MultiModal LLMs for Fall Detection

Prompting Techniques, RAG, & Fine-Tuning

DATA 266 Group 4
Final Presentation
Chelsea Jaculina - 011274913
Himani Shah - 017411407

Problem Definition

Falls lead to serious injury and are a leading cause of death among seniors.

Traditional approaches like wearables and video surveillance often compromise comfort or raise concerns in sensitive settings.

Proposed Methodology

A) Prompt Engineering
1. One-Shot
2. Few-Shot (5)
3. Chain-of-Thought (CoT)
4. Tree-of-Thought (ToT)

B) RAG

C) Fine-Tuning Technique
1. LoRA - PEFT
2. Reward Modeling
3. RLHF

UR Fall Dataset Description

Why this Dataset?

1. Publicly available and widely used in fall detection research.
2. Captured using depth cameras → non-intrusive, privacy-friendly, and safe.

Data Types

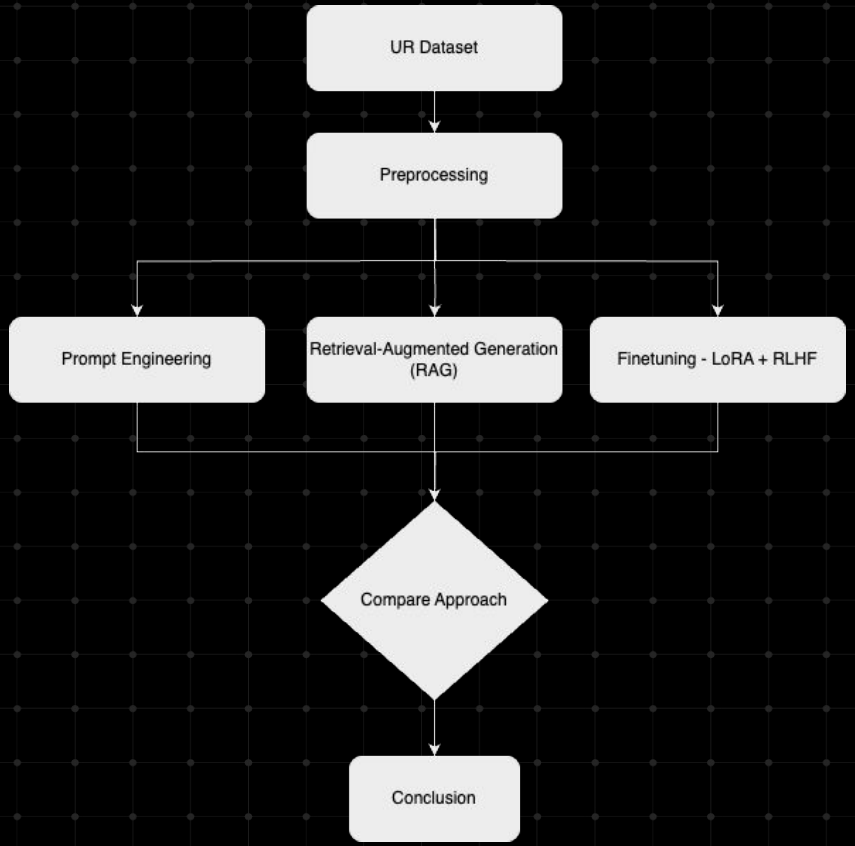
1. Falls Activity
2. Normal Activity

`ID`	Unique number for each activity session
`Frame`	Frame number from the video
`Label`	1 = Fall, -1 = Normal Activity
`HeightWidthRatio`	Tells if the person is standing tall or lying down
`MajorMinorRatio`	Shape of the person's body – wide vs tall
`BoundingBoxOccupancy`	How much space the person takes up in the frame
`MaxStdXZ`	Movement variation in horizontal directions (X and Z axes)
`HHmaxRatio`	Height change — lower values mean the person is likely on the ground
`H`	Height of the person in mm
`D`	Distance from the person to the floor (in mm)
`P40`	Percentage of the body below a certain height (shows closeness to floor)

Preprocessing steps

1. Column Naming
2. Duplicate Removal
3. Data Type Casting
4. Upscaling minority class with stratified method

Experimental Design



Prompting Techniques

Tree of Thought

Model: *microsoft/phi-2*

```
def create_tot_prompt(query):
    return f"""
You are a fall detection expert analyzing  motion features extracted from a single video frame.

Feature Descriptions:
- HHmaxRatio: Change in vertical posture
- H: Estimated height
- D: Distance from camera
- P40: Motion percentile feature

Your task:
1. Generate three distinct interpretations (thoughts) of the subject's activity.
2. For each, explain how the features suggest a specific activity.
3. Choose one final activity per thought from:
   Walking, Standing, Sitting, Lying Down, Falling
4. At the end, pick the most likely final activity based on your thoughts.

Respond using only the exact format shown in the example. Do not add any commentary.

---

Example:

Features: 1.12 | 1980.0 | 1020.5 | 0.05000

Thought 1:
- Reasoning: High MaxStdXZ suggests sudden, uncontrolled movement. HHmaxRatio > 1 indicates a posture drop. This is consistent with falling.
- Final activity: Falling

Thought 2:
- Reasoning: MajorMinorRatio and HWR are high, indicating upright posture. Moderate motion and posture stability could suggest standing.
- Final activity: Standing

Thought 3:
- Reasoning: High MaxStdXZ with consistent height and balanced posture suggests intentional movement, possibly walking.
- Final activity: Walking

Final decision (only one word): Falling

---

Now classify the following:

Features: {query}

Thought 1:
- Reasoning:
- Final activity:

Thought 2:
- Reasoning:
- Final activity:

Thought 3:
- Reasoning:
- Final activity:

Final decision (only one word):
"""
```

