AN EXTENSION OF THE OUTLIER MAP FOR VISUALIZING THE CLASSIFICATION RESULTS OF THE MULTI-CLASS SUPPORT VECTOR MACHINE

Mandana Mohammadi[1], Majid Sarmad[2*], Naser Reza Arghami[3]

[1,2,3]Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

E-mail: Manda.Mohammadi@mail.um.ac.ir[1], sarmad@um.ac.ir[2*](corresponding author)

DOI: https://doi.org/10.22452/mjcs.vol34no3.5

ABSTRACT

The main objective of this study is a graphical display of the results of the high (as well as the low) dimensional multi-class support vector machine classification. Additionally, we will visually be able to detect the outliers and misclassified observations by using this graphical tool. The "outlier map" as a successful graphical outlier detection tool of robust statistics is extended in this paper. In fact, this is a bilateral extension concerning the misclassified and outlying observations recognition. The most important feature of this extension is creating two types of discriminative boundaries to segregate the data and detect the outlying observations. For this purpose, we employed the simple but efficient concept of the "confidence interval", which is computed for the mean of decision function of support vector machine and then, "thresholding" technique. After that, the efficiency of the outlier map in terms of the preciseness of the correct outlier identification has been tested by the classification accuracy. Moreover, we deployed the margin width "before" and "after" outlier detection as the other criterion to assess the preciseness of the correct outlier identification. We conducted an empirical study based on the proposed method on the simulated and several well-known real datasets. It shows the effectiveness of our proposed method by increasing the "margin width" and gaining a higher classification accuracy.

Keywords*: *Multi-class Support Vector Machine, High-dimensional data analysis, Robust statistics, Outlier/ misclassified detection, Thresholding, Data visualization

1.0 INTRODUCTION

Support Vector Machine (SVM) which is invented and developed by Cortes and Vapnik [1], was initially designed as a binary classification tool. It has recently become one of the most common approaches for the classification and supervised learning [2, 3]. This algorithm is well suited for the high-dimensional data. The basic idea of SVM is to find out the optimal hyperplane for the linearly separable data that can maximize the margin between two classes. This importance is obtained by solving the following minimization problem:

$$\min_{\mathbf{w},\xi} \ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \tag{1}$$

$$\text{s.t.} \ \ y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i; \qquad \xi_i \geq 0; \ i = 1, \ldots, n.$$

Where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are the training samples and $y_i \in \{-1, +1\}$ are the classes labels. $\mathbf{w}$ and b are the hyperplane normal vector and the bias term, respectively. ξ_i are the slack variables which allow for the penalized constraint violation. The cost parameter (C) is the penalty parameter for controlling the trade-off between maximizing the margin and minimizing the training error. Usually, k–folds cross validation, is used to estimate the best cost parameter (C). The optimal hyperplane in Equation (1) can be found based on transforming the primal formulation of SVM to the dual space formulation of the quadratic programming (QP) problem.

$$\max_{\boldsymbol{\alpha}} \ \sum_{i=1}^{n}\boldsymbol{\alpha} i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j \tag{2}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \boldsymbol{\alpha}_i y_i = 0; \qquad 0 \leq \boldsymbol{\alpha}_i \leq C; \;\; i = 1, \ldots, n,$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ is the vector of non-negative Lagrange multipliers. By solving the above quadratic optimization problem, $\boldsymbol{\alpha}_i$ and consequently, $\mathbf{w}^* = \sum_{i=1}^{n} \boldsymbol{\alpha}_i y_i \mathbf{x}_i$ are obtained. Based on the KKT condition, the bias term is also computed as $b^* = \frac{1}{SVs} \sum_{i=1}^{SVs} ({\mathbf{w}^*}^T \mathbf{x}_i - y_i)$, where the SVs are the set of support vectors (the observations that their corresponding $\boldsymbol{\alpha}_i$ are greater than zero). The decision function of SVM is obtained according to the following formula,

$$f(\mathbf{x}) = \text{sign}({\mathbf{w}^*}^T \mathbf{x} + b^*) = \text{sign}(\sum_{i=1}^{SVs} \boldsymbol{\alpha}_i y_i \mathbf{x}_i^T \mathbf{x} + b^*). \tag{3}$$

Moreover, for the non-linearly separable case, the kernel function $(K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j))$ can be incorporated to SVM to find an optimal hyperplane in the higher dimensional space. Φ maps the data into a higher dimensional space and does not need to be explicitly specified. K is any desired kernel function that computes the dot product in some higher dimensional feature space [4, 5]. This function avoid computationally intensive operations. The decision function in the feature space is as follows,

$$f(\mathbf{x}) = \text{sign}(\sum_{i=1}^{SVs} \boldsymbol{\alpha}_i y_i \Phi(\mathbf{x}_i^T) \Phi(\mathbf{x}) + b^*) = \text{sign}(\sum_{i=1}^{SVs} \boldsymbol{\alpha}_i y_i K(\mathbf{x}_i, \mathbf{x}) + b^*). \tag{4}$$

However, existence of outlier in the data (called unclean data) is a classical problem in the pattern recognition [6]. This presence could sometimes affect the prediction accuracies of the SVM [7, 9, 8, 10, 11] (see Fig.1). The commonly used definition of outliers is provided by Hawkins [12] as *"An outlier is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism"*. In the simulation study (Section 4.0), the outliers were generated based on the above definition. To be more specific, in high-dimensional data, there is high chance of the presence of outliers [13, 2, 14, 15]. This phenomenon is leading to the violation of independently identically distributed (i.i.d.) assumption, which also causes the decrement of the classification accuracy [16, 17, 2, 18].

It is important to note that, the margin width and consequently generalization ability of SVM directly depends on

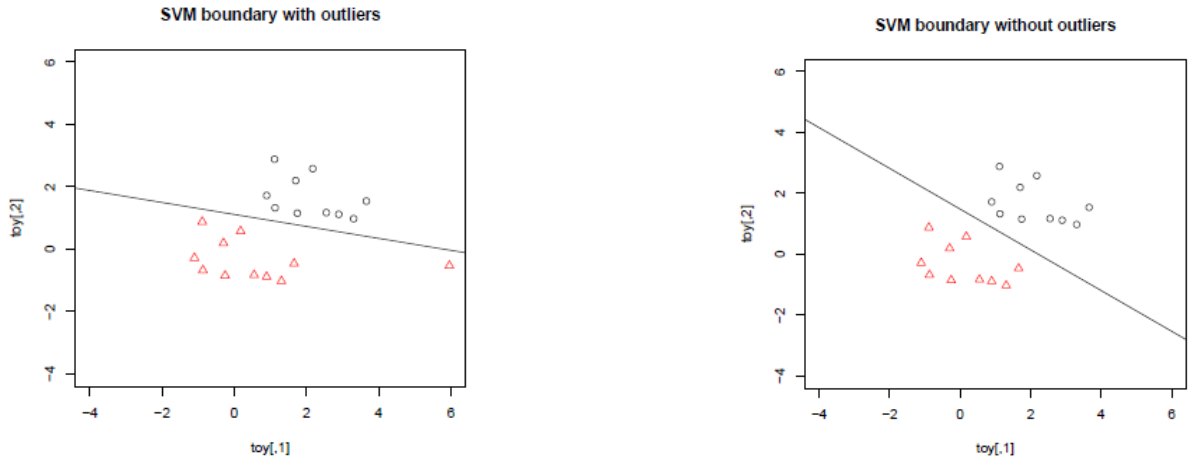

Fig. 1: Sensitivity of SVM to the outliers.

the C parameter. The number of noisy points that the model can tolerate during the training phase, is tuned by C parameter. However, these outlying points related to the misclassified observations, not the data points which are far from the regular observations. Therefore, our proposed method can find the outlying points which cannot be identified by adjusting the margin using the C parameter.

