Journal of Applied Science, Engineering, Technology, and Education *ISSN: 2685-0591 (online)*
Vol. 7 No. 2 (2025)
https://doi.org/10.35877/454RI.asci3948

Robust Principal Component Analysis in Multivariate Applications

Sharifah Sakinah Syed Abd Mutalib[a], Wan Nur Syahidah Wan Yusoff[b,*], Angelina Prima Kurniati[c], Nurul Aida Osman[d], & Zulhelmy[e]

[a]*Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia.*
[b*]*Centre for Mathematical Science, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Gambang, Pahang, Malaysia.*
[c]*School of Computing, Telkom University, Bandung, Indonesia.*
[d]*Department of Computer & Information Sciences, Universiti Teknologi Petronas, Perak, Malaysia.*
[e]*Faculty of Ecnomoy and Business Universitas Islam Riau Riau, Indonesia.*

Abstract

Principal component analysis (PCA) is one of the multivariate methods that aims to construct components, each of which contains a maximal amount of variation from the data unexplained by the other components. Classical PCA (CPCA) is based on the classical covariance, which is easily affected by outliers. Additionally, the outliers may distort the PCA result. Robust statistics is one of the methods to overcome the outliers. Therefore, in this study, an alternative technique for robust principal component analysis (RPCA) is presented. Index Set Equality (ISE) is used as a robust estimator to robustify the CPCA and hence produce robust PCA (RPCA). The Hawkins-Bradu Kass dataset is used to illustrate the robustness of RPCA towards outliers compared to CPCA. The score plot, diagnostic plot, and performance measures are used to evaluate and illustrate the robustness of the RPCA. From the plots and performance measures, RPCA successfully identifies all outliers and is unaffected by masking and swamping effects. However, CPCA can only detect 0.2857 outliers, has 0.7143 masking effects, but does not have swamping effects. This study shows that RPCA, by using ISE as the robust estimator, is a promising approach.

Keywords: Index Set Equality; multivariate data, outliers; principal component analysis; robust principal component analysis.

Received: 30 May 2025 Revised: 25 August 2025 Published: 31 August 2025

1. Introduction

One technique in multivariate data analysis called Principal Component Analysis (PCA) uses a few components to try and explain the covariance structure of the data (Hubert et al., 2005). PCA aims to determine a smaller number, k, of linear combinations of the initially observed, p variables while maintaining the majority of the data's variability (Filzmoser & Todorov, 2013; Hubert et al., 2005). The first principal component is the linear combination, which explains the highest variation in the data (Filzmoser & Todorov, 2013; Hubert et al., 2005; Rencher, 2002). In a direction orthogonal to the first principal component, the linear combination with maximal variance is the second principal component, and so on (Filzmoser & Todorov, 2013; Hubert et al., 2005; Rencher, 2002).

Classical PCA (CPCA) uses classical covariance or correlation matrices to compute eigenvalues and eigenvectors. Nevertheless, outliers can affect these classical covariance and correlation matrices. Masking and swamping effects can arise from even a single outlier distorting the computation of the classical covariance and correlation matrices. In CPCA, the existence of outliers leads to the first components being frequently drawn to outliers and might not capture the variation of inliers (Chen et al., 2020). Thus, data reduction becomes unreliable using CPCA in the presence of outliers.

Robust methods are used in PCA to overcome outlier problems, such as in Chen et al. (2020), Wang & Pham (2011), and Hubert et al. (2005). Obtaining principle components that are unaffected by outliers is the aim of robust PCA (RPCA) (Hubert et al., 2005). There are two approaches to RPCA, which are: (i) utilising a robust covariance

* Corresponding author.
E-mail address: wnsyahidah@umpsa.edu.my

estimator to replace the classical covariance and (ii) using projection pursuit (PP) techniques (Hubert et al., 2005). This study is going to focus on the first approach, which substitutes the classical covariance with a robust estimation of covariance.

