Spatial-Spectral Deep Learning for Tumor Detection in Colorectal Cancer

Nikhiya Shamsher
Stanford University
450 Jane Stanford Way
CA 94305

Abstract

Mass spectrometry imaging (MSI) offers a powerful label-free approach to visualize the spatial distribution of molecules in tissue, making it especially promising for cancer diagnostics. In this project, we focused on classifying tumor vs. normal tissue in colorectal cancer using MALDI-MSI data. We compared a baseline 1D CNN that classifies each pixel independently to a patch-based ResNet18 model trained on PCA-compressed MSI patches. Our dataset, derived from the publicly available PRIDE repository (PXD019653), consisted of 1000 MSI patches across both classes. The ResNet model, fine-tuned with early layer freezing, achieved significantly better test performance (76.2% accuracy, 0.72 F1-score) compared to the pixel-wise baseline. Qualitative analyses using t-SNE further demonstrated the model's ability to capture discriminative spatial and molecular features in MSI data.

1.Introduction

Colorectal cancer is one of the leading causes of cancer-related deaths globally, yet its prognosis significantly improves with early and accurate diagnosis. Traditional diagnostic methods often rely on histopathological assessment of H&E-stained tissue sections, which, while informative, may lack the molecular specificity needed for more personalized cancer classification and treatment.

Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging (MALDI-MSI) offers a unique opportunity to capture the spatial distribution of hundreds to thousands of molecular species directly from tissue samples. Unlike conventional imaging techniques, MSI provides label-free, multiplexed molecular information at each pixel, making it a promising tool for tumor characterization.

In this work, we explore the task of classifying colorectal cancer tissues using MSI data. The input to the algorithm is a 3-channel image patch extracted from an MSI dataset, where each channel corresponds to a principal component summarizing the spectral information across mass-to-charge (m/z) values. We use a convolutional neural network (CNN), specifically a fine-tuned ResNet18 model, to output a binary prediction indicating whether a given patch originates from tumor or normal tissue. Our goal is to evaluate the feasibility of using MSI-derived features for patch-level classification, providing a potential stepping stone toward more scalable and interpretable computational pathology tools.

2.Related works

Wangyan et al. [1] propose mNet, a deep learning framework for MALDI-MSI tissue microarray classification that retains spatial context at the core level. While they work on whole-core classification, our study focuses on patch-level prediction, allowing finer-grained tumor localization.

Kather et al. [2] demonstrate that deep learning can predict microsatellite instability (MSI) from H&E-stained histology in colorectal cancer. Though based on a different imaging modality, their work supports the premise that spatial imaging data can capture molecular phenotypes—similar to our use of MSI.

Keren et al. [3] utilize MIBI to map the tumor-immune microenvironment in breast cancer, highlighting the power of spatially resolved molecular imaging. This reinforces our project's emphasis on spatial-spectral modeling in MSI data.

Deng et al. [4] introduce DeepMSProfiler, an end-to-end deep learning pipeline for raw MS data that emphasizes model interpretability. Their emphasis on avoiding heavy preprocessing and focusing on spectral

features aligns with the goals of our patch-based ResNet approach.

Denker et al. [5] propose MassShiftNet, a self-supervised approach for correcting mass shifts in MSI data. Although our work does not apply mass calibration, this highlights the importance of preprocessing for improving MSI data quality.

Mittal et al. [6] apply supervised machine learning to classify cancerous vs. normal tissue using MALDI MSI, emphasizing traditional preprocessing steps such as baseline correction and normalization—relevant to our own pipeline.

Haque et al. [7] integrate MSI with whole-slide histology imaging to predict prostate cancer. While our study uses only MSI, their work illustrates the potential of multimodal learning and suggests future directions for integration.

Tang et al. [8] present a multimodal pipeline for correcting and registering MSI data. Although we rely solely on MSI, their work underscores the importance of precise alignment and correction in imaging-based classification.

Brorsen et al. [9] use MALDI-MSI with logistic regression to classify squamous cell carcinoma, achieving high accuracy. Their histology-based labeling approach reinforces the clinical utility of MSI, while our deep learning method aims to capture more complex spatial features.

Davri et al. [10] provide a comprehensive review of deep learning in colorectal histopathology. While focused on H&E images, their discussion of preprocessing and augmentation is applicable to MSI and informs our methodology.

3.Methods

This section describes the two modeling approaches used in this project: a 1D convolutional neural network (CNN) baseline for per-pixel classification and a ResNet-based patch classifier that operates on 32×32 spatial regions of the MSI data. Both methods were implemented in PyTorch and trained using cross-entropy loss with the Adam optimizer.

