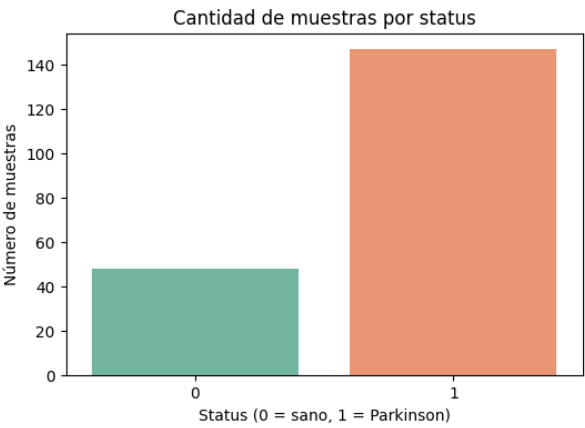


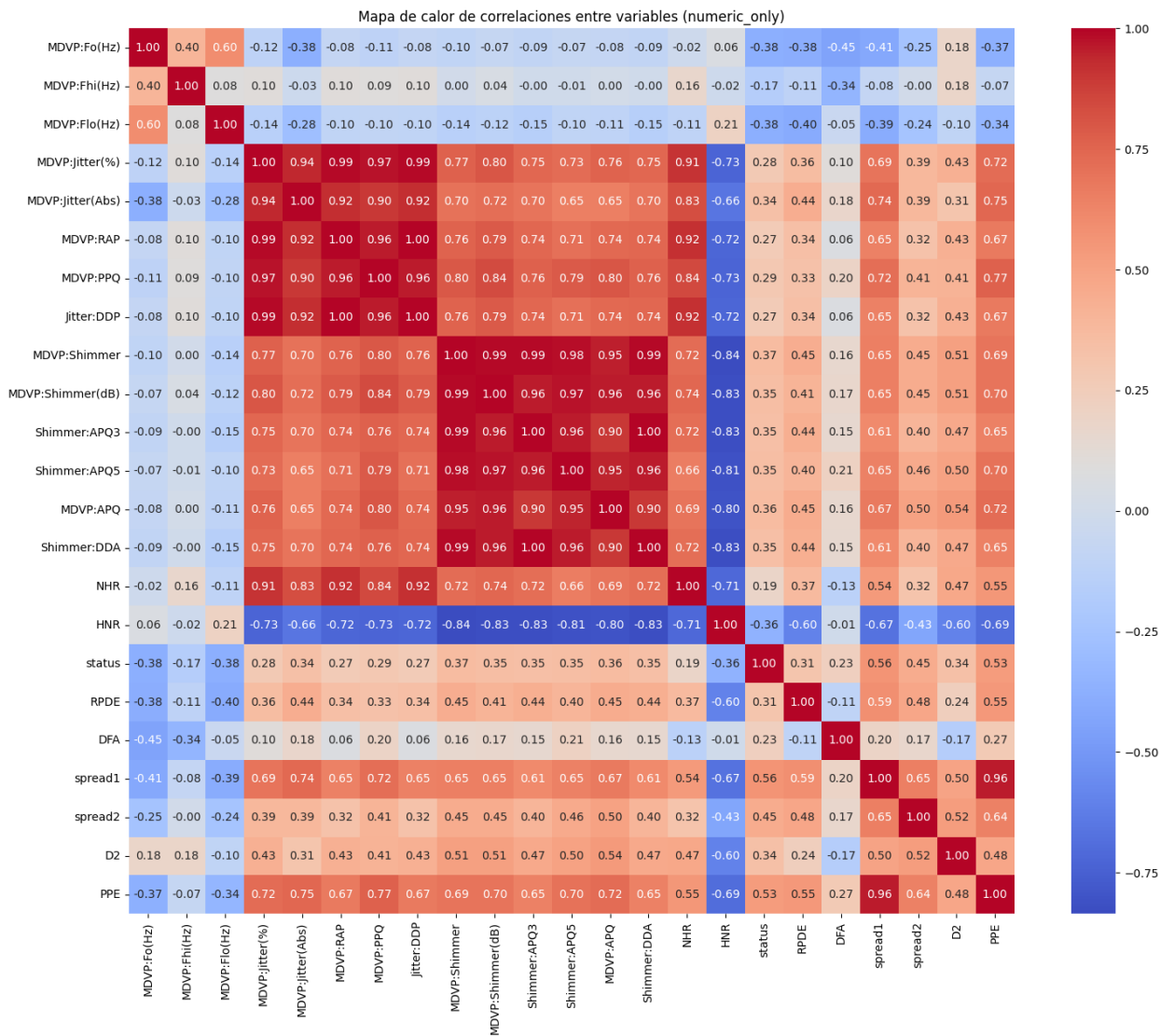
# PREDICCION PARKISTON

Reporte técnico-estadístico del software

Distribución por target estatus:



Mapa de correlación de Pearson entre variables:



## Validación de supuestos y selección de prueba estadística

Es fundamental verificar los **supuestos** que sostienen las pruebas paramétricas clásicas. Si esos supuestos no se cumplen, optamos por métodos no paramétricos más robustos a desviaciones en la distribución.

### Normalidad (Shapiro–Wilk)

Para cada variable, dividimos las observaciones en los dos grupos según status (0 = sano, 1 = Parkinson) y aplicamos la **prueba de Shapiro–Wilk** por separado a cada subconjunto.

- **Hipótesis nula ( $H_0$ ):** los datos provienen de una distribución normal.
- **Decisión:** si  $p \geq 0.05$ , no rechazamos  $H_0$  y consideramos que la muestra es “suficientemente normal”; si  $p < 0.05$ , detectamos desviaciones significativas de la normalidad.

	Variable	0	Normalidad	1	Normalidad
0	MDVP:F0(Hz)	<0.001	No normal	<0.001	No normal
1	MDVP:Fhi(Hz)	<0.001	No normal	<0.001	No normal
2	MDVP:Flo(Hz)	<0.001	No normal	<0.001	No normal
3	MDVP:Jitter(%)	<0.001	No normal	<0.001	No normal
4	MDVP:Jitter(Abs)	<0.001	No normal	<0.001	No normal
5	MDVP:RAP	<0.001	No normal	<0.001	No normal
6	MDVP:PPQ	<0.001	No normal	<0.001	No normal
7	Jitter:DDP	<0.001	No normal	<0.001	No normal
8	MDVP:Shimmer	<0.001	No normal	<0.001	No normal
9	MDVP:Shimmer(dB)	<0.001	No normal	<0.001	No normal
