

# Survival regression with accelerated failure time model in XGBoost: Supplementary Material

Avinash Barnwal  
Department of Statistics, Stony Brook University

and

Hyunsu Cho

NVIDIA

and

Toby Hocking

School of Informatics, Computing, and Cyber Systems,  
Northern Arizona University

January 27, 2022

## 1 Full proof for Definition 3, the gradient and hessian of the AFT loss

We first derive the first- and second-order partial derivatives of  $\ell_{\text{AFT}}$  for the uncensored case, using basic rules of Calculus:

$$\begin{aligned}
 \ell(y, u) &= -\ln \left[ f_Z(s(y)) \cdot \frac{1}{\sigma y} \right] \\
 \frac{\partial \ell}{\partial u} &= -\frac{\partial}{\partial u} \ln \left[ f_Z(s(y)) \cdot \frac{1}{\sigma y} \right] \\
 &= -\frac{1}{f_Z(s(y)) \cdot \frac{1}{\sigma y}} \cdot \frac{\partial}{\partial u} \left[ f_Z(s(y)) \cdot \frac{1}{\sigma y} \right] && \text{Chain Rule} \\
 &= -\frac{\sigma y}{f_Z(s(y))} \cdot \left[ f'_Z(s(y)) \cdot \frac{\partial}{\partial u} \frac{\ln y - u}{\sigma} \cdot \frac{1}{\sigma y} \right] && \text{Chain Rule} \\
 &= -\frac{\sigma y}{f_Z(s(y))} \cdot \left[ f'_Z(s(y)) \cdot -\frac{1}{\sigma} \cdot \frac{1}{\sigma y} \right] \\
 &= \frac{f'_Z(s(y))}{\sigma f_Z(s(y))} \\
 \frac{\partial^2 \ell}{\partial u^2} &= \frac{\partial}{\partial u} \frac{\partial \ell}{\partial u} \\
 &= \frac{\partial}{\partial u} \frac{f'_Z(s(y))}{\sigma f_Z(s(y))} \\
 &= \frac{\partial/\partial u[f'_Z(s(y))] \cdot \sigma f_Z(s(y)) - f'_Z(s(y)) \cdot \sigma \partial/\partial u[f_Z(s(y))]}{\sigma^2 f_Z(s(y))^2} && \text{Quotient Rule} \\
 &= \frac{f''_Z(s(y)) \cdot (-1/\sigma) \cdot \sigma f_Z(s(y)) - f'_Z(s(y)) \cdot \sigma f'_Z(s(y)) \cdot (-1/\sigma)}{\sigma^2 f_Z(s(y))^2} && \text{Chain Rule} \\
 &= -\frac{f''_Z(s(y)) f_Z(s(y)) - f'_Z(s(y))^2}{\sigma^2 f_Z(s(y))^2}
 \end{aligned}$$

The censored case proceeds similarly:

$$\begin{aligned}
\ell(y, u) &= -\ln [F_Z(s(\bar{y})) - F_Z(s(\underline{y}))] \\
\frac{\partial \ell}{\partial u} &= -\frac{\partial}{\partial u} \ln [F_Z(s(\bar{y})) - F_Z(s(\underline{y}))] \\
&= -\frac{1}{F_Z(s(\bar{y})) - F_Z(s(\underline{y}))} \cdot \frac{\partial}{\partial u} [F_Z(s(\bar{y})) - F_Z(s(\underline{y}))] && \text{Chain Rule} \\
&= -\frac{1}{F_Z(s(\bar{y})) - F_Z(s(\underline{y}))} \cdot \left[ f_Z(s(\bar{y})) \cdot -\frac{1}{\sigma} - f_Z(s(\underline{y})) \cdot -\frac{1}{\sigma} \right] && \text{Chain Rule; } f_Z = F'_Z \\
&= \frac{f_Z(s(\bar{y})) - f_Z(s(\underline{y}))}{\sigma[F_Z(s(\bar{y})) - F_Z(s(\underline{y}))]} \\
\frac{\partial^2 \ell}{\partial u^2} &= \frac{\partial}{\partial u} \frac{\partial \ell}{\partial u} \\
&= \frac{\partial}{\partial u} \frac{f_Z(s(\bar{y})) - f_Z(s(\underline{y}))}{\sigma[F_Z(s(\bar{y})) - F_Z(s(\underline{y}))]} \\
&= \frac{\partial/\partial u[f_Z(s(\bar{y})) - f_Z(s(\underline{y}))] \cdot \sigma[F_Z(s(\bar{y})) - F_Z(s(\underline{y}))]}{\sigma^2[F_Z(s(\bar{y})) - F_Z(s(\underline{y}))]^2} \\
&\quad - \frac{[f_Z(s(\bar{y})) - f_Z(s(\underline{y}))] \cdot \sigma \partial/\partial u[F_Z(s(\bar{y})) - F_Z(s(\underline{y}))]}{\sigma^2[F_Z(s(\bar{y})) - F_Z(s(\underline{y}))]^2} && \text{Quotient Rule} \\
&= \frac{[f'_Z(s(\bar{y})) - f'_Z(s(\underline{y}))] \cdot (-1/\sigma) \cdot \sigma[F_Z(s(\bar{y})) - F_Z(s(\underline{y}))]}{\sigma^2[F_Z(s(\bar{y})) - F_Z(s(\underline{y}))]^2} \\
&\quad - \frac{[f_Z(s(\bar{y})) - f_Z(s(\underline{y}))] \cdot \sigma[f_Z(s(\bar{y})) - f_Z(s(\underline{y}))] \cdot (-1/\sigma)}{\sigma^2[F_Z(s(\bar{y})) - F_Z(s(\underline{y}))]^2} && \text{Chain Rule; } f_Z = F'_Z \\
&\quad - [F_Z(s(\bar{y})) - F_Z(s(\underline{y}))][f'_Z(s(\bar{y})) - f'_Z(s(\underline{y}))] \\
&= \frac{+ [f_Z(s(\bar{y})) - f_Z(s(\underline{y}))]^2}{\sigma^2[F_Z(s(\bar{y})) - F_Z(s(\underline{y}))]^2}
\end{aligned}$$

## 2 Effect of hyperparameters on model performance

The goal of the experiment is to quantify the impact of the hyperparameter choice on the generalization performance. For each model produced by each hyperparameter combination, we compute the test accuracy using the held-out test set.

The following tables show the validation and test accuracy of all models produced by the hyperparameter search. See Section 3.3 for detailed instructions.

Is the AFT method very sensitive to the hyperparameter choice? If yes, then we will need to run many trials of hyperparameter search and the performance benefit of XGBoost will be negated. If not, then we will be able to run a comparatively small number of trials and we will be able to obtain an optimal model in short amount of time.

The answer to the question of the preceding paragraph is yes. For all data sets, running 1000 trials of hyperparameter search did not produce significant advantage over running only 100 trials, when measured in the test accuracy measure. Thus, we can simply run 100 rounds of hyperparameter search and obtain a close-to-optimal model, at least when measured in the aggregate with the test accuracy.

Unfortunately, this analysis has a gaping hole: a model being optimal in the aggregate does not mean it is well calibrated in individual predictions. Two models may have similar test accuracy but produce wildly differing prediction for some inputs. See discussion in Section 4.

Table 1: Comparison of hyperparameter search methods. Both validation and test accuracy are averages over the 4 outer cross-validation folds.

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
Dataset ATAC_JV_adipose, normal distribution					
0	Grid	max_depth, reg_lambda	54	0.715249	0.448685
1	Grid	learning_rate, max_depth	36	0.710944	0.446252
2	Grid	reg_alpha	6	0.686967	0.445603
3	Grid	max_depth, min_child_weight	45	0.708613	0.443379
4	Random	All six	1000	0.724057	0.441571
5	Grid	max_depth	9	0.704368	0.440111
6	Baseline	None (all defaults)	1	0.677217	0.435522
7	Random	All six	100	0.718862	0.434572
8	Grid	min_child_weight	5	0.684872	0.432649
9	Grid	max_depth, reg_alpha	54	0.710192	0.427622
10	Grid	reg_lambda	6	0.698796	0.425557
11	Grid	max_depth, aft_loss_distribution_scale	54	0.719006	0.425557
12	Grid	learning_rate, reg_alpha	24	0.704978	0.423541
13	Grid	reg_alpha, aft_loss_distribution_scale	36	0.709728	0.420552
14	Grid	min_child_weight, reg_alpha	30	0.694428	0.420089
15	Grid	reg_lambda, aft_loss_distribution_scale	36	0.709096	0.418952
16	Grid	learning_rate, min_child_weight	20	0.709576	0.418629
17	Grid	reg_alpha, reg_lambda	36	0.700208	0.416635
18	Grid	learning_rate, aft_loss_distribution_scale	24	0.707424	0.415268
19	Grid	learning_rate, reg_lambda	24	0.708276	0.413947
20	Grid	aft_loss_distribution_scale	6	0.700577	0.412047
21	Grid	learning_rate	4	0.698327	0.411630
22	Grid	min_child_weight, reg_lambda	30	0.703793	0.400971
23	Grid	min_child_weight, aft_loss_distribution_scale	30	0.706980	0.395038
Dataset ATAC_JV_adipose, logistic distribution					
0	Grid	learning_rate, max_depth	36	0.717829	0.451651
1	Grid	max_depth, aft_loss_distribution_scale	54	0.718995	0.445811
2	Grid	aft_loss_distribution_scale	6	0.705980	0.439856
3	Grid	reg_alpha, aft_loss_distribution_scale	36	0.711729	0.439092
4	Grid	min_child_weight, aft_loss_distribution_scale	30	0.706704	0.437747
5	Grid	learning_rate	4	0.707096	0.430030
6	Random	All six	100	0.720789	0.426439
7	Grid	reg_lambda	6	0.704542	0.425743
8	Grid	learning_rate, min_child_weight	20	0.709754	0.424283
9	Grid	min_child_weight, reg_lambda	30	0.707443	0.415964
10	Grid	learning_rate, reg_lambda	24	0.712631	0.415662
11	Grid	learning_rate, reg_alpha	24	0.708575	0.409893
12	Grid	reg_lambda, aft_loss_distribution_scale	36	0.716163	0.407899
13	Grid	learning_rate, aft_loss_distribution_scale	24	0.711026	0.407691
14	Random	All six	1000	0.727451	0.406277
15	Grid	max_depth, reg_lambda	54	0.716056	0.406069
16	Grid	reg_alpha, reg_lambda	36	0.710365	0.405976
17	Grid	min_child_weight, reg_alpha	30	0.437865	0.016986
18	Grid	reg_alpha	6	0.437119	0.015943
19	Grid	max_depth, reg_alpha	54	0.437119	0.015943
20	Baseline	None (all defaults)	1	0.021481	0.000000

