Dynamic Classifier Selection based on Complexity Analysis

André Luiz Brun
Supervisor: Prof. Dr. Alceu S. Britto Jr.
Co-Supervisor: Prof. Dr. Robert Sabourin

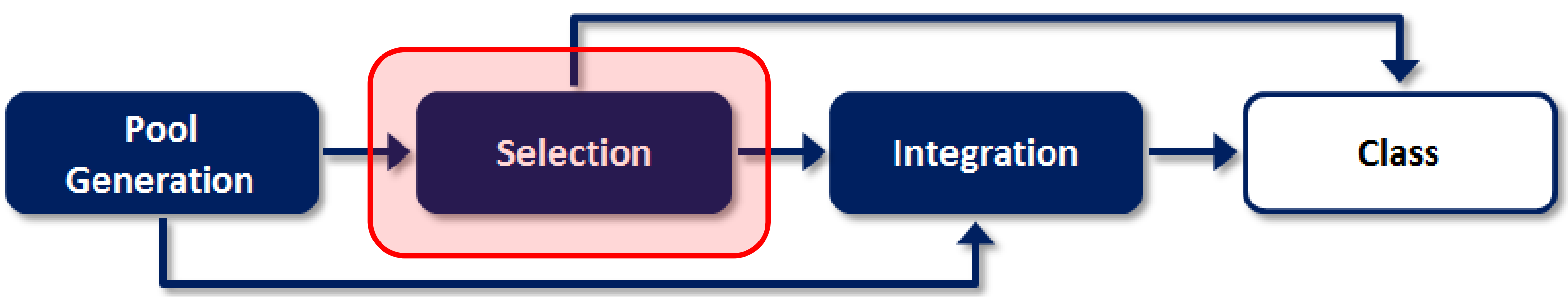
PUCPR

Introduction

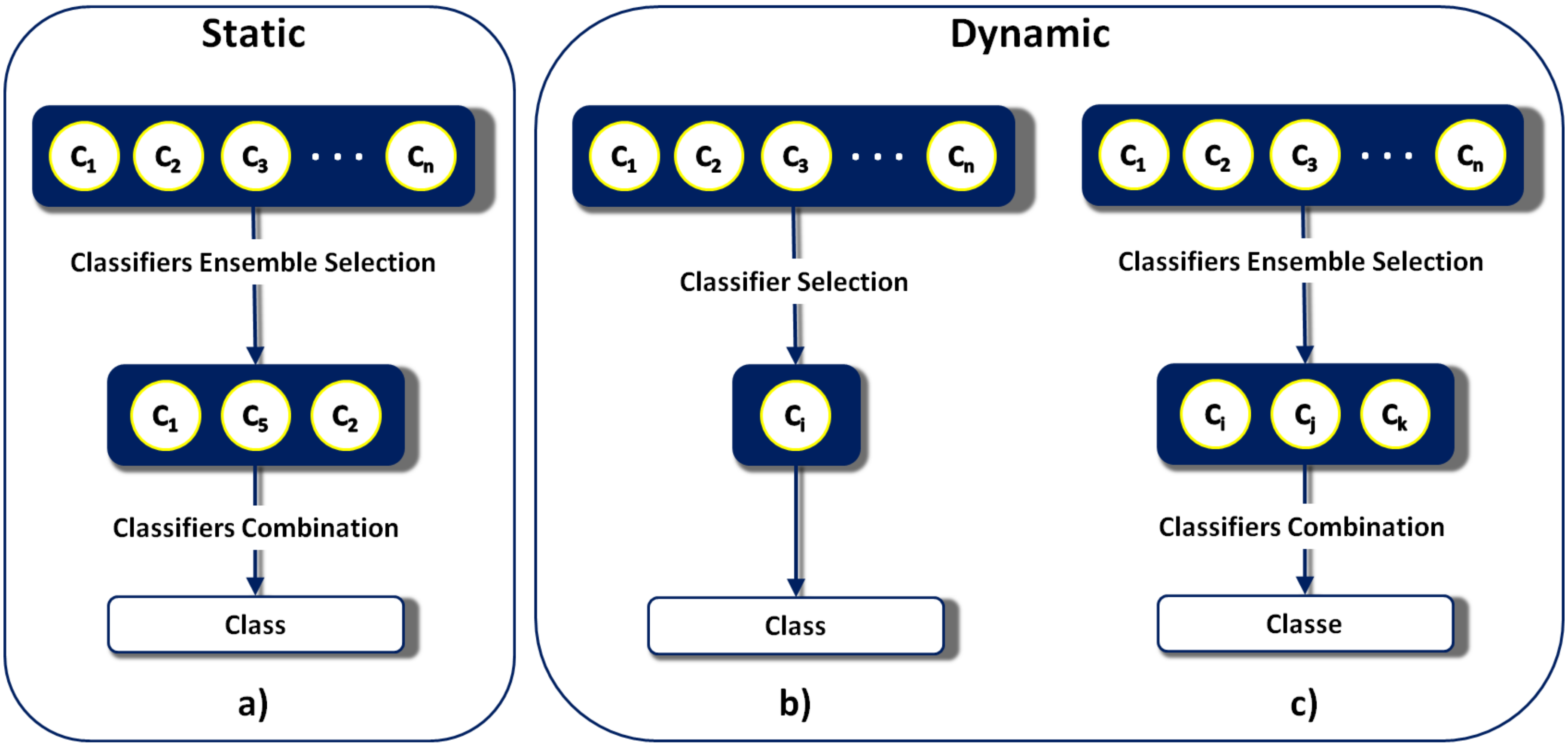
- Classification: most important task on pattern recognition
- Train a single classifier to be capable of learning the wide variability usually found in a pattern recognition is a challenging task (sometimes infeasible)

Introduction

- An interesting alternative is the use of Multiple Classifier Systems (MCSs)
  - Need to commit different erros
  - Diversity among the members



Introduction



PUCPR

Introduction

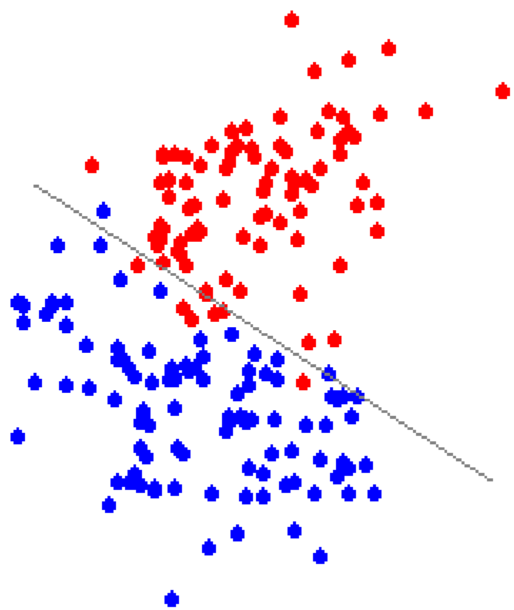
- Selection Criteria
  - Local Accuracy
  - Diversity
  - Behavioral
  - Ambiguity
  - Ranking
  - ...

