, the spatial coordinate matrix (with columns representing the X and Y coordinates) is processed to obtain a distance matrix. For cell 'X', we rank the cells based on their expression similarity (from highest to lowest) and their spatial proximity (from farthest to nearest). (b) Using these ranked cell indices, we perform Contrastive Multi-Sentence Generation. Here, the positive pairs consist of the top-k cells in terms of both expression similarity and proximity to cell 'X', while the negative cases use the top-k cells with the most dissimilar expressions and that are most distant from cell 'X'. (c) Finally, equipped with both positive and negative pairs, we prompt these pairs into LLMs to leverage their capability in handling proteins and spatial information within a multi-sentence framework that captures interactions among different cells. In the illustration, the top-1 case is used as an example; however, any k can be used.

based on the expression profiles and the Euclidean distance based on the spatial coordinates.

Expression Similarity. Let $V \in \mathbb{R}^{N\times M}$ represent the matrix of protein expression profiles, where each row v_i is the expression vector of cell i across M proteins. The cosine similarity between two cells i and j is computed as:

$$g_{i,j} = \text{CosSim}(i,j) = \frac{v_i \cdot v_j}{\|v_i\|_2 \|v_j\|_2} \tag{3}$$

The resulting expression similarity matrix $\mathbf{G} \in \mathbb{R}^{N\times N}$ captures the pairwise cosine similarities between the expression profiles of all cells, and each entry $g_{i,j}$ represents the cosine similarity between cells i and j.

$$\mathbf{G} = \begin{bmatrix} 1 & g_{1,2} & \cdots & g_{1,N} \\ g_{2,1} & 1 & \cdots & g_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ g_{N,1} & g_{N,2} & \cdots & 1 \end{bmatrix} \tag{4}$$

Spatial Proximity. Given the matrix of spatial coordinates $C \in \mathbb{R}^{N\times 2}$ where each row corresponds to the spatial coordinates $\{x_i, y_i\}$ of cell i, the Euclidean distance between cells i and j is

$$d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{5}$$

which is done pairwise for all cells, resulting in the distance matrix $\mathbf{D} \in \mathbb{R}^{N\times N}$, where each entry $D_{i,j}$ represents the Euclidean distance between the spatial coordinates of cells i and j. We compute this for all pairs of cells, which output the full spatial distance matrix:

$$\mathbf{D} = \begin{bmatrix} 0 & d_{1,2} & \cdots & d_{1,N} \\ d_{2,1} & 0 & \cdots & d_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ d_{N,1} & d_{N,2} & \cdots & 0 \end{bmatrix} \tag{6}$$

Ranking Cells Sentences. Once we compute $\mathbf{D}$ and $\mathbf{G}$, we rank the cells in terms of their proximity and similarity. For each cell i, we generate two ordered lists: ① Expression Similarity Ranked Sentences: This list contains the indices of cells ordered by their cosine similarity to cell i, indicating

how similar their protein expression profiles are. ② Spatial Proximity Ranked Sentences: This list contains the indices of cells ordered by their Euclidean distance to cell i, indicating how physically close the cells are in the tissue.

3.3 Contrastive Multi-sentence Generation & Prompting

Motivation. Traditional single-cell approaches often represent each cell by a single sentence derived from its feature expression profile. While this method captures a cell's individual characteristics, it fails to account for the complex interrelationships between cells or provide a comprehensive understanding of cellular behavior. In particular, such representations overlook the potential nuances in how cells relate to each other in a biological context, especially in the presence of heterogeneous tissue environments. Therefore, relying on only one sentence per cell limits the model's ability to discern subtle differences or similarities between cells, which is crucial for tasks such as cell-type identification or disease status prediction.

Multi-sentence Prompt. To address this limitation, we introduce the concept of multi-sentence prompts, where each prompt includes the protein expression profiles of multiple cells. Instead of relying on a single sentence, the model processes pairs (or more) of sentences from different cells, which enhances its ability to capture both similarities and differences in protein expression across different cell types.

