






Supplementary Materials: An Extensive Performance Comparison between Feature Reduction and Feature Selection Algorithm Preprocessing on Unbalanced Wide Data

Ismael Ramos-Pérez , José Antonio Barbero-Aparicio , Antonio Canepa-Oneto ,
Álvar Arnaiz-González  and and Jesús M. Maudes-Raedo 

This document compiles the tables not included in the main document to simplify the results. These tables are obtained using all five metrics during the experimental process, all of them detailed in the main document.

The tables are divided into three groups:

- Tables [S1–S5](#) show the average rank of all feature reducers used in this study.
- The previous ranks divided by the five classifiers are shown in Tables [S6–S10](#) to see which feature reducer perform better on each classifier.
- Tables [S11–S15](#) show the average rank of each of the seven balancing strategy for both the best feature reducer and feature selector algorithms we compare at the end of the study.

All the table groups are ordered by the metric used to calculate it (MCC, F_1 -Score, Kappa, AUC and G-Mean). In all groups, two slightly different behaviours can be found depending on the metric in use: F_1 -Score, Kappa as first group, AUC and G-Mean as second group. This difference is portrayed in the main document Bayesian tests, where for the first group (F_1 -Score, MCC, and Kappa) the feature reduction option performed significantly better than the feature selection. Meanwhile, the other group (AUC and G-Mean) does not show enough significant difference on the used datasets.

Table S1. Comparison of average ranks using the MCC metric, the 90 possible configurations when mixing our 5 classifiers and the 18 FR preprocessing methods, including as baseline the non-preprocessing option. The color code indicates the type of algorithm, linear unsupervised, linear supervised or non-linear unsupervised .

Classifier	FR algorithm	Avg. Rank
KNN	MMC	1.96
SVM-G	No	2.54
SVM-G	FSCORE	9.21
KNN	FSCORE	9.29
KNN	No	10.71
SVM-G	LLE	10.71
SVM-G	MDS	10.71
SVM-G	MMDS	10.71
RF	FSCORE	13.39
KNN	LLE	13.57
NBayes	FSCORE	14.79
NBayes	No	15.36
KNN	MDS	15.43
KNN	MMDS	15.43
SVM-G	NPE	15.43
KNN	PCA	16.64
SVM-G	SNE	19.36
RF	No	19.39
KNN	NPE	21.00
NBayes	LLE	21.64
SVM-G	Autoencoder	21.86
KNN	SNE	22.07
KNN	LPE	22.79
NBayes	MDS	26.50
NBayes	MMDS	26.50
C4.5	FSCORE	28.11
NBayes	NPE	29.57
NBayes	SNE	29.93
KNN	Autoencoder	30.36
C4.5	No	30.68
C4.5	NPE	30.86
NBayes	PCA	31.36
C4.5	MDS	34.14
SVM-G	LSLS	34.21
KNN	SAVE	34.36
RF	NPE	34.43
NBayes	Autoencoder	34.64
C4.5	PCA	35.21
C4.5	LLE	35.57
C4.5	MMDS	37.00
RF	Autoencoder	38.71
SVM-G	LPE	38.79
C4.5	LPE	40.07
KNN	LSLS	40.50
NBayes	LSLS	42.86

Classifier	FR algorithm	Avg. Rank
RF	LSLS	43.93
C4.5	Autoencoder	46.00
C4.5	LSLS	49.64
C4.5	MMC	49.93
RF	LPE	50.43
SVM-G	SAVE	52.14
RF	SAVE	52.36
SVM-G	LEA	53.43
RF	MMDS	56.64
SVM-G	RNDPROJ	56.64
C4.5	SNE	57.21
RF	MDS	57.64
RF	PCA	58.50
NBayes	LPE	59.43
RF	LLE	59.50
NBayes	LEA	61.29
NBayes	SAVE	62.14
RF	LEA	62.21
RF	MMC	62.50
KNN	LEA	63.21
C4.5	SAVE	65.00
KNN	RNDPROJ	67.43
NBayes	RNDPROJ	67.50
NBayes	MMC	67.93
RF	RNDPROJ	70.50
KNN	LFDA	71.93
RF	SNE	73.89
C4.5	RNDPROJ	76.07
KNN	SLPE	77.00
C4.5	SLPE	78.11
RF	LFDA	78.14
RF	SLPE	78.21
KNN	PFLPP	78.43
RF	PFLPP	78.64
SVM-G	SLPE	78.75
C4.5	LEA	79.29
C4.5	PFLPP	79.29
NBayes	PFLPP	79.29
SVM-G	LFDA	79.29
SVM-G	MMC	79.29
SVM-G	PCA	79.29
SVM-G	PFLPP	79.29
NBayes	SLPE	79.43
C4.5	LFDA	79.79
NBayes	LFDA	80.11

