

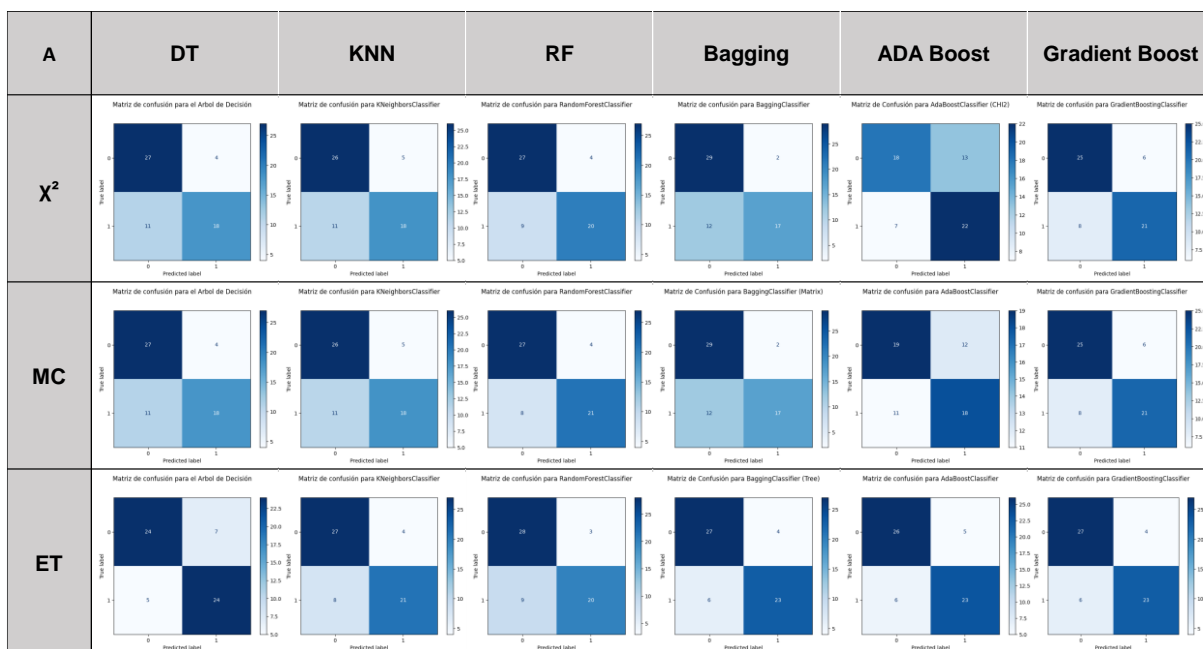
Apéndice de matrices de confusión de todas las pruebas

En este documento se encuentran las matrices de confusión de todas las pruebas realizadas, siendo los modelos (Árbol de decisión o DT, K-Vecinos más cercanos o KNN, Bosque Aleatorio o RF, *Bagging*, *ADA Boost* y *Gradient Boost*) distribuidos horizontalmente y los clasificadores (Chi-Cuadrado o χ^2 , Matriz de Correlación o MC y *Extra Tree* o ET) distribuidos verticalmente.

En la Figura 1 se observan las matrices de confusión correspondiente a las pruebas realizadas con el conjunto de datos de la vocal A del corpus PC-GITA, en donde la clase 0 es para pacientes con controles sanos y la clase 1 para pacientes con la Enfermedad de Parkinson (EP). En todos los modelos entrenados para este conjunto de datos, se aprecia que los subconjuntos de datos con las características seleccionadas con ayuda clasificador *Extra Tree* obtiene los mejores resultados, destacando los modelos Gradient Boost y Bagging, debido a su bajo número de datos mal clasificados. Muy de cerca, se encuentra el modelo ADA Boost. *Decision Tree* o DT es el modelo que menor rendimiento cuenta en combinación con el subconjunto de datos del clasificador *Extra Tree*.

Figura 1

Matrices de confusión correspondientes a las pruebas con el conjunto de datos de la vocal A.

