Research Experience (Major Two Research Fields)

I am a machine-learning researcher with 5+ years of experience developing multimodal AI systems for *biomedical imaging and non-invasive health monitoring*. My work combines *generative models, graph neural networks, large language models, and signal processing,* expertise that aligns closely with your lab's focus on probabilistic and deep generative frameworks for cancer systems biology.

- **MASc at University of Waterloo (Specialized in Artificial Intelligence)**
  - **Title:** *Advanced AI for Histopathological Whole Slide Image Classification and Captioning*
    Keyword: *Digital Pathology, Whole Slide Image (WSI), Breast Cancer, Graph Neural Networks, Vision Transformers, Large Language Models, Image Captioning*

Summary:

This research addressed challenges in automatic classification and captioning of microscopic histopathological images for computer-aided cancer diagnosis. Whole Slide Images (WSIs) pose issues such as redundant patches and variable patch locations due to subjective sampling by pathologists. To tackle these, the work is structured into two main contributions:

- **TransUAAE-CapGen**: A hybrid framework combining a UNet-based Adversarial Autoencoder (AAE) for extracting high-level tissue features and a transformer for caption generation from low-dimensional embeddings of histopathological patches.
- **GNN-ViTCap**: A novel GNN-ViTCap framework is introduced for classification and caption generation from histopathological microscopic images. A visual feature extractor is used to extract feature embeddings. The redundant patches are then removed by dynamically clustering images using deep embedded clustering and extracting representative images through a scalar dot attention mechanism. The graph is formed by constructing edges from the similarity matrix, connecting each node to its nearest neighbors. Therefore, a graph neural network (GNNs) is utilized to extract and represent contextual information from both local and global areas. The aggregated image embeddings are then projected into the language model's input space using a linear layer and combined with input caption tokens to fine-tune the large language models (LLMs) for caption generation.

Evaluations on **BreakHis** and **PatchGastric** datasets, along with ablation studies, demonstrate that both methods outperform state-of-the-art techniques. The frameworks have the potential for integration into clinical diagnostic pipelines to improve early cancer detection and treatment planning.

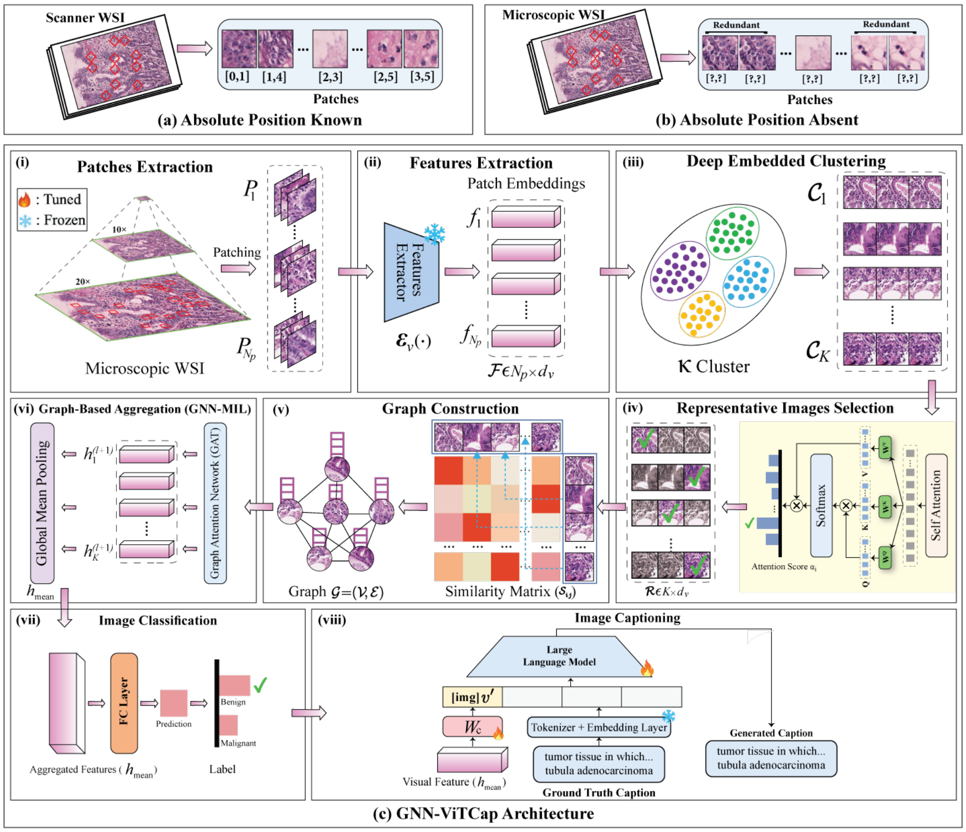

Fig. 1: Overview of the GNN-ViTCap framework for microscopic whole slide images classification and captioning. (a) Scanner WSI where the absolute position of each patch is known. (b) Microscopic WSI lacks patch position and contains redundant patches due to multiple captures from a pathologist's subjective perspective. **(c) GNN-ViTCap architecture:** (i) extracting the patches from whole slide image, (ii) extracting image embeddings using a visual feature extractor, (iii) removing redundancy through deep embedded clustering, (iv) extracting representative images with scalar dot attention mechanism, (v) constructing a graph neural network (GNN) using the similarity of representative patches to capture contextual information within clusters (local) and between different clusters (regional), (vi) applying global mean pooling, which aggregates all node representations, (vii) classifying microscopic WSI using aggregated image embeddings and, (viii) projecting the aggregated image embeddings into the language model's input space using a linear layer, and combining these projections with input caption tokens fine-tunes the LLMs for caption generation.

