Evaluation of Multivariate Outlier Detection Methods with Benchmark Medical Datasets

Zahra Nazari† and Dongshik Kang††

†Graduate School of Engineering & Science, University of the Ryukyus, Okinawa, 903-0213 Japan
††Department of Information Engineering, University of the Ryukyus, Okinawa, 903-0213 Japan

Summary
Outliers are unusual data points which are inconsistent with other observations in a dataset. Outlier detection method has been researched in diverse application domains and recently it has been realized that there is a direct mapping between outliers in data and real world anomalies. The importance of outlier detection is due to the fact that outliers in data sometimes interpret to significant information in a wide variety of application domains (Chandola et al. 2007). Several types of outlier detection methods are developed and a number of surveys and reviews are performed to distinguish their advantages and disadvantages. Outlier detection methods are highly domain oriented; therefore an evaluation is needed to find an appropriate one for the intended domain. In this study we evaluate widely used multivariate outlier detection methods namely distance based, statistical based and clustering based for medical datasets. Five benchmark medical datasets of Heart disease, Breast Cancer Pima Indian Diabetes, Liver Disorders and Thyroid Gland are used for experiments. To identify the effectiveness of mentioned outlier detection methods, the above datasets are classified and their total variances are calculated before and after outlier detection. Eight well-known individual and ensemble classifiers are used for data classification. Finally a comparative review is performed to distinguish the advantages and disadvantages of each method and their respective effects on accuracy of classifiers.

Key words:
Outlier Detection, Data Mining, Machine Learning, Data Clustering, Pattern Recognition

1. Introduction

Real world data are not as good as we would like them to be; often they are corrupted or misplaced in the wrong category that may affect data interpretations, data processing, classifiers and model generated from data as well as decisions which are made based on data. The corruption can be due to several factors, such as measurement error, coding/recording error, ignorance and human error, rounding error, transcription error, inherent variability of the domain, instrument malfunction and biases. Due to existence of unusual data in most datasets it is necessary to develop techniques that allow us to identify them. The aim of an outlier detection method sometimes called novelty detection method is to find patterns in data that do not conform to expected behavior. Outlier detection methods have extensive use in a large number of applications such as military surveillance for enemy activities, intrusion detection in cyber security, fraud detection for credit cards, insurance or health care and fault detection in safety critical systems, etc. [1-4]. There are many types of outlier detection methods which are used for different application domains and a number of studies are performed either to compare or provide a coarse classification of these methods. Escalante has provided a comparison of outlier detection methods applied on benchmark data with inserted artificial noise and outliers [1]. Chendola et al. have performed a comprehensive overview of existing outlier detection methods by classifying them along different dimensions [2]. A survey of outlier detection methodologies to distinguish their advantages and disadvantages is done by Hodge and Austin [3]. Ben-Gal has done a survey to distinguish between univariate vs. multivariate outlier detection techniques [4]. Another short review of outlier detection methods using data mining techniques is presented by Petrovskiy [5].

Frequently, it has been proved that presence of outliers is one of factors that affect accuracy of the most classifiers. Hence, in this study we evaluate the effectiveness of multivariate outlier detection methods on performance of eight classifiers applied on medical datasets. Distance based, Statistical based and Clustering based outlier detection methods are used for outlier detection. Linear, Quadratic and Gaussian Support Vector Machines (SVM), Decision Tree Classifier, Linear Discriminant Analysis (LDA), and K-Nearest Neighbor (KNN) as individual classifiers, Bagged Tree and Subspace KNN as ensemble classifiers are used before and after outlier detection. Five benchmark medical datasets of Heart disease, Breast cancer, Pima Indian diabetes, Liver disorder and Thyroid gland from UCI repository are used for the experiments [6]. To realize how well outliers are detected by each outlier detection methods, we calculated variance of each datasets

Manuscript received April 5
Manuscript revised April 20

before and after outlier detection. The generalization ability (accuracy) of each classifier before and after outlier detection is also compared to evaluate the effectiveness of outlier detection methods which are used in this study.

