Geometric SMOTENC
A geometrically enhanced drop-in replacement for SMOTENC

Joao Fonseca[1*], Fernando Bacao[1]

[1]NOVA Information Management School, Universidade Nova de Lisboa

*Corresponding Author

Postal Address: NOVA Information Management School, Campus de Campolide, 1070–312 Lisboa, Portugal

Telephone: +351 21 382 8610

This is an abstract.

1. Introduction

This is text [1].

2. Related Work

SMOTE-ENC [2]

3. Motivation

C_{maj} set of majority class observations (most common class found in the target variable)

C_{min} set of minority class observations (least common class found in the target variable)

4. Proposed Method

5. Methodology

This section describes how the evaluation of G-SMOTENC was performed. We describe the datasets used in the experiment, their source and preprocessing steps carried out in Section 5.1. We describe the resampling and classifications methods used for comparing the performance of G-SMOTENC with other relevant oversampling and undersampling mthods in Section 5.2. The performance metrics used are defined in Section 5.3. Finally, the experimental procedure is described in Section 5.4.

5.1. Experimental Data

The datasets used in this experiment were extracted from the UC Irvine Machine Learning Repository. All of the datasets are publicly available and cover a range of different domains. The selection of datasets was done to ensure that all datasets are imbalanced and contained non-metric features (*i.e.*, whether ordinal, nominal or binary). These datasets will be used to show how the performance of different classifiers varies according to the used over/undersampling method.

At an initial stage, all datasets were preprocessed manually with minimal manipulations, to avoid the application of preprocessing methods beyond the scope of this paper. This step was conducted to remove features and/or observations with missing values and identifying the non-metric features. The second stage of our preprocessing was done systematically. The resulting datasets are shown in Table 1.

Table 1: Description of the datasets collected after data preprocessing. The sampling strategy is similar across datasets. Legend: (IR) Imbalance Ratio

Dataset	Metric	Non-Metric	Obs.	Min. Obs.	Maj. Obs.	IR	Classes
Abalone	1	7	4139	15	689	45.93	18
Adult	8	6	5000	1268	3732	2.94	2
Adult (10)	8	6	5000	451	4549	10.09	2
Annealing	4	6	790	34	608	17.88	4
Census	24	7	5000	337	4663	13.84	2
Contraceptive	4	5	1473	333	629	1.89	3
Contraceptive (10)	4	5	1036	62	629	10.15	3
Contraceptive (20)	4	5	990	31	629	20.29	3
Contraceptive (31)	4	5	973	20	629	31.45	3
Contraceptive (41)	4	5	966	15	629	41.93	3
Covertype	2	10	5000	20	2449	122.45	7
Credit Approval	9	6	653	296	357	1.21	2
German Credit	13	7	1000	300	700	2.33	2
German Credit (10)	13	7	770	70	700	10.00	2
German Credit (20)	13	7	735	35	700	20.00	2
German Credit (30)	13	7	723	23	700	30.43	2
German Credit (41)	13	7	717	17	700	41.18	2
Heart Disease	5	5	740	22	357	16.23	5
Heart Disease (21)	5	5	735	17	357	21.00	5

The second part of the data preprocessing pipeline starts with the generation of artificially imbalanced datasets with different Imbalance Ratios (IR). For each original dataset, we create its more imbalanced versions at intervals of 10, while ensuring that $|C_{min}| \geq 15$. The sampling strategy was determined for class $n \in \{1, \ldots, n, \ldots, m\}$ as a linear interpolation using $|C_{maj}|$ and $|C'_{min}| = \frac{|C_{maj}|}{IR}$, as shown in equation 1.

$$|C_n|^{imb} = \frac{\frac{|C_{maj}|}{IR} - |C_{maj}|}{|\{1, \ldots, n, \ldots, m\}| - 1}.|C_n| + |C_{max}| \tag{1}$$

The new, artificially imbalanced dataset, is formed by randomly removing observations from each C_n such that $C'_n \subseteq C_n, |C'_n| = |C_n|^{imb}$. The artificially imbalanced datasets are marked with its imbalance ratio as a suffix in Table 1.

The datasets (both original and artificially imbalanced versions) are then filtered to ensure all datasets have a minimum of 500 observations. The remaining datasets whose number of observations is larger than 5000 are randomly sampled to match this number of observations. Afterwards, for each remaining dataset we remove all observations from target classes whose frequency is lower than 15 observations. Finally, the continuous and discrete features are scaled to the range $[0, 1]$ to ensure a common range between all features.

5.2. Machine Learning Algorithms

The choice of classifiers used in the experimental procedure were based on their type (tree-based, nearest neighbors-based, linear model and ensemble-based), popularity and consistency in performance. We used Decision Tree (DT), a K-Nearest Neighbors (KNN) classifier, a Logistic Regression (LR) and a Random Forest (RF).

Given the lack of existing oversamplers that address imbalanced learning problems with mixed data types, the amount of benchmark methods used is also limited. We used the well known methods that are compatible with this type of datasets: SMOTENC, Random Undersampling (RUS) and Random Oversampling (ROS). Table 2 shows the hyperparameters used for the parameter search described in Section 5.4.

5.3. Performance Metrics

Although the typical performance metrics, *e.g.*, Overall Accuracy (OA), are intuitive to interpret, they are often inappropriate to measure a classifier's performance in an imbalanced learning context **[CITATION]**. For example, to estimate an event that occurs in 1% of the dataset, a constant classifier would obtain an OA of 0.99 and still be unusable. However, this metric is still reported in some of our results to maintain a metric that is easier to interpret.