Previous studies have offered a variety of robust estimators, including the Minimum Volume Ellipsoid (MVE) estimator (Rousseeuw, 1985), the Fast Minimum Covariance Determinant (FMCD) estimator (Rousseeuw & Driessen, 1999), the M-estimator (Devlin et al., 1981), the S-estimator (Rousseeuw & Yohai, 1984), and the MM-estimator (Maronna & Yohai, 2017). Among the robust estimators, the FMCD is a well-known robust estimator and is used widely. However, when dealing with large datasets featuring many dimensions, the FMCD is still not computationally efficient because the complexity of computation grows exponentially with the dimension. Hence, the problem in FMCD became the motivation for (Salleh, 2013) to develop robust estimators named Covariance Matrix Equality (CME) and Index Set Equality (ISE). According to Salleh (2013), the computational complexity of CME and ISE is only $O(p \ln p)$ compared to FMCD, which is $O(p^3)$. It is discovered that ISE performs better than CME (Lim & Midi, 2016). Recent contributions to robust PCA, such as Zahariah & Midi (2023) and Cai et al. (2021), have been developed primarily for high-dimensional contexts in which the number of variables exceeds the sample size. While these approaches are valuable for such settings, they do not directly address scenarios considered in this study, where the focus lies on datasets where the sample size exceeds the dimensionality.

Hence, in this study, the ISE is used as the robust estimator to replace the classical estimator in PCA and yield RPCA. The objective of this study is to present an alternative technique for robust principal component analysis (RPCA) using ISE. The robustness of RPCA towards outliers compared to CPCA is illustrated using the Hawkins-Bradu Kass (HBK) dataset. The remainder of this paper is organised as follows. The next section outlines the methodology, which is divided into three phases. Section 3 reports the findings of the study using graphical illustrations and performance measures. The last section presents the conclusion of this study.

2. Materials and Methods

The methodology for this study consists of three phases. Phase I is to obtain the robust estimator, the Index Set Equality (ISE) estimator. Next, the sample mean and covariance matrix of ISE are used to robustify Principal Component Analysis (PCA). The final phase involves performance evaluation, in which RPCA and CPCA are compared using the Hawkins-Bradu-Kass (HBK) dataset. R programming is utilized in this study for the analysis. The complete phases of the methodology are shown in Figure 1.

