FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 3, March, 2025, pp 247 - 252
DOI: https://doi.org/10.33003/fjs-2025-0903-3245

EFFECTS OF DATA DELETION AND WEIGHTING ON FISHER'S LINEAR CLASSIFICATION METHOD: A ROBUSTIFICATION APPROACH

***[1,2]Okwonu, Friday Zinzendoff and [2,3]Ahad Aishah Nor**

[1]Department of Mathematics, Faculty of Science, Delta State University, P.M.B. 1, Abraka, Nigeria
[2]Institute of Strategic Industrial Decision Modeling, School of Quantitative Science, Universiti Utara Malaysia,06010 UUM Sintok, Kedah Malaysia
[3]School of Quantitative Sciences, College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia

*Corresponding authors' email: fokwonu@gmail.com; okwonufz@delsu.edu.ng

ABSTRACT

The classical supervised classification model's performance is hampered by the effects of influential observations (IOs). The influential observations(IO's) when deleted, weighted,Winsorized, truncated and retained have enormous effects in making replicative inferences in different classification models. Due to the influence of IO's on classical supervised classification models, different methods such as IO deletion or weighting have been introduced to reduce the influence of IO's. Some of these influential observations reduction or deletion methods have resulted in information loss of various degrees. In this study, we investigated the effects of IOs deletion and weighting using the Mahalanobis distance as a plug in to enhance the robustness of the Fisher linear classification method (FLCM). We proposed an F-weight plug in method to robustify the FLCM. We compared the performance of these methods to determine whether IO deletion or IO weighting retards or enhances the classification accuracy of the FLCM. The study affirmed that IO weighting using the F-weight minimizes information loss more than the IO deletion using the Mahalanobis distance. This study concludes that the variant of FLCM based on the F-weight method showed improved classification accuracy, and efficiency more than the Mahalanobis distance based FLCM.

Keywords: Influential observations, Mahalanobis distance, F-weight, Fisher classification, Robustness

INTRODUCTION

In practice, primary data often contains influential observations (IO). The IO may originate from different sources such as during data recording or data collections. Influential observations or outliers may hamper the performance of different classical methods. However, IO may contain useful information regarding the subject of investigation. The capability to reveal the existence of an IO in a data set plays a significant role which is the focus of robust statistics(Hubert et al., 2008; Rousseeuw & Hubert, 2018). To identify the presence of IO in a data set is unique and is the focusof robust statistical techniques which involve removing IOs from the main data collection. However, removing IO may result in loss of vital information and retaining it may also result to underperformance of the classical methods(Hubert et al., 2008; Maechler et al., 2021; Maronna et al., 2006; Nursalam, 2016 et al., 2019; Seheult et al., 1989; Tyler, 2008, Okwonu, et al.,2022; Apanapudor, et al.,2023).

When the data set is very similar in distance, the numerical values of the sample mean, and covariance matrix should be accepted. On the other hand, if the data set are far apart, the numerical value of the sample mean, and matrix should be investigated for further insight with respect to data imputation. This is because the sample mean $(\bar{x})$ is always susceptible to influential observations(Rousseeuw& Hubert, 2018). In practice, the IOs are not problematic but it may result that the IOs originated from a different class other than the class to which it been considered. IOs contain unique and useful information which might demonstrate unusual information as such it should not be deleted rather it should be accommodated by way of data transformation.

In some cases, the median which is robust with high breakdown is occasionally used as plug in for the mean to enhance robustness (Hubert &Debruyne, 2010; Rousseeuw& Hubert, 2018,Apanapudor, et al., 2020). The sample mean is not robust which reflect the zero-breakdown value and unboundedness due to its susceptibility to IOs (Hubert & Debruyne, 2009; Jennison et al., 1987; Law et al., 1986). The mean absolute deviation (MAD) is very robust like the median, both are high breakdown estimators. Different estimators such as the MCD, S and M, (minimum volume ellipsoid) MVE (Croux&Dehon, 2001; Ghosh et al., 2021; He & Fung, 2000; Lim et al., 2018;Okwonu et al., 2020; Okwonu, et al.,2021) have been applied as plug in to improve the classification performance of the Fisher linear classification method (FLCM) (Qin et al., 2020; Wang et al., 2014; Okwonu, and Othman 2013).

