

## A. Otras pruebas

A lo largo de todo el proceso de pruebas y experimentación se han probado distintos conjuntos de features basados en los empleados en algunos estudios previos. Para poder reconocer cada uno de ellos, se les va a proporcionar un alias: ODE-RGRU [1], EBM [2], Minimal-RNN [3], Sarica [4] y ADNIMERGE (conjunto de datos completo), siendo el de Sarica el utilizado finalmente para el análisis de los métodos estudiados en este trabajo.

El método ODE-RGRU (Ordinary Differential Equation-Recurrent Gated Recurrent Unit) utiliza un framework de aprendizaje geométrico basado en ecuaciones diferenciales ordinarias junto con una red neuronal recurrente. El método EBM (Explainable Boosting Machine) combina la capacidad de predicción de métodos de *boosting* con la interpretabilidad de los modelos lineales utilizando un enfoque basado en Generalized Additive Models (GAMs). El enfoque Minimal-RNN se basa en una red neuronal recurrente simplificada, menos propensa al sobreajuste.

	ODE-RGRU	EBM	Minimal-RNN	Sarica	ADNIMERGE
AGE		✓		✓	✓
PTGENDER		✓		✓	✓
APOE4		✓		✓	✓
FDG			✓	✓	✓
AV45			✓		✓
ABETA		✓	✓	✓	✓
TAU			✓	✓	✓
PTAU			✓	✓	✓
CDRSB		✓	✓	✓	✓
ADAS11	✓	✓	✓	✓	✓
ADAS13	✓	✓	✓	✓	✓
ADASQ4				✓	✓
MMSE	✓	✓	✓	✓	✓
RAVLT_immediate			✓	✓	✓
RAVLT_learning			✓	✓	✓
RAVLT_forgetting			✓	✓	✓
RAVLT_perc_forgetting			✓	✓	✓
LDELTOTAL				✓	✓
DIGITSCOR				✓	✓
TRABSCOR				✓	✓
FAQ		✓	✓	✓	✓
MOCA			✓		✓
EcogPtMem			✓		✓
EcogPtLang			✓		✓
EcogPtVisspat			✓		✓
EcogPtPlan			✓		✓
EcogPtOrgan			✓		✓
EcogPtDivatt			✓		✓
EcogPtTotal			✓		✓
EcogSPMem			✓		✓
EcogSPLang			✓		✓
EcogSPVisspat			✓		✓
EcogSPPlan			✓		✓
EcogSPOrgan			✓		✓
EcogSPDivatt			✓		✓
EcogSPTotal			✓		✓
Ventricles	✓	✓	✓	✓	✓
Hippocampus	✓	✓	✓	✓	✓
WholeBrain	✓	✓	✓	✓	✓
Entorhinal	✓	✓	✓	✓	✓
Fusiform	✓	✓	✓	✓	✓
MidTemp	✓	✓	✓	✓	✓
ICV			✓	✓	✓
mPACCdigit				✓	✓
mPACCtrailsB				✓	✓

**Tabla A.1:** Subconjuntos de features de ADNIMERGE evaluados

## A.1. Random Survival Forests

### A.1.1. Métricas de evaluación

En la siguiente tabla se muestra una comparación de las métricas obtenidas para bl, m12, m24 y bl+m12 con los distintos conjuntos de features analizados.