10	Shimmer:APQ3	<0.001	No normal	<0.001	No normal
11	Shimmer:APQ5	<0.001	No normal	<0.001	No normal
12	MDVP:APQ	0.015	No normal	<0.001	No normal
13	Shimmer:DDA	<0.001	No normal	<0.001	No normal
14	NHR	<0.001	No normal	<0.001	No normal
15	HNR	0.013	No normal	0.002	No normal
16	RPDE	0.630	Normal	<0.001	No normal
17	DFA	<0.001	No normal	0.090	Normal
18	spread1	0.371	Normal	0.009	No normal
19	spread2	0.640	Normal	0.520	Normal
20	D2	0.917	Normal	0.025	No normal
21	PPE	0.030	No normal	<0.001	No normal

## 2. Homogeneidad de varianzas (Levene)

Incluso si ambas muestras son normales, las pruebas paramétricas como el ANOVA requieren que las varianzas de los dos grupos sean similares. Con la **prueba de Levene** comprobamos:

- **H<sub>0</sub>:** las varianzas de los dos grupos son iguales.
- **Decisión:** igual que arriba,  $p \geq 0.05 \rightarrow$  varianzas homogéneas;  $p < 0.05 \rightarrow$  varianzas distintas.

## 3. Elección de la prueba

Con los resultados de las pruebas de normalidad y de homogeneidad:

- Si ambos grupos pasan Shapiro ( $p \geq 0.05$ ) y Levene ( $p \geq 0.05$ ), entonces usamos la **prueba ANOVA-F**. El **estadístico F** mide la razón entre la variabilidad “entre grupos” y la “intragrupo”: cuanto mayor sea, más discriminante es la variable.
- **En caso contrario**, empleamos la **prueba de Kruskal–Wallis**, que no asume normalidad ni igualdad de varianzas. Su p-valor nos indica si las medianas de los dos grupos difieren de forma significativa.

