Robust Linear Discriminant Analysis with Highest Breakdown Point Estimator

Yai-Fung Lim, Sharipah Soaad Syed Yahaya, Hazlina Ali
School of Quantitative Sciences, UUM College Arts & Sciences, 06010 Universiti Utara Malaysia Sintok, Kedah, Malaysia
lim_yai@ahsgs.uum.edu.my

***Abstract*— Linear Discriminant Analysis (LDA) is a supervised classification technique concerned with the relationship between a categorical variable and a set of interrelated variables. The main objective of LDA is to create a rule to distinguish between populations and allocating future observations to previously defined populations. The LDA yields optimal discriminant rule between two or more groups under the assumptions of normality and homoscedasticity. Nevertheless, the classical estimates, sample mean and sample covariance matrix, are highly affected when the ideal conditions are violated. To abate these problems, a new robust LDA rule using high breakdown point estimators has been proposed in this article. A winsorized approach used to estimate the location measure while the multiplication of Spearman's rho and the rescaled median absolute deviation were used to estimate the scatter measure to replace the sample mean and sample covariance matrix, respectively. Simulation and real data study were conducted to evaluate the performance of the proposed model measured in terms of misclassification error rates. The computational results showed that the proposed LDA is always better than the classical LDA and were comparable with the existing robust LDAs.**

***Index Terms*— Linear Discriminant Analysis; Misclassification Error Rates; Robust Estimator, Winsorization.**

I. INTRODUCTION

LDA is a multivariate technique which is apt when the dependent variable is a categorical variable and the predictor variables are numerical variables. It focuses on separating distinct sets of objects into two or more groups and allocates new observations to previously defined groups. Generally, LDA is the process of constructing rules to assign a new individual observation point into one of the known populations via discriminant rules. This discriminant rules are constructed based on information (such as variables and groups) in the training data. Classification is done by allocating new observations into this discriminant rule and obtaining the group membership to which the new observation belongs [1]. A good discriminant rule is when it can provide low misclassification error rates.

The Linear Discriminant Rule (LDR) performs well for data that follow normal distribution with identical population covariance matrices. However, this rule is deemed unstable when any of these assumptions is violated [2]. This is due to the fact that the classical estimators, the mean and covariance, are known to be sensitive to deviation from the assumptions. The performance of the classical estimators can be dramatically affected if the data deviate from normality [3]. Unfortunately, ideal data set having normal distribution is hardly attainable in real life situation. To circumvent this problem, some works that are related to the robustness issues of LDA are addressed by several authors.

A number of Robust LDRs (RLDRs) which can deal with non-normality have been developed by replacing the classical mean and covariance matrix with some robust estimators of location and scatter respectively. Robust estimators such as *M*-estimators [4], *S*-estimators [5, 6, 7], Minimum Covariance Determinant (MCD) estimators [5, 8, 9], Minimum Volume Ellipsoid (MVE) estimators [10], Coordinatewise Trimming (CT) estimators [3], Feasible Solution Algorithm (FSA) [11] and automatic trimmed mean estimators [12] were used to alleviate the sensitivity problem of discrimination analysis rules. However, these robust estimators cannot guarantee the precision of the performance in all situations. Some estimators are good on certain conditions only but perform badly on other conditions.

In this paper, we propose winsorization approach paired with robust covariance matrix in an effort to create a discriminant rule that is robust to the violation of assumptions. Winsorization is a strategy that pays more attention to the central portion of a distribution by transforming the tails of the distribution [13]. This winsorization approach was chosen based on their great performance in constructing robust Hotelling's T^2 control chart [14]. This paper is the extended study from Lim, Syed Yahaya and Ali [15], which considers RLDR in higher dimension and compare the performance (misclassification error rates) with RLDR using *S*-estimator as well as MCD estimators.

The performance of the proposed RLDR was observed through simulation and real data. A comparison among the classical LDR, existing RLDR with *S*-estimators, existing RLDR with MCD estimators and proposed RLDR was done to evaluate the classification efficiency of these rules. This study focuses on two-group discrimination problem with particular interest in the influence of outliers towards classification error.

This paper is structured as follows. Section II contains a brief review of classical LDR and proposed RLDR. Then simulation and real data study are described and presented in Section III and Section IV, respectively. Finally, the conclusion is provided in Section V.

