Final Project

Objective

In this project, you will work in **groups of 2–3 students** to develop an **Image Caption Generator** that can generate descriptive sentences for images. You will combine a **Convolutional Neural Network (CNN)** for extracting image features with a **Long Short-Term Memory (LSTM)** network for generating captions.

Dataset

- Dataset contains 8,000 images, each paired with 5 human-annotated captions.
- Available at: LINK
- Use a 70/15/15 train-validation-test split.
- Preprocess both images and captions appropriately.

Project Requirements

1. Model Architecture

- Use a **CNN** to extract image feature vectors.
- Connect these features to an **LSTM decoder** to generate captions.
- Use word embeddings with the LSTM.

2. Preprocessing

- Resize and normalize images as needed by your CNN.
- Clean, tokenize, and pad the captions.
- Construct a vocabulary and handle out-of-vocabulary tokens.

3. Training & Evaluation

- Train your model to generate captions.
- Evaluate your model using:
  - **BLEU scores** (at least BLEU-1 to BLEU-4)
- Include **qualitative results**: Show example images with both predicted and reference captions.

4. Bonus Enhancements (Optional)

- Try **attention mechanisms** (Transformers).

Report Format (Two-Column PDF)

Each group must submit a **two-column PDF report** (IEEE style recommended) with the following sections:

1. Introduction
- Define the task and explain its relevance.

2. Preprocessing
- Explain how images and captions were processed and prepared.

3. Methods
- Describe your model architecture (CNN + LSTM), training pipeline, and any special techniques used.

4. Results
- Present BLEU scores and sample generated captions.
- Include **at least 3–5 images with predicted captions** in comparison to the original ground truth captions.

5. Conclusion
- Summarize your findings, discuss model strengths/weaknesses, and suggest potential improvements.

Grading Breakdown

Component	Weight
Model implementation	30%
Preprocessing & methodology	15%
Evaluation & BLEU scores	20%
Results & visual examples	15%
Report quality & organization	20%