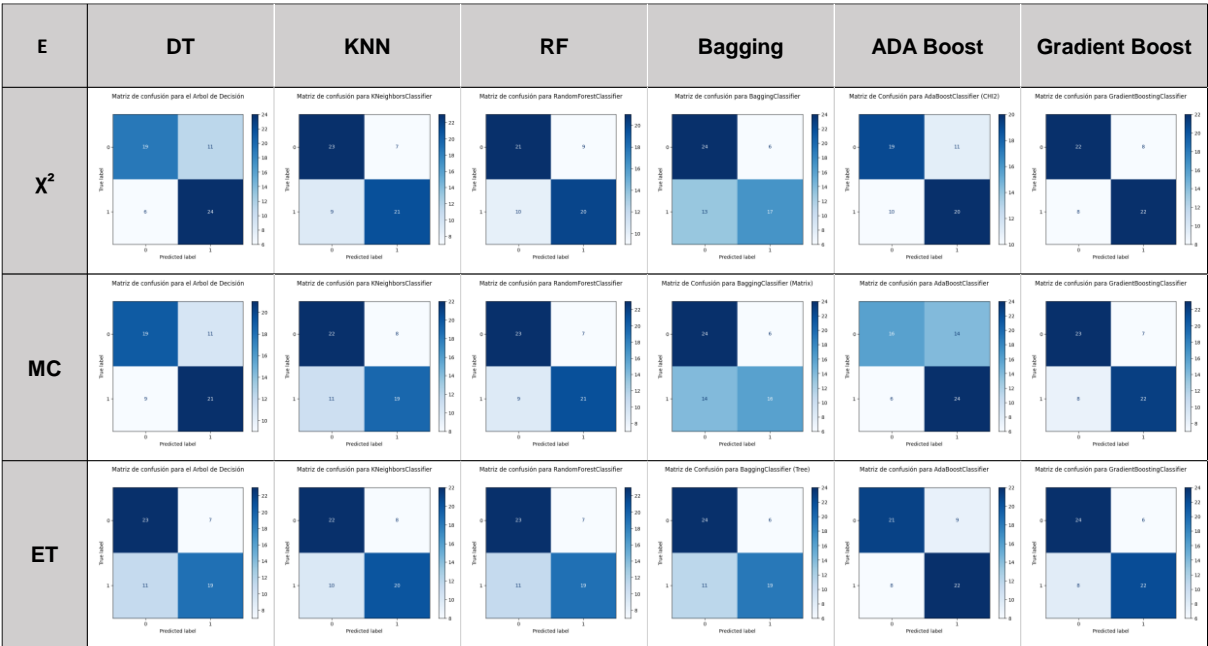


La Figura 2 contiene las matrices de confusión correspondiente a las pruebas realizadas con el conjunto de datos de la vocal E del corpus PC-GITA, en donde la clase 0 es para pacientes con controles sanos y la clase 1 para pacientes con la Enfermedad de Parkinson (EP). El modelo Gradient Boost se destaca en los tres subconjuntos de datos, siendo el subconjunto del clasificador *Extra Tree* el que mejor resultados obtiene, también, los modelos ADA Boost y Bagging obtienen sus mejores resultados con ET. Los modelos

DT y KNN obtienen su mejor resultado con el subconjunto X^2 , mientras que el modelo RF obtiene su mejor resultado con el subconjunto MC.

Figura 2

Matrices de confusión correspondientes a las pruebas con el conjunto de datos de la vocal E.

