Question:

We will use a nutrition dataset collected in a research study. This dataset contains breakfast and lunch information, data on motion, and photographs of meals, across over 40 participants and up to 10 days (note not all participants had 10 full days of data). Our goal is to build a multimodal model to estimate the calories of the lunches consumed from the sensing data, photographs, micro gut health, demographics, and the data from the breakfasts.

(a)   **Data preprocessing:** Preprocess each modality appropriately. For example: the length of some modalities is variable but most models required fixed-length input; some modalities are categorical whereas others are continuous numerical.

(b)   **Data preparation:** Creating a multimodal dataset by merging all data modalities and labels and saving to a DataLoader (https://pytorch.org/ tutorials/beginner/basics/data_tutorial.html).

(c)   **Multimodal model implementation**: Implement a multimodal model which encodes at least two data modalities into a joint embedding and predicts label of interest from the joint embedding. Your are required to use Pytorch for model implementation.

(d)   **Model training:** Declare root mean square relative error (RMSRE) as loss function and appropriate optimizer. Train the model and print the training curve, where the x- axis is the training epochs, and the y-axis is the training performance (RMSRE).

$$RMSRE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\frac{\hat{y}_i - y_i}{y_i})^2}$$

(e)   **Result analysis:** Finetune the hyperparameters and use the best trained model predict on the test set ('[modality]_test.csv').