

La investigación reportada en esta tesis es parte de los programas de investigación del CICESE (Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California).

La investigación fue financiada por el SECIHTI (Secretaría de Ciencia, Humanidades, Tecnología e Innovación).

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México). El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo o titular de los Derechos de Autor.

CICESE © 2025, Todos los Derechos Reservados, CICESE

Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California



Maestría en Ciencias en Nanociencias

Predicción de nanotriángulos tipo Janus basados en dicalcogenuros de metales de transición (MXY) mediante primeros principios y aprendizaje máquina

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de
Maestro en Ciencias

Presenta:

Jair Othoniel Domínguez Godínez

Ensenada, Baja California, México

2025

Tesis defendida por

Jair Othoniel Domínguez Godínez

y aprobada por el siguiente Comité

Dr. Jonathan Guerrero Sánchez

Codirector de tesis

Dr. Luis Miguel Pellegrin Zazueta

Codirector de tesis

Dr. Aldo Humberto Romero

Dr. José Israel Paez Ornelas

Dr. Carlos Alberto Brizuela Rodríguez

Dr. Héctor Noé Fernández Escamilla



Dra. Catalina López Bastidas

Coordinadora del Posgrado en Nanociencias

Dra. Ana Denise Re Araujo

Directora de Estudios de Posgrado

Resumen de la tesis que presenta Jair Othoniel Domínguez Godínez como requisito parcial para la obtención del grado de Maestro en Ciencias en Nanociencias.

Predicción de nanotriángulos tipo Janus basados en dicalcogenuros de metales de transición (MXY) mediante primeros principios y aprendizaje máquina

Resumen aprobado por:

Dr. Jonathan Guerrero Sánchez

Codirector de tesis

Dr. Luis Miguel Pellegrin Zazueta

Codirector de tesis

El interés en los dicalcogenuros de metales de transición (TMD) ha crecido casi de forma exponencial desde el descubrimiento del grafeno, pues, a diferencia de este, la mayoría de TMD son semiconductores y resultan idóneos para nanodispositivos electrónicos, plataformas catalíticas, aplicaciones biomédicas y adsorción de contaminantes. En particular, los nanotriángulos de TMD son valiosos en hidrodesulfurización, donde la actividad catalítica se localiza en los vértices. La fase Janus, con fórmula MXY (M = metal de transición; X, Y = calcógenos distintos), introduce asimetría química que puede potenciar dicha actividad; sin embargo, la exploración experimental de su vasto espacio composicional es lenta y costosa. Para acelerar el descubrimiento se usó una estrategia híbrida que integra teoría del funcional de la densidad y aprendizaje de máquina. Primero construimos una base de datos de ejemplos mediante cálculos de energía de formación de superficie para 344 nanotriángulos Janus con bordes α y β , abarcando 29 metales y cuatro calcógenos en tamaños desde $n = 4$ hasta $n = 10$ (n = número de M en el eje externo del nanotriángulo). A partir de descriptores exclusivamente composicionales entrenamos dos modelos: *Gradient Boosting* para el borde α y *AdaBoost* para el borde β . Ambos alcanzan $R^2 \approx 0.80$ – 0.85 y errores de ~ 0.01 eV/átomo, suficientes para discriminar nanotriángulos estables sin recurrir a nuevos cálculos de primeros principios. El cribado exhaustivo de todas las combinaciones posibles predice que más del 90 % son químicamente estables. Validamos la predicción en nanotriángulos $MoSO$ ($n = 4$ y $n = 10$) mediante dinámica molecular a temperatura ambiente, confirmando su estabilidad termodinámica. El análisis de isosuperficies de potencial electrostático revela un momento dipolar fuera del plano, con la cara de oxígeno enriquecida en carga negativa y la de azufre relativamente positiva. Esta redistribución genera sitios activos adicionales en el centro, además de los bordes, lo que sugiere un posible aumento de la eficiencia catalítica para la remoción de azufre en hidrocarburos. En conjunto, el enfoque DFT-ML demuestra ser una vía rápida y confiable para predecir la estabilidad de nanotriángulos Janus, allanando el camino hacia su síntesis y aplicación en remediación ambiental.

Palabras clave: Teoría del funcional de la densidad, dicalcogenuros de metales de transición, materiales Janus, nanotriángulos, energía de formación, minería de datos de materiales, aprendizaje de máquina supervisado

Abstract of the thesis presented by Jair Othoniel Domínguez Godínez as a partial requirement to obtain the Master of Science degree in Nanoscience.

Prediction of Janus Nanotriangles Based on Transition Metal Dichalcogenides (MXY) Using First-Principles and Machine Learning

Abstract approved by:

Dr. Jonathan Guerrero Sánchez

Thesis Co-Director

Dr. Luis Miguel Pellegrin Zazueta

Thesis Co-Director

Interest in transition-metal dichalcogenides (TMDs) has risen almost exponentially since the discovery of graphene because, unlike graphene, most TMDs are semiconductors and are suitable for electronic nanodevices, catalytic platforms, biomedical applications, and pollutant adsorption. In particular, TMD nanotriangles are valuable for hydrodesulfurization, where catalytic activity is confined to the vertices. The Janus phase, with formula MXY (M = transition metal; X, Y = different chalcogens), introduces chemical asymmetry that can enhance this activity; however, experimental exploration of its vast compositional space is slow and costly. To accelerate discovery, we designed a hybrid strategy that integrates density-functional theory (DFT) with machine learning (ML). First, we built a reference database via DFT surface-formation-energy (EFS) calculations for 344 Janus nanotriangles with α and β edges, covering 29 metals and four chalcogens in sizes $n = 4$ –10. Using purely compositional descriptors, we trained two models: Gradient Boosting for the α edge and AdaBoost-SVR for the β edge. Both reach $R^2 \approx 0.80$ –0.85 with errors of ~ 0.01 eV/atom, sufficient to discriminate stable nanotriangles without additional first-principles calculations. Exhaustive screening of every possible combination predicts that more than 90 % are thermodynamically stable. We validated the extrapolation on *MoSO* nanotriangles ($n = 4$ and $n = 10$) using DFT and *ab initio* molecular dynamics at room temperature, confirming structural stability. Electrostatic-potential maps reveal an out-of-plane dipole moment, with the oxygen face enriched in negative charge and the sulfur face relatively positive. This redistribution produces extra active sites in the triangle's center, in addition to its edges, suggesting higher catalytic efficiency for sulfur removal from hydrocarbons. Overall, the combined DFT-ML approach proves to be a fast and reliable route for predicting the stability and reactivity of Janus nanotriangles, paving the way for their synthesis and environmental catalytic applications.

Keywords: Density functional theory, transition metal dichalcogenides, formation energy, machine learning, supervised learning

Dedicatoria

A mi familia: el amor de mi vida, Corina; mi madre, Violeta; mis hermanas, Salma y Selene; y mi sobrino Izzay. Los amo hoy y siempre.

Agradecimientos

Primero, debo agradecer el apoyo de mi familia. A mi esposa, Corina Olvera, gracias por tu apoyo emocional, económico y espiritual. A mi mamá, la señora Violeta Ornelas, por su incondicional respaldo en mi crecimiento tanto humanístico como profesional. A mis hermanas, Salma Violeta y Luz María Selene, por su constante apoyo a pesar de las adversidades. A mi sobrino Izzay, por su cariño y curiosidad inagotable. A todos ellos les debo y les agradezco profundamente su apoyo a lo largo de esta etapa.

Expreso mi sincero agradecimiento al profesor Jonathan Guerrero, por confiar en este proyecto desde el inicio y por compartir su entusiasmo por la ciencia; al profesor Luis Pellegrin, cuya orientación en ciencia de datos fue clave y sumamente enriquecedora; y al profesor José Páez, por su mirada crítica y su colaboración constante, que aportaron solidez en cada etapa de la investigación.

Asimismo, reconozco al profesor Noé Escamilla, quien me introdujo al mundo de las simulaciones computacionales y despertó en mí el interés por esta área interdisciplinaria durante la licenciatura. Las valiosas observaciones de los profesores Aldo Romero y Carlos Brizuela contribuyeron a afinar cada una de las etapas de este trabajo.

Extiendo mi más sincero agradecimiento al Centro de Nanociencias y Nanotecnología, por brindarme el espacio y las herramientas para realizar esta investigación; al Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California por acogerme y favorecer mi desarrollo académico; y a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación. Finalmente agradezco a los proyectos GAPA-UNAM IG101124 y IA100624 por el soporte financiero. Los cálculos fueron hechos en el centro de supercómputo DGTIC-UNAM con los proyectos LANCAD-UNAM-DGTIC-368 y LANCAD-UNAM-DGTIC-422

No sé si la ausencia de alguna de estas personas hubiera cambiado el rumbo de mi camino; lo único que sé es que les agradezco a cada uno de ellos por haber estado presentes en este recorrido que, sin duda, apenas comienza.

Tabla de contenido

	Página
Resumen en español	ii
Resumen en inglés	iii
Dedicatoria	iv
Agradecimientos	v
Lista de figuras	viii
Lista de tablas	xii
 Capítulo 1. Introducción	
1.1. Antecedentes	5
1.1.1. Evidencia experimental de los nanotriángulos	5
1.1.2. Enfoque teórico-computacional	7
1.1.3. El aprendizaje de máquina para la predicción de nanomateriales	10
1.2. Justificación	11
1.3. Hipótesis	12
1.4. Objetivos	12
1.4.1. Objetivo general	12
1.4.2. Objetivos específicos	12
 Capítulo 2. Metodología	
2.1. Simulaciones de primeros principios	14
2.1.1. Teoría del Funcional de la Densidad	16
2.1.2. Formalismo de energía de formación de superficie	19
2.1.3. Parámetros de simulación de DFT	20
2.2. Aprendizaje de máquina	23
2.2.1. Regresión	24
2.2.2. Algoritmos para regresión supervisada	26
2.2.3. Proceso de extracción de conocimiento de los datos	36
 Capítulo 3. Resultados	
3.1. Energías de Formación de Superficie	42
3.2. Extracción y selección de características	44
3.3. Entrenamiento y evaluación	47
3.4. Modelos en producción	48
3.5. Validación de primeros principios del conjunto <i>MoSO</i>	50
3.5.1. Estabilidad química	50
3.5.2. Estabilidad termodinámica	52
3.5.3. Propiedades electrónicas	53

Capítulo 4. Discusión**Capítulo 5. Conclusiones**

Literatura citada	60
Anexos	65

Lista de figuras

Figura

Página

1. Clasificación de los TMDs como prístinos y Janus. De cada tipo es posible obtener diferentes tipos de nanoestructuras: $0D$, $1D$ y $2D$ 4
2. Microscopías de nanotriángulos basados en TMDs: **(a)** STM de cúmulo de MoS_2 antes de exposición a H atómico, mostrando dímeros de azufre (Helveg et al. (2000)); **(b)** STM de gran área de cúmulos de MoS_2 sobre Au(111) con tamaños controlados (Lauritsen et al. (2007)); **(c)** STM de nanotriángulo pequeño de MoS_2 tras tratamiento con H atómico, con defectos en bordes y esquinas (Tuxen et al. (2010)); **(d)** AFM de monocapa triangular Janus SMoSe sobre SiO_2/Si , confirmando su grosor de 1 nm (Zhang et al. (2017)). 7
3. Modelos geométricos considerados en este trabajo. (a) y (b) muestran los nanotriángulos (de tamaño $n = 4$) tipo Janus en configuraciones α y β , respectivamente, mientras que (c) y (d) corresponden a los nanotriángulos prístinos en configuraciones α y β . Los átomos de metal de transición se representan en color negro, mientras que los calcógenos X y Y aparecen en verde y morado, respectivamente. 9
4. Flujo de trabajo del proyecto, generación de datos, preprocesamiento y creación de los modelos de ML 14
5. Distribución de elementos en los nanotriángulos Janus de fórmula general $MX Y$. Se incluyen los metales de transición de los grupos III–XII y los calcógenos del grupo XVI. En color verde se resaltan los elementos empleados en la generación de la base de datos, y en naranja aquellos no utilizados. 21
6. Planteamiento general del problema de regresión se busca ajustar una función $f(X)$ a la distribución de datos que minimice el error entre la predicción y el valor real. *Imagen elaborada por el autor.* 24
7. Los datos pueden separarse utilizando una línea recta $w x - b = 0$, con un margen definido por la distancia a dos rectas paralelas que sirven como referencia para la función de decisión. También se consideran casos atípicos mediante dos vectores de soporte que apuntan a los datos de la clase 1 (puntos en color azul) y clase 2 (puntos en color naranja). La función a optimizar es la distancia entre las dos rectas $w x - b = 1$ y $w x - b = -1$, la cual puede calcularse como $\frac{2}{\|w\|}$ o, de forma equivalente, minimizar $\frac{1}{2}\|w\|^2$. Imagen adaptada del libro Bishop (2006). 28
8. Anatomía de una red neuronal multicapa (perceptrón), figura adaptada del libro Bishop (2006) 34
9. Esquema del proceso de extracción de características 41
10. Evolución estructural de los nanotriángulos Janus de fórmula general $MX Y$ para los sistemas MoSeO (arriba) y MoSTe (abajo). En cada bloque, las subfiguras (a) y (c) muestran el modelo α —con un dímero en el borde— para tamaños $n = 4$ a $n = 10$, mientras que las subfiguras (b) y (d) corresponden al modelo β —con dos dímeros— en los mismos valores de n . Los átomos de Mo se representan en violeta, los de O en rojo, S en amarillo, Se en verde y los de Te en marrón. 43

11. Distribución (Variable objetivo EFS_{α} y EFS_{β}) de las energías de formación de superficie (EFS) para los modelos de borde α (izquierda) y β (derecha). En cada panel se muestra el histograma de los valores de EFS junto con su estimación de densidad de probabilidad (línea continua), y un recuadro circular que indica el porcentaje de nanotriángulos estables ($EFS < 0$, verde) y no estables ($EFS > 0$, rojo). Estos son los valores de EFS que se utilizaron en la base de datos de ejemplo. 44
12. Importancia de las características determinada por *Lasso* para predecir la EFS de nanotriángulos Janus de TMD. El panel superior corresponde al modelo α y el inferior al β . Las barras representan los coeficientes normalizados de *Lasso*, ordenados de mayor a menor. Entre los descriptores más influyentes destacan el rango de electronegatividad, el rango de radios covalentes y atributos de valencia, subrayando el papel de la diversidad química en la estabilidad de la EFS . Para una descripción completa del significado de cada variable véase Tabla 6 y Tabla 7. 46
13. Diagramas de dispersión entre los valores reales de la energía de formación de superficie (EFS) y las predicciones de los modelos de aprendizaje automático para nanotriángulos tipo Janus. El panel (a) muestra el desempeño de *Gradient Boosting* para el borde α ; el panel (b), el del ensamble *AdaBoost* con SVR como estimador base para el borde β . En cada panel, los triángulos representan las predicciones sobre el conjunto de entrenamiento y los círculos las de la validación cruzada ($k = 5$); la línea discontinua negra indica la tendencia ideal $y = x$. En los recuadros internos se reportan las medias y desviaciones estándar de R^2 , $RMSE$ y MAE , tanto para entrenamiento como para validación cruzada. 49
14. Desempeño comparado de los modelos de borde α (barras rojas) y β (barras azules) para la predicción de la energía de formación de superficie. (a) Coeficiente de determinación R^2 : el modelo *Bagging* (NuSVR) alcanza el mayor R^2 en α , mientras que *AdaBoost* (NuSVR) lo maximiza en β . (b) Error cuadrático medio (RMSE): *Bagging* (NuSVR) consigue el menor RMSE en α , y *AdaBoost* (NuSVR) en β 49
15. Distribución de valores predichos de EFS por los modelos de ML para los modelos de borde α en color rojo y β en color azul, se puede observar que de acuerdo a los modelos de ML la mayoría de valores de EFS son negativos lo que indica que la estabilidad química es favorable de acuerdo a esta cantidad. 50
16. Nanoestructuras optimizadas mediante DFT de nanotriángulos Mo–S–O en geometrías de borde α y β . (a) Modelo α : composiciones $Mo_{10}S_{18}O_{18}$ ($n = 4$) y $Mo_{55}S_{75}O_{75}$ ($n = 10$). (b) Modelo β : composiciones $Mo_{10}S_{15}O_{15}$ ($n = 4$) y $Mo_{55}S_{66}O_{66}$ ($n = 10$). Las esferas moradas representan átomos de molibdeno, las amarillas de azufre y las rojas de oxígeno. 51
17. Resultados de las simulaciones AIMD en ensamble NVT a 300 K para nanotriángulos Janus MoSO en el modelo de borde α . Se presentan dos tamaños distintos: $n = 4$ (panel superior) y $n = 10$ (panel inferior). En cada panel, la curva roja muestra la evolución de la temperatura (eje izquierdo) y la curva azul la evolución de la energía total (eje derecho) durante 5 ps de simulación. A la derecha e izquierda de cada gráfico se muestran las configuraciones atómicas en el paso inicial y en el paso final respectivamente, evidenciando que, a pesar de las fluctuaciones, la estructura mantiene su forma triangular y converge hacia valores medios estables de temperatura y energía. 53

