

# Informal identification of outliers in medical data

Jorma Laurikkala<sup>1</sup>, Martti Juhola<sup>1</sup> and Erna Kentala<sup>2</sup>

**Abstract.** Informal box plot identification of outliers in real-world medical data was studied. Box plots were used to detect univariate outliers directly whereas the box plotted Mahalanobis distances identified multivariate outliers. Vertigo and female urinary incontinence data were used in the tests. The removal of outliers increased the descriptive classification accuracy of discriminant analysis functions and nearest neighbour method, while the predictive ability of these methods reduced somewhat. Outliers were also evaluated subjectively by expert physicians, who found most of the multivariate outliers to truly be outliers in their area. The experts sometimes disagreed with the method on univariate outliers. This happened, for example, in heterogeneous diagnostic groups where also extreme values are natural. The informal method may be used for straightforward identification of suspicious data or as a tool to collect abnormal cases for an in-depth analysis.

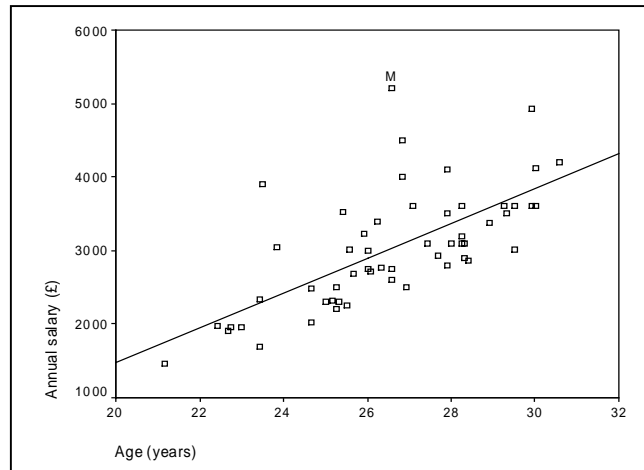
## 1 INTRODUCTION

There are many definitions for outliers which differ in words [1-3]. We use the one of Barnett and Lewis [1 pp. 4], who defined an outlier in a set of data to be "*an observation (or subsets of observations) which appears to be inconsistent with the remainder of that set of data*". This type of observations are often a problem in statistical analysis where outliers may especially lower the model fit. To illustrate, consider fitting a linear regression equation to data shown in Figure 1 [1 pp. 261-263]. The regression line runs nicely through the scatter plot of ages and salaries of electrical engineers ( $N=55$ , United Kingdom, 1974), but the most extreme observation M has a surprisingly strong impact on the analysis. Dropping M and refitting the regression line elevates the goodness of fit statistics  $R^2$  from 0.452 to 0.526.

There are various origins of outliers. Human error often produces unintentional outliers. Data entry may be incorrect and missing value codes are sometimes used as real data. Outliers are frequently generated as the result of the natural variation of population or process one cannot control. These outliers are from the intended population, but their values are unusual in comparison with the normal values. It is also possible to have an outlier that is not a member of population due to a sampling error [1,3].

Machine learning researchers often use the concept of noise rather than that of outliers. Noise is defined as mislabeled examples (class noise) or errors in the values of attributes (attribute noise) [4 pp. 92]. Outlier is, therefore, a broader concept which includes errors, as well as discordant data produced by the natural

variation of population [5 pp. 175]. Examples with class noise are outliers produced by sampling error [6], while attribute noise may or may not show in the data as outliers.



**Figure 1.** Ages and salaries of electrical engineers (UK, 1974)

Outlier identification (and consequent removal or accommodation) is a part of the data screening process which should be done routinely before statistical analyses [2,3]. The simplest and the most researched case is the identification of univariate outliers, where the distribution of a single variable is examined [1]. Extreme data values are obvious outlier candidates. When the distribution is symmetric, we suspect that candidate outliers are the extremes of the left or right tail. Correspondingly, the identified outliers are referred to as the lower and upper univariate outliers. In a skewed distribution, the suspect outliers are likely to be the extremes of the longer tail (see Figure 2). Multivariate outlier detection is more difficult, because the multivariate distribution has no tails [1,7]. Multivariate outliers, such as engineer M, are sometimes also univariate outliers. However, multivariate outliers are not necessarily univariate outliers, because unusual combinations of normal values may cause the case to be a multivariate outlier.