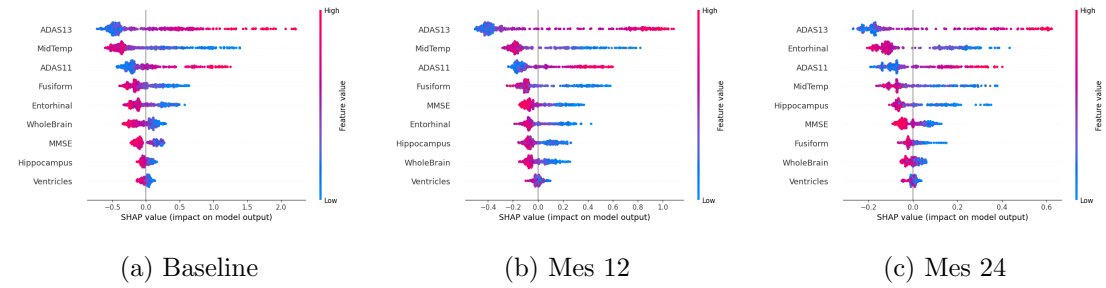
		c-index	accuracy	sensitivity	specificity	f1-score
ODE-RGRU	Baseline	0.805 $\pm$ 0.041	0.747 $\pm$ 0.046	0.703 $\pm$ 0.124	0.775 $\pm$ 0.050	0.655 $\pm$ 0.085
	Mes 12	0.853 $\pm$ 0.037	0.808 $\pm$ 0.024	0.729 $\pm$ 0.053	0.838 $\pm$ 0.020	0.667 $\pm$ 0.046
	Mes 24	0.824 $\pm$ 0.042	0.785 $\pm$ 0.013	0.571 $\pm$ 0.189	0.845 $\pm$ 0.057	0.483 $\pm$ 0.097
	Baseline + mes 12	0.837 $\pm$ 0.048	0.793 $\pm$ 0.019	0.730 $\pm$ 0.100	0.821 $\pm$ 0.033	0.674 $\pm$ 0.031
EBM	Baseline	0.837 $\pm$ 0.028	0.769 $\pm$ 0.041	0.725 $\pm$ 0.073	0.797 $\pm$ 0.049	0.684 $\pm$ 0.066
	Mes 12	0.874 $\pm$ 0.026	0.815 $\pm$ 0.033	0.761 $\pm$ 0.062	0.835 $\pm$ 0.028	0.685 $\pm$ 0.059
	Mes 24	0.850 $\pm$ 0.040	0.824 $\pm$ 0.025	0.612 $\pm$ 0.114	0.882 $\pm$ 0.053	0.558 $\pm$ 0.079
	Baseline + mes 12	0.874 $\pm$ 0.028	0.815 $\pm$ 0.037	0.761 $\pm$ 0.069	0.835 $\pm$ 0.031	0.685 $\pm$ 0.067
Minimal-RNN	Baseline	0.844 $\pm$ 0.027	0.796 $\pm$ 0.033	0.735 $\pm$ 0.053	0.831 $\pm$ 0.058	0.714 $\pm$ 0.047
	Mes 12	0.861 $\pm$ 0.032	0.809 $\pm$ 0.031	<b>0.767 <math>\pm</math> 0.065</b>	0.824 $\pm$ 0.046	0.694 $\pm$ 0.062
	Mes 24	0.857 $\pm$ 0.047	0.821 $\pm$ 0.034	0.650 $\pm$ 0.150	0.887 $\pm$ 0.049	0.603 $\pm$ 0.094
	Baseline + mes 12	0.855 $\pm$ 0.032	0.805 $\pm$ 0.023	0.742 $\pm$ 0.095	0.846 $\pm$ 0.042	0.704 $\pm$ 0.029
Sarica	Baseline	0.844 $\pm$ 0.031	0.777 $\pm$ 0.028	0.738 $\pm$ 0.061	0.802 $\pm$ 0.058	0.697 $\pm$ 0.037
	Mes 12	<b>0.875 <math>\pm</math> 0.032</b>	0.813 $\pm$ 0.021	0.760 $\pm$ 0.070	0.831 $\pm$ 0.040	0.682 $\pm$ 0.035
	Mes 24	0.872 $\pm$ 0.035	<b>0.836 <math>\pm</math> 0.030</b>	0.653 $\pm$ 0.150	0.888 $\pm$ 0.042	0.591 $\pm$ 0.058
	Baseline + mes 12	0.869 $\pm$ 0.032	0.811 $\pm$ 0.040	0.746 $\pm$ 0.076	0.839 $\pm$ 0.069	0.701 $\pm$ 0.037
ADNIMERGE	Baseline	0.846 $\pm$ 0.022	0.798 $\pm$ 0.034	0.747 $\pm$ 0.041	0.828 $\pm$ 0.069	<b>0.720 <math>\pm</math> 0.046</b>
	Mes 12	0.874 $\pm$ 0.031	0.830 $\pm$ 0.018	0.746 $\pm$ 0.090	0.862 $\pm$ 0.029	0.697 $\pm$ 0.049
	Mes 24	0.866 $\pm$ 0.034	0.827 $\pm$ 0.034	0.596 $\pm$ 0.150	<b>0.891 <math>\pm</math> 0.033</b>	0.556 $\pm$ 0.037
	Baseline + mes 12	0.865 $\pm$ 0.029	0.811 $\pm$ 0.037	0.694 $\pm$ 0.073	0.860 $\pm$ 0.063	0.686 $\pm$ 0.032

**Tabla A.2:** Media y desviación típica de las métricas obtenidas en RSF con los diferentes conjuntos de features considerados

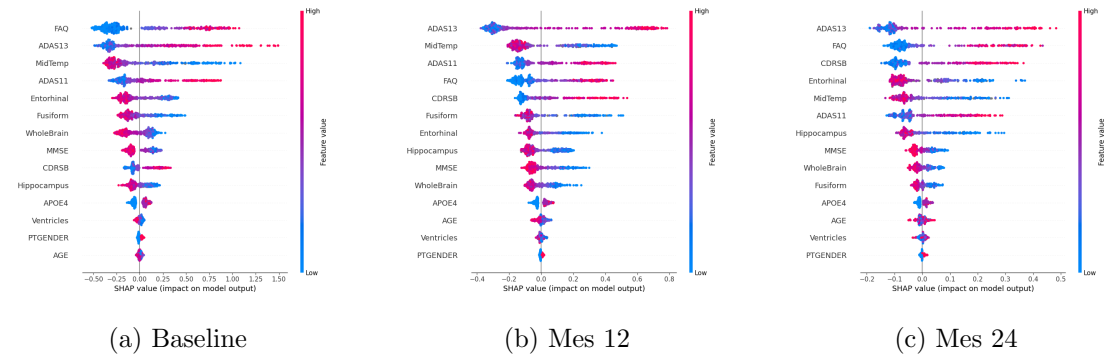
Como ya se ha comentado anteriormente en la parte principal de este documento, el modelo de Sarica es el que mejores resultados proporciona, especialmente en m12 y con datos longitudinales, mostrando resultados consistentes y equilibrados entre folds y visitas. Además, destaca también por mantener buenos valores de sensitivity a lo largo de todas las pruebas, una característica que se ha considerado de gran importancia en el contexto de este problema para favorecer el diagnóstico temprano y la reducción de falsos negativos.

