

# Ensembles for predicting structured outputs

Tests for statistical significance: regression and HMLC

Dragi Kocev, Celine Vens, Jan Struyf, Sašo Džeroski

July 19, 2010

Table 1: Properties of the datasets with multiple targets;  $N$  is number of instances,  $D/C$  number of descriptive attributes (discrete/continous), and  $T$  number of target attributes.

	$N$	$D/C$	$T$
Collembola	393	8/39	3
EDM - 1	154	0/16	2
Forestry-Kras	60607	0/160	11
Forestry-Slivnica-LandSat	6218	0/150	2
Forestry-Slivnica-IRS	2731	0/29	2
Forestry-Slivnica-SPOT	2731	0/49	2
Sigma real	817	0/4	2
Soil quality 1	1944	0/142	3
Solar-flare 1	323	10/0	3
Solar-flare 2	1066	10/0	3
Vegetation Clustering	29679	0/65	11
Vegetation Condition	16967	1/39	7
Water quality	1060	0/16	14

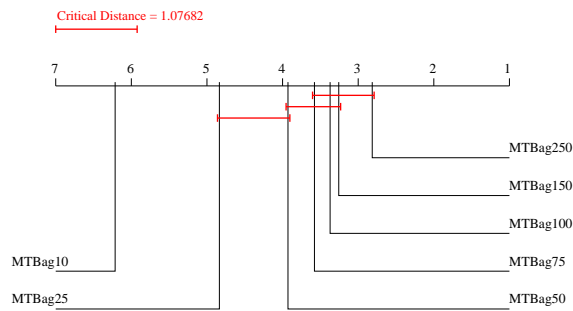
Table 2: Properties of the datasets with hierarchically organized multiple labels;  $N_{train}$  is number of instances in the training dataset,  $N_{test}$  is number of instances in the testing dataset,  $D/C$  is number of descriptive attributes (discrete/continous),  $HierarchySize$  is number of classes in the hierarchy, and  $\overline{Label}$  is average number of labels per example.

Domain	$N_{train}$	$N_{test}$	$D/C$	$HierarchySize$	$\overline{Label}$
ImageCLEF2007(Directional)	10000	1006	0/80	XX	1.0
ImageCLEF2007(Anatomical)	10000	1006	0/80	XX	1.0
Diatoms	2065	1054	0/371	XX	1.0
Enron	988	660	0/1001	XX	YY
Reuters	3000	3000	0/47236	XX	YY
WIPO	1352	358	0/74435	XX	YY
Expression (GO)	2485	1288	0/551	XX	YY
SCOP (GO)	6507	3336	0/2003	XX	YY
Sequence (FunCat)	2455	1264	2/4448	XX	YY
Yeast (GO)	2310	1155	5588/342	XX	YY

# 1 Multi-Target Regression - Correlation Coefficient

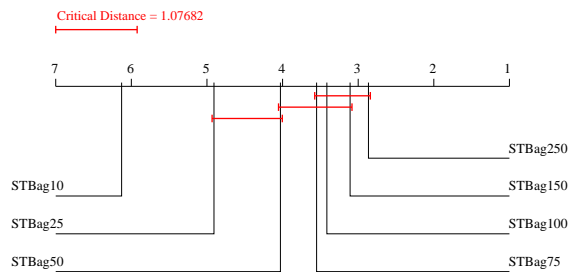
## MTRegression - CC: Multi-Target Bagging

Method	MTBag10	MTBag25	MTBag50	MTBag75	MTBag100	MTBag150	MTBag250
Average Rank	6.21	4.84	3.93	3.58	3.37	3.26	2.81



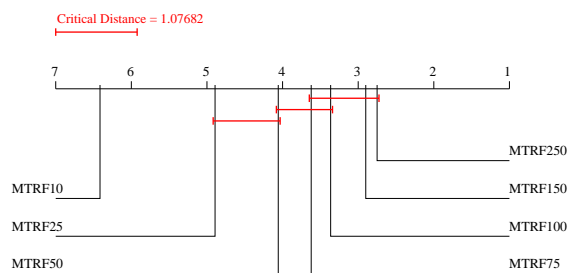
## MTRegression - CC: Single-Target Bagging

Method	STBag10	STBag25	STBag50	STBag75	STBag100	STBag150	STBag250
Average Rank	6.13	4.91	4.03	3.55	3.41	3.11	2.86



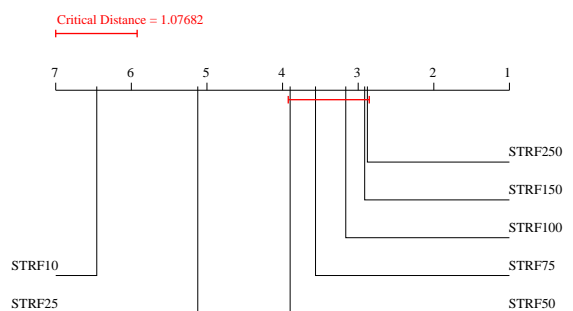
## MTRegression - CC: Multi-Target Random Forest

Method	MTRF10	MTRF25	MTRF50	MTRF75	MTRF100	MTRF150	MTRF250
Average Rank	6.41	4.89	4.06	3.62	3.36	2.90	2.75



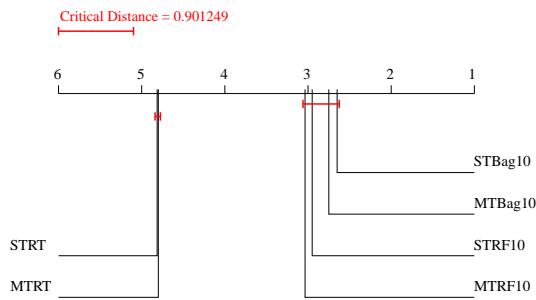
## MTRegression - CC: Single-Target Random Forest

Method	STRF10	STRF25	STRF50	STRF75	STRF100	STRF150	STRF250
Average Rank	6.46	5.12	3.90	3.56	3.16	2.91	2.88



