



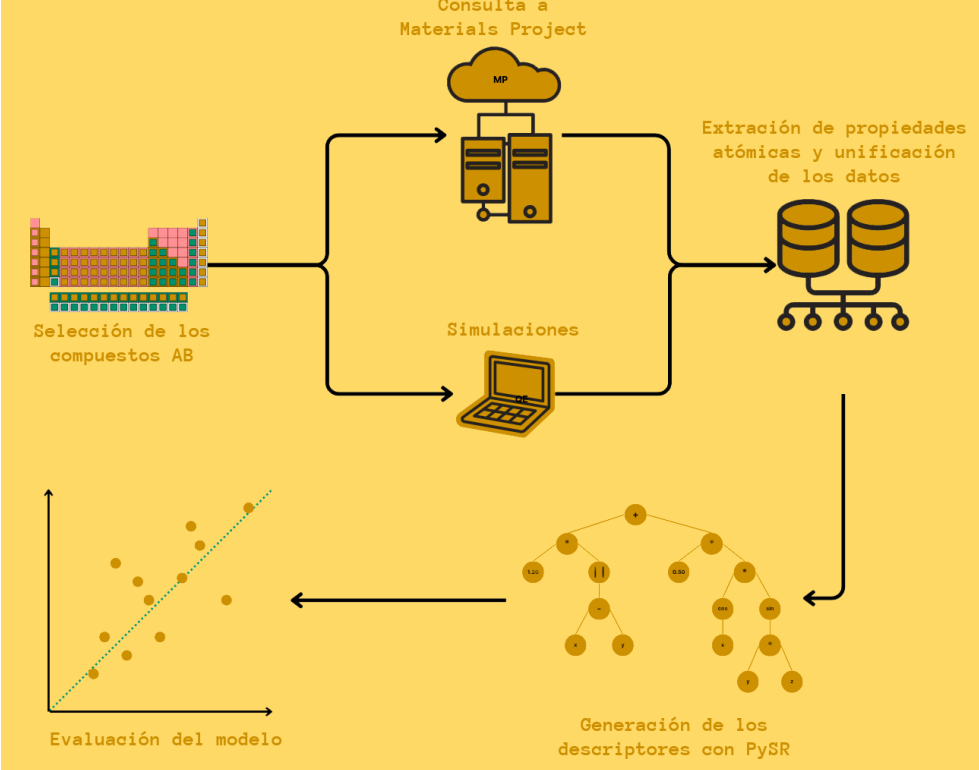
Resumen

El band gap es una propiedad fundamental en materiales semiconductores. En este trabajo se aplicó regresión simbólica para generar descriptores físicos interpretables que predican el band gap de compuestos binarios tipo AB con estructura cúbica tipo zinc blenda. Se utilizaron dos conjuntos de datos: uno obtenido del repositorio Materials Project (MP) y otro construido a partir de simulaciones propias (ZB). Se aplicó el algoritmo PySR y se validó el desempeño mediante validación cruzada. Los descriptores obtenidos muestran buen ajuste a los datos y revelan relaciones consistentes con propiedades atómicas fundamentales.

Introducción

La predicción del band gap es esencial para el diseño de materiales semiconductores en aplicaciones fotovoltaicas, optoelectrónicas y termoeléctricas. Los métodos tradicionales, como la teoría del funcional de la densidad (DFT), son precisos pero costosos. Por ello, el uso de inteligencia artificial, y en particular de modelos interpretables como la regresión simbólica, ha ganado relevancia. Este trabajo busca generar descriptores matemáticos simbólicos que relacionen propiedades atómicas con el band gap, permitiendo interpretabilidad y bajo costo computacional.

Metodología



Resultados

Compuestos III-V				
BN	BP	Ba	BSb	BBi
AlN	AlP	AlAs	AlSb	AlBi
GaN	GaP	GaAs	GaSb	GaBi
InN	InP	InAs	InSb	InBi
Compuestos II-VI				
BeO	BeS	BeSe	BeTe	MgO
MgS	MgSe	MgTe	CaO	CaS
CaSe	CaTe	SrO	SrS	SrSe
SrTe	BaO	BaS	BaSe	BaTe
ZnO	ZnS	ZnSe	ZnTe	CdO
CdS	CdSe	CdTe		

Tabla 1. Listado de los 48 compuestos binarios tipo AB utilizado, clasificados por su tipo químico.

