Exploring the Possibilities of The Most Effective Models for Exercise Tracking

Prepared by:
Matin Rahmani Neishabour
Mohamad Lebdeh
Mostafa Aboutaam
Arturo Rebolledo

Advisor:
Dr. Tim H. Lin

California State polytechnic University, Pomona

Abstract:
This paper explores the use of context-aware applications in strength training, focusing on using wristband accelerometers and gyroscopes to track free weight exercises. By analyzing data from 5 participants performing barbell exercises, the project aims to develop models that can track exercises, count repetitions, and detect improper form, similar to a personal trainer. The paper employs supervised machine learning algorithms, comparing their accuracies to identify the most effective models for exercise tracking.This paper explores the use of context-aware applications in strength training, specifically utilizing wristband accelerometers and gyroscopes to track free weight exercises. By analyzing data from five participants performing barbell exercises, the project aims to develop models that can track exercises, count repetitions, and detect improper form (similar to a personal trainer). The paper employs supervised machine learning algorithms, comparing their accuracies to identify the most effective models for exercise tracking.

Abstract:......1
Chapter 1: Introduction......3
Chapter 2: Approach...... 3
platform used:......3
Tasks and Contributions:......3
Data Collection (Arturo):...... 3
Experimental Setup (Mostafa):...... 3
Data Processing (Mohamad):......4
Model Selection and Training (Matin):...... 4
Coding Collaboration (All Participants):......4
Chapter 3: Progress and Achievements (Phase 1)......5
3.1 Data Collection...... 4
3.2 Combination of Raw Data......4
3.3 Overview of Dataset Composition...... 5
3.4 . Sensor Data Distribution......6
3.5. Activity Insights......6
3.6. Participant Behavior......7
3.7. Intensity Analysis......8
3.8. Conclusion:......8
Chapter 4: Refinement and Analysis (Phase 2)...... 11
4.1 Mark outliers Chauvenet......9
4.2 Low-pass Filter...... 9
4.3 Principal Component Analysis...... 10
4.4 Frequency Domain: Fourier Transformation......10
4.5 New dataset......10
4.6 Clustering......10
4.7 Predictive Modeling...... 12
4.8 Accuracy of the Models...... 13
4.9 Confusion Matrix......14
Chapter 5: Schedule......17
4820 (First Semester of 2023):...... 15
4830 (Second Continued Semester of 2023):...... 15
Chapter 6: Cost......18
Itemized Costs:...... 16
1. Equipment (Wristband Kit):......16
2. Gym Membership:......16
Chapter 7: Lessons learned......19
Chapter 8: Conclusions...... 20

Chapter 1: Introduction

Recent developments in wearable sensors, such as gyroscopes and accelerometers, have accelerated the growth of activity monitoring via smartwatches and related devices, especially in the areas of machine learning and pattern recognition.
By investigating the possibilities of wristband accelerometers and gyroscopes in monitoring strength training exercises, this project tracks exercises, counts repetitions, and identifies incorrect form in an effort to create models that mimic personal trainers. Data from five participants performing barbell exercises are analyzed. In order to identify the best models for tracking exercise, the research compares various machine learning algorithms using a supervised learning methodology.
A review of related work, an explanation of data collection and processing techniques, feature extraction from the data, model construction and evaluation, and a summary of major findings round out the paper's structure.

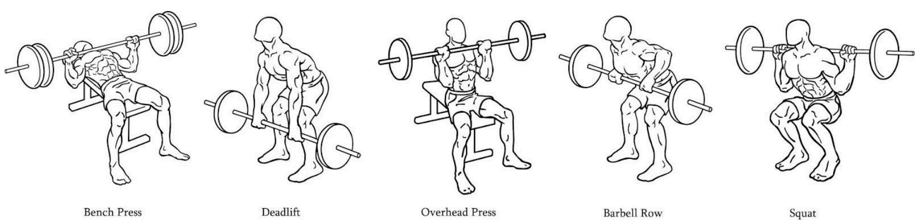


Fig. 1. Basic Barbell Exercise

Chapter 2: Approach

Group Organization and Task Assignments:

Our project depends on the efforts of Matin, Mostafa, Mohamad, and Arturo, each of whom makes a unique contribution to guarantee the project's overall success.

platform used:

- Google Colab

Tasks and Contributions:

Data Collection (Arturo):

- Ensuring the inclusion of barbell exercises performed by participants.
- Collaborated with Mostafa to integrate code functionality for automated recognition of exercise start and end points, optimizing data collection efficiency.

Experimental Setup (Mostafa):

- Designed and implemented Python scripts to automate the configuration of the experimental setup based on the Starting Strength program.
- Developed algorithms to synchronize the sensor data with the planned sets and repetitions, ensuring accurate labeling for subsequent analysis.

Data Processing (Mohamad):

- Led the implementation of Python scripts for preprocessing raw sensor data, including noise reduction, filtering.
- Implemented algorithms for the augmentation of the dataset, introducing variations in data points.

Model Selection and Training (Matin):

- Led the feature engineering phase, implementing Python code for Fourier transformation and k-means clustering to enrich the dataset.
- Explored and implemented machine learning algorithms, focusing on ensemble methods such as Random Forest.

Coding Collaboration (All Participants):

wrote, optimized, and reviewed Python code in a team environment while adhering to best coding practices and project-specific guidelines. Regular code review sessions were held to enhance the implemented algorithms' quality, efficacy, and maintainability.