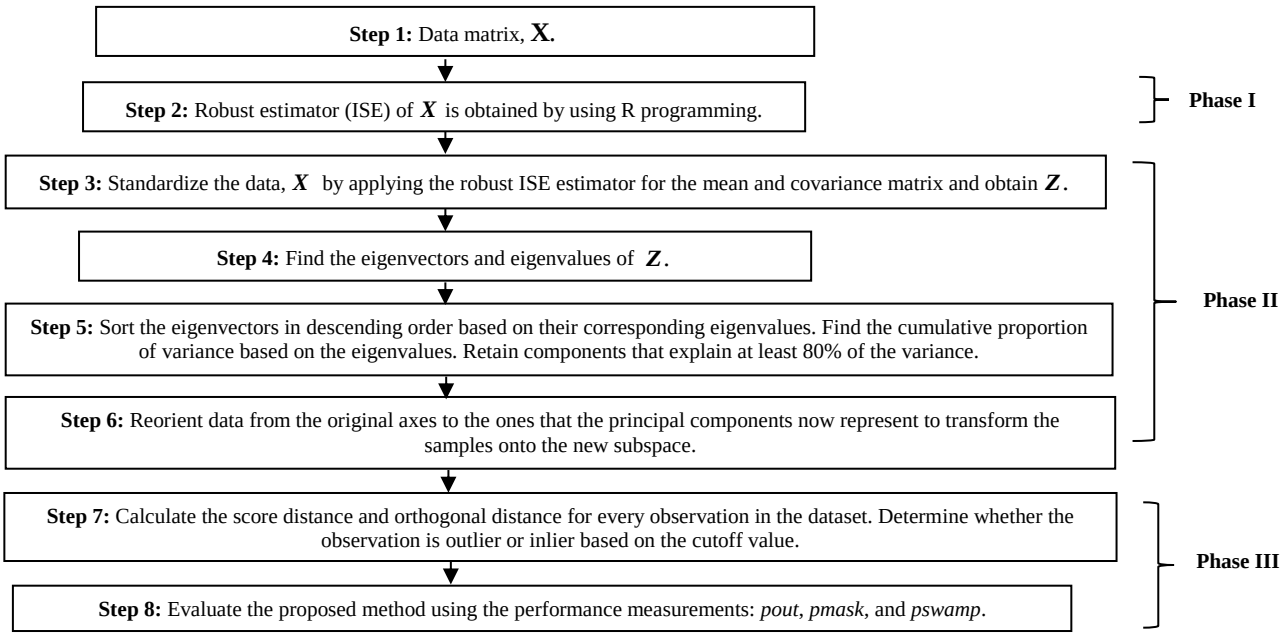


Fig. 1. Flow chart for three phases of the methodology.

Phase I: Index Set Equality (ISE) estimator

To address the issue of Minimum Vector Variance (MVV), Salleh (2013) introduced two robust estimators: Covariance Matrix Equality (CME) and Index Set Equality (ISE). Moreover, MVV has been proposed to solve issues in the FMCD (Fast Minimum Covariance Determinant) estimator. CME, ISE, and MVV introduced modifications to the final step of the FMCD algorithm. In the final step of the FMCD, CME, MVV, and ISE algorithms, the equivalence of the two covariance structures is assessed.

The steps of the ISE algorithm are as follows (Salleh, 2013; Salleh & Djauhari, 2011). In the ISE algorithm, the key steps that allow it to overcome the influence of outliers are the subset selection. The algorithm identifies a stable subset of observations through index set equality as shown in Step 6. This procedure effectively selects the portion of the data that is internally consistent, while observations that behave differently (i.e., potential outliers) are excluded from contributing to the estimation. Intuitively, this means that the influence of outliers is reduced because they are not part of the equality set, allowing the algorithm to capture the main structure of the data more reliably.

Step 1: Choose an arbitrary subset H_{old} comprising h distinct observations, with h being the smallest integer $\geq (n+p+1)/2$, where n is the sample size and p is the number of variables. H_{old} represents an initial subset of h observations selected among n observations.

Step 2: Calculate the covariance matrix $S_{H_{old}}$ and mean vector $\bar{X}_{H_{old}}$ of all observations included in H_{old}.

Step 3: Calculate $d^2_{H_{old}}(i) = \left(X_i - \bar{X}_{H_{old}}\right)' S^{-1}_{H_{old}} \left(X_i - \bar{X}_{H_{old}}\right)$ for $i = 1,2,\ldots,n$. $d^2_{H_{old}}(i)$ is the distance for each observation in H_{old}.

Step 4: Arrange $d^2_{H_{old}}(i)$ for $i = 1,2,\ldots,n$ from smallest to largest,

$$d^2_{H_{old}}(\pi(1)) \leq d^2_{H_{old}}(\pi(2)) \leq \ldots \leq d^2_{H_{old}}(\pi(n))$$

where π is a permutation on $\{1,2,\ldots,n\}$.

Step 5: Define $H_{new} = \left\{X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(h)}\right\}$ and then calculate $\bar{X}_{H_{new}}$, $S_{H_{new}}$ and $d^2_{H_{new}}(i)$ for $i = 1,2,\ldots,n$. H_{new} is a new subset that was created in Step 4, where only h observations are selected after $d^2_{H_{old}}(i)$ is sorting. $\bar{X}_{H_{new}}$ and $S_{H_{new}}$ is the sample mean and covariance matrix of H_{new}. While $d^2_{H_{new}}(i)$ is the distance for each observation in H_{new}.

Step 6$_{ISE}$: If $I_{new} \neq I_{old}$, let $H_{old} := H_{new}$, calculate $\bar{X}_{H_{new}}$ and let $H_{old} := H_{new}$, $\bar{X}_{H_{old}} := \bar{X}_{H_{new}}$ and $S_{H_{old}} := S_{H_{new}}$. Subsequently, go to Step 3. The procedure ends otherwise. I_{old} and I_{new} is the index sets corresponding to the observations in H_{old} and H_{new}, respectively.

