Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

New Framework of Location Model for Mixed Variables Classification in the Presence of Outliers

Kartini Kasim[1,2,*], Hashibah Hamid[1], Ayu Abdul-Rahman[1]

[1] Department of Mathematics, School of Quantitative Sciences, UUM College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia
[2] School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Kedah Branch, Sungai Petani Campus, 08400 Sungai Petani, Kedah, Malaysia

ARTICLE INFO

Keywords:
Framework; Location model; Mixed variable classification; Outliers; Median estimator; Robust covariance

ABSTRACT

Classification that contains a mix of binary and continuous variables is a common challenge in statistical modelling. The Location Model (LM) was developed to address this challenge by discriminating groups based on continuous variables classified by multinomial cells derived from binary variables. However, a key limitation of the LM is its susceptibility to outliers, which can significantly degrade its classification accuracy, leading to high misclassification rates. To address this issue, this study aims to develop a robust framework for classifying mixed variables in the presence of outliers, with a focus on improving the LM. The developed framework introduces the Robust Location Model (RLM_{med}) based on the median, which enhances the model's resistance to outliers by employing a median estimator combined with a robust covariance matrix. The RLM_{med} framework is anticipated to outperform existing models, offering superior classification performance in datasets that contain mixed variables and are affected by outliers.

1. Introduction

Real-world applications in medical [1-4], engineering [5-8], psychology [9,10] and agriculture [11,12] serve as instances of classification studies dealing with a mixed of continuous and categorical variables. Numerous statistical techniques are employed specifically for the classification task containing mixed variables. One of the technique is the Location Model (LM), a parametric approach that effectively classifies objects consisting of mixed variable types concurrently [13-15]. The LM has demonstrated considerable success in both methodological and model development, as well as its practical application in solving the problem of classification involving a mixture of variables [16-18].

On the other hand, Olkin and Tate [19] was initially introduced the LM by deriving the mixed variables distribution. Later, Chang and Afifi [20] proposed a parametric classification approach that specifically dealt with mixed variables problems. The authors implemented the LM in the

* *Corresponding author.*
E-mail address: ms.kartini@gmail.com

https://doi.org/10.37934/araset.62.2.186193

classification model for one binary and one continuous variable. In the approach, categorical variables are treated as binary variables with values of zero or one. Consequently, Krzanowski [21] extended the model for more than two variables, followed by another generalization of mixed variables (categorical and continuous) of the two groups' problems [14-22].

Previous research has extensively explored classification procedures utilizing the LM. Table 1 provides examples of such studies, which delve into challenges including high dimensionality, variable selection, and handling a substantial number of binary variables.

Table 1
Studies conducted based on the LM

Author/Year	Description
Hamid *et al*., [23]	Managed high dimensionality problems arising from the binary variables by combining two kinds of Multiple Correspondence Analysis (MCA).
Mbina *et al*., [17]	Proposed a method for variable selection in discriminant analysis with mixed continuous and binary variables based on a criterion that permits the reduction of the variable selection problem to a problem of estimating suitable permutation and dimensionality.
Hamid *et al*., [24-25]	Managed the issue of a large number of binary variables involved in the study using a combination of two variable extraction techniques, namely Principal Component Analysis (PCA) and MCA.
Mahat *et al.,* [26]	Proposed a method for performing variable selection on the basis of distance between groups as measured by smoothed Kullback-Leibler divergence.
Franco *et al.,* [27]	The procedure is used in the context of a two-stage clustering strategy in which initial groups are first defined using hierarchical clustering. Consequently, the procedure is applied to the formed groups.

All the studies demonstrated great achievement in method and model expansion, as well as successfully solving classification problems regarding mixed variables. However, despite these successes, the classification procedure based on LM only performs well under non-contaminated data scenarios, i.e., free from outliers [26-27]. The definition of outlier has been defined by various perspectives depending on the applications. As exhibited in Table 2, the definitions highlight deviations of data from most of the observed data.

