Figure 6 – Graph showing precision and recall for all sub-classes obtained using level 3 random forest classifier.

Precision and recall for most sub-classes is around similar range, besides seven sub-classes that have a higher recall. Minimum precision is 74.07% (sub-class 3.7) while minimum recall is 57.05% (sub-class 2.3).

Results from experiments using Model 2

Model 2 (see Figure 2) is a direct single step approach to predicting the sub-class of enzymes. As in previous cases, we first sought to find optimal values of the random forest parameters. The random forest classifier reports the lowest OOB error when $mtry = 5$ and $ntree = 200$ (see Figure 7). Using these values, we performed the classification experiment, the results of which are summarised in Table 5.

Figure 7 – Model 2 random forest classifier OOB error for different $ntree$ and $mtry$ values.

The graph shows OOB error obtained from different runs of the random forest classifier during training phase. The least value of OOB error is obtained when $mtry = 5$ and $ntree = 200$. 