18. ISPE (isovalor = 20) para nanotriángulos de *MoSO*: modelo α de tamaño (a) $n = 4$ y (b) $n = 10$; modelo β de tamaño (c) $n = 4$ y (d) $n = 10$. En la escala de colores, el rojo corresponde a valores de potencial negativos y el azul a valores de potencial positivos. 55
22. Coeficientes de correlación de Pearson entre las características seleccionadas y la *EFS* para los modelos de borde α (escala de rojos) y β (escala de azules). En el modelo α la mayor correlación positiva corresponde al mínimo peso atómico ($r = 0.48$) y la mayor correlación negativa al rango del radio covalente ($r = -0.76$). En el panel b la característica con correlación más alta es la moda de la banda prohibida ($r = 0.69$), mientras que la variable «Clase» (óxido vs. no óxido) presenta la asociación negativa más fuerte ($r = -0.69$). 66
23. Comparación de los tiempos promedio de simulación DFT frente al coste computacional de los modelos de ML para los nanotriángulos Janus. Cada marcador (círculo rojo para α , cuadrado azul para β) indica el tiempo medio de un cálculo DFT en función de n , el número de átomos de metales de transición en el borde exterior del nanotriángulo. Las líneas rojas y azules continuas (escala logarítmica, eje izquierdo) muestran cómo crece el tiempo de DFT al aumentar n . Las líneas punteadas horizontales en rojo (α) y azul marino (β), trazadas en escala lineal (eje derecho), representan el tiempo total de entrenar y evaluar cada modelo de ML (validación cruzada 5-fold) para α y β , respectivamente, evidenciando la drástica reducción de tiempo al usar ML en lugar de DFT 68
24. Resultados de las simulaciones AIMD en ensamble NVT a 300 K para nanotriángulos Janus *MoSO* en el modelo de borde β . Se muestran dos tamaños distintos: $n = 4$ (panel superior) y $n = 10$ (panel inferior). En cada panel, la curva roja indica la evolución de la temperatura (eje izquierdo) y la curva azul la evolución de la energía total (eje derecho) durante 5 ps de simulación. A la derecha e izquierda de cada gráfico aparecen las configuraciones atómicas inicial y final respectivamente, demostrando que, pese a las oscilaciones térmicas y energéticas, la forma triangular se conserva y las magnitudes oscilan alrededor de valores medios estables. 69
25. Isosuperficies de potencial electrostático para nanotriángulos prístinos de *MoS₂* en dos modelos de borde: α (paneles a y b) y β (paneles c y d), para tamaños $n = 4$ y $n = 10$. El código de colores indica potencial negativo (exceso de electrones, en rojo) y potencial positivo (déficit de electrones, en azul). En ambos tamaños, el modelo α muestra una clara localización de potencial positivo en los bordes, lo que indica que estos sitios son los más activos y, por tanto, los más susceptibles a procesos como arranque de átomos o catálisis, mientras que el centro (plano basal) exhibe una mayor acumulación de carga negativa (potencial negativo). Para la geometría β , se observa una distribución más uniforme del potencial, con una mayor presencia de exceso de electrones en toda la superficie y menos diferenciación en los bordes, sugiriendo una menor actividad relativa en comparación con el modelo α . El aumento de tamaño del nanotriángulo refuerza estos patrones, intensificando la diferenciación entre bordes y superficies basales en ambos modelos. 70

19. Distribución de los valores absolutos de las diferencias en la energía de formación de superficie (ΔEFS) entre pares de predicciones para los modelos de borde α (panel superior, color rojo) y β (panel inferior, color azul). Ambas distribuciones se presentan en escala logarítmica; las franjas sombreadas muestran $\pm\sigma$ (desviación estándar) y las líneas verticales discontinuas señalan la media μ . La mayoría de las diferencias se concentra en el rango de 10^{-1} a 10^0 eV/átomo. Los valores mínimos de diferencia son $\Delta EFS_{\min}^{\alpha} = 3.80 \times 10^{-5}$ eV/átomo y $\Delta EFS_{\min}^{\beta} = 4.00 \times 10^{-6}$ eV/átomo. 71
20. Relación entre el número de átomos de metal de transición en el borde (n) y la EFS para los modelos α (izquierda) y β (derecha). Los puntos están coloreados según la clase: oxidados (azul) y no oxidados (naranja). Al aumentar n , las EFS tienden a valores más negativos (mayor estabilidad), especialmente en las estructuras oxidadas. 72
21. Comparativa de las métricas de validación cruzada (R^2 , RMSE y MAE) promedio de 5 pliegues para distintos umbrales de selección con Lasso en los modelos α y β . Los umbrales óptimos se destacan con anotaciones y líneas verticales. 73

Lista de tablas

Tabla	Página
1. Desempeño de los modelos α y β en regresión lineal simple con validación cruzada (5 pliegues), aplicando distintos métodos de selección de características. Se reportan los valores promedio de R^2 , RMSE y MAE sobre el conjunto de prueba; en cada caso se mantuvieron sólo las características de importancia positiva. El método Lasso (sin escalado) logra las mejores métricas en ambos modelos.	45
2. Desempeño de los modelos de borde α y β en la predicción de la <i>EFs</i> , empleando los hiperparámetros óptimos hallados para cada algoritmo (véase la Tabla 8). Se muestra la media $\bar{m}$ y la desviación estándar σ_m de las métricas R^2 , <i>RMSE</i> y <i>MAE</i> ($\bar{m} \pm \sigma_m$), calculadas mediante validación cruzada de 5 pliegues sobre el conjunto de evaluación.	48
3. Comparación de las energías de formación superficie predichas por ML y calculadas mediante DFT, con la diferencia absoluta.	51
4. Nanotriángulos Janus simulados que formaron la BD de ejemplo en total se obtuvieron 172 nanotriángulos del modelo α y 172 para el modelo β . Para cada fórmula general <i>MX₂Y</i> se indica el rango de <i>n</i>	65
8. Hiperparámetros optimizados para cada algoritmo de ML de los modelos de borde α y β de acuerdo a la nomenclatura usada en la librería <i>scikit-learn</i> desarrollada por Pedregosa et al. (2011)	67
5. Nanotriángulos con energías de formación de superficie positivas (formación no favorable según los cálculos de DFT), estos datos pertenecen al conjunto de ejemplos para entrenar los algoritmos de ML.	72
6. Descripción de las características utilizadas en el modelo α	74
7. Descripción de las características utilizadas en el modelo β	74

Capítulo 1. Introducción

La utilidad de la ciencia, además de responder preguntas fundamentales sobre la naturaleza, debe contribuir a la solución de problemas, sean estos provocados o no por el ser humano. Es decir, la ciencia debe ser una herramienta que permita a la humanidad avanzar y mejorar la calidad de vida. En este sentido, uno de los objetivos de esta investigación es la búsqueda y predicción de nuevos materiales para afrontar desafíos como la generación de fuentes de energía más amigables con el medio ambiente, el desarrollo de materiales con propiedades únicas que ayuden a reducir la contaminación en el agua, el aire o el suelo y, en general, identificar materiales útiles para la vida humana. Las nanociencias emergen como un campo de estudio integral que combina la física, la química, la biología y las matemáticas para entender la naturaleza a escalas que nuestros sentidos no pueden percibir. Desde esta perspectiva se desarrolla la presente investigación, integrando las nanociencias con la computación para descubrir y analizar nuevos nanomateriales. Un nanomaterial es una porción de materia que puede manipularse a escala atómica. Recordemos que un material es aquello que sirve como bloque de construcción para crear utensilios, productos o artefactos utilizados para satisfacer alguna necesidad (Aguilar, 2002). Cuando manipulamos la materia a la escala nanométrica, es decir, entre 1 y 100 nanómetros (un nanómetro equivale a la milmillonésima parte de un metro), emergen propiedades que no se presentan en la escala macroscópica. Así, utilizando los átomos (los elementos de la tabla periódica) como piezas básicas, es posible formar diferentes arreglos y estructuras. Dependiendo de la composición de los elementos seleccionados, la disposición en el espacio y el tamaño de las estructuras, se obtendrán propiedades específicas. Podemos imaginar estos átomos como piezas de Lego con las cuales construir variadas estructuras, aprovechando así una propiedad fascinante de la materia para desarrollar nanomateriales (Schaefer, 2010). Existen distintos tipos de nanomateriales, que se clasifican según sus propiedades geométricas y de tamaño o, de manera más específica, según la periodicidad que presentan a escala nanométrica:

- **Materiales cero-dimensionales (0D):** También conocidos como puntos cuánticos, nanopartículas o nanocúmulos, son estructuras con forma definida que no presentan periodicidad en ninguna dirección y cuyas propiedades varían en función de su tamaño.
- **Materiales unidimensionales (1D):** Poseen periodicidad en una sola dirección. Ejemplos de estos son los nanohilos o nanotubos, cuya estructura alargada les confiere propiedades especiales.
- **Materiales bidimensionales (2D):** Son estructuras con periodicidad en dos direcciones, como las monocapas. La disposición de sus átomos en un plano les otorga características físicas singulares.

Esta clasificación en función de la dimensionalidad tiene un impacto directo en las propiedades de los

nanomateriales. En resumen, la diferencia en la dimensionalidad se traduce en variaciones sustanciales en sus propiedades fisicoquímicas. Dichas diferencias se evidencian en su comportamiento electrónico, como la movilidad de los portadores de carga y la presencia o ausencia de bandas de energía prohibidas, es decir, la energía necesaria para que un electrón en la banda de valencia (el electrón más alejado del núcleo) salte a la banda de conducción. También se manifiestan en sus características mecánicas (resistencia, flexibilidad y anisotropía), así como en su reactividad química (capacidad de funcionalización y reactividad en zonas específicas). Cada uno de estos materiales ofrece ventajas particulares que los hacen idóneos para aplicaciones en campos tan diversos como la electrónica, los compuestos avanzados, la energía y la biomedicina (Wang et al., 2018). Los **dicalcogenuros de metales de transición (TMDs**, por sus siglas en inglés), han atraído atención creciente debido a su estructura en capas, sus propiedades electrónicas ajustables y su aplicabilidad en dispositivos optoelectrónicos. Presentan propiedades altamente dependientes de su forma, tamaño y composición química (Wilson & Yoffe, 1969). Su fórmula general es MX_2 (en fase prístina), donde M representa un metal de transición (perteneciente a los grupos 3 al 12 de la tabla periódica, con excepción de los elementos del periodo 7) y X corresponde a un calcógeno (elemento del grupo 16, de los periodos 2 al 5). Estos nanomateriales han captado una atención creciente por parte de la comunidad científica debido a su amplia gama de aplicaciones en nanotecnología. A continuación, se enumeran algunas de las más relevantes:

- **Nanoelectrónica:** por ejemplo, en el trabajo de Tong et al. (2015) se usaron monocapas de disulfuro de molibdeno (MoS_2), demostrando que pueden emplearse en dispositivos electrónicos como *transistores de efecto de campo* (FET por sus siglas en inglés), permitiendo escalabilidad a dimensiones nanométricas con excelente control sobre el canal de conducción. Incluso tienen aplicaciones en paneles solares, mostrando un buen rendimiento como dispositivo fotovoltaico de acuerdo al trabajo de Ramos et al. (2019).
- **Catálisis:** como en la reducción de moléculas contaminantes, tales como el CO_2 , reportado en la investigación hecha por Zhang et al. (2020), o en procesos de *hidrodesulfurización* (HDS). La HDS es un proceso químico catalítico que se utiliza principalmente en la refinación de petróleo para eliminar compuestos de azufre de los combustibles. En este proceso, se hace reaccionar el petróleo con hidrógeno en presencia de un catalizador (usualmente temperatura y presión altas), lo que permite transformar los compuestos sulfurados en sulfuro de hidrógeno (H_2S) y otros productos menos nocivos. Esto es fundamental para reducir las emisiones contaminantes (como el dióxido de azufre) y evitar la corrosión en los equipos de combustión. Los cuales permiten obtener combustibles más limpios mediante la reducción del contenido de azufre en hidrocarburos

(Lauritsen et al., 2007).

- **Producción de hidrógeno:** los TMDs pueden ser usados en reacciones de evolución de hidrógeno (HER, por sus siglas en inglés), como se muestra en el trabajo de An et al. (2019), donde se explora su potencial uso como fuente de energía limpia.
- **Optoelectrónica:** los TMDs se podrían utilizar en dispositivos optoelectrónicos para la detección y emisión de luz (Wang et al., 2020).
- **Aplicaciones biomédicas:** como la liberación controlada de fármacos o agentes terapéuticos (Wang et al., 2018).

Los TMDs pueden sintetizarse en diversas nanoestructuras, incluyendo fases en bulto (Wilson & Yoffe, 1969), materiales bidimensionales (Chhowalla et al., 2013), nanotubos (Aftab et al., 2023) y nanopartículas con geometrías específicas, como las de forma triangular reportadas experimentalmente en el artículo de Helveg et al. (2000). Véase la Figura 1 para ejemplos de cada nanoestructura de TMDs prístinos. Uno de los primeros TMDs estudiados fue el **disulfuro de molibdeno** (MoS_2), que se encuentra naturalmente en el mineral *molibdenita* reportado en el trabajo de Dickinson & Pauling (1923), en el cual se estudiaron por primera vez sus propiedades estructurales. El interés moderno en esta familia de materiales fue impulsado a partir de los estudios pioneros de Wilson & Yoffe (1969), quienes reportaron la síntesis y caracterización de múltiples TMDs en su fase en bulto, destacando sus propiedades semiconductoras y la presencia de estructuras cristalinas hexagonales o trigonales. No obstante, fue tras el descubrimiento del grafeno por Novoselov et al. (2004) que se revitalizó significativamente el estudio de materiales bidimensionales, lo que dio lugar a una auténtica explosión en el desarrollo e investigación de nuevas fases 2D, incluyendo los TMDs, debido a su gran versatilidad y funcionalidad a nanoescala. El MoS_2 , que puede ser sintetizado mediante exfoliación física a partir de la molibdenita (Chhowalla et al., 2013) o por métodos químicos (Li & Zhu, 2015) como la deposición en fase vapor (CVD por sus siglas en inglés). También se han logrado sintetizar otros TMDs prístinos (Wang et al., 2018) como el disulfuro de tungsteno (WS_2), el disulfuro de titanio (TiS_2) y el diseleniuro de niobio (NbSe_2). Una pregunta interesante es: ¿qué ocurre si en estos nanomateriales sustituimos uno de los calcógenos? Es decir, si reemplazamos completamente una capa de átomos X por una capa diferente de átomos Y , se obtiene lo que se conoce como un **material Janus**, en referencia al dios romano de las dos caras, Janus. En este caso, la fórmula química del TMD cambia a MXY , donde M es un metal de transición, X un calcógeno y Y otro calcógeno diferente de X . Al igual que los TMDs prístinos, estos *Janus TMDs* pueden ser sintetizados en diversas nanoestructuras de acuerdo a lo reportado en Zhang et al. (2017)

(véase la Figura 1 para ejemplos de posibles nanoestructuras TMDs). Hasta ahora, ha sido posible la síntesis experimental de materiales Janus en 2D, como el $MoSSe$, mediante técnicas como la CVD (Zhang & Sun, 2019).

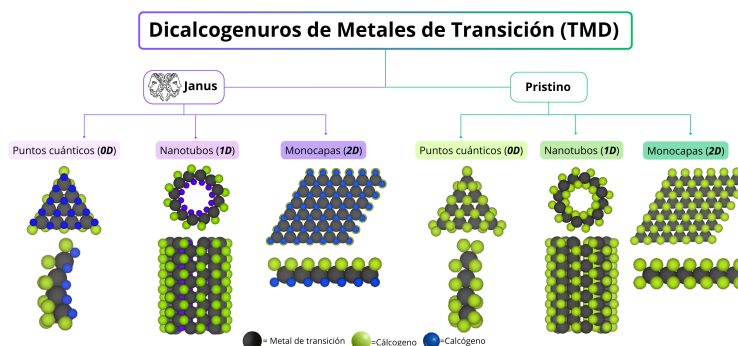


Figura 1. Clasificación de los TMDs como prístinos y Janus. De cada tipo es posible obtener diferentes tipos de nanoestructuras: 0D, 1D y 2D.

A pesar de sus múltiples aplicaciones, aún existen muchas combinaciones de TMDs Janus que no han sido sintetizadas experimentalmente, sobre todo en nanoestructuras 0D, como puntos cuánticos (PC), nanopartículas o nanocúmulos. La exploración de nuevos nanomateriales mediante métodos experimentales conlleva altos costos en recursos humanos y operativos, además de estar limitada por la escalabilidad de los procedimientos, ya que investigar un solo material puede requerir un tiempo considerable. Por esta razón, las simulaciones computacionales ofrecen una alternativa rápida y de alto rendimiento para estudiar una amplia gama de nanomateriales. Estas simulaciones replican fenómenos naturales mediante cálculos computacionales, con el fin de predecir, comprender y describir sistemas físicos a nivel atómico y molecular. Para obtener una descripción fundamental de la materia a escala atómica, es necesario utilizar simulaciones de primeros principios (o *ab initio*), las cuales se basan en la teoría cuántica e incorporan cálculos de la estructura electrónica de moléculas y nanoestructuras. Uno de los métodos más populares en la ciencia computacional de materiales es la teoría funcional de la densidad (DFT, por sus siglas en inglés), que permite predecir propiedades electrónicas —como la estructura de bandas—, así como propiedades estructurales, mecánicas, ópticas e incluso de reactividad (Ohno et al., 2018). Con el advenimiento de la inteligencia artificial (AI, por sus siglas en inglés), el campo de las simulaciones computacionales de nanomateriales no ha quedado al margen en el uso de esta tecnología. Existen numerosas bases de datos (BD) de nanomateriales como Materials Project que almacena propiedades físico-químicas de materiales cristalinos desarrollada en conjunto por Jain et al. (2013). Otras BD importantes en el área de materiales bidimensionales son C2DB que almacena cálculos de propiedades electrónicas y estructurales (Hastrup et al., 2018), JARVIS que es un repositorio automatizado que integra BD de nanomateriales y subrutinas para simulación (Choudhary et al., 2020) o Alexandria que

almacena BD de materiales $2D$ con una diversidad de geometrías y composiciones (Wang et al., 2023). Las BD pueden aprovecharse para predecir nuevos nanomateriales usando AI (Wei et al., 2019). Además, el alto rendimiento y la escalabilidad de las simulaciones permiten generar grandes volúmenes de datos en un periodo de tiempo relativamente corto. De este modo, al combinar los datos obtenidos mediante simulaciones computacionales con técnicas de AI, es posible predecir nuevas estructuras con potenciales aplicaciones en nanotecnología acelerando la identificación de nuevos nanomateriales y sus propiedades. En consecuencia, se decidió adoptar un enfoque de informática de materiales asistido por ML, en el que se emplean algoritmos predictivos entrenados sobre bases de datos de propiedades electrónicas calculadas mediante DFT. La informática de materiales es un campo de investigación interdisciplinario que combina simulaciones computacionales de física y química, aprovechando las BD existentes o generando nuevas a partir de simulaciones, como es el caso de los nanotriángulos tipo Janus basados en TMDs. Esta metodología integra herramientas informáticas para procesar, almacenar y administrar dichos datos, aprovechando técnicas de minería de datos y aprendizaje de máquina (ML, por sus siglas en inglés) para extraer conocimiento de las BD de nanomateriales y predecir nuevas propiedades (Rajan, 2005). En este trabajo se ha optado por estudiar, mediante simulaciones de DFT, un conjunto de nanoestructuras $0D$ con geometría triangular, es decir, nanotriángulos tipo Janus basados en TMDs, con el objetivo de generar una base de datos (BD) que permita utilizar algoritmos de ML para predecir propiedades, tales como la energía de formación, propiedades que aún no han sido exploradas mediante DFT y tampoco han sido sintetizados experimentalmente a excepción de algunas fases prístinas por el grupo de investigadores Helveg et al. (2000) y un estudio basado en primeros principios hecho por Paez-Ornelas et al. (2021).

