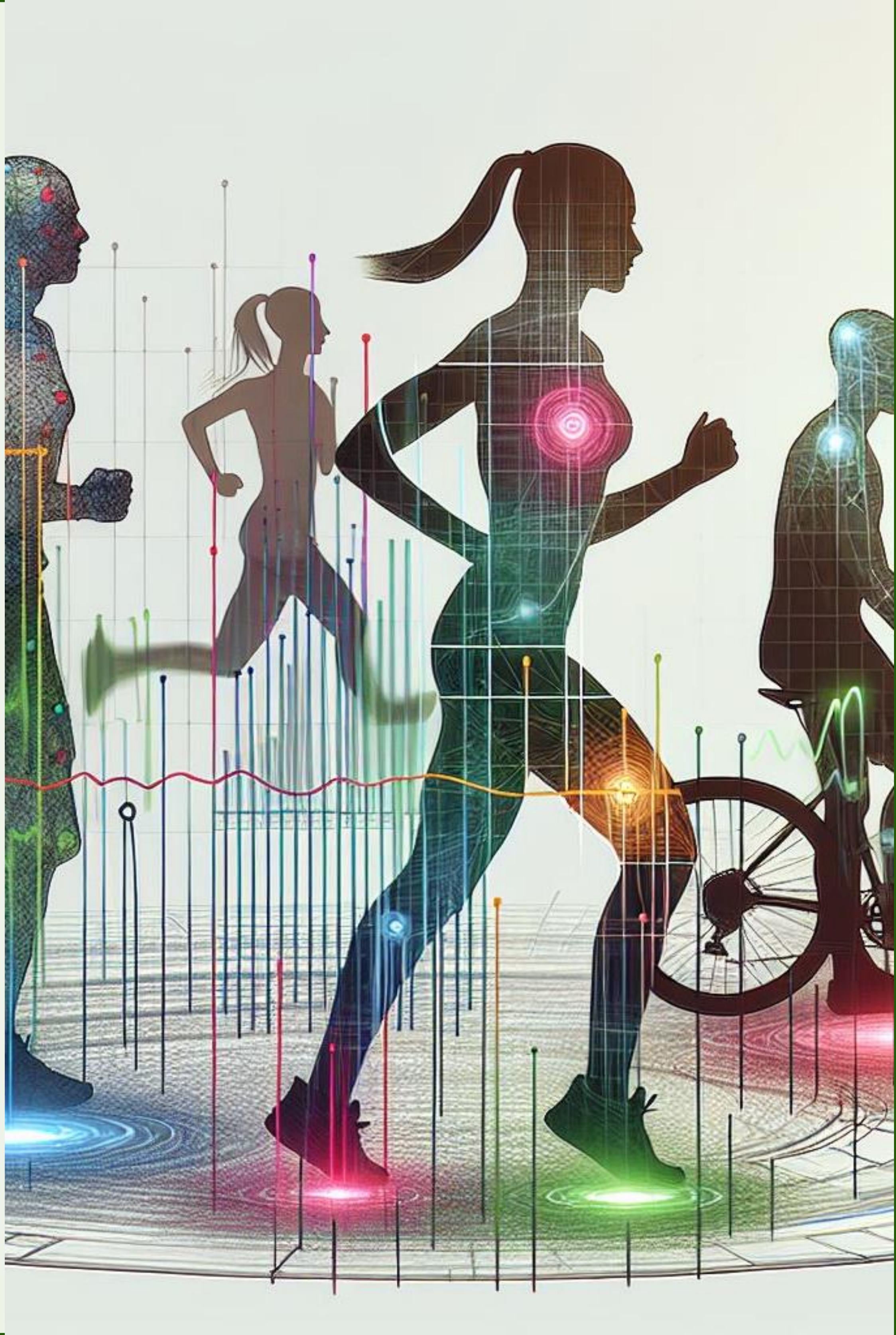
Human Activity Recognition Using Time-Series Wearable Sensor Data