Filtering examples before the analysis seems to be a less studied area in the machine learning [8]. Brodley *et al.* eliminated outliers (mislabeled examples) from the training data by using ensemble filters, before passing the data to the final learning algorithm [6]. Wilson filtered the examples misclassified by nearest neighbour classifier ( $k = 3$ ) to another nearest neighbour classifier ( $k = 1$ ) [9]. Majority of the machine learning methods deals with irrelevant examples within the algorithm itself (embedded ap-

<sup>1</sup> Department of Computer and Information Sciences, FIN-33014 University of Tampere, Finland, email: jpl@cs.uta.fi

<sup>2</sup> Department of Otorhinolaryngology, FIN-00029 Helsinki University Central Hospital, Helsinki, Finland

proach) [e.g. 5] or apply some suitable method as a sub-routine during the learning (wrapper approach) [8].

We research machine learning methods, such as genetic algorithms and decision trees, in the context of descriptive and predictive analysis of medical data [10-12]. These methods seem to be quite robust and, therefore, they perform well with the data containing missing values and outliers [10-12]. Identification of outliers has recently begun to interest us for two reasons. Firstly, we consider balancing the imbalanced class distribution [e.g. 13] by reducing the largest classes before analysis. Outliers of the major classes seem to be worthwhile candidates for removal. In this line of work, the outliers are treated as poor data which may be removed without further analysis. Secondly, during the enlargement of the vertigo data [11,12], we noticed that the outliers may give us some additional insight of data. Outliers are not outright dropped from data, instead they are presented to expert physicians for further consideration.

In this paper, we identify both univariate and multivariate outliers with box plots [14] which are an informal method for outlier detection. The functioning of the informal method was tested in this work with two medical data sets. The results were evaluated objectively by performing discriminant analysis and nearest neighbour classification for the reduced data. Subjective evaluation was done by the expert physicians who studied the outliers manually.

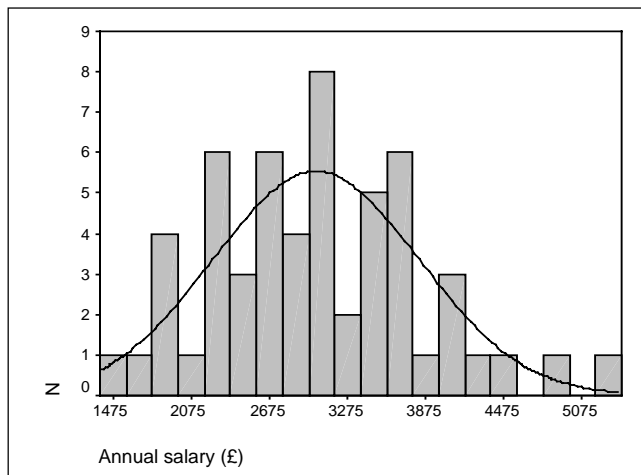


Figure 2. A histogram for annual salary

## 2 METHODS

Test of discordancy, formal or informal, is needed to declare extreme values as outliers. Formal testing requires a test statistic, which usually assumes some well-behaving distribution, on the basis of which the extremes are possibly declared outliers. Most of the test statistics, for example many Dixon-type tests, are designed to identify a single univariate outlier or an outlier pair using a normal distribution [1]. Unfortunately, the medical data sets we study are problematic in the statistical point of view. The data sets may be mixed, i.e. they contain both quantitative and qualitative variables, and the distribution of the continuous variables is frequently skewed or non-normal. Application of various test statistics would require identification of the distributions, transforma-

tions and possibly estimation of distribution parameters. For large data sets this process would be very difficult and tedious. Considering the practical aims of our research, we decided to test discordancy informally using regular box plots [14].