	Variable	Prueba	Estadístico	p-valor
1	spread2	ANOVA–F	50.34	<0.001
2	PPE	Kruskal–Wallis	68.08	<0.001
3	spread1	Kruskal–Wallis	68.08	<0.001
4	MDVP:APQ	Kruskal–Wallis	45.88	<0.001
5	MDVP:Jitter(Abs)	Kruskal–Wallis	36.87	<0.001
6	MDVP:PPQ	Kruskal–Wallis	35.63	<0.001
7	MDVP:Shimmer(dB)	Kruskal–Wallis	35.11	<0.001
8	MDVP:Shimmer	Kruskal–Wallis	34.53	<0.001
9	MDVP:Jitter(%)	Kruskal–Wallis	33.32	<0.001
10	Jitter:DDP	Kruskal–Wallis	33.25	<0.001
11	MDVP:RAP	Kruskal–Wallis	33.13	<0.001
12	NHR	Kruskal–Wallis	32.24	<0.001
13	Shimmer:APQ5	Kruskal–Wallis	31.47	<0.001
14	Shimmer:APQ3	Kruskal–Wallis	28.05	<0.001
15	Shimmer:DDA	Kruskal–Wallis	28.02	<0.001
16	HNR	Kruskal–Wallis	24.46	<0.001
17	D2	Kruskal–Wallis	21.85	<0.001
18	RPDE	Kruskal–Wallis	18.55	<0.001
19	MDVP:F0(Hz)	Kruskal–Wallis	17.40	<0.001
20	MDVP:F10(Hz)	Kruskal–Wallis	16.81	<0.001

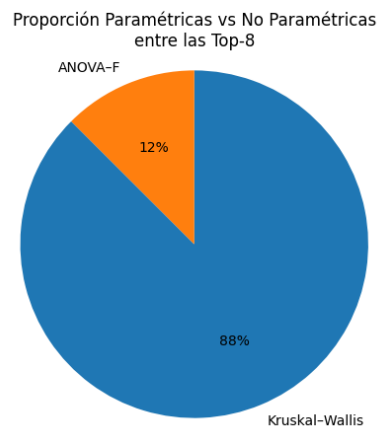
21	MDVP:Fhi(Hz)	Kruskal–Wallis	13.21	<0.001
22	DFA	Kruskal–Wallis	9.69	0.002

**Para las variables paramétricas** (aquellas que pasaron Shapiro–Wilk y Levene), es el **estadístico F** de la ANOVA.

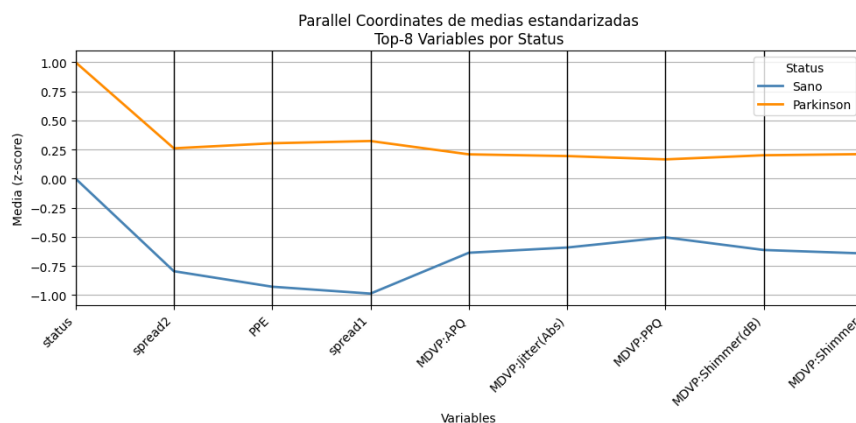
- Matemáticamente,  $F = (\text{variabilidad entre grupos}) / (\text{variabilidad intra grupo})$ .
- Cuanto más alto sea F, mayor es la diferencia relativa de medias entre sanos y Parkinson, y más discriminante resulta la variable.

