UPDATED RESULTS 3

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 was made by **leaving one out CV.**

In this table I used a method of **leaving only the above 60% and under 40% of train set** based on **improvement in HDRS21** which gave the **best results in both specificity and sensitivity**.

Model	Accuracy	Sensitivity	Specificity
XGBoost	0.636364	0.676471	0.604651
RandomForest	0.662338	0.558824	0.744186
Catboost	0.61039	0.588235	0.627907
TPF	0.636364	0.647059	0.627907
TPF_top3	0.584416	0.794118	0.418605
TPF_top5	0.649351	0.823529	0.511628
TPF_top7	0.584416	0.735294	0.465116
TPF_top9	0.519481	0.705882	0.372093
TPF_top11	0.480519	0.647059	0.348837
TPF_top13	0.506494	0.764706	0.302326
TPF_top15	0.493506	0.735294	0.302326

Table 2: 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 was made by **leaving one out CV.**

In this table I used a method of **leaving only the above 60% and under 40% of train set** based on improvement in HDRS21 which gave **best results on sensitivity**.

Model	Accuracy	Sensitivity	Specificity
XGBoost	0.596296	0.607843	0.585965
RandomForest	0.537037	0.545098	0.529825
Catboost	0.566667	0.580392	0.554386
TPF	0.531481	0.521569	0.540351
TPF_top3	0.551852	0.686275	0.431579
TPF_top5	0.553704	0.701961	0.421053
TPF_top7	0.542593	0.686275	0.414035
TPF_top9	0.542593	0.654902	0.442105
TPF_top11	0.488889	0.619608	0.37193
TPF_top13	0.492593	0.584314	0.410526
TPF_top15	0.477778	0.596078	0.37193

Table 3: 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 was made by **leaving one out CV.**

In this table I used a method of **leaving only the above 65% and under 35% of train set** based on **improvement in HDRS21** which gave **best results on specificity.**

Model	Accuracy	Sensitivity	Specificity
XGBoost	0.537634	0.384615	0.684211
RandomForest	0.5	0.335165	0.657895
Catboost	0.553763	0.395604	0.705263
TPF	0.478495	0.456044	0.5
TPF_top3	0.5	0.274725	0.715789
TPF_top5	0.502688	0.274725	0.721053
TPF_top7	0.521505	0.307692	0.726316
TPF_top9	0.518817	0.296703	0.731579
TPF_top11	0.510753	0.263736	0.747368
TPF_top13	0.5	0.236264	0.752632
TPF_top15	0.491935	0.247253	0.726316

Discussion:

After testing the new models based on leave one out CV, we can still see the H1 model gets better specificity results and H7 model gets better sensitivity results, but we can see it also depends on percentage from the median input as training data. As a matter of fact, I don't have a strong intuition on why this is happening but I believe that more data from H1 subject will give us better understanding.