For the conventional statistical models, swamping and masking of the data set are often unnoticeable problems. Swamping simply means when an inlier data point is categorized as IO and masking is the inability of the statistical methods to detect IO thereby treating such data points as inliers. Over the years, researchers have advanced different ways of resolving data masking and swamping. The process of deleting and weighting IO's is a usual concept in robust statistics.

In this paper, we consider the following plug in, the median absolute deviation (MAD), enhanced median absolute deviation (EMAD), the Mahalanobis distance (MD) which is based on identifying and deleting IOs while the F-Weight (FW) is designed to transform IOs to inliers using weight concept with minimum information loss. We applied the above plug in to investigate the effects of data deletion and weighting on the FLCM classification accuracy. We also investigated the classifiers efficiency and the computational time of the FLCM variants based on the plug in. Finally, we looked at the comparative performance of the variants of the plug in applied on the Fisher's classifiers and the proposed F-weight for the FLCM

The objective of this study is to determine whether IOs deletion and weighting affect the classification performance of the Fisher's linear classification method (FLCM) using three real data set. The rest of this article is as follows. The methods are discussed the Section 2 followed by data collection, results, and discussion in Section 3. Conclusions follow in Section 4.

MATERIALS AND METHODS

Fisher linear classification method (FLCM)

The component of the FLCM coefficient is the sample mean $(\bar{x})$and covariance(S^2). These components are susceptible to a single IO which retard the performance of any statistical methods that apply $\bar{x}$ and S^2 as a plug in (Okwonu, 2012).

Let $X_{n\times p}$ be the data matrix such that the sample mean, and covariance are defined as

$$\bar{X}_i = \frac{\sum_{j=1}^{n_i} X_{i,n\times p}}{n_i} \text{ and } S_i = \frac{\sum_{j=1}^{n_i}(X_{i,n\times p}-\bar{X}_i)^2}{n_i-1} \quad (1)$$

The components of Equation (1) have zero breakdown points due to the sample mean. This implies that the minimum number of IO's could hamper it and as such would affect the classification performance of the FLCM(Okwonu & Othman, 2013; Ghosh et al., 2021). The components of Equation (1) are applied to formulate the coefficient of the FLCM. The classifier's benchmark from Equation (1) are compared to the classification score to determine the group predictions, that is,

$$w = (\bar{X}_1 - \bar{X}_2)S_p^{-1}X_i \quad (2)$$

$$\overline{w} = (\bar{X}_1 + \bar{X}_2)S_p^{-1} \quad (3)$$

From Equations (2-3), a new object can be classified to group one if

$$w \geq \overline{w} \quad (4)$$

Otherwise to group two if

$$w < \overline{w}. \quad (5)$$

Mahalanobis FLCM (M-FLCM)

From Equation (1), we can define the multivariate Mahalanobis distance as follows

$$mMD_i = SQRT\left(\left(X_{i,n\times p} - \bar{X}_i\right)' S_p^{-1}\left(X_{i,n\times p} - \bar{X}_i\right)\right) \quad (6)$$

where S_p^{-1} denote the inverse pooled sample covariance matrix. An IO can be detected if mMD_i is compared with a defined benchmark based on the degree of freedom of the Chi square, that is

$$mMD_i \geq \sqrt{\chi^2_{p,0.975}} \quad (7)$$

From Equation (7), the weighting process could be applied to delete the contribution of the data point before the sample mean and covariance matrix could be computed. When the data set are weighted, the weighted data set are applied to compute the weight sample mean and covariance matrix and subsequently to plug in into Equations (2) and (3) to obtain the M-FLCM classifiers. Therefore, a new object is assigned to group one if Equation (4) is satisfied otherwise to group two.

