An AI-Driven Camera-Based Platform for Patient Ambulation Assessment

Mostafa Habibi, Mehrdad Nourani
Predictive Analytics & Technologies Laboratory
The University of Texas at Dallas
Richardson, TX 75080, USA
{mostafa.habibidehsheikhi, nourani}@utdallas.edu

Dennis H. Sullivan
Geriatric Research, Education and Clinical Center
Central Arkansas Veterans Healthcare System
North Little Rock, AR 72114, USA
dennis.sullivan3@va.gov

Outline

Motivation and Contribution

Proposed Methodology

- Person & Head Detection and Tracking
- Head Coordinates in 3D and Movement Measurement
- Body Landmark Estimation and Feature Extraction
- Posture Classification

Experimental Result

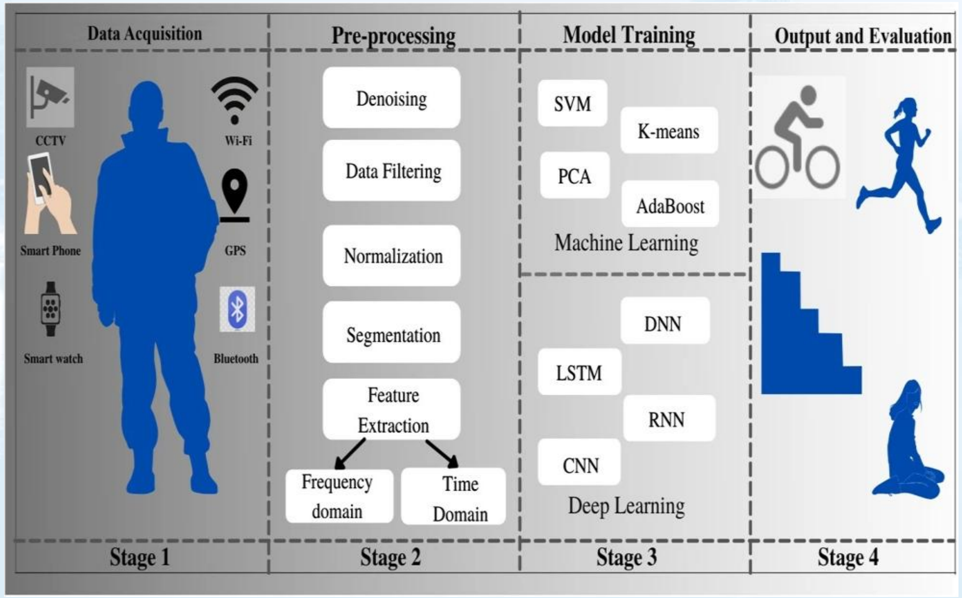
Comparison and Conclusion

Motivation

- **Patient Ambulation**
  - **Definition:** The act of assisting a patient to move around (simple bedside activities to more extensive walking exercises) within a healthcare facility or at home.
  - **Challenges:** Existing technologies for monitoring patient ambulation in a clinical setting suffering from low adoption due to high cost, inaccuracy, and/or difficulty in use.
- **Significance**
  - Facilitates Faster Recovery
  - Reduces Hospital Stay Duration
  - Improves Blood Circulation
  - Enhances Respiratory Function
  - Prevents Muscle Atrophy

Human Action Recognition (HAR)

- **Definition:** The automatic identification of physical activities performed by individuals.
- **Applications:** HAR is used in various fields such as healthcare, sports, security, and human-computer interaction.
- HAR helps improve patient movement by providing continuous, real-time monitoring to ensure safety and offer useful data for personalized care and rehabilitation.

