BMSeg: An Automated System for Cell Segmentation and Classification in Bone Marrow Biopsies

Lang Xiong*, Joshua Liu*, Vivek Pujara BS, Khang Le PhD, Louis J. Vaickus MD PhD, Joshua J. Levy PhD
Emerging Diagnostic and Investigative Technologies, Department of Pathology, Dartmouth Hitchcock Medical Center
*equal contribution

Dartmouth Health

ABSTRACT

- The manual pathological review of bone marrow biopsies is a time-consuming and subjective process prone to misdiagnosis.
- Many bone marrow biopsy whole slide images contain manual annotations, resulting in many unidentified cells.
- Cell niches are increasingly important to diagnose diseases
- Graph neural network (GNN) could utilize the spatial relationship between cells to classify cell types and identify cell niches.

INTRODUCTION

Bone Marrow Analysis: A Critical Diagnostic Tool

- Forms the cornerstone for diagnosing hematologic diseases.
- The manual review process is time-consuming, labor-intensive, and subjective, leading to misdiagnosis in up to 12% of cases.

Challenges in Current Analytical Methods:

- While H&E stained tissue biopsies allow for crucial spatial assessment, they have significant annotation costs, creating a data bottleneck.
- Most current AI tools focus on bone marrow *aspirates*, which lose the tissue architecture that is preserved in biopsies.
- There is no general-purpose, large-scale model for comprehensive bone marrow *biopsy* analysis.

Machine Learning Approaches:

- Convolutional Neural Networks (CNNs) have been popular for cell classification but do not leverage the surrounding spatial architecture, treating each cell in isolation.
- There is a need for a methodology that can model the complex cellular neighborhoods found in bone marrow tissue.

Our Approach: Graph Neural Networks (GNNs)

- GNNs are neural networks that can pass messages between adjacent cells (nodes) in a graph.
- This allows for node-level inference (cell type classification) that utilizes information from neighboring cells and the broader tissue micro-architecture.
- By modeling these spatial relationships, GNNs are better suited to handle the challenges of complex biopsy images and can learn effectively from sparsely annotated data through techniques like self-supervised pre-training.

Figure 1: Left - Aspirate; Right - Biopsy

METHODS

Experimental Design:

- Dataset of 23 H&E stained bone marrow biopsy Whole Slide Images (WSIs).
- Biopsies were chosen over aspirates to preserve tissue architecture.
- Slides were sparsely annotated, with ~300 ground-truth labels per WSI across 14 cell types.
- A dense Region of Interest (ROI) was extracted from each WSI based on annotations

Algorithmic Methods:

- **Cell Segmentation:** A pre-trained HoverNet model was used to detect all cell nuclei within the ROIs, creating a comprehensive cellular map.
- **Feature Extraction:** A pre-trained ResNet18 CNN was used to generate a rich, 512-dimensional feature vector for each detected cell.
- **Graph Neural Network:**
  - The GNN uses the CNN feature vectors as node features and connects neighboring cells within a 150-pixel radius to form the graph.
  - A self-supervised pre-training step was used to learn from all detected cells, followed by fine-tuning on the sparse ground-truth labels.
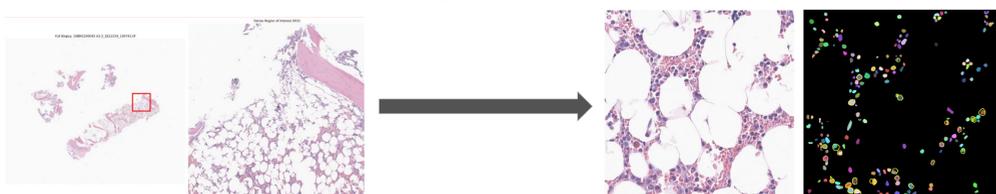

Figure 2: Data Preparation Pipeline
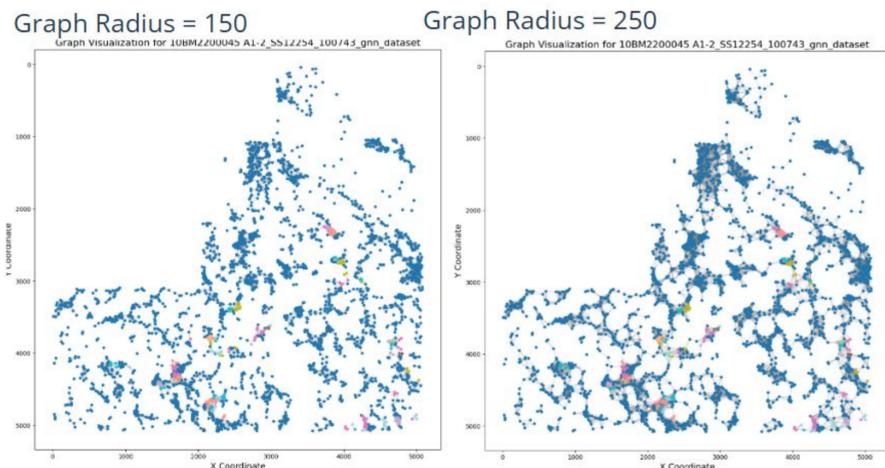
RESULTS

Figure 3: WSI patch, immune mask, ground truth images on upper row with predictions from Dectron2, CNN, and GNN models on bottom row

Table 1: GNN model outperform other approaches

	No pre-training	20 epochs	40 epochs
7 features	18.70%	16.30%	**19.57%**
512 features	28.29%	28.47%	**28.54%**

- Extracting 512 features greatly improved performance from 19.57% to 28.54%
- Unsupervised pre-training of GNN had a negligible impact on overall accuracy Per-Class Performance: The model showed highly varied performance across different cell types. It excelled at identifying Class 6 (e.g., Megakaryocytes) with an accuracy of 70.9% (642/906), but struggled significantly with rarer or morphologically similar types, scoring 0% on critical classes like Class 1 (Band) and Class 2 (Blast).
- Cell type accuracy greatly depended on graph radius hyperparameter setting

Figure 4: Class-based accuracy Left: 7 features; Right - 512 features

CONCLUSION

Potential for Clinical Impact

- Reduces time, subjectivity, and potential for misdiagnosis (up to 12%) in manual bone marrow analysis.
- Increases diagnostic accuracy and efficiency through standardized, quantitative cell analysis.
- Enables objective analysis of the cellular microenvironment and diagnostic patterns often missed by the human eye.

Limitations

- Preliminary classification accuracy is modest at 28.29% for a complex 14-class problem.
- Moderate segmentation performance (Dice Score ≈ 0.7) may introduce upstream errors into the pipeline.
- Overall performance is heavily dependent on the quality of the upstream feature extraction step.

Future Directions

- Improve segmentation (Dice Score) and enhance CNN-based feature extraction to boost overall accuracy.
- Implement a second GNN to analyze spatial cell arrangements and automatically identify clinically significant cellular niches.

edit.