International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 21(2), 2021, pp 45-60

Robustification of Quadratic Discriminant Analyzer and Its Application for Gene Expression Data Analysis

Md. Matiur Rahaman[1] and Md. Nurul Haque Mollah[2, *]

[1]Department of Statistics, Faculty of Science, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj-8100, Bangladesh

[2]Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh

[*]Correspondence should be addressed to Md. Nurul Haque Mollah

(Email: mollah.stat.bio@ru.ac.bd)

[Received July 14, 2021; Revised August 20, 2021; Accepted September 7, 2021]

Abstract

There are some robust quadratic discriminant analyzers (QDAs) in the literature based on the robust estimation of location and scatter matrix. In this paper, we showed that their performance not only depends on the robust estimation of mean vector and covariance matrix, but also depends on the diagnosis of query objects. To show it, we compared the performance of our proposed β-divergence based quadratic discriminant analyzer (QDA) with three popular robust QDA based on Minimum Volume Ellipsoid (MVE), Minimum Covariance Determinant (MCD) and Orthogonalized Gnanadesikan-Kettenring (OGK) estimators of mean vector and covariance matrix. The performance of robust QDA was investigated using both simulated and real gene expression data analysis. We observed that proposed robust β-divergence based QDA improves the performance over the traditional QDA as well as the MVE, MCD and OGK estimators based robust QDA. The performance of proposed QDA depends on the value of the tuning parameter β. It reduces to the traditional QDA, when $\beta \to 0$.

Keywords: Classification, Quadratic discriminant analyzers (QDA), minimum β-divergence estimators, gene expression data, robustness.

AMS Subject Classification: 62H30.

1. Introduction

Classification is a supervised learning approach concerned with separating distinct sets of objects and with allocating new objects to previously defined groups. It has important role in data mining, pattern recognition and life science research. In the literature, there are several approaches for classifications, where different variants

of Bayes classifiers including linear discriminant analyzer (LDA), quadratic discriminant analyzer (QDA) and Naïve Bayes classifiers are widely using as popular statistical classifiers (Anderson, 2003; Johnson and Wichern, 2007, Ahmad et al. 2017; Rahaman and Mollah, 2019). However, most of them produce misleading results in presence of outlying observations. To overcome this problem, Ahmed et al. (Ahmed et al. 2017) robustify Naïve Bayes Classifier and Rahaman and Mollah (Rahaman and Mollah, 2019) robustify Bayesian LDA for robust classification. But both LDA and Naïve Bayes have some other drawbacks. For example, LDA is not suitable when population variance-covariance matrices are unequal and Naïve Bayes is not appropriate when feature variables are correlated. However, in both cases, QDA is appropriate and it is widely using for gene expression analysis (Arevalillo and Navarro, 2011, Asyali et al., 2006, Koçhan, et al., 2019; Soukup and Lee, 2004; Zararsiz, et al., 2015). Cancer/tumor patients classification or gene classification based on gene expression on normal cell and cancer cell playing the important role in biological research (Veer *et al*., 2002; Wuju and Momiao, 2002; Wright et al., 2003; Wang et al., 2008). But, classical QDA is very much sensitive to outliers and gene expression data are often contaminated by outlying observations, since several steps involve during the sample collection to image processing of gene expression. In this article, we considered to discuss the minimum $\beta-$divergence method robust QDA (Rahaman and Mollah, 2019, Mollah et al., 2007, 2010). Before going to the robust QDA, let us first introduce classical QDA under multivariate normal populations, let Π_k denote kth multivariate normal population with density functions $\boldsymbol{f_k(y) = N(y|m_k, V_k)}$, where $\boldsymbol{m}_k$ is the mean vector for kth normal population ($k = 1, \ldots, p$) and $\boldsymbol{V_k}$ is the covariance matrix of order $t \times t$ for all p normal populations, where $\boldsymbol{V_1 \neq V_2 \neq \ldots \neq V_p}$. Suppose a training sample or dataset $\{\boldsymbol{y}_{(k)}^{\boldsymbol{i}}\}; i = 1, 2, \ldots, N_k$ obtained from $\boldsymbol{f_k(y)}$ for $k = 1, \ldots, p$. The target is to classify an unclassified observation $\boldsymbol{y} = (y_1, y_2, \ldots, y_t)$ into one of p populations Π_k ($k = 1, 2, \ldots, p$) based on the training dataset as early considered. Obviously, the pdf of the unclassified observation $\boldsymbol{y}$ can be defined as the mixture of multivariate normal distributions as follows:

$$f(\boldsymbol{y}) = \sum_{k=1}^{p} q_k f_k(\boldsymbol{y}), \tag{1.1}$$

where q_k is the mixing proportion or prior probability of $\boldsymbol{y} \in \Pi_k$ such that $q_1+q_2+\ldots+q_p = 1$. Then the posterior pdf of $\boldsymbol{y} \in \Pi_k$ is given by

$$h(\Pi_k|\boldsymbol{y}) = \frac{q_k f_k(\boldsymbol{y})}{\sum_{l=1}^{p} q_l f_l(\boldsymbol{y})} \tag{1.2}$$

Then the classification rule for classical QDA is defined to classify an object $\boldsymbol{y}$ in population Π_k as

$$y^{(k)} = \arg \max_{j=1,2,\dots p} \delta_j(\boldsymbol{y}),\ k = 1,2,\dots,p \tag{1.3}$$

where

$$\delta_j(\boldsymbol{y}) = -\frac{1}{2} log|V_j| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{m}_j)^T V_j^{-1}(\boldsymbol{y} - \boldsymbol{m}_j) + logq_j \tag{1.4}$$