Chapter 3: Progress and Achievements (Phase 1)

3.1 Data Collection

We collected data using MbientLab's wristband sensor research kit, which was selected due to its usefulness and similarity to a smartwatch, featuring a gyroscope and accelerometer. To test the model's generalization across varying exercise intensities, five participants—three sets of five repetitions for barbell exercises and three sets of ten repetitions for another—participated in the study. This method produced 150 sets of exercise data in total. To ensure that the dataset is comprehensive and representative of typical workout scenarios, data was also collected during "resting" periods, which included sitting, walking, and standing. This allowed for the analysis of transitions from rest to exercise.

3.2 Combination of Raw Data

The raw dataset included 69,677 entries, each of which included the wristband's x, y, and z sensor values along with an epoch timestamp. For coherence, the data—which had previously been stored in distinct files with distinct timestamps—was combined. The cleaned dataset made for more impactful visualizations. For instance, distinct exercise patterns and the ability to clearly identify repetitions from data peaks were shown in accelerometer data that was visualized for a heavy set of each exercise. The difference in y-acceleration between the medium and heavy weight squat sets was another thing to notice. The medium weight sets had deeper drops (because of the higher load causing faster downward motion) and higher peaks (because of lower resistance and faster upward motion)

3.3 Overview of Dataset Composition

The dataset comprises data from five unique participants (A, B, C, D, and E) who engaged in six different activities: Bench, Overhead Press (OHP), Squat, Deadlift (Dead), Row, and Rest. A total of 91 unique sets were recorded.

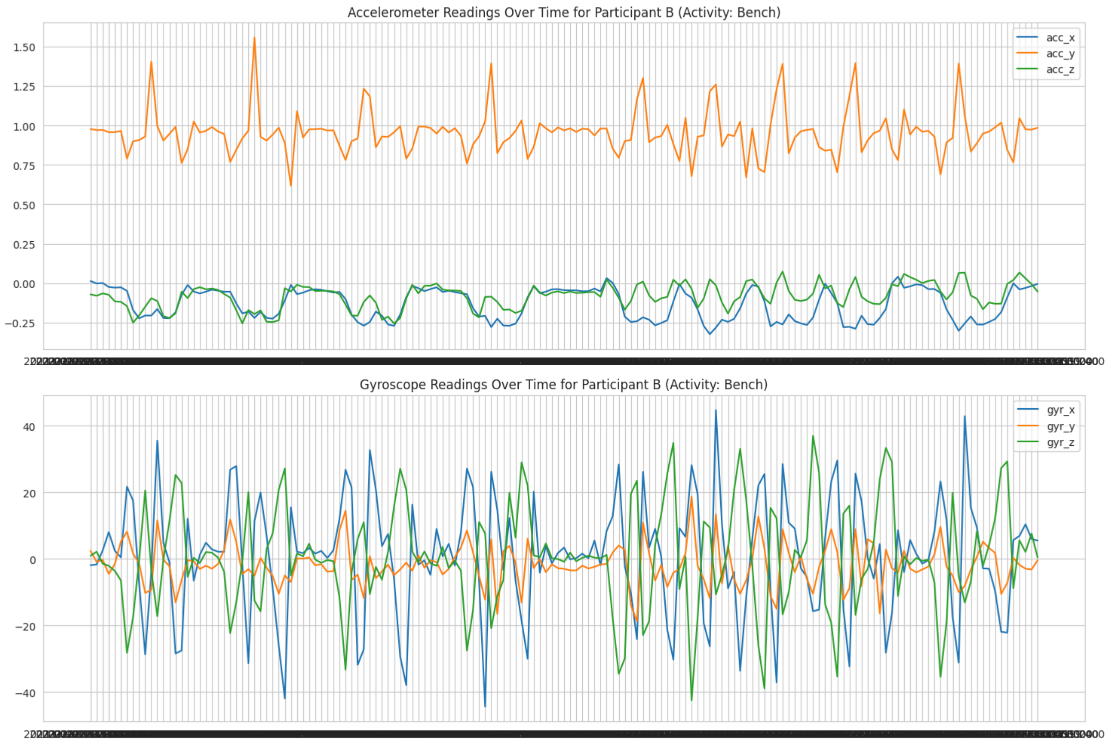


Fig. 2. Accelerometer and Gyroscope Data

3.4 . Sensor Data Distribution

The histograms provided a glimpse into the general behavior of the accelerometer and gyroscope readings:

Accelerometer:

- Readings along the X-axis are centered around zero, displaying a fairly symmetric distribution.
- The Y-axis readings are a bimodal distribution.
- The Z-axis readings are right-skewed, with a majority of the data points around zero.

Gyroscope:

- The readings across all three axes show distributions centered around zero, suggesting rotational stability for most of the data points.