More recent surveys have found the Geometric-mean ($\textit{G-mean} = \sqrt{Sensitivity \times Specificity}$), F1-score ($\textit{F-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$), $Sensitivity = \frac{TP}{FN+TP}$ and $Specificity = \frac{TN}{TN+FP}$ to be commonly used performance metrics in imbalanced learning contexts [3]. These metrics are calculated as a function

Table 2: Hyperparameter definition for the classifiers and resamplers used in the experiment.

Classifier		
DT	min. samples split	2
	criterion	gini
	max depth	3, 6
LR	maximum iterations	10000
	multi-class	One-vs-All
	solver	saga
	penalty	None, L1, L2
KNN	# neighbors	3, 5
	weights	uniform
	metric	euclidean
RF	min. samples split	2
	# estimators	50, 100
	Max depth	3, 6
	criterion	gini
Resampler		
SMOTENC	# neighbors	3, 5
G-SMOTENC	# neighbors	3, 5
	deformation factor	0.0, 0.25, 0.5, 0.75, 1.0
	truncation factor	-1.0, -0.5, 0.0, 0.5, 1.0
	selection strategy	"combined", "minority", "majority"
RUS	replacement	False
ROS	(no applicable parameters)	

of the number of False/True Positives (FP and TP) and False/True Negatives (FN and TN), having $Precision = \frac{TP}{TP+FP}$ and $Recall = \frac{TP}{TP+FN}$. This finding is consistent with other well-known recommendations on the usage of performance metrics [4]. This led us to adopt, along with OA, both F-score and G-mean as the main performance metrics for this study.

5.4. Experimental Procedure

The experimental procedure was applied similarly to all combinations of resamplers, classifiers and hyperparameter combinations across all datasets. The evaluation of the models' performance was tested using a 5-fold Cross Validation (CV) approach. The mean performance in the test set is calculated over the 5 folds and 3 different runs of the experimental procedure for each combination resampling/classifier hyperparameters. For each dataset, results of the hyperparameters that optimize the performance of a resampler/classifier are selected. These results were then used for analysis and are shown in Table 7 (see Appendix). Figure 1 shows a diagram of the experimental procedure described.

A CV run consists of a stratified partitioning (*i.e.*, each partition contains the same relative frequencies of target labels) of the dataset into five parts. A given resampler/classifier combination with a specific set of hyperparameters is fit and tested five times, using one of the partitions as a test set and the remaining ones as training set. The estimated performance consists of the average classification performance across the five different test sets.

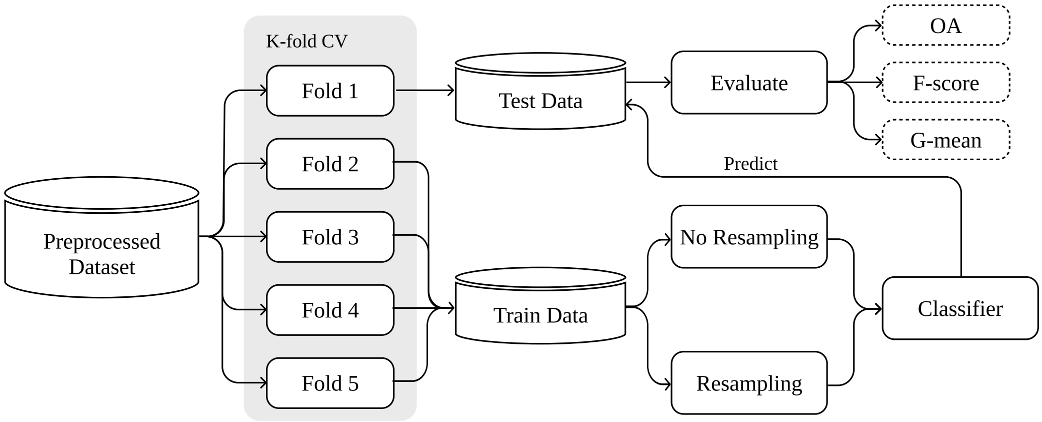

Figure 1: Experimental procedure used in this study.

5.5. Software Implementation

The algorithmic implementation of G-SMOTENC was written using the Python programming language and is available in the open-source package ML-Research [5], along with other utilities used to produce the experiment and outputs used in Section 6. In addition, the packages Scikit-Learn [6], Imbalanced-Learn [7] and Research-Learn were also used in the experimental procedure to get the implementations of the classifiers, benchmark over/undersamplers and run the experimental procedure. The Latex code, Python scripts (including data pulling and preprocessing, experiment setup and results' analysis), as well as the datasets used are available in this GitHub repository.

6. Results and Discussion

In this section we present the experimental results. We focus on the comparison of classification performance using oversamplers whose generation mechanism is compatible with datasets containing both continuous and categorical features.

The analysis of our experimental results were developed in two stages: (1) analysis of mean ranking and absolute performance and (2) statistical analysis. In Section 6.3 we discuss the main insights extracted by analysing the results reported in Sections 6.1 and 6.2.

6.1. Results

Table 3 presents the mean rankings of cross validation scores across the different combinations of oversamplers, metrics and classifiers. These results were calculated by assigning a ranking score for each oversampler from 1 (best) to 4 (worst) for each dataset, metric and classifier, based on the results reported in Table 7 (see Appendix).

Table 3: Mean rankings over the different datasets, folds and runs used in the experiment.

