FUSION OF TECHNIQUES FOR HANDLING THE IMBALANCED TRAINING SAMPLE PROBLEM

R. BARANDELA[1], J. S. SÁNCHEZ[2], V. GARCÍA[1], E. RANGEL[1]
[1]Laboratorio de Reconocimiento de Patrones, Instituto Tecnológico de Toluca, Ave. Tecnológico s/n, 52140 Metepec, México
rbarandela@hotmail.com
[2]Dept. Lenguajes y Sistemas Informáticos, Universidad Jaume I, E-12071 Castellón, España
sanchez@uji.es

Abstract. *The problem of imbalanced training data in supervised pattern recognition methods is receiving growing attention. Imbalanced data means that one class is much more represented than the others in the training sample. It has been observed that this situation, which arises in several practical domains, may produce an important deterioration of the classification accuracy, in particular with patterns belonging to the less represented classes. In the present paper, we review several aspects related to this subject. We report experimental results that point at the convenience of correctly downsizing the majority class while simultaneously employing a weighted distance function. Novel procedures for editing-condensing the training sample are also introduced.*

1. INTRODUCTION

Design of supervised pattern recognition methods is based on a training sample: a collection of examples previously analyzed by a human expert. Performance of the resulting classification system depends on the quantity and the quality of the information contained in the training sample. This dependence is particularly strong in the case of non-parametric classifiers since these systems do not rest upon any probabilistic assumption about the class models. Researchers in the pattern recognition area have very early realized that the training sample must satisfy some requirements in order to guarantee good classification results. From the start, two assumptions were established:

1. The set of c classes considered in the training set covers the whole space of the relevant classes.
2. The training instances used to teach the classifier how to identify each class are actually members of that class.

As the number of practical applications of these methods grows, obtained experience has gradually indicated the necessity of some other requisites for the system to reach satisfactory results. Among them:

3. The training sample represents the population (all strata in the population must be represented in the sample and in the same proportion as they occur in the population).
4. The considered features must permit discrimination.
5. The size/dimensionality rate of the sample is high enough (5-10 training patterns for each attribute).

Recently, concern has arisen about the complications produced by imbalance in the training sample. A training sample is said to be imbalanced when one of the classes (the minority one) is heavily under-represented in comparison to the other (the majority) class. For simplicity, and following the common published practice, we consider here only two-class cases. This imbalanced situation is usual in several real domains where the cost of misclassification of patterns from the minority class far outweighs the other type of cost [1].

It has been observed that imbalanced training samples may cause a significant deterioration in the performance attainable by supervised methods. Most of the attempts at addressing this problem can be sorted into three categories [2]:

a) Over-sampling (re-sampling some training patterns multiple times) the minority class so as to match the size of the other class.
b) Downsizing the majority class so as to match the size of the other class.
c) Internally biasing the discrimination based process so as to compensate for the class imbalance.

It is also currently argued whether accuracy is the best criterion to assess the classifier's performance in imbalanced domains and several other criteria have been proposed. Perhaps the most widely accepted is the geometric mean, $g=(a^+ \cdot a^-)^{½}$, where a^+ is the accuracy on cases from the minority class and a^- is the accuracy on cases from the majority one.

In the present work, we present preliminary results of a more extensive research that we are conducting to explore all the issues linked to the imbalanced training samples. We employ the widespread used 1-NN (Nearest Neighbor) rule, because of its well-known performance and its flexibility. Experimental results with real datasets are reported.

2. EXPERIMENTED STRATEGIES

The NN rule is a well-known supervised nonparametric classifier that combines its conceptual simplicity and an asymptotic error rate conveniently bounded in terms of the optimal Bayes error. In its classical manifestation, given a set of n previously labeled prototypes or training sample (TS), this classifier assigns any given sample to the class indicated by the label of the closest prototype in the TS. More generally, the *k*-NN rule maps any sample to the pattern class most frequently represented among its k closest neighbors.

