Presented By : Vaishnavi Manjunatha
Namratha Renjal

HUMAN STRESS PREDICTION USING ACTIVITY DATA

Supervisor : Prof Hyowon Lee

Introduction

- Prevalence of Stress: Common in modern life, impacting both physical and mental health.
- Effects of Excessive Stress: Can lead to issues like high blood pressure, anxiety, and weakened immune function.
- Importance of Stress Prediction: Early intervention allows for better management and prevention of serious health problems.
- Project Focus: Analyzing daily activities to predict stress, enabling individuals to manage stress proactively.
- Methodology Overview: Uses activity data, to infer stress levels based on surveys completed by users.

Literature Review

Human Activity Recognition

1. Non parametric discovery of human routines from sensor data(F.-T. Sun, Y.-T. Yeh Et.al, 2014)
2. Activities of daily living monitoring via a wearable camera: Toward real-world applications(Cartas, A., Radeva Et.al,2020)
3. Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning.(M. Z. Uddin and A. SoyluEt.al,2021)

Stress Prediction

1. Prediction of daily happiness using supervised learning of multimodal lifelog data (T. Yamamoto, J. Yoshimoto Et.al ,2019)
2. Mood detection and prediction based on user daily activities (P. Soleimaninejadian, M. Zhang, Y. Liu, and S. Ma Et.al, 2018).
3. A lifelogging platform towards detecting negative emotions in everyday life using wearable devices(C. Dobbins, S. Fairclough Et.al,2018)

Insights from literature Review

- SVM is commonly used for activity and stress recognition due to its effectiveness in feature-based classification.
- LSTM is proven to provide the best accuracy in some stress prediction cases using sensor data.
- Current Research Focuses on sensor-based stress prediction (e.g., heart rate).
- HAR is mostly performed for basic physical activities.
- Traditional studies rely on user input (e.g., surveys, mood logs), which requires manual effort.
- Research Gap: Limited exploration of the impact of user activities on stress prediction in a technical context.

Research Questions

1) Can the accuracy of stress prediction be enhanced by incorporating activity duration, activity frequency, or both for users?

2) Can the accuracy of stress prediction be improved by utilizing data collected closer to the time when the ground truth was captured, as opposed to using data from the entire day?

Dataset Description

- The ETRI dataset includes 570 days of experimental sessions from 22 subjects.

- Dataset contains-

1. Sensor data- Acc,gyr,mag,gps,eda,temp,bvp,hr
2. Activity label details with timestamp
3. User survey with stress data

Activity labels data

Column	Options (Descriptions)
ts	timestamp
action	sleep, personal_care, work, study, household, care_housemem (caregiving), recreation_media, entertainment, outdoor_act (sports), hobby, recreation_etc (free time), shop, communitiy_interaction (regular activity), travel (includes commute), meal (includes snack), socialising

Sensor Details

Sensor	Readings
Acc	Mobile phone accelerometer (30 Hz) and E4 accelerometer (32 Hz)
mGyr	Mobile phone gyroscope (30 Hz)
mMag	Mobile phone magnetometer (30 Hz)
mGps	Mobile phone GPS (every 5 seconds)
e4Bvp	E4 photoplethysmography (PPG) sensor (64 Hz)
e4Eda	E4 EDA sensor (4 Hz)
e4Hr	Average heart rate values E4 (1 Hz)
e4Temp	Skin temperature from E4 infrared thermopile (4 Hz)

Data Preprocessing and EDA

1. **Cleaning and consolidation of Sensor Data:** Involves integrating diverse sensor data into a unified dataset. This includes sensor readings and activity-related data, which are merged into a single file to ensure comprehensive access to all relevant information.

