

# Paper Review

## Physical activity classification in free-living conditions using smartphone accelerometer data and exploration of predicted results - Kangjae Lee, Mei-Po Kwan

**Name:** Kavana Manvi Krishnamurthy

**ID:** 2158984

### Part 1: Problem being addressed.

Decreasing physical activity has emerged as a significant health concern in recent decades, as modern lifestyles increasingly favour sedentary behaviours at home and work. This shift poses heightened risks for chronic diseases such as obesity, diabetes, and cardiovascular conditions. Personally, as someone who has owned a Fitbit since 2015 (trying to hit 10k steps everyday for the past 10 years), I can vouch that the endorphin release from regular exercise not only improved my mental health but also strengthened my commitment to an active lifestyle. This underscores the importance of understanding and addressing sedentary behaviour to create healthier routines.

To tackle this issue, the study employed machine learning techniques to classify physical activities using data from smartphone accelerometers, which are ubiquitous and non-intrusive. By leveraging random forest and gradient boosting classifiers, it successfully identified four distinct activity types: jogging, walking, sedentary behaviour, and standing. The gradient boosting classifier achieved an impressive predictive accuracy of 99.10%, demonstrating its capability to distinguish these activities in real-world conditions. Because the published accelerometer data used to train the models were collected under controlled conditions, the study also assessed the models using data recorded in uncontrolled, daily life scenarios. Additionally, for a more thorough examination, a visual exploration of classified physical activity types over space and time was performed on a map using GPS and accelerometer data collected from the smartphones of two subjects. These findings highlight the potential of smartphone-based systems not only for accurate activity tracking but also for enabling personalized interventions that encourage an active, balanced lifestyle, ultimately reducing the risks associated with inactivity.

### Part 2: ML methods evaluated and results found.

**Related Work:** Smartphone accelerometers have been utilized in machine learning to classify various physical activities (PA), including walking, standing, and sitting. Support Vector Machines (SVMs) have been applied for this purpose, while multilayer perceptrons achieved a predictive accuracy of 91.7%. Additionally, using optimal features derived from raw accelerometer data with correlation coefficients increased the accuracy to 97% when employing K-nearest neighbour algorithms.

**Pre-processing:** The acceleration values along the x, y, and z axes are initially smoothed using a low-pass filter to reduce noise and clean the signals effectively. Next, raw accelerometer data samples are grouped into examples, preparing the data for feature computation. Finally, features for each example are extracted using a moving window, ensuring a structured and comprehensive representation of the data.

**Feature Extraction:** A set of 59 features are extracted from each example of 200 in-stances by incorporating mathematical and statistical calculations in this study. Fast Fourier Transform (FFT) is used to calculate the mean dominant frequency and mean energy of frequency. This study especially adds two more features – 1) standard deviation of the total acceleration and 2) min-max mean value of the total acceleration. The magnitude of the total acceleration is calculated by the square root of the sum of squared acceleration of three axes.

**Training model:** Random Forest is an aggregate method that constructs multiple decision trees and classifies based on majority voting, allowing for robust generalisation and resistance to overfitting. Its randomness, with feature subsets selected at each node, ensures diversity among trees and improves accuracy. On the other hand, Gradient Boosting, with its stage-wise optimisation, focuses on learning from errors by sequentially adding weak classifiers, making it highly effective at minimising residual errors and enhancing performance. R packages Random forest and XG Boost were used specifically.

**Evaluate Performance:** While training the model, 10-fold cross-validation is performed - 10% is randomly sampled. The total number of examples is 10,243 and while doing cv, the original proportion of each PA type is accounted for. Random forest and gradient boosting classifiers are run and predictive training accuracy is 99.03% and 99.22 respectively.

The test accuracy results (free-living test data) show that random forest and gradient boosting classifiers perform well, the accuracy is slightly lower than the training accuracy, with random forest achieving 95.10% and gradient boosting reaching 99.10%. Seeing the confusion matrix, there is a small misclassification in walking examples (especially brisk walking misclassified as jogging).

When these 2 algorithms are applied to examine GPS and accelerometer data for categorizing physical activities (PA) in real-world settings, gradient boosting algorithm seems to misclassify in-vehicle activities, while indoor movements may be misclassified due to poor GPS signal quality.

### **Part 3: Critique of what the authors did both good and bad.**

What I liked is how each section is clearly explained. I appreciate how the author adheres to the machine learning workflow taught to us in Class. I expressed my opinion for each section below:

**Related Work:** I appreciate that the author thoughtfully explored multiple models, including SVM, multilayer perceptron, and K-nearest neighbour, before transitioning to more complex models like random forest and gradient boosting. This progression provides a solid foundation for understanding the capabilities of simpler classifiers and their limitations. However, it would have been valuable to see a more in-depth attempt to enhance the multilayer perceptron through parameter tuning or exploring different kernel methods for the SVM, as these approaches might have yielded improved results with less complexity. Incorporating ablation techniques systematically could have further strengthened the study by isolating and understanding the impact of individual features or preprocessing steps. Such efforts would provide a more comprehensive evaluation of the simpler models before committing to more sophisticated ensemble methods.