In this paper, what is necessary to consider is the concept of the margin width. In addition to the classification accuracy, the width of the margin is another criterion for evaluating the performance of SVM. It is the distance between the hyperplane and the nearest point of each class, which is equal to $\frac{2}{\|\mathbf{w}\|}$. In fact, the greater certainty of correct classification, largely depends on the margin w idth. The larger margin is leading to gain the higher

generalization ability of the discriminating hyperplane [19, 5, 4, 20, 21].
Although classification accuracy is an important criterion for comparing and evaluating the performance of a classifier, in some areas understanding the model is equally important as the classification accuracy. As a result, the transparency of the prediction process is one important requirement for predictors. More details on visualization analysis have been provided in the next section.
The outlier map as a successful graphical outlier detection tool of robust statistics is used in a different field such as linear regression [22], linear principal component analysis [23, 24], and binary classification [2, 25]. Nevertheless, the outlier map in Debruyne [2], can visually identify the abnormal points such as outliers and misclassified points for the high-dimensional data. Our previous work [25] has focused on the low dimensional data structure. We designed a graphical method to identify the outlying and misclassified points by using the minimum covariance determinant (MCD) estimator, which is inherently suitable for the low dimensional data.
However, the lack of a method that, in addition to visualizing high-dimensional multi-class data, can create an innovative boundary for the segregation of data in two-dimensional space motivates this work. The other requirements of this method are its significant conformity to the separation boundary created by SVM and the ability to detect outliers. This extension was inspired by Debruyne [2], which is used for the identification of atypical observations in the binary classification problem. In this research, we extended the binary outlier map to the multi-class problem. In fact, the binary outlier map was extended based on two perspectives, which produced two types of segregation criteria with the aim of the outlier and misclassified detection. In respect of the binary outlier map, the vertical line passing through zero is the segregating criteria for the misclassified detection. There is no such thing for the multi-class problem. The most notable part of this research is creating the boundary for segregating the classes of data. To this end, we utilized the simple but functional concept of the confidence interval. Furthermore, the extended outlier map takes the advantage of the Stahel–Donoho outlyingness measure similar to the work of Debruyne [2]. However, due to the non-existence of any theoretical threshold in the binary outlier map [2], the identification of outliers was mainly based upon a visual interpretation and was not truly trustworthy. The difference of our proposed method from Debruyne's work [2] is producing a criterion (boundary) for recognizing the outliers from the clean data (regular observations) by the employment of the thresholding methods. Different approaches for this technique exist in the literature [26, 27, 28]. But, we used the threshold based on the parametric approach which is explained in Section 3.1.
In short words we can say that the main contribution of the present study is

- extending one visual outlier detection tool from the binary classification to the multi-class problem called the outlier map by using the concept of confidence interval
- proposing theoretical threshold for the sake of precise outlier detection
- gaining insight on different types of outliers.

The rest of the paper is organized in the following manner. Section 2.0, contains three parts, including the overview of visualization analysis, multi-class SVM, and Stahel–Donoho outlyingness measure. Section 3.0 devoted to the construction of the extended outlier map. In Section 4.0, we applied our methods to a simple but illustrative, simulated example and also to a few real data. The detailed experimental results and performance assessment are presented in this section. We have also compared the proposed method with the related work in Section 5.0. Finally, Section 6.0 concludes the paper.

2.0 SOME PRELIMINARY CONCEPTS FOR CONSTRUCTING THE ROBUST MULTI-CLASS SVM CLASSIFIER

In this section, we present a background on the visualization of SVM. We also provide a short introduction to the multi-class classification problem and the Stahel–Donoho outlyingness measure. We implement these concepts to design a graphical method for the identification of observations that affect the performance of the multi-class SVM classification, especially for the high-dimensional data.

2.1 Visualization

The background of most visualization methods comes from the exploratory analysis with the invention of the box-plot to detect outliers [29]. Data exploration is still one of the main thrust of modern visualization techniques. A

plethora of visualization technique has been proposed over the last decade by different researchers e.g. [30, 31, 32, 33, 34, 35, 36, 37, 38]. At the same time, some data visualization research has focused specifically on classification issues. The visualization is very helpful in understanding the boundaries between classes, the degree of classes overlapping, as well as the degree of importance of each feature in different classes. The transparent explanatory analysis of the data gives the analyst higher confidence in the interpretation of the relationship between features and classifier, and also the identification of the trends and suspicious patterns in the training phase.
Visualization of the decision boundary is one of the most informative elements of the classification system, which is performed by various researchers. Jakulin et al. [30] presented an approach for visualizing the SVM models by extending one graphical tool called nomograms. They offered further clarification on the impact of the different feature values on the probability that the corresponding sample will be assigned to a specific class. The other category of visualizing SVM is restricted to the projections method that transforms the high-dimensional data into two-dimensional space. Wang et al. [39] attempted to visualize the binary classification results by the employment of self-organizing maps. Moreover, self-organizing maps from unsupervised learning have been used by Hamel et al. [35] to visualize the support vectors and decision surfaces. This method also conveys visually the dependency of the final decision to the data cluster. On the other hand, Caragea et al. [40] deployed the projection-based tour methods to visualize the SVM classifiers with the linear kernel. In addition to the above, Hansen et al. [32] studied the effect of the data futures on the decision surfaces. They visualized the most relevant training sample along with a predicted value to assess the prediction quality. More recently, Barbosa et al. [33] presented a visual kernel-based multi-dimensional projection technique for analyzing kernelized data. Their proposed technique help to determine the effect of kernel choice on the neighborhood structure and SVM decision boundaries. Additionally, a color-based Nomogram is proposed by Van Belle et al. [37]. They also quantify the impact of each feature on the prediction. In other recent literature, Ma et al. [41] proposed a visualization tool EasySVM to facilitate the users in tuning parameters, managing the training data, and extracting rules as a component of the SVM training process.
Another crucial factor that is important to display in addition to the decision boundaries is the outlying points. There are various methods available to visualize the high-dimensional data through dimension reduction. However, some outliers cannot be depictable by such methods [13, 15, 42, 43]. As stated by Onderwater [42], when the dimensionality is reduced by projecting all points onto a line, they might be placed in the middle part of the reduced data. This will lead to a mistaken inference about the nature of the regular observations and the outliers. In other words, outliers might be hidden among regular observations and scape from the detection by the implementation of the dimensionality reduction methods. This will cause problems for applications that rely on the detection of outliers. Based on this statement, a proper tool for visualizing the high-dimensional data is essential.

2.2 Multi-class classification

The classification problem involving multiple classes, known as multi-class classification are frequently used in real-world problems. Naturally extending the binary classification technique can efficiently solve many state-of-the-art multi-class classification approaches. Multi-class SVM uses the most popular schemes such as one-vs-all and one-vs-one strategies. In this paper, the one-vs-one strategy (henceforth, abbreviated as OvO) is chosen to construct the extended outlier map.
The goal is to learn a classification rule (decision function $f(\mathbf{x}_i)$) from the data such that correctly predicts the class label (y_i). The new unseen observations will assign to one of the possible classes. In the OvO reduction approach a set of binary classifiers is trained; each classifier separates a pair of classes while discarding the rest of the classes. This approach transforms a problem with M classes into $L = M(M-1)/2$ binary classifiers, where each one is trained on the data from two classes. As a result, for each observation, there would be L decision function $f(\mathbf{x}_i)$. Subsequently, according to the voting strategy, add a vote for the winning and the class with more votes will be the output. However, for the case with identical votes for two classes, the new unseen observations are simply assigned to the class with the smaller index.

