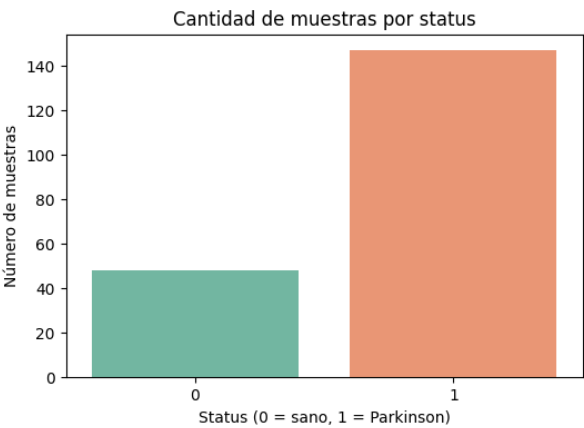


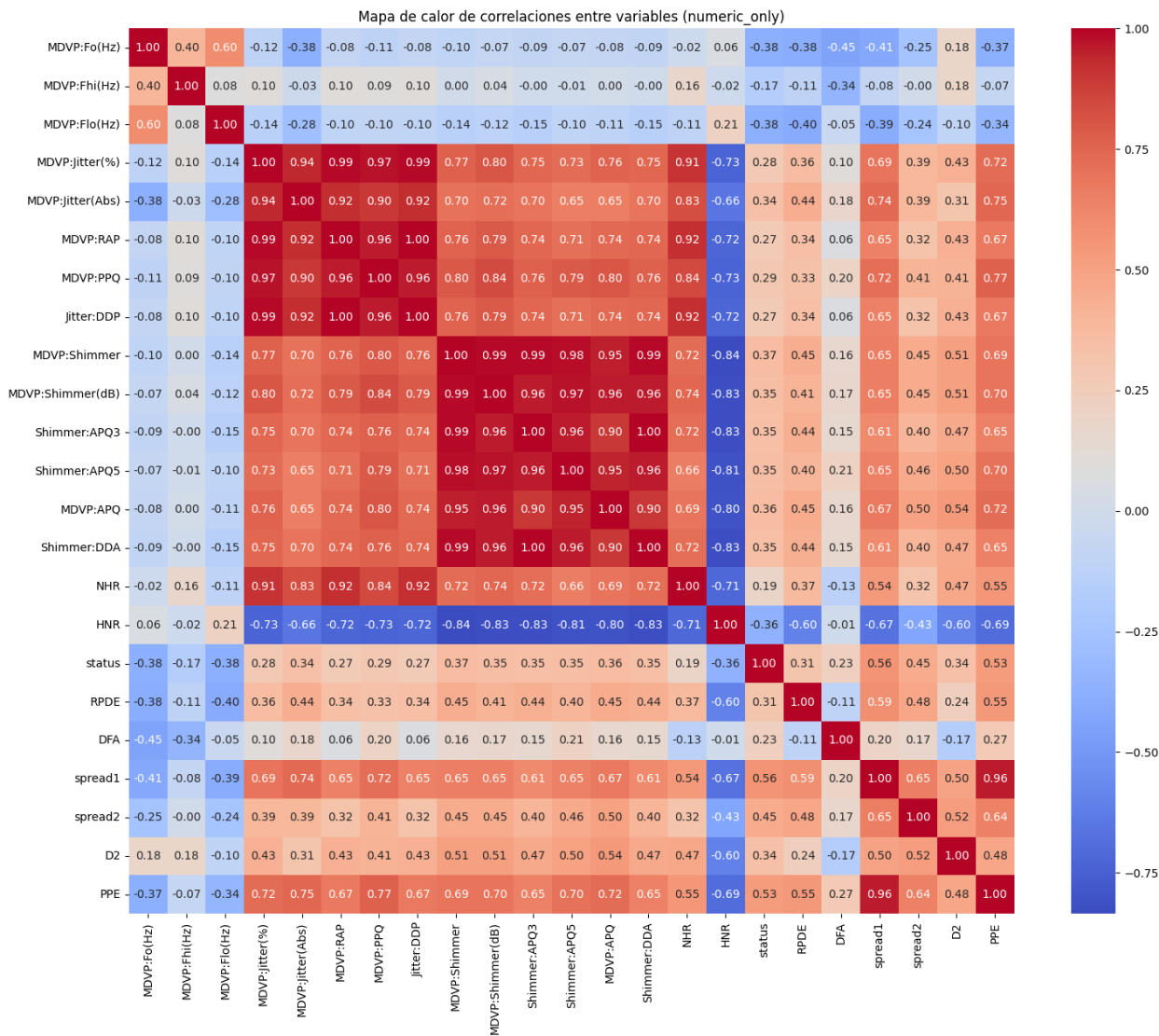
PARKINSON'S PREDICTION

Technical-Statistical Report of the Software

Distribution by Target Status:



Pearson Correlation Map between Variables:



Validation of Assumptions and Statistical Test Selection

It is fundamental to verify the assumptions underlying classical parametric tests. If these assumptions are not met, we opt for non-parametric methods, which are more robust against deviations in distributions.

Normality (Kolmogorov–Smirnov):

For each variable, we divide observations into two groups according to status (0 = healthy, 1 = Parkinson's) and apply the Kolmogorov–Smirnov test separately to each subgroup.

- **Null Hypothesis (H_0):** Data comes from a normal distribution.
- **Decision:** If $p \geq 0.05$, we do not reject H_0 and consider the sample sufficiently normal. If $p < 0.05$, we detect significant deviations from normality.

	Variable	0	Normalidad	1	Normalidad
0	MDVP:Fo(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

Homogeneity of Variances (Levene):

Even if both samples are normal, parametric tests such as ANOVA require similar variances across groups. With the Levene test we verify:

- **H₀:** Variances are equal across groups.
- **Decision:** $p \geq 0.05 \rightarrow$ homogeneous variances; $p < 0.05 \rightarrow$ unequal variances.

Choice of Statistical Test:

Based on results from normality and homogeneity tests:

- If both groups pass Shapiro ($p \geq 0.05$) and Levene ($p \geq 0.05$), we use the ANOVA-F test. The F-statistic measures the ratio between inter-group variability and intra-group variability; a higher value indicates more discriminative power.
- Otherwise, we apply the Kruskal–Wallis test, which does not assume normality nor equal variances. Its p-value indicates if medians significantly differ between the groups.

	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

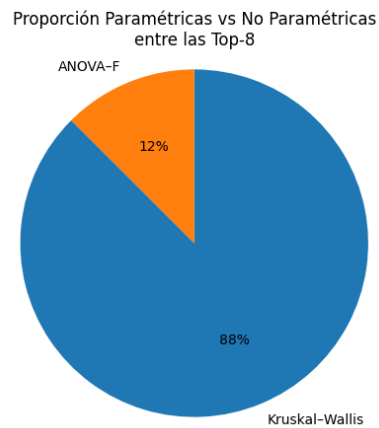
16	HNR	Kruskal–Wallis	24.46	<0.001
17	D2	Kruskal–Wallis	21.85	<0.001
18	RPDE	Kruskal–Wallis	18.55	<0.001
19	MDVP:Fo(Hz)	Kruskal–Wallis	17.40	<0.001
20	MDVP:Flo(Hz)	Kruskal–Wallis	16.81	<0.001
21	MDVP:Fhi(Hz)	Kruskal–Wallis	13.21	<0.001
22	DFA	Kruskal–Wallis	9.69	0.002

For parametric variables (those passing Shapiro–Wilk and Levene), we use the F-statistic from ANOVA.

