# MULTI-TASK GRAPH LEARNING FOR CELL–CELL COMMUNICATION INFERENCE IN SPATIAL TRANSCRIPTOMICS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Intercellular communication is fundamental to multicellular life, orchestrating tissue development, maintaining homeostasis, and driving the pathogenesis of complex diseases. Recent advances in spatial transcriptomics have enabled high-resolution, in situ mapping of intricate ligand-receptor (L-R) interactions within native tissue contexts. However, existing computational approaches often underutilize spatial topological information and overlook the distinct regulatory functions of different L-R interaction types, thereby limiting their capacity to capture the dynamic heterogeneity of cellular communication in complex microenvironments. To address these limitations, we propose CellMM, a novel multi-task learning framework powered by multi-relational graph neural networks. CellMM constructs a multi-relational graph that integrates gene expression profiles, spatial coordinates, and L-R interaction types. It employs relational graph convolutional networks to learn expressive node embeddings and leverages a multi-head attention mechanism to adaptively fuse multi-layer graph features, enabling the identification of active L-R signaling events between spatially adjacent cells. Experiments on three simulated datasets and applications to three real spatial transcriptomics datasets show that CellMM outperforms baseline methods in simulated evaluations and recovers biologically plausible communication patterns in real data.

## 1 INTRODUCTION

The highly coordinated nature of multicellular life is based on intricate intercellular signaling networks (Su et al., 2024). At the core of cross-cell communication is the selective binding of specific signaling molecules to recognition molecules on target cell surfaces (Li et al., 2025). Dysregulation of signaling pathways can disrupt the behavior of individual cells and trigger cascading effects, altering the fate of neighboring cells and reshaping the entire local tissue ecology. These biological signaling pathways are activated by specific ligand-receptor (L-R) interactions, which subsequently regulate intracellular signal transduction processes in the recipient cells, influencing their gene expression and functional states (Azhar et al., 2025). Ultimately, these interactions reshape the local microenvironment through cell-cell communication. For instance, in the tumor microenvironment, tumor cells can release exosomes rich in programmed PD-L1. Upon uptake by dendritic cells, these exosomes inhibit antigen-presenting functions and induce T cell exhaustion, thereby enabling effective evasion of immune surveillance and clearance. Therefore, systematically decoding the complex dynamics of intercellular communication is not only crucial for elucidating the underlying causes of major diseases but also forms the foundation for developing novel intervention strategies (Iqbal, 2024).

Cell-cell interactions are mediated through diverse, specialized signaling mechanisms, primarily encompassing autocrine, paracrine, juxtacrine, and endocrine signaling (Aykan, 2013). Autocrine and juxtacrine signaling rely on self-stimulation or direct cell-to-cell contact, respectively, exerting regulatory functions within the local microenvironment. In contrast, paracrine and endocrine signaling operate over longer spatial scales, facilitating coordination and communication between distant cells or tissues. These signaling mechanisms differ in their range, temporal dynamics, and biological functions, collectively forming a multilayered regulatory network for intercellular communication



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

(Kholodenko, 2006). Significant progress has been made in inferring L-R interactions from single-cell transcriptomic data (Shao et al., 2026). Representative methods, such as CellChat (Jin et al., 2021), quantify L-R associations by leveraging differentially expressed ligand and receptor genes in conjunction with the law of mass action. CellPhoneDB (Efremova et al., 2020) identifies potential interactions by evaluating the specific enrichment of L-R interactions between distinct cell populations. Connectome (Hou et al., 2020), on the other hand, defines the weight of L-R interactions as the product of normalized expression values of ligands and receptors at the cell-type level. However, these methods typically do not incorporate spatial information about cell locations, and the lack of modeling for spatial proximity or cell-cell contact relationships may lead to an increased rate of false-positive results (Tang et al., 2023). With the rapid advancement of spatial transcriptomics technologies, recent years have witnessed the emergence of numerous computational methods that effectively integrate spatial positional information to infer L-R interactions and cell-cell communication (Meng et al., 2026). Although these approaches have made substantial strides in modeling spatial architectures, most fail to adequately consider the specific regulatory roles played by distinct L-R interaction types in the cell communication process. This limitation continues to challenge existing models in elucidating the dynamic diversity and functional nuances of intercellular communication within complex tissue microenvironments, impeding the systematic revelation of synergistic, competitive, or cascading regulatory relationships among different signaling modalities in spatial contexts.

In this study, we introduce CellMM, a multi-task learning framework that integrates multi-relational graph neural networks for inferring L-R interactions in spatial transcriptomics data. By constructing a multi-relational graph that fuses gene expression profiles, L-R interaction types, and spatial topological information, the model enables in-depth dissection of cell-cell communication patterns. The primary contributions of this work are summarized as follows:

- To the best of our knowledge, CellMM represents the first computational framework that leverages interdependencies among distinct L-R interactions to predict L-R links, offering a novel modeling perspective for cell communication research.
- We propose a multi-task learning model incorporating multi-relational graph neural networks, which employs relational graph convolutional networks (RGCN) to derive node embeddings, utilizes a multi-head attention mechanism for adaptive fusion of multi-layer features, and incorporates an adaptive multi-task loss function to jointly optimize link prediction and relation classification tasks.
- Experiments on three simulated datasets and three real spatial transcriptomics datasets show that CellMM achieves competitive performance in simulated evaluations and identifies biologically plausible communication patterns in real-data analyses.