This paper is constructed as follows. In section 2, we present the definitions of outlier and outlier detection. In section 3, we briefly review and distinguish the multivariate outlier detection methods under evaluation. In section 4, we provide our detailed evaluation results. Finally in section 5 we conclude the paper with discussion and future research directions.

2. Preliminaries

2.1 Outlier

Outlier in earlier times were regarded as noisy data, but in recent years has turned out to be an important problem which being researched in various fields of research and application domains. Conceptually, an outlier is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism [5]. Outlier is an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data [7]. An outlier may also be 'surprising veridical data', a point belonging to class A but actually situated inside class B so the true (veridical) classification of the point is surprising to the observer [8]. A further outlier definition from Aggarwal and Yu is: outliers may be considered as noise points lying outside a set of defined clusters or alternatively outliers may be defined as the points that lie outside of the set of clusters but are also separated from the noise [9]. These outliers behave differently from the norm. Fig. 1 illustrates outliers in a simple two dimensional dataset. The data has two normal regions, R_1 and R_2, since most observations lie in these two regions. There are some points O_1, O_2, and O_3 that are sufficiently far away from the regions, hence they are called outliers.

2.1 Outlier Detection

Detecting outliers is an important data mining task. Outlier detection is to find patterns in data that do not conform to expected behaviors and it is one of the first steps towards obtaining a coherent analysis. Detected outliers are candidates for aberrant data that may otherwise adversely lead to model misspecification and incorrect results. Outlier detection techniques have an extensive use in wide variety of areas such as machine learning, data mining, data cleansing, data warehousing, pattern recognition and applications as: military surveillance for enemy activities, intrusion detection in cyber security, fraud detection for credit cards, insurance or health care and fault detection in safety critical systems, etc. [10-12].

There are many types of outlier detection methods and researchers according to different criteria have divided them to many categories. The taxonomy of *Univariate* and *Multivariate* methods is proposed in earlier works in this field. The *Parametric* (*Statistical*) and *Non-parametric* methods are another fundamental taxonomy of outlier detection methods. Within the class of non-parametric methods, *distance based methods* are different. These methods are based local distance measures and often are suitable for large datasets [4, 11]. Another category of outlier detection methods is founded on *clustering techniques*, where a cluster of small size can be considered as clustered outliers. Another related class of methods consists of detection techniques for *spatial outliers*. These methods search for extreme observations or local instabilities with respect to neighboring values, although these observations may not be significantly different from the entire population. Almost all the studies that consider outlier identification as their primary objective are in the field of statistics. Outlier detection for data mining is often based on distance measures, clustering and spatial methods. Other categorizations of outlier detection methods can be found in Barnett and Lewis (1994), Acuna and Rodriguez (2004), and Hu and Sung (2003) [13, 14].

In this study we have considered the taxonomy of univariate and multivariate and our focus is on multivariate outlier detection methods.

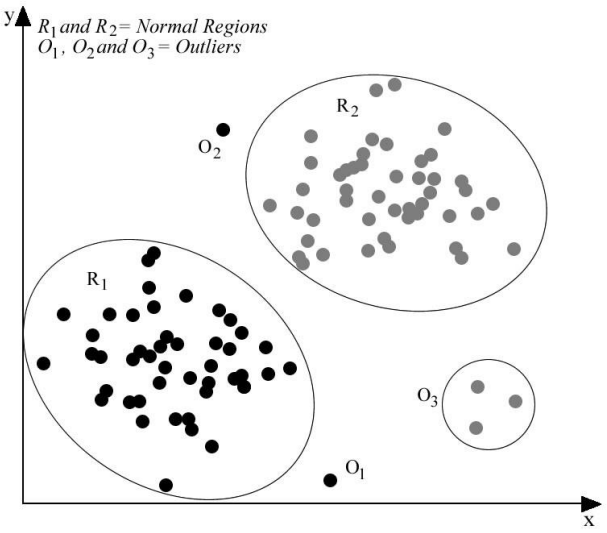

Fig. 1 A simple example of outliers in a 2-Dimensional dataset.