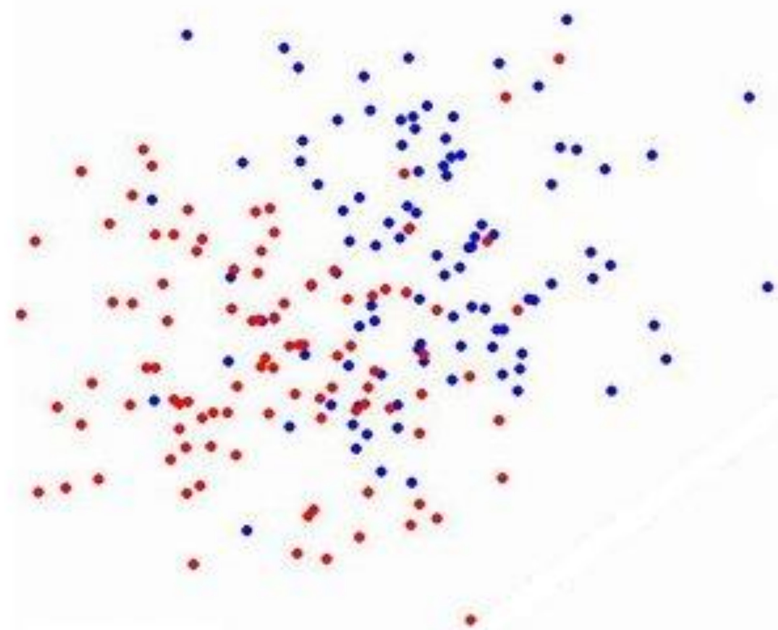
- Our hypothesis: use the problem complexity to evaluate the competence of each classifier in a pool

PUCPR

Complexity Measures

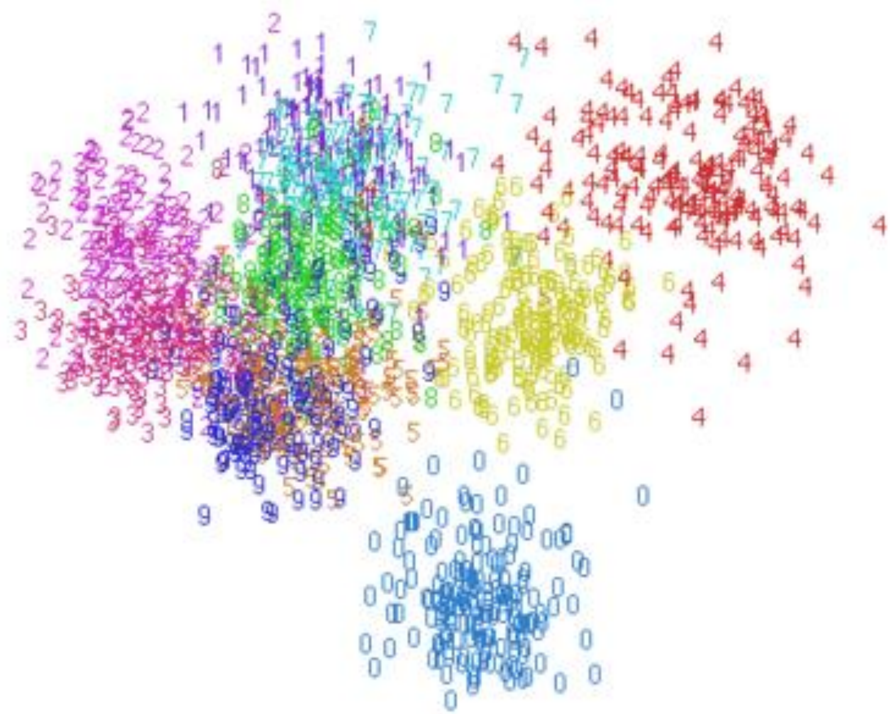
- Relationship between data characteristics and performance of classifiers





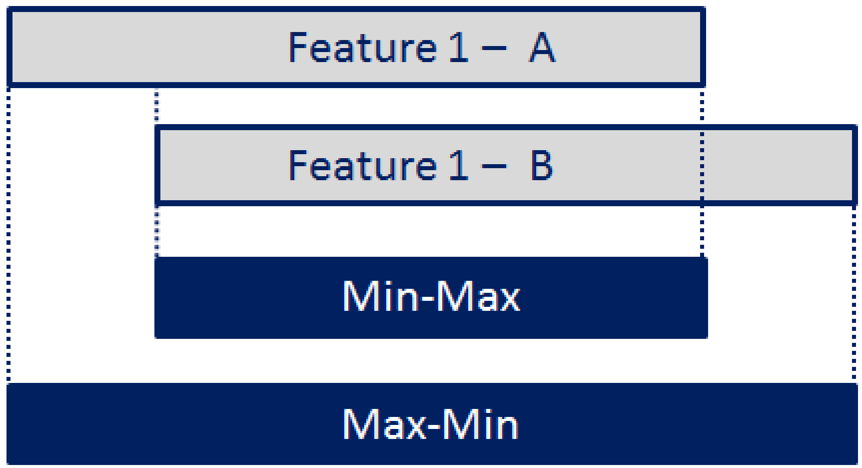
Complexity Measures

- Number of Classes, Instances, Features, etc...



