

## 1. Resultados CC

El análisis estadístico se hizo entre 21 modelos (poblaciones) con 100 muestras emparejadas.

El nivel de significancia para los test globales (*family-wise*) es de  $\alpha = 0,050$ .

No hemos podido rechazar la hipótesis nula de que la población sea normal para ninguna población (mínimo  $p = 0,024$ ). Por lo tanto asumimos que las poblaciones son normales.

Hemos aplicado el test Bartlett para evaluar la homogeneidad de los datos y rechazamos la hipótesis nula ( $p \approx 0$ ) de que los datos son homocedásticos. Por lo que asumimos que los datos son heteroscedásticos.

Dado que tenemos más de dos poblaciones, donde todas son normales pero heterocedásticas, usamos el test no paramétrico de Friedman como test *omnibus* para determinar si hay alguna diferencia significativa entre las medias de las poblaciones. Utilizamos el test de Nemenyi como *post-hoc* para saber que diferencias son significativas. Mostramos, en la tabla 1, la media ( $M$ ), la desviación estándar ( $SD$ ), y el ranking medio ( $MR$ ) de cada modelo a partir de todas las muestras. Consideraremos significativas la diferencias entre modelos si la diferencia entre rankings medios es mayor que la distancia crítica de  $CD = 3,132$  dada por el test de Nemenyi.

	MR	M	SD	CI	$d$	Magnitudo
<b>DUMMY</b>	15.330	0.500	0.000	[0.500, 0.500]	-	large
PLS + SVC	14.200	0.505	0.074	[0.482, 0.529]	-0.101	negligible
whole + SVCSScaler	13.185	0.514	0.073	[0.491, 0.537]	-0.265	small
PCA + SVC	13.005	0.516	0.073	[0.493, 0.539]	-0.315	small
PCA + SVCSScaler	12.700	0.519	0.069	[0.498, 0.541]	-0.395	small
PCA + KNNSScaler	12.470	0.528	0.065	[0.507, 0.548]	-0.601	medium
mRMR + KNNSScaler	12.460	0.524	0.061	[0.505, 0.543]	-0.551	medium
PLS + KNNSScaler	12.435	0.519	0.060	[0.500, 0.538]	-0.446	small
mRMR + KNN	12.285	0.524	0.061	[0.505, 0.543]	-0.559	medium
PLS + KNN	11.920	0.530	0.070	[0.509, 0.552]	-0.612	medium
PLS + SVCSScaler	11.900	0.527	0.065	[0.507, 0.547]	-0.582	medium
whole + SVC	10.860	0.540	0.072	[0.517, 0.562]	-0.786	medium
PLS + LR	9.855	0.553	0.071	[0.531, 0.575]	-1.050	large
mRMR + SVC	9.270	0.558	0.068	[0.537, 0.579]	-1.204	large
mRMR + SVCSScaler	9.230	0.556	0.063	[0.536, 0.575]	-1.242	large
PCA + LR	9.015	0.560	0.072	[0.538, 0.583]	-1.187	large
whole + KNNSScaler	8.765	0.566	0.078	[0.542, 0.590]	-1.200	large
PCA + KNN	8.605	0.559	0.060	[0.541, 0.578]	-1.409	large
whole + LR	8.400	0.562	0.062	[0.542, 0.581]	-1.414	large
whole + KNN	8.215	0.567	0.072	[0.544, 0.589]	-1.305	large
mRMR + LR	6.895	0.578	0.071	[0.556, 0.600]	-1.550	large

Tabla 1: Comparativa del ROC AUC de los modelos con el dataset CC

Rechazamos la hipótesis nula ( $p = 2,46 \times 10^{-20}$ ) del test de Friedman de que no hay diferencia entre las medidas de tendencia central de los modelos. Por lo que asumimos que hay una diferencia estadísticamente significativa entre las medianas de las poblaciones.