Typically, we don't know these parameters of equation (1.3); we just have the training data. In that case we simply estimate the parameters and plug them into the rule. Let N the total number of training observations, and N_k the number of training observations in the k-th class. The following estimates are used:

$$\hat{q}_k = \frac{N_k}{N}$$

$$\hat{\boldsymbol{m}}_{\boldsymbol{k}} = \frac{1}{N_k} \sum_{i=1}^{N_k} \boldsymbol{y}_{ik}, \tag{1.5}$$

and

$$\hat{\boldsymbol{V}}_{\boldsymbol{k}} = \frac{1}{N_k} \sum_{i=1}^{N_k} (\boldsymbol{y}_{ik} - \hat{\boldsymbol{m}}_{\boldsymbol{k}})(\boldsymbol{y}_{ik} - \hat{\boldsymbol{m}}_{\boldsymbol{k}})^T \tag{1.6}$$

It is obvious from equations (1.3) that classical QDA depends on the unclassified feature vector, mean vectors and covariance matrices those are estimated by the non-robust MLEs (1.5 & 1.6) based on the training dataset. Therefore, traditional QDA produce misleading results in presence of outliers in the training dataset or test dataset or in both datasets. To improve the results, MLEs can be replaced by the robust estimators like Minimum Volume Ellipsoid (MVE), Minimum Covariance Determinant (MCD) (Todorov et al. 1990, 1994), or Orthogonalized Gnanadesikan-Kettering (OGK) (Maronna and Zamar 2002). These three popular robust estimation procedures for the location and scatter matrix were also utilized in an r-package for multivariate analysis (Todorov and Filzmoser, 2009). In the case of high dimensional dataset the performance of these robust procedures is not good enough (Rahaman and Mollah, 2019). These estimators may not control the influence of contaminated unclassified feature vector ($\boldsymbol{y}$). To overcome this problem, an attempt is made to robustify the classical QDA by the minimum $\beta-$divergence method (Rahaman and Mollah, 2019, Mollah et al., 2007, 2010).

We investigated the performance of the robust method through comprehensive simulation study (the datasets were generated from multivariate normal populations, and simulated gene expression datasets were generated from a gene expression data generating model) and a real microarray gene expression data analysis.

2. Material and Methods

2.1. Minimum β–divergence Based Quadratic Discriminant Analysis (QDA) Method (proposed)

Rahaman and Mollah (2019) robustify Gaussian Bayes classifier using β divergence method and discussed in detailed robust classification procedure for constant covariance matrices. By the several simulation study it has compared and proved that robust Gaussian Bayes classifier shows better performance when it contains constant variance matrices than inconstant variance matrices and vice versa (Rahaman and Mollah, 2019). For minimum β –divergence based robust QDA classifier parameters are obtained by the minimum β-divergence estimators $\widehat{\boldsymbol{m}}_{k,\beta}$ and $\widehat{\boldsymbol{V}}_{k,\beta}$ for the mean vector $\boldsymbol{m}_k$ and the covariance matrix $\boldsymbol{V}_k$ respectively are estimated iteratively as follows:

$$\boldsymbol{m}_{\boldsymbol{k},(\boldsymbol{r}+\boldsymbol{1})} = \frac{\sum_{i=1}^{N_k} \boldsymbol{\omega}_{\boldsymbol{k},\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{y}_{\boldsymbol{ik}}}{\sum_{i=1}^{N_k} \boldsymbol{\omega}_{\boldsymbol{k},\boldsymbol{\beta}}(\boldsymbol{y})}\,, \tag{2.1}$$

and

$$\boldsymbol{V}_{\boldsymbol{k},(\boldsymbol{r}+\boldsymbol{1})} = \frac{\sum_{i=1}^{N_k} \boldsymbol{\omega}_{\boldsymbol{k},\boldsymbol{\beta}}(\boldsymbol{y})(\boldsymbol{y}_{\boldsymbol{ik}}-\boldsymbol{m}_{\boldsymbol{k},\boldsymbol{r}})(\boldsymbol{y}_{\boldsymbol{ik}}-\boldsymbol{m}_{\boldsymbol{k},\boldsymbol{r}})^{T}}{(\boldsymbol{\beta}+\boldsymbol{1})^{-1}\sum_{i=1}^{N_k} \boldsymbol{\omega}_{\boldsymbol{k},\boldsymbol{\beta}}(\boldsymbol{y})} \tag{2.2}$$

where $\boldsymbol{\omega}_{\boldsymbol{k},\boldsymbol{\beta}}(\boldsymbol{y}_i) = exp\left\{-\frac{\beta}{2}d^2\left(\boldsymbol{y}_{ik}|\boldsymbol{m}_{k,r},\boldsymbol{V}_{k,r}\right)\right\}$, which we call β-weight function (Mollah et al., 2007, 2010) and $d\left(\boldsymbol{y}_{ik}|\boldsymbol{m}_{k,r},\boldsymbol{V}_{k,r}\right) = \sqrt{\left(\boldsymbol{y}_{ik}-\boldsymbol{m}_{\boldsymbol{k},\boldsymbol{r}}\right)^{T}\boldsymbol{V}_{k,r}{}^{-1}\left(\boldsymbol{y}_{\boldsymbol{ik}}-\boldsymbol{m}_{\boldsymbol{k},\boldsymbol{r}}\right)}$ is the Mahalanobis distance. If $V_{k,r}{}^{-1}$ does not exist, then Moore-Penrose generalized inverse of $\boldsymbol{V}_{\boldsymbol{k},\boldsymbol{r}}$ is used during iteration. If β tends to 0, then (2.1) and (2.2) reduces to the classical non-iterative estimates of mean and variance as given in (1.5) and (1.6), respectively. However, the β-weight function plays the significant role for robust QDA. The detail classification rule by the diagnosis of query objects based on β-weight function is as follows:

Step-(a): First, we have to calculate β-weight for the unclassified feature vector ($\boldsymbol{y}$) using the β-weight function

$$\boldsymbol{\omega_{k,\beta}}(\boldsymbol{y}) = exp\left\{-\frac{\beta}{2}d^2\left(\boldsymbol{y}|\widehat{\boldsymbol{m}}_{k,\beta}\ ,\widehat{\boldsymbol{V}}_{k,\beta}\right)\right\}, \tag{2.3}$$

The values of this weight function lie between 0 and 1. This weight function produces larger weight (but less than 1) if $\boldsymbol{y} \in \Pi_k$ and smaller weight (but greater than 0) if $\boldsymbol{y} \notin \Pi_k$ or contaminated by outlier. Then we construct a criterion to test of a feature or data vector whether it is contaminated or not as follows:

$$\boldsymbol{\omega_{\beta}}(\boldsymbol{y}) = \sum_{k=1}^{p} \boldsymbol{\omega_{k,\beta}}(\boldsymbol{y}) = \begin{cases} \geq \theta, & \text{if } \boldsymbol{y} \text{ is not contaminated} \\ < \theta, & \text{if } \boldsymbol{y} \text{ is contaminated} \end{cases} \tag{2.4}$$

It should be noted here that the data vector $\boldsymbol{x}$ is said to be contaminated by outliers if it does not belongs to any one of p populations. Using the proposed test criteria, we take the decision that an unclassified feature vector $\boldsymbol{x}$ is said to be contaminated by outliers if

$$\boldsymbol{\omega_{\beta}}(\boldsymbol{y}) \leq \theta,$$

where we choose the threshold value of θ by

$$\theta = (1\text{-}\xi) \min_{\boldsymbol{y}_i \in \boldsymbol{\mathcal{D}}} \boldsymbol{\omega}_{\beta}(\boldsymbol{y}_i) + \xi \max_{\boldsymbol{y_i} \in \boldsymbol{\mathcal{D}}} \boldsymbol{\omega}_{\beta}(\boldsymbol{y}_i)$$

with heuristically ξ =0.1, where, $\boldsymbol{\mathcal{D}}$ is the training dataset including the unclassified data vector $\boldsymbol{y}$. It was also used in Mollah et al. (2010) for choosing the threshold value. If the unclassified data vector $\boldsymbol{y}$ is uncontaminated by outliers, we should follow the classification rule (1.4) using the minimum β-divergence estimators $\left\{\widehat{\boldsymbol{m}}_{k,\beta}\ ,\widehat{\boldsymbol{V}}_{k,\beta}\right\}$ of $\{\boldsymbol{m_k}, \boldsymbol{V_k}\}$, here $\widehat{\boldsymbol{V}}_{k,\beta}$ is computed using equation (1.6) replacing $\boldsymbol{V_k}$ by $\widehat{\boldsymbol{V}}_{k,\beta}$. If the unclassified data vector $\boldsymbol{x}$ is contaminated by outliers, we classify it by replacing its contaminated components by their mean components as discussed in the following step-(b).

Step- (b): If the unclassified data vector $\boldsymbol{y}$ is contaminated by outliers, we calculate the absolute difference between the contaminated vector and each mean vector as

$$\boldsymbol{d}_k = abs\left(\boldsymbol{y} - \widehat{\boldsymbol{m}}_{k,\beta}\right) = (d_{k1}, d_{k2}, \ldots, d_{kt})^T, k = 1, 2, \ldots, p \tag{2.6}$$

Compute sum of the smallest r components of $\mathbf{d}_k$ as $\varphi_k = d_{k(1)} + d_{k(2)} + \ldots + d_{k(r)}$, where r=round(t/2). Then find the tentative class or population for the unclassified data vector $\boldsymbol{y}$ as

$$j = \arg \min_k \varphi_k$$

and some or all components of the unclassified contaminated data vector $\boldsymbol{y}$ corresponding to $d_{k(r+1)}, d_{k(r+2)}, \ldots, d_{k(t)}$ are assumed to be corrupted by outliers. We update $\boldsymbol{y}$ by replacing its outlying components with the corresponding mean components from the mean vector $\widehat{\boldsymbol{m}}_{\beta j}$ of jth population. Let $\boldsymbol{y}^*$ be the updated vector of the contaminated data vector $\boldsymbol{y}$. Then we use $\boldsymbol{y}^*$ instead of $\boldsymbol{y}$ for classifying $\boldsymbol{y}$. Similar steps was followed to classify contaminated data vector(s) for other robust QDA using their corresponding weight function.

3. Results and Discussion

3.1. Simulation Study with the Multivariate Normal Populations

To investigate the performance of the proposed minimum β –divergence based robust QDA in a comparison of the traditional QDA as well as other Bayesian robustified quadratic classifier based on MVE, MCD and OGK estimators, we generated several training and test datasets from m=3 multivariate normal distributions with different mean vectors ($\boldsymbol{m}_k$, k = 1, 2, 3) of length t=10 variables/features and covariance matrix $\boldsymbol{V}_k$; k = 1, 2, 3. In this simulation study, we generated N_1=40 samples from the first population, N_2=42 samples from the second population and N_3=43 samples from the third populations for both training and test datasets. The data structure as follows:

$$\pi_1 \sim 40N_{10}\,(\boldsymbol{m_1}, \boldsymbol{V_1});\ \pi_2 \sim 42N_{10}\,(\boldsymbol{m_2}, \boldsymbol{V_2}) \text{ and } \pi_3 \sim 43N_{10}\,(\boldsymbol{m_3}, \boldsymbol{V_3})$$

Then we contaminated data vectors randomly in both training and test datasets and compute the training error as well as test error for all classifiers as early mentioned including the minimum β–divergence based robust QDA. Similarly, we compute the training error as well as test error for all five classifiers based on the contaminated datasets those are obtained by changing the mean vectors with $\{(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3) + c;\ c = 1, \ldots, 9)\}$ keeping fixed other parameters as before for each dataset. For convenience of presentation, we increase each mean component from each mean vector by c. We repeated this simulation study 100 times and computed average misclassification error rate (MER) *i.e.*, training error/test error.