3. Multivariate Outlier Detection Methods

Let us consider a set of N instances with d features, objective is to detect all the instances that seem to be unusual; they will be called multivariate outliers. One might think that these outliers can be detected based on the univariate outliers on each feature, but in many cases an unusual multivariate instance cannot be detected as outlier when each variable is considered separately. Detecting outliers is possible only when multivariate analysis is performed and the interactions among variables are compared with the class of data. In the other words, an instance can be a multivariate outlier but a usual data in each feature and an instance can have values that are outliers in several features but the whole instance might be a usual multivariate data. There are many types of multivariate outlier detection methods. Distance-based outlier detection, statistical-based and clustering-based outlier detection methods are evaluated in this paper [4].

3.1 Distance Based Outlier Detection Method

This method removes outliers from the training data to purify the data and improves the performance of prediction remarkably. Three measures are proposed to describe the extent to which an instance is likely to be an outlier compared to others. The first measure is used to describe the distance of an instance from the center vector of all instances with the same class label. The second measure is the probability of the class label (PCL) of the instance. The third measure describes the importance of within-class and between-class (IWB) instances. By integrating the three measures outlier score can be computed. Instances with large score will be treated as outliers and subsequently should be removed from the training data [15]. The three measures integrated for outlier detection are as follows:

- K-Distance of an Instance: The K-distance of an instance x is the mean distance between the instance x and its K (positive integer) nearest neighbors. It describes how far the K nearest neighbors is away from x on average.

$$K_{\text{dist}}(x) = \frac{1}{K}\sum_{m=1}^{K} d\,(x,x_m) \qquad (1)$$

  where $d(x, x_m)$ is the Euclidean distance measurement, and x_m is one of the K nearest neighbors of x.

- Probability of the Class Label (PCL) of an Instance: PCL is related to the probability of the class label of an instance in terms of its K nearest neighbors. For example the PCL of the instance x (the black one within the circle in Fig. 2), denoted by PCL(x), is defined as the ratio of the number of instances with class label 1 to the total number of instances in the circle in terms of its K nearest neighbors, including the instance itself. Therefore, PCL (x) =1/4 if K=4 for instance x in Fig 2.

- Importance of Within-Class and Between-Class (IWB): IWB measures the importance of within-class and between-class changes for an instance. The IWB (x) is defined as the change in the ratio of between-class scatter S_b to within-class scatter S_w before and after excluding instance x. In the two class cases S_b stands for the subtraction of the mean values of the classes from each other. In contrast, S_w denotes the summation of the two scatters calculated within the same class. In general, a within class scatter is equivalent to the variance in the same class computed as:

$$S_j = \left(\sum_{x_1 \in C_j} (x_1 - m_1)^2\right)^{1/2} \qquad (2)$$

  where j=1 or 2, and C_j is the instance set of class j. In particular, $\tilde{S}_b$ and $\tilde{S}_w$ denote between-class scatter and within-class scatter after excluding the instance x, respectively.

$$\text{IWB}(x) = \frac{S_b}{S_w} - \frac{\tilde{S}_b}{\tilde{S}_w} = \frac{|m_1 - m_2|^2}{S_1 + S_2} - \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{S}_1 + \tilde{S}_2}$$

$$= \frac{|m_1 - m_2|^2}{\left(\sum_{x_1 \in C_1}(x_1 - m_1)^2\right)^{1/2} + \left(\sum_{x_2 \in C_2}(x_2 - m_2)^2\right)^{1/2}} \qquad (3)$$

$$- \frac{|\tilde{m}_1 - \tilde{\text{m}}_2|^2}{\left(\sum_{x_1 \in C_1}(x_1 - \tilde{m}_1)^2\right)^{1/2} + \left(\sum_{x_2 \in C_2}(x_2 - \tilde{m}_2)^2\right)^{1/2}}$$

  where m_1 and m_2 are sample means for particular classes, and $\tilde{m}_1$ and $\tilde{m}_2$ are sample means for the two corresponding classes excluding the instance x.

