# STAT 4051 Project Report

Kevin Tran, Gavin Bulthuis, Ivan Fierros-Pliego
May 2025

**Abstract**

This report analyzes motion sensor data from the UCI Machine Learning Repository to cluster 19 human activities using unsupervised learning techniques. We compare K-Means, Spectral Clustering, and Hierarchical Clustering with validation metrics including Silhouette scores and Dunn indices. Our results demonstrate that hierarchical clustering best captures natural groupings of activities like sedentary behaviors, walking patterns, and high-intensity movements. The methodology and findings provide insights for applications in fitness tracking and rehabilitation.

# 1 Introduction

## 1.1 Problem Statement

Human activity recognition using motion sensors has become critical for applications ranging from healthcare monitoring to athletic performance analysis. This project addresses the challenge of automatically classifying 19 distinct physical activities (e.g., walking, cycling, jumping) from wearable sensor data. Traditional supervised approaches require labeled training data, but unsupervised clustering can reveal natural movement patterns without predefined categories. Our work evaluates whether sensor data alone can group activities by biomechanical similarity, which could enable more adaptive fitness trackers and rehabilitation tools.

Key challenges include:
- High-dimensional data (45 sensors × 125 readings/activity)
- Similar sensor signatures across activities (e.g., walking vs. running)
- Variability in individual movement patterns

## 1.2 Dataset

We analyze the UCI Daily and Sports Activities Dataset collected from 8 subjects (4 female, 4 male, aged 20-30) performing 19 activities for 5 minutes each. Key characteristics:

| **Attribute** | **Details** |
|---|---|
| Activities | 19 (e.g., sitting, rowing, basketball) |
| Subjects | 8 adults |

| | |
|---|---|
| Segments | 60 per activity (5-sec windows at 25 Hz) |
| Sensor Units | 5 body locations (torso, arms, legs) |
| Sensors/Unit | 9 (accelerometer, gyroscope, magnetometer) |

The dataset's unconstrained movement recording (subjects performed activities "naturally") makes it ideal for testing real-world generalizability. Figure 1 shows sample sensor signals from three torso accelerometers during sitting.


*Figure 1: Raw Sensor Data Plot*

# 2 Methodology

## 2.1 Preprocessing

The raw motion sensor data was collected from 19 distinct human activities, each stored in a separate folder. Each activity folder contained multiple .txt files with sensor readings from different body units: Torso, RightArm, LeftArm, RightLeg, and LeftLeg. These sensor streams were labeled with suffixes _S1 through _S9 to represent different axes and metrics. We used a custom function to read and combine all files associated with a given activity into a single dataframe using read_csv().

We extracted raw and engineered features to prepare the data for modeling. First, we scaled the raw sensor data and applied Principal Component Analysis (PCA) to reduce the dimensionality, preserving the top 20 principal components for each activity. This reduced redundancy and helped focus on the most significant patterns in the sensor signals, while maintaining approximately 80% of the cumulative variance.

The raw data didn’t hold many distinguishable features that allowed for true cluster separation. To solve this, we then performed feature engineering to compute statistical summaries from each sensor unit (mean, standard deviation, min, max, first quartile, third quartile, and median). These features were derived using the feature_extraction() function and attached to each observation. The resulting final feature vector for each activity included both PCA-reduced raw data and engineered features, providing a comprehensive representation of motion dynamics. Although this added more features, the usage of PCA was still needed to reduce 25 features from the raw data. This method supported better clustering than applying PCA to both the raw data and the engineered features together.

## 2.2 Clustering Methods

We explored three clustering algorithms to group similar motion patterns based on the compiled feature vectors:

- **K-Means Clustering**: This method was run with k = 9 clusters, based on the number of major activity groups we wanted to discover. We initialized the algorithm with multiple random starts (nstart = 100) to ensure stability.

- **Spectral Clustering**: Implemented using the specc() function from the kernlab package, this approach used graph-based similarity to cluster the activities. We set the number of clusters to 9, consistent with the K-means experiment, and assigned cluster labels based on the spectral embedding.

- **Hierarchical Clustering**: Using the Ward.D2 method, we performed agglomerative clustering on the pairwise distances between activity feature vectors. The resulting dendrogram was cut at k = 9 clusters to match the other methods. Activity labels were manually attached to the resulting clusters for interpretability.