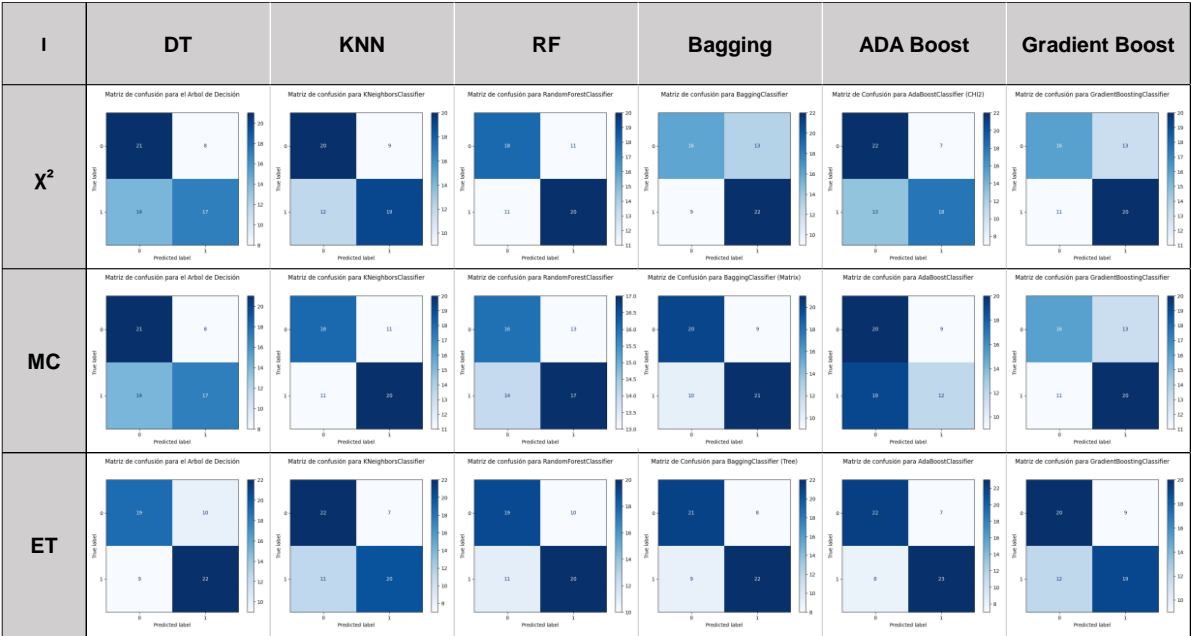


La Figura 3**Figura 2** contiene las matrices de confusión correspondiente a las pruebas realizadas con el conjunto de datos de la vocal I del corpus PC-GITA, en donde la clase 0 es para pacientes con controles sanos y la clase 1 para pacientes con la Enfermedad de Parkinson (EP). Para este conjunto de datos, el modelo ADA Boost es el que más destaca con el subconjunto de datos del clasificador ET, que, en contraste, muestra un pobre rendimiento con los otros subconjuntos de datos, especialmente con el subconjunto de datos MC. Esta situación se repite en los demás modelos, que obtienen los mejores

resultados con el subconjunto de datos del clasificador ET, mientras que presentan resultados inferiores con los demás subconjuntos.

Figura 3

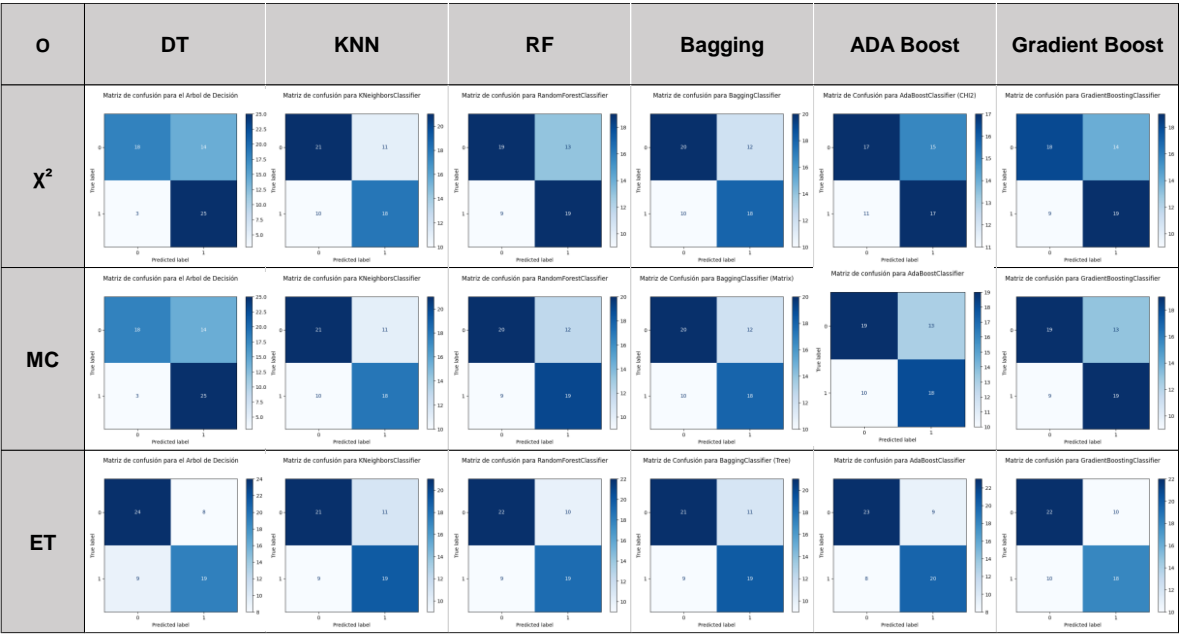
Matrices de confusión correspondientes a las pruebas con el conjunto de datos de la vocal I.