MAD based FLCM (MAD-FLCM)

The median absolute deviation (MAD) is a robust measure that can resist approximately 50% of IO's before it completely break down. The MAD denoted as ∂ in this paper is simply defined as

$$\partial = \emptyset \# MEDIAN\left|X_{ij} - MEDIAN\left(X_{ij}\right)\right| \quad (8)$$

where $\emptyset = 1.4826$ is the correction factor that allows this estimator to be consistent with the normal distribution (Rousseeuw& Hubert, 2018). The outputs from Equation (8) are applied to compute Equation (1) used as a plug in for Equations (2) and (3) to obtain MAD-FLCM. The decision criteria are similar to that described in Equations (4-5).

Enhanced MAD- FLCM(EMAD-FLCM)

The enhanced MAD-FLCM computational process follows a similar procedure as shown in Equation (8). However, Equation (8) is the denominator of the enhanced MAD-FLCM, and the numerator is the data deviation from the median, that is, Equation (10).

$$D_{ij} = \left(X_{ij} - MEDIAN\left(X_{ij}\right)\right) \quad (9)$$

$$W_{ij} = \frac{D_{ij}}{\partial} \quad (10)$$

Equation (10) is used to compute the components in Equation (1), the output from Equation (1) are applied as plugged in to develop Equations (2) and (3) and subsequently Equations (4) and (5).

F-Weight -FLCM(FW-FLCM)

Let $X = \sum_{j=1}^{n_i} X_{i,n\times p}, i = 1,2$ be the data set for two groups. Then the data weight Equation (11) can be described as follows

$$p_i = \frac{X_{i,n\times p}}{\sum_{i=1}^{k} X_{i,n\times p}} \quad (11)$$

From Equation (11) we can weight the data points such that the IO's is assigned the minimum weights whereas the non-IO's are assigned corresponding weight that allows the weighted data points to retain the status of inliers. Equation (12) describes this process

$$X_i = X_{ij}p_i \quad (12)$$

The component of Equation (1) can be rewritten as Equation (13) and Equation (14) as follow

$$\bar{X}_{gi} = \frac{\sum_{j=1}^{n_i} X_{ij}p_i}{n_i} \quad (13)$$

$$S_{gi} = \frac{\sum_{j=1}^{n_i}(X_{ij}p_i-\bar{X}_{gi})^2}{\sum_{i=1}^{k} n_i-k} \quad (14)$$

Therefore, the pooled sample covariance can be written as

$$S_p = \frac{\sum_{i=1}^{k}(n_i-1)S_{gi}}{\sum_{i=1}^{k} n_i-k} \quad (15)$$

Then the above Equations (12) and (15) are plugged in into Equations (2) and (3) formulate the proposed *FW- FLCM* classifier. The rules of group classification defined in Equations (4) and (5) are adopted for the proposed classifier. From the different classifiers designed based on the modified sample means and covariance matrix, the weighted data sets may reduce the influence of the IO's thereby displaying a higher breakdown value compared to the conventional sample mean and covariance matrix with zero breakdown values. In the following section, we aim to investigate the effects of data weighting on the classifier's prediction accuracy.

Evaluation criteria

To determine the performance of these classifiers, we define the probability of correct classification based on the optimal probability of correct classification $(h)$and the efficiency of the classifiers (eff)*(Okwonu, et.al., 2022, Okwonu et. Al. 2023*) as follows

$$h = 1 - q, \quad (16)$$

$$q = \left(\frac{1-\pi}{2\times\pi}\right) \times \pi \quad (17)$$

$$eff = \left[\frac{\pi}{h}\right] \times 100. \quad (18)$$

where q denotes the optimal probability of misclassification.