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
21	Grid	max_depth	9	0.021481	0.000000
22	Grid	min_child_weight	5	0.021481	0.000000
23	Grid	max_depth, min_child_weight	45	0.021481	0.000000
Dataset ATAC_JV_adipose, extreme distribution					
0	Grid	learning_rate, max_depth	36	0.717760	0.451024
1	Grid	max_depth, aft_loss_distribution_scale	54	0.716240	0.450167
2	Grid	max_depth, reg_lambda	54	0.717171	0.444699
3	Grid	max_depth	9	0.713462	0.440481
4	Grid	reg_alpha, aft_loss_distribution_scale	36	0.711823	0.435939
5	Random	All six	1000	0.724805	0.430863
6	Grid	aft_loss_distribution_scale	6	0.702618	0.430748
7	Grid	learning_rate	4	0.706831	0.423656
8	Grid	learning_rate, aft_loss_distribution_scale	24	0.712637	0.423587
9	Grid	min_child_weight	5	0.704203	0.423078
10	Random	All six	100	0.721105	0.422962
11	Baseline	None (all defaults)	1	0.699781	0.422798
12	Grid	max_depth, min_child_weight	45	0.718588	0.422521
13	Grid	learning_rate, min_child_weight	20	0.712486	0.420204
14	Grid	learning_rate, reg_lambda	24	0.712615	0.419810
15	Grid	max_depth, reg_alpha	54	0.719171	0.419763
16	Grid	reg_alpha, reg_lambda	36	0.708990	0.417446
17	Grid	min_child_weight, reg_alpha	30	0.717606	0.414851
18	Grid	reg_lambda	6	0.706155	0.413622
19	Grid	min_child_weight, aft_loss_distribution_scale	30	0.712047	0.410332
20	Grid	reg_alpha	6	0.706762	0.409381
21	Grid	reg_lambda, aft_loss_distribution_scale	36	0.714897	0.408896
22	Grid	min_child_weight, reg_lambda	30	0.711983	0.400645
23	Grid	learning_rate, reg_alpha	24	0.711072	0.397679
Dataset CTCF_TDH_ENCODE, normal distribution					
0	Grid	max_depth, aft_loss_distribution_scale	54	0.882398	0.807692
1	Grid	learning_rate, aft_loss_distribution_scale	24	0.867323	0.798077
2	Grid	min_child_weight, aft_loss_distribution_scale	30	0.863787	0.798077
3	Random	All six	1000	0.884321	0.788462
4	Grid	learning_rate, reg_lambda	24	0.854172	0.788462
5	Grid	aft_loss_distribution_scale	6	0.861864	0.788462
6	Grid	learning_rate, max_depth	36	0.875066	0.788462
7	Grid	reg_alpha, aft_loss_distribution_scale	36	0.868936	0.788462
8	Grid	learning_rate, reg_alpha	24	0.858018	0.783654
9	Grid	max_depth, reg_alpha	54	0.859049	0.783654
10	Random	All six	100	0.876939	0.778846
11	Grid	reg_lambda, aft_loss_distribution_scale	36	0.874965	0.774038
12	Grid	learning_rate	4	0.840710	0.759615
13	Grid	reg_alpha, reg_lambda	36	0.847821	0.759615
14	Grid	learning_rate, min_child_weight	20	0.850326	0.754808
15	Grid	max_depth, reg_lambda	54	0.857487	0.730769
16	Grid	max_depth	9	0.845948	0.721154
17	Grid	min_child_weight, reg_lambda	30	0.842412	0.721154
18	Grid	max_depth, min_child_weight	45	0.852710	0.697115
19	Grid	reg_lambda	6	0.840439	0.692308

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
20	Grid	min_child_weight, reg_alpha	30	0.836062	0.692308
21	Baseline	None (all defaults)	1	0.826446	0.682692
22	Grid	min_child_weight	5	0.826446	0.677885
23	Grid	reg_alpha	6	0.828369	0.673077
Dataset CTCF_TDH_ENCODE, logistic distribution					
0	Grid	min_child_weight, aft_loss_distribution_scale	30	0.871530	0.831731
1	Random	All six	100	0.884321	0.826923
2	Grid	max_depth, aft_loss_distribution_scale	54	0.884321	0.826923
3	Grid	learning_rate, reg_alpha	24	0.869246	0.822115
4	Grid	learning_rate	4	0.865400	0.817308
5	Grid	max_depth	9	0.877299	0.817308
6	Grid	max_depth, reg_alpha	54	0.880785	0.812500
7	Grid	reg_alpha	6	0.860561	0.807692
8	Grid	reg_alpha, aft_loss_distribution_scale	36	0.874345	0.807692
9	Grid	learning_rate, max_depth	36	0.880475	0.802885
10	Grid	learning_rate, reg_lambda	24	0.869246	0.802885
11	Baseline	None (all defaults)	1	0.857076	0.798077
12	Random	All six	1000	0.889780	0.798077
13	Grid	reg_alpha, reg_lambda	36	0.863838	0.798077
14	Grid	aft_loss_distribution_scale	6	0.867684	0.798077
15	Grid	max_depth, min_child_weight	45	0.884321	0.793269
16	Grid	max_depth, reg_lambda	54	0.878862	0.793269
17	Grid	reg_lambda	6	0.861864	0.788462
18	Grid	min_child_weight, reg_lambda	30	0.865710	0.788462
19	Grid	reg_lambda, aft_loss_distribution_scale	36	0.873092	0.783654
20	Grid	learning_rate, min_child_weight	20	0.869246	0.783654
21	Grid	min_child_weight	5	0.862485	0.759615
22	Grid	learning_rate, aft_loss_distribution_scale	24	0.875016	0.750000
23	Grid	min_child_weight, reg_alpha	30	0.869556	0.725962
Dataset CTCF_TDH_ENCODE, extreme distribution					
0	Grid	max_depth, reg_alpha	54	0.850996	0.798077
1	Grid	max_depth	9	0.845227	0.788462
2	Grid	learning_rate	4	0.841381	0.783654
3	Grid	learning_rate, reg_lambda	24	0.845227	0.783654
4	Grid	learning_rate, max_depth	36	0.847201	0.778846
5	Grid	learning_rate, aft_loss_distribution_scale	24	0.848763	0.778846
6	Random	All six	1000	0.870859	0.774038
7	Grid	learning_rate, min_child_weight	20	0.850996	0.774038
8	Grid	learning_rate, reg_alpha	24	0.850996	0.769231
9	Grid	reg_alpha, reg_lambda	36	0.852249	0.769231
10	Random	All six	100	0.854843	0.764423
11	Grid	max_depth, min_child_weight	45	0.858278	0.759615
12	Grid	min_child_weight	5	0.847150	0.754808
13	Baseline	None (all defaults)	1	0.835612	0.750000
14	Grid	reg_alpha	6	0.847150	0.745192
15	Grid	min_child_weight, reg_alpha	30	0.863787	0.745192
16	Grid	max_depth, aft_loss_distribution_scale	54	0.861915	0.740385
17	Grid	reg_alpha, aft_loss_distribution_scale	36	0.856044	0.740385
18	Grid	reg_lambda, aft_loss_distribution_scale	36	0.854532	0.740385