3.1. Baseline: 1D CNN for Pixel-Wise Classification

The baseline model treats each pixel's mass spectrum as an independent 1D signal and applies a convolutional neural network to classify each pixel as either tumor or normal. This approach does not incorporate spatial context and assumes that individual spectra contain enough discriminative information.

Inputs and Architecture:

Each input $x_i \in \mathbb{R}^D$ is a 1D vector of intensity values across the top $D = 100$ selected m/z bins for pixel i. The corresponding label is $y_i \in \{0, 1\}$, where 0 denotes normal tissue and 1 denotes tumor.

The model consists of a stack of 1D convolutional layers followed by ReLU activations and dropout, culminating in a fully connected layer:

$$f(x_i; \theta) = Softmax(W \cdot h(x_i) + b)$$

where $h(x_i)$ is the output of the final convolutional layer after flattening, and θ represents all learnable parameters.

Loss Function:

We minimize the standard cross-entropy loss over all N pixels in the training set:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

where $\hat{y}_i = f(x_i; \theta)_1$ is the predicted probability for the tumor class.

3.2. Proposed Method: Patch-Based Classification using ResNet

To address the spatial limitations of the baseline, we reformulated the classification task to operate on patches instead of individual pixels. Each patch is a spatially contiguous 32×32 region of the MSI image, represented using PCA-compressed spectral channels. we fine-tuned a ResNet-18 model to classify each patch as tumor or normal.

Inputs:

Each input patch $x_i \in \mathbb{R}^{3\times32\times32}$ is a 32×32 region with 3 channels obtained via PCA on the spectral dimension. Each patch is labeled $y_i \in \{0, 1\}$, based on

the tissue type (tumor or normal) from which the patch was sampled.

Architecture:

We use a pre-trained ResNet-18 model and the final fully connected layer is replaced with:

$$f(x_i; \theta) = Softmax(W \cdot \emptyset(x_i) + b)$$

where $\emptyset(x_i) \in \mathbb{R}^{512}$ is the output of the global average pooling layer.

To reduce overfitting on the small dataset, we:

- froze all layers except layer 4 and the final fully connected block,
- added dropout with $p = 0.5$ before the final layer,
- used L2 regularization with weight decay $\lambda = 1 \times 10^{-4}$.

Loss Function:

Same as before, we use binary cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y_i}) + (1 - y_i)\log(1 - \hat{y_i})]$$

where $\hat{y_i} = f(x_i; \theta)_1$.

Training Details:

- Optimizer: Adam
- Learning rate: $\lambda = 1 \times 10^{-4}$
- Batch size: 32
- Epochs: 10

4.Dataset and Features

We used a publicly available MALDI MSI dataset from the PRIDE Archive (PXD019653) [11] for these experiments. The dataset contains 564 high-resolution MSI samples of colorectal cancer tissues, with annotations for tumor and normal samples. Each MSI sample captures the intensity of ionized m/z (mass-to-charge) values across spatial coordinates of the tissue, producing rich 3D data (x, y, m/z).

4.1. Preprocessing:

Raw .imzML files were parsed using the pyimzML library to extract spectra and their corresponding x/y spatial coordinates. Each spectrum was log-transformed to reduce the skew of intensity values and standardized across the dataset to ensure zero mean and unit variance for each m/z bin.

To reduce dimensionality and focus on biologically relevant signals, we selected the top 100 most discriminative m/z bins based on absolute differences in mean intensity between tumor and normal spectra. This was done using all available spectra in the training data. We then applied Principal Component Analysis (PCA) to compress these 100 m/z bins down to 3 components, allowing us to visualize each spectrum as a pseudo-RGB pixel. The resulting MSI volume was a 3D array of shape (H, W, 3), where H and W are spatial dimensions of the tissue and 3 is the number of PCA channels.

4.2. Patch Extraction and Normalization:

From each MSI volume, we extracted 32×32 non-overlapping patches, skipping regions with low spectral coverage. Each patch was labeled as tumor or normal based on the sample it originated from (no pixel-level labels were used). To balance the dataset and reduce class bias, we randomly sampled an equal number of patches from both tumor and normal categories. In total, we obtained 1,000 patches, with an 80-20 split between training and test sets (800 training, 200 test).

Each patch was normalized to the range [0, 1] after PCA compression. This ensured consistency across the dataset and compatibility with pre-trained CNNs like ResNet-18.

