

## Abstract

Various feature selection techniques have been proposed in the field of machine learning. The performance of wrapper methods are often found comparable to filter methods in feature selection due to its ability to take feature dependencies and the interaction between feature subset search and model selection, if cost is ignored. Feature selection techniques typically depend on ways how they compute feature-feature correlation and feature-class correlation. These two computations are highly governed by the correlation measure being used.

In this work, an effective correlation measure named *Effective Normalized Mean Residue Similarity* (ENMRS) has been proposed for determining feature-feature correlation and it is validated statistically as well as biologically on gene expression dataset. A deterministic statistical measure named *Core Range based Relevance* (crRelevance) for assessing feature-class correlation have been proposed and it is validated on several UCI datasets. Finally, a method is designed named *Feature Pair Enhanced Correlation* (FPEC) to effectively employ these feature-feature and feature-class correlation to extract relevant feature subset from multi-class gene expression data as well as machine learning datasets. The performance of FPEC in association with decision tree classifier, random forest classifier and KNN classifier has been found highly satisfactory over several benchmark datasets.

**Keywords:** *Feature, correlation, relevance, featur-feature, feature-class*