During the algorithm's final stage, Step 6$_{ISE}$, the final robust estimator of ISE for sample means, $\bar{x}_{j,ISE}$ and the covariance matrix, $\mathbf{S}_{ISE}$ are produced. These robust estimators of $\bar{x}_{j,ISE}$ and $\mathbf{S}_{ISE}$ are used to robustify Principal Component Analysis.

Phase II: Robust Principal Component Analysis (RPCA)

Next, the formulation of an alternative technique by incorporating the principal component analysis with Index Set Equality (ISE) is performed.

Step 1: The robust estimator of ISE for the sample means, $\bar{x}_{j,ISE}$ and covariance matrix, $\mathbf{S}_{ISE}$ is obtained.

Step 2: Standardise the data so that each variable contributes equally to the analysis. In Eq. (1), z_j is the standardize variable, x_{ij} is the original data, $\bar{x}_{j,ISE}$ and $\mathbf{S}_{ISE}$ are the robust estimators of mean and standard deviation for each variable.

$$z_j = \frac{x_{ij} - \bar{x}_{j,ISE}}{s_{j,ISE}} \qquad i = 1,2,\ldots,n \quad j = 1,2,\ldots,p. \tag{1}$$

Step 3: Find the eigenvectors and eigenvalues of the sample covariance matrix. Let $\mathbf{S}_{ISE}$ obtain the eigenvalue-eigenvector pairs $(\lambda_1,\mathbf{e}_1),(\lambda_2,\mathbf{e}_2),\ldots,(\lambda_p,\mathbf{e}_p)$ where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ (Johnson & Wichern, 2002). To find the eigenvector and eigenvalue, Eq. (2) and (3) are used.

$$\left(\mathbf{S}_{ISE} - \lambda_j \mathbf{I}\right)U_j = 0 \qquad j = 1,2,\ldots,p \tag{2}$$

$$\det\left(\mathbf{S}_{ISE} - \lambda \mathbf{I}\right) = 0 \tag{3}$$

where $\mathbf{I}$ is the identity matrix.

Step 4: The eigenvectors are sorted in decreasing order based on the corresponding eigenvalues. This step is crucial as it helps prioritise the principal components that explain the most variance. Decide how many principal components to retain. A common approach is to select the top k eigenvectors that correspond to the k largest eigenvalues. The proportion of variance explained can guide this decision; for example, retaining components that explain 80% or 90% of the variance (Filzmoser & Todorov, 2013; Hubert et al., 2005; Rencher, 2002).

Step 5: In the last step, we need to reorient data from the original axes to the ones that the principal components now represent to transform the samples onto the new subspace. Finally, we projected the data points using the new axes and calculated the principal components.

Phase III: Performance Evaluation

The performance of RPCA will be evaluated by comparing its performance with CPCA. The HBK dataset is used in this study to illustrate the robustness of RPCA compared to CPCA. HBK dataset is a benchmark dataset in outlier studies and has been used in many studies, such as in Abd Mutalib et al. (2021), Midi et al. (2020), Affindi et al. (2019) and Pan et al. (2000)(Pan et al., 2000). A visual comparison of RPCA and CPCA is provided through a diagnostic plot or outlier map. The diagnostic plot consists of the score distance, SD_i, of every observation on the horizontal axis and the orthogonal distance, OD_i, of every observation on the vertical axis (Hubert et al., 2005).