Classifier	Metric	G-SMOTENC	NONE	SMOTENC	ROS	RUS
DT	OA	1.66 ± 0.13	**1.55 ± 0.22**	3.16 ± 0.16	4.00 ± 0.08	4.63 ± 0.19
DT	F-Score	**1.11 ± 0.07**	3.21 ± 0.30	2.58 ± 0.18	3.53 ± 0.16	4.58 ± 0.19
DT	G-Mean	**1.53 ± 0.21**	4.89 ± 0.07	2.53 ± 0.18	2.47 ± 0.23	3.58 ± 0.23
KNN	OA	2.39 ± 0.12	**1.32 ± 0.23**	3.58 ± 0.16	2.97 ± 0.26	4.74 ± 0.17
KNN	F-Score	**1.37 ± 0.16**	3.37 ± 0.28	2.68 ± 0.20	2.95 ± 0.27	4.63 ± 0.17
KNN	G-Mean	**1.74 ± 0.17**	4.84 ± 0.12	2.63 ± 0.17	3.26 ± 0.25	2.53 ± 0.35
LR	OA	2.47 ± 0.15	**1.32 ± 0.23**	2.76 ± 0.17	3.66 ± 0.21	4.79 ± 0.16
LR	F-Score	**1.89 ± 0.21**	3.84 ± 0.28	2.05 ± 0.24	2.79 ± 0.25	4.42 ± 0.21
LR	G-Mean	1.97 ± 0.23	5.00 ± 0.00	3.29 ± 0.17	**1.89 ± 0.17**	2.84 ± 0.30
RF	OA	1.76 ± 0.09	**1.24 ± 0.09**	3.37 ± 0.11	3.66 ± 0.12	4.97 ± 0.03
RF	F-Score	**1.26 ± 0.13**	4.21 ± 0.25	2.68 ± 0.17	2.42 ± 0.22	4.42 ± 0.12
RF	G-Mean	**1.68 ± 0.22**	4.84 ± 0.16	2.89 ± 0.21	2.26 ± 0.23	3.32 ± 0.25

Table 4 presents the mean cross validation scores. With exception to the OA metric, G-SMOTENC either outperformed or matched the the remaining oversamplers.

Table 4: Mean scores over the different datasets, folds and runs used in the experiment

Classifier	Metric	G-SMOTENC	NONE	SMOTENC	ROS	RUS
DT	OA	0.74 ± 0.05	**0.75 ± 0.04**	0.68 ± 0.04	0.66 ± 0.04	0.58 ± 0.04
DT	F-Score	**0.56 ± 0.04**	0.52 ± 0.04	0.54 ± 0.04	0.52 ± 0.04	0.48 ± 0.04
DT	G-Mean	**0.69 ± 0.03**	0.60 ± 0.02	0.68 ± 0.03	0.67 ± 0.03	0.65 ± 0.03
KNN	OA	0.69 ± 0.04	**0.73 ± 0.05**	0.67 ± 0.04	0.69 ± 0.05	0.57 ± 0.04
KNN	F-Score	**0.53 ± 0.04**	0.50 ± 0.04	0.52 ± 0.04	0.52 ± 0.04	0.46 ± 0.04
KNN	G-Mean	**0.66 ± 0.03**	0.58 ± 0.03	0.64 ± 0.03	0.62 ± 0.03	0.65 ± 0.03
LR	OA	0.68 ± 0.05	**0.75 ± 0.04**	0.68 ± 0.05	0.66 ± 0.05	0.58 ± 0.04
LR	F-Score	**0.54 ± 0.04**	0.52 ± 0.04	**0.54 ± 0.04**	0.53 ± 0.04	0.48 ± 0.04
LR	G-Mean	**0.69 ± 0.02**	0.60 ± 0.03	0.68 ± 0.02	**0.69 ± 0.03**	0.67 ± 0.03
RF	OA	0.74 ± 0.04	**0.76 ± 0.04**	0.69 ± 0.04	0.69 ± 0.04	0.59 ± 0.04
RF	F-Score	**0.57 ± 0.04**	0.48 ± 0.04	0.55 ± 0.04	0.55 ± 0.04	0.49 ± 0.04
RF	G-Mean	**0.70 ± 0.02**	0.57 ± 0.02	0.68 ± 0.03	0.69 ± 0.03	0.68 ± 0.03

6.2. Statistical Analysis

To conduct an appropriate statistical analysis in an experiment with multiple datasets, it is necessary to use methods that account for the multiple comparison problem. Based on the recommendations found in [8], we applied the Friedman test along with the Holm-Bonferroni test for a post-hoc analysis.

In Section 5.3 we explained that OA, although easily interpretable, is not an appropriate performance metric for imbalanced learning problems. Therefore, the statistical analysis was developed using the two imbalance-appropriate metrics used in the study: F-Score and G-Mean. The statistical analysis started with the assessment of a statistically significant difference in performance across resampling methods using a Friedman test [9]. The results of this test are shown in Table 5. The null hypothesis is rejected in all cases.

Table 5: Results for Friedman test. Statistical significance is tested at a level of $\alpha = 0.05$. The null hypothesis is that there is no difference in the classification outcome across resamplers.

Classifier	Metric	p-value	Significance
DT	F-Score	3.6e-10	True
DT	G-Mean	3.4e-10	True
KNN	F-Score	1.5e-08	True
KNN	G-Mean	2.5e-08	True
LR	F-Score	1.7e-07	True
LR	G-Mean	6.9e-10	True
RF	F-Score	9.8e-11	True
RF	G-Mean	6.7e-09	True

We performed a Holm-Bonferroni test to understand whether the difference in performance of G-SMOTENC is statistically significant to the remaining resampling methods. The results of this test are shown in Table 6. The null hypothesis was rejected in 27 out of 32 tests.

Table 6: Adjusted p-values using the Holm-Bonferroni test. Statistical significance is tested at a level of $\alpha = 0.05$. The null hypothesis is that the benchmark methods perform similarly compared to the control method (G-SMOTENC).

