Alternative Assignment (50%)

1 Instruction

Medical images such as X-rays, CT scans, and MRIs contain complex visual information that is difficult to interpret automatically. Traditional image-only models can classify abnormalities but cannot understand clinical questions or provide interactive responses. To address this, recent AI systems combine vision and language to perform Medical Visual Question Answering (Med-VQA), where a model answers natural-language questions based on a medical image (e.g., "Is there pneumonia?", "What organ is shown?", "Is there a mass?").

In this assignment, you will build and evaluate deep learning models that can answer questions about medical images. You may use any suitable techniques taught in the course, such as neural networks, generative AI, or hybrid deep learning approaches such as Visual-Language Models (VLMs). Your goal is to compare at least two different methods (e.g., a CNN-based baseline vs. a VLM-based model) and analyze their ability to answer both closed-ended and open-ended medical questions.

You will work with an open-source Med-VQA dataset (e.g., VQA-RAD, SLAKE, or PathVQA) containing medical images paired with question–answer annotations. Your final output should include your models, experimental results, and a discussion of their strengths and limitations in performing medical image understanding.

2 Evaluation

2.1 Preliminary Project Report (10% - Due on Week 9):

Prepare a report with at least 5 pages that describe the following:

1. 5% Background. Provide context on what problem you are solving.
2. 5% Objective.
3. 5% Method: Dataset description, data preparation and preprocessing detailing how you prepare and preprocess the datasets for modelling.
4. 5% Method: What algorithm do you plan to use or build (model architecture), and justify your selection? Support with past evidence.
5. 5% Preliminary results to support your proposal.

Note: 25% will be converted to 10%.

2.2 Final Report and Code (40% - Week 13,14)

Write a report of 15 pages min. Please use the report template in Spectrum. In the report, you should clearly.

1. 15% Method: Explain the experiment setup clearly so that other able to reproduce your results.
2. 25% Results: Critically analyse results, and compare your work with the baseline model. What interesting findings you have found?
3. 15% Discussion: Discuss the findings and limitations.
4. 2% Conclusion.
5. 2% GitHub code. Share the link in your report.
6. 1% Reference - APA style
7. Member contributions

Note: 60% will be converted to 40%.

Our faculty will provide cloud tokens to support fine-tuning of state-of-the-art (SOTA) models. If your project requires access to any cloud computing resources or tokens for training or deployment, these will be made available upon request. More information about this can check with Prof. Ir. Dr. Chan Chee Seng - cs.chan@um.edu.my.