## 2 Related Work

Current methods for inferring cell-cell communication from spatial transcriptomics data can be broadly categorized into two primary technical paradigms: probabilistic statistical models and machine learning approaches. In the probabilistic statistical paradigm, optimal transport theory has emerged as a pivotal tool. For instance, COMMOT (Cang et al., 2023) employs collective optimal transport strategies to infer the directionality of spatial signaling and identify downstream genes regulated by these signals. SpaOTsc (Cang & Nie, 2020) leverages structured optimal transport models to transport information from signal-sending cells to spatially corresponding receiver cells, thereby reconstructing intercellular communication relationships. Other statistical strategies are also widely utilized in this domain. Giotto (Dries et al., 2021) constructs cell-cell communication strength scores by computing weighted average expression levels of L-R pairs across different cell types. SpatialDM (Li et al., 2023b) incorporates bivariate Moran's spatial autocorrelation statistics to identify L-R pairs exhibiting significant spatial co-expression patterns. NICHES (Raredon et al., 2023) effectively characterizes signal activity states within local cellular microenvironments by embedding L-R signal proxies. Within the machine learning paradigm, graph neural networks predominate. SpaTalk (Shao et al., 2022) employs a computational framework that integrates graph neural networks with knowledge graphs to systematically model and quantitatively score signaling networks between spatially adjacent cells. spaCI (Tang et al., 2023) utilizes an adaptive graph neural network equipped with an attention mechanism to detect active L-R interactions between neighboring cells.

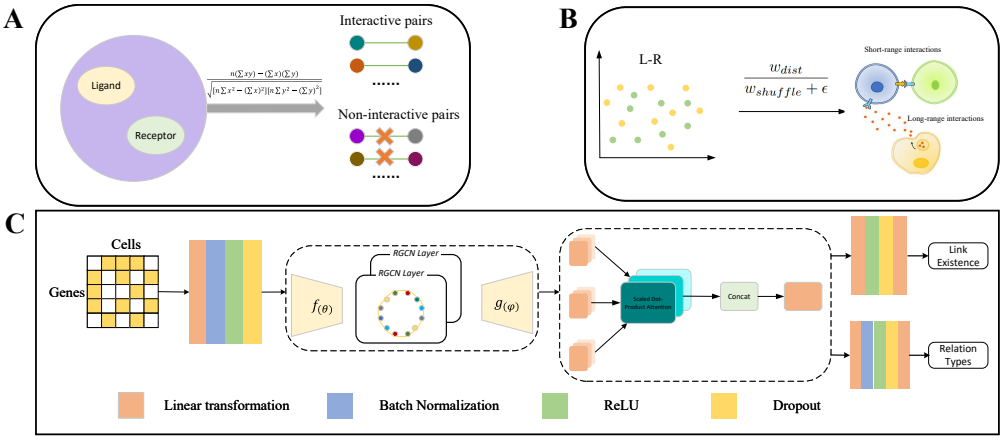


Figure 1: Schematic overview of CellMM. (A) Gene pairs exhibiting significant positive or negative correlations were selected based on Pearson's correlation coefficients. (B) The spatial distribution relationships of L-R pairs are quantified, while statistical tests are employed to identify interactions exhibiting significant spatial distribution patterns. (C) CellMM employs RGCN to perform differentiated information aggregation across diverse types of L-R interactions. It incorporates a multi-head attention mechanism to facilitate multi-scale information fusion, ranging from local neighborhood features to global graph structures. Through a multi-task learning framework, the model jointly optimizes two closely interrelated tasks: L-R link prediction and relation classification.

SpaCcLink (Liu et al., 2025) integrates intracellular downstream signaling responses with spatial co-expression information, employing graph attention networks to identify biologically plausible L-R relationships and effectively associate them with relevant signaling pathways.

# 3 METHOD

## 3.1 COLLECTION OF L-R PAIRS

For real-world data analysis, we integrated data from seven publicly available databases: International Union of Pharmacology (Sharman et al., 2012), CellTalkDB (Shao et al., 2021), Human Plasma Membrane Receptome (Ben-Shlomo et al., 2003), HPRD (Prasad et al., 2009), Fantom5 (Ramilowski et al., 2015), Connectome (Hou et al., 2020), and Database of L-R Partners (Graeber & Eisenberg, 2001). The dataset includes 2,734 mouse L-R pairs and 3,446 human L-R pairs, which were used to screen candidate L-R interactions in CellMM.

## 3.2 GENE PAIRS USED FOR MODEL TRAINING

To construct a training dataset for predicting gene interactions and their types, we calculated the Pearson Correlation Coefficient (PCC) for all gene pairs based on spatial transcriptomic data (Fig. 1A). Based on the distribution of PCC values, we selected gene pairs demonstrating significant positive and negative correlations to construct the positive and negative sample sets for model training. Specifically, gene pairs in the top and bottom 1% quantiles of the PCC distribution were designated as high co-expression and low co-expression candidates, respectively, to represent potential biological interactions and the absence of direct functional associations.

## 3.3 CALCULATION OF GENE INTERACTION TYPES

All ligand and receptor genes are first extracted from the gene expression matrix, with redundant entries removed through merging. Based on this gene set, the expression levels of corresponding genes are retrieved from the spatial expression matrix $X$ to generate a submatrix. Subsequently, a Euclidean distance matrix is computed based on the spatial coordinates of all sample points. For

each L-R pair, the expression vectors of the corresponding ligand and receptor genes are extracted from the submatrix across all spatial points. Points with zero expression are filtered, yielding sets of spatial points positive for ligands $P_l$ and receptors $P_r$. $C \in \mathbf{R}^{|P_l| \times |P_r|}$ is extracted from the distance matrix, where each element $C_{ij} = D(P_{l_i}, P_{r_j})$ represents the Euclidean distance between points $P_{l_i}$ and $P_{r_j}$. $C$ serves as the cost matrix for the optimal transport problem. To enhance numerical stability in subsequent computations, the cost matrix is normalized.