II. LDR

A brief description of a statistical discriminant analysis problem is presented in this section. In a two-group discrimination problem, suppose that n observations of a training data with d-dimensional features where the n observations are obtained from two different populations, π_1 and π_2, with the corresponding sample sizes, n_1 and n_2. The Classical LDR (CLDR) with plug in method is given in equation (1) [16].

$$\begin{aligned} &\text{If} \quad (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\, \boldsymbol{\Sigma}^{-1}\left[\mathbf{x}_0-\frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)\right] \geq \ln\left(\frac{p_2}{p_1}\right) \\ &\text{then} \quad \mathbf{x}_0 \in \pi_1, \\ &\text{else} \quad \mathbf{x}_0 \in \pi_2. \end{aligned} \tag{1}$$

where p_1 and p_2 are the prior probability that an individual comes from population π_1 and π_2 respectively. This CLDR is built to be optimal in classifying the new observation $\mathbf{x}_0$ under the assumptions that π_1 and π_2 are both multivariate normal distributions with different location but equal covariance matrices [2]. In particular, π_1 and π_2 are $N_d(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $N_d(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ respectively and under the assumption that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$. It is a known fact that this CLDR is not robust. Glèlè Kakaï, Pelz and Palm (2010) proved that non-normality and/or heteroscedasticity will negatively impacted the performance of the CLDR [17].

As a solution to the sensitivity of the LDA, a RLDR is constructed using robust estimators with highest breakdown point. The proposed RLDR used the winsorization approach to obtain robust location measure and then paired with the robust covariance matrix. A robust location estimator namely Winsorized Modified One-step *M*-estimators (WMOM), is proposed in this study. Basically, WMOM follows an automatic trimming approach which takes into consideration the shape of data distribution during the trimming process. This location estimator gives more attention to the centre rather than weighted in the tails of a data distribution. Only outliers will be trimmed away through this automatic trimming approach [12]. However, the trimmed values will be then replaced by the lowest and highest remaining data, rather than just omitting them. The problem of losing information due to trimming process can be reduced since winsorization always retain the original sample size. WMOM estimate of location for each population can be defined as equation (2).

$$\hat{\boldsymbol{\mu}}_k = \sum_{i=1}^{n_{jk}} \frac{W_{ij}}{n_{jk}} \qquad j = 1, \ldots, d; \quad k = 1, 2 \tag{2}$$

where W_{ij} is the winsorization of a random sample. The construction of the winsorized sample is based on an automatic trimmed mean proposed by Wilcox and Kelseman [18], and then follows the winsorization process introduced by Wilcox [13]. The detail equation of automatic trimmed mean can be referred in Syed Yahaya et al. [12]. Meanwhile, the covariance is replaced by the multiplication of spearman correlation coefficients (ρ) and rescaled median absolute deviation ($MADn$). The robust covariance matrix is represented by equation (3).

$$\hat{\boldsymbol{\Sigma}}_k = \begin{bmatrix} MADn_{1k}^2 & \cdots & \rho_{1dk} MADn_{1dk} \\ \rho_{21k} MADn_{21k} & & \rho_{2dk} MADn_{2dk} \\ \vdots & \ddots & \vdots \\ \rho_{d1k} MADn_{d1k} & \cdots & MADn_{dk}^2 \end{bmatrix} \tag{3}$$

where

$$MADn_{jk} = 1.4826 \text{ Median } \left\{ \left| x_{(1)jk} - \hat{M}_{jk} \right|, \ldots, \left| x_{(n)jk} - \hat{M}_{jk} \right| \right\}$$

The robust location (2) and robust covariance matrix (3) will replace the classical mean, $\boldsymbol{\mu}$ and covariance matrix, $\boldsymbol{\Sigma}$, to form a new robust discriminant rule denoted as RLDR_W.

III. SIMULATION STUDY

Since the main assumptions of LDA are normality and homoscedasticity, therefore manipulating a few variables that would likely influence the two assumptions is a good way to investigate on the optimality of the proposed RLDR against the CLDR and the existing RLDRs. The performance in terms of misclassification error rate for the proposed RLDR_W was assessed on various simulation settings and compared to CLDR, RLDR with *S*-estimators and RLDR with MCD estimators. Various conditions generated from manipulating the variables, which are deemed capable of highlighting the strengths and weaknesses of the discriminant rule are presented in Table 1. The training data were generated from normal distribution, but differ in the means and the shape of group populations. The data were contaminated as in equation (4).

$$\begin{aligned} \pi_1 &: (1-\varepsilon) n_1 N_d(0, I_d) + \varepsilon\, n_1 N_d(0+\mu, \kappa I_d) \\ \pi_2 &: (1-\varepsilon) n_2 N_d(1, I_d) + \varepsilon\, n_2 N_d(1-\mu, \kappa I_d) \end{aligned} \tag{4}$$

The combination of various variable settings produced 612 different data distributions (36 uncontaminated, 144 location contamination, 144 shape contamination and 288 location and shape contamination). The simulation started off by generating training samples of the given sizes which were used to formulate the discrimination rules. Testing sample of size 2000 from each uncontaminated population was then generated and the misclassification error rates determined by calculating the proportion of misclassified testing sample observations in each population. This process was repeated 2000 times for each condition.