Each clustering method was applied to the same feature matrix, ensuring comparability.

## 2.3 Cluster Choices

For these clustering methods, the number of clusters has to be controlled by the user. We noticed immediately that the activities fall into three main branches of intensity.

| Branch | Exercises | Classifier |
|---|---|---|
| Low Intensity: Zone 0 | Lying on Back, Sitting, etc | <50% of max heart rate |
| Medium Intensity: Zone 1-2 | Walking in Parking Lot, Rowing, etc | 50-70% of max heart rate |
| High Intensity: Zone 3-5 | Playing Basketball, Running, etc | >70% of max heart rate |

From these main branches, we assigned a constant number of clusters that would group similar activities together that would fall into the three branches. The number arbitrarily chosen was 9, with the approximate cluster assignments below.

**Low Intensity:**

1. Stationary Movements:
    a. Lying on Side, Lying on Back, Standing, Sitting, Standing in Elevator

**Medium Intensity:**

2. Biking Movements:
    a. Stationary Bike in Horizontal/Vertical Position
3. Rowing Movements:
    a. Rowing
4. Flat Walking Movements:
    a. Walking on a Treadmill, Walking in a Parking Lot
5. Incline/Decline Movements:
    a. Ascending Stairs, Descending Stairs, Walking on Incline Treadmill
6. Other Machine-Based Movements:
    a. Exercising on an Elliptical, Moving in an Elevator, Walking on Stairmaster

**High Intensity:**

7. Playing Basketball
8. Running on a Treadmill
9. Jumping

The high-intensity branch could be combined into one cluster, but the movements are different enough and/or combinations of each other, and can be separated. For example, playing basketball requires both running and jumping movements, and strictly running and jumping aren't heavily correlated outside of the intensity level itself.

## 2.4 Validation Metrics

To assess the quality of our clustering results, we employed two internal validation metrics:

- **Silhouette Score**: This metric measures how well-separated the clusters are by comparing intra-cluster cohesion with inter-cluster separation.
- **Dunn Index**: The Dunn index evaluates compactness and separation of clusters by considering the ratio between the smallest inter-cluster distance and the largest intra-cluster distance.

# 3 Results

## 3.1 Cluster Performance

To evaluate the performance of our clusters, we utilized the average Silhouette Score and Dunn Indices across all of the subjects, as mentioned previously.

The table below summarizes these results:

| Method | Num Clusters | Silhouette Score | Dunn Index | Description |
|---|---|---|---|---|
| K-Means | 9 | 0.2057 | 0.6864 | Centroid-based partitioning |
| Spectral | 9 | 0.1132 | 0.2508 | Graph-based kernel method |
| Hierarchical | 9 | **0.2086** | **0.7064** | Dendrogram-based approach |

Based on both validation metrics, hierarchical clustering offered the best performance and was the most reliable for uncovering meaningful structure in the motion data. It is important to note that these metrics can fluctuate depending on the number of clusters chosen as well as the dis(similarity) or distance measures in respective methods.

## 3.2 Activity Groupings

Due to Hierarchical Clustering having the best performance, we evaluated and analyzed the results of this method by visualizing the dendograms for each of the eight subjects. This uncovered more information on how cluster assignments were determined and which activities have similar sensor readings (on a per-person basis).

Figure 2 shows the dendrogram for each of the eight subjects. Again, nine clusters were chosen, and the respective clusters and labels are color-coded.