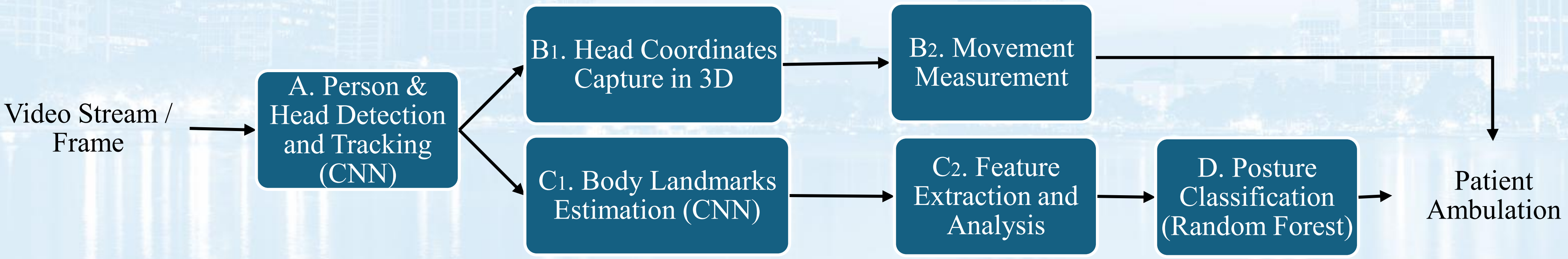


Prior Works

Technique	Pros	Cons
Steel et al. [19]	High precision	Impractical, Requires infrastructure
S. K. Yasav et al. [4]	High precision	Inconvenient for patients, Expensive
I. Cheol Jeong et al. [6]	Reasonable estimates	Expensive to install
Z. You et al. [18]	High accuracy, Supports telerehabilitation	Requires special equipment

Proposed Methodology

- Our platform includes four main modules:
  A. Person and head detection and tracking
  B. The head's 3D coordinate, and movement measurement
  C. Landmark estimation to provide features
  D. Twelve postures classifications

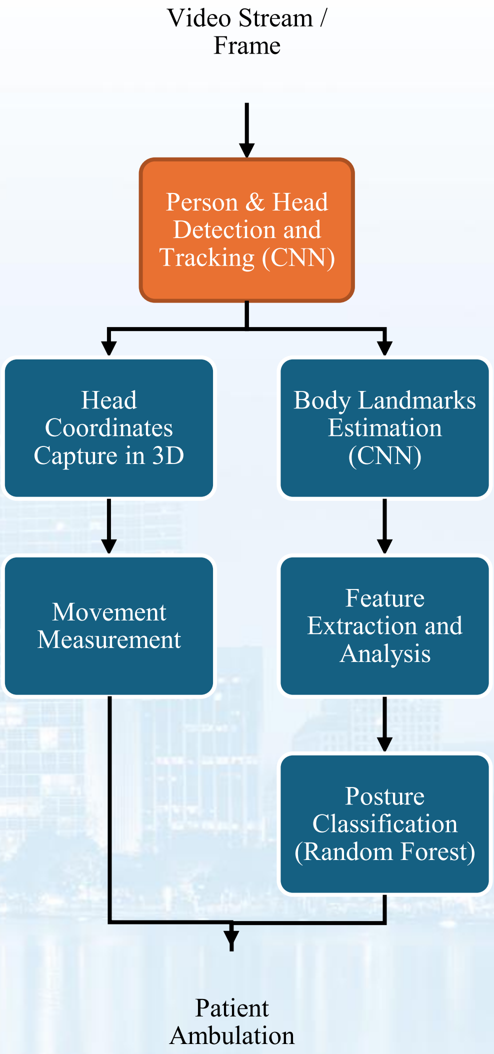


Person & Head Detection and Tracking

- **Person and Head Detection**
    1. Utilize CNN architecture inspired by YOLO.
    2. Training involves feeding annotated images to adjust model weights and biases.
    3. Customized CrowdHuman dataset used, focusing on head and body bounding boxes.

- **Head Tracking**

  Employ BoT-SORT for tracking object to estimate object states over time (position, velocity, acceleration).
    - Objects in consecutive frames are located based on previous positions.
    - Object re-identification ensures consistent tracking with unique IDs assigned to each object.