Table 1
Simulation settings

Variable Settings	Parameters
Dimensions, d	2, 6, 10
Percentage of Contamination, ε	0, 0.1, 0.2
Sample Size of Training Data, (n_1, n_2)	(20,20), (50,50), (100, 100)
Shift in Location of the Populations, μ	0, 3, 5
Shift in Shape of the Populations, κ	0, 9, 25

Figure 1 presents the average of the misclassification error rates for each rule under the clean (uncontaminated) data. The misclassification error rates for each rule seem to decrease when the dimensional of variables as well as the sample sizes increase. In short, all the LDR perform equally well but CLDR always provide the lowest misclassification error rates in the case of clean data, such that $\varepsilon = 0$, $\mu = 0$ and $\kappa = 0$. These results concur with the theory that the optimality of CLDR can be guaranteed once all the assumptions of LDA are met. Figure 1 reveals that all the RLDRs closely follow CLDR under different dimensions and sample sizes. Moreover, the misclassification error rates of the proposed RLDR_W are almost overlapping the CLDR across various dimensions and sample sizes. The average misclassification error rates for each LDR with various simulation conditions are recorded in Table 2 – Table 4

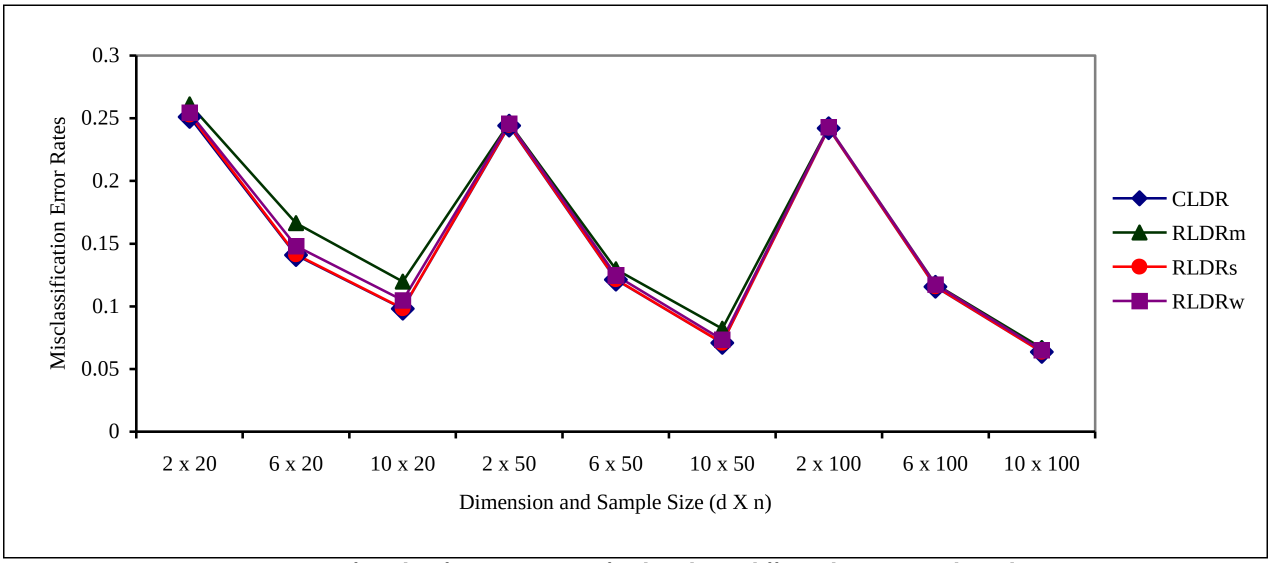


Figure 1: Average of misclassification error rates for clean data in different dimensions and sample sizes

Table 2
Average of the misclassification error rates for various linear discriminant rules with *d=2*.

