Lungs Dataset Comparison Chart

Data Shape :

(205, 12601)

Preprocessing :

Standard Scaler Used

Test – Train Split :

Test -size : 20%

((162, 12600), (41, 12600))

Assumptions :

Total Features : 12600 (Excluding Target Class)

Features Extracted by Filter Methods(For A) : 2394 (Around 19%)

For B and C : Cascade Filtering : N → 2N/3 → N/3 = 12600 → 8400 → 4200

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	**2109.199**	**95.12**	**0.9512**
F2	**2158.398**	**85.36**	**0.8536**
F3	**147.65**	**97.56**	**0.9756**
F1 U F2 U F3	**4415.257**	**85.36**	**0.8536**
F1→ F2 →F3	**2027.608**	**97.56**	**0.9756**
F2→ F3→ F1	**3277.77**	**97.56**	**0.9756**
F3→ F1→ F2	**1019.67**	**95.12**	**0.9512**
SFS	**1978.713**	**82.92**	**0.8292**
SBS	**5694.713**	**85.36**	**0.8536**

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	**2109.199**	**87.80**	**0.8780**
F2	**2158.398**	**85.36**	**0.8536**
F3	**147.65**	**87.80**	**0.8780**
F1 U F2 U F3	**4415.257**	**85.36**	**0.8536**
F1→ F2 →F3	**2027.608**	**90.24**	**0.9024**
F2→ F3→ F1	**3277.77**	**90.24**	**0.9024**
F3→ F1→ F2	**1019.67**	**87.80**	**0.8780**
SFS	**1978.713**	**82.92**	**0.8292**
SBS	**5694.713**	**80.48**	**0.8048**

Inferences:

- We get the best model for the F3- T-Test with KNN which performs pretty good with low execution time.
- Overall KNN Classifier performs better than SVM Classifier
- Cascaded Hybrid model F1→ F2 →F3 performs good with high accuracy, high F-Score and low Execution Time among the other cascaded methods
- Wrapper methods are highly time consuming where SFS and SBS performances are quite close, with SFS taking lesser execution time comparatively.