- Class Outlier Score (COS): The class outlier score of an instance x stands for the degree of an instance being outlier with respect to a particular class [15].

$$\text{COS}(x) = \alpha * PCL(x) + \beta * K_{dist}(x) + \gamma * IWB(x) \qquad (4)$$

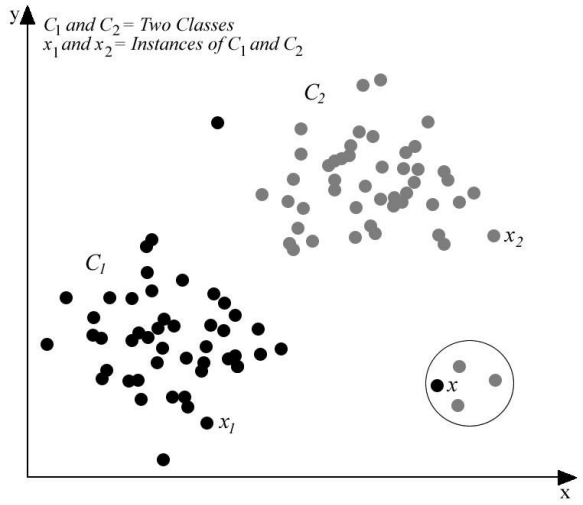

Fig. 2 An example of detecting class outliers.

where α, β and γ are parameters to trade off the probability of class label, K-distance, and IWB, respectively. In this work α, β and γ are normalized in the range [-1, 1]. The outlier detection was performed by a grid search. An instance x is assigned as an outlier if $\mathrm{COS}(x) \geq c$, where c is a threshold according to experimental results. Therefore the larger the values of the three measures, the more likely the instance x could be an outlier. Instances with the top scores are treated as outliers, where the exact number of outliers depends on a specific dataset [15].

3.2 Statistical Based Outlier Detection

The basic concept which is used in this method is Population Density. The concept of population density is very close to population distribution but is different, as distribution is based on location while density is a ratio. Population density is the ratio of people to physical space. It is the way of measuring the population per unit of area or volume. In other words, the concept of population density indicates the relationship between number of population and the occupied area by them. By using this concept the densely populated and less populated areas can be determined. Considering the training points as the population, those points placed in areas with low population densities can be treated as outliers, illustrated in Fig. 1 [16].

The Mahalanobis distance is the distance from x_i to the quantity μ. This distance is based on the correlation between variables or the variance-covariance matrix. Mahalanobis distance measure includes the inter-attribute dependencies so the system can compare attribute combination [2]. This measure is unit less and it takes into accounts the correlation of the dataset and does not depend on the scale of measurement [17]. Mahalanobis distance is computationally expensive to calculate for large high dimensional data sets compared to the Euclidean distance as it requires a pass through the entire data set to identify the attribute correlations [2]. The mahalanobis distance of point x_i to the mean of distribution can be calculated by Eq. 5. The first step of this method is to calculate the mahalanobis distance between each data points and the quantity μ; afterwards the average distance (AD) Eq. 6. should be calculated and each data point with $D_i \geq (\mathrm{AD})^2$ goes to the group of outliers.

$$D_m(x) = \sqrt{(x-\mu)^T S^{-1}(x-\mu)} \tag{5}$$

$$\mathrm{AD} = \frac{\sum_{i=1}^{N} \sqrt{(x_i-\mu)^T S^{-1}(x_i-\mu)}}{n} \tag{6}$$

$$\text{if } D_i > (AD)^2 \rightarrow x_i = \text{outlier}$$

3.3 Clustering Based Outlier Detection

A general definition of clustering is organizing a group of objects that share similar characteristics. Clustering is a widely used method in different areas such as business and retails, computational biology, social media network analysis, etc. Hierarchical, partitioning, grid based and density based are different types of clustering methods that each of them uses a different induction principle. Hierarchical and partitioning are widely used methods and we briefly explain these two in the following. The hierarchical method is a well-known clustering method that can be thought of as a set of flat clustering methods organized in a tree structure. This method constructs the clusters by recursively partitioning the data in either a top-down or bottom-up fashion. This clustering method is subdivided to Agglomerative hierarchical clustering and Divisive hierarchical clustering [18].