ε	μ	κ	$n_1 = 20$ $n_2 = 20$				$n_1 = 50$ $n_2 = 50$				$n_1 = 100$ $n_2 = 100$			
			CLDR	**$RLDR_M$**	**$RLDR_S$**	**$RLDR_W$**	**CLDR**	**$RLDR_M$**	**$RLDR_S$**	**$RLDR_W$**	**CLDR**	**$RLDR_M$**	**$RLDR_S$**	**$RLDR_W$**
0.1	3	0	0.3389	0.2621	0.2670	0.2866	0.2960	0.2468	0.2508	0.2646	0.2741	0.2431	0.2455	0.2542
0.1	5	0	0.4987	0.2577	0.2539	0.2862	0.4986	0.2455	0.2449	0.2658	0.5010	0.2425	0.2424	0.2566
0.1	0	9	0.3178	0.2579	0.2533	0.2579	0.2759	0.2457	0.2449	0.2472	0.2587	0.2426	0.2423	0.2438
0.1	0	25	0.4205	0.2577	0.2530	0.2579	0.3863	0.2456	0.2448	0.2474	0.3447	0.2425	0.2424	0.2439
0.1	3	9	0.3884	0.2578	0.2532	0.2602	0.3610	0.2456	0.2449	0.2487	0.3270	0.2426	0.2425	0.2446
0.1	3	25	0.4527	0.2577	0.2529	0.2587	0.4441	0.2456	0.2448	0.2479	0.4234	0.2425	0.2423	0.2441
0.1	5	9	0.4548	0.2581	0.2535	0.2631	0.4732	0.2457	0.2449	0.2502	0.4804	0.2426	0.2424	0.2455
0.1	5	25	0.4755	0.2577	0.2529	0.2593	0.4870	0.2456	0.2448	0.2483	0.4917	0.2425	0.2423	0.2444
0.2	3	0	0.5770	0.2870	0.4409	0.4753	0.6202	0.2615	0.4011	0.5297	0.6542	0.2504	0.3698	0.5772
0.2	5	0	0.6530	0.2562	0.5471	0.4442	0.6911	0.2455	0.5870	0.5179	0.7124	0.2425	0.6150	0.6010
0.2	0	9	0.3624	0.2560	0.2584	0.2628	0.3055	0.2459	0.2468	0.2499	0.2745	0.2428	0.2433	0.2451
0.2	0	25	0.4637	0.2552	0.2547	0.2622	0.4277	0.2455	0.2457	0.2499	0.3929	0.2425	0.2427	0.2454
0.2	3	9	0.5083	0.2559	0.2590	0.2735	0.5334	0.2459	0.2471	0.2561	0.5678	0.2427	0.2434	0.2489
0.2	3	25	0.5041	0.2550	0.2551	0.2652	0.5062	0.2455	0.2457	0.2515	0.5237	0.2425	0.2427	0.2461
0.2	5	9	0.6039	0.2561	0.2597	0.2865	0.6795	0.2458	0.2474	0.2665	0.7158	0.2427	0.2435	0.2565
0.2	5	25	0.5310	0.2551	0.2550	0.2678	0.5590	0.2454	0.2456	0.2530	0.6061	0.2426	0.2427	0.2469

Unlike in the case of clean data, the misclassification error rates of CLDR inflate considerably above the other RLDRs when contamination occurs. Across Table 2 to Table 4, the result reveals that misclassification error rates have negative relationship with dimensional (d) of variables under most of the simulation conditions. As the number of variables increase, the misclassification error rates decrease. The performance of the robust rules; $RLDR_M$, $RLDR_S$ and $RLDR_W$, are directly affected by the sample sizes. The performance of robust rules improves when sample sizes of training data increase. Nonetheless, this pattern changes when the shift in location for groups differ especially under 20% contamination. The misclassification error rates for $RLDR_W$ ranging from 6.66% to 60.10% as compared to $RLDR_M$ (6.59% to 52.80%) and $RLDR_S$ (6.42% to 62.70%). Although the range for the proposed $RLDR_W$ is wider than $RLDR_M$ but it is narrower than $RLDR_S$, not to mention the range for the CLDR is 10.78% to 76.69%.

At d = 2 and ε = 0.1, although $RLDR_M$ and $RLDR_S$ seems to perform better than $RLDR_W$ irrespective of shift in location and/or shape, the disparities among these robust rules are quite small, that is not more than 0.04. When number of variables d increases to 6, combined with shift in location, the performance for $RLDR_M$ is the best followed by $RLDR_W$ and $RLDR_S$. Overall, $RLDR_M$ produces constant misclassification error rates irrespective to the contamination percentage, shift in location and/or shape for $n_1 = n_2 = 50$, 100. Besides, $RLDR_S$ also yields almost constant discrimination performance at $n_1 = n_2 = 50$, 100 with 10% contamination, but not including the cases of location contamination. This pattern continues at d = 10 for both $RLDR_M$ and $RLDR_S$. Meanwhile, $RLDR_W$ performs well with smaller misclassification error rates at most of the conditions for small sample sizes, $n_1 = n_2 = 20$ with d = 10 and ε = 0.2.

