Ilia State University

Faculty of Natural Sciences and Medicine

Master's program: Applied Genetics

MASTER'S THESIS

HistoMamba: Predicting Spatial Transcriptomics from Histological WSIs Using Mamba State Space Models

Author: Demetre Mchedlishvili M.D

Supervisor: Vincenzo Lagani Ph.D

June 22, 2025

Abstract

Histological analysis has long served as the foundation of diagnostic pathology, yet standard hematoxylin and eosin (H&E) staining provides only morphological information, leaving the molecular and spatial heterogeneity of tissues largely unexplored. Spatial transcriptomics (ST) addresses this limitation by mapping gene expression in situ, but its high experimental cost and technical demands prevent widespread clinical adoption. A major computational challenge, therefore, is to accurately predict spatial gene expression landscapes directly from routine H&E whole slide images (WSIs), while ensuring both generalizability and biological interpretability.

In this thesis, I present a novel hybrid deep learning architecture—**HistoMamba**—that integrates *Mamba Structured State Space Models* (SSMs) with multi-scale *Convolutional Neural Networks* (CNNs). My approach leverages CNNs to extract detailed local morphometric features, while employing the Mamba SSM to model long-range spatial dependencies through an input-adaptive state-space mechanism. By synergistically combining these components, I enable the model to learn both fine-grained histological patterns and global spatial context, allowing for robust inference of gene expression profiles at high spatial resolution.

I trained model on HEST1-k datasets lung samples and validated it on the lung organoid subset of the Villacampa collection within the SpaRED dataset, which provides matched WSIs and spot-resolved transcriptomic data. The model's predictions were evaluated using clustering and correlation metrics, with K-means clustering of predicted gene expression profiles yielding a highest silhouette score of 0.5913 for two well-separated clusters, indicating strong biological coherence. Throughout the project, I was constrained by computational and financial resources, which limited the scope of benchmarking and hyperparameter exploration. Nevertheless, ablation experiments confirmed the added value of the state space path for modeling tissue-wide dependencies.

My model is relatively small in size especially compared to foundational models and yet it demonstrates the feasibility and advantages of hybrid SSM–CNN architectures for spatial transcriptomics prediction and highlight a scalable computational pipeline for non-destructive molecular annotation of tissue specimens. This work contributes to the advancement of computational histopathology and demonstrates the potential of artificial intelligence in personalized medicine, biomarker discovery, and spatial tissue mapping.

Disclosure of AI Tool Usage

During the preparation of this thesis, various AI-based tools were utilized to enhance productivity and accuracy. Specifically, AI-assisted coding tools such as VSCode,PyCharm,were employed to facilitate code development and debugging; Grammarly was used for language editing and grammatical corrections; and Jenni AI assisted with reference management and formatting. All AI-generated outputs were critically reviewed, verified, and adapted by the author to ensure accuracy and originality. The scientific content, data analysis, and interpretation remain the sole responsibility of the author.