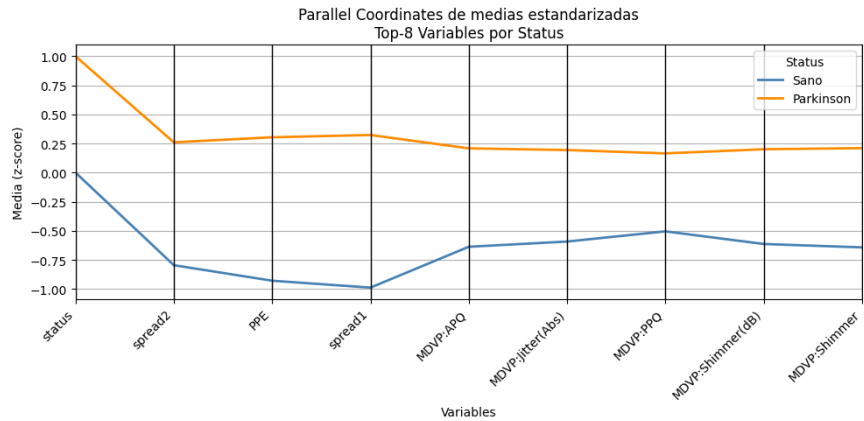
- Mathematically, $F = (\text{between-groups variability}) / (\text{within-group variability})$.
- A higher F-value indicates a greater relative difference in means between healthy and Parkinson's groups, making the variable more discriminative.

For non-parametric variables (those failing assumptions), we use the H-statistic from Kruskal–Wallis.

- H measures dispersion in ranks of observations between groups.
- A higher H-value indicates a greater difference in distributions between the two samples.



After ranking variables, we selected the eight most discriminative features, standardized them to z-scores to equalize scales, and calculated their means for each group (healthy vs. Parkinson's). We then plotted a parallel coordinate graph, where the Parkinson's line (orange) consistently appears above the healthy line (blue) across these eight variables, notably showing the largest separation between groups in the spread2 and spread1 variables.



Training:

We applied an 80% hold-out of the data for training and 20% for testing.

Cross-validation Table (CV):

The table lists each model's average training time (TrainTime) and evaluation metrics (Accuracy, Precision, Recall, F1, and MCC), averaged across a 5-fold Stratified CV scheme. The models are sorted by MCC in descending order to highlight the model with the best balanced classification capability between classes.

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

System Characteristics Table:

Describes the environment where experiments were performed: operating system, processor, and RAM. This provides context regarding training times and ensures reproducibility of results.

Recurso	Especificación
OS	Linux 6.1.123+
CPU	x86_64
RAM	12.67 GB

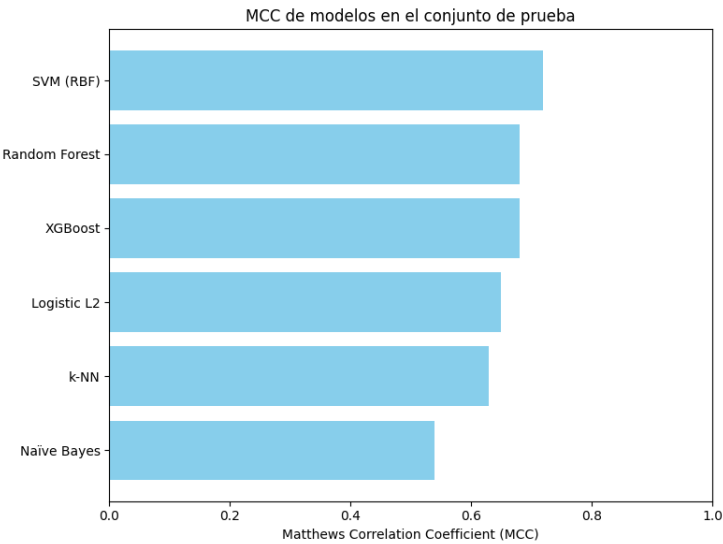
Testing:

In the final phase, we trained each complete pipeline using 80% of data, serialized it to disk for reproducibility, and evaluated its performance on the remaining 20% (“test set”). For each model, we calculated Accuracy, Precision, Recall, F1, and MCC, and plotted its confusion matrix as an annotated heatmap with True/False Positives/Negatives. Finally, all metrics were gathered into a table sorted by MCC, reflecting overall class balance.

