In the *Gradient Boosting for Human Activity Recognition from Smartphone Sensors* (Lee, 2018) paper, the authors proposed an approach to classify Physical Activity (PA) in natural everyday settings in people's daily lives. The data used for research was publicly released wireless sensor data mining (WISDM)'s accelerometer data version 1.1, which were collected from 36 person using Android smartphones under laboratory conditions with controlled settings. Originally there were 1,098,207 instances and a set of 200 consecutive raw instances formed one example. 10,243 examples were ultimately used in modeling and cross-validation. Another test dataset was accelerometer data which have 49,016 instances recorded at 20 Hz using an LG G3 smartphone.

Two classification algorithms were used to train the models, i.e. random forest and gradient boosting. In the preprocessing steps, the authors merged two types (upstairs and downstairs) into walking that reduced the target class from 6 to 4 levels. They also used a low pass filter to smooth x, y and z acceleration values, grouped raw instances to compute features and used 50% overlap moving windows to calculate features. 59 features were extracted from calculation.

In classification, random forest grew multiple trees and classified labels based on the votes of all the trees. Among all features, a subset of features was randomly selected for each node, and the best split on the subset of features was chosen to split node. Gradient boosting as a forward stage-wise optimization algorithm used votes of each weak classifier, which is learned at every iteration, to generate a strong classifier. In order to unveil how the accelerometer data collected in a laboratory setting provide convincing predicted results in various daily activities, the authors used 10-fold cross-validation and tested accelerometer data collected from a smartphone of one adult subject in free-living conditions.

Random forest and gradient boosting classifiers were run, and predictive accuracy is 99.03% and 99.22% respectively using 10-fold cross-validation. In the confusion matrix of the random forest classifier, walking and jogging had the lowest classification accuracy. 46 walking examples and 48 jogging examples were incorrectly predicted as jogging and walking respectively. The use of a 50% overlap window helped to increase the prediction accuracy of both random forest and gradient boosting by 0.54% and 0.44% respectively.

The two learning models generated were tested using accelerometer data collected from one adult subject in free-living conditions to validate the applicability of the generated learning models. The averaged predictive accuracy of the random forest classifier on free living condition dataset in the 200 iterations was 95.10%, which is almost 4% lower than the accuracy of its training model, whereas gradient boosting had a 99.10% predictive accuracy. The confusion matrix of random forest indicated that almost 9% of the walking examples were wrongly classified as jogging. It was caused by some instances recorded during brisk walking at a fast pace with the walking label, confirmed by exploration of GPS trajectories coupled with classified types. On the other hand, in the classification using gradient boosting, walking examples including walking up and down stairs are classified with high accuracy (97.80%). For sitting and standing, both random forest and gradient boosting achieve perfect classification accuracy.

GPS trajectories coupled with predicted types were visualized on OpenStreetMap to assess the prediction results with movement patterns on GPS tracks and geographic contexts. 4 types of PA behaviors were correctly classified by random forest and gradient boosting. Characteristics of some walking examples that were wrongly classified as jogging were visually examined. However, indoor GPS predictions had lower accuracy due to signal loss and barriers.

In addition to the four basic PA types, other kinds of movement behaviors were explored to assess the predictive accuracy of the two classifiers for different daily activities. Both random forest and gradient boosting yield similar and plausible results for sports and shopping activities. Gradient boosting predicted slightly more standing examples in the sports and shopping activities while only random forest showed consistent results in the prediction of the subjects' postures when they were in a vehicle.

The authors pointed out that if the smartphone was carried in other pockets, like a jacket pocket, the collected accelerometer data from the smartphones were not be able to predict types correctly. If smartphones are out of their pant pockets for phone calls or playing games, PA types were likely to be incorrectly classified. Fast walking trips in the middle of walking trips, supplementary data using daily activity diaries and transportation modes as additional PA types were suggested for future research.

Personally, I like the idea of using GPS trajectories as a complementary source to validate machine learning models. There are a few questions I have after reading this paper.

1. The authors merged upstairs and downstairs with walking, which increased the percentage of walking from 39% to 59%. This action caused an obvious imbalance between classified class. Walking and stair climbing are different categories in terms of exercise. If we don't merge those types, will the model predict a different result?

2. Data was collected on 36 people using smartphone with Android system. I was curious 1) if the group of 36 people or 2-adult subject in the free-living assessment would really represent the target users; 2) if the operating system is switching to iOS. The authors mentioned the intention of using two different devices to identify whether there is any difference in predicted types using accelerometer data collected from the two smartphones, but they didn't provide a clear statement or conclusion.

3. Random forest and gradient boosting seemed to perform well on their data. Interpretation is not the primary goal for this project. Only 2 classification algorithms applied on the dataset and parallel comparison are not enough for a broader application.

4. It is true that we rely heavily on smartphones, but we don't necessarily carry the phone on us every time. As the authors mentioned that if the smartphones were carried in pockets other than pants, the models lose the ability to predict well. In recent years, smartwatches gradually found its position as fitness tracker and may collect more accurate data of our bodies. If such information could be used for analysis, I think the results would be more exciting and practical.