For two cells i and j, each with their respective protein expression sentences S_i and S_j, the multi-sentence prompt consists of:

$$\begin{aligned} S_i &= \{b_{r_{i,1}}, b_{r_{i,2}}, \ldots, b_{r_{i,M}}\}, \\ S_j &= \{b_{r_{j,1}}, b_{r_{j,2}}, \ldots, b_{r_{j,M}}\} \end{aligned} \tag{7}$$

Here, $r_{i,k}$ and $r_{j,k}$ denote the rank of the k-th protein in cell i and cell j, respectively. This design encourages the LLMs to consider relationships between cells and can help identify commonalities or distinctions in cell types or disease status.

Once we have computed the spatial proximity and expression similarity for each cell, we design the multi-sentence contrastive prompting framework. This involves using both `positive-pair` and `negative-pair` prompts to provide the model with the necessary context for learning relationships between cells.

`Positive-pair Prompt.` In the positive-pair prompt, the goal is to guide the model to identify common characteristics between cells that share similar expression profiles and spatial proximities. We use cells that are close both in terms of expression similarity and spatial proximity. Specifically, for each cell i, we generate a positive-pair prompt by selecting the top K most similar cells in terms of expression and the top K spatially closest cells. The input prompt is formatted as {`Prompt`, S_i, $S_{\mathsf{TopK}(D_i)}$, $S_{\mathsf{TopK}(G_i)}$, `Task (Predict cell type)`} (See Figure 1 for details). This prompt encourages the model to consider both molecular similarity and spatial proximity in making its predictions.

`Negative-pair Prompt.` For the negative-pair prompt, the objective is to contrast cells that exhibit dissimilar expression profiles and spatial distances. This allows the model to learn to distinguish between cells that are fundamentally different in terms of their biological properties. For each cell i, we generate a negative-pair prompt by selecting the top K most dissimilar cells in terms of expression and the top K spatially furthest cells. The negative-pair prompt is formatted as {`Prompt`, S_i, $S_{\mathsf{TopK}(D_{\mathsf{far}})}$, $S_{\mathsf{TopK}(G_{\mathsf{dissim}})}$, `Task`} (See Figure 1 for details). This prompt helps the model learn the distinctions between cells that are biologically or spatially divergent. With the positive-pair prompt and negative-pair prompt, the objective is to effectively distinguish between similar and dissimilar cells by leveraging both expression profiles and spatial contexts. The model is trained to predict cell types based on the information from the positive-pair and negative-pair prompts, enabling it to understand cellular behavior in the context of tissue heterogeneity.

4 Experiments

4.1 Datasets

To assess the model's ability to predict cell types and status, as well as to summarize cell-type abundances, we applied it to two multi-sample IMC image datasets, which are described below (Damond et al., 2019; Karimi et al., 2023).

Diabetes dataset. This dataset profiles 67 diabetic and non-diabetic donors longitudinally from human pancreatic tissue. Among these, 33 samples are from non-diabetic donors, while 34 samples represent donors who developed long-term Type 1 diabetes. Long-term Type 1 diabetes refers to individuals who have lived with the disease for an extended period, typically several years after diag-

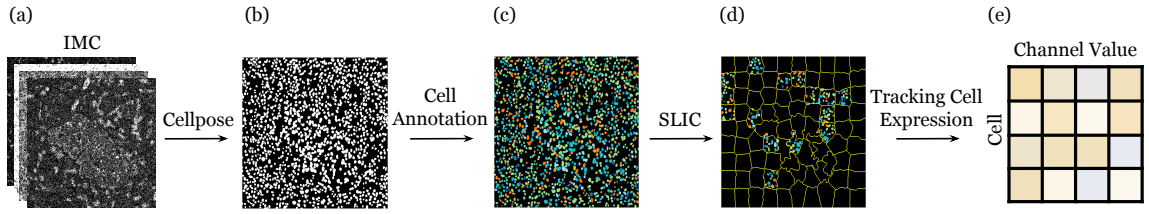

Figure 2: Given a (**a**) multi-sample IMC dataset (**b**), we used Cellpose to detect cell centers within superpixels (**c**) and extracted cells from IMC images for cell-type annotation. (**d**) We then applied the SLIC algorithm to segment images into superpixel region, (**e**) generating a cell × protein feature matrix for analysis.

Table 1: Comparison of different models including scGPT, Geneformer, C2S, GenePT, scELMo, LangCell, and `Spatial2Sentence` on diabetes and brain tumor datasets. Table shows classification accuracy (%) across various settings, where *Single-Task* (*Multi-Task*) denotes single-tasking (multi-tasking), *Type* represents cell-type classification, and *Status* indicates clinical status prediction. Best results for each column are bolded.

