

Empowering One-vs-One Decomposition with Ensemble Learning for Multi-Class Imbalanced Data

Zhongliang Zhang^a, Bartosz Krawczyk^{b,*}, Salvador García^c, Alejandro Rosales-Pérez^d, Francisco Herrera^{c,e}

^a*School of Information Science and Engineering, Key Laboratory of Integrated Automation of Process Industry, Northeastern University, Shenyang 110819, China.*

^b*Department of Systems and Computer Networks, Wrocław University of Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland.*

^c*Department of Computer Science and Artificial Intelligence, University of Granada, P.O. Box 18071, Granada, Spain.*

^d*Instituto Nacional de Astrofísica, Óptica y Electrónica, Computer Science Department, Luis E. Erro No. 1, Santa María Tonantzintla, Puebla, 72840, Mexico.*

^e*Faculty of Computing and Information Technology, King Abdulaziz University, 21589, Jeddah, Saudi Arabia.*

Abstract

Multi-class imbalance classification problems occur in many real-world applications, which suffer from the quite different distribution of classes. Decomposition strategies are well-known techniques to address the classification problems involving multiple classes. Among them binary approaches using one-vs-one and one-vs-all has gained a significant attention from the research community. They allow to divide multi-class problems into several easier-to-solve two-class sub-problems. In this study we develop an exhaustive empirical analysis to explore the possibility of empowering the one-vs-one scheme for multi-class imbalance classification problems with applying binary ensemble learning approaches. We examine several state-of-the-art ensemble learning methods proposed for addressing the imbalance problems to solve the pairwise tasks derived from the multi-class data set. Then the aggregation strategy is employed to combine the binary ensemble outputs to

*Corresponding author

Email addresses: zz119860210@126.com (Zhongliang Zhang), bartosz.krawczyk@pwr.edu.pl (Bartosz Krawczyk), salvag1@decsai.ugr.es (Salvador García), arosales@inaoep.mx (Alejandro Rosales-Pérez), herrera@decsai.ugr.es (Francisco Herrera)

Preprint submitted to Knowledge-Based Systems

April 1, 2016

reconstruct the original multi-class task. We present a detailed experimental study of the proposed approach, supported by the statistical analysis. The results indicate the high effectiveness of ensemble learning with one-vs-one scheme in dealing with the multi-class imbalance classification problems.

Keywords: Multi-class classification, Imbalanced data, Ensemble learning, Binary decomposition, Classifier combination

1. Introduction

In machine learning and data mining, while one or more classes are underrepresented in the data set, it is called as class imbalance classification. Many real-world classification tasks suffer from the class imbalance problem, which is considered as one of the important challenges for the data mining community [18]. The main difficulty of these problems is that the skewed distribution makes conventional classification algorithms less effective, since standard learning algorithms consider a balanced training data set, which result in making it harder to predict minority class examples [50].

In recent years, many efforts have been focused on the binary class imbalance problems [31, 41], which only contain two classes. However, multi-class imbalance classification, is widely applied in many areas, such as text categorization [47], human activity recognition [1] and medical diagnosis [35]. Unfortunately, it may be invalid to directly apply the solutions proposed for the two-class problems to the multi-class imbalance problems, and some algorithms cannot be used to solve the multi-class imbalance problems directly [18].

Fortunately, in the research community, decomposition strategies turn up to deal with multi-class classification problem. In this solution framework, the multi-class classification problems are transformed into binary class sub-problems, which are much easier to discriminate [61, 53]. Such well-known approaches are the one versus one (OVO) [33, 25] and one versus all (OVA) [7]. As OVA introduces an artificial class imbalance (e.g., for 10 class problem with roughly equally represented classes, the binary sub-problem will have an imbalance ratio 1:9), it is not advisable to use it for handling problems with initially skewed distributions [46].

In this paper, we focus on multi-class imbalance classification problems and develop a complete empirical study to explore the effectiveness of ensemble learning methods [62] in the multi-class imbalanced datasets with

30 OVO scheme, where binary-class classifiers are trained from the subset con-
31 taining each pair of classes by ensemble learning approaches based on data
32 preprocessing [23]. Our initial works in this domain showed that empower-
33 ing binary decomposition with pairwise ensemble learning can significantly
34 improve mining imbalanced multi-class problems [34].

35 Regarding ensemble learning methods, six state-of-the-art approaches
36 are selected to carry out the experiment: UnderBagging [3], SMOTEBag-
37 ging [56] [15], RUSBoost [49], SMOTEBoost [10], SMOTE+AdaBoost [40],
38 EasyEnsemble [40]. Additionally, to show the efficiency of ensemble learn-
39 ing with OVO scheme for addressing the multi-class imbalance problems,
40 the original data preprocessing strategies, including random under-sampling
41 (RUS) [4], random over-sampling (ROS) [4] and synthetic minority oversam-
42 pling technique (SMOTE) [8], are also implemented in the OVO scheme for
43 our comparative analysis.

44 Finally, we carry out a thorough experimental study that supports the
45 effectiveness of our methodology. Concretely, 20 multi-class imbalanced data
46 sets are selected from the UCI repository in our experiment. The average
47 accuracy rate [20] is used as the performance measures in this study. In
48 order to analyze the results obtained from the different solutions, statistical
49 analysis suggested in [28] is given to support the significance of the results.

50 The main contributions of this paper with respect to previous studies are
51 as following:

- 52 • We propose to enhance the OVO scheme for multi-class imbalanced
53 data by using ensemble techniques for each sub-problem.
- 54 • We show, how to extend the area of applicability of binary imbalanced
55 ensemble classifiers to handling far more challenging multi-class imbal-
56 anced scenarios.
- 57 • We develop a complete experimental study of comparison of the state-
58 of-the-art ensemble learning techniques with conventional resampling
59 methods with OVO strategy and state-of-the-art solutions for multi-
60 class imbalance problems.
- 61 • In order to obtain the impacts of the base classifier used in our sce-
62 nario, we choose three different algorithms, including Classification
63 and Regression tree (CART) [6], Back Propagation Neural Network
64 (BPNN) [17] and Support Vector Machine (SVM) [54].

65 The rest of this paper is organized as follows. The background of this
66 study is introduced in Section 2, including multi-class imbalance classifica-
67 tion problems and decomposition strategies. Next, in Section 3 we present
68 the framework of our methodology of ensemble learning with OVO scheme
69 for dealing with multi-class imbalance classification problems. In section 4,
70 the experimental framework is given, including the data sets, the base clas-
71 sifiers and the relative parameters setting, the performance measures and
72 the statistical tests. The complete empirical study is presented in Section 5.
73 Lessons learned from the paper are given in Section 6, while conclusions and
74 potential directions for future works are to be found in the final Section.

75 2. Background

76 In this section, we first introduce the problem of multi-class imbalance
77 classification. Then, we present the solutions for addressing the imbalance
78 problems. Finally, we describe the decomposition strategy for dealing with
79 multi-class classification problems.

80 2.1. Multi-class Imbalanced Data Analysis

81 Multi-class imbalanced data sets, where there are much more instances
82 of some classes (referred to as the majority classes) than others (referred to
83 as the minority classes), is one of the most challenging problems with data
84 quality that always reduces classification performance in machine learning
85 and data mining [57]. The minority classes are usually the most important
86 concepts to recognize, since they represent the rare cases [59]. Additionally,
87 it is expensive or hard to select these examples [58].

88 However, standard classification algorithms are designed with the premise
89 of a balanced training set [42]. With such a precondition, it is much more
90 difficult for the classical classification algorithms to deal with class imbalance
91 problems, especially for identifying the minority class instance [9]. Addition-
92 ally, most of the methods however are specific to address the binary class
93 imbalance problems. Obviously, multi-class imbalance problems are far more
94 complex, since these issues are involved with large number of classes and the
95 relationships among the classes are complicated. Furthermore, it is hard to
96 distinguish between minority classes and noise examples and the minority
97 classes can be ignored by the classifier as the noise examples.

98 *2.2. Solutions for Imbalanced Classification Problems*

99 To overcome the dilemma of skewed class distribution, a large amount of
100 techniques have been developed to deal with such problem. These proposals
101 can be roughly categorized into four groups:

- 102 • Data level: the origin of the problem is the class distribution in the data
103 sets, therefore, it is natural to consider of rebalancing by sampling the
104 data space to reduce the impact of class imbalance, known as an external
105 approach. One of the advantages of such solution is independent
106 from the classifier used, so they are also considered as pre-sampling
107 method [27, 51].
- 108 • Algorithm level: these solutions try to adopt appropriate decision thresh-
109 old to reinforce the learning towards the minority class instances. The
110 proposed algorithms that take the class imbalance into consideration
111 belong to such techniques. They are defined as internal approaches
112 in some papers [52, 11], since the effect depends on the problems and
113 the classifier [13]. One of the most well-known solutions is the direct
114 modification of the learning procedure for a selected algorithm [45].
- 115 • Cost-sensitive level: these approaches consider higher costs for misclas-
116 sifying the minority classes with respect to the majority classes, that
117 is, misclassification of minority class is much more expensive [44]. The
118 learning process turns to minimize the cost errors instead of maximiza-
119 tion of accuracy rate [63].
- 120 • Ensemble level: these solutions combine the efficient ensemble learning
121 solutions [62] with one of the three previously mentioned strategies in
122 order to create a balanced training sets for base classifier and at the
123 same time introduce diversity into the pool of base learners. Special
124 attention should be paid to recent combination of intelligent and di-
125 rected data-level approaches with Bagging solution [5] or randomized
126 oversampling [15], hybrid combination of algorithm-level methods [55]
127 and cost-sensitive pruning for decision tree ensembles [36].

128 Due to the advantage of the data level solutions (as pointed out by a
129 recent tutorial on data preprocessing [29]) we focus on such methods in this
130 study.

131 RUS [51] is the basic under-sampling, which randomly removes the major-
132 ity class instances to balance the class distribution. This approach is efficient
133 for dealing with class imbalance problems, since most of the majority class
134 instances are redundant. Additionally, RUS makes the training process be-
135 come much faster, since the training set contains less instances than original
136 data set. However, some potential useful information contained in the ma-
137 jority class instance may be neglected, since RUS randomly generates the
138 subset without considering the relationship among the instances.