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
19	Grid	reg_lambda	6	0.843304	0.735577
20	Grid	min_child_weight, reg_lambda	30	0.854842	0.735577
21	Grid	aft_loss_distribution_scale	6	0.841381	0.706731
22	Grid	max_depth, reg_lambda	54	0.852919	0.706731
23	Grid	min_child_weight, aft_loss_distribution_scale	30	0.854843	0.682692
Dataset H3K27ac-H3K4me3_TDHAM_BP, normal distribution					
0	Grid	max_depth, aft_loss_distribution_scale	54	0.884932	0.738933
1	Grid	learning_rate, min_child_weight	20	0.881360	0.726499
2	Grid	reg_alpha, aft_loss_distribution_scale	36	0.883761	0.725599
3	Grid	max_depth, reg_alpha	54	0.879121	0.720361
4	Random	All six	100	0.885043	0.719415
5	Grid	learning_rate, aft_loss_distribution_scale	24	0.883147	0.718623
6	Grid	learning_rate	4	0.880459	0.718509
7	Grid	learning_rate, reg_alpha	24	0.883937	0.718318
8	Grid	min_child_weight, reg_lambda	30	0.879320	0.717944
9	Grid	learning_rate, max_depth	36	0.885600	0.716705
10	Grid	learning_rate, reg_lambda	24	0.883249	0.715441
11	Grid	max_depth	9	0.877523	0.714167
12	Random	All six	1000	0.889028	0.713855
13	Grid	max_depth, reg_lambda	54	0.880891	0.713815
14	Grid	max_depth, min_child_weight	45	0.879840	0.711777
15	Grid	reg_lambda, aft_loss_distribution_scale	36	0.884314	0.710863
16	Grid	aft_loss_distribution_scale	6	0.881473	0.710385
17	Grid	reg_alpha	6	0.875224	0.709079
18	Grid	min_child_weight, aft_loss_distribution_scale	30	0.881970	0.708373
19	Grid	reg_alpha, reg_lambda	36	0.878228	0.707810
20	Grid	min_child_weight, reg_alpha	30	0.876971	0.707595
21	Grid	reg_lambda	6	0.876496	0.703961
22	Baseline	None (all defaults)	1	0.874224	0.702230
23	Grid	min_child_weight	5	0.875278	0.692765
Dataset H3K27ac-H3K4me3_TDHAM_BP, logistic distribution					
0	Grid	max_depth, aft_loss_distribution_scale	54	0.883494	0.736861
1	Grid	learning_rate, reg_alpha	24	0.881191	0.734715
2	Grid	max_depth, reg_lambda	54	0.883060	0.729967
3	Grid	min_child_weight, reg_lambda	30	0.881664	0.726794
4	Grid	reg_lambda	6	0.880549	0.724545
5	Grid	reg_alpha, aft_loss_distribution_scale	36	0.881595	0.722549
6	Grid	learning_rate, reg_lambda	24	0.883224	0.722238
7	Grid	learning_rate, min_child_weight	20	0.880681	0.720969
8	Grid	learning_rate, aft_loss_distribution_scale	24	0.883757	0.719983
9	Grid	learning_rate	4	0.879262	0.718029
10	Random	All six	1000	0.887514	0.717477
11	Grid	aft_loss_distribution_scale	6	0.880184	0.717038
12	Grid	reg_alpha, reg_lambda	36	0.881584	0.716748
13	Random	All six	100	0.883953	0.716571
14	Grid	learning_rate, max_depth	36	0.883329	0.716231
15	Grid	reg_lambda, aft_loss_distribution_scale	36	0.883250	0.715480
16	Grid	max_depth, reg_alpha	54	0.837738	0.714114
17	Grid	max_depth, min_child_weight	45	0.818940	0.713289

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
18	Grid	reg_alpha	6	0.835391	0.713041
19	Grid	min_child_weight, aft_loss_distribution_scale	30	0.882020	0.711488
20	Grid	min_child_weight, reg_alpha	30	0.837452	0.708518
21	Grid	min_child_weight	5	0.817225	0.707281
22	Baseline	None (all defaults)	1	0.812914	0.698842
23	Grid	max_depth	9	0.816306	0.686587
Dataset H3K27ac-H3K4me3_TDHAM_BP, extreme distribution					
0	Grid	max_depth, reg_lambda	54	0.878618	0.724906
1	Grid	min_child_weight, reg_alpha	30	0.876197	0.723198
2	Random	All six	1000	0.881604	0.717300
3	Grid	max_depth, reg_alpha	54	0.878082	0.715187
4	Grid	learning_rate, reg_lambda	24	0.877669	0.711786
5	Grid	min_child_weight	5	0.873195	0.711241
6	Grid	reg_alpha, reg_lambda	36	0.877408	0.710815
7	Grid	reg_alpha	6	0.874107	0.709582
8	Grid	min_child_weight, reg_lambda	30	0.876320	0.708710
9	Grid	min_child_weight, aft_loss_distribution_scale	30	0.880395	0.706504
10	Baseline	None (all defaults)	1	0.869358	0.705427
11	Grid	reg_lambda, aft_loss_distribution_scale	36	0.880276	0.704131
12	Grid	learning_rate, max_depth	36	0.877551	0.703992
13	Grid	learning_rate, min_child_weight	20	0.877098	0.703003
14	Grid	reg_lambda	6	0.873835	0.702990
15	Grid	max_depth, aft_loss_distribution_scale	54	0.880390	0.702945
16	Grid	learning_rate	4	0.875194	0.701294
17	Grid	max_depth	9	0.876619	0.701220
18	Grid	aft_loss_distribution_scale	6	0.879035	0.701024
19	Grid	learning_rate, aft_loss_distribution_scale	24	0.879532	0.700521
20	Random	All six	100	0.879358	0.698957
21	Grid	reg_alpha, aft_loss_distribution_scale	36	0.879035	0.698509
22	Grid	max_depth, min_child_weight	45	0.876620	0.698202
23	Grid	learning_rate, reg_alpha	24	0.875916	0.697114
Dataset H3K27ac_TDHSome, normal distribution					
0	Grid	min_child_weight	5	0.544659	0.355263
1	Grid	min_child_weight, reg_alpha	30	0.563655	0.322368
2	Grid	reg_alpha	6	0.548655	0.315789
3	Grid	reg_alpha, reg_lambda	36	0.570076	0.315789
4	Random	All six	1000	0.609261	0.309211
5	Grid	reg_lambda	6	0.556364	0.309211
6	Grid	learning_rate, reg_alpha	24	0.565663	0.309211
7	Grid	min_child_weight, reg_lambda	30	0.574413	0.309211
8	Grid	max_depth	9	0.548447	0.302632
9	Grid	max_depth, reg_alpha	54	0.567955	0.302632
10	Grid	min_child_weight, aft_loss_distribution_scale	30	0.570795	0.302632
11	Grid	learning_rate, min_child_weight	20	0.557746	0.296053
12	Grid	reg_lambda, aft_loss_distribution_scale	36	0.587462	0.296053
13	Grid	max_depth, min_child_weight	45	0.565947	0.289474
14	Grid	reg_alpha, aft_loss_distribution_scale	36	0.572708	0.289474
15	Baseline	None (all defaults)	1	0.522481	0.282895
16	Grid	learning_rate	4	0.540284	0.282895

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
17	Grid	learning_rate, max_depth	36	0.570530	0.276316
18	Random	All six	100	0.598258	0.250000
19	Grid	aft_loss_distribution_scale	6	0.559167	0.250000
20	Grid	learning_rate, aft_loss_distribution_scale	24	0.565795	0.243421
21	Grid	learning_rate, reg_lambda	24	0.586174	0.236842
22	Grid	max_depth, aft_loss_distribution_scale	54	0.584716	0.230263
23	Grid	max_depth, reg_lambda	54	0.568201	0.217105
Dataset H3K27ac_TDH_some, logistic distribution					
0	Grid	min_child_weight, reg_lambda	30	0.564413	0.375000
1	Grid	reg_lambda, aft_loss_distribution_scale	36	0.585076	0.375000
2	Grid	min_child_weight, reg_alpha	30	0.562708	0.361842
3	Grid	min_child_weight	5	0.558163	0.348684
4	Grid	reg_lambda	6	0.556913	0.335526
5	Grid	learning_rate, min_child_weight	20	0.566042	0.328947
6	Random	All six	100	0.597633	0.309211
7	Grid	learning_rate	4	0.562292	0.302632
8	Grid	reg_alpha	6	0.555208	0.302632
9	Grid	learning_rate, reg_alpha	24	0.562292	0.302632
10	Random	All six	1000	0.601174	0.296053
11	Grid	max_depth	9	0.562083	0.289474
12	Grid	aft_loss_distribution_scale	6	0.562121	0.289474
13	Grid	max_depth, reg_alpha	54	0.565000	0.289474
14	Grid	min_child_weight, aft_loss_distribution_scale	30	0.576458	0.289474
15	Grid	reg_alpha, reg_lambda	36	0.570417	0.289474
16	Grid	reg_alpha, aft_loss_distribution_scale	36	0.577955	0.289474
17	Grid	learning_rate, aft_loss_distribution_scale	24	0.574621	0.282895
18	Grid	max_depth, min_child_weight	45	0.579205	0.282895
19	Grid	learning_rate, reg_lambda	24	0.576875	0.276316
20	Grid	max_depth, reg_lambda	54	0.572083	0.276316
21	Baseline	None (all defaults)	1	0.550663	0.269737
22	Grid	learning_rate, max_depth	36	0.569583	0.256579
23	Grid	max_depth, aft_loss_distribution_scale	54	0.578750	0.250000
Dataset H3K27ac_TDH_some, extreme distribution					
0	Grid	max_depth, reg_alpha	54	0.569413	0.440789
1	Grid	max_depth	9	0.555208	0.427632
2	Grid	max_depth, min_child_weight	45	0.564792	0.427632
3	Grid	min_child_weight	5	0.553788	0.421053
4	Grid	learning_rate, max_depth	36	0.570208	0.407895
5	Grid	min_child_weight, reg_alpha	30	0.566042	0.401316
6	Grid	learning_rate, min_child_weight	20	0.570208	0.368421
7	Grid	min_child_weight, aft_loss_distribution_scale	30	0.583087	0.368421
8	Baseline	None (all defaults)	1	0.526913	0.355263
9	Grid	learning_rate, reg_alpha	24	0.566667	0.355263
10	Grid	reg_lambda, aft_loss_distribution_scale	36	0.602670	0.355263
11	Grid	reg_alpha	6	0.555455	0.348684
12	Grid	learning_rate	4	0.560208	0.342105
13	Grid	reg_lambda	6	0.569792	0.342105
14	Grid	max_depth, aft_loss_distribution_scale	54	0.586629	0.328947
15	Grid	learning_rate, reg_lambda	24	0.578958	0.315789