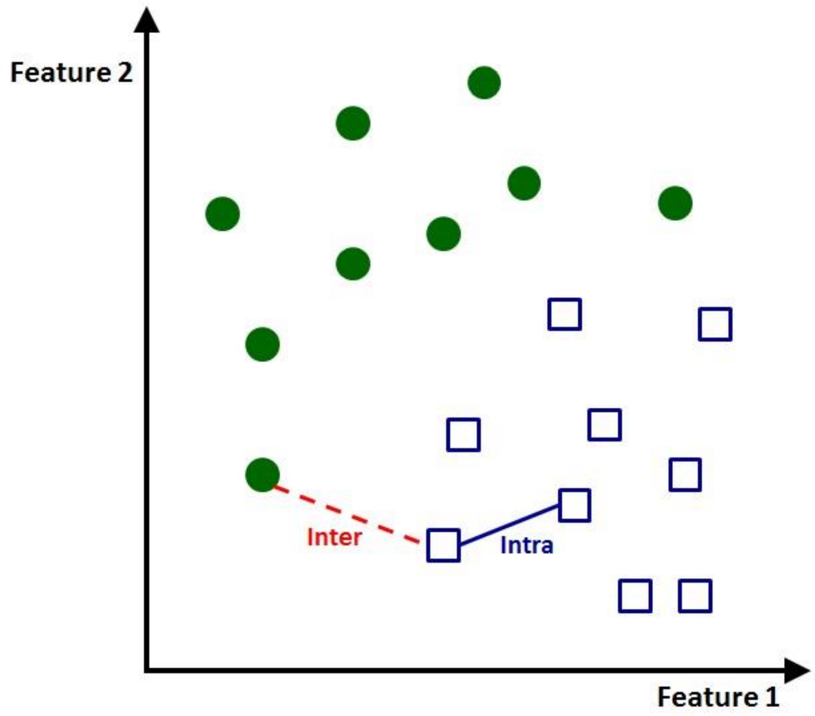
Complexity Measures

- Class Overlap
  - F1, F2, F3, F4



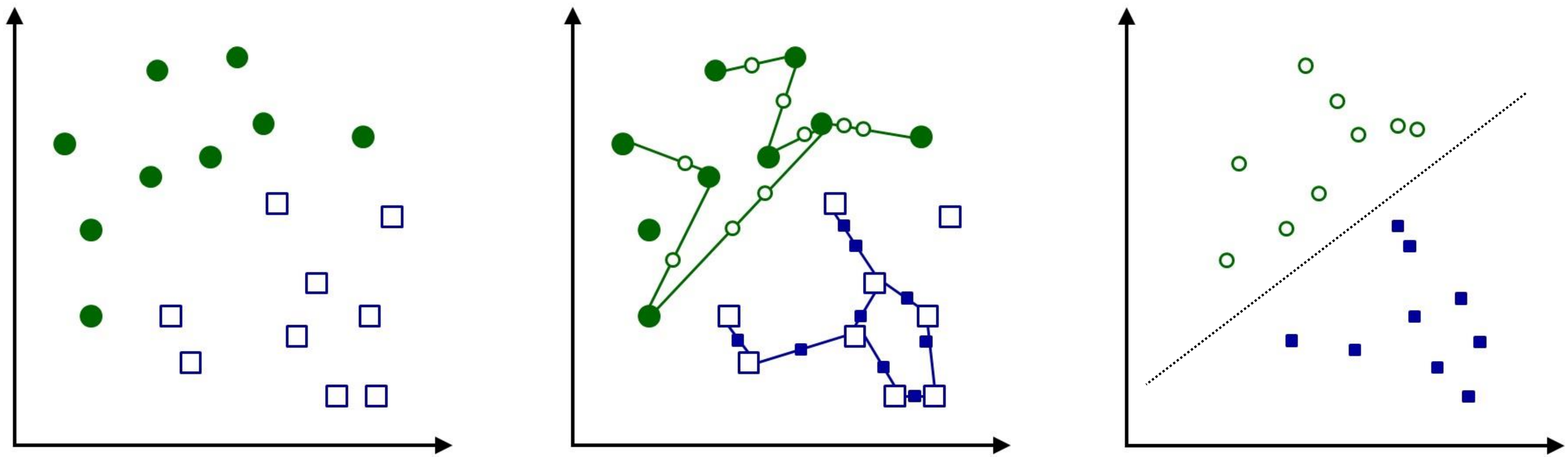
Complexity Measures

- Separability of classes
  - L1, L2, N1, N2, N3



Medidas de Complexidade

- Geometry, Topology and Density Measures
  - L3, N4, T1, T2, C1, D1, D2, D3

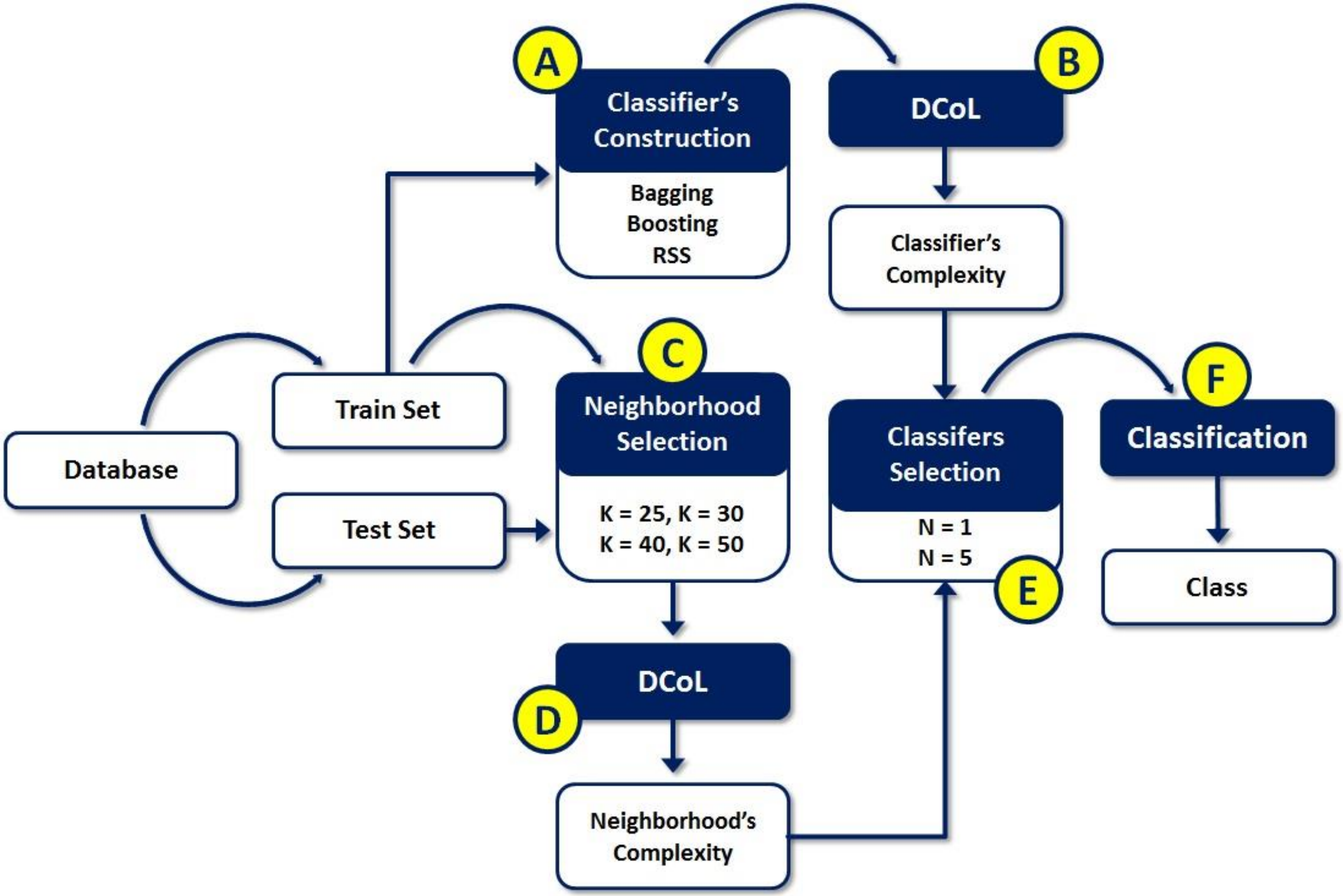


Proposed Method

- Idea: dynamically select the classifier(s) trained on data with similar complexity than that observed in the local region where the test pattern is located.