Classifier	Metric	NONE	RAND-OVER	RAND-UNDER	SMOTENC
DT	F-Score	**1.0e-04**	**5.5e-06**	**1.0e-06**	**1.0e-04**
DT	G-Mean	**4.5e-07**	**2.8e-02**	**2.9e-04**	**1.8e-03**
KNN	F-Score	**6.4e-04**	**7.2e-04**	**6.4e-04**	**2.2e-04**
KNN	G-Mean	**1.3e-05**	**6.5e-03**	2.0e-01	**9.6e-03**
LR	F-Score	**3.0e-03**	**6.2e-03**	**2.9e-04**	6.1e-01
LR	G-Mean	**1.3e-07**	8.6e-01	2.4e-01	**3.0e-04**
RF	F-Score	**1.4e-06**	**4.0e-03**	**1.4e-06**	**1.6e-04**
RF	G-Mean	**3.1e-06**	2.5e-01	**2.3e-02**	**8.8e-03**

6.3. Discussion

The results reported in Section 6.1 show that...

7. Conclusion

This is a conclusion.

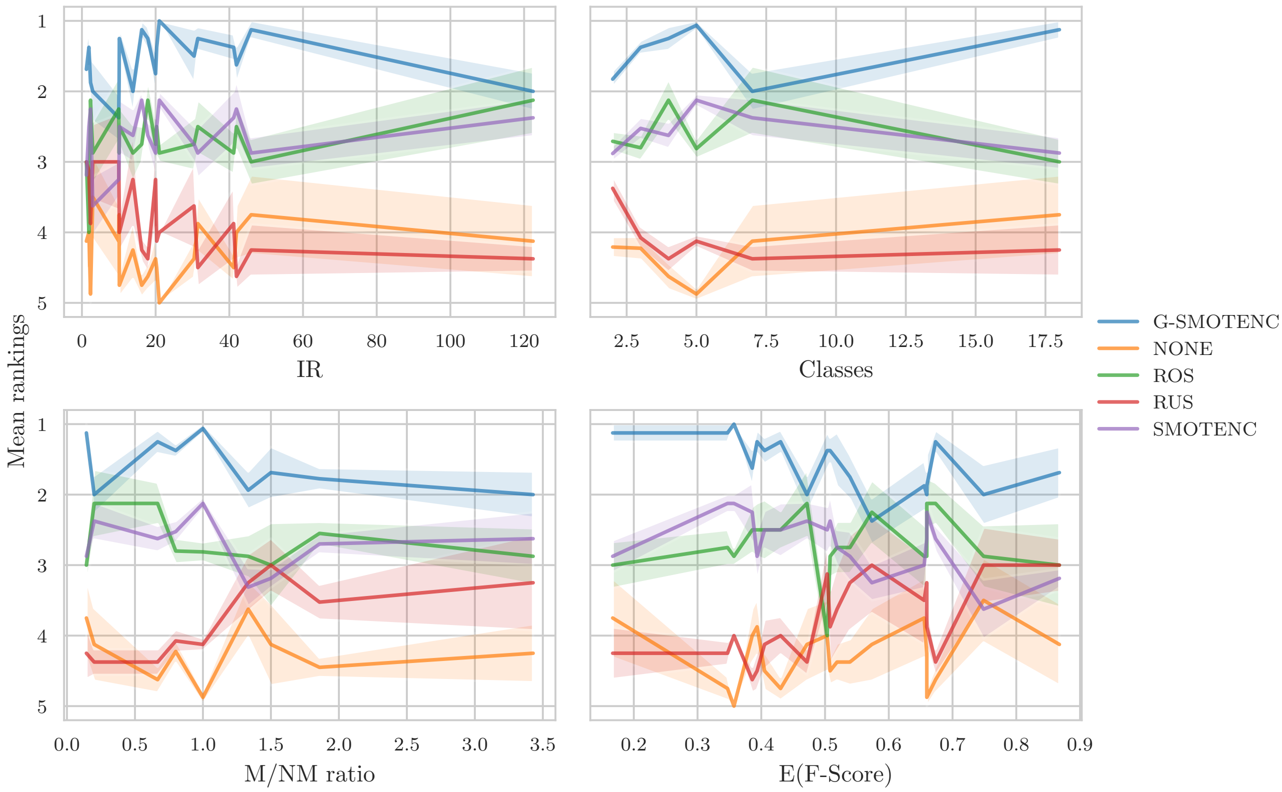

Figure 2: Average ranking of oversamplers over different characteristics of the datasets used in the experiment. Legend: IR — Imbalance Ratio, Classes — Number of classes in the dataset, M/NM ratio — ratio between the number of metric and non-metric features, E(F-Score) — Mean F-Score of dataset across all combinations of classifiers and oversamplers.

References

[1] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Oversampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, jun 2002.

[2] M. Mukherjee and M. Khushi, "Smote-enc: A novel smote-based method to generate synthetic data for nominal and continuous features," *Applied System Innovation*, vol. 4, no. 1, p. 18, 2021.

[3] N. Rout, D. Mishra, and M. K. Mallick, "Handling imbalanced data: a survey," in *International proceedings on advances in soft computing, intelligent systems and applications*, pp. 431–443, Springer, 2018.

[4] L. A. Jeni, J. F. Cohn, and F. De La Torre, "Facing imbalanced data–recommendations for the use of performance metrics," in *2013 Humaine association conference on affective computing and intelligent interaction*, pp. 245–251, IEEE, 2013.