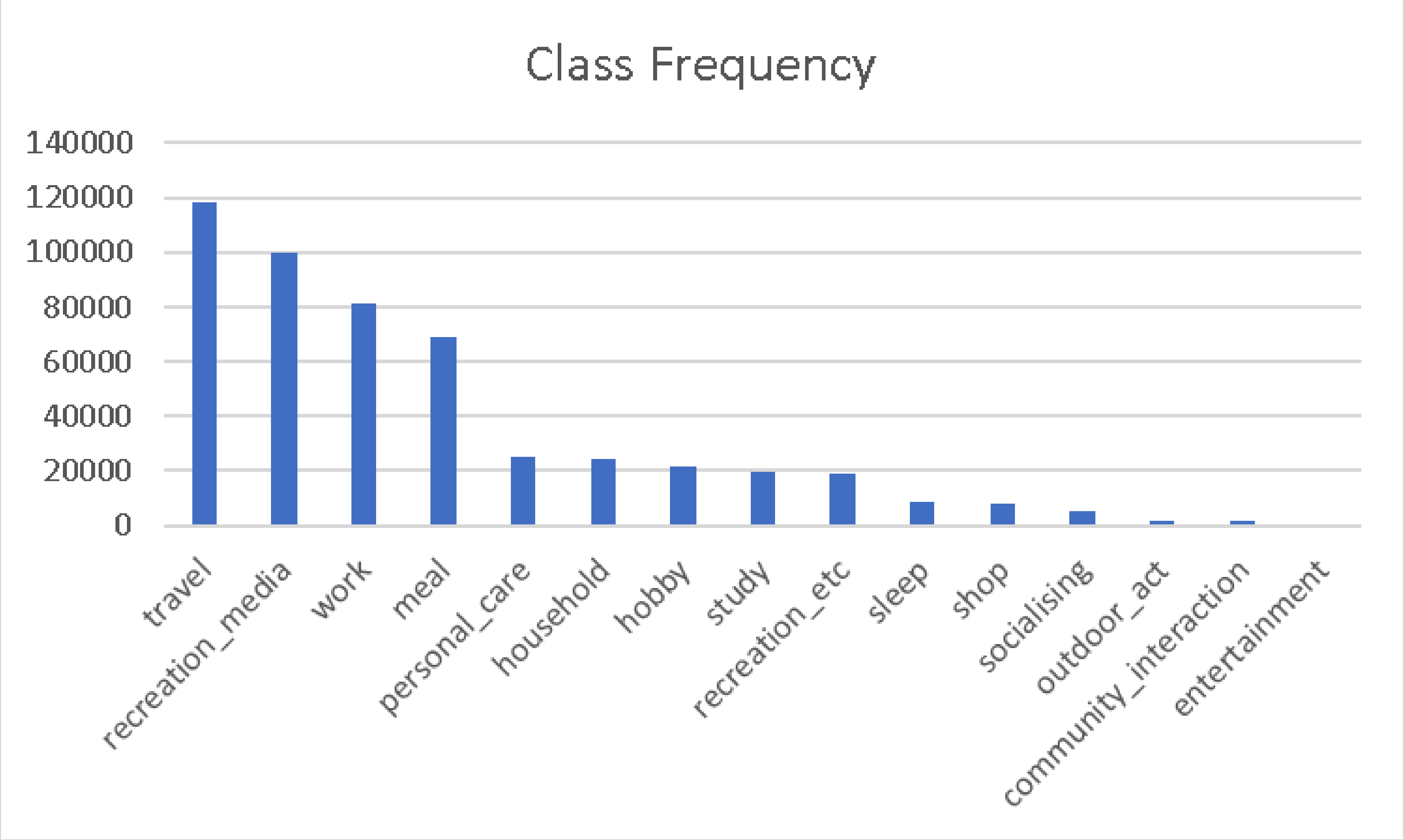
2. **Frequency Normalization:** Adjustments such as upsampling or downsampling are performed to standardize the data, ensuring compatibility across different sensor streams.