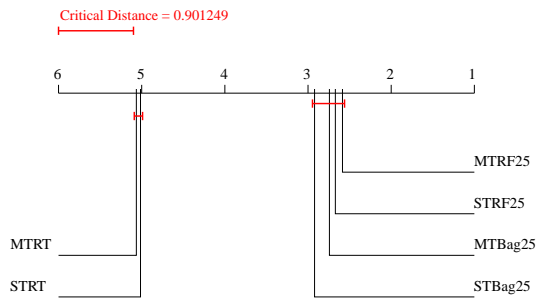
MTRegression - CC: Ensemble size 10

Method	MTRT	STRT	MTBag10	STBag10	MTRF10	STRF10
Average Rank	4.80	4.81	2.75	2.65	3.04	2.95



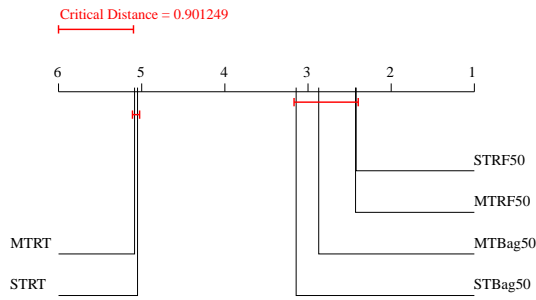
MTRegression - CC: Ensemble size 25

Method	MTRT	STRT	MTBag25	STBag25	MTRF25	STRF25
Average Rank	5.06	5.01	2.74	2.92	2.59	2.67



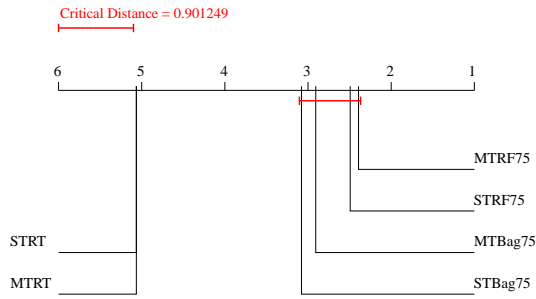
MTRegression - CC: Ensemble size 50

Method	MTRT	STRT	MTBag50	STBag50	MTRF50	STRF50
Average Rank	5.09	5.05	2.87	3.14	2.43	2.42



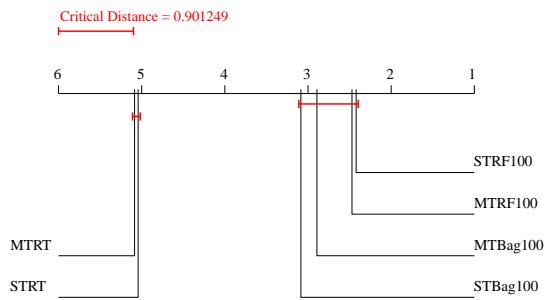
MTRegression - CC: Ensemble size 75

Method	MTRT	STRT	MTBag75	STBag75	MTRF75	STRF75
Average Rank	5.06	5.06	2.91	3.08	2.39	2.49



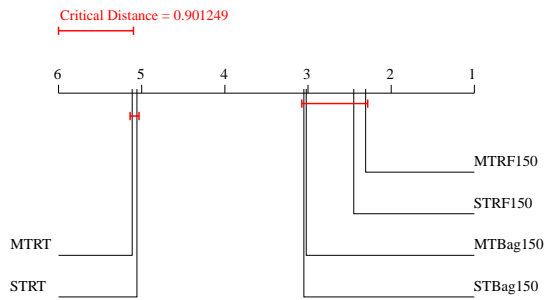
MTRegression - CC: Ensemble size 100

Method	MTRT	STRT	MTBag100	STBag100	MTRF100	STRF100
Average Rank	5.09	5.04	2.89	3.09	2.47	2.42



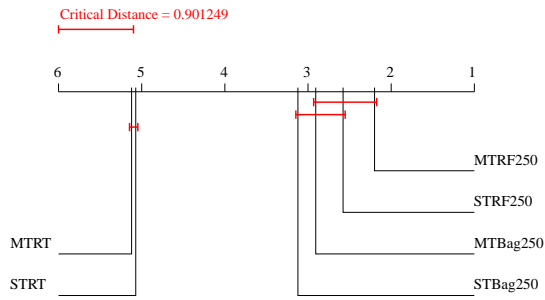
MTRegression - CC: Ensemble size 150

Method	MTRT	STRT	MTBag150	STBag150	MTRF150	STRF150
Average Rank	5.11	5.06	3.02	3.05	2.31	2.45



MTRegression - CC: Ensemble size 250

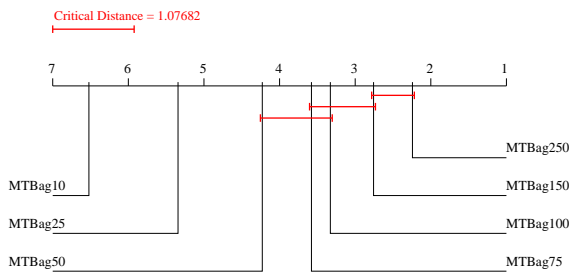
Method	MTRT	STRT	MTBag250	STBag250	MTRF250	STRF250
Average Rank	5.12	5.07	2.91	3.12	2.20	2.58



## 2 Multi-Target Regression - RMSE

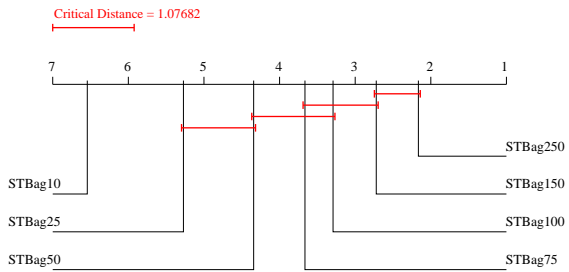
### MTRRegression - RMSE: Multi-Target Bagging