RESULTS AND DISCUSSION

Data Collection and Results

Let n denotes the overall sample size $(n = n_1 + n_2)$and p denotes the data dimension. Henceforth in this discussion we

will simply refer n as the sample size. The objective of this study is to investigate the classification efficiency, the real time and the CPU time of the various classifiers. The data set applied to investigate the performance of these different methods are well known data set culled from the UCI machine learning repository apart from the first data set taken from the classical example of the Salmon data (Johnson & Wichern, 1992), page 604 Chapter 11). This data consists of the growth ring diameters between Alaska and Canadian Salman classified into Male and Female groups. Each group consist of $n = 50, p = 2$ respectively. The second data set consist of the Algerian forest fire data for the Bejaia region(Abid &Izeboudjen, 2020). The data set was culled from (https://archive.ics.uci.edu/ml/machine-learning-databases/00547/). This version of the data set consists of $n = 121, p = 10$ for the two groups. The third data set consist of the Benign and Malignant culled from (ftp://ftp.cs.wisc.edu/math-prog/cpo-dataset/machine-learn/cancer/).This data consists of $n = 210, p = 30$ (https://www.openml.org/d/1510).

Table 1 and Figure 1 illustrate the different computational times of the different classifiers. Table 1 further demonstrated that the FW-FLCM achieved the best computational time followed by the FLCM classifier. Meanwhile, the MAD-FLCM and FW-FLCM achieved the best computational efficiency.

Table 1: Classification performance of different classifiers on Freshwater and marine Alaska and Canadian Salmon (Abid &Izeboudjen, 2020)

Classifiers	Accuracy (π)	Efficiency	Time (RT)	Time (CPU)
FLCM	0.9300	0.9591	0.04	0.03
M-FLCM	0.9100	0.9453	0.21	0.06
MAD-FLCM	0.9500	0.9720	0.27	0.09
EMAD-FLCM	0.9300	0.9591	0.34	0.09
FW- FLCM	0.9500	0.9720	0.02	0.00

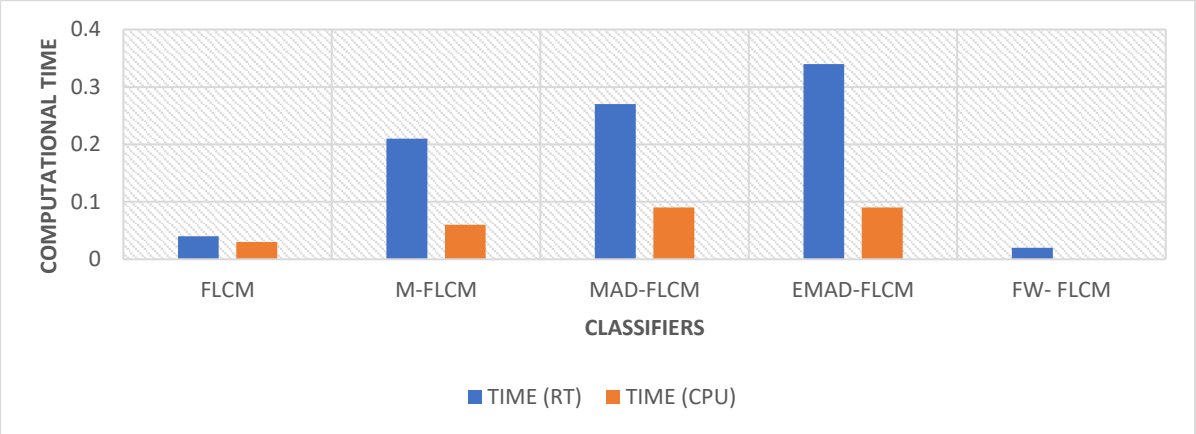

Figure 1: Comparative analysis of classifiers computational time

Table 2 present the classification results for the Algeria forest fire dataset. The results revealed that the proposed FW-FLCM outperformed the other classifiers with 100% followed by the EMAD-FLCM with 97.62% accuracy. Figure 2 illustrates the computational time of the different classifiers. The M-FLCM achieved the best computational time followed by the proposed FW-FLCM classifier.