Table S2. Comparison of average ranks using the F_1 -Score metric, the 90 possible configurations when mixing our 5 classifiers and the 18 FR preprocessing methods, including as baseline the non-preprocessing option. The color code indicates the type of algorithm, linear unsupervised, linear supervised or non-linear unsupervised.

Classifier	FR algorithm	Avg. Rank
KNN	MMC	1.89
SVM-G	No	2.71
SVM-G	FSCORE	8.86
KNN	FSCORE	9.00
KNN	No	10.36
SVM-G	LLE	11.64
SVM-G	MDS	11.64
SVM-G	MMDS	11.64
KNN	LLE	12.21
NBayes	No	13.57
NBayes	FSCORE	13.93
RF	FSCORE	14.46
KNN	MDS	15.36
KNN	MMDS	15.36
KNN	PCA	16.21
SVM-G	NPE	18.93
NBayes	LLE	20.36
SVM-G	Autoencoder	21.29
KNN	NPE	21.64
KNN	LPE	22.00
KNN	SNE	22.21
SVM-G	SNE	23.57
NBayes	MDS	24.79
NBayes	MMDS	24.79
C4.5	FSCORE	26.04
RF	No	26.57
NBayes	NPE	26.64
KNN	Autoencoder	28.29
C4.5	No	28.32
NBayes	PCA	29.14
NBayes	Autoencoder	30.93
SVM-G	LSLS	31.43
NBayes	SNE	31.93
C4.5	PCA	33.14
KNN	SAVE	33.71
C4.5	MDS	35.29
C4.5	NPE	35.43
SVM-G	LPE	36.79
C4.5	LPE	38.00
C4.5	MMDS	38.00
C4.5	LLE	38.21
KNN	LSLS	38.79
NBayes	LSLS	38.86
C4.5	Autoencoder	42.57
RF	NPE	42.64

Classifier	FR algorithm	Avg. Rank
SVM-G	SAVE	46.64
C4.5	LSLS	47.43
RF	Autoencoder	47.43
RF	LSLS	52.21
RF	MMC	54.57
RF	SAVE	55.57
C4.5	MMC	56.00
NBayes	RNDPROJ	56.64
SVM-G	MMC	57.29
NBayes	LPE	57.86
C4.5	SAVE	58.00
NBayes	LEA	59.79
KNN	LEA	61.07
SVM-G	LFDA	61.18
KNN	LFDA	61.21
NBayes	MMC	61.79
RF	LFDA	61.79
SVM-G	RNDPROJ	62.11
RF	LPE	62.14
C4.5	SNE	63.07
SVM-G	LEA	63.07
KNN	RNDPROJ	63.93
C4.5	LFDA	64.36
RF	MMDS	67.00
RF	SLPE	67.43
RF	LEA	67.71
RF	MDS	67.86
RF	RNDPROJ	68.43
RF	PCA	68.93
RF	LLE	70.82
RF	PFLPP	71.00
NBayes	SAVE	72.25
SVM-G	PFLPP	73.39
KNN	SLPE	74.29
KNN	PFLPP	74.57
SVM-G	SLPE	75.57
C4.5	RNDPROJ	79.71
RF	SNE	84.00
NBayes	SLPE	84.21
NBayes	LFDA	84.79
C4.5	SLPE	84.82
SVM-G	PCA	85.18
C4.5	LEA	86.25
C4.5	PFLPP	86.25
NBayes	PFLPP	86.25

Table S3. Comparison of average ranks using the Kappa metric, the 90 possible configurations when mixing our 5 classifiers and the 18 FR preprocessing methods, including as baseline the non-preprocessing option. The color code indicates the type of algorithm, linear unsupervised, linear supervised or non-linear unsupervised.

