

International Conference on **ONCOLOGY AND RADIOLOGY**
&
International Conference on **NANOTECHNOLOGY**

December 03-04, 2018 | Edinburgh, Scotland

An algorithm to identify common genes among diseases to construct predictive models through data mining

Dayana Carla de Macedo, ECM Ishikawa, CB Santos, SN Matos, HB Borges and AC Francisco
Midwest University of Parana, Brazil

This research proposes a new method with an algorithm DRM_F to reduce the dimensionality of gene expression data. Dimensionality reduction methods are applied in various domains, however, the area involving gene expression data was opted for. The new method of data dimension reduction is called DRM-F and it is able to identify in n bases of a gene domain the most relevant attributes, using the concepts of equivalence and generalization. The excessive use of attributes may affect the search for patterns and extraction of useful knowledge, because they harm the performance of learning of algorithms in both speed and success rate. The use of dimensionality reduction methods becomes an important alternative; however, these methods do not deal with the reduction of attributes in a specific area. For this experiment three databases of gene expression,

were used which deals with cancer disease. The bases are called DLBCL, DLBCL – tumor and DLBCL ALL / AML. The Attribute Selection was also applied in the three databases for the comparison of the results. Analyses of the results using the criterion of cross-validation revealed that the employment of the methods resulted in the improvement of the success rates compared to the bases containing the full range of attributes. The algorithm is composed of three steps. With the defined disease, the second step is prepare the basis with the genes to ensure its integrity and apply the concept of the proposed algorithm to search patterns common among the genes. The last step the algorithm to evaluate the basis by mining algorithm.

dayanacarlademacedo@gmail.com