**Para las variables no paramétricas** (las que no cumplían los supuestos), es el **estadístico H** de Kruskal–Wallis.

- H mide la dispersión de los rangos de las observaciones entre los dos grupos.
- Un H mayor indica una mayor diferencia de las distribuciones de las dos muestras.



Tras el ranking de variables, tomamos las ocho características más discriminantes, las estandarizamos a z-scores para igualar su escala y calculamos la media de cada una por grupo (sanos vs. Parkinson); luego trazamos un gráfico de coordenadas paralelas donde la línea de Parkinson (naranja) recorre esas ocho variables siempre por encima de la de sanos (azul), mostrando especialmente en spread2 y spread1 la mayor separación entre ambos grupos.



## Entrenamiento:

Se empleó un **hold-out** del 80 % de los datos para entrenamiento y 20 % para prueba

## Tabla de validación cruzada (CV)

La tabla recoge para cada modelo el tiempo medio de entrenamiento (TrainTime) y las métricas de evaluación (Accuracy, Precision, Recall, F1 y MCC) promediadas en un esquema de 5-fold Stratified CV. Está ordenada por MCC en orden descendente para resaltar qué modelo ofreció la mejor capacidad de clasificación equilibrada entre clases.

Modelo	Tiempo (s)	Accuracy	Precision	Recall	F1	MCC
Random Forest	0.58	0.85	0.89	0.92	0.9	0.59
Naïve Bayes	0.01	0.76	0.97	0.71	0.82	0.54
XGBoost	0.12	0.83	0.89	0.88	0.88	0.54
k-NN	0.01	0.81	0.87	0.88	0.87	0.47
SVM (RBF)	0.01	0.82	0.85	0.93	0.89	0.46
Logistic L2	0.02	0.81	0.85	0.91	0.88	0.44

## Tabla de características del sistema:

Detalla el entorno donde se ejecutaron los experimentos: sistema operativo, procesador y memoria RAM. Esto aporta contexto sobre los tiempos de entrenamiento y permite reproducibilidad de los resultados.

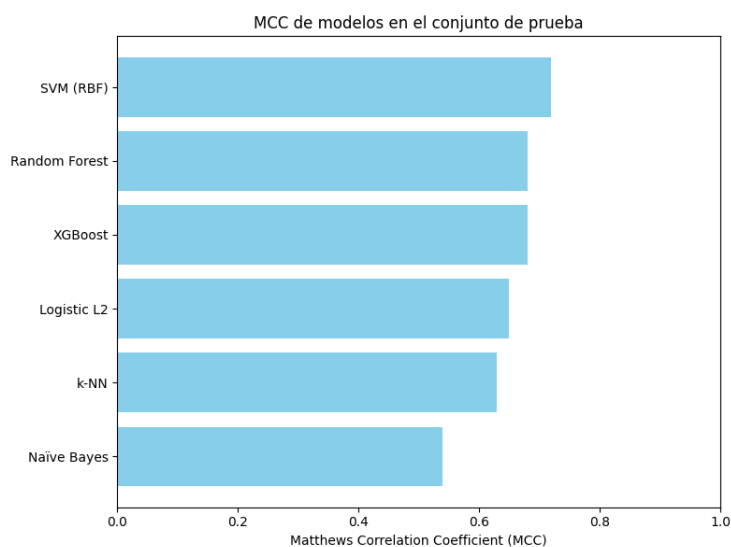
Recurso	Especificación
OS	Linux 6.1.123+
CPU	x86_64
RAM	12.67 GB

