Abstract No: 218 ***ICT, Mathematics, and Statistics***

REDUCED ORDER MODEL FOR BREAST CANCER DETECTION

H.D.A.W. Subawickrama[1*], U.G.I.G.K. Udagedara[2] and S.A.A. Nishantha[3]

[1]Department of Physical Science, Faculty of Applied Sciences, Rajarata University of Sri Lanka, Mihintale, Sri Lanka
[2]Department of Engineering Computing and Mathematical Sciences, Lewis University, IL, USA
[3]Department off Mathematics, Clarkson University, NY, USA
[*]amandisubawickrama586@gmail.com

Breast cancer often starts in the milk ducts or glandular lobules. According to the World Health Organization, there are about 2.3 million new cases and over 600,000 deaths annually. In Sri Lanka, around 3,000 new cases are diagnosed yearly, highlighting the need for better detection methods. Imaging techniques have limitations, like reduced sensitivity in dense breast tissue and potential false positives, which means dataset features may not always effectively capture crucial information for detecting breast cancer. This study presents a dimensional reduction process using Principal Component Analysis (PCA), a powerful technique for feature selection and dimensionality reduction, particularly effective with large datasets. This study aimed to create a low-dimensional model that can be used to classify breast cancers accurately. A benchmark dataset called Wisconsin Diagnostic Breast Cancer (WDBC) was used for data preprocessing and creating a reduced-order model using PCA. The performance of the reduced order model was evaluated by reconstructing the original data using the selected principal components and assessing the accuracy of this reconstructed data. The reconstruction error was measured using the Mean Squared Error (MSE) metric. The scree plot revealed a notable difference in the percentage of correctly reconstructed benign and malignant cases when selecting principal components 1 and 2, highlighting a substantial drop in variance. However, when choosing between 3 to 30 components, the percentages of correctly reconstructed benign and malignant cases remained consistent at 70.78% and 29.22%, respectively. These values are identical to the class distributions observed from the original data. Ten principal components were selected for model reduction, resulting in a reconstruction error of 0.0002 MSE. This low error value reflects a very accurate reconstruction from the reduced dimensions. These findings suggest the capability of the reduced order model to identify benign and malignant cases, which can be further used for investigations with high dimensional and complex data.

Keywords: Breast cancer detection, Feature extraction, Principal component analysis, Reduced order modelling