Divisive clustering which is introduced by Kaufmann and Rousseeuw (1990) is a top-down approach that all observations initially belongs to a single root cluster and iteratively partitions existing cluster into sub-clusters. The agglomerative clustering is a bottom-up approach and in this method each instance initially represents a cluster of its own, and then similar clusters are iteratively merged until the desired cluster structure is obtained. This method for N instances begins with N clusters and each cluster contains a single instance. Minimum distance, maximum distance, average and center distance are the criteria which are used to calculate the similarity between clusters.

Among hierarchical methods, bottom-up approaches tend to be more accurate, but have higher computational cost than top-down approaches [18, 19].

Embedded flexibility with regard to the level of granularity, ease of handling any forms of similarity and distance and applicability to any attributes type are some properties of hierarchical clustering method which make it great interest for a number of application domains and to produce better quality clusters. In hierarchical clustering, one can represent the clusters of data as a tree called a dendrogram, in which the nodes represent subsets of the input dataset. The leaf nodes correspond to the individual elements of the dataset, and the root corresponds to the entire dataset. Each edge in the dendrogram represents an inclusion relationship (Fig. 3) [18, 19].

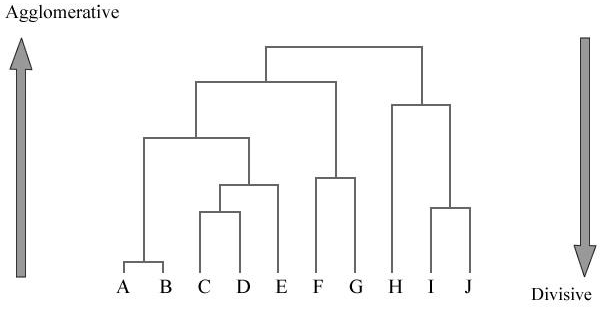

Fig. 3 An example of hierarchical clustering illustrating agglomerative and divisive methods. A - J are observations.

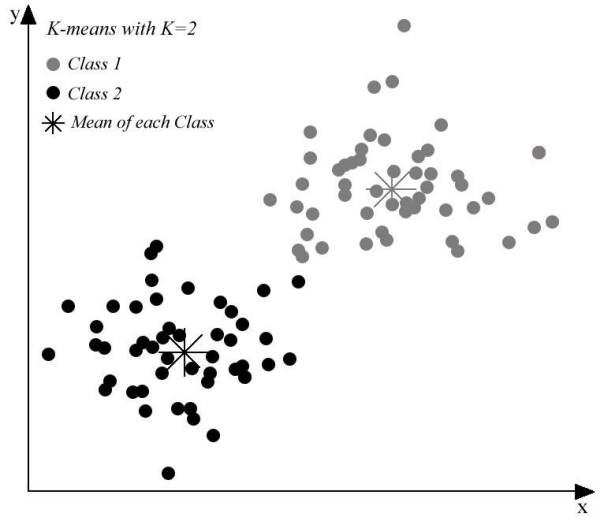

Fig. 4 An example of partitioning clustering (K-means).

Partitioning clustering is the simplest form of clustering which splits a dataset into K (an arbitrary number) partitions, where each partition represents a cluster. Unlike the hierarchical method, partitioning method creates one-level partitions of data; if K is the desired number of clusters, then partitioning methods find all K clusters at once. The famous K-means, Bisecting K-means method, PAM (Partitioning Around Medoids), CLARA and the Probabilistic Clustering are different types of partitioning clustering method [19]. Fig. 4 illustrates K-means clustering with K=2.

4. Experiments and Results

In this section the experimental results of above described outlier detection methods under evaluation are presented. Five benchmark medical datasets of Heart Disease, Breast Cancer, Pima Indian Diabetes, Liver Disorder and Thyroid Gland from the UCI repository are used. Besides distance based and statistical based outlier detection methods, agglomerative hierarchical clustering is also used to detect outliers in data. In this method observations in small clusters are tend to be dissimilar to most others; thus they will be considered as outliers and should be discarded from data. In the following, Fig. 5 shows the agglomerative hierarchical clustering result of Heart disease dataset.