The Sinkhorn algorithm is applied to compute the regularized Wasserstein distance $w_{dist}$ between the ligand and receptor expression distributions, quantifying their dissimilarity (Fig. 1B). To assess the statistical significance of the observed distance, a permutation test is conducted. While preserving the spatial coordinates, the rows of the entire gene expression matrix are randomly permuted. For each permutation, the Wasserstein distance for the L-R pair is recalculated, generating a permuted distance distribution. For each L-R pair, the mean Wasserstein distance $w_{shuffle}$ is computed over n permutations, serving as the expected random distance under the null hypothesis that gene expression is independent of spatial position. The spatial interaction distance ratio can be calculated as:

$$d_{\text{ratio}} = \frac{w_{\text{dist}}}{w_{\text{shuffle}} + 10^{-5}} \tag{1}$$

Based on the values and distribution of $d_{ratio}$, we categorized the spatial interaction modes of L-R pairs into three classes: pairs with $d_{ratio}$ ¡ 1 and falling within the lowest 10% quantile were defined as short-range interactions, indicative of a co-localization tendency; those with $d_{ratio} > 1$ and in the highest 10% quantile were classified as long-range interactions, signifying spatial separation; all remaining pairs were designated as medium-range interactions.

### 3.4 Node Embedding Learning

Spatial gene expression data can be represented as $x \in \mathbb{R}^{m \times n}$, where $m$ represents the number of genes and $n$ denotes the number of cells. We employed a sequential module for gene feature extraction, which consists of a linear transformation, batch normalization, a ReLU activation function, and a dropout regularization layer (Fig. 1C). The primary function of this module is to map the input features from the input_dim-dimensional space $d_{in}$ to the hidden_dim-dimensional space $d_{hidden}$. The process can be represented by formula 2:

$$h_u^{(0)} = Dropout(ReLU(\gamma \frac{(x_u W^T + b) - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta)) \tag{2}$$

$\mathbf{W} \in \mathbb{R}^{d_{hidden} \times d_{in}}$ and $\mathbf{b} \in \mathbb{R}^{d_{hidden}}$ are learnable parameters of the linear layer. $\mu$ and $\sigma^2$ represent the mean and variance of the features, while $\gamma$ and $\beta$ denote the learnable scaling and shifting parameters.

Subsequently, node embeddings are refined across $L$ RGCN layers. For each layer $(1, \ldots, L)$, the embedding of node $u$ is updated by aggregating messages from its neighbors under different relation types:

$$\begin{aligned} \mathbf{h}(u)^{(l)} = \text{ReLU}(&\sum_{r \in \mathcal{R}} \\ &\sum_{v \in \mathcal{N}_r(u)} \frac{1}{|\mathcal{N}_r(u)|} \mathbf{W}_r^{(l)} \mathbf{h}_v^{(l-1)} + \mathbf{W}_0^{(l)} \mathbf{h}_u^{(l-1)}) \end{aligned} \tag{3}$$

In the formulation 3, $\mathcal{N}_r(u)$ represents the neighbors of node $u$ connected via relation $r$, $\mathbf{W}_r^{(l)}$ is the weight matrix specific to relation $r$ at layer $l$, and $\mathbf{W}_0^{(l)}$ accounts for self-loops. To mitigate parameter explosion, we employ basis decomposition, constraining the number of basis matrices to $\min(30, |\mathcal{R}|)$. Following each convolution, dropout is applied to prevent overfitting.

To exploit information from varying depths, we collect embeddings from all RGCN layers and fuse them using a multi-head attention mechanism. The embeddings $(\mathbf{h}(u)^{(1)}, \ldots, \mathbf{h}(u)^{(L)})$ are stacked

into a tensor and processed by a four-head attention layer. The resulting outputs are averaged across the layer dimension to yield the final node embedding $\mathbf{z}_u$:

$$\mathbf{z}_u = \text{Mean}(\text{MultiHeadAttention}(\mathbf{h}(u)^{(1)}, \ldots, \mathbf{h}(u)^{(L)})) \tag{4}$$

### 3.5 Decoding for Multi-Task Prediction

We employ two task-specific decoders that share the node embeddings produced by the graph encoder but use different input representations. The link decoder scores a node pair under a specified candidate relation, whereas the auxiliary relation decoder predicts the spatial category without receiving the target relation index or its embedding as decoder inputs.

**Relation-conditioned link prediction.** Given node embeddings $\mathbf{z}_u, \mathbf{z}_v \in \mathbb{R}^d$ and a candidate relation embedding $\mathbf{e}_r \in \mathbb{R}^d$, we construct the link-prediction feature vector as

$$\mathbf{c}^{\text{link}}_{u,v,r} = [\mathbf{z}_u; \mathbf{z}_u \odot \mathbf{e}_r \odot \mathbf{z}_v; \mathbf{z}_v], \tag{5}$$

where $\odot$ denotes element-wise multiplication and $[\cdot;\cdot]$ denotes concatenation. The link decoder outputs a scalar logit:

$$\hat{y}^{\text{link}}_{u,v,r} = f_{\text{link}}\left(\mathbf{c}^{\text{link}}_{u,v,r}\right). \tag{6}$$

The decoder consists of a linear layer mapping the input to the hidden dimension, a LeakyReLU activation with negative slope $0.1$, dropout, and a final linear layer producing a single output. Applying a sigmoid to this logit gives the predicted link probability under the specified relation $r$.