En cuanto al resto de modelos, ODE-RGRU tiene el peor rendimiento debido a su limitación a ADAS11, ADAS13 y features MRI, perdiendo información valiosa de otras features como mPACC, FAQ o FDG. EBM es el mejor modelo con datos longitudinales, equilibrando bien el número de features utilizadas y los resultados obtenidos, a pesar de usar menos features que otros modelos. Minimal-RNN ofrece resultados similares a Sarica, superándolo en specificity y f1-score en algunos casos, pero tiene un rendimiento inferior en análisis de supervivencia (c-index) con datos distintos de bl. Por último, utilizar el conjunto ADNIMERGE completo da muy buenos resultados en general, aunque su rendimiento en m24 disminuye y es inferior a Sarica y EBM en datos longitudinales, destacando negativamente en sensitivity.

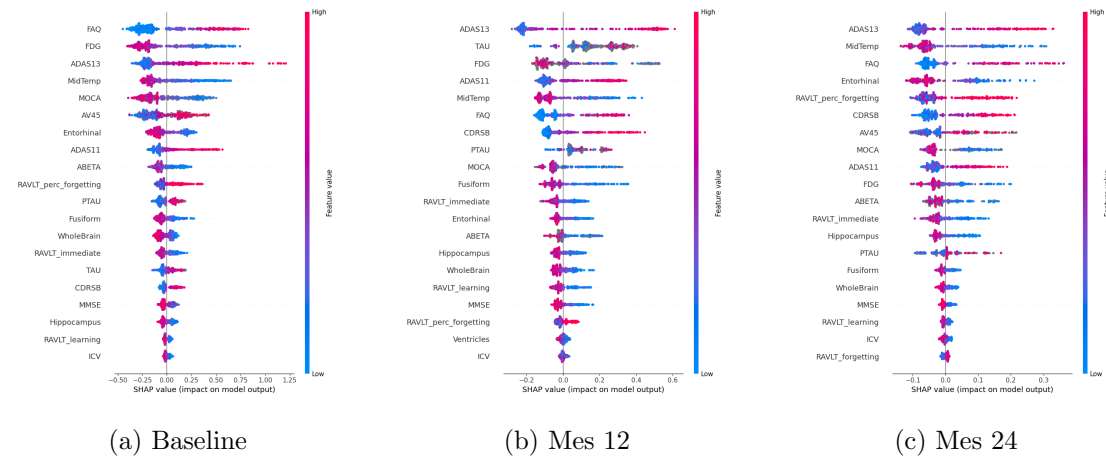
### A.1.2. Explicabilidad global



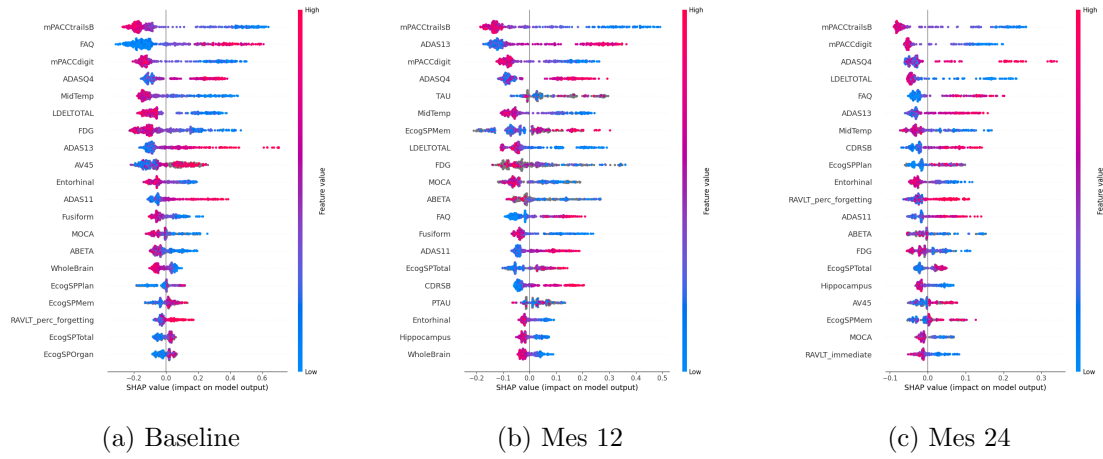
**Figura A.1:** Gráficas SHAP de explicabilidad global del modelo RSF con el conjunto de features ODE-RGRU