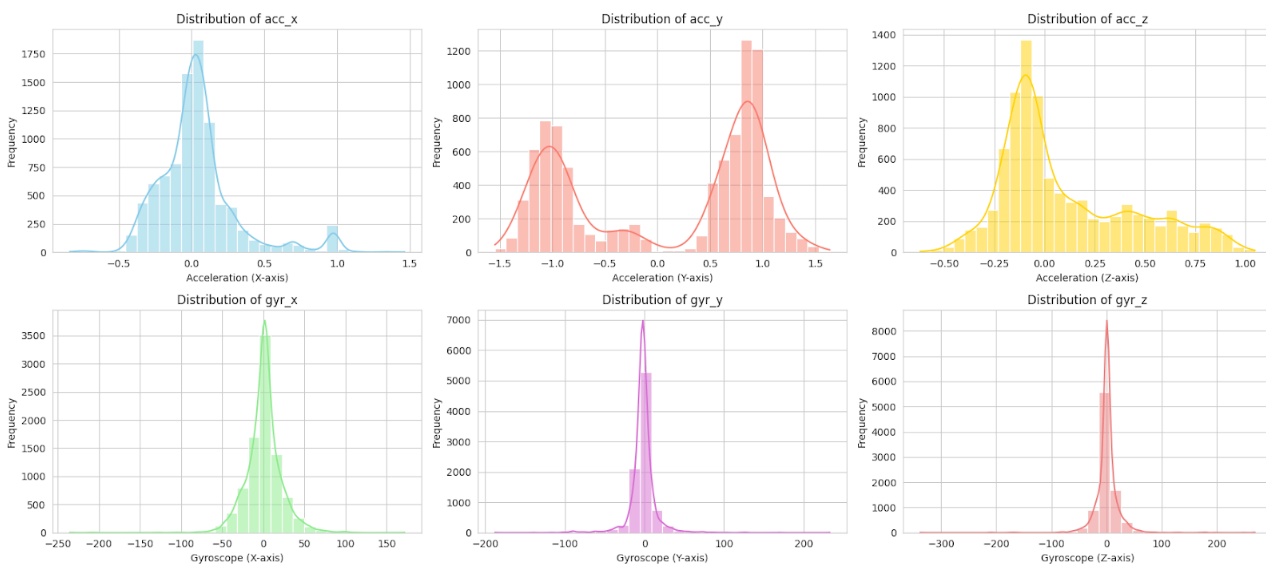


Fig. 3. Distribution of Accelerometer and Gyroscope Data Components

3.5. Activity Insights

Box plots were utilized to compare the spread and central tendency of accelerometer and gyroscope readings for each activity:

Activities like "rest" typically had compact ranges, indicating consistent measurements.

In contrast, activities like "squat" and "bench" displayed wider spreads and varied sensor readings, hinting at the diverse movements involved in these exercises.

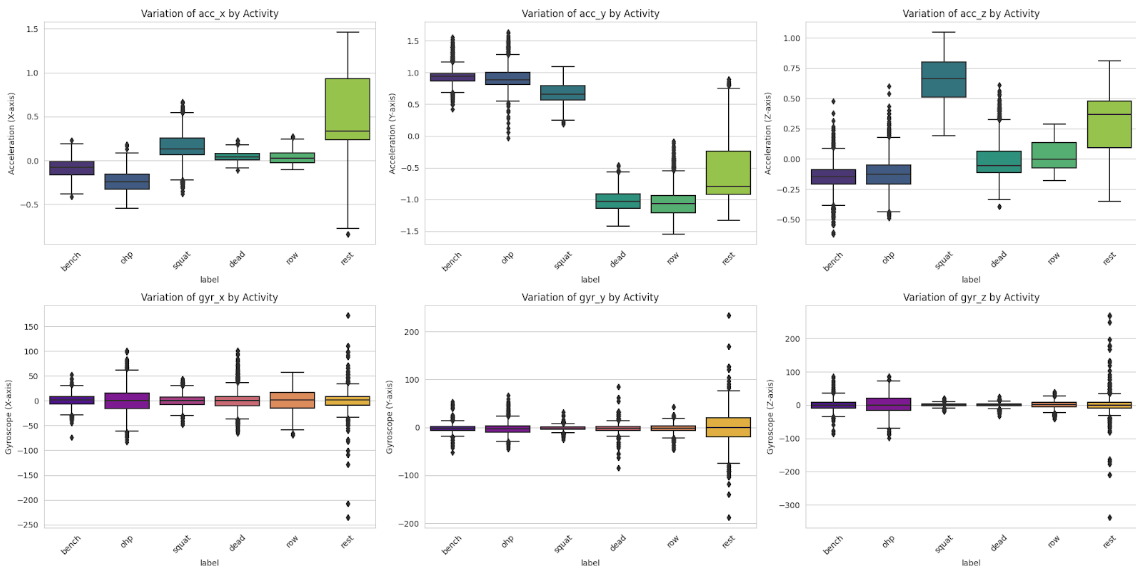


Fig. 4. Box Plots Comparing Accelerometer and Gyroscope Readings Across Activities

3.6. Participant Behavior

Each participant showcased unique movement patterns, especially evident when analyzing the "bench" activity:

Participants had distinct characteristics in their accelerometer readings, indicating different movement techniques or postures.

Variations in gyroscope readings emphasized the individualized rotational movements during the bench exercise.