Figure 1A & 1B shows training error and test error, respectively for all the classifiers. It represents that the minimum β–divergence based robust QDA method has produced smaller error rates than other four methods in presence of only 5% data contamination (when both training and test datasets are contaminated). To investigate the performance of the β–divergence based robust QDA method in presence of huge data contamination, we contaminated 10% and 20% data vectors by outliers in the previous datasets, separately. Then compute the training error as well as test error for all five classifiers. Both figures (Figure 1C-D & Figure 3A-B) also showed that the proposed classifier error rate is smaller in a comparison with other four classifiers, where datasets contain 10% and 20% outliers, respectively. We also observed that, if any one of the dataset either training dataset or test dataset are contaminated by the outliers, then our proposed method also produce better results than other four methods as early discussed (Figure 2). Figure 2A & 2B presented that when only training dataset was contaminated, the MER for test dataset is same for all the robust QDA classifiers except classical QDA.

However, from Figure 2C & 2D we observed that if only test dataset is contaminated, then other methods produce much higher MER *i.e.*, prediction accuracy is very low comparatively our proposed method. Hence, our proposed methods performance is well in different scenarios of the simulation study. Figure 3C & 3D represent training and test MER with equal performance of the methods when datasets are not contaminated by the outliers. Thus, we conclude that, our proposed method perform better when datasets are contaminated. Otherwise, it keeps equal performance (Figure 3C-D).

3.2. Simulated Gene Expression Data Analysis

Let us consider a model for generating differentially expressed (DE) genes displayed in Table 1 which is also used by Nowak and Tibshirani (2008). Table 1 first column represents the gene expression of normal patient ($\boldsymbol{P_1}$) individual, second column represents the gene expression of moderate patient ($\boldsymbol{P_2}$) individual and third column represents the gene expression of severe patient ($\boldsymbol{P_3}$) individual. First row represents the gene from group *A*, second row represents the gene from group *B* and third row represents the gene from group *C*. To randomize the gene expression, Gaussian noise is added from $N(0, \sigma^2)$. First we generate a training

gene set using gene expression data generating model (Table 1) with parameters d = 2, α=0.1 and σ^2 =1, where N_1 = 10 genes

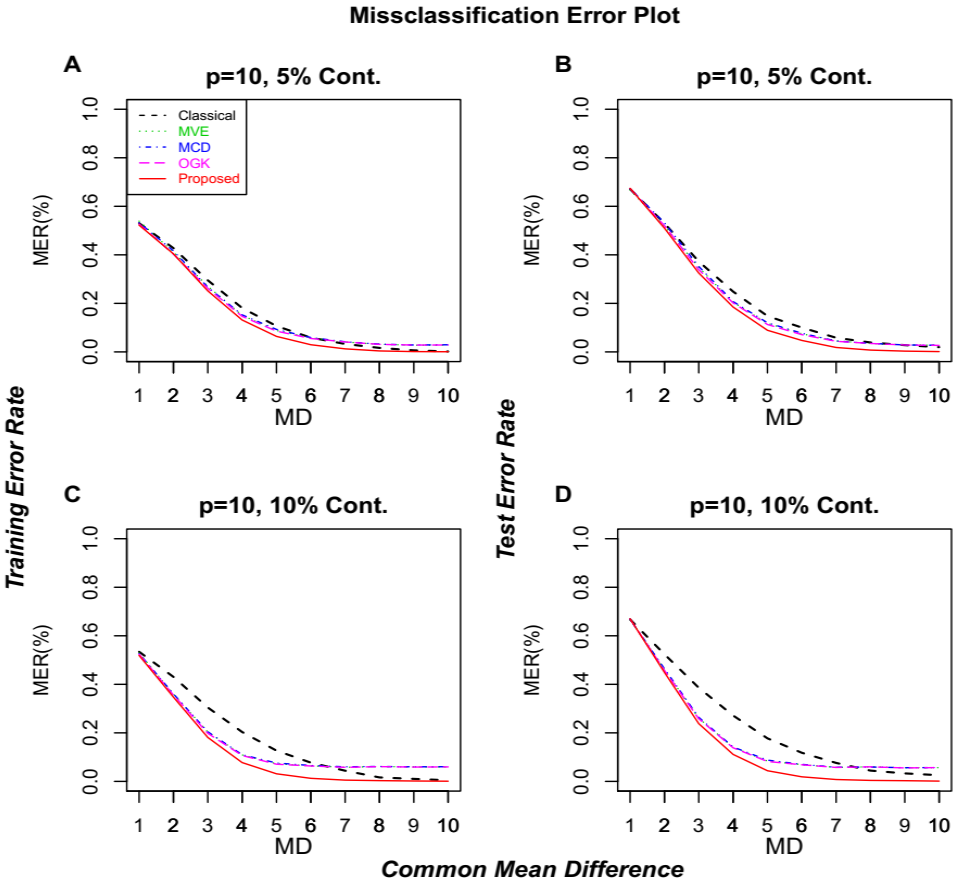


Figure 1: MER plot when both training and test datasets are contaminated. (A) Taring MER having 5 percent contamination of the data, (B) test MER having 5 percent contamination of the data, (C) training MER having 10 percent contamination of the data and (D) test MER having 10 percent contamination of the data.

denoted by $\{A1, A2, \ldots, A10\} \in A$, N_2 = 10 genes denoted by $\{B1, B2, \ldots, B10\} \in B$ and N_3 = 10 genes denoted by $\{C1, C2, \ldots, C10\} \in C$ are generated under n_1 = 5 normal individuals, n_2 = 5 patient (e.g., cancer or any other disease with moderate condition) individuals and n_3 = 5 patient (e.g., cancer or any other disease with

severe condition) individuals. Then we generate a test gene set using the gene expression data generating model in a similar way. Randomly allocated test genes set is displayed in Figure 4B.

