5ARE0, Academic Year 2025 – 2026

Assignment 1

Human Activity Recognition for Healthy Lifestyle Monitoring

Early detection and monitoring of daily physical activities is critical for promoting healthy lifestyles and supporting wellness interventions, particularly among elderly populations. Traditional methods rely on subjective self-reporting or expensive specialized equipment, but emerging wearable sensor technologies offer the potential for objective, continuous monitoring of human activity patterns.

In this hands-on assignment, you will explore the intersection of biomedical engineering and digital health by building an automated classification system for common daily activities. You'll capture real sensor data using smartphones and develop models that can distinguish between fundamental human movements and postures.

This project mirrors real-world challenges in digital health and wearable technology: working with noisy sensor data, extracting meaningful movement features, and creating robust classification algorithms that could eventually support activity tracking and health monitoring applications. By the end of this assignment, you'll have hands-on experience with the complete data science pipeline—from raw sensor acquisition to model evaluation—while contributing to the growing field of human activity recognition.

📌 This group assignment constitutes **20%** of your final grade and should be submitted by **October 5th, 20:00**.

---

1. Background and Relevance

Human Activity Recognition (HAR) using wearable sensors has become a cornerstone technology in modern health monitoring systems. Commercial devices like smartwatches and fitness trackers rely on similar sensor-based classification systems to automatically detect and log daily activities, encouraging users to maintain active lifestyles. Understanding the underlying principles of activity detection provides valuable insights into how these technologies work and their limitations. This assignment focuses on five fundamental activities that represent the spectrum of human movement patterns:

1. **Sitting down (transition from standing position to sitting position)**
2. **Standing up (transition from sitting position to standing position)**
3. **Walking**
4. **Running**

5. **Climbing stairs**

---

2. Aim and Research Questions

Primary Objective:

To develop and evaluate machine learning models that can automatically classify common daily activities using smartphone sensor data, with the goal of creating tools that support healthy lifestyle monitoring.

Research Questions:

1. Which sensor-based features best discriminate between the five target activities (sitting down, standing up, walking, running, climbing stairs)?
2. How do supervised and unsupervised learning compare for this multi-class activity recognition problem?
3. Which machine learning model yields the highest classification accuracy when combined with appropriate feature extraction and selection techniques?
4. What are the practical limitations and considerations for deploying such a system in real-world health monitoring applications?

---

3. Protocol

3.1 Participants & Activity Selection

- **Groups of 3 students** collaborate on data collection and analysis
- Each group focuses on the **five core activities**: sitting down, standing up, walking, running, and climbing stairs
- Justify your experimental design choices and hyperparameters.

Inclusion Criteria:

- Students who are able to perform basic daily activities safely
- No mobility restrictions or injuries that would prevent safe participation

Exclusion Criteria:

- Any condition preventing safe performance of walking, running, or stair climbing
- Recent injuries affecting movement patterns

3.2 Measurements

Use the free **Sensor Logger** app (iOS/Android) configured to record:

- **Accelerometer** (captures linear acceleration in 3D space)
- **Gyroscope** (captures angular velocity and rotational movements)
- **Gravity sensor** (provides device orientation relative to Earth's gravity)

All other sensors **OFF**. Use **default sampling rates of 100 HZ**. Ensure **Standardisation** option is activated to ensure consistency among different operating systems.

> **Tip:** Secure the phone consistently—e.g., in a tight pocket, armband, or waistband—so that orientation is reproducible across trials and participants. Document your placement strategy as it will affect data interpretation.

3.3 Procedures

Pre-Data Collection:

1. **Safety Assessment:** Ensure all participants can safely perform each activity
2. **Environment Setup:** Use consistent environments
3. **Phone Placement:** Standardize device placement

Data Collection Protocol:

For each of the five target activities:

1. **Setup Recording**

- Launch Sensor Logger with specified sensor configuration
- Attach phone to designated body location consistently

1. **Activity Performance**

- **Sitting Down:** Sit down slowly on a standard chair with normal posture without additional movements.
- **Standing up:** Stand up slowly with normal posture.
- **Walking:** Walk at a comfortable, self-selected pace in a straight line.
- **Running:** Jog/run at a comfortable pace in a straight line. seconds
- **Climbing stairs:** Ascend one flight of stairs.

1. **Rest Periods**

- Allow 2-3 minutes rest between physcially demanding activities to prevent fatigue effects

1. **Repetitions**

- Repeat each activity. Think about how many repetitions you need and motivate your choices.

- Each group collects data from **all participants**.