Nevertheless, the NN classifiers also suffer from certain drawbacks. The performance of these rules, as with any non-parametric method, is extremely sensitive to incorrectness or imperfections in the training sample. On the other hand, its applicability to real-time problems, with a large set of training patterns of high dimensionality, can become prohibitive because of the immense computational loads required for searching the nearest neighbors of each new pattern in the training sample.

In this initial phase of our research, we focus on techniques of the categories b) and c), as above-mentioned. We are also assessing the benefits of employing a combination of methods of the two types of strategies.

For downsizing the majority class we have tried with three different algorithms. Two of them are in the group of the Editing techniques: the now classical Wilson's proposal [3] and the k-NCN (Nearest Centroid Neighborhood; [4]). Tomek [5] proposed to repeatedly apply Wilson's method while k-NCN is used here for the first time in that way. These two techniques are aimed at filtering the training sample by removal of the noisy or atypical elements. A noisy pattern is defined as that element whose label does not coincide with the most represented class in its neighborhood. Wilson's idea and k-NCN editing differ to each other in the way they define the neighborhood concept. Benefits of the two editing techniques by increasing the classification accuracy of the NN rule have been corroborated (e.g., [6]). However, they do not produce an important amount of removed patterns.

On the other hand, Modified Selective [7] is an algorithm specifically designed for reducing the training sample size. The algorithm is based on the idea of a consistent subset [8] and, with a rather simple algorithm (in terms of computation time and memory requirements), it guarantees a satisfactory approximation to the decision boundaries as they are defined by the whole training sample. Experimental comparison with several other pruning algorithms [7] corroborated this statement.

Tomek [5] very early mentioned the convenience of combining Edition with the consistent subset: "when the edited set is processed by CNN (Hart's algorithm) or other methods, the result in reduction is much more significant than that attainable on the original set". Employment of these compound techniques has been often promoted and proved in several practical applications. For instance, it has been employed with the Nearest Neighbor rule [9] and with the Multi-Layer Perceptron [10]. In the experiments reported here we have employed the combination Wilson + Modified Selective and, for the first time, k-NCN + Modified Selective.

As a technique for internally biasing the discrimination procedure, we have experimented with a modification of the Euclidean metric that can be regarded as a weighted distance function. With this modification, when classification of a new pattern Y is attempted, and in the search through the training sample of its nearest neighbor, the following quantity must be computed for each training instance:

$$d_W(Y, x_0) = (n_i/n)^{1/m}\, d_E(Y, x_0)$$

where:

x_0 is a training pattern representing class i.
n_i is the number of training patterns from class i.
n is the training sample size
m is the dimensionality of the feature space
$d_E(\,.\,)$ is the Euclidean metric.

The idea is to compensate for the imbalance in the training sample without actually altering the imbalance. Weights are assigned, unlike in the usual weighted k-NN rule proposals, to the respective classes and not to the individual prototypes. In that way, since the weighting factor is greater for the majority class than for the minority one, distance values to training instances of the minority class are reduced much more than the distance values to training examples of the majority class. This produces a tendency for the new patterns to find their nearest neighbor among the cases of the minority class, increasing the accuracy in that class.

The three downsizing algorithms above-mentioned are explained in more detail in the following subsections.

2.1 Wilson's editing

This corresponds to the first proposal to edit the NN rule. In a few words, it consists of applying the k-NN classifier to estimate the class label of all prototypes in the TS and discard those samples whose class label does not agree with the class associated with the largest number of the k neighbors.

Thus, the Wilson's editing procedure can be written as follows:

- Let $S = X$. (X is the original TS, and S will be the resulting or edited TS)
- For each x_i in X do:
    a) Find the k-NN of x_i in $X - \{x_i\}$.
    b) Discard x_i from S if there is a majority of NNs from a different class.

2.2 k-NCN editing

This is a way of Wilson's algorithm particularized for the case of using the k-NCN classification rule to estimate the class label of prototypes.