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
16	Random	All six	1000	0.617254	0.302632
17	Random	All six	100	0.613712	0.302632
18	Grid	max_depth, reg_lambda	54	0.598125	0.302632
19	Grid	reg_alpha, reg_lambda	36	0.593296	0.256579
20	Grid	aft_loss_distribution_scale	6	0.576212	0.250000
21	Grid	learning_rate, aft_loss_distribution_scale	24	0.576212	0.250000
22	Grid	min_child_weight, reg_lambda	30	0.597330	0.243421
23	Grid	reg_alpha, aft_loss_distribution_scale	36	0.589546	0.223684
Dataset H3K27me3_RL_cancer, normal distribution					
0	Grid	max_depth, reg_lambda	54	0.646010	0.444079
1	Random	All six	1000	0.675466	0.421053
2	Grid	reg_lambda, aft_loss_distribution_scale	36	0.646969	0.388158
3	Random	All six	100	0.655305	0.375000
4	Grid	min_child_weight, reg_lambda	30	0.623847	0.375000
5	Grid	min_child_weight	5	0.615952	0.358553
6	Grid	learning_rate, reg_alpha	24	0.627534	0.355263
7	Grid	min_child_weight, reg_alpha	30	0.633020	0.351974
8	Grid	learning_rate, reg_lambda	24	0.639905	0.351974
9	Grid	reg_lambda	6	0.614571	0.348684
10	Grid	reg_alpha	6	0.622270	0.345395
11	Grid	learning_rate, min_child_weight	20	0.623352	0.332237
12	Grid	max_depth, aft_loss_distribution_scale	54	0.642759	0.332237
13	Baseline	None (all defaults)	1	0.609307	0.332237
14	Grid	min_child_weight, aft_loss_distribution_scale	30	0.634842	0.328947
15	Grid	learning_rate	4	0.620126	0.319079
16	Grid	max_depth, reg_alpha	54	0.644709	0.319079
17	Grid	reg_alpha, reg_lambda	36	0.631237	0.309211
18	Grid	max_depth	9	0.626917	0.305921
19	Grid	max_depth, min_child_weight	45	0.629549	0.305921
20	Grid	learning_rate, aft_loss_distribution_scale	24	0.631610	0.302632
21	Grid	reg_alpha, aft_loss_distribution_scale	36	0.629485	0.302632
22	Grid	learning_rate, max_depth	36	0.639296	0.296053
23	Grid	aft_loss_distribution_scale	6	0.623847	0.276316
Dataset H3K27me3_RL_cancer, logistic distribution					
0	Grid	aft_loss_distribution_scale	6	0.623945	0.437500
1	Grid	min_child_weight, aft_loss_distribution_scale	30	0.623945	0.437500
2	Grid	reg_alpha	6	0.615499	0.427632
3	Grid	learning_rate, reg_lambda	24	0.633127	0.424342
4	Grid	min_child_weight, reg_alpha	30	0.620984	0.421053
5	Grid	max_depth, aft_loss_distribution_scale	54	0.634811	0.411184
6	Grid	min_child_weight	5	0.614923	0.411184
7	Grid	learning_rate, reg_alpha	24	0.631806	0.401316
8	Grid	min_child_weight, reg_lambda	30	0.622968	0.398026
9	Grid	reg_alpha, aft_loss_distribution_scale	36	0.636061	0.394737
10	Baseline	None (all defaults)	1	0.609154	0.391447
11	Grid	reg_lambda	6	0.615561	0.384868
12	Grid	learning_rate, max_depth	36	0.646977	0.384868
13	Grid	reg_alpha, reg_lambda	36	0.624317	0.384868
14	Grid	max_depth	9	0.642827	0.384868

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
15	Grid	max_depth, min_child_weight	45	0.642827	0.384868
16	Grid	max_depth, reg_alpha	54	0.644679	0.384868
17	Grid	max_depth, reg_lambda	54	0.642827	0.384868
18	Random	All six	1000	0.659736	0.378289
19	Grid	learning_rate	4	0.625585	0.378289
20	Grid	learning_rate, min_child_weight	20	0.628217	0.371711
21	Grid	reg_lambda, aft_loss_distribution_scale	36	0.644768	0.371711
22	Random	All six	100	0.645898	0.358553
23	Grid	learning_rate, aft_loss_distribution_scale	24	0.642394	0.358553
Dataset H3K27me3_RL_cancer, extreme distribution					
0	Grid	learning_rate, aft_loss_distribution_scale	24	0.642351	0.460526
1	Grid	max_depth, aft_loss_distribution_scale	54	0.649722	0.460526
2	Grid	reg_alpha, reg_lambda	36	0.635641	0.460526
3	Grid	min_child_weight, reg_alpha	30	0.632692	0.453947
4	Baseline	None (all defaults)	1	0.606034	0.450658
5	Grid	reg_alpha	6	0.629199	0.447368
6	Grid	aft_loss_distribution_scale	6	0.629934	0.447368
7	Grid	min_child_weight, aft_loss_distribution_scale	30	0.639285	0.447368
8	Grid	reg_alpha, aft_loss_distribution_scale	36	0.647543	0.440789
9	Grid	reg_lambda	6	0.626211	0.434211
10	Grid	reg_lambda, aft_loss_distribution_scale	36	0.643589	0.430921
11	Grid	learning_rate	4	0.618405	0.421053
12	Grid	max_depth	9	0.630837	0.421053
13	Grid	min_child_weight	5	0.622225	0.414474
14	Random	All six	1000	0.653055	0.414474
15	Random	All six	100	0.644059	0.411184
16	Grid	max_depth, reg_alpha	54	0.637652	0.407895
17	Grid	max_depth, min_child_weight	45	0.634170	0.407895
18	Grid	max_depth, reg_lambda	54	0.637635	0.407895
19	Grid	learning_rate, max_depth	36	0.632575	0.401316
20	Grid	learning_rate, reg_alpha	24	0.631050	0.401316
21	Grid	learning_rate, min_child_weight	20	0.628988	0.394737
22	Grid	min_child_weight, reg_lambda	30	0.636605	0.384868
23	Grid	learning_rate, reg_lambda	24	0.628063	0.348684
Dataset H3K27me3_TDH_some, normal distribution					
0	Grid	max_depth, reg_alpha	54	0.572917	0.476190
1	Grid	max_depth, reg_lambda	54	0.581845	0.464286
2	Grid	learning_rate, reg_alpha	24	0.590774	0.455357
3	Random	All six	100	0.594643	0.434524
4	Grid	max_depth	9	0.547917	0.434524
5	Baseline	None (all defaults)	1	0.539583	0.413690
6	Grid	min_child_weight	5	0.539583	0.413690
7	Grid	reg_alpha	6	0.547917	0.413690
8	Grid	max_depth, min_child_weight	45	0.578869	0.413690
9	Random	All six	1000	0.633036	0.398810
10	Grid	learning_rate, max_depth	36	0.575000	0.398810
11	Grid	reg_alpha, reg_lambda	36	0.581845	0.398810
12	Grid	min_child_weight, reg_alpha	30	0.556250	0.392857
13	Grid	learning_rate, aft_loss_distribution_scale	24	0.613393	0.383929

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
14	Grid	reg_lambda, aft_loss_distribution_scale	36	0.587202	0.383929
15	Grid	learning_rate, reg_lambda	24	0.598512	0.377976
16	Grid	learning_rate	4	0.558631	0.372024
17	Grid	max_depth, aft_loss_distribution_scale	54	0.625595	0.363095
18	Grid	min_child_weight, aft_loss_distribution_scale	30	0.580060	0.363095
19	Grid	reg_alpha, aft_loss_distribution_scale	36	0.608631	0.363095
20	Grid	learning_rate, min_child_weight	20	0.565774	0.351190
21	Grid	reg_lambda	6	0.558036	0.336310
22	Grid	aft_loss_distribution_scale	6	0.578869	0.321429
23	Grid	min_child_weight, reg_lambda	30	0.575893	0.315476
Dataset H3K27me3_TDH_some, logistic distribution					
0	Grid	reg_alpha	6	0.555060	0.532738
1	Grid	reg_alpha, reg_lambda	36	0.564583	0.532738
2	Grid	min_child_weight, reg_alpha	30	0.557441	0.511905
3	Grid	min_child_weight	5	0.544941	0.482143
4	Grid	min_child_weight, reg_lambda	30	0.550893	0.461310
5	Grid	min_child_weight, aft_loss_distribution_scale	30	0.598810	0.446429
6	Grid	learning_rate, reg_lambda	24	0.579464	0.440476
7	Grid	max_depth, min_child_weight	45	0.578572	0.425595
8	Grid	reg_alpha, aft_loss_distribution_scale	36	0.594048	0.425595
9	Grid	learning_rate, min_child_weight	20	0.575893	0.419643
10	Grid	reg_lambda, aft_loss_distribution_scale	36	0.587798	0.419643
11	Grid	reg_lambda	6	0.550893	0.419643
12	Grid	learning_rate, max_depth	36	0.583036	0.419643
13	Grid	learning_rate, aft_loss_distribution_scale	24	0.588095	0.404762
14	Grid	max_depth, aft_loss_distribution_scale	54	0.588095	0.404762
15	Random	All six	1000	0.622321	0.401786
16	Baseline	None (all defaults)	1	0.541369	0.398810
17	Grid	max_depth	9	0.557143	0.398810
18	Grid	aft_loss_distribution_scale	6	0.571429	0.383929
19	Grid	learning_rate, reg_alpha	24	0.580952	0.383929
20	Grid	max_depth, reg_lambda	54	0.572619	0.383929
21	Grid	learning_rate	4	0.567560	0.377976
22	Grid	max_depth, reg_alpha	54	0.563095	0.342262
23	Random	All six	100	0.616964	0.321429
Dataset H3K27me3_TDH_some, extreme distribution					
0	Baseline	None (all defaults)	1	0.552083	0.523810
1	Grid	min_child_weight, reg_lambda	30	0.598512	0.523810
2	Grid	min_child_weight	5	0.582143	0.482143
3	Grid	learning_rate, min_child_weight	20	0.589286	0.482143
4	Grid	reg_lambda, aft_loss_distribution_scale	36	0.629464	0.476190
5	Grid	reg_lambda	6	0.591369	0.467262
6	Grid	max_depth	9	0.566369	0.461310
7	Grid	max_depth, min_child_weight	45	0.597619	0.461310
8	Grid	learning_rate, reg_alpha	24	0.583036	0.455357
9	Grid	min_child_weight, reg_alpha	30	0.582143	0.446429
10	Grid	reg_alpha, reg_lambda	36	0.604464	0.425595
11	Grid	min_child_weight, aft_loss_distribution_scale	30	0.609524	0.404762
12	Grid	reg_alpha	6	0.561607	0.383929