According to Hubert et al. (2005) and Zahariah & Midi (2023) the score distance is defined as in Eq. (4),

$$SD_i = \sqrt{\sum_{j=1}^{k} \frac{t_{ij}^2}{l_j}}, \tag{4}$$

where the scores t_{ij} are obtained from Eq. (5) and l_j is the eigenvalues.

$$\mathbf{T}_{n,k} = \left(\mathbf{X}_{n,p} - \mathbf{1}_n \hat{\boldsymbol{\mu}}'\right)\mathbf{P}_{p,k}, \tag{5}$$

where $\mathbf{X}_{n,p}$ is the $n \times p$ original data, $\mathbf{1}_n$ is the column vector of length n with all entries equal to 1, $\hat{\boldsymbol{\mu}}'$ and $\mathbf{P}_{p,k}$ are the mean and loading matrix for the original $\mathbf{X}_{n,p}$. The cutoff value for SD_i is equal to $\sqrt{\chi^2_{k,0.975}}$ (Zahariah & Midi, 2023). The orthogonal distance is given as in Eq. (6),

$$OD_i = \left\| \mathbf{x}_i - \hat{\boldsymbol{\mu}} - \mathbf{P}_{p,k}\mathbf{t}_i' \right\|, \tag{6}$$

where the ith observation is represented as the p-variate column vector $\mathbf{x}_i$ and $\mathbf{t}_i'$ is the ith row of $\mathbf{T}_{n,k}$. The cutoff value for OD_i is equal to $\left(\hat{\mu} + \hat{\sigma} z_{0.975}\right)^{3/2}$, where $\hat{\mu}$ and $\hat{\sigma}$ can be estimated by the median and MAD of the values $OD_i^{2/3}$ (Hubert & Engelen, 2004; Todorov & Filzmoser, 2009).

Next, the performance of CPCA and RPCA is evaluated by using *pout*, *pmask* and *pswamp*. According to Hubert & Engelen (2004), the score distance, SD_i exactly matches the Mahalanobis distance when employing CPCA. Therefore, we used SD_i to further investigate the performance of CPCA and RPCA in classifying outliers. The outliers are identified if the value of SD_i is more than $\sqrt{\chi^2_{k,0.975}}$. The *pout*, *pmask* and *pswamp* are computed as in Eq. (7) – (9) (Abd Mutalib et al., 2023; Zulkipli et al., 2022).

$$pout = \frac{"success"}{out} \tag{7}$$

where *out* is the number of outliers, and *"success"* is the number of data sets in which the method was able to identify every outlier. Eq. (8) represents the probability of misclassifying outliers as inliers (*pmask*).

$$pmask = \frac{"\ failure"}{out} \tag{8}$$

where *"failure"* is the total of outliers that are found to be inliers in all datasets. Eq. (9) gives the probability of misidentifying inliers as outliers (*pswamp*).

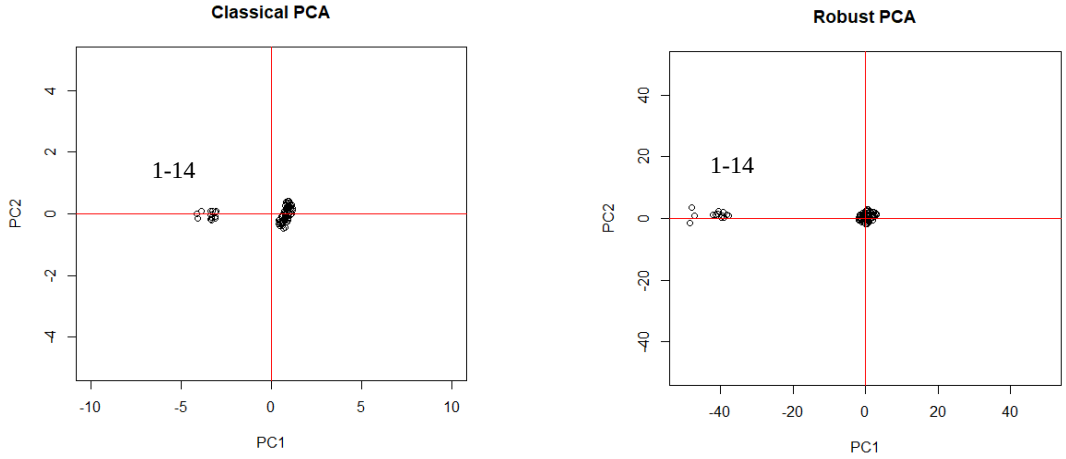
$$pswamp = \frac{"\ false"}{(n - out)} \tag{9}$$

where *n* represents the sample size and *"false"* denotes the total number of inliers in all data sets that are misclassified as outliers.