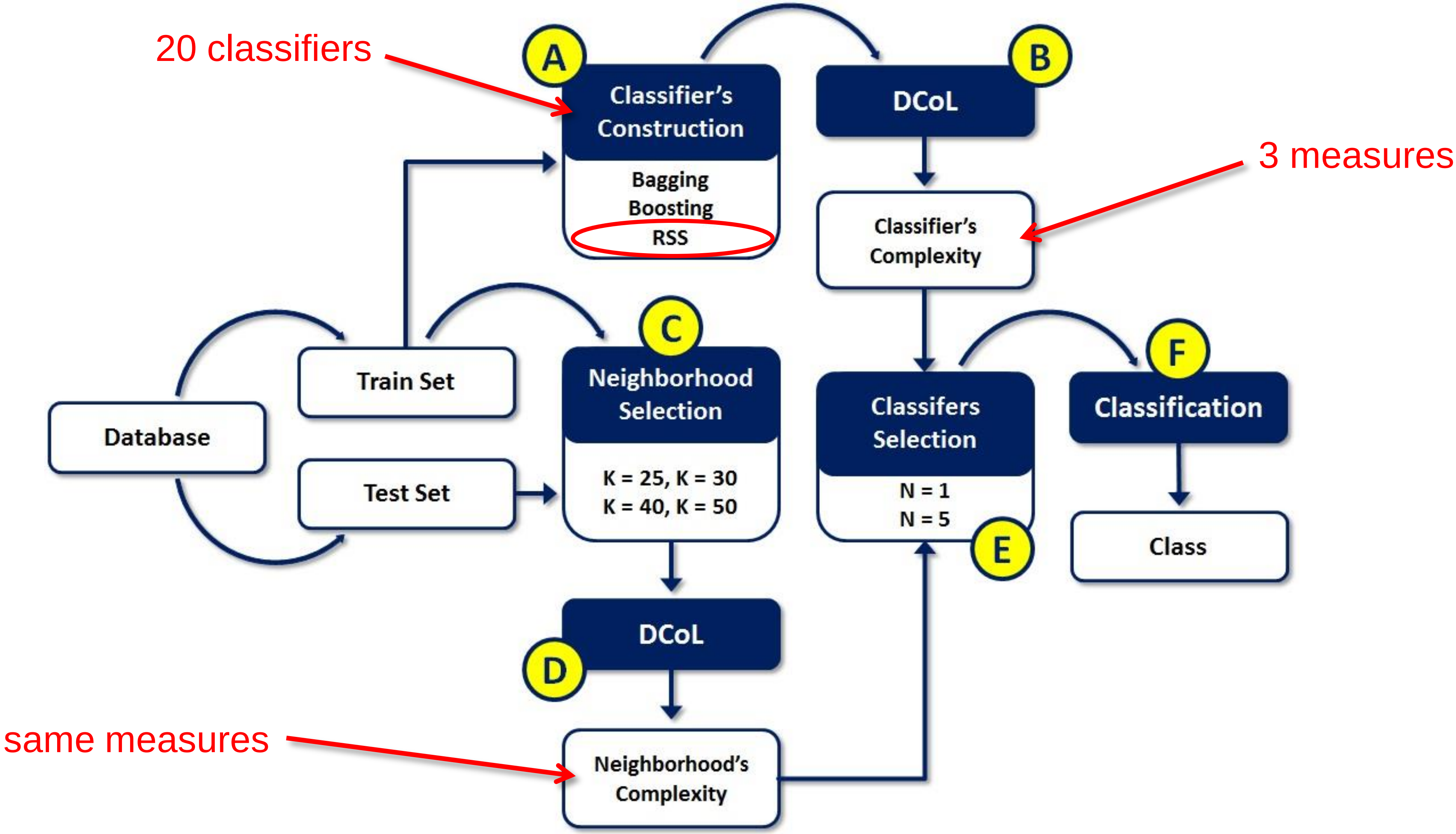
PUCPR

Proposed Method



PUCPR

Proposed Method



PUCPR

Proposed Method

- Characterization of Databases

Base	Classes	Tr	Ts	Features	RS-Card.	Bagging (%)	Boosting (%)
Blood	2	374	374	4	-	66	66
Diabetes	2	383	383	8	4	66	66
Haberman	2	154	152	3	-	66	66
Image	7	231	2079	19	4	66	66
Iris	3	75	75	4	-	66	66
Letter	26	10007	9993	16	12	66	66
Liver	2	172	173	6	3	66	66
Sonar	2	105	103	60	12	66	66
Vehicle	4	424	422	18	6	66	66
WBC	2	284	285	30	5	66	66
Wine	3	88	88	13	6	66	66
Yeast	10	745	739	8	5	66	66
Ionosphere	2	176	175	34	8	66	66

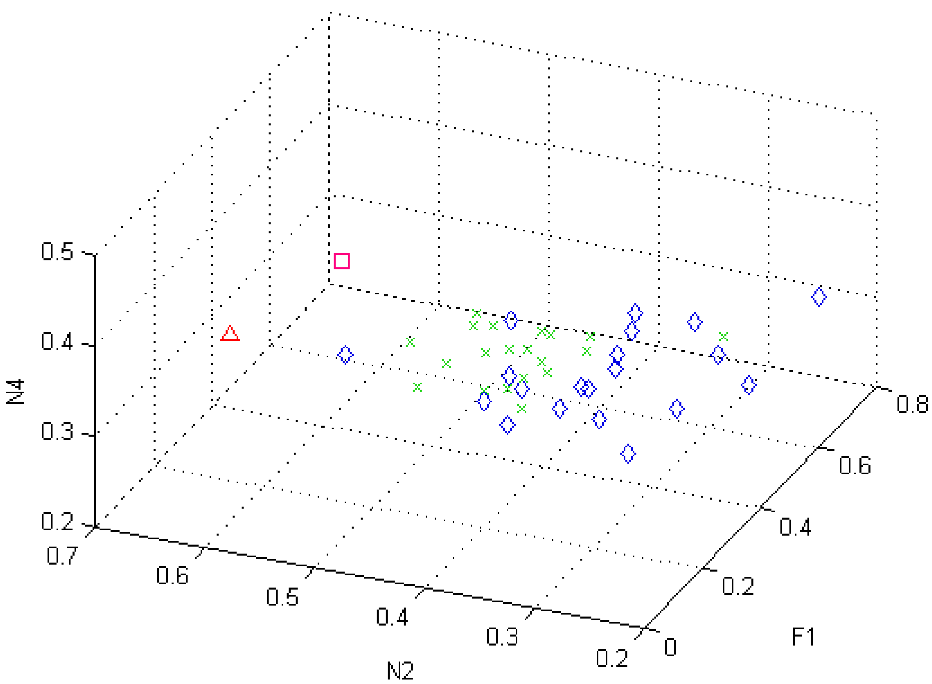
PUCPR

Results

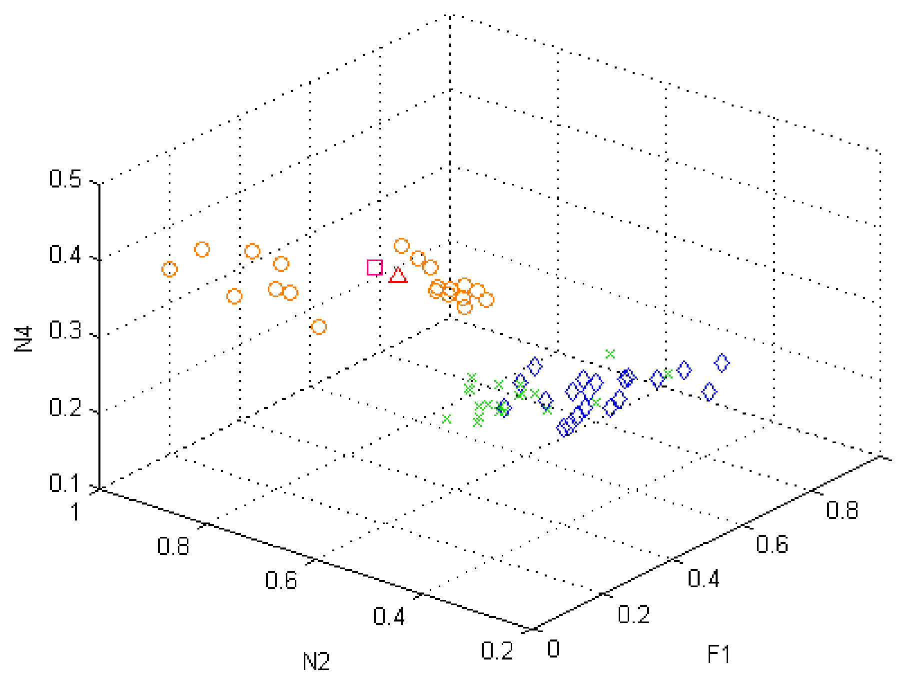
- Spacial representation (F1 x N2 x N4)
  - Bagging ◇
  - Boosting X
  - RSS O
  - Test □
  - Train △