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
13	Grid	learning_rate, reg_lambda	24	0.612798	0.383929
14	Grid	learning_rate, aft_loss_distribution_scale	24	0.611012	0.383929
15	Grid	max_depth, reg_lambda	54	0.598512	0.363095
16	Random	All six	1000	0.665774	0.357143
17	Grid	aft_loss_distribution_scale	6	0.596429	0.342262
18	Grid	reg_alpha, aft_loss_distribution_scale	36	0.597322	0.342262
19	Grid	max_depth, aft_loss_distribution_scale	54	0.634524	0.327381
20	Grid	max_depth, reg_alpha	54	0.588988	0.321429
21	Grid	learning_rate	4	0.572619	0.321429
22	Grid	learning_rate, max_depth	36	0.596131	0.294643
23	Random	All six	100	0.602381	0.258929
Dataset H3K36me3_AM_immune, normal distribution					
0	Grid	min_child_weight, reg_lambda	30	0.939712	0.881349
1	Random	All six	100	0.945227	0.872817
2	Grid	max_depth, reg_alpha	54	0.942738	0.872222
3	Grid	learning_rate, max_depth	36	0.940332	0.867659
4	Grid	learning_rate, min_child_weight	20	0.940506	0.864683
5	Grid	aft_loss_distribution_scale	6	0.938072	0.863889
6	Grid	min_child_weight, aft_loss_distribution_scale	30	0.938818	0.863889
7	Grid	reg_alpha, aft_loss_distribution_scale	36	0.938072	0.863889
8	Grid	max_depth	9	0.939503	0.863095
9	Grid	learning_rate, aft_loss_distribution_scale	24	0.940347	0.860714
10	Grid	learning_rate, reg_alpha	24	0.938024	0.859722
11	Random	All six	1000	0.948306	0.859722
12	Grid	reg_lambda, aft_loss_distribution_scale	36	0.940513	0.858135
13	Grid	max_depth, reg_lambda	54	0.941155	0.857937
14	Baseline	None (all defaults)	1	0.923607	0.856548
15	Grid	max_depth, min_child_weight	45	0.942839	0.850992
16	Grid	reg_alpha	6	0.934849	0.850198
17	Grid	reg_lambda	6	0.933234	0.844246
18	Grid	reg_alpha, reg_lambda	36	0.938024	0.840675
19	Grid	max_depth, aft_loss_distribution_scale	54	0.943565	0.839683
20	Grid	learning_rate	4	0.930826	0.838492
21	Grid	min_child_weight	5	0.935744	0.828968
22	Grid	min_child_weight, reg_alpha	30	0.938077	0.822421
23	Grid	learning_rate, reg_lambda	24	0.938822	0.822024
Dataset H3K36me3_AM_immune, logistic distribution					
0	Random	All six	1000	0.945865	0.909921
1	Grid	reg_lambda, aft_loss_distribution_scale	36	0.938072	0.890873
2	Grid	reg_lambda	6	0.935093	0.875397
3	Grid	reg_alpha, reg_lambda	36	0.935839	0.872421
4	Grid	max_depth, reg_lambda	54	0.935839	0.868254
5	Grid	learning_rate	4	0.932492	0.867262
6	Grid	learning_rate, reg_alpha	24	0.932492	0.867262
7	Grid	learning_rate, min_child_weight	20	0.932503	0.866667
8	Grid	learning_rate, max_depth	36	0.936521	0.856349
9	Grid	min_child_weight, aft_loss_distribution_scale	30	0.934246	0.851786
10	Random	All six	100	0.939702	0.819444
11	Grid	min_child_weight, reg_lambda	30	0.935839	0.655159

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
12	Grid	aft_loss_distribution_scale	6	0.933387	0.653373
13	Grid	reg_alpha, aft_loss_distribution_scale	36	0.934235	0.651389
14	Grid	learning_rate, aft_loss_distribution_scale	24	0.941100	0.627183
15	Grid	max_depth, aft_loss_distribution_scale	54	0.939963	0.445437
16	Grid	learning_rate, reg_lambda	24	0.941930	0.441468
17	Grid	reg_alpha	6	0.689409	0.404762
18	Grid	min_child_weight, reg_alpha	30	0.692814	0.404762
19	Grid	max_depth, reg_alpha	54	0.689409	0.392857
20	Baseline	None (all defaults)	1	0.013979	0.000000
21	Grid	max_depth	9	0.013979	0.000000
22	Grid	min_child_weight	5	0.013979	0.000000
23	Grid	max_depth, min_child_weight	45	0.013979	0.000000
Dataset H3K36me3_AM_immune, extreme distribution					
0	Baseline	None (all defaults)	1	0.936525	0.902183
1	Grid	learning_rate, reg_lambda	24	0.941356	0.900198
2	Random	All six	100	0.946934	0.898413
3	Grid	min_child_weight	5	0.936525	0.894444
4	Grid	learning_rate	4	0.940548	0.893849
5	Grid	reg_alpha	6	0.938007	0.893254
6	Grid	aft_loss_distribution_scale	6	0.941194	0.890675
7	Grid	reg_lambda	6	0.938167	0.890278
8	Grid	min_child_weight, reg_alpha	30	0.940456	0.889683
9	Grid	min_child_weight, reg_lambda	30	0.940568	0.888095
10	Grid	learning_rate, max_depth	36	0.942076	0.885516
11	Grid	max_depth, aft_loss_distribution_scale	54	0.945163	0.884127
12	Grid	min_child_weight, aft_loss_distribution_scale	30	0.943575	0.883532
13	Random	All six	1000	0.950760	0.882937
14	Grid	learning_rate, min_child_weight	20	0.941308	0.880754
15	Grid	learning_rate, reg_alpha	24	0.942823	0.880159
16	Grid	reg_alpha, reg_lambda	36	0.942876	0.878968
17	Grid	reg_lambda, aft_loss_distribution_scale	36	0.945163	0.878770
18	Grid	reg_alpha, aft_loss_distribution_scale	36	0.945163	0.874008
19	Grid	max_depth, reg_lambda	54	0.942938	0.873810
20	Grid	max_depth	9	0.939696	0.867857
21	Grid	learning_rate, aft_loss_distribution_scale	24	0.943575	0.864881
22	Grid	max_depth, min_child_weight	45	0.940518	0.861111
23	Grid	max_depth, reg_alpha	54	0.942137	0.857937
Dataset H3K36me3_TDH_ENCODE, normal distribution					
0	Grid	reg_alpha, reg_lambda	36	0.717657	0.586538
1	Grid	learning_rate, reg_lambda	24	0.719895	0.576923
2	Grid	min_child_weight, reg_alpha	30	0.721049	0.576923
3	Grid	reg_alpha, aft_loss_distribution_scale	36	0.720804	0.567308
4	Grid	min_child_weight, aft_loss_distribution_scale	30	0.728986	0.567308
5	Grid	learning_rate, min_child_weight	20	0.709511	0.557692
6	Grid	max_depth, aft_loss_distribution_scale	54	0.724895	0.557692
7	Random	All six	100	0.758672	0.538462
8	Grid	aft_loss_distribution_scale	6	0.702867	0.528846
9	Grid	reg_lambda	6	0.704511	0.528846
10	Grid	min_child_weight	5	0.709511	0.519231

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
11	Grid	max_depth, reg_lambda	54	0.713811	0.519231
12	Random	All six	1000	0.775909	0.500000
13	Grid	learning_rate, aft_loss_distribution_scale	24	0.715804	0.500000
14	Grid	reg_lambda, aft_loss_distribution_scale	36	0.732797	0.480769
15	Grid	min_child_weight, reg_lambda	30	0.714511	0.471154
16	Grid	learning_rate, max_depth	36	0.696329	0.461538
17	Baseline	None (all defaults)	1	0.651574	0.451923
18	Grid	learning_rate	4	0.672028	0.442308
19	Grid	reg_alpha	6	0.685874	0.442308
20	Grid	max_depth	9	0.675874	0.403846
21	Grid	max_depth, reg_alpha	54	0.699720	0.394231
22	Grid	learning_rate, reg_alpha	24	0.694021	0.375000
23	Grid	max_depth, min_child_weight	45	0.717203	0.326923
Dataset H3K36me3_TDH_ENCODE, logistic distribution					
0	Grid	min_child_weight, reg_lambda	30	0.737588	0.586538
1	Grid	reg_lambda	6	0.694266	0.557692
2	Grid	aft_loss_distribution_scale	6	0.716049	0.557692
3	Grid	max_depth	9	0.690420	0.548077
4	Grid	reg_alpha, reg_lambda	36	0.713357	0.548077
5	Grid	reg_alpha, aft_loss_distribution_scale	36	0.728741	0.548077
6	Baseline	None (all defaults)	1	0.667727	0.538462
7	Random	All six	100	0.747133	0.538462
8	Grid	learning_rate, reg_lambda	24	0.729895	0.538462
9	Grid	max_depth, reg_lambda	54	0.707658	0.528846
10	Grid	learning_rate, aft_loss_distribution_scale	24	0.729895	0.509615
11	Grid	max_depth, reg_alpha	54	0.690420	0.509615
12	Grid	min_child_weight, aft_loss_distribution_scale	30	0.723741	0.500000
13	Grid	max_depth, aft_loss_distribution_scale	54	0.721049	0.490385
14	Grid	reg_lambda, aft_loss_distribution_scale	36	0.726049	0.490385
15	Grid	reg_alpha	6	0.668636	0.442308
16	Grid	max_depth, min_child_weight	45	0.717658	0.423077
17	Random	All six	1000	0.759371	0.413462
18	Grid	learning_rate, max_depth	36	0.704266	0.394231
19	Grid	min_child_weight	5	0.717658	0.394231
20	Grid	learning_rate, min_child_weight	20	0.717658	0.394231
21	Grid	min_child_weight, reg_alpha	30	0.722203	0.394231
22	Grid	learning_rate, reg_alpha	24	0.691574	0.384615
23	Grid	learning_rate	4	0.673182	0.365385
Dataset H3K36me3_TDH_ENCODE, extreme distribution					
0	Baseline	None (all defaults)	1	0.682273	0.548077
1	Grid	max_depth, reg_lambda	54	0.731748	0.538462
2	Grid	max_depth	9	0.700210	0.519231
3	Grid	reg_lambda	6	0.727203	0.519231
4	Grid	min_child_weight, aft_loss_distribution_scale	30	0.736049	0.519231
5	Grid	reg_alpha, reg_lambda	36	0.731748	0.519231
6	Random	All six	100	0.736049	0.509615
7	Grid	aft_loss_distribution_scale	6	0.731049	0.509615
8	Grid	reg_alpha, aft_loss_distribution_scale	36	0.736049	0.509615
9	Grid	learning_rate	4	0.699965	0.500000

