Figure 1. (left) Classification errors for all classes of activities is shown together with total misclassification error (OOB) as a function of number of trees used in the model. It can be seen that different activities have different sensitivity of classification error to number of trees used. For example ‘laying’ activity is easiest to predict and around 50 trees would be sufficient to get 0% classification errors in training data set. In contrary “standing” activity classification continues to improve up to 500 trees and it might benefit even more with higher number of trees. The overall classification error OOB reaches minimum for around 200 trees and doesn’t seem to improve with higher number of trees used. Contrary, “sitting” and “walkdown” activities seem to have higher classification errors for number of trees higher than around 200. (right) Value of a mean decrease of GINI coefficient [8] is shown for the 30 most important variables as classified by “randomForest” algorithm.

Gini criterion is the splitting criterion used in random forest (see p. 11 in [5]). Mean decrease of GINI coefficient is one of the metrics used by “randomForest” algorithm to classify importance of each variable in contribution to the final model. It is apparent from the plot that there are 9 covariates listed in Table 2 that have particularly high mean decrease of GINI coefficient value above 120. For all other variables the mean decrease of GINI coefficient values are below 80.