Anshul Bankey
22BCG10013

Link: https://github.com/nipun1509/CAD-Flutter

1. Introduction

Dataset preparation is a crucial step in training an AI model for Coronary Artery Calcification (CAC) detection. This project utilizes CT scan images from the MMWHS dataset, focusing on preprocessing, segmentation, and structuring the data to ensure high-quality input for model training.

2. Dataset Description

The dataset consists of CT scan images stored in .nii.gz format, requiring preprocessing before model training. The dataset was divided into:

- Training Set: Used for model learning.
- Test Set: Used to evaluate the model's performance.

Key parameters:

- Resolution: Images were resized to 256x256 pixels for uniformity.
- Intensity Range: Normalized between 0 and 1 for better model learning.
- Segmentation Masks: Each CT image has a corresponding mask highlighting CAC regions.

Datasets used:

- MMWHS Dataset (CT scans Only) – Publicly available dataset for medical image analysis.
- Pre-processed Dataset – Normalized and structured version for model training.

3. Feature Engineering (Preprocessing Steps)

My contributions to the dataset part is:

Segmentation Mask Processing:

- Masks were binarized to **0 (background) and 1 (CAC region)**.

Image Resizing & Reshaping:

- Original images had inconsistent sizes; resized all images to **256x256 pixels**.

4. Machine Learning Models Used

Although dataset preparation doesn't involve model training, this step was essential for U-Net, the deep learning model used for CAC segmentation.

Preprocessing ensured:

Optimized input quality for deep learning models.
Efficient training by removing noise and inconsistencies.

5. Model Performance

Preprocessing improved model convergence and reduced noise in segmentation outputs. While accuracy depends on the model, structured input significantly impacts performance.

6. Conclusion

Dataset preparation was a critical step in ensuring reliable AI predictions. By normalizing and structuring the data, we enabled efficient training and improved segmentation accuracy.

7. Future Work

- Implement data augmentation (flipping, rotation) to enhance model generalization.
- Use additional datasets to improve robustness.
- Optimize preprocessing to handle real-world CT scan variations.

Further improvements can be made by:

- Incorporating real-world crash data for better model generalization.
- Implementing real-time crash prediction in drone flight software.
- Exploring deep learning models for improved predictive performance.

This project lays the foundation for enhanced drone safety through data-driven insights and machine learning solutions.