📁 3.4 Data Management

A clear and consistent approach to data management ensures reproducibility and reliability. Students should systematically document their activities, maintain uniform sensor setups and trial conditions, and have all group members contribute data. The team should consolidate the datasets with clear labeling and timestamps. Raw and processed files should be stored separately, named consistently, and anonymized to protect privacy.

---

4. Analysis & Modeling

Your analysis must demonstrate a complete data science pipeline with the following components:

4.1 Data Preprocessing

Make use of the techniques learned during the course and instruction sessions. Remember that features are concatenated horizontally while samples/datapoints are concatenated vertically.

Tip: visualizing your features helps in guiding your preprocessing.

4.2 Feature Engineering & Extraction

Consider windowing strategy for time series-analysis. Transform raw sensor signals into meaningful representations using techniques covered in the course. Apply feature extraction and feature selection techniques (filter + wrapper methods) to identify the most discriminative features.

4.3 Supervised Learning Approach

Train logistic regression, decision tree, Naive Bayes and K-nearest neighor.

4.4 Unsupervised Learning Approach

Train (hard) k-means, fuzzy C-means and Gaussian mixture models. Make use of PCA for visualization.

4.5 Model Comparison & Analysis

Supervised vs. Unsupervised Comparison:

- Compare the effectiveness of both approaches for activity recognition.
- Analyze advantages and limitations of each method.
- Discuss scenarios where each approach might be preferred.

- Identify one model as your optimal model of choice.
- Identify optimal features from implemented feature selection techniques.

Model Justification:
Provide reasoning grounded in:

- **Problem Suitability:** How well does the model match the characteristics of activity recognition?
- **Data Requirements:** Is the model appropriate for your dataset size and feature complexity?
- **Interpretability:** Can the results be understood and trusted by end users?
- **Computational Efficiency:** Is the model practical for real-time or mobile deployment?
- **Performance:** How does accuracy compare across different models?

Evaluation Metrics:
Evaluate all models using:

- **Accuracy**
- **Cohen's kappa**
- **F1-Score**

Include tables and visualizations (plots, confusion matrices, feature selection plots) to support your findings and make results as clear as possible.

---

5. Reporting

Single ZIP Submission:
Create one ZIP archive named `groupX_submission.zip` containing:

1. **Jupyter Notebook** (`groupX_activity_recognition.ipynb`)
2. **Data collection and analysis report** (`groupX_Report.pdf`)
3. **Data Folder** (`data/`) with organized raw sensor files
4. **README file** (`README.md`) with setup instructions and dataset description

Notebook Requirements:

- **Complete Pipeline:** The notebook must execute end-to-end, including:

1. **Data Loading**
2. **Preprocessing**
3. **Supervised Learning**
4. **Unsupervised Learning**

- **Reproducibility:** Include Python version and required packages (with versions) at the top.

- **Code Quality:** Use clear structure, meaningful variable names, adequate comments, and section organization
- **Documentation:** Provide clear explanations of methodology and rationale for each step
- **Do not use other python packages besides the ones used during the instruction sessions.**

Write your report based on CRISP-DM Framework:
Conclude your notebook with a comprehensive **4-5 page** analysis covering:

Business/Objective Understanding

- Problem definition and relevance to health monitoring

Data Understanding

- Dataset characteristics, collection methodology
- Exploratory data analysis and key insights
- Data quality assessment and limitations

Data Preparation

- Preprocessing pipeline and rationale.
- Feature engineering/extraction/selection strategy.

Modeling

- Unsupervised and supervised approaches implemented
- Model selection rationale and hyperparameter tuning
- Comparison methodology between approaches

Evaluation

- Performance metrics and validation strategy
- Model comparison results
- Analysis of classification errors and failure modes

Deployment Considerations & Recommendations

- Practical implications for real-world implementation
- Limitations and potential improvements
- Recommendations for future work and clinical applications

Submission Guidelines:

- Upload **only** the `groupX_HAR_submission.zip` file to Canvas
- Ensure all group member names appear on the notebook title page
- **Reproducibility Check:** Instructors will run your notebook to validate results. Any failure in code execution or missing dependencies may impact your grade

6. Ethical and Privacy Considerations

- All activities performed by **consenting adult participants** with full understanding of the research purpose
- Ensure **participant safety** at all times—stop any trial if discomfort or injury risk arises
- **Data confidentiality:** All sensor data remains anonymous and is used exclusively for this educational assignment
- **Informed participation:** All group members understand data collection procedures and analysis goals
- **Data retention:** Raw data will be deleted after course completion and final grade assignment