Table 3
Average of the misclassification error rates for various linear discriminant rules with d = 6

ε	μ	κ	$n_1 = 20$ $n_2 = 20$				$n_1 = 50$ $n_2 = 50$				$n_1 = 100$ $n_2 = 100$			
			CLDR	**$RLDR_M$**	**$RLDR_S$**	**$RLDR_W$**	**CLDR**	**$RLDR_M$**	**$RLDR_S$**	**$RLDR_W$**	**CLDR**	**$RLDR_M$**	**$RLDR_S$**	**$RLDR_W$**
0.1	3	0	0.3915	0.1670	0.3394	0.2733	0.3286	0.1277	0.2612	0.2123	0.2740	0.1175	0.2069	0.1759
0.1	5	0	0.4998	0.1587	0.3927	0.2758	0.5004	0.1277	0.3516	0.2184	0.4991	0.1174	0.3170	0.1855
0.1	0	9	0.2108	0.1584	0.1452	0.1529	0.1812	0.1277	0.1230	0.1276	0.1505	0.1175	0.1164	0.1189
0.1	0	25	0.2543	0.1584	0.1446	0.1535	0.2696	0.1277	0.1230	0.1280	0.2252	0.1174	0.1164	0.1192
0.1	3	9	0.2679	0.1584	0.1453	0.1631	0.2757	0.1277	0.1230	0.1338	0.2414	0.1174	0.1164	0.1224
0.1	3	25	0.2655	0.1584	0.1445	0.1557	0.3288	0.1277	0.1230	0.1298	0.3142	0.1174	0.1164	0.1201
0.1	5	9	0.3253	0.1584	0.1449	0.1754	0.3809	0.1277	0.1230	0.1412	0.4000	0.1175	0.1164	0.1267
0.1	5	25	0.2783	0.1584	0.1445	0.1581	0.3812	0.1277	0.1230	0.1313	0.4072	0.1174	0.1164	0.1210
0.2	3	0	0.5365	0.2874	0.5214	0.4659	0.5611	0.1329	0.5293	0.5070	0.5866	0.1176	0.5380	0.5399
0.2	5	0	0.5668	0.1719	0.5575	0.4436	0.6101	0.1256	0.5917	0.4896	0.6526	0.1175	0.6270	0.5459
0.2	0	9	0.2514	0.1487	0.1878	0.1603	0.1980	0.1256	0.1368	0.1321	0.1587	0.1175	0.1205	0.1212
0.2	0	25	0.3613	0.1486	0.1723	0.1607	0.3534	0.1256	0.1250	0.1327	0.2921	0.1175	0.1173	0.1218
0.2	3	9	0.3933	0.1487	0.2173	0.1842	0.4948	0.1256	0.1438	0.1507	0.5381	0.1175	0.1217	0.1330
0.2	3	25	0.4204	0.1486	0.1769	0.1657	0.4977	0.1256	0.1250	0.1366	0.5044	0.1175	0.1173	0.1242
0.2	5	9	0.4956	0.1486	0.2502	0.2167	0.6776	0.1256	0.1508	0.1805	0.7669	0.1175	0.1217	0.1554
0.2	5	25	0.4625	0.1486	0.1807	0.1711	0.5911	0.1256	0.1249	0.1407	0.6490	0.1175	0.1173	0.1266

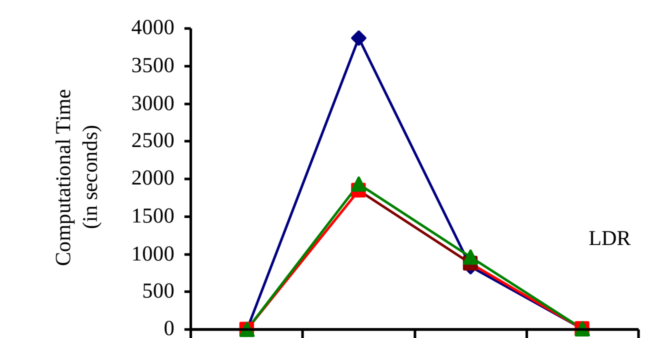
Table 4
Average of the misclassification error rates for various linear discriminant rules with d = 10