### 2.1 Box plot outlier identification

The box plot is a well-known simple display of the five-number summary (lower extreme, lower quartile, median, upper quartile, upper extreme) [14]. Box plots are most suitable for exploring both symmetric and skewed quantitative data, but they can also identify infrequent values from categorical data. Unlike in the quick box plot, the extremes of the box plots are not the smallest and largest data values, but the most extreme data values that are not extreme enough to be considered outliers [14]. Figure 3 shows a box plot for the salaries of the electrical engineers discussed earlier.

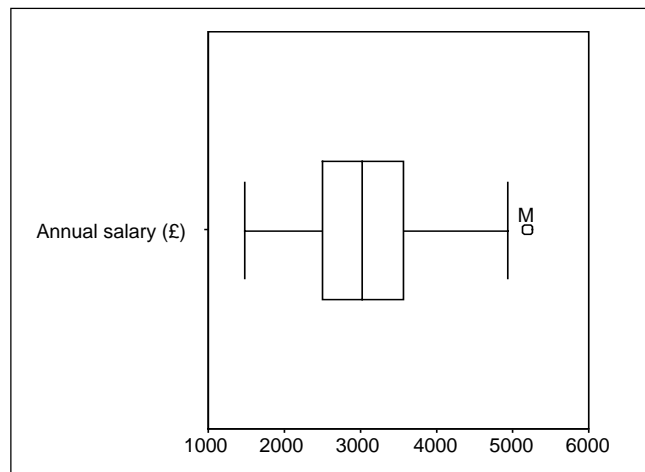


Figure 3. A box plot for annual salary

The thresholds for lower and upper outliers are defined as follows: lower threshold = lower quartile - step and upper threshold = upper quartile + step. Step is 1.5 times the interquartile range (upper quartile - lower quartile) which contains 50% of the data. Value  $x$  is a lower outlier, if  $x < \text{lower threshold}$  and an upper outlier, if  $x > \text{upper threshold}$ . Box plot identifies engineer M's salary as an upper univariate outlier (see Figure 3).

### 2.2 Univariate outliers

Univariate outliers were identified for each variable within classes. Unfortunately, when there is a large number of univariate outliers, removal of all the identified outliers may cause a large portion of data to be excluded. For this reason, we decided to rank the non-multivariate outliers according to the frequencies of univariate outlier values in these examples and to discard (or trim) examples with the highest frequencies. Since often a small fraction of the extreme values are trimmed [1,7,15], we decided to trim 10% of the worst examples within each class.

## 2.3 Multivariate outliers

Methods for identifying univariate outliers are based on unarguable order of data values. For example, in the box plot method salaries are sorted in ascending order and, on the basis of the order, extremes, quartiles and outliers can be found. There is no unambiguous total ordering for  $N$  multivariate observations, but different sub-orderings have been suggested [1,16], of which the reduced sub-ordering is the most often used in the outlier study [1].

Reduced sub-ordering is established in two phases [1,16]. Firstly, a set of scalars  $R = \{ r_i \} (i=1,\dots,N)$  is produced by transforming each multivariate observation  $\mathbf{x}_i$  into a scalar  $r_i$ . Then,  $R$  is sorted to produce the actual ordering of the multivariate data. The transformation is often done with a distance metric [16] and, therefore, the extremes are those multivariate observations associated with the largest values in  $R$ .

We used in this study sub-ordering based on the generalised distance metric [1,7,16]

$$r_i^2 = (\mathbf{x}_i - \mathbf{x}_0)' \mathbf{\Gamma}^{-1} (\mathbf{x}_i - \mathbf{x}_0), \quad (1)$$

where  $\mathbf{x}_0$  indicates the location of the data set and  $\mathbf{\Gamma}^{-1}$  weights variables inversely to their scatter. Different choices of these parameters result in different distance metrics. For example, when  $\mathbf{\Gamma}$  is the identity matrix  $\mathbf{I}$ , (1) defines the Euclidean distance of  $\mathbf{x}_i$  to the location of the data set.