La Figura 4Figura 2 contiene las matrices de confusión correspondiente a las pruebas realizadas con el conjunto de datos de la vocal O del corpus PC-GITA, en donde la clase 0 es para pacientes con controles sanos y la clase 1 para pacientes con la Enfermedad de Parkinson (EP). Para este conjunto de datos, el clasificador *Extra Tree* obtuvo los mejores resultados en todos los modelos, destacando los modelos *ADA Boost* y *Decision Tree*, que obtienen un número considerable de datos correctamente clasificados. Los demás modelos presentan resultados similares, con uno o dos datos de más incorrectamente clasificados.

Figura 4

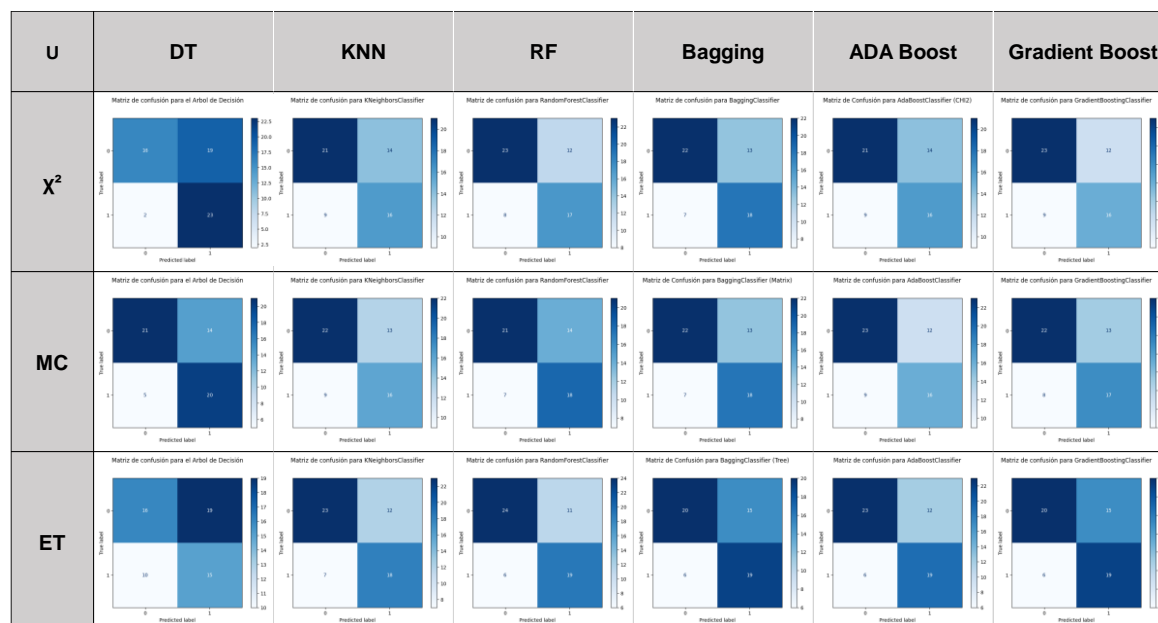
Matrices de confusión correspondientes a las pruebas con el conjunto de datos de la vocal O.



La Figura 5Figura 4Figura 2 contiene las matrices de confusión correspondiente a las pruebas realizadas con el conjunto de datos de la vocal U del corpus PC-GITA, en donde la clase 0 es para pacientes con controles sanos y la clase 1 para pacientes con la Enfermedad de Parkinson (EP). Este conjunto de datos obtuvo en promedio los peores resultados de los conjuntos de datos utilizados, siendo el modelo *Random Forest* con el subconjunto de datos ET el que mejor resultado obtuvo, seguido del modelo ADA Boost con el mismo subconjunto de datos.