It is worth mentioning that the k-NCN classifier is thought to obtain a more accurate information about prototypes and more specially, for those close to decision boundaries. In general, this is expected to result in a practical improvement of the corresponding editing procedure. The k-NCN editing algorithm consists of the following steps:

- Let $S = X$.
- For each x_i in X do:
    a) Find the k-NCN of x_i in $X - \{x_i\}$.

b) Discard x_i from S if there is a majority of NCNs from a different class.

The k-NCN of x_i are searched in T such that:

a) They are as near to x_i as possible

b) Their centroid is also as close to x_i as possible.

Both conditions can be satisfied through an iterative procedure [11] in the following way:

--The first neighbor of x_i is its nearest neighbor, q_1

--The jth neighbor q_j, $j > 1$, is such that the centroid of this and previously selected neighbors, $q_1, \ldots, q_{j-1}$, is the closest to x_i

This definition gives rise to a neighborhood in which both closeness and spatial distribution of neighbors are taken into account because of the centroid criterion.

2.3 Modified Selective algorithm

According to Hart's statement, the Condensed Subset (CS) is a subset S of the TS such that every member of TS is closer to a member of S of the same class than to a member of S of a different class. Ritter et al. [12] have changed this concept in their Selective Subset (SS). They defined the SS as that subset S such that every member of TS must be closer to a member of S of the same class than to a member of TS (instead of S) of a different class. Their purpose is to eliminate the order-dependence of the building algorithm. For getting the SS, an algorithm too extensive and complicated (in memory and execution time) to be described here is proposed. As Ritter et al. have recognized, their algorithm does not necessarily conduct to a unique solution. Moreover, although they stated the importance of selecting "samples near the decision boundaries", this requisite is not included in the criteria serving as a basis for their SS. After describing R_i as the set of all related neighbors to the training pattern x_i, that is, the set of all x_j in TS such that x_j is of the same class of x_i and is closer to x_i than the nearest neighbor of x_i in TS of a different class than x_i, they merely defined their SS as "the smallest subset containing at least one member of R_i for each training pattern x_i. As a natural consequence, the algorithm they established does not guarantee the best approximation to the decision boundary.

The reduction technique employed in this work, the Modified Selective (MSS), rests upon a modification of the definition in the preceding paragraph. The MSS is defined as that subset of the TS containing, for every x_i in TS, that element of its R_i that is the <u>nearest</u> to a class different than that of x_i. Although the main purpose of this modification is to strengthen the condition to be fulfilled by the reduced subset in order to attain a best approximation to the decision boundaries, as a byproduct the resulting algorithm is much simpler than that proposed in the Selective approach. This algorithm, for the two-class case, consists of the following steps. Only the class 1 is considered, class 2 being afterwards processed in a similar way:

1. Let $S = \emptyset$.
2. Let $x_1, x_2, \ldots, x_{n1}$ be the training cases of class 1.
   These instances are ordered such that $D_1 < D_2 < \ldots < D_{ni}$ where D_i stands for the minimum distance from x_i to the class 2.
3. Place x_1 in S. Put $KN = n_1 - 1$.
   For $i = 2$ to n_1 do:
   If $d(x_1, x_i) < D_i$ then $\{K_i = 0;\ KN = KN - 1;\}$
   Else $K_i = 1$;
4. For $i = 2$ to n_1 begin I
   a) Put $IND = 0$; If $KN = 0$ then exit;
   Else {If $K_i = 1$ then $\{K_i = 0;\ KN = KN - 1;$ Put x_i in S; $IND = 1;\}\}$
   b) For $j = i + 1$ to n_1 begin II
   If ($K_j = 1$ and $d(x_i, x_j) < D_j$) then
   $K_j = 0;\ KN = KN - 1;$
   If $IND = 0$ then $\{IND = 1;$ Put x_i in $S;\}\}$
   End II;
   End I;

The first training instance, after they have been ordered in step 2, is always selected for the pruned subset (step 3) since it is the nearest to the other class and, at least, its own related neighbor. Also in step 3, all the other sets R_i where x_1 is present are detected and marked by letting $K_i = 0$. KN is used for the current number of training cases still not represented (that is, none of its related neighbors is currently included) in S. In step 4 it is searched through the rest of the training prototypes (in increasing order of their distances to class 2). The purpose is to look after those that must be incorporated to S either because they have not been represented by any previous instance or because they are related neighbors of a posterior (more far to the other class) not yet represented training case. IND is a flag that prevents duplication of an instance in S.