**Figura A.2:** Gráficas SHAP de explicabilidad global del modelo RSF con el conjunto de features EBM



**Figura A.3:** Gráficas SHAP de explicabilidad global del modelo RSF con el conjunto de features Minimal-RNN



**Figura A.4:** Gráficas SHAP de explicabilidad global del modelo RSF con el conjunto de features completo de ADNIMERGE

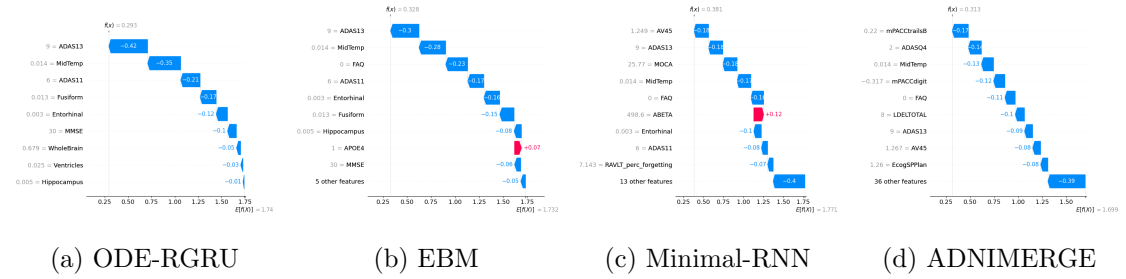
Los resultados de explicabilidad global obtenidos con cada uno de los conjuntos de features son similares a los analizados en la parte principal de la memoria. La mayor parte de las diferencias entre modelos se deben principalmente a la presencia de nuevas features en algunos conjuntos o a la ausencia de algunas de las features más importantes del conjunto de Sarica.

En RSF, debido a la ausencia de las features mPACC, FDG y LDELTOTAL al utilizar los conjuntos ODE-RGRU y EBM, en estos modelos adquieren una mayor importancia ADAS13 y ADAS11 (ver Figuras A.1 y A.2). En el caso de Minimal-RNN (ver Figura A.3) sí aparece FDG, por lo que las features con mayor impacto son más similares a las de Sarica (FAQ, FDG, MidTemp y ADAS13).

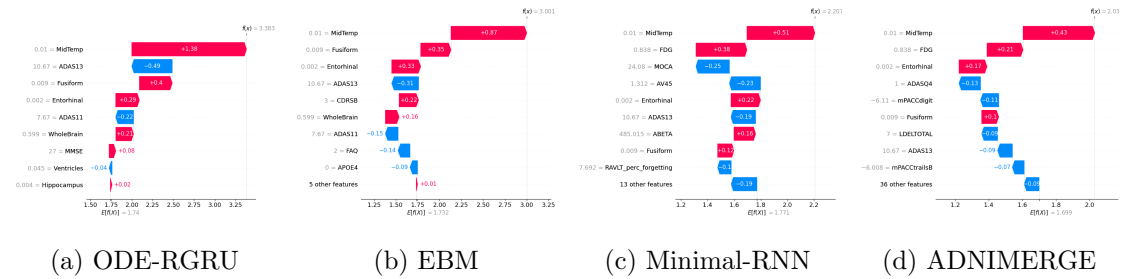
Por otra parte, al utilizar el conjunto ADNIMERGE completo, así como también Minimal-RNN, se introducen algunas features que no se tenían en cuenta en el conjunto de Sarica. Es el caso de MOCA y AV45, que destacan especialmente en Minimal-RNN (ver Figura A.3), o ADASQ4, que tiene una gran influencia en los modelos que utilizan el conjunto completo (ver Figura A.4). Además, con el conjunto ADNIMERGE completo se obtienen unos resultados que coinciden mayormente con los del conjunto de Sarica, pues entre las features más importantes se observan FAQ, MidTemp, LDELTOTAL y los tests mPACC. No obstante, también destaca la aparición de ADASQ4 entre las 5 features más importantes, así como la presencia de las features Ecog y algunas diferencias sutiles en el orden.

No obstante, es preciso mencionar que, al igual que ocurría con los biomarcadores CSF, algunas de estas nuevas features como MOCA y AV45 tienen porcentajes considerablemente altos de valores faltantes. De hecho, se tuvo que prescindir por completo de AV45 en m12 debido a que no había suficientes datos en esta visita como para poder llevar a cabo la imputación. Asimismo, las features PIB y FBB, que estaban presentes en el conjunto completo de ADNIMERGE, no han sido consideradas para ninguna de las visitas por este mismo motivo.

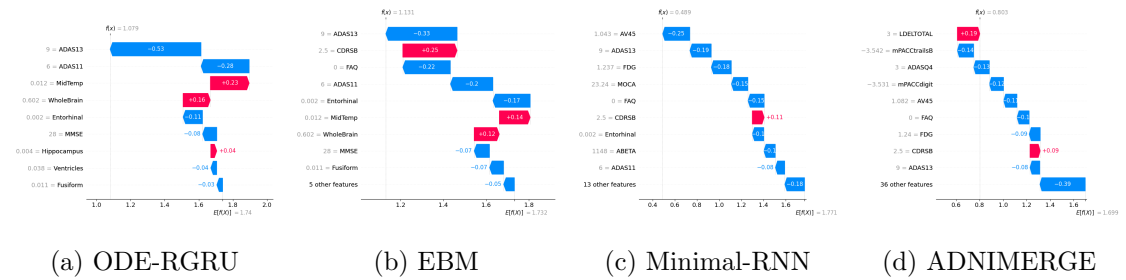
### A.1.3. Explicabilidad local



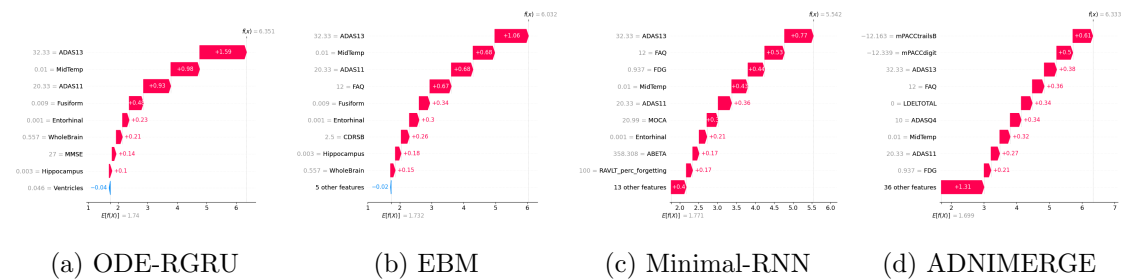
**Figura A.5:** Gráficas SHAP de explicabilidad local de los modelos RSF para el paciente sMCI#1



