Table 4: OpenML binary classification data set details

Data	rows	columns	class sizes	Balanced accuracy on default	Room for improvement
EEG eye state	14980	14	(8257, 6723)	90.28%	4.38%
Electricity	45312	8	(26075, 19237)	87.78%	5.14%
Heart statlog	270	13	(150, 120)	79.42%	6.17%
Oil spill	937	49	(896, 41)	63.22%	11.36%
Pollen	3848	5	(1924, 1924)	48.86%	3.35%
Sonar	208	61	(111, 97)	87.43%	3.82%
PC3	1563	37	(1403, 160)	58.99%	4.82%

A Appendix

A.1 Dataset details

The details of the binary classification data sets used in our evaluation is reported in Table 4. We report the 10-fold cross-validated balanced accuracy of the default HP configuration on each of data sets with centralized training. The "Room for improvement" column in Table 4 denotes the difference between the best 10-fold cross-validated balanced accuracy obtained via centralized HPO and the 10-fold cross-validated balanced accuracy of the default HP configuration.

A.2 Search space

We use the search space definition used in the NeurIPS 2020 Black-box optimization challenge (`https://bbochallenge.com/`), described in details in the API documentation[1]. Given this format for defining the HPO search space, we utilize the following precise search space for the `HistGradientBoostingClassifier` in `scikit-learn`:

```
api_config = {
    'max_iter': 'type': 'int', 'space': 'linear', 'range': (10, 200),
    'learning_rate': 'type': 'real', 'space': 'log', 'range': (1e-3, 1.0),
    'min_samples_leaf': 'type': 'int', 'space': 'linear', 'range': (1, 40),
    'l2_regularization': 'type': 'real', 'space': 'log', 'range': (1e-4, 1.0),
}
```

The HP configuration we consider for the single-shot baseline described in §4 is as follows:

```
api_config = {
    'max_iter': 100,
    'learning_rate': 0.1,
    'min_samples_leaf': 20,
    'l2_regularization': 0,
}
```

[1] `https://github.com/rdturnermtl/bbo_challenge_starter_kit/#configuration-space`