1.1. Antecedentes

1.1.1. Evidencia experimental de los nanotriángulos

El primer reporte de la síntesis PC basados en TMDs fue realizado por Helveg et al. (2000), donde se logró el crecimiento epitaxial molecular de MoS_2 sobre un sustrato de oro. Mediante microscopía de efecto túnel (STM, por sus siglas en inglés Scanning Tunneling Microscopy), se caracterizó la morfología y la estructura de los ejes de los nanotriángulos de MoS_2 , demostrando que estos ejes exhiben sitios activos para posibles reacciones de HDS, véase la Figura 2a donde se muestra la imagen STM. Otro estudio relevante, llevado a cabo por Lauritsen et al. (2007), investigó la estructura atómica y electrónica de nanotriángulos de MoS_2 y el cambio de propiedades en función de su tamaño. Utilizando imágenes

STM (Figura 2b) y simulaciones basadas en DFT, se mostró que los nanotriángulos más grandes (con $n \geq 6$, donde n es el número de átomos de Mo por lado) presentan bordes de Mo cubiertos con dímeros de S_2 , lo cual les confiere propiedades metálicas. En contraste, los nanotriángulos más pequeños ($n \leq 5$) exhiben bordes terminados en S . Se identificaron ciertos tamaños como estructuralmente estables, especialmente aquellos con n par, debido a efectos electrónicos y a la reconstrucción de los bordes; además, se concluyó que la actividad catalítica, como en la HDS, está fuertemente ligada a estas propiedades de borde y tamaño. El trabajo de Tuxen et al. (2010) explora cómo el tamaño y la estructura de los nanotriángulos de MoS_2 afectan su capacidad para adsorber moléculas de dibenzotiofeno (DBT), un contaminante difícil de eliminar en procesos de HDS. En este estudio se sintetizaron nanotriángulos de MoS_2 sobre una superficie de Au(111) bajo condiciones de vacío, y se analizaron mediante STM (Figura 2) antes y después de exponerlos a hidrógeno atómico (para inducir vacantes de S) y a moléculas de DBT. Los investigadores concluyen que la actividad catalítica y la adsorción selectiva de DBT dependen fuertemente del tamaño del nanotriángulo, siendo las vacantes en las esquinas de nanotriángulos pequeños los sitios activos más eficientes para la desulfuración directa de DBT. Se propone que estos sitios de esquina sin obstrucciones podrían optimizarse en catalizadores industriales, tales como $CoMoS$ o $NiMoS$, para mejorar la eficiencia de la HDS. Estos trabajos aportan al entendimiento de las propiedades estructurales y electrónicas de los nanotriángulos y de cómo estas afectan su capacidad catalítica. Otro trabajo relevante sobre la síntesis de nanotriángulos y sus aplicaciones es el de Lu et al. (2023), donde se reporta la obtención, mediante CVD, de nanotriángulos concéntricos de diseleniuro de tungsteno (WSe_2). Estos nanotriángulos presentan una alta sensibilidad a gases a temperatura ambiente, aunque los sensores basados en TMDs suelen mostrar respuesta baja y recuperación incompleta tras la exposición. En el estudio, el sensor de WSe_2 exhibió una señal elevada frente al dióxido de nitrógeno (NO_2) y la trietilamina (TEA) a temperatura ambiente. Al aplicar iluminación ultravioleta (UV), el rendimiento mejoró: la respuesta aumentó, la recuperación fue completa y el límite de detección alcanzó el rango de partes por billón (ppb). Estos resultados demuestran que diseñar intencionadamente la nanoestructura de los TMDs permite optimizar su desempeño como sensores de gases. Si bien aún no se ha reportado la síntesis de nanotriángulos o PC tipo Janus basados en TMDs, sí se ha logrado la síntesis de monocapas en fases Janus con morfología triangular. En el trabajo de Zhang et al. (2017) se sintetizó la monocapa Janus $MoSSe$, en la cual una capa de azufre (S) sustituye selectivamente a la cara superior de una monocapa de $MoSe_2$, generando una estructura tipo sándwich ($S-Mo-Se$) que rompe la simetría espacial a lo largo del eje z . La síntesis controlada de $MoSSe$ se logra mediante un proceso de sulfurización parcial por CVD, que permite la sustitución únicamente de los átomos de selenio en la superficie expuesta al vapor de azufre, mientras que la cara inferior permanece inalterada.

debido a las barreras de difusión en la interfaz van der Waals con el sustrato. De acuerdo a Zhang et al. (2017) esta configuración asimétrica da lugar a una estructura estable con propiedades optoelectrónicas modificadas, tales como un aumento en la banda prohibida, la aparición de modos Raman característicos y la generación de un campo eléctrico interno. Además de su interés fundamental, las monocapas Janus han demostrado un gran potencial en catálisis. En particular, *MoSSe* presenta una actividad catalítica mejorada para la reacción de HER en el plano basal —normalmente inactivo en TMDs convencionales—, atribuida a tensiones internas, efectos de simetría rota y a la presencia de vacantes atómicas que favorecen la adsorción de hidrógeno. Esta combinación de propiedades convierte a los materiales Janus en plataformas promisorias para futuras aplicaciones en catálisis, energía, sensores y electrónica flexible, véase la Figura 2d para una referencia de la imagen microscopia de fuerza atómica obtenida en dicho trabajo.

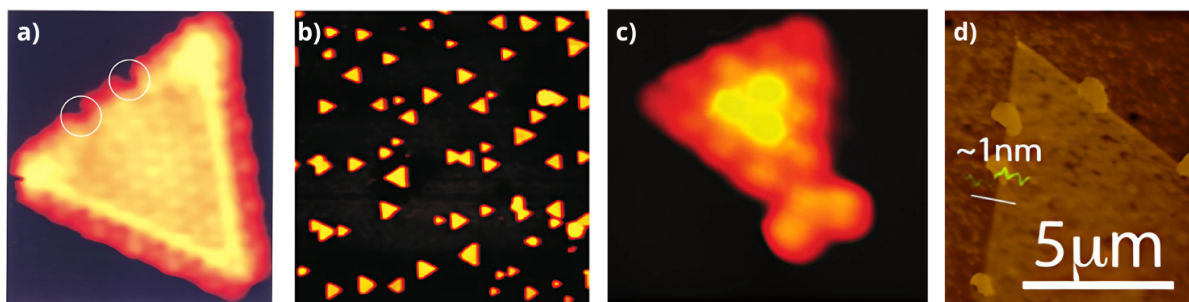


Figura 2. Microscopías de nanotriángulos basados en TMDs: **(a)** STM de cúmulo de *MoS₂* antes de exposición a H atómico, mostrando dímeros de azufre (Helveg et al. (2000)); **(b)** STM de gran área de cúmulos de *MoS₂* sobre Au(111) con tamaños controlados (Lauritsen et al. (2007)); **(c)** STM de nanotriángulo pequeño de *MoS₂* tras tratamiento con H atómico, con defectos en bordes y esquinas (Tuxen et al. (2010)); **(d)** AFM de monocapa triangular Janus SMOSe sobre SiO₂/Si, confirmando su grosor de 1 nm (Zhang et al. (2017)).

1.1.2. Enfoque teórico-computacional

Uno de los paradigmas más utilizados para predecir y comprender las propiedades fundamentales de nuevos nanomateriales es la DFT, desarrollada inicialmente en los trabajos de Hohenberg y Kohn Hohenberg & Kohn (1964) y formalizada posteriormente por Kohn y Sham Sham & Kohn (1965). Este método de modelado computacional, basado en principios de la mecánica cuántica, permite determinar la estructura electrónica de sistemas de muchos cuerpos compuestos por átomos o moléculas en fases condensadas, y ha demostrado su utilidad en física, química y ciencia de materiales. La DFT se ha consolidado como una

de las técnicas más empleadas en fisicoquímica computacional y ciencia computacional de materiales, tal como se evidencia en el análisis presentado en Dumaz et al. (2023). Gracias a ello, es posible identificar teóricamente nanomateriales de alto rendimiento mediante paquetes computacionales que implementan esta teoría, como el *Vienna Ab initio Simulation Package* (VASP) desarrollado por Kresse y Furthmüller G. Kresse (1996). Una variedad de trabajos teóricos basados en la DFT han aportado al entendimiento de las propiedades físico-químicas de los nanotriángulos basados en TMDs, específicamente en los prístinos, como el MoS_2 . El trabajo de Bollinger et al. (2001) delucida cómo las terminaciones en monómeros de S y en dímeros de S_2 influyen en las propiedades electrónicas de los nanotriángulos y en la reactividad de sus bordes. La geometría de los bordes en los nanotriángulos de MoS_2 tiene un impacto decisivo en sus propiedades químicas (véase las Figura 3 para una referencia visual de dichos modelos), ya que determina la aparición de estados electrónicos unidimensionales localizados que influyen directamente en su reactividad. En particular, la presencia de terminaciones específicas, como bordes de azufre o de molibdeno, con o sin adsorción de monómeros o dímeros de azufre, puede inducir comportamientos metálicos o semiconductores, modificando la afinidad por especies como el hidrógeno. Estas variaciones estructurales controlan la formación de enlaces $S-H$, claves en procesos catalíticos como la HDS, y permiten sintonizar la actividad química mediante condiciones externas, como la temperatura y las presiones parciales de gases. Así, la geometría a escala atómica no solo define la estabilidad termodinámica de los bordes, sino también su capacidad catalítica, posicionando a los nanotriángulos de MoS_2 como plataformas ajustables para la catálisis de precisión. Además del MoS_2 , también se han estudiado las propiedades catalíticas de nanotriángulos de WS_2 para la reacción de HER. En la investigación de Lu et al. (2023) se presenta un estudio sobre la mejora en el rendimiento de nanotriángulos de WS_2 y de otros TMDs dopados con cromo, manganeso y cobalto. Los resultados muestran que el dopado con metales en los bordes de estas estructuras mejora significativamente su desempeño catalítico, al aumentar el número de sitios activos y favorecer la transferencia electrónica. El análisis de la estructura electrónica revela que la incorporación de metales dopantes reduce considerablemente la brecha de energía al introducir estados electrónicos de dichos átomos y de los monómeros de azufre cercanos al nivel de Fermi. Esta modificación permite una interacción óptima entre los estados p del azufre y el estado $1s$ del hidrógeno, favoreciendo así una adsorción adecuada para la HER. Además, se observa que tanto la geometría atómica —especialmente la configuración de los bordes— como las propiedades electrónicas dependen estrechamente del tipo de metal dopante, lo que permite diseñar y optimizar estas nanoestructuras como catalizadores altamente eficientes. Una reciente línea de investigación hecha por Paez-Ornelas et al. (2021) en nanomateriales 0D ha centrado su atención en los PC triangulares de $MoSSe$ con estructura *Janus*, los cuales presentan una asimetría fuera del plano que induce una distribución electrónica polar y propiedades catalíticas me-

poradas. Este estudio teórico mediante DFT ha demostrado que la estabilidad estructural y la actividad catalítica de estos PC dependen fuertemente tanto de la **terminación de borde** como del **tamaño del sistema**. En particular, las configuraciones con dímeros y monómeros *S-Se* (véase Figura 3a y Figura 3b para una representación de los modelos estructurales) decorando los bordes resultan ser las más estables, mientras que las distorsiones estructurales inducidas por la curvatura favorecen la aparición de enlaces Mo-S/Se elongados, potencialmente activos en procesos catalíticos. Además, simulaciones STM y análisis de isosuperficies de potencial electrostático han revelado que la cara de azufre presenta una mayor densidad electrónica, lo cual sugiere una **actividad catalítica preferencial del plano basal**, desafiando el paradigma clásico que consideraba únicamente activos a los bordes en los TMDs. Estos hallazgos posicionan a los PC *Janus* como plataformas prometedoras para el diseño de nanocatalizadores altamente eficientes y selectivos.

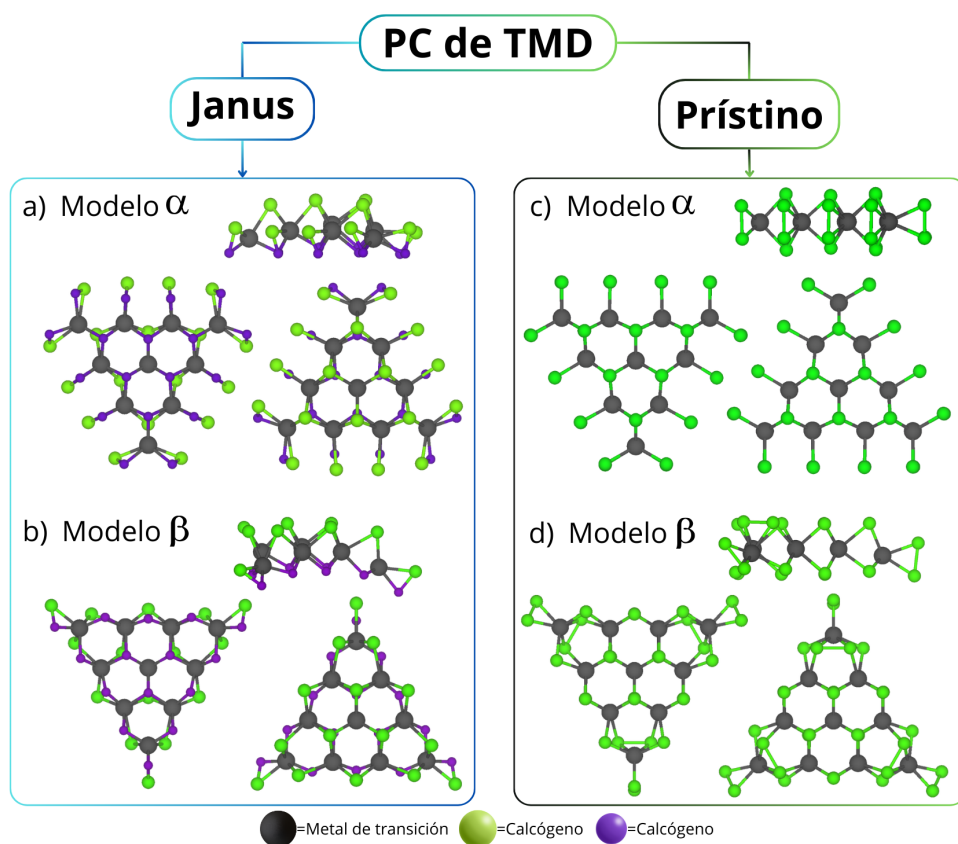


Figura 3. Modelos geométricos considerados en este trabajo. (a) y (b) muestran los nanotriángulos (de tamaño $n = 4$) tipo Janus en configuraciones α y β , respectivamente, mientras que (c) y (d) corresponden a los nanotriángulos prístinos en configuraciones α y β . Los átomos de metal de transición se representan en color negro, mientras que los calcógenos X y Y aparecen en verde y morado, respectivamente.

1.1.3. El aprendizaje de máquina para la predicción de nanomateriales

Aunque los métodos teórico-computacionales permiten una identificación menos costosa y más rápida en comparación con los métodos experimentales, como el CVD, aún se requieren recursos computacionales elevados, por ejemplo, mediante el uso de supercomputadoras. Con el advenimiento de la AI, el campo de las simulaciones en ciencias se ha beneficiado enormemente, ya que la integración de esta tecnología ha incrementado tanto la velocidad como la capacidad para identificar nuevos nanomateriales, reduciendo a su vez el tiempo de cálculo de propiedades físico-químicas. Por ejemplo, de acuerdo con el trabajo de Rajan et al. (2018), se logró reducir el tiempo de cálculo de la banda prohibida en TMDs semiconductores de días a segundos. La predicción de propiedades es una de las áreas más beneficiadas por la AI y el ML. En particular, de acuerdo al análisis de aplicaciones de la AI en nanociencias hecho por Asghari & Naderi (2024), se reporta que en el caso de las nanopartículas (0D), esta herramienta es utilizada con mayor frecuencia. Según estos autores, aunque actualmente el número de estudios en esta área es limitado, se evidencia una tendencia al alza notable. En la actualidad, la contribución de la AI a las aplicaciones de nanomateriales es inferior al 0.5 %, pero se prevé que esta participación crezca a medida que su adopción se generalice en los campos científicos y de investigación. Cabe destacar que países como China, Estados Unidos e India lideran en el uso de esta tecnología para investigar y aplicar nanomateriales. Existen numerosos trabajos relacionados con las nanociencias y la ciencia de materiales en los que se aplica la IA, tanto experimentales como teóricos, que van desde el análisis de imágenes experimentales hasta la predicción de nanoestructuras utilizando redes neuronales de acuerdo a la revisión hecha por Wei et al. (2019). Sin embargo, concentremos la atención en el uso de la AI para estudiar y predecir propiedades de los TMDs. Modelos de ML, incluyendo redes neuronales, análisis de componentes principales y técnicas de agrupamiento, han sido aplicados para analizar datos experimentales —por ejemplo, videos de microscopía de efecto túnel—, permitiendo extraer patrones y filtrar información relevante (Kim et al., 2023), como la contribución exclusiva de las capas activas de los TMDs. En el artículo de Kumar et al. (2022b) a través de simulaciones computacionales de primeros principios y ML se diseñó y analizó la banda prohibida de materiales 2D basados en TMDs y sus óxidos. En este trabajo se usaron características basadas en la composición del metal de transición y del calcógeno para predecir la banda prohibida (directa o indirecta) de TMDs. Se logró identificar que la banda prohibida depende fuertemente de las características del momento magnético, la energía de formación y el volumen atómico de los compuestos. Se construyeron modelos de clasificación de la banda prohibida (0: sin banda prohibida, 1: con banda prohibida), alcanzando un rendimiento del 92 % para identificar TMDs que posean banda prohibida. Asimismo, se emplearon modelos de regresión para predecir dicho valor con

un rendimiento del 72 %, lo que contribuye a la predicción de la banda prohibida en esta familia de nanomateriales 2D. Otro trabajo importante en la predicción de propiedades de materiales 2D basados en TMDs es el realizado por Kumar et al. (2022a), en el que se construyó una BD propia de TMDs 2D con fases geométricas hexagonales y tetragonales, y se desarrollaron modelos de ML capaces de predecir la preferencia de fase de estos nanomateriales a partir de la energía de formación obtenida por DFT. Se simuló un conjunto de monocapas MXY y MX_2 , y se extrajeron los vectores característicos de cada material a partir de atributos elementales de composición, tales como la electronegatividad y el radio iónico, formando en total 30 atributos fisico-químicos. A través del coeficiente de correlación de Pearson, que mide la linealidad entre características, se filtraron aquellas con coeficiente bajo, logrando predecir la energía de formación con un coeficiente de determinación mayor al 90 %, lo que indica una alta capacidad explicativa del modelo basado en características elementales de fácil acceso para predecir nuevas composiciones de TMDs, tanto prístinos como Janus. A pesar de estos importantes avances, aún no existen reportes teóricos o experimentales que analicen las propiedades de los nanotriángulos tipo Janus basados en TMDs.