Method	MTBag10	MTBag25	MTBag50	MTBag75	MTBag100	MTBag150	MTBag250
Average Rank	6.52	5.34	4.23	3.58	3.33	2.76	2.24



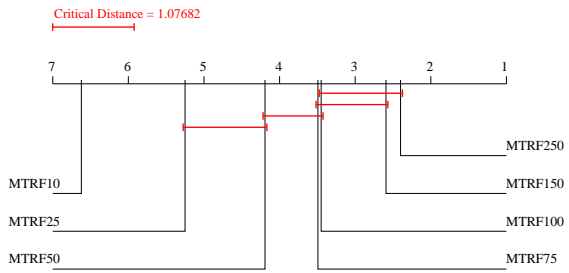
### MTRRegression - RMSE: Single-Target Bagging

Method	STBag10	STBag25	STBag50	STBag75	STBag100	STBag150	STBag250
Average Rank	6.54	5.27	4.34	3.66	3.29	2.72	2.16



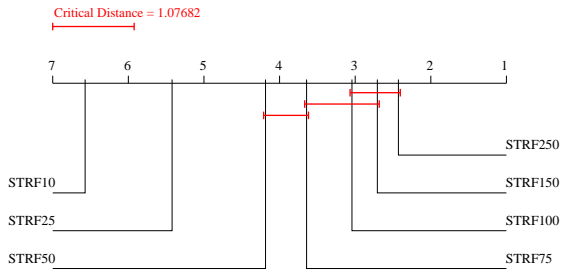
### MTRRegression - RMSE: Multi-Target Random Forest

Method	MTRF10	MTRF25	MTRF50	MTRF75	MTRF100	MTRF150	MTRF250
Average Rank	6.62	5.25	4.19	3.49	3.45	2.59	2.40



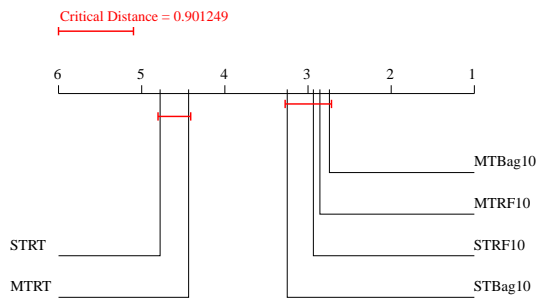
### MTRRegression - RMSE: Single-Target Random Forest

Method	STRF10	STRF25	STRF50	STRF75	STRF100	STRF150	STRF250
Average Rank	6.57	5.42	4.19	3.64	3.04	2.71	2.43



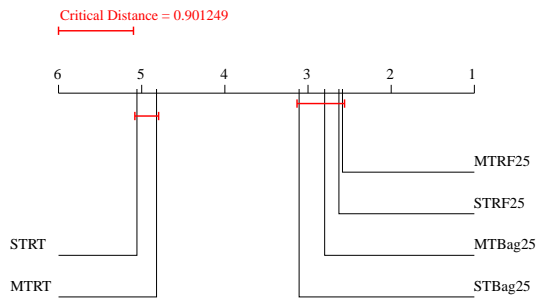
MTRegression - RMSE: Ensemble size 10

Method	MTRT	STRT	MTBag10	STBag10	MTRF10	STRF10
Average Rank	4.44	4.78	2.74	3.25	2.86	2.94



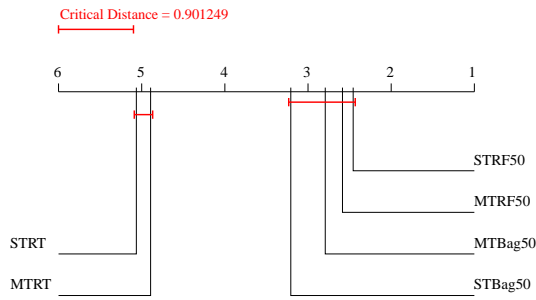
MTRegression - RMSE: Ensemble size 25

Method	MTRT	STRT	MTBag25	STBag25	MTRF25	STRF25
Average Rank	4.82	5.06	2.80	3.11	2.59	2.63



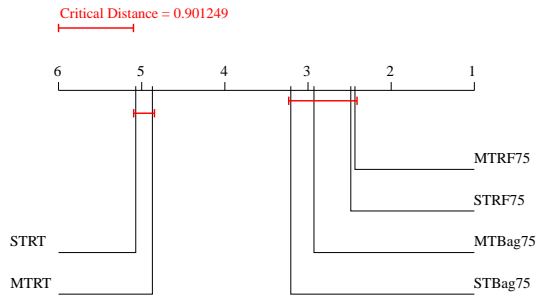
MTRegression - RMSE: Ensemble size 50

Method	MTRT	STRT	MTBag50	STBag50	MTRF50	STRF50
Average Rank	4.89	5.06	2.79	3.21	2.59	2.46



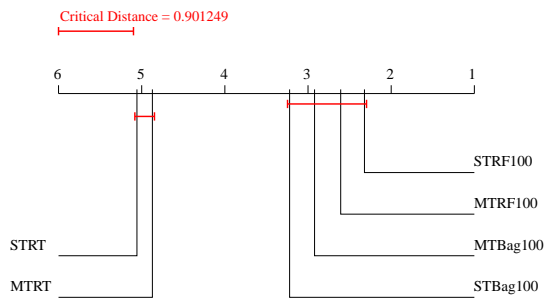
MTRegression - RMSE: Ensemble size 75

Method	MTRT	STRT	MTBag75	STBag75	MTRF75	STRF75
Average Rank	4.87	5.07	2.93	3.21	2.44	2.49