Table 2
Definition of outlier

Author/Year	Definition of outliers
Raymaekers & Rousseeuw [28]	Outliers are cases that do not obey the fit suggested by the majority of the data.
Etaga *et al*., [29]	Outliers are data values that do not fit the pattern of the data.
Okhli & Jabbari Nooghabi [30]	Outliers usually refer to extreme observations that are far from the other observed data points.
Rousseeuw & Zomeren [31]	Outliers are observations that do not follow the pattern of the majority of the data.
Hawkins [32]	Outliers are observations that deviate so much from other observations as to arouse suspicions that they were generated from different mechanisms.
Grubbs [33]	Outliers are considered as observations that are to be inconsistent.

The outliers will disrupt the parametric nature of LM, as it relies on the assumption of normality, i.e., $(y \sim MVN(\mu_{\mathrm{im}}, \sum)$. The presence of outliers in the dataset violate this assumption, affecting key parameters such as the sample mean and covariance. Under this circumstance, the outlier affects the parameters, i.e., sample mean $\left(\mu_{im}\right)$ and covariance$\left(\Sigma\right)$. Consequently, the sample mean may fail to represent the true center of the distribution, and the covariance may inaccurately capture data variability. This misrepresentation can lead to underestimated or overestimated parameter values, ultimately disrupting the normality assumption and increasing the misclassification rate, resulting in poor performance of the LM [34-36]. To address this gap, the objective of this study is to propose a new framework, the Robust Location Model (RLM_{med}), which integrates a median estimator and a robust covariance matrix to replace the classical mean and covariance in the LM, thereby mitigating the impact of outliers on parameter estimation and enhancing the model performance.

2. The Classical Location Model

In the classical LM, there are two groups denoted as π_i (*i* = 1, 2) for Groups 1 and 2, respectively, with n_i represents the group sample size. The structure of the LM is $s = 2^b$ in which s represent the number of cells and b is the number of binary variables. In the structure, the binary variables drop as cells, and each cell contains only c continuous data. Figure 1 illustrates the structure of the LM, with x and y representing binary and continuous variables, respectively. Let $b = 2,$ which then yields $s = 4$ per group.

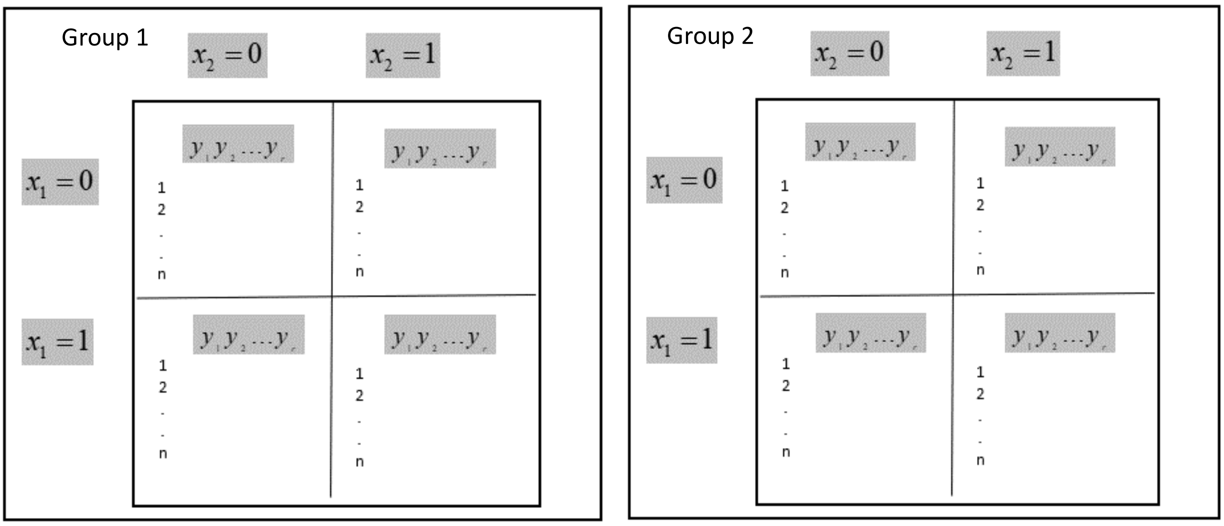

Fig. 1. The Structure of the LM when $b = 2$

The continuous data in each cell follows normal distributions with different means and equal covariance within the groups. The probability of obtaining an object in the cell m of π_i is p_{im}. All objects can be observed as a vector in the form of $z^t = \left(x^T, y^T\right)$, where $x^T = \left(x_1, x_2, \dots, x_b\right)$ is the vector of b binary variables, while $y^T = \left(y_1, y_2, \dots, y_c\right)$ is the vector of c continuous variables. Hence, the conditional distribution of vector y given x for π_i is $(y|x = m \sim MVN(\mu_{im}, \sum))$.Thus, the optimal rule of the classical LM is written as Eq. (1). The future objects, y are classified into π_1 if

