Supplementary File for Paper: BO-SMOTE: A Novel Bayesian-Optimization-based Synthetic Minority Oversampling Technique

This supplementary gives the detailed comparison among our method and other oversamplers by using evaluation metrics F1 and GMean.

Table A gives the average F1 scores for all 20 datasets based on different combinations of classifiers and ten classic oversamplers. For different classifiers, all oversamplers are ranked under F1 scores. "None" in the table indicates a case where no oversampler is used. It serves as a baseline (BL) along with original SMOTE.

TABLE A
F1 RESULTS OF COMPARISON EXPERIMENTS WITH CLASSIC OVERSAMPLERS

Classifier	DT		kNN		SVM		MLP	
Rank	Oversampler	F1	Oversampler	F1	Oversampler	F1	Oversampler	F1
1	**BO-SMOTE**	**0.7476**	**BO-SMOTE**	**0.7764**	**BO-SMOTE**	0.7117	**BO-SMOTE**	0.7900
2	STL	0.7230	Lee	0.7498	GS	0.6841	SIPF	0.7396
3	ProWSyn	0.7160	PFS	0.7468	CCR	0.6796	ProWSyn	0.7367
4	SIPF	0.7149	STL	0.7465	PFS	0.6780	GS	0.7359
5	SMOBD	0.7109	CCR	0.7460	ProWSyn	0.6778	CCR	0.7351
6	AS	0.7094	GS	0.7452	SMOBD	0.6776	PFS	0.7339
7	CCR	0.7036	AS	0.7443	Lee	0.6762	SMOBD	0.7320
8	PFS	0.6971	SIPF	0.7435	SIPF	0.6741	AS	0.7251
9	LVQS	0.6963	SMOBD	0.7362	AS	0.6710	Lee	0.7219
10	GS	0.6946	ProWSyn	0.7321	STL	0.6655	LVQS	0.6829
11	Lee	0.6940	LVQS	0.7242	LVQS	0.6313	STL	0.6767
BL	SMOTE	0.7271	SMOTE	0.7435	SMOTE	0.6784	SMOTE	0.7485
BL	None	0.7220	None	0.7135	None	0.2635	None	0.7541

Table B gives the average GMean scores for all 20 datasets based on different combinations of classifiers and ten classic oversamplers. For different classifiers, all oversamplers are ranked under GMean scores. "None" in the table indicates a case where no oversampler is used. It serves as a baseline (BL) along with original SMOTE.

TABLE B
GMEAN RESULTS OF COMPARISON EXPERIMENTS WITH CLASSIC OVERSAMPLERS

Classifier	DT		kNN		SVM		MLP	
Rank	Oversampler	$\bar{G}$	Oversampler	$\bar{G}$	Oversampler	$\bar{G}$	Oversampler	$\bar{G}$
1	LVQS	0.8540	SMOBD	0.8793	AS	0.8329	**BO-SMOTE**	**0.8747**
2	**BO-SMOTE**	**0.8413**	SIPF	0.8782	STL	0.8314	SIPF	0.8416
3	STL	0.8377	CCR	0.8763	SMOBD	0.8312	ProWSyn	0.8391
4	ProWSyn	0.8350	GS	0.8749	PFS	0.8301	PFS	0.8390
5	AS	0.8285	PFS	0.8744	SIPF	0.8298	LVQS	0.8388
6	SIPF	0.8265	AS	0.8741	Lee	0.8281	AS	0.8387
7	CCR	0.8232	Lee	0.8727	GS	0.8273	CCR	0.8367
8	SMOBD	0.8207	LVQS	0.8724	CCR	0.8271	GS	0.8361
9	PFS	0.8093	ProWSyn	0.8715	ProWSyn	0.8269	SMOBD	0.8357
10	Lee	0.8071	STL	0.8531	LVQS	0.8142	Lee	0.8228
11	GS	0.8059	**BO-SMOTE**	**0.8502**	**BO-SMOTE**	**0.8112**	STL	0.8103
BL	SMOTE	0.8362	SMOTE	0.8771	SMOTE	0.8303	SMOTE	0.8456
BL	None	0.8133	None	0.7606	None	0.2731	None	0.8337

Table C gives the average F1 scores for all 20 datasets based on different combinations of classifiers and four recent oversamplers. For different classifiers, all oversamplers are ranked under F1 scores.

TABLE C
F1 RESULTS OF COMPARISON EXPERIMENTS WITH RECENT OVERSAMPLERS

Classifier	DT		kNN		SVM		MLP	
Rank	Oversampler	F1	Oversampler	F1	Oversampler	F1	Oversampler	F1
1	**BO-SMOTE**	**0.7476**	**BO-SMOTE**	**0.7764**	**BO-SMOTE**	**0.7117**	**BO-SMOTE**	**0.7900**
2	RSMOTE	0.7313	RSMOTE	0.7699	RSMOTE	0.6672	KNNOR	0.7242
3	GeSMOTE	0.7208	GeSMOTE	0.7422	KNNOR	0.6502	RSMOTE	0.7168
4	KNNOR	0.7195	KNNOR	0.7416	GeSMOTE	0.6354	IWSMOTE	0.7091
5	IWSMOTE	0.6838	IWSMOTE	0.6894	IWSMOTE	0.5716	GeSMOTE	0.6863

Table D gives the average GMean scores for all 20 datasets based on different combinations of classifiers and four recent oversamplers. For different classifiers, all oversamplers are ranked under GMean scores.

TABLE D
GMEAN RESULTS OF COMPARISON EXPERIMENTS WITH RECENT OVERSAMPLERS

Classifier	DT		kNN		SVM		MLP	
Rank	Oversampler	$\bar{G}$	Oversampler	$\bar{G}$	Oversampler	$\bar{G}$	Oversampler	$\bar{G}$
1	IWSMOTE	0.8496	GeSMOTE	0.8654	GeSMOTE	0.8200	**BO-SMOTE**	**0.8747**
2	GeSMOTE	0.8481	IWSMOTE	0.8601	**BO-SMOTE**	**0.8112**	IWSMOTE	0.8606
3	**BO-SMOTE**	**0.8413**	RSMOTE	0.8598	KNNOR	0.8005	GeSMOTE	0.8522
4	RSMOTE	0.8240	**BO-SMOTE**	**0.8502**	IWSMOTE	0.7897	KNNOR	0.8302
5	KNNOR	0.8189	KNNOR	0.8442	RSMOTE	0.7705	RSMOTE	0.7082