PUCPR

Results

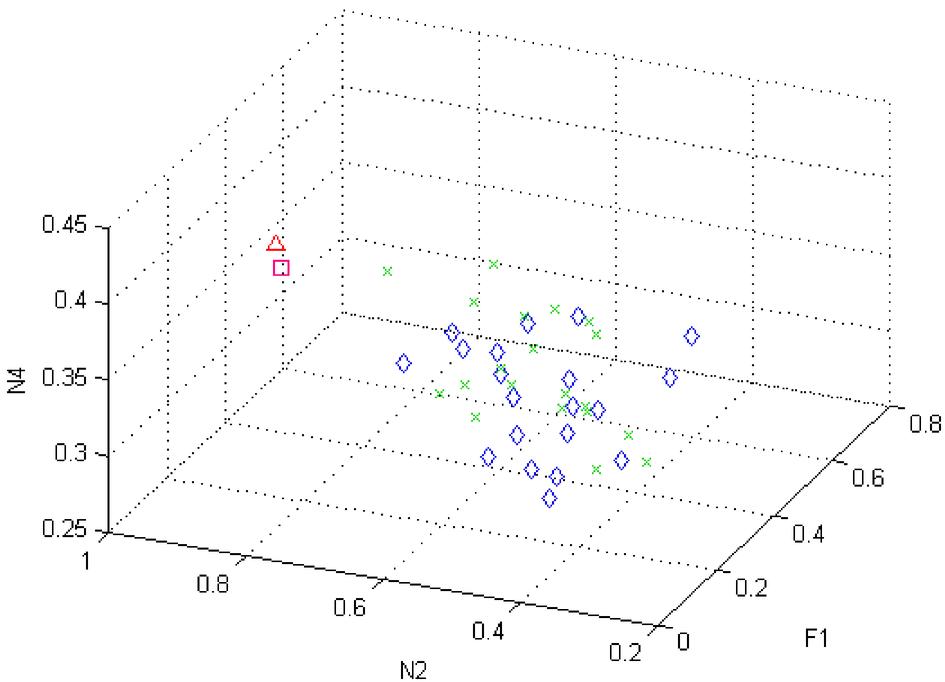


Blood

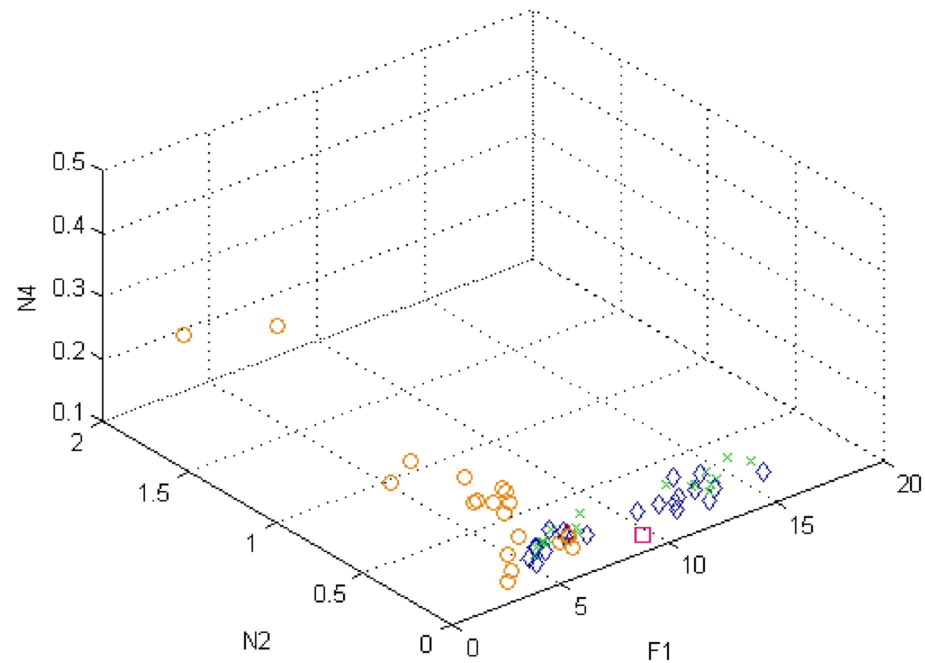


Diabetes

Results



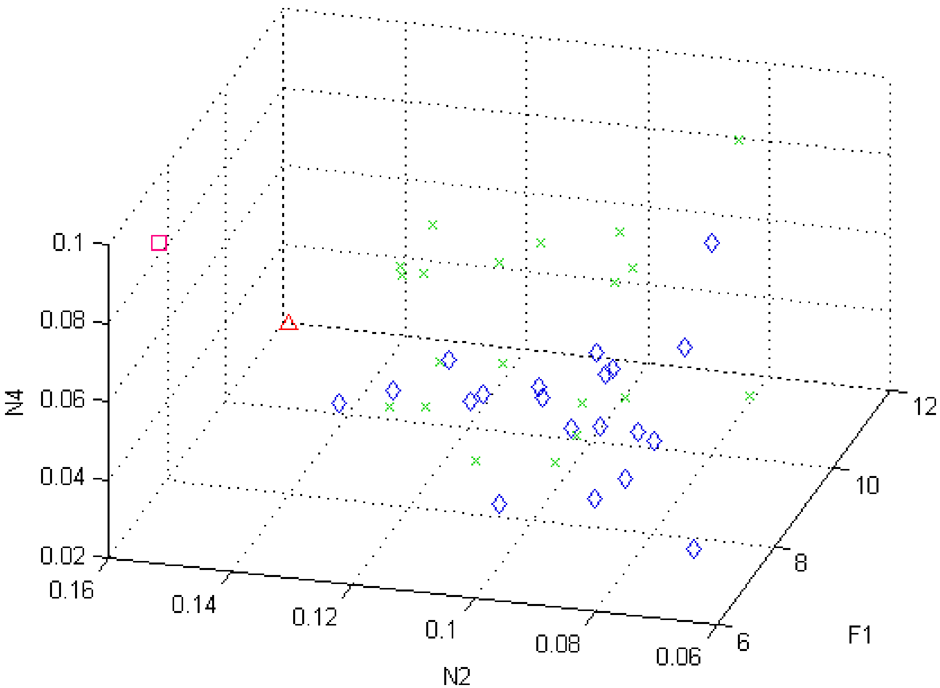
Haberman



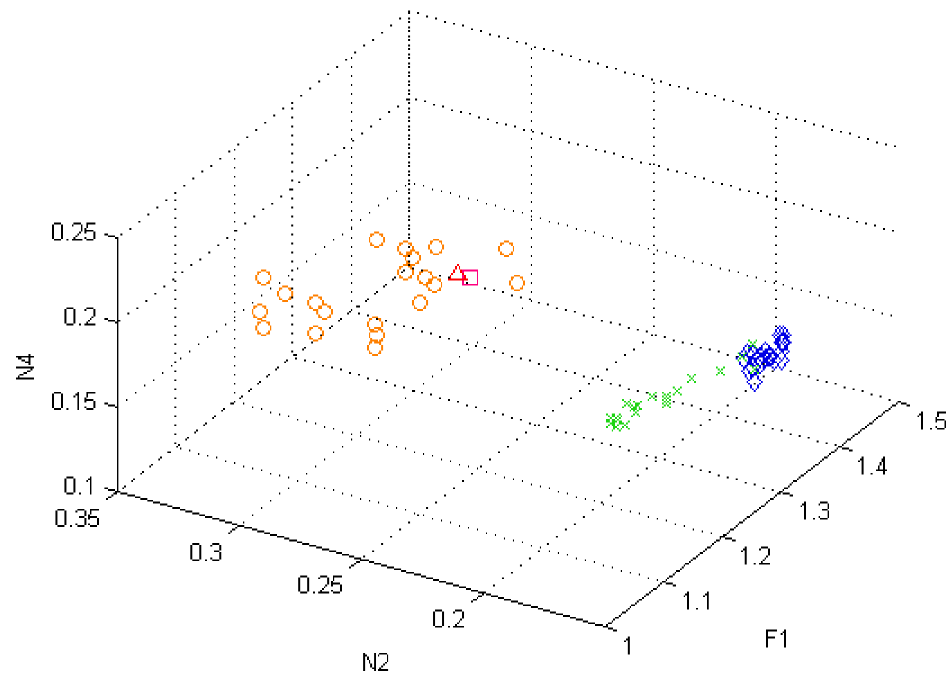
Image

PUCPR

Results



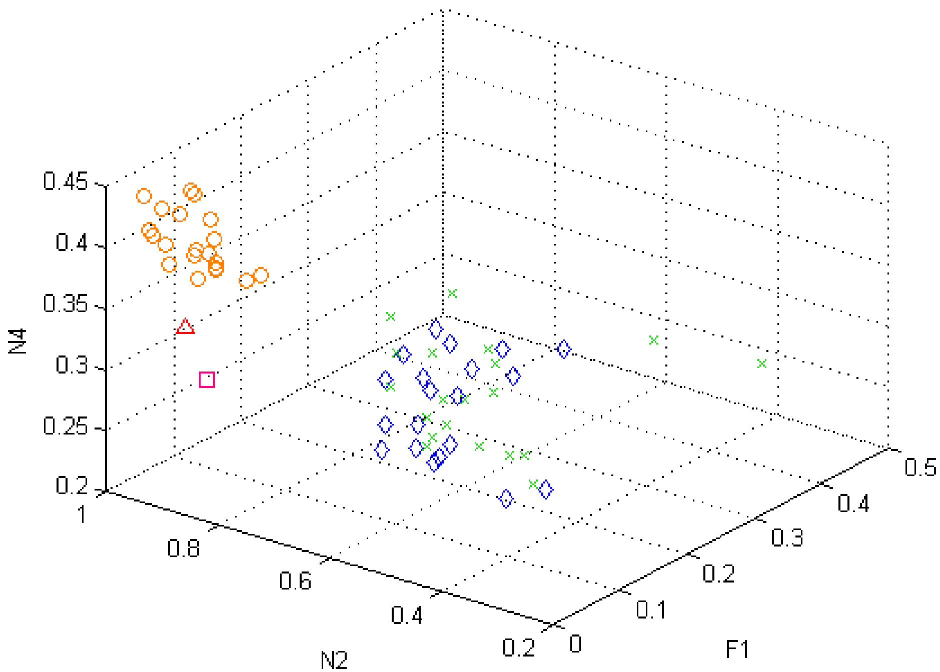
Iris



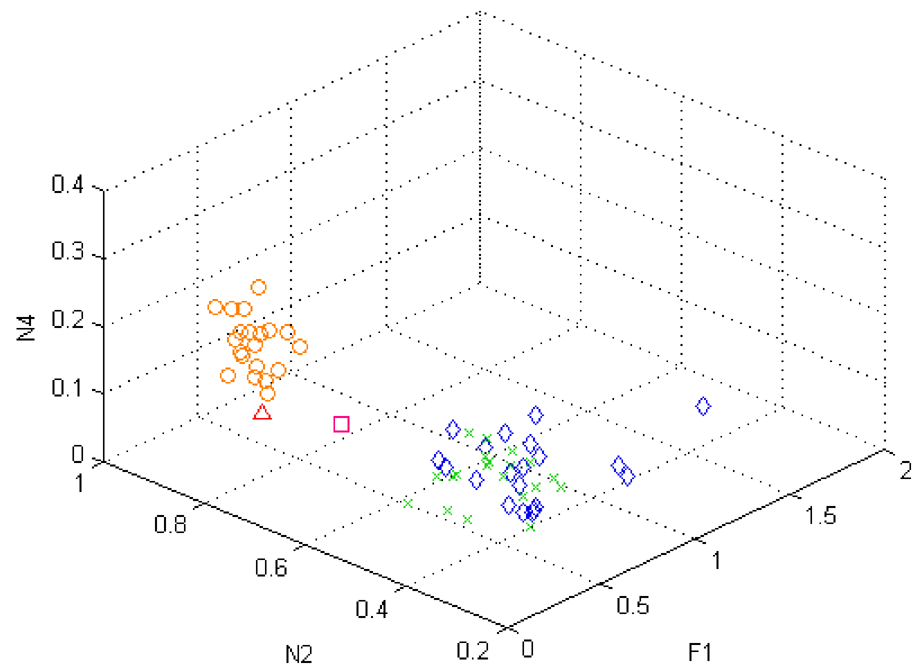
Letter

PUCPR

Results



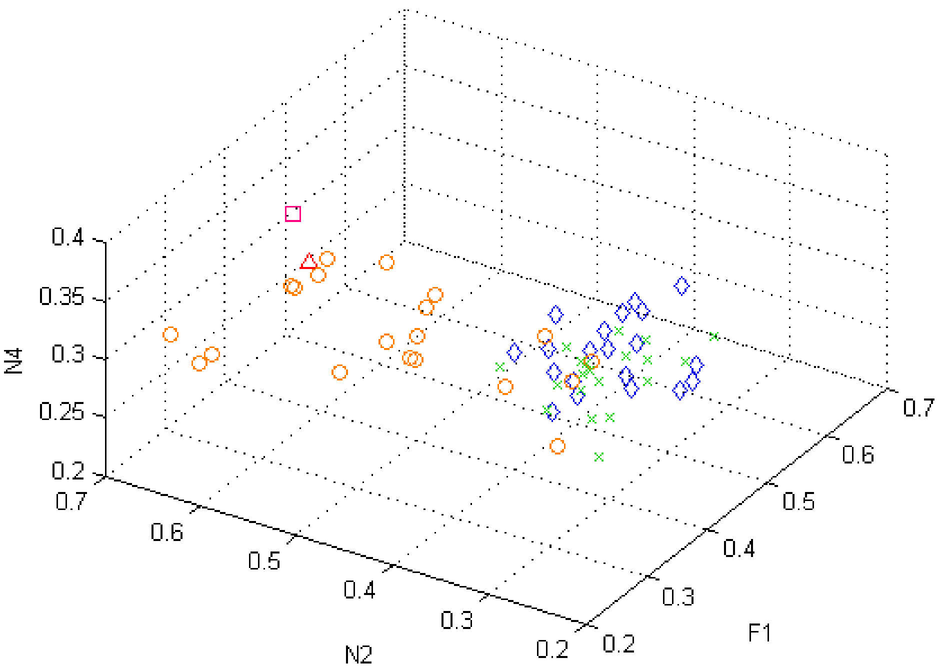
Liver



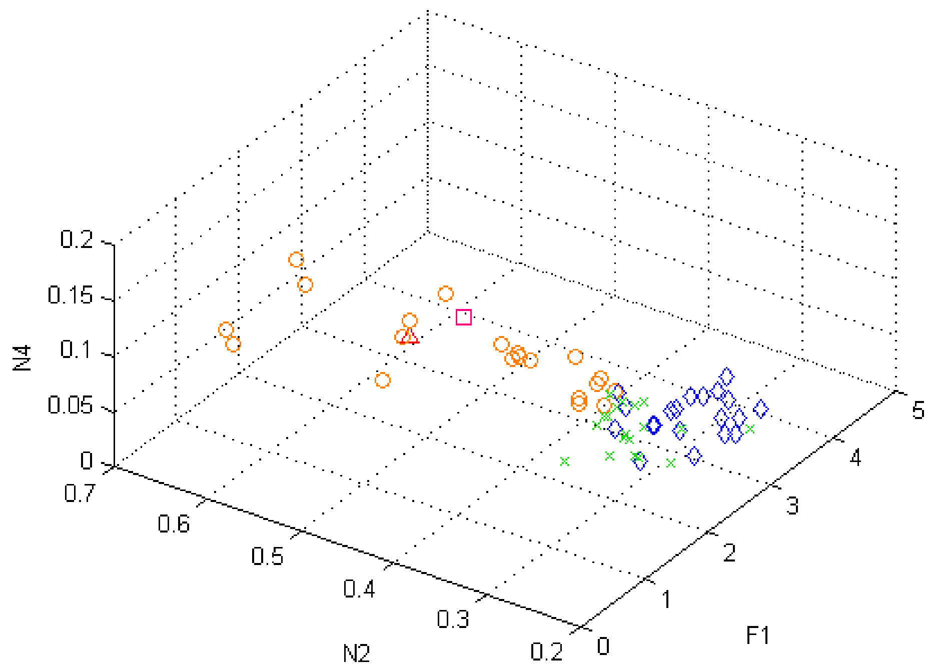
Sonar

PUCPR

Results



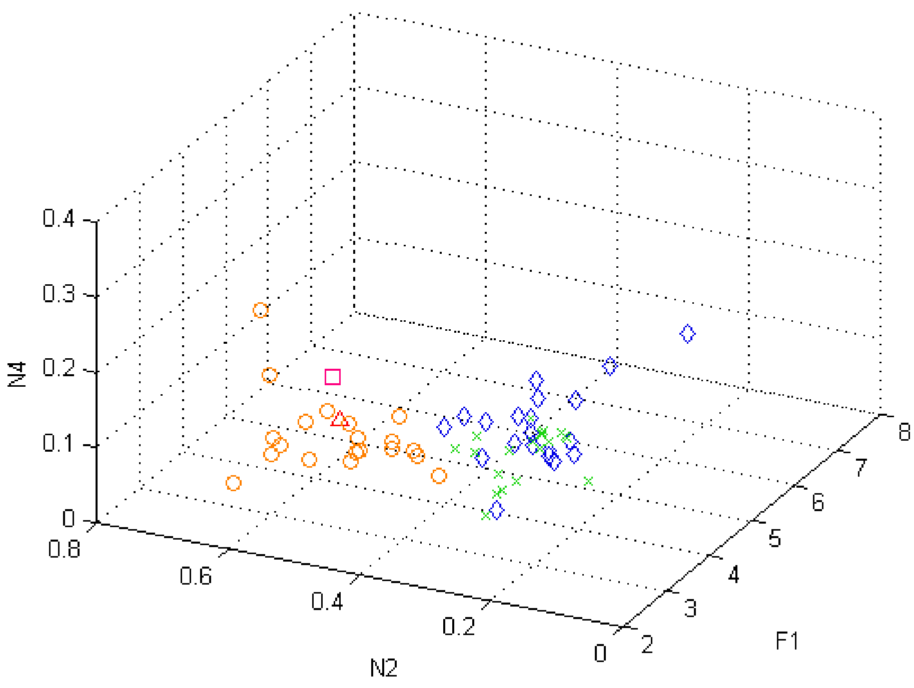
Vehicle



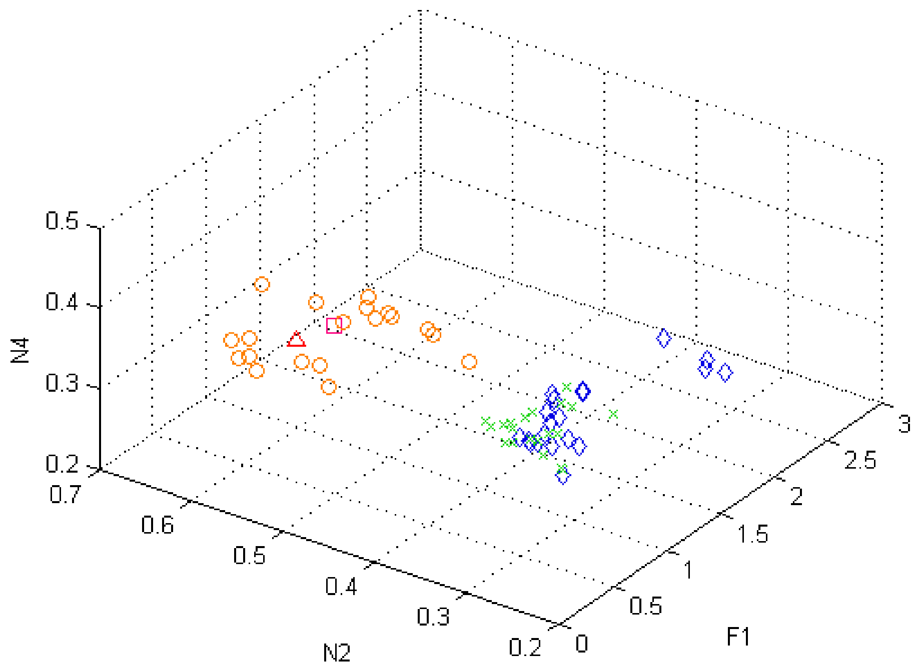
WBC

PUCPR

Results

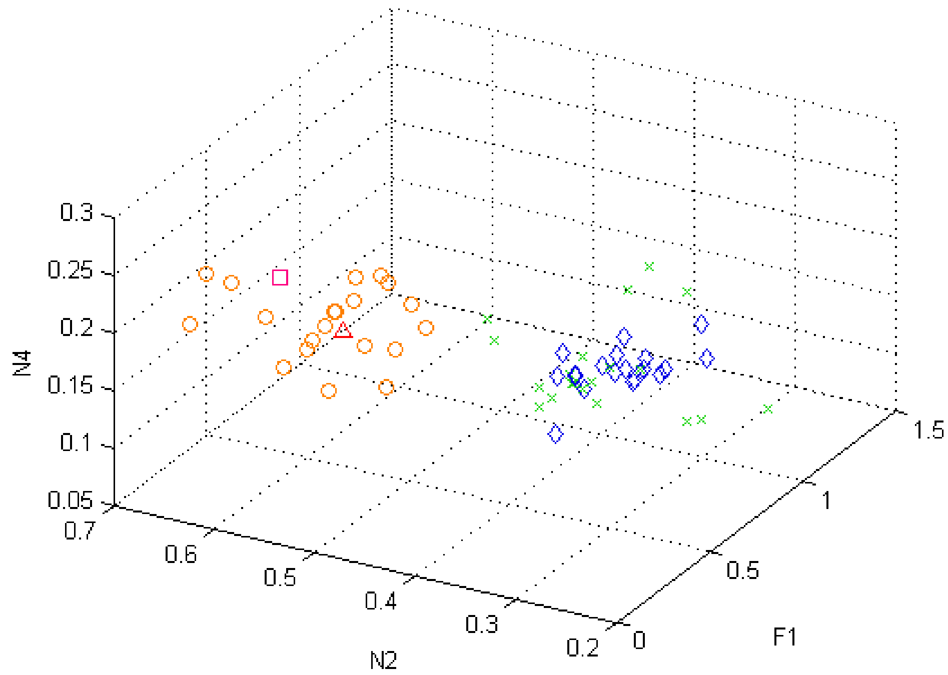


Wine



Yeast

Results



Ionosphere

Results

- Analysis of average dispersion
  - relative to classifiers centroid
  - relative to test set

  - From the 12 tested bases:
    - Yeast, Image, Iris, Vehicle e Ionosphere shown best accuracy for the most disperse pools
    - Blood, Haberman, Sonar e Wine presented best performance for the most compact pools

PUCPR

Results

- From the 12 tested bases:
  - To Haberman, Iris and Wine bases, the best pool was the farther to the test set