**Auxiliary spatial-pattern classification.** For relation classification, we construct a separate pair representation that does not contain the target relation embedding:

$$\mathbf{c}^{\text{rel}}_{u,v} = [\mathbf{z}_u; \mathbf{z}_u \odot \mathbf{z}_v; \mathbf{z}_v]. \tag{7}$$

The relation decoder maps this representation to logits over the three spatial categories defined in Section 3.3:

$$\hat{\mathbf{y}}^{\text{rel}}_{u,v} = f_{\text{rel}}\left(\mathbf{c}^{\text{rel}}_{u,v}\right) \in \mathbb{R}^3. \tag{8}$$

This decoder comprises a linear layer to the hidden dimension, batch normalization, a ReLU activation, dropout, and a final linear layer with three outputs. A softmax is applied when class probabilities are required.

The spatial categories provide data-derived pseudo-labels for the auxiliary classification task. The target category supervises the relation decoder but is not supplied as a decoder input. This design requires the classifier to infer the spatial category from node features and the remaining graph context, rather than directly using an encoding of the target label. Both decoders jointly update the shared encoder during training.

### 3.6 Multi-Task Loss Function

The model is trained using a composite loss function that balances the objectives of link prediction and auxiliary spatial-pattern classification.

For link prediction, we use binary cross-entropy with logits, adjusted by a positive weight to account for class imbalance:

$$\mathcal{L}_{\text{link}} = \text{BCEWithLogits}(\hat{y}_{\text{link}}, y_{\text{link}}, \text{pos_weight}) \tag{9}$$

where $y_{\text{link}} \in [0, 1]$ is the ground truth label, and pos_weight is dynamically computed from the training batch.

For auxiliary spatial-pattern classification, let $\mathcal{D}_{\text{rel}}$ denote the set of supervised node pairs with valid spatial-category labels. Let $\tilde{r}_{u,v}$ denote the pseudo-label assigned to pair $(u, v)$ by the procedure in Section 3.3. We minimize the mean cross-entropy loss:

$$\mathcal{L}_{\text{rel}} = \frac{1}{|\mathcal{D}_{\text{rel}}|} \sum_{(u,v)\in\mathcal{D}_{\text{rel}}} \text{CE}\left(\hat{\mathbf{y}}^{\text{rel}}_{u,v}, \tilde{r}_{u,v}\right). \tag{10}$$

The total loss combines both components with an adaptive weighting scheme:

$$\mathcal{L} = \mathcal{L}_{\text{link}} + \alpha \cdot \mathcal{L}_{\text{rel}} \tag{11}$$

The weighting coefficient $\alpha$ is dynamically computed for each training batch and treated as a constant during gradient backpropagation, thereby enabling adaptive balancing of the losses from the two tasks. By detaching the losses in the weighting factor, we ensure that the gradient computation focuses on optimizing each task independently while dynamically adjusting their relative importance based on their current magnitudes. This approach stabilizes training and promotes balanced performance across both tasks.

# 4 EXPERIMENT

## 4.1 DATASETS

To validate the effectiveness of the proposed method, we conducted extensive and systematic experiments on three publicly available spatial transcriptomics datasets and three simulated datasets. The real datasets comprise human melanoma (Li et al., 2023b), human breast cancer (Li et al., 2023a), and mouse brain (Palla et al., 2022). Table 1 presents the statistics of these datasets. Details of the simulated dataset generation process and experiment settings are provided in the Supplementary Notes 1 and 2.

Table 1: Statistics of Datasets.

| Datasets | Cell number | Gene number | Cell type |
|---|---|---|---|
| melanoma | 294 | 12473 | 7 |
| mouse brain | 2689 | 18078 | 8 |
| human breast | 3799 | 21129 | 8 |

## 4.2 EVALUATION OF CELLMM USING SIMULATED DATA

To evaluate the performance of different methods in identifying interacting L-R pairs, we systematically compared four methods, including CellMM, spaCI, SpaTalk and CellChat, across three simulated datasets (Table 2). The evaluation utilized multiple metrics capturing various aspects of classification performance, specifically accuracy (ACC), recall, F1 score, and the area under the receiver operating characteristic curve (AUROC). These metrics collectively provide a comprehensive assessment of the methods' detection capabilities. Our proposed method CellMM achieved best or near-best performance across nearly all datasets and evaluation metrics. On the low-density, high signal-to-noise ratio Simulated Dataset 1, it attained an ACC of 85.81% and an AUROC of 91.73%. On the medium-quality Simulated Dataset 2, the method achieved an ACC of 87.16% and an AUROC of 93.81%. Performance further improved on the high cell density, high-noise Simulated Dataset 3, with an ACC of 92.12% and an AUROC of 96.19%. The F1 scores suggest that CellMM maintained a favorable balance between precision-oriented and recall-oriented performance in these simulated evaluations. The spaCI method performed exceptionally well on Simulated Dataset 3 (ACC: 84.21%, Recall: 90.88%), surpassing SpaTalk in certain metrics; however, its performance declined significantly on Simulated Dataset 1 (Recall: 69.84%), suggesting that spaCI's algorithmic mechanism may be more sensitive to data quality or structure. The SpaTalk method exhibited high sensitivity, achieving the highest recall (79.75%) among all methods on Simulated Dataset 1,

indicating its effectiveness in detecting potential cell-cell interactions. However, its relatively lower F1-score suggests a higher false-positive rate. We employed independent-samples t-tests to assess the differences in performance among the different methods (Supplementary Table S1).

