

**A COMPARATIVE STUDY ON PRINCIPAL COMPONENT ANALYSIS AND
RANDOM PROJECTION FOR NEAREST NEIGHBOR CLASSIFICATION**

D.S. DEEGALLA,^{1,3} K.S. WALGAMA,² AND H. BOSTRÖM,³

¹*Department of Computer Engineering, Faculty of Engineering, University of Peradeniya.*

²*Department of Engineering Mathematics, Faculty of Engineering, University of Peradeniya.*

³*Department of Computer and Systems Sciences, Stockholm University and Royal Institute of
Technology, Sweden*

Nearest neighbor is known as a simple effective classifier for most classification problems. However, the computational cost of using nearest neighbor classification often prevents the method from being applied in practice when dealing with high-dimensional data, such as images and microarrays. One possible solution to this problem is to reduce the dimensionality of the data, ideally without losing predictive performance.

In this study, we have investigated the use of two dimensionality reduction methods namely Principal Component Analysis (PCA) and Random Projection (RP) for this purpose and compared with respect to the performance of the resulting nearest neighbor classifier on five image data sets and five microarray data sets.

The experimental results show that PCA generally results in a higher degree of accuracy than RP. PCA is more sensitive to the choice of the number of reduced dimensions compared to RP. After reaching a peak, the accuracy degrades with the number of dimensions for PCA, while the accuracy for RP increases with the number of dimensions. The experiments also show that the use of PCA resulted in the highest accuracy for 9 of the 10 data sets. For 6 cases out of 10, we have noticed that both PCA and RP outperform using all features for classification. Therefore, both methods result in more efficient and more accurate, nearest neighbor classifications.

Financial Assistance by the SIDA/SAREC is acknowledged.