3. EXPERIMENTAL RESULTS

The experiments here reported were conducted with four real datasets taken from the UCI Repository [13]. In each dataset, five-fold cross validation was employed (80% for the training sample and 20% for a test set). Results to be presented hereafter represent the averaged values of the five replications. To facilitate comparison with other published results [14], in the Glass set the problem was transformed for discriminate class 7 against all the other classes and in the Vehicle dataset the task is to classify class 1 against all the others. Satimage dataset was also mapped to configure a two-class case, the training patterns of classes 1, 2, 3, 5 and 6 were joined to form a unique class and the original class 4 was left as the minority one. Descriptions of the datasets are in Table 1.

Several experiments were conducted with each dataset. Most of them are aimed at downsizing the majority class while filtering its noisy training elements and removing the redundant examples.

The following techniques were applied to the majority class:

1. Modified Selective
2. Wilson's Editing
3. Wilson's Editing repeatedly
4. Wilson's Editing (once and repeatedly) in combination with Modified Selective
5. k-NCN Editing

6. *k*-NCN Editing repeatedly
7. *k*-NCN Editing (once and repeatedly) in combination with Modified Selective.

The two latter techniques have been employed for the first time in the present work. Besides, and as a complementary exploration, Modified Selective, Wilson's Editing and Wilson's Editing combined with Modified Selective were applied to both classes in three of the datasets (minority class of the Glass dataset was excepted of the editing application for reasons to be discussed below).

In all the mentioned experiments, after processing the majority class (or both classes), patterns of the test set were classified according to the NN rule. Usual Euclidean metric as well as the proposed weighted distance was employed for these classification tasks. Averaged results of the g criterion are shown in Table 2.

The most interesting facts arising from the lecture of Table 2 are:

1. The best results are always obtained when the weighted distance is employed for classification. In all datasets, this technique alone produces improvement in the g values.
2. Benefits of the Wilson's Editing are well known for increasing the classification accuracy. This beneficial effect is corroborated for the geometric mean. The same can be stated about the performance obtained when processing the training sample with the combination Wilson's Editing + Modified Selective.
3. Repeated application of *k*-NCN Editing shows similar or better results than those of the classical technique. In general, this editing algorithm allows more repetitions (the procedure is stopped when no further removals are produced or when a class becomes empty of training instances). However, these additional repetitions not always cause important changes. Combination *k*-NCN Editing +Modified Selective yields also good results.
4. Glass dataset suffers not only from the imbalance issue. In addition to that, the minority class is too small. Adequacy of the training sample size must be measured by considering the number of training cases of the smallest class and not that of the whole training sample. For the minority class in Glass dataset, the size/dimensionality rate is very low: 2.7 examples for each attribute. Either more cases of this class are added to the training sample or the number of attributes is reduced by a convenient selection. For this reason, we did not try to carry out any kind of filtering of the minority class in this dataset. Nevertheless, the weighted distance produces an increase in the g value, albeit not very significant.
5. In the other three datasets, exploration was done to evaluate the convenience of processing too the minority class, for removing noisy and redundant examples. Improvement was obtained in the Phoneme dataset when Wilson's Editing was applied in both classes. A deeper study of this subject is required.

Table 3 presents a detailed analysis of the behavior of the Vehicle dataset with all the experimented variants. This dataset was selected for allowing comparison with the report in [14].