## Pruebas:

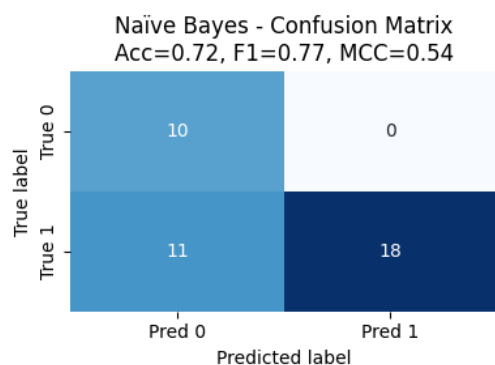
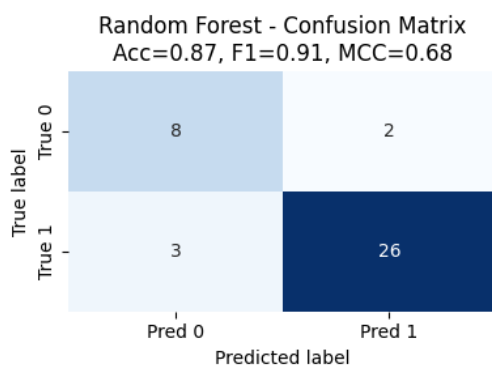
En la última fase entrenamos cada pipeline completo con el 80 % de los datos, lo serializamos a disco para poder reproducirlo y luego evaluamos su rendimiento en el 20 % restante (“test set”). Para cada modelo calculamos Accuracy, Precision, Recall, F1 y MCC, y pintamos su matriz de confusión como un heatmap anotado con los valores de True/False Positives/Negatives. Finalmente, recogimos todas esas métricas en una tabla ordenada por MCC, que refleja el equilibrio global entre clases.

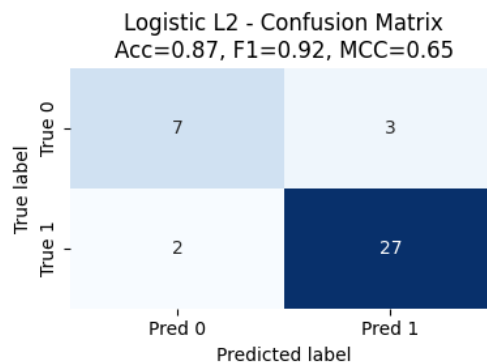
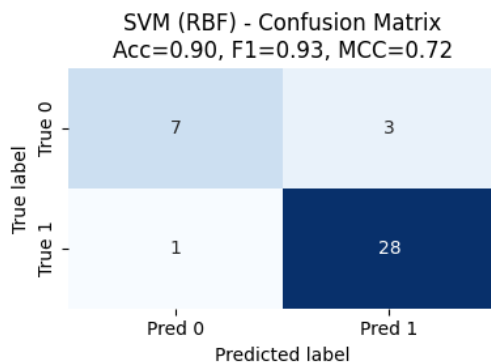
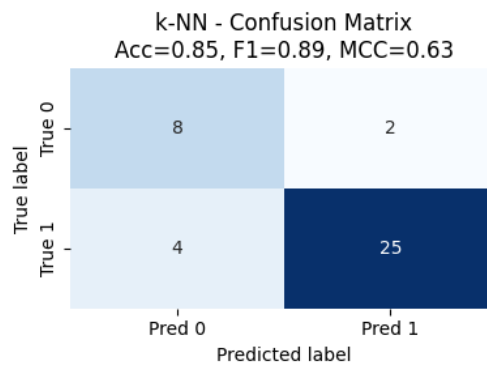
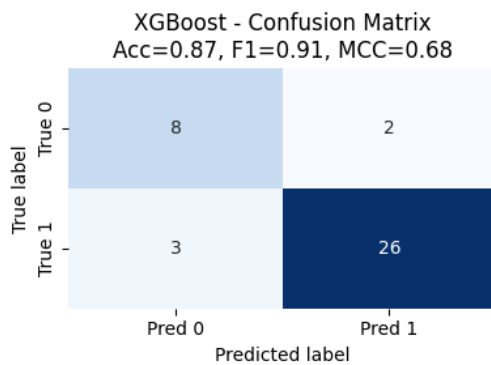
Modelo	Nombre archivo	Accuracy	Precision	Recall	F1	MCC
SVM (RBF)	SVM_(RBF).h	0.9	0.9	0.97	0.93	0.72
Random Forest	Random_Forest.h	0.87	0.93	0.9	0.91	0.68
XGBoost	XGBoost.h	0.87	0.93	0.9	0.91	0.68
Logistic L2	Logistic_L2.h	0.87	0.9	0.93	0.92	0.65
k-NN	k-NN.h	0.85	0.93	0.86	0.89	0.63
Naïve Bayes	Naïve_Bayes.h	0.72	1.0	0.62	0.77	0.54

**MCC (Matthews Correlation Coefficient):** métrica de correlación para clasificación binaria que considera TP, TN, FP y FN, y proporciona un único valor interpretativo (−1 a +1). Se elige criterio principal porque equilibra la eficacia de la predicción en datasets con posible desbalance de clases.