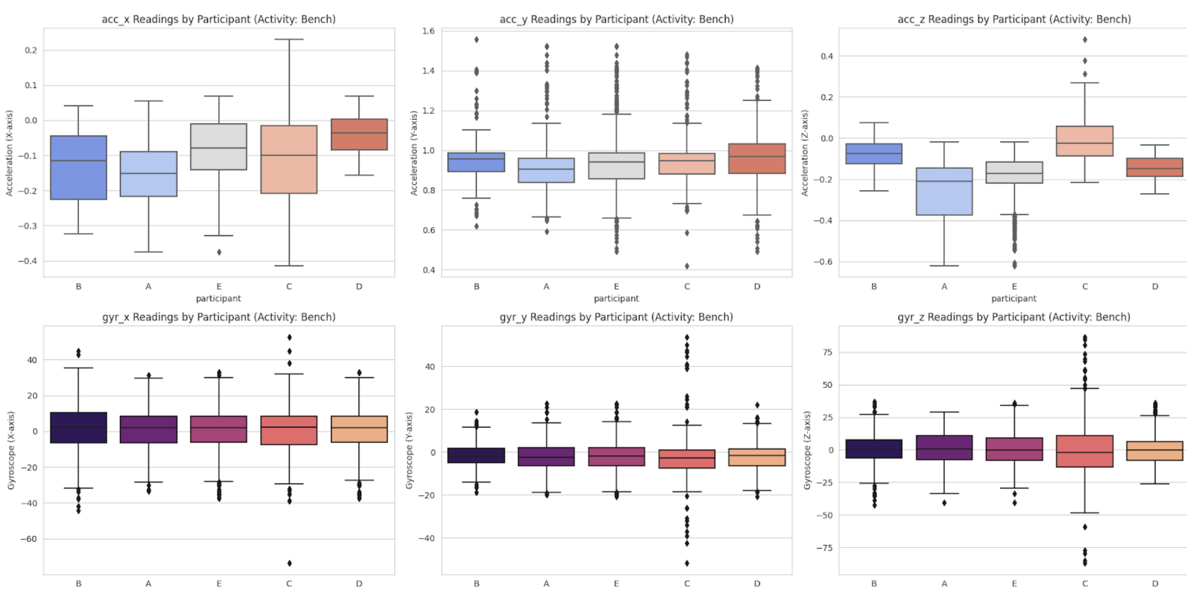


Fig. 5. Participant Behavior - Unique Movement Patterns in "Bench" Activity

3.7. Intensity Analysis

The exercise intensity, classified as "heavy" or "light," had a profound effect on the sensor readings:

"Heavy" exercises generally displayed a more compact range, suggesting controlled and consistent movements.
"Medium" intensity exercises showed wider spreads in readings, indicating varied movements.

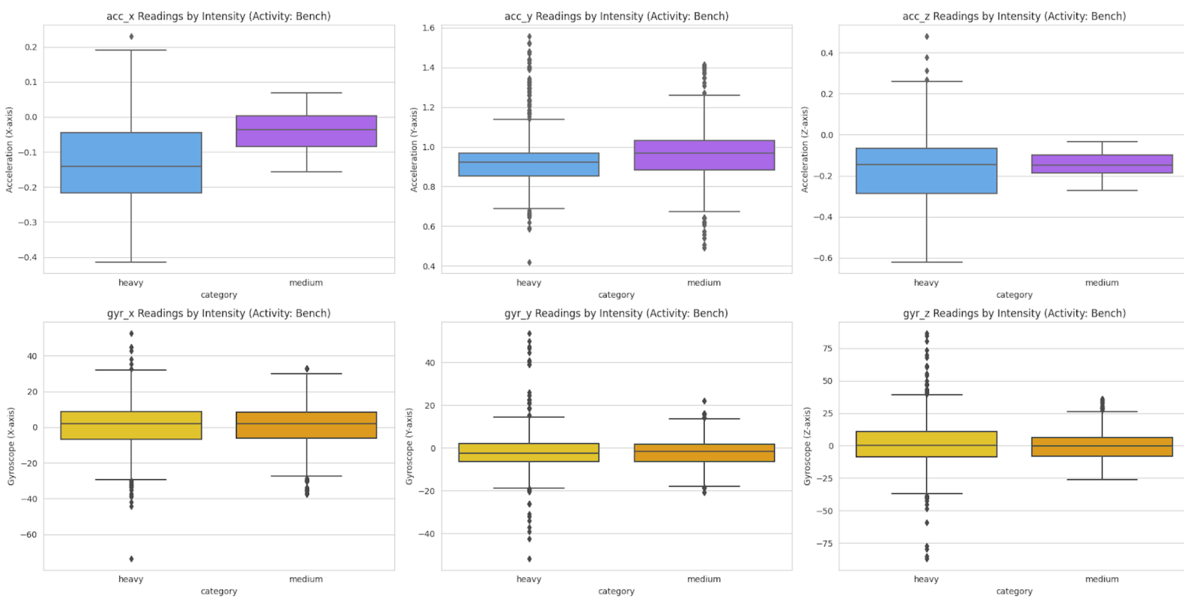


Fig. 6. Intensity Analysis of Sensor Readings

3.8. Conclusion:

Through this analysis, we've uncovered patterns in movement based on activities, participant behaviors, and exercise intensities. These insights can guide fitness trainers, athletes, or individuals in tailoring training regimes, making informed decisions, or detecting inconsistencies in workout techniques.