The first thing to note in Table 3 is the fact that increase in a^+ and in the g value means, in general, decrease of the values in a^- and in overall accuracy. A similar situation is reported in [14]. These results are in contradiction with what should be expected: a classifier with good performance on cases from both classes. This issue deserves careful consideration. In some real applications, it may not be convenient to get an increase in the classification accuracy of the minority class (and in the geometric mean) at the cost of reducing this accuracy in the majority class and in the overall accuracy. Of course, that would depend on the misclassification costs of both types.

Downsizing the majority class cause radical changes in the proportions in which both classes are represented into the training sample. However, these modifications permit to obtain better results, although the original proportions are kept in the corresponding test sets. That is, accuracy is improved when working with classifiers that have been trained with different class priors than the priors of the test cases.

As it usually happens, normal Wilson's Editing procedure (in both classes) and the combination Wilson + Modified Selective improve classification accuracy. The combination, besides, considerably reduces the sample size (only 12.8% of the original training sample for the Vehicle dataset). However, an important deterioration of the geometric mean value is produced with these techniques when used in the traditional way.

Superiority of the g values when employing the weighted distance over those obtained with the traditional Euclidean metric becomes less pronounced as the proportion of the minority class in the training sample is increased.

4. PRELIMINARY CONCLUSIONS

Replicating the minority class to eliminate imbalance in the training sample does not add new information to the system. Moreover, to work in that direction means to worsen the known computational burden of learning algorithms such as the Nearest Neighbor rule and the Multi-Layer Perceptron.

On the other hand, since downsizing the majority class can result in throwing away some useful information, this kind of process must be done carefully. Editing and pruning algorithms offer a good alternative for using this opportunity and to remove noisy and redundant training examples. In this work, experiments with some of these algorithms were, in general, successful. It has been shown that *k*-NCN editing, a relatively new technique, produces good results when applied repeatedly and that it yields a good combination with the Modified Selective. The problem with all these techniques is that they do not permit control on the number of training cases to be removed. In that way, eliminated patterns can be too many or too few to adequately solve the imbalance problem. Genetic algorithms could be of interest, in particular that of Kuncheva and Jain [15] that address the selection of patterns and features simultaneously. Of course, it will be required to transform the algorithm to adapt it to the imbalance situation.

The Weighted distance showed itself as a good resource to transform the classification procedure for taking into account the disproportion of the different class representations. Other weighting factors must be studied. We have reported here evidence enough to advocate the necessity of researching the joint employment of strategies of different sorts.

One of the most promising research lines is based on the employment of a classifier ensemble [16]. That is, creating an ensemble with several classifiers and distributing the training sample to reach balance in each one of the resulting learning samples. We have started to explore in this direction and already with a very simple design and a majority-voting schema, interesting results have begun to appear. Geometric mean values have improved to 79.6 for the Satimage dataset and to 70.5 for the Vehicle one (see Table 2 for comparison). The study of this subject involves a great variety of possibilities that we will attempt to cover in the next future.

Acknowledgements.

This work has been partially supported by grants 32016-A and 744.99P from the *Mexican CONACyT* and *COSNET,* respectively, and by grant TIC2000-1703-C03-03 from the *Spanish CICYT.*

References

[1] M. Kubat, R. Holte and S. Matwin. Detection of Oil-Spills in Radar Images of Sea Surface. *Machine Learning*, 30, 195-215, 1998.

[2] T. Eavis and N. Japkowicz. A Recognition-based Alternative to Discrimination-based Multi-Layer Perceptrons. Workshop on Learning from Imbalanced Data Sets. *Technical Report WS-00-05*, AAAI Press, 2000.

[3] D. L. Wilson. Asymptotic Properties of Nearest Neighbor Rules using Edited Data, *IEEE Transactions on Systems, Man and Cybernetics*, 2(3), 408-421, 1972.

[4] F. J. Ferri, J.S. Sánchez and F. Pla. Editing prototypes in the finite sample size case using alternative neighborhoods. In: Advances in Pattern Recognition, Lecture Notes in Computer Science 1451, Springer Verlag, 620-629, 1998.