## Matrix de confusión de todos los modelos





En la última etapa elegimos el **mejor modelo** según el MCC en el conjunto de prueba (en este caso, SVM con kernel RBF,  $MCC = 0.720$ ) y guardamos su pipeline. Para verificar que todo funciona, cargamos ese pipeline y realizamos una **inferencia de ejemplo** sobre una muestra de test, obteniendo tanto la etiqueta predicha como la probabilidad asociada.

Luego, para asegurarnos de que su superioridad no fuese casual, aplicamos la **prueba de McNemar** entre el mejor modelo y cada uno de los demás: construimos tablas de contingencia a partir de sus predicciones en el mismo set de test, calculamos el p-value y determinamos si las diferencias en el número de aciertos/fallos eran estadísticamente significativas ( $\alpha = 0.05$ ).

#### McNemar Test (mejor vs. otros):

SVM (RBF) vs Random Forest: p-value = 0.250 → diferencia significativa: no

SVM (RBF) vs XGBoost: p-value = 0.375 → diferencia significativa: no

SVM (RBF) vs Logistic L2: p-value = 1.000 → diferencia significativa: no

SVM (RBF) vs k-NN: p-value = 0.219 → diferencia significativa: no

SVM (RBF) vs Naïve Bayes: p-value = 0.000 → diferencia significativa: sí

## Optimización de hiperparámetros

Empleamos **GridSearchCV** con 5-fold estratificado (mismo que en la validación previa) para cada una de las métricas clave: AUC-ROC, Accuracy, Precision, Recall, F1 y MCC.

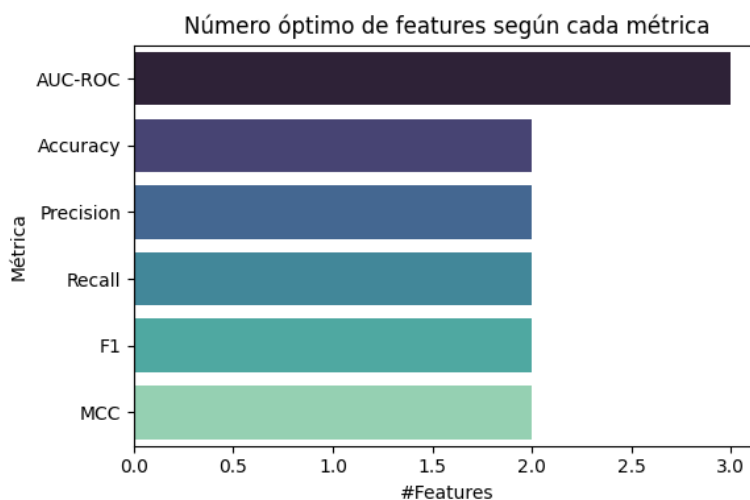
Métricas	Parámetros	CV_score
AUC-ROC	{'clf__C': 0.1, 'clf__gamma': 'scale', 'clf__kernel': 'linear'}	0.90
Accuracy	{'clf__C': 100, 'clf__gamma': 1, 'clf__kernel': 'rbf'}	0.84
Precision	{'clf__C': 100, 'clf__gamma': 1, 'clf__kernel': 'rbf'}	0.88
Recall	{'clf__C': 0.1, 'clf__gamma': 'scale', 'clf__kernel': 'rbf'}	1.0
F1	{'clf__C': 1, 'clf__gamma': 1, 'clf__kernel': 'rbf'}	0.89
MCC	{'clf__C': 100, 'clf__gamma': 1, 'clf__kernel': 'rbf'}	0.56

## Optimización por MCC y Selección de Variables con RFECV

Para refinar nuestro SVM, realizamos dos pasos encadenados:

1. **GridSearchCV** optimizando la métrica MCC (Matthews Correlation Coefficient), que equilibra verdaderos/falsos positivos y negativos.
2. **RFECV** con SVM lineal, para determinar el número óptimo de variables según cada métrica (AUC-ROC, Accuracy, Precision, Recall, F1 y MCC).