Figura 5

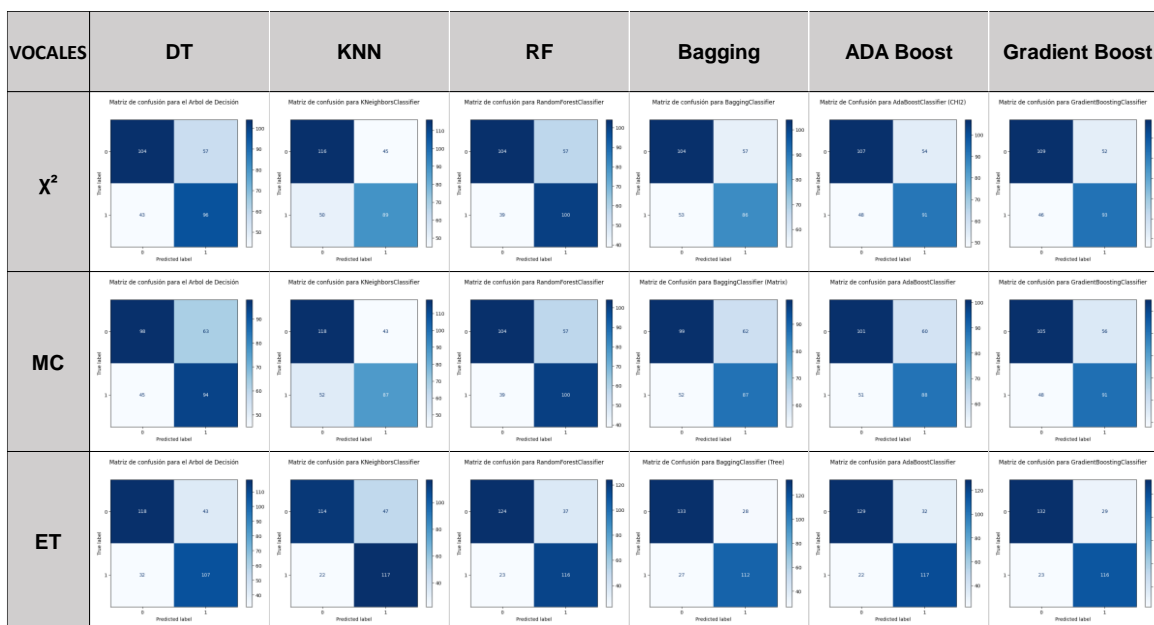
Matrices de confusión correspondientes a las pruebas con el conjunto de datos de la vocal U.



La Figura 6**Figura 2** contiene las matrices de confusión correspondiente a las pruebas realizadas con el conjunto de datos de todas las VOCALES juntas del corpus PC-GITA, en donde la clase 0 es para pacientes con controles sanos y la clase 1 para pacientes con la Enfermedad de Parkinson (EP). En este conjunto de datos el modelo Gradient Boost, usando el subconjunto de datos ET, obtuvo el mejor resultado de datos correctamente clasificados. Siguiendo la tendencia, los modelos de este conjunto de datos obtuvieron los mejores resultados con el subconjunto de datos ET.

Figura 6

Matrices de confusión correspondientes a las pruebas con el conjunto de datos de todas las VOALES juntas.



La Figura 7 contiene las matrices de confusión correspondiente a las pruebas realizadas con el conjunto de datos PARKINSONS, en donde la clase 0 es para pacientes con controles sanos y la clase 1 para pacientes con la Enfermedad de Parkinson (EP). En estas matrices de confusión se hace evidente el desbalance de datos presente en el conjunto de datos, siendo los modelos entrenados propensos a clasificar pacientes con controles sanos como pacientes con EP. Los modelos que se destacan en este conjunto son *Decision Tree* o *DT* y Gradient Boost con el subconjunto de datos X^2 , teniendo un comportamiento opuesto a lo visto con los conjuntos de datos con las vocales del corpus PC-GITA.

Figura 7

Matrices de confusión correspondientes a las pruebas con el conjunto de datos de PARKINSONS.