Classifier	FR algorithm	Avg. Rank
KNN	MMC	1.82
SVM-G	No	2.71
SVM-G	FSCORE	8.43
KNN	FSCORE	8.79
KNN	No	10.07
SVM-G	LLE	11.00
SVM-G	MDS	11.00
SVM-G	MMDS	11.00
KNN	LLE	12.93
RF	FSCORE	13.25
NBayes	No	14.86
NBayes	FSCORE	15.00
KNN	MDS	15.46
KNN	MMDS	15.46
SVM-G	NPE	16.43
KNN	PCA	17.00
SVM-G	SNE	20.21
KNN	NPE	20.93
SVM-G	Autoencoder	21.07
KNN	SNE	21.50
RF	No	21.71
NBayes	LLE	22.43
KNN	LPE	22.86
NBayes	MDS	26.29
NBayes	MMDS	26.29
C4.5	FSCORE	27.25
KNN	Autoencoder	28.21
C4.5	No	28.82
NBayes	SNE	30.71
NBayes	NPE	31.07
NBayes	PCA	31.57
C4.5	NPE	32.00
SVM-G	LSLS	32.71
KNN	SAVE	33.43
NBayes	Autoencoder	33.71
C4.5	PCA	33.93
C4.5	MDS	34.07
C4.5	LLE	35.93
RF	NPE	37.00
C4.5	MMDS	37.21
SVM-G	LPE	37.43
C4.5	LPE	38.93
KNN	LSLS	39.57
RF	Autoencoder	40.64
NBayes	LSLS	41.86

Classifier	FR algorithm	Avg. Rank
C4.5	Autoencoder	44.21
RF	LSLS	45.43
C4.5	LSLS	47.71
SVM-G	SAVE	50.29
RF	SAVE	51.07
RF	LPE	52.86
SVM-G	LEA	52.93
C4.5	MMC	55.00
SVM-G	RNDPROJ	57.00
C4.5	SNE	57.07
RF	MMDS	59.00
NBayes	LPE	59.21
RF	MDS	59.57
RF	PCA	59.57
NBayes	LEA	60.57
RF	LLE	61.21
RF	MMC	61.79
KNN	LEA	62.00
RF	LEA	62.07
NBayes	SAVE	63.36
C4.5	SAVE	64.29
KNN	RNDPROJ	66.00
NBayes	RNDPROJ	67.29
NBayes	MMC	67.50
RF	RNDPROJ	69.00
KNN	LFDA	71.29
RF	SNE	74.75
C4.5	RNDPROJ	75.43
C4.5	SLPE	77.96
RF	SLPE	78.07
RF	PFLPP	78.43
KNN	PFLPP	78.50
KNN	SLPE	78.50
SVM-G	SLPE	78.75
C4.5	LEA	79.21
C4.5	PFLPP	79.21
NBayes	PFLPP	79.21
RF	LFDA	79.21
SVM-G	LFDA	79.21
SVM-G	MMC	79.21
SVM-G	PCA	79.21
SVM-G	PFLPP	79.21
NBayes	SLPE	79.29
C4.5	LFDA	79.71
NBayes	LFDA	80.04

Table S4. Comparison of average ranks using the AUC metric, the 90 possible configurations when mixing our 5 classifiers and the 18 FR preprocessing methods, including as baseline the non-preprocessing option. The color code indicates the type of algorithm, linear unsupervised, linear supervised or non-linear unsupervised.

Classifier	FR algorithm	Avg. Rank
RF	No	3.32
SVM-G	No	3.57
KNN	MMC	4.14
KNN	FSCORE	7.11
RF	FSCORE	7.54
SVM-G	FSCORE	8.82
KNN	No	11.46
RF	LSLS	15.86
KNN	LLE	16.14
KNN	PCA	17.00
SVM-G	LLE	17.39
SVM-G	MDS	17.39
SVM-G	MMDS	17.39
KNN	MDS	19.14
KNN	MMDS	19.14
NBayes	No	19.36
RF	LPE	19.71
NBayes	LLE	21.36
KNN	LPE	22.00
SVM-G	Autoencoder	25.07
SVM-G	NPE	25.39
KNN	SNE	25.50
NBayes	FSCORE	25.64
NBayes	MDS	27.00
NBayes	MMDS	27.00
KNN	NPE	27.14
KNN	SAVE	28.79
C4.5	FSCORE	29.32
SVM-G	SNE	30.11
NBayes	NPE	31.00
C4.5	No	31.25
C4.5	PCA	31.79
KNN	Autoencoder	32.50
SVM-G	LSLS	33.21
NBayes	SNE	33.50
C4.5	LPE	34.07
NBayes	Autoencoder	34.14
RF	PCA	34.43
NBayes	PCA	36.86
C4.5	NPE	38.00
KNN	LSLS	38.07
C4.5	MDS	39.14
RF	SAVE	40.00
C4.5	MMDS	41.07
C4.5	LLE	41.14