ε	μ	κ	$n_1 = 20$ $n_2 = 20$				$n_1 = 50$ $n_2 = 50$				$n_1 = 100$ $n_2 = 100$			
			CLDR	**$RLDR_M$**	**$RLDR_S$**	**$RLDR_W$**	**CLDR**	**$RLDR_M$**	**$RLDR_S$**	**$RLDR_W$**	**CLDR**	**$RLDR_M$**	**$RLDR_S$**	**$RLDR_W$**
0.1	3	0	0.4202	0.1692	0.4064	0.3042	0.3629	0.0791	0.3270	0.2214	0.3102	0.0661	0.2620	0.1675
0.1	5	0	0.4996	0.1179	0.4839	0.3076	0.5003	0.0790	0.4756	0.2321	0.4995	0.0661	0.4601	0.1829
0.1	0	9	0.1421	0.1112	0.1155	0.1089	0.1426	0.0790	0.0724	0.0765	0.1078	0.0661	0.0642	0.0666
0.1	0	25	0.1521	0.1112	0.1114	0.1095	0.2256	0.0790	0.0724	0.0769	0.1745	0.0661	0.0642	0.0668
0.1	3	9	0.1979	0.1112	0.1360	0.1256	0.2392	0.0790	0.0724	0.0856	0.2223	0.0662	0.0642	0.0720
0.1	3	25	0.1616	0.1112	0.1135	0.1132	0.2563	0.0790	0.0724	0.0789	0.2549	0.0661	0.0642	0.0681
0.1	5	9	0.2581	0.1112	0.1563	0.1460	0.3294	0.0790	0.0725	0.0982	0.3637	0.0662	0.0642	0.0799
0.1	5	25	0.1747	0.1112	0.1158	0.1171	0.2869	0.0790	0.0724	0.0810	0.3404	0.0662	0.0642	0.0694
0.2	3	0	0.5237	0.5046	0.5197	0.4643	0.5436	0.1438	0.5311	0.4967	0.5616	0.0664	0.5413	0.5231
0.2	5	0	0.5432	0.5280	0.5409	0.4431	0.5787	0.0749	0.5713	0.4777	0.6115	0.0659	0.5997	0.5220
0.2	0	9	0.1977	0.1277	0.1503	0.1175	0.1470	0.0747	0.1004	0.0806	0.1083	0.0659	0.0706	0.0692
0.2	0	25	0.2575	0.1309	0.1447	0.1180	0.2858	0.0747	0.0987	0.0814	0.2469	0.0659	0.0653	0.0698
0.2	3	9	0.3049	0.1498	0.2022	0.1535	0.4063	0.0747	0.1333	0.1067	0.4972	0.0659	0.0737	0.0865
0.2	3	25	0.2798	0.1336	0.1474	0.1252	0.4314	0.0747	0.1019	0.0866	0.4937	0.0659	0.0653	0.0731
0.2	5	9	0.3826	0.1796	0.2573	0.2003	0.5863	0.0747	0.1778	0.1501	0.7423	0.0659	0.0763	0.1219
0.2	5	25	0.3030	0.1395	0.1566	0.1338	0.5366	0.0747	0.1072	0.0920	0.6630	0.0659	0.0653	0.0770

Generally, the simulation study indicates that the robust rules outperform CLDR. The performance for all robust rules are quite equivalent except under location contamination. Their differences in misclassification error rates are small, i.e. not more than 0.08. Under the location contamination, $RLDR_M$ provides the lowest misclassification error rate which slightly outperforms $RLDR_W$. Meanwhile, $RLDR_S$ shows the worst performance among the three robust rules under the influence of location contamination.

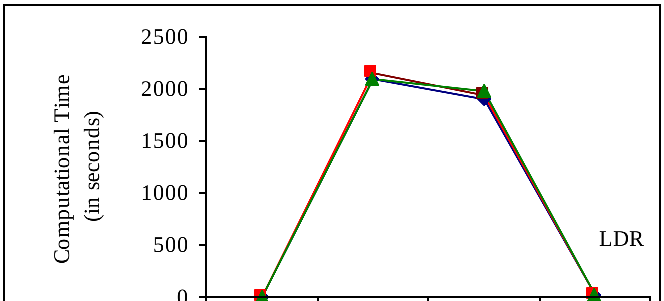
An advantage of $RLDR_W$ over $RLDR_S$ and $RLDR_M$ is the computational time. The average computational time (in seconds) for each LDR with d = 2, 6, 10 are presented in Figure 2 to Figure 4, respectively.

On average, the computing time for $RLDR_W$ is very much smaller than $RLDR_S$ and $RLDR_M$ for all the investigated conditions. Although the computational time for CLDR is the lowest among all the models, the performance in terms of misclassification error rates of CLDR is the worst when contamination occurs.