Table 2: Performance comparison with state-of-the-art methods on three simulated datasets. The note* means our model significantly outperforms the baselines based on t-test.

| Dataset | Metric | CellChat | SpaTalk | spaCI | CellMM |
|---|---|---|---|---|---|
| Simulated Dataset 1 | ACC | 42.14 | 74.01 | 71.78 | **85.81*** |
| | Rec | 46.76 | **79.75** | 69.84 | 78.51 |
| | F1 | 52.61 | 67.16 | 67.95 | **74.44*** |
| | AUROC | 60.28 | 81.14 | 78.20 | **91.73*** |
| Simulated Dataset 2 | ACC | 58.01 | 76.51 | 80.48 | **87.16*** |
| | Rec | 52.14 | 75.60 | 84.39 | **85.55*** |
| | F1 | 57.89 | 68.21 | 81.21 | **82.64*** |
| | AUROC | 63.45 | 83.24 | 85.16 | **93.81*** |
| Simulated Dataset 3 | ACC | 62.24 | 80.12 | 84.21 | **92.12*** |
| | Rec | 60.32 | 83.21 | 90.88 | **91.60*** |
| | F1 | 63.65 | 79.08 | 85.20 | **90.38*** |
| | AUROC | 68.21 | 86.12 | 87.42 | **96.19*** |

### 4.3 Evaluation of CellMM Using Real-World Spatial Transcriptomic Data

To further evaluate the performance of the CellMM model, we compared it against several existing methods using three real-world spatial transcriptomics datasets: human melanoma sequencing data, mouse brain sequencing data, and human breast cancer sequencing data. Due to the absence of a gold-standard benchmark for L-R interactions in real datasets, we adopted a widely used evaluation strategy that assesses inter-method consistency by measuring the overlap of identified interactions across methods. Specifically, we quantified inter-method consistency using the Jaccard index, defined as the ratio of L-R pairs commonly identified by two methods to the total number of L-R pairs identified by at least one method. The calculation formula is as follows:

$$J(M_i, M_j) = \frac{|M_i \cap M_j|}{|M_i \cup M_j|} \tag{12}$$

$M_i$ and $M_j$ represent the sets of L-R pairs identified by methods i and j, respectively. A Jaccard index closer to 1 indicates higher consistency between the two methods, reflecting greater concordance of the identified results. The CellMM framework demonstrated high consistency with other leading methods across all three datasets, supporting its robustness across diverse tissue types and species (Fig. 2). Specifically, in the human melanoma dataset, CellMM exhibited the highest Jaccard index with SpaTalk (0.471), reflecting strong concordance in L-R interaction inference, followed by moderate consistency with spaCI (0.381). In the mouse brain dataset, CellMM showed the highest Jaccard consistency with spaCI (0.454). Notably, its consistency with SpaTalk was somewhat reduced in this dataset compared to the human melanoma data, suggesting that tissue microenvironment or species differences may influence the degree of consensus between methods. In the human breast cancer dataset, CellMM again achieved the highest Jaccard index with spaCI (0.437). In contrast, CellChat consistently yielded lower Jaccard indices across all three datasets.

### 4.4 Ablation Study

To validate the effectiveness of key design choices in the CellMM framework, we conducted ablation studies. First, to assess the impact of the multi-task learning mechanism, we compared the full model against a simplified variant that performs only the link prediction task (OLP). Second, to evaluate the contribution of the feature fusion strategy, we tested a no-feature-fusion variant model (NFV) that omits the multi-head attention mechanism for integrating embeddings from all graph neural network layers, instead feeding only the output of the final RGCN into the decoder. The experimental results (Fig. 3) demonstrate that both simplified variants exhibited significant performance declines across all evaluation metrics. Using the F1-score as an example metric, ablation studies conducted on three

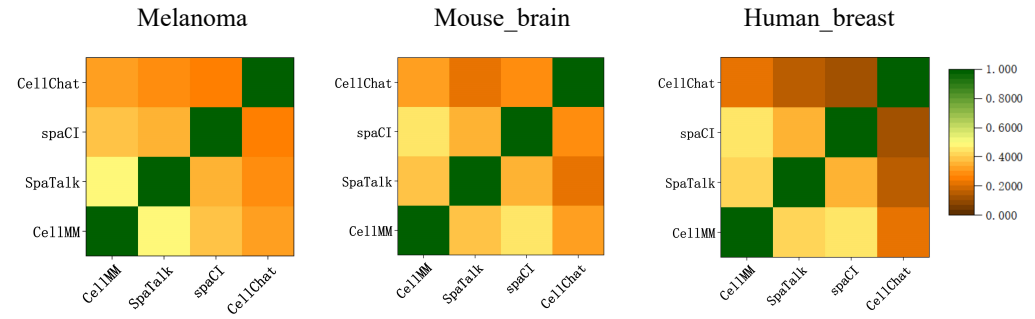


Figure 2: The heatmap quantitatively assesses the pairwise consistency of inferred L-R interaction results across different methods in the three real datasets analyzed, employing the Jaccard index. The figure utilizes a continuous color mapping, wherein color intensity is proportional to the Jaccard index values, thereby visually illustrating the extent of overlap in interactions among the methods.

simulated datasets revealed that the OLP variant exhibited a significant performance decline, with reductions ranging from 3.66% to 9.57%. Similarly, the NFV also demonstrated notable performance degradation, with F1-score decreases ranging from 1.68% to 6.62%. These results establish the pivotal role of the multi-task learning mechanism and the multi-layer feature fusion strategy in enhancing model performance. Notably, the multi-task learning component was identified as the most pivotal element of our framework, with its removal resulting in particularly pronounced performance degradation.