Basándonos en el test *post-hoc* de Nemenyi, asumimos que no hay diferencia significativa

dentro de los siguientes grupos:

- **DUMMY**, PLS + SVC, whole + SVCSScaler, PCA + SVC, PCA + SVCSScaler, PCA + KNNSScaler, mRMR + KNNSScaler, PLS + KNNSScaler y mRMR + KNN
- PLS + SVC, whole + SVCSScaler, PCA + SVC, PCA + SVCSScaler, PCA + KNNSScaler, mRMR + KNNSScaler, PLS + KNNSScaler, mRMR + KNN, PLS + KNN y PLS + SVCSScaler
- whole + SVCSScaler, PCA + SVC, PCA + SVCSScaler, PCA + KNNSScaler, mRMR + KNNSScaler, PLS + KNNSScaler, mRMR + KNN, PLS + KNN, PLS + SVCSScaler y whole + SVC
- PCA + SVCSScaler, PCA + KNNSScaler, mRMR + KNNSScaler, PLS + KNNSScaler, mRMR + KNN, PLS + KNN, PLS + SVCSScaler, whole + SVC y PLS + LR
- mRMR + KNN, PLS + KNN, PLS + SVCSScaler, whole + SVC, PLS + LR, mRMR + SVC y mRMR + SVCSScaler
- PLS + KNN, PLS + SVCSScaler, whole + SVC, PLS + LR, mRMR + SVC, mRMR + SVCSScaler y PCA + LR
- whole + SVC, PLS + LR, mRMR + SVC, mRMR + SVCSScaler, PCA + LR, whole + KNNSScaler, PCA + KNN, whole + LR y whole + KNN
- PLS + LR, mRMR + SVC, mRMR + SVCSScaler, PCA + LR, whole + KNNSScaler, PCA + KNN, whole + LR, whole + KNN y mRMR + LR

Todas las demás diferencias son significativas.

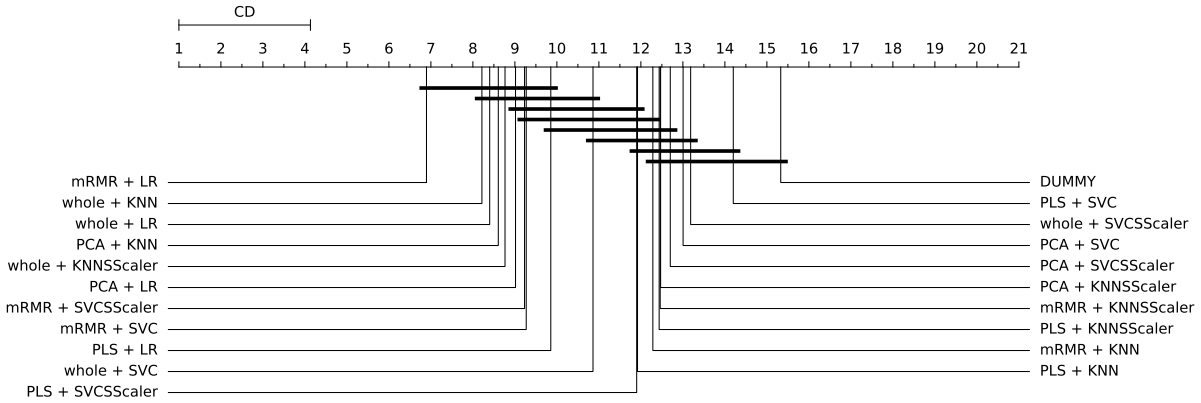


Figura 1: El diagrama CD representa los resultados del test Nemenyi para el ranking basado en el ROC AUC de ls modelos con el dataset CC. Las líneas horizontales representan diferencias no significativas

## 2. Resultados DCOR

El análisis estadístico se hizo entre 21 modelos (poblaciones) con 100 muestras emparejadas.

El nivel de significancia para los test globales (*family-wise*) es de  $\alpha = 0,050$ .

No hemos podido rechazar la hipótesis nula de que la población sea normal para ninguna población (mínimo  $p = 0,018$ ). Por lo tanto asumimos que las poblaciones son normales.