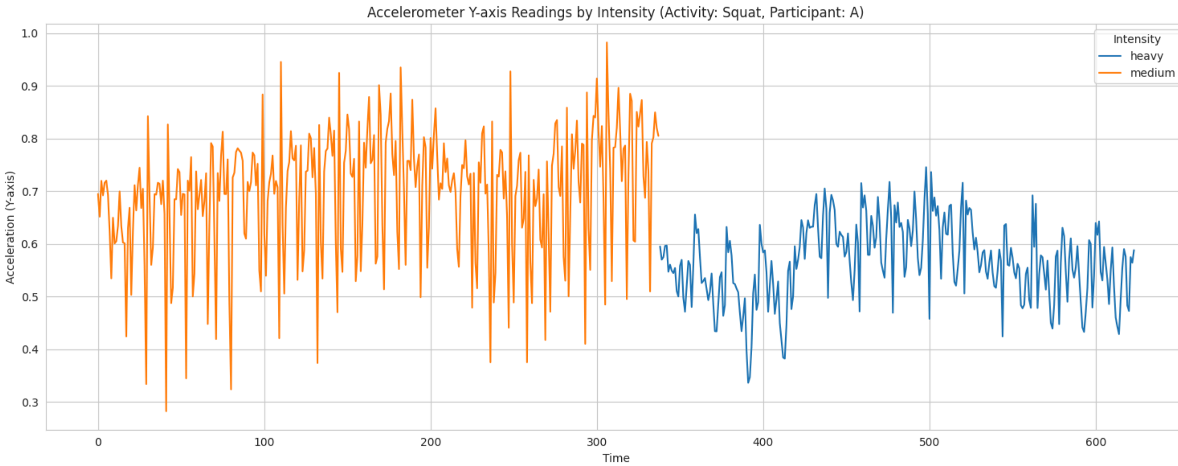


Fig. 7. Squat Activity - Accelerometer Y-axis Reading

Chapter 4: Refinement and Analysis (Phase 2)

Building upon the collected dataset, Phase 2 involved refining the data and conducting in-depth analyses to derive valuable insights. This phase includes several key steps:

4.1 Mark outliers Chauvenet

Finds outliers in the specified column of datatable and adds a binary column with the same name extended with '_outlier' that expresses the result per data point.

4.2 Low-pass Filter

The low-pass filter operates on the assumption that there is some sort of periodicity and can be applied to temporal data. Some of the high frequency noise in the dataset that might impede learning can be eliminated by the Butterworth low-pass filter. Everything was filtered out except the desired features. The fraction of the Nyquist frequency, or half of the sampling frequency, is how cutoff frequencies are expressed.

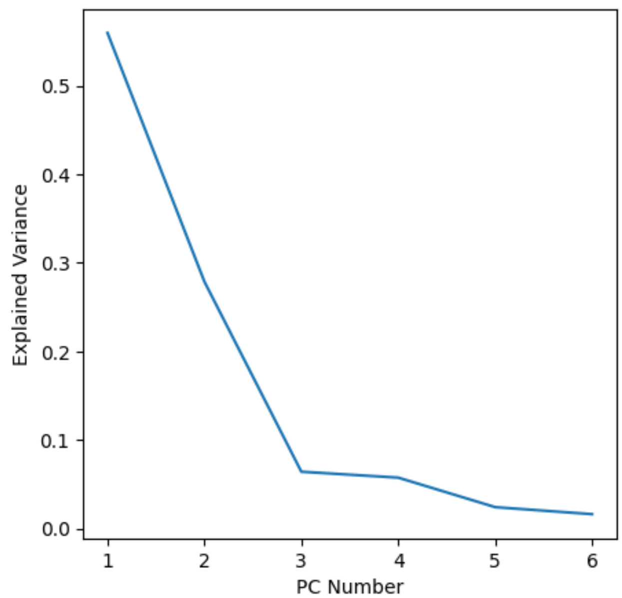


Fig. 8. Low-pass Filter and Principal Component Number

4.3 Principal Component Analysis

To identify the characteristics that could account for the majority of the variance, a principal component analysis (PCA) was performed. All features, with the exception of the target columns, underwent PCA. Figure 8 presents the results and illustrates how, after three components, the explained variance sharply drops. As a result, three components are chosen, and the dataset contains their values.

4.4 Frequency Domain: Fourier Transformation

Utilizing a fast fourier transform, determine the amplitudes of the various frequencies. In this case, the sampling rate (i.e., frequency, which is the dataset's Hertz) indicates the quantity of samples per second. Obtain frequencies within a specific window.

4.5 New dataset

Create new columns for the frequency data. Pass over the dataset (we cannot compute it when we do not have enough history) and compute the values.

4.6 Clustering

A label may be predicted in part by belonging to a particular cluster. The results indicated that the gyroscope data was not useful, so the focus will be on clustering the acceleration data. Following some experiments, the best approach was k-means clustering (k=5). Because they best represented the various labels, five clusters were selected.