Classifier	FR algorithm	Avg. Rank
SVM-G	LPE	44.18
SVM-G	RNDPROJ	44.29
RF	NPE	44.46
C4.5	Autoencoder	45.57
NBayes	LSLS	45.86
RF	Autoencoder	46.00
NBayes	SAVE	47.86
C4.5	LSLS	49.86
C4.5	MMC	52.57
SVM-G	LEA	55.14
C4.5	SNE	57.36
SVM-G	SAVE	57.43
NBayes	LEA	58.43
RF	MMDS	58.86
RF	MDS	59.29
RF	RNDPROJ	60.00
RF	LLE	60.36
KNN	LEA	60.57
RF	LEA	60.79
KNN	RNDPROJ	63.71
NBayes	MMC	65.07
NBayes	LPE	66.14
C4.5	RNDPROJ	67.43
RF	SNE	68.79
C4.5	SAVE	70.93
C4.5	LEA	71.68
C4.5	PFLPP	71.68
NBayes	PFLPP	71.68
NBayes	SLPE	71.68
C4.5	SLPE	72.57
NBayes	LFDA	72.79
SVM-G	PCA	73.54
NBayes	RNDPROJ	76.07
RF	MMC	77.43
KNN	PFLPP	78.50
KNN	SLPE	78.93
KNN	LFDA	79.29
C4.5	LFDA	82.25
RF	PFLPP	82.50
RF	SLPE	83.29
RF	LFDA	83.57
SVM-G	SLPE	83.86
SVM-G	PFLPP	86.04
SVM-G	LFDA	88.46
SVM-G	MMC	89.29

Table S5. Comparison of average ranks using the G-mean metric, the 90 possible configurations when mixing our 5 classifiers and the 18 FR preprocessing methods, including as baseline the non-preprocessing option. The color code indicates the type of algorithm, linear unsupervised, linear supervised or non-linear unsupervised.

Classifier	FR algorithm	Avg. Rank
RF	No	3.32
SVM-G	No	3.50
KNN	MMC	4.14
KNN	FSCORE	7.04
RF	FSCORE	7.25
SVM-G	FSCORE	8.68
KNN	No	11.46
KNN	LLE	15.79
KNN	PCA	15.86
NBayes	No	18.00
SVM-G	LLE	18.36
SVM-G	MDS	18.36
SVM-G	MMDS	18.36
KNN	MDS	19.29
KNN	MMDS	19.29
RF	LSLS	20.43
NBayes	LLE	20.50
KNN	LPE	21.07
NBayes	FSCORE	23.71
SVM-G	Autoencoder	24.64
KNN	SNE	25.71
SVM-G	NPE	26.07
NBayes	MDS	26.29
NBayes	MMDS	26.29
KNN	NPE	26.71
KNN	SAVE	26.86
RF	LPE	27.21
C4.5	FSCORE	27.32
C4.5	No	29.68
C4.5	PCA	30.07
SVM-G	SNE	31.07
SVM-G	LSLS	31.29
NBayes	NPE	31.36
KNN	Autoencoder	31.64
NBayes	Autoencoder	32.00
C4.5	LPE	32.64
NBayes	PCA	33.86
NBayes	SNE	34.79
KNN	LSLS	35.00
C4.5	NPE	39.57
RF	SAVE	39.79
C4.5	MDS	40.00
C4.5	MMDS	41.21
NBayes	LSLS	42.07
C4.5	LLE	43.00