2.3 The Stahel–Donoho outlyingness

The Stahel–Donoho outlyingness measure [44, 45] is based on the one-dimensional projection of data on the axis in which the observation is the most outlying. The underlying idea is that every multivariate outlier must be a univariate outlier in some projection. Suppose the set of n observations with p variables denoted by $\mathbf{X}$, as an element of $\mathbb{R}^p$. Let m and s be the location and scale functionals in $\mathbb{R}^1$. Due to non-robustness of classical estimators to the presence of outliers, they are considered as the median and the median absolute deviation (MAD), respectively [22].

The outlyingness of $\mathbf{x}_i$ with respect to $\mathbf{X}$ can be defined as

$$\text{outl}(\mathbf{x}_i, \mathbf{X}) = \sup_{\mathbf{a} \in S_p} \mid \frac{\mathbf{a}^t \mathbf{x}_i - m(\mathbf{a}^t \mathbf{X})}{s(\mathbf{a}^t \mathbf{X})} \mid, \tag{5}$$

where $S_p = \{\mathbf{a} : \|\mathbf{a}\| = 1\}$ is a set of p directions in $\mathbb{R}^p$. According to Zuo [46], the exact computation of the supremum in the outlyingness (Equation (5)) is not feasible. Therefore, a random search algorithm based on sub-sampling is usually used (for more details, see [47, 48, 2]). Further, according to Debruyne [2], the outlyingness measure can be computed in an arbitrary kernel-induced feature space.

3.0 EXTENSION OF OUTLIER MAP FOR THE MULTI-CLASS CLASSIFICATION PROBLEM

The extended outlier map similar to its binary version, is a visual method which can simplify the recognition of different types of outlying points. Whether, regarding Hawkins [12] definition, the observation potentially being an outlier or it is a mislabelled point. In the binary outlier map [2] there is not any specified or theoretical criteria to identify what observations qualify as potential outliers. This is upon the user's opinion to decide an observation as outliers regarding its distance to the bulk of data. Due to the nature of non-precise of detected outliers which rely on the visual perception of the user, the decision will not be realistic. Meaning that, the detected outliers may vary depending on the user's opinion. By the nature of the outlier map, that is bilaterally capable of detecting anomalous data, its recognition ability is based on two kinds of separation lines namely the "threshold" and "class boundary". We utilized the thresholding technique for identification of outliers and the concept of confidence interval for recognition of misclssified data which will be explained in the next subsections.

3.1 Outlying points recognition criteria

We utilized the thresholding technique to binaries the outlyingness of data. In fact, this technique makes the distinction between data with low outlyingness value (regular data) versus the data with the high amount of outlyingness (outlier). After computing the outlyingness of data based on Equation (5), a discriminating method is required to make a threshold for segregation of outliers from regular data. For this purpose, we used the idea of the thresholding technique which is widely applied in the field of image processing. Thresholding is used in image processing for partitioning an image into a foreground and background. This technique basically discriminates two types of normally distributed object based on the specific criteria. Different approaches for this technique exist but, we used the parametric approach which is proposed by Kittler and Illingworth [49]. In this approach, the probability density function of outlyingness of outliers and regular points (Equation (5)) is estimated to model each class.
The Gaussian mixture model is used to model the distribution of the outlyingness of the data. In this method, we assumed that the distribution of the outlyingness of the data is mixture of two normal distribution as $f_k(x_i; \boldsymbol{\theta}_k)$ for $k = 1, 2$ (where f_1 is related to the outlyingness of the regular points and f_2 is related to the outlyingness of the outlier points.) with distinct mean μ_k, standard deviation σ_k and prior probability π_k (where $\pi_1 + \pi_2 = 1$ and $0 \leq \pi_k \leq 1$) as,

$$f(x_i; \boldsymbol{\theta}) = \pi_1 f_1(x_i; \boldsymbol{\theta}_1) + \pi_2 f_2(x_i; \boldsymbol{\theta}_2), \tag{6}$$

where,

$$f_k(x_i; \boldsymbol{\theta}_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}). \tag{7}$$

The parameters of this model $\theta_k = \{\mu_k, \sigma_k\}$ and π_k can be estimated from the training data by an iterative Expectation-Maximization (EM) algorithm.
After estimating the parameters of the above mentioned mixture model, we assumed that the two sub-populations associated with the outlyingness of data can be segregate using the threshold T. This threshold can be defined as follows

$$\pi_1 f(x_i; \boldsymbol{\theta}_1) < \pi_2 f(x_i; \boldsymbol{\theta}_2); \qquad x_i \leq T \tag{8}$$

$$\pi_1 f(x_i; \boldsymbol{\theta}_1) > \pi_2 f(x_i; \boldsymbol{\theta}_2); \qquad x_i > T. \tag{9}$$

By applying the logarithmic function on both sides of the Inequalities (8) and (9),

$$\frac{(x_i-\mu_1)^2}{\sigma_1^2}+\log\sigma_1^2-2\log\pi_1<\frac{(x_i-\mu_2)^2}{\sigma_2^2}+\log\sigma_2^2-2\log\pi_2; \qquad x_i\leq T \tag{10}$$

$$\frac{(x_i-\mu_1)^2}{\sigma_1^2}+\log\sigma_1^2-2\log\pi_1>\frac{(x_i-\mu_2)^2}{\sigma_2^2}+\log\sigma_2^2-2\log\pi_2; \qquad x_i>T. \tag{11}$$

The threshold can be obtained by solving the quadratic equation. We can get it by equating both sides of Inequalities (8) and (9) and ignoring the negative answer. T must satisfy the following equation,

$$AT^2+BT+C=0, \tag{12}$$

where

$$\begin{aligned} A&=\sigma_1^2-\sigma_2^2, \\ B&=2(\mu_1\sigma_2^2-\mu_2\sigma_1^2), \\ C&=\sigma_1^2\mu_2^2-\sigma_2^2\mu_1^2+2\sigma_1^2 2\sigma_2^2 ln(\frac{\sigma_2 P_1}{\sigma_1 P_2}) \end{aligned} \tag{13}$$

After finding the threshold T, this two sub-population can be segregated based on T, which is represented in Fig.2 by the vertical line. The thresholding technique is used to identify outliers based on the computed outlyingness

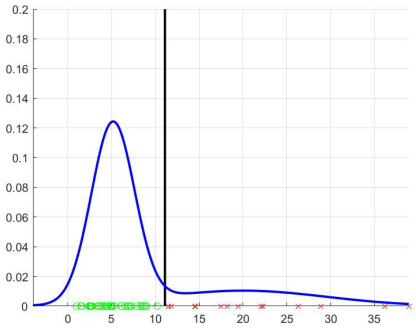

Fig. 2: Binarisation of dataset based on the produced threshold line.

(outl($\mathbf{x}_i$)). Each observation is classified as outliers if outl($\mathbf{x}_i$) $> T$ and vice versa. Therefore, based on the above-mentioned technique, the dataset divided into two parts based on their outlyingness. The major part which is called clean (the regular points) and the rest as the outliers part.

