**Part 1: Understanding the Data**

* **What is the objective of the data collection process?**

The objective of data collection process for HAR is to observe the actions of humans when a set of activities are provided to them and their surrounding environment. The recognition could be achieved by obtaining information from sources like environment or body worn sensors. Smartphones are bringing up new opportunities for such recognition applications where huge amount of information is obtained from user while the phone is a firsthand sensing tool. This has enabled us to observe human behavior in a very progressive manner.

* **What human activity types does this dataset have? How many subjects/people have performed these activities?**

The human activity types are as follows**-**

* Standing
* Sitting
* Laying down
* Walking
* Walking downstairs
* Walking upstairs

30 volunteers between the age of 19 to 48 years have performed the above mentioned activities.

* **How many instances are available in the training and test sets? How many features are used to represent each instance? Summarize the type of features extracted in 2-3 sentences.**

The training dataset has 7352 instances while the testing dataset has 2947 instances. While 561 features have been used to represent these instances.

Features have been extracted using inertial data from smartphone accelerometers and gyroscopes at a sampling rate of 50Hz. These signals were then preprocessed to reduce noise by the use of a median and low pass Butter worth filter. Standard measures such as the mean, correlation, signal magnitude area (SMA) and autoregression coefficients were also employed for the feature mapping.

* **Describe briefly what machine learning model is used in this paper for activity recognition and how is it trained. How much is the maximum accuracy achieved?**

Support Vector Machine [SVM] algorithm has been used as a machine learning model. The SVM hyperparameters are selected using 10-fold Cross Validation procedure and Gaussian kernels. An overall maximum accuracy of 96% has been obtained for the test data.

**Part 6: Discussion**

**Write a brief discussion about which classification method achieved the best performance. Your thoughts on the reason behind this. What method performed the worst? Could you do better or worse than the results in the dataset paper? Do you have any suggestions to further improve model performances?**

As per the results obtained for all the 4 classifiers, SVM has turned out to be the best classifier with an accuracy of 96%. While the worst performing classifier was KNN having an accuracy of 89.6%

We can say that the model performance can be greatly increased by using Principal Component Analysis(PCA) so as to reduce the number of features by removing unwanted and less important features.