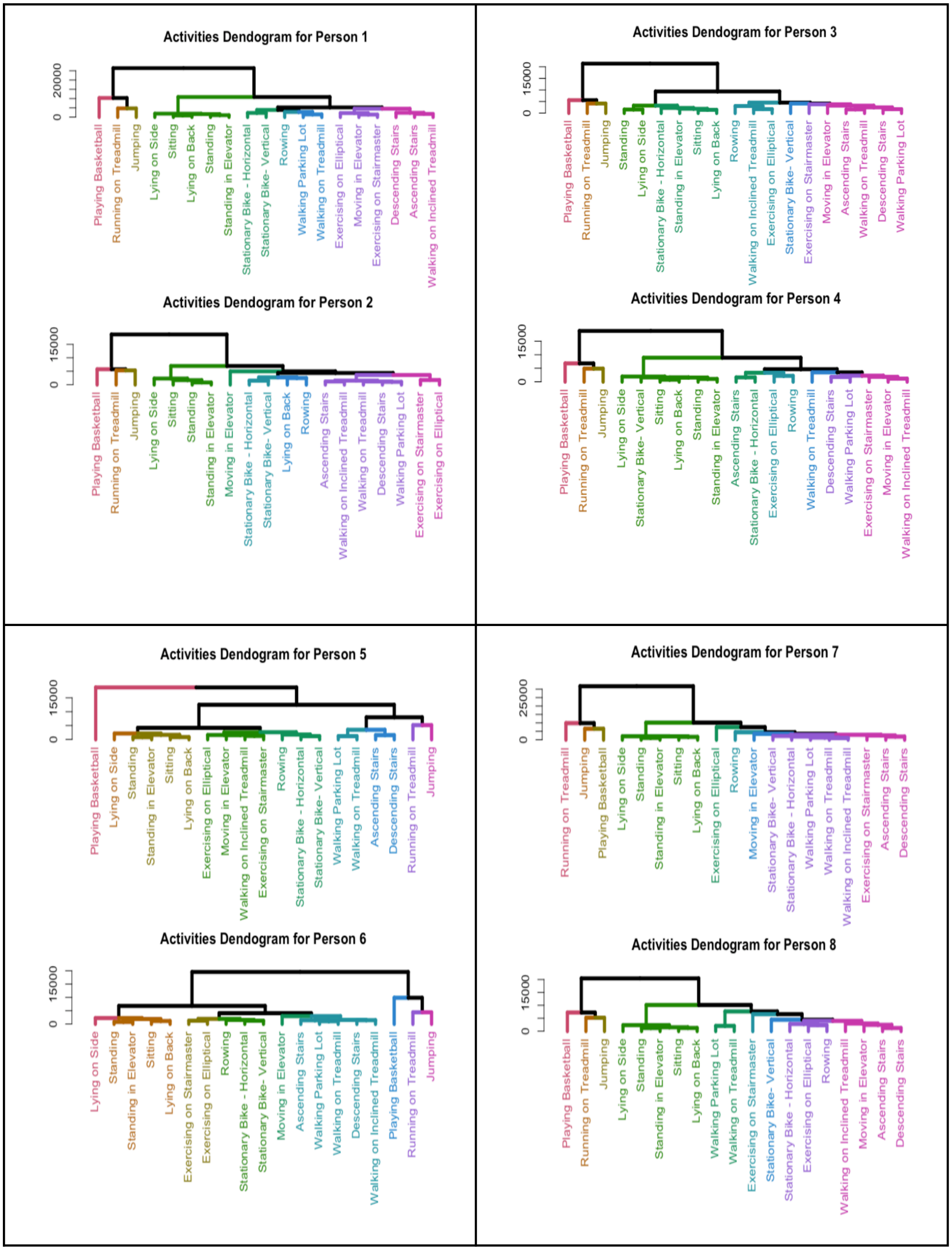


Figure 2: Activity Dendograms for each of the 8 Subjects with 9 Clusters

By visualizing the dendrograms, it can be seen that all eight subjects for the most part maintained those three main branches of activities in the high, medium, and low intensity movements. In a couple of cases, a low intensity movement ends up in the medium intensity branch and vice versa, but the high intensity branch remains strong throughout. In seven out of the eight instances, the three high intensity movements are all within the same branch and separate clusters, which remains 100% accurate with our pre-determined clustering.

To evaluate the accuracy of the clustering results, we compared the formed clusters against our expected groupings. A cluster was considered correct if it contained all the activities that were expected to be grouped, regardless of whether it also included additional, unrelated activities. However, if a cluster was missing even a single activity that should have been included, it was deemed incorrect. This strict criterion ensured that the completeness of the expected grouping was prioritized in our accuracy assessment. Ultimately, we were looking for the branch accuracy to be high and ideally accurate cluster assignments, although not as important.