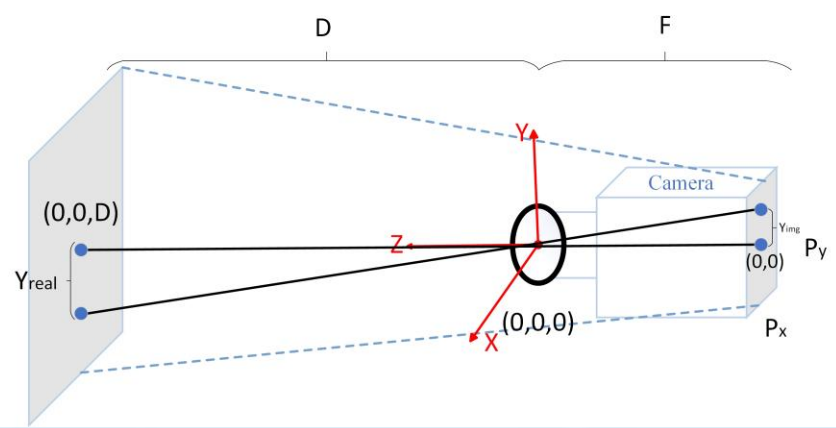
Head Coordinates Capture

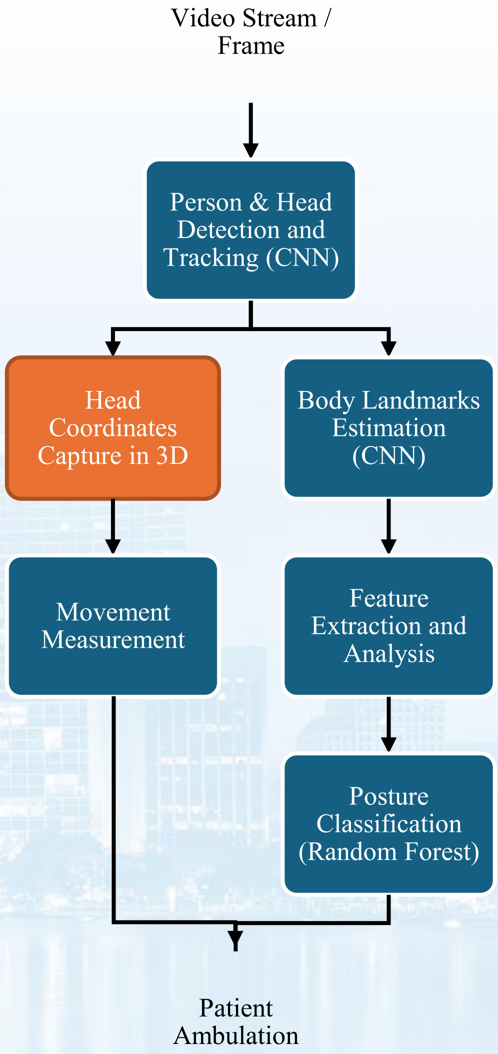
- Head Distance Calculation
  - Head's center-to-camera distance (D_{Head}) estimated using head width in image ($W_{Head\ Img}$), real width ($W_{Head\ Real}$) and focal length of camera (F).

$$D_{Head} = \frac{W_{Head\ Real} \times F}{W_{Head\ Img}}$$

- Head Coordinates in 3D
  - Head position calculated in both real-world (3D) and image (2D) coordinates.

$$\begin{cases} X_{real} = \frac{D_{Head} \times X_{img}}{F} \\ Y_{real} = \frac{D_{Head} \times Y_{img}}{F} \\ Z_{real} = D_{Head} \end{cases}$$