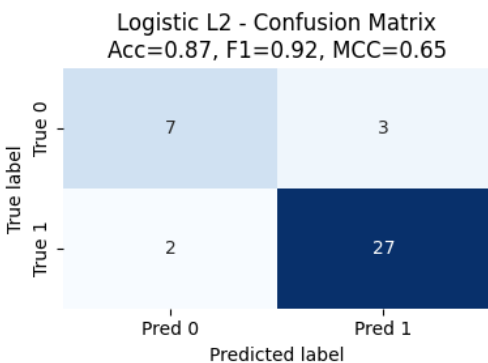
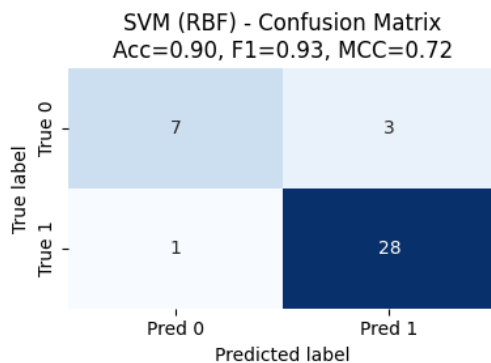
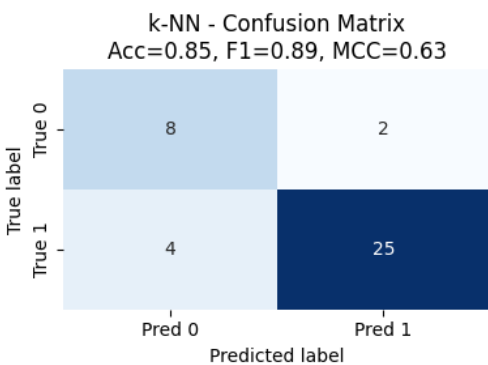
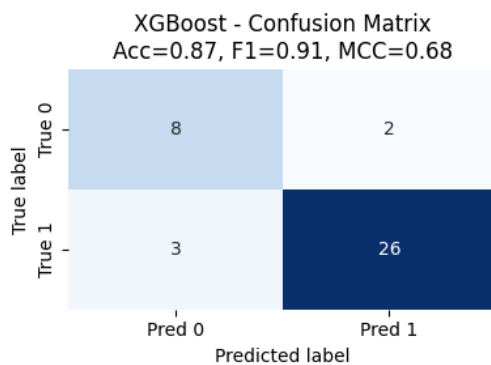
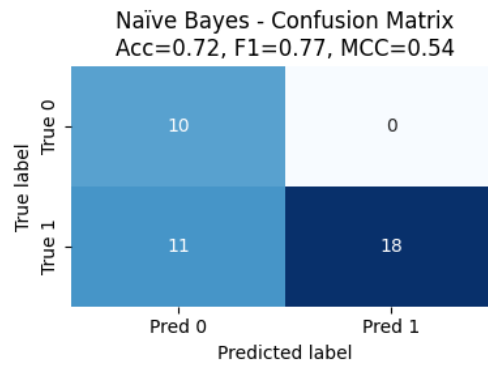
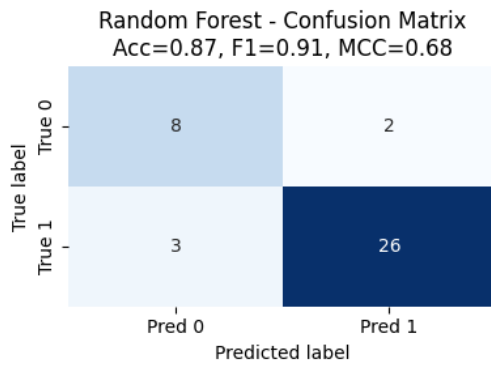
Modelo	Nombre archivo	Accuracy	Precision	Recall	F1	MCC
SVM (RBF)	SVM_(RBF).h	0.9	0.9	0.97	0.93	0.72
Random Forest	Random_Forest.h	0.87	0.93	0.9	0.91	0.68
XGBoost	XGBoost.h	0.87	0.93	0.9	0.91	0.68
Logistic L2	Logistic_L2.h	0.87	0.9	0.93	0.92	0.65
k-NN	k-NN.h	0.85	0.93	0.86	0.89	0.63
Naïve Bayes	Naïve_Bayes.h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

Matthews Correlation Coefficient (MCC):

MCC is a correlation metric for binary classification that takes into account TP, TN, FP, and FN, providing a single interpretable value (ranging from -1 to $+1$). It was selected as the primary criterion because it balances prediction effectiveness in datasets with potential class imbalance.