The following dendrogram represents the similarities among the instances of Heart disease dataset. Each branch of a dendrogram is called a clade and the arrangement of clades shows which leaves are most similar to each other. The branch's height indicates how similar or different they are from each other; the greater the height, the greater the

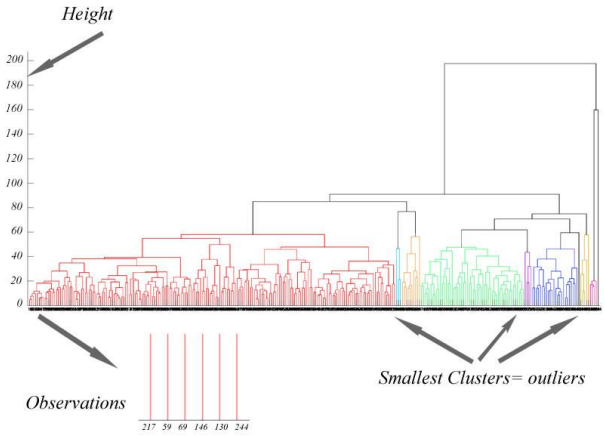

Fig. 5 The result of agglomerative hierarchical clustering applied on Heart disease dataset. Clusters are shown by different colors. The vertical axis shows height of branches and horizontal axis shows observations of Heart disease dataset.

difference [20]. Those clusters pointed out with arrows are smallest clusters containing most dissimilar instances with rest of data which are considered as outliers.

To confirm the advantages of outlier detection, we classified data two times, before and after applying outlier detection methods. All experiments are done in MATLAB environment in 10-folds cross validation form. Datasets characteristics are listed in Table 1. The following tables (Table 2-6) show the classification results of Heart Disease, Breast Cancer, Pima Indian Diabetes, Liver Disorder and Thyroid Gland datasets before and after outlier detection. It should be noted that in this study, heuristic search is used to determine values for K in KNN, α, β and γ as well as number of outliers in distance based and clustering based outlier detection methods.

Table 1: Datasets characteristics.

Dataset Name	Number of Instances	Number of Attributes	Number of Categories
Heart Disease	303	14	5
Breast Cancer	683	11	2
Pima Indian Diabetes	768	9	3
Liver Disorder	345	7	2
Thyroid Gland	215	6	3

Table 2: Classification results of Heart Disease dataset before and after outlier detection.

Classifiers	Before Outlier Detection	After Outlier Detection by		
		Statistical Based	Distance Based	Clustering Based
Linear SVM	60%	**63%**	56%	60%
Quadratic SVM	58%	**61%**	53%	57%
Gaussian SVM	61%	**63%**	55%	57%
Decision Tree	55%	**58%**	51%	55%
KNN	58%	**63%**	55%	59%
LDA	57%	**59%**	58%	58%
Bagged Tree	**58%**	**58%**	**58%**	56%
Subspace KNN	58%	**60%**	57%	57%

Table 3: Classification results of Breast Cancer dataset before and after outlier detection.

Classifiers	Before Outlier Detection	After Outlier Detection by		
		Statistical Based	Distance Based	Clustering Based
Linear SVM	96%	**98%**	97%	97%
Quadratic SVM	95%	**98%**	97%	97%
Gaussian SVM	**97%**	**97%**	**97%**	**97%**
Decision Tree	**96%**	95%	95%	94%
KNN	97%	97%	97%	97%
LDA	95%	**97%**	96%	96%
Bagged Tree	**98%**	**98%**	97%	97%
Subspace KNN	**98%**	**98%**	97%	97%

Table 4: Classification results of Indian Pima Diabetes dataset before and after outlier detection.

Classifiers	Before Outlier Detection	After Outlier Detection by		
		Statistical Based	Distance Based	Clustering Based
Linear SVM	**78%**	**78%**	77%	77%
Quadratic SVM	**76%**	**76%**	75%	**76%**
Gaussian SVM	76%	76%	76%	76%
Decision Tree	73%	73%	**75%**	73%
KNN	71%	**74%**	**74%**	**74%**
LDA	72%	**75%**	**75%**	**75%**
Bagged Tree	**76%**	**76%**	75%	75%
Subspace KNN	71%	**74%**	72%	73%

Table 5: Classification results of Liver Disorder dataset before and after outlier detection.