Hemos aplicado el test Bartlett para evaluar la homogeneidad de los datos y rechazamos la hipótesis nula ( $p \approx 0$ ) de que los datos son homocedásticos. Por lo que asumimos que los datos son heteroscedásticos.

Dado que tenemos más de dos poblaciones, donde todas son normales pero heterocedásticas, usamos el test no paramétrico de Friedman como test *omnibus* para determinar si hay alguna diferencia significativa entre las medias de las poblaciones. Utilizamos el test de Nemenyi como *post-hoc* para saber que diferencias son significativas. Mostramos, en la tabla 2, la media ( $M$ ), la desviación estándar ( $SD$ ), y el ranking medio ( $MR$ ) de cada modelo a partir de todas las muestras. Consideraremos significativas la diferencias entre modelos si la diferencia entre rankings medios es mayor que la distancia crítica de  $CD = 3,132$  dada por el test de Nemenyi.

Rechazamos la hipótesis nula ( $p = 2,41 \times 10^{-37}$ ) del test de Friedman de que no hay diferencia entre las medidas de tendencia central de los modelos. Por lo que asumimos que hay una diferencia estadísticamente significativa entre las medianas de las poblaciones.

	MR	M	SD	CI	$d$	Magnitude
PCA + SVCSScaler	14.025	0.453	0.069	[0.431, 0.474]	0.000	negligible
PCA + SVC	13.780	0.459	0.061	[0.440, 0.478]	-0.095	negligible
PCA + KNNSScaler	13.375	0.464	0.063	[0.445, 0.484]	-0.175	negligible
PLS + SVC	13.200	0.467	0.066	[0.447, 0.488]	-0.215	small
PLS + SVCSScaler	12.625	0.472	0.064	[0.452, 0.492]	-0.291	small
PLS + LR	11.580	0.485	0.071	[0.463, 0.508]	-0.465	small
PCA + LR	11.490	0.485	0.065	[0.464, 0.505]	-0.471	small
whole + SVC	11.260	0.484	0.064	[0.464, 0.504]	-0.473	small
whole + SVCSScaler	11.200	0.482	0.067	[0.461, 0.503]	-0.431	small
mRMR + KNNSScaler	11.025	0.484	0.059	[0.466, 0.503]	-0.490	small
PLS + KNNSScaler	10.955	0.488	0.065	[0.468, 0.509]	-0.529	medium
mRMR + SVCSScaler	10.745	0.495	0.071	[0.473, 0.517]	-0.603	medium
PLS + KNN	10.615	0.494	0.068	[0.473, 0.515]	-0.601	medium
mRMR + KNN	10.295	0.494	0.058	[0.476, 0.512]	-0.647	medium
whole + LR	9.830	0.499	0.057	[0.481, 0.517]	-0.729	medium
mRMR + SVC	9.715	0.499	0.066	[0.478, 0.520]	-0.682	medium
<b>DUMMY</b>	9.640	0.500	0.000	[0.500, 0.500]	-0.962	large
PCA + KNN	9.600	0.502	0.066	[0.481, 0.523]	-0.726	medium
whole + KNNSScaler	8.915	0.511	0.071	[0.489, 0.533]	-0.829	large
whole + KNN	8.745	0.510	0.072	[0.487, 0.532]	-0.802	large
mRMR + LR	8.385	0.516	0.067	[0.495, 0.536]	-0.919	large

Tabla 2: Comparativa del ROC AUC de los modelos con el dataset DCOR

Basándonos en el test *post-hoc* de Nemenyi, asumimos que no hay diferencia significativa dentro de los siguientes grupos:

- PCA + SVCSScaler, PCA + SVC, PCA + KNNSScaler, PLS + SVC, PLS + SVCSScaler, PLS + LR, PCA + LR, whole + SVC, whole + SVCSScaler, mRMR + KNNSScaler y PLS + KNNSScaler