Classifier	FR algorithm	Avg. Rank
C4.5	Autoencoder	44.21
SVM-G	LPE	44.46
RF	PCA	45.57
RF	NPE	46.93
C4.5	LSLS	47.29
RF	Autoencoder	48.86
SVM-G	RNDPROJ	49.54
C4.5	MMC	49.64
SVM-G	SAVE	51.14
NBayes	SAVE	55.89
NBayes	MMC	56.79
KNN	LEA	56.93
RF	RNDPROJ	57.71
KNN	RNDPROJ	58.21
C4.5	SAVE	60.14
NBayes	LEA	60.71
C4.5	SNE	60.79
NBayes	LPE	61.36
RF	MMC	61.36
NBayes	RNDPROJ	62.86
SVM-G	LEA	64.14
KNN	LFDA	64.29
RF	LEA	65.00
RF	LFDA	67.07
RF	MMDS	67.21
RF	MDS	67.71
C4.5	LFDA	68.07
RF	LLE	70.39
RF	SLPE	72.00
KNN	PFLPP	73.14
RF	PFLPP	74.00
KNN	SLPE	74.21
C4.5	RNDPROJ	75.68
RF	SNE	80.64
NBayes	SLPE	81.86
C4.5	SLPE	82.71
NBayes	LFDA	83.07
SVM-G	SLPE	83.36
C4.5	LEA	84.21
C4.5	PFLPP	84.21
NBayes	PFLPP	84.21
SVM-G	LFDA	84.21
SVM-G	MMC	84.21
SVM-G	PCA	84.21
SVM-G	PFLPP	84.21

Table S6. Comparison of average ranks using the MCC metric, the 18 FR preprocessing methods, including as baseline the non preprocessing option. Each ranking is performed by a different classifier in order to detect what is more suitable. The color code indicates the type of algorithm, **linear unsupervised**, **linear supervised** or **non-linear unsupervised**.

(a) KNN

Feature reducer	Avg. Rank
MMC	1.07
FSCORE	3.86
No	4.57
LLE	5.29
MDS	5.93
MMDS	5.93
PCA	6.29
SNE	7.50
LPPE	7.86
NPE	8.07
Autoencoder	10.79
SAVE	11.29
LSLS	12.57
LEA	14.57
RNDPROJ	15.43
LFDA	16.14
SLPE	16.79
PFLPP	17.07

(b) SVM-G

Feature reducer	Avg. Rank
No	1.07
FSCORE	3.86
LLE	4.21
MDS	4.21
MMDS	4.21
NPE	5.93
Autoencoder	6.64
SNE	7.21
LSLS	8.79
LPPE	9.71
SAVE	11.57
LEA	11.86
RNDPROJ	12.43
SLPE	15.71
LFDA	15.89
MMC	15.89
PCA	15.89
PFLPP	15.89

(c) C4.5

Feature reducer	Avg. Rank
FSCORE	3.46
NPE	3.57
MDS	4.43
No	4.54
LLE	4.71
PCA	5.36
MMDS	5.64
LPPE	6.14
Autoencoder	8.79
LSLS	9.64
MMC	9.93
SNE	11.93
SAVE	13.36
RNDPROJ	15.25
SLPE	15.82
LEA	16.04
PFLPP	16.04
LFDA	16.36

(d) RF

Feature reducer	Avg. Rank
FSCORE	1.29
No	2.07
NPE	3.50
Autoencoder	4.50
LSLS	6.14
LPPE	7.43
SAVE	8.36
MMDS	9.36
MDS	9.64
PCA	10.14
LEA	11.00
LLE	11.00
MMC	11.36
RNDPROJ	12.93
SNE	14.93
LFDA	15.64
SLPE	15.64
PFLPP	16.07

(e) NBayes

Feature reducer	Avg. Rank
FSCORE	2.86
No	3.14
LLE	3.79
MDS	5.71
MMDS	5.71
NPE	6.14
SNE	6.21
PCA	6.29
Autoencoder	7.00
LSLS	9.43
LPPE	12.50
SAVE	12.71
LEA	13.00
MMC	13.86
RNDPROJ	14.00
PFLPP	16.14
SLPE	16.21
LFDA	16.29

Table S11. Average ranks using the MCC metric of the balancing strategies for (a) the best configuration that uses an FS method and (b) the best configuration that uses an FR method.

(a) SVM-RFE + SVM-G

Balacing		Avg. Rank
Prior	Posterior	
ROS	No	3.11
No	ROS	3.11
No	SMOTE	3.46
No	No	3.57
SMOTE	No	3.86
No	RUS	4.25
RUS	No	6.64

(b) MMC + KNN

Balacing		Avg. Rank
Prior	Posterior	
No	No	3.50
No	ROS	3.50
No	SMOTE	3.50
No	RUS	3.68
SMOTE	No	4.00
ROS	No	4.29
RUS	No	5.54

Table S7. Comparison of average ranks using the F₁-Score metric, the 18 FR preprocessing methods, including as baseline the non preprocessing option. Each ranking is performed by a different classifier in order to detect what is more suitable. The color code indicates the type of algorithm, linear unsupervised, linear supervised or non-linear unsupervised.

