

# DLBCL Dataset Comparison Chart

---

## Data Shape :

(79, 7071)

## Preprocessing :

Standard Scaler Used

## Test – Train Split :

Test -size : 20%

((61, 7070), (16, 7070))

## Assumptions :

**Total Features :** 7069 (Excluding Target Class)

**Features Extracted by Filter Methods( For A) :** 1343 (Around 19%)

**For B and C :** Cascade Filtering :  $N \rightarrow 2N/3 \rightarrow N/3 = 7069 \rightarrow 4713 \rightarrow 2356$

**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	229.06	93.75	0.9375
F2	927.994	100	1
F3	14.19	100	1
F1 U F2 U F3	1171.244	100	1
F1 → F2 → F3	349.56	93.75	0.9375
F2 → F3 → F1	543.84	93.75	0.9375
F3 → F1 → F2	143.98	93.75	0.9375
SFS	1139.42	87.5	0.875
SBS	1159.426	93.75	0.9375

### NOTE :

- Considering only the best results for the Accuracy and F-Score.
- F-Score measure is derived from the confusion matrix and 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	229.06	75	0.75
F2	927.994	68.75	0.6875
F3	14.19	68.75	0.6875
F1 U F2 U F3	1171.244	68.75	0.6875
F1 → F2 → F3	349.56	68.75	0.6875
F2 → F3 → F1	543.84	68.75	0.6875
F3 → F1 → F2	143.98	68.75	0.6875
SFS	1139.42	68.75	0.6875
SBS	1159.426	68.75	0.6875

### Inferences:

- We get the best model for F3 – T-Test having both high accuracy and high F-Score and low Computation Time.
- The accuracy and F-Score values for SVM classifier are not much affected by changing Filter of Wrapper methods. KNN Classifier performs better than SVM. The best SVM model is given by filter method F1- Mutual Information.
- For KNN, the best Hybrid Cascading model is F3 → F1 → F2 having both high accuracy, high F-Score and low Computation time.
- The wrapper methods are highly time consuming.