Model	Diabetes				Brain Tumor			
	Single-Task		Multi-Task		Single-Task		Multi-Task	
	Type	Status	Type	Status	Type	Status	Type	Status
Geneformer	31.62	55.78	34.29	62.02	48.10	54.09	50.14	58.16
GenePT	37.33	60.15	39.98	64.02	51.90	52.20	53.67	58.31
scELMo	34.33	56.14	37.95	62.03	49.44	50.26	51.66	55.27
LangCell	40.83	**64.64**	41.17	72.51	53.38	55.70	55.16	63.80
scGPT	32.45	58.12	34.50	65.38	47.92	56.27	50.11	62.04
scGPT w/ Spatial Info	33.98	57.78	36.35	67.23	49.68	56.16	52.98	62.98
C2S	36.37	60.45	37.54	72.55	49.03	52.06	51.86	54.04
C2S w/ Spatial Info	36.98	62.15	38.03	**74.11**	51.12	53.93	53.22	56.09
`Spatial2Sentence` w/o Spatial Info	37.89	62.08	38.26	72.05	51.50	55.48	52.98	**65.19**
`Spatial2Sentence`	**41.35**	64.12	**41.98**	74.02	**53.89**	**57.25**	**55.67**	63.31

nosis. For those patients in the advanced stages of Type 1 diabetes are typically marked by prolonged autoimmune responses and extensive destruction of some particular cell-type, which lead to irreversible impairment of endogenous insulin production. Each cell in the dataset was characterized by 38 measured proteins. Moreover, we associated a clinical outcome with each sample in the dataset as ① long-duration diabetes or ② non-diabetic control. We annotated the cells into seven cell types, including T cell, Helper T cell, CD8 T/Cytotoxic T cell, Neutrophils, Monocytes/Macrophages, Immune cell, and other cell types.

Brain tumor dataset. The glioblastoma IMC dataset includes samples from 118 glioblastoma patients, allowing detailed characterization of the tumor microenvironment (TME), and 46 brain metastasis (BrM) tumors from distinct patients. A total of 21 protein markers were selected for analysis across both conditions. After excluding samples with missing information or that lack all common markers, we balanced the dataset to include samples from 37 glioblastoma donors and from 37 brain metastasis donors. Glioblastoma patients have a primary brain tumor that originates in the brain, typically from glial cells, and is known for being highly aggressive and fast-growing. In other condition, brain metastasis patients have secondary tumors that spread to the brain from other cancers in the body, such as lung or breast cancer. While both affect the brain, their origins, progression, and treatment strategies are fundamentally different. The images cover multiple tissue regions, including the tumor core, tumor margin, and tumor-adjacent normal tissue. Some samples contain multiple tissue regions, resulting in a final collection of 100 brain metastasis samples and 72 glioblastoma samples. For cell annotation, the cells were phenotyped into six distinct cell types, such as Tc cell, B cell, Astrocytes, M1-like MDMs, M2-like MDMs, and undefined cells.

For the data pre-processing, we utilized Cellpose to detect cell centers (Stringer et al., 2021). After extracting cells from the IMC images, we performed downstream analyses, including cell clustering and cell-type annotation (Stanley et al., 2020). Cell-type annotation was manually defined at the cluster level rather than on a per-cell ba-

Table 2: **Ablation study** of model components, showing the accuracy of the model with different components removed on diabetes and brain tumor datasets.

Component	Diabetes		Brain Tumor	
	Type	Status	Type	Status
`Spatial2Sentence`	**41.35**	**64.12**	**53.89**	**57.25**
w/o Multi-sentence Prompting	36.37	60.45	49.03	52.06
w/o Negative Pair	38.78	62.12	51.87	56.35
w/o Positive Pair	36.97	59.13	50.23	54.05
w/o Expression Similar Sentences	39.86	60.24	52.12	52.14
w/o Spatial Proximal Sentences	40.16	61.53	53.04	54.45
w/o Cosine Similarity (Random Select)	36.02	62.13	49.45	53.82
w/o Euclidean Distance (Random Select)	35.54	58.68	48.12	51.57

sis. Following cell detection, we employed the SLIC algorithm to segment the image into multiple super-pixel regions(Achanta et al., 2010). This ultimately produced a matrix of cells × protein features within the defined superpixels for subsequent analysis. The overview of data preprocessing shown in Fig. 2.