$$\left(\mu_{1m} - \mu_{2m}\right)^T \sum^{-1} \left(y_r - \frac{1}{2}\left(\mu_{1m} + \mu_{2m}\right)\right) \geq \log\left(\frac{p_{2m}}{p_{1m}}\right), \tag{1}$$

otherwise, y will be classified into π_2.

In constructing the LM, the presented outliers will distort the values of parameters [35-37]. In this case, the parameters are underestimated (or overestimated), yielding inaccurate classification rules and leading to a high misclassification rate [30,34-36]. Therefore, outliers in the dataset must be tackled prior to constructing the LM to obtain accurate and optimal results.

3. RLM$_{med}$

3.1 Median and Robust Covariance as Alternative to Classical Estimator

One way to alleviate the effect of outliers is by estimating the parameters using robust estimators [38-42]. Robust estimators are used in place of classical estimators since the former is able to reduce the effect of outliers [43-46]. Furthermore, implementing robust estimators in the model helps minimize the misclassification rate [47-49]. On the other hand, Bickel [50] highlighted that medians are alternative location estimators to the classical mean. Its robustness is evident as it remains unaffected by outliers in up to 50% of the data [51-52]. Alternatively, Croux and Dehon [53] asserted that using a robust covariance matrix in classification analysis can effectively reduce misclassification rates for low and high-contaminated data. Therefore, employing a paired robust covariance and location estimator is recommended to solve the sensitivity toward outliers [54-55]. For this purpose, the mean vectors and covariance matrix in the classical LM are substituted with the median and robust covariance matrix.

3.2 New Framework of LM

In developing the new framework of LM, the outliers are managed prior to its construction. The entire procedure is structured as follows.

3.2.1 Phase I: Handling Outliers in the LM

For the first phase, the raw data are split into testing and training sets. In the testing dataset, an object, k where $k=1,2,...,n$ is omitted sequentially, while the remaining object, $n-k$ is treated as the training set. The training set is used to estimate the median and the robust covariance matrix. The median estimator are computed using the following Eq. (2).

$$\tilde{y}=\begin{cases} y_{\frac{n+1}{2}} & \text{if } n \text{ is odd} \\ \frac{1}{2}\left(y_{\left(\frac{n}{2}\right)}+y_{\left(\frac{n}{2}\right)+1}\right) & \text{if } n \text{ is even} \end{cases} \tag{2}$$

In this study, the robust covariance matrix is estimated based on the product of Spearman's correlation $\left(\rho_s\right)$ and the median absolute deviation about the median (MAD*n*). The computation of ρ_s is computed following Eq. (3).

$$\rho_s = 1-\frac{6\sum d^2}{n\left(n^2-1\right)}, \tag{3}$$

where d^2 is the square of the difference between the ranks of the r^{th} objects of y_{r1} and y_{r2}. Meanwhile, the MAD*n* is computed as in Eq. (4).

$$\text{MAD}n = h \text{ med } \left| y_r - \text{med}(y_r) \right|, \tag{4}$$

where h is a constant set at 1.4826 to make the estimator consistent and unbiased for the estimation of standard deviation under a normal distribution [56-57]. The robust covariance matrix $\left(S_R\right)$ is estimated by multiplying the ρ_s and MAD*n* as indicated in Eq. (5).

$$S_R = \rho_s \left(\text{MAD}n_1\right)\left(\text{MAD}n_2\right). \tag{5}$$

3.2.3 Phase II: Construction of new LM

In constructing the new LM, the median estimator and the robust covariance matrix are fused into the LM, in which the fusing estimator becomes a new approach. This new approach is applied to the LM, creating a new LM, RLM_{med}. The RLM_{med} is computed as indicated in Eq. (6).

