CS 640 Homework

1 Introduction

This assignment requires you to design, implement, and evaluate a machine learning model to classify microscopic images of blood cells for the presence of malaria. The task is a binary classification: determine if a cell is **Positive (Infected/Parasitized)** or **Negative (Uninfected)**.

1.1 Dataset Structure

The dataset is provided as a zip file. The unzipped directory structure as follows:

```
/malaria_dataset
|-- /train
|   |-- /negative (Uninfected images)
|   |-- /positive (Parasitized images)
|-- /val
|   |-- /negative
|   |-- /positive
```

Note: All images are TIFF or PNG files of varying sizes.

2 Classification Tasks

2.1 Data Preprocessing and Loading (20 Points)

1. **Data Loading:** Write code to load the images from the specified directory structure for both the `train` and `val` sets.
2. **Image Standardization:** Implement resizing of all images to a uniform size and normalize pixel values (e.g., to the range $[0, 1]$).
3. **Data Augmentation:** Apply at least **one** data augmentation techniques to the `train` set only. Briefly explain why augmentation is beneficial for this task.

2.2 Model Implementation (40 Points)

Design and implement a **machine learning model** for the binary classification task. Explain why you choose the specific machine learning algorithm and model.

2.3 Training and Evaluation (40 Points)

1. **Compile and Train:** Compile your model using an appropriate **Loss Function** (justify your choice) and **Optimizer** (e.g., Adam). Train the model for an adequate number of epochs, monitoring both training and validation loss and accuracy.

2. **Performance Metrics:** After training, evaluate your model's performance on the `val` dataset. Report the following metrics:
    - **Accuracy**
    - **Precision**
    - **Recall**
    - **F1-Score**
3. **Analysis:** Generate and clearly display the **Confusion Matrix** for the validation set. Discuss the results, specifically addressing the trade-offs between Precision and Recall in the context of a medical diagnosis task like malaria detection.

3 Submission Guidelines

Be ready to demo your project during the last week of the semester if this is chosen for your demo. Submit a single PDF report including:

1. Your full name and student ID.
2. Responses to all the above parts of the assignment.
3. Your full, commented source code included in an Appendix.
4. Plots showing the training and validation **Accuracy** and **Loss** over epochs.