[5] J. Fonseca, G. Douzas, and F. Bacao, "Increasing the effectiveness of active learning: Introducing artificial data generation in active learning for land use/land cover classification," *Remote Sensing*, vol. 13, no. 13, p. 2619, 2021.

[6] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[7] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, no. 17, pp. 1–5, 2017.

[8] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[9] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the american statistical association*, vol. 32, no. 200, pp. 675–701, 1937.

A. Appendix

Table 7: Wide optimal results

Dataset	Classifier	Metric	G-SMOTENC	NONE	SMOTENC	ROS	RUS
Abalone	DT	OA	0.221	**0.256**	0.190	0.203	0.207
Abalone	DT	F-Score	0.168	**0.170**	0.156	0.154	0.132
Abalone	DT	G-Mean	**0.460**	0.413	0.445	0.457	0.421
Abalone	KNN	OA	0.215	**0.237**	0.186	0.197	0.188
Abalone	KNN	F-Score	**0.167**	0.157	0.150	0.151	0.140
Abalone	KNN	G-Mean	**0.429**	0.391	0.409	0.397	0.421
Abalone	LR	OA	0.235	**0.272**	0.228	0.229	0.195
Abalone	LR	F-Score	**0.189**	0.180	0.186	0.179	0.166

Continued on next page

Table 7: Wide optimal results

Dataset	Classifier	Metric	G-SMOTENC	NONE	SMOTENC	ROS	RUS
Abalone	LR	G-Mean	**0.473**	0.415	0.466	0.456	0.441
Abalone	RF	OA	0.237	**0.276**	0.221	0.224	0.197
Abalone	RF	F-Score	**0.194**	0.174	0.180	0.184	0.162
Abalone	RF	G-Mean	**0.486**	0.416	0.461	0.465	0.448
Adult	DT	OA	0.830	**0.835**	0.785	0.800	0.785
Adult	DT	F-Score	**0.767**	0.763	0.754	0.755	0.744
Adult	DT	G-Mean	**0.809**	0.747	0.808	0.806	0.801
Adult	KNN	OA	0.786	**0.805**	0.781	0.763	0.761
Adult	KNN	F-Score	**0.738**	0.732	0.735	0.718	0.728
Adult	KNN	G-Mean	0.766	0.724	0.762	0.757	**0.780**
Adult	LR	OA	0.803	**0.839**	0.803	0.804	0.801
Adult	LR	F-Score	0.768	**0.773**	0.767	0.771	0.769
Adult	LR	G-Mean	0.813	0.758	0.805	**0.815**	**0.815**
Adult	RF	OA	0.820	**0.832**	0.757	0.755	0.753
Adult	RF	F-Score	**0.769**	0.739	0.727	0.729	0.728
Adult	RF	G-Mean	0.796	0.711	0.787	**0.797**	**0.797**
Adult (10)	DT	OA	**0.930**	0.928	0.822	0.789	0.775
Adult (10)	DT	F-Score	**0.711**	0.708	0.656	0.641	0.630
Adult (10)	DT	G-Mean	0.812	0.663	0.807	**0.815**	0.808
Adult (10)	KNN	OA	0.864	**0.909**	0.854	0.851	0.745
Adult (10)	KNN	F-Score	**0.667**	0.652	0.658	0.648	0.602
Adult (10)	KNN	G-Mean	0.745	0.629	0.747	0.722	**0.783**
Adult (10)	LR	OA	0.836	**0.925**	0.837	0.815	0.791
Adult (10)	LR	F-Score	0.666	**0.705**	0.667	0.663	0.647
Adult (10)	LR	G-Mean	0.804	0.663	0.787	0.811	**0.814**
Adult (10)	RF	OA	0.899	**0.924**	0.773	0.763	0.743
Adult (10)	RF	F-Score	**0.718**	0.615	0.620	0.624	0.610
Adult (10)	RF	G-Mean	**0.809**	0.579	0.786	0.806	0.806
Annealing	DT	OA	0.824	**0.843**	0.742	0.733	0.694
Annealing	DT	F-Score	**0.736**	0.643	0.732	0.724	0.683
Annealing	DT	G-Mean	**0.914**	0.738	0.909	0.906	0.880
Annealing	KNN	OA	0.849	0.847	0.829	**0.854**	0.508
Annealing	KNN	F-Score	0.780	0.724	0.747	**0.783**	0.476
Annealing	KNN	G-Mean	0.901	0.781	0.867	**0.909**	0.814
Annealing	LR	OA	0.572	**0.814**	0.573	0.566	0.510
Annealing	LR	F-Score	**0.620**	0.540	0.617	0.615	0.496
Annealing	LR	G-Mean	**0.851**	0.663	0.843	0.848	0.811
Annealing	RF	OA	**0.868**	**0.868**	0.729	0.733	0.637
Annealing	RF	F-Score	**0.800**	0.644	0.730	0.736	0.641
Annealing	RF	G-Mean	**0.917**	0.727	0.904	0.910	0.873
Census	DT	OA	0.942	**0.943**	0.894	0.844	0.795
Census	DT	F-Score	**0.733**	0.731	0.693	0.652	0.617
Census	DT	G-Mean	0.813	0.698	0.800	0.814	**0.817**
Census	KNN	OA	0.874	**0.933**	0.867	0.878	0.731
Census	KNN	F-Score	0.652	0.648	**0.655**	0.640	0.567
Census	KNN	G-Mean	0.767	0.620	0.768	0.733	**0.794**
Census	LR	OA	0.940	**0.949**	0.938	0.940	0.815
Census	LR	F-Score	0.760	0.743	0.760	**0.762**	0.639