(a) KNN

Feature reducer	Avg. Rank
MMC	1.07
FSCORE	3.93
No	4.57
LLE	4.79
MDS	5.79
MMDS	5.79
PCA	6.07
SNE	8.00
LPE	8.21
NPE	8.21
Autoencoder	10.64
SAVE	11.43
LSLS	12.50
LFDA	14.79
LEA	14.93
RNDPROJ	15.57
SLPE	17.29
PFLPP	17.43

(b) SVM-G

Feature reducer	Avg. Rank
No	1.18
FSCORE	3.25
LLE	4.36
MDS	4.36
MMDS	4.36
Autoencoder	6.29
NPE	6.50
SNE	7.57
LSLS	8.50
LPE	9.79
SAVE	11.21
MMC	12.79
LFDA	13.68
RNDPROJ	13.89
LEA	13.93
PFLPP	15.54
SLPE	16.07
PCA	17.75

(c) C4.5

Feature reducer	Avg. Rank
FSCORE	2.82
No	3.68
PCA	4.71
MDS	4.93
NPE	5.07
LLE	5.86
MMDS	5.86
LPE	5.93
Autoencoder	7.71
LSLS	9.14
MMC	11.64
SAVE	11.93
SNE	12.86
LFDA	13.07
RNDPROJ	15.32
SLPE	16.61
LEA	16.93
PFLPP	16.93

(d) RF

Feature reducer	Avg. Rank
FSCORE	1.21
No	2.93
NPE	3.93
Autoencoder	5.50
MMC	7.00
LSLS	7.29
SAVE	7.71
LFDA	9.43
LPE	9.79
SLPE	11.64
MDS	11.93
MMDS	12.00
LEA	12.21
RNDPROJ	12.21
PCA	12.57
PFLPP	12.64
LLE	13.50
SNE	17.50

(e) NBayes

Feature reducer	Avg. Rank
FSCORE	2.86
No	3.07
LLE	3.71
MDS	5.50
MMDS	5.50
NPE	6.14
PCA	6.36
Autoencoder	6.86
SNE	7.21
LSLS	9.00
RNDPROJ	12.14
MMC	12.21
LPE	12.50
LEA	12.79
SAVE	14.50
SLPE	16.71
LFDA	16.86
PFLPP	17.07

Table S8. Comparison of average ranks using the Kappa metric, the 18 FR preprocessing methods, including as baseline the non preprocessing option. Each ranking is performed by a different classifier in order to detect what is more suitable. The color code indicates the type of algorithm, linear unsupervised, linear supervised or non-linear unsupervised.

(a) KNN

Feature reducer	Avg. Rank
MMC	1.07
FSCORE	3.86
No	4.57
LLE	4.93
MDS	5.89
MMDS	5.89
PCA	6.21
SNE	7.50
NPE	8.14
LPE	8.29
Autoencoder	10.79
SAVE	11.29
LSLS	12.57
LEA	14.64
RNDPROJ	15.43
LFDA	16.14
SLPE	16.79
PFLPP	17.00

(b) SVM-G

Feature reducer	Avg. Rank
No	1.18
FSCORE	3.25
LLE	4.43
MDS	4.43
MMDS	4.43
NPE	6.07
Autoencoder	6.29
SNE	7.29
LSLS	8.64
LPE	9.71
SAVE	11.43
LEA	12.00
RNDPROJ	12.57
SLPE	15.71
LFDA	15.89
MMC	15.89
PCA	15.89
PFLPP	15.89

(c) C4.5

Feature reducer	Avg. Rank
FSCORE	3.32
NPE	4.07
No	4.11
MDS	4.64
LLE	5.00
PCA	5.14
MMDS	5.57
LPE	5.93
Autoencoder	8.36
LSLS	9.21
MMC	11.43
SNE	11.64
SAVE	13.21
RNDPROJ	15.18
SLPE	15.82
LEA	16.04
PFLPP	16.04
LFDA	16.29