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
10	Grid	learning_rate, min_child_weight	20	0.717902	0.500000
11	Grid	min_child_weight	5	0.700420	0.490385
12	Grid	learning_rate, reg_lambda	24	0.732203	0.480769
13	Grid	learning_rate, reg_alpha	24	0.709965	0.471154
14	Grid	max_depth, min_child_weight	45	0.709511	0.471154
15	Grid	max_depth, reg_alpha	54	0.704511	0.471154
16	Grid	reg_lambda, aft_loss_distribution_scale	36	0.736049	0.471154
17	Grid	learning_rate, max_depth	36	0.704965	0.461538
18	Grid	learning_rate, aft_loss_distribution_scale	24	0.731049	0.442308
19	Grid	max_depth, aft_loss_distribution_scale	54	0.735595	0.432692
20	Random	All six	1000	0.748986	0.403846
21	Grid	reg_alpha	6	0.699965	0.384615
22	Grid	min_child_weight, reg_lambda	30	0.731748	0.384615
23	Grid	min_child_weight, reg_alpha	30	0.714266	0.336538
Dataset H3K36me3_TDH_immune, normal distribution					
0	Grid	min_child_weight, reg_alpha	30	0.925321	0.761905
1	Grid	max_depth, min_child_weight	45	0.937180	0.738095
2	Baseline	None (all defaults)	1	0.870513	0.726190
3	Grid	reg_alpha, reg_lambda	36	0.920833	0.702381
4	Grid	reg_lambda	6	0.912500	0.690476
5	Grid	learning_rate, min_child_weight	20	0.941026	0.690476
6	Grid	min_child_weight, aft_loss_distribution_scale	30	0.940705	0.690476
7	Grid	reg_lambda, aft_loss_distribution_scale	36	0.936859	0.690476
8	Grid	max_depth	9	0.901603	0.678571
9	Grid	reg_alpha	6	0.885577	0.678571
10	Grid	reg_alpha, aft_loss_distribution_scale	36	0.928526	0.678571
11	Grid	max_depth, reg_lambda	54	0.932372	0.666667
12	Grid	max_depth, aft_loss_distribution_scale	54	0.932692	0.666667
13	Grid	learning_rate	4	0.936859	0.654762
14	Grid	aft_loss_distribution_scale	6	0.920833	0.654762
15	Grid	min_child_weight	5	0.921154	0.642857
16	Grid	learning_rate, reg_lambda	24	0.944551	0.642857
17	Grid	min_child_weight, reg_lambda	30	0.941026	0.642857
18	Grid	learning_rate, max_depth	36	0.941026	0.630952
19	Grid	learning_rate, reg_alpha	24	0.937180	0.630952
20	Grid	max_depth, reg_alpha	54	0.913141	0.630952
21	Random	All six	100	0.941026	0.607143
22	Random	All six	1000	0.953205	0.571429
23	Grid	learning_rate, aft_loss_distribution_scale	24	0.944872	0.511905
Dataset H3K36me3_TDH_immune, logistic distribution					
0	Grid	reg_alpha, aft_loss_distribution_scale	36	0.941026	0.726190
1	Grid	learning_rate, min_child_weight	20	0.924359	0.714286
2	Grid	learning_rate, reg_alpha	24	0.927885	0.714286
3	Grid	max_depth, aft_loss_distribution_scale	54	0.941026	0.714286
4	Grid	reg_alpha	6	0.924038	0.702381
5	Grid	min_child_weight, aft_loss_distribution_scale	30	0.937180	0.702381
6	Grid	max_depth	9	0.924359	0.690476
7	Grid	min_child_weight, reg_alpha	30	0.924038	0.690476
8	Grid	learning_rate, max_depth	36	0.928205	0.678571

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
9	Grid	min_child_weight, reg_lambda	30	0.927885	0.678571
10	Grid	max_depth, min_child_weight	45	0.924359	0.666667
11	Baseline	None (all defaults)	1	0.920192	0.666667
12	Grid	learning_rate	4	0.920192	0.666667
13	Grid	min_child_weight	5	0.920192	0.666667
14	Grid	max_depth, reg_lambda	54	0.932372	0.654762
15	Grid	reg_lambda	6	0.927885	0.654762
16	Grid	aft_loss_distribution_scale	6	0.936859	0.654762
17	Grid	max_depth, reg_alpha	54	0.932051	0.654762
18	Grid	reg_alpha, reg_lambda	36	0.927885	0.654762
19	Grid	learning_rate, aft_loss_distribution_scale	24	0.940705	0.630952
20	Grid	reg_lambda, aft_loss_distribution_scale	36	0.941026	0.630952
21	Random	All six	1000	0.956731	0.464286
22	Random	All six	100	0.948398	0.440476
23	Grid	learning_rate, reg_lambda	24	0.932372	0.309524
Dataset H3K36me3_TDH_immune, extreme distribution					
0	Random	All six	1000	0.957051	0.773810
1	Grid	learning_rate, aft_loss_distribution_scale	24	0.940385	0.738095
2	Random	All six	100	0.949039	0.726190
3	Grid	reg_lambda	6	0.932372	0.726190
4	Grid	reg_alpha, aft_loss_distribution_scale	36	0.943910	0.726190
5	Grid	learning_rate, reg_alpha	24	0.932692	0.702381
6	Grid	learning_rate, reg_lambda	24	0.936539	0.690476
7	Grid	reg_alpha, reg_lambda	36	0.940385	0.690476
8	Grid	reg_lambda, aft_loss_distribution_scale	36	0.949039	0.690476
9	Grid	max_depth, reg_lambda	54	0.936859	0.678571
10	Grid	min_child_weight, reg_lambda	30	0.948398	0.666667
11	Grid	learning_rate, max_depth	36	0.941026	0.654762
12	Grid	learning_rate, min_child_weight	20	0.944872	0.642857
13	Grid	min_child_weight, aft_loss_distribution_scale	30	0.944551	0.607143
14	Grid	aft_loss_distribution_scale	6	0.928846	0.607143
15	Grid	max_depth, min_child_weight	45	0.937500	0.607143
16	Grid	max_depth, reg_alpha	54	0.936539	0.607143
17	Grid	reg_alpha	6	0.932372	0.595238
18	Grid	max_depth, aft_loss_distribution_scale	54	0.948398	0.595238
19	Grid	min_child_weight, reg_alpha	30	0.944872	0.595238
20	Baseline	None (all defaults)	1	0.916346	0.571429
21	Grid	learning_rate	4	0.932372	0.571429
22	Grid	max_depth	9	0.928846	0.571429
23	Grid	min_child_weight	5	0.925000	0.535714
Dataset H3K36me3_TDH_other, normal distribution					
0	Grid	min_child_weight	5	0.887143	0.765625
1	Random	All six	100	0.926905	0.734375
2	Grid	reg_alpha, reg_lambda	36	0.928095	0.734375
3	Baseline	None (all defaults)	1	0.885952	0.703125
4	Grid	max_depth, min_child_weight	45	0.918571	0.703125
5	Grid	min_child_weight, reg_alpha	30	0.917381	0.687500
6	Random	All six	1000	0.943571	0.671875
7	Grid	reg_alpha	6	0.917381	0.656250