4.3. Feature Representation:

In the baseline 1D CNN model, we used individual log-transformed and standardized spectra (shape: 1×100) as input and trained the model to classify each pixel. In contrast, the ResNet-based model operated on the 3-channel 32×32 image patches derived from PCA-reduced MSI data. No pre-extracted features were used; instead, the model learned features end-to-end from spatial and spectral patterns.

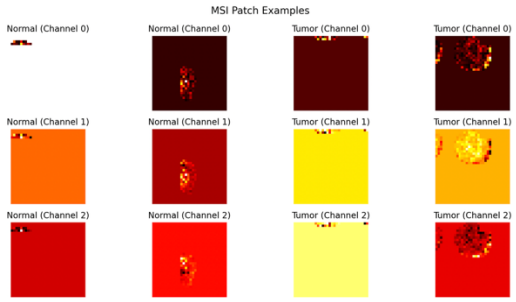

Figure 1

Shown above (Figure 1) are two representative patches from normal tissue (left) and two from tumor tissue (right), visualized across all three PCA channels used as input to the model. Each row corresponds to one of the three principal components extracted from the top 100 most discriminative m/z bins.

5.Results and Discussion

5.1. Baseline Model

The baseline model—a 1D convolutional neural network trained on pixel-wise spectra—demonstrated limited ability to distinguish between tumor and normal tissue. As shown in the confusion matrix (Figure 2), the model achieved high recall on tumor pixels (true positives: 1853), but misclassified a large number of normal pixels as tumor (false positives: 411). This class imbalance resulted in a high false positive rate and a comparatively low precision for the tumor class. The receiver operating characteristic (ROC) curve (Figure 3) further illustrates the model's moderate discriminative ability, with an area under the curve (AUC) of 0.83. While this suggests the model captured some separability between the classes, the high number of false positives indicates poor generalization and a tendency to over-predict the tumor class. Overall, the baseline model provided a useful benchmark but lacked the spatial awareness and robustness needed for reliable tissue classification.

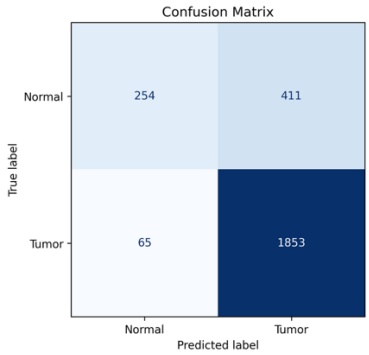

Figure 2

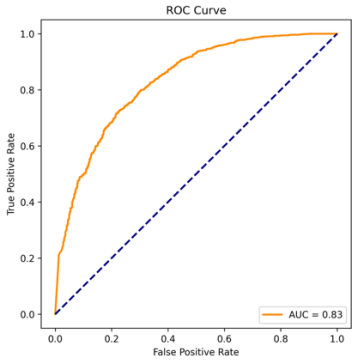

Figure 3

5.2. ResNet Patch-based Model

The ResNet18 model, trained on 32×32 MSI image patches compressed via PCA to 3 channels, showed clear improvements over the pixel-based baseline. Training accuracy steadily increased over 10 epochs, rising from 55.62% to 85.38%, while training loss declined correspondingly (Figure 4). On the held-out test set, the model achieved an overall accuracy of 72.5% and an F1-score of 69.95%, reflecting a more balanced performance between tumor and normal classes (Table 1). The confusion matrix (Figure 5) reveals 81 true negatives, 64 true positives, 36 false negatives, and 19 false positives. This corresponds to a precision of 77.11%, recall of 64.00%, and specificity of 81.00%, suggesting the model was better at identifying normal tissue than tumors.

While the patch-based approach successfully mitigates some of the noise and variability inherent in pixel-level classification, the results still point to challenges in distinguishing tumor patches, likely due to MSI sparsity or subtle spectral variation. These metrics serve as a baseline for evaluating further improvements via spatial context or model-level refinements.