1.2. Justificación

A pesar de su alto potencial en catálisis, optoelectrónica y almacenamiento de energía, aún no se han sintetizado experimentalmente PC o nanotriángulos tipo Janus MXY basados en TMD, y únicamente existen algunos estudios teóricos aislados —por ejemplo, sobre $MoSSe$ —. Esta carencia de datos experimentales y teóricos limita la exploración sistemática de sus propiedades fisicoquímicas y su posible aplicación nanotecnológica. En este contexto, proponemos generar, mediante cálculos de DFT, una BD de nanotriángulos Janus que permita evaluar su estabilidad química a través de la energía de formación de superficie (E_{FS}). Considerando la fórmula general $M_aX_bY_c$, 29 metales de transición y cuatro calcógenos (oxígeno, azufre, selenio y telurio), basta con un único tipo de terminación de calcógeno y tamaño para formar 174 combinaciones de TMDs tipo Janus; si además variamos la configuración de borde (2 tipos de terminación) y el tamaño de los nanotriángulos (7 tamaños diferentes), el espacio de búsqueda se amplía significativamente hasta 2436 posibles combinaciones que incluyen diferentes composiciones, tamaños y geometrías de borde. El uso de recursos de supercómputo y técnicas de simulación de alto rendimiento nos permitió optimizar y caracterizar estas estructuras de manera sistemática. A partir de los resultados de DFT construimos modelos de ML capaces de predecir con alta precisión la energía de formación de nuevas combinaciones, reduciendo el número de cálculos explícitos necesarios.

De este modo, nuestro enfoque interdisciplinario acelera el descubrimiento y diseño de nanotriángulos Janus estables, contribuyendo al avance de la nanotecnología y la ciencia de materiales.

1.3. Hipótesis

Es posible predecir la EFS de nanotriángulos tipo Janus a partir de propiedades elementales por composición $M_aX_bY_c$. Se espera que los modelos de ML alcancen rendimientos superiores al 80 %, lo que reducirá significativamente el costo de cálculo de dicha propiedad de días a segundos.

1.4. Objetivos

1.4.1. Objetivo general

Crear una BD de nanotriángulos tipo Janus basados en TMDs en sus configuraciones geométricas α y β (véase la Figura 3), con el fin de determinar su estabilidad química mediante el cálculo de la EFS , utilizando DFT para generar la BD de ejemplos y entrenar algoritmos de ML que sean capaces de predecir la EFS que permita hacer una identificación rápida y escalable de nuevos nanotriángulos que no estuvieron presentes en la BD de ejemplo.

1.4.2. Objetivos específicos

- Simular mediante DFT 344 nanotriángulos (BD de ejemplo para los algoritmos de ML) tipo Janus ($M_aX_bY_c$) basados en TMDs, considerando diferentes tamaños, composiciones y bordes. El conjunto incluirá 172 configuraciones tipo α y 172 tipo β (Figura 3).
- Calcular la EFS para cada nanotriángulo usando el formalismo de energía de formación para determinar la estabilidad química.
- Compilar una BD con los resultados obtenidos, utilizando EFS como variable objetivo para cada modelo y extraer los vectores característicos para cada nanotriángulo usando una representación

basada en composición.

- Entrenar modelos de ML para predecir EFS a partir de los descriptores extraídos.
- Evaluar predicciones mediante validación cruzada y seleccionar los modelos más exitosos según las métricas de regresión.
- Llevar los modelos de ML a producción para predecir el resto de composiciones no consideradas en la BD de ejemplo.

Capítulo 2. Metodología

La metodología empleada en esta investigación consta de dos etapas principales: en primer lugar, simulaciones de primeros principios basadas en DFT; y, en segundo lugar, modelado de datos mediante técnicas de ML. En la fase de DFT se generó y optimizó un conjunto de nanotriángulos tipo α y β de diferentes tamaños (número de átomos de metales de transición n en el borde externo), obteniendo para cada estructura la energía del estado base necesaria para calcular su EFS y evaluar así su estabilidad química. A continuación, estos resultados formaron la BD de ejemplo con la que se siguió un flujo de trabajo típico de ML: extracción de descriptores fisicoquímicos basados en la composición atómica, limpieza y preprocesamiento de datos, selección de características relevantes, ajuste de hiperparámetros, entrenamiento y validación cruzada de los modelos, y finalmente generación de predicciones de EFS para nuevas combinaciones de nanotriángulos, este flujo se resume en la Figura 4.

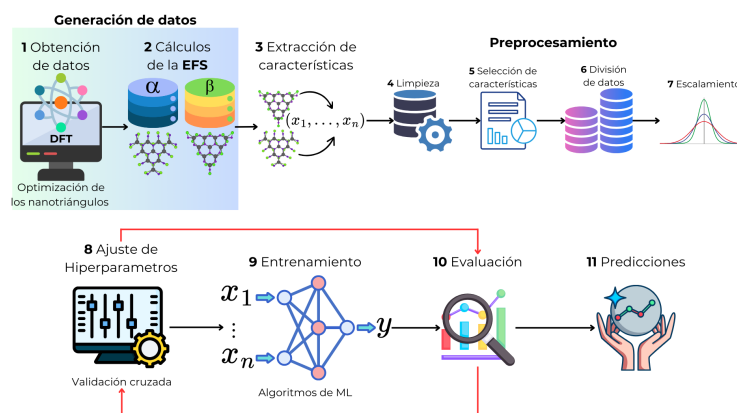


Figura 4. Flujo de trabajo del proyecto, generación de datos, preprocesamiento y creación de los modelos de ML

2.1. Simulaciones de primeros principios

Cuando se habla de simulaciones computacionales en ciencias, se hace referencia a la implementación de modelos físico-químicos en computadoras a través de ecuaciones matemáticas que se resuelven numéricamente. El término *simulación* implica la posibilidad de imitar, modelar o reproducir un sistema físico real, con el fin de explorar su comportamiento, apariencia o propiedades bajo condiciones controladas.

Existen diferentes tipos de simulaciones. En particular, las simulaciones *ab initio* o de *primeros principios* se distinguen porque no requieren parámetros empíricos externos: se fundamentan exclusivamente en

leyes físicas fundamentales bien establecidas, como la mecánica cuántica. Estas simulaciones permiten estudiar sistemas a escala atómica o molecular, y son especialmente relevantes en el contexto de la ciencia de materiales.

La base formal de estas simulaciones se encuentra en la mecánica cuántica de sistemas de muchos cuerpos. El objetivo principal es resolver el problema cuántico asociado a un conjunto de núcleos y electrones interactuantes, es decir, calcular la energía del estado base y otras propiedades electrónicas del sistema. Esta tarea comienza con la ecuación de Schrödinger independiente del tiempo, la cual describe el estado estacionario de un sistema cuántico, para una sola partícula la ecuación se expresa como:

$$\hat{H}\psi(x, y, z) = \epsilon\psi(x, y, z) \quad (1)$$

Donde $\hat{H}$ representa el hamiltoniano del sistema, que contiene toda la información sobre las interacciones entre partículas, ψ es la función de onda del sistema y ϵ es la energía del estado base. Para un sistema compuesto por M núcleos y N electrones, el hamiltoniano completo se expresa como:

$$\hat{H} = - \sum_{i=1}^N \frac{\nabla_i^2}{2} + \sum_{i>j}^N \frac{1}{r_{ij}} - \sum_{i,I}^N \frac{Z_I}{r_{iI}} - \sum_{I=1}^M \frac{\nabla_I^2}{2M_I} + \sum_{I>J}^M \frac{Z_I Z_J}{R_{IJ}}. \quad (2)$$

Aquí se incluyen términos de energía cinética de electrones y núcleos, interacciones de tipo Coulomb electrón-electrón, electrón-núcleo y núcleo-núcleo. No obstante, resolver exactamente la ecuación de Schrödinger para sistemas con más de unos pocos electrones resulta inabordable debido a la complejidad computacional que crece exponencialmente con el número de partículas ($3N + 3M$ grados de libertad). Para superar esta limitación, se introdujo la **aproximación de Born-Oppenheimer**, también conocida como *aproximación de núcleos fijos*, que asume que los núcleos se mueven mucho más lentamente que los electrones y, por tanto, su energía cinética puede ignorarse inicialmente. Esto reduce el problema al movimiento de los electrones en un campo generado por posiciones fijas de los núcleos:

$$\hat{H} = - \sum_i^N \frac{1}{2} \nabla_i^2 + \sum_{i>j}^N \frac{1}{r_{ij}} - \sum_{i,I}^N \frac{Z_I}{r_{iI}} + \text{constante} \quad (3)$$

Uno de los primeros intentos exitosos para aproximar soluciones al problema de muchos cuerpos fue la **teoría de Hartree-Fock (HF)**, desarrollada en los años 1920s y 1930s. En este método, se introduce

una aproximación de *campo medio* donde cada electrón se mueve en el potencial promedio generado por los demás electrones. Además, se incorpora el principio de exclusión de Pauli mediante el uso de determinantes de Slater, lo que garantiza la antisimetría de la función de onda. A pesar de que la teoría de Hartree-Fock permitió avances significativos, también presenta limitaciones importantes: no considera de manera exacta la correlación electrónica (interacciones instantáneas entre electrones), lo cual restringe su precisión para muchos sistemas reales.

2.1.1. Teoría del Funcional de la Densidad

Estas limitaciones motivaron el desarrollo de teorías más eficientes, siendo una de las más influyentes la **DFT** cuyas bases fueron establecidas por los teoremas de Hohenberg & Kohn (1964) y el enfoque práctico propuesto por Sham & Kohn (1965). Cabe destacar que en 1998, el Premio Nobel de Química se otorgó a Walter Kohn y John A. Pople por el desarrollo de la DFT y métodos computacionales en química cuántica, respectivamente. Kohn recibió la mitad del premio por la DFT, una teoría que permite calcular la estructura electrónica de la materia. A diferencia de HF, en DFT la variable fundamental no es la función de onda, sino la densidad electrónica del sistema, lo que reduce dramáticamente la complejidad computacional. DFT permite obtener con buena precisión propiedades fundamentales de sólidos, moléculas y superficies con un costo computacional accesible, y por esta razón ha sido ampliamente adoptada en simulaciones de materiales. Esta teoría toma en cuenta el intercambio entre electrones por lo que introduce el término de correlación e intercambio este es un potencial que tiene que ser fijado mediante diferentes niveles de aproximación conocidos como potenciales de correlación e intercambio, algunas de las aproximaciones más utilizadas son:

- Aproximación de densidad local (LDA, por sus siglas en inglés *Local Density Approximation*).
- Aproximación de gradiente generalizado (GGA, por sus siglas del inglés *Generalized Gradient Approximation*).
- Meta-GGA.
- Funcionales híbridos.
- Funcionales de van der Waals (no locales).
- Corrección de Hubbard (DFT + U).

Cada uno de estos funcionales de intercambio–correlación aporta mejoras en la precisión y en la descripción fisicoquímica del sistema; sin embargo, a medida que se emplean potenciales más sofisticados, también aumenta el costo computacional. Uno de los funcionales más utilizados en DFT es el propuesto por Perdew et al. (1996, 1997) (PBE, por las siglas en inglés de los nombres de los desarrolladores del funcional), perteneciente a la familia GGA, que ofrece un excelente equilibrio entre precisión y costo para sólidos y moléculas. En cuanto a las propiedades estructurales, PBE es una opción “todoterreno”: presenta errores típicamente bajos en estas propiedades, por lo que suele ser el primer funcional que se prueba en estudios de sólidos y catálisis superficial. Para las energías de formación de superficie, PBE tiende a subestimar ligeramente los valores absolutos, pero reproduce con fidelidad las tendencias relativas entre distintas orientaciones y fases cristalinas. Para mejorar la cuantitatividad en sistemas con interacciones débiles, se recomienda complementarlo con correcciones de tipo Van der Waals (por ejemplo, vdW-DF). De esta manera Conociendo el tipo de átomos, el número de partículas y sus posiciones podemos calcular la energía del estado base del sistema. Esto se enfatiza en los teoremas de H-K:

Teorema 2.1 (Primer teorema de Hohenberg–Kohn) *Para un sistema de electrones sometido a un potencial externo $v_{\text{ext}}(\mathbf{r})$, la densidad de estado fundamental $\rho_0(\mathbf{r})$ determina de modo único, salvo una constante aditiva, dicho potencial. En consecuencia, existe una correspondencia biunívoca*

$$v_{\text{ext}}(\mathbf{r}) \longleftrightarrow \rho_0(\mathbf{r}),$$

y toda propiedad observacional en el estado fundamental es un funcional de ρ_0 .

De aquí se deriva la existencia del funcional universal

$$F[\rho] = T[\rho] + U[\rho],$$

independiente de v_{ext} .

Teorema 2.2 (Segundo teorema de Hohenberg–Kohn) *Definido el funcional de energía total*

$$E[\rho] = F[\rho] + \int v_{\text{ext}}(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r},$$

la densidad correcta ρ_0 minimiza $E[\rho]$. Es decir, para cualquier otra densidad ρ' normalizada ($\int \rho' = N$),

$$E[\rho'] \geq E[\rho_0] = E_0.$$

Este principio variacional justifica el método de Kohn–Sham, donde se introducen orbitales que reproducen la misma ρ_0 . Estos resultados pueden ser introducidos en un esquema iterativo conocido como campo autoconsistente para calcular la energía del estado base del sistema, véase el Algoritmo 1 para consultar el pseudoalgoritmo de referencia. La solución es proponer una densidad inicial que podría ser la de los átomos superpuestos, se calculan los potenciales: externo V_{ext} debido a los núcleos, Hartree V_H , el de correlación e intercambio $V_{XC}[\rho]$ y finalmente se calcula el potencial efectivo V_{eff} que es la suma de los tres anteriores. Se resuelve la ecuación de K-S que es una ecuación tipo Schrodinger mediante diagonalización, se calcula una nueva densidad, se hace una mezcla de las densidades para estabilizar la convergencia, se evalúa la convergencia mediante la diferencia de energía total $\Delta E = E(n+1) - E(n)$ y/o la norma de la diferencia de densidad. Finalmente si la diferencia en energía entre el paso actual y el anterior es menor a una cierta tolerancia el algoritmo se detiene; en caso contrario vuelve al paso donde se calcula el potencial efectivo.

Algorithm 1 Ciclo SCF para cálculo de la energía de estado fundamental

- 1: **Entrada:** densidad inicial $\rho^{(0)}(\mathbf{r})$, tolerancias $\epsilon_\rho, \epsilon_E$, máximo de iteraciones $N_{\text{máx}}$
- 2: **for** $n = 0 \dots N_{\text{máx}} - 1$ **do**
- 3: Construir el potencial efectivo

$$V_{\text{eff}}^{(n)} = V_{\text{ext}} + V_H[\rho^{(n)}] + V_{\text{XC}}[\rho^{(n)}]$$

- 4: Resolver las ecuaciones de Kohn–Sham:

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{eff}}^{(n)}\right] \psi_i^{(n)} = \varepsilon_i^{(n)} \psi_i^{(n)}$$

- 5: Calcular la densidad nueva:

$$\rho^{\text{new}}(\mathbf{r}) = \sum_i f_i |\psi_i^{(n)}(\mathbf{r})|^2$$

- 6: Mezclar densidades:

$$\rho^{(n+1)} = \text{Mix}(\rho^{(n)}, \rho^{\text{new}})$$

- 7: Evaluar convergencia:

$$\Delta\rho = \|\rho^{(n+1)} - \rho^{(n)}\|, \quad \Delta E = |E[\rho^{(n+1)}] - E[\rho^{(n)}]|$$

- 8: **if** $\Delta\rho < \epsilon_\rho$ **y** $\Delta E < \epsilon_E$ **then**
 - 9: **salir del bucle**
 - 10: **end if**
 - 11: **end for**
 - 12: **Salida:** densidad auto-consistente ρ_0 , energía E_0
-

De esta manera es posible calcular la energía del estado base E del sistema nanométrico. De acuerdo con el análisis bibliométrico realizado por Dumaz et al. (2023), el uso de la DFT ha crecido de manera exponencial en diversas áreas de la ciencia, como la ciencia de materiales, la nanotecnología, la catálisis, la óptica y el almacenamiento de energía. Su estudio demuestra que la comunidad científica que emplea DFT es amplia y diversa, y que esta metodología se aplica en una gran variedad de contextos interdisciplinarios. Estos resultados destacan la versatilidad y popularidad de DFT como herramienta para predecir propiedades estructurales, electrónicas, ópticas y magnéticas, entre otras.

2.1.2. Formalismo de energía de formación de superficie

La energía libre de Gibbs es una ecuación de estado termodinámica que representa la formación espontánea de una reacción química a temperatura y presión constantes

$$G = E + PV - TS \quad (4)$$

En DFT obtenemos la energía total de estado fundamental E a temperatura cero ($T = 0$) Kelvin, que en este límite equivale a la energía interna U del sistema. Bajo esta aproximación, el cambio de energía libre de Gibbs ΔG puede aproximarse por la diferencia de energías internas ΔE , al despreciar las contribuciones entálpicas y entrópicas:

$$\Delta G \approx \Delta E \quad (T = 0 \text{ K}).$$

La energía de formación E_f cuantifica la energía absorbida o liberada al crear una fase química a partir de sus referencias (átomos aislados, fases prístinas, etc.). Su signo indica la viabilidad de la formación:

- Si $E_f > 0$, hace falta energía externa para formar la fase (no espontáneo).
- Si $E_f = 0$, el sistema está en equilibrio térmico (ni libera ni consume energía).
- Si $E_f < 0$, la formación es termodinámicamente favorable y ocurre de modo espontáneo.