3.2 Misclassified points recognition criteria

Besides detection of the outliers, the ability to find the misclassified observations is the other important expected property of outlier map. Therefore, a visual criterion capable of distinguishing the observations which are wrongly classified is required. As mentioned before, due to the chosen OvO scheme, there would be L (where $l=1,\ldots,L$) set of classifiers output $(f(\mathbf{x}_i))$, where $f(\mathbf{x}_i)$ is defined as,

$$f(\mathbf{x}_i)=\sum_{i=1}^{SVs}\boldsymbol{\alpha}_i y_i \mathbf{x}_i^T\mathbf{x}+b, \tag{14}$$

Thus, in order to have a single criterion as the boundary of each class, the suggested method is as follows.
Firstly, for each observation $\mathbf{x}_i$, there is L classifying function $f_{ljk}(\mathbf{x}_i)$, where j and k are different class labels. Such that $j<k$ ($j=1,\ldots,M-1$ and $k=2,\ldots,M$). The measure of central tendency is used to give us a single value instead of a set of f. The *bar* notation is used to denote the average over the all set of classifier of each observation ($\mathbf{x}_i$), which is denoted by $\bar{f}(\mathbf{x}_i)$.

$$\bar{f}(\mathbf{x}_i)=\frac{1}{L}\sum_{l=1}^{L}f_{ljk} \tag{15}$$

At this step, there will be a single value for the classifying function corresponding to each observation.
Secondly, for the observations of each class, the mean of the respective $\bar{f}(\mathbf{x}_i)$ which is obtained from the previous step is computed as follows. Suppose, $\mathbf{x}_{i_m} = \{\mathbf{x}_i | \mathbf{x}_i \in m\}$, in which $m = \{1, \ldots, M\}$. Note that, $\bar{\bar{f}}(\mathbf{x}_{i_m})$ is the mean of the mean of the classifying functions of the observations that belong to the m-th class ($\bar{f}(\mathbf{x}_{i_m})$).

$$\bar{\bar{f}}(\mathbf{x}_{i_m}) = \frac{1}{\sum_{i=1}^{n} I_m(\mathbf{x}_i)} \sum_{\mathbf{x}_i \in \mathbf{x}_{i_m}} \bar{f}(\mathbf{x}_{i_m}), \tag{16}$$

where,

$$I_m(\mathbf{x}_i) = \left\{ \begin{array}{lll} 1 & \text{for} & \mathbf{x}_i \in m \\ 0 & \text{for} & \mathbf{x}_i \notin m. \end{array} \right\}$$

Then, $100(1-\alpha)\%$ confidence interval related to the $\bar{\bar{f}}(\mathbf{x}_{i_m})$ is computed as $[\bar{X}_{\bar{\bar{f}}(\mathbf{x}_{i_m})} \pm z_{\alpha/2} \frac{\sigma_{\bar{\bar{f}}(\mathbf{x}_{i_m})}}{\sqrt{n_{\bar{\bar{f}}(\mathbf{x}_{i_m})}}}]$, in which $\bar{X}_{\bar{\bar{f}}(\mathbf{x}_{i_m})}$, $\sigma_{\bar{\bar{f}}(\mathbf{x}_{i_m})}$ and $n_{\bar{\bar{f}}(\mathbf{x}_{i_m})}$ are the corresponding mean, standard deviation and sample size of $\bar{\bar{f}}(\mathbf{x}_{i_m})$, respectively. Afterward, the classes are reordered by $\bar{\bar{f}}(\mathbf{x}_{i_m})$. For example, $\bar{\bar{f}}_{(1)}(\mathbf{x}_{i_m})$ is the class with the lowest $\bar{\bar{f}}(\mathbf{x}_{i_m})$. In this stage, we can compute the boundary of the classes. Such that the mean of the upper bound of any previous class and the lower bound of its next class is the boundary line of each class.

3.3 Extended outlier map

According to the previous Subsections 3.1 and 3.2, we gained the ability to recognize atypical observations such as outlying points and misclassified observations. We can identify misclassified observations with the class boundary and the effective outliers on the SVM by utilizing the related threshold of the outlier map. In the following, we briefly explain the steps involved in making and using the extended outlier map.
First, we computed the outlyingness (outl($\mathbf{x}_i$)) and the decision function ($\bar{f}(\mathbf{x}_i)$) of each observation of the training set. Where, outl($\mathbf{x}_i$) is the Stahel–Donoho outlyingness of observation i, (using Equation (5)). Then we sketched the scatter plot of outlyingness outl($\mathbf{x}_i$) versus the value of the $\bar{f}(\mathbf{x}_i)$. Thus, for $i = 1, \ldots, n$, pairs ($\bar{f}(\mathbf{x}_i)$, outl($\mathbf{x}_i$)) are plotted. Subsequently, we displayed the observations of each class with a specific symbol that is distinct from the other classes. By using Equation (16) we calculated $\bar{\bar{f}}(\mathbf{x}_{i_m})$ and sort ascendingly, the amounts of $\bar{\bar{f}}(\mathbf{x}_{i_m})$. After that, we computed the $100(1-\alpha)\%$ confidence interval of the corresponding $\bar{\bar{f}}(\mathbf{x}_{i_m})$ of each class separately. A solid vertical line is added at the mean of the upper and lower bound of $100(1-\alpha)\%$ confidence interval of the $\bar{\bar{f}}(\mathbf{x}_{i_m})$ for any previous class and its next class. Also, we added a solid horizontal line at the produced threshold to the graph for the sake of dividing the outliers from the regular points. We have summarized the above-mentioned steps in the algorithmic format and graphically in Algorithm 1 and Fig. 3.

Algorithm 1 Extended outlier map algorithm

Input:A data matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^T$, labels vector $\mathbf{y} = [y_1, y_2, \ldots, y_n]$
Output: A diagnostic plot to detect the outliers and misclassified points
Algorithm:

1. Calculate outl($\mathbf{x}_i, \mathbf{X}$) using Equation (5)
2. Fit a two component Gaussian mixture model to the outlyingness of data
3. Find threshold T using Equation (12)
4. Find the clean portion of data based on T
5. Train a SVM model on clean train set using OvO strategy
6. Calculate $\bar{f}(\mathbf{x}_i)$ using Equation (15)
7. Plot the obtained amount from previous steps 1 and 6 versus each other
8. Sketch the vertical and horizontal line respectively at the mean of the upper and lower bound of the $100(1-\alpha)\%$ confidence interval of $\bar{\bar{f}}(\mathbf{x}_{i_m})$ (using Equation (16)) for any previous class and its next class and the threshold line T.

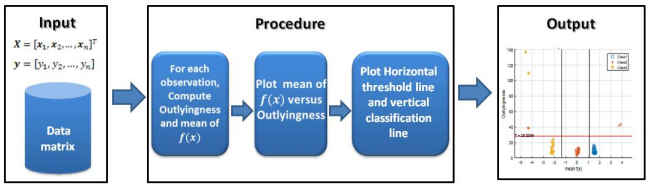

Fig 3: Stepwise procedure for producing the extended outlier map.

The clean dataset is utilized to train the SVM classifying function to see whether the detected outliers really affect the performance of SVM or not. For this purpose, the accuracy of classification is calculated on the testing dataset.

We iterated the experiments 100 times. In each iteration, we have drawn the training and a testing set. Such that 70% is chosen for the training set and the remaining of the observation has been assigned to the testing set. Then we calculated and presented the average of the accuracy obtained in each iteration. In all the examples of this paper, 10-fold cross-validation was used to optimize parameter C. According to Debruyne [2] the linear kernel has been used, due to the high-dimensional structure of the data sets in this study.

4.0 EXPERIMENT ON THE VISUALIZATION OF SVM RESULTS

In this section, a simulation study was carried out to assess the performance of the extended outlier map. Apart from that, some real-world datasets are used to test the performance of the proposed method.