139 ROS [27] is another basic resampling approach, which randomly dupli-
140 cates the minority class instances to make the training set balanced. With
141 such consideration, the size of minority class examples tends to match the
142 size of the other classes. However, it might lead to two important defects.
143 One of the problems is overfitting, since ROS typically replicates examples of
144 the minority class so that the interface between minority class and majority
145 class is too prone to the former. Another problem is that ROS increases the
146 training time. Suppose there is an imbalance data set with 20000 majority
147 class examples and 100 minority class examples, to generate an equal size
148 of minority class and majority class, the ROS will create a training data set
149 with 40000 examples. Obviously, it must take much more time if ROS is
150 used as resampling technique.

151 SMOTE is an intelligent oversampling approach proposed by Chawla et
152 al. [8]. Unlike ROS duplicates the minority examples, SMOTE produces
153 synthetic minority class examples by k nearest neighbors, augmented with
154 randomized interpolation. However, the noise might incorporate in the syn-
155 thetic minority class examples.

156 In the above reviewed methods, sampling process is independently carried
157 out before the training process. In these approaches, the training process is
158 unchanged and the classifiers are trained by the balanced data sets produced
159 in the resampling process. To overcome the drawbacks of classical resam-
160 pling approaches mentioned above, some ensemble learning methods [37] are
161 proposed to devote to combine the resampling process and training process.
162 These techniques developed for dealing with class imbalance problems follow
163 the architecture of bagging or boosting.

164 In the framework of bagging, the base classifier is trained by using the
165 resampling examples. That is, the diversity is obtained through resampling
166 minority class instances, since bootstrapped instances of the original data set
167 are randomly duplicated. Many proposals are introduced in the research com-
168 munity, such as OverBagging, UnderBagging, and UnderOverBagging [3].

169 Regarding to the scenario of boosting, the resampling techniques are em-
170 bed into bootstrap. In such a manner, the resampling methods are introduced
171 in each iteration to train a classifier toward the minority class. In this family,
172 SMOTEBoost [10] and RUSBoost [49] are the representative proposals. Be-
173 sides these techniques, some algorithms combine both bagging and boosting
174 to obtain an ensemble of ensembles [40].

175 However, when we refer to multi-class imbalance classification problems,
176 the conventional solutions may not be feasible or perform a lower achieve-
177 ment. There are still only few approaches for handling this task. Static-
178 SMOTE [19] applied resampling procedure in M steps, where M is the
179 number of classes. In each iteration, the resampling procedure selects the
180 minimum size class, and duplicates the number of instances of the class in
181 the original data-set. An ensemble learning algorithm for multi-majority
182 and multi-minority cases was proposed in [57]. Authors combine AdaBoost
183 with negative correlation learning, where starting weights of examples are
184 calculated in inverse proportion to the number of objects in this class. A
185 combination of binary decomposition and pre-processing methods was pro-
186 posed as an efficient solution when the number of classes is high [18].

187 More recent studies on this problem propose to combine pairwise modified
188 SVMs with boundary shift asymmetric regularization costs using one-vs-all
189 technique [14]. Additionally, two hybrid ensemble techniques with embedded
190 feature selection were proposed to deal with the problem of skewed distri-
191 butions among multiple classes [30, 38]. The most recent work study in this
192 domain reports the high importance of considering the individual types of
193 minority classes examples and their learning difficulty when performing over-
194 sampling for multi-class imbalanced data and proposes a data-driven univer-
195 sal strategy that can be embedded in any data-level multi-class solution [48].

196 As one can see, there are many efficient ensemble techniques dedicated
197 to binary problems. However, it is not straightforward to extend them for
198 scenarios with higher number of classes. Therefore, the ensemble learning
199 methods are considered in this study to learn the pattern from the data set
200 derived from paired classes, and then the binary ensembles are combined by
201 aggregation strategy to be a final multi-class classifier.

202 *2.3. One-vs-One Scheme for Multi-Class Classification Problems*

203 In the OVO decomposition scheme, a m -class problem is divided into
204 $m(m - 1)/2$ binary subproblems. Each problem is faced by independent
205 base classifiers, which are responsible for distinguishing the instances from

206 the different pairs of classes. With such appropriate consideration, the much
 207 more complex multi-class problem is translated into the simpler binary class
 208 subproblems, which is expected to receive better results or address multi-class
 209 problems with binary classification techniques. An example of binarization
 210 technique for decomposing the multi-class problem is shown in Figure 1.

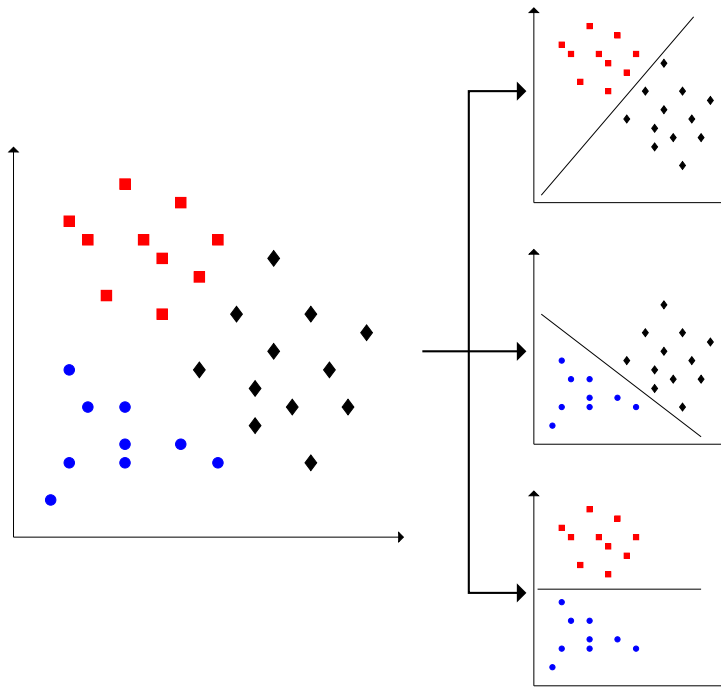


Figure 1: An example of one-versus-one (OVO) decomposition of a three-class problem into three two-class problems.

211 In order to predict a new pattern, there are two phases required to com-
 212 plete the task. The first phase is to learn the classifiers that are trained by
 213 the original instances with pairwise classes, that is, the task of each classifier
 214 is to distinguish a pair of classes $\{C_i, C_j\}$. In the integration phase, a confi-
 215 dence degree $r_{ij} \in [0, 1]$ in favor of C_i is given by a classifier to discriminate
 216 the class i from class j . The confidence in favor of j -th class is computed
 217 as $r_{ji} = 1 - r_{ij}$, if the classifier considers the output class as the class with
 218 the largest confidence value. To be clearer, all the confidence degree can be
 219 represented by a score matrix R :

$$R_{m,n} = \begin{pmatrix} - & r_{1,2} & \cdots & r_{1,m} \\ r_{2,1} & - & \cdots & a_{m,n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m,1} & r_{m,2} & \cdots & - \end{pmatrix}. \quad (1)$$

220 There are a number of aggregations proposed in the literature [26, 39] to
 221 infer the final class. The voting strategy (VOTE) is the simplest but powerful
 222 aggregation, therefore, it is considered as the aggregation approach in this
 223 study. VOTE, which is also called binary voting and Max-Wins rule [21],
 224 considers a vote for the predicted class by the binary classifier. Votes received
 225 by each class are counted and the final class obtaining the largest number of
 226 votes is predicted as follows:

$$Class = \arg \max_{i=1,\dots,m} \sum_{1 \leq j \neq i \leq m} s_{ij}, \quad (2)$$

227 where

$$s_{ij} = \begin{cases} 1 & r_{ij} > r_{ji} \\ 0 & \text{otherwise} \end{cases}. \quad (3)$$

228 3. Combining OVO with Ensemble Learning for Multi-class Imbal- 229 anced Data

230 As mentioned above, on the one hand, ensemble learning proposed for
 231 imbalanced datasets may be not effective or even impossible for dealing with
 232 multi-class problems. On the other hand, decomposition strategy is effec-
 233 tive direction to handle the multi-class classification problems. Therefore, it
 234 is interesting for us to consider their respective advantages in the scenario
 235 of multi-class imbalanced datasets. In this section, we firstly describe our
 236 methodology of employing the ensemble learning approaches in the OVO
 237 scheme. Then, we present the well-known ensemble learning methods em-
 238 ployed in this study.

239 3.1. Solving Multi-Class Imbalanced Problems with Binary Decomposition

240 Multi-class imbalance classification problems are extremely complex tasks,
 241 which suffer from more than two classes and class imbalance distribution. Ac-
 242 cording to the suggestion of decomposition strategy, which aims at solving

243 the complex problem by decomposing it into series of much easier subprob-
244 lems, we propose to explore the effectiveness of combination of ensemble
245 learning and OVO scheme to address the multi-class imbalance classification
246 problems.

247 Firstly, the multi-class imbalance classification problems are decomposed
248 into a series of pairwise datasets. One should note that there are some
249 pairwise classes after decomposition will have roughly equal size. For these
250 cases, the subproblems are handled as normal binary classification problems
251 by the normal ensemble learning algorithms. Then, the ensemble learning
252 methods based on the resampling techniques devote to address the binary
253 class imbalance datasets.

254 In this work we focus only on OVO approach as recent studies on decom-
255 position techniques clearly proved its superiority over OVA methods [22, 24].
256 Additionally, studies on combination of resampling strategies with single clas-
257 sifiers for multi-class imbalanced data showed clearly that OVA displays in-
258 ferior performance [18]. Finally, OVA introduces an artificial class imbalance
259 which may further damage the learning process in scenarios with multiple
260 skewed distributions.

261 An example of the proposed approach is shown in Figure 2.

262 There are two strategies employed in the ensemble learning process.

263 As shown in Figure 2 (left), one of the ensemble learning algorithms com-
264 bines the single classifiers which are derived from the balanced datasets by
265 reducing the size of the majority class. Concretely, the multi-class classi-
266 fication problem is split into several binary class subproblems. Then, the
267 ensemble learning method combining under-sampling strategy is employed
268 to learn a binary classifier for each subproblem. Finally, the aggregation
269 strategy is considered to obtain the final output class.

270 As described in Figure 2 (right), oversampling approach is adopted to
271 increase the number of the minority class to balance the class distribution.
272 Firstly, the binary class subproblems are derived from the original multi-
273 class classification problem. Secondly, ensemble learning method considering
274 SMOTE to create the synthetic minority class to balance the distribution for
275 the training process is used to achieve a binary classifier for each pairwise
276 class data set. Finally, once each binary classifier derived from the ensemble
277 learning, aggregation strategy will be employed to provide the final output
278 from the score matrix.

