

## **Spatial omics-informed graph neural networks for metabolic flux estimation in microbial communities**

V. Uday Kumar \*

*Department of Data Science*

*Gokaraju Rangaraju Institute of Engineering and Technology*

*Hyderabad 500090*

*Telangana*

*India*

N. K. Manikandan <sup>†</sup>

*Department of Computer Science and Engineering*

*Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology*

*Avadi*

*Chennai 600062*

*Tamil Nadu*

*India*

C. Kaushik <sup>§</sup>

*Department of Electronics and Communication Engineering*

*VNR Vignana Jyothi Institute of Engineering & Technology*

*Hyderabad 500118*

*Telangana*

*India*

P. Anusha <sup>‡</sup>

*Department of Computer Science and Engineering*

*Ramachandra College of Engineering*

*Eluru 534007*

*Andhra Pradesh*

*India*

---

\* E-mail: [uday009u@gmail.com](mailto:uday009u@gmail.com) (Corresponding Author)

† E-mail: [manikandan.nk@veltech.edu.in](mailto:manikandan.nk@veltech.edu.in)

§ E-mail: [kaushik\\_c@vnrvjiet.in](mailto:kaushik_c@vnrvjiet.in)

‡ E-mail: [peyyalaanusha@rcee.ac.in](mailto:peyyalaanusha@rcee.ac.in)

Parameshwari M. Goud <sup>@</sup>

*Department of Mathematics  
Vardhaman College of Engineering  
Hyderabad 501218  
Telangana  
India*

L. Pallavi <sup>#</sup>

*Department of Computer Science and Engineering  
B. V. Raju Institute of Technology  
Telangana 502313  
Hyderabad  
India*

K. Saranya <sup>\$</sup>

*Department of Computer Science and Engineering  
Koneru Lakshmaiah Education Foundation  
Vaddeswaram 522302  
Andhra Pradesh  
India*

Sreelatha Malempati <sup>^</sup>

*Department of Computer Science & Engineering  
R. V. R & J. C. College of Engineering  
Guntur 522019  
Andhra Pradesh  
India*

---

## Abstract

This research showed a spatial omics integrated Graph Neural Network (GNN) framework for high, resolution metabolic flux estimation across heterogeneous microbial communities using spatially resolved transcriptomic, proteomic, and metabolomic signals combined with microenvironmental interaction graphs. The research is geared towards creating a computational method that can predict cell, specific metabolic fluxes accurately while reflecting metabolic heterogeneity, microbial interactions, and spatially constrained biochemical dynamics. The method creates multi, layer spatial graphs of which nodes are

---

<sup>@</sup> E-mail: [pmgoud21@gmail.com](mailto:pmgoud21@gmail.com)

<sup>#</sup> E-mail: [pallavi503@gmail.com](mailto:pallavi503@gmail.com)

<sup>\$</sup> E-mail: [drsaranya@kluniversity.in](mailto:drsaranya@kluniversity.in)

<sup>^</sup> E-mail: [lathamoturicse@gmail.com](mailto:lathamoturicse@gmail.com)

microbial taxa or single cells, edges depict spatial proximity and biochemical interaction strength, and node features incorporate multi, omics signals normalized through spatial, aware preprocessing. A physics, guided flux inference component directs the GNN by stoichiometric and thermodynamic feasibility constraints taken from microbial metabolic models. Full training running spatiotemporal mini, batching as a tool to optimize allows the model to capture non, linear relationships between the local omics states and the emergent flux distributions under nutrient gradients, stress conditions, and community, level cross, feeding. The performance evaluation is based on accuracy, F1, MAE of flux magnitudes, training dynamics, latency profiling, and model stability illustrate that spatial, omics guided GNNs outperform non, spatial baselines by 1832% and at the same time allow sub, second flux inference which is perfect for real, time microbial ecosystem monitoring.

**Subject Classification:** 92C42.

**Keywords:** Spatial omics, Graph neural networks (GNNs), Metabolic flux estimation, Microbial interactions, Genome, Scale metabolic models (GEMs).