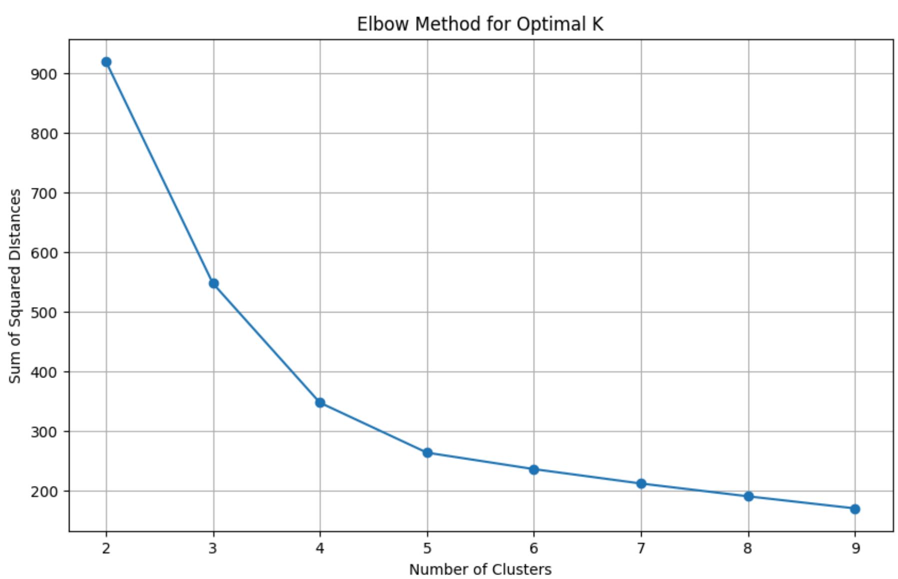
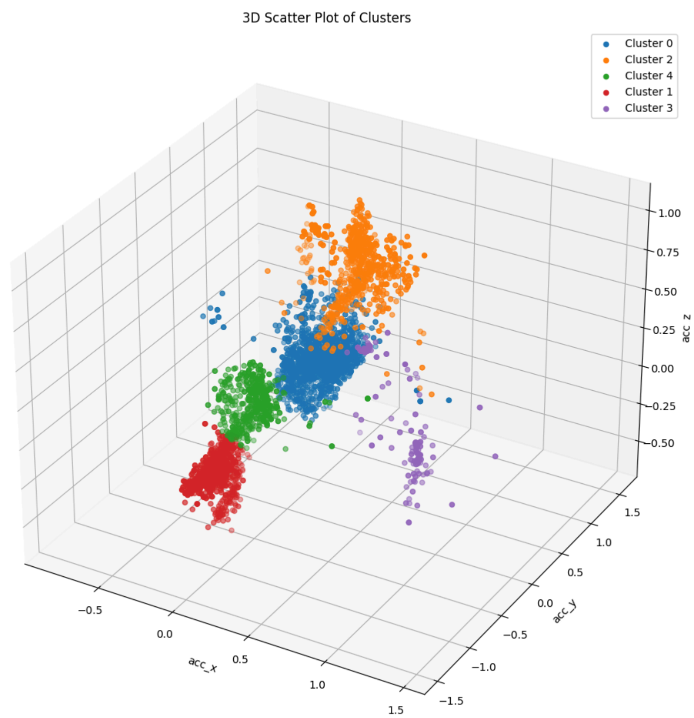


Fig. 9. KMeans

Fig. 10. Clusters



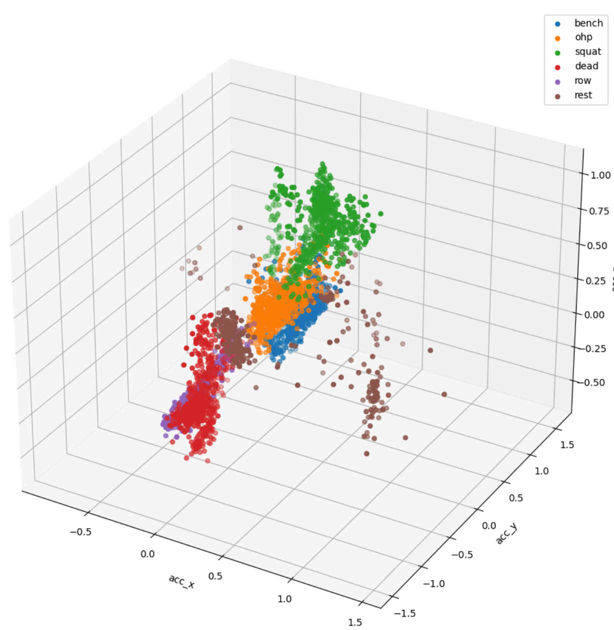


Fig. 11. Clusters labels

4.7 Predictive Modeling

Different sets of features were defined and used in our data analysis to capture different aspects of the sensor data. The raw gyroscope (gyr_x, gyr_y, and gyr_z) and accelerometer (acc_x, acc_y, and acc_z) data are included in the basic_features set. Square_features, which include squared values of accelerometer and gyroscope readings (acc_r, gyr_r) were added to provide additional dimensions and improve the ability to detect movements with varying magnitudes. Furthermore, pca_features (pca_1, pca_2, pca_3) were added to the data in order to highlight the principal components and lower dimensionality. Additionally, we took into account frequency-based (freq_features) and temporal (time_features) characteristics, which were taken from the dataset as features with the terms "temp," "freq," or "_pse," respectively. In order to capture possible clusters in the data, cluster_features was added in the end. These elements are gradually integrated into the feature sets, which begin with f_set1's fundamental features and progress to f_set4's comprehensive feature set for strong data analysis.

```python
basic_features = ['acc_x', 'acc_y', 'acc_z', 'gyr_x', 'gyr_y', 'gyr_z']
square_features = ['acc_r', 'gyr_r']
pca_features = ['pca_1', 'pca_2', 'pca_3']
time_features = [f for f in df.columns if '_temp_' in f]
freq_features = [f for f in df.columns if ('_freq_' in f) or ('_pse' in f)]
cluster_features = ['cluster']

f_set1 = list(set(basic_features))
f_set2 = list(set(basic_features + square_features + pca_features))
f_set3 = list(set(f_set2 + time_features))
f_set4 = list(set(f_set3 + freq_features + cluster_features))
```

Fig. 12. Code for Feature sets

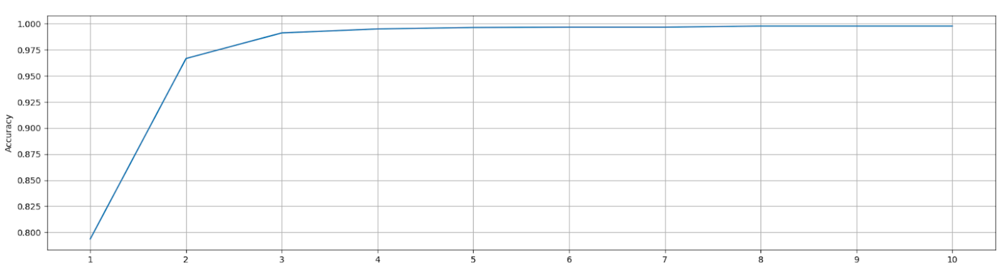


Fig. 13. number of Features

First, a preliminary test run was conducted to evaluate a range of models and features. Neural networks, Random Forests, K-nearest Neighbors, Decision Trees, and Naive Bayes were among the models used in this test.