AAI-510 Final Project

Sandeep Kumar Jakkaraju

Harish Kapettu Acharya

John Kalaiselvan



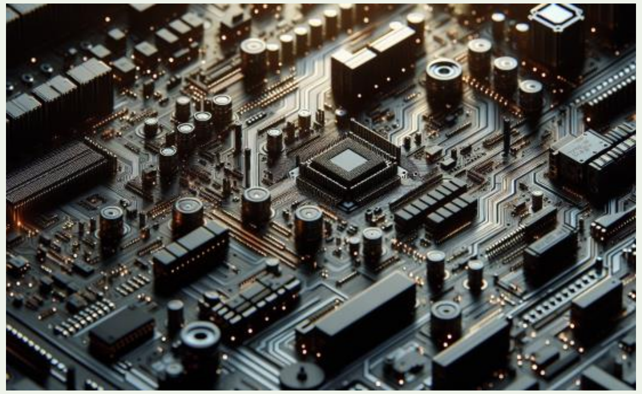
Business Understanding: Need for Human Activity Recognition (HAR)

Significance of HAR in Health Monitoring

Human Activity Recognition (HAR) enables automatic identification of physical activities using sensor data, crucial for remote healthcare by detecting falls, gait abnormalities, and breathing disorders, thereby enhancing patient safety and autonomy.

Necessity of Real-Time Activity Classification

Real-time HAR supports immediate feedback and intervention in applications like fall detection and emergency response, requiring efficient algorithms to process sensor data with minimal latency for timely preventive actions.



Key Challenges in HAR Implementation

HAR systems face challenges including data variability across users, sensor noise, data imbalance, and the need for scalable, computationally efficient models suitable for deployment on wearable and edge devices.

Data Understanding: Exploring the Dataset

1

Dataset Overview

The dataset includes sensor data from 9 subjects performing 18 physical activities, captured via three IMUs and a heart rate monitor for comprehensive human activity recognition research.