279 Let us now discuss ensemble learning algorithms used in this study.

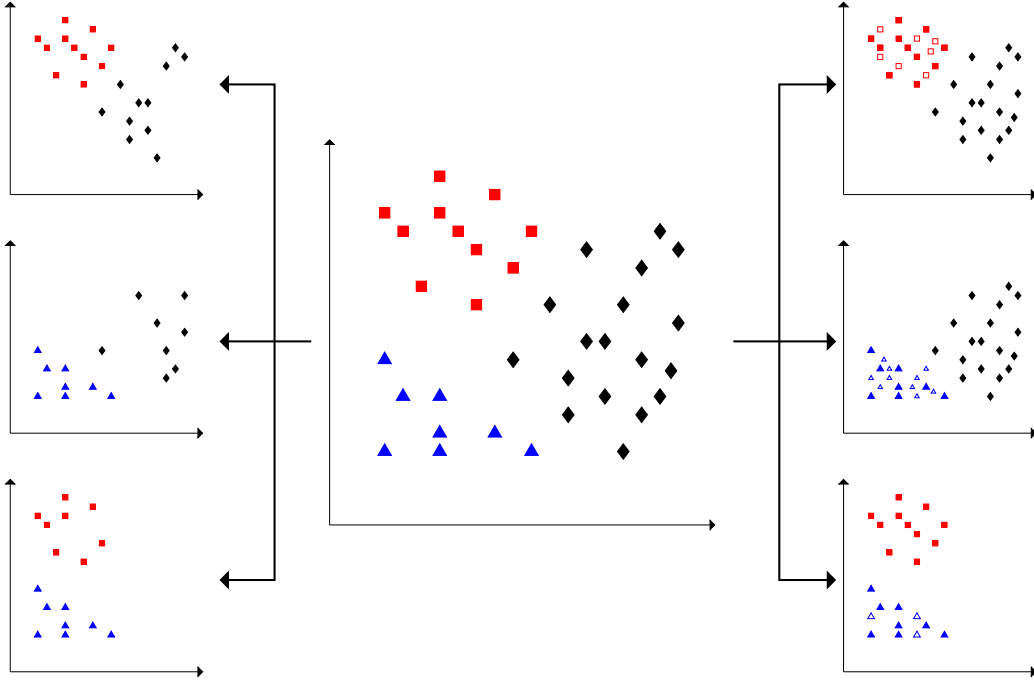


Figure 2: An example of pre-processing methods applied over a decomposed multi-class imbalanced dataset. (*Left*) Undersampling applied to each subproblem. (*Right*) Oversampling applied to each subproblem.

280 *3.2. Ensemble Learning Algorithms for Binary Imbalanced Data*

281 As described above, ensemble learning is applied in our study to classify
 282 the binary classes which are derived from the multiclass data sets. There-
 283 fore, several well-known ensemble learning algorithms proposed for imbalance
 284 problems were selected to train the binary classifiers. Specifically, the ensemble
 285 learning algorithms employed in this study are as follows:

- 286 • UnderBagging (UBA). RUS is applied to under-sampling the majority
 287 class in each bag of the ensemble. In this way, we can obtain the
 288 balanced data set for each classifier of the bag and the diversity for
 289 each bag is produced with different majority class instances and the
 290 same minority class instances. Additionally, the strategy of resampling
 291 with replacement is adopted in order to increase the diversity among
 292 the bags.
- 293 • SMOTEBagging (SBA). SMOTE algorithm is employed to insert ar-

294 tificial minority class in each bag. This way aims at balancing the
295 distribution of the data set by increasing the number of minority class
296 instances. The diversity of each bag of the ensemble is obtained with
297 mutative synthetic samples.

298 • RUSBoost (RBO). This strategy applies RUS of minority class to bal-
299 ance the distribution of the data set before ensemble learning. The
300 AdaBoost is employed to construct a classifier from the balanced data
301 set. One should note that this method handle the imbalance problem
302 before the ensemble learning process.

303 • SMOTEBoost (SBO). SMOTE algorithm is employed to balance the
304 distribution of the data set in each iteration of AdaBoost. Specifically,
305 a training set is obtained by resampled from weighted samples. Then,
306 synthetic samples are generated to increasing the number of minority
307 class. Therefore, the distribution of the training set for each single
308 classifier in AdaBoost is balanced.

309 • SMOTE+AdaBoost (SMB). Considering to this method, we apply SMOTE
310 algorithm to generate synthetic majority class to balance the data
311 set before AdaBoost. That is, the SMOTE is independent from Ad-
312 aBoost, which is different from SBO employing SMOTE algorithm dur-
313 ing Boosting progress.

314 • EasyEnsemble (Easy). Under-sampling is simple but effective for han-
315 dling imbalance problem, since the training data set is much more bal-
316 anced and the training process is much faster. However, many majority
317 class examples are ignored, which leads to loss potentially useful infor-
318 mation. In order to overcome the drawback, this method constructs
319 classifier ensembles from all of the minority class and a subset of the
320 majority class. Then, the final ensemble combines the outputs of clas-
321 sifiers which are built by using AdaBoost algorithm. Therefore, it can
322 be considered as an ensemble of ensembles.

323 4. Experimental Framework

324 In this section, we present the set-up of the experimental framework used
325 in the experiment of Section 5. The data sets chosen to test the algorithms
326 are described in Section 4.1, while single and ensemble classifiers with their

Table 1: Summary description of the data sets used in the experimental study.

| id | Dataset | #Ex. | #Atts. | #Num. | #Nom. | #Cl. | #Dc. | IR |
|-----|------------------|------|--------|-------|-------|------|-------------------------------|--------|
| Aut | Automobile | 159 | 25 | 15 | 10 | 6 | 46/29/13/48/20/3 | 16.00 |
| Bal | Balance | 625 | 4 | 4 | 0 | 3 | 288/49/288 | 5.88 |
| Car | Car | 1728 | 6 | 0 | 6 | 4 | 384/69/1210/65 | 18.62 |
| Cle | Cleveland | 297 | 13 | 5 | 8 | 5 | 160/54/35/35/13 | 12.31 |
| Con | Contraceptive | 1473 | 9 | 6 | 3 | 3 | 629/333/511 | 1.89 |
| Der | Dermatology | 358 | 34 | 1 | 33 | 6 | 111/60/71/48/48/20 | 5.55 |
| Fla | Flare | 1066 | 11 | 0 | 11 | 6 | 147/211/239/95/43/331 | 7.70 |
| Gla | Glass | 214 | 9 | 9 | 0 | 6 | 70/76/17/29/13/9 | 8.44 |
| Hay | Hayes-roth | 160 | 4 | 4 | 0 | 3 | 65/64/31 | 2.10 |
| Led | Led7digit | 500 | 7 | 0 | 7 | 10 | 45/37/51/57/52/52/47/57/53/49 | 1.54 |
| Lym | Lymphography | 148 | 18 | 3 | 15 | 4 | 4/61/81/2 | 40.50 |
| New | New-thyroid | 215 | 5 | 5 | 0 | 3 | 150/35/30 | 5.00 |
| Pag | Page-blocks | 5472 | 10 | 10 | 0 | 5 | 4913/329/28/87/115 | 175.46 |
| Sat | Satimage | 6435 | 36 | 36 | 0 | 6 | 1533/703/1358/626/707/1508 | 2.45 |
| Shu | Shuttle | 2175 | 9 | 9 | 0 | 5 | 1706/2/6/338/123 | 853.00 |
| Spl | Splice | 3190 | 60 | 0 | 60 | 3 | 767/768/1655 | 2.16 |
| Thy | Thyroid | 720 | 21 | 6 | 15 | 3 | 17/37/666 | 39.18 |
| Win | Wine | 178 | 13 | 13 | 0 | 3 | 59/71/48 | 1.48 |
| Wqr | Wine-Quality-Red | 1599 | 11 | 11 | 0 | 6 | 10/53/681/638/199/18 | 68.10 |
| Zoo | Zoo | 101 | 16 | 0 | 16 | 7 | 41/20/5/13/4/8/10 | 10.25 |

327 parameters are described in Section 4.2. The measure to evaluate the per-
328 formance of the approaches in this study are presented in Section 4.3. Final
329 Section describes the statistical test applied to compare the results obtained
330 from experiments.

331 4.1. Data Sets

332 In this study, twenty data sets from the UCI repository were selected to
333 test the methodology. The properties of the data sets were showed in Table 1.
334 For each data set, it includes the number of examples (#Ex.), the number of
335 attributes (#Atts.), the number of numerical (#Num) and nominal (#Nom)
336 attributes, the number of classes (#Cl.), the distribution of class (#Dc) and
337 the imbalance rate (IR). For the missing values instances in the data sets
338 (Cleveland and Dermatology), we removed them before doing the partitions.

339 The results of the average accuracy was obtained by means of 5 times
340 5-fold stratified cross-validation (SCV) [43]. That is, each data set is split
341 into 5 folds and each fold contained 20% of the instances of the data set.
342 For each fold, the algorithm is trained with the instances contained in the
343 remaining folds (80% of the instances of the data set) and then tested by the
344 current fold. The reason why 5-fold SCV is more appropriate than a 10-fold
345 SCV in such framework was explained in [42]. If we use smaller partitions,

346 there would be some test sets without containing any instance from some of
347 the classes.

348 *4.2. Classification Algorithms*

349 In this section, the base classification algorithms compared in the exper-
350 iment are presented. Specially, CART, BPNN and SVM are selected for the
351 study, and they are described as follows.

352 CART is a tree induction technique. CART is a binary recursive parti-
353 tioning methodology to generate a tree, that is, a node in the tree always
354 split the attribute into only two parts (nodes). In CART, the Gini index is
355 used to measure the impurity of node. The attribute and the corresponding
356 binary split on that maximizes the reduction in impurity are chosen as the
357 splitting point. The process will continue until reach the stopping criterion,
358 including there is no possible split point or the maximal tree is obtained.
359 Due to the noise or outlier in the training data set, some branches reflect
360 abnormal information, which may lead to overfit the data. Pruning is an
361 effective way to avoid overfitting. In CART, the cost complexity pruning
362 algorithm which is a post-pruning method is used to prune the tree. In the
363 approach, the cost complexity of a tree is measured by a function of error
364 rate, which is the percentage of instances misclassified by the tree.

365 BPNN is a typical feed forward neural network, which has input layer,
366 hidden layer and output layer. Back propagation learns by iteratively pro-
367 cessing a set of training samples, comparing the networks prediction for each
368 sample with the actual known target value. For each training sample, the
369 weights are modified so as to minimize the mean-square error between the
370 networks prediction and the actual target value. These modifications are
371 made in the backwards direction through each hidden layer down to the first
372 hidden layer.