- **MSc. at Khulna University of Engineering & Technology (Computer Science & Engineering)**
  - **Title:** *Hemoglobin and Glucose Levels Estimation Techniques Using Optimal PPG Characteristic Features of Smartphone Videos*

    Keyword: *Photoplethysmography (PPG), Smartphone Video, Hemoglobin, Glucose, Non-Invasive Measurement, Deep Neural Networks, Feature Selection*

Summary:
This research proposes a non-invasive, cost-effective technique for estimating hemoglobin (Hb) and glucose (Gl) levels using fingertip videos captured by smartphones. The key contributions include:

- Invoking a non-invasive system for monitoring blood glucose, hemoglobin and creatinine using smartphone which is easy to use and comfortable to patients
- Constructing a wearable device using NIR LED for collecting fingertip video and Generating PPG signal from smartphone-captured fingertip videos of 93 subjects under NIR LED illumination.

- Feature engineering: 34 statistical, temporal, and frequency domain features were extracted from raw and derivative PPG signals, along with age and gender.
- Selecting appropriate features using a correlation based feature selection method to avoid redundancy in features and over-fitting problem.
- Development of Deep Neural Network (DNN) models for Hemoglobin and Glucose level estimation, benchmarked against traditional regression methods.

The proposed DNN-based method shows promising results for real-time, non-invasive health monitoring. This approach can potentially be deployed in resource-constrained or point-of-care environments.

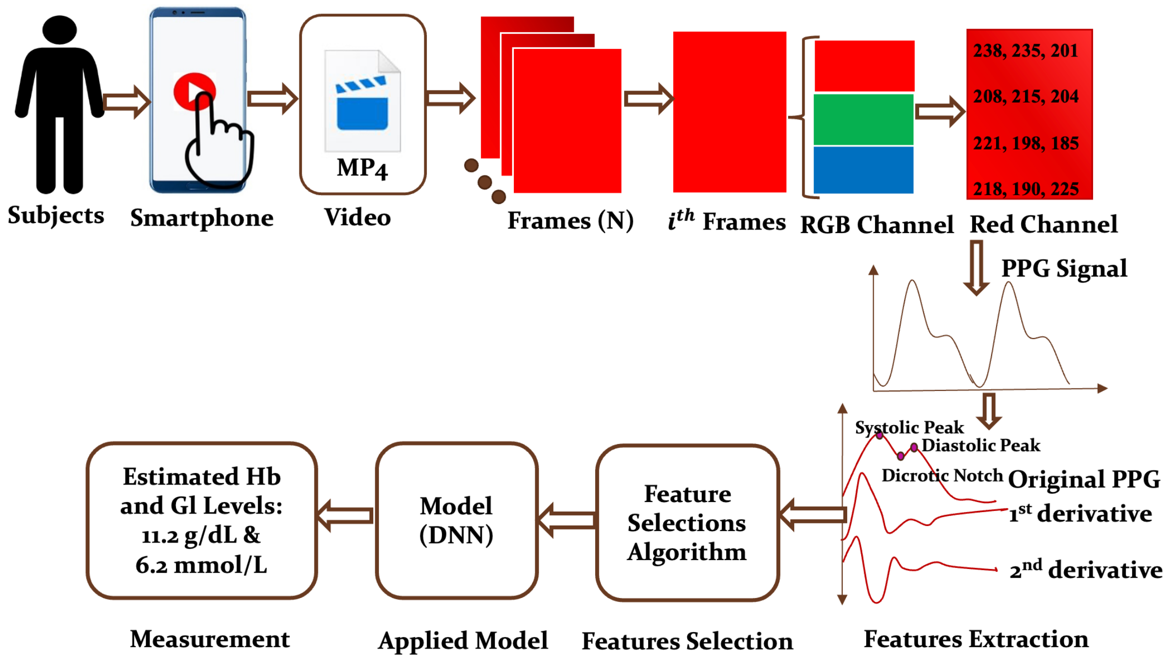

Fig. 2: The overall system architecture of blood Hemoglobin, and Glucose levels measurement

Data Collection Device:

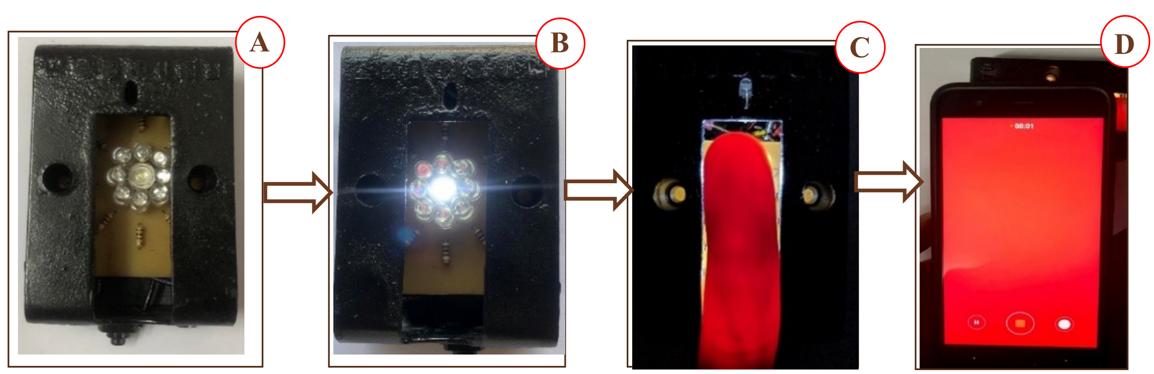

Fig. 3: The Fingertip video data collection kit/device: (A) NIR-LED device with power off, (B) NIR-LED device in turned on condition, (C) Index finger on the device while turned on, and (b) Video recorded with a Nexus-6p smartphone