Bibliography


In this work, we have used eleven benchmark cancer data sets of microarray gene expression, which are obtained from http://csse.szu.edu.cn/staff/zhuzx/Datasets.html[21] to evaluate our proposed approaches. The datasets used in the experiments have large number of dimensions (thousands of genes) and consist of two and more than two classes. The cancer data sets are of Colon Tumor, Leukemia 2-class, Leukemia 3-class, Leukemia 4-class, Central Nervous System, Breast, Lung, Ovarian, Lymphoma, mixed-lineage leukemia (MLL), and small round blue cell tumors (SRBCT). The detailed description of these microarray datasets are given below:

- **Central Nervous System (CNS) dataset**: The dataset contains the study of embryonal tumours of the central nervous system (CNS) originally analyzed by Pomeroy et al. [96]. It consists of 7129 gene expressions from 60 samples including 39 medulloblastoma survivors and 21 treatment failures. The sample of this dataset is shown in Table A.1.

- **Breast cancer dataset**: The Breast cancer data set originally studied by van’t Veer et al. [110]. It includes analyses of 97 patient samples, 46 of which are relapse patients who had developed distance metastases within 5 years, and 51 patients who are non-relapsed who remained healthy for at least 5 years from the distance after their initial diagnosis. It has 24,481 gene expressions. The sample of this dataset is shown in Table A.2.

- **Leukemia 2-class dataset**: The Golub et al. (1999) [10] data of comprises of bone marrow samples of 72 patients in which 47 patients diagnosed with acute lymphoblastic leukemia (ALL) and 25 patients with acute myeloid leukemia (AML) for 7129 gene expressions. The sample of this dataset is shown in Table A.3.

- **Leukemia 3-class dataset**: This dataset comprises of samples of 72 patients, divided into 3 classes. There are 38 patients of ALL B-cell, 9 patients of ALL T-cell, and 25
patients with acute myeloid leukemia (AML). Each sample has 7129 gene expressions. The sample of this dataset is shown in Table A.4

- **Leukemia 4-class dataset**: This dataset comprises of samples of 72 patients, divided into 4 classes in which 38 patients of ALL B-cell, 9 patients of ALL T-cell, and 20 patients with bone-marrow samples of AML and 5 patients with peripheral blood samples of AML. Each sample has 7129 gene expressions. The sample of this dataset is shown in Table A.5

- **Lymphoma dataset**: The Lymphoma dataset [51] consists of expressions of 4026 genes for 62 patients diagnosed with three types of cancers of lymphatic system. Here, we have 11 patients of Chronic Lymphocytic Lymphoma (CLL), 9 patients of Follicular Lymphoma (FL) and 42 patients of Diffuse Large B-cell Lymphoma (DLBCL). The sample of this dataset is shown in Table A.6

- **Mixed-lineage leukemia (MLL) dataset**: The MLL dataset [97] contains in total 72 samples in three classes, acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), and mixed-lineage leukemia gene (MLL), which have 24, 28, 20 samples, respectively. It has 12,582 gene expressions. The sample of this dataset is shown in Table A.7

- **Ovarian dataset**: This dataset contains 253 samples among which 91 samples are of healthy persons and 162 samples are of patients suffering from ovarian cancer. Each sample contains expression information of 15,154 genes and first analyzed by Petricoin et al.[148]. The sample of this dataset is shown in Table A.8

- **Small Round Blue Cell Tumors (SRBCT) dataset**: The SRBCT dataset [85] contains gene expressions of 2308 genes for 83 samples of patients of four classes of this childhood cancer. Here, 29 samples from Ewing sarcoma (EWS(1)), 11 samples from Burkitt lymphoma (BL(2)), 18 samples from neuroblastoma (NB(3)) and 25 samples from rhabdomyosarcoma (RMS(4)) have been studied. The sample of this dataset is shown in Table A.9

- **Colon dataset**: Colon tumor (Colorectal cancer) refers to the malignant development