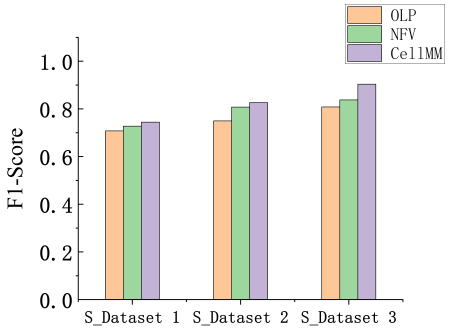


Figure 3: Ablation studies conducted on the multi-task learning and feature fusion variants across three simulated datasets demonstrate that all simplified variants exhibit declines in F1-scores on these datasets.

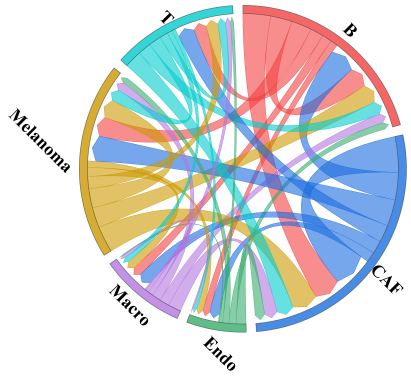


Figure 4: Summary chord diagram of the cell–cell communication network of major cell types in the human melanoma dataset.

### 4.5 Application of CellMM to Human Melanoma Data

To visualize communication patterns among cell types, we constructed an aggregated interaction network wherein edge widths are proportional to the communication strengths between distinct cell types (Fig. 4). Our analysis revealed pronounced interactions between CAF cells and cancer cells, with connection strengths substantially exceeding those observed among other cell types, suggesting the formation of intimate signaling exchanges between these two populations within the tumor microenvironment. We conducted similar analyses on mouse brain and human breast cancer datasets (Supplementary Notes 3 and 4).

To validate the biological significance of the method proposed in this study, we applied CellMM to a human melanoma spatial transcriptomics dataset. Fig. 5 highlights eight L-R interactions with the highest interaction probabilities, which exhibited no significant correlations in the original gene

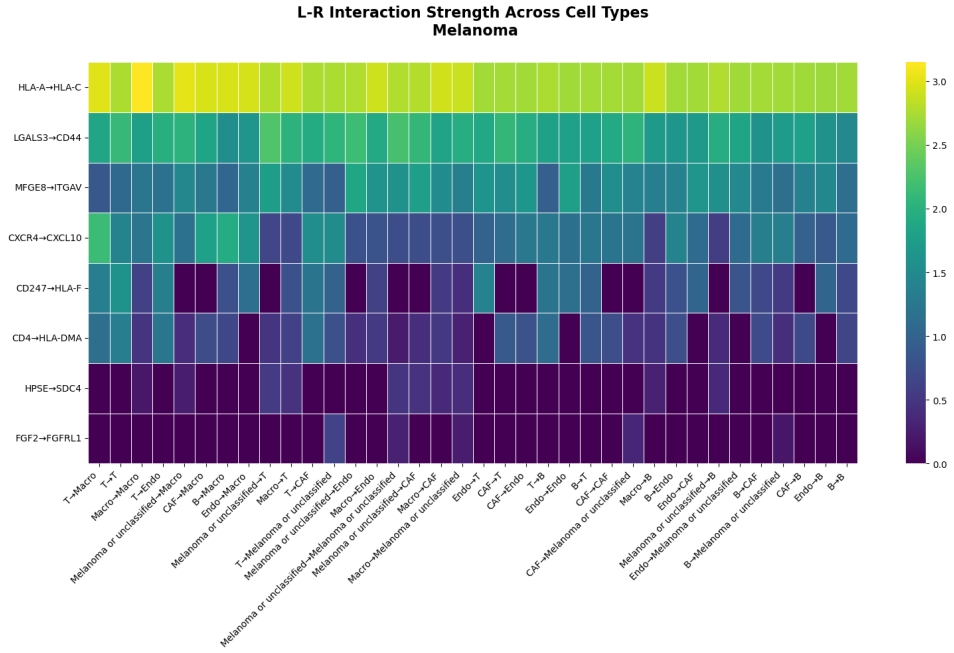


Figure 5: Heatmap shows the interaction strengths of L-R pairs across different cell types in the human melanoma dataset.

expression data. A review of the relevant literature supported the roles of these L-R pairs in key biological processes, including the tumor immune microenvironment and regulation of cell proliferation. We extracted ligand expression levels from sender cells and receptor expression levels from receiver cells, computed the average expression levels of ligands and receptors across each cell type, and thereby generated a detailed interaction strength matrix for cell-type combinations. The interaction between LGALS3 and CD44 is well-established, contributing to cell adhesion, migration, and signal transduction processes, and is commonly implicated in cancer and immune regulation (Lakshminarayan et al., 2014). The interaction between MFGE8 and ITGAV mediates the clearance of apoptotic cells and intercellular communication (Wang et al., 2025). CD4 interacts with HLA-DMA, contributing to antigen presentation and T cell activation (Álvaro-Benito et al., 2020). FGF2 binds to FGFRL1 or FGFRs, promoting angiogenesis and melanoma growth (Kähkönen et al., 2018). HPSE interacts with SDC4, a heparan sulfate proteoglycan receptor, facilitating ECM degradation and melanoma invasion/metastasis (Lopez-Bergami et al., 2008). Low scores for lower-ranked pairs, such as FGF2-FGFRL1 and HPSE-SDC4, imply limited proliferative or invasive signaling in CAFs and endothelial cells, possibly indicating a fibrotic or vascular barrier role rather than active tumor promotion.