373 SVM is an effective machine learning method, which is based on Vapnik-
374 Chervonenkis structural risk minimization instead of the empirical risk. SVM
375 maps the original input feature space into a high dimensional feature space in
376 order to construct an optimal separating hyperplane with maximal margin.
377 The choice of a proper kernel has a strong effect on the final quality of the
378 obtained model [12].

379 These classifiers are used as base learners for six ensemble methods de-
380 scribed in Section 3.2.

381 Additionally, to offer a fair comparison with other ensemble techniques
382 dedicated to multi-class imbalanced learning we have selected two state-of-

383 the-art methods: Near-Bayesian Support Vector Machines (NBSVM) [14]
384 and multi-class imbalanced Boosting [57].

385 The first approach modifies the SVM classifier by using samples from two
386 classes in order to achieve boundary shift and combines it with the asymmet-
387 ric regularization costs. Additionally, authors propose a modification of the
388 popular SMO algorithm to train NBSVM. To handle multi-class scenarios a
389 binary aggregation scenario with OVA technique is being used.

390 AdaBoost.NC combines the multi-class AdaBoost algorithm with nega-
391 tive correlation learning, where starting weights of examples are calculated
392 in inverse proportion to the number of objects in this class. This way it is
393 possible to capture varying relations between classes.

394 Detailed parameters of used methods are given in Table 2.

395 *4.3. Performance Measures*

396 There is a large amount of measures for the performance of algorithms
397 in the imbalance classification problems, for example, precision, sensitivity,
398 G-mean [3], F-measure [2], or AUC [16]. However, all of them are designed
399 especially for the binary class problems. Standard metrics such as classi-
400 fication rate (accuracy rate) is an unreasonable measure in the multi-class
401 imbalance classification problem, as it does not differentiate the classifica-
402 tion rates from different classes. For example, in the data set of thyroid,
403 a classifier can achieve a high accuracy rate of 92.5%, if it recognizes all the
404 instances as class 3. There are some proposals for measures displaying a
405 balanced performance on multiple classes, like multi-class AUC [32]. For this
406 study we have decided to use the average accuracy metric.

407 The average accuracy gives the same weight to each class. It achieves
408 the accuracy rate of each class independently, and then the final result is
409 obtained by the average value. The average accuracy is computed as follows:

$$\text{AveAcc} = \frac{1}{m} \sum_{i=1}^m \text{TRP}_i, \quad (4)$$

410 where m is the number of classes and TRP_i stands for the True Positive
411 Rate of the i -th class.

412 *4.4. Statistical Analysis*

413 Statistical tests are important for analyzing the experimental results to
414 extract the findings. In this paper the hypothesis testing techniques, which

Table 2: Parameters setting for single and ensemble classifiers used in the study.

| Methods | Parameters |
|---|--|
| Single classifiers | |
| CART | Prune = true Method = classification Impure nodes must have 10 or more examples to be split The number of nodes in hidden layer = 10 |
| BPNN | Transfer function of hidden layer = logsig Transfer function of output layer = logsig Train epochs = 100 C = 1.0 |
| SVM | Tolerance parameter = 0.001 $\gamma = 0.0046$ Kernel type = Gaussian radial basis function Optimization method = SMO |
| Ensemble classifiers | |
| UBA | The number of bags = 40 The number of resampling majority class = the number of minority class |
| SBA | The number of bags = 40 The number of nearest neighbors in SMOTE = 5 |
| RBO | The number of iterations in AdaBoost = 40 The number of resampling majority class = the number of minority class |
| SBO | The number of iterations in AdaBoost = 40 The number of nearest neighbors in SMOTE = 5 |
| SMB | The number of iterations in AdaBoost = 40 The number of nearest neighbors in SMOTE = 5 |
| Easy | The number of subsets is 4 The number of iteration in each AdaBoost ensemble is 10 |
| Algorithms for multi-class imbalance | |
| OVA-NBSVM | C = 2.0 $\sigma \in [1, 2, 3, 4, 5, 10, 20, 30, 40, 50]$ - best selected for each dataset Optimization method = modified SMO classifier combination = OVA The number of iterations = 51 |
| AdaBoost.NC | $\alpha = 2$ base classifier = C4.5 |

415 are recommended in [28], are used to provide statistical support for the analy-
416 sis of the results obtained by the experiment. Specially, for the pairwise com-
417 parisons, Wilcoxon signed-rank test [60] is adopted as a non-parametric sta-
418 tistical procedure to perform pairwise comparisons between ensemble learn-
419 ing and resampling approaches in OVO scheme and comparisons of the repre-
420 sentation ensemble learning with OVO scheme and the state-of-the-art meth-
421 ods. Then, for the multiple comparisons, average aligned-rankings [28] of
422 each method are adopted to compare the behavior of each method with re-
423 spect to the others. Moreover, Friedman aligned-ranks test is used to check
424 whether the best method (the control one) is significant better than others.

425 5. Experimental Study

426 In this section we develop a thorough empirical analysis in addressing
427 the multi-class imbalance classification problems. We want to verify the
428 effectiveness of the combination of ensemble learning and OVO scheme for
429 the multi-class imbalance classification compared with classical approaches.
430 Therefore, we develop the pairwise comparative study on analysis of ensemble
431 learning approaches versus resampling techniques in the OVO scheme with
432 different base classifiers, including CART, BPNN and SVM. Additionally, to
433 show that using ensembles as base classifiers in OVO can efficiently empower
434 learning from multi-class imbalanced data we present comparison with two
435 state-of-the-art algorithms dedicated to this problem.

436 5.1. Evaluating Ensemble Approaches with CART as Base Classifier

437 The study for the CART decision tree is shown in Table 3, where we
438 present all the results of average accuracy.

439 According to the best result stressed through bold-face, we can clearly find
440 that the methods considering ensemble learning in the OVO scheme always
441 receive better results in each data set. Observing the average performance,
442 SBO obtains the best performance, followed by SMB and SBA.

443 The statistical study based on Wilcoxon test for CART is developed in
444 Table 4.

445 Results of Wilcoxon test show that for CART classifier ensemble systems
446 can significantly outperform resampling-based strategies with single classi-
447 fier. However, there is no single ensemble strategy that is statistically sig-
448 nificantly better than all of three resampling approaches tested at the same
449 time. For example SBO method achieves excellent p -values in comparison

Table 3: Average accuracy [%] results for resampling and ensemble learning methods with CART as the base classifier. The best result is highlighted in bold.

| Data sets | Resampling techniques | | | Ensemble learning | | | | | |
|-----------|-----------------------|--------------|-------|-------------------|-------|--------------|--------------|--------------|--------------|
| | RUS | ROS | SMOTE | UBA | SBA | RBO | SBO | SMB | Easy |
| Aut | 10.00 | 83.93 | 78.43 | 10.00 | 79.97 | 67.38 | 81.60 | 69.85 | 51.63 |
| Bal | 54.49 | 58.24 | 56.43 | 64.90 | 56.51 | 71.40 | 58.35 | 66.27 | 73.13 |
| Car | 93.87 | 94.64 | 91.93 | 95.72 | 93.87 | 96.53 | 97.34 | 97.56 | 97.57 |
| Cle | 31.92 | 30.19 | 28.82 | 31.92 | 29.22 | 32.98 | 31.65 | 33.18 | 31.89 |
| Con | 48.54 | 48.11 | 47.96 | 51.97 | 47.96 | 48.31 | 47.85 | 48.96 | 50.79 |
| Der | 94.36 | 95.15 | 95.37 | 95.45 | 95.37 | 94.49 | 94.94 | 94.65 | 94.41 |
| Fla | 63.28 | 62.52 | 61.20 | 63.46 | 61.11 | 62.65 | 60.58 | 62.60 | 63.03 |
| Gla | 64.08 | 66.66 | 67.90 | 71.38 | 69.39 | 69.68 | 72.17 | 72.59 | 71.78 |
| Hay | 86.07 | 85.86 | 85.76 | 85.86 | 85.76 | 85.54 | 85.15 | 85.24 | 86.15 |
| Led | 69.77 | 69.95 | 70.13 | 69.69 | 70.13 | 70.20 | 70.48 | 70.78 | 70.83 |
| Lym | 11.00 | 79.05 | 69.98 | 11.00 | 69.98 | 36.10 | 76.10 | 62.51 | 19.83 |
| New | 92.80 | 89.70 | 90.37 | 93.48 | 91.75 | 93.24 | 91.72 | 93.69 | 94.43 |
| Pag | 91.98 | 81.19 | 84.77 | 94.24 | 84.31 | 94.09 | 85.33 | 93.63 | 94.68 |
| Sat | 83.82 | 84.25 | 84.23 | 87.97 | 86.36 | 90.02 | 89.41 | 90.20 | 89.83 |
| Shu | 78.23 | 85.85 | 85.11 | 79.69 | 85.10 | 37.39 | 85.94 | 41.89 | 25.98 |
| Spl | 92.74 | 92.68 | 92.89 | 94.50 | 93.33 | 95.51 | 95.65 | 95.64 | 95.64 |
| Thy | 96.24 | 93.77 | 93.77 | 98.14 | 93.77 | 96.42 | 92.95 | 98.61 | 98.16 |
| Win | 91.77 | 92.60 | 92.29 | 94.18 | 92.29 | 97.32 | 91.34 | 96.60 | 95.63 |
| Wqr | 33.18 | 32.60 | 34.61 | 37.80 | 36.21 | 39.27 | 39.15 | 38.86 | 39.12 |
| Zoo | 48.01 | 85.67 | 72.15 | 49.23 | 72.15 | 48.08 | 87.48 | 62.64 | 23.67 |
| Avg. | 66.81 | 75.63 | 74.21 | 69.03 | 74.73 | 71.33 | 76.76 | 73.80 | 68.41 |

Table 4: Wilcoxon tests for comparison of ensemble learning and resampling techniques in the OVO scheme with CART as the base classifier. R^+ corresponds to the sum of the ranks for ensemble learning and R^- for the resampling techniques.

| Comparison | R^+ | R^- | p -value | Comparison | R^+ | R^- | p -value |
|---------------|--------|--------|------------|----------------|--------|-------|------------|
| UBA vs. RUS | 197.00 | 13.00 | 0.000629 | SBO vs. RUS | 157.00 | 53.00 | 0.052222 |
| UBA vs. ROS | 135.50 | 74.50 | 0.295424 | SBO vs. ROS | 143.00 | 67.00 | 0.156004 |
| UBA vs. SMOTE | 135.00 | 75.00 | 0.262722 | SBO vs. SMOTE | 179.00 | 31.00 | 0.005734 |
| SBA vs. RUS | 144.00 | 66.00 | 0.145400 | SMB vs. RUS | 185.00 | 25.00 | 0.002821 |
| SBA vs. ROS | 91.50 | 118.50 | 0.629162 | SMB vs. ROS | 131.00 | 79.00 | 0.331723 |
| SBA vs. SMOTE | 158.00 | 52.00 | 0.022909 | SMB vs. SMOTE | 137.00 | 73.00 | 0.232226 |
| RBO vs. RUS | 172.00 | 38.00 | 0.012374 | Easy vs. RUS | 167.00 | 43.00 | 0.020633 |
| RBO vs. ROS | 127.00 | 83.00 | 0.411465 | Easy vs. ROS | 133.00 | 77.00 | 0.295878 |
| RBO vs. SMOTE | 131.00 | 79.00 | 0.331723 | Easy vs. SMOTE | 133.00 | 77.00 | 0.295878 |

450 with RUS and SMOTE, but achieves p -value slightly above the significance
451 level for comparison with ROS. Similar situation can be observed for all
452 of highlighted best-performing ensemble techniques. At the same time one
453 must notice that the OVO-based ensemble learning does not damage the
454 accuracies when compared to OVO single model learning with resampling.
455 Therefore, we can summarize that ensemble learning with OVO decompo-
456 sition is a preferred choice over OVO resampling techniques, never leading
457 to reduced classification rates and often offering significant improvement for
458 the multi-class imbalance classification problems with CART classifier.