Continued on next page

Table 7: Wide optimal results

Dataset	Classifier	Metric	G-SMOTENC	NONE	SMOTENC	ROS	RUS
Census	LR	G-Mean	0.807	0.707	0.782	0.801	**0.837**
Census	RF	OA	0.876	**0.933**	0.819	0.740	0.714
Census	RF	F-Score	**0.679**	0.483	0.636	0.580	0.562
Census	RF	G-Mean	**0.827**	0.500	0.818	0.822	0.814
Contraceptive	DT	OA	**0.563**	0.538	0.537	0.512	0.525
Contraceptive	DT	F-Score	**0.549**	0.518	0.529	0.507	0.520
Contraceptive	DT	G-Mean	**0.661**	0.630	0.646	0.630	0.641
Contraceptive	KNN	OA	0.465	**0.478**	0.455	0.435	0.468
Contraceptive	KNN	F-Score	0.460	**0.462**	0.450	0.432	0.461
Contraceptive	KNN	G-Mean	0.588	0.580	0.579	0.566	**0.590**
Contraceptive	LR	OA	**0.515**	0.514	0.514	0.510	0.510
Contraceptive	LR	F-Score	**0.512**	0.492	0.509	0.505	0.506
Contraceptive	LR	G-Mean	**0.635**	0.604	0.631	0.628	0.627
Contraceptive	RF	OA	0.553	**0.557**	0.540	0.534	0.526
Contraceptive	RF	F-Score	**0.545**	0.524	0.535	0.529	0.522
Contraceptive	RF	G-Mean	**0.659**	0.634	0.653	0.649	0.643
Contraceptive (10)	DT	OA	**0.645**	**0.645**	0.568	0.528	0.487
Contraceptive (10)	DT	F-Score	**0.479**	0.452	0.478	0.454	0.414
Contraceptive (10)	DT	G-Mean	0.644	0.584	**0.648**	0.637	0.610
Contraceptive (10)	KNN	OA	0.524	**0.570**	0.508	0.495	0.451
Contraceptive (10)	KNN	F-Score	**0.419**	0.404	0.410	0.404	0.368
Contraceptive (10)	KNN	G-Mean	**0.576**	0.529	0.561	0.569	0.561
Contraceptive (10)	LR	OA	0.516	**0.622**	0.506	0.489	0.476
Contraceptive (10)	LR	F-Score	**0.431**	0.375	0.426	0.425	0.411
Contraceptive (10)	LR	G-Mean	0.619	0.526	0.609	**0.624**	0.618
Contraceptive (10)	RF	OA	0.648	**0.651**	0.569	0.550	0.494
Contraceptive (10)	RF	F-Score	**0.500**	0.387	0.473	0.471	0.425
Contraceptive (10)	RF	G-Mean	**0.656**	0.542	0.639	0.650	0.625
Contraceptive (20)	DT	OA	**0.671**	0.659	0.612	0.556	0.456
Contraceptive (20)	DT	F-Score	**0.475**	0.430	0.459	0.428	0.371
Contraceptive (20)	DT	G-Mean	**0.643**	0.570	0.626	0.632	0.605
Contraceptive (20)	KNN	OA	0.556	**0.600**	0.529	0.541	0.442
Contraceptive (20)	KNN	F-Score	**0.399**	0.375	0.384	0.389	0.345
Contraceptive (20)	KNN	G-Mean	**0.565**	0.519	0.544	0.537	0.549
Contraceptive (20)	LR	OA	0.506	**0.641**	0.508	0.486	0.440
Contraceptive (20)	LR	F-Score	**0.397**	0.375	**0.397**	0.389	0.358
Contraceptive (20)	LR	G-Mean	0.608	0.523	0.604	**0.613**	0.585
Contraceptive (20)	RF	OA	0.668	**0.674**	0.588	0.562	0.475
Contraceptive (20)	RF	F-Score	**0.473**	0.384	0.450	0.436	0.389
Contraceptive (20)	RF	G-Mean	0.659	0.535	0.641	**0.670**	0.633
Contraceptive (31)	DT	OA	0.667	**0.670**	0.608	0.604	0.440
Contraceptive (31)	DT	F-Score	**0.454**	0.441	0.438	0.453	0.346
Contraceptive (31)	DT	G-Mean	0.642	0.577	0.605	**0.655**	0.592
Contraceptive (31)	KNN	OA	0.563	**0.633**	0.545	0.550	0.405
Contraceptive (31)	KNN	F-Score	**0.403**	0.385	0.384	0.378	0.298
Contraceptive (31)	KNN	G-Mean	**0.574**	0.527	0.544	0.531	0.511
Contraceptive (31)	LR	OA	0.500	**0.656**	0.508	0.483	0.423
Contraceptive (31)	LR	F-Score	**0.379**	0.376	**0.379**	0.374	0.336