[5] I. Tomek. An experiment with the Edited Nearest Neighbor rule. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-6, 448-452, 1976.

[6] J. S. Sánchez and R. Barandela. Experiments with a new family of edited nearest neighbor rules. Submitted to *Pattern Recognition, 2001*.

[7] R. Barandela, N. Cortés and A. Palacios. The nearest neighbor rule and the reduction of the training sample size. In: *Proceedings of the 9th Spanish Symposium on Pattern Recognition and Image Analysis* I, 103-108, 2001.

[8] P. E. Hart. The Condensed Nearest Neighbor rule. *IEEE Transactions on Information Theory*, 6, 4, 515-516, 1968.

[9] R. Barandela. Una metodología para el reconocimiento de patrones en tareas geólogo-geofísicas. *Geofísica Internacional*, 34, 4, 399-405, 1995.

[10] E. Gasca and R. Barandela. Influencia del preprocesamiento de la muestra de entrenamiento en el poder de generalización del Perceptron Multicapa. In *Proc. 6th Brazilian Symposium on Neural Networks*, Rio de Janeiro, 2000.

[11] B. B. Chaudhuri. A new definition of neighborhood of a point in a multi-dimensional space. *Pattern Recognition Letters*, 17, 11-17, 1996.

[12] G. L. Ritter, H.B. Woodruff, S.R. Lowry and T.J. Isenhour. An Algorithm for Selective Nearest Neighbor Decision Rule. *IEEE Transactions on Information Theory*, 21, 6, 665-669, 1975.

[13] C.J. Merz and P.M. Murphy. *UCI Repository of Machine Learning Databases*, University of California at Irvine, Department of Information and Computer Science. http://www.csi.uci.edu/~mlearn, 1998.

[14] M. Kubat and S. Matwin. Addressing the Curse of Imbalanced Training Sets: One-Sided Selection. *Proceedings of the 14th International Conference on Machine Learning*, 179-186, 1997.

[15] L. I. Kuncheva and L.C. Jain. Nearest neighbor classifier: Simultaneous editing and feature selection. *Pattern Recognition Letters*, 20, 1149-1156, 1999.

[16] D. Opitz and R. Maclin. Popular Ensemble Methods: An Empirical Study. *Journal of Artificial Intelligence Research*, 11, 169-198, 1999.

		Training sample		Test sample	
Dataset	**Features**	**Class 1**	**Class 2**	**Class 1**	**Class 2**
Phoneme	5	1268	3054	318	764
Satimage	36	500	4647	126	1162
Glass	9	24	150	5	35
Vehicle	18	170	508	42	126

Table 1. Characterization of the datasets employed in the experiments (all of them with two classes)

	Phoneme		Sat image		Glass		Vehicle	
	A	**B**	**A**	**B**	**A**	**B**	**A**	**B**
Original TS	73.8	76.0	70.9	75.9	86.7	**88.2**	55.8	59.6
Wilson's Editing								
1st application	74.9	75.7	72.6	76.1	86.2	87.9	62.8	64.9
2nd application	74.8	75.5	72.9	**76.2**	-	-	62.8	64.8
3rd application	74.6	75.3	73.0	**76.2**	-	-	64.0	65.8
4th application	74.6	75.3	-	-	-	-	-	-
Wilson + Selective								
1st application	74.5	72.4	74.0	74.2	86.1	86.2	65.7	65.6
2nd application	74.6	72.6	74.2	74.3	-	-	65.6	65.2
3rd application	74.5	72.5	74.2	74.3	-	-	65.8	65.7
4th application	74.5	72.5	-	-	-	-	-	-
***k*-NCN Editing**								
1st application	75.0	75.9	71.9	**76.2**	86.2	87.9	62.1	64.5
2nd application	74.7	75.4	72.3	**76.2**	86.2	87.9	64.1	66.3
3rd application	74.7	75.4	72.3	**76.2**	-	-	65.2	67.4
4th application	74.6	75.3	72.3	**76.2**	-	-	66.2	**68.3**
5th application	74.6	75.3	72.3	**76.2**	-	-	65.8	68.0
***k*-NCN + Selective**								
1st application	74.9	71.9	73.7	74.2	85.7	86.2	65.6	64.8
2nd application	71.6	71.9	74.4	74.8	85.7	86.2	66.2	65.8
3rd application	74.6	72.1	74.4	74.8	-	-	66.5	66.8
4th application	74.6	72.1	74.4	74.8	-	-	67.6	67.8
5th application	74.6	72.1	74.4	74.8	-	-	67.1	67.5
Selective class 2	74.0	70.0	72.3	73.0	86.2	86.6	60.3	60.3
Selective class 1-2	72.2	72.8	70.1	73.3	86.4	87.1	59.7	59.9
Wilson both classes	73.8	**76.7**	66.4	68.8	-	-	47.5	51.5
W+S both classes	72.4	72.8	65.7	67.1	-	-	50.1	51.5