459 5.2. Evaluating Ensemble Approaches with BPNN as Base Classifier

460 The complete results for the BPNN version are shown in Table 5.

461 In this case the trend is quite similar with the segment of CART, since
462 most of the best results obtained in each data set are acquired by the ensemble
463 learning approaches. Additionally, for the average results, SMB receives the
464 best performance, followed closely by SBO. This is consistent with those
465 obtained by CART. Therefore, we must highlight that the methodology of
466 ensemble learning methods used to train the binary classifier in the OVO
467 scheme achieve quite high quality of the average performance.

468 The Wilcoxon tests are carried out in Table 6, when BPNN is used as the
469 base classifier.

470 In this case, all the null hypotheses of equivalence are rejected for the
471 ensemble learning methods, since the largest p -value is equal to 0.043804,

Table 5: Average accuracy [%] results for resampling and ensemble learning methods with BPNN as the base classifier. The best result is highlighted in bold.

| Data sets | Resampling techniques | | | Ensemble learning | | | | | |
|-----------|-----------------------|--------------|-------|-------------------|-------|--------------|--------------|--------------|--------------|
| | RUS | ROS | SMOTE | UBA | SBA | RBO | SBO | SMB | Easy |
| Aut | 47.12 | 64.99 | 56.34 | 61.40 | 70.44 | 70.83 | 77.67 | 71.91 | 66.88 |
| Bal | 70.81 | 89.11 | 85.93 | 88.18 | 91.07 | 85.80 | 94.12 | 86.08 | 88.28 |
| Car | 61.78 | 72.10 | 66.98 | 77.17 | 80.41 | 90.27 | 97.98 | 95.93 | 91.79 |
| Cle | 33.03 | 35.18 | 32.18 | 32.62 | 30.07 | 32.68 | 30.70 | 30.31 | 31.16 |
| Con | 45.97 | 47.12 | 38.27 | 50.72 | 36.71 | 51.90 | 51.15 | 50.70 | 51.96 |
| Der | 91.60 | 93.81 | 92.49 | 96.87 | 96.81 | 96.41 | 96.94 | 96.90 | 96.79 |
| Fla | 56.96 | 60.07 | 54.11 | 65.24 | 59.61 | 63.38 | 62.28 | 63.75 | 65.05 |
| Gla | 48.91 | 52.67 | 51.39 | 63.21 | 60.02 | 71.09 | 63.92 | 70.91 | 70.02 |
| Hay | 60.17 | 61.84 | 53.36 | 71.06 | 66.53 | 77.78 | 82.49 | 78.38 | 78.71 |
| Led | 68.48 | 67.28 | 67.25 | 73.64 | 73.04 | 71.54 | 72.15 | 70.81 | 72.02 |
| Lym | 70.41 | 85.53 | 78.14 | 86.79 | 88.83 | 83.22 | 87.86 | 84.08 | 86.17 |
| New | 90.04 | 94.89 | 92.80 | 97.70 | 96.52 | 97.07 | 96.22 | 97.19 | 97.93 |
| Pag | 78.57 | 82.26 | 83.52 | 88.52 | 89.49 | 91.50 | 84.29 | 91.08 | 92.08 |
| Sat | 86.24 | 86.58 | 85.47 | 87.38 | 86.99 | 89.30 | 88.30 | 89.71 | 89.55 |
| Shu | 68.10 | 84.87 | 89.46 | 80.71 | 93.07 | 85.79 | 93.56 | 91.40 | 87.88 |
| Spl | 85.76 | 86.70 | 83.67 | 90.34 | 90.32 | 89.73 | 91.32 | 90.55 | 89.78 |
| Thy | 74.98 | 80.84 | 75.70 | 86.45 | 70.54 | 95.80 | 78.36 | 96.19 | 96.81 |
| Win | 94.08 | 92.09 | 95.50 | 98.70 | 98.27 | 97.94 | 98.00 | 97.87 | 98.57 |
| Wqr | 33.05 | 34.80 | 35.36 | 39.25 | 36.93 | 39.36 | 31.40 | 40.42 | 41.96 |
| Zoo | 73.04 | 81.19 | 79.61 | 90.24 | 90.10 | 86.60 | 89.52 | 88.86 | 89.39 |
| Avg. | 66.96 | 72.70 | 69.88 | 76.31 | 75.29 | 78.40 | 78.41 | 79.15 | 79.14 |

Table 6: Wilcoxon tests for comparison of ensemble learning and resampling techniques in the OVO scheme with BPNN as the base classifier. R^+ corresponds to the sum of the ranks for ensemble learning and R^- for the resampling techniques.

| Comparison | R^+ | R^- | p -value | Comparison | R^+ | R^- | p -value |
|---------------|-------|-------|------------|----------------|-------|-------|------------|
| UBA vs. RUS | 209 | 1 | 0.000103 | SBO vs. RUS | 206 | 4 | 0.000163 |
| UBA vs. ROS | 186 | 24 | 0.002495 | SBO vs. ROS | 186 | 24 | 0.002495 |
| UBA vs. SMOTE | 197 | 13 | 0.000593 | SBO vs. SMOTE | 201 | 9 | 0.000338 |
| SBA vs. RUS | 189 | 21 | 0.001713 | SMB vs. RUS | 208 | 2 | 0.00012 |
| SBA vs. ROS | 159 | 51 | 0.043804 | SMB vs. ROS | 196 | 14 | 0.000681 |
| SBA vs. SMOTE | 194 | 16 | 0.000892 | SMB vs. SMOTE | 208 | 2 | 0.00012 |
| RBO vs. RUS | 209 | 1 | 0.000103 | Easy vs. RUS | 209 | 1 | 0.000103 |
| RBO vs. ROS | 194 | 16 | 0.000892 | Easy vs. ROS | 199 | 11 | 0.000449 |
| RBO vs. SMOTE | 205 | 5 | 0.000189 | Easy vs. SMOTE | 207 | 3 | 0.00014 |

472 which is lower than our α -value (0.05). This is much better than in the case
473 of CART classifiers, as here we observe a globally statistically significant
474 improvement regardless of the ensemble method used. This is a highly inter-
475 esting observation, as commonly it was assumed that decision trees as weak
476 learners should benefit the most from multiple classifier systems. However,
477 one can see that combination of BPNNs works much better than combination
478 of CARTs. Therefore, we can conclude that the ensemble learning approaches
479 with OVO scheme for the multi-class imbalance classification problems out-
480 perform the OVO resampling techniques for BPNN classifier.

481 5.3. Evaluating Ensemble Approaches with SVM as Base Classifier

482 Finally, we show the complete results for the version of SVM in Table 7.

483 The average results obtained in this version are somewhat lower than in
484 the case of BPNN, but it is consistent that the ensemble learning approaches
485 improve the performance for the multi-class imbalance classification problems
486 with the OVO scheme, comparing with the original resampling techniques.
487 Additionally, once again SMB is considered as the best average performance,
488 followed by Easy and RBO respectively.

489 Observing the statistical analysis shown in Table 8, the behavior of the
490 ensemble learning techniques in the OVO scheme is generally better than
491 those obtained by the resampling approaches.

492 According to the statistical results of Wilcoxon test, we can observe that
493 RBO, SMB and Easy clearly outperform all of the conventional resampling

Table 7: Average accuracy [%] results for resampling and ensemble learning methods with SVM as the base classifier. The best result is highlighted in bold.

| Data sets | Resampling techniques | | | Ensemble learning | | | | | |
|-----------|-----------------------|--------------|-------|-------------------|--------------|--------------|--------------|--------------|--------------|
| | RUS | ROS | SMOTE | UBA | SBA | RBO | SBO | SMB | Easy |
| Aut | 47.86 | 56.74 | 58.01 | 52.30 | 57.63 | 60.71 | 67.34 | 67.51 | 60.41 |
| Bal | 61.78 | 57.43 | 73.70 | 64.12 | 73.66 | 78.77 | 64.60 | 80.59 | 79.64 |
| Car | 79.92 | 73.71 | 86.23 | 80.41 | 86.23 | 91.6 | 89.18 | 91.65 | 91.20 |
| Cle | 30.78 | 24.20 | 33.23 | 33.46 | 33.99 | 32.07 | 31.61 | 32.13 | 34.53 |
| Con | 50.37 | 46.32 | 48.40 | 50.76 | 48.40 | 50.12 | 41.61 | 50.35 | 50.85 |
| Der | 97.46 | 95.49 | 96.21 | 97.66 | 96.21 | 97.49 | 96.53 | 97.72 | 97.34 |
| Fla | 62.16 | 56.43 | 62.45 | 64.53 | 62.72 | 64.11 | 58.33 | 63.95 | 65.72 |
| Gla | 47.83 | 53.14 | 50.81 | 49.53 | 51.86 | 68.67 | 60.81 | 68.37 | 66.18 |
| Hay | 53.24 | 53.25 | 53.66 | 53.56 | 52.89 | 64.24 | 62.38 | 64.65 | 57.64 |
| Led | 70.95 | 67.93 | 70.11 | 73.52 | 70.11 | 71.78 | 72.88 | 72.10 | 72.31 |
| Lym | 83.12 | 89.26 | 85.16 | 85.94 | 88.50 | 83.35 | 87.80 | 82.96 | 87.35 |
| New | 85.25 | 90.00 | 92.20 | 87.30 | 91.91 | 96.13 | 95.66 | 95.43 | 97.70 |
| Pag | 71.30 | 76.84 | 79.63 | 73.15 | 80.46 | 89.87 | 83.84 | 90.95 | 90.74 |
| Sat | 86.24 | 71.59 | 86.81 | 86.32 | 86.86 | 84.94 | 84.32 | 85.30 | 86.21 |
| Shu | 68.27 | 94.15 | 94.73 | 74.64 | 94.70 | 77.97 | 99.33 | 88.72 | 85.46 |
| Spl | 95.96 | 95.43 | 96.30 | 96.16 | 96.43 | 95.94 | 95.86 | 96.07 | 96.20 |
| Thy | 68.27 | 79.49 | 82.61 | 70.48 | 82.74 | 91.11 | 78.74 | 92.46 | 93.67 |
| Win | 96.37 | 96.18 | 96.57 | 97.07 | 96.57 | 97.25 | 98.05 | 97.88 | 97.62 |
| Wqr | 37.20 | 35.53 | 38.64 | 38.57 | 38.83 | 40.63 | 24.67 | 39.41 | 38.92 |
| Zoo | 87.16 | 88.47 | 90.29 | 88.77 | 90.87 | 90.14 | 91.68 | 88.38 | 90.07 |
| Avg. | 69.07 | 70.08 | 73.79 | 70.91 | 74.08 | 76.34 | 74.26 | 77.33 | 76.99 |