4.1 Simulation study

We are mostly interested in generating a simple multi-class high-dimensional example. The classes labelled as I, II and III. Two different scenarios are considered to check the performance of the proposed method. Thirty-three observations were generated from a multivariate Gaussian distribution $MVN(\mu, I_p)$ for each class, where $p = 70$ and $\mu_I = -3$ (for class I), $\mu_{II} = 0$ (for class II) and $\mu_{III} = 2$ (for class III). In the next step, some observations substituted to illustrate several types of outliers such as outlying and mislabelled data. Therefore, in the first scenario, a small portion of data is substituted by the generated outliers from $MVN(\mu^*, I_p)$. Where μ^* and $\mu^*{}_{II}$ are equal -6 and 5, respectively. As, it is evident from the indices, μ^* and $\mu^*{}_{II}$ are the mean of the contaminated observation of the first and third classes. The second scenario is a combination of the existence of the mislabelled and outlying points. These scenarios are presented in Table 1.

Table 1: Different scenario for simulation study.

Case	Tot. no. of obs.	Class *I*			Class *II*			Class *III*		
		Label	Mean	Total	Label	Mean	Total	Label	Mean	Total
Contaminated	99	I	μ_I	31	-	-	-	-	-	-
		I	$\mu^*{}_I$	2	II	μ_{II}	33	III	$\mu^*{}_{III}$	3
		-	-	-	-	-	-	III	μ_{III}	30
Mislabelled-Contaminated	99	I	μ_I	30	-	-	-	II	$\mu^*{}_{III}$	2
		II	$\mu^*{}_I$	3	II	μ_{II}	33	III	$\mu^*{}_{III}$	3
		-	-	-	-	-	-	III	μ_{III}	28

Obviously, from Fig. 4(a) the extended outlier map detects two outlying observations in the first class; and three outliers in the third class.

The second part of the simulation study devoted to the presence of the mislabelled and outlying observation, simultaneously. According to Fig. 4(b), there are two mislabelled observations in the first class. The existence of the homogeneous observations is shown in the second class, while in the third class, the presence of the two types of outliers (misclassified and outlying observations) is visible.

As previously mentioned, due to the improvement of the generalization capacity of SVM, the separation hyperplane with the largest margin is desired. With OvO strategy of the multi-class, there are L binary classifiers and consequently L set of normal vectors ($\mathbf{w}$). Therefore, in order to have a comparison criterion, we take the average over all the margins.

For the sake of the comparison, the criterion for the datasets including outliers calculated. Later on, the criterion will present after omitting the detected outliers by the extended outlier map. The criterion before omitting the detected outliers for the contaminated and mislabelled-contaminated are respectively equal to 38.18 and 22.10 which is equal to 38.18 and 37.92 after outlier removal.

It is worth noting that, for the contaminated data this criterion does not change, due to the non-existence of any

misclassified observations and; but, it increased for the second scenario. Therefore, the wider separation area is leading to gain the better generalization ability.

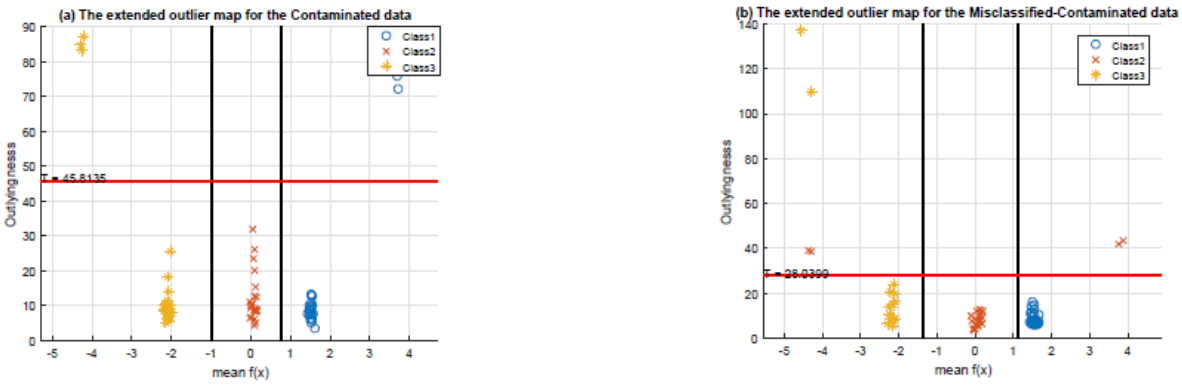

Fig. 4: The plot of the Stahel–Donoho outlyingness versus the value of the classifier for contaminated (a) and mislabelled–contaminated data (b).

4.2 Real-world data examples

We used some benchmark datasets to evaluate the performance of the extended outlier map. Fisher's Iris data, Leukemia cancer (MLL), Lung cancer, Gastric cancer, Small-blue-round-cell (SRBCT) tumor and Glioblastoma cancer are the respective datasets that are publicly available [1]. These datasets and their characteristics are presented in Table 2. As shown in Fig. 4 and Fig. 5, the outlying points are recognizable by the respective threshold of each outlier map. The misclassified observations in each plot are also clear based on their different relevant symbol and color.

Table 2: The benchmark dataset.

Dataset	No. of observation	No. of features	No. of classes
Fisher's Iris	150	4	3
Leukemia cancer (MLL)	56	12582	3
Lung cancer	34	10541	3
Gastric cancer	30	4522	3
Small-blue-round-cell (SRBCT) tumor	83	2308	4
Glioblastoma cancer	50	12625	4

Thresholding: Fig. 6 shows the corresponding produced threshold for the simulated and real dataset. There is no visual method able to recognize the atypical points in high-dimensional multi-class data. Thus, we have only presented our results.

Note that it is not our aim to compare the classifier's accuracies but rather to visualize high-dimensional multi-class data, which leads to the identification of outlying observations. Nevertheless, the accuracy criterion in Table 3 and the average margin width in Table 4 show the significance of the outlier identification by improving the classification results. The increase of margin width and classification accuracy after omitting the detected outliers by the extended outlier map on the simulated and real data illustrated that our proposed visual method satisfies the claims. Obviously, if we mistakenly considered clean data instead of outlying points, the classification accuracy would be reduced as the sample size decreased. Recalling from the SVM theory, the classification decision has been made based on a small part of the data named the support vectors placed in the margin boundary. The changing of one of them will necessarily change the margin and the decision boundary of the model. Therefore, reducing the number of non-SVs train data will not affect the SVM model. Especially when linear kernels are used for the classification, the SVs make up a small percentage of the data. Even if it includes some SVs, reducing the amount of data is still not significant, as confirmed by not reduction of the model accuracy.

The confusion matrix was computed to verify the performance of the extended outlier map from the perspective of misclassified identification. Regarding the simulated data, the number of generated mislabelled points exactly