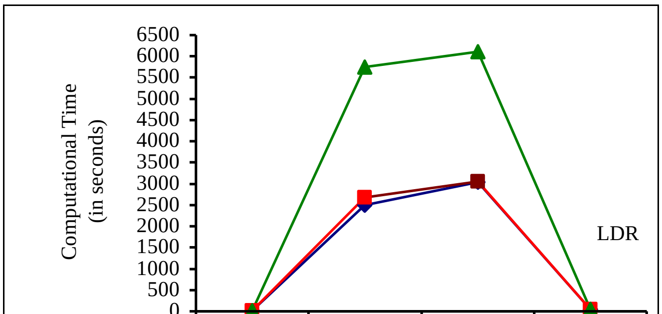
	CLDR	RLDRm	RLDRs	RLDRw
n1=n2=20	2	3871	838	6
n1=n2=50	2	1846	877	6
n1=n2=100	2	1933	964	6

Figure 2: Average computational time (in seconds) for each LDR with *d* = 2



	CLDR	RLDRm	RLDRs	RLDRw
n1=n2=20	5	2095	1907	19
n1=n2=50	5	2154	1937	20
n1=n2=100	6	2088	1988	21

Figure 3: Average computational time (in seconds) for each LDR with *d* = 6



	CLDR	RLDRm	RLDRs	RLDRw
n1=n2=20	9	2495	3046	40
n1=n2=50	9	2685	3051	41
n1=n2=100	9	5743	6105	43

Figure 4: Average computational time (in seconds) for each LDR with *d* = 10

IV. Real Data Study

Real data was also used to evaluate the optimality of the proposed $RLDR_W$. All the discriminant rules were tested using the real data items namely bank level, profit and loss of banking institutions to classify financially distressed and non-distressed banking institutions in Malaysia. The financial data were extracted from selected balance sheet in the annual report of 27 commercial banks from year 1988 to 1999. Among the 27 commercial banks, 17 of them are classified as non-distressed bank while 10 are classified as distressed bank. Two independent variables, namely ratio of total shareholder's fund to total assets (CA) and ratio of total shareholder's fund to total equity (EQ) were used to capture variation in financial crisis. Table 5 displays the results of normality test for both variables in each group and the results indicate that normality assumption is violated when *p*-value < 0.05.

Table 5
Results of the Lilliefor normality test

Group	*p*-value CA	*p*-value EQ
Distress	0.0066	0.0214
Non-distress	0.1321	0.0011

Two types of error rates, Apparent Error Rate (AER) and estimate error rate using leave-one-out Cross Validation (CV) were calculated to evaluate the performance of each rule and documented in Table 6. The real data results indicate that the proposed $RLDR_W$, existing $RLDR_S$ and $RLDR_M$ are able to detect and eliminate outliers, then provide smaller misclassification error rates as compared to the CLDR. Moreover, the proposed $RLDR_W$ and the existing $RLDR_M$ which produce similar results are found to be the best with smallest misclassification error rates via AER as well as CV. The findings from the simulation and real data study prove that the proposed $RLDR_W$ is able to provide a comparable or sometimes better performance in classification problems.

Table 6
Misclassification error rates for the each discriminant rule

Discriminant Rules	AER	CV
CLDR	0.1111	0.1111
$RLDR_S$	0.0741	0.1111
$RLDR_M$	0.0370	0.0741
$RLDR_W$	0.0370	0.0741

V. Conclusion

In this paper, winsorization approach was used to eliminate the outliers of the data and then paired with robust covariance to form a robust discriminant rule, $RLDR_W$, in order to alleviate the sensitivity problem in classification. The simulation and real data study show that the proposed $RLDR_W$ is comparable or sometimes better when compared to the existing RLDRS, not to mention the CLDR. The proposed $RLDR_W$ produces low misclassification error rates as well as computational time. Thus, the findings suggested that $RLDR_W$ can be an alternative to solve the classification problems even under the influence of non-normality and various cases of contamination in data sets. In general, this study will contribute towards knowledge development in classification problems especially when dealing with supervised data. Frequent users of LDA are aware that the LDR depends on assumption of normality. However, in the real world, data are not always normally distributed. However, with this new finding, researchers will not be constrained to the assumption of normality and can instead work with the original data without having to worry about the shape of the distributions and still be able to achieve accurate and appropriate classification rule, thus safeguarding the quality of the end results.

ACKNOWLEDGMENT

The authors would like to acknowledge the work that led to this paper, which was fully funded by the Fundamental Research Grant Scheme (S/O 12801) of Ministry of Higher Education.