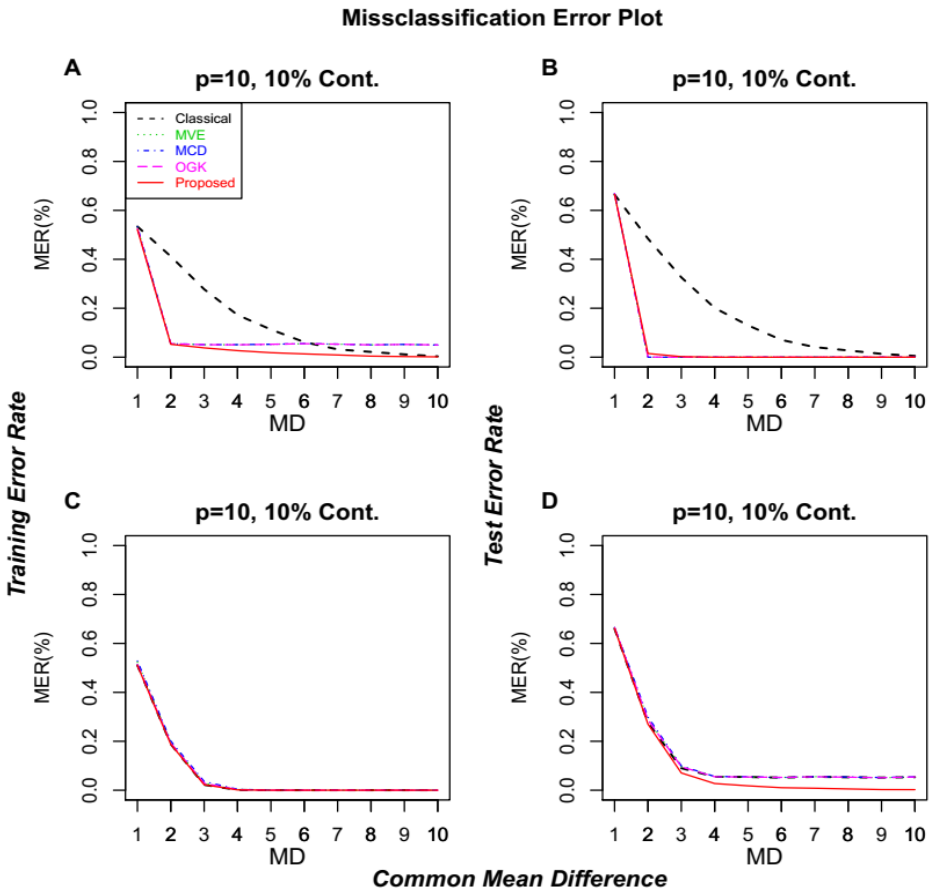


Figure 2: MER plot when any one of training or test datasets are contaminated. (A) Training MER having 10 percent contamination when only training datasets are contaminated, (B) test MER having 10 percent contamination when only training datasets are contaminated, (C) training MER having 10 percent contamination when only test datasets are contaminated and (D) test MER having 10 percent contamination when only test datasets are contaminated.