$$\left(\tilde{\mathbf{y}}_{1m} - \tilde{\mathbf{y}}_{2m}\right)^T \sum\nolimits_{S_R}^{-1} \left[y - \frac{1}{2}\left(\tilde{\mathbf{y}}_{1m} + \tilde{\mathbf{y}}_{2m}\right) \right] \geq \log\left(\frac{p_{2m}}{p_{1m}}\right). \tag{6}$$

3.2.3 Phase III: Model validation

In this phase, the newly constructed model, RLM_{med}, is compared to other classification methods for validation purposes. The performance of RLM_{med} is measured using the error rate through the leave-one-out method, where the method with the lowest error is considered the best. The RLM_{med} will validated using a real dataset, i.e., an echocardiogram dataset. The echocardiogram dataset consists of 64 patients, of which 41 have not survived less than a year $\left(n_1\right)$, and the remaining being 23 $\left(n_2\right)$. This data contains nine continuous variables with two binary variables.

4. Conclusion

From this study, the findings are expected to be obtained:

i. An estimator that is able to handle outlier’s problem, which became a new approach in the LM;
ii. A new LM that can be used as an alternative to other mixed variables classification models. This new LM is expected to perform better than the classical LM.

Acknowledgement
This research was not funded by any grant.

References

[1] Balaji, K., K. Lavanya, and A. Geetha Mary. "Clustering algorithm for mixed datasets using density peaks and Self-Organizing Generative Adversarial Networks." *Chemometrics and Intelligent Laboratory Systems* 203 (2020): 104070. https://doi.org/10.1016/j.chemolab.2020.104070

[2] Hasanpour, Hesam, Ramak Ghavamizadeh Meibodi, Keivan Navi, and Sareh Asadi. "Dealing with mixed data types in the obsessive-compulsive disorder using ensemble classification." *Neurology, Psychiatry and Brain Research* 32 (2019): 77-84. https://doi.org/10.1016/j.npbr.2019.04.004

[3] Hu, Chun-Song, Qing-Hua Wu, and Da-Yi Hu. "Cardiovascular, diabetes, and cancer strips: evidences, mechanisms, and classifications." *Journal of Thoracic Disease* 6, no. 9 (2014): 1319.

[4] Esmail, Fahd Sabry, Mohamed Badr Senousey, and Mohamed Ragaie Sayed. "Selecting the best model predicting based data mining classification algorithms for leukemia disease infection." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 7, no. 1 (2017): 1-10.

[5] Zamani, Hadi, and Muhamad Kamal Mohammed Amin. "Classification of phishing websites using machine learning techniques." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 5, no. 2 (2016): 12-19.

[6] Wahid, N. S. A., P. Saad, and M. Hariharan. "Automatic infant cry classification using radial basis function network." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 4, no. 1 (2016): 12-28.

[7] Liu, Xiao-Xiao, and Yuansheng Wang. "A novel seismic risk analysis method for structures with both random and convex set mixed variables: case study of a RC bridge." *Mathematical Problems in Engineering* 2019, no. 1 (2019): 3613651. https://doi.org/10.1155/2019/3613651

[8] Yang, Jiawei, Zeping Wu, Wenjie Wang, Weihua Zhang, Hailong Zhao, and Jingbo Sun. "A surrogate-based optimization method for mixed-variable aircraft design." *Engineering Optimization* 54, no. 1 (2022): 113-133. https://doi.org/10.1080/0305215X.2020.1855156

[9] Yapan, Saadet, M. Hakan Türkçapar, and Murat Boysan. "Rumination, automatic thoughts, dysfunctional attitudes, and thought suppression as transdiagnostic factors in depression and anxiety." *Current Psychology* 41, no. 9 (2022): 5896-5912. https://doi.org/10.1007/s12144-020-01086-4

[10] Hsu, Yu Cheng, Zhiyu Ye, Lisha Dai, Yaqin Jing, Kwok-Leung Tsui, Paul SF Yip, Wentian Li, and Qingpeng Zhang. "Understanding MMPI-2 response structure between schizophrenia and healthy individuals." *Frontiers in Psychiatry* 13 (2022): 918999. https://doi.org/10.3389/fpsyt.2022.918999

