

## INTRODUCCIÓN

Los modelos de aprendizaje profundo suelen ser considerados “cajas negras”, debido a su problema de interpretabilidad. En este trabajo se investigó la interpretación de un modelo de redes neuronales recurrentes (RNNs) mediante el uso de valores SHAP (SHapley Additive exPlanations). El objetivo fue comprender la contribución de las principales variables satelitales (198 variables usadas) en la predicción de la concentración de material particulado fino (PM<sub>2.5</sub>) en la provincia de Córdoba, Argentina.

## METODOLOGÍA

Ejemplo -> Modelo con tres variables (A,B,C) -> **Stacked RNNs con GRUs**

Cada valor SHAP representa la contribución marginal (CM) de una variable a una predicción individual.

	A	B	C	Predicción
Instancia 1	10	6.5	0.5	24.2

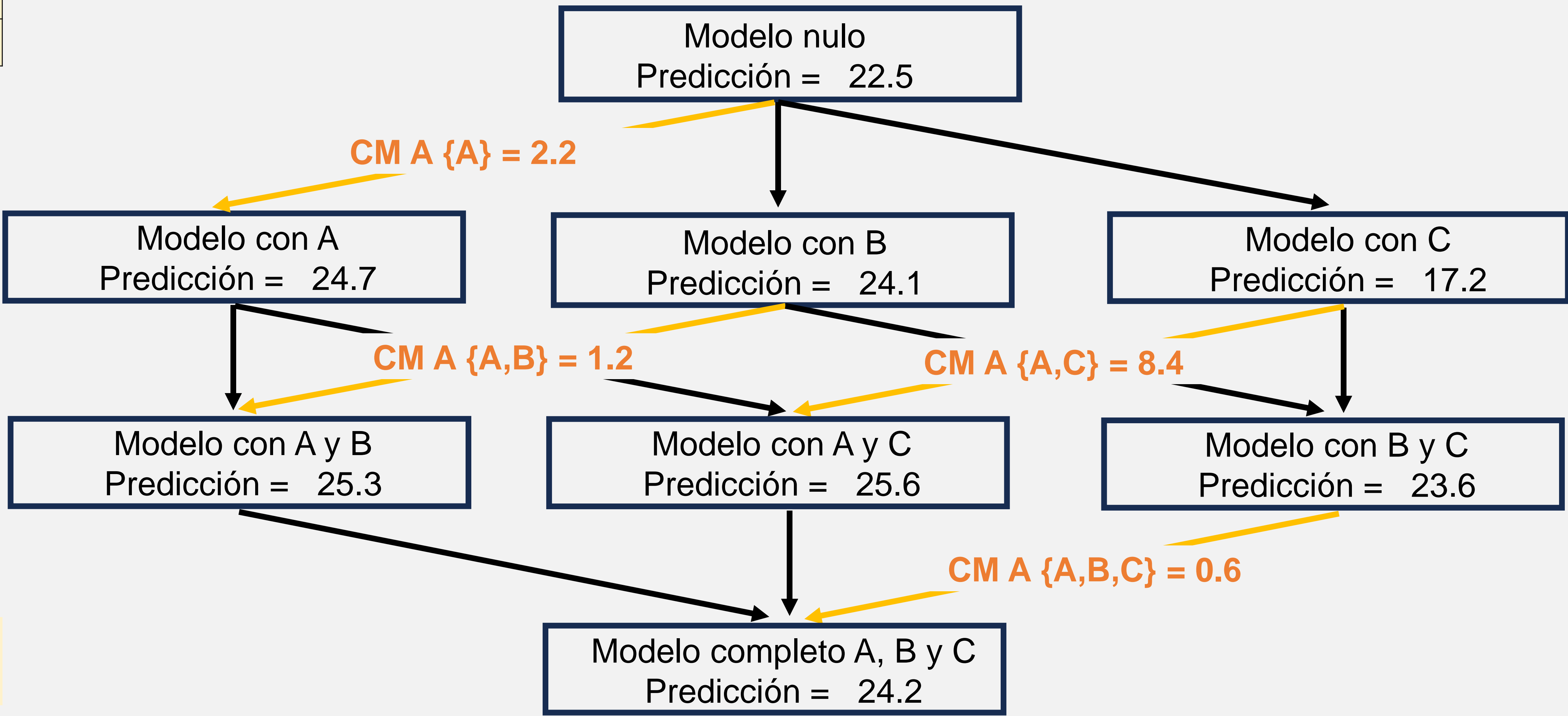
$$\text{Pred.}(\text{instancia1}) = b + \text{shapleyA}(\text{instancia1}) + \text{shapleyB}(\text{instancia1}) + \text{shapleyC}(\text{instancia1})$$