Classifiers	Before Outlier Detection	After Outlier Detection by		
		Statistical Based	Distance Based	Clustering Based
Linear SVM	69%	69%	**71%**	**71%**
Quadratic SVM	71%	70%	71%	**72%**

Gaussian SVM	69%	69%	69%	**70%**
Decision Tree	66%	**67%**	65%	65%
KNN	60%	**64%**	63%	60%
LDA	62%	**64%**	62%	61%
Bagged Tree	70%	70%	**72%**	**72%**
Subspace KNN	65%	**67%**	61%	65%

Table 6: Classification results of Thyroid Gland dataset before and after outlier detection.

Classifiers	Before Outlier Detection	After Outlier Detection by		
		Statistical Based	Distance Based	Clustering Based
Linear SVM	96%	96%	**98%**	**98%**
Quadratic SVM	96%	96%	95%	96%
Gaussian SVM	96%	**98%**	96%	96%
Decision Tree	**92%**	90%	90%	91%
KNN	90%	**92%**	90%	90%
LDA	90%	**91%**	**91%**	**91%**
Bagged Tree	**97%**	**97%**	96%	95%
Subspace KNN	97%	**98%**	**98%**	**98%**

In the above tables (Table 2-6) the difference between accuracies of each classifier before and after outlier detection are presented in detail. The below tables (Table 7 and 8) respectively show the variance values of each dataset which are also calculated before and after applying outlier detection methods and the detected outliers from each dataset.

Table 7: Total variance values of each dataset before and after outlier detection.

Dataset	Before Outlier Detection	After Outlier Detection by		
		Statistical Based	Distance Based	Clustering Based
Heart Disease	6.5126	**6.2920**	6.5111	6.4521
Breast Cancer	0.9705	**0.9279**	0.9692	0.9433
Pima Indian Diabetes	0.4086	**0.3899**	0.4077	0.3956
Liver Disorder	0.2336	**0.1953**	0.1980	0.2322
Thyroid Gland	0.3087	**0.2043**	0.2357	0.2474

Table 8: All outliers detected by described methods. Underlined numbers show outliers which are detected by more than one method.

Detected outliers in each dataset	
Heart Disease	
Statistical Based	10, 46, 70, 92, 124, , 178, 182, 185, 242, 259, 292
Distance Based	10, 13, 21, 41, 42, 49, 59, 84, 91, 113, 118, 126, 152
Clustering Based	21, 42, 84, 91, 113, 126, 146, 152, 182, 185, 209, 242, 292
Breast Cancer	
Statistical Based	64, 70, 84, 128, 145, 159, 162, 420, 480, 674
Distance Based	64, 97, 162, 257, 294, 344, 574
Clustering Based	52, 64, 70, 84, 97, 128, 162, 348, 556
Pima Indian Diabetes	
Statistical Based	14, 229, 248, 372, 446, 454, 580, 707
Distance Based	5, 229, 358, 372, 446
Clustering Based	5, 14, 229, 248, 358, 372, 454, 580, 446
Liver Disorder	
Statistical Based	36, 85, 190, 233, 300, 316, 317, 323
Distance Based	192, 300, 301, 308, 316, 323, 331, 337, 342
Clustering Based	36, 85, 115, 190, 233, 300, 316, 323, 331, 342
Thyroid Gland	
Statistical Based	14, 50, 111, 179, 189, 195
Distance Based	156, 167, 195, 196, 199, 208
Clustering Based	156, 167, 170, 179, 189, 190, 195, 204, 208

Considering all the results presented in the above tables, the accuracy of classifiers after outlier detection by statistical based method is mostly higher than other methods. The variance values calculated after applying this method are also smaller than others. Hence we can claim this method detects those observations that are dissimilar to other observations in a dataset. But, there is no big difference between the variance of datasets before and after outlier detection by other methods. Although numbers of detected outliers by all three methods are approximately same, experiments results prove that outlier detected by statistical method are more likely to be real unusual observations. Eventually the statistical based method with very simple algorithm performs well in

detecting outliers and accuracy of classifier increase after outlier detection by this method compared to other methods. Method complexity and determination of K for KNN, α, β and γ values to calculate COS in distance based method and determination of number of clusters and outliers in clustering based method are considerable points in these methods.