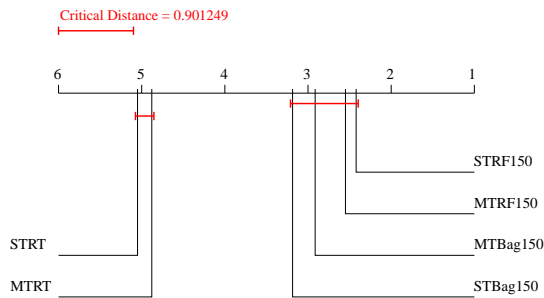
MTRegression - RMSE: Ensemble size 100

Method	MTRT	STRT	MTBag100	STBag100	MTRF100	STRF100
Average Rank	4.87	5.06	2.92	3.22	2.61	2.32



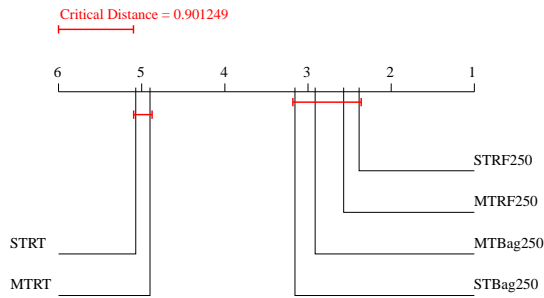
MTRegression - RMSE: Ensemble size 150

Method	MTRT	STRT	MTBag150	STBag150	MTRF150	STRF150
Average Rank	4.88	5.05	2.91	3.19	2.55	2.42



MTRegression - RMSE: Ensemble size 250

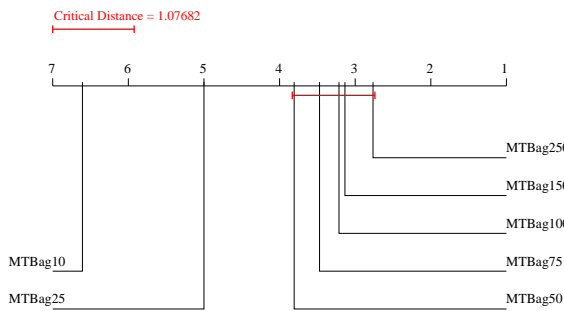
Method	MTRT	STRT	MTBag250	STBag250	MTRF250	STRF250
Average Rank	4.90	5.07	2.91	3.16	2.57	2.39



### 3 Multi-Target Regression - RRMSE

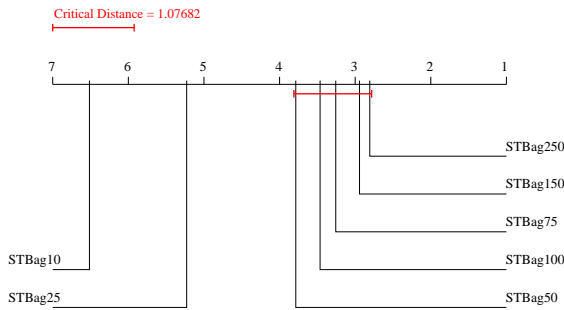
#### MTRegression - RRMSE: Multi-Target Bagging

Method	MTBag10	MTBag25	MTBag50	MTBag75	MTBag100	MTBag150	MTBag250
Average Rank	6.61	5.00	3.81	3.47	3.21	3.14	2.76



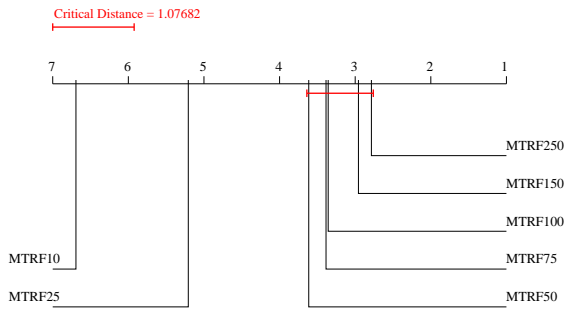
#### MTRegression - RRMSE: Single-Target Bagging

Method	STBag10	STBag25	STBag50	STBag75	STBag100	STBag150	STBag250
Average Rank	6.51	5.23	3.79	3.26	3.46	2.94	2.81



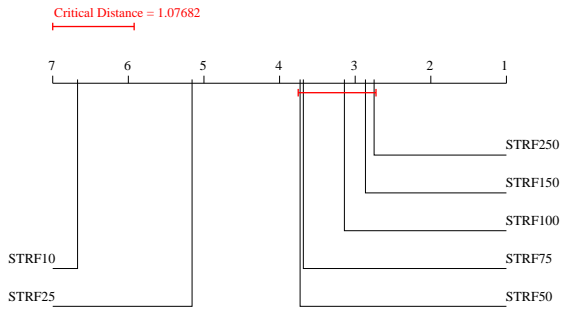
#### MTRegression - RRMSE: Multi-Target Random Forest

Method	MTRF10	MTRF25	MTRF50	MTRF75	MTRF100	MTRF150	MTRF250
Average Rank	6.69	5.21	3.61	3.39	3.36	2.96	2.79



#### MTRegression - RRMSE: Single-Target Random Forest

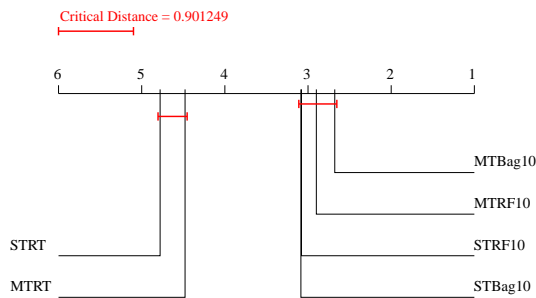
Method	STRF10	STRF25	STRF50	STRF75	STRF100	STRF150	STRF250
Average Rank	6.67	5.16	3.73	3.69	3.14	2.86	2.75