4.2 Experimental Details

Baselines. We compare our proposed method with several state-of-the-art and relevant models. Our comparisons include the widely-used large foundation model scGPT (Cui et al., 2024), a deep learning-based approach for cell-type identification in single-cell RNA sequencing; we also evaluate a variant, scGPT w/ Spatial Info, where spatial coordinates are explicitly provided as input. Additionally, we benchmark against Geneformer (Lan et al., 2024), a transformer-based model for multi-omics data integration, and C2S (Levine et al., 2023), which generates cell-level textual representations from protein expression profiles. An extension, C2S w/ Spatial Info (Walsh and Quail, 2023), which incorporates spatial data, is also considered. Furthermore, we include GenePT (Chen and Zou, 2024), an embedding model for single-cell biology focusing on feature-level representations; scELMo (Liu et al., 2023), which combines metadata and expression profiles using embeddings from language models; and LangCell (Zhao et al., 2024), a pre-training framework integrating gene ranks with metadata for language-cell understanding. All these baseline models are evaluated on the diabetes and brain tumor datasets to assess their performance in cell-type classification and clinical status prediction tasks.

Tasks. We perform both single-task and multi-task predictions to comprehensively assess our model's performance. Specifically, we evaluate cell Type and Status at the individual cell level. The performance metrics is classification accuracy.

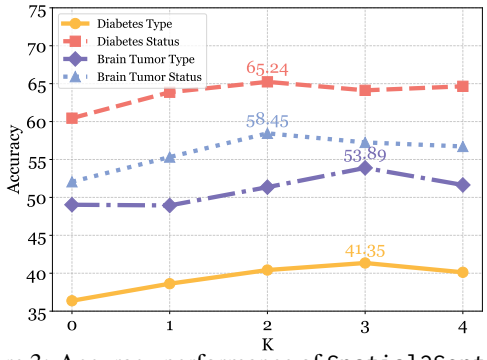

Figure 3: Accuracy performance of `Spatial2Sentence` across different values of the hyperparameter K.

Experiment Details. We use the Llama-3.2-1B (Dubey et al., 2024) model for our experiments by default, fine-tuning it with the following training parameters: the batch size is set to 8 per device, and we apply a learning rate of 2e-4 with a cosine learning rate scheduler. The model is trained for 5 epochs, with a warm-up ratio of 0.05. The training is carried out using PyTorch on an NVIDIA RTX 6000 Ada Generation 48GB GPU. The data set is divided into 90% for training and 10% for validation, with a separate test set reserved for the final evaluation. To ensure reproducibility, each experiment is repeated three times with different random seeds, and the results are averaged for reporting.

4.3 Primary Results

Comparison to State-of-the-Art. Table 1 provides a detailed comparison of our method with several state-of-the-art and relevant models, now including scGPT with added spatial information, GenePT (Chen and Zou, 2024), scELMo (Liu et al., 2023), and LangCell (Zhao et al., 2024), Geneformer (Theodoris et al., 2023), alongside the baselines like scGPT (Cui et al., 2024), and C2S (Levine et al., 2023) with and without spatial information. While the performance landscape varies across different tasks and datasets, `Spatial2Sentence` demonstrates leading results in several key areas, particularly in cell-type classification tasks and certain clinical status prediction scenarios, underscoring the benefits of its spatial-aware multi-sentence framework (further details in Table 1). For instance, `Spatial2Sentence` achieves top performance in Diabetes single-task cell-type classification (41.35%) and Brain Tumor single-task cell-type classification (53.89%). The inclusion of spatial information within the `Spatial2Sentence` framework generally enhances its ability to capture relevant patterns effectively.

Table 3: Classification accuracy comparison between the C2S method and `Spatial2Sentence` across various LLMs on diabetes and brain tumor datasets. Single and Multi denote single-task and multi-task learning, respectively.