Para superficies, y en nuestro caso para nanotriángulos tipo Janus, definimos la energía de formación de superficie (EFS) extendiendo el concepto anterior con referencias adecuadas. La expresión general es:

$$EFS = \frac{E - E_{\text{ref}} - \sum_i N_i \mu_i}{N} \quad (5)$$

donde:

- E es la energía total del sistema (el nanotriángulo Janus MXY) obtenida por DFT.
- E_{ref} es la energía total de la fase de referencia (por ejemplo, el nanotriángulo en fase prístina MX_2).
- μ_i es el potencial químico de la especie atómica i en su estado estándar o fase de referencia.
- N_i es el número de átomos de la especie i en el sistema.
- N es el número de total de átomos del nanotriángulo.

Debido a que en estos nanotriángulos el número de metales de transición es constante se consideró tomar la EFS sólo en términos de los potenciales atómicos de los calcógenos (oxígeno, azufre, selenio y telurio) y el oxígeno. Permitiendo así disminuir el tiempo de cálculo de las EFS ya que no requiere calcular el μ de los metales de transición.

2.1.3. Parámetros de simulación de DFT

Los cálculos realizados en este trabajo fueron hechos en el paquete computacional **VASP**, uno de los programas más robustos y utilizados en la actualidad para simulaciones de materiales mediante DFT, desarrollado en conjunto por G. Kresse (1996); Kresse & Furthmüller (1996); Grimme et al. (1999). Para simular los nanotriángulos se consideraron dos modelos de terminación de borde, denominados α y β (véase la Figura 3 para observar los modelos atomísticos). La diferencia principal entre ambos radica en la terminación del borde. En el modelo α , los átomos de metal de transición en los bordes del nanotriángulo están enlazados a un dímero $X-Y$, mientras que en el modelo β los metales de transición del borde están enlazados a dos dímeros $X-Y$, esta diferencia en la terminación del borde juega un papel crucial para las propiedades de los nanotriángulos en términos de su estabilidad. La diferencia en la coordinación de borde afecta sustancialmente las propiedades electrónicas de los nanotriángulos, ya que el modelo β presenta una mayor saturación de calcógenos en los bordes, lo cual incrementa la densidad electrónica disponible para interacciones químicas o electrónicas. Además, para cada modelo se consideraron diferentes tamaños de nanotriángulo, definidos por el número de átomos de metal de transición presentes en los bordes externos. Dicho parámetro, representado por n , varió entre $n = 4$ y

$n = 10$. En la Figura 5 se muestra la distribución de elementos tanto de metales de transición como de calcógenos que fueron utilizados para crear la base de datos.



Figura 5. Distribución de elementos en los nanotriángulos Janus de fórmula general MXY . Se incluyen los metales de transición de los grupos III–XII y los calcógenos del grupo XVI. En color verde se resaltan los elementos empleados en la generación de la base de datos, y en naranja aquellos no utilizados.

Se utilizó el funcional de intercambio y correlación del tipo GGA, específicamente en su forma PBE. Esta aproximación mejora el tratamiento del funcional LDA al incluir dependencias del gradiente de la densidad electrónica, y es ampliamente utilizada en simulaciones de materiales por su equilibrio entre precisión y eficiencia computacional. Para describir la interacción entre los núcleos y los electrones de valencia, se emplearon pseudopotenciales del tipo PAW (por sus siglas en inglés *Projector Augmented-Wave*) propuesto en el trabajo de Blöchl (1994); Kresse & Joubert (1999). Esta técnica adaptada en el código *VASP*, permite una descripción precisa de los electrones de valencia al tiempo que elimina la necesidad de describir explícitamente los electrones del núcleo, que no participan directamente en los enlaces químicos. El método PAW reconstruye la función de onda completa a partir de una representación pseudo en regiones no nodales y una proyección de corrección cerca del núcleo, lo que permite una excelente precisión manteniendo tiempos de cómputo razonables. Para la expansión de las funciones de onda en una base de ondas planas, se estableció una energía de corte de 500 eV, valor que garantiza la convergencia adecuada en los cálculos para la mayoría de los elementos incluidos. Dado que las interacciones de Van der Waals (vdW) pueden desempeñar un papel relevante en sistemas 0D y en la estabilización de estructuras a nanoescala, se incluyó una corrección de tipo vdW en los cálculos, con el objetivo de capturar adecuadamente estas fuerzas de dispersión débiles. El criterio de convergencia para las optimizaciones estructurales se fijó en 1×10^{-4} eV, asegurando que la energía total del sistema esté suficientemente estabilizada en cada paso de la simulación. Para evitar interacciones entre las imágenes periódicas de los nanotriángulos —dado que se emplea una celda periódica en las simulaciones— se

introdujo un vacío en las tres direcciones espaciales. Esta separación garantiza que cada nanotriángulo se comporte como un sistema aislado, emulando adecuadamente su naturaleza 0D. El tamaño del vacío se ajustó de forma proporcional al tamaño del nanotriángulo, con valores que oscilaron entre:

- $a = 25 \text{ \AA}$, $b = 23 \text{ \AA}$, $c = 13 \text{ \AA}$ para los sistemas más pequeños ($n = 4$),
- hasta $a = 44 \text{ \AA}$, $b = 40 \text{ \AA}$, $c = 16 \text{ \AA}$ para los sistemas de mayor tamaño ($n = 10$).

Estos parámetros permiten modelar con precisión las propiedades estructurales y electrónicas de los nanotriángulos tipo Janus y prístinos, sin introducir errores asociados a condiciones periódicas artificiales.

Para evaluar la estabilidad termodinámica de los nanotriángulos a temperatura ambiente se decidió llevar a cabo estudios de dinámica molecular de primeros principios o AIMD por sus sigla en inglés. Estas simulaciones se basan en el principio de Born–Oppenheimer, en el cual los electrones se relajan instantáneamente al mínimo de energía electrónica en cada paso, mientras que los núcleos evolucionan según la mecánica clásica.

- **Ensamble termodinámico:** Utilizamos un ensamble NVT (constante Número de partículas, Volumen y Temperatura) controlado mediante un termostato desarrollado por Nosé (1984); Hoover (1985), manteniendo la temperatura en 300 K con una tolerancia de 5 K.
- **Paso de integración:** El paso de tiempo se fijó en 1 fs, un valor que garantiza un equilibrio entre precisión y costo computacional.
- **Duración de la simulación:** Cada nanotriángulo se sometió a 5 ps de evolución dinámica, lo que permitió observar cambios estructurales y fluctuaciones térmicas.
- **Configuración electrónica:** Se empleó el mismo funcional GGA–PBE y pseudopotenciales PAW utilizados en los cálculos estáticos, con una convergencia de energía de 10^{-4} eV en cada paso.
- **Análisis de resultados:** Se monitorearon parámetros como la energía total, la temperatura instantánea. La ausencia de roturas de enlace o reordenamientos mayores durante la simulación confirma la robustez estructural de la mayoría de los nanotriángulos.

Estas simulaciones AIMD proporcionan una validación adicional de la estabilidad química predicha por la EFS y revelan la respuesta dinámica de los bordes y caras Janus ante la agitación térmica. Los cálculos de DFT se realizaron en la supercomputadora de Miztli, es la supercomputadora de quinta generación de

la Universidad Nacional Autónoma de México (UNAM), ubicada en Ciudad Universitaria y operada por la Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC). Su nombre en náhuatl significa “felino mayor” o “puma”. Esta infraestructura de alto rendimiento cuenta con más de 5,000 núcleos de procesamiento distribuidos en 344 servidores, y una capacidad de cómputo de 118 Teraflops. Miztli es utilizada anualmente en más de 100 proyectos científicos, entre ellos simulaciones de materiales, dinámica molecular, astrofísica y estudios sobre sismos, demostrando su relevancia como una herramienta fundamental para el desarrollo de la ciencia en México.

2.2. Aprendizaje de máquina

El ML es un área de la AI que se ocupa de responder a la pregunta: ¿cómo aprenden las máquinas? Esta área ataca problemas dentro de los campos de estudio de la visión por computadora, la robótica, ML y el procesamiento del lenguaje natural (Russell & Norvig, 2010). En esencia, el ML comprende un conjunto de métodos capaces de extraer patrones de los datos de forma automática y utilizar dichos patrones para predecir datos futuros (Murphy, 2012). Según el tipo de datos a utilizar para el aprendizaje, el ML emplea tres grandes enfoques: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo. En este trabajo hemos optado por el aprendizaje supervisado (AS). En el AS, los datos de entrada se organizan en una base de datos estructurada —es decir, relacional— compuesta por filas (observaciones) y columnas (atributos o características). Cada columna representa una dimensión del *espacio de características* (con dimensión D), y la matriz X almacena los vectores característicos correspondientes a cada observación. La variable que se desea predecir se denomina variable objetivo Y . Dependiendo de la naturaleza de Y , distinguimos dos tipos de problemas de aprendizaje supervisado:

- **Clasificación:** Y es una etiqueta discreta (clase). El objetivo es aprender una función que asigne a cada vector X la clase correcta.
- **Regresión:** Y es una variable continua (número decimal). Se busca una función $f(X)$ que ajuste los datos de entrenamiento (véase la Figura 6) y permita predecir valores futuros con un cierto grado de error o incertidumbre.

En este caso, como la propiedad a predecir es la *EFS* de los nanotriángulos con bordes α y β , estamos ante un problema de regresión supervisada.

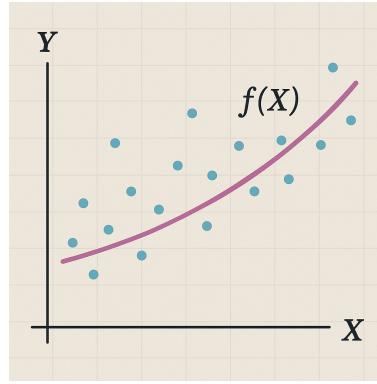


Figura 6. Planteamiento general del problema de regresión se busca ajustar una función $f(X)$ a la distribución de datos que minimice el error entre la predicción y el valor real. *Imagen elaborada por el autor.*

2.2.1. Regresión

La primera aproximación para resolver este problema consiste en suponer que la variable objetivo y puede calcularse como una combinación lineal de las características (variables independientes) x . Esto nos lleva al modelo de regresión lineal, cuya formulación general es la siguiente:

$$\hat{y} = w_0 + \sum_{i=1}^N w_i x_i \quad (6)$$

donde w_i son los pesos o coeficientes que posteriormente se buscarán ajustar, x_i es el vector característico, y el término w_0 se conoce como sesgo o *bias* en inglés. Estos parámetros pueden optimizarse a partir de los datos de entrenamiento con el fin de ajustar el mejor hiperplano que modele la relación entre las variables. Para visualizar este concepto, consideremos un problema tridimensional donde y depende únicamente de dos características x_1 y x_2 . En este caso, el modelo toma la siguiente forma:

$$y = w_0 + w_1 x_1 + w_2 x_2 \quad (7)$$

El ajuste del mejor plano se realiza minimizando una función de error J , definida como la suma de los errores cuadráticos entre las predicciones $\hat{y}_i$ del modelo y los valores verdaderos y_i para cada observación:

$$J = \sum_{i=0}^N (y_i - \hat{y}_i)^2 \quad (8)$$

Sin embargo, un modelo puramente lineal puede ser una simplificación excesiva del problema, especialmente cuando la relación entre las variables no es lineal. Para abordar este caso, se puede generalizar la

regresión utilizando una función no lineal $\phi(\mathbf{X}, \mathbf{W})$ que transforma las características originales mediante una base funcional más expresiva. El modelo se expresa entonces como:

$$\hat{y}(\mathbf{X}, \mathbf{W}) = w_0 + \sum_{i=1}^N \mathbf{w}_i \phi_i(\mathbf{x}_i) \quad (9)$$

Suponiendo que $\phi_0 = 1$, esta expresión puede escribirse de forma compacta como un producto escalar:

$$y(\mathbf{X}, \mathbf{W}) = \mathbf{W}^T \phi(\mathbf{X}) \quad (10)$$

donde $\mathbf{W} = (w_0, w_1, \dots, w_{N-1})^T$ y $\phi(\mathbf{X}) = (\phi_0, \phi_1(\mathbf{X}), \dots, \phi_{N-1}(\mathbf{X}))^T$. Esta formulación permite utilizar diferentes tipos de funciones base para representar la no linealidad, como:

- Polinomios: $\phi_i(x) = x^i$
- Funciones gaussianas: $\phi_i(x) = \exp \left\{ -\frac{(x-\mu_i)^2}{2s^2} \right\}$
- Funciones sigmoides: $\phi_i(x) = \frac{1}{1+e^{-x}}$

No obstante, estas no son las únicas posibles elecciones de funciones base; existen muchas otras dependiendo del problema específico (Bishop, 2006). Una vez formulado el modelo de regresión, el siguiente paso es encontrar los valores óptimos de los parámetros $\mathbf{W}$ que minimicen el error cometido en las predicciones. Esta tarea se plantea como un problema de optimización de una función de costo, comúnmente la suma de los errores cuadráticos (ECM):

$$J(\mathbf{W}) = \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N (y_i - \mathbf{W}^T \phi(\mathbf{X}_i))^2 \quad (11)$$

Existen diversas técnicas para minimizar esta función. A continuación se presentan dos enfoques ampliamente utilizados. La solución cerrada mediante mínimos cuadrados es posible cuando el modelo es lineal y la función de base es simplemente la identidad (es decir, $\phi_i(\mathbf{X}) = X_i$), el error puede minimizarse analíticamente resolviendo un sistema de ecuaciones matriciales. La solución óptima se obtiene derivando la función de costo con respecto a los parámetros $\mathbf{W}$ e igualando a cero:

$$\mathbf{W} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{Y} \quad (12)$$

donde $\Phi \in R^{N \times M}$ es la matriz de diseño con las N muestras y M funciones de base, y $\mathbf{Y} \in R^N$ es el vector columna de las observaciones. Esta solución es computacionalmente eficiente cuando el número

de características no es elevado y no existe multicolinealidad en los datos.

Cuando se utilizan funciones no lineales o cuando el tamaño del conjunto de datos es muy grande, la solución analítica puede volverse impracticable. En estos casos, se recurre a métodos iterativos como el descenso por gradiente, el cual ajusta los parámetros en la dirección opuesta al gradiente de la función de error:

$$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} - \eta \nabla_{\mathbf{W}} J(\mathbf{W}^{(t)}) \quad (13)$$

donde η es la tasa de aprendizaje (*learning rate*), y $\nabla_{\mathbf{W}} J$ representa el gradiente del error con respecto a los pesos. El algoritmo se detiene cuando el cambio en el error es suficientemente pequeño o cuando se alcanza un número máximo de iteraciones. Este método es ampliamente utilizado en ML moderno debido a su simplicidad y escalabilidad, especialmente cuando se entrena sobre grandes volúmenes de datos o redes neuronales.

2.2.2. Algoritmos para regresión supervisada

Existe una gran variedad de algoritmos para regresión supervisada disponibles en `scikit-learn`, desde algoritmos lineales simples hasta variantes que agregan términos de penalización a la función de costo, lo que los hace más robustos que la regresión lineal (RL) simple, cuya función de costo considera únicamente los mínimos cuadrados, de acuerdo con la ecuación 8.

Particularmente, algoritmos como *Lasso* agregan un término a la función de costo de mínimos cuadrados llamado operador de contracción y selección por valor absoluto (L_1), lo que permite anular los pesos w_i de características no informativas (véase la ecuación 14). El algoritmo *Ridge*, en contraste, agrega el operador cuadrático L_2 a la función de costo, lo que permite reducir de forma proporcional los coeficientes de la regresión sin que estos lleguen a cero (véase la ecuación 15). Finalmente, el algoritmo *Elastic Net* combina los dos términos L_1 y L_2 para construir un modelo de RL aún más robusto, mediante una función de costo (ecuación 16) que incorpora ambos términos de regularización, donde η es el número de muestras, χ controla la fuerza de la regularización, y ρ (con valores en el intervalo $[0, 1]$) pondera la contribución relativa de las normas L_1 y L_2 .

$$J_{Lasso}(\mathbf{w}) = \sum_{i=1}^N (y_i - \mathbf{w}^\top \mathbf{x}_i)^2 + \chi \|\mathbf{w}\|_1 \quad (14)$$

$$J_{Ridge}(\mathbf{w}) = \sum_{i=1}^N (y_i - \mathbf{w}^\top \mathbf{x}_i)^2 + \chi \|\mathbf{w}\|_2^2 \quad (15)$$

$$J_{ElasticNet}(\mathbf{w}) = \frac{1}{2\eta} \sum_{i=1}^n (y_i - \mathbf{x}_i^\top \mathbf{w})^2 + \chi \left(\rho \|\mathbf{w}\|_1 + \frac{1-\rho}{2} \|\mathbf{w}\|_2^2 \right) \quad (16)$$

Los hiperparámetros clave son el coeficiente de regularización χ en la ecuación 16 y, en el caso de *Elastic Net*, la proporción $\rho \in [0, 1]$. Sin embargo, los modelos lineales en ML pueden resultar poco exitosos, ya que suponen relaciones estrictamente lineales entre las variables de entrada $\mathbf{x}_i$ y la variable de salida y , lo cual no necesariamente se cumple en la práctica. Para solventar este problema, se pueden emplear algoritmos de ML más complejos, capaces de identificar patrones no lineales en los datos.

