**Table S1. Description of Different Filter Methods**

| **Filter Method** | **Definition** | **Description** |
| --- | --- | --- |
| Modified Fisher score (Gu 2011) |  | Here modified Fisher score of feature is presented.  Here represents the number of samples of class , and are the mean and standard deviation of th feature for th class respectively, while and are the mean and standard deviation of th feature . |
| Modified T-test (Zhou 2007) | t( (2) | Here modified t-test of feature is presented.  Here, and are the mean of th feature for th class and mean of th feature respectively. is the sum of within class standard deviation and is represented as,  While  , is the number of samples for th class, is lth sample’s value for th feature in th class. Here and are total number of samples and total number of classes respectively. |
| Chi Square (Das 2019) |  | Chi Square is a statistical test which is commonly used to compare observed data with the expected data according to a specific hypothesis. This test is also used as a measure to find class discrimination capability of a gene with respect to class vector. The chi square value of every gene vector is calculated in equation (3).  Here, is the set of all distinct values present in a gene, if the gene is discretized. and represent the number of observed and expected co-occurrence of value and appeared in gene and class vector respectively. |
| Mutual Information (Das 2019) | Mutual information is also represented using entropy measure shown in equation (5). | Mutual information is an important measure based on information theory. Mutual information between a gene and a class vector can be defined as in equation (4).  Here,  is the entropy or amount of information present in and is the entropy or amount of information present in in presence of . Here, is the set of all distinct values present in a gene vector, if the gene vector is discretized. |
| Relief-F (Das 2019) |  | Relief-F measures class discrimination power of a feature according to its distinguishing capability between near instances. For any random instance , ReliefF finds the nearest instance from same class to find a hit and a miss from different class and according to that is increased or decreased using equation (6). This process is run for w number of times for different instances. |
| Pearson Correlation Coefficient (Leung 2010) | (7) | Pearson Correlation is used for detecting the linear relationship between two vectors. The equation (7) is used to calculate the PC (ρ) between the independent vector x and dependent vector y. For this paper x is considered as and y as class vector C. |
| Signal-to-noise ratio (Leung 2010) |  | The signal to noise ratio test is a feature selection method that selects significant features according to their expression levels using S/R test.  Here and are the mean and standard deviation of th feature for th class respectively, while and are the mean and standard deviation of th feature |