[1] http://www.biolab.si/supp/bi-cancer/projections/index.html

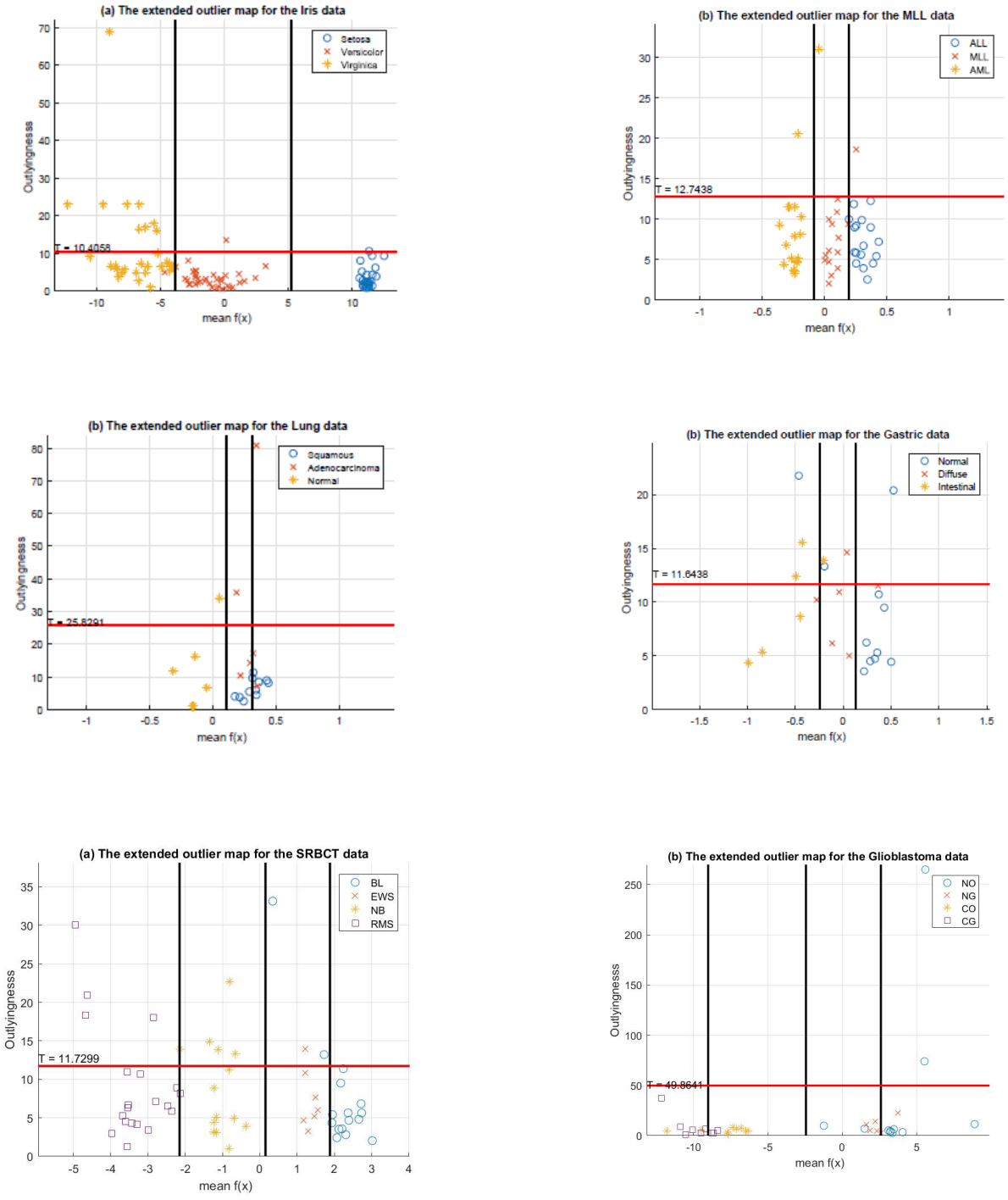

Fig. 5: The plot of the Stahel–Donoho outlyingness versus the value of the classifier for the Iris, MLL, Lung, Gastric, SRBCT and Glioblastoma data.

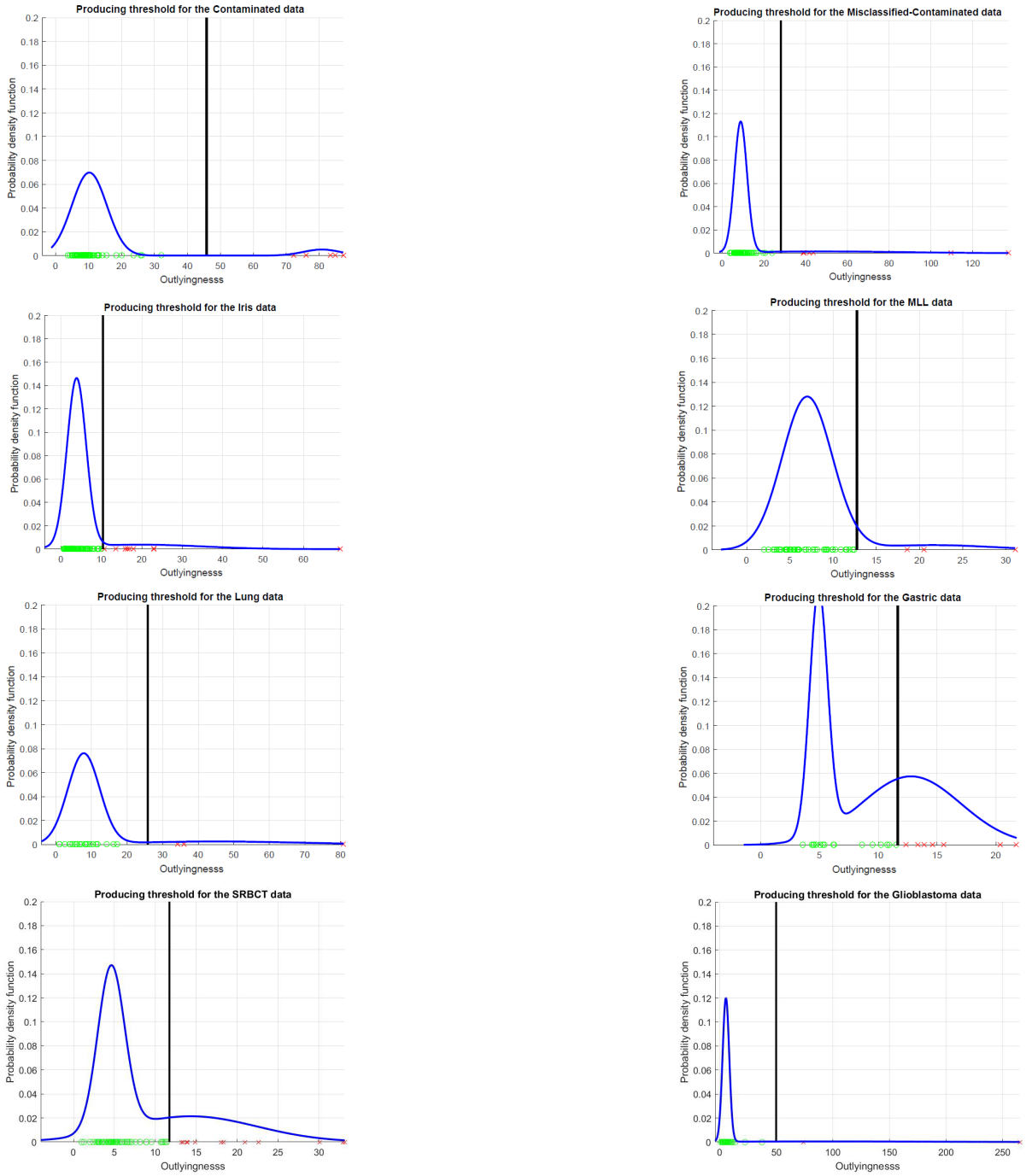

Fig. 6: Simulated and real data approximated by two Gaussian distribution and partitioned based on the produced threshold

Table 3: Mean(standard division) of the classification accuracy obtained without (with) outlier omission based on the result of the extended outlier map.

Dataset	Without outlier omission	With outlier omission
Contaminated	$100(0.01)$	$100(0.01)$
Mislabelled–Contaminated	$76.67(0.05)$	$96.67(0.04)$
Fisher's Iris	$93.33(0.06)$	$95.56(0.08)$
Leukemia cancer (MLL)	$81.82(0.05)$	$90.91(0.33)$
Lung cancer	$63.64(0.07)$	$81.82(0.02)$
Gastric cancer	$54.54(0.15)$	$63.64(0.04)$
Small-blue-round-cell (SRBCT) tumor	$90.12(0.02)$	$92.57(0.03)$
Glioblastoma cancer	$72.08(0.05)$	$73.33(0.03)$

coincides with the obtained confusion matrix (Table 5). The misclassified points shown by the outlier map of the real data are very similar to the classification confusion matrix. In the real data, we saw that the vertical line did not exactly coincide with SVM results. It should be taking into account that it is a visual perspective of high-dimensional multi-class SVM classification. This method aims to have a clear picture of the problem at hand. It does not mean that we should expect the outlier map to achieve similar results with SVM. Since we aim to visualize the SVM classification results based on the proposed scheme, whereas SVM's purpose is to segregate the data as much as possible without considering the visualization concept.
According to all of the figures, the proposed method has operated well for separable data like Fisher Iris. It has also depicted the intertwining of the inherently intertwined data. Consequently, it can be deduced that the extended outlier map worked properly.