2

Dataset Structure and Features

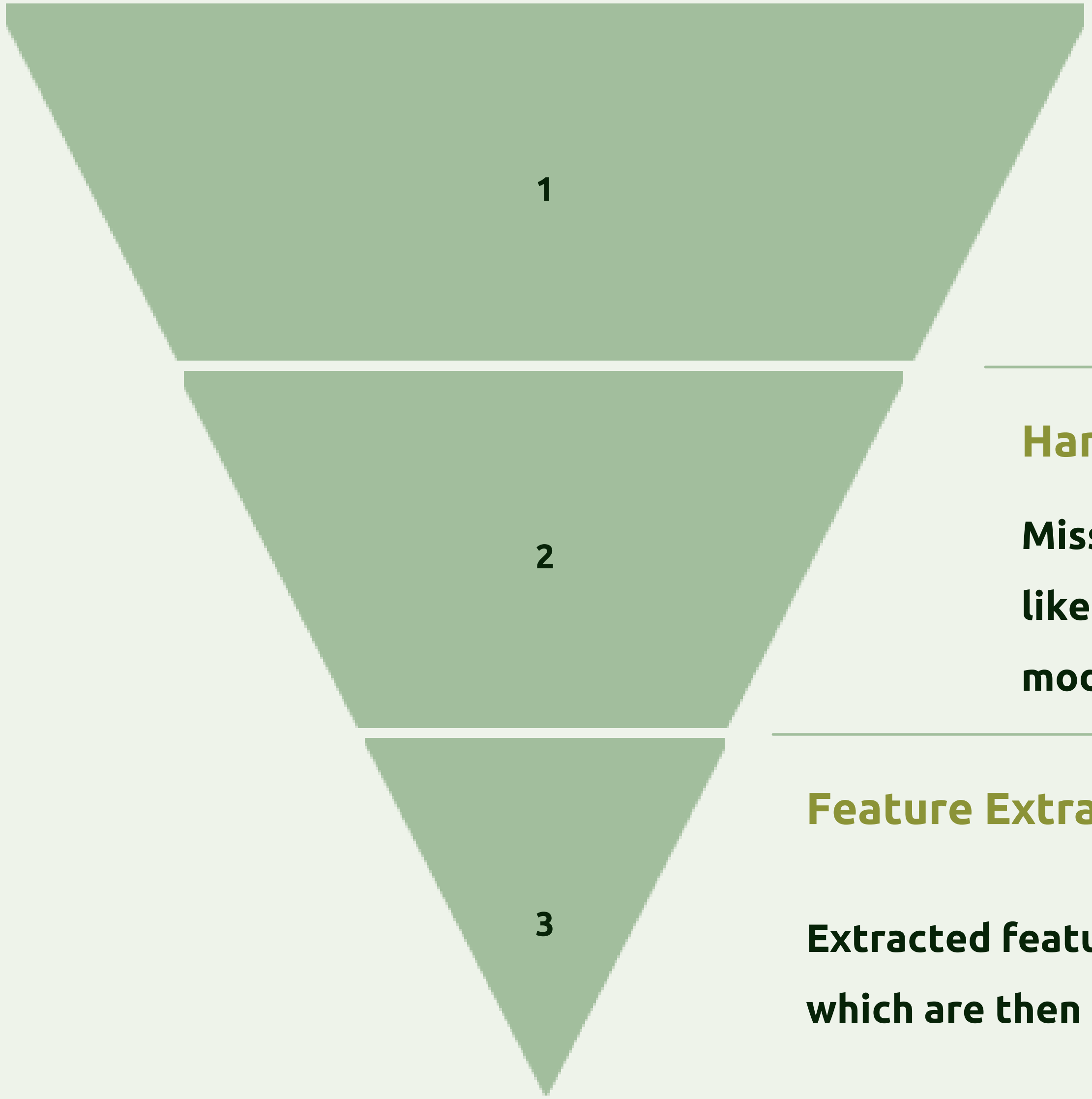
Each data file contains 54 columns including triaxial accelerometer, gyroscope, magnetometer data, and heart rate, enabling detailed multivariate time-series analysis of human movements.

3

Data Quality and Preprocessing

Preprocessing must address sensor noise, missing values, synchronization issues, and subject variability to ensure data integrity and improve model reliability in activity recognition.

Data Preparation: Cleaning and Organizing Data for Modeling



Segmenting Data into Time Windows

Continuous sensor data is divided into fixed-length time windows to capture activity patterns and facilitate feature extraction for machine learning.

Handling Missing Values

Missing sensor data is imputed using interpolation or statistical estimates like mean or median to maintain data integrity and ensure reliable modeling.