  - WBC, Yeast, Blood, Vehicle, Ionosphere and Liver performed best for the closest sets of test

Results

- Bagging

Base	SB	CMB	KU	KE	OLA	LCA	APR	APO	ORA	Short1	Short5
Blood	74,87	75,94	**76,20**	**76,20**	73,26	72,19	71,66	71,66	95,99	75,40	75,40
Diabetes	71,02	67,10	70,24	67,36	67,62	69,45	66,84	67,62	93,21	68,93	**71,80**
Haberman	71,05	69,08	71,71	70,4	70,4	68,41	70,4	70,4	97,37	**75,66**	**75,66**
Image	83,55	**85,28**	84,37	84,8	83,41	78,79	83,41	83,46	97,40	80,57	83,36
Iris	**96**	94,67	94,67	94,67	**96**	93,33	92	92	98,67	90,67	93,33
Liver	67,05	63,58	**67,63**	63,01	64,16	65,9	53,76	60,69	99,42	55,49	64,16
Sonar	**76,7**	**76,7**	**76,7**	72,82	**76,7**	67,96	67,96	67,96	95,15	66,02	73,79
Vehicle	61,61	59,01	60,66	60,19	57,82	54,27	56,4	56,4	92,89	57,11	**66,66**
WBC	**92,63**	91,93	91,58	91,579	90,89	88,77	90,88	90,88	98,95	90,53	91,23
Wine	75,56	72,22	70	70	73,33	72,22	68,89	68,89	93,33	**78,89**	72,22
Yeast	51,69	52,23	52,77	52,50	49,12	48,17	49,12	49,12	85,93	51,15	**55,35**
Ionosphere	85,71	81,14	84	83,43	83,43	**91,43**	83,44	83,43	92	84,57	84