We chose to use Mahalanobis distance [17,18] in the multivariate outlier identification. Mahalanobis distance is obtained from (1) by selecting  $\mathbf{\Gamma}$  to be the population covariance matrix  $\mathbf{\Sigma}$ . As usual, the population mean  $\boldsymbol{\mu}$  was used as the location parameter [1,3,7]. Often the population values are unknown and they are estimated with sample mean vector  $\mathbf{m}$  and sample covariance matrix  $\mathbf{S}$

$$r_i^2 = (\mathbf{x}_i - \mathbf{m})' \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{m}). \quad (2)$$

Mahalanobis distance incorporates the dependencies between the attributes. This property is essential in multivariate outlier identification, where the goal is to detect unusual value combinations. Many distance metrics, including Euclidean distance, utilise only location information and are, therefore, unsuitable for this task. Another advantage of Mahalanobis distance is that the unit of variable has no influence on the distance, because each variable is standardised to mean of zero and variance of one [17,18].

Gamma-type probability plots are useful for informal outlier detection with generalised distances [1 pp. 274-275,7]. These graphical displays are produced by plotting the ordered reduced univariate measures  $r_i$  against the quantiles of a gamma distribution. If the multivariate observations are from a normal distribution, then the reduced measures follow approximately the gamma distribution. As a result, the points should cluster around a straight line and points that lie clearly off the linear relationship are considered to be outliers.

However, we applied again box plots, because we cannot assume that multivariate observations come from a normal distribution. Also, the gamma probability plots require a human, who must evaluate whether the anomalous points are really outliers. Box plots use objective rules for outlier identification.

## 3 MATERIALS

Outliers were searched from two medical data sets. The female urinary incontinence data (see Table 1) was collected retrospectively in the Department of Obstetrics and Gynaecology of Kuopio University Hospital, Finland [10]. The examples are described with 16 variables of which 7 are binary and 9 quantitative. Two variables (uroflowmetry and cystoscopy) were dropped from the analysis, because they had extremely high missing value rates.

The vertigo data (see Table 2) was collected in the vestibular unit of the Helsinki University Central Hospital, Finland. The patients, referred to the vestibular laboratory, filled out a questionnaire concerning their symptoms, earlier diseases, accidents, use of medicine, tobacco and alcohol [19]. The information was stored in the patient database of the expert system ONE [19]. The diagnoses were confirmed by an experienced specialist in the field of otoneurology. In this study, we focused on the six largest patient groups with vertigo and used the 38 most important variables of all the 170 available variables [11,12]. The subset of variables consisted of 16 quantitative variables, 10 ordinal variables and 12 nominal variables, 11 of which were dichotomous.

The missing values were replaced in both data sets with modes (nominal variables), medians (ordinal variables) and means within diagnostic classes. The imputed values of discrete variables were rounded.

## 4 EXPERIMENTAL SETUP

Multivariate and univariate outliers were identified separately with box plots by each diagnostic class, as usual [3]. Nearest neighbour classification ( $k = 1$ ) with the heterogeneous value difference metric [20] and discriminant analysis were used for the objective evaluation. These methods were selected, because they are classical methods for classification in the areas of machine learning and statistics, respectively.

Four versions of both the medical data sets were evaluated: the original data  $A_0$  ( $|A_0| = N$ ) and three reduced data sets  $A_1$ ,  $A_2$  and  $A_3$ . Reduced sets were created by excluding 1) the multivariate outliers ( $|A_1| = N - N_m$ ), 2) the multivariate outliers and 10% of the non-multivariate outlier examples with the most univariate outlier values ( $|A_2| = N - N_m - N_u$ ) and 3) a random sample from the original data. The random removal was a baseline method where the number of removed examples was the same as the number of outliers excluded from the original data to produce the data set  $A_2$ .

Outlier removal was studied from two viewpoints. Firstly, we considered the descriptive analysis where statistical or machine learning methods are used to model the data. The goal is to find a reasonably simple model which fits the data well. This was studied by classifying data sets  $A_i$  with nearest neighbour method and discriminant analysis. Secondly, we studied the predictive analysis in which the aim is to build a model for accurate classification of new data which contains outliers. To assess the predictive ability of the models, the original data sets  $A_0$  were classified with the nearest neighbour method, using the reduced data sets  $A_1$ ,  $A_2$  and  $A_3$  as classifiers, and with the discriminant functions obtained from the reduced data sets.