[11] Mohamad, Siarudin, Ronggo Sadono, and Priyono Suryanto. "The pattern recognition of small-scale privately-owned forest in Ciamis Regency, West Java, Indonesia." *Forest and Society* 6, no. 1 (2022): 104-120. https://doi.org/10.24259/fs.v6i1.17997

[12] Apostolou, Konstantinos, Alexandra Staikou, Smaragda Sotiraki, and Marianthi Hatziioannou. "An assessment of snail-farm systems based on land use and farm components." *Animals* 11, no. 2 (2021): 272. https://doi.org/10.3390/ani11020272

[13] Hosmer, Trina, David Hosmer, and Lloyd Fisher. "A comparison of the maximum likelihood and discriminant function estimators of the coefficients of the logistic regression model for mixed continuous and discrete variables." *Communications in Statistics-Simulation and Computation* 12, no. 1 (1983): 23-43. https://doi.org/10.1080/03610918308812298

[14] Krzanowski, Wojtek J. "Mixtures of continuous and categorical variables in discriminant analysis." *Biometrics* (1980): 493-499. https://doi.org/10.2307/2530217

[15] Daudin, J. J., and A. Bar-Hen. "Selection in discriminant analysis with continuous and discrete variables." *Computational statistics & data analysis* 32, no. 2 (1999): 161-175. https://doi.org/10.1016/S0167-9473(99)00027-4

[16] Mahat, Nor Idayu, W. J. Krzanowski, and A. Hernandez. "Strategies for non-parametric smoothing of the location model in mixed-variable discriminant analysis." *Modern Applied Science* 3, no. 1 (2009): 151-163. https://doi.org/10.5539/mas.v3n1p151

[17] Mbina Mbina, Alban, Guy Martial Nkiet, and Fulgence Eyi Obiang. "Variable selection in discriminant analysis for mixed continuous-binary variables and several groups." *Advances in Data Analysis and Classification* 13 (2019): 773-795. https://doi.org/10.1007/s11634-018-0343-0

[18] Hamid, Hashibah. "A Conceptual Framework in Developing a New Location Model." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12, no. 3 (2021): 2631-2635. https://doi.org/10.17762/turcomat.v12i3.1265

[19] Olkin, Ingram, and Robert F. Tate. "Multivariate correlation models with mixed discrete and continuous variables." *The Annals of Mathematical Statistics* (1961): 448-465. https://doi.org/10.1214/aoms/1177705052