Model	Method	Diabetes				Brain Tumor			
		Single		Multi		Single		Multi	
		Type	Status	Type	Status	Type	Status	Type	Status
GPT-2-Small	C2S	36.56	58.24	37.01	72.24	53.23	55.23	52.56	53.34
	Ours	38.43	59.70	39.54	71.40	56.02	57.26	55.48	52.78
GPT-2-XL	C2S	37.13	62.34	37.81	**76.96**	57.68	60.34	59.26	61.83
	Ours	37.07	62.72	35.30	74.83	59.28	61.21	60.78	**63.51**
Llama-2-7B	C2S	37.45	62.89	38.17	75.68	53.25	54.12	55.67	58.98
	Ours	38.17	59.72	39.54	74.40	56.02	57.26	55.48	56.78
Llama-3.2-1B	C2S	36.37	60.45	37.54	72.55	49.03	52.06	51.86	54.04
	Ours	**41.35**	64.12	**41.98**	74.02	53.89	57.25	55.67	63.31
Gemma-2-9B	C2S	38.34	65.12	38.01	76.78	59.53	57.39	60.32	59.32
	Ours	40.45	**67.24**	40.31	75.89	**62.13**	**63.23**	**62.45**	61.23

Table 4: **Sensitivity** on different spatial distances.

Component	Diabetes		Brain Tumor	
	Type	Status	Type	Status
L1 Norm	**41.35**	63.11	53.19	56.03
Cosine Distance	38.25	62.24	50.56	53.31
Euclidean Distance	**41.35**	**64.12**	**53.89**	**57.25**

Table 5: **Sensitivity** on different expression similarities.

Component	Diabetes		Brain Tumor	
	Type	Status	Type	Status
Pearson Correlation	39.16	**65.56**	49.76	56.86
Euclidean Distance	39.22	63.87	48.94	54.32
Cosine Similarity	**41.35**	64.12	**53.89**	**57.25**

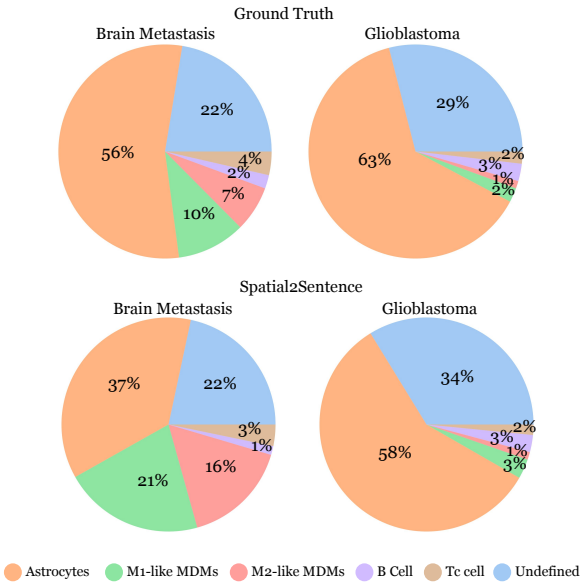

Figure 4: Cell-type distribution in brain tumor dataset between ground truth and `Spatial2Sentence` result.

Effect of Multi-task learning. The results in Table 1 demonstrate the effectiveness of multi-task learning in improving model performance across both the diabetes and brain tumor datasets. For instance, on the diabetes dataset, the multi-task setting improves cell-type prediction accuracy by 3.44% and status prediction accuracy by 0.91% over the single-task approach. We argue that this performance boost can be attributed to the model's ability to learn shared representations across related tasks, enhancing its generalization capability. Furthermore, patient-level status information can aid in improving cell-type prediction, as the model learns broader context about disease states. Similarly, knowing the cell types contributes to better status prediction, as it provides crucial biological insights into the patient's condition, reinforcing the interdependence between these tasks.

Cell-Type Frequency Analysis. To summarize the cell type distribution in different disease statuses, we conduct experiment on brain dataset shown in Fig. 4. The finding from Spatial2Sentence indicates a sharp increase in M1-like MDMs in brain metastasis patients, which aligns with the trend observed in the ground truth pie chart. Additionally, previous studies have reported that brain metastases appear in higher proportion of M1-like monocyte-derived macrophages (MDMs) which further supports our finding (Karimi et al., 2023; Schreurs et al., 2025).

Table 6: Ablation study on contrastive negative pair selection strategies for Spatial2Sentence.