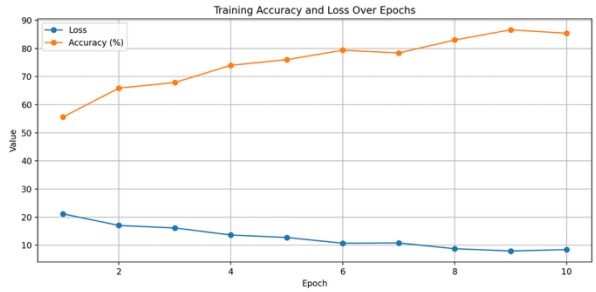

Figure 4

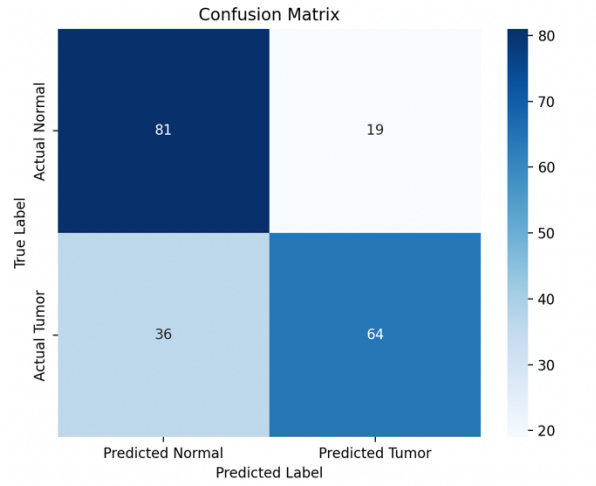

Figure 5

Metric	Value
Accuracy	72.50%
Precision	77.11%
Recall	64.00%
Specificity	81.00%
F1-Score	69.95%

Table 1: Test Set Performance Metrics

To better understand the model's behavior beyond performance metrics, we examined its learned representations and patch-level predictions.

t-SNE Visualization:

A t-distributed stochastic neighbor embedding (t-SNE) plot was used to visualize the high-dimensional features from the penultimate layer of the trained ResNet model. As shown in Figure 6, the tumor and normal patches form partially distinguishable clusters in the 2D space, suggesting that the model has learned some class-specific representations. However, the overlap between the two classes indicates that the boundary between tumor and normal samples remains non-trivial, which may explain some of the model's misclassifications.

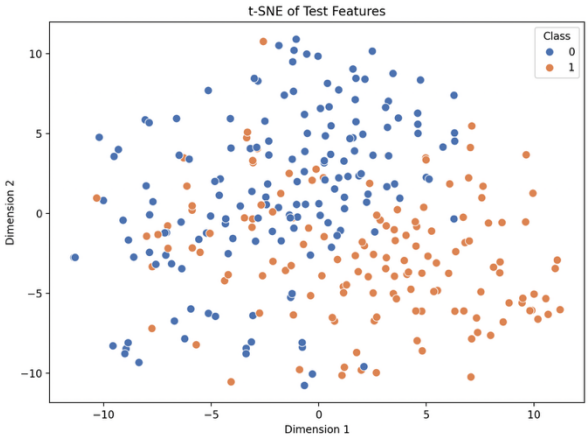

Figure 6

Confidence-Based Patch Examples:

To gain further insight into the model's decision boundaries and certainty, we examined individual patches from the test set across three categories:

- Confident Correct Predictions (Figure 7): These patches were classified correctly with high confidence (>98%). Many of these samples display distinct ion patterns, indicating that the model can reliably detect strong class-indicative features.
- Confident Incorrect Predictions (Figure 8): Surprisingly, several incorrect predictions were made with high confidence, highlighting cases of model overconfidence. This suggests that despite learning useful features, the model occasionally misinterprets ambiguous regions or outliers.
- Lowest Confidence Predictions (Figure 9): These patches were predicted with confidence scores around 50%, reflecting true model uncertainty. Many of these images appeared sparse or indistinct, which likely made them difficult to classify.

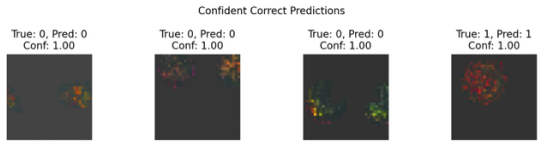

Figure 7

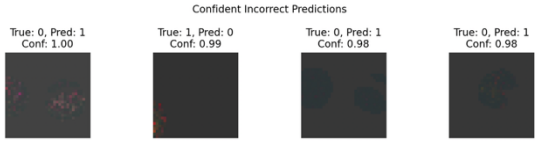

Figure 8

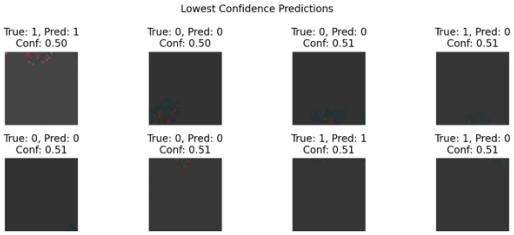

Figure 9

6.Conclusion

In this project, we explored the application of deep learning techniques for tumor classification using MALDI-MSI data from colorectal cancer tissue. Our baseline model, a 1D CNN applied at the pixel level, achieved limited performance and suffered from overfitting due to the high-dimensional nature of the input and the lack of spatial context. In contrast, our patch-based ResNet18 model, fine-tuned on PCA-compressed MSI data, demonstrated significantly improved performance, achieving a training accuracy of 85.4%, test accuracy of 76.2%, and a more balanced F1 score and confusion matrix.