PUCPR

Results

- Boosting

Base	SB	CMB	KU	KE	OLA	LCA	APR	APO	ORA	Short1	Short5
Blood	76,41	78,08	**76,74**	76,20	76,48	72,73	71,64	75,67	95,46	73	76,20
Diabetes	68,67	68,41	69,19	67,89	67,89	69,71	67,89	67,89	95,04	67,36	**69,97**
Haberman	**75**	70,4	75	73,68	69,74	73,68	69,08	70,4	96,05	74,34	74,34
Image	82,64	**86,39**	86	86,53	82,4	76,67	81,48	81,48	96,54	81,53	84,75
Iris	**96**	**96**	**96**	**96**	**96**	**96**	**96**	**96**	98,67	90,67	**96**
Liver	61,27	58,38	63,01	62,43	61,27	**67,05**	59,54	59,54	99,42	60,12	63,01
Sonar	78,64	77,67	**80,58**	78,64	75,73	72,82	67,96	72,82	98,06	74,76	77,67
Vehicle	61,14	61,37	62,79	**63,03**	59,01	54,03	55,45	55,45	92,18	60,9	61,85
WBC	**92,98**	89,47	90,88	90,88	90,88	87,02	90,88	90,88	98,6	91,58	90,88
Wine	**77,78**	71,11	71,11	71,11	75,56	70	61,11	61,11	97,78	73,33	75,56
Yeast	49,53	52,64	51,96	52,23	49,26	45,74	49,26	49,26	87,15	51,56	**54,13**
Ionosphere	83,43	80	82,86	82,86	82,86	**88,57**	82,86	82,86	89,14	84	81,14

PUCPR

Results

- RSS

Base	SB	CMB	KU	KE	OLA	LCA	APR	APO	ORA	Short1	Short5