Feature Extraction and Normalization

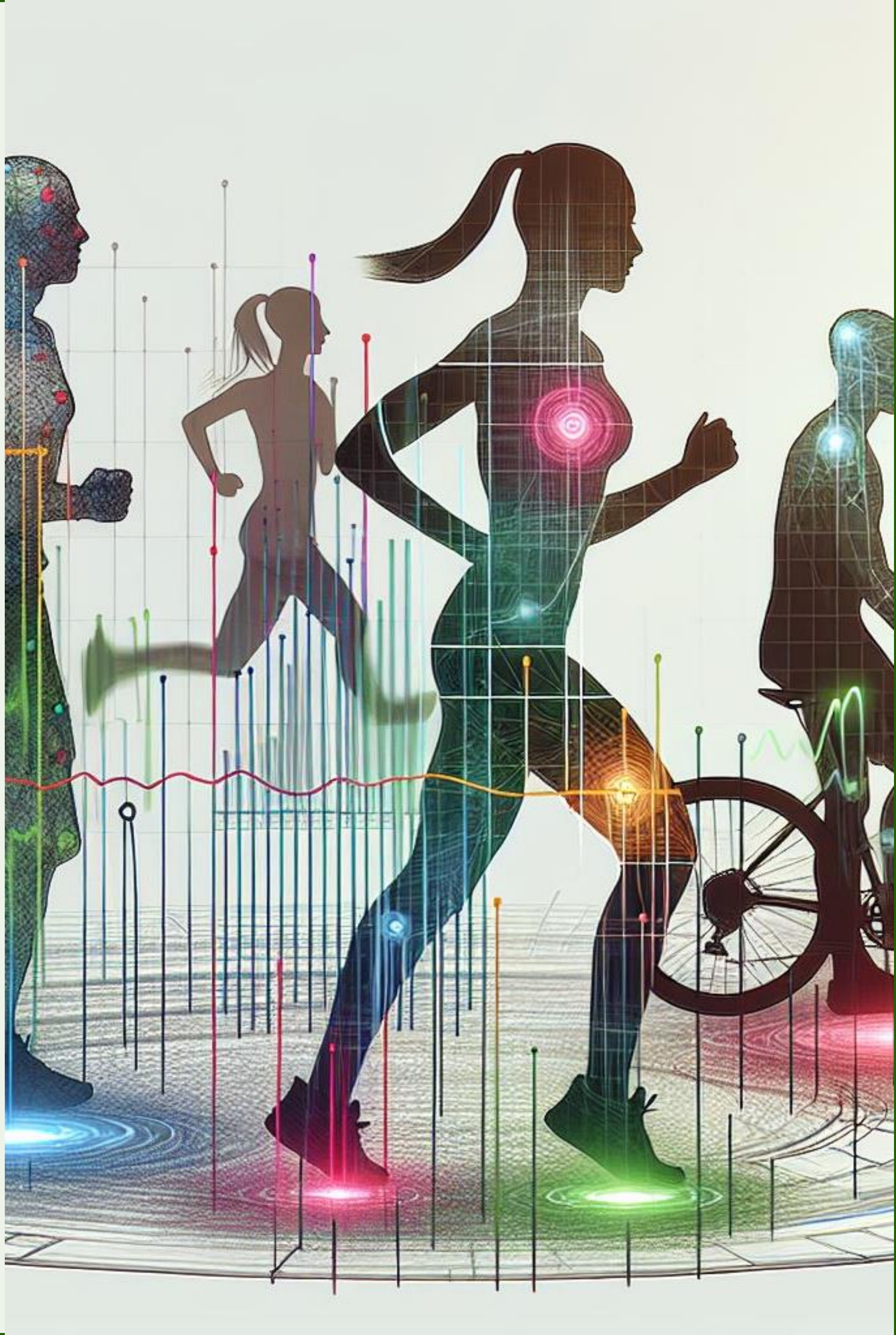
Extracted features include statistical, temporal, and frequency domain characteristics, which are then normalized to zero mean and unit variance to improve model stability.