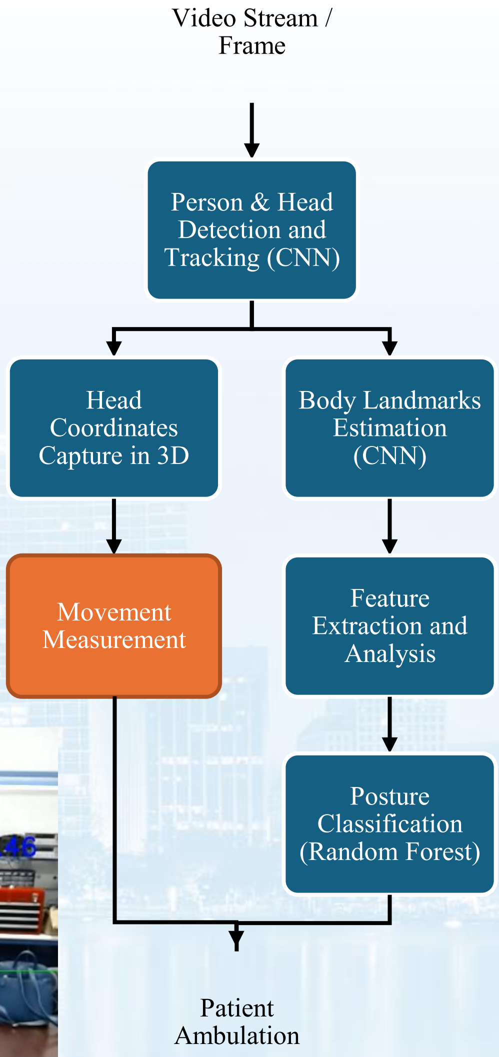




Movement Measurement

- Calculation of Movement between Frames
  - Euclidean formula is used to determine total displacement between two consecutive frames.
- Total Human Movement Calculation
  - By tracking head and summing distance between N consecutive frames, total human movement for a time window is calculated.

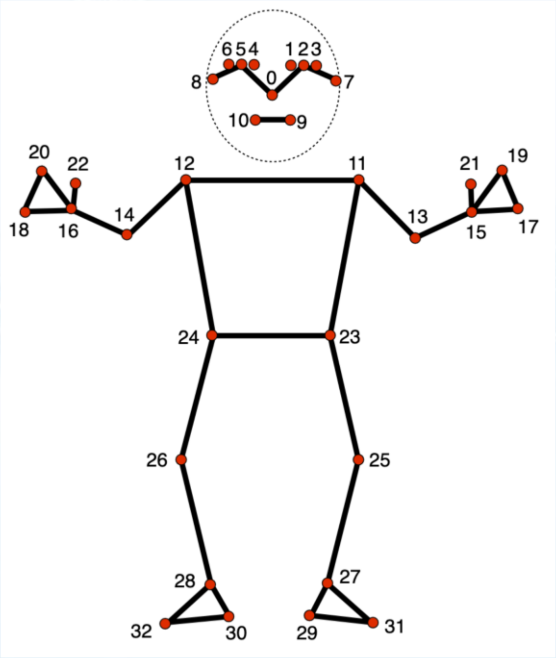
$$d_{total} = \sum_{i=1}^{N} [(X_{i+1} - X_i)^2 + (Y_{i+1} - Y_i)^2 + (Z_{i+1} - Z_i)^2]^{\frac{1}{2}}$$

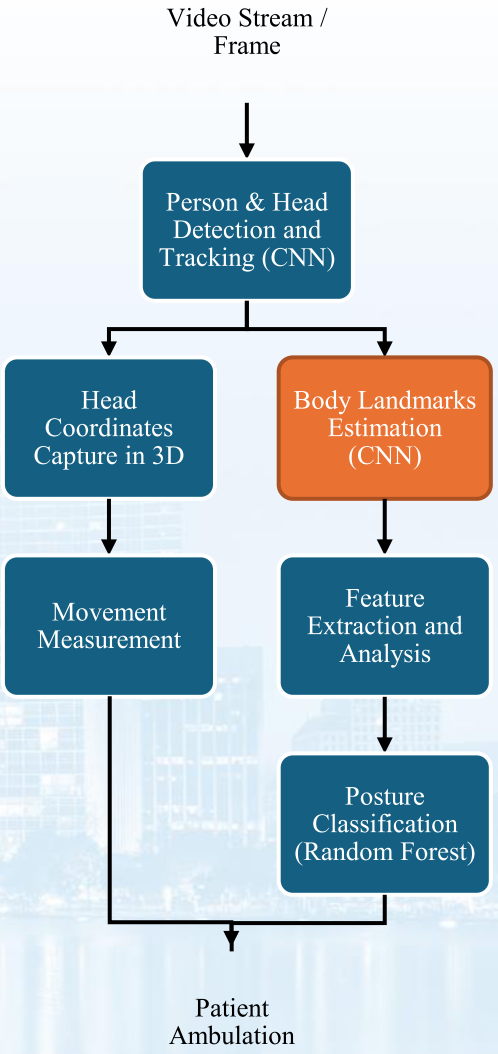


Body Landmarks Estimation

- We have employed the **Mediapipe**’s general pose model to track the approximate locations of 33 body landmarks.
  - An open-source framework for building pipelines to perform computer vision.
  - Inference over arbitrary sensory data such as video or audio.

