

https://dspace.org/

Cluster analysis of gene expression data on cancerous tissue samples.

Dinger, Steven Conrad

2012-01-25
http://hdl.handle.net/10539/11173

Downloaded from a DSpace repository. https://dspace.org/

ABSTRACT

The cluster analysis of gene expression data is an important unsupervised learning method that is commonly used to discover the inherent structure in the large amounts of data generated by microarray measurements. The focus of this research is to develop a novel clustering algorithm that adheres to the definition of unsupervised learning whilst minimising any sources of bias. The developed diffractive clustering algorithm is based on the fundamental diffraction properties of light, which presents a novel view and framework for clustering data. The algorithm is tested on multiple cancerous tissue data sets that are well established in the literature. The overall result is a clustering algorithm that outperforms the conventional clustering algorithms, such as k-means and fuzzy c-means, by 10% in terms of accuracy and more than 30% in terms of cluster validity. The diffraction-based clustering algorithm is also independent of any parameters and is able to automatically determine the correct number of clusters in the data.