4.8 Accuracy of the Models

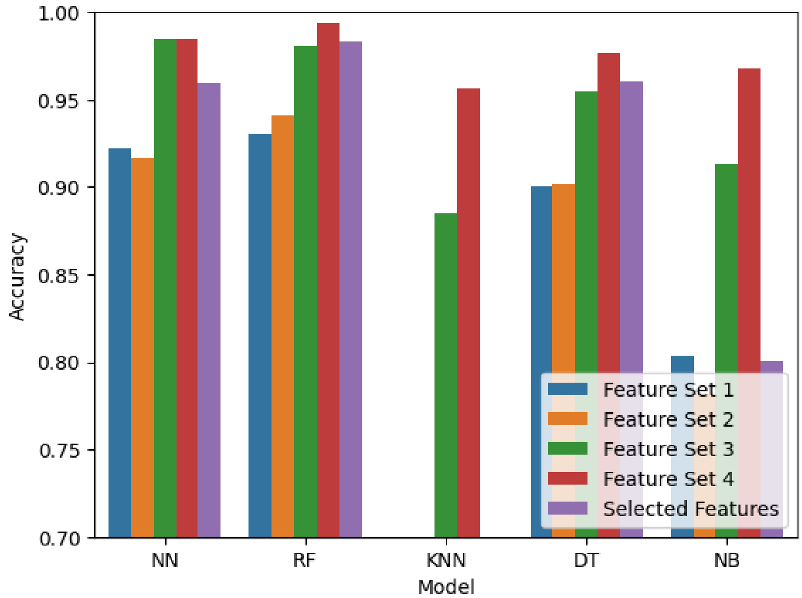


Fig. 14. Accuracy of the Models

```
acc = accuracy_score(y_test, class_test_y)
print(f'Accuracy of the model is {acc}')
```

```
Accuracy of the model is 0.9937888198757764
```

Fig. 15. Accuracy

4.9 Confusion Matrix

We used a confusion matrix, represented as a heatmap, to assess our classification model. Different classes were identified by comparing the test labels with the predicted labels in the matrix. The heatmap was improved with a color bar and blue color scheme, which made it easier to read the true labels on the y-axis and the predicted labels on the x-axis. To help with the rapid identification of the model's accurate and inaccurate classifications, we highlighted the matrix values using contrasting text colors based on a threshold calculation. The effectiveness of the model's performance in various categories was shown by this visual aid.

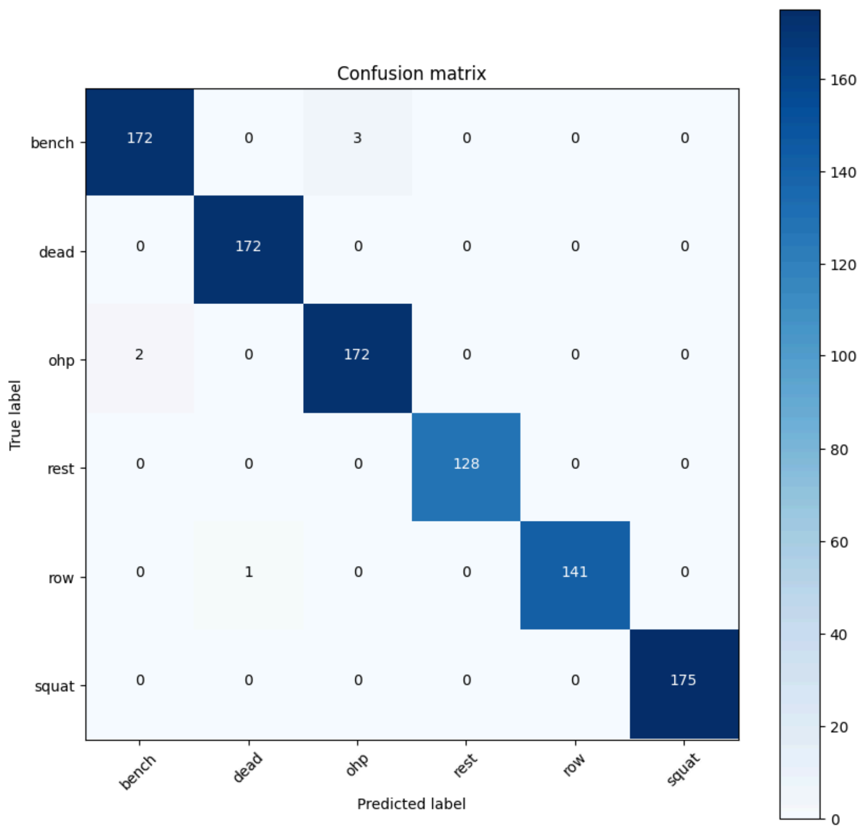


Fig. 16. Confusion Matrix

Chapter 5: Schedule

4820 (First Semester of 2023):

- **Week 3-10:**
  - Introduction: Provide an overview of the project, including its goals and the concept of quantified self.
  - Goal Setting: Define the objectives and expected outcomes of the exercise tracking models.
  - Quantified Self: Explore the concept of quantified self, emphasizing the role of wearable technology.
  - MetaMotion Sensor: Introduce the MbientLab's wristband sensor research kit and its features.
  - Dataset Preparation: Begin collecting raw data from participants during barbell exercises.
- **Week 10-16:**
  - Data Processing: Convert raw sensor data into a usable format by reading CSV files, splitting data, and performing necessary cleaning procedures.