Confusion Matrix for All Models:



In the final stage, we selected the best-performing model according to MCC on the test set (in this case, SVM with RBF kernel, $MCC = 0.720$), and saved its pipeline. To verify proper functionality, we loaded the pipeline and performed a test inference on a sample from the test set, obtaining both predicted labels and associated probabilities.

To confirm that its superiority was not due to chance, we applied the McNemar test between the best model and each of the other models: constructing contingency tables based on their predictions on the same test set, calculating the p-value, and determining if differences in the number of correct/incorrect predictions were statistically significant ($\alpha = 0.05$).

McNemar Test (Best vs. Others):

- SVM (RBF) vs Random Forest: p-value = 0.250 → significant difference: No
- SVM (RBF) vs XGBoost: p-value = 0.375 → significant difference: No
- SVM (RBF) vs Logistic L2: p-value = 1.000 → significant difference: No
- SVM (RBF) vs k-NN: p-value = 0.219 → significant difference: No
- SVM (RBF) vs Naïve Bayes: p-value = 0.000 → significant difference: Yes

Hyperparameter Optimization:

We applied GridSearchCV with stratified 5-fold cross-validation (same as previously) for each key metric: AUC-ROC, Accuracy, Precision, Recall, F1, and MCC.

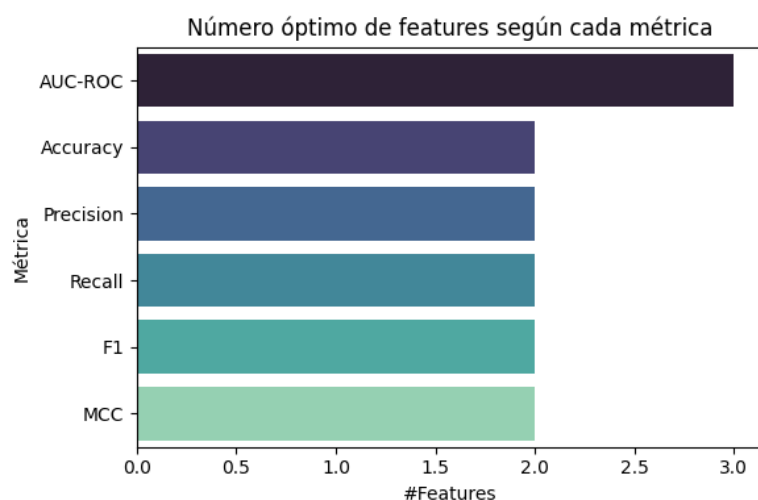
Métricas	Parámetros	CV_score
AUC-ROC	{'clf__C': 0.1, 'clf__gamma': 'scale', 'clf__kernel': 'linear'}	0.90
Accuracy	{'clf__C': 100, 'clf__gamma': 1, 'clf__kernel': 'rbf'}	0.84
Precision	{'clf__C': 100, 'clf__gamma': 1, 'clf__kernel': 'rbf'}	0.88
Recall	{'clf__C': 0.1, 'clf__gamma': 'scale', 'clf__kernel':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

Optimization by MCC and Feature Selection with RFECV:

To refine our SVM, we performed two sequential steps:

1. GridSearchCV optimizing the MCC metric (Matthews Correlation Coefficient), balancing true/false positives and negatives.
2. RFECV with linear SVM to determine the optimal number of features for each metric (AUC-ROC, Accuracy, Precision, Recall, F1, and MCC).

