Leukemia Dataset Comparison Chart

Data Shape :

(74, 5148)

Preprocessing :

Standard Scaler Used

Test – Train Split :

Test -size : 20%

((57, 5147), (15, 5147))

Assumptions :

Total Features : 5147 (Excluding Target Class)

Features Extracted by Filter Methods(For A) : 977 (Around 19%)

For B and C : Cascade Filtering : $N \rightarrow 2N/3 \rightarrow N/3 = 5147 \rightarrow 3431 \rightarrow 1715$

For D : As Wrapper Methods are more time consuming then filter methods so I have extracted 500 features from the actual feature list using F- Classification Filter Method, followed by applying the wrapper methods to extract :

- Sequential Forward Search : 100 Features
- Sequential Backward Search : 400 Features

Given,

F1 – Mutual Information

F2 – F- Classification

F3 - T-Test

SFS - Sequential Forwards Search

SBS - Sequential Backward Search

Comparison Chart of KNN Classifier

Parameters / Methods	Execution Time (seconds)	Accuracy	F-Score
F1	**27.77**	**93.33**	**0.93**
F2	**488.727**	**100**	**0.8235**
F3	**3.398**	**93.33**	**0.93**
F1 U F2 U F3	**519.88**	**100**	**1**
F1→ F2 →F3	**307.09**	**86.6**	**0.87**
F2→ F3→ F1	**574.36**	**86.6**	**0.87**
F3→ F1→ F2	**140.34**	**93.33**	**0.933**
SFS	**703**	**93.33**	**0.9333**
SBS	**713**	**93.33**	**0.9333**

NOTE :

- Considering only the best results for the Accuracy and F-Score.
- F-Score measure is derived from Confusion Matrix and it defines its characteristics.
- The KNN classifier was run for k=1 to 20 and the k value for which maximum accuracy and fscore has been found is mentioned above.

Comparison Chart of SVM Classifier

Parameters / Methods	Execution Time (seconds)	Accuracy	F-Score
F1	**27.77**	**80.0**	**0.823**
F2	**488.727**	**80.0**	**0.823**
F3	**3.398**	**80.0**	**0.823**
F1 U F2 U F3	**519.88**	**93.33**	**0.9333**
F1→ F2 →F3	**307.09**	**80.0**	**0.8235**
F2→ F3→ F1	**574.36**	**80.0**	**0.8235**
F3→ F1→ F2	**140.34**	**80.0**	**0.8235**
SFS	**703**	**80.0**	**0.8235**
SBS	**713**	**80.0**	**0.8235**

Inferences:

- We get the best model for F1 U F2 U F3 , having the highest accuracy and F-Score when KNN classifier is used.But it takes a lot of time to compute.
- Based on time T-Test takes the least amount of time and provides a good accuracy and F-Score when KNN is applied
- Among the Filter methods F-Classif has the highest accuracy(KNN Classifier),although it takes a high computation time
- The Wrapper Methods are resulting in a good model with high F-Score and high Accuray for KNN but takes high computation time.
- Overall KNN performs better than SVM
- Among hybrid method F3→ F1→ F2 performs well in less time and results into high accuracy and F-Score