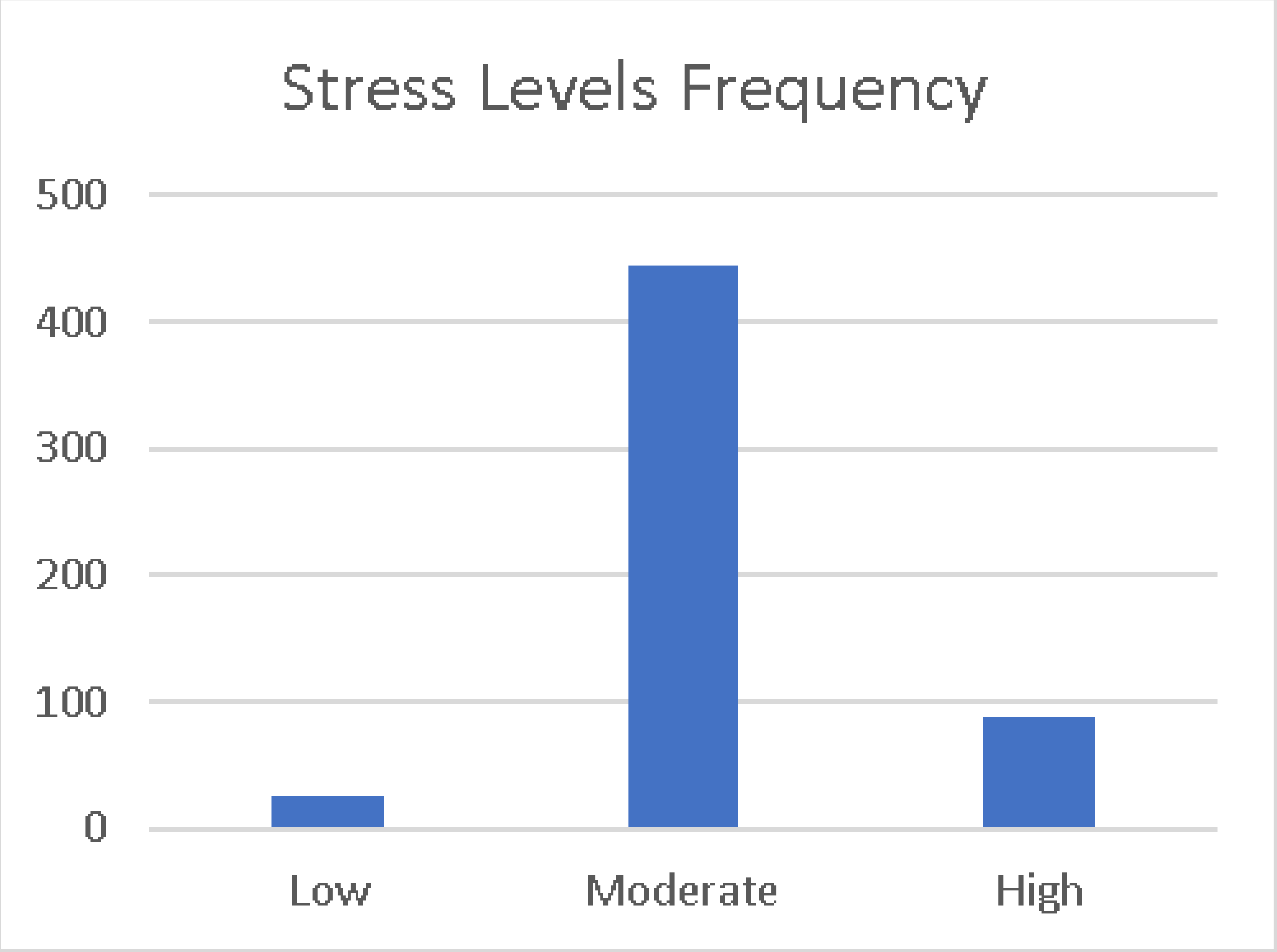
3. **Dimensionality Reduction:** A correlation matrix is used to identify and eliminate highly correlated features, reducing redundancy and improving model efficiency.

4. **Formatting for Stress Prediction:** The data is formatted to calculate and store the duration and frequency of each activity, integrating this with users' emotion and stress scores collected via surveys.