[20] Chang, P. C., and A. A. Afifi. "Classification based on dichotomous and continuous variables." *Journal of the American Statistical Association* 69, no. 346 (1974): 336-339. https://doi.org/10.1080/01621459.1974.10482949
[21] Krzanowski, Wojtek J. "Discrimination and classification using both binary and continuous variables." *Journal of the American statistical association* 70, no. 352 (1975): 782-790. https://doi.org/10.1080/01621459.1975.10480303
[22] Krzanowski, W. J. "Mixtures of continuous and categorical variables in discriminant analysis: A hypothesis-testing approach." *Biometrics* (1982): 991-1002. https://doi.org/10.2307/2529879
[23] Hamid, Hashibah, Nor Idayu Mahat, and Safwati Ibrahim. "Adaptive Variable Extractions with LDA for Classification of Mixed Variables, and Applications to Medical Data." *Journal of Information and Communication Technology* 20, no. 03 (2021): 305-327. https://doi.org/10.32890/jict2021.20.3.2
[24] Hamid, Hashibah, Nazrina Aziz, and Penny Ngu Ai Huong. "Variable extractions using principal component analysis and multiple correspondence analysis for large number of mixed variables classification problems." *Global Journal of Pure and Applied Mathematics* 12, no. 6 (2016): 5027-5038.
[25] Hashibah, Hamid. "Integrated smoothed location model and data reduction approaches for multi variables classification." PhD diss., Universiti Utara Malaysia, 2014.
[26] Mahat, Nor Idayu, Wojtek Janusz Krzanowski, and Adolfo Hernandez. "Variable selection in discriminant analysis based on the location model for mixed variables." *Advances in Data Analysis and Classification* 1 (2007): 105-122. https://doi.org/10.1007/s11634-007-0009-9
[27] Franco, Jorge, José Crossa, Suketoshi Taba, and Steve A. Eberhart. "The modified location model for classifying genetic resources: II. Unrestricted variance–covariance matrices." *Crop Science* 42, no. 5 (2002): 1727-1736. https://doi.org/10.2135/cropsci2002.1727
[28] Raymaekers, Jakob, and Peter J. Rousseeuw. "The cellwise minimum covariance determinant estimator." *Journal of the American Statistical Association* (2023): 1-12. https://doi.org/10.1080/01621459.2023.2267777
[29] Etaga Harrison, O., and Okoro Ifeanyichukwu. "Methods of estimating correlation coefficients in the presence of influential outlier (S)." *Methods* 4, no. 3 (2021): 157-185. https://doi.org/10.52589/AJMSS-LLNZXUOZ
[30] Okhli, Kheirolah, and Mehdi Jabbari Nooghabi. "On the contaminated exponential distribution: A theoretical Bayesian approach for modeling positive-valued insurance claim data with outliers." *Applied Mathematics and Computation* 392 (2021): 125712. https://doi.org/10.1016/j.amc.2020.125712
[31] Rousseeuw, Peter J., and Bert C. Van Zomeren. "Unmasking multivariate outliers and leverage points." *Journal of the American Statistical association* 85, no. 411 (1990): 633-639. https://doi.org/10.1080/01621459.1990.10474920
[32] Hawkins, Douglas M. *Identification of outliers*. Vol. 11. London: Chapman and Hall, 1980. https://doi.org/10.1007/978-94-015-3994-4
[33] Grubbs, Frank E. "Procedures for detecting outlying observations in samples." *Technometrics* 11, no. 1 (1969): 1-21. https://doi.org/10.1080/00401706.1969.10490657
[34] Hamid, Hashibah. "A Framework of Mixed Variables Classification in The Presence of Outliers: A Robust Location Model." In *AIP Conference Proceedings*, vol. 50019, no. 2017. 1905. https://doi.org/10.1063/1.5012238
[35] Hamid, Hashibah. "New location model based on automatic trimming and smoothing approaches." *Journal of Computational and Theoretical Nanoscience* 15, no. 2 (2018): 493-499. https://doi.org/10.1166/jctn.2018.7148
[36] Oyeyemi, Gafar Matanmi, George Chinanu Mbaeyi, Saheed Ishola Salawu, and Bernard Olagboyega Muse. "On discrimination procedure with mixtures of continuous and categorical variables." *Journal of Applied Statistics* 43, no. 10 (2016): 1864-1873. https://doi.org/10.1080/02664763.2015.1125859
[37] Hamid, Hashibah. "New algorithm of Location Model based on robust estimators and smoothing approach." In *2017 International Conference on Applied System Innovation (ICASI)*, pp. 1091-1093. IEEE, 2017. https://doi.org/10.1109/ICASI.2017.7988183
[38] Ghosh, Abhik, Rita SahaRay, Sayan Chakrabarty, and Sayan Bhadra. "Robust generalised quadratic discriminant analysis." *Pattern Recognition* 117 (2021): 107981. https://doi.org/10.1016/j.patcog.2021.107981
[39] Hubert, Mia, Peter J. Rousseeuw, and Tim Verdonck. "A deterministic algorithm for robust location and scatter." *Journal of Computational and Graphical Statistics* 21, no. 3 (2012): 618-637. https://doi.org/10.1080/10618600.2012.672100
[40] Moheghi, H. R., R. Noorossana, and O. Ahmadi. "Phase I and phase II analysis of linear profile monitoring using robust estimators." *Communications in Statistics-Theory and Methods* 51, no. 5 (2022): 1252-1269. https://doi.org/10.1080/03610926.2020.1758724
[41] Saraceno, Giovanni, Claudio Agostinelli, and Luca Greco. "Robust estimation for multivariate wrapped models." *Metron* 79, no. 2 (2021): 225-240. https://doi.org/10.1007/s40300-021-00214-9
[42] Vranckx, Iwein, Jakob Raymaekers, Bart De Ketelaere, Peter J. Rousseeuw, and Mia Hubert. "Real-time discriminant analysis in the presence of label and measurement noise." *Chemometrics and Intelligent Laboratory*