Table 2: Classification performance on the Algeria forest fire data set(Abid &Izeboudjen, 2020)

Classifiers	Accuracy (π)	Efficiency	Time (RT)	Time (CPU)
FLCM	0.9365	0.9633	0.22	0.12
M-FLCM	0.9048	0.9415	0.06	0.04
MAD FLCM	0.9286	0.9576	0.33	0.03
EMAD -FLCM	0.9762	0.9877	0.26	0.06
FW- FLCM	1.0000	1.0000	0.07	0.04

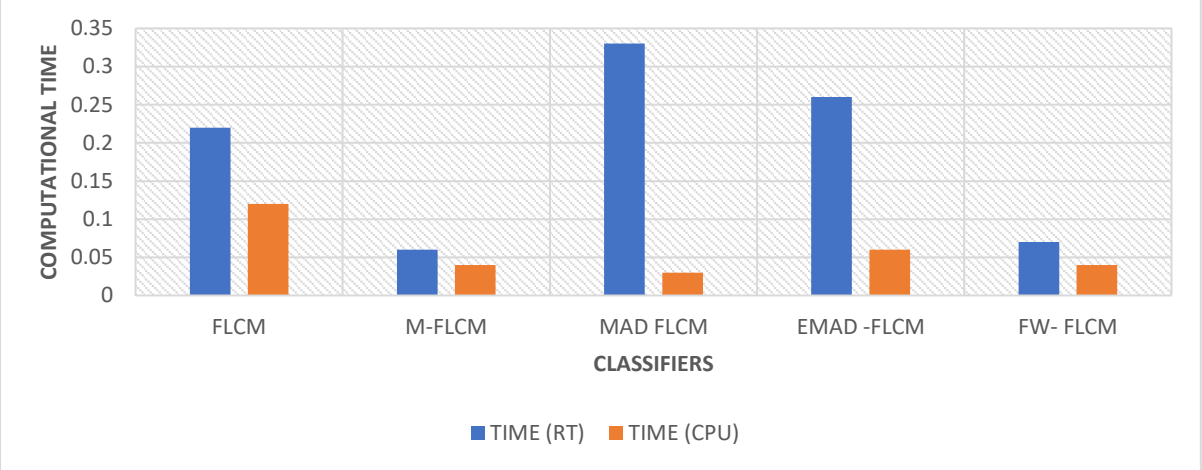

Figure 2: Comparative analysis of classifiers' computational time

Table 3 presents the classification results of the different classifiers and their computational times. For this data set, the proposed classifier's (FW-FLCM) efficiency is 100% followed by the conventional FLCM (98.51%) and the MAD-FLCM (98.30%) respectively. The M-FLCM has the best computational time followed by the FW-FLCM classifier.

Table 3: Classification performance on the Benign and Malignant data set (N=210, P=30)

Classifiers	Accuracy (π)	Efficiency	Time (RT)	Time (CPU)
FLCM	0.9714	0.9851	0.22	0.18
M-FLCM	0.8524	0.9096	0.11	0.01
MAD FLCM	0.9667	0.9830	0.27	0.18
EMAD -FLCM	0.9452	0.9703	0.21	0.17
FW- FLCM	1.0000	1.0000	0.18	0.12