3. Results And Discussions

An illustration of robustness between classical PCA (CPCA) and robust PCA (RPCA) is shown by using the Hawkins-Bradu Kass (HBK) dataset presented in this section. The HBK dataset consists of 75 observations, $n = 75$, and three variables, $p = 3$. Observations 1 – 14 are identified as outliers in previous studies, whereas observations 15 – 75 are identified as inliers (Abd Mutalib et al., 2021, 2023; Affindi et al., 2019; Atkinson & Mulira, 1993; Lim & Midi, 2016).

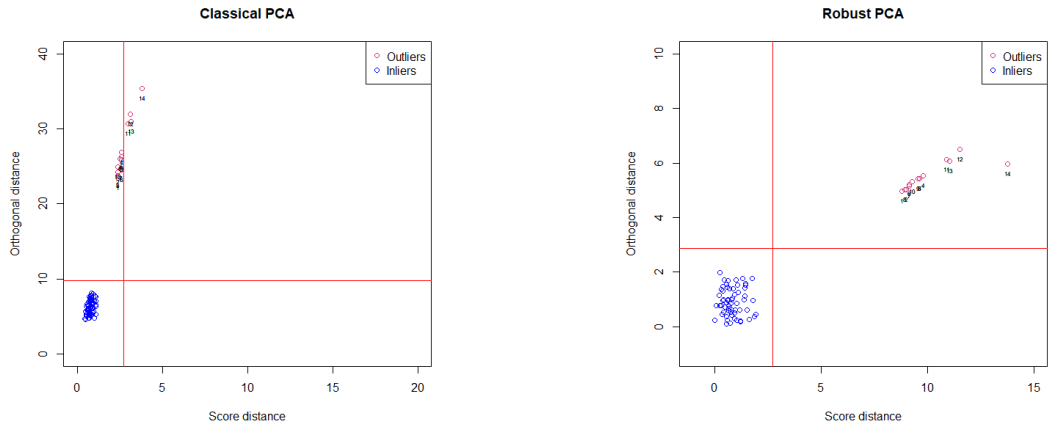
The HBK dataset's score plot, which is based on robust and classical principal components, is displayed in Figure 2. Figure 2(a) and (b) show the separation of outliers and inliers. However, in Figure 2(a), the inliers are far from zero even though the scores have a zero mean. This occurs because the mean of the data points provides an inaccurate estimate of the true center of the data when outliers are present. Figure 2(b) presents the score plot for robust PCA, demonstrating that the outliers do not affect the center estimate, with all inliers grouped around the zero mean.



(a) (b)

Fig. 2. Score plot of the Hawkins– Bradu–Kass dataset obtained with (a) Classical PCA (CPCA) and (b) Robust PCA (RPCA).

Figure 3 shows the diagnostic plot for CPCA and RPCA. Orthogonal outliers, bad leverage points, and good leverage points can be identified based on the diagnostic plot. From Figures 3(a) and (b), the separation of inliers and outliers is clear by CPCA and RPCA. From Figure 3(a), CPCA identified observations 1-10 as orthogonal outliers and 11 – 14 as bad leverage points. Figure 3(b) illustrates the diagnostic plot of the HBK dataset using robust PCA. All outliers from Figure 3(b) are classified as bad leverage points, while observations 15 through 75 are classified as inliers. From Figures 2(b) and 3(b), it is shown that RPCA using ISE as the robust estimator is not sensitive towards outliers compared to CPCA.



(a) (b)

Fig. 3. Diagnostic plot of the Hawkins– Bradu–Kass dataset obtained with (a) Classical PCA (CPCA) and (b) Robust PCA (RPCA).

Table 1 presents the results for *pout*, *pmask*, and *pswamp* of CPCA and RPCA. The results demonstrate that RPCA successfully detects all outliers in the HBK dataset, outperforming CPCA. Additionally, RPCA does not exhibit any masking effect, while CPCA shows a significant masking effect, with a high probability of misclassifying outliers as inliers. Regarding *pswamp*, neither CPCA nor RPCA shows any swamping effect.