Table 2. Experimental results of the four datasets with the experimented techniques. Averaged values of the g criterion. Figures in columns A: results obtained when employing Euclidean distance. Figures in columns B: results obtained when employing the Weighted distance. Best results for each dataset in bold.

		Euclidean distance				Weighted distance			
	minority class (%)	a+	a-	g	accuracy	a+	a-	g	accuracy
Original TS	25.1	37.6	82.8	55.8	71.5	45.2	78.5	59.6	70.2
Wilson's Editing									
1st application	27.8	50.5	78.2	62.8	71.3	56.2	75.0	64.9	70.3
2nd application	28.5	52.4	75.2	62.8	69.5	57.6	72.8	64.8	69.0
3rd application	28.6	54.8	74.7	64.0	69.7	59.5	72.7	65.8	69.4
Wilson + Selective									
1st application	57.4	61.4	70.3	65.7	68.1	59.0	73.0	65.6	69.5
2nd application	59.7	63.3	67.9	65.6	66.7	60.5	70.3	65.2	67.8
3rd application	60.3	64.8	66.9	65.8	66.4	61.9	69.8	65.7	67.8
***k*-NCN Editing**									
1st application	27.8	48.6	79.4	62.1	71.8	54.8	75.9	64.5	70.7
2nd application	28.5	53.8	76.3	64.1	70.8	60.0	73.3	66.3	70.1
3rd application	28.6	56.7	75.4	65.2	70.8	62.4	72.7	67.4	70.2
4th application	29.5	59.0	74.9	66.2	71.0	64.3	72.5	68.3	70.6
5th application	29.6	59.0	74.1	65.8	70.4	64.3	71.9	68.0	70.0
***k*-NCN + Selective**									
1st application	57.3	60.5	71.1	65.6	68.5	57.1	73.5	64.8	69.5
2nd application	59.9	64.3	68.1	66.2	67.2	61.0	71.0	65.8	68.5
3rd application	62.0	66.2	66.8	66.5	66.7	64.3	69.5	66.8	68.3
4th application	62.7	69.0	66.3	67.6	67.0	66.7	69.0	67.8	68.5
5th application	62.8	69.0	65.6	67.1	66.3	66.7	68.3	67.5	67.9
Selective class 2	49.9	47.6	76.3	60.3	69.1	47.6	76.3	60.3	69.1
Selective class 1-2	42.0	44.8	79.5	59.7	70.8	46.2	77.7	59.9	69.8
Wilson both classes	12.7	25.2	89.7	47.5	73.5	31.0	85.7	51.5	72.0
W+S both classes	33.9	28.6	87.8	50.1	72.9	31.0	85.5	51.5	71.9

Table 3. A more detailed report of the experimental results with the Vehicle dataset. a+ is the accuracy on cases from the minority class. a- is the accuracy on cases from the majority class. accuracy is the overall accuracy on the whole test set.