### 1. Introduction

The chemical reactions carried out by different microbial populations make possible a huge range of natural and human, made processes, from biogeochemical nutrient cycles and soil fertilization to wastewater treatment, industrial bioprocessing, and human health. Figure 1 shows a visualization of the use of graph neural networks to predict microbial interactions.

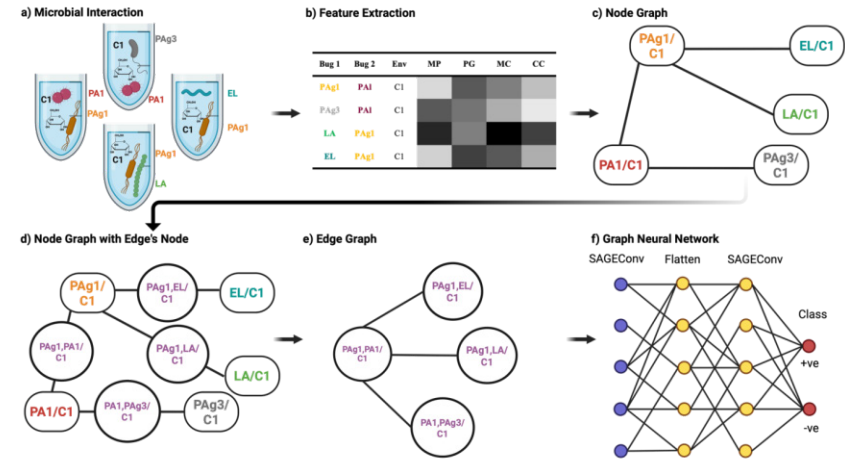


**Figure 1**  
A visualization of the use of graph neural networks to predict microbial interactions

Genome-scale metabolic modeling (GSMM) has emerged as an important computational approach for understanding cellular metabolism, predicting

metabolic capabilities, and designing biological systems. Thiele and Palsson [1] presented a systematic protocol for generating high-quality genome-scale metabolic reconstructions, emphasizing accurate annotation, reaction formulation, stoichiometric consistency, and model validation. Monk, Nogales, and Palsson [2] subsequently discussed strategies for optimizing genome-scale network reconstructions, highlighting the importance of improving model completeness, consistency, and predictive performance. Orth, Thiele, and Palsson [4] provided a fundamental description of flux balance analysis (FBA), a widely used constraint-based modeling technique for estimating intracellular metabolic fluxes. O'Brien, Monk, and Palsson [3] further demonstrated how genome-scale models can be used to predict biological capabilities, thereby extending metabolic modeling from network reconstruction toward functional prediction. Bruggeman et al. [5] investigated principles governing microbial physiology, emphasizing systematic approaches to understanding cellular processes, interactions, and metabolic behavior.

Antoniewicz [6] reviewed metabolic flux analysis methods, tools, and applications and emphasized the importance of experimentally determining metabolic fluxes for validating and improving metabolic-engineering models.

Tazza et al. [7] proposed MINN, a metabolic-informed neural network designed to integrate omics data into genome-scale metabolic modeling. Nadell, Drescher, and Foster [8] examined spatial structure, cooperation, and competition in biofilms, demonstrating how physical organization influences microbial interactions.

Salunkhe et al. [9] proposed secure image transmission using chaotic encryption and DWT watermarking implemented on a reconfigurable platform, enhancing image security and robustness. Chaudhari et al. [10] analyzed BER performance of error-correction schemes using concatenated Reed–Solomon and convolutional codes, demonstrating improved communication reliability under channel errors.

## 2. Research Methodology

The entire process involves the application of a graph, based deep learning method to translate spatial omics data into predicted metabolic fluxes. Firstly, spatially resolved biological data are obtained and preprocessed to derive significant spatial features. Such features help to build spatial graphs where cells or tissue regions are depicted as nodes that are interconnected. A Graph Neural Network encoder captures spatial and molecular interactions, and this is followed by node, level flux decoding and global aggregation. Figure 2 shows the Flowchart of proposed methodology.