- PCA + SVC, PCA + KNNSScaler, PLS + SVC, PLS + SVCSScaler, PLS + LR, PCA + LR, whole + SVC, whole + SVCSScaler, mRMR + KNNSScaler, PLS + KNNSScaler y mRMR + SVCSScaler
- PCA + KNNSScaler, PLS + SVC, PLS + SVCSScaler, PLS + LR, PCA + LR, whole + SVC, whole + SVCSScaler, mRMR + KNNSScaler, PLS + KNNSScaler, mRMR + SVCSScaler, PLS + KNN y mRMR + KNN
- PLS + SVCSScaler, PLS + LR, PCA + LR, whole + SVC, whole + SVCSScaler, mRMR + KNNSScaler, PLS + KNNSScaler, mRMR + SVCSScaler, PLS + KNN, mRMR + KNN, whole + LR, mRMR + SVC, **DUMMY** y PCA + KNN
- PLS + LR, PCA + LR, whole + SVC, whole + SVCSScaler, mRMR + KNNSScaler, PLS + KNNSScaler, mRMR + SVCSScaler, PLS + KNN, mRMR + KNN, whole + LR, mRMR + SVC, **DUMMY**, PCA + KNN, whole + KNNSScaler y whole + KNN
- PCA + LR, whole + SVC, whole + SVCSScaler, mRMR + KNNSScaler, PLS + KNNSScaler, mRMR + SVCSScaler, PLS + KNN, mRMR + KNN, whole + LR, mRMR + SVC, **DUMMY**, PCA + KNN, whole + KNNSScaler, whole + KNN y mRMR + LR

Todas las demás diferencias son significativas.

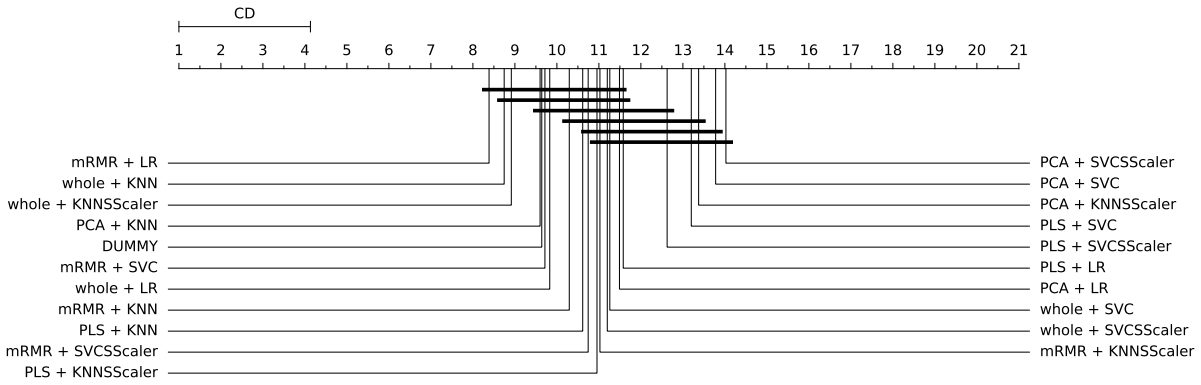


Figura 2: El diagrama CD representa los resultados del test Nemenyi para el ranking basado en el ROC AUC de los modelos con el dataset DCOR. Las líneas horizontales representan diferencias no significativas

### 3. Resultados FFT

El análisis estadístico se hizo entre 5 modelos (poblaciones) con 100 muestras emparejadas.

El nivel de significancia para los test globales (*family-wise*) es de  $\alpha = 0,050$ .

No hemos podido rechazar la hipótesis nula de que la población sea normal para ninguna población (mínimo  $p = 0,031$ ). Por lo tanto asumimos que las poblaciones son normales.