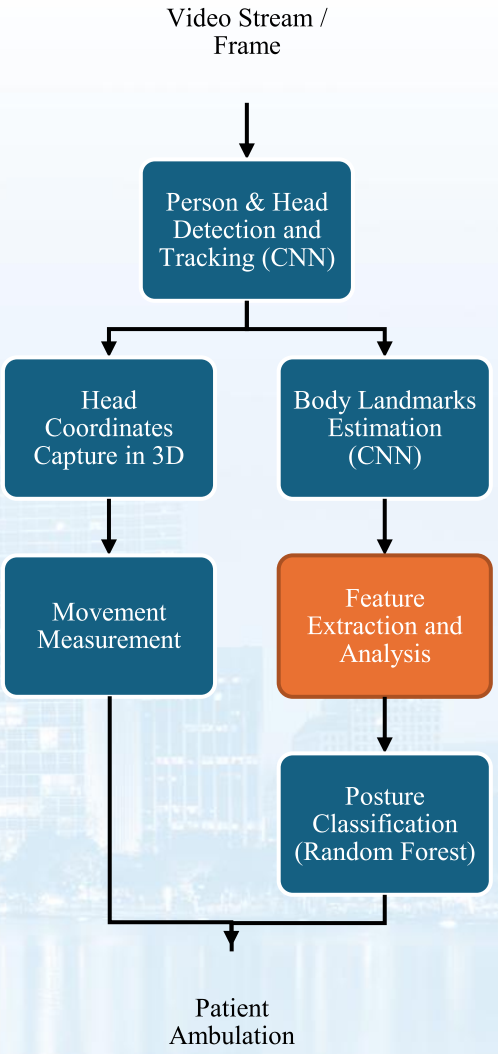
- **Customizations for our application**
  - Estimating 3D coordinates of landmarks
  - Tracking the relevant landmarks effectively
  - Using ML to extract useful feature vectors
  - Dealing with incompleteness and inaccuracies
  - Training the ML-based techniques



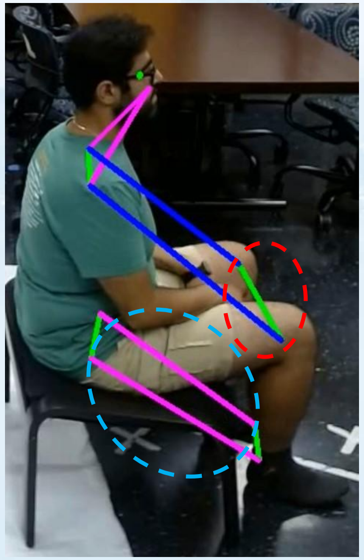


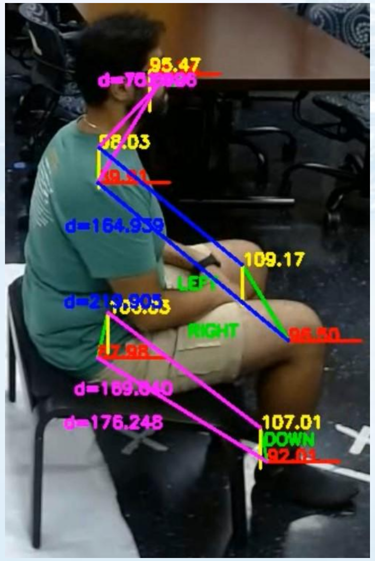
Feature Extraction and Analysis

- Eleven specific landmarks are selected.
- For feature extraction, using selected landmarks we construct eleven vectors.
  - Five main-body vectors
  - Six custom vectors
- From this 11 vectors we extract 33 features (6 lengths and 27 angles).



Main Vector	Start	End	Custom Vector	Start	End
Eye Line	4	1	Nose → L_Shoulder	0	11
Shoulder Line	12	11	Nose → R_Shoulder	0	12
Hip Line	24	23	L_Shoulder → L_Knee	11	25
Knee Line	26	25	R_Shoulder → R_Knee	12	26
Ankle Line	28	27	L_Hip → L_Ankle	23	27
-	-	-	R_Hip → R_Ankle	24	28