Table 4: The percentage of detected outliers by the extended outlier map and the average of margins width before and after outlier removal.

Dataset	Margin Without outlier	Percent of detected outliers	Margin With outlier
Contaminated	38.18	0.07	38.18
Mislabelled-Contaminated	22.10	0.08	37.92
Fisher's Iris	2.89	0.20	3.11
Leukemia cancer (MLL)	535468.2	0.06	655776.17
Lung cancer	342117.02	0.13	399043.90
Gastric cancer	325386.31	0.30	337777.82
Small-blue-round-cell (SRBCT) tumor	60.97	0.32	64.36
Glioblastoma cancer	72.08	0.12	73.33

5.0 COMPARISON TO THE RELATED WORK

In this section, we attempted to summarize the relevant research and compare our proposed method with them. We have categorized them in terms of perspective and specific usage. In contrast with the approaches found in the literature, this work attends to present a visual tool with some specific differences and purpose from the existing methods. One distinct difference between the proposed method and the existing methods is related to separator boundaries and functionality. In the proposed method, the separator boundaries are, in fact, a function of the decision function $f(x)$, which is tried to be as close as possible to the separator boundaries of SVM. In contrast, the existing methods have illustrated the boundaries of SVM separators. It should be noted that the binary problem is considered in all these methods, and the multi-class problem is not addressed. The most notable difference is that the extended outlier map is designed for the multi-class classification problem. Furthermore, the existing methods focused on the quantification of the impact of each feature on the prediction, while the extended outlier map deals with the effect of outliers on the classification system. The second important element of our method is identifying outlying data, which is applied in a couple of consecutive stages. The first criterion is the Stahel-Donoho outlyingness measure, and the second is the thresholding technique. Among the references, only references [40] and [41] have considered the outliers, which are similar to Debruyne's work [2], are based on the user's opinion, and no theoretical criteria have been proposed for outlier detection. Moreover, unlike references [30, 39], our proposed method can deploy any kind of kernel. Nonetheless, most of the existing classification algorithms [31, 37, 39] required the pre-processing step to reduce the dimensionality. The other advantages of our proposed method are needless of any pre-processing steps such as dimensionality reduction. The similarity of the extended outlier map and existing methods [41] is their

Table 5: Confusion table for the Contaminated, Mislabelled–Contaminated, Iris, MLL, Lung, Gastric, SRBCT and Glioblastoma data.

(a) Contaminated

		Actual		
		I	*II*	*III*
Fitted	*I*	23	0	0
	II	0	23	0
	III	0	0	23

(b) Mislabelled-Contaminated

		Actual		
		I	*II*	*III*
Fitted	*I*	28	0	0
	II	2	21	2
	III	0	0	16

(c) Iris

		Actual		
		I	*II*	*III*
Fitted	*I*	50	0	0
	II	0	49	1
	III	0	0	50

(d) MLL

		Actual		
		I	*II*	*III*
Fitted	*I*	22	2	0
	II	0	20	0
	III	0	1	27

(e) Lung

		Actual		
		I	*II*	*III*
Fitted	*I*	11	0	0
	II	2	4	0
	III	1	0	5

(f) Gastric

		Actual		
		I	*II*	*III*
Fitted	*I*	9	1	1
	II	0	6	0
	III	0	0	6

(g) SRBCT

		Actual			
		I	*II*	*III*	*IV*
Fitted	*I*	15	1	0	0
	II	0	8	0	0
	III	0	0	13	1
	IV	0	0	0	20

(h) Glioblastoma

		Actual			
		I	*II*	*III*	*IV*
Fitted	*I*	10	1	0	0
	II	2	4	0	0
	III	0	0	9	0
	IV	0	0	2	7

incorporation ability to other classifiers. Obviously, the proposed method is only an approximate solution, and there is no real solution to visualize the high-dimensional data in only two-dimensions.

6.0 CONCLUSION

In this paper, we aimed at generating a visual boundary for the multi-class classification problem that is close to the SVM boundary as much as possible. To this end, we have proposed an extended outlier map for the multi-class support vector machine classification. The strength of this study is producing two types of boundaries. The proposed boundaries have the ability to recognize misclassified and outlying points. For this purpose, the confidence interval and the thresholding methods are deployed. If the extended outlier map implies the homogeneousness of the classes, the analysis can be safely conducted without fear of the outlier's effect. Otherwise, the analyst may attempt to delete or modify the outliers depending on the problem. It is important to know that by removing the true outliers, the SVM margins width, as well as the classification accuracy, increased. This visual diagnostic tool as the preprocessing operation can incorporate with other multi-class classifiers such as random forest or boosting the classification methods with little modification to gain higher classification accuracy. To the best of our knowledge, no other visual methods were found to deal with the issue of outlier detection in high-dimensional multi-class classification problems using SVM classifier.

7.0 ACKNOWLEDGEMENT

This paper is part of the Ph.D thesis submitted to Ferdowsi University of Mashhad by Mandana Mohammadi (No. 37943).

REFERENCES

[1] C. Cortes, V. Vapnik, "Support-vector networks". Machine learning, Vol. 20, No. 3, 1995, pp. 273-97.

[2] M. Debruyne, "An outlier map for support vector machine classification". The Annals of Applied Statistics, 2009, 1566-1580.

[3] D. Hwang, D. Kim, "Near-Boundary Data Selection for Fast Suppor Vector Machines". Malaysian Journal of Computer Science, Vol. 25, No. 1, 2012, pp. 23-37.

[4] B. E. Boser , I. M. Guyon , V. N. Vapnik, " A training algorithm for optimal margin classifiers". In: Proceedings of the fifth annual workshop on Computational learning theory. New York, NY, USA: ACM, 1992, pp. 144-152.

[5] V. Vapnik, "The Nature of Statistical Learning Theory".Springer science & business media, 2013.

[6] M. Goldstein, S. Uchida, "A Comparative Study on Outlier Removal from a Large-scale Dataset using Unsupervised Anomaly Detection". ICPRAM, 2016, pp. 263-269.

[7] S. Ertekin, L. Bottou, C. L. Giles, "Nonconvex online support vector machines". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 2, 2010, pp.368-381.

[8] Kicanaoglu, Berkay, Alexandros Iosifidis, and Moncef Gabbouj. "Maximum margin binary classifiers using intrinsic and penalty graphs". In 2016 24th European Signal Processing Conference (EUSIPCO), pp. 2270-2274. IEEE, 2016.

[9] Kim, Sunwha, and Choongrak Kim. "Influence Diagnostics in Support Vector Machines". Journal of the Korean Statistical Society, Vol. 49, No. 3, 2020, pp. 757-778.

[10] W. Li et al., "Outlier detection and removal improves accuracy of machine learning approach to multispectral burn diagnostic imaging". Journal of biomedical optics, Vol. 20, No. 12, 2015, pp. 121305.

[11] Gu, Xiaoqing, Tongguang Ni, and Yiqing Fan. "A fast and robust support vector machine with anti-noise convex hull and its application in large-scale ncRNA data classification". IEEE Access. Vol. 7, 2019, pp. 134730-134741.