Modeling: Applying Machine Learning Techniques for Activity Recognition

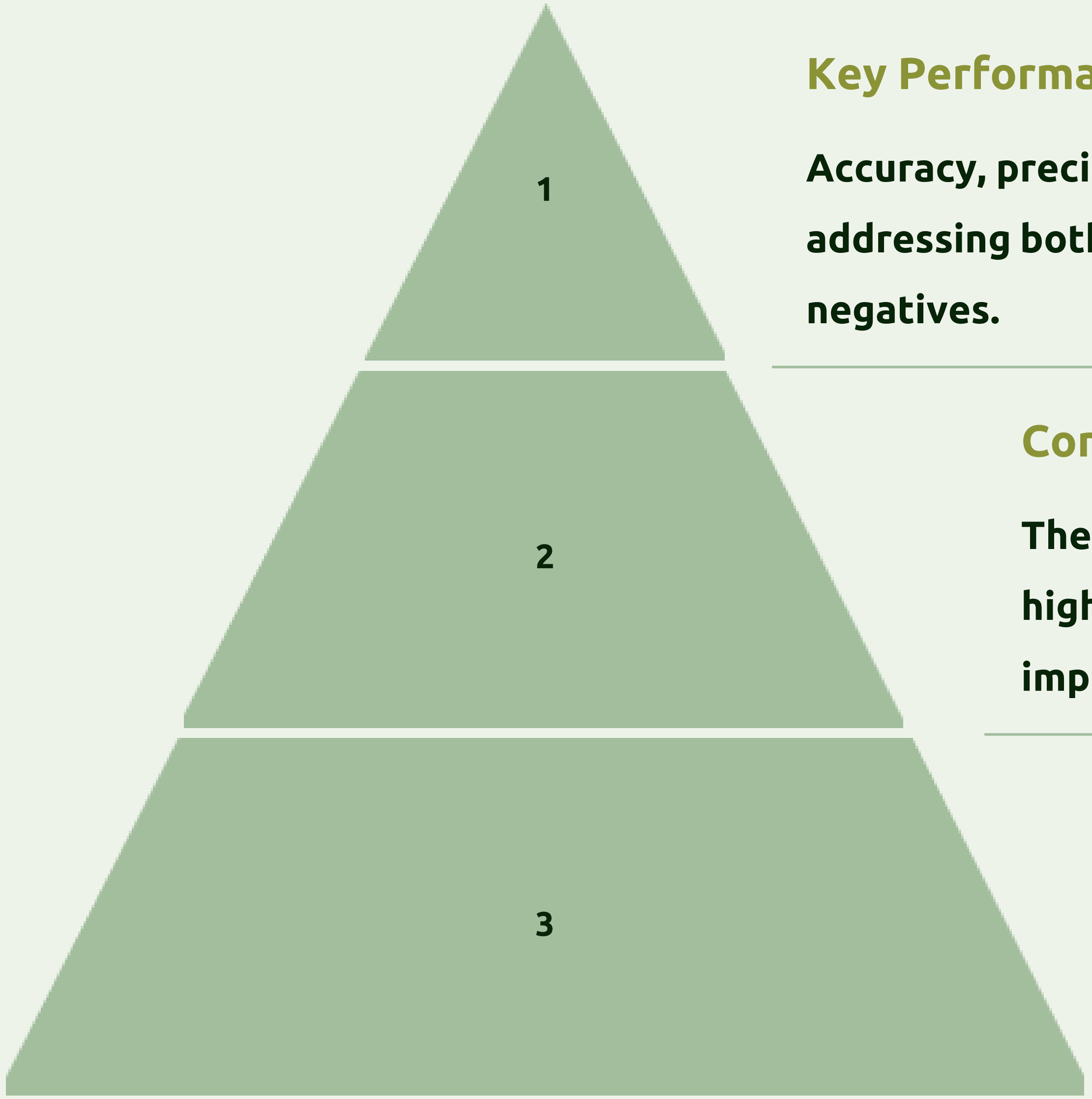
Classical Machine Learning Models : Decision Trees, Support Vector Machines, and Random Forests are widely used for human activity recognition due to their strong classification performance and interpretability using features from wearable sensors.

Deep Learning Architectures : Convolutional Neural Networks extract spatial features, while Recurrent Neural Networks, including LSTMs, capture temporal dependencies. Combining these, models like DeepConvLSTM improve accuracy by learning spatiotemporal patterns.

Rationale for Model Selection : Choosing models balances accuracy, complexity, and real-time needs. Deep learning with attention mechanisms offers superior performance on multimodal time-series data, while classical models provide efficiency and interpretability.



Evaluation: Assessing Model Performance and Robustness



Key Performance Metrics

Accuracy, precision, recall, and F1 score collectively evaluate classification quality, addressing both overall correctness and balance between false positives and false negatives.