Las máquinas de soporte vectorial (*Support Vector Regression*, SVR) extienden la idea del margen máximo de las SVR para clasificación al caso de regresión. El objetivo geométrico es hallar el hiperplano que más se ajuste al espacio de características (es decir, con norma de pesos mínima) de modo que la mayor parte de los puntos de entrenamiento queden contenidos en un ε -tubo alrededor de dicho hiperplano, tolerando pequeñas desviaciones sin penalización. Dados los pares $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ y una transformación no lineal $\phi : R^d \rightarrow \mathcal{H}$ (generalmente de dimensión muy alta o incluso infinita) inducida por un núcleo o kernel $k(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle$, la formulación algebraica se escribe como

$$\min_{\mathbf{w}, b, \xi_i, \xi_i^*} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad \text{sueto a} \quad \begin{cases} y_i - (\mathbf{w}^\top \phi(\mathbf{x}_i) + b) \leq \varepsilon + \xi_i, \\ (\mathbf{w}^\top \phi(\mathbf{x}_i) + b) - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, \end{cases} \quad (17)$$

donde $\mathbf{w}$ y b definen el hiperplano, $C > 0$ controla la penalización por no cumplir el ancho del tubo y ε fija su anchura. Desde un punto de vista geométrico:

- Minimización de $\|\mathbf{w}\|$: asegura que el hiperplano sea lo más plano posible, análogo a maximizar el margen en clasificación.
- Soporte del ε -tubo: sólo los puntos cuyo error excede ε (los *vectores de soporte*) contribuyen al costo. En el espacio de características, cada tal punto “empuja” la pared interna del tubo; el balance entre “presión” de los datos y “tensión” del margen está gobernado por C .
- Dualidad y núcleo: al pasar a la forma dual aparecen los multiplicadores de Lagrange α_i, α_i^* . La

predicción para un punto nuevo $\mathbf{x}$ queda

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(\mathbf{x}_i, \mathbf{x}) + b,$$

y solo un subconjunto de índices con $\alpha_i - \alpha_i^* \neq 0$ (los vectores de soporte) interviene en la suma, proporcionando un modelo *disperso*.

Intuición geométrica. Visualice el ε -tubo como dos hiperplanos paralelos equidistantes a la hiperplana central. A medida que ε aumenta, el tubo se ensancha y engloba más datos sin penalización, sacrificando ajuste fino pero ganando robustez al ruido. Disminuir ε hace que el tubo sea más estrecho: se incrementa la cantidad de vectores de soporte y el modelo puede capturar detalles sutiles (con riesgo de sobreajuste si C también es alto). En dos dimensiones, los datos pueden dividirse mediante una línea recta, este tubo sería aumentar o disminuir la distancia entre las dos rectas y los vectores de soporte como se observa en la Figura 7.

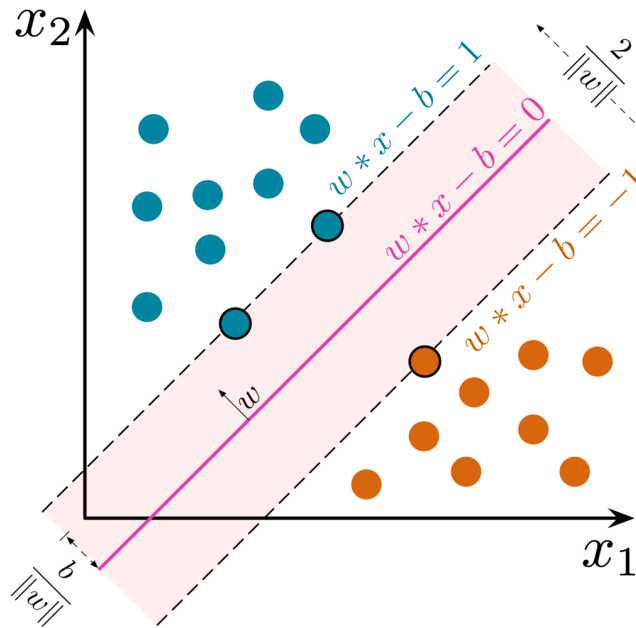


Figura 7. Los datos pueden separarse utilizando una línea recta $w\mathbf{x} - b = 0$, con un margen definido por la distancia a dos rectas paralelas que sirven como referencia para la función de decisión. También se consideran casos atípicos mediante dos vectores de soporte que apuntan a los datos de la clase 1 (puntos en color azul) y clase 2 (puntos en color naranja). La función a optimizar es la distancia entre las dos rectas $w\mathbf{x} - b = 1$ y $w\mathbf{x} - b = -1$, la cual puede calcularse como $\frac{2}{\|w\|}$ o, de forma equivalente, minimizar $\frac{1}{2}\|w\|^2$. Imagen adaptada del libro Bishop (2006).

- **Caso lineal.** Con $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top \mathbf{x}'$ la SVR produce una regresión lineal robusta que tolera observaciones atípicas dentro del margen. Útil cuando la relación entrada-salida es aproximadamente lineal o $d \gg n$ es decir el número de variables de entrada d (características) es mucho mayor que

el número de observaciones (muestras) disponibles n .

- **Caso no lineal.** Emplear núcleos proyecta los datos a espacios donde la relación se vuelve lineal; la curvatura del hiperplano en el espacio original depende de k . Uno de los núcleos más usados son las bases de funciones radiales o RBF por sus siglas en inglés, que suele ser la primera elección por su capacidad de modelar variaciones locales complejas.

Núcleos habituales.

Lineales Separación lineal de los datos, costo computacional bajo; preferible con muchos atributos poco ruidosos.

Polinomial Separación usando polinomios de grado j . Permite curvaturas globales controladas por el grado d ; útil cuando existen interacciones polinómicas.

RBF Introduce vecindades locales de influencia con parámetro γ ; versátil para relaciones no lineales y tamaño de muestras $\sim 10^2$.

Sigmoide Es una función sigmoide, véase la ecuación 22. Similar a una neurona logística, sin embargo, menos estable numéricamente.

Hiperparámetros clave.

- C : penalización de errores; alto C busca un incumplimiento mayor con riesgo de sobreajuste, bajo C tolera más incumplimientos con mayor grado de generalización.
- ε : tolerancia dentro del tubo; gobernar “precisión deseada” frente a ruido experimental.
- γ (sólo en RBF/polinomiales): define la extensión radial de la influencia de un ejemplo; γ grande $\Rightarrow$ cada punto afecta sólo su vecindad inmediata, γ pequeño $\Rightarrow$ influencia amplia.

Esta descripción geométrica y práctica posiciona la SVR como una herramienta flexible para predecir propiedades de nanomateriales, siempre que se ajuste cuidadosamente la tríada (C, ε, γ) y se disponga de un conjunto de validación representativo.

Uno de los algoritmos más utilizados en ML son los modelos basados en árboles de decisión. Un árbol de decisión particiona recursivamente el espacio de características R^d en regiones ortogonales hasta que

cada hoja contenga observaciones “suficientemente homogéneas”. Para regresión se usa típicamente el criterio *CART* (*Classification and Regression Trees* por sus siglas en inglés), que en cada nodo t selecciona la división (j, s) —atributo j y umbral s — que minimiza la **pérdida residual cuadrática**:

$$\arg \min_{j,s} \left[\sum_{i \in \mathcal{D}_{\text{izq}}(j,s)} (y_i - \bar{y}_{\text{izq}})^2 + \sum_{i \in \mathcal{D}_{\text{der}}(j,s)} (y_i - \bar{y}_{\text{der}})^2 \right],$$

donde $\bar{y}_{\text{izq}}$ y $\bar{y}_{\text{der}}$ son las medias de las respuestas en los subconjuntos resultantes. La predicción $\hat{y}(\mathbf{x})$ para un punto nuevo es simplemente la media de y en la hoja que lo contiene. **Ventajas clave.**

- Interpretabilidad: las reglas de división se leen como “si-entonces”.
- Manejo de variables mixtas y valores faltantes mínimos.
- Capturan no linealidades y relaciones de interacción sin preprocesamiento complejo.

Limitaciones. *Gran varianza*: pequeños cambios en los datos pueden alterar fuertemente la estructura del árbol, conduciendo a sobreajuste. *Frentes de decisión ortogonales*: pueden ser ineficientes para límites oblicuos. **Hiperparámetros esenciales.**

- `max_depth`: profundidad máxima; controla sesgo-varianza.
- `min_samples_split` y `min_samples_leaf`: tamaño mínimo para dividir un nodo o dejarlo como hoja.
- `max_features`: número de variables candidatas en cada división (clave en ensambles).

Para reducir la varianza de un único árbol, se recurre a la agregación de múltiples árboles crecidos sobre subconjuntos ligeramente distintos de los datos y/o atributos, promediando luego sus salidas, conocidos como modelos de ensamble. El bosque aleatorio introducido por Breiman (2001), combina dos fuentes de aleatoriedad:

1. **Bootstrap** (*bagging*): cada árbol se entrena con una muestra diferente de tamaño n (número de muestras) seleccionada con reemplazo.
2. **Submuestreo de atributos**: al dividir un nodo, sólo se consideran $\sqrt{d}$, $\log_2 d$ o un número fijo de características d .

Esto fuerza desacoplar los árboles para que sus errores individuales estén menos correlacionados; el promedio reduce la varianza sin aumentar demasiado el sesgo. **Hiperparámetros típicos.**

- `n_estimators`: número de árboles; varianza ↓ al aumentar.
- `max_features`: fracción o número de atributos revisados en cada división.
- `bootstrap`: Verdadero (*bagging* clásico) o Falso (submuestreo sin reemplazo).
- Los mismos parámetros de un árbol base (`max_depth`, `min_samples_leaf`, etc.).

Ventajas prácticas: alta precisión, robusto al ruido y al escalado de variables, estimación gratuita fuera de bolsa del error generalizado y de la importancia de cada variable (Pedregosa et al., 2011). Otro esquema interesante son los árboles extremadamente aleatorios o en inglés conocidos como *Extremely Randomized Trees Regressor* (Bishop, 2006) avanza un paso más:

1. Se evita el *bootstrap*; cada árbol usa todo el conjunto de entrenamiento (o una fracción fija).
2. Para cada nodo, además de submuestrear atributos, el *umbral* se selecciona al azar dentro del rango del atributo, sin optimizar la división mediante la función de costo.

Consecuencias.

- *Varianza* todavía menor que en *Random Forest*, pues los árboles son aún más independientes.
- *Sesgo* ligeramente mayor, ya que muchos nodos no son óptimos localmente.
- Entrenamiento más *rápido*: no se busca el mejor punto de corte exacto.

Hiperparámetros distintivos.

- `splitter = 'random'` (implícito): umbrales aleatorios.
- Misma lista de `n_estimators`, `max_features`, `max_depth`, etc.; el parámetro `bootstrap` suele fijarse a `False`.

Resumen comparativo.

	Varianza	Sesgo
Árbol individual	Alta	Bajo-medio
Random Forest	Media-baja	Ligeramente mayor
Extra Trees	Baja	Moderada

La elección entre *Random Forest* y *Extra Trees* depende del compromiso deseado entre sesgo y varianza, así como de los recursos computacionales disponibles. En ciencia de materiales y nanociencia, donde n suele ser pequeño y los descriptores muy ruidosos, ambos ensambles ofrecen predicciones más estables que un árbol único; Extra Trees brinda velocidad adicional y puede ser preferible cuando se requieren múltiples ejecuciones (p. ej. en búsqueda de hiperparámetros) o cuando los datos presentan alta colinealidad.

Las redes neuronales multicapa (o *Multilayer Perceptron*, MLP) son modelos de aprendizaje supervisado inspirados en la arquitectura y el funcionamiento del cerebro humano. Su origen se remonta al trabajo de Warren McCulloch y Walter Pitts (1943) sobre neuronas formales, así como al perceptrón de Frank Rosenblatt (1958), que implementó un clasificador lineal entrenable. El MLP extiende el perceptrón sencillo añadiendo una o más capas ocultas de neuronas, lo que le permite aproximar funciones no lineales arbitrarias. Un MLP para regresión consta de:

- **Capa de entrada:** recibe el vector de características $\mathbf{X} \in R^D$, donde D representa el número de atributos. Cada neurona de esta capa calcula una combinación lineal de las entradas más un sesgo (bias), y la pasa a través de una función de activación f :

$$y(\mathbf{X}, \mathbf{W}) = f(W_i \phi_i(x)). \quad (18)$$

Si consideramos M neuronas ocultas, el valor de activación de la j -ésima neurona se obtiene como

$$a_j = \sum_{i=1}^D w_{ji}^{(1)} x_i + w_{j0}^{(1)}, \quad j = 1, \dots, M, \quad (19)$$

donde el superíndice (1) indica la primera capa. Aquí $w_{ji}^{(1)}$ son los pesos y $w_{j0}^{(1)}$ el sesgo; ambos son parámetros ajustables durante el entrenamiento. A continuación, cada a_j se transforma mediante una función diferenciable y no lineal h , conocida como función de activación:

$$z_j = h(a_j). \quad (20)$$

- **Capas ocultas:** las salidas z_j actúan como base de funciones no lineales. Entre las funciones de activación más comunes se incluyen:

$$\text{ReLU}(x) = \max(0, x) \quad (21)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (22)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (23)$$

Estas transformaciones introducen la no linealidad necesaria para que la red pueda separar patrones complejos. A su vez, las salidas de las unidades ocultas se combinan linealmente en la siguiente capa:

$$a_k = \sum_{j=1}^M w_{kj}^{(2)} z_j + w_{k0}^{(2)}, \quad k = 1, \dots, K, \quad (24)$$

donde K es el número de neuronas de la capa de salida y $w_{k0}^{(2)}$ el correspondiente sesgo.

- **Capa de salida:** finalmente, las activaciones a_k de la última capa se transforman mediante una función g adecuada al problema de regresión, que suele ser la identidad (neurona lineal). El resultado es la predicción $\hat{y}$:

$$\hat{y} = g(a_k). \quad (25)$$

La elección de g depende de la naturaleza de la variable objetivo y de la distribución estadística que asumamos para el error.

Combinando estos pasos obtenemos la función completa de la red. Por ejemplo, si tomamos $g = \sigma$ (sigmoideal) en la capa de salida, la predicción se expresa como

$$y(\mathbf{X}, \mathbf{W}) = \sigma\left(\sum_{j=1}^M w_{kj}^{(2)} h\left(\sum_{i=1}^D w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right). \quad (26)$$

El conjunto de todos los pesos y sesgos puede agruparse en un único vector $\mathbf{W}$. De esta manera, el MLP define una función no lineal parametrizada $\hat{y} = f(\{x_i\}; \mathbf{W})$, cuya estructura gráfica se muestra en la Figura 8. El entrenamiento de la red se realiza mediante el algoritmo de *retropropagación* (*backpropagation*), en combinación con métodos de optimización basados en gradiente (por ejemplo, descenso de gradiente estocástico o adaptativo estocástico Adam). Se minimiza una función de pérdida, como el error cuadrático medio, ajustando iterativamente $\mathbf{W}$ para mejorar la calidad de las predicciones.

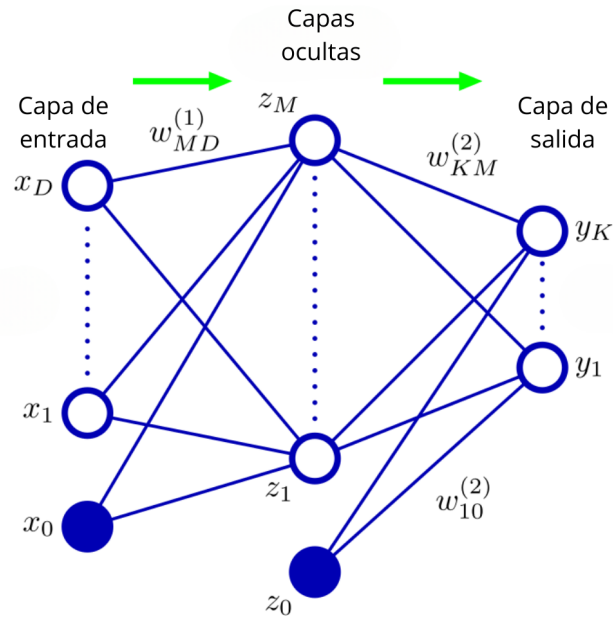


Figura 8. Anatomía de una red neuronal multicapa (perceptrón), figura adaptada del libro Bishop (2006)

Para ajustar los hiperparámetros de un MLP típicamente se exploran (Pedregosa et al., 2011):

- **Arquitectura:** número de capas ocultas (1, 2 o 3) y número de neuronas en cada capa.
- **Función de activación:** por ejemplo, *identity*, *tanh* o *relu*.
- **Algoritmo de optimización:** *adam*, *lbfgs*, *sgd*, etc.
- **Tasa de aprendizaje inicial (η):** controla el tamaño de paso en cada actualización de peso.
- **Regularización (α):** coeficiente de penalización L_2 o L_1 para evitar sobreajuste.

Los modelos de ensamble, conocidos también como meta-aprendices, combinan múltiples modelos de aprendizaje de máquina base con el objetivo de mejorar la precisión y robustez en comparación con un único modelo. Los dos enfoques principales en este tipo de algoritmos son *bagging* y *boosting*, ambos fundamentados en principios distintos. El *Bagging* (*Bootstrap Aggregating*) es una técnica de ensamble que mejora la estabilidad y la precisión de los modelos mediante la combinación de múltiples estimadores entrenados independientemente sobre subconjuntos de datos generados por muestreo aleatorio con reemplazo. Cada estimador individual aporta una predicción, y la predicción final se obtiene como promedio (en problemas de regresión) o mediante votación mayoritaria (en clasificación). La idea subyacente del *bagging* es reducir la varianza del modelo al promediar múltiples estimaciones independientes. El algoritmo más común basado en esta técnica es el `BaggingRegressor` disponible en bibliotecas como

scikit-learn. Los hiperparámetros ajustables más importantes del *bagging* son (Pedregosa et al., 2011):

- **Estimador base** (*estimator*): Modelo base que será replicado en el ensamble.
- **Número de estimadores** (*n_estimators*): Cantidad de modelos base que se combinarán.
- **Número máximo de muestras** (*max_samples*): Tamaño de cada subconjunto generado por *bootstrap*.
- **Número máximo de características** (*max_features*): Cantidad de características seleccionadas aleatoriamente para entrenar cada modelo base.
- **Uso de muestreo con reemplazo** (*bootstrap*): Indica si las muestras son seleccionadas con o sin reemplazo.
- **Muestreo de características con reemplazo** (*bootstrap_features*).
- **Evaluación fuera de bolsa** (*oob_score*): Permite estimar el desempeño del modelo usando las muestras no incluidas en cada subconjunto.