Hemos aplicado el test Bartlett para evaluar la homogeneidad de los datos y rechazamos la hipótesis nula ( $p \approx 0$ ) de que los datos son homocedásticos. Por lo que asumimos que los datos son heteroscedásticos

Dado que tenemos más de dos poblaciones, donde todas son normales pero heterocedásticas, usamos el test no paramétrico de Friedman como test *omnibus* para determinar si hay alguna diferencia significativa entre las medias de las poblaciones. Utilizamos el test de Nemenyi como *post-hoc* para saber que diferencias son significativas. Mostramos, en la tabla 3, la media ( $M$ ), la desviación estándar ( $SD$ ), y el ranking medio ( $MR$ ) de cada modelo a partir de todas las muestras. Consideraremos significativas la diferencias entre modelos si la diferencia entre rankings medios es mayor que la distancia crítica de  $CD = 3,807$  dada por el test de Nemenyi.

Rechazamos la hipótesis nula ( $p = 5,15 \times 10^{-161}$ ) del test de Friedman de que no hay diferencia entre las medidas de tendencia central de los modelos. Por lo que asumimos que hay una diferencia estadísticamente significativa entre las medianas de las poblaciones.

	MR	M	SD	CI	\$d\$	Magnitude
whole + LR	19.255	0.461	0.074	[0.437, 0.485]	0.000	negligible
PLS + KNNSScaler	18.480	0.479	0.055	[0.462, 0.497]	-0.278	small
PCA + KNN	18.255	0.484	0.064	[0.463, 0.504]	-0.323	small
whole + KNNSScaler	17.770	0.485	0.059	[0.467, 0.504]	-0.363	small
whole + LRSScaler	17.165	0.488	0.066	[0.467, 0.509]	-0.379	small
<b>DUMMY</b>	17.000	0.500	0.000	[0.500, 0.500]	-0.741	medium
whole + SVC	17.000	0.500	0.000	[0.500, 0.500]	-0.741	medium
PLS + KNN	16.950	0.497	0.059	[0.478, 0.516]	-0.535	medium
PCA + SVC	15.955	0.499	0.085	[0.472, 0.526]	-0.474	small
PLS + SVCSScaler	14.945	0.509	0.074	[0.486, 0.533]	-0.650	medium
PLS + SVC	14.550	0.520	0.065	[0.499, 0.540]	-0.846	large
mRMR + KNN	13.830	0.524	0.068	[0.503, 0.546]	-0.890	large
whole + KNN	13.545	0.530	0.065	[0.509, 0.551]	-0.984	large
PCA + KNNSScaler	13.320	0.529	0.065	[0.509, 0.550]	-0.976	large
mRMR + KNNSScaler	12.950	0.533	0.066	[0.512, 0.554]	-1.020	large
PLS + LRSScaler	10.325	0.561	0.079	[0.536, 0.587]	-1.307	large
PLS + LR	10.320	0.562	0.078	[0.537, 0.586]	-1.318	large
mRMR + SVC	10.110	0.565	0.070	[0.543, 0.587]	-1.439	large
whole + SVCSScaler	9.885	0.571	0.063	[0.551, 0.591]	-1.594	large
PCA + SVCSScaler	9.220	0.574	0.066	[0.553, 0.595]	-1.610	large
PCA + LRSScaler	9.215	0.575	0.085	[0.548, 0.602]	-1.427	large
PCA + LR	8.620	0.582	0.077	[0.558, 0.607]	-1.599	large
mRMR + SVCSScaler	5.880	0.610	0.073	[0.587, 0.633]	-2.015	large
mRMR + LRSScaler	5.340	0.622	0.079	[0.597, 0.647]	-2.106	large
mRMR + LR	5.115	0.624	0.071	[0.602, 0.647]	-2.253	large

Tabla 3: Comparativa del ROC AUC de los modelos con el dataset FFT

Basándonos en el test *post-hoc* de Nemenyi, asumimos que no hay diferencia significativa dentro de los siguientes grupos:

- whole + LR, PLS + KNNSScaler, PCA + KNN, whole + KNNSScaler, whole + LRSScaler, **DUMMY**, whole + SVC, PLS + KNN, and PCA + SVC
- PLS + KNNSScaler, PCA + KNN, whole + KNNSScaler, whole + LRSScaler, **DUMMY**, whole + SVC, PLS + KNN, PCA + SVC, and PLS + SVCSScaler
- PCA + KNN, whole + KNNSScaler, whole + LRSScaler, **DUMMY**, whole + SVC, PLS + KNN, PCA + SVC, PLS + SVCSScaler, and PLS + SVC