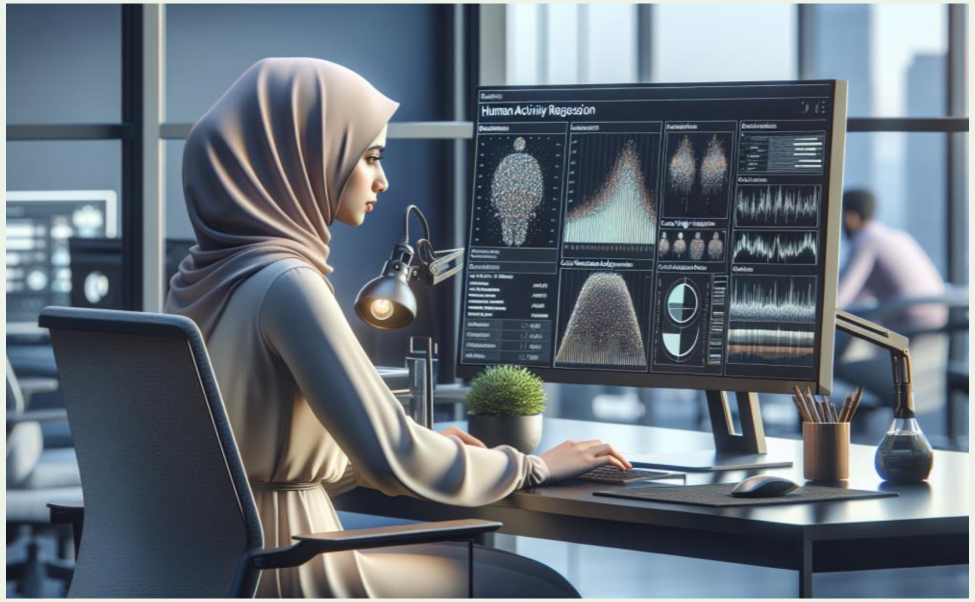
Confusion Matrix Insights

The confusion matrix visualizes prediction errors across classes, highlighting common misclassifications and guiding targeted model improvements.

Cross-Validation for Robustness

Cross-validation techniques ensure model generalizability by systematically testing performance on multiple data subsets, reducing overfitting risks.

Addressing Challenges in HAR



Managing Data Imbalance in HAR

Imbalanced datasets in HAR can bias model training, reducing accuracy for underrepresented activities. Techniques like data augmentation, over-sampling minority classes, and class-weight adjustments improve classification fairness across all activity types in the PAMAP2 dataset.

Reducing Sensor Noise for Accurate Recognition

Wearable sensor data often contain noise due to device inaccuracies and environmental factors. Applying filtering methods such as complementary filters and smoothing algorithms, along with attention mechanisms in deep learning models, enhances signal quality and improves activity recognition reliability.

Deployment: Strategies for Integrating HAR Models into Production

1. Batch vs. Real-time Inference

Batch inference processes large data volumes periodically, optimizing resource use but with limited responsiveness. Real-time inference enables immediate processing of streaming data, essential for applications requiring low latency such as fall detection.

2. Hosting Options for HAR Models

Cloud deployment offers scalability and ease of management, suitable for large-scale processing. Edge deployment reduces latency by processing data near the source but faces resource constraints. Hybrid deployment balances latency, cost, and computational capacity.

3. Monitoring and Maintenance Strategies

Continuous monitoring detects data drift and model degradation. Automated retraining pipelines improve adaptability. Version control ensures traceability and rollback capabilities, supporting robust model lifecycle management.

4. Deployment Strategies and Best Practices

Shadow, canary, and A/B testing deployments validate new models without impacting users. Containerization and API exposure simplify scaling and integration, ensuring efficient and reliable production deployment.

Conclusion and Future Directions

01

Summary of HAR Solution

Human Activity Recognition using the dataset demonstrates the effectiveness of wearable sensor data combined with machine learning and deep learning models to accurately classify diverse physical activities.

02

Potential Enhancements

Future improvements focus on addressing data imbalance and sensor noise through advanced augmentation and filtering, as well as enabling efficient real-time processing for practical deployment.

03

Encouragement for Continued Innovation

Ongoing research should explore scalable feature extraction, attention-based architectures, and multimodal sensor fusion to enhance accuracy and adaptability in real-world HAR applications.

Thank You