The effect of removing the outliers was measured with the classification accuracy of the nearest neighbour classifiers and discriminant analysis functions. Classification accuracy  $ACC$  in per cents is  $ACC = 100\% N_c / N$ , where  $N_c$  is the number of cor-

rectly classified examples and  $N$  is the number of all examples in the data set. Nearest neighbour classification was performed with 10-fold cross-validation which was repeated 3 times, while discriminant analysis was not cross-validated. Therefore, the accuracies of nearest neighbour method were more realistic than those of the discriminant analysis.

## 5 RESULTS

Tables 1 and 2 show the frequencies of the multivariate outliers ( $N_m$ ) and examples with the most univariate outliers ( $N_u$ ) identified from the female urinary incontinence and vertigo data sets, respectively. The tables also show the sizes of diagnostic groups and, for the comparison with the number of outliers, the absolute frequencies corresponding to the 10% portion of each diagnostic group ( $N_{10\%}$ ). The outlier frequencies behaved as expected. The largest classes had the highest number of outliers and the overall number of outliers was reasonably small in both the data sets.

**Table 1.** Frequencies of the original female urinary incontinence data ( $N$ ) and its outliers by the diagnostic classes ( $N_m$  = multivariate outliers,  $N_u$  = examples with the most univariate outlier values).

Diagnosis	Original data		Outliers		Sum
	$N$	$N_{10\%}$	$N_m$	$N_u$	
Stress	323	32	29	20	49
Mixed	140	14	9	9	18
Sensory urge	33	3	2	2	4
Motor urge	15	2	0	1	1
Normal	18	2	0	1	1
Sum	529	53	40	33	73

**Table 2.** Frequencies of the original vertigo data ( $N$ ) and its outliers by the diagnostic classes ( $N_m$  = multivariate outliers,  $N_u$  = examples with the most univariate outlier values).

Diagnosis	Original data		Outliers		Sum
	$N$	$N_{10\%}$	$N_m$	$N_u$	
Vestibular schwannoma	128	13	5	9	14
Benign positional vertigo	59	6	2	4	6
Meniere's disease	243	24	13	14	27
Sudden deafness	21	2	1	1	2
Traumatic vertigo	53	5	1	4	5
Vestibular neuritis	60	6	0	4	4
Sum	564	56	22	36	58

The descriptive accuracies of nearest neighbour method and the discriminant analysis functions are reported in Table 3. There was a clear improvement in the classification ability of both methods in the female urinary incontinence data. Also, removal of a randomly selected sample produced less accurate results, than excluding the identified outliers. The removal of outliers helped the classification of the vertigo data only slightly.

**Table 3.** Descriptive accuracies of the nearest neighbour method (NN) and discriminant analysis functions (DA) in different versions of the data.

Version of data	Accuracy (%)			
	Incontinence		Vertigo	
	NN	DA	NN	DA
Original ( $A_0$ )	83.5	84.3	90.9	94.3
No multivariate outliers ( $A_1$ )	85.9	86.7	92.7	94.6
No multi- and univariate outliers ( $A_2$ )	87.6	89.5	91.2	94.3
Random sample removed ( $A_3$ )	82.1	84.2	90.9	95.1

Table 4 shows the prediction accuracies of the two methods. Reduced data set  $A_2$ , from which the multivariate outliers and the examples with the most univariate outlier values were removed, was the worst nearest neighbour classifier. Also, discriminant functions produced from this data classified the original data with the lowest prediction accuracy.

**Table 4.** Prediction accuracies of the nearest neighbour method (NN) and discriminant analysis functions (DA) in the classification of original data.