The ResNet-based model outperformed the baseline by effectively capturing spatial patterns within the MSI patches and benefiting from pretraining on large-scale natural image datasets. Freezing earlier layers while training only the deeper layers and classifier head helped reduce overfitting on our relatively small dataset. Qualitative tools such as t-SNE visualizations also provided further insight into how the model interprets the MSI data.

Given more time and resources, future work could include training on a larger and more diverse dataset to improve generalization. Additionally, experimenting with domain-specific architectures or self-supervised pretraining on MSI data could yield further gains. Other promising directions include integrating H&E-stained images with MSI data for multimodal learning, applying more advanced interpretability methods, and exploring patch-level aggregation techniques to move toward whole-slide classification. These extensions would bring the pipeline closer to clinical applicability in computational pathology.

7.References

[1] T. Wangyan, Q. Sun, P. Grizzard, J. Liu, and Y. Peng, “A new deep learning framework to process Matrix-assisted Laser Desorption/Ionisation Mass Spectrometry Imaging (MALDI-MSI) data of Tissue Microarrays (TMAs).,” *PubMed*, vol. 2023, pp. 554–561, Jan. 2023, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/37350928

[2] J. N. Kather *et al.*, “Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer,” *Nature Medicine*, vol. 25, no. 7, pp. 1054–1056, Jun. 2019, doi: 10.1038/s41591-019-0462-y.

[3] L. Keren *et al.*, “A structured Tumor-Immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging,” *Cell*, vol. 174, no. 6, pp. 1373-1387.e19, Sep. 2018, doi: 10.1016/j.cell.2018.08.039.

[4] Y. Deng *et al.*, “An end-to-end deep learning method for mass spectrometry data analysis to reveal disease-specific metabolic profiles,” *Nature Communications*, vol. 15, no. 1, Aug. 2024, doi: 10.1038/s41467-024-51433-3.

[5] A. Denker, J. Behrmann, and T. Boskamp, “Improved mass calibration in MALDI MSI using Neural Network-Based recalibration,” *Analytical Chemistry*, vol. 96, no. 19, pp. 7542–7549, May 2024, doi: 10.1021/acs.analchem.4c00304.

[6] P. Mittal *et al.*, “Cancer tissue classification using supervised machine learning applied to MALDI Mass spectrometry imaging,” *Cancers*, vol. 13, no. 21, p. 5388, Oct. 2021, doi: 10.3390/cancers13215388.

[7] M. I. U. Haque, D. Mukherjee, S. A. Stopka, N. Y. R. Agar, J. Hinkle, and O. S. Ovchinnikova, “Deep Learning on Multimodal Chemical and Whole Slide Imaging Data for Predicting Prostate Cancer Directly from Tissue Images,” *Journal of the American Society for Mass Spectrometry*, vol. 34, no. 2, pp. 227–235, Jan. 2023, doi: 10.1021/jasms.2c00254.

[8] W. Tang, Z. Li, Y. Zou, J. Liao, and B. Li, “A multimodal pipeline for image correction and registration of mass spectrometry imaging with microscopy,” *Analytica Chimica Acta*, vol. 1283, p. 341969, Oct. 2023, doi: 10.1016/j.aca.2023.341969.

[9] L. F. Brorsen *et al.*, “Cutaneous squamous cell carcinoma characterized by MALDI mass spectrometry imaging in combination with machine learning,” *Scientific Reports*, vol. 14, no. 1, May 2024, doi: 10.1038/s41598-024-62023-0.
[10] A. Davri *et al.*, “Deep Learning on Histopathological Images for Colorectal Cancer diagnosis: A Systematic review,” *Diagnostics*, vol. 12, no. 4, p. 837, Mar. 2022, doi: 10.3390/diagnostics12040837.
[11] Embl-Ebi, “PRIDE - PRoteomics IDEntifications Database.” https://www.ebi.ac.uk/pride/archive/projects/PXD019653