Table 8: Wilcoxon tests for comparison of ensemble learning and resampling techniques in the OVO scheme with SVM as the base classifier. R^+ corresponds to the sum of the ranks for ensemble learning and R^- for the resampling techniques.

| Comparison | R^+ | R^- | p -value | Comparison | R^+ | R^- | p -value |
|---------------|-------|-------|------------|----------------|-------|-------|------------|
| UBA vs. RUS | 210 | 0 | 0.000089 | SBO vs. RUS | 166 | 44 | 0.022769 |
| UBA vs. ROS | 129 | 81 | 0.370261 | SBO vs. ROS | 178 | 32 | 0.006425 |
| UBA vs. SMOTE | 62 | 148 | 0.108427 | SBO vs. SMOTE | 122 | 88 | 0.525653 |
| SBA vs. RUS | 186 | 24 | 0.002495 | SMB vs. RUS | 201 | 9 | 0.000338 |
| SBA vs. ROS | 196 | 14 | 0.000681 | SMB vs. ROS | 191 | 19 | 0.001325 |
| SBA vs. SMOTE | 145.5 | 64.5 | 0.111769 | SMB vs. SMOTE | 166 | 44 | 0.022769 |
| RBO vs. RUS | 197 | 13 | 0.000593 | Easy vs. RUS | 207 | 3 | 0.00014 |
| RBO vs. ROS | 183 | 27 | 0.003592 | Easy vs. ROS | 193 | 17 | 0.001019 |
| RBO vs. SMOTE | 165 | 45 | 0.025094 | Easy vs. SMOTE | 186 | 24 | 0.002495 |

Table 9: Friedman aligned-rank tests comparing the different ensemble learning methods with OVO scheme.

| Ensemble learning | CART | BPNN | SVM |
|-------------------|-----------------|-----------------|-----------------|
| UBA | 66.10(0.151106) | 71.10(0.168688) | 82.25(0.002173) |
| SBA | 75.15(0.018658) | 76.45(0.058981) | 73.60(0.025194) |
| RBO | 61.60(0.285839) | 58.35(1.000000) | 51.50(0.939695) |
| SBO | 61.25(0.285839) | 58.50(1.000000) | 67.35(0.091477) |
| SMB | 43.25 | 49.85(1.000000) | 43.55 |
| Easy | 55.65(0.285839) | 48.75 | 44.75(0.939695) |

494 techniques. In addition, RUS is defeated by any ensemble learning approach.
 495 For ROS only UBA method does not deliver significant improvement. With
 496 regard to SMOTE the p -values observed for comparing with RBO, SMB and
 497 Easy are lower than our α -value (0.05). Therefore, in the scenario of SVM we
 498 also can conclude that ensemble learning approaches outperform resampling
 499 techniques in the OVO scheme for the multi-class imbalance classification
 500 problems.

501 5.4. Comparison of Different Used Ensemble Learning Techniques

502 The goal of this study is to explore the validity of the ensemble learn-
 503 ing approaches with OVO decomposition in the application of multi-class
 504 imbalance classification problems, since these techniques combine the single

505 classifiers to be an ensemble, which is useful for addressing the binary-class
506 imbalance problems. The facts found in this study support the goodness of
507 the methodology of ensemble learning methods based on resampling tech-
508 niques for improving the performance in dealing with the multi-class imbal-
509 ance classification problems.

510 However, we also need to compare the different examined ensemble learn-
511 ing techniques among themselves in order to find the statistically best method
512 for each base classifier. Using only averaged accuracy for such a selection can
513 be misleading as we do not take into account the ranks among these meth-
514 ods for each dataset. Therefore, we need to conduct a statistical test over
515 multiple datasets to choose the best combined classifier.

516 For this purpose we use the Friedman aligned-rank test in order to de-
517 tect the highest-ranked methods. The results from this test are depicted in
518 Table 9.

519 When analyzing the aligned ranks for CART classifier we can clearly see
520 that SMB method achieves the best results. Its rank is clearly lower than
521 other methods. This is especially interesting when considering the fact of
522 SBO scoring the best average accuracy results. But according to statistical
523 analysis SMB method displays the most efficient and stable performance over
524 all of datasets used.

525 For BPNN classifier we can see that the lowest ranks were obtained by
526 Easy approach. However, differences between it and SMB are practically
527 insignificant (48.75 vs. 49.85) thus allowing us to conclude that both of
528 these methods display excellent performance for neural networks in mining
529 multi-class imbalanced data.

530 Similar situation can be observed for SVM classifier. Here SMB method
531 obtains the lowest ranks but differences between it and Easy method are
532 negligible (43.55 vs. 44.75), showing that both of these methods are highly
533 suitable for working with SVMs.

534 *5.5. Comparison with Ad-hoc Approaches for Multi-Class Imbalanced Data*

535 In order to put the obtained results into context we cannot refer only to
536 OVO-based resampling techniques. There is a number of solutions already
537 proposed in the literature for addressing skewed distributions in multiple
538 classes. We decided to select two ensemble-based techniques from them as
539 reference methods for our proposal. We have selected popular AdaBoost.NC
540 [57], which is considered as one of the best approaches for handling multi-
541 class imbalanced data. Additionally, we present results for recently intro-

542 duced OVA-NBSVM [14] that combines binary modified SVMs using one-vs-
543 all technique.

544 For each base classifier examined we have selected a single best-performing
545 ensemble technique according to the Friedman aligned-rank tests discussed
546 in Section 5.4. This means that for comparison SMB for CART / SVM and
547 Easy for BPNN are being used.

548 Accuracy results for examined benchmarks and methods are depicted in
549 Table 10.

550 We will focus our discussion on comparison with reference methods, as
551 accuracies of ensemble techniques for different base classifiers were discussed
552 in previous sections.

553 OVA-NBSVM returned surprisingly the worst results from all of the ex-
554 amined methods. Only for a single dataset (Zoo) it was able to outperform
555 all other approaches. This proves the lack of usefulness of OVA decompo-
556 sition for multi-class problems with skewed distributions. OVA for some
557 of class combinations additionally boosts the disproportion between classes
558 thus making learning from small-sample classes very difficult. This can be
559 observed for cases with high imbalance ratios and small classes like Pag and
560 Wqr datasets. Here some of classes are very small when compared to others,
561 which leads to a extremely high binary imbalance ratio when these classes
562 are used as a positive and aggregation of remaining ones as a negative. For
563 these cases OVA-NBSVM drops highly in comparison to all other discussed
564 solutions.

565 AdaBoost.NC is a more demanding reference method, achieving best re-
566 sults on 4 datasets. However, on 16 remaining benchmarks one of the pro-
567 posed ensemble techniques outperforms this boosting scheme. It allows us to
568 conclude that empowering OVO decomposition with ensemble learning can
569 be a highly efficient solution for the discussed problem, being able to de-
570 liver generally improved performance over most popular solutions from the
571 literature.

572 To gain an additional insight into the performance of examined methods
573 we have conducted a Wilcoxon test, outcomes of which are presented in
574 Table 11.

575 Obtained p -values show that proposed ensemble techniques with BPNN
576 and SVM are significantly better than OVA-NBSVM. For comparison with
577 AdaBoost.NC only BPNN-based ensemble offer significantly superior results.
578 When analyzing obtained p -values and accuracies for remaining methods we
579 can see that they offer small but visible improvement over reference algo-

Table 10: Complete average accuracy test results for the representative combinations (SMB-CART, Easy-BPNN and SMB-SVM) and the state-of-the-art approaches for mining multi-class imbalanced data (AdaBoost.NC and OVA-NBSVM). The best result is highlighted in bold.

| Dataset | AdaBoost.NC | OVA-NBSVM | SMB-CART | Easy-BPNN | SMB-SVM |
|---------|--------------|--------------|--------------|--------------|--------------|
| Aut | 76.53 | 70.73 | 69.85 | 66.88 | 67.51 |
| Bal | 66.92 | 85.81 | 66.27 | 88.28 | 80.59 |
| Car | 85.08 | 54.83 | 97.56 | 91.79 | 91.65 |
| Cle | 30.33 | 31.11 | 33.18 | 31.16 | 32.13 |
| Con | 47.36 | 47.21 | 48.96 | 51.96 | 50.35 |
| Der | 95.00 | 96.45 | 94.65 | 96.79 | 97.72 |
| Fla | 60.57 | 54.45 | 62.60 | 65.05 | 63.95 |
| Gla | 69.89 | 61.96 | 72.59 | 70.02 | 68.37 |
| Hay | 85.39 | 76.42 | 85.24 | 78.71 | 64.65 |
| Led | 72.10 | 59.45 | 70.78 | 72.02 | 72.10 |
| Lym | 78.33 | 69.86 | 62.51 | 86.17 | 82.96 |
| New | 88.42 | 89.05 | 93.69 | 97.93 | 95.43 |
| Pag | 80.75 | 48.97 | 93.63 | 92.08 | 90.95 |
| Sat | 87.49 | 88.16 | 90.20 | 89.55 | 85.30 |
| Shu | 89.82 | 82.07 | 41.89 | 87.88 | 88.72 |
| Spl | 94.14 | 74.10 | 95.64 | 89.78 | 96.07 |
| Thy | 95.46 | 64.97 | 98.61 | 96.81 | 92.46 |
| Win | 93.13 | 97.58 | 96.60 | 98.57 | 97.88 |
| Wqr | 35.73 | 31.84 | 38.86 | 41.96 | 39.41 |
| Zoo | 86.67 | 90.62 | 62.64 | 89.39 | 88.38 |
| Avg. | 75.96 | 68.78 | 73.80 | 79.14 | 77.33 |