Table 1. Results of performance measurements for CPCA and RPCA.

	CPCA	**RPCA**
pout	0.2857	1.0000
pmask	0.7143	0.0000
pswamp	0.0000	0.0000

4. Conclusions

This study intends to introduce an alternative technique for robust principal component analysis (RPCA) using ISE. The robustness of RPCA is compared with CPCA by using a benchmark dataset, which is the Hawkins-Bradu Kass (HBK) dataset. The score plot, diagnostic plot, and performance measures are used to evaluate and illustrate the robustness of the RPCA. From the plots and performance measures, RPCA effectively finds every outlier and does not have masking and swamping effects. This is because misclassifying outliers as inliers is a greater problem than misclassifying inliers as outliers. This alternative method of RPCA by using ISE shows promising results and can be used to robustify the PCA. However, according to Salleh (2013), ISE can only be used when the sample size is more than the dimension. Practically, the proposed RPCA may be employed in diverse fields, including microarray studies, genetics, and chemometrics. In this study, only the HBK dataset is used to investigate the performance of RPCA using ISE. For future studies, the performance of RPCA using ISE should be investigated by using a large and high-dimensional dataset.

Acknowledgements: This research was funded by a grant from the International Matching Grant, RDU222708/ UIC221521 of Universiti Malaysia Pahang Al-Sultan Abdullah and Universiti Malaysia Terengganu.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

Contribution: Sharifah Sakinah Syed Abd Mutalib: Analysis, Writing -Original Draft. Wan Nur Syahidah Wan Yusoff: Conceptualization, Methodology, Writing-Reviewing and Editing. Angelina Prima Kurniati: Conceptualization, Methodology, Writing-Reviewing and Editing. Nurul Aida Osman: Conceptualization,

Methodology, Writing-Reviewing and Editing. Zulhelmy: Conceptualization, Methodology, Writing-Reviewing and Editing.

References

Abd Mutalib, S. S. S., Satari, S. Z., & Wan Yusoff, W. N. S. (2021). Comparison of robust estimators for detecting outliers in multivariate datasets. *Journal of Physics: Conference Series*, *1988*(1), 1–9. https://doi.org/10.1088/1742-6596/1988/1/012095

Abd Mutalib, S. S. S., Satari, S. Z., & Yusoff, W. N. S. W. (2023). A New Single Linkage Robust Clustering Outlier Detection Procedures for Multivariate Data. *Sains Malaysiana*, *52*(8), 2431–2451. https://doi.org/10.17576/jsm-2023-5208-19

Affindi, A. N., Ahmad, S., & Mohamad, M. (2019). A Comparative Study between Ridge MM and Ridge Least Trimmed Squares Estimators in Handling Multicollinearity and Outliers. *Journal of Physics: Conference Series*, *1366*(1), 012113. https://doi.org/10.1088/1742-6596/1366/1/012113

Atkinson, A. C., & Mulira, H. M. (1993). The stalactite plot for the detection of multivariate outliers. *Statistics and Computing*, *3*(1), 27–35. https://doi.org/10.1007/BF00146951

Cai, H. Q., Liu, J., & Yin, W. (2021). Learned Robust PCA: A Scalable Deep Unfolding Approach for High-Dimensional Outlier Detection. *Advances in Neural Information Processing Systems*, *21*(NeurIPS), 16977–16989.