Solutions

Candidate Thoughts & Activities:
Thought 1: Falling
Thought 2: Falling
Thought 3: Falling

Final ToT Predicted Activity: Falling

Prompting Techniques

Prompting Technique	Accuracy	Precision	Recall	F1-score
Tree-of-Thought (ToT)	93.0	94.0	93.0	93.0
Chain-of-Thought (CoT)	89.0	90.0	88.0	89.0
Few-Shot	86.0	87.0	85.0	86.0
One-Shot	84.0	85.0	83.0	84.0
Zero-Shot	82.0	83.0	80.0	81.0

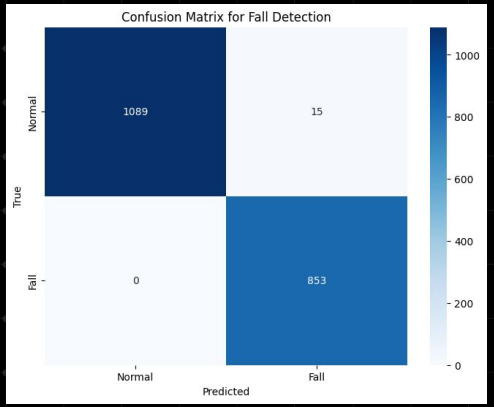
RAG

- Convert posture features to natural language (text format)
- Generate embeddings using *multi-qa-MiniLM-L6-cos-v1*
- Store embeddings in a FAISS index
- For any new activity:
  Convert to text → encode → retrieve top-k similar examples
- LLM (GPT-3.5) + rule-based logic classify as Fall or Normal
- GPT gives both decision and justification.

```
prompt = "You are a fall detection expert AI assistant.\n\n"
prompt += "Classify each activity as 'Fall' or 'Normal' based on posture features and past examples:\n"
prompt += "- Distance to floor (D) < 500mm suggests a Fall.\n"
prompt += "- Height < 850mm might indicate falling.\n"
prompt += "- HHmaxRatio < 0.4 could suggest height reduction.\n"
prompt += "- P40 > 0.4 might mean the body is near the floor.\n\n"
```

```
Enter posture readings for a new activity

HeightWidthRatio: 6
MajorMinorRatio: 7
BoundingBoxOccupancy: 5
MaxStdXZ: 230
HHmaxRatio: 0.6
Height (mm): 450
Distance to floor (mm): 430
P40: 0.3

--- AI Response ---

Decision: Fall

Explanation: The new activity has a distance to the floor of 430mm, which is less than 500mm, suggesting a fall. Additionally, the Height of 450mm is significantly lower than the typical threshold values, indicating a posture that aligns with a fall scenario.
```

Accuracy	0.75
Precision	0.6734
Recall	0.8414
F1 score	0.7481



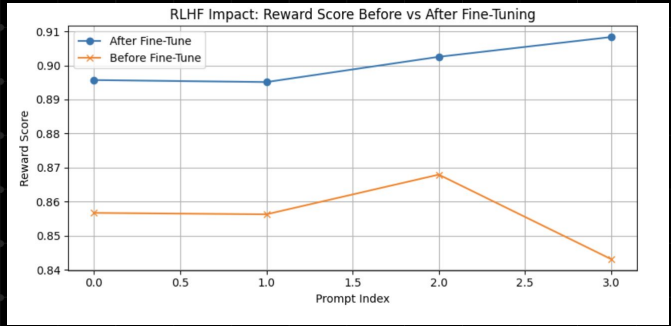
Fine Tuning - LoRA + RLHF

LoRA Fine Tuning
Model - Google/flan-t5-small
Format - Instruction - Input - Response

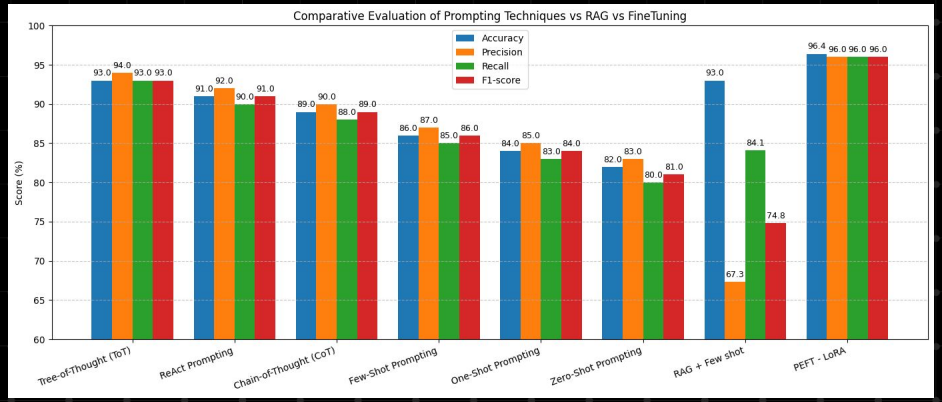
Reward Modeling
Model - bert-base-uncased(binary classifier)
Input - Prompt + Response
Label: 1= good, 0= poor response

RLHF:
Generator: gpt2, 4 responses per prompt
Score with reward model
Select top response
Fine-tune gpt2 on high-reward outputs

Class	Precision	Recall	F1 score
Falling	0.96	0.96	0.96
Lying	0.97	0.97	0.97
Sitting	0.94	0.94	0.94
Standing	0.95	0.95	0.95



Conclusion and Future Work



Key Findings

- Our system can detect falls accurately by combining video and sensor data.
- Using ToT prompts and RAG technique helps the model make better, clearer decisions.

Future Directions

- Work on large annotated datasets.
- Explore quantization to make model lightweight for edge deployment.