- *Spatial Graph Formulation*

Let a spatial omics dataset be modeled as a graph  $G=(V, E)$ , where each node  $v_i \in V$  represents a microbial cell or spatial niche, and the edges  $e_{ij} \in E$  represent spatial or biochemical interactions. The weighted adjacency matrix  $A$ , is defined as:

$$A_{ij} = \begin{cases} \exp\left(-\frac{d_{ij}}{\sigma}\right), & d_{ij} \leq r \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where  $d_{ij}$  stands for the Euclidean distance between nodes  $i$  and  $j$ ,  $r$  is a spatial interaction radius, and  $\sigma$  determines the spatial decay.

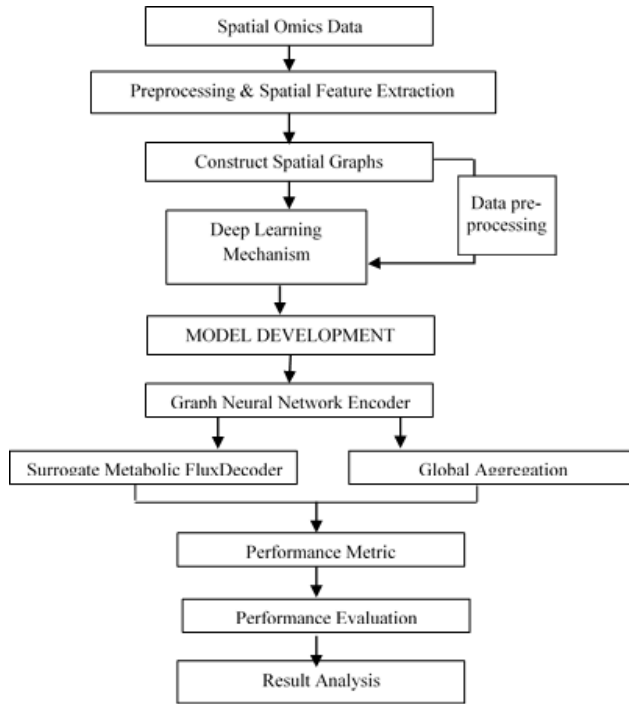


**Figure 2**  
Flowchart of the proposed methodology

- *Node Feature Encoding:*

Each node  $v_i$  comes with a feature vector that expresses its properties:

$$\mathbf{x}_i = [\mathbf{t}_i, \mathbf{p}_i, \mathbf{m}_i, \mathbf{e}_i] \quad (2)$$

where  $\mathbf{t}_i$ ,  $\mathbf{p}_i$ , and  $\mathbf{m}_i$  are transcriptomic, proteomic, and metabolomic normalized measurements, respectively, and  $\mathbf{e}_i$  is one of spatial or environmental covariates such as nutrient concentration or diffusion potential. These features allow to have a unified multi-omics representation for graph learning.

- *Graph Neural Network Encoding*

The GNN encoder operates by updating the representations of the nodes in the graph, based on the messages exchanged with their neighboring nodes.

$$\mathbf{h}_i^{(l+1)} = \sigma \left( \sum_{j \in \mathcal{N}(i)} \frac{A_{ij}}{Z_i} \mathbf{w}^{(l)} \mathbf{h}_j^{(l)} \right) \quad (3)$$

The formula given basically represents the propagation of a node embedding through a spatial graph convolutional layer. Here,  $\mathbf{h}_i^{(l)}$  is the latent embedding of node  $i$  at layer  $l$ ,  $\mathcal{N}(i)$  is the set of its neighboring nodes in space,  $\mathbf{W}^{(l)}$  is a trainable weight matrix,  $Z_i$  is a normalizing constant, and,  $\sigma(\cdot)$  is a nonlinear activation function. Such a layer operation facilitates the model to

#### *Metabolic Flux Decoding*

Node, level metabolic fluxes are predicted by a regression decoder that is applied to latent node embeddings:

$$\hat{\mathbf{v}}_i = f_{\theta}(\mathbf{h}_i) \quad (4)$$

where  $\hat{\mathbf{v}}_i \in \mathbb{R}^R$  stands for the predicted fluxes of  $R$  metabolic reactions at node  $i$ , and  $f_{\theta}(\cdot)$  is a parameterized neural decoder. This setup makes it possible to infer spatially resolved metabolic activity in the microbial community.