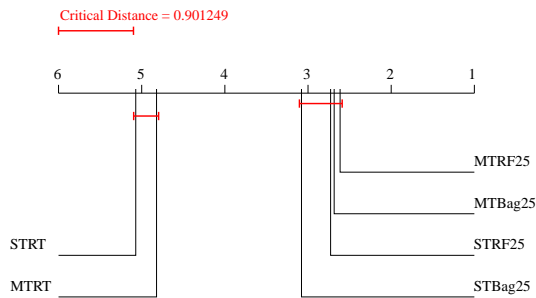
MTRegression - RRMSE: Ensemble size 10

Method	MTRT	STRT	MTBag10	STBag10	MTRF10	STRF10
Average Rank	4.48	4.78	2.68	3.09	2.90	3.08



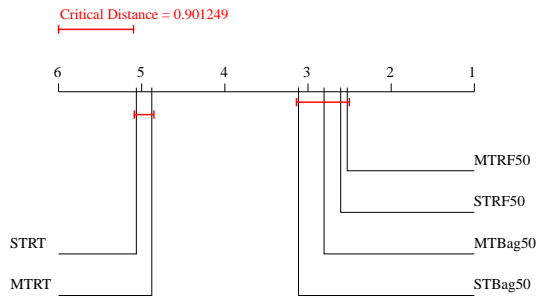
MTRegression - RRMSE: Ensemble size 25

Method	MTRT	STRT	MTBag25	STBag25	MTRF25	STRF25
Average Rank	4.82	5.07	2.69	3.08	2.61	2.73



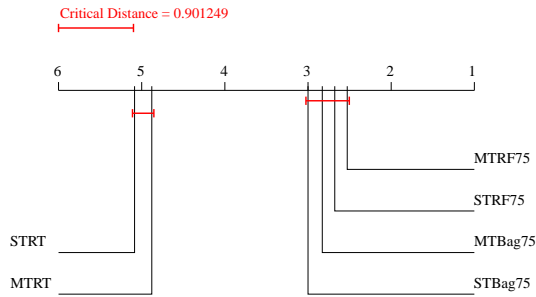
MTRegression - RRMSE: Ensemble size 50

Method	MTRT	STRT	MTBag50	STBag50	MTRF50	STRF50
Average Rank	4.88	5.06	2.81	3.11	2.53	2.61



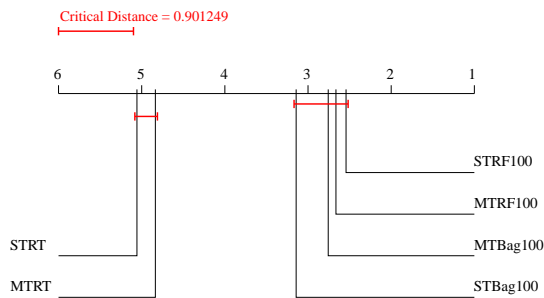
MTRegression - RRMSE: Ensemble size 75

Method	MTRT	STRT	MTBag75	STBag75	MTRF75	STRF75
Average Rank	4.88	5.09	2.83	3.00	2.53	2.68



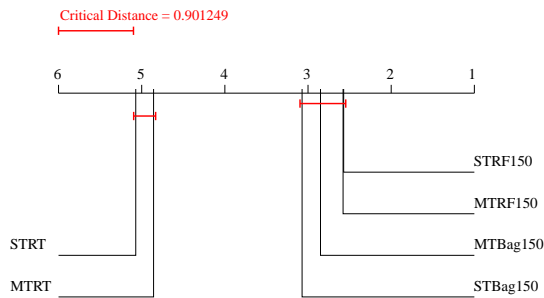
MTRegression - RRMSE: Ensemble size 100

Method	MTRT	STRT	MTBag100	STBag100	MTRF100	STRF100
Average Rank	4.84	5.06	2.76	3.14	2.66	2.54



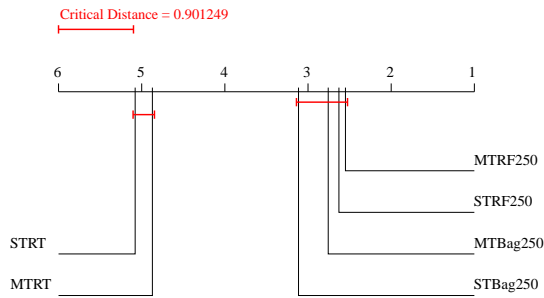
MTRegression - RRMSE: Ensemble size 150

Method	MTRT	STRT	MTBag150	STBag150	MTRF150	STRF150
Average Rank	4.86	5.07	2.85	3.07	2.58	2.57



MTRegression - RRMSE: Ensemble size 250

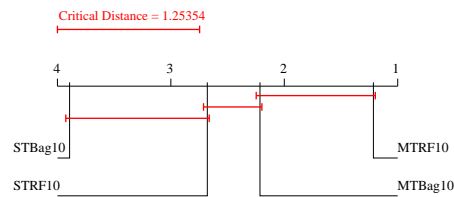
Method	MTRT	STRT	MTBag250	STBag250	MTRF250	STRF250
Average Rank	4.87	5.08	2.76	3.11	2.55	2.63



## 4 Multi-Target Regression - Time Comparison

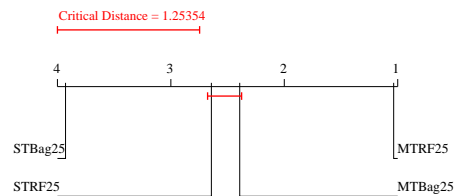
### MTRegression - Time: Ensemble size 10

Method	MTBag10	STBag10	MTRF10	STRF10
Average Rank	2.21	3.89	1.21	2.68



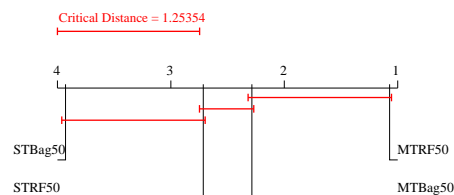
### MTRegression - Time: Ensemble size 25

