Acronyms

A1 = Results from Mutual Information Filter
A2 = Results from F_Classif Filter
A3 = Results from t-test Filter

B = Result from Part B of Assignment

C1 = Result from (a) of Part C (F1 -> F2 -> F3)
C2 = Result from (b) of Part C (F2 -> F3 -> F1)
C3 = Result from (c) of Part C (F3 -> F1 -> F2)

D = Result from Part D of Assignment

Parameters & Stats

Number of Features

Dataset	Number of Features
DBCL	7070
Leukemia	5147
Lung	12600

Number of Selected Features (N)

Dataset	Number of Features
DBCL	1413
Leukemia	1028
Lung	2519

Parameters for Classifiers

Classifier:	kNN	SVM
Parameter:	N_neighbor = 3	Kernel = "RBF"

DBCL Dataset (D1)

Accuracy

Method	A1	A2	A3	B	C1	C2	C3	D
kNN	0.94	1.00	1.00	0.8	0.90	0.95	0.95	0.95
SVM	0.94	0.90	0.9	1.0	0.95	0.95	0.95	1.00

Weighted F-Score

Method	A1	A2	A3	B	C1	C2	C3	D
kNN	0.94	1.00	1.00	0.8	0.89	0.94	0.94	0.94
SVM	0.94	0.89	0.89	1.0	0.94	0.94	0.94	1.00

Leukemia Dataset (D2)

Accuracy

Method	A1	A2	A3	B	C1	C2	C3	D
kNN	0.944	0.944	0.944	0.888	1.0	1.0	1.0	1.0
SVM	0.944	0.944	0.944	0.888	1.0	1.0	1.0	1.0

Weighted F-Score

Method	A1	A2	A3	B	C1	C2	C3	D
kNN	0.943	0.943	0.943	0.886	1.0	1.0	1.0	1.0
SVM	0.943	0.943	0.943	0.886	1.0	1.0	1.0	1.0

Lung Dataset (D3)

Accuracy

Method	A1	A2	A3	B	C1	C2	C3	D
kNN	0.862	0.823	0.843	0.941	0.882	0.960	0.921	0.95
SVM	0.823	0.862	0.803	0.882	0.843	0.921	0.843	1.00

Weighted F-Score

Method	A1	A2	A3	B	C1	C2	C3	D
kNN	0.836	0.784	0.818	0.932	0.869	0.959	0.912	0.957
SVM	0.782	0.833	0.749	0.852	0.809	0.894	0.812	0.916

Observations

Part A
From the results obtained on various Datasets, it can be said that these filters give a similar performance on small Datasets (e.g. D1 & D2). However, the results on Larger Dataset like D3 have a varying result.

Part B
It was expected that the union of features from all the 3 filters will give even better results. However, this wasn't the case in small Datasets like D1 & D2. However, it worked well on a larger dataset like D3.
So, it might be concluded that the method explained in B performs well on larger datasets.

Part C

Results for Part C show that they generally perform equally or better than individual filters, irrespective of the size of the dataset.

Part D

It is observed that the performance of Wrapper methods as shown in Part D gives the best result. However, one disadvantage of this method is that the features extracted are not Classifier agnostic, unlike filter methods. Hence, the features extracted using one classifier might not work well on other classifiers.