**Seminar Report: 9/13/2024**

**Speaker**: Dr. Junzhou Huang, Ph.D., Jenkins Garrett Professor, Department of Computer Science and Engineering

**About the Speaker**

Dr. Junzhou Huang holds a strong academic background in computer science and statistics. As a leading researcher, he has authored over 100 papers on machine learning (ML), artificial intelligence (AI), and related fields, garnering more than 29,000 citations.

**Topic: Multimodal Large Language Models (MLLM) for Biomedical Applications**

Dr. Huang presented a comprehensive talk on the use of **Multimodal Large Language Models (MLLMs)** in biomedical applications, focusing on how ML techniques can be applied to diverse biomedical modalities such as text, images, audio, and video. Below is a summary of the key concepts discussed.

**Key Topics Covered**

**1. Machine Learning Algorithms and Medical Modalities**

Dr. Huang began by explaining the use of **unimodal** vs. **multimodal** data, emphasizing that integrating multiple modalities such as **text, images, audio, and video** provides a richer and more nuanced understanding. He discussed how **transformers**, specifically **Vision Transformers** and **Twin Tower Models**, are vital to processing this data.

**2. Biomedical Applications of MLLM**

1. **Computational Pathology**
   * Challenges in handling **gigapixel whole-slide images (WSIs)** and how to efficiently represent this data by leveraging redundancy and consistency.
   * Techniques for **multimodal data decomposition** to capture information at varying granularities.
   * Addressing **multimodal heterogeneity** in biomedical datasets.
2. **Image-Gene Learning**
   * The fusion and alignment of multimodal data to predict gene functions from images.
   * Applying **contrastive learning** techniques to train models to distinguish between data pairs.
3. **Vision-Language Contrastive Learning**
   * **Patch-level learning**: Fine-tuning models like **CLIP** for better inference, including **zero-shot inference** and **few-shot inference** on WSIs.
   * **WSI-level learning**: Inputting WSIs into CLIP models and cleaning WSI reports to improve model predictions.

**MLLM Steps**

1. **Data Generation**: Extracting image data and utilizing **GPT** to generate instructed data.
2. **Domain Alignment**: Aligning the histopathology domain by training only the **MLM layer**.
3. **Instruction Tuning**: Fine-tuning MLLM using the instructed data.

**PathM3 Model**

* **Training**: Incorporates tasks like **LLM, Captioning**, and **Classification**.
* **Inference**: Focuses on classification, with input WSIs undergoing correlation and redundancy reduction through a **Query-based Transformer**.
* **Method**: 1. Input:   
  “””””””””””””””¿WSIs

2. Correlation: Reduce redundancy

3. Query based transformer: Alignment

4. Output: Classification prediction

**Dataset Used:**

The **Patch Gastric ADC22 Dataset** was used, consisting of 693 WSIs from three subtypes, with techniques applied to extract multigranularity information from incomplete data.

**Gene Pathology Applications of MLLM**

Dr. Huang outlined how MLLMs can be used to **computationally predict gene functions**, reducing reliance on costly large-scale wet lab experiments.

* **Gene Ontology and Annotation**: Described how gene functions are annotated through a process involving DNA, RNA, gene products, and relationships captured via **GO Terms**.
* **Deep Learning for Gene Prediction**: Highlighted limitations such as incomplete gene-to-protein sequences and described the **GoBERT pre-training** model, which achieved a **76.15% accuracy** in predicting novel functions.

**Computational Immunology**

* Explored applications such as **BCR Epitope Prediction**, **TCR**, and **Antibody Design**, with references to leading models in the field.

**Current State-of-the-Art (SOTA) Models**

1. **ESM-3** (98B size)
2. **Alpha Fold 2**: Known for its protein structure prediction capabilities, widely used in the 14th Community-wide experiment.
3. **Alpha Fold 3**: The latest iteration that expands predictions on various types of gene products.

Dr. Huang also discussed applications related to **T-cell receptors and epitope binding** and referenced the **GTE Dataset** for benchmarking model performance.

**Conclusion**

Dr. Huang’s seminar provided a detailed exploration of how MLLMs are revolutionizing biomedical research. Through multimodal learning, these models are enabling breakthroughs in fields such as computational pathology, gene prediction, and immunology, thus demonstrating the immense potential of AI in advancing healthcare and biomedical sciences.