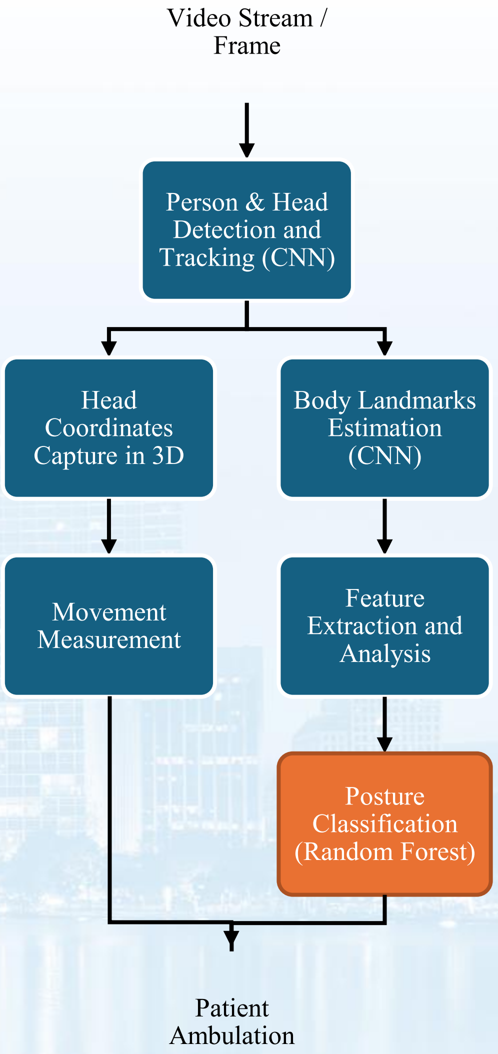




Posture Classification

- **Random Forest Tree:** A machine learning method that combines the strengths of many decision trees. Builds multiple decision trees, each using a random subset of the data and random features for splitting decisions.
- We employed a Random Forest classifier known for
  - Reduces variance and avoids overfitting the data
  - High accuracy
  - Robustness
  - Flexibility
  - Low sensitivity to noise and outliers

Posture	Classes
Standing	**1. Front, 2. Back, 3. Right Side, 4. Left Side**
Sitting	**5. Front, 6. Back, 7. Right Side, 8. Left Side**
Lying	**9. Supine, 10. Prone, 11. Right Side, 12. Left Side**



Experiment Setup

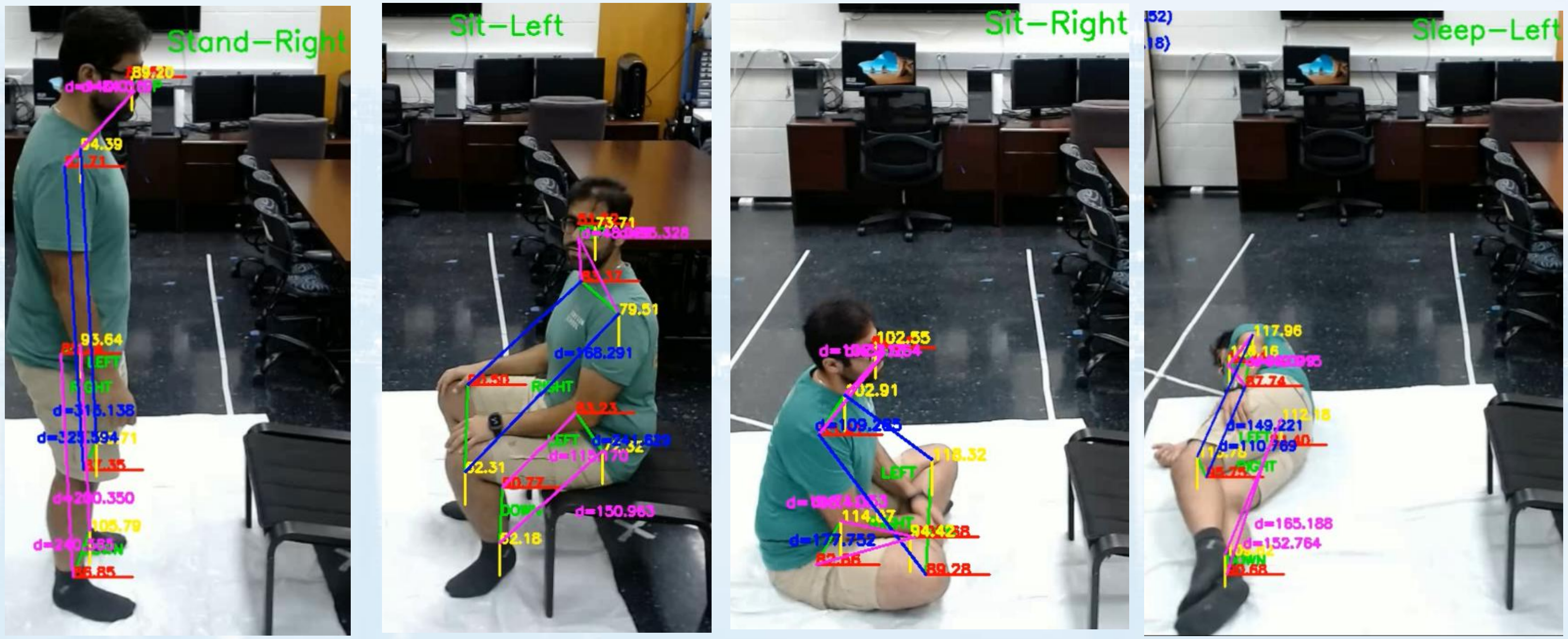
Hardware	
Webcam	Logitech C920S
Resolution	1920 × 1080 pixels
Frame Rate	30 FPS
RAM	32GB
CPU	Intel i9-12900HK
GPU	NVIDIA GeForce RTX 3050
Deep Learning Frameworks	CUDA toolkit, cuDNN module