Systems 208 (2021): 104197. https://doi.org/10.1016/j.chemolab.2020.104197

[43] Singh, G. N., D. Bhattacharyya, and A. Bandyopadhyay. "Robust estimation strategy for handling outliers." *Communications in Statistics-Theory and Methods* 53, no. 15 (2024): 5311-5330. https://doi.org/10.1080/03610926.2023.2218567

[44] Alomair, Abdullah Mohammed, and Usman Shahzad. "Neutrosophic mean estimation of sensitive and non-sensitive variables with robust hartley–ross-type estimators." *Axioms* 12, no. 6 (2023): 578. https://doi.org/10.3390/axioms12060578

[45] Sharif, Shamshuritawati, Nuraimi Ruslan, and Tareq AM Atiany. "Type 1 Error Rate Comparison Between Classical and Modified Box M-Statistic." *Journal of Engineering and Applied Sciences* 13, no. 5 (2018): 1246-1252.

[46] Kudraszow, Nadia L., and M. Victoria Fasano. "Robust approaches to redundancy analysis." *Communications in Statistics-Theory and Methods* 53, no. 2 (2024): 607-626. https://doi.org/10.1080/03610926.2022.2087882

[47] Lim, Yai-Fung, Sharipah Soaad Syed Yahaya, and Hazlina Ali. "Robust Linear Discriminant Analysis with Highest Breakdown Point Estimator." *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* 10, no. 1-11 (2018): 7-12.

[48] Lim, Yai-Fung, Sharipah Soaad Syed Yahaya, Faoziah Idris, Hazlina Ali, and Zurni Omar. "Robust linear discriminant models to solve financial crisis in banking sectors." In *AIP Conference Proceedings*, vol. 1635, no. 1, pp. 794-798. American Institute of Physics, 2014. https://doi.org/10.1063/1.4903673

[49] Melik, Hameedah Naeem, Nor Aishah Ahad, and Sharipah Soaad Syed Yahaya. "Modified One-Step M-Estimator with Robust Scale Estimator for Multivariate Data." *Journal of Engineering and Applied Sciences* 13, no. 24 (2018): 10396-10400.

[50] Bickel, Peter J. "On some alternative estimates for shift in the p-variate one sample problem." *The Annals of Mathematical Statistics* (1964): 1079-1090. https://doi.org/10.1214/aoms/1177703266

[51] Rousseeuw, Peter J., and Christophe Croux. "Alternatives to the median absolute deviation." *Journal of the American Statistical association* 88, no. 424 (1993): 1273-1283. https://doi.org/10.1080/01621459.1993.10476408

[52] Wilcox, Rand R. *Introduction to robust estimation and hypothesis testing*. Academic press, 2011. https://doi.org/10.1016/B978-0-12-386983-8.00001-9

[53] Croux, Christophe, and Catherine Dehon. "Robust linear discriminant analysis using S-estimators." *Canadian Journal of Statistics* 29, no. 3 (2001): 473-493. https://doi.org/10.2307/3316042

[54] Herwindiati, Dyah E., and Sani M. Isa. "The new measure of robust principal component analysis." *Electronic Engineering and Computing Technology* (2010): 397-408. https://doi.org/10.1007/978-90-481-8776-8_34

[55] Pang, Y. S., N. A. Ahad, SS Syed Yahaya, and Y. F. Lim. "Robust linear discriminant rule using novel distance-based trimming procedure." *Journal of Adv Research in Dynamical & Control Systems (JARDCS)* 11 (2019): 969-978.

[56] Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical association*, *88*(424), 1273-1283. https://doi.org/10.1080/01621459.1993.10476408

[57] Rousseeuw, Peter J., and Mia Hubert. "Robust statistics for outlier detection." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1, no. 1 (2011): 73-79. https://doi.org/10.1002/widm.2