Method	MTBag25	STBag25	MTRF25	STRF25
Average Rank	2.39	3.93	1.04	2.64



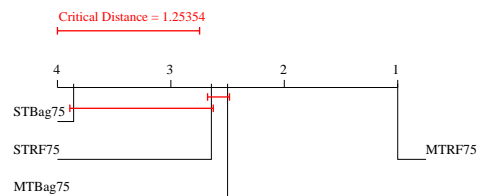
### MTRegression - Time: Ensemble size 50

Method	MTBag50	STBag50	MTRF50	STRF50
Average Rank	2.29	3.93	1.07	2.71



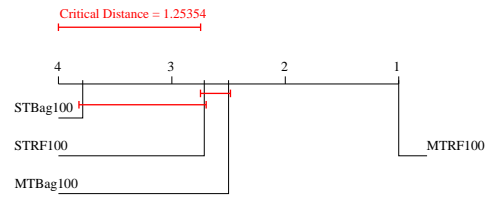
### MTRegression - Time: Ensemble size 75

Method	MTBag75	STBag75	MTRF75	STRF75
Average Rank	2.50	3.86	1.00	2.64



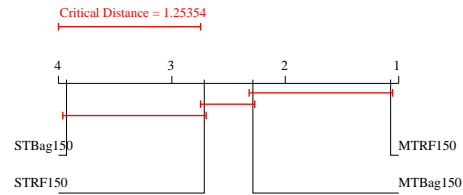
MTRegression - Time: Ensemble size 100

Method	MTBag100	STBag100	MTRF100	STRF100
Average Rank	2.50	3.79	1.00	2.71



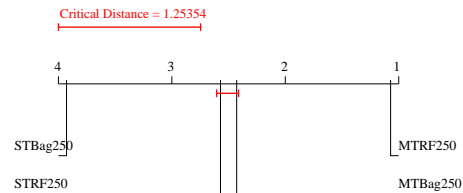
MTRegression - Time: Ensemble size 150

Method	MTBag150	STBag150	MTRF150	STRF150
Average Rank	2.29	3.93	1.07	2.71



MTRegression - Time: Ensemble size 250

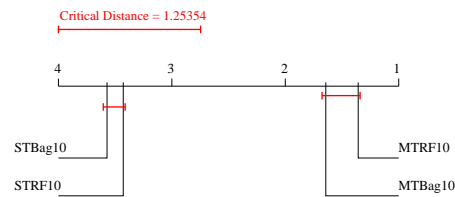
Method	MTBag250	STBag250	MTRF250	STRF250
Average Rank	2.43	3.93	1.07	2.57



## 5 Multi-Target Regression - Size Comparison

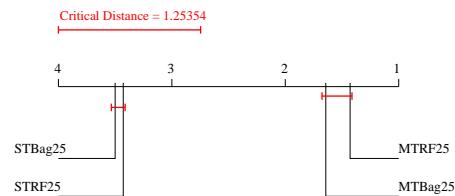
### MTRegression - Size: Ensemble size 10

Method	MTBag10	STBag10	MTRF10	STRF10
Average Rank	1.64	3.57	1.36	3.43



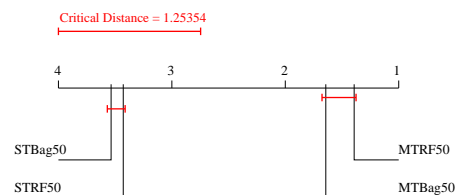
### MTRegression - Size: Ensemble size 25

Method	MTBag25	STBag25	MTRF25	STRF25
Average Rank	1.64	3.50	1.43	3.43



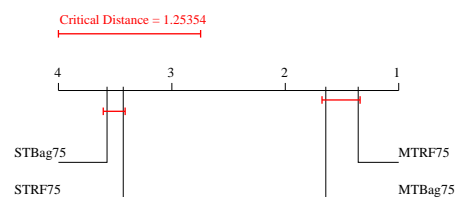
### MTRegression - Size: Ensemble size 50

Method	MTBag50	STBag50	MTRF50	STRF50
Average Rank	1.64	3.54	1.39	3.43



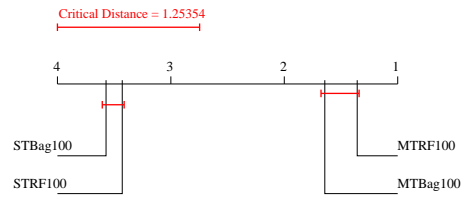
### MTRegression - Size: Ensemble size 75

Method	MTBag75	STBag75	MTRF75	STRF75
Average Rank	1.64	3.57	1.36	3.43



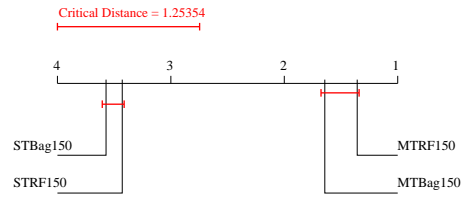
MTRegression - Size: Ensemble size 100

Method	MTBag100	STBag100	MTRF100	STRF100
Average Rank	1.64	3.57	1.36	3.43



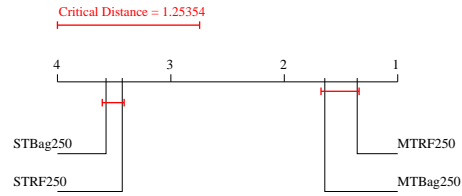
MTRegression - Size: Ensemble size 150

Method	MTBag150	STBag150	MTRF150	STRF150
Average Rank	1.64	3.57	1.36	3.43



MTRegression - Size: Ensemble size 250

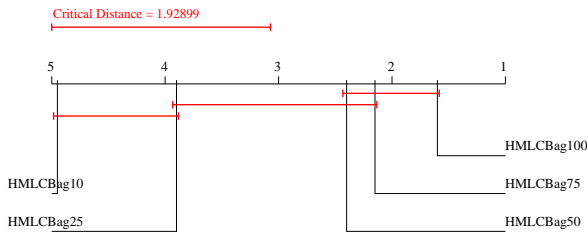
Method	MTBag250	STBag250	MTRF250	STRF250
Average Rank	1.64	3.57	1.36	3.43