$$\text{shapleyA}(\text{instancia1}) = 1/3 \times 2.2 + 1/6 \times 1.2 + 1/6 \times 8.4 + 1/3 \times 0.6 = 2.55 \text{ (+)}$$

$$\text{shapleyB}(\text{instancia1}) = 1.250 \text{ (+)}$$

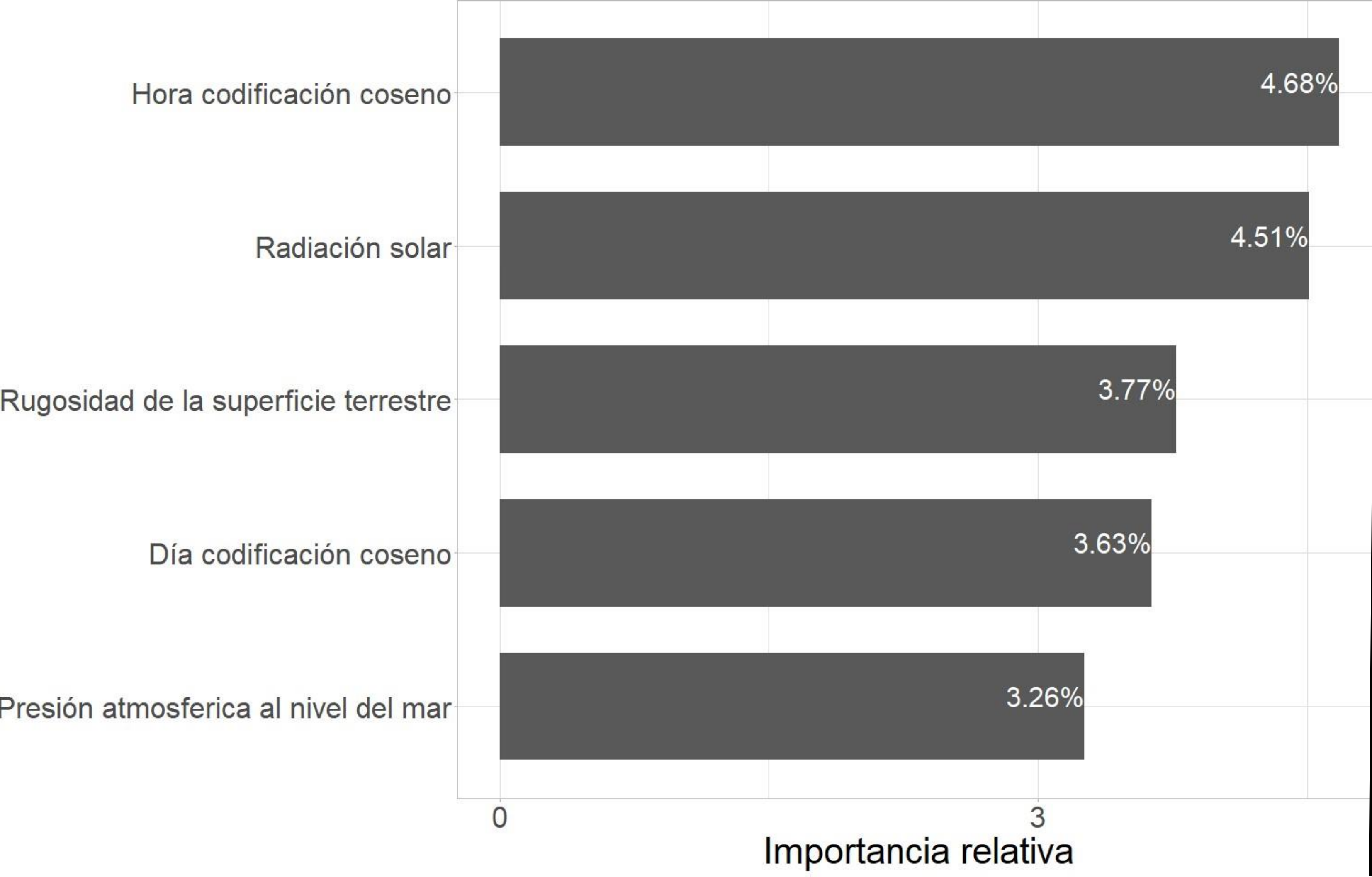
$$\text{shapleyC}(\text{instancia1}) = -2.100 \text{ (-)}$$