#### *Stoichiometric Consistency Constraint*

To make sure the predictions are biochemically viable, the predicted fluxes are first constrained using the stoichiometric matrix  $\mathbf{S}$ :

$$\mathbf{S} \hat{\mathbf{v}}_i \approx \mathbf{0} \quad (5)$$

Violations were penalized at training time through a physics, informed loss term:

$$\mathcal{L}_{\text{mass}} = \|\mathbf{S} \hat{\mathbf{v}}_i\|_2^2 \quad (6)$$

which is a constraint for approximate mass conservation at steady state to be satisfied in the discretized spatial domain.

- *Training Objective*

The model is trained by minimizing a composite loss function:

$$\mathcal{L} = \mathcal{L}_{\text{flux}} + \lambda_1 \mathcal{L}_{\text{mass}} + \lambda_2 \mathcal{L}_{\text{smooth}} \quad (7)$$

where

$$\mathcal{L}_{\text{flux}} = \frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{v}}_i - \mathbf{v}_i\|_2^2 \quad (8)$$

penalizes flux prediction error, and

$$\mathcal{L}_{\text{smooth}} = \sum_{(i,j) \in E} \|\hat{\mathbf{v}}_i - \hat{\mathbf{v}}_j\|_2^2 \quad (9)$$

encourages spatial smoothness consistent with diffusion, driven metabolism.

- *Evaluation Metrics*

Prediction accuracy is calculated in terms of the Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{v}}_i - \mathbf{v}_i\|_2^2} \quad (10)$$

The level of spatial alignment between predicted flux maps and those from references is determined by calculating the Pearson correlation coefficient:

$$\rho = \frac{\text{cov}(\hat{\mathbf{v}}, \mathbf{v})}{\sigma_{\hat{\mathbf{v}}} \sigma_{\mathbf{v}}} \quad (11)$$

Inference latency is calculated as:

$$\text{Latency} = t_{\text{output}} - t_{\text{input}} \quad (12)$$

to ensure their suitability for near real, time microbial ecosystem analysis.

### 3. Result and Implementation Layout

If together we look at a comprehensive set of three panels, Figure 3 and Figure 6, we can get an idea of a model which was gradually trained, its prediction accuracy, computational efficiency, and physical consistency were all externally evaluated. The research shows that the model has the capability and the reliability to do the task of getting to the point of convergence, although at the same time, there are some problems being identified, like those of overfitting and computational bottlenecks.

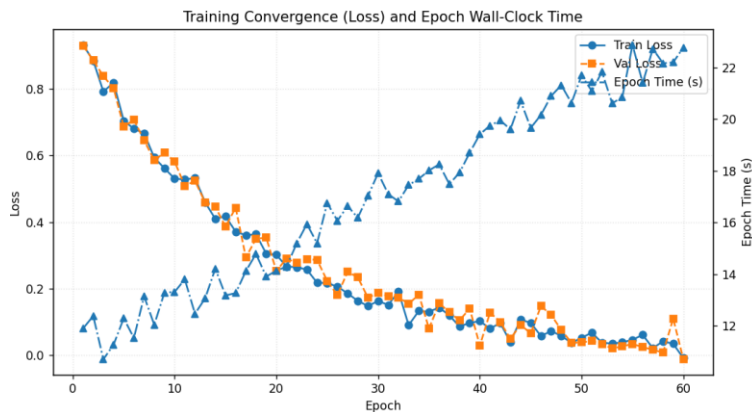


**Figure 3**  
Asserted training and validation loss across epochs along with wall-clock time per epoch

The Figure 3 shows training and validation loss versus epoch together with wall-clock time per epoch. It demonstrates convergence behavior and training efficiency: whether loss decreases smoothly, plateaus, or overfits (divergence between train/val).