4830 (Second Continued Semester of 2023):

- **Week 1-3:**
  - Data Visualization: Use Python scripts to visualize the collected data and plot time series data.

- **Week 3-5:**
  - Outlier Detection: Implement outlier detection techniques, including Chauvenet's criterion and local outlier factor.

- **Week 5-8:**
  - Feature Engineering: Develop algorithms for feature engineering, including frequency analysis, low-pass filtering, PCA, and clustering.

- **Week 8-11:**
  - Predictive Modeling: Explore and implement predictive models such as Naive Bayes, *Random Forest, and Neural Network.

- **Week 10-16:**
  - Project Paper: Compile the findings, methodology, and results into the final project paper.

Chapter 6: Cost

In this part, we explain the financial components of our project:

Itemized Costs:

1. **Equipment (Wristband Kit):**

The use of the Wristband Kit by MbientLab is a critical component of our research. This package, which includes accelerometers and gyroscopes, is essential for gathering complete sensor data during strength training activities. The price of the Wristband Kit is $175.34.

2. **Gym Membership:**

The crew obtained gym memberships collectively to enable the gathering of strength training workout data. The following is a breakdown of the gym membership costs:

A monthly price of $14.99 is charged to each team member, an annual fee of $49 is charged to each person, and a one-time sign-up fee of $99 is charged to each team member. To get the total cost for one person, multiply the monthly rate by 12, then add the yearly fee plus the sign-up fee. Extending this to cover the entire team of five members,
($14.99 x 12 + $49 + $99) x 5 = $1639.40

The total cost for gym memberships comes to $1639.40.

Chapter 7: Lessons learned

The process by which our team created wristband-sensor-based exercise tracking models was a worthwhile educational opportunity. We improved our teamwork abilities by focusing on clear roles and efficient communication. As we overcame unforeseen obstacles, adaptability proved to be essential, highlighting the value of flexibility in project management. Each team member gained a variety of skills on their own, including analytical rigor, machine learning proficiency, experimental planning, and data collection expertise. This project improved our technical proficiency while emphasizing the value of interdisciplinary expertise in developing a thorough and cohesive strategy.

The process of creating wristband-sensor-based exercise tracking models was a valuable learning experience for our team. We honed our teamwork skills by emphasizing clear roles and effective communication. As we overcame unexpected challenges, adaptability proved to be essential, underscoring the importance of flexibility in project management. Each team member developed a variety of skills, including analytical rigor, machine learning proficiency, experimental planning, and data collection expertise. This project improved our technical skills while highlighting the value of interdisciplinary expertise in developing a comprehensive and cohesive strategy.

We began by defining the scope of our project and identifying the key tasks that needed to be completed. We then divided the work into smaller, more manageable subtasks and assigned each team member a specific role. This helped us to ensure that everyone was clear on their responsibilities and that we were working together efficiently. Throughout the project, we maintained open communication with each other. We met regularly to discuss our progress, share ideas, and troubleshoot problems. This helped us to stay on track and to avoid costly mistakes. We also had to be adaptable as we encountered unexpected challenges. For example, we originally planned to use a specific type of wristband sensor, but we were unable to obtain the necessary hardware. We quickly adapted our plan and used a different type of sensor, which ultimately worked just as well.

This project was a valuable learning experience for our team. We developed our teamwork skills, honed our technical skills, and gained a deeper understanding of the value of interdisciplinary expertise. We are confident that these skills will be beneficial to us in our future endeavors.

Chapter 8: Conclusions

To sum up, this project has shown how context-aware applications can significantly improve the efficacy and efficiency of strength training. We have demonstrated that it is feasible to precisely track and analyze strength training exercises, giving users real-time feedback and customized recommendations, through the creative use of wristband accelerometers and gyroscopes. This method guarantees correct exercise form, which helps to prevent injuries in addition to optimizing workout regimens. Additionally, the incorporation of context-aware technology into fitness applications creates new opportunities for customized fitness regimens, increasing the accessibility and appeal of strength training to a wider range of users. With the potential to revolutionize the way we approach strength training and physical fitness in general, the future of context-aware applications in fitness appears bright as wearable technology and data analytics continue to progress.To summarize, this project has demonstrated how context-aware applications can significantly improve the efficacy and efficiency of strength training.

We have shown that it is possible to precisely track and analyze strength training exercises using wristband accelerometers and gyroscopes. This data can then be used to provide users with real-time feedback on their form and progress, as well as customized recommendations for exercises and sets. This method has been shown to help users improve their form, which can reduce the risk of injury and help them to achieve their fitness goals more effectively. In addition to improving form, context-aware applications can also be used to create customized fitness regimens for users. This is done by taking into account factors such as the user's fitness level, goals, and available time. This can make strength training more accessible and appealing to a wider range of users, as it can be tailored to meet their individual needs.

The future of context-aware applications in fitness is bright, as wearable technology and data analytics continue to progress. These technologies will make it possible to collect even more detailed data on users' workouts, which can be used to provide even more personalized and effective feedback and recommendations. This has the potential to revolutionize the way we approach strength training and physical fitness in general.

Acknowledgments:

We express our gratitude to our advisor, Dr. Tim H. Lin, for his guidance and support throughout this project.