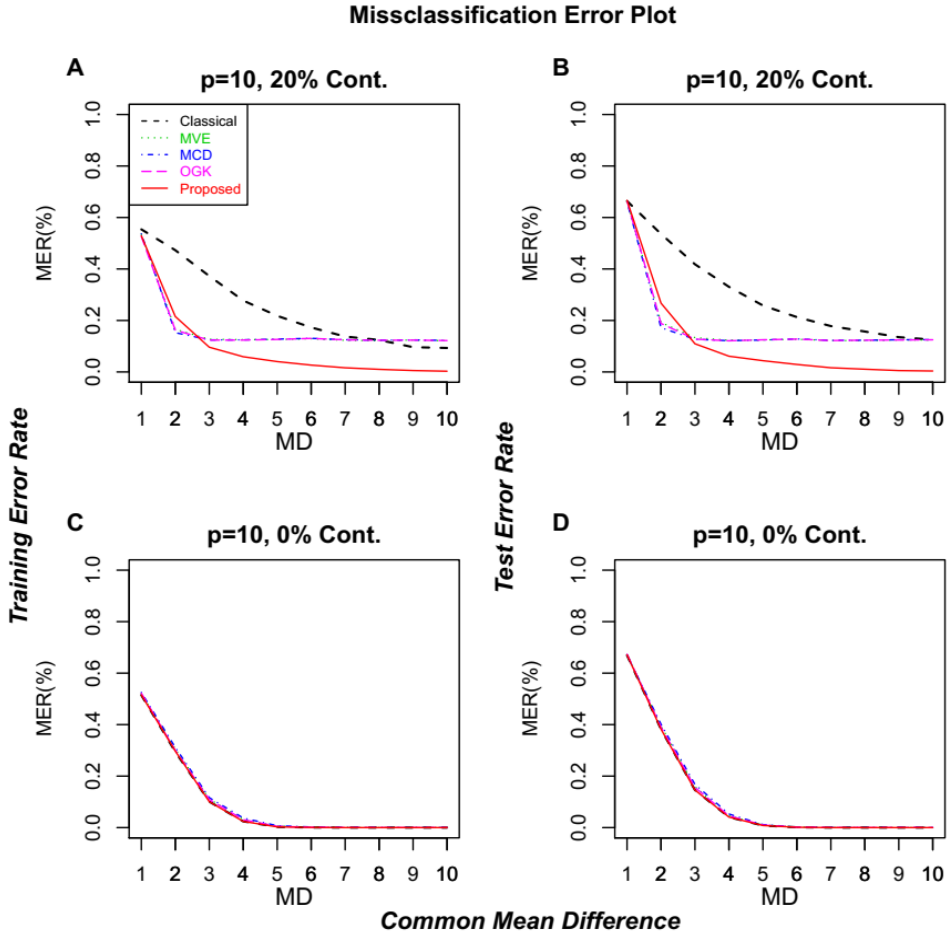


Figure 3: MER plot when 20% training and test (both) datasets are contaminated and without contamination. (A) Training MER having 20% contamination, (B) test MER having 20% contamination, (C) training MER having no contamination and (D) test MER having no contamination.

To investigate the robustness performance of the proposed method in a comparison with the classical QDA for classification, we contaminated 10% genes by outliers in both training and test gene sets (Figure 4A & 4B). To classify genes

from the contaminated test gene set, we apply both the traditional QDA procedure and proposed procedure, and the classification results are viewed in Figure 4C & 4D, respectively. We observed that MER is more than 20% with the traditional QDA procedure and unable to recover the original gene group structure (Figure 4C). Figure 4D shows the original gene group structure obtained by proposed procedure with correct classification of genes. When both training and test gene set are not contaminated by outliers, both the traditional QDA procedure and proposed robust QDA procedure has produced almost the same results (Figure 5). The appropriate value for the tuning parameter β in the proposed method is selected using cross validation (Mollah et al., 2007).

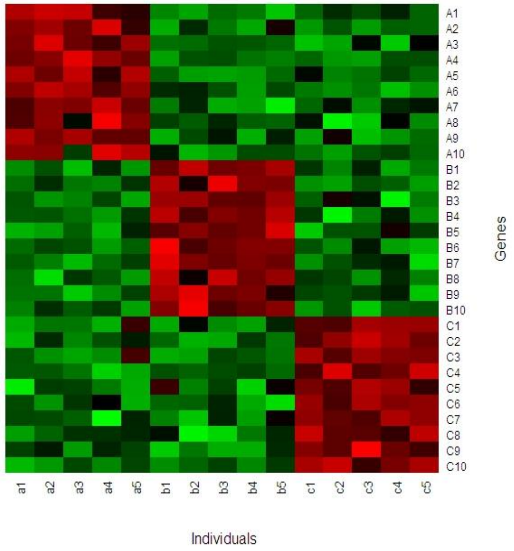
Table 1: Gene expression data generating model

Gene Group	Sample Group			
	P_1	P_2	P_3	
A	$+d$	$-d$	$-(\alpha+d)$	$+ N(0, \sigma^2+0.1)$
B	$-(\alpha+d)$	$+d$	$-d$	$+ N(0, \sigma^2+0.2)$
C	$-d$	$-(\alpha+d)$	$+d$	$+ N(0, \sigma^2+0.3)$

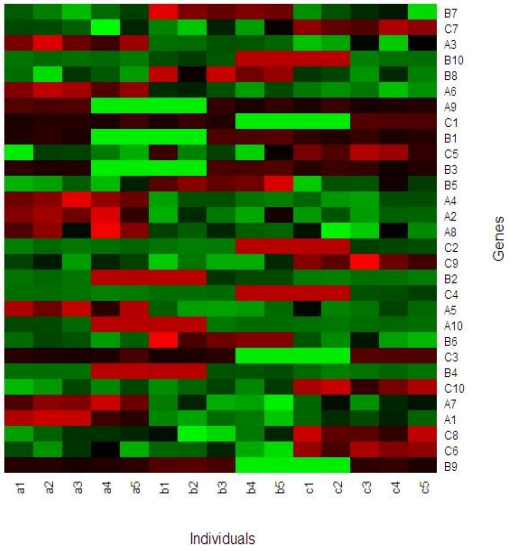
3.3. Real Gene Expression Microarray Data Analysis

We analysis gene expression data from the microarray experiments of colon tissue samples (Alon et., al. 1999). This data set contains 62 samples with 2000 genes: 40 tumor tissues, coded 2 and 22 normal tissues, coded 1. The data set is available at http://microarray.princeton.edu/oncology/. Using *t*-test we find the differentially expressed genes, and then predict patients group using classical and proposed method. We perform leave-one-out cross validation (LOOCV) technique during patient classification for both methods. Figure 6 represents the prediction accuracy of each method for patient classification. It indicates that the proposed method prediction accuracy is much better than the traditional method. Hence, we may say that our proposed method performed well than the classical method in real data analysis also.

