Understanding MIL

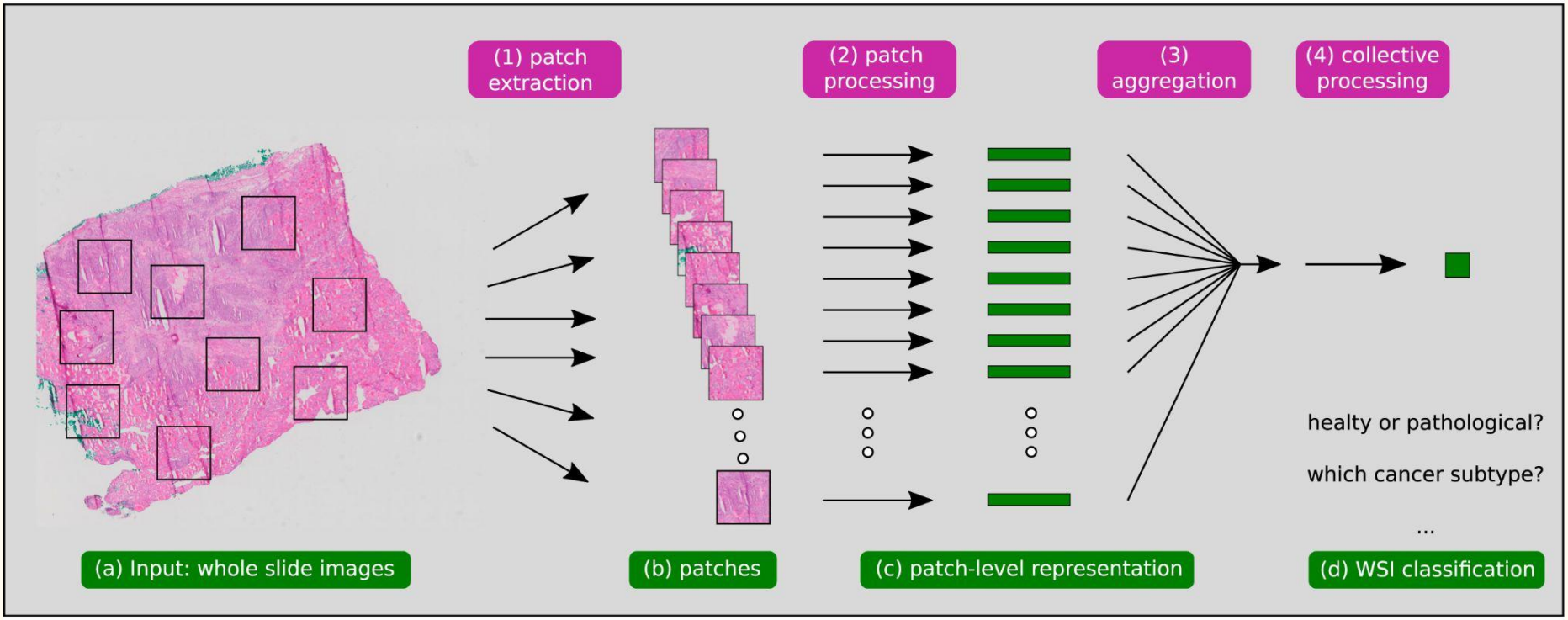
Divya Bhardwaj

Overall Context

- Digital pathology involves extremely large whole slide images (WSIs) that are hard to annotate at fine granularities
  - Time-consuming
- Opportunity: MIL allows model training using weak labels (i.e., WSI-level labels) without requiring pixel or region annotations
  - A WSI is treated as a “bag” of image patches (instances), and the label is assigned to the bag, not to individual patches.
- Sharp increase in MIL-related publications in digital pathology, especially post-2020

Source: “Multiple Instance Learning for Digital Pathology” by Michael Gadermayr & Maximilian Tschuchnig (2024) in Computerized Medical Imaging and Graphics

General Architecture



Source: "Multiple Instance Learning for Digital Pathology" by Michael Gadermayr & Maximilian Tschuchnig (2024) in Computerized Medical Imaging and Graphics

Ways to Refine: Pooling Strategies

- Basic: Max, Average
- Attention-based
  - Learns patch importance weights based on features
  - Computes a weighted average of patch features or scores
  - Focuses on key patches, easily interpretable
  - Does not capture relationships between patches
- Transformer-based
  - Models interactions between patches
  - Instead of computing importance for each patch independently (like attention-based pooling), it uses self-attention to understand how patches influence each other
  - Captures global context/patterns across multiple patches
  - More complex + computationally expensive

Source: "Multiple Instance Learning for Digital Pathology" by Michael Gadermayr & Maximilian Tschuchnig (2024) in Computerized Medical Imaging and Graphics

Ways to refine: MIL type

Instance-Based MIL

- Computes one score/probability per patch
- Pools the scores
- Predicts the case-level label using the sigmoid function to obtain the probability
- Enables us to see predictions at the patch level

Embedding-Based MIL

- Creates a feature vector for each patch
- Pools the vectors into a single bag-level label
- Feeds this vector into a classifier (typically a small neural network called a Multi-Layer Perceptron or MLP) which gives a sigmoid output to predict the case-level label
- Typically more accurate

Source: "Multiple Instance Learning for Digital Pathology" by Michael Gadermayr & Maximilian Tschuchnig (2024) in Computerized Medical Imaging and Graphics

Next Steps

- Start with Embedding-Based MIL + Attention Pooling
  - Would generate a heatmap showing which patches are most important
- Could try Instance-Based MIL
  - Would generate a heatmap showing the probability that each patch is benign/high grade
    - Would enable us to see if it is identifying the lesions accurately
- Could explore transformer pooling
  - TransMIL framework
  - Source: TransMIL: Transformer based Correlated Multiple Instance Learning for Whole Slide Image Classification by Shao et al. in Computer Vision and Pattern Recognition