(d) RF

Feature reducer	Avg. Rank
FSCORE	1.21
No	2.21
NPE	3.50
Autoencoder	4.50
LSLS	6.00
SAVE	7.36
LPE	7.71
MDS	10.14
MMDS	10.21
PCA	10.29
LEA	10.86
MMC	10.86
LLE	11.14
RNDPROJ	12.79
SNE	15.07
SLPE	15.57
LFDA	15.71
PFLPP	15.86

(e) NBayes

Feature reducer	Avg. Rank
No	2.71
FSCORE	2.93
LLE	4.00
MDS	5.71
MMDS	5.71
PCA	6.36
NPE	6.50
SNE	6.57
Autoencoder	6.79
LSLS	9.21
LPE	12.29
LEA	12.86
SAVE	13.00
MMC	13.71
RNDPROJ	14.00
PFLPP	16.14
SLPE	16.21
LFDA	16.29

Table S9. Comparison of average ranks using the AUC metric, the 18 FR preprocessing methods, including as baseline the non preprocessing option. Each ranking is performed by a different classifier in order to detect what is more suitable. The color code indicates the type of algorithm, linear unsupervised, linear supervised or non-linear unsupervised.

(a) KNN

Feature reducer	Avg. Rank
MMC	1.07
FSCORE	3.93
No	4.57
LLE	4.79
MDS	5.79
MMDS	5.79
PCA	6.07
SNE	8.00
LPE	8.21
NPE	8.21
Autoencoder	10.64
SAVE	11.43
LSLS	12.50
LFDA	14.79
LEA	14.93
RNDPROJ	15.57
SLPE	17.29
PFLPP	17.43

(b) SVM-G

Feature reducer	Avg. Rank
No	1.18
FSCORE	3.25
LLE	4.36
MDS	4.36
MMDS	4.36
Autoencoder	6.29
NPE	6.50
SNE	7.57
LSLS	8.50
LPE	9.79
SAVE	11.21
MMC	12.79
LFDA	13.68
RNDPROJ	13.89
LEA	13.93
PFLPP	15.54
SLPE	16.07
PCA	17.75

(c) C4.5

Feature reducer	Avg. Rank
FSCORE	2.82
No	3.68
PCA	4.71
MDS	4.93
NPE	5.07
LLE	5.86
MMDS	5.86
LPE	5.93
Autoencoder	7.71
LSLS	9.14
MMC	11.64
SAVE	11.93
SNE	12.86
LFDA	13.07
RNDPROJ	15.32
SLPE	16.61
LEA	16.93

(d) RF

Feature reducer	Avg. Rank
FSCORE	1.21
No	2.93
NPE	3.93
Autoencoder	5.50
MMC	7.00
LSLS	7.29
SAVE	7.71
LFDA	9.43
LPE	9.79
SLPE	11.64
MDS	11.93
MMDS	12.00
LEA	12.21
RNDPROJ	12.21
PCA	12.57
PFLPP	12.64
LLE	13.50
SNE	17.50

(e) NBayes

Feature reducer	Avg. Rank
FSCORE	2.86
No	3.07
LLE	3.71
MDS	5.50
MMDS	5.50
NPE	6.14
PCA	6.36
Autoencoder	6.86
SNE	7.21
LSLS	9.00
RNDPROJ	12.14
MMC	12.21
LPE	12.50
LEA	12.79
SAVE	14.50
SLPE	16.71
LFDA	16.86
PFLPP	17.07

Table S10. Comparison of average ranks using the G-mean metric, the 18 FR preprocessing methods, including as baseline the non preprocessing option. Each ranking is performed by a different classifier in order to detect what is more suitable. The color code indicates the type of algorithm, linear unsupervised, linear supervised or non-linear unsupervised.

(a) KNN

Feature reducer	Avg. Rank
MMC	1.50
FSCORE	2.71
No	4.36
PCA	5.36
LLE	5.71
LPE	6.64
MDS	7.36
MMDS	7.36
SNE	8.21
SAVE	9.57
NPE	9.71
Autoencoder	11.21
LSLS	11.29
LEA	14.50
RNDPROJ	14.71
LFDA	15.79
PFLPP	17.43
SLPE	17.57

(b) SVM-G

Feature reducer	Avg. Rank
No	1.18
FSCORE	2.46
LLE	4.79
MDS	4.79
MMDS	4.79
Autoencoder	6.21
NPE	7.07
LSLS	7.57
SNE	8.36
LPE	9.79
RNDPROJ	10.54
SAVE	11.14
LEA	12.57
SLPE	15.75
LFDA	16.00
MMC	16.00
PCA	16.00
PFLPP	16.00