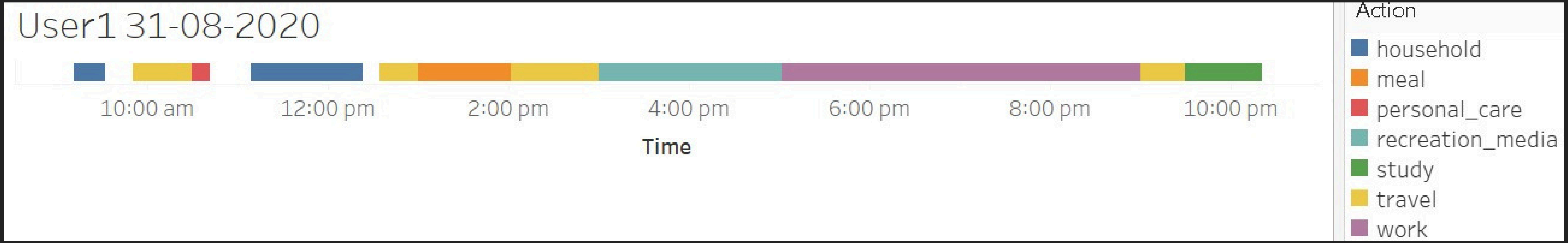
Exploratory Data Analysis

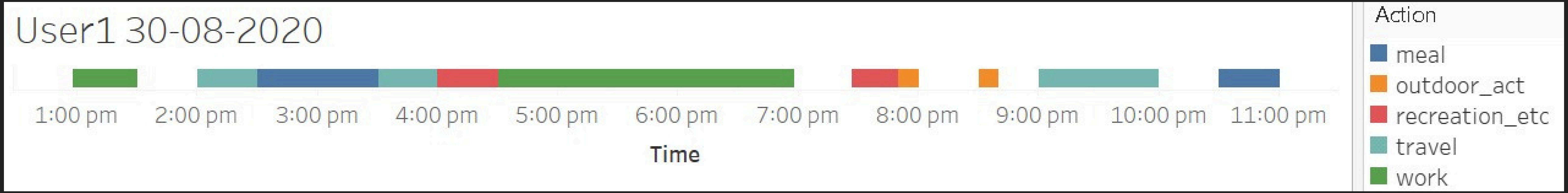
We observed the presence of class imbalance in activity labels. Certain activities, like work or sleep, were recorded more frequently, while others, like recreational activities or shopping, were underrepresented.



