Data Collection and Model Optimization Report

Part One: Additional Data Collection and Initial Decision Tree Training

1. **Method**: Combined data from both assignments and used a decision tree classifier. Automated ARFF file creation and Weka access through code.
2. **Result**:
    - Accuracy with original data: 88.4%
    - Accuracy with increased dataset: 91.3%
3. **Observation**: Accuracy improved by 2.9% with the additional data, showing the benefit of an expanded training set.

Part Two: Experimenting with Time Slices and Sliding Windows

1. **Method**: Tested 2, 3, and 4-second time slices with a 1-second sliding window. Retrained the decision tree using the expanded dataset from Part One.
2. **Result**:
    - 1-second slice accuracy: 91.3%
    - 2-second slice accuracy: 92.7%
    - 3-second slice accuracy: 90.6%
    - 4-second slice accuracy: 92.8%
3. **Best Time Interval**: The 4-second time slice provided the highest accuracy, indicating an optimal time window for data segmentation.
4. **Observation**: Accuracy increased by 1.5% compared to the 1-second interval.

Part Three: Adding Median and Root Mean Square Features

1. **Method**: Added median and RMS for each axis, creating 12 total features. Used the best time interval (4 seconds) from Part Two.
2. **Result**:
    - Accuracy with added features: 96.5%
    - Previous best accuracy (from Part Two): 92.8%
3. **Observation**: Accuracy improved by 3.7%, indicating that additional statistical features improved the classifier's performance.

Part Four: Sequential Feature Selection with Decision Tree

1. **Method**: Applied Sequential Feature Selection with a decision tree to find the best subset of features.
2. **Selected Features**: `Best feature subset: ['mean_y', 'std_y', 'mean_z', 'std_z', 'median_x', 'root_mean_square_z']`
3. **Result**:

    - Accuracy with selected features: 95.17%
    - Accuracy with all features (from Part Three): 96.5%
4. **Observation**: Accuracy worsened by 1.33% likely because SFS is a greedy algorithm that isn't guaranteed to find a globally optimal set of features.

Part Five: Comparing Classifiers (Random Forest, SVM, Decision Tree)

1. **Method**: Repeated Sequential Feature Selection for Random Forest and SVM.
2. **Results**:
    - Decision Tree:
        i. Selected features - ["mean_x", "mean_z", "root_mean_square_y", "median_z", "median_x"]
        ii. Final accuracy - 95.86%
    - Random Forest:
        i. Selected features - ["mean_x", "median_z", "mean_y", "mean_z", "std_z"]
        ii. Final accuracy - 96.2%
    - SVM:
        i. Selected features - ["median_x", "mean_z", "std_y"]
        ii. Final accuracy - 93.8%
3. **Best Classifier**: Random Forest achieved the highest accuracy of 96.2%.
4. **Observation**: Random Forest provided superior performance, likely due to its ensemble nature, which captures complex patterns better than single-model classifiers.