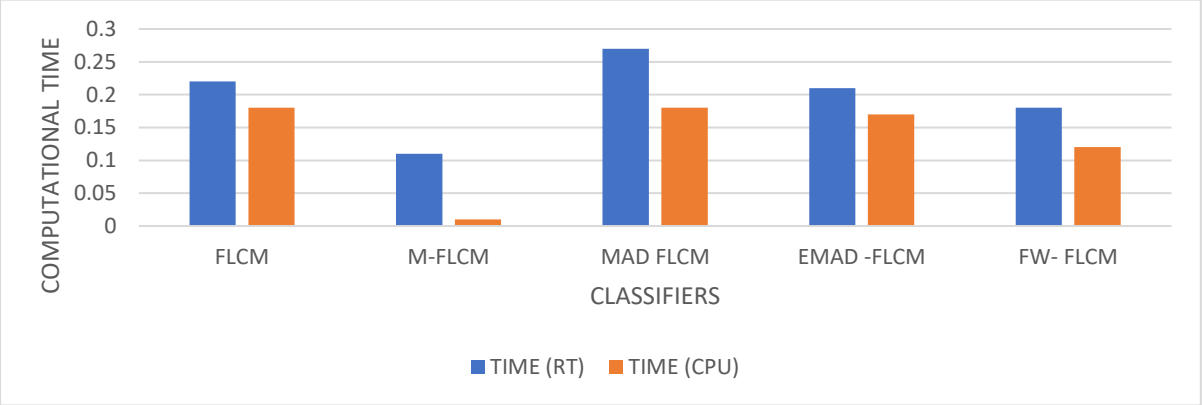

Figure 3: Comparative analysis of classifiers computational time

This study revealed that deleting IO's reduces the performance of the classifiers especially for the Mahalanobis approach. From this study we observed that due to IO's weighting and deletion, the M-FLCM classification efficiency is the lowest compared to the data weighting procedure. We also observed that the median variants also affect the classifier's performance compared to the direct data weighting method proposed. Therefore, this study demonstrated that IO's deletion has significant effects on the classifier's efficiency. The non-robustness of the OI's deletion method might be due to loss of information during data processing. This study therefore demonstrated that OI's weighting has extremely minimum information loss compared to the deletion method.

Discussion

From Table 4, we observed that group 1 probability value is higher than the probability for group 2 for all the data set. This summarizes that the classification output in Table 1 to Table 3 maybe comparable for some classifiers. Note that if the data set is normally distributed, the variation between the probabilities of the classifiers for the original data set, treated (Mahalanobis deletion) and weighted (F-weighted) data would be extremely small depending on the numerical strength for each group. On the contrary, if the data set contain influential observations, the probability difference between the classifiers for the original data, treated and weighted data set would vary. From this analysis, we observed that the different methods have various degrees of variation from the original data based on the percentage of IOs in the data set. We also noted that the higher the probability of group one the smaller the probability of group 2. Based on classification results and efficiency, this study affirmed that data deletion affects classification performance, and the weighting procedure enhances performance.

Table 4: Probabilities of data deletion and weighting

Data type	Methods	Probabilities	
		Group 1	Group 2
Freshwater and marine Alaska and Canadian Salmon (Abid &Izeboudjen, 2020)	Conventional data process	*0.5116*	*0.4884*
	Mahalanobis /deletion	0.5195	0.4805
	F-weight	0.5476	0.4524
Algeria forest fire data set(Abid &Izeboudjen, 2020)	Conventional data process	*0.6069*	*0.3931*
	Mahalanobis /deletion	0.5975	0.4025
	F-weight	0.5601	0.4399
Beign and Malaign Data	Conventional data process	*0.6799*	*0.3201*
	Mahalanobis /deletion	0.7070	0.2930
	F-weight	0.7345	0.2655

Comparing the classification outputs from Table 1 to Table 3, group probability plays a significant role in robust classification because the decision rules are based on the strength of group probabilities. The group probability comparison in Table 4 simply revealed that group classification accuracy depends strictly on the strength of the probability in each group which measures the probability of contribution for each object in the group in comparison with the total group probability. This concept is also true for Baye's probability rule. This analysis strictly affirmed the data dependency theory which concurred that classification performance depend on the nature of the data set(Okwonu et al., 2022).

