ICMASE 2026

PERFORMANCE EVALUATION OF ROBUST PRINCIPAL COMPONENT ANALYSIS-BASED MACHINE LEARNING MODELS FOR BREAST CANCER CLASSIFICATION

Sharifah Sakinah bt Syed Abd Mutalib[1,*], Maharani Abu Bakar[2], Muhamad Safiih bin Lola[3]

[1,2,3]*Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia*

ABSTRACT

Breast cancer classification remains an important problem in medical data analysis due to its high dimensionality and the presence of outliers that may affect model performance. Dimensionality reduction techniques such as Principal Component Analysis (PCA) are commonly used in high-dimensional data to reduce dimensionality while still retaining as much information as possible from the original data. Therefore, in this study, a performance evaluation of machine learning models incorporating PCA and Robust PCA for breast cancer classification is performed. Baseline models classifiers consisting of K-means, Support Vector Machine (SVM), and Artificial Neural Network (ANN) are compared with their corresponding PCA-based and Robust PCA-based classifiers. The objective of this study is to assess the impact of dimensionality reduction techniques on classification performance and robustness. All models are evaluated using accuracy, precision, recall, and F1-score. The experimental results show that SVM and Robust PCA-SVM have similar performance and are the best classifiers compared to other classifiers. In particular, Robust PCA-based machine learning models consistently demonstrate improved classification performance when compared to PCA-based approaches. These findings suggest that incorporating robust dimensionality reduction techniques enhances the reliability of classification outcomes in breast cancer data analysis. The study provides a comparison of baseline, PCA-based, and Robust PCA-based classifiers and highlights the benefits of robust dimensionality reduction in medical classification tasks.

***K*eywords** Principal Component Analysis · Robust Principal Component Analysis · Machine Learning · K-means · Support Vector Machine · Artificial Neural Network.

References

[1] Adiwijaya, Wisesty, U. N., Lisnawati, E., Aditsania, A., and Kusumo, D. S., Dimensionality reduction using Principal Component Analysis for cancer detection based on microarray data classification, Journal of Computer Science, 14(11):1521–1530, 2018.

[2] An, Q., Rahman, S., Zhou, J., and Kang, J. J., A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges, Sensors, 23(9), 2023.

[3] Boudt, K., Rousseeuw, P. J., Vanduffel, S., and Verdonck, T., The minimum regularized covariance determinant estimator, Statistics and Computing, 30(1):113–128, 2020.

[4] Chen, X., Zhang, B., Wang, T., Bonni, A., Zhao, G., Robust principal component analysis for accurate outlier sample detection in RNA-Seq data, BMC Bioinformatics, 21(1):1–20., 2020.

[5] Chhajer, P., Shah, M., Kshirsagar, A., The applications of artificial neural networks , support vector machines , and long – short term memory for stock market prediction, Decision Analytics Journal, 2(November 2021):100015, 2021.

[*] *Corresponding Author's E-mail: s.sakinah@umt.edu.my*

[6] Dhomse Kanchan, B., and Mahale Kishor, M., Study of machine learning algorithms for special disease prediction using principal of component analysis, Proceedings - International Conference on Global Trends in Signal Processing, Information Computing and Communication, ICGTSPICC 2016, (February), 5–10, 2017.

[7] Hubert, M., Rousseeuw, P. J., and Branden, K. Vanden., ROBPCA: A new approach to robust principal component analysis, Technometrics, 47(1):64–79, 2005.

[8] Zahariah, S., and Midi, H, 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, 2023.