TABLE VIII
INTRINSIC CHARACTERISTICS OF OVERSAMPLING METHODS. THE SIGN "●" INDICATES THE PRESENCE OF A SPECIFIC PROPERTY, WHILE "○" INDICATES ITS ABSENCE.

Properties	ROS	SMOTE	SMOTE+TL	SMOTE+ENN	CBO+Random	Borderline-SMOTE1	Borderline-SMOTE-2	AHC
Replication/ Synthesization of examples	Replication	Synthesization	Synthesization	Synthesization	Replication	Synthesization	Synthesization	Synthesization
Takes into account the majority examples neighbourhood	○	○	●	●	Not directly, but through clustering	●	●	Not directly, but through clustering
Considers a taxonomy of minority data	○	○	○	○	○	Noise, Danger, Safe	Noise, Danger, Safe	○
Overlapping is performed in specific area(s)	○	○	○	○	○	Borderline Regions	Borderline Regions	○
Cluster-based Oversampling	○	○	○	○	●	○	○	●
Oversampling of minority class	●	●	●	●	●	●	●	●
Oversampling of majority class	○	○	○	○	●	○	○	○
Minority examples are assigned different weights	○	○	○	○	○	○	○	○
Neighbourhood-based oversampling	○	●	●	●	●	●	●	●
Includes a cleaning-based procedure	○	○	●	●	○	○	○	○
SMOTE-based synthesization	○	●	●	●	○	●	SMOTE-like, but also considering the nearest majority neighbour	○
Performs a filtering procedure	○	○	○	○	○	Noise and Safe examples are not oversampled	Noise and Safe examples are not oversampled	○
Provides perfect balancing	●	●	●	●	●	●	●	●
Advantages	Simplest of oversampling techniques	Allows generation of synthetic examples, creating larger and less specific decision regions	Alleviates SMOTE's problem of overgeneralization	Alleviates SMOTE's problem of overgeneralization. Provides a deeper cleaning than SMOTE+TL.	Eases the problem of small disjuncts	Strengthens the borderline minority examples		Considers the structure of data (both minority and majority examples), through clustering.
Disadvantages	Prone to overfitting, due to replication of a random subset of minority examples.	Overgeneralization. May generate instances in overlapping and noise regions. Definition of k-neighbourhood	May augment unnecessary safe examples while also enlarging noisy regions.		Prone to overfitting, due to ROS. Definition of the number of clusters	May generate instances in overlapping and noise regions. The criterion to identify borderline examples may fail in some scenarios. Definition of k-neighbourhood		Computationally expensive

Table VIII: Continued from previous page.

Properties	ADASYN	SPIDER1	SPIDER2	ADOMS	Safe-Level-SMOTE	CBO+SMOTE	MWMOTE
Replication/ Synthesization of examples	Synthesization	Replication	Replication	Synthesization	Synthesization	Synthesization	Synthesization
Takes into account the majority examples neighbourhood	●	●	●	○	●	Not directly, but through clustering	●
Considers a taxonomy of minority data	○	Both minority and majority examples are flagged as Noise or Safe	Both minority and majority examples are flagged as Noise or Safe	○	Safe and Noise	○	Noise, Borderline, Sparse and Dense clusters
Overlapping is performed in specific area(s)	○	○	○	○	Safe Regions	○	●
Cluster-based Oversampling	○	○	○	○	○	●	●
Oversampling of minority class	●	●	●	●	●	●	●
Oversampling of majority class	○	○	○	○	○	●	○
Minority examples are assigned different weights	w_i	○	○	○	sl_{ratio}	○	S_w
Neighbourhood-based oversampling	●	●	●	Computes PCA of local data distribution	●	●	●
Includes a cleaning-based procedure	○	●	●	○	○	○	○
SMOTE-based synthesization	●	○	○	●	●	●	SMOTE-like, in clusters
Performs a filtering procedure	○	○	○	○	○	○	Noise examples are not oversampled
Provides perfect balancing	●	○	○	●	●	●	●
Advantages	Minority examples surrounded by majority examples are oversampled more often: decision boundary is more focused on these difficult examples	When relabelling is used, the oversampling procedure is similar to SMOTE, without the problem of overgeneralization	Addresses the deterioration of majority class found in SPIDER	Considers the k-neighbourhood of minority data more properly.	Strengthens the safe minority examples, easing the problem of small disjuncts. Avoids the augmentation of noise regions.	Eases the problem of small disjuncts. Eases the problem of overgeneralization.	Weights of minority examples depend on their importance for classification. Alleviates the problem of small disjuncts. Avoids the problem of SMOTE-based sintetization of samples
Disadvantages	Parameter used to define weights for minority class could be inappropriate. Definition of k-neighbourhood	Choice of amplification type: may augment noisy regions or cause a deterioration in the majority class. Replication of existing minority examples. Re-labelling examples might not be acceptable in some domains.	Replication of existing minority examples. Re-labelling examples might not be acceptable in some domains. May replicate undesired noise.	Same issues of SMOTE by not considering the distribution of majority examples	Definition of k-neighbourhood May generate inconsistent data.	Definition of the number of clusters	Need to specify a threshold for clustering procedure. Definition of k-neighbourhood