Continued on next page

Table 7: Wide optimal results

Dataset	Classifier	Metric	G-SMOTENC	NONE	SMOTENC	ROS	RUS
Contraceptive (31)	LR	G-Mean	**0.597**	0.523	0.579	0.585	0.580
Contraceptive (31)	RF	OA	0.681	**0.683**	0.608	0.583	0.442
Contraceptive (31)	RF	F-Score	**0.450**	0.378	0.434	0.435	0.349
Contraceptive (31)	RF	G-Mean	**0.647**	0.531	0.630	0.640	0.600
Contraceptive (41)	DT	OA	0.651	**0.666**	0.588	0.566	0.433
Contraceptive (41)	DT	F-Score	**0.459**	0.426	0.408	0.409	0.336
Contraceptive (41)	DT	G-Mean	**0.622**	0.573	0.579	0.589	0.555
Contraceptive (41)	KNN	OA	0.563	**0.611**	0.546	0.538	0.395
Contraceptive (41)	KNN	F-Score	**0.393**	0.373	0.381	0.370	0.289
Contraceptive (41)	KNN	G-Mean	0.542	0.515	**0.550**	0.526	0.515
Contraceptive (41)	LR	OA	0.525	**0.658**	0.524	0.504	0.435
Contraceptive (41)	LR	F-Score	0.389	0.375	**0.393**	0.387	0.336
Contraceptive (41)	LR	G-Mean	0.606	0.520	0.604	**0.627**	0.569
Contraceptive (41)	RF	OA	0.665	**0.681**	0.598	0.588	0.415
Contraceptive (41)	RF	F-Score	**0.444**	0.378	0.418	0.429	0.323
Contraceptive (41)	RF	G-Mean	0.612	0.528	**0.616**	**0.616**	0.566
Covertype	DT	OA	0.580	**0.705**	0.587	0.567	0.450
Covertype	DT	F-Score	0.484	**0.490**	0.481	0.475	0.361
Covertype	DT	G-Mean	**0.769**	0.671	0.758	0.758	0.700
Covertype	KNN	OA	0.690	**0.700**	0.683	0.699	0.454
Covertype	KNN	F-Score	0.532	0.457	0.535	**0.561**	0.367
Covertype	KNN	G-Mean	0.745	0.642	0.753	**0.763**	0.691
Covertype	LR	OA	0.637	**0.721**	0.640	0.611	0.472
Covertype	LR	F-Score	0.516	0.507	**0.526**	0.492	0.353
Covertype	LR	G-Mean	**0.792**	0.678	0.786	0.790	0.697
Covertype	RF	OA	0.598	**0.704**	0.583	0.587	0.485
Covertype	RF	F-Score	0.517	0.360	0.507	**0.519**	0.394
Covertype	RF	G-Mean	0.800	0.572	0.799	**0.804**	0.737
Credit Approval	DT	OA	**0.867**	0.847	0.862	0.861	0.865
Credit Approval	DT	F-Score	**0.867**	0.845	0.862	0.861	0.865
Credit Approval	DT	G-Mean	**0.874**	0.848	0.869	0.867	0.872
Credit Approval	KNN	OA	**0.870**	0.865	0.868	**0.870**	0.865
Credit Approval	KNN	F-Score	**0.869**	0.864	0.867	**0.869**	0.864
Credit Approval	KNN	G-Mean	**0.871**	0.865	0.868	**0.871**	0.866
Credit Approval	LR	OA	0.873	0.868	0.871	**0.874**	0.873
Credit Approval	LR	F-Score	0.873	0.868	0.871	**0.874**	0.873
Credit Approval	LR	G-Mean	0.877	0.873	0.877	**0.879**	0.878
Credit Approval	RF	OA	0.876	**0.877**	0.871	0.868	0.868
Credit Approval	RF	F-Score	0.876	**0.877**	0.871	0.868	0.868
Credit Approval	RF	G-Mean	**0.879**	**0.879**	0.876	0.872	0.873
German Credit	DT	OA	0.704	**0.713**	0.702	0.660	0.644
German Credit	DT	F-Score	**0.662**	0.608	0.654	0.633	0.623
German Credit	DT	G-Mean	**0.681**	0.608	0.667	0.663	0.660
German Credit	KNN	OA	0.681	**0.718**	0.682	0.670	0.641
German Credit	KNN	F-Score	**0.653**	0.628	0.650	0.636	0.616
German Credit	KNN	G-Mean	**0.675**	0.621	0.668	0.656	0.642
German Credit	LR	OA	0.727	**0.751**	0.729	0.724	0.712
German Credit	LR	F-Score	0.695	0.681	**0.697**	**0.697**	0.686