(c) C4.5

Feature reducer	Avg. Rank
FSCORE	3.18
No	3.32
PCA	3.57
LPE	4.79
MDS	6.14
NPE	6.14
MMDS	6.71
LLE	6.93
Autoencoder	7.79
LSLS	8.29
MMC	9.86
SAVE	12.29
SNE	12.57
LFDA	13.86
RNDPROJ	15.04
SLPE	16.68
LEA	16.93
PFLPP	16.93

(d) RF

Feature reducer	Avg. Rank
No	1.61
FSCORE	2.39
LSLS	4.00
LPE	4.64
SAVE	5.79
NPE	6.43
Autoencoder	7.29
PCA	8.64
RNDPROJ	9.57
MMC	10.14
LEA	11.71
LFDA	12.29
MMDS	12.79
MDS	12.86
LLE	14.07
SLPE	14.64
PFLPP	15.07
SNE	17.07

(e) NBayes

Feature reducer	Avg. Rank
LLE	3.50
No	3.57
MDS	5.14
MMDS	5.14
FSCORE	5.21
NPE	6.07
Autoencoder	6.43
PCA	6.86
SNE	7.07
LSLS	9.14
SAVE	11.57
MMC	12.00
LEA	12.71
LPE	13.00
RNDPROJ	13.14
SLPE	16.50
LFDA	16.86
PFLPP	17.07

Table S12. Average ranks using the F₁-Score metric of the balancing strategies for (a) the best configuration that uses an FS method and (b) the best configuration that uses an FR method.

(a) SVM-RFE + SVM-G

Balacing		Avg. Rank
Prior	Posterior	
ROS	No	2.96
No	ROS	3.11
No	SMOTE	3.61
No	No	3.86
SMOTE	No	3.86
No	RUS	4.11
RUS	No	6.50

(b) MMC + KNN

Balacing		Avg. Rank
Prior	Posterior	
No	No	3.29
No	ROS	3.29
No	SMOTE	3.29
No	RUS	3.46
ROS	No	4.57
SMOTE	No	4.57
RUS	No	5.54

Table S13. Average ranks using the Kappa metric of the balancing strategies for (a) the best configuration that uses an FS method and (b) the best configuration that uses an FR method.

(a) SVM-RFE + SVM-G

Balacing		Avg. Rank
Prior	Posterior	
ROS	No	3.04
No	ROS	3.18
No	SMOTE	3.46
No	No	3.71
SMOTE	No	3.86
No	RUS	4.25
RUS	No	6.50

(b) MMC + KNN

Balacing		Avg. Rank
Prior	Posterior	
No	No	3.43
No	ROS	3.43
No	SMOTE	3.43
No	RUS	3.61
SMOTE	No	4.00
ROS	No	4.57
RUS	No	5.54

Table S14. Average ranks using the AUC metric of the balancing strategies for (a) the best configuration that uses an FS method and (b) the best configuration that uses an FR method.

(a) SVM-RFE + SVM-G

Balacing		Avg. Rank
Prior	Posterior	
No	No	3.04
No	ROS	3.25
No	SMOTE	3.46
ROS	No	3.68
RUS	No	4.68
SMOTE	No	4.86
No	RUS	5.04

(b) MMC + KNN

Balacing		Avg. Rank
Prior	Posterior	
SMOTE	No	1.57
ROS	No	2.68
No	No	4.11
No	ROS	4.11
No	SMOTE	4.11
No	RUS	4.61
RUS	No	6.82

Table S15. Average ranks using the G-Mean metric of the balancing strategies for (a) the best configuration that uses an FS method and (b) the best configuration that uses an FR method.

(a) SVM-RFE + SVM-G

Balacing		Avg. Rank
Prior	Posterior	
No	No	3.04
No	ROS	3.25
No	SMOTE	3.46
ROS	No	3.68
RUS	No	4.68
SMOTE	No	4.86
No	RUS	5.04

(b) MMC + KNN

Balacing		Avg. Rank
Prior	Posterior	
SMOTE	No	1.57
ROS	No	2.68
No	No	4.11
No	ROS	4.11
No	SMOTE	4.11
No	RUS	4.61
RUS	No	6.82