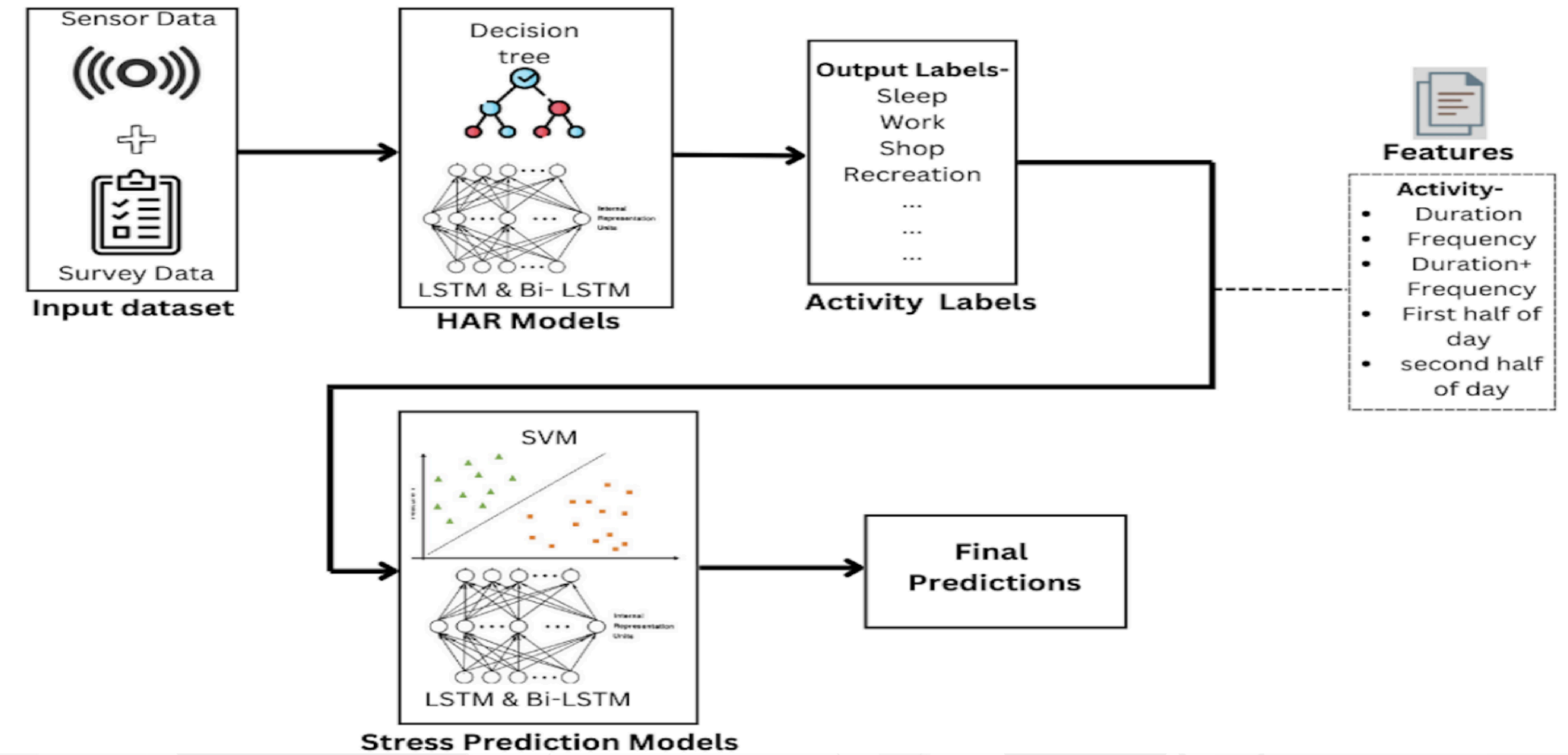


Gantt chart of activities





Proposal



Sensor Data
Survey Data
Input dataset
Decision tree
LSTM & Bi- LSTM
HAR Models
Output Labels-
Sleep
Work
Shop
Recreation
...
...
...
Activity Labels
Features
Activity-
• Duration
• Frequency
• Duration+ Frequency
• First half of day
• second half of day
SVM
LSTM & Bi-LSTM
Stress Prediction Models
Final Predictions

Data Augmentation

To address the class imbalance in activity labels observed during exploratory analysis, SMOTE (Synthetic Minority Over-sampling Technique) is applied to generate synthetic samples for the minority class.

Uses-
Overfitting Reduction
Enhanced model performance and accuracy

ML Techniques

Human Activity Recognition (HAR) Models

Decision Tree:

- Recursive splitting of data based on significant features.
- Handles non-linear relationships in sensor data.
- Interpretable model, useful for extracting meaningful insights.

Long Short-Term Memory (LSTM):

- Recurrent neural network variant, effective for time-series data.
- Captures temporal dependencies in sequential sensor data.
- Suitable for predicting activities over time.

Bi-Directional LSTM (Bi-LSTM):

- Processes input data in both forward and backward directions.
- Enhances accuracy by considering past and future context.
- Outperformed other models for HAR.

Using Predicted Activity Labels the following variation were considered for Stress prediction

Feature Variations:

- Activity duration
- Activity frequency
- Combined frequency and duration
- Activities segmented into first and second half of the day

Stress Prediction Models

SVM :
Baseline model for stress prediction, leveraging its interpretability.

LSTM :

Bi-LSTM :
Highest accuracy (72.5%) for stress prediction using second-half activities.

Results

To evaluate the performance of our classification models we used the metrics:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Results – HAR

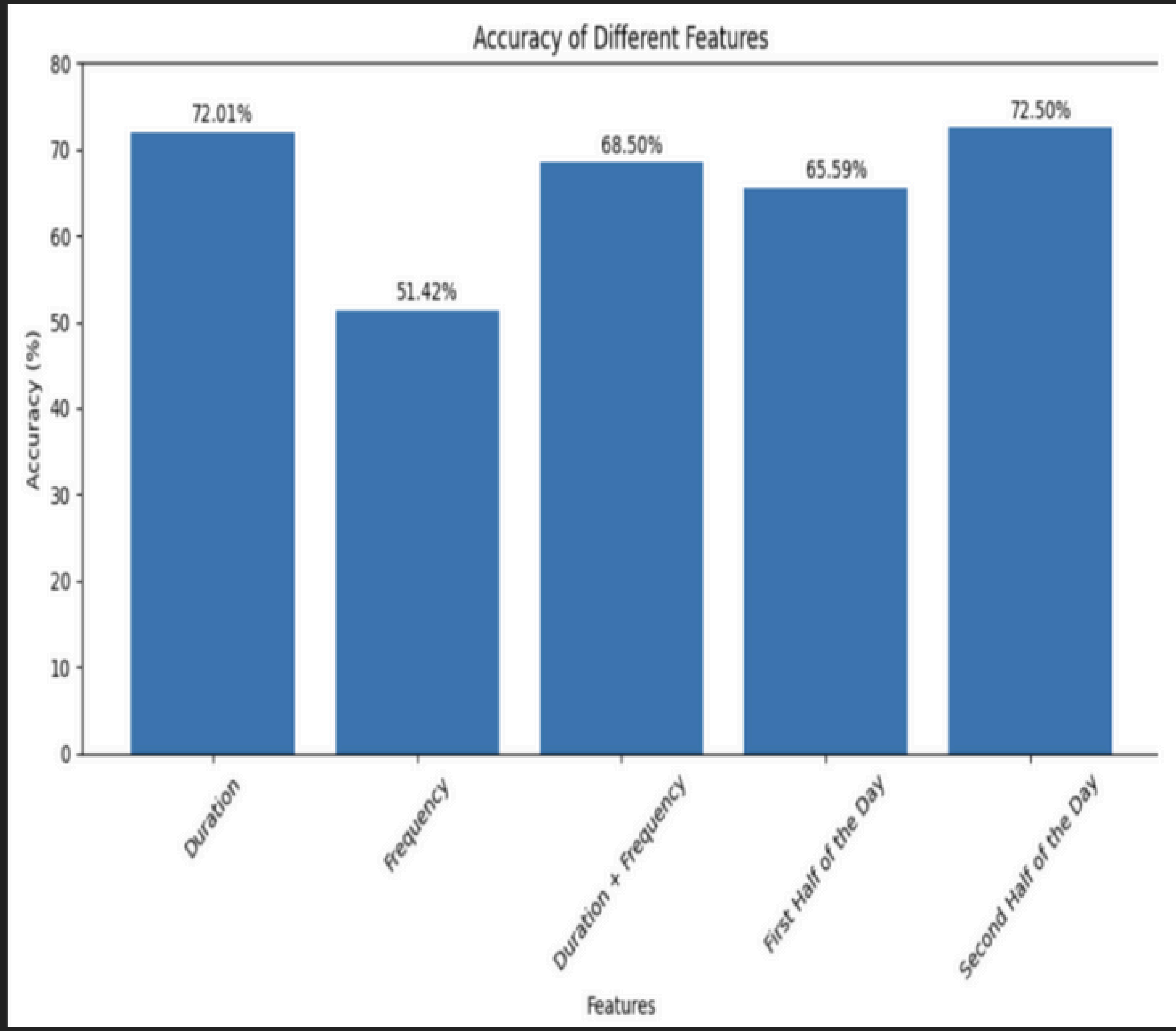
Model	Accuracy	Precision	Recall
Decision Tree(with GridSearchCV)	69.22%	72%	69%
LSTM	85.74%	90.04%	82.81%
Bi-LSTM	**92.18%**	**93.72%**	**90.92%**