CONCLUSION

The analysis based on three real data set demonstrated that the proposed FW-FLCM outperformed the conventional FLCM classifier and the median based FLCM respectively. The analysis revealed that the computational time for the proposed classifier is moderate compared to the conventional FLCM variants. The result affirmed that the proposed FW-FLCM classifier could be applied to perform classification tasks with data set containing IO's. This study concludes that the proposed classifier based on weighting improves the classifier's efficiency thereby validating the claim that IOs weighting results in minimum information loss compared to deleting methods which enhances relatively higher information loss. This study revealed that IOs deletion affect the performance of the FLCM. Therefore, this study demonstrated that data deletion and weighting play a significant role in classification performance.

REFERENCES

Abid, F., &Izeboudjen, N. (2020). Predicting Forest Fire in Algeria Using Data Mining Techniques: Case Study of the Decision Tree Algorithm. *Advances in Intelligent Systems and Computing*, *1105 AISC*. https://doi.org/10.1007/978-3-030-36674-2_37

Croux, C., & Dehon, C. (2001). Robust linear discriminant analysis using S-estimators. *Canadian Journal of Statistics*, *29*(3). https://doi.org/10.2307/3316042

Ghosh, A., SahaRay, R., Chakrabarty, S., & Bhadra, S. (2021). Robust generalised quadratic discriminant analysis. *Pattern Recognition*, *117*. https://doi.org/10.1016/j.patcog.2021.107981

He, X., & Fung, W. K. (2000). High Breakdown Estimation for Multiple Populations with Applications to Discriminant Analysis. *Journal of Multivariate Analysis*, *72*(2). https://doi.org/10.1006/jmva.1999.1857

Apanapudor, JS; Umukoro, J; Okwonu, FZ; Okposo, N,(2023), Optimal solution techniques for control problem of evolution equations, Science World Journal,18(3): 503-508

Hubert, M., & Debruyne, M. (2009). Breakdown value. *Wiley Interdisciplinary Reviews: Computational Statistics*, *1*(3). https://doi.org/10.1002/wics.34

Okwonu, F. Z; Ahad, N. A.; Okoloko, I. E.; Apanapudor, J. S.; Kamaruddin, S. A; Arunaye, F. I. (2022). Robust hybrid classification methods and applications, Pertanika J. Sci. & Technol. 30 (4): 2831 - 2850

Apanapudor, J. S; Aderibigbe, FM; Okwonu, FZ, (2020). An optimal penalty constant for discrete optimal control regulator problems, Journal of Physics: Conference Series, 1529(4): 042073

Okwonu, F. Z.(2024), Nonpharmaceutical And Pharmaceutical Covid-19 Prediction Models, FUDMA Journal Of Sciences,Vol.8(3): 309-313.

Hubert, M., & Debruyne, M. (2010). Minimum covariance determinant. *WIREs Computational Statistics*, *2*(1), 36–43. https://doi.org/10.1002/wics.61

Hubert, M., Rousseeuw, P. J., & van Aelst, S. (2008). High-breakdown robust multivariate methods. *Statistical Science*, *23*(1). https://doi.org/10.1214/088342307000000087

Jennison, C., Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1987). Robust Statistics: The Approach Based on Influence Functions. *Journal of the Royal Statistical Society. Series A (General)*, *150*(3). https://doi.org/10.2307/2981480

Johnson, R. A., & Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis* (3rd ed.). Prentice-Hall, Inc, Englewood Cliffs.

Okwonu, F. Z.; Ahad, N.A., Apanapudor J.S.; Arunaye I.F.(2021). Robust Multivariate Correlation Techniques: A Confirmation Analysis using Covid-19 Data Set. Pertanika J. Sci. & Technol. 29 (2): 999 - 1015 (2021). DOI: https://doi.org/10.47836/pjst.29.2.16

Law, J., Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). Robust Statistics-The Approach Based on Influence Functions. *The Statistician*, *35*(5). https://doi.org/10.2307/2987975

Lim, Y. F., Yahaya, S. S. S., & Ali, H. (2018). Robust linear discriminant analysis with highest breakdown point estimator. *Journal of Telecommunication, Electronic and Computer Engineering*, *10*(1–11).