- whole + LRSScaler, **DUMMY**, whole + SVC, PLS + KNN, PCA + SVC, PLS + SVCSScaler, PLS + SVC, mRMR + KNN, and whole + KNN
- **DUMMY**, whole + SVC, PLS + KNN, PCA + SVC, PLS + SVCSScaler, PLS + SVC, mRMR + KNN, whole + KNN, and PCA + KNNSScaler
- PCA + SVC, PLS + SVCSScaler, PLS + SVC, mRMR + KNN, whole + KNN, PCA + KNNSScaler, and mRMR + KNNSScaler
- mRMR + KNN, whole + KNN, PCA + KNNSScaler, mRMR + KNNSScaler, PLS + LRSScaler, PLS + LR, and mRMR + SVC
- whole + KNN, PCA + KNNSScaler, mRMR + KNNSScaler, PLS + LRSScaler, PLS + LR, mRMR + SVC, and whole + SVCSScaler
- mRMR + KNNSScaler, PLS + LRSScaler, PLS + LR, mRMR + SVC, whole + SVCSScaler, PCA + SVCSScaler, and PCA + LRSScaler
- PLS + LRSScaler, PLS + LR, mRMR + SVC, whole + SVCSScaler, PCA + SVCSScaler, PCA + LRSScaler, and PCA + LR
- PCA + SVCSScaler, PCA + LRSScaler, PCA + LR, and mRMR + SVCSScaler
- PCA + LR, mRMR + SVCSScaler, mRMR + LRSScaler, and mRMR + LR

Todas las demás diferencias son significativas.

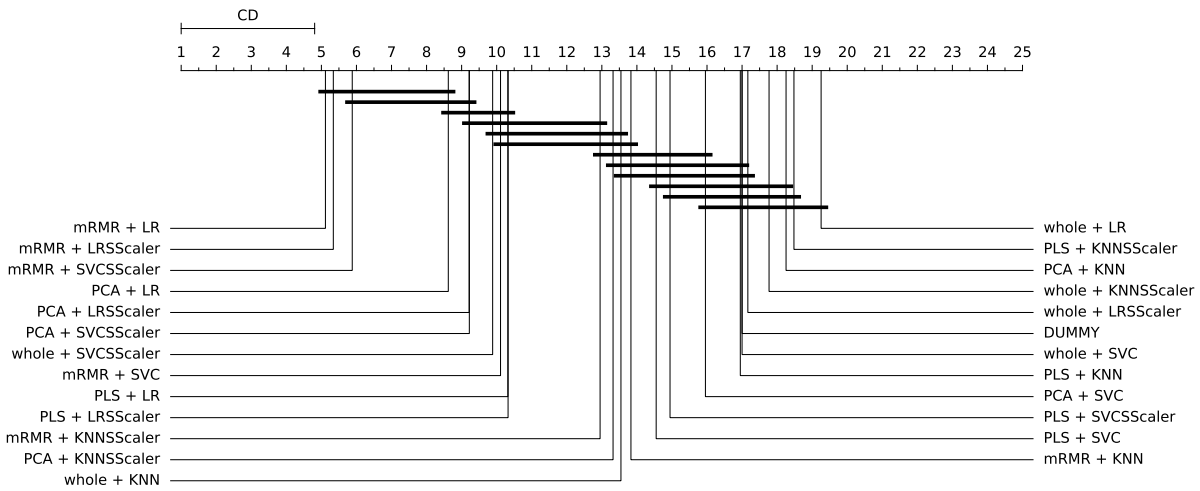


Figura 3: El diagrama CD representa los resultados del test Nemenyi para el ranking basado en el ROC AUC de los modelos con el dataset FFT. Las líneas horizontales representan diferencias no significativas