Continued on next page

Table 7: Wide optimal results

Dataset	Classifier	Metric	G-SMOTENC	NONE	SMOTENC	ROS	RUS
German Credit	LR	G-Mean	**0.722**	0.672	0.713	0.720	0.713
German Credit	RF	OA	**0.760**	0.741	0.739	0.737	0.700
German Credit	RF	F-Score	0.701	0.580	0.702	**0.709**	0.680
German Credit	RF	G-Mean	0.715	0.588	0.716	**0.730**	0.719
German Credit (10)	DT	OA	**0.909**	0.906	0.804	0.713	0.696
German Credit (10)	DT	F-Score	**0.575**	0.539	0.572	0.526	0.511
German Credit (10)	DT	G-Mean	0.628	0.535	0.629	**0.644**	0.631
German Credit (10)	KNN	OA	0.787	**0.913**	0.757	0.835	0.684
German Credit (10)	KNN	F-Score	0.578	**0.581**	0.558	0.573	0.528
German Credit (10)	KNN	G-Mean	0.662	0.559	0.643	0.588	**0.667**
German Credit (10)	LR	OA	0.839	**0.904**	0.831	0.799	0.682
German Credit (10)	LR	F-Score	0.619	0.596	0.610	**0.620**	0.550
German Credit (10)	LR	G-Mean	0.683	0.578	0.675	0.716	**0.722**
German Credit (10)	RF	OA	**0.910**	0.909	0.865	0.877	0.696
German Credit (10)	RF	F-Score	0.624	0.476	0.614	**0.661**	0.557
German Credit (10)	RF	G-Mean	0.653	0.500	0.646	0.709	**0.729**
German Credit (20)	DT	OA	**0.952**	**0.952**	0.875	0.795	0.668
German Credit (20)	DT	F-Score	**0.573**	0.525	0.559	0.522	0.457
German Credit (20)	DT	G-Mean	0.666	0.529	0.679	**0.690**	0.629
German Credit (20)	KNN	OA	0.856	**0.952**	0.826	0.905	0.679
German Credit (20)	KNN	F-Score	**0.561**	0.535	0.528	0.556	0.491
German Credit (20)	KNN	G-Mean	0.692	0.527	0.635	0.570	**0.709**
German Credit (20)	LR	OA	0.913	**0.952**	0.910	0.838	0.680
German Credit (20)	LR	F-Score	**0.596**	0.534	0.593	0.553	0.473
German Credit (20)	LR	G-Mean	0.651	0.531	0.627	0.661	**0.682**
German Credit (20)	RF	OA	**0.954**	0.952	0.920	0.931	0.709
German Credit (20)	RF	F-Score	**0.597**	0.488	0.574	0.572	0.493
German Credit (20)	RF	G-Mean	0.681	0.500	0.625	0.674	**0.691**
German Credit (30)	DT	OA	**0.968**	0.963	0.885	0.856	0.628
German Credit (30)	DT	F-Score	**0.558**	0.509	0.526	0.506	0.413
German Credit (30)	DT	G-Mean	**0.686**	0.509	0.631	0.602	0.565
German Credit (30)	KNN	OA	0.902	**0.968**	0.849	0.935	0.697
German Credit (30)	KNN	F-Score	**0.530**	0.492	0.512	0.519	0.473
German Credit (30)	KNN	G-Mean	0.681	0.500	0.588	0.536	**0.705**
German Credit (30)	LR	OA	0.921	**0.967**	0.918	0.877	0.611
German Credit (30)	LR	F-Score	**0.578**	0.516	0.577	0.537	0.421
German Credit (30)	LR	G-Mean	0.649	0.510	0.650	**0.661**	0.660
German Credit (30)	RF	OA	**0.968**	**0.968**	0.942	0.954	0.705
German Credit (30)	RF	F-Score	**0.592**	0.492	0.563	0.589	0.474
German Credit (30)	RF	G-Mean	**0.689**	0.500	0.601	0.606	0.679
German Credit (41)	DT	OA	**0.976**	0.971	0.916	0.905	0.635
German Credit (41)	DT	F-Score	**0.563**	0.493	0.544	0.502	0.408
German Credit (41)	DT	G-Mean	**0.636**	0.497	0.615	0.520	0.524
German Credit (41)	KNN	OA	0.929	**0.976**	0.876	0.944	0.674
German Credit (41)	KNN	F-Score	**0.524**	0.494	0.500	0.502	0.440
German Credit (41)	KNN	G-Mean	0.593	0.500	0.558	0.516	**0.630**
German Credit (41)	LR	OA	0.940	**0.976**	0.943	0.927	0.641
German Credit (41)	LR	F-Score	0.546	0.494	**0.552**	0.515	0.420

Continued on next page

Table 7: Wide optimal results

Dataset	Classifier	Metric	G-SMOTENC	NONE	SMOTENC	ROS	RUS
German Credit (41)	LR	G-Mean	**0.602**	0.500	0.592	0.598	0.597
German Credit (41)	RF	OA	**0.976**	**0.976**	0.961	0.969	0.636
German Credit (41)	RF	F-Score	**0.598**	0.494	0.566	0.591	0.413
German Credit (41)	RF	G-Mean	0.621	0.500	**0.622**	0.614	0.572
Heart Disease	DT	OA	0.532	**0.566**	0.509	0.473	0.430
Heart Disease	DT	F-Score	**0.371**	0.322	0.342	0.331	0.295
Heart Disease	DT	G-Mean	**0.588**	0.534	0.563	0.545	0.515
Heart Disease	KNN	OA	0.538	**0.564**	0.535	0.534	0.504
Heart Disease	KNN	F-Score	**0.363**	0.287	0.360	0.352	0.341
Heart Disease	KNN	G-Mean	**0.571**	0.509	**0.571**	0.560	0.557
Heart Disease	LR	OA	0.558	**0.584**	0.557	0.536	0.480
Heart Disease	LR	F-Score	**0.397**	0.329	0.395	0.374	0.333
Heart Disease	LR	G-Mean	0.601	0.539	0.601	**0.603**	0.567
Heart Disease	RF	OA	0.553	**0.601**	0.546	0.539	0.480
Heart Disease	RF	F-Score	**0.385**	0.314	0.366	0.360	0.326
Heart Disease	RF	G-Mean	**0.600**	0.531	0.580	0.569	0.566
Heart Disease (21)	DT	OA	0.532	**0.566**	0.512	0.486	0.431
Heart Disease (21)	DT	F-Score	**0.376**	0.296	0.341	0.336	0.311
Heart Disease (21)	DT	G-Mean	**0.598**	0.509	0.558	0.562	0.538
Heart Disease (21)	KNN	OA	0.561	**0.569**	0.543	0.541	0.491
Heart Disease (21)	KNN	F-Score	**0.385**	0.312	0.365	0.363	0.334
Heart Disease (21)	KNN	G-Mean	**0.589**	0.520	0.570	0.566	0.546
Heart Disease (21)	LR	OA	0.573	**0.592**	0.565	0.547	0.525
Heart Disease (21)	LR	F-Score	**0.408**	0.331	0.405	0.387	0.343
Heart Disease (21)	LR	G-Mean	**0.638**	0.540	0.610	0.602	0.583
Heart Disease (21)	RF	OA	0.577	**0.608**	0.565	0.561	0.517
Heart Disease (21)	RF	F-Score	**0.417**	0.323	0.390	0.383	0.337
Heart Disease (21)	RF	G-Mean	**0.621**	0.536	0.596	0.593	0.567