The results of this analysis can be found in the table below:

| **Activity** | **Activity Level** | **Correct Branch?** | **Correct Cluster?** |
|---|---|---|---|
| Playing Basketball | High | 100% | 100% |
| Running on Treadmill | High | 87.50% | 100% |
| Jumping | High | 87.50% | 100% |
| Bike - Horizontal | Medium | 75% | 62.50% |
| Bike - Vertical | Medium | 75% | 62.50% |
| Moving in Elevator | Medium | 87.50% | 25% |
| Exercising on Stairmaster | Medium | 87.50% | 25% |
| Exercising on Elliptical | Medium | 87.50% | 25% |
| Rowing | Medium | 87.50% | 25% |
| Walking on Inclined Treadmill | Medium | 87.50% | 50% |
| Ascending Stairs | Medium | 100% | 50% |
| Descending Stairs | Medium | 100% | 50% |
| Walking in Parking Lot | Medium | 100% | 87.50% |
| Walking on Treadmill | Medium | 100% | 87.50% |
| Lying on Side | Low | 100% | 62.50% |
| Lying on Back | Low | 87.50% | 75% |
| Standing | Low | 100% | 87.50% |
| Sitting | Low | 100% | 87.50% |
| Standing in Elevator | Low | 100% | 87.50% |

The table results show an average branch accuracy of 92% and an average cluster accuracy of 66%. The sensor data excelled in being able to identify whether the activities are of high, medium, or low intensity. This was an important step towards being able to use this sensor data to at least group activities at a large scale.

The cluster accuracies, however, were a bit lower due largely to all of the similar activities in the medium intensity grouping. The activities in this branch require a lot of the same actions to perform. For example, walking in a parking lot is essentially the same as walking on an incline treadmill with a slight difference in the walking surface and angle. It may be difficult to use only sensor data to truly detect subtle distinctions. This idea can also be applied to movements that are performed on machines versus normal conditions. For instance, distinguishing between climbing a physical staircase and using a stairmaster presents a challenge, as both generate highly similar sensor patterns. Those factors, among others, cause a lot of overlap and mixing of clusters in the medium intensity branches. This is where supervised deep learning methods may be a valuable resource to assist in making correct classifications on extremely similar activities.

Despite these challenges, the results still demonstrate that raw motion sensor data captures meaningful patterns. The clustering was generally successful at grouping similar activities, both at the macro level (large branches) and micro level (specific activity clusters), supporting the potential of unsupervised methods in activity recognition.

# 4 Discussion

## 4.1 Limitations

The results should be interpreted considering the following:

1. High Dimensionality: Despite reducing the dimensionality of the raw data using PCA, the dataset remained large after adding the summary statistic features. Regardless, though the dimensionality was going to be high. Processing the high-dimensional data requires notable computing power and memory. It is important to note that we had to run our code for each person one at time, as our computers couldn't handle the volume of data of all eight subjects at once.
2. Subject Variability: Each individual moves differently, leading to differences in how an activity is performed. This creates inconsistent clustering across subjects.
3. Cluster Interpretability: Clusters were not always interpretable across different clustering methods.
4. Cluster Choice: The selection of clusters and distance measurements for each person can alter the results between individuals.
5. Subject Knowledge: Having more information about the subjects would've been helpful in the analysis. Currently, all we knew was that either a male or female adult was performing the movements. Information such as knowing the exact gender, age, and activity level would have helped us uncover more insights.

### 4.2 Future Work

To extend the scope of this work, one area we'd explore would include enhancing cluster visualization by developing a visualization tool to inspect the cluster decision boundaries. This would help clarify why certain activities are clustered together. Moreover, it may be valuable to explore how variations in sensor type, quantity, and placement affect the quality and interpretability of the resulting clusters. This could potentially reduce data volume and improve usability in wearable devices. Lastly, we could attempt to build supervised learning models to classify movements being performed, which could lead to major improvements in fitness tracking and rehabilitation devices.

## 5 Conclusion

In this project, we successfully demonstrated the application of unsupervised learning techniques, such as K-Means, Spectral Clustering, and Hierarchical Clustering, to identify groupings of 19 physical activities using motion sensor data. Our analysis revealed that hierarchical clustering most effectively identified activity groupings, particularly for distinguishing between sedentary and high-intensity exercises.

The use of feature extraction and dimensionality reduction (PCA) was critical in addressing the high-dimensional sensor data, ensuring the clustering algorithms performed efficiently. Our evaluation using validation metrics, such as Silhouette scores and Dunn indices, confirmed that the hierarchical approach outperformed the alternatives for uncovering meaningful structure in the motion data.

