**Human Activity Recognition**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Project Mentor: Dr. Namitha K

Project Lead: Girish S

Team Members

|  |  |
| --- | --- |
| Girish S | AM.EN.U4AIE22044 |
| Anuvind MP | AM.EN.U4AIE22010 |
| R S Harish Kumar | AM.EN.U4AIE22043 |
| Harishankar Binu Nair | AM.EN.U4AIE22023 |

**Objective:** Addressing a specific domain for research on Human Activity Recognition

# **Overview:**

On thorough review on various applications of HAR (Human Activity Recognition). Several ideas were proposed in a recent session, each proposed idea was looked into and had a thorough review on the details such as the main task, previous papers published on the ideas, datasets available, challenges faced in the problems, etc.

The titles of the various proposed ideas are mentioned below in section 1

* **Parkinson's Disease tremor detection**
* **Assistive rehabilitation of stroke patients in rehabilitation centers**
* **Application of optimization techniques in sports based on HAR**
* **enhance athletic performance in matches based on HAR**
* **HAR systems for gesture-based interaction for the disabled**
* **HAR on focusing on understanding student engagement, classroom dynamics, or designing adaptive learning environments.**
* **HAR on driver safety recognition**
* **HAR on workflow optimization**
* **HAR on sensorless VR/AR recognition**
* **HAR on behavior analysis using body language and pose estimation**
* **Sign Language Interpretation**
* **Assistive Navigation for the Visually Impaired using skeletal pose estimation**

# **Review:**

1. **Parkinson's Disease tremor detection**

ASD & PD

---------

- ASD and PD are respectively neuro-developmental and neuro-degenerative disorders, each

with different symptoms involving atypical motor movements.

- PD affects the motor system causing motor symptoms such as tremors, bradykinesia (slowness), Freezing of Gait (FOG), and muscle

rigidity.

\* FOG increases the risk of falling generally in elderly PD patients

- ASD has also some specific motor behavior symptoms such as Stereotypical Motor Movements (SMMs)

\*SMMs are the major group of abnormal repetitive behaviors, e.g., hand flapping and body rocking, in children

\*decrease the performance of children while learning new skills or using learned skills

----------------------------------------------------------------------------------------------------------------------------------

-- Inertial Measurement Unit

(IMU) sensors, have provided an effective platform for remote monitoring of patients with motor

malfunctions such as Parkinson’s Disease (PD) [1] and Autism Spectrum Disorder (ASD) [2]. IMUs

contain built-in accelerometers, gyroscopes and magnetometer sensors allowing one to measure the

angular velocity and linear acceleration of body parts during movement.

- DEEP NORMATIVE MODELING USED

- PROBABLISTIC NOVELTY DETECTION :

\* novelty detection is defined as the task of learning the overall characteristics of available normal samples in the training

phase and then using these characteristics to recognize novel samples that differ in some respects from the normal samples at test time

\* probabilistic policy in novelty detection enables us to estimate the generative probability density function of the normal data, which can cover a wide and heterogeneous spectrum of normal samples

\*In general, a normative model is constructed in the training phase by estimating a

mapping function between two different data modalities, e.g., behavioral covariates and biological

measurements.

--- DENOISING AUTOENCODER (DAE)

\* In some applications, such as ours, only one modality of data is available. To overcome

this barrier, we use the denoising autoencoder (DAE) to reconstruct the original IMU signals of normal

movements from their noisy versions.

\* the model implicitly learns the distribution of the

normal movements. Using dropout layers in the DAE architecture enables us to estimate also the

variance of predictions (which is necessary for normative modeling) in addition to mean predictions.

--> DAE :

\* A denoising autoencoder is a type of artificial neural network that is trained to reconstruct clean input data from corrupted or noisy versions of that data. The primary goal of a denoising autoencoder is to learn a robust representation of the input data by removing the effects of noise.

\* An autoencoder is a neural network architecture consisting of an encoder and a decoder. The encoder compresses the input data into a lower-dimensional representation (encoding), and the decoder reconstructs the original input from this encoding.

\* In the training phase of a denoising autoencoder, the input data is intentionally corrupted by adding noise or introducing distortions. This process forces the autoencoder to learn a more robust and meaningful representation of the data.

\* The objective of training a denoising autoencoder is to minimize the reconstruction error between the clean input and the output produced by the network. Common loss functions include mean squared error (MSE) or binary cross-entropy, depending on the nature of the data.

\* During training, the autoencoder learns to map the corrupted input to a clean reconstruction. The weights of the network are adjusted through backpropagation to minimize the reconstruction error.

----------------------------------------------------------------------------------------------------------------------------------

-> METHODOLOGY :

\* novelty detection is defined as detecting atypical movements in the test phase while only normal movements are available

in the training phase.

\* normal models are modeled using cnn

-- probabilistic novelty detection approach consisting of the following three steps :

1. learning the distribution of normal movements using a probabilistic denoising autoencode

2. quantifying the deviation of each test sample from the distribution of normal movements, the so-called Normative Probability Map (NPM), in the normative modeling framework

3. computing the degree of novelty of each test sample by fitting a generalized extreme value distribution on summary statistics of its NPM.