Chen, X., Zhang, B., Wang, T., Bonni, A., & Zhao, G. (2020). Robust principal component analysis for accurate outlier sample detection in RNA-Seq data. *BMC Bioinformatics*, *21*(1), 1–20. https://doi.org/10.1186/s12859-020-03608-0

Devlin, S. J., Gnanadesikan, R., & Kettenring, J. R. (1981). Robust estimation of dispersion matrices and principal components. *Journal of the American Statistical Association*, *76*(374), 354–362. https://doi.org/10.1080/01621459.1981.10477654

Filzmoser, P., & Todorov, V. (2013). Robust tools for the imperfect world. *Information Sciences*, *245*, 4–20. https://doi.org/10.1016/j.ins.2012.10.017

Hubert, M., & Engelen, S. (2004). Robust PCA and classification in biosciences. *Bioinformatics*, *20*(11), 1728–1736. https://doi.org/10.1093/bioinformatics/bth158

Hubert, M., Rousseeuw, P. J., & Branden, K. Vanden. (2005). ROBPCA: A new approach to robust principal component analysis. *Technometrics*, *47*(1), 64–79. https://doi.org/10.1198/004017004000000563

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (Fifth Edit). New Jersey: Hall.

Lim, H. A., & Midi, H. (2016). Diagnostic Robust Generalized Potential based on Index Set Equality (DRGP (ISE)) for the identification of high leverage points in linear model. *Computational Statistics*, *31*(3), 859–877. https://doi.org/10.1007/s00180-016-0662-6

Maronna, R. A., & Yohai, V. J. (2017). Robust and efficient estimation of multivariate scatter and location. *Computational Statistics and Data Analysis*, *109*, 64–75. https://doi.org/10.1016/j.csda.2016.11.006

Midi, H., Hendi, H. T., Arasan, J., & Uraibi, H. (2020). Fast and robust diagnostic technique for the detection of high leverage points. *Pertanika Journal of Science and Technology*, *28*(4), 1203–1220. https://doi.org/10.47836/pjst.28.4.05

Mutalib, S. S. S. A., Satari, S. Z., & Yusoff, W. N. S. W. (2021). Comparison of Robust Estimators for Detecting Outliers in Multivariate Datasets. *Journal of Physics: Conference Series*, *1988*(1). https://doi.org/10.1088/1742-6596/1988/1/012095

Pan, J.-X., Fung, W.-K., & Fang, K.-T. (2000). Multiple outlier detection in multivariate data using projection pursuit techniques. *Journal of Statistical Planning and Inference*, *83*(1), 153–167. https://doi.org/10.1016/s0378-3758(99)00091-9

Rencher, A. C. (2002). *Methods of Multivariate Analysis*. https://doi.org/10.2307/2669873

Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. *Mathematical Statistics and*

Applications, *8*, 283–297.

Rousseeuw, P. J., & Driessen, K. Van. (1999). A fast algorithm for the Minimum Covariance Determinant estimator. *Technometrics*, *41*(3), 212–223.

Rousseeuw, P., & Yohai, V. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis* (pp. 256–272). https://doi.org/10.1007/978-1-4615-7821-5_15

Salleh, R. M. (2013). *A robust estimation method of location and scale with application in monitoring process variability* [Universiti Teknologi Malaysia]. https://doi.org/10.1190/segam2013-0137.1

Salleh, R. M., & Djauhari, M. A. (2011). Robust Hotelling's T^2 Control Charting in Spike Production Process. *International Seminar on the Application of Science & Mathematics 2011 (ISASM 2011)*, 1–8.

Todorov, V., & Filzmoser, P. (2009). An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software*, *32*(3), 1–47. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.150.3305&rep=rep1&type=pdf

Wang, Y., & Pham, H. (2011). Analyzing the effects of air pollution and mortality by generalized additive models with robust principal components. *International Journal of System Assurance Engineering and Management*, *2*(3), 253–259. https://doi.org/10.1007/s13198-011-0072-7

Zahariah, S., & Midi, H. (2023). Minimum regularized covariance determinant and principal component analysis-based method for the identification of high leverage points in high dimensional sparse data. *Journal of Applied Statistics*, *50*(13), 2817–2835. https://doi.org/10.1080/02664763.2022.2093842

Zulkipli, N. S., Satari, S. Z., & Wan Yusoff, W. S. (2022). The effect of different similarity distance Measures in Detecting Outliers Using Single-Linkage Clustering Algorithm for Univariate Circular Biological Data. *Pakistan Journal of Statistics and Operation Research*, *18*(3), 561–573.