However, our work also showed the challenges of clustering human activities, including subject variability and the difficulty of selecting or deciding an optimal number of clusters depending on the method used.

Looking ahead, implementing supervised learning techniques could enhance clustering accuracy. Future research may also explore clustering techniques that can generalize across different subjects or optimize sensor location for more informative clusters.

Overall, this research showcased how unsupervised learning can reveal patterns in human motion data, providing a foundation for applications in fitness tracking, rehabilitation, personalized health monitoring, general motion sensors, and more..

## Appendix

### Code Availability

R code available at: https://github.com/gavin-bulthuis/Activity-Clustering-using-Motion-Sensors

## Additional Visualizations

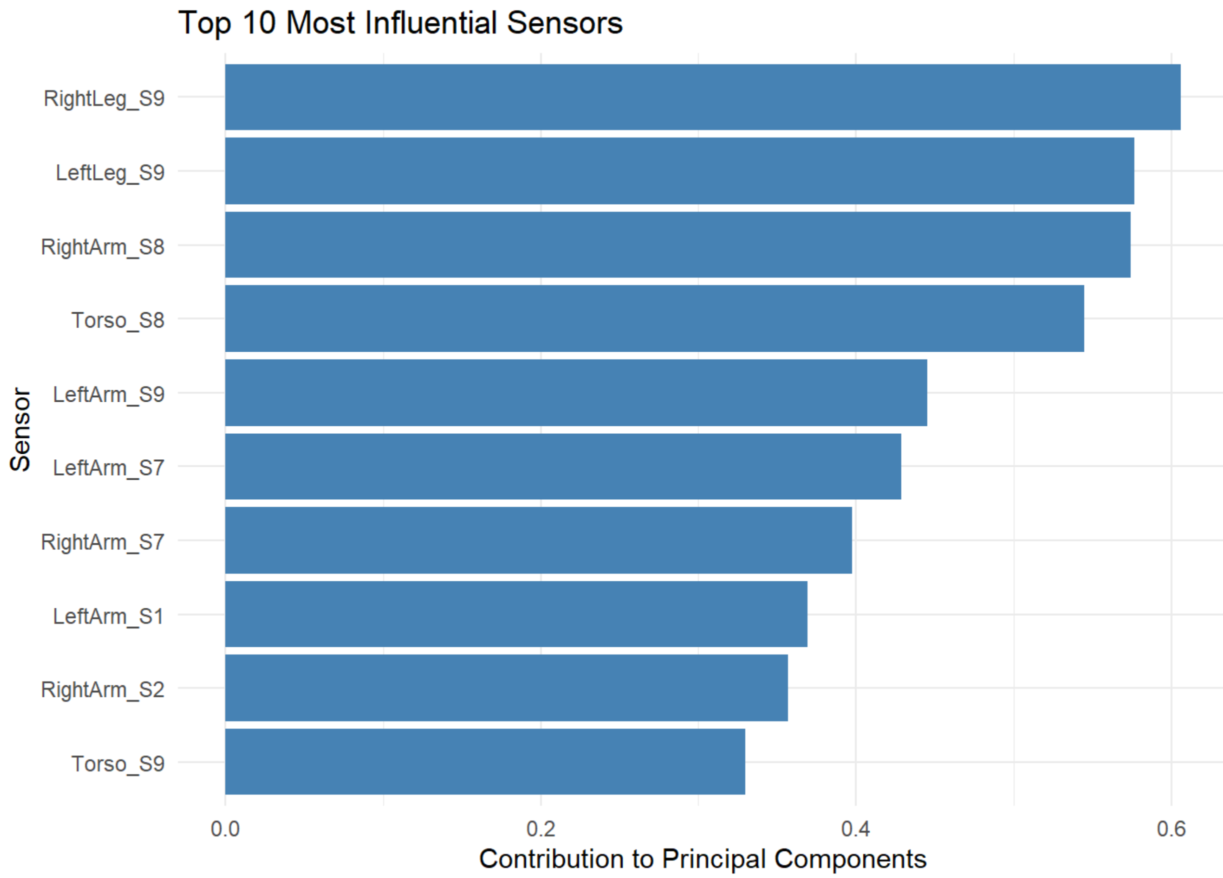


*Figure A1: Top 10 Most Influential Sensors for Person 1 After Applying PCA*