## 5 Conclusion

In this study, we introduce CellMM, a computational framework designed to infer spatial cellular communication patterns from spatial transcriptomics data. By incorporating prior knowledge of L–R interaction types, CellMM constructs a multi-relational graph that jointly encodes gene expression profiles and interaction modalities, thereby enabling multidimensional modeling of cell-cell communication patterns. Experimental results suggest that CellMM can improve ligand-receptor interaction inference in simulated settings and provide useful candidate communication patterns for downstream biological analysis. Although these results are encouraging, the existing modeling of L–R interactions in CellMM mainly depends on spatial co-localization patterns, which do not fully capture the underlying biological processes. In future work, we plan to explore the low-dimensional relation embedding vectors of each edge in the learning graph to represent spatial distribution, expression patterns, and functional features.

### AI USE STATEMENT

(This section is **required** and does not count toward the page limit.)

We used an LLM to assist in refining the language, improving readability, and ensuring clarity in various sections of the paper. The model helped with tasks such as sentence rephrasing, grammar checking, and enhancing the overall flow of the text. It is important to note that the LLM was not involved in the ideation, research methodology, or experimental design. All research concepts, ideas, and analyses were developed and conducted by the authors. The contributions of the LLM were solely focused on improving the linguistic quality of the paper, with no involvement in the scientific content or data analysis. The authors take full responsibility for the content of the manuscript, including any text generated or polished by the LLM. We have ensured that the LLM-generated text adheres to ethical guidelines and does not contribute to plagiarism or scientific misconduct.

### ETHICS STATEMENT

(This section is **recommended** and does not count toward the page limit.)

This work adheres to the ICLR Code of Ethics. In this study, no human subjects or animal experimentation was involved. All datasets used were sourced in compliance with relevant usage guidelines, ensuring no violation of privacy. We have taken care to avoid any biases or discriminatory outcomes in our research process. No personally identifiable information was used, and no experiments were conducted that could raise privacy or security concerns. We are committed to maintaining transparency and integrity throughout the research process.

### REPRODUCIBILITY STATEMENT

(This section is **recommended** and does not count toward the page limit.)

We have made every effort to ensure that the results presented in this paper are reproducible. All code and datasets have been made publicly available in an anonymous repository to facilitate replication and verification. The experimental setup, including training steps, model configurations, and hardware details, is described in detail in the paper.

## REFERENCES

Miguel Álvaro-Benito, Eliot Morrison, Friederike Ebner, Esam T Abualrous, Marie Urbicht, Marek Wieczorek, and Christian Freund. Distinct editing functions of natural hla-dm allotypes impact antigen presentation and cd4+ t cell activation. *Cellular & Molecular Immunology*, 17(2):133–142, 2020.

Nuri Faruk Aykan. Intercellular communication. *Current Frontiers and Perspectives in Cell Biology*, 12:361–376, 2013.

Maleeha Azhar, Aroob Hasan Khan, Aroosa Younis Nadeem, Areej Fatima, Ebtesam A Al-Suhaimi, and Adeeb Shehzad. Cellular receptors and cell signaling. In *Cell Signaling*, pp. 46–63. CRC Press, 2025.

Izhar Ben-Shlomo, Sheau Yu Hsu, Rami Rauch, Haili W Kowalski, and Aaron JW Hsueh. Signaling receptome: a genomic and evolutionary perspective of plasma membrane receptors involved in signal transduction. *Science's STKE*, 2003(187):re9–re9, 2003.

Zixuan Cang and Qing Nie. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nature communications*, 11(1):2084, 2020.

Zixuan Cang, Yanxiang Zhao, Axel A Almet, Adam Stabell, Raul Ramos, Maksim V Plikus, Scott X Atwood, and Qing Nie. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nature methods*, 20(2):218–228, 2023.

Ruben Dries, Qian Zhu, Rui Dong, Chee-Huat Linus Eng, Huipeng Li, Kan Liu, Yuntian Fu, Tianxiao Zhao, Arpan Sarkar, Feng Bao, et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome biology*, 22(1):78, 2021.

Mirjana Efremova, Miquel Vento-Tormo, Sarah A Teichmann, and Roser Vento-Tormo. Cellphonedb: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nature protocols*, 15(4):1484–1506, 2020.

Thomas G Graeber and David Eisenberg. Bioinformatic identification of potential autocrine signaling loops in cancers from gene expression profiles. *Nature genetics*, 29(3):295–300, 2001.

Rui Hou, Elena Denisenko, Huan Ting Ong, Jordan A Ramilowski, and Alistair RR Forrest. Predicting cell-to-cell communication networks using natmi. *Nature communications*, 11(1):5011, 2020.

Mudassir Iqbal. Systems biology approaches to deciphering complex biological networks: Implications for health and disease. *Multidisciplinary Journal of Biochemistry*, 1(2):74–83, 2024.

Suoqin Jin, Christian F Guerrero-Juarez, Lihua Zhang, Ivan Chang, Raul Ramos, Chen-Hsiang Kuan, Peggy Myung, Maksim V Plikus, and Qing Nie. Inference and analysis of cell-cell communication using cellchat. *Nature communications*, 12(1):1088, 2021.

TE Kähkönen, KK Ivaska, M Jiang, KG Büki, HK Väänänen, and PL Härkönen. Role of fibroblast growth factor receptors (fgfr) and fgfr like-1 (fgfrl1) in mesenchymal stromal cell differentiation to osteoblasts and adipocytes. *Molecular and Cellular Endocrinology*, 461:194–204, 2018.

Boris N Kholodenko. Cell-signalling dynamics in time and space. *Nature reviews Molecular cell biology*, 7(3):165–176, 2006.