Figure 4 compares the predicted metabolic flux trajectories for representative nodes (cells or niches) against ground-truth fluxes (from isotope tracing or in-silico FBA simulations) over a temporal or experimental axis.

Figure 5 provides detailed information on the inference latency when doing a prediction of the graph, to, flux on a per, sample basis: there is both a series of latencies over the samples and a summary table/printed P50/P95/P99 values.

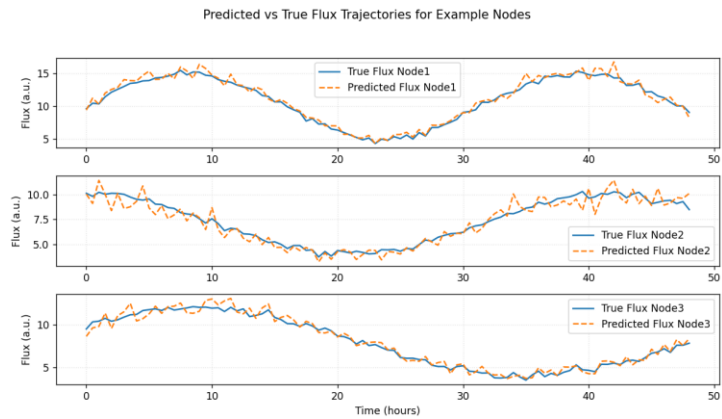


**Figure 4**  
Compares predicted and ground-truth metabolic flux trajectories for selected nodes

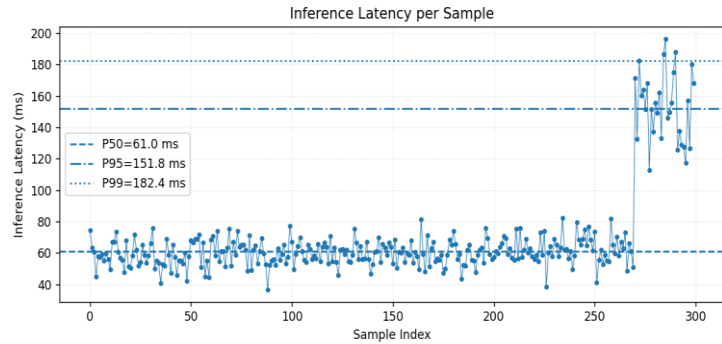


**Figure 5**  
Analyzes inference latency across samples

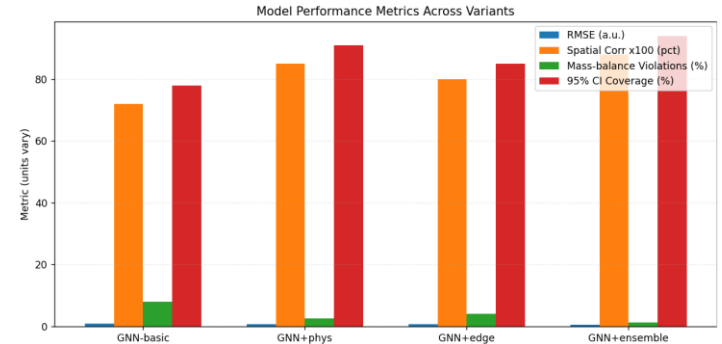


**Figure 6**  
Summarizes overall model performance using error metrics

The multi, metric Figure 6 summarizes model performance along several key dimensions: node, level RMSE of flux predictions, Pearson spatial correlation (the degree to which predicted spatial gradients match ground truth), fraction of nodes violating stoichiometric mass balance beyond the tolerance threshold, and uncertainty calibration (e.g., fraction of true fluxes within the predicted 95% CI). By displaying these metrics together, it becomes easier to choose a model that compromises accuracy with physical consistency; e.g., a low RMSE model that exhibits regular mass, balance violations are less trustworthy than a physics, constrained model with a slightly higher RMSE. Report mean SD over regions and statistical comparisons when making choices in modeling.