El *Boosting* es una técnica iterativa que construye secuencialmente un conjunto de modelos base, dando mayor peso en cada iteración a las observaciones que fueron mal predichas por los modelos anteriores. El objetivo principal es reducir el sesgo del modelo final, mejorando gradualmente su desempeño. El algoritmo más representativo de *boosting* es el *Gradient Boosting*, que minimiza iterativamente una función de pérdida utilizando un enfoque basado en descenso de gradiente. En cada paso, un nuevo modelo es entrenado para corregir los errores residuales del modelo anterior. Los hiperparámetros principales en algoritmos como *Gradient Boosting Regressor*, son (Pedregosa et al., 2011):

- **Función de pérdida** (*loss*): Determina cómo se penalizan los errores durante el entrenamiento (por ejemplo, error cuadrático, absoluto, etc.).
- **Tasa de aprendizaje** (*learning_rate*): Controla la contribución de cada modelo base en el ensamble. Valores menores requieren más estimadores para converger.
- **Número de estimadores** (*n_estimators*): Cantidad de modelos base añadidos secuencialmente.
- **Submuestras** (*subsample*): Fracción de muestras usadas para entrenar cada modelo base, introduciendo aleatoriedad adicional al proceso (*Boosting* Estocástico).

- **Criterio de división** (*criterion*): Define la calidad de una división en modelos base como árboles de decisión.
- **Muestras mínimas para división** (*min_samples_split*): Controla el número mínimo de muestras requeridas para dividir un nodo.
- **Muestras mínimas en hoja** (*min_samples_leaf*): Número mínimo de muestras que debe tener un nodo hoja.

Ambos enfoques, *bagging* y *boosting*, presentan fortalezas particulares y son útiles dependiendo del problema específico y del conjunto de datos. *Bagging* tiende a reducir la varianza del modelo, siendo especialmente útil con modelos que sufren de alta varianza (como árboles profundos), mientras que *boosting* es efectivo en reducir el sesgo y mejorar el rendimiento en problemas más complejos.

2.2.3. Proceso de extracción de conocimiento de los datos

La minería de datos consiste en la extracción de conocimiento a partir de grandes volúmenes de información. Esta metodología, introducida en la ciencia de materiales, busca descubrir patrones ocultos en los datos para acelerar el diseño y la predicción de nuevos materiales, concepto introducido en el artículo de Rajan (2005). En inglés, este proceso se denomina KDD (*Knowledge Discovery in Databases*) y se estructura en varias etapas bien definidas (Han et al., 2012):

- **Recolección de datos:** Consiste en construir o reunir la base de datos (BD) sobre la cual se extraerá conocimiento. Los datos pueden ser experimentales, teóricos (por ejemplo, resultados de simulaciones de DFT), obtenidos mediante interfaces de programación de aplicaciones (*API* por sus siglas en inglés) o incluso recolectados por extracción automática de sitios web conocida como *web scraping*. En esta tesis, generamos una base de datos propia a partir de cálculos sistemáticos de DFT sobre nanotriángulos Janus.
- **Extracción de características (ingeniería de descriptores):** Una vez construida la base de datos y definida la variable objetivo y (*EFS*), diseñamos y calculamos los descriptores que constituyen la matriz de características X . En esta etapa tomaron en cuenta los siguientes puntos:
 1. **Representación por composición atómica:** Cada nanotriángulo se describe mediante su fórmula general $M_aX_bY_c$, donde M es el metal de transición, X y Y son los dos calcógenos

distintos, y a, b, c sus respectivos recuentos atómicos (véase la Figura 9). Este esquema es universal y de bajo costo computacional para todo el conjunto de estructuras lo que indica ser una buena elección para usarse como representación de acuerdo a la propuesta de Damewood et al. (2023).

2. **Cálculo de fracciones atómicas:** Sea $N = a + b + c$ el número total de átomos en el nanotriángulo. Definimos las fracciones γ para cada especie atómica como:

$$\gamma_M = \frac{a}{N}, \quad \gamma_X = \frac{b}{N}, \quad \gamma_Y = \frac{c}{N}.$$

3. **Descriptores basados en propiedades elementales:** Este esquema está implementado en Matminer (acrónimo de *Materials Data Mining*) desarrollado por Ward et al. (2018); convertimos la cadena fórmula en un objeto `Composition` y aplicamos la clase `ElementProperty` para extraer estadísticas de cada propiedad química f (electronegatividad, radio covalente, volumen atómico, etc.); estos valores fueron recolectados a través de la plataforma *Materials Agnostic Platform for Informatics and Exploration* (Magpie). Para cada f calculamos:

$$\bar{f} = \sum_{i \in \{M, X, Y\}} \gamma_i f_i, \quad (\text{media ponderada}) \quad (27)$$

$$\hat{f} = \sum_{i \in \{M, X, Y\}} \gamma_i |f_i - \bar{f}|, \quad (\text{desviación media}) \quad (28)$$

$$\begin{aligned} f_{\min} &= \min_{i \in \{M, X, Y\}} f_i, & f_{\max} &= \max_{i \in \{M, X, Y\}} f_i, \\ f_{\text{rango}} &= f_{\max} - f_{\min}, & f_{\text{moda}} &= f_j \quad \text{tal que} \quad j = \arg \max_{i \in \{M, X, Y\}} \gamma_i. \end{aligned} \quad (29)$$

Esta representación fue propuesta en el trabajo de Ward et al. (2016) logrando alcanzar buenos rendimientos en métricas de regresión para considerar que esta metodología acelera el diseño y predicción de materiales inorgánicos.

4. **Características adicionales:** Además, se incorporaron como descriptores:

- La clase del nanotriángulo (oxidado = 1 si contiene oxígeno, 0 en caso contrario).
- El tamaño n (número de átomos de M en el vértice externo).
- El número total de átomos N .

De este modo, cada muestra queda representada por un vector de características que resume, de forma compacta y comparable, la composición y las propiedades químicas elementales de la

nanoestructura. Esto facilita la aplicación de modelos de ML con bajo costo computacional y buena capacidad predictiva de acuerdo a reportes similares como el de Kalidindi & De Graef (2015).

- **Limpieza de la base de datos:** Se detectan y corrigen valores nulos, duplicados o inconsistentes. Este paso es esencial para evitar que errores o ausencias de datos introduzcan sesgos o reduzcan la calidad del entrenamiento de los modelos de ML.
- **Selección de características (SC):** Con el fin de eliminar descriptores redundantes o poco informativos, se aplicaron tanto técnicas de filtrado como eliminación de columnas con varianza cero y métodos integrados como la selección mediante Lasso. En este tipo de método de selección Lasso, está basado en una regresión lineal, el operador de contracción y selección por valor absoluto penaliza los coeficientes de menor importancia y fuerza muchos de ellos a cero; se debe ajustar un modelo de regresión lineal Lasso cuya función de costo se vio en la ecuación 14, donde el segundo término —la norma L_1 — induce en el vector de pesos $\mathbf{w} = (w_1, \dots, w_D)$ los términos de menor importancia se vayan a cero. Aquellos coeficientes w_j que resultan exactamente cero (o muy cercanos a cero) se consideran no informativos y se eliminan del conjunto de descriptores. Tras experimentar con distintos valores de χ , establecimos un umbral de selección $|w_\alpha| > 0.02$ y $|w_\beta| > 0.07$; todas las características con pesos por debajo de este umbral fueron descartadas, reduciendo la dimensionalidad del espacio de características, la complejidad del modelo y el costo computacional (VanderPlas, 2022).
- **Estandarización de datos:** Muchos algoritmos de ML (regresión lineal, perceptrones, máquinas de soporte vectorial) son sensibles a la escala de las características. Por ello, normalizamos cada columna restando su media μ y dividiendo por su desviación estándar s :

$$X' = \frac{X - \mu}{s},$$

garantizando que todas las variables tengan media cero y varianza uno.

- **Ajuste de hiperparámetros:** A diferencia de los parámetros internos de la red (los pesos w_{ji} y los sesgos w_{j0}), los hiperparámetros no se actualizan durante el entrenamiento, sino que se fijan *a priori*. Ejemplos típicos son la tasa de aprendizaje o la profundidad de un árbol de decisión. Para optimizarlos se empleó la búsqueda por cuadrícula (*grid search*) combinada con validación cruzada (VC), lo que permite evaluar de forma sistemática todas las combinaciones posibles y seleccionar la que maximice las métricas de regresión. El buscador interno prueba cada conjunto de hiperparámetros mediante VC de k pliegues y elige aquel que ofrece el mejor desempeño

predictivo, garantizando así un ajuste robusto y poco sensible a la partición de los datos.

- Se reduce la varianza debida a una partición específica de los datos.
 - Se obtiene una estimación más realista de la capacidad predictiva del modelo.
 - Permite comparar directamente diferentes configuraciones de hiperparámetros, usando la misma estrategia de particionado.
- **Entrenamiento:** Una vez realizado el preprocesamiento de datos, la base está lista para entrenar un algoritmo de ML. Típicamente, el conjunto se divide en **datos de entrenamiento** y **datos de evaluación**. Los primeros son aquellos de los cuales el modelo va a *aprender* o extraer los patrones, mientras que los datos de evaluación sirven para medir su capacidad predictiva. Esta separación asegura que el modelo no se *sobreajuste*: la mayor parte de los datos se emplea para entrenar y un pequeño porcentaje para evaluar. No obstante, el porcentaje de la división y la selección específica de ejemplos en cada subconjunto pueden influir en la capacidad predictiva del modelo. Para solventar este problema se empleó la VC, que permite estimar el rendimiento de forma más robusta al promediar los resultados obtenidos en múltiples particiones. La VC es una estrategia de re-muestreo que permite estimar el rendimiento real de un modelo y seleccionar hiperparámetros sin incurrir en sobreajuste. A grandes rasgos, el conjunto de datos $\mathcal{D}$ se divide en varias particiones (conjuntos) mutuamente exclusivas; el modelo se entrena con algunos conjuntos y se evalúa en los restantes, repitiendo el proceso hasta que cada partición haya sido usada como validación exactamente una vez.

1. **Particionado.** Se elige un entero K (en este trabajo $K = 5$) y se parten las N observaciones en subconjuntos $\mathcal{D}_1, \dots, \mathcal{D}_K$ de tamaños lo más similares posible.

2. **Entrenamiento y evaluación.** Para cada $k = 1, \dots, K$:

a) Definir el *conjunto de entrenamiento* $\mathcal{D}_{\text{entrenamiento}}^{(k)} = \mathcal{D} \setminus \mathcal{D}_k$.

b) Ajustar el modelo $f^{(k)}$ con $\mathcal{D}_{\text{entrenamiento}}^{(k)}$.

c) Calcular la métrica de rendimiento (p.ej. R^2 o $RMSE$) sobre el *conjunto de prueba* $\mathcal{D}_{\text{Evaluación}}^{(k)} = \mathcal{D}_k$, obteniendo el valor $E^{(k)}$.

3. **Estimación final.** Se reporta la media $\bar{m} = \frac{1}{K} \sum_{k=1}^K m^{(k)}$ y, opcionalmente, la desviación

$$\text{típica } \sigma_m = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (m^{(k)} - \bar{m})^2}.$$

Ventajas

- Aprovecha todo el conjunto de datos para entrenar y validar, proporcionando una estimación menos sesgada que un único entrenamiento.
- Disminuye la varianza de la métrica al promediar múltiples pruebas.

Limitaciones

- Aumenta el coste computacional (entrenar el modelo k veces).
- Si los datos presentan dependencia temporal o espacial, deben usarse esquemas específicos para evitar fuga de información.

En esta tesis se emplea principalmente la validación cruzada estratificada de 5 pliegues para evaluar el desempeño de los diferentes modelos de regresión sobre la *EFs*, y la versión anidada para la búsqueda óptima de hiperparámetros.

- **Métricas de regresión:** Una vez entrenados, se compararon distintos algoritmos de regresión empleando métricas estándar para evaluar su **capacidad predictiva**. El coeficiente de determinación R^2 (ecuación 30) mide la precisión de las predicciones al comparar el *promedio de las diferencias cuadráticas* entre el valor real y el valor predicho con el promedio de las diferencias entre el valor real y la media aritmética de las predicciones; se mide entre 0 y 1, donde 0 indica nula capacidad descriptiva y 1 una perfecta descripción: se busca maximizar este valor. Para cuantificar el error del modelo se utilizan dos métricas: la **raíz del error cuadrático medio** (*RMSE*, ecuación 31), que resume la desviación cuadrática media entre el valor real y el predicho, y el **error absoluto medio** (*MAE*, ecuación 32), que promedia el valor absoluto de dichas diferencias. Estas métricas permiten *definir y seleccionar* el algoritmo de ML que ofrece los mejores resultados al contrastar los valores reales de la BD con los valores estimados por cada modelo. Los errores se miden en las mismas unidades que la variable objetivo y se buscan los valores mínimos.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (30)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (31)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (32)$$

Las métricas permiten cuantificar la precisión y robustez de las predicciones de la *EFs*, y seleccionar el modelo óptimo para su aplicación en diseño de nanomateriales.

Este flujo de trabajo (véase la Figura 4) combina simulaciones de DFT con modelado de ML para identificar y predecir nuevos nanomateriales con gran eficiencia, acelerando así el descubrimiento de nanoestructuras con potencial en nanotecnología y ciencia de materiales. La fase de ML se implementó en un equipo de propósito general usando Python: `scikit-learn` librería desarrollada por el equipo de Pedregosa et al. (2011) para el preprocesamiento y la construcción de modelos, y `Matminer` (Ward et al., 2011) para la extracción automatizada de descriptores compuestos. El ajuste de hiperparámetros se llevó a cabo mediante validación cruzada de 5 pliegues, explorando cuadrículas de parámetros específicas para cada algoritmo. Se compararon diversos métodos supervisados de regresión, desde modelos lineales y máquinas de soporte vectorial, hasta árboles de decisión, perceptrones multicapa y técnicas de ensamble (*Boosting* y *Bagging*), seleccionando el modelo con mayor rendimiento según las métricas definidas.

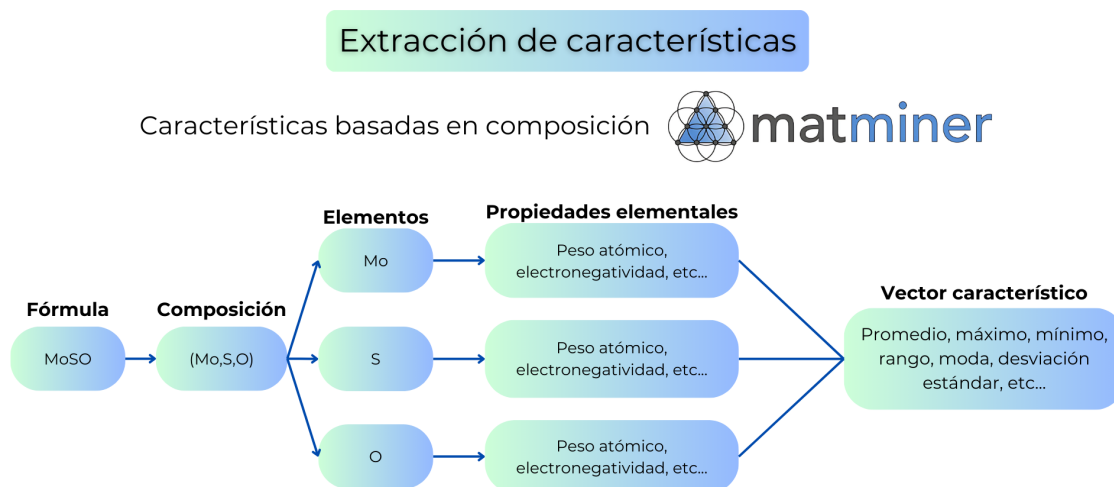


Figura 9. Esquema del proceso de extracción de características

Capítulo 3. Resultados

En total se simularon 172 nanotriángulos para cada modelo de borde, α y β (véase la Tabla 4), esta fue la BD de ejemplo que se usó para entrenar los algoritmos de ML, que incluyen nanotriángulos de diferentes tamaños en los que varía n . Se utilizaron los metales y calcógenos indicados en la Figura 5, y además se simularon las fases prístinas MX_2 para servir de referencia al calcular la EFS . Un aspecto relevante es que, en las familias oxidadas (e.g., $MoSO$, $MoSeO$, $MoTeO$, WSO , $WSeO$, $WTeO$), para tamaños mayores emerge una curvatura pronunciada. En contraste, sus contrapartes no oxidadas (e.g., $MoSSe$, $MoSeTe$, $WSSe$, $WSeTe$) y las fases prístinas permanecen prácticamente planas. Dicha curvatura se atribuye a la diferencia de electronegatividad entre los dos calcógenos y a los efectos de confinamiento en la nanoestructura 0D. En la Figura 10 se muestra la evolución estructural de los nanotriángulos $MoSeO$ y $MoSeTe$ a medida que crece n , ilustrando con claridad la distinción entre fases oxidadas y no oxidadas. La presencia de deformaciones en TMDs puede modificar localmente el comportamiento electrónico, lo cual podría influir en su desempeño catalítico bajo ciertas condiciones experimentales. No obstante, el control de dichas deformaciones en entornos reales de síntesis permanece como un reto abierto. De este modo, hemos generado una base de datos exhaustiva de nanotriángulos Janus —abarcando diversas composiciones, tamaños y terminaciones de borde— y recogido sus energías del estado base.

3.1. Energías de Formación de Superficie

Para calcular la EFS se utilizó la ecuación 5 en ambos modelos de borde, α y β . Esta es la BD de ejemplo que se usó como variable objetivo para cada composición. Como referencia se tomó siempre el nanotriángulo prístino con el calcógeno de menor electronegatividad: por ejemplo, para $MoSO$ se empleó MoS_2 con el respectivo tamaño, y para $WSTe$ se utilizó WTe_2 . Esta elección se basa en que la transición desde la fase Janus a la fase prístina con el calcógeno más electronegativo implicaría un costo energético mayor. Se obtuvieron 172 valores de EFS para cada modelo, los cuales indican la viabilidad termodinámica de formar cada nanotriángulo a partir de la fase prístina: un $EFS < 0$ señala estabilidad química y un $EFS > 0$ implica que la formación no es favorable. En la Figura 11 se muestran los histogramas de los valores de EFS para los modelos α y β , junto con su estimación de densidad y el porcentaje de estructuras estables versus no estables. Los resultados revelan que más del 90 % de los nanotriángulos son químicamente estables ($EFS < 0$) en ambos modelos, lo cual sugiere una viabilidad termodinámica bajo las condiciones específicas del cálculo, aunque no se garantiza su síntesis sin considerar condiciones experimentales y efectos cinéticos. Sólo un pequeño porcentaje presenta

$EFS > 0$ —es decir, su formación resulta desfavorable—; dichos casos se detallan en la Tabla 5. Las distribuciones de EFS_α y EFS_β constituyen las variables objetivo definidas para el posterior modelado de ML. El hecho de que más del 90 % de los nanotriángulos Janus de TMD tengan $EFS < 0$ implica que su formación es químicamente favorable. Esto sugiere que, bajo condiciones adecuadas como la selección de precursores, temperatura de reacción controlada y atmósfera rica en especies de calcógenos (S, Se o Te), las estructuras Janus podrían formarse de manera espontánea o mantenerse estables una vez creadas. Sin embargo, hacen pruebas más rigurosas para evidenciar su estabilidad termodinámica y dinámica por ejemplo a temperatura ambiente o altas como es requerido en catálisis. Esta evaluación sistemática de múltiples composiciones (M_a, X_b, Y_c) ofrece un mapa predictivo para priorizar combinaciones Janus de alto rendimiento y descartar aquellas menos estables, apuntando a aplicaciones a largo plazo en optoelectrónica, sensores y catálisis, donde la robustez química bajo condiciones ambientales es esencial.