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
8	Grid	learning_rate, max_depth	36	0.918571	0.656250
9	Grid	max_depth, reg_alpha	54	0.918571	0.656250
10	Grid	max_depth	9	0.908571	0.640625
11	Grid	max_depth, reg_lambda	54	0.918571	0.640625
12	Grid	min_child_weight, aft_loss_distribution_scale	30	0.931667	0.546875
13	Grid	learning_rate, reg_alpha	24	0.917381	0.531250
14	Grid	aft_loss_distribution_scale	6	0.931667	0.515625
15	Grid	reg_lambda	6	0.916190	0.484375
16	Grid	max_depth, aft_loss_distribution_scale	54	0.932857	0.484375
17	Grid	min_child_weight, reg_lambda	30	0.916190	0.484375
18	Grid	reg_alpha, aft_loss_distribution_scale	36	0.932857	0.484375
19	Grid	learning_rate	4	0.909048	0.468750
20	Grid	learning_rate, reg_lambda	24	0.917381	0.468750
21	Grid	learning_rate, aft_loss_distribution_scale	24	0.931667	0.468750
22	Grid	learning_rate, min_child_weight	20	0.909048	0.437500
23	Grid	reg_lambda, aft_loss_distribution_scale	36	0.931667	0.421875
Dataset H3K36me3_TDH_other, logistic distribution					
0	Baseline	None (all defaults)	1	0.918571	0.515625
1	Grid	learning_rate	4	0.931667	0.515625
2	Grid	max_depth	9	0.925714	0.515625
3	Grid	min_child_weight	5	0.918571	0.515625
4	Grid	aft_loss_distribution_scale	6	0.919762	0.515625
5	Grid	learning_rate, max_depth	36	0.932857	0.515625
6	Grid	learning_rate, min_child_weight	20	0.931667	0.515625
7	Grid	learning_rate, aft_loss_distribution_scale	24	0.925714	0.515625
8	Grid	max_depth, min_child_weight	45	0.925714	0.515625
9	Grid	max_depth, reg_alpha	54	0.925714	0.515625
10	Grid	max_depth, aft_loss_distribution_scale	54	0.926905	0.515625
11	Grid	min_child_weight, reg_alpha	30	0.924524	0.515625
12	Grid	min_child_weight, reg_lambda	30	0.918571	0.515625
13	Random	All six	100	0.924524	0.500000
14	Grid	reg_alpha	6	0.924524	0.500000
15	Grid	learning_rate, reg_alpha	24	0.931667	0.500000
16	Grid	learning_rate, reg_lambda	24	0.931667	0.484375
17	Grid	reg_alpha, aft_loss_distribution_scale	36	0.925714	0.484375
18	Grid	min_child_weight, aft_loss_distribution_scale	30	0.919762	0.453125
19	Random	All six	1000	0.934048	0.437500
20	Grid	reg_lambda	6	0.918571	0.437500
21	Grid	reg_alpha, reg_lambda	36	0.924524	0.437500
22	Grid	reg_lambda, aft_loss_distribution_scale	36	0.931667	0.437500
23	Grid	max_depth, reg_lambda	54	0.926905	0.406250
Dataset H3K36me3_TDH_other, extreme distribution					
0	Random	All six	1000	0.941190	0.671875
1	Grid	min_child_weight	5	0.909048	0.640625
2	Grid	max_depth, min_child_weight	45	0.909048	0.640625
3	Baseline	None (all defaults)	1	0.886429	0.593750
4	Grid	max_depth	9	0.901905	0.593750
5	Grid	min_child_weight, reg_lambda	30	0.924524	0.578125
6	Grid	reg_alpha	6	0.917381	0.515625

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
7	Grid	learning_rate, aft_loss_distribution_scale	24	0.931667	0.515625
8	Grid	min_child_weight, reg_alpha	30	0.924524	0.515625
9	Random	All six	100	0.940000	0.484375
10	Grid	learning_rate, min_child_weight	20	0.924524	0.484375
11	Grid	learning_rate	4	0.917381	0.468750
12	Grid	reg_lambda	6	0.917381	0.468750
13	Grid	max_depth, reg_lambda	54	0.924524	0.468750
14	Grid	learning_rate, reg_alpha	24	0.917381	0.453125
15	Grid	learning_rate, max_depth	36	0.924524	0.437500
16	Grid	reg_alpha, reg_lambda	36	0.924524	0.437500
17	Grid	learning_rate, reg_lambda	24	0.917381	0.421875
18	Grid	max_depth, reg_alpha	54	0.924524	0.421875
19	Grid	max_depth, aft_loss_distribution_scale	54	0.931667	0.375000
20	Grid	aft_loss_distribution_scale	6	0.931667	0.359375
21	Grid	min_child_weight, aft_loss_distribution_scale	30	0.931667	0.359375
22	Grid	reg_alpha, aft_loss_distribution_scale	36	0.932857	0.359375
23	Grid	reg_lambda, aft_loss_distribution_scale	36	0.931667	0.281250
Dataset simulated.abs, normal distribution					
0	Baseline	None (all defaults)	1	0.0	0.1375
1	Random	All six	1000	0.0	0.1375
2	Random	All six	100	0.0	0.1375
3	Grid	learning_rate	4	0.0	0.1375
4	Grid	max_depth	9	0.0	0.1375
5	Grid	min_child_weight	5	0.0	0.1375
6	Grid	reg_alpha	6	0.0	0.1375
7	Grid	reg_lambda	6	0.0	0.1375
8	Grid	aft_loss_distribution_scale	6	0.0	0.1375
9	Grid	learning_rate, max_depth	36	0.0	0.1375
10	Grid	learning_rate, min_child_weight	20	0.0	0.1375
11	Grid	learning_rate, reg_alpha	24	0.0	0.1375
12	Grid	learning_rate, reg_lambda	24	0.0	0.1375
13	Grid	learning_rate, aft_loss_distribution_scale	24	0.0	0.1375
14	Grid	max_depth, min_child_weight	45	0.0	0.1375
15	Grid	max_depth, reg_alpha	54	0.0	0.1375
16	Grid	max_depth, reg_lambda	54	0.0	0.1375
17	Grid	max_depth, aft_loss_distribution_scale	54	0.0	0.1375
18	Grid	min_child_weight, reg_alpha	30	0.0	0.1375
19	Grid	min_child_weight, reg_lambda	30	0.0	0.1375
20	Grid	min_child_weight, aft_loss_distribution_scale	30	0.0	0.1375
21	Grid	reg_alpha, reg_lambda	36	0.0	0.1375
22	Grid	reg_alpha, aft_loss_distribution_scale	36	0.0	0.1375
23	Grid	reg_lambda, aft_loss_distribution_scale	36	0.0	0.1375
Dataset simulated.abs, logistic distribution					
0	Random	All six	1000	0.875000	0.82500
1	Random	All six	100	0.842187	0.81875
2	Grid	learning_rate, max_depth	36	0.782813	0.76875
3	Grid	learning_rate, aft_loss_distribution_scale	24	0.801562	0.76875
4	Grid	min_child_weight, aft_loss_distribution_scale	30	0.796875	0.75625
5	Grid	max_depth, min_child_weight	45	0.764062	0.73750

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
6	Grid	max_depth, reg_alpha	54	0.743750	0.73750
7	Grid	max_depth, aft_loss_distribution_scale	54	0.765625	0.72500
8	Grid	learning_rate, min_child_weight	20	0.693750	0.72500
9	Grid	max_depth	9	0.731250	0.70000
10	Grid	learning_rate, reg_alpha	24	0.696875	0.70000
11	Grid	max_depth, reg_lambda	54	0.731250	0.70000
12	Grid	min_child_weight, reg_alpha	30	0.685937	0.70000
13	Grid	aft_loss_distribution_scale	6	0.743750	0.69375
14	Grid	reg_alpha, aft_loss_distribution_scale	36	0.751563	0.68750
15	Grid	reg_lambda, aft_loss_distribution_scale	36	0.745313	0.68125
16	Grid	learning_rate	4	0.684375	0.67500
17	Grid	learning_rate, reg_lambda	24	0.734375	0.66250
18	Grid	min_child_weight, reg_lambda	30	0.689063	0.65625
19	Grid	min_child_weight	5	0.676562	0.65000
20	Grid	reg_alpha	6	0.667188	0.63750
21	Baseline	None (all defaults)	1	0.659375	0.62500
22	Grid	reg_alpha, reg_lambda	36	0.689063	0.55625
23	Grid	reg_lambda	6	0.676562	0.54375
Dataset simulated.abs, extreme distribution					
0	Grid	max_depth, aft_loss_distribution_scale	54	0.868750	0.86250
1	Random	All six	1000	0.896875	0.86250
2	Grid	min_child_weight, aft_loss_distribution_scale	30	0.868750	0.85000
3	Random	All six	100	0.868750	0.83750
4	Grid	reg_alpha, aft_loss_distribution_scale	36	0.843750	0.83750
5	Grid	max_depth, min_child_weight	45	0.834375	0.82500
6	Grid	aft_loss_distribution_scale	6	0.820312	0.81875
7	Grid	learning_rate, aft_loss_distribution_scale	24	0.851562	0.81875
8	Grid	learning_rate, min_child_weight	20	0.820312	0.81250
9	Grid	min_child_weight, reg_lambda	30	0.825000	0.81250
10	Grid	min_child_weight	5	0.815625	0.80000
11	Grid	min_child_weight, reg_alpha	30	0.818750	0.79375
12	Grid	max_depth	9	0.740625	0.78750
13	Grid	max_depth, reg_alpha	54	0.753125	0.78750
14	Grid	reg_lambda, aft_loss_distribution_scale	36	0.839063	0.78750
15	Grid	max_depth, reg_lambda	54	0.748437	0.78125
16	Grid	learning_rate, max_depth	36	0.790625	0.76875
17	Grid	learning_rate, reg_lambda	24	0.773438	0.75000
18	Grid	learning_rate	4	0.709375	0.74375
19	Baseline	None (all defaults)	1	0.662500	0.72500
20	Grid	learning_rate, reg_alpha	24	0.715625	0.72500
21	Grid	reg_lambda	6	0.676562	0.68750
22	Grid	reg_alpha, reg_lambda	36	0.690625	0.68125
23	Grid	reg_alpha	6	0.675000	0.67500
Dataset simulated.linear, normal distribution					
0	Baseline	None (all defaults)	1	0.0	0.1
1	Random	All six	1000	0.0	0.1
2	Random	All six	100	0.0	0.1
3	Grid	learning_rate	4	0.0	0.1
4	Grid	max_depth	9	0.0	0.1