**Figura A.6:** Gráficas SHAP de explicabilidad local de los modelos RSF para el paciente sMCI#2



**Figura A.7:** Gráficas SHAP de explicabilidad local de los modelos RSF para el paciente pMCI#1



**Figura A.8:** Gráficas SHAP de explicabilidad local de los modelos RSF para el paciente pMCI#2

## A.2. Gradient Booster

### A.2.1. Métricas de evaluación

En la siguiente tabla se muestra una comparación de las métricas obtenidas para bl, m12, m24 y bl+m12 con los distintos conjuntos de features analizados.

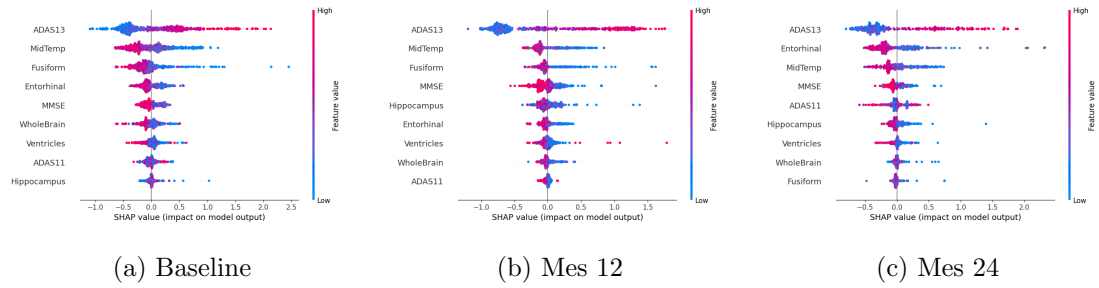
		c-index	accuracy	sensitivity	specificity	f1-score
ODE-RGRU	Baseline	0.799 $\pm$ 0.055	0.749 $\pm$ 0.036	0.727 $\pm$ 0.106	0.766 $\pm$ 0.056	0.665 $\pm$ 0.068
	Mes 12	0.845 $\pm$ 0.038	0.797 $\pm$ 0.059	0.675 $\pm$ 0.165	0.844 $\pm$ 0.027	0.633 $\pm$ 0.126
	Mes 24	0.811 $\pm$ 0.023	0.798 $\pm$ 0.009	0.467 $\pm$ 0.136	0.877 $\pm$ 0.047	0.453 $\pm$ 0.106
	Baseline + mes 12	0.826 $\pm$ 0.051	0.801 $\pm$ 0.043	0.732 $\pm$ 0.108	0.832 $\pm$ 0.042	0.683 $\pm$ 0.063
EBM	Baseline	0.831 $\pm$ 0.043	0.775 $\pm$ 0.063	0.730 $\pm$ 0.083	0.802 $\pm$ 0.063	0.693 $\pm$ 0.087
	Mes 12	0.849 $\pm$ 0.039	0.819 $\pm$ 0.041	0.690 $\pm$ 0.122	0.865 $\pm$ 0.024	0.664 $\pm$ 0.097
	Mes 24	0.839 $\pm$ 0.035	0.810 $\pm$ 0.038	0.532 $\pm$ 0.145	0.882 $\pm$ 0.069	0.502 $\pm$ 0.112
	Baseline + mes 12	<b>0.862 <math>\pm</math> 0.040</b>	0.819 $\pm$ 0.057	0.720 $\pm$ 0.102	0.860 $\pm$ 0.063	0.702 $\pm$ 0.076
Minimal-RNN	Baseline	0.843 $\pm$ 0.054	0.775 $\pm$ 0.059	0.691 $\pm$ 0.091	0.827 $\pm$ 0.080	0.683 $\pm$ 0.066
	Mes 12	0.844 $\pm$ 0.045	0.784 $\pm$ 0.044	0.707 $\pm$ 0.083	0.814 $\pm$ 0.045	0.635 $\pm$ 0.069
	Mes 24	0.861 $\pm$ 0.051	<b>0.824 <math>\pm</math> 0.019</b>	0.525 $\pm$ 0.142	<b>0.904 <math>\pm</math> 0.054</b>	0.518 $\pm$ 0.061
	Baseline + mes 12	0.861 $\pm$ 0.041	0.796 $\pm$ 0.045	0.663 $\pm$ 0.126	0.854 $\pm$ 0.058	0.655 $\pm$ 0.070
Sarica	Baseline	0.843 $\pm$ 0.025	0.788 $\pm$ 0.031	<b>0.745 <math>\pm</math> 0.077</b>	0.816 $\pm$ 0.061	<b>0.708 <math>\pm</math> 0.052</b>
	Mes 12	0.860 $\pm$ 0.050	0.815 $\pm$ 0.048	0.727 $\pm$ 0.064	0.845 $\pm$ 0.059	0.678 $\pm$ 0.057
	Mes 24	0.833 $\pm$ 0.032	0.804 $\pm$ 0.028	0.456 $\pm$ 0.100	0.889 $\pm$ 0.049	0.458 $\pm$ 0.116
	Baseline + mes 12	0.843 $\pm$ 0.033	0.791 $\pm$ 0.051	0.713 $\pm$ 0.095	0.824 $\pm$ 0.067	0.670 $\pm$ 0.056
ADNIMERGE	Baseline	0.841 $\pm$ 0.042	0.792 $\pm$ 0.058	0.701 $\pm$ 0.064	0.845 $\pm$ 0.084	0.701 $\pm$ 0.084
	Mes 12	0.839 $\pm$ 0.061	0.813 $\pm$ 0.030	0.652 $\pm$ 0.172	0.870 $\pm$ 0.036	0.638 $\pm$ 0.107
	Mes 24	0.832 $\pm$ 0.034	0.801 $\pm$ 0.035	0.455 $\pm$ 0.082	0.884 $\pm$ 0.041	0.458 $\pm$ 0.096
	Baseline + mes 12	0.842 $\pm$ 0.029	0.786 $\pm$ 0.039	0.625 $\pm$ 0.085	0.853 $\pm$ 0.037	0.632 $\pm$ 0.060

