**Yavar Internship Selection Assignment**

Problem Statement

**Problem 1: People Fall Detection**

People's fall is a serious concern, especially in areas such as staircases, escalators, and steps, as it can lead to injuries or even fatalities. The task is to develop a solution for accurately detecting falls in offline videos captured in mp4 format. The solution should aim for high accuracy in predictions while minimizing false positives.

**Approach Overview**

To address the problem of fall detection, we propose a multi-stage approach utilizing various computer vision and deep learning techniques. The approach consists of the following key components:

**Preprocessing:**

* Resize video frames: Resize each frame to a standard size for consistency in processing.
* Normalize pixel values: Ensure uniformity in pixel intensity across frames.
* Object Detection with YOLO (You Only Look Once):
* Utilize YOLO, a state-of-the-art real-time object detection system, to detect people in video frames.
* YOLO identifies bounding boxes around people, enabling subsequent analysis.

**Pose Estimation:**

* Apply pose estimation algorithms to analyze human postures and movements.
* Mediapipe Pose model is used to detect key points of human body landmarks.
* Drawing utilities are employed to visualize the detected pose landmarks on the frames.

**Open-Vocabulary Object Detection with OWL-ViT:**

* Incorporate OWL-ViT (Open-Vocabulary Visual Transformer) model for object detection.
* OWL-ViT enables detection of a wide range of objects, including falls, in an open-vocabulary manner.
* Use OWL-ViT's capability to process both text and image inputs for improved fall detection.

**Evaluation and Validation**:

* Evaluate the proposed solution based on accuracy of predictions and minimization of false positives.
* Assess the model's performance on various scenarios, including different environments and types of falls.
* Fine-tune the model parameters to optimize performance metrics.

**Implementation Details**

The proposed approach is implemented using Python and several libraries, including OpenCV, Mediapipe, and Hugging Face Transformers. **Key implementation steps include:**

* Preprocessing of video frames to ensure uniformity and consistency.
* Utilization of YOLO for person detection, with bounding box visualization.
* Integration of Mediapipe Pose model for pose estimation and landmark detection.
* Incorporation of OWL-ViT model for open-vocabulary object detection, specifically targeting fall detection.
* Evaluation of the model's performance on test videos, with emphasis on accuracy and false positive rates.