Nombre	Abreviación	Unidad
Número atómico	AN	-
Número de electrones de valencia	NVE	-
Energía de ionización	IE	kJ/mol
Pesó atómico	AW	a.m.u.
Radio atómico calculado	ARC	pm
Radio atómico empírico	ARE	pm
Radio covalente	CR	pm
Afinidad electrónica	EA	kJ/mol
Electronegatividad de Pauling	EN	-

Tabla 2. Propiedades atómicas utilizadas para el entrenamiento del modelo de IA

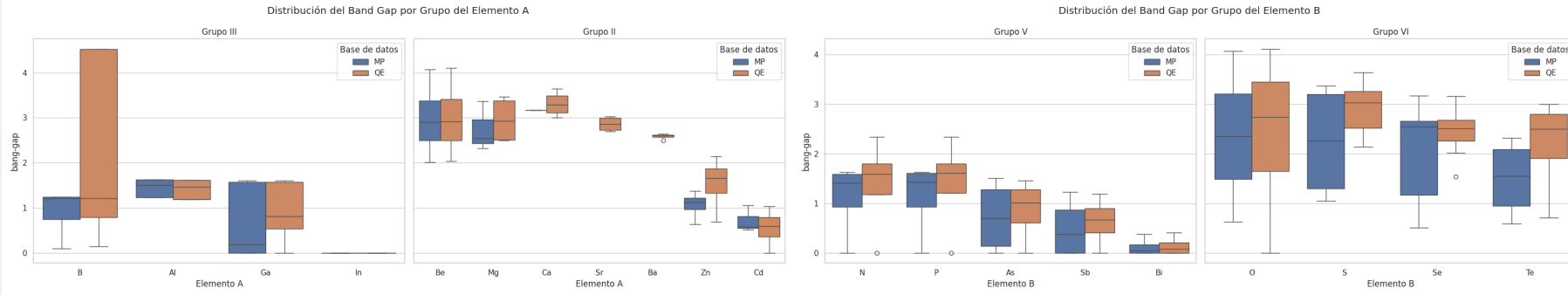


Figura 1. Distribución del band-gap por elemento en ambos conjuntos de datos. Materials Project (MP) y simulaciones con Quantum Espresso (QE).

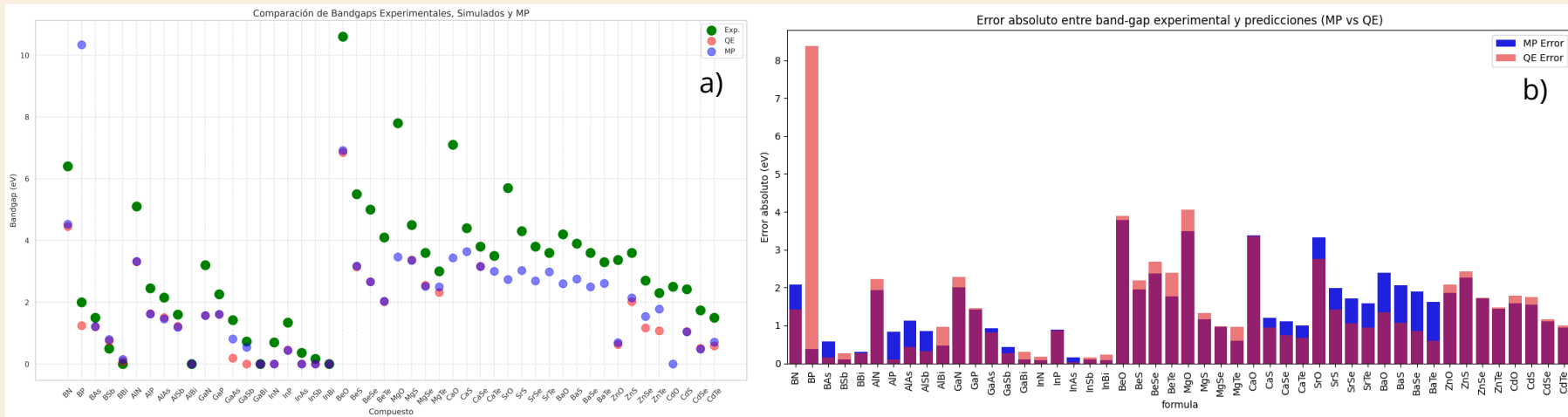


Figura 2. a) Distribución del band-gap y b) error absoluto de ambos conjuntos de datos con respecto a valores experimentales.