5. Conclusion

Outlier detection is a very important problem which is used in a wide variety of areas. In this paper three well-known multivariate outlier detection methods namely Distance based, Statistical based and Clustering based methods were evaluated. Five benchmark datasets of Heart disease, Breast cancer, Pima Indian diabetes, Liver disorder and Thyroid gland are used to perform experiments. To realize how well outliers are detected by each outlier detection method, we calculated variance of each dataset before and after outlier detection. Eight widely used individual and ensemble classifiers are also used to classify data before and after outlier detection. The generalization abilities (accuracy) of each classifier are also compared to evaluate the effectiveness of outlier detection methods which are used in this paper. Finally, experiment results shows that Statistical based method outperforms other outlier detection methods which are evaluated in this study.

Acknowledgment

The authors would like to thank The Mitsubishi Corporation and university of the Ryukyus for their supports.

References

[1] H. J. Escalante, "A Comparison of Outlier Detection Methods for Machine Learning", Congress International on Computation- IPN, 2005.

[2] V. Chandola, R. Banerjee, and V. Kumar, "Outlier Detection: A Survey", 2007, Technical Report, University of Minnesota.

[3] V. S. Hodge and J. Austin, "A Survey of Outlier Detection Methodologies," Kluwer Academic Publishers, 2004.

[4] I. Ben-Gal, "Outlier Detection," Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, Kluwer Academic Publishers, 2005.

[5] D. M. Hawkins, "Identification of Outliers," Springer 1980.

[6] M. Lichman (2013), UCI Machine Learning Repository [http://archive.ics.uci.edu/ml] Irvine, CA: University of the California, School of Information and Computer Science.

[7] V. Barnett and T. Lewis, "Outliers in Statistical Data," Third Edition, J. Wiley & Sons, 1994.

[8] G. H. John, "Robust Decision Trees: Removing Outliers from Database," Association for Advancement of Artificial Intelligence, 1995.

[9] C. C. Aggarwal and P. S. Yu, "Outlier Detection for High Dimensional Data," ACM SIGMOD International Conference on Management of Data, USA 2001.

[10] Y. Kou, C. T. Lu, and D. Chen, "Spatial weighted outlier detection", In proceedings of SIAM Conference on Data Mining, 2006.

[11] E. M. Knorr and R. T. Ng, "Methods for Mining Distance Based Outliers in Large Datasets," Proceeding of the 24th VLDB Conference, New York, USA 1998.

[12] G. Williams, R. A. Baxter, H. X. He, S. Hawkins and L. Gu, "A Comparative Study of RNN for Outlier Detection in Data Mining," IEEE International Conference on Data Mining (ICDM), Japan 2002.

[13] H. Liu, S. Shah and W. Jiang, "Online Outlier Detection and Data Cleaning," Computers and Chemical Engineering, 2004.

[14] T. Hu and S. Y. Sung, "Detecting Pattern Based Outliers," Pattern Recognition Letters, 24,3059-3068.

[15] P. Chen, L. Wong and J. Li, "Detection of Outlier Residues for Improving Interface Prediction Protein Heterocompleses," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol 9. No. 4. August 2012.

[16] Z. Nazari and D. Kang, "Density Based Support Vector Machines for Classification," International Journal of Advacnced Research in Artificial Intelligence (IJARAI), vol 4, Issue. 4, 2015.

[17] Y. Doge, "The Concise Encyclopedia of Statistics", Springer 2008.

[18] Z. Nazari, Dongshik Kang, M. Reza Asharif, Yulwan Sung and Seiji Ogawa, "A New Hierarchical Clustering Method", International Conference on Intelligent Informatics and Biomedical Sciences, Japan 2015.

[19] A. K. Mann and N. Kuar, "Review Paper on Clustering Techniques", Global Journal of Computer Science and Technology Software and Data Engineering. vol 13. Issue. 5 2013.

[20] How to read a dendrogram, http://wheatoncollege.edu/lexomics/files/2012/08/How-to-Read-a-Dendrogram-Web-Ready.pdf