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
5	Grid	min_child_weight	5	0.0	0.1
6	Grid	reg_alpha	6	0.0	0.1
7	Grid	reg_lambda	6	0.0	0.1
8	Grid	aft_loss_distribution_scale	6	0.0	0.1
9	Grid	learning_rate, max_depth	36	0.0	0.1
10	Grid	learning_rate, min_child_weight	20	0.0	0.1
11	Grid	learning_rate, reg_alpha	24	0.0	0.1
12	Grid	learning_rate, reg_lambda	24	0.0	0.1
13	Grid	learning_rate, aft_loss_distribution_scale	24	0.0	0.1
14	Grid	max_depth, min_child_weight	45	0.0	0.1
15	Grid	max_depth, reg_alpha	54	0.0	0.1
16	Grid	max_depth, reg_lambda	54	0.0	0.1
17	Grid	max_depth, aft_loss_distribution_scale	54	0.0	0.1
18	Grid	min_child_weight, reg_alpha	30	0.0	0.1
19	Grid	min_child_weight, reg_lambda	30	0.0	0.1
20	Grid	min_child_weight, aft_loss_distribution_scale	30	0.0	0.1
21	Grid	reg_alpha, reg_lambda	36	0.0	0.1
22	Grid	reg_alpha, aft_loss_distribution_scale	36	0.0	0.1
23	Grid	reg_lambda, aft_loss_distribution_scale	36	0.0	0.1
Dataset simulated.linear, logistic distribution					
0	Grid	learning_rate, aft_loss_distribution_scale	24	0.806250	0.83125
1	Random	All six	100	0.871875	0.81875
2	Grid	max_depth, aft_loss_distribution_scale	54	0.864062	0.81875
3	Grid	reg_lambda	6	0.782813	0.81250
4	Random	All six	1000	0.889062	0.80625
5	Grid	aft_loss_distribution_scale	6	0.800000	0.80625
6	Grid	reg_alpha, reg_lambda	36	0.792188	0.80000
7	Grid	max_depth, reg_lambda	54	0.831250	0.79375
8	Grid	min_child_weight, reg_alpha	30	0.825000	0.79375
9	Grid	reg_alpha, aft_loss_distribution_scale	36	0.817187	0.79375
10	Grid	max_depth	9	0.817187	0.78750
11	Grid	learning_rate, max_depth	36	0.817187	0.78750
12	Grid	learning_rate, reg_lambda	24	0.792188	0.78750
13	Grid	min_child_weight	5	0.809375	0.78125
14	Grid	max_depth, reg_alpha	54	0.826562	0.78125
15	Grid	min_child_weight, aft_loss_distribution_scale	30	0.846875	0.78125
16	Grid	max_depth, min_child_weight	45	0.829688	0.77500
17	Grid	reg_lambda, aft_loss_distribution_scale	36	0.840625	0.76875
18	Grid	min_child_weight, reg_lambda	30	0.835938	0.76250
19	Grid	learning_rate, min_child_weight	20	0.817187	0.70625
20	Grid	reg_alpha	6	0.762500	0.70000
21	Grid	learning_rate	4	0.742188	0.68125
22	Grid	learning_rate, reg_alpha	24	0.771875	0.66250
23	Baseline	None (all defaults)	1	0.737500	0.64375
Dataset simulated.linear, extreme distribution					
0	Grid	reg_lambda, aft_loss_distribution_scale	36	0.859375	0.86875
1	Random	All six	1000	0.906250	0.85625
2	Grid	min_child_weight, aft_loss_distribution_scale	30	0.896875	0.85625
3	Grid	min_child_weight, reg_lambda	30	0.870313	0.85000

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
4	Grid	max_depth, aft_loss_distribution_scale	54	0.889062	0.84375
5	Random	All six	100	0.898438	0.83750
6	Grid	reg_alpha, aft_loss_distribution_scale	36	0.862500	0.83125
7	Grid	aft_loss_distribution_scale	6	0.840625	0.82500
8	Grid	min_child_weight, reg_alpha	30	0.862500	0.82500
9	Grid	max_depth	9	0.854688	0.81875
10	Grid	max_depth, min_child_weight	45	0.867188	0.81875
11	Grid	learning_rate, max_depth	36	0.859375	0.81250
12	Grid	learning_rate, reg_alpha	24	0.800000	0.81250
13	Grid	learning_rate, aft_loss_distribution_scale	24	0.856250	0.80000
14	Grid	max_depth, reg_lambda	54	0.871875	0.80000
15	Grid	reg_lambda	6	0.801562	0.79375
16	Grid	learning_rate, reg_lambda	24	0.812500	0.79375
17	Baseline	None (all defaults)	1	0.768750	0.78750
18	Grid	learning_rate	4	0.768750	0.78750
19	Grid	max_depth, reg_alpha	54	0.862500	0.78750
20	Grid	reg_alpha	6	0.795312	0.77500
21	Grid	min_child_weight	5	0.845313	0.76875
22	Grid	learning_rate, min_child_weight	20	0.850000	0.76875
23	Grid	reg_alpha, reg_lambda	36	0.823438	0.76250
Dataset simulated.sin, normal distribution					
0	Baseline	None (all defaults)	1	0.0	0.325
1	Random	All six	1000	0.0	0.325
2	Random	All six	100	0.0	0.325
3	Grid	learning_rate	4	0.0	0.325
4	Grid	max_depth	9	0.0	0.325
5	Grid	min_child_weight	5	0.0	0.325
6	Grid	reg_alpha	6	0.0	0.325
7	Grid	reg_lambda	6	0.0	0.325
8	Grid	aft_loss_distribution_scale	6	0.0	0.325
9	Grid	learning_rate, max_depth	36	0.0	0.325
10	Grid	learning_rate, min_child_weight	20	0.0	0.325
11	Grid	learning_rate, reg_alpha	24	0.0	0.325
12	Grid	learning_rate, reg_lambda	24	0.0	0.325
13	Grid	learning_rate, aft_loss_distribution_scale	24	0.0	0.325
14	Grid	max_depth, min_child_weight	45	0.0	0.325
15	Grid	max_depth, reg_alpha	54	0.0	0.325
16	Grid	max_depth, reg_lambda	54	0.0	0.325
17	Grid	max_depth, aft_loss_distribution_scale	54	0.0	0.325
18	Grid	min_child_weight, reg_alpha	30	0.0	0.325
19	Grid	min_child_weight, reg_lambda	30	0.0	0.325
20	Grid	min_child_weight, aft_loss_distribution_scale	30	0.0	0.325
21	Grid	reg_alpha, reg_lambda	36	0.0	0.325
22	Grid	reg_alpha, aft_loss_distribution_scale	36	0.0	0.325
23	Grid	reg_lambda, aft_loss_distribution_scale	36	0.0	0.325
Dataset simulated.sin, logistic distribution					
0	Grid	min_child_weight, aft_loss_distribution_scale	30	0.850000	0.82500
1	Grid	max_depth, aft_loss_distribution_scale	54	0.853125	0.81875
2	Random	All six	100	0.823438	0.81250

(Continued on next page)

Table 1 (Continued from previous page)

Rank	Method	Hyperparameters selected	# trials	Valid. acc.	Test acc.
3	Random	All six	1000	0.851562	0.80625
4	Grid	max_depth, reg_alpha	54	0.746875	0.78125
5	Grid	reg_alpha, aft_loss_distribution_scale	36	0.807813	0.77500
6	Grid	reg_lambda, aft_loss_distribution_scale	36	0.804688	0.77500
7	Grid	learning_rate, aft_loss_distribution_scale	24	0.800000	0.77500
8	Grid	aft_loss_distribution_scale	6	0.793750	0.76250
9	Grid	min_child_weight, reg_alpha	30	0.750000	0.75625
10	Grid	learning_rate, max_depth	36	0.745313	0.75000
11	Grid	min_child_weight, reg_lambda	30	0.753125	0.74375
12	Grid	max_depth	9	0.740625	0.73750
13	Grid	learning_rate, min_child_weight	20	0.751563	0.73125
14	Grid	max_depth, reg_lambda	54	0.745313	0.73125
15	Grid	max_depth, min_child_weight	45	0.762500	0.73125
16	Grid	min_child_weight	5	0.742188	0.72500
17	Baseline	None (all defaults)	1	0.643750	0.66250
18	Grid	reg_alpha, reg_lambda	36	0.676562	0.65625
19	Grid	reg_alpha	6	0.654687	0.63750
20	Grid	reg_lambda	6	0.668750	0.63750
21	Grid	learning_rate	4	0.654687	0.62500
22	Grid	learning_rate, reg_lambda	24	0.668750	0.62500
23	Grid	learning_rate, reg_alpha	24	0.676562	0.45625
Dataset simulated.sin, extreme distribution					
0	Grid	max_depth, aft_loss_distribution_scale	54	0.884375	0.87500
1	Grid	reg_lambda, aft_loss_distribution_scale	36	0.865625	0.87500
2	Grid	aft_loss_distribution_scale	6	0.862500	0.86875
3	Grid	reg_alpha, aft_loss_distribution_scale	36	0.862500	0.86875
4	Random	All six	100	0.868750	0.86250
5	Grid	min_child_weight, aft_loss_distribution_scale	30	0.881250	0.85625
6	Random	All six	1000	0.887500	0.85625
7	Grid	learning_rate, aft_loss_distribution_scale	24	0.867188	0.85000
8	Grid	max_depth, min_child_weight	45	0.835938	0.78125
9	Grid	reg_alpha, reg_lambda	36	0.782813	0.78125
10	Grid	max_depth, reg_alpha	54	0.817187	0.77500
11	Grid	min_child_weight, reg_lambda	30	0.834375	0.77500
12	Grid	max_depth	9	0.810937	0.77500
13	Grid	min_child_weight	5	0.829688	0.77500
14	Grid	max_depth, reg_lambda	54	0.815625	0.76875
15	Grid	min_child_weight, reg_alpha	30	0.832812	0.76875
16	Grid	learning_rate, max_depth	36	0.815625	0.76875
17	Grid	learning_rate, min_child_weight	20	0.834375	0.76250
18	Grid	reg_alpha	6	0.776563	0.75625
19	Grid	reg_lambda	6	0.778125	0.75625
20	Baseline	None (all defaults)	1	0.764062	0.75625
21	Grid	learning_rate, reg_lambda	24	0.793750	0.74375
22	Grid	learning_rate	4	0.775000	0.73125
23	Grid	learning_rate, reg_alpha	24	0.782813	0.72500