Version of data	Accuracy (%)			
	Incontinence		Vertigo	
	NN	DA	NN	DA
No multivariate outliers ( $A_1$ )	96.2	84.3	98.9	94.0
No multi- and univariate outliers ( $A_2$ )	93.6	83.4	97.9	92.6
Random sample removed ( $A_3$ )	98.3	83.9	98.6	94.9

## 6 DISCUSSION

Outlier identification was studied with informal box plot method using as the test material two real-world data sets. There were two motivations for the identification. Firstly, outliers can be considered to be suspicious data whose removal before applying inductive machine learning methods is reasonable. Especially, dropping of the outliers of the largest classes balances the class distribution and, consequently, makes the classification of the members of the smaller classes easier [13]. Secondly, the knowledge carried by the outliers may be valuable for the domain experts, who may gain additional insight into the data by examining them.

The suitability of the method for the straightforward data reduction was studied objectively with discriminant analysis and nearest neighbour method which are well-known classification methods in statistics and machine learning. Removal of the identified outliers from the female urinary incontinence data improved clearly the classification ability of discriminant analysis functions and nearest neighbour method (see Table 3). The descriptive accuracy of the discriminant functions improved from 84.3% to 89.5% when both multivariate outliers and examples with the largest numbers of univariate outliers (14% of the data) were removed from the original data set. The prediction accuracy of the nearest neighbour classifier raised from 83.5% to 87.6%. However, the improvement was only marginal in the vertigo data set.

The most probable explanation for the differences is the characteristics of the data sets. All the female urinary incontinence data was collected retrospectively from the patient records [10], while the main body of vertigo data was obtained carefully in prospective fashion [19]. In addition, the vertiginous patients used in this study were selected to meet the definitions of the six diagnostic classes [19]. As a result, there were not much improvement left in the classification of this data. The earlier machine learning experiments [11,12] lend additional support for this conclusion. The best results were obtained with the decision trees [12] whose classification accuracy ranged from 94% (Meniere's disease) to 100% (benign positional vertigo and vestibular neuritis).

The prediction accuracy behaved oppositely to descriptive accuracy. Exclusion of outliers lowered the prediction accuracies. Again, the effect was stronger in the female urinary incontinence data than in the vertigo data. Nearest neighbour classification suffered from the outlier exclusion clearly in the incontinence data. This result suggest that the removed data were indeed outliers. Nearest neighbour classification should be more difficult, when

data having unusual examples is classified with cleaned data from which these examples have been removed. In addition, Quinlan [4 pp. 96] has shown experimentally that, for higher levels of attribute noise, a decision tree built from cleaned data is inferior to a tree built from the noisy data, when the data to be classified has same noise level. Discriminant analysis is in comparison with nearest neighbour method a highly advanced tool which produces more robust classifiers. For this reason, the prediction accuracies of the discriminant functions remained close to the descriptive accuracies of the original data set as in [15].

The identified univariate and multivariate outliers were presented to the expert physicians, who evaluated the suspect data to decide whether it was truly outlying. In their opinion, most of the multivariate outliers were abnormal cases. For example, in one case the post voiding residuals and urgency score were abnormally high for a stress-incontinent woman. However, closer examination revealed that the diagnosis was correct, because she had to drink excessively due to bowel problems. The experts sometimes disagreed with the box plot method on univariate outliers. The most frequent reason was the natural variation in diagnostic parameters between patients. The upper or lower outliers were extreme, but yet reasonable values for a parameter in a particular diagnostic class. This happened, especially in heterogeneous diagnostic groups, where also extreme values are natural. Vestibular schwannoma and Meniere's disease are examples of the heterogeneous diagnostic groups [19]. Both diseases worsen during the time and the extreme values come often from the patients who have had these diseases for a long time [19].

The experimental results suggest that box plots can be used for data reduction, but the benefit obtained of excluding outliers is data set dependent. Outlier removal helps the descriptive analysis, but the predictive analysis, i.e. classification of unseen cases, may suffer. Therefore, the usefulness of the method depends also the final goal of the analysis. The behaviour of predictive analysis needs to be studied further, because machine learning applications are often used to classify new data. The subjective evaluation by the experts gave controversial, but sound, results. There were real abnormalities in the multivariate outliers and the disagreement on the univariate outliers resulted from the method, which does not utilise any prior knowledge in the outlier identification.