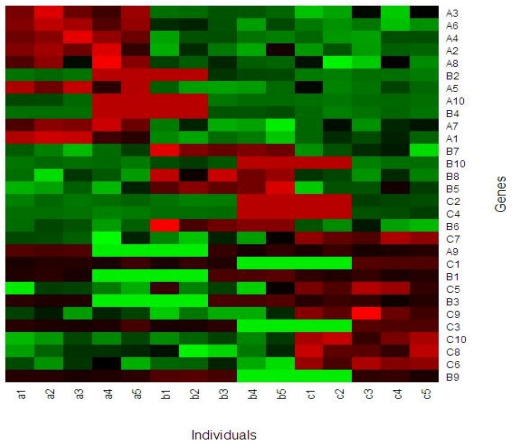
(A) Structured genes (Training dataset)



(B) Allocated genes (Test dataset)



(C) Recovered structured by Classical QDA



(D) Recovered structured by proposed method

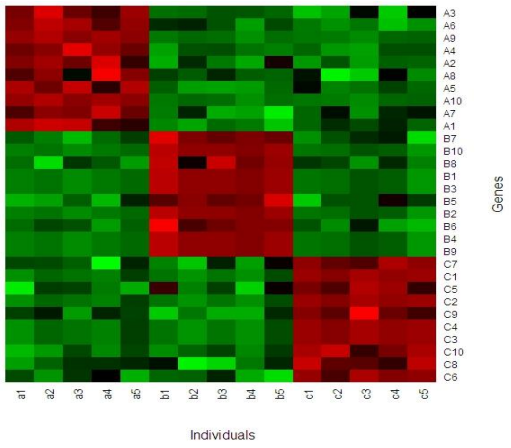
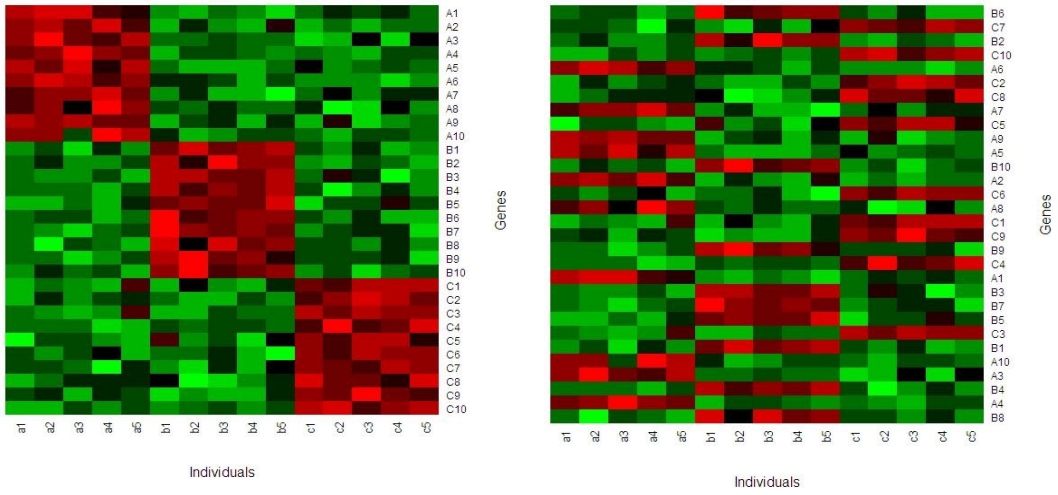


Figure 4: Gene classification in presence of outliers. (A) Training gene expression dataset with 10% contaminated genes, (B) test gene expression dataset with 10% contaminated genes, (C) classification of test gene expression dataset by classical QDA method and (D) classification of test gene expression dataset by robust QDA method (Proposed).

(A) Structured genes (Training dataset) **(B) Allocated genes (Test dataset)**



(C) Recovered structured by Classical QDA **(D) Recovered structured by proposed QDA**

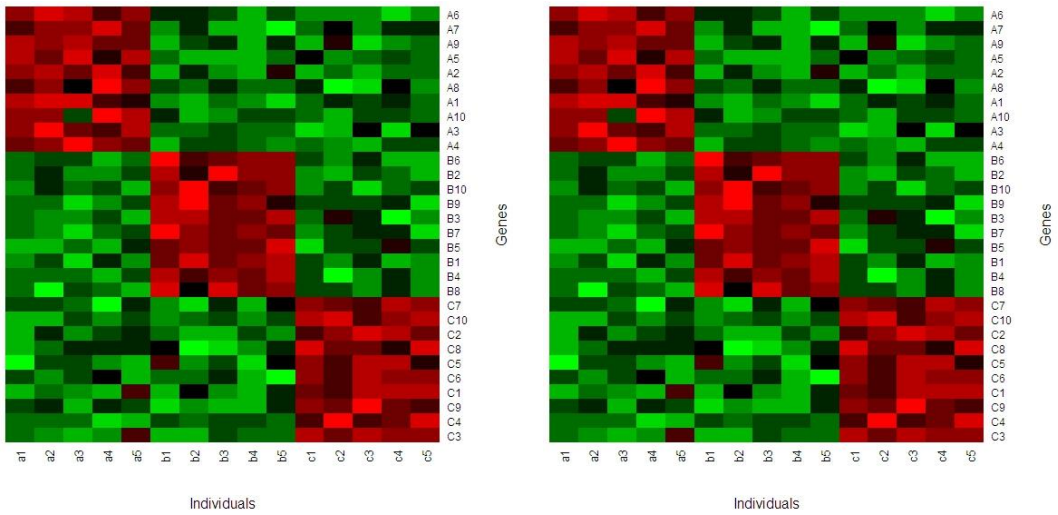


Figure 5: Gene classification in absence of outliers. (A) Training gene expression dataset with uncontaminated genes, (B) test gene expression dataset with uncontaminated genes, (C) classification of test gene expression dataset by classical QDA method and (D) classification of test gene expression dataset by robust QDA method (Proposed).