**Tabla A.3:** Media y desviación típica de las métricas obtenidas en GBSA con los diferentes conjuntos de features considerados

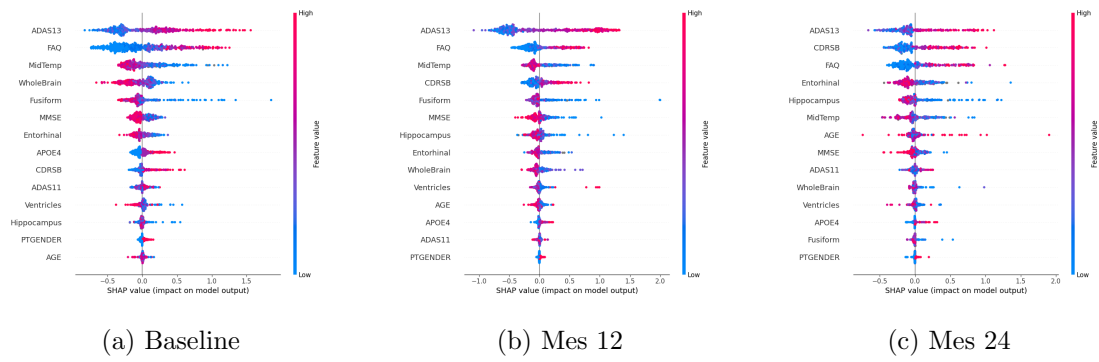
En el análisis de Gradient Booster se observa una tendencia similar a la analizada en Random Survival Forests (ver Tabla A.2). En este caso, el modelo de Sarica muestra un desempeño superior en todas las métricas en bl y m12 salvo en specificity. Sin embargo, en m24 su rendimiento es el más bajo, especialmente en sensitivity y f1-score.

Respecto al resto de modelos, ODE-RGRU persiste como el de peor rendimiento general, aunque muestra métricas de clasificación notables en datos longitudinales, siendo el mejor en sensitivity en este caso específico. Por otro lado, EBM sigue siendo el mejor modelo con datos longitudinales, mostrando además resultados muy equilibrados en todas las visitas. Minimal-RNN destaca positivamente en m24, y muestra un desempeño similar a Sarica en bl pero inferior en m12 respecto a Sarica y EBM. Por último, el uso del conjunto ADNIMERGE completo ofrece un buen rendimiento en bl, pero su eficacia disminuye en visitas más cercanas al evento de interés y con datos longitudinales, donde se posiciona como el peor modelo. Además, los valores de sensitivity obtenidos se encuentran entre los más bajos de todas las pruebas realizadas.