Métrica	#Features	CV_Score	Test_Score	Features
AUC-ROC	3	0.912	0.921	spread1, MDVP:APQ, MDVP:Shimmer
Accuracy	2	0.847	0.846	MDVP:APQ, MDVP:Shimmer
Precision	2	0.876	0.897	MDVP:APQ, MDVP:Shimmer
Recall	2	0.933	0.897	MDVP:APQ, MDVP:Shimmer
F1	2	0.902	0.897	MDVP:APQ, MDVP:Shimmer
MCC	2	0.573	0.597	MDVP:APQ, MDVP:Shimmer

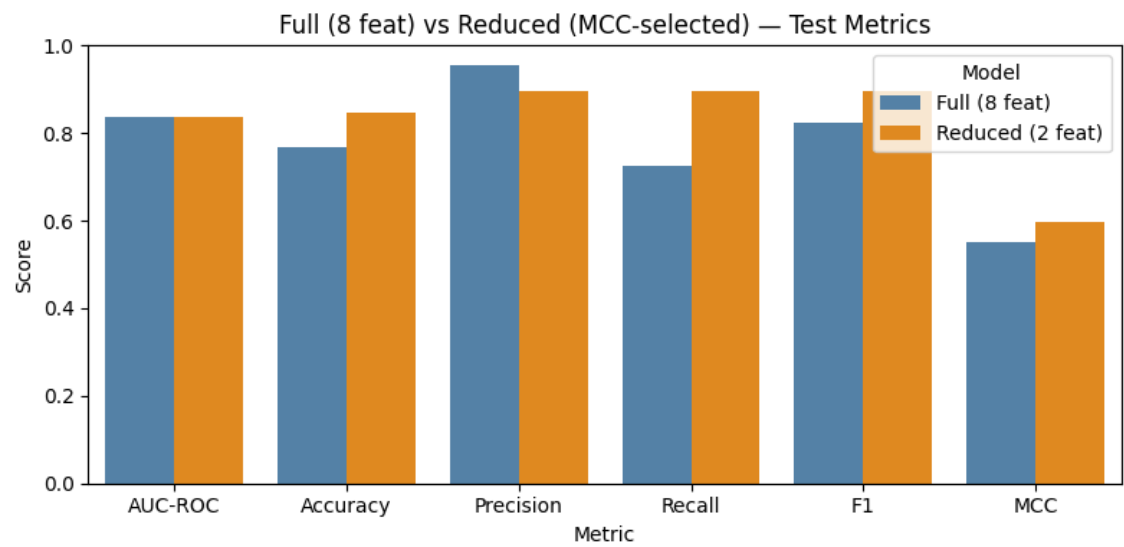




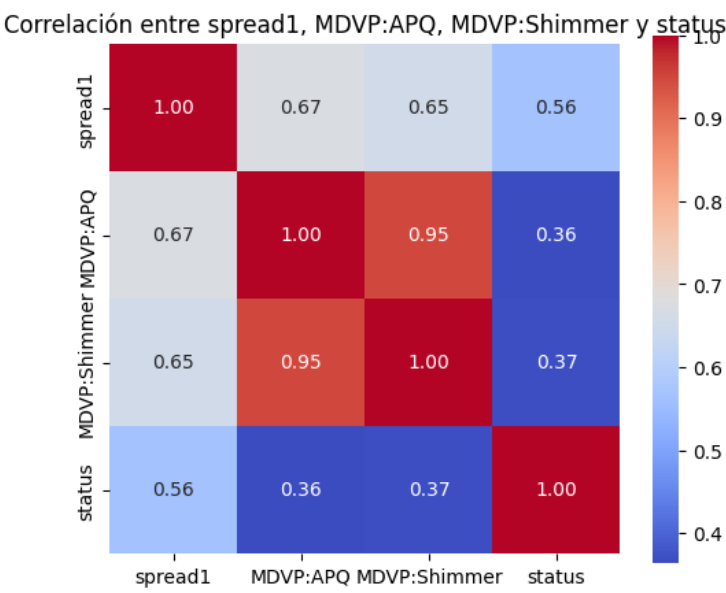
Comparativa de rendimiento en Test: Full (8 features) vs Reduced (2 features)

