UPDATED RESULTS 2

Table 1: model results for different classification models on above **median improvement in HDRS21 score (H7)**. Features from **Frequency Power** abstractions, subject records are full records. Data imbalance is 45% above median and 55 % below median (many subject has the median value). The train test ratio is 80/20.

Each model results were calculated 10 times with different train test split and the results above are mean ± std. Data size contains 77 records.

Model	Accuracy	Sensitivity	Specificity
XGBoost	0.67 ± 0.13	0.72 ± 0.10	0.62 ± 0.21
RandomForest	0.57 ± 0.13	0.57 ± 0.17	0.57 ± 0.21
Catboost	0.64 ± 0.08	0.62 ± 0.13	0.65 ± 0.19
TPF	0.59 ± 0.08	0.58 ± 0.13	0.60 ± 0.22
TPF_top3	0.57 ± 0.08	0.74 ± 0.09	0.38 ± 0.16
TPF_top5	0.63 ± 0.09	0.75 ± 0.12	0.49 ± 0.14
TPF_top7	0.61 ± 0.06	0.68 ± 0.12	0.52 ± 0.16
TPF_top9	0.59 ± 0.08	0.68 ± 0.12	0.49 ± 0.22

Table 5: model results for different classification models on above **median improvement in HDRS21 score (H1)**. Features from **Frequency Power** abstractions, subject records are divided into **1 minute with 30 seconds overlap** records assuring no data leakage (same subject in train and test). Data imbalance is 49% above median and 51 % below median. The train test ration is 80/20.

Each model results were calculated 10 times with different train test split and the results above are mean ± std. Data size contains 464 records.

Model	Accuracy	Sensitivity	Specificity
XGBoost	0.54 ± 0.11	0.42 ± 0.19	0.65 ± 0.26
RandomForest	0.61 ± 0.10	0.38 ± 0.18	0.83 ± 0.15
Catboost	0.60 ± 0.09	0.41 ± 0.11	0.78 ± 0.16
TPF	0.50 ± 0.15	0.43 ± 0.18	0.57 ± 0.22
TPF_top3	0.62 ± 0.11	0.43 ± 0.24	0.80 ± 0.15
TPF_top5	0.60 ± 0.12	0.38 ± 0.24	0.80 ± 0.13
TPF_top7	0.60 ± 0.13	0.37 ± 0.23	0.82 ± 0.14
TPF_top9	0.61 ± 0.14	0.37 ± 0.24	0.83 ± 0.14
TPF_top11	0.62 ± 0.13	0.37 ± 0.21	0.87 ± 0.14
TPF_top13	0.62 ± 0.11	0.35 ± 0.20	0.88 ± 0.14
TPF_top15	0.59 ± 0.08	0.34 ± 0.15	0.83 ± 0.12

Conclusions:

We can see that the best model on H7 treatment response is good on sensitivity, while the best model on H1 treatment response is good on specificity. That means we can detect the subjects who will respond well to H7 and the subjects that will not respond well to H1 treatment.

Important note: Even though the models are not exactly the same (one is with full entities and the other with entities divided into parts), the results and the conclusion from before remains the same. When I tested the same models on the data of H1 and H7 separately, H7 still gets better sensitivity than specificity and H1 still gets better specificity than sensitivity. The reason I selected these models is because they perform best. Probably because the lack of records in H1 treatment (53 subjects) relative to H7 treatment (77 subjects).