Negative Pair Size	Diabetes (Multi-Task)		Brain Tumor (Multi-Task)	
	Type	Status	Type	Status
Top-1	41.08	73.72	54.25	62.78
Top 1-3	**41.98**	**74.02**	**55.67**	**63.31**
Top 4-6	40.42	73.15	54.93	63.64
Top 7-9	39.09	73.37	53.56	62.18

4.4 Diagnostic Analysis

Ablation Study. In parallel, we conduct ablation experiments to assess the impact of various model components on the overall performance. The results in Table 2 shows that ① removing multi-sentence prompting decreases contextual understanding, while excluding either pair type (positive or negative) reduces ability to differentiate cells, with positive pairs being more critical. ② Removing expression similarity harms performance more than spatial proximity, highlighting the importance of molecular features. ③ Replacing structured similarity with random selection leads to a significant performance drop, emphasizing the importance of these measures.

Sensitive Study. ① Hyperparameter K determines amount of cell sentences for generating the positive-pair and negative-pair multi-sentence prompts. We evaluated Spatial2Sentence across various K values in Figure 3. Spatial2Sentence ($K \geq 1$) consistently ourperforms baseline ($K = 0$). ② Sensitivity studies on spatial distance methods (Table 4) and expression similarity methods (Table 5) show that Euclidean Distance and Cosine Similarity yield the best results.

Performance Across Different LLMs. We evaluate the performance of our method across different LLMs, including GPT-2-Small (Radford et al., 2019), GPT-2-XL (Radford et al., 2019), Llama-2-7B (Touvron et al., 2023), Llama-3.2-1B (Dubey et al., 2024), and Gemma-2-9B (Team et al., 2024). Table 3 presents the classification accuracy of both the C2S model and our proposed method under each LLM configuration on the diabetes and brain tumor datasets. We find that ① Our method outperforms C2S in cell-type classification and disease status prediction for most LLM configurations. ② On the diabetes dataset, our method achieves 41.35% accuracy in cell-type classification using Llama-3.2-1B, compared to C2S's 36.37%, but some models like GPT-2-Small show less improvement. We argue this is because smaller models like GPT-2-Small have relatively limited capacity to handle longer prompts.

Impact of Negative Pair Selection Strategy. The construction of informative negative pairs is crucial for contrastive learning. Our framework selects the top-K most dissimilar cells (in terms of expression and spatial distance) to form negative pairs, rather than relying on random sampling or a single extreme case. To evaluate the impact of this selection, we performed an ablation study by varying the criteria for dissimilar cells. Table 6 shows the performance of Spatial2Sentence when using different top-K ranges for selecting negative cells for the multi-sentence prompts. The results indicate that using the top 1-3 most dissimilar and distant cells provides the most effective contrastive signal, leading to the best performance in multi-task settings. This suggests that carefully chosen, highly contrasting negative examples are more beneficial than moderately or extremely dissimilar ones beyond a certain threshold.

5 Conclusion

We proposed Spatial2Sentence, a novel framework that jointly integrates spatial information and cell expression by modeling cells as sentences. Spatial2Sentence effectively captures cellular similarities and distinctions across diverse cell types, enhancing the biological context encoded in language. Experiments on two IMC datasets show that Spatial2Sentence consistently outperforms existing methods. By introducing a multi-sentence contrastive prompting strategy that leverages both spatial proximity and cell-level similarity, we demonstrate that even lightweight LLMs can gain substantial performance boosts from spatially informed prompts. These findings suggest that scaling to stronger models could further advance this direction, offering a promising bridge between spatial biology and large language models.

Limitations and Future Works

However, there are still substantial opportunities to expand this methodology to multi-model framework, incorporating multi-omic data such as genomics, transcriptomics, or proteomics. A full biological view of cell heterogeneity and disease mechanisms. Additionally, enhancing our model into large LLMs, e.g., with 70B-80B parameters will further improve its ability to capture more complex biological insights. Furthermore, we plan to explore robust learning strategies for scenarios when some modalities are missing, ensuring the model remains effective even given incomplete data. By addressing these challenges, we aim to develop flexible and powerful framework for biology analysis.

Ethical Statement

To the best of our knowledge, the Diabetes and Brain datasets used in this study have been compiled from publicly available sources, ensuring compliance with ethical guidelines and avoiding the inclusion of sensitive or private information. The primary focus of Spatial2Sentence is to enhance the representation of single-cell expression and spatial interactions by leveraging a multi-sentence natural language approach. This method is specifically designed for biomedical research, distinguishing it from general-purpose language models used in dialogue systems. However, we acknowledge potential ethical concerns, including biases in dataset representation and the risk of misinterpretation of biological findings. While our framework inherently reduces the likelihood of generating harmful content, we emphasize the need for responsible usage and validation to prevent potential misuse, particularly in clinical or diagnostic applications.