**Loading Data:** The author begins with the WISDM accelerometer dataset, ensuring the use of a reliable and well-documented dataset as a foundation. The WISDM dataset is imbalanced, with walking and jogging comprising 70% of labels, while standing accounts for only 4%. This imbalance risks biasing the model toward majority classes, reducing accuracy for minority classes. Merging activities like upstairs and downstairs into walking may further obscure distinctions, potentially limiting generalization. Addressing this through resampling or class weighting could enhance fairness and performance.

**Preprocessing Data:** The pre-processing steps effectively apply domain knowledge to the WISDM data, using a low-pass filter to reduce noise and a 50% overlapping moving window to improve predictive accuracy. These tailored methods, including grouping raw data into examples, show a strong understanding of signal processing and dataset-specific requirements.

**Feature Selection:** The feature extraction process is thoughtfully executed, incorporating key statistical and mathematical calculations without overwhelming the analysis with unnecessary complexity, as noted in Lipton's paper. The inclusion of both established and novel features, like the standard deviation and min-max mean of total acceleration, ensures a comprehensive yet accessible approach to characterizing physical activity.

**Model Training:** Random Forest and Gradient Boosting are both excellent choices for multiclass classification due to their powerful ensemble-based approaches, each bringing distinct advantages. Random Forest is an aggregate method that constructs multiple decision trees and classifies based on majority voting, allowing for robust generalisation and resistance to overfitting. Its randomness, with feature subsets selected at each node, ensures diversity among trees and improves accuracy. On the other hand, Gradient Boosting, with its stage-wise optimisation, focuses on learning from errors by sequentially adding weak classifiers, making it highly effective at minimising residual errors and enhancing performance. While Random Forest's strength lies in parallelisation and handling a variety of features well, Gradient Boosting excels at refining predictions over time, particularly for datasets with complex relationships. The contrast between these methods — Random Forest's simplicity and speed versus Gradient Boosting's precision and iterative refinement — makes them complementary in their approach to tackling multiclass classification.

I find it intriguing that XGBoost employs parallel computing, which greatly boosts its performance while cutting down computation time. As a passionate enthusiast of distributed systems, I am eager to explore the mechanisms behind this effective parallelization and its role in XGBoost's exceptional performance in machine learning tasks.

**Model Evaluation:** 10-fold cross-validation and real-world testing validate the models' effectiveness, achieving accuracies above 99% while identifying minor misclassifications.

As for training results, it's remarkable how both random forest and gradient boosting classifiers achieve such high predictive accuracy, 99.03% and 99.22%, respectively with gradient boosting slightly outperforming random forest. The fact that the accuracy is above 99% for both models highlights their effectiveness in handling the multiclass classification task. Additionally, the use of a 50% overlap sliding window is a brilliant strategy, as it not only preserves important contextual information between data points but also boosts the accuracy by 0.54% for random forest and 0.44% for gradient boosting, showcasing its value in improving model performance. The models were evaluated using 10-fold cross-validation, where 10% of the data was randomly sampled for testing in each iteration, ensuring reliable performance estimates. Since the original proportion of each PA type is accounted for, this creates a balanced sample. This approach, combined with the high predictive accuracy achieved, demonstrates the robustness and generalization capabilities of both classifiers in a multiclass classification setting.

To get the accuracy for the test set- data from real world was gathered. Author refers to it as free-living conditions. The test accuracy results show that while both random forest and gradient boosting classifiers perform well, the accuracy is slightly lower than the training accuracy, with random forest achieving 95.10% and gradient boosting reaching 99.10%. This slight drop, especially in random forest, reflects the natural challenge of applying the model to new, real-world data, highlighting its generalization ability. Despite a small misclassification in walking examples (especially brisk walking misclassified as jogging), both models still demonstrate excellent overall performance, with perfect accuracy in sitting and standing classifications. These results suggest that the models are fairly robust and generalised, with only minor inaccuracies in distinguishing between certain walking and jogging instances.

Although the train and test accuracy were impressive, with both models achieving over 95% on test data, the lack of free-living data in the training set may have limited the models' ability to fully capture the variability of real-world conditions. Incorporating free-living data during training could potentially reduce misclassifications, such as brisk walking being labelled as jogging, and further enhance the models' generalization to diverse, real-world scenarios.

**Real-World Validation:** This study effectively employs random forest and gradient boosting classifiers to examine GPS and accelerometer data for categorizing physical activities (PA) in real-world settings, providing valuable insights into activity patterns. A significant strength is the incorporation of GPS accuracy metrics (HDOP) and interpolation techniques, which enhances the reliability of location data. The study's use of OpenStreetMap to visualize predicted activities offers an intuitive method for validating classifier outputs against geographical contexts. However, the research also uncovers limitations. Gradient boosting shows inconsistent classification of in-vehicle activities, while indoor movements may be misclassified due to poor GPS signal quality. Although random forest demonstrates higher reliability for certain scenarios, such as in-vehicle activities, the study could be improved by addressing gradient boosting's weaknesses and exploring alternative methods to enhance indoor activity classification.