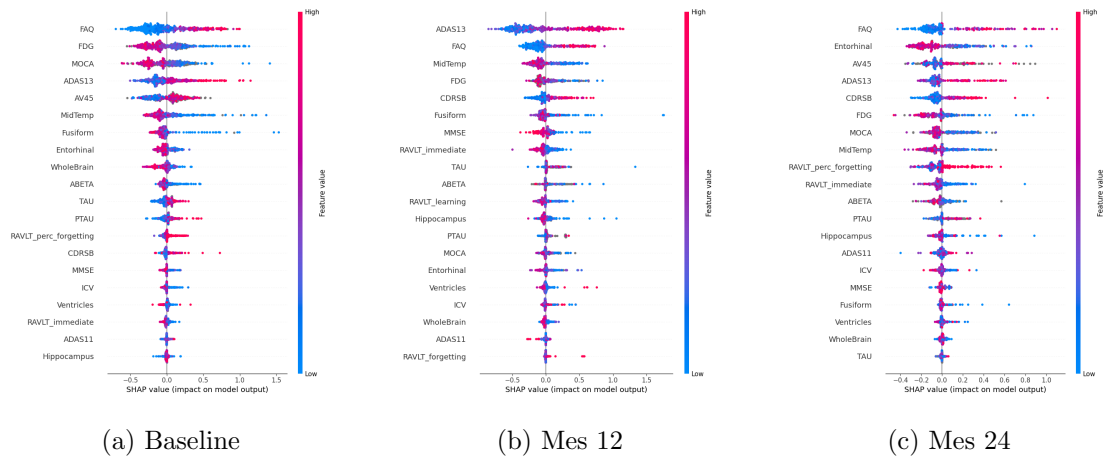
### A.2.2. Explicabilidad global



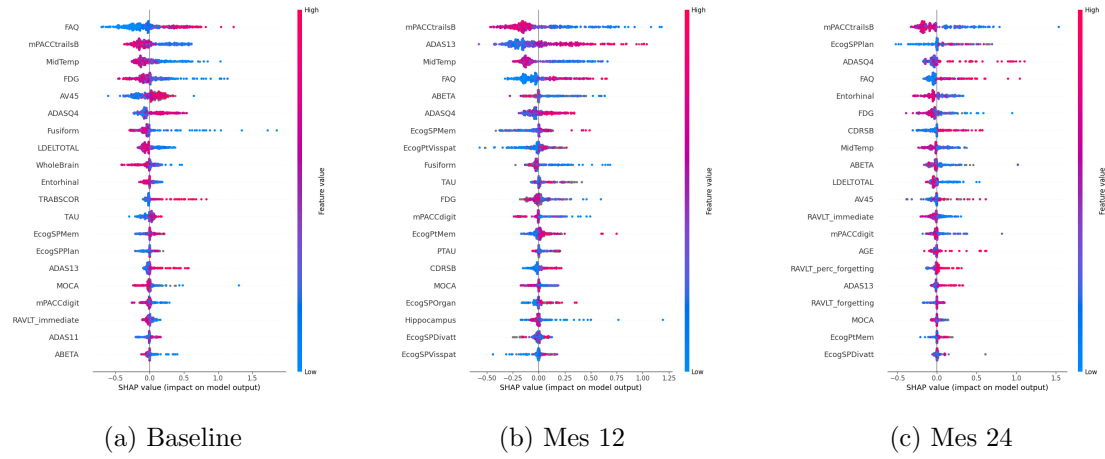
**Figura A.9:** Gráficas SHAP de explicabilidad global del modelo GBSA con el conjunto de features ODE-RGRU



**Figura A.10:** Gráficas SHAP de explicabilidad global del modelo GBSA con el conjunto de features EBM



**Figura A.11:** Gráficas SHAP de explicabilidad global del modelo GBSA con el conjunto de features Minimal-RNN

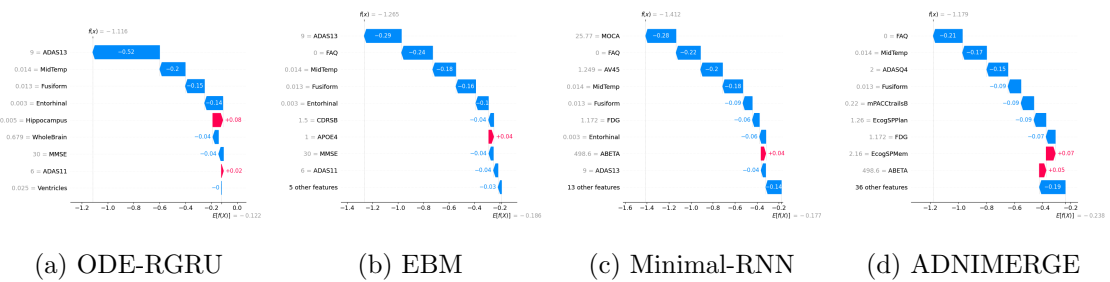


**Figura A.12:** Gráficas SHAP de explicabilidad global del modelo GBSA con el conjunto de features completo de ADNIMERGE

De nuevo, los resultados de explicabilidad global son similares a los obtenidos con el conjunto de Sarica, si bien hay algunas pequeñas diferencias respecto a la tendencia observada en RSF para estos conjuntos. Por ejemplo, destaca la disminución en la importancia de la feature ADAS11, especialmente en los modelos que utilizan los conjuntos ODE-RGRU y EBM, en los que tenía una mayor influencia por la ausencia de otras features importantes. Sin embargo, en GBSA tiene un impacto menor que la mayor parte de las features MRI (ver Figuras A.9 y A.10). En el conjunto Minimal-RNN vuelven a destacar MOCA y AV45 (ver Figura A.11), así como en el conjunto completo de ADNIMERGE se reduce ligeramente la importancia de ADASQ4 y destaca EcogSPPlan en m24 (ver Figura A.12).

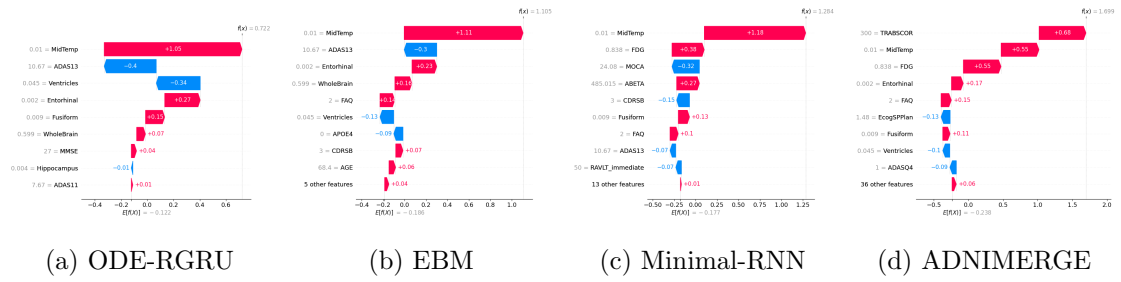
Globalmente se observa que las features con mayor influencia sobre los modelos son mPAC-CtrailsB, FAQ, FDG, MidTemp y ADAS13, lo que coincide mayormente con el análisis de explicabilidad realizado en la parte principal de esta memoria.