Acknowledgments

This research was, in part, funded by the National Institutes of Health (NIH) under other transactions 1OT2OD03804501. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of the NIH. This work was also supported by the National Institute of Aging of the NIH award number 5R21AG084251-02 (NS) and the National Institute of Allergy and Infectious Diseases of the NIH award number 5R21AI171745-02 (NS).

References

Tamim Abdelaal, Vincent van Unen, Thomas Höllt, Frits Koning, Marcel J.T. Reinders, and Ahmed Mahfouz. 2018. Predicting cell types in single cell mass cytometry data. *bioRxiv*, 19(6):759–769.

Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. 2010. Slic superpixels.

Sean C Bendall, Garry P Nolan, Mario Roederer, and Pratip K Chattopadhyay. 2012. A deep profiler's guide to cytometry. *Trends in immunology*, 33(7):323–332.

Petter Brodin, Darragh Duffy, and Lluis Quintana-Murci. 2019. A call for blood—in human immunology. *Immunity*, 50(6):1335–1336.

Yiqun Chen and James Zou. 2024. Genept: a simple but effective foundation model for genes and cells built from chatgpt. *bioRxiv*, pages 2023–10.

Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. 2024. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature Methods*, pages 1–11.

Nicolas Damond, Stefanie Engler, Vito RT Zanotelli, Denis Schapiro, Clive H Wasserfall, Irina Kusmartseva, Harry S Nick, Fabrizio Thorel, Pedro L Herrera, Mark A Atkinson, et al. 2019. A map of human type 1 diabetes progression by imaging mass cytometry. *Cell metabolism*, 29(3):755–768.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Chen Fang, Yidong Wang, Yunze Song, Qingqing Long, Wang Lu, Linghui Chen, Guihai Feng, Yuanchun Zhou, and Xin Li. 2024. How do large language models understand genes and cells. *ACM Transactions on Intelligent Systems and Technology*.

Charlotte Giesen, Hao AO Wang, Denis Schapiro, Nevena Zivanovic, Andrea Jacobs, Bodo Hattendorf, Peter J Schüffler, Daniel Grolimund, Joachim M Buhmann, Simone Brandt, et al. 2014. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature methods*, 11(4):417–422.

Felix J Hartmann and Sean C Bendall. 2020. Immune monitoring using mass cytometry and related high-dimensional imaging approaches. *Nature Reviews Rheumatology*, 16(2):87–99.

Karthik A Jagadeesh, Kushal K Dey, Daniel T Montoro, Rahul Mohan, Steven Gazal, Jesse M Engreitz, Ramnik J Xavier, Alkes L Price, and Aviv Regev. 2022. Identifying disease-critical cell types and cellular processes by integrating single-cell rna-sequencing and human genetics. *Nature genetics*, 54(10):1479–1492.

Vijayakumar R Kakade, Marlene Weiss, and Lloyd G Cantley. 2021. Using imaging mass cytometry to define cell identities and interactions in human tissues. *Frontiers in Physiology*, 12:817181.

Elham Karimi, Miranda W Yu, Sarah M Maritan, Lucas JM Perus, Morteza Rezanejad, Mark Sorin, Matthew Dankner, Parvaneh Fallah, Samuel Doré, Dongmei Zuo, et al. 2023. Single-cell spatial immune landscapes of primary and metastatic brain tumours. *Nature*, 614(7948):555–563.

Leeat Keren, Marc Bosse, Diana Marquez, Roshan Angoshtari, Samir Jain, Sushama Varma, Soo-Ryum Yang, Allison Kurian, David Van Valen, Robert West, et al. 2018. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell*, 174(6):1373–1387.

Wei Lan, Guohang He, Mingyang Liu, Qingfeng Chen, Junyue Cao, and Wei Peng. 2024. Transformer-based single-cell language model: A survey. *Big Data Mining and Analytics*, 7(4):1169–1186.