Diabetes	70,76	71,54	**72,32**	**72,32**	61,88	65,8	67,89	61,88	99,74	67,36	70,24
Image	**87,78**	82,68	82,49	82,49	64,84	49,26	69,61	64,84	98,9	74,99	81,29
Liver	67,05	63,58	67,63	67,63	67,63	**68,21**	61,85	67,63	100	56,64	60,69
Sonar	80,58	77,67	80,58	80,58	76,7	77,67	76,7	76,7	100	75,73	**81,55**
Vehicle	**70,14**	67,3	68,01	68,01	54,27	50,24	61,85	54,27	98,58	60,66	67,77
WBC	92,28	92,63	**94,04**	**94,04**	89,12	59,3	88,87	89,12	100	85,61	90,52
Wine	**93,33**	80	76,67	76,67	67,78	58,89	90	67,78	100	87,78	90
Yeast	48,04	**55,21**	53,32	53,32	43,03	36,13	35,59	40,87	94,86	45,60	51,56
Ionosphere	88,57	89,14	**89,72**	89,14	84,57	85,71	85,14	85,14	98,29	84	89,71

Conclusions

- Proposed DSOC method has shown to be an interesting strategy for classifier/ensemble selection
- Experimental results were similar to that of related works available in the literature.
- Ensenble selection was superior to classifier selection
- Further work
  - Better understand the complexity analysis
  - Generate pools of classifiers based on complexity
  - To consider combining class accuracy and data complexity

PUCPR