----------------------------------------------------------------------------------------------------------------------------------

-> Network architecture :

1. DAE architecture for FOG dataset :

\*CNN used for encoding the signal into a lower dimensional representation

\* This architecture contains :

fourconvolutional layers, alternating convolution, batch normalization, Rectified Linear Units(ReLU) and max-pooling layers to map the large input space to a lower dimensional feature space

\*

2. DAE architecture for SMM datasets:

\* the encoder architecture consists of

three convolutional layers,whichalternates convolution, batchnormalization, ReLUs and average-pooling layers to transform the raw feature space into a lower dimensional set of features

-------------------------------------------------------------------------------------------------------------------------------

1. **HAR on driver safety recognition**

\*\*\*DAR\*\*\*

-----------------------------

=> a multiview, multimodal vision framework in order to characterize driver activity based on head, eye, and hand cues.

=> work leverages two views for driver activity analysis, a camera looking at the driver’s hand and another looking at the head

------------------------------------------------------------------------------------------------------------------------------------

=>HOG (histogram oriented gradient) descriptor:

------------------------------------------------------------------------------------------------------------------------------------

FEATURE EXTRACTION MODULE :

1. Hand cues:

\* hand activities may be characterized by zones or regions of interest. These zones are important for understanding driver activities and secondary tasks.

\* structure in the scene can be captured by leveraging information from the multiple salient regions.

\*edge, color, texture, and motion features were studied for the purpose of hand activity recognition, found that edge

features were particularly successful

\* in this work we employ a pyramidal representation for each region using Histogram of Oriented Gradients (HOG), with cell sizes 1(over the entire region), 4, and 8 for a 8+128+512 = 648 dimensional feature vector.

\*\*pyramidal representation using hog : capture information at various resolutions, allowing the detection of objects at different sizes. The original image is often resized to create a set of images at different scales. This set of images is referred to as an "image pyramid." Each level of the pyramid corresponds to a different scale of the image.For each level of the image pyramid, HOG features are computed independently.

2. head and eye cues:

\* In our implementation, the eye state at time t is estimated using two variables: area of the eye and area of the face.

Area of the eye is the area of a polygon whose vertices are the detected facial landmarks around the left or right eye.

Similarly, the area of the face is the area of the smallest polygon that encompass all the detected facial landmarks. To

compute the level of eye opening, we divide area of the eye by the area of the face at every time t. This normalization

will allow the computation of eye opening to be invariable to driver’s physical distance to the camera, where closer distances

makes the face appear larger in the image plane. Finally, a normalization constant learned for each driver representing

his or her normal eye-opening state is used such that after normalization values < 1 represent downward glances and values > 1 represent upward glances

\*\*activity classification :

{i)wheel

(ii)gear

(iii) instrument cluster

----------------------------------------------------------------------------------------------------------------------------------  
\*\*\*ACTIVIY RECOGNITION FRAMEWORK\*\*\*

=> classifier used : linear kernel svm

----------------------------------------------------------------------------------------------------------------------------------

The methodology for developing a "**Posture Analysis for Stress or Anxiety Detection**" AI system involves several key steps:

* **Data Collection:**
  + Acquire a diverse dataset of video recordings that capture individuals in various contexts, including situations that may induce stress or anxiety.
  + Annotate the dataset with labels indicating the stress or anxiety levels, ensuring a balanced representation of different scenarios.
* **Preprocessing:**
  + Extract relevant features from the video data, such as key points or joints in the body, using a pose estimation algorithm.
  + Normalize and preprocess the data to remove noise, ensuring consistency across different video sources.
* **Pose Estimation:**
  + Implement a pose estimation model to identify the 2D or 3D positions of key body joints in each frame of the video.
  + Choose a suitable pose estimation algorithm, such as OpenPose or PoseNet, based on the project requirements.
* **Feature Extraction:**
  + Derive meaningful features from the pose data that can represent body posture and movement patterns.
  + Consider features such as angles between joints, body sway, or the relative position of key body parts.
* **Labeling for Training:**
  + Annotate the dataset with labels indicating the stress or anxiety levels corresponding to each video segment.
  + Ensure that the labeling process involves experts or individuals familiar with stress and anxiety cues.
* **Model Training:**
  + Select or design a machine learning model suitable for the task, such as a neural network or a classic machine learning algorithm.
  + Train the model using the labeled dataset, optimizing it to recognize patterns associated with stress or anxiety based on posture and movement features.
* **Validation and Testing:**
  + Split the dataset into training, validation, and test sets to assess the model's performance.
  + Validate the model's effectiveness on unseen data and fine-tune hyperparameters if necessary.
* **Threshold Setting:**
  + Establish a threshold for stress or anxiety detection based on the model's output probabilities.
  + Determine a balance to avoid false positives and false negatives, considering the impact on end-users.
* **Real-time Processing:**
  + Implement real-time processing capabilities for the AI system to analyze posture and movements in live or recorded video streams.
  + Optimize the system for low latency to enable timely stress or anxiety detection.
* **User Interface Integration:**
  + Develop a user-friendly interface that communicates stress or anxiety detection results to end-users.
  + Consider incorporating visualization tools to display detected postures or movements.
* **Ethical Considerations:**
  + Address privacy concerns by anonymizing and securing video data.
  + Ensure transparency in communicating the limitations and potential biases of the model.
* **User Feedback and Iteration:**
  + Gather feedback from users and experts to iteratively improve the model's performance and user experience.
  + Consider ongoing updates and refinements to enhance the system's accuracy and applicability.