Hao-Chih Lee, Roman Kosoy, Christine E Becker, Joel T Dudley, and Brian A Kidd. 2017. Automated cell type discovery and classification through knowledge transfer. *Bioinformatics*, 33(11):1689–1695.

Daniel Levine, Syed Asad Rizvi, Sacha Lévy, Nazreen Pallikkavaliyaveetil, David Zhang, Xingyu Chen, Sina Ghadermarzi, Ruiming Wu, Zihe Zheng, Ivan Vrkic, et al. 2023. Cell2sentence: teaching large language models the language of biology. *BioRxiv*, pages 2023–09.

Tianyu Liu, Tianqi Chen, Wangjie Zheng, Xiao Luo, and Hongyu Zhao. 2023. scelmo: Embeddings from language models are good learners for single-cell data analysis. *bioRxiv*, pages 2023–12.

Vivien Marx. 2021. Method of the year: spatially resolved transcriptomics. *Nature methods*, 18(1):9–14.

Vladan Milosevic. 2023. Different approaches to imaging mass cytometry data analysis. *Bioinformatics Advances*, 3(1):vbad046.

Nitya Nair, Henrik E Mei, Shih-Yu Chen, Matthew Hale, Garry P Nolan, Holden T Maecker, Mark Genovese, C Garrison Fathman, and Chan C Whiting. 2015. Mass cytometry as a platform for the discovery of cellular biomarkers to guide effective rheumatic disease therapy. *Arthritis research & therapy*, 17:1–9.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Anjali Rao, Dalia Barkley, Gustavo S França, and Itai Yanai. 2021. Exploring tissue architecture using spatial transcriptomics. *Nature*, 596(7871):211–220.

Amy Reece, Bingzhao Xia, Zhongliang Jiang, Benjamin Noren, Ralph McBride, and John Oakey. 2016. Microfluidic techniques for high throughput single cell analysis. *Current opinion in biotechnology*, 40:90–96.

Luca D Schreurs, Alexander F Vom Stein, Stephanie T Jünger, Marco Timmer, Ka-Won Noh, Reinhard Buettner, Hamid Kashkar, Volker Neuschmelting, Roland Goldbrunner, and Phuong-Hien Nguyen. 2025. The immune landscape in brain metastasis. *Neuro-Oncology*, 27(1):50–62.

Aya M Shaaban, Nancy M Salem, and Lamees N Mahmoud. 2024. Cutting-edge approaches to cell segmentation in imaging mass cytometry: A detailed review. In *2024 6th Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, pages 469–474. IEEE.

Natalie Stanley, Ina A Stelzer, Amy S Tsai, Ramin Fallahzadeh, Edward Ganio, Martin Becker, Thanaphong Phongpreecha, Huda Nassar, Sajjad Ghaemi, Ivana Maric, et al. 2020. Vopo leverages cellular heterogeneity for predictive modeling of single-cell data. *Nature communications*, 11(1):3738.

Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. 2021. Cellpose: a generalist algorithm for cellular segmentation. *Nature methods*, 18(1):100–106.

Gemma Team, Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, et al. 2024. Gemma 2: Improving open language models at a practical size. *arXiv preprint arXiv:2408.00118*.

Christina V Theodoris, Ling Xiao, Anant Chopra, Mark D Chaffin, Zeina R Al Sayed, Matthew C Hill, Helene Mantineo, Elizabeth M Brydon, Zexian Zeng, X Shirley Liu, et al. 2023. Transfer learning enables predictions in network biology. *Nature*, 618(7965):616–624.

Luyi Tian, Fei Chen, and Evan Z Macosko. 2023. The expanding vistas of spatial transcriptomics. *Nature Biotechnology*, 41(6):773–782.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Logan A Walsh and Daniela F Quail. 2023. Decoding the tumor microenvironment with spatial technologies. *Nature Immunology*, 24(12):1982–1993.

Sukwon Yun, Jie Peng, Alexandro E Trevino, Chanyoung Park, and Tianlong Chen. 2024. Mew: Multiplexed immunofluorescence image analysis through an efficient multiplex network. In *European Conference on Computer Vision*, pages 127–144. Springer.

Suyuan Zhao, Jiahuan Zhang, Yushuai Wu, Yizhen Luo, and Zaiqing Nie. 2024. Langcell: Language-cell pre-training for cell identity understanding. *arXiv preprint arXiv:2405.06708*.