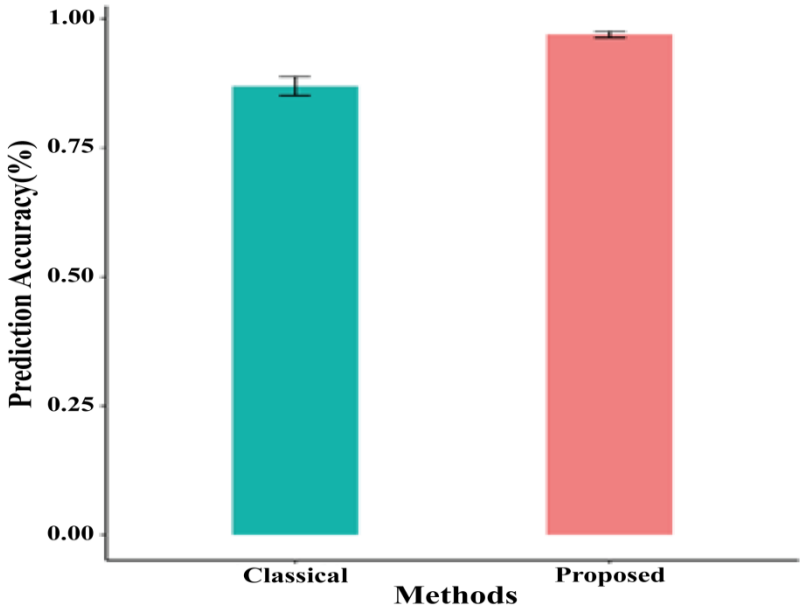


Figure 6: Performance evaluation of the proposed method for cancer prediction in comparison with classical QDA.

4. Conclusion

In this paper, we investigated the performance of our proposed robust QDA compared to the traditional QDA as well as MVE, MCD and OGK based robust QDA approaches by using the simulated datasets. The MVE, MCD and OGK based robust QDA approaches were implemented by using the *r-package* 'rrcov'. We observed that the proposed robust β-divergence based QDA improves the performance over the traditional QDA as well as the MVE, MCD and OGK estimators based robust QDA approaches. Thus, we showed that the performance of robust QDA not only depends on the robust estimation of mean vectors and covariance matrices, but also depends on the diagnosis of query objects. Finally, we demonstrated the performance of the proposed QDA in the case of gene expression data analysis and observed good performance in both presence and absence of outliers in the datasets.

References

[1] Ahmed, M. S., Shahjaman, M., Rana, M. M. and Mollah, M. N. H. (2017). Robustification of Naïve Bayes Classifier and Its Application for Microarray Gene Expression Data Analysis. BioMed Research International. Volume 2017, Article ID 3020627, 17 pages.

[2] Alon, U., Barkai, N., Notterman, D.A., Gish, K, Mack, S. Y. D and Levine.(1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. Natl. Acad. Sci. USA, 96, 6745-6750.

[3] Anderson, T. W. (2003). An Introduction to Multivariate Statistical Analysis. Wiley Interscience.

[4] Arevalillo, J. M., and Navarro, H. (2011). A new method for identifying bivariate differential expression in high dimensional microarray data using quadratic discriminant analysis. BMC Bioinformatics, 12 (Suppl 12): S6.

[5] Asyali, M. H., Colak, D., Demirkaya, O., and Mehmet S. Inan, M. S. (2006). Gene Expression Profile Classification: A Review. Current Bioinformatics, 1, 55-73.

[6] Johnson, R. A., and Wichern, D. W. (2007). Applied multivariate statistical analysis. Sixth edition, Prentice-Hall.

[7] Koçhan, N., et al. (2019). qtQDA: quantile transformed quadratic discriminant analysis for high-dimensional RNA-seq data, PeerJ, 7, e8260.

[8] Maronna, R., and Zamar, R. (2002). Robust estimation of location and dispersion for high-dimensional datasets. Technometrics, 44, 307–317

[9] Mollah, M. N. H., Minami, M. and Eguchi, S. (2007). Robust prewhitening for ICA by minimizing beta-divergence and its application to FastICA. Neural processing Letters, 25(2), 91-110.

[10] Mollah, M. N. H., Sultana, N., Minami, M. and Eguchi, S. (2010). Robust extraction of local structures by the minimum β-divergence method. Neural Networks, 23, 226-238.

[11] Nowak, G. and Tibshirani, R. (2008) Complementary Hierarchical Clustering. Biostatistics. 9, 3, 467-483.

[12] Rahaman, M. M. and Mollah, M. N. H. (2019). Robustification of Gaussian Bayes Classifier by the Minimum β-Divergence Method. Journal of Classification, 36, 113–139.

[13] Soukup, M. and Lee, J. K. (2004). Developing optimal prediction models for cancer classification using gene expression data. Journal of Bioinformatics and Computational Biology, 1, 681-694.

[14] Todorov, V., Neykov, N., Neytchev, P. (1990). Robust selection of variables in the discriminant analysis based on mve and mcd estimators. In Proceedings in computational statistics COMPSTAT. Heidelberg: Physica.

[15] Todorov, V., Neykov, N., Neytchev, P. (1994). Robust two-group discrimination by bounded influence regression. Computational Statistics and Data Analysis, 17, 289–302.

[16] Todorov, V and Filzmoser, P. (2009). An Object-Oriented Framework for Robust Multivariate Analysis. October. Journal of Statistical Software, Volume 32, Issue 3.

[17] Veer, L. J. et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature, 415, 530-536.

[18] Wang, S., Gui, j. and Li, X. (2008). Factor analysis for cross-platform tumour classification based on gene expression profiles. Journal of Circuits, Systems, and Computers, 19, 243-258.

[19] Wright G., Tan B., Rosenwald A., Hurt E., Wiestner A. and Staudt L. (2003). A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci USA, 2003, 100:9991-9996.

[20] Wuju L. and Momiao X. (2002). Tumour classification system based on gene expression profile. Bioinformatics, 18(2), 325-326.

[21] Zararsiz, G., et al. (2015). Diagonal discriminant analysis for gene-expression based tumor classification. Journal of Advances in Information Technology Vol, 6.