Data Collection and Preprocessing Phase

Date	7 November 2024
Team ID	team-739994
Project Title	Virtual Eye - Life Guard for Swimming Pools to Detect Active Drowning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

The data collection plan for the Virtual Eye Lifeguard project focuses on gathering a diverse, high-quality dataset to train a YOLOv5 model for detecting active drowning in swimming pools. Primary data sources include custom images and video footage of pools under various conditions (e.g., different lighting, crowding levels), along with augmented data to enhance model robustness. Each image will be annotated with bounding boxes distinguishing drowning from non-drowning scenarios. Supplementary data from public aquatic datasets and synthetic data may also be used to increase variety.

Data Collection Plan Template

Section	Description
Project Overview	To detect active drowning events in swimming pools using YOLOv5, a machine learning model for object detection, with high accuracy and real-time alerts. By using diverse and labeled image data, the model will learn to identify signs of drowning in real-time, providing immediate alerts to help lifeguards and improve swimmer safety. This project combines computer vision and real-time monitoring to create a reliable and efficient pool safety solution.
Data Collection Plan	The data for the Virtual Eye Lifeguard project has been downloaded from Kaggle. This dataset includes images from various pool environments, providing a foundation for training the YOLOv5 model to detect drowning scenarios effectively. Additional preprocessing and augmentation will be applied to enhance model performance.

Raw Data Sources Identified	1. **Kaggle Dataset**: A collection of images from various pool environments, providing labeled examples of swimming and potential drowning scenarios. This dataset serves as the primary source for training and validating the YOLOv5 model. 2. **Augmented Data**: Synthetic variations of the Kaggle dataset images, created through techniques like flipping, rotation, and scaling. These augmentations increase data diversity, helping the model generalize to varied real-world conditions. 3. **Synthetic Data (Optional)**: If needed, artificially created or simulated images of pool environments to cover edge cases or less common drowning scenarios, further enhancing the dataset's completeness.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle	The Kaggle dataset has labeled pool images with annotations for detecting swimming and drowning incidents, essential for training the YOLOv5 model in real-time detection.	https://www.kaggle.com/datasets/alanoudawaji/swimming-and-drowning-dataset	Image	380mb	Private (with access)