## 6 Hierarchical Multi-Label Classification - HMLC

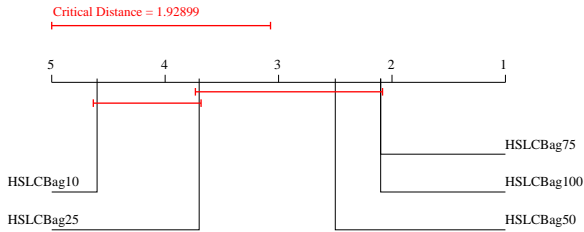
### HMLC: Bagging

Method	HMLCBag10	HMLCBag25	HMLCBag50	HMLCBag75	HMLCBag100
Average Rank	4.95	3.9	2.4	2.15	1.6



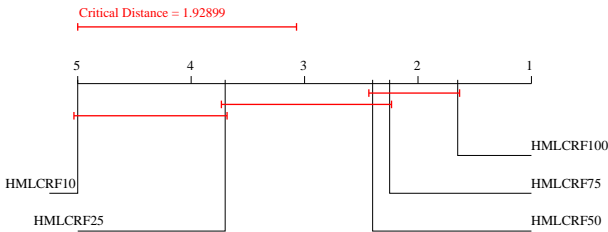
### HSLC: Bagging

Method	HSLCBag10	HSLCBag25	HSLCBag50	HSLCBag75	HSLCBag100
Average Rank	4.6	3.7	2.5	2.1	2.1



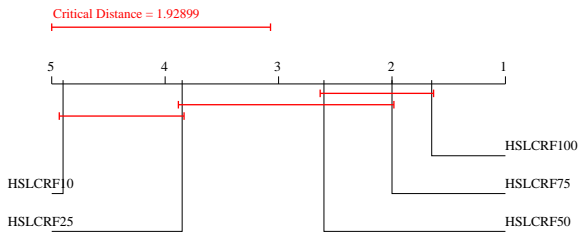
### HMLC: Random Forest

Method	HMLCRF10	HMLCRF25	HMLCRF50	HMLCRF75	HMLCRF100
Average Rank	5.0	3.7	2.4	2.25	1.65



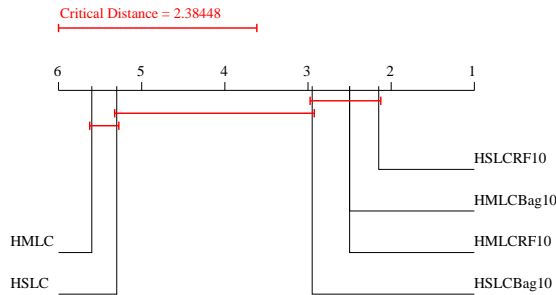
### HSLC: Random Forest

Method	HSLCRF10	HSLCRF25	HSLCRF50	HSLCRF75	HSLCRF100
Average Rank	4.9	3.85	2.6	2.0	1.65



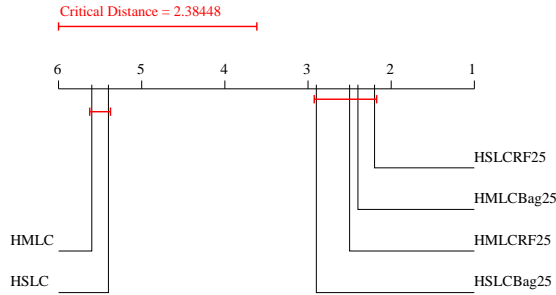
HMLC: Ensemble size 10

Method	HMLC	HSLC	HMLCBag10	HSLCBag10	HMLCRF10	HSLCRF10
Average Rank	5.6	5.3	2.5	2.95	2.5	2.15



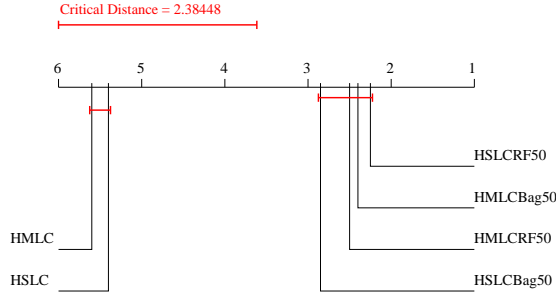
HMLC: Ensemble size 25

Method	HMLC	HSLC	HMLCBag25	HSLCBag25	HMLCRF25	HSLCRF25
Average Rank	5.6	5.4	2.4	2.9	2.5	2.2



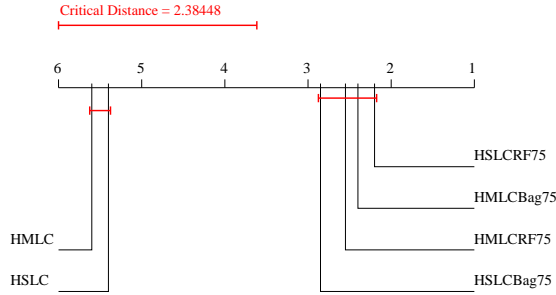
HMLC: Ensemble size 50

Method	HMLC	HSLC	HMLCBag50	HSLCBag50	HMLCRF50	HSLCRF50
Average Rank	5.6	5.4	2.4	2.85	2.5	2.25



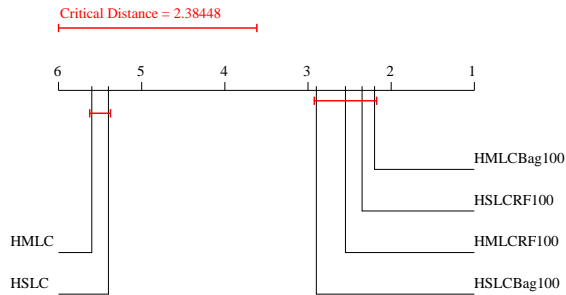
HMLC: Ensemble size 75

Method	HMLC	HSLC	HMLCBag75	HSLCBag75	HMLCRF75	HSLCRF75
Average Rank	5.6	5.4	2.4	2.85	2.55	2.2