Table 11: Wilcoxon tests for comparison of the representative ensemble learning with OVO scheme and the state-of-the-art methods for multi-class imbalanced data. R^+ corresponds to the sum of the ranks for the representative ensemble learning and R^- for the state-of-the-art methods.

| Comparison | R^+ | R^- | p -value |
|---------------------------|-------|-------|------------|
| SMB-CART vs. AdaBoost.NC | 128 | 82 | 0.390533 |
| Easy-BPNN vs. AdaBoost.NC | 162 | 48 | 0.033340 |
| SMB-SVM vs. AdaBoost.NC | 150 | 60 | 0.092963 |
| SMB-CART vs. OVA-NBSVM | 146 | 64 | 0.125859 |
| Easy-BPNN vs. OVA-NBSVM | 198 | 12 | 0.000517 |
| SMB-SVM vs. OVA-NBSVM | 172 | 38 | 0.012374 |

580 rithms. In conclusion it must be noted that the proposed combination of
 581 OVO and pairwise ensemble learning can outperform state-of-the-art meth-
 582 ods for multi-class imbalance on a variety of datasets, thus being a worthwhile
 583 choice for such problems.

584 6. Lessons Learned

585 In order to summarize this manuscript let us present three main points
 586 capturing the most important research findings.

- 587 1. **The role of ensemble learning method.** Obtained results allow us
 588 to observe the existence of a trend among six examined ensemble learn-
 589 ing techniques. Regardless of the used base classifier SMB and Easy
 590 methods delivered the best performance both in terms of averaged accu-
 591 racy and ranking statistical tests. This makes them the most universal
 592 ensemble techniques that should always be examined in combination
 593 with OVO decomposition when imbalanced multi-class problems are
 594 being faced. Additionally, RBO and SBO techniques tend to deliver
 595 good results for some of the datasets, thus making them a reasonable
 596 second choice if time allows for a more compound experimental study.
- 597 2. **The role of base classifier.** The choice of a base classifier had a sig-
 598 nificant impact on the observed classification accuracies. Surprisingly,
 599 used decision tree model (CART) returned least satisfactory improve-
 600 ment when used in ensemble setting. This is contrary to numerous
 601 statements in the literature, reporting excellent properties of tree mod-
 602 els as weak classifiers for combination. This may be explained by high

603 sensitivity of CART model to skewed distributions. SVM worked very
604 well with most of the ensemble techniques, achieving improved classi-
605 fication rates in most of scenarios. BPNN was found to be the best
606 working model in ensemble setting, achieving statistically superior re-
607 sults for any kind of used committee approach.

608 **3. Comparison with resampling and ad-hoc solutions.** The pro-
609 posed empowering of OVO decomposition with pairwise ensemble learn-
610 ing achieved highly competitive results when compared with standard
611 OVO with resampling techniques applied. Using compound learners
612 for each pair of classes can lead to a better capturing of their local
613 specifics, higher robustness to imbalance and improved final multi-class
614 recognition rates. Additionally, proposed ensemble-based OVO is able
615 to outperform state-of-the-art methods for multi-class imbalanced data,
616 especially those using OVA solutions.

617 **7. Conclusions and Future Works**

618 In this paper we have proposed to improve the performance of binary
619 decomposition used for multi-class imbalanced problems by applying ensem-
620 ble learning paradigm to each sub-problem. This way we are able to exploit
621 highly efficient combined classification approaches that were so far restricted
622 to binary scenarios. Their proven efficacy in two-class imbalanced tasks mo-
623 tivated us to extend their area of applicability to much more challenging
624 scenarios, where multiple majority and minority classes are present and the
625 relationships between them are no longer obvious. As we wanted to capture
626 pairwise relations between objects we focused one OVO decomposition, as it
627 does not affect the distributions of positive and negative classes.

628 To check the flexibility of the proposed approach we have tested it with
629 three different base classifiers: CART, BPNN and SVM. Experimental study,
630 backed-up by a thorough statistical analysis indicate that it is possible to
631 significantly boost the OVO approach performance for multi-class imbalanced
632 data when enhancing it with ensemble classifiers. Regardless of the used
633 committee approach in most cases we were able to outperform traditional
634 OVO approach utilizing single classifiers with pre-processing algorithms, as
635 well as state-of-the-art multi-class solutions for skewed data.

636 Obtained results allow us to formulate recommendations for selecting
637 ensemble schemes. SMB technique is the best choice when considering CART
638 and SVM as base classifiers. At the same time BPNN should be used with

639 Easy approach. As result differences between several methods are not high,
640 we point out the high effectiveness of Easy, SMB and RBO ensemble methods
641 for empowering OVO techniques in multi-class imbalance scenarios.

642 We highlighted the effectiveness of synergy between decomposition strat-
643 egy and ensemble learning in the multi-class imbalanced datasets. Above all,
644 we must emphasize that our work provides the basis for addressing the multi-
645 class imbalance classification problems with a combination of decomposition
646 and multiple classifier approaches.

647 In our study we only presented the VOTE aggregation strategy. There-
648 fore, it is interesting to develop the analysis of the other aggregations in the
649 OVO scheme and trained combiners like Error-Correcting Output Codes or
650 Decision Templates. Additionally, we would like to extend our proposal to
651 include recent findings in OVO decomposition that take into consideration
652 the dynamic classifier selection [24] and competence-based weighting [26] to
653 remove the non-competent classifiers from the pool.

654 **Acknowledgments**

655 Zhongliang Zhang was supported by the National Science Foundation of
656 China (NSFC Proj. 61273204) and CSC Scholarship Program (CSC NO.
657 201406080059).

658 Bartosz Krawczyk was supported by the Polish National Science Center
659 under the grant no. DEC-2013/09/B/ST6/02264.

660 Salvador García and Francisco Herrera were partially supported by the
661 Spanish Ministry of Education and Science under Project TIN2014-57251-P
662 and the Andalusian Research Plan P10-TIC-6858, P11-TIC-7765.

663 Alejandro Rosales-Pérez was supported by the CONACyT grant 329013.

664 **References**

- 665 [1] M.H.B. Abidine and B. Fergani. A new multi-class wsvm classification
666 to imbalanced human activity dataset. *Journal of computers*, 9(7):1560–
667 1565, 2014.
- 668 [2] R. Baeza-Yates and B. Ribeiro-Neto. *Modern information retrieval*.
669 ACM press, New York, 1999.
- 670 [3] R. Barandela, R.M. Valdovinos, and J.S. Sanchez. New applications of
671 ensembles of classifiers. *Pattern Analysis & Applications*, 6(3):245–256,
672 2003.

- 673 [4] G.E.A.P.A. Batista, R.C. Prati, and M.C. Monard. A study of the be-
674 havior of several methods for balancing machine learning training data.
675 *ACM SIGKDD Explorations Newsletter*, 6(1):20–29, 2004.
- 676 [5] J. Blaszczynski and J. Stefanowski. Neighbourhood sampling in bagging
677 for imbalanced data. *Neurocomputing*, 150:529–542, 2015.
- 678 [6] L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen. *Classification*
679 *and regression trees*. CRC press, 1984.
- 680 [7] L. Cerf, D. Gay, N. Selmaoui-Folcher, B. Cremilleux, and J.F. Boulicaut.
681 Parameter-free classification in multi-class imbalanced data sets. *Data*
682 *& Knowledge Engineering*, 87:109–129, 2013.
- 683 [8] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer. Smote:
684 synthetic minority over-sampling technique. *Journal of artificial intelli-*
685 *gence research*, 16:321–357, 2002.
- 686 [9] N.V. Chawla, N. Japkowicz, and A. Kotcz. Editorial: special issue
687 on learning from imbalanced data sets. *ACM SIGKDD Explorations*
688 *Newsletter*, 6(1):1–6, 2004.
- 689 [10] N.V. Chawla, A. Lazarevic, L.O. Hall, and K.W. Bowyer. *Knowledge*
690 *Discovery in Databases: PKDD 2003*, volume 2838 of *Lecture Notes in*
691 *Computer Science*, chapter SMOTEBoost: Improving prediction of the
692 minority class in boosting, pages 107–119. Springer-Verlag, 2003.
- 693 [11] D.A. Cieslak, T.R. Hoens, N.V. Chawla, and W.P. Kegelmeyer.
694 Hellinger distance decision trees are robust and skew-insensitive. *Data*
695 *Mining and Knowledge Discovery*, 24(1):136–158, 2012.
- 696 [12] W.M. Czarnecki and J. Tabor. Two ellipsoid support vector machines.
697 *Expert Syst. Appl.*, 41(18):8211–8224, 2014.
- 698 [13] W.M. Czarnecki and J. Tabor. Multithreshold entropy linear classifier:
699 Theory and applications. *Expert Syst. Appl.*, 42(13):5591–5606, 2015.
- 700 [14] S. Datta and S. Das. Near-bayesian support vector machines for imbal-
701 anced data classification with equal or unequal misclassification costs.
702 *Neural Networks*, 70:39–52, 2015.