Okwonu, F. Z.; Ahad, N.A. Apanapudor, J. S. Arunaye, F. I.(2020). Covid-19 Prediction Model (COVID-19-PM) For Social Distancing: The Height Perspective: COVID-19 Prediction Model for Social Distancing: The Height Perspective. Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences,57(4): 93-98.

Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M., Conceicao, E., & di Palma, M. A. (2021). robustbase: Basic Robust Statistics R package version 0.93-7. In *cran.stat.upd.edu.ph*.

Maronna, R. A., Martin, R. D., &Yohai, V. J. (2006). Robust Statistics: Theory and Methods. In *Robust Statistics: Theory and Methods*. https://doi.org/10.1002/0470010940

Okwonu, F.Z. and Othman, A.R.(2013). Comparative performance of classical Fisher linear discriminant analysis and robust Fisher linear discriminant analysis, Matematika,29: 213-220,https://doi.org/10.11113/matematika.v29.n.594

Nursalam, 2016, metode penelitian, Maronna, R. A., Martin, R. D., Yohai, V. J., & Salibián-Barrera, M. (2019). Robust Statistics Theory and Methods (with R) Second. In *Wiley Series in Probability and Statistics* (Vol. 53, Issue 9).

Okwonu, F. Z. (2012). *Several Robust Techniques in Two-Groups Unbiased Linear Classification*. https://core.ac.uk/download/pdf/199245931.pdf.

Okwonu, F. Z., Ahad, N. A., Ogini, N. O., Okoloko.I.E., & Wan, W. Z. (2022). Comparative Performance Evaluation of Efficiency for High Dimensional Classification Methods. *Journal Of Information and Communication Technology*, *Vol.21(3): 437-464*

Qin, X., Wang, S., Chen, B., & Zhang, K. (2020). Robust Fisher Linear Discriminant Analysis with Generalized Correntropic Loss Function. *Proceedings - 2020 Chinese Automation Congress, CAC 2020*. https://doi.org/10.1109/CAC51589.2020.9326644.

Okwonu, F.Z. Dieng, H.; Othman, A. R.; Hui, O. S.(2012), Classification of aedes adults mosquitoes in two distinct groups based on fisher linear discriminant analysis and FZOARO techniques, Mathematical Theory and Modeling,2(6): 22-30

Rousseeuw, P. J., & Hubert, M. (2018). Anomaly detection by robust statistics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(2). https://doi.org/10.1002/widm.1236

Okwonu, F. Z. Ahad, N. A.Hamid, H.; Muda, N. & Olimjon, S.S. (2023). Enhanced robust univariate classification methods for solving outliers and overfitting problems, Journal of Information and Communication TechnologyVol.22(1):1-30.

Seheult, A. H., Green, P. J., Rousseeuw, P. J., & Leroy, A. M. (1989). Robust Regression and Outlier Detection. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *152*(1). https://doi.org/10.2307/2982847

Okwonu, F. Z.; Othman, A. R.(2013): Heteroscedastic variancecovariance matrices for unbiased two groups linear classification methods, Applied Mathematical Sciences,7(138): 6855-6865. http://dx.doi.org/10.12988/ams.2013.39486

Tyler, D. E. (2008). Robust Statistics: Theory and Methods. *Journal of the American Statistical Association*, *103*(482). https://doi.org/10.1198/jasa.2008.s239

Wang, H., Lu, X., Hu, Z., & Zheng, W. (2014). Fisher discriminant analysis with L1-norm. *IEEE Transactions on Cybernetics*, *44*(6). https://doi.org/10.1109/TCYB.2013.2273355