REFERENCES

[1] E. A. Maharaj and A. M. Alonso, "Discriminant Analysis of Multivariate Time Series Using Wavelets," Statistics and Economies Series (03), Working Papers 12-06, 2012.

[2] C. Croux, P. Filzmoser and K. Joossen, "Classification Efficiencies for Robust Linear Discriminant Analysis," Statistica Sinica., vol. 18, no. 2, pp. 581-599, Apr. 2008.

[3] T.T. Sajobi, L.M. Lix, B.M. Dansu, W. Laverty and L. Li, "Robust Descriptive Discriminant Analysis for Repeated Measures Data," Computational Statistic and Data Analysis, vol. 56, no. 9, pp. 2782-2794, Mar. 2012.

[4] D. Wang and J. A. Romagnoli, "Robust Multi-scale Principal Components Analysis with Applications to Process Monitoring," Journal of Process Control, vol. 15, no. 8, pp. 869-882, Dec. 2005.

[5] v and Z. Omar Z, "Robust Linear Discriminant Models to solve Financial Crisis in Banking Sector," in Proceedings of the 3rd International Conference on Quantitative Sciences and Its Applications, AIP Conference Proceedings 1635, Kedah, 2014, pp: 794–798.

[6] C. Croux and C. Dehon, "Robust Linear Discriminant Analysis using S-estimators," Canad. J. Statist., vol. 29, no.3, pp. 473–493, Sept. 2001.

[7] X. He and W.K. Fung, "High Breakdown Estimation for Multiple Populations with Applications to Discriminant Analysis," J. Multivariate Anal., vol. 72, no. 2, pp. 151–162, Feb. 2000.

[8] M.J. Alrawashdeh, S.R. Muhammad Sabri and M.T. Ismail, "Robust Linear Discriminant Analysis with Financial Ratios in Special Interval," Applied Mathematical Sciences, vol. 6, no. 121, pp. 6021-6034, Jun. 2012.

[9] M. Hubert and K. Driessen, "Fast and Robust Discriminant Analysis," Comput. Statist. Data Anal., vol. 45, pp. 301–320, Mar. 2004.

[10] C.Y. Chorl and P.J. Rousseeuw, "Integrating a High-Breakdown option into Discriminant Analysis in Exploration Geochemistry," Journal of Geochemical Exploration, vol. 43, no. 3, pp. 191-203, Jun. 1992.

[11] Wina, D. E. Herwindiati and S. M. Isa, "Robust Discriminant Analysis for Classification of Remote Sensing Data," in International Conference on Advanced Computer Science and Information Systems, IEEE, Jakarta, 2014, pp: 454-458.

[12] S. S. Syed Yahaya, Y. F. Lim, H. Ali and Z. Omar, "Robust Linear Discriminant Analysis with Automatic Trimmed Mean," Journal of Telecommunication, Electronic and Computer Engineering, vol. 8, no. 10, pp. 1-3, Dec. 2016.

[13] F. S. Haddad, S. S. Syed-Yahaya and J. L. Alfaro, "Alternative Hotelling's T2 Charts using Winsorized Modified One-Step M-Estimator," Quality and Reliability Engineering International. vol. 29, pp. 583-593, Jun. 2013.

[14] F.S. Haddad, J.L. Alfaro and M. K. Alsmadi, "Hotelling's T2 Charts using Winsorized Modified One-Step M-Estimator for Individual Non Normal Data," Journal of Theoretical and Applied Information Technology, vol. 72, no. 2, pp. 215-226, Feb. 2015.

[15] Y.F. Lim, S.S. Syed Yahaya and H. Ali, "Winsorization on Linear Discriminant Analysis," in Proceedings of the 4th International Conference on Quantitative Sciences and Its Applications, AIP Conference Proceedings 1782, Kuala Lumpur, 2016, pp: 050010-1-7.

[16] R.A. Johnson and D.W. Wichern, Applied Multivariate Statistical Analysis, Prentice Hall International Edition, New Jersey, 2002, pp: 581-612.

[17] R.M. GlèlèKakaï, D. Pelz and R. Palm, "On the Efficiency of the Linear Classification Rule in Multi-group Discriminant Analysis," African Journal of Mathematics and Computer Science Research, vol. 3, no. 1, pp. 019-025, Jan. 2010.

[18] R.R. Wilcox and H.J. Keselman. "Repeated Measures ANOVA Based on a Modified One-Step M-Estimator," Journal of British Mathematical and Statistical Psychology, vol. 56, no. 1, pp. 15 – 26, May 2003.