Person and Head Detection	
Model	Customized CNN (YOLO)
Training Data	Customized CrowdHuman dataset (24,370 images)
Train-Validation-Test Split	80% - 10% - 10%
Validation F1 Score	Head 0.87, Person 0.90

Posture Detection	
Custom Dataset	22,750 images from webcam recordings
Classifier	Random Forest Tree
Cross-validation	5-fold (80% training, 20% testing)

Experimental Results

- **Classification**
  - We used our random forest classifier and achieved 98% accuracy, 0.8% Mean error, and 98% F1 score based on the average values for all 12 posture classes.



Experimental Results

- **Walking Experiment**
  - Ten volunteers walked along a 13-meter path at three speeds (slow, normal, and fast).
  - Average metrics over a window length (e.g. 20 frames equivalent to 0.7 sec for 30 FPS video) to avoid error accumulation.

Size		Average Distance Error (%)			
Window	Overlap	Fast	Normal	Slow	Avg.
30	15	-11.46	-7.92	-2.23	7.2
20	**15**	**-9.46**	**-4.38**	**3.08**	**5.65**
15	10	-9.08	-3.62	5.23	5.97

- **Observation**
  - For fast walking, a smaller window size and overlap produced better results.
  - For slow walking, a larger window and overlap sizes showed better results.
  - Empirically, window size of 20 frames (0.7 sec) and overlap of 15 frames (0.5 sec) provided the minimum error for the platform using a 30 FPS camera.

Comparison

Reference	Technology	F1 Score	Key Features
Z. You et al.[18]	Depth Camera	93.7%	Equipment and computationally expensive
I. Choeol Jeong et al.[6]	RTLS	N / A	Requires IR beacons and sensors. Only targets distance.
Steele et al.[19]	RGB Camera	95.8%	Attach AprilTag on target
Our Work	**RGB Camera**	**98.0%**	**High accuracy and inexpensive equipment**

- **Our work**
  - High Accuracy
  - No Specialized Equipment
  - Ease of Implementation
  - Cost-Effective

Conclusion

- We utilized customized datasets to train CNN for human detection and tracking.
- We mapped 2D image coordinates to real-world 3D locations.
- Feasibility in identifying complex features and accurately classifying postures have been studied.
- Our platform enhances care delivery and cost-effectiveness in home healthcare and rehabilitation settings.
- Proposed methodology has application in patient identification, monitoring, and tailored rehabilitation plans.

Thanks for Your Attention!

Mostafa Habibi, Mehrdad Nourani
Department of ECE
The University of Texas at Dallas
{mostafa.habibidehsheikhi, nourani}@utdallas.edu

Dennis H. Sullivan
Geriatric Research, Education and Clinical Center
Central Arkansas Veterans Healthcare System
dennis.sullivan3@va.gov