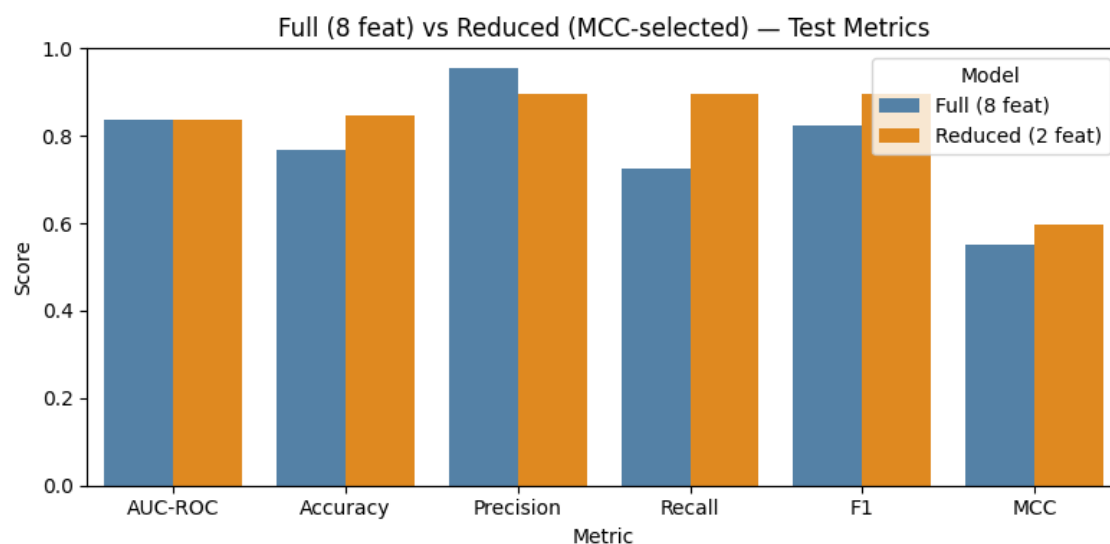
Métrica	#Features	CV_Score	Test_Score	Features
AUC-ROC	3	0.912	0.921	spread1, MDVP:APQ, MDVP:Shimmer
Accuracy	2	0.847	0.846	MDVP:APQ, MDVP:Shimmer
Precision	2	0.876	0.897	MDVP:APQ, MDVP:Shimmer
Recall	2	0.933	0.897	MDVP:APQ, MDVP:Shimmer
F1	2	0.902	0.897	MDVP:APQ, MDVP:Shimmer
MCC	2	0.573	0.597	MDVP:APQ,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



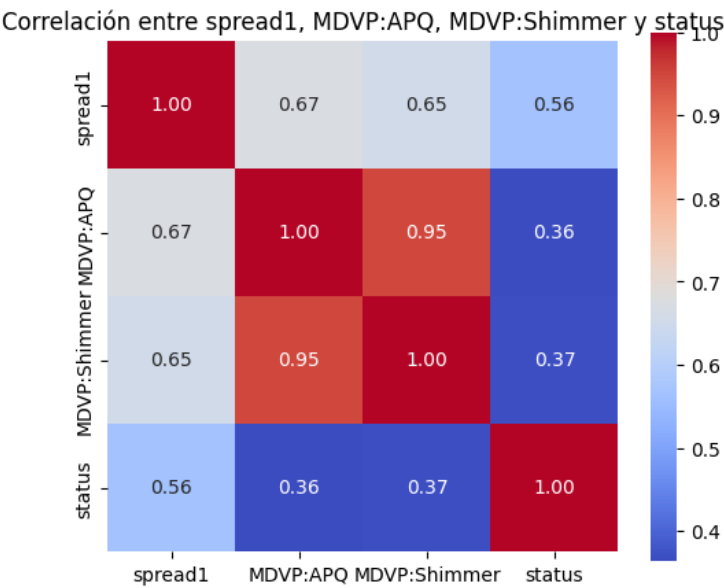
Comparativa de rendimiento en Test: Full (8 features) vs Reduced (2 features)