$$-0.171EN_b + NVE_b - \log \left(AW_a + 6.78\sqrt{AW_b}EN_a + 6.78EA_b - 10.1NVE_b \right) + \frac{2.20}{NVE_a}$$

Ecuación 1. Descriptor generado a partir del conjunto de datos obtenidos a partir de Materials Project (MP).

$$-3.54 \log \left(EN_a \right) + 1.80 + \frac{92.3}{AN_b + AW_a - \frac{14.2}{AN_a - AN_b + AW_a}}$$

Ecuación 2. Descriptor generado a partir del conjunto de datos obtenidos a partir simulaciones con Quantum Espresso (QE).

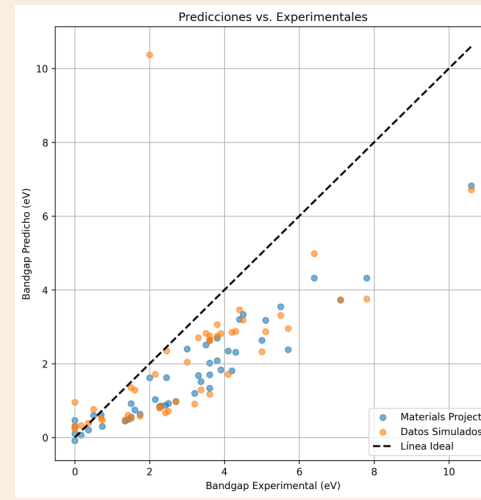


Figura 3. Comparación entre los valores de band-gap predichos por el modelo y los valores experimentales para las dos bases de datos evaluadas.

Conclusiones

Los resultados obtenidos en este estudio demuestran que el uso de propiedades atómicas prístinas como descriptores en modelos de regresión simbólica permite predecir con alta precisión el band-gap de compuestos semiconductores binarios. El modelo entrenado con datos de Materials Project superó al basado en simulaciones, lo que resalta la importancia de la calidad de los datos. La correlación con propiedades periódicas como la electronegatividad y afinidad electrónica respalda su capacidad para capturar tendencias electrónicas relevantes. Este enfoque representa una vía eficiente y físicamente fundamentada para acelerar el descubrimiento de materiales funcionales.

Referencias

- [1] C. Fritz, J. Fernandez-Serra, M. Soler, J. Chem. Phys. 144, 224101 (2016).
- [2] L. Fiedler, K. Shah, M. Bussman, A. Cangi, Phys. Rev. Materials 6, 040301 (2022).
- [3] W. Kohn, L. J. Sham, Phys. Rev. 140, A1133-A1138 (1965).
- [4] S.I. Simak, E.K. Delczeg-Czirjak, O. Eriksson, Sci Rep 15, 17212 (2025).
- [5] A. Mazheika, Y.-G. Wang, R. Valero, F. Viñes, F. Illas, L. M. Ghiringhelli, S. V. Levchenko and M. Scheffler, Nat. Commun., 13, 419 (2022).
- [6] Fu, X., Wang, J., Hu, X., Xu, W., Levchenko, S. V., & Han, Z.-K. Chem. Commun., 61, 5122-5125, (2025).
- [7] M. Cranmer, Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl. arXiv [Astro-Ph.IM]. Retrieved from <http://arxiv.org/abs/2305.01582> (2023).
- [8] Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., ... Persson, K. A. (07 2013). Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. APL Materials, 1(1), 011002. doi:10.1063/1.4812323.
- [9] P. Giannozzi, O. Baseggio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. de Gironcoli, P. Delugas, F. Ferrari Ruffino, A. Ferretti, N. Marzari, I. Timrov, A. Urru, S. Baroni; J. Chem. Phys. 152, 154105 (2020).
- [10] X. Hua, X. Chen, W. A. Goddard, Phys. Rev. B 55, 16103 (1997).

Agradecimientos

En agradecimiento a la Universidad de Guanajuato, al Verano Científico.