### A.2.3. Explicabilidad local

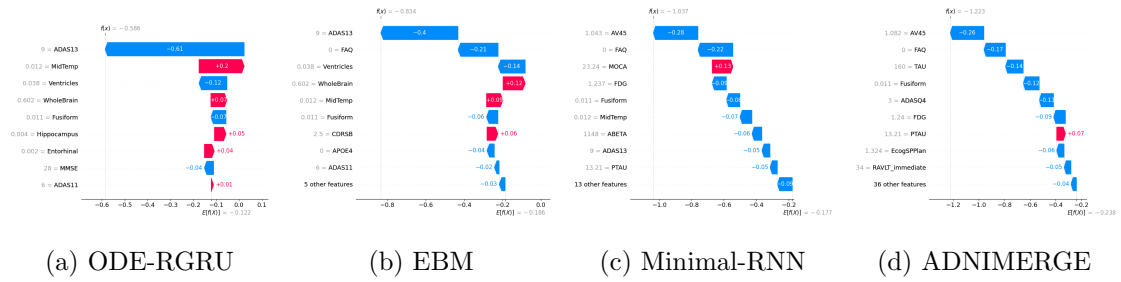


**Figura A.13:** Gráficas SHAP de explicabilidad local de los modelos GBSA para el paciente sMCI#1

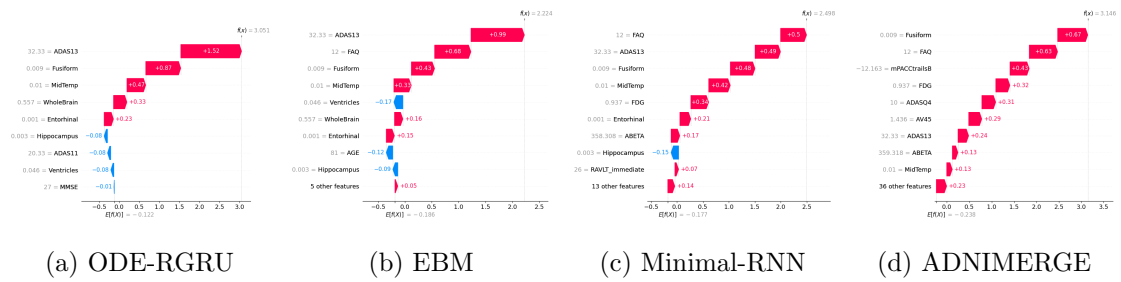




**Figura A.14:** Gráficas SHAP de explicabilidad local de los modelos GBSA para el paciente sMCI#2



**Figura A.15:** Gráficas SHAP de explicabilidad local de los modelos GBSA para el paciente pMCI#1



**Figura A.16:** Gráficas SHAP de explicabilidad local de los modelos GBSA para el paciente pMCI#2

## B. Código desarrollado

Todo el código desarrollado en este trabajo se encuentra disponible en el repositorio de GitHub <https://github.com/diegodomingo/adsurv>, donde además se incluyen instrucciones para su uso y reproducibilidad de los resultados analizados en este documento.

## Referencias

- [1] S. Jeong, W. Jung, J. Sohn y H. I. Suk, “Deep geometrical learning for Alzheimer’s disease progression modeling,” en *2022 IEEE International Conference on Data Mining (ICDM)*, IEEE, 2022, págs. 211-220. DOI: [10.1109/ICDM54844.2022.00031](https://doi.org/10.1109/ICDM54844.2022.00031).
- [2] J. Cai, W. Hu, J. Ma, A. Si, S. Chen, L. Gong et al., “Explainable machine learning with pairwise interactions for predicting conversion from mild cognitive impairment to Alzheimer’s disease utilizing multi-modalities data,” *Brain Sciences*, vol. 13, n.º 11, 2023. DOI: [10.3390/brainsci13111535](https://doi.org/10.3390/brainsci13111535).
- [3] M. Nguyen, T. He, L. An, D. C. Alexander, J. Feng y B. T. T. Yeo, “Predicting Alzheimer’s disease progression using deep recurrent neural networks,” *NeuroImage*, vol. 222, pág. 117 203, 2020. DOI: [10.1016/j.neuroimage.2020.117203](https://doi.org/10.1016/j.neuroimage.2020.117203).
- [4] A. Sarica, F. Aracri, M. G. Bianco, F. Arcuri, A. Quattrone, A. Quattrone et al., “Explainability of random survival forests in predicting conversion risk from mild cognitive impairment to Alzheimer’s disease,” *Brain Informatics*, vol. 10, n.º 1, pág. 31, 2023. DOI: [10.1186/s40708-023-00211-w](https://doi.org/10.1186/s40708-023-00211-w).