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

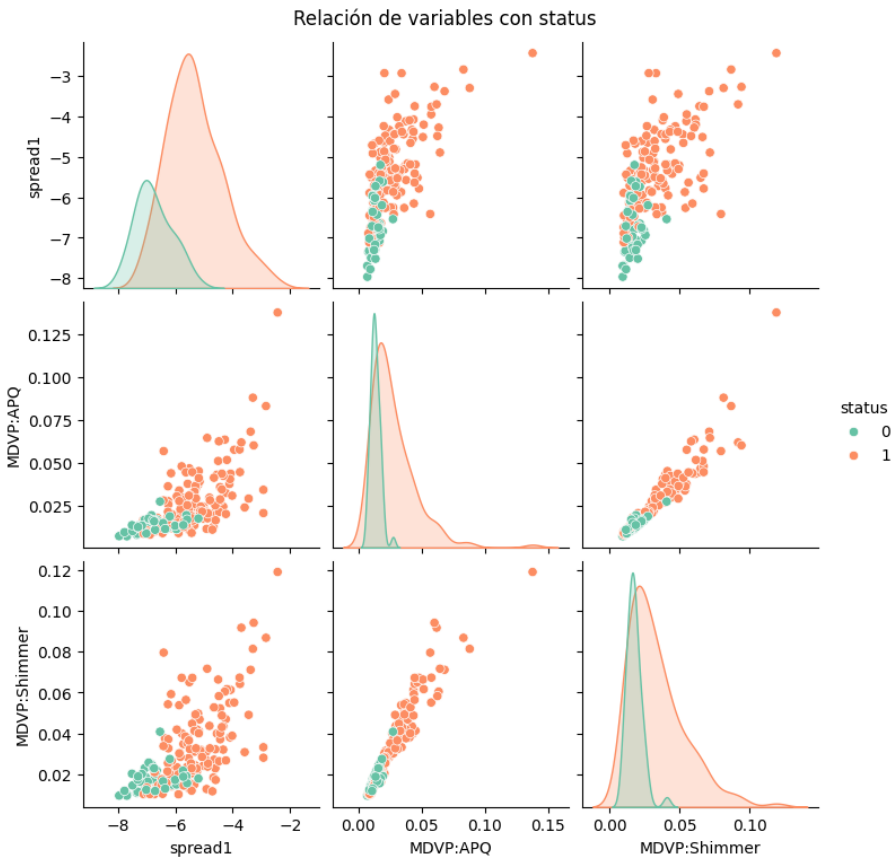
Reducing the feature set to just MDVP:APQ and MDVP:Shimmer not only simplifies the model by reducing variables by 75%, but also enhances its balanced classification capability, reflected in increased MCC. This confirms that these two features capture most of the discriminative information without sacrificing overall classifier reliability.



Correlation Map of Final Variables by Status:



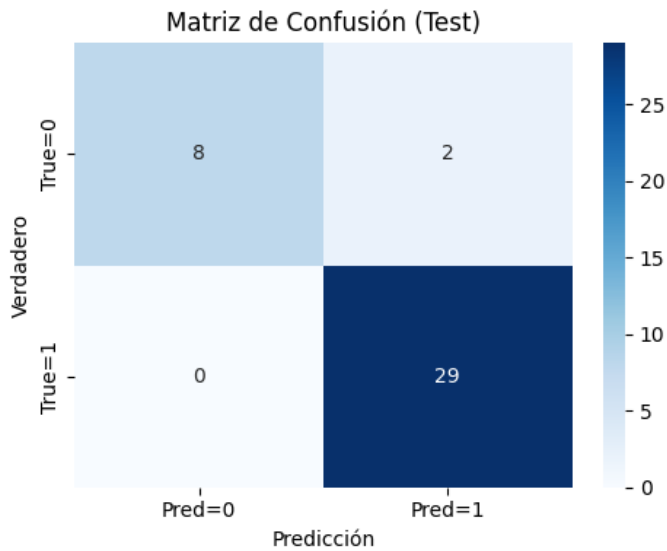
Discrimination of Variables between Patients with Parkinson's vs. Healthy Individuals:



Using the final optimized model (SVM-RBF optimized by MCC) trained on the three final variables (spread1, MDVP:APQ, and MDVP:Shimmer):**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

Final Confusion Matrix:



Hybrid Model Selection:

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