Ramya Lakshminarayan, Christian Wunder, Ulrike Becken, Mark T Howes, Carola Benzing, Senthil Arumugam, Susanne Sales, Nicholas Ariotti, Valerie Chambon, Christophe Lamaze, et al. Galectin-3 drives glycosphingolipid-dependent biogenesis of clathrin-independent carriers. *Nature cell biology*, 16(6):592–603, 2014.

Haochen Li, Tianxing Ma, Minsheng Hao, Wenbo Guo, Jin Gu, Xuegong Zhang, and Lei Wei. Decoding functional cell–cell communication events by multi-view graph learning on spatial transcriptomics. *Briefings in Bioinformatics*, 24(6):bbad359, 2023a.

Yongxing Li, Fengshuo Liu, Qingjin Cai, Lijun Deng, Qin Ouyang, Xiang H-F Zhang, and Ji Zheng. Invasion and metastasis in cancer: molecular insights and therapeutic targets. *Signal transduction and targeted therapy*, 10(1):57, 2025.

Zhuoxuan Li, Tianjie Wang, Pentao Liu, and Yuanhua Huang. Spatialdm for rapid identification of spatially co-expressed ligand–receptor and revealing cell–cell communication patterns. *Nature communications*, 14(1):3995, 2023b.

Jingtao Liu, Litian Ma, Fen Ju, Chenguang Zhao, and Liang Yu. Spacclink: exploring downstream signaling regulations with graph attention network for systematic inference of spatial cell–cell communication. *BMC biology*, 23(1):44, 2025.

Pablo Lopez-Bergami, Boris Fitchman, and Ze'ev Ronai. Understanding signaling cascades in melanoma. *Photochemistry and photobiology*, 84(2):289–306, 2008.

Yangyang Meng, Minhao Yao, Tong Han, Huandong Zhao, Zhonghua Liu, and Baoshan Ma. Inferring directed gene regulatory networks from single-cell ribonucleic acid sequencing data via multi-view contrastive learning. *Engineering Applications of Artificial Intelligence*, 172:114350, 2026.

Giovanni Palla, Hannah Spitzer, Michal Klein, David Fischer, Anna Christina Schaar, Louis Benedikt Kuemmerle, Sergei Rybakov, Ignacio L Ibarra, Olle Holmberg, Isaac Virshup, et al. Squidpy: a scalable framework for spatial omics analysis. *Nature methods*, 19(2):171–178, 2022.

TS Keshava Prasad, Kumaran Kandasamy, and Akhilesh Pandey. Human protein reference database and human proteinpedia as discovery tools for systems biology. In *Reverse Chemical Genetics: Methods and Protocols*, pp. 67–79. Springer, 2009.

Jordan A Ramilowski, Tatyana Goldberg, Jayson Harshbarger, Edda Kloppmann, Marina Lizio, Venkata P Satagopam, Masayoshi Itoh, Hideya Kawaji, Piero Carninci, Burkhard Rost, et al. A draft network of ligand–receptor-mediated multicellular signalling in human. *Nature communications*, 6(1):7866, 2015.

Micha Sam Brickman Raredon, Junchen Yang, Neeharika Kothapalli, Wesley Lewis, Naftali Kaminski, Laura E Niklason, and Yuval Kluger. Comprehensive visualization of cell–cell interactions in single-cell and spatial transcriptomics with niches. *Bioinformatics*, 39(1):btac775, 2023.

Chunfeng Shao, Yuanshen Zhao, Yinsheng Chen, Jingxian Duan, Zeyu Zhang, Rongpin Wang, Dong Liang, Dongyue Chen, and Zhi-Cheng Li. Integrating whole-slide images and transcriptomic data for survival analysis using multimodal attention networks. *Engineering Applications of Artificial Intelligence*, 174:114461, 2026.

Xin Shao, Jie Liao, Chengyu Li, Xiaoyan Lu, Junyun Cheng, and Xiaohui Fan. Celltalkdb: a manually curated database of ligand–receptor interactions in humans and mice. *Briefings in bioinformatics*, 22(4):bbaa269, 2021.

Xin Shao, Chengyu Li, Haihong Yang, Xiaoyan Lu, Jie Liao, Jingyang Qian, Kai Wang, Junyun Cheng, Penghui Yang, Huajun Chen, et al. Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with spatalk. *Nature communications*, 13(1): 4429, 2022.

Joanna L Sharman, Helen E Benson, Adam J Pawson, Veny Lukito, Chidochangu P Mpamhanga, Vincent Bombail, Anthony P Davenport, John A Peters, Michael Spedding, Anthony J Harmar, et al. Iuphar-db: updated database content and new features. *Nucleic acids research*, 41(D1): D1083–D1088, 2012.

Jimeng Su, Ying Song, Zhipeng Zhu, Xinyue Huang, Jibiao Fan, Jie Qiao, and Fengbiao Mao. Cell–cell communication: new insights and clinical implications. *Signal transduction and targeted therapy*, 9(1):196, 2024.

Ziyang Tang, Tonglin Zhang, Baijian Yang, Jing Su, and Qianqian Song. spaci: deciphering spatial cellular communications through adaptive graph model. *Briefings in Bioinformatics*, 24(1): bbac563, 2023.

Wenhui Wang, Jiming Chen, Shibo Wang, Xinhai Sun, Jie Yang, Pengfei Yu, Guinv Hu, Jiang Wang, Jing Zhang, Shuya Qiao, et al. Mfge8 induces anti-pd-1 therapy resistance by promoting extracellular vesicle sorting of pd-l1. *Cell Reports Medicine*, 6(2), 2025.