Key Finding: Bi-LSTM is most effective for HAR, capturing both past and future contexts more efficiently.

Stress prediction model performance

Models Tested: SVM, LSTM, Bi-LSTM
Bi-LSTM: Consistently superior across all feature variations
Best Performance: Precision of 75.94% when analyzing activities performed in the second half of the day.

Features	Evaluation metrics	SVM	LSTM	Bi-LSTM
Activity duration	Accuracy	70.40%	69.79%	**72.01%**
	Precision	71.54%	70.75%	72.23%
	Recall	70.40%	58.36%	63.80%
Activity duration + Frequency	Accuracy	68.16%	62.91%	68.50%
	Precision	69.88%	66.13%	72.42%
	Recall	68.16%	50.07%	60.22%
Activity frequency	Accuracy	50.65%	51.57%	51.42%
	Precision	49.82%	57.17%	56.60%
	Recall	50.65%	24.55%	27.24%
Activities in the first half of the day	Accuracy	43.49%	61.84%	65.59%
	Precision	51.20%	68.64%	68.63%
	Recall	43.49%	51.91%	61.10%
Activities in the second half of the day	Accuracy	56.02%	65.96%	**72.50%**
	Precision	69.69%	70.65%	**75.94%**
	Recall	56.02%	59.71%	68.16%



Experimental Findings

Data Augmentation:

- Used SMOTE to address class imbalance.
- Improved model accuracy by generating synthetic samples for underrepresented classes.

Human Activity Recognition (HAR):

- Expanded beyond physical activities to include behavioral and contextual activities (e.g., work, travel, personal care).
- Models effectively recognized a diverse range of activities, enhancing prediction accuracy.

Temporal Activity Impact:

- Activities in the second half of the day are more predictive of stress.
- Bi-LSTM model achieved 72.5% accuracy and 75.94% precision for stress prediction.

Conclusion

Bi-LSTM Model Performance: Achieved the highest accuracy in both Human Activity Recognition (92.18%) and stress prediction (72.5%).

Effectiveness of Activity Data: Demonstrated that daily activity data is a valuable resource for accurately predicting stress levels through machine learning techniques.

Impact on Well-being: Early detection of stress allows for timely intervention, potentially improving overall mental and physical health.

Future Directions: Potential to enhance the model with real-time monitoring and diverse datasets, improving the robustness and applicability of stress prediction systems.

Combining activity data with contextual information could give a better overall understanding of what influences stress, leading to more accurate predictions.

Thank You

Any Questions?