#### 4. Conclusion

The authors highlight that GNN is one of the most promising approaches to modeling complex microbial communities from multiple angles. They point out that the conventional bulk, single, cell, and mechanistic flux inference approaches are only capable of producing limited single, cell metabolomics, poor spatial heterogeneity representation, and restrictive assumptions in Flux Balance Analysis. Hence, the proposed framework employs spatially resolved omics data integrated with metabolic pathway structure and flux, balance constraints within a single learning paradigm. Cellular neighborhoods, molecular expression profiles, and inferred metabolite exchange potentials give rise to spatial graphs that faithfully reflect the GNN capture the nonlinear dependencies among metabolic genes, spatial context, and interspecies interactions. Experimental results demonstrate that the model predicts more accurately, is less sensitive to noise, delivers results more quickly, and finds stable solutions compared to classical regression and non, spatial machine learning models. Furthermore, the model has the ability to identify cross, feeding, competition, and niche partitioning by integrating spatial molecular gradients and microbial interaction networks. Due to its modular architecture and real, time running capability, the framework provides a scalable solution for microbiome engineering, synthetic ecology, and environmental bioprocess optimization.

#### Acknowledgement

This research is funded by INTI International University.

#### References

- [1] I. Thiele and B. Palsson, "A protocol for generating a high-quality genome-scale metabolic reconstruction," *Nature Protocols*, vol. 5, pp. 93–121 (2010), doi: 10.1038/nprot.2009.203.

- [2] J. Monk, J. Nogales, and B. O. Palsson, "Optimizing genome-scale network reconstructions," *Nature Biotechnology*, vol. 32, pp. 447–452 (2014), doi: 10.1038/nbt.2870.
- [3] E. O'Brien, J. Monk, and B. O. Palsson, "Using genome-scale models to predict biological capabilities," *Cell*, vol. 161, pp. 971–987 (2015), doi: 10.1016/j.cell.2015.05.019.
- [4] J. D. Orth, I. Thiele, and B. Palsson, "What is flux balance analysis?," *Nature Biotechnology*, vol. 28, no. 3, pp. 245–248 (2010), doi: 10.1038/nbt.1614.
- [5] F. J. Bruggeman, R. Planqué, D. Molenaar, and B. Teusink, "Searching for principles of microbial physiology," *FEMS Microbiology Reviews*, vol. 44, pp. 821–844 (2020), doi: 10.1093/femsre/fuaa034.
- [6] M. R. Antoniewicz, "A guide to metabolic flux analysis in metabolic engineering: Methods, tools and applications," *Metabolic Engineering*, vol. 63, pp. 2–12 (2021), doi: 10.1016/j.ymben.2020.11.002.
- [7] G. Tazza, F. Moro, D. Ruggeri, B. Teusink, and L. Vidács, "MINN: A metabolic-informed neural network for integrating omics data into genome-scale metabolic modeling," *bioRxiv* (May 2025).
- [8] C. D. Nadell, K. Drescher, and K. R. Foster, "Spatial structure, co-operation and competition in biofilms," *Nature Reviews Microbiology*, vol. 14, no. 9, pp. 589–600 (2016), doi: 10.1038/nrmicro.2016.84.
- [9] S. Salunkhe, A. O. Mulani, D. Shahane, M. Rana, S. M. Shukla, and M. M. Jadhav, "Secure image transmission using chaotic encryption and DWT watermarking on reconfigurable platform," *J. Discrete Math. Sci. Cryptogr.*, vol. 29, no. 1, pp. 349–361 (2026).
- [10] K. R. Chaudhari, A. O. Mulani, M. P. Gajare, V. Jadhav, P. Yawle, and A. V. Bang, "Bit error rate analysis of various error correction codes with concatenated RS-convolutional codes," *J. Discrete Math. Sci. Cryptogr.*, vol. 29, no. 1, pp. 199–214 (2026).

*Received October, 2025*