- 703 [15] J.-F. Díez-Pastor, J.-J. Rodríguez Díez, C. García-Osorio, and L. I.
704 Kuncheva. Random balance: Ensembles of variable priors classifiers
705 for imbalanced data. *Knowl.-Based Syst.*, 85:96–111, 2015.
- 706 [16] T. Fawcett. An introduction to roc analysis. *Pattern recognition letters*,
707 27(8):861–874, 2006.
- 708 [17] R. Feraud and F. Clerot. A methodology to explain neural network
709 classification. *Neural Networks*, 15(2):237–246, 2002.
- 710 [18] A. Fernandez, V. Lopez, M. Galar, M.J. del Jesus, and F. Her-
711 rera. Analysing the classification of imbalanced data-sets with multiple
712 classes: Binarization techniques and ad-hoc approaches. *Knowledge-
713 Based Systems*, 42:97–110, 2013.
- 714 [19] F. Fernandez-Navarro, C. Hervás-Martínez, and P.A. Gutiérrez. A dy-
715 namic over-sampling procedure based on sensitivity for multi-class prob-
716 lems. *Pattern Recognition*, 44(8):1821–1833, 2011.
- 717 [20] C. Ferri, J. Hernández-Orallo, and R. Modroiu. An experimental com-
718 parison of performance measures for classification. *Pattern Recognition
719 Letters*, 30(1):27–38, 2009.
- 720 [21] J.H. Friedman. Another approach to polychotomous classification. *Techni-
721 cal report: <http://www.stat.stanford.edu/jhf>*, 1996.
- 722 [22] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera.
723 An overview of ensemble methods for binary classifiers in multi-class
724 problems: Experimental study on one-vs-one and one-vs-all schemes.
725 *Pattern Recognition*, 44(8):1761–1776, 2011.
- 726 [23] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Her-
727 rera. A review on ensembles for the class imbalance problem: bagging-
728 , boosting-, and hybrid-based approaches. *IEEE Transactions on
729 Systems, Man, and Cybernetics-Part C: Applications and Reviews*,
730 42(4):463–484, 2012.
- 731 [24] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Her-
732 rera. Dynamic classifier selection for one-vs-one strategy: Avoiding
733 non-competent classifiers. *Pattern Recognition*, 46(12):3412–3424, 2013.

- 734 [25] M. Galar, A. Fernandez, E. Barrenechea, and F. Herrera. Empower-
735 ing difficult classes with a similarity-based aggregation in multi-class
736 classification problems. *Information Sciences*, 264:135–157, 2014.
- 737 [26] M. Galar, A. Fernndez, E. Barrenechea, and F. Herrera. Drcw-ovo:
738 Distance-based relative competence weighting combination for one-vs-
739 one strategy in multi-class problems. *Pattern Recognition*, 48(1):28–42,
740 2015.
- 741 [27] S. Garcia, J. Derrac, I. Triguero, C.J. Carmona, and F. Herrera.
742 Evolutionary-based selection of generalized instances for imbalanced
743 classification. *Knowledge-Based Systems*, 25(1):3–12, 2012.
- 744 [28] S. Garcia, A. Fernandez, J. Luengo, and F. Herrera. Advanced non-
745 parametric tests for multiple comparisons in the design of experiments
746 in computational intelligence and data mining: Experimental analysis
747 of power. *Information Sciences*, 180(10):2044–2064, 2010.
- 748 [29] S. García, J. Luengo, and F. Herrera. Tutorial on practical tips of
749 the most influential data preprocessing algorithms in data mining.
750 *Knowledge-Based Systems*, 98:1 – 29, 2016.
- 751 [30] H. Guo, Y. Li, Y. Li, X. Liu, and J. Li. Bpso-adaboost-knn ensemble
752 learning algorithm for multi-class imbalanced data classification. *Eng.*
753 *Appl. of AI*, 49:176–193, 2016.
- 754 [31] H. Haibo and E.A. Garcia. Learning from imbalanced data. *IEEE Trans-*
755 *actions on Knowledge and Data Engineering*, 21(9):1263–1284, 2009.
- 756 [32] D.J. Hand and R.J. Till. A simple generalisation of the area under the
757 ROC curve for multiple class classification problems. *Machine Learning*,
758 45(2):171–186, 2001.
- 759 [33] S. Kang, S. Cho, and P. Kang. Constructing a multi-class classifier using
760 one-against-one approach with different binary classifiers. *Neurocomput-*
761 *ing*, 149:677–682, 2015.
- 762 [34] B. Krawczyk. Combining one-vs-one decomposition and ensemble learn-
763 ing for multi-class imbalanced data. In R. Burduk, K. Jackowski,
764 M. Kurzyński, M. Woźniak, and A. Żolnierek, editors, *Proceedings of the*

- 765 *9th International Conference on Computer Recognition Systems CORES*
766 *2015*, pages 27–36, 2016.
- 767 [35] B. Krawczyk, M. Galar, Ł Jeleń, and F. Herrera. Evolutionary under-
768 sampling boosting for imbalanced classification of breast cancer malign-
769 nancy. *Applied Soft Computing*, 2015.
- 770 [36] B. Krawczyk, M. Woźniak, and G. Schaefer. Cost-sensitive decision tree
771 ensembles for effective imbalanced classification. *Appl. Soft Comput.*,
772 14:554–562, 2014.
- 773 [37] M. Kurzyński and M. Woźniak. Combining classifiers under probabilistic
774 models: experimental comparative analysis of methods. *Expert Systems*,
775 29(4):374–393, 2012.
- 776 [38] Y. Li, H. Guo, X. Liu, Y. Li, and J. Li. Adapted ensemble classification
777 algorithm based on multiple classifier system and feature selection for
778 classifying multi-class imbalanced data. *Knowl.-Based Syst.*, 94:88–104,
779 2016.
- 780 [39] B. Liu, Z.F. Hao, and E.C.C. Tsang. Nesting one-against-one algorithm
781 based on svms for pattern classification. *IEEE Transactions on Neural*
782 *Networks*, 19(12):2044–2052, 2008.
- 783 [40] X.Y. Liu, J.X. Wu, and Z.H. Zhou. Exploratory undersampling for
784 class-imbalance learning. *IEEE Transactions on Systems, Man, and*
785 *Cybernetics-Part B: Cybernetics*, 39(2):539–550, 2009.
- 786 [41] V. Lopez, A. Fernandez, S. Garcia, V. Palade, and F. Herrera. An
787 insight into classification with imbalanced data: Empirical results and
788 current trends on using data intrinsic characteristics. *Information Sci-*
789 *ences*, 250:113–141, 2013.
- 790 [42] V. Lopez, A. Fernandez, and F. Herrera. On the importance of the val-
791 idation technique for classification with imbalanced datasets: Address-
792 ing covariate shift when data is skewed. *Information Sciences*, 257:1–13,
793 2014.
- 794 [43] J.G. Moreno-Torres, J.A. Saez, and F. Herrera. Study on the impact of
795 partition-induced dataset shift on k-fold cross-validation. *IEEE Transac-*
796 *tions on Neural Networks and Learning Systems*, 23(8):1304–1312, 2012.

- 797 [44] W. Penar and M. Woźniak. Cost-sensitive methods of constructing hi-
798 erarchical classifiers. *Expert Systems*, 27(3):146–155, 2010.
- 799 [45] M. D. Pérez-Godoy, A. J. Rivera, C. J. Carmona, and del Jesús.
- 800 [46] W. Prachuabsupakij and N. Soonthornphisaj. Clustering and combined
801 sampling approaches for multi-class imbalanced data classification. *Ad-
802 vances in Information Technology and Industry Applications*, 136:717–
803 724, 2012.
- 804 [47] P. Pramokchon and P. Piamsanga. *Recent Advances in Information and
805 Communication Technology*, volume 265 of *Advances in Intelligent Sys-
806 tems and Computing*, chapter Reducing Effects of Class Imbalance Dis-
807 tribution in Multi-class Text Categorization, pages 263–272. Springer-
808 Verlag, 2014.
- 809 [48] J. A. Sáez, B. Krawczyk, and M. Woźniak. Analyzing the oversam-
810 pling of different classes and types of examples in multi-class imbalanced
811 datasets. *Pattern Recognition*. 10.1016/j.patcog.2016.03.012.
- 812 [49] C. Seiffert, T.M. Khoshgoftaar, J. Van Hulse, and A. Napolitano. Rus-
813 boost: A hybrid approach to alleviating class imbalance. *IEEE Transac-
814 tions on Systems, Man and Cybernetics-Part A: Systems and Humans*,
815 40(1):185–197, 2010.
- 816 [50] Y. Sun, A.K.C. Wong, and M.S. Kamel. Classification of imbalanced
817 data: A review. *Journal of Pattern Recognition and Artificial Intelli-
818 gence*, 23(4):687–719, 2009.
- 819 [51] M.A. Tahir, J. Kittler, and F. Yan. Inverse random under sampling for
820 class imbalance problem and its application to multi-label classification.
821 *Pattern Recognition*, 45(10):3738–3750, 2012.
- 822 [52] Y.C. Tang, Y.Q. Zhang, N.V. Chawla, and S. Krasser. Svms modeling
823 for highly imbalanced classification. *IEEE Transactions on Systems,
824 Man, and Cybernetics-Part B: Cybernetics*, 39(1):281–288, 2009.
- 825 [53] P. Trajdos and M. Kurzyński. An extension of multi-label binary rel-
826 evance models based on randomized reference classifier and local fuzzy

- 827 confusion matrix. In *Intelligent Data Engineering and Automated Learning - IDEAL 2015 - 16th International Conference Wroclaw, Poland, October 14-16, 2015, Proceedings*, pages 69–76, 2015.
- 828
- 829
- 830 [54] V. Vapnik. *Statistical learning theory*. Wiley, New York, 1998.
- 831 [55] P. Vorraboot, S. Rasmeequan, K. Chinnasarn, and C. Lursinsap. Improving classification rate constrained to imbalanced data between overlapped and non-overlapped regions by hybrid algorithms. *Neurocomputing*, 152:429–443, 2015.
- 832
- 833
- 834
- 835 [56] S. Wang and X. Yao. Diversity analysis on imbalanced data sets by using ensemble models. In *IEEE Symposium on Computational Intelligence and Data Mining, 2009*, pages 324–331, 2009.
- 836
- 837
- 838 [57] S. Wang and X. Yao. Multiclass imbalance problems: Analysis and potential solutions. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 42(4):1119–1130, 2012.
- 839
- 840
- 841 [58] G. Weiss and Y. Tian. Maximizing classifier utility when there are data acquisition and modeling costs. *Data Mining and Knowledge Discovery*, 17(2):253–282, 2008.
- 842
- 843
- 844 [59] G.M. Weiss. Mining with rarity: a unifying framework. *ACM SIGKDD Explorations Newsletter*, 6(1):7–19, 2004.
- 845
- 846 [60] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83, 1945.
- 847
- 848 [61] T. Wilk and M. Woźniak. Soft computing methods applied to combination of one-class classifiers. *Neurocomputing*, 75(1):185–193, 2012.
- 849
- 850 [62] M. Woźniak, M. Graña, and E. Corchado. A survey of multiple classifier systems as hybrid systems. *Information Fusion*, 16:3–17, 2014.
- 851
- 852 [63] Z.H. Zhou and X.Y. Liu. On multi-class cost-sensitive learning. *Computational Intelligence*, 26(3):232257, 2010.
- 853