**Predicción final**  
 $24.2 = 22.5 + 2.550 + 1.250 - 2.100$

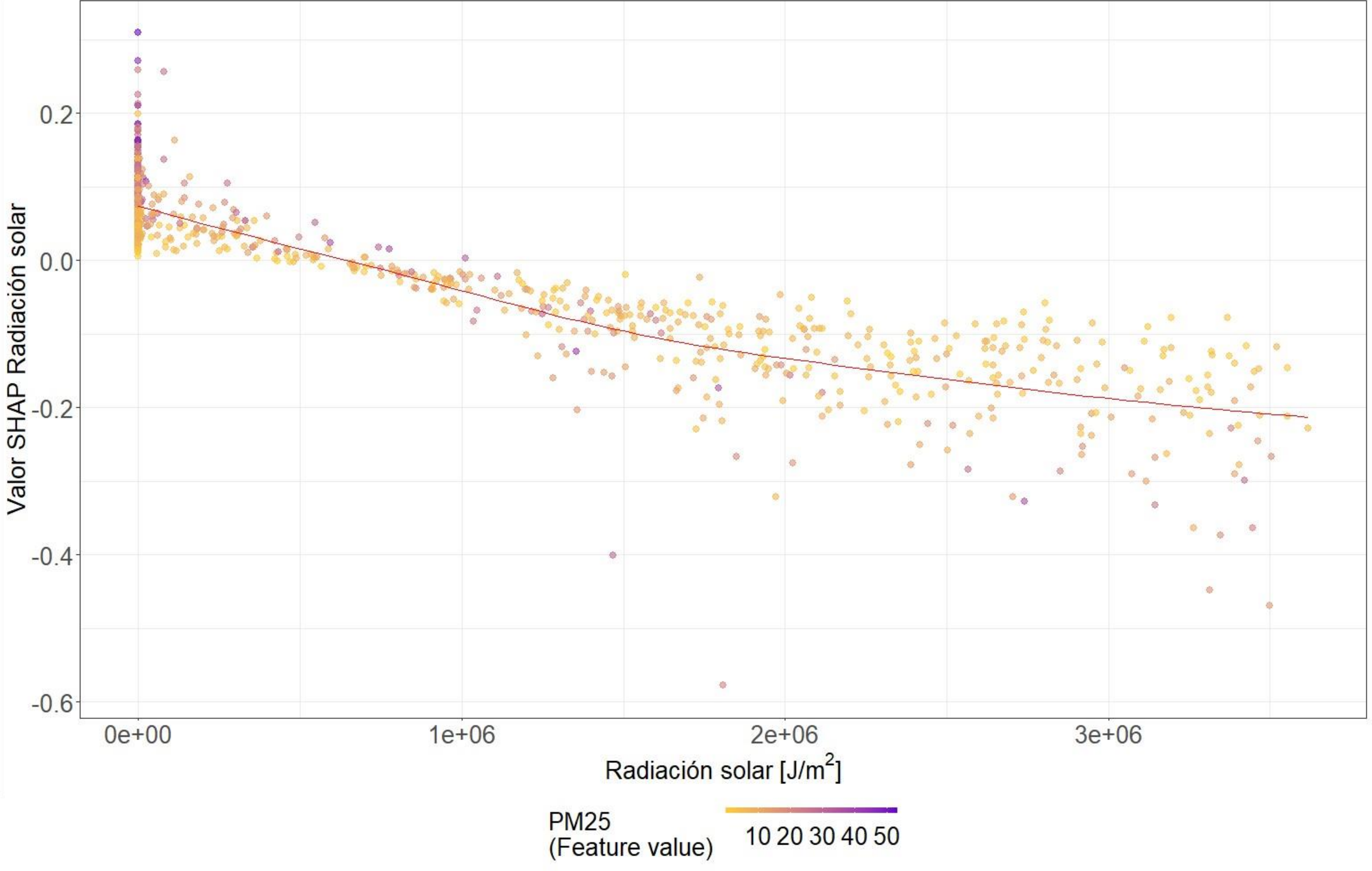


## RESULTADOS

### Importancia global de las variables en el modelo (Top 5)



### Contribución de la radiación solar en la predicción de PM<sub>2.5</sub>



## CONCLUSIÓN

Los valores SHAP representan una poderosa herramienta para la interpretación de modelos de aprendizaje profundo, permitiendo una comprensión más precisa sobre su funcionamiento y las relaciones subyacentes interpretadas por el algoritmo.