Modelo	AUC-ROC	Accuracy	Precision	Recall	F1	MCC
Full (8 feat)	0.838	0.769	0.955	0.724	0.824	0.55
Reduced (2 feat)	0.838	0.846	0.897	0.897	0.897	0.597

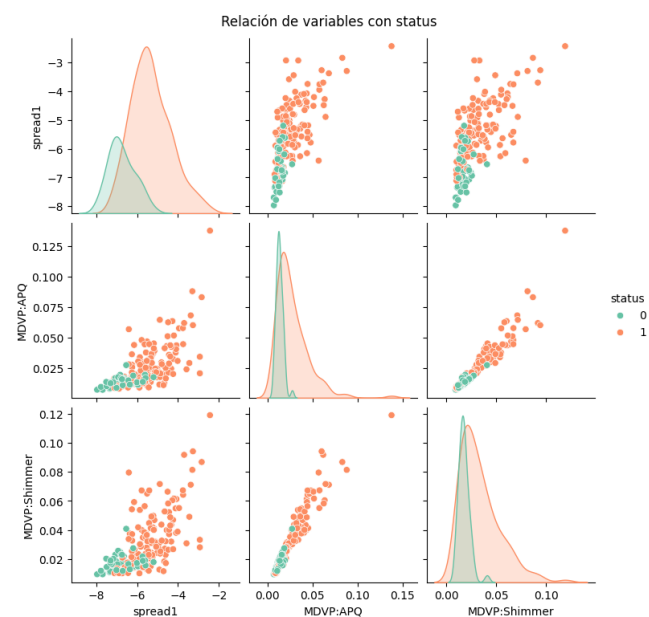
Reducir a solo MDVP:APQ y MDVP:Shimmer no solo simplifica el modelo en un 75 % menos de variables, sino que también **mejora su capacidad de clasificación balanceada**, reflejado en el aumento de MCC. Esto confirma que estas dos características capturan la mayor parte de la información discriminativa sin sacrificar la fiabilidad global del clasificador.



Mapa de correlación de las variables finales por estatus:



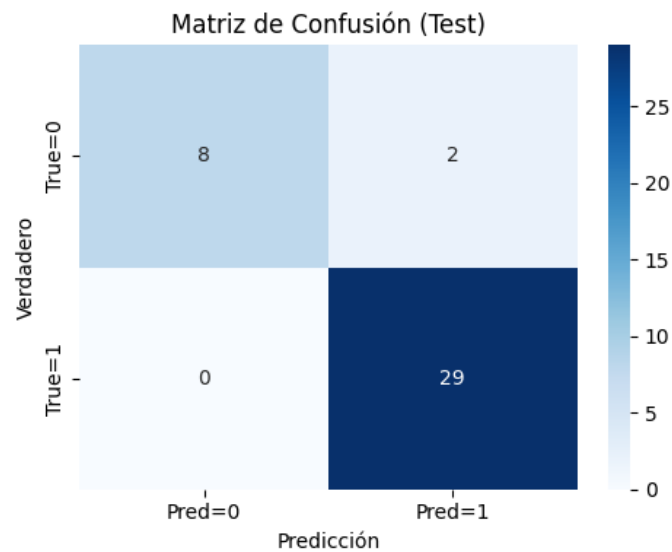
**Discriminación de variables entre pacientes que tiene vs con lo que no tiene Parkinson:**



Con el modelo final (SVM-RBF optimizado por MCC) entrenado sobre las **tres variables finales** (spread1, MDVP:APQ y MDVP:Shimmer):

Métrica	Train	Test
AUC-ROC	0.951	0.921
Accuracy	0.885	0.949
Precision	0.879	0.935
Recall	0.983	1.0
F1	0.928	0.967
MCC	0.669	0.865

**Matrix de confusión final:**



### Selección de modelo Híbrido

<b>Model</b>	<b>Auc</b>	Accuracy	Precision	Recall	F1	Mcc
<b>Soft_voting_parkinson</b>	0.917	0.846	0.897	0.897	0.897	0.597
<b>Stacking_parkinson</b>	0.914	0.821	0.867	0.897	0.881	0.515