[12] D. M. Hawkins, "Identification of outliers". London: Chapman and Hall, 1980.

[13] C. C. Aggarwal. "Outlier Analysis". Springer Publishing Company, Incorporated, 2nd edition, 2016.

[14] Zimek, Arthur, Erich Schubert, and Hans-Peter Kriegel. "A survey on unsupervised outlier detection in high-dimensional numerical data". Statistical Analysis and Data Mining: The ASA Data Science Journal, Vol, 5, No. 5, 2012, pp. 363-387.

[15] Ro, Kwangil, Changliang Zou, Zhaojun Wang, and Guosheng Yin. "Outlier detection for high-dimensional data". Biometrika, Vol. 102, No. 3, 2015, pp. 589-599.

[16] I. Steinwart, A. Christmann, "Support vector machines". Springer Science & Business Media, 2008.

[17] Shilton, Alistair, Sutharshan Rajasegarar, and Marimuthu Palaniswami. "Combined multiclass classification and anomaly detection for large-scale wireless sensor networks". In 2013 IEEE eighth international conference on intelligent sensors, sensor networks and information processing, 2013, pp. 491-496. IEEE.

[18] Cao, Longbing. "Non-iidness learning in behavioral and social data". The Computer Journal, Vol. 57, No. 9, 2014, pp. 1358-1370.

[19] V. N. Vapnik, S. Kotz, "Estimation of dependences based on empirical data". New York: Springer-Verlag, 1982.

[20] J. Shawe-Taylor, "Classification accuracy based on observed margin". Algorithmica, Vol. 22, No. 1-2, 1998, pp. 157-172.

[21] W.R., Huang, Z., Emam, M., Goldblum, L., Fowl, J.K., Terry, F. Huang, and T., Goldstein. "Understanding generalization through visualizations". 2020, pp. 87-97.

[22] P. J. Rousseeuw, B. C. Van Zomeren, "Unmasking multivariate outliers and leverage points". Journal of the American Statistical Association, Vol. 85, No. 411, 1990, pp. 633-9.

[23] M. Hubert, S. Engelen, "Robust PCA and classification in biosciences". BIOINFORMATICS, Vol. 20, No. 11, 2004, pp. 1728-1736.

[24] M. Hubert, P. Rousseeuw, T. Verdonck, "Robust PCA for skewed data and its outlier map". Computational Statistics & Data Analysis, Vol. 53, No. 6, 2009, pp. 2264-2274.

[25] M. Mohammadi, M. Sarmad, "Outlier Detection for Support Vector Machine using Minimum Covariance Determinant Estimator". Journal of AI and Data Mining, Vol. 7, No. 2, 2019, pp. 299-309.

[26] H. P. Narkhede, "Review of image segmentation techniques". International Journal of Science and Modern Engineering, Vol. 1, No. 8, 2013, pp. 54-61.

[27] N.M. Zaitoun, and M.J., Aqel, "Survey on image segmentation techniques". Procedia Computer Science Vol. 65, 2015, pp. 797-806.

[28] D. Kaur, and K. Yadwinder. "Various image segmentation techniques: a review". International Journal of Computer Science and Mobile Computing, Vol. 3, No. 5, 2014, pp. 809-814.

[29] B. J. Ferdosi, M. M. Tarek, "Visual Verification and Analysis of Outliers Using Optimal Outlier Detection Result by Choosing Proper Algorithm and Parameter". Emerging Technologies in Data Mining and Information Security, Springer, 2019, pp. 507-517.

[30] Jakulin, A., Možina, M., Demšar, J., Bratko, I., & Zupan, B., "Nomograms for visualizing support vector machines". In Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, 2005, pp. 108-117.

[31] Poulet, F., "Towards effective visual data mining with cooperative approaches". In Visual Data Mining, Springer, Berlin, Heidelberg, 2008, pp. 389-406.

[32] Hansen, K., Baehrens, D., Schroeter, T., Rupp, M., & Mŭller, K. R., "Visual Interpretation of Kernel-based prediction models". Molecular Informatics, Vol.30, No.9, 2011, pp. 817-826.

[33] Barbosa, A., Paulovich, F. V., Paiva, A., Goldenstein, S., Petronetto, F., & Nonato, L. G., "Visualizing and interacting with kernelized data". IEEE transactions on visualization and computer graphics, Vol.22, No.3, 2015, pp. 1314-1325.

[34] Robnik-Šikonja, M., & Bohanec, M., "Perturbation-based explanations of prediction models". In Human and machine learning, Springer, Cham, 2018, pp. 159-175.

[35] Hamel, L., "Visualization of support vector machines with unsupervised learning". In 2006 IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology, 2006, pp. 1-8.

[36] Migut, M. A., Worring, M., & Veenman, C. J., "Visualizing multi-dimensional decision boundaries in 2D". Data Mining and Knowledge Discovery, Vo.29, No.1, 2015, pp. 273-295.

[37] Van Belle, V., Van Calster, B., Van Huffel, S., Suykens, J. A., & Lisboa, P., "Explaining support vector machines: a color based nomogram". PloS one, Vol.11, No.10, 2016, e0164568.

[38] France, S. L., & Akkucuk, U., "A review, framework, and R toolkit for exploring, evaluating, and comparing visualization methods". The Visual Computer, 2020, pp.1-19.

[39] Wang, X., Wu, S., Wang, X., & Li, Q., "SVMV–a novel algorithm for the visualization of SVM classification results". In International Symposium on Neural Networks, Springer, Berlin, Heidelberg, 2006, pp. 968-973.

[40] Caragea, D., Cook, D., Wickham, H., & Honavar, V., "Visual methods for examining svm classifiers". In Visual Data Mining, Springer, Berlin, Heidelberg, 2008, pp. 136-153.

[41] Ma, Y., Chen, W., Ma, X., Xu, J., Huang, X., Maciejewski, R., & Tung, A. K., "EasySVM: A visual analysis approach for open-box support vector machines". Computational Visual Media, Vol. 3, No.2, 2017, pp. 161-175.

[42] M. Onderwater, "Outlier preservation by dimensionality reduction techniques". IJDATS, Vol. 7, No. 3, 2015, pp. 231-252.

[43] Q. Ye, W. Zhi, "Outlier detection in the framework of dimensionality reduction". International Journal of Pattern Recognition and Artificial Intelligence, Vol. 29, No. ,4 2015, pp. 1550017.

[44] W. A. Stahel, "Robuste schatzungen: infinitisimale optimalitat und schatzunguen von kovarianzmatrizen". Eidgenossischent Technischen Hochschule, 1981.

[45] D. L. Donoho, "Breakdown properties of multivariate location estimators". Technical report, Harvard University, Boston. URL http://www-stat. stanford. edu/ donoho/Reports/Oldies/BPMLE. pdf, 1982.

[46] Y. Zuo, S. Lai, "Exact computation of bivariate projection depth and the Stahel–Donoho estimator". Computational Statistics & Data Analysis, Vol. 55, No. 3, 2011, pp. 1173-9.

[47] R. A. Maronna, V. J. Yohai, "The behavior of the Stahel-Donoho robust multivariate estimator". Journal of the American Statistical Association, Vol. 90, No. 429, 1995, pp. 330-41.

[48] Van Aelst, Stefan, Ellen Vandervieren, and Gert Willems. "A Stahel–Donoho estimator based on huberized outlyingness." Computational Statistics & Data Analysis, Vol. 56, No. 3, 2012, pp. 531-542.

[49] J. Kittler, J. Illingworth, "Minimum error thresholding". Pattern recognition, Vol. 19, No. 1, 1986, pp. 41-7.