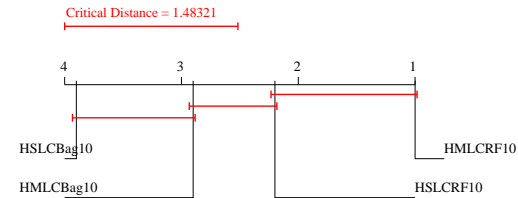


<b>HMLC: Ensemble size 100</b>						
Method	HMLC	HSLC	HMLCBag100	HSLCBag100	HMLCRF100	HSLCRF100
Average Rank	5.6	5.4	2.2	2.9	2.55	2.35

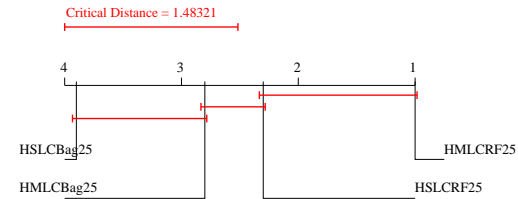


7 HMLC - Time Comparison

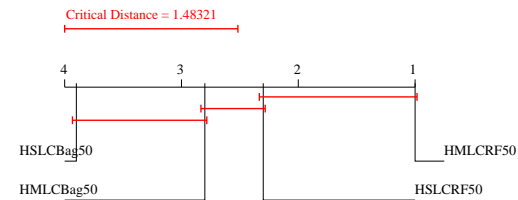
<b>HMLC - Time: Ensemble size 10</b>				
Method	HMLCBag10	HSLCBag10	HMLCRF10	HSLCRF10
Average Rank	2.9	3.9	1.0	2.2



<b>HMLC - Time: Ensemble size 25</b>				
Method	HMLCBag25	HSLCBag25	HMLCRF25	HSLCRF25
Average Rank	2.8	3.9	1.0	2.3

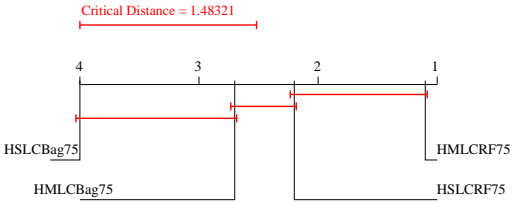


<b>HMLC - Time: Ensemble size 50</b>				
Method	HMLCBag50	HSLCBag50	HMLCRF50	HSLCRF50
Average Rank	2.8	3.9	1.0	2.3



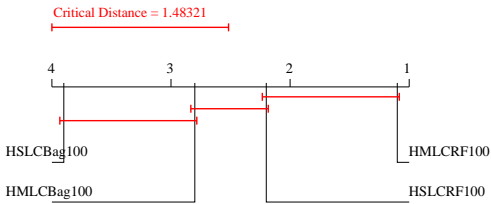
HMLC - Time: Ensemble size 75

Method	HMLCBag75	HSLCBag75	HMLCRF75	HSLCRF75
Average Rank	2.7	4.0	1.1	2.2



HMLC - Time: Ensemble size 100

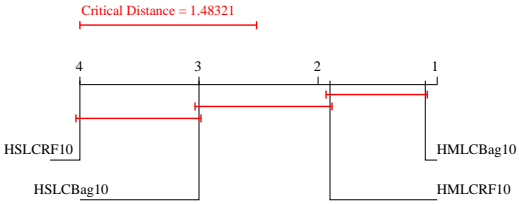
Method	HMLCBag100	HSLCBag100	HMLCRF100	HSLCRF100
Average Rank	2.8	3.9	1.1	2.2



8 HMLC - Size Comparison

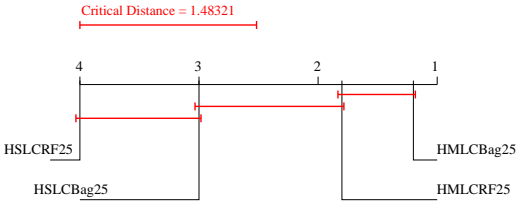
HMLC - Size: Ensemble size 10

Method	HMLCBag10	HSLCBag10	HMLCRF10	HSLCRF10
Average Rank	1.1	3.0	1.9	4.0



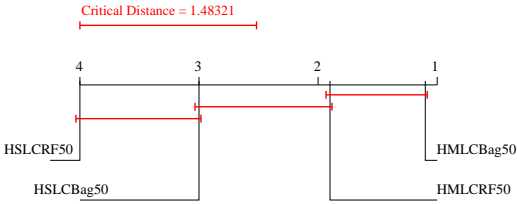
HMLC - Size:
Ensemble size 25

Method	HMLCBag25	HSLCBag25	HMLCRF25	HSLCRF25
Average Rank	1.2	3.0	1.8	4.0



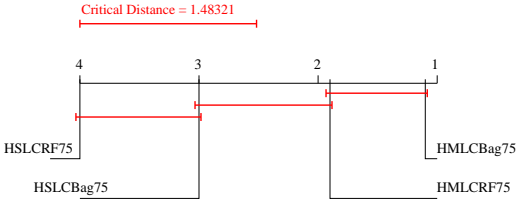
HMLC - Size:
Ensemble size 50

Method	HMLCBag50	HSLCBag50	HMLCRF50	HSLCRF50
Average Rank	1.1	3.0	1.9	4.0



HMLC - Size:
Ensemble size 75

Method	HMLCBag75	HSLCBag75	HMLCRF75	HSLCRF75
Average Rank	1.1	3.0	1.9	4.0



HMLC - Size:
Ensemble size 100

Method	HMLCBag100	HSLCBag100	HMLCRF100	HSLCRF100
Average Rank	1.1	3.0	1.9	4.0