The major limitation of this work is the use of the informal box plot method. Discordancy test statistics identify outliers on the basis of the solid theory. Unfortunately, these methods make many assumptions which should be met in order the tests to be applicable. Therefore, one can also argue that a well-known and widely used informal method may be a more appropriate choice for large practical applications, where these assumption are not usually fulfilled. However, the future work should also address the formal test statistics. There is also two limitation in the multivariate outlier identification with the Mahalanobis distance. Firstly, the Mahalanobis distance works best with quantitative normally distributed data [18]. Secondly, missing values must be treated before distance computation. Possibly some type of heterogeneous distance function [20] could address these problems.

## REFERENCES

- [1] V. Barnett and T. Lewis, *Outliers in Statistical Data*, John Wiley & Sons, Norwich, 2nd edn., 1987.
- [2] R.J. Beckman and R.D Cook, 'Outliers', *Technometrics*, **25**, 119-149, (1983).
- [3] B.G. Tabachnick and L.S. Fidell, *Using Multivariate Statistics*, HarperCollins, New York, 1996.
- [4] J.R. Quinlan, 'Induction of decision trees', *Machine Learning*, **1**, 81-106, (1986).
- [5] G.H. John, *Robust Decision Trees: Removing Outliers from Databases*, 174-179, Proceedings of the First International Conference on Knowledge Discovery and Data Mining, AAAI Press, 1995.
- [6] C.E. Brodley and M.A. Friedl, *Identifying and Eliminating Mislabelled Training Instances*, 799-805, Proceedings of the Thirteenth National Conference on Artificial Intelligence, AAAI Press, Menlo Park, 1996.
- [7] R. Gnanadesikan and J.R. Kettenring, 'Robust estimates, residuals, and outlier detection with multiresponse data', *Biometrics*, **28**, 81-124, (1972).
- [8] A.L. Blum and P. Langley, 'Selection of relevant features and examples in machine learning', *Artificial Intelligence*, **97**, 245-271, (1997).
- [9] D.L. Wilson, 'Asymptotic properties of nearest neighbor rules using edited data', *IEEE Transactions on Systems, Man, and Cybernetics*, **2**, 408-421, (1972).
- [10] J. Laurikkala and M. Juhola, 'A genetic-based machine learning system to discover the diagnostic rules for female urinary incontinence', *Computer Programs in Biomedicine*, **55**, 217-228, (1998).
- [11] E. Kentala, J. Laurikkala, I. Pyykkö, and M. Juhola, 'Discovering diagnostic rules from a neurotologic database with genetic algorithms', *Annals of Otolaryngology & Laryngology*, **108**, 948-954, (1999).
- [12] E. Kentala, K. Viikki, I. Pyykkö, and M. Juhola, 'Production of diagnostic rules from a neurotologic database with decision trees', *Annals of Otolaryngology & Laryngology*, **109**, 170-176, (2000).
- [13] M. Kubat and S. Matwin, *Addressing the Curse of Imbalanced Training Sets: One-sided Selection*, 179-186, Proceedings of the Fourteenth International Conference on Machine Learning, Morgan Kaufmann, San Francisco, 1997.
- [14] A.F. Siegel, *Statistics and Data Analysis: An Introduction*, Wiley, New York, 1988.
- [15] E. Krusinska and J. Liebhart, 'The influence of outliers on discrimination of chronic obstructive lung disease', *Methods of Information in Medicine*, **27**, 167-176, (1988).
- [16] V. Barnett, 'The ordering of multivariate data (with Discussion)', *Journal of the Royal Statistical Society A*, **139**, 318-354, (1976).
- [17] A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data*, Prentice Hall, New Jersey, 1988.
- [18] J. Boberg, *Cluster Analysis: A Mathematical Approach with Applications to Protein Structures*, Academic dissertation, Turku Centre for Computer Science, Turku, Finland, 1999.
- [19] E. Kentala, *A Neurotologic Expert System for Vertigo and Characteristics of Six Otologic Diseases Involving Vertigo*, MD thesis, University of Helsinki, Helsinki, Finland, 1996.
- [20] D.R. Wilson and T.R. Martinez, 'Improved heterogeneous distance functions', *Journal of Artificial Intelligence Research*, **6**, 1-34, (1997).