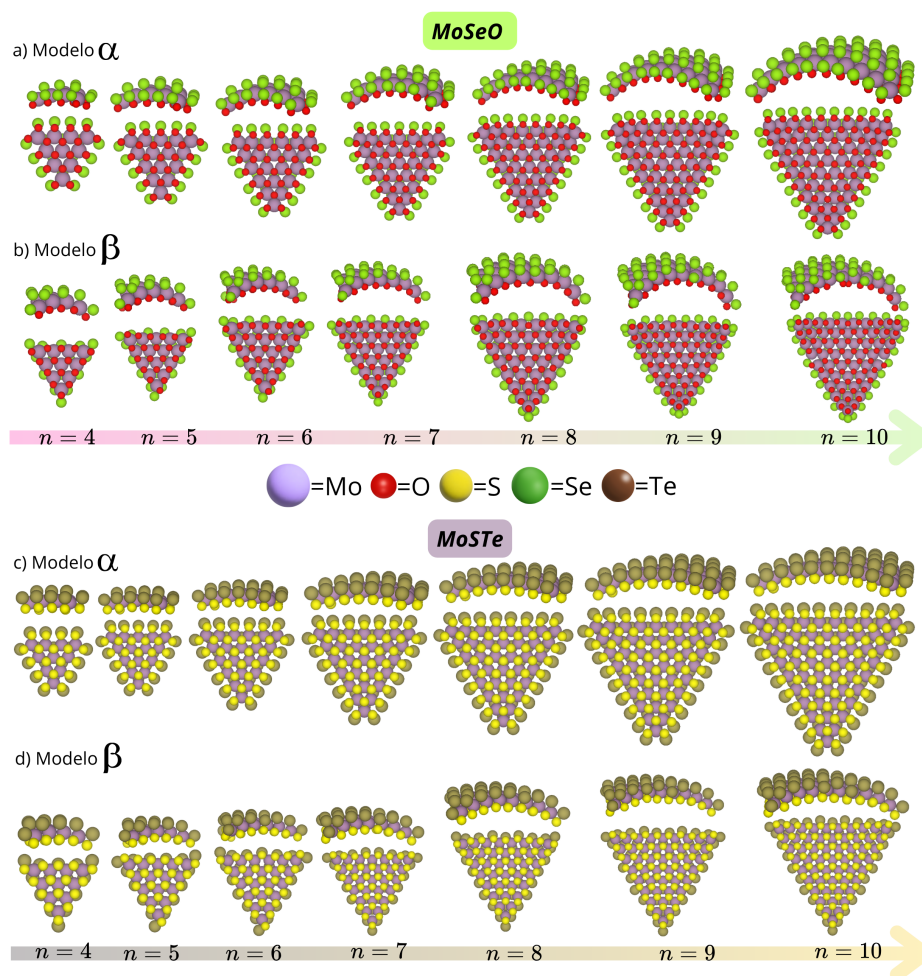


Figura 10. Evolución estructural de los nanotriángulos Janus de fórmula general $MX Y$ para los sistemas MoSeO (arriba) y MoSTe (abajo). En cada bloque, las subfiguras (a) y (c) muestran el modelo α —con un dímero en el borde— para tamaños $n = 4$ a $n = 10$, mientras que las subfiguras (b) y (d) corresponden al modelo β —con dos dímeros— en los mismos valores de n . Los átomos de Mo se representan en violeta, los de O en rojo, S en amarillo, Se en verde y los de Te en marrón.

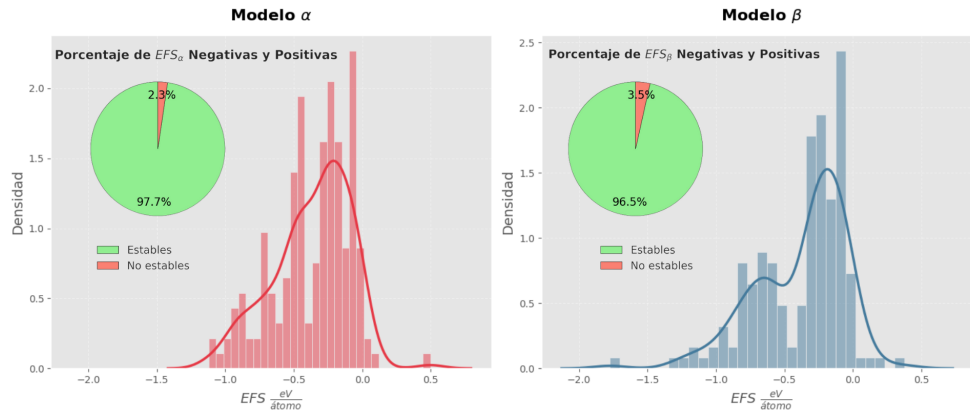


Figura 11. Distribución (Variable objetivo EFS_{α} y EFS_{β}) de las energías de formación de superficie (EFS) para los modelos de borde α (izquierda) y β (derecha). En cada panel se muestra el histograma de los valores de EFS junto con su estimación de densidad de probabilidad (línea continua), y un recuadro circular que indica el porcentaje de nanotriángulos estables ($EFS < 0$, verde) y no estables ($EFS > 0$, rojo). Estos son los valores de EFS que se utilizaron en la base de datos de ejemplo.

3.2. Extracción y selección de características

Mediante el método de extracción de características basado en composición descrito en el capítulo 2, se obtuvieron inicialmente 132 descriptores elementales —por ejemplo, electronegatividad, radio covalente, número de valencia, volumen atómico, etc.— a partir de la fórmula de cada nanotriángulo $M_aX_bY_c$, donde a , b y c indican el número de átomos de cada especie. Como primer paso de preprocesamiento, se detectaron y eliminaron columnas con valores nulos, duplicados o con varianza cero (valores constantes), así como aquellas basadas en propiedades periódicas de fases más estables por composición. Con ello se redujo el espacio de descriptores a 99 para cada modelo de borde. A continuación, se aplicarán las técnicas de SC para conservar únicamente los descriptores más informativos de cara al modelado de la EFS . Se evaluaron tres métodos:

- Comparación de información mutua para regresión (CIM).
- Selección basada en bosque aleatorio (BA).
- Contracción y selección por norma absoluta mínima (*Lasso*).

Cada método se comparó midiendo su impacto en las métricas de evaluación (R^2 , $RMSE$ y MAE) de un modelo lineal simple entrenado mediante VC de 5 pliegues. Los resultados, que aparecen en la Tabla 1, muestran que Lasso (sin escalado) alcanza las mejores métricas medias, por lo que se seleccionó

como método definitivo. Con Lasso se conservaron únicamente los descriptores de importancia no nula (coeficientes mayores que cero en valor absoluto), resultando en 22 características para el modelo α y 30 para el modelo β .

Tabla 1. Desempeño de los modelos α y β en regresión lineal simple con validación cruzada (5 pliegues), aplicando distintos métodos de selección de características. Se reportan los valores promedio de R^2 , RMSE y MAE sobre el conjunto de prueba; en cada caso se mantuvieron sólo las características de importancia positiva. El método Lasso (sin escalado) logra las mejores métricas en ambos modelos.

Método	α			β		
	R^2	RMSE	MAE	R^2	RMSE	MAE
Todas las características	0.61	0.16	0.10	0.48	0.21	0.14
CIM (Información mutua)	0.61	0.16	0.10	0.48	0.21	0.14
Bosque Aleatorio (BA)	0.72	0.14	0.08	0.60	0.19	0.14
Lasso (sin escalado)	0.74	0.13	0.08	0.68	0.18	0.13

De este modo, la estrategia recomendada para predecir la $EF\bar{S}$ en nanotriángulos Janus es:

Extracción por composición $M_aX_bY_c \rightarrow$ Selección con *Lasso* $\rightarrow$ Entrenamiento con VC.

Para optimizar aún más los modelos de ML, se realizó un estudio iterativo variando el umbral importancia de Lasso entre 0.01 y 0.10 (partiendo de 0), cuyos resultados se muestran en la Figura 21. Finalmente, se fijaron los umbrales óptimos de selección en 0.02 para α y 0.05 para β , obteniendo 16 y 18 características respectivamente, véase la Tabla 6 y Tabla 7 para una descripción completa de cada característica. Los descriptores más relevantes fueron el radio covalente, la electronegatividad, el peso atómico y el número de valencia, seleccionados por los modelos Lasso y respaldados por el análisis de correlación de Pearson en la Figura 22. Estas características son importantes porque remarcan el hecho de que las nanoestructuras, con menor radio covalente y mayor electronegatividad, tienden a minimizar la $EF\bar{S}$ y presentan curvaturas más pronunciadas (véase la Figura 12). El método *Lasso* revela una relación lineal entre la $EF\bar{S}$ y características como el tamaño atómico, la electronegatividad de los elementos en la composición y la clase de nanotriángulo. En particular, los nanotriángulos oxidados presentan valores de $EF\bar{S}$ más bajos, como se aprecia en la Figura 20. En el modelo de borde α , se observa una correlación positiva entre $EF\bar{S}_\alpha$ y el valor máximo del peso atómico: al aumentar el peso atómico de la composición, crece $EF\bar{S}_\alpha$, lo que sugiere que seleccionar átomos más ligeros favorece la estabilización química de los nanotriángulos. De forma inversa, tanto el radio covalente como la electronegatividad muestran una correlación negativa con $EF\bar{S}_\alpha$: al incrementar estos descriptores, disminuye la $EF\bar{S}$, véase la Figura 22. Para el modelo de borde β , la correlación positiva más alta se da con la moda de la banda prohibida y el volumen atómico máximo de las especies en su fase más estable, indicando que átomos más grandes

y bandas energéticas mayores tienden a elevar EFS_{β} . Por otra parte, la clase del nanotriángulo (óxido vs. no óxido) y la electronegatividad presentan correlaciones negativas con EFS_{β} , lo que implica que las estructuras oxidadas y con mayor electronegatividad resultan más estables (menor EFS_{β}), véase la Figura 22. Estos resultados permiten definir pautas de diseño para la estabilidad química de nanotriángulos Janus: la ingeniería de composición, orientada a especies más ligeras y electronegativas, favorece la formación espontánea y la estabilidad química de estas nanoestructuras.

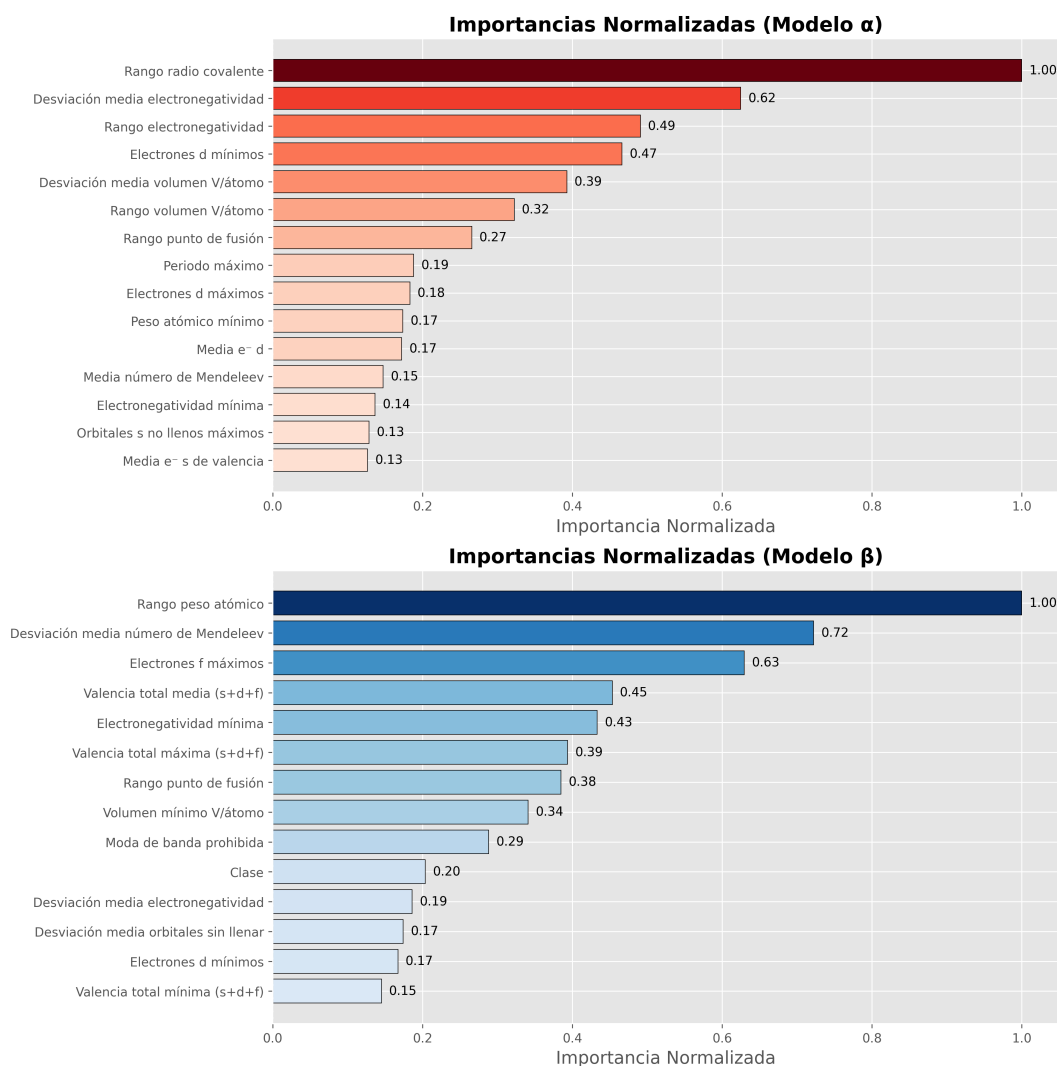


Figura 12. Importancia de las características determinada por *Lasso* para predecir la EFS de nanotriángulos Janus de TMD. El panel superior corresponde al modelo α y el inferior al β . Las barras representan los coeficientes normalizados de *Lasso*, ordenados de mayor a menor. Entre los descriptores más influyentes destacan el rango de electronegatividad, el rango de radios covalentes y atributos de valencia, subrayando el papel de la diversidad química en la estabilidad de la EFS . Para una descripción completa del significado de cada variable véase Tabla 6 y Tabla 7.

3.3. Entrenamiento y evaluación

Tras la selección de características, se entrenaron diversos algoritmos de regresión supervisada. El conjunto de modelos incluyó:

- *Modelos lineales: Ridge, Lasso y Elastic Net.*
- *Modelos no lineales: MLP, SVR y árboles de decisión.*
- *Técnicas de ensamble: boosting y bagging.*

Todos los datos se estandarizaron mediante la transformación descrita en la sección 2.2.2, garantizando media cero y varianza unitaria en cada descriptor. Para cada algoritmo se optimizaron los hiperparámetros mediante VC estratificada de cinco pliegues, evaluando una cuadrícula predefinida con el fin de maximizar las métricas de regresión (ecuaciones 30, 31, 32). Los valores óptimos hallados se resumen en la Tabla 8 del anexo. Una vez ajustados los hiperparámetros, se seleccionó la configuración óptima de cada modelo y se realizaron experimentos finales aplicando nuevamente VC ($k = 5$). Esta estrategia evita tanto el sobreajuste como el subajuste, puesto que los algoritmos son sensibles al porcentaje de división entre entrenamiento y prueba. El subconjunto de prueba, nunca expuesto durante el aprendizaje, permite evaluar objetivamente la capacidad predictiva de cada modelo. En los nanotriángulos con terminación α , el algoritmo con mejor desempeño fue el ensamble *Gradient Boosting*, con un poder predictivo promedio de $R^2 = 0.81 \pm 0.21$; sus errores fueron $RMSE = 0.10 \pm 0.05$ y $MAE = 0.06 \pm 0.02$ (eV/átomo). Para la terminación β , el modelo más destacado fue el ensamble *AdaBoost* con estimador base SVR, alcanzando $R^2 = 0.85 \pm 0.08$, $RMSE = 0.12 \pm 0.05$ y $MAE = 0.07 \pm 0.02$ (eV/átomo). La Tabla 2 y la Figura 14 presentan las métricas promedio de evaluación para todos los algoritmos. La Figura 13 muestra los diagramas de dispersión de los mejores modelos para cada geometría de borde, con los valores reales (DFT) frente a los valores predichos por el modelo de ML; la línea negra punteada representa la predicción ideal. La proximidad de los puntos a dicha línea confirma la alta capacidad de generalización, incluso utilizando únicamente descriptores basados en composición. El aumento de la complejidad del modelo mejora sistemáticamente la precisión. En conjunto, la combinación de extracción y selección de características, junto con el ajuste exhaustivo de hiperparámetros, permitió obtener un desempeño predictivo robusto para la *EFS* en nanotriángulos Janus con terminaciones α y β .

Tabla 2. Desempeño de los modelos de borde α y β en la predicción de la EFS , empleando los hiperparámetros óptimos hallados para cada algoritmo (véase la Tabla 8). Se muestra la media $\bar{m}$ y la desviación estándar σ_m de las métricas R^2 , $RMSE$ y MAE ($\bar{m} \pm \sigma_m$), calculadas mediante validación cruzada de 5 pliegues sobre el conjunto de evaluación.

Algoritmo	Modelo	R^2	$RMSE$ (eV/átomo)	MAE (eV/átomo)
<i>Ridge</i>	α	0.76 ± 0.21	0.12 ± 0.04	0.07 ± 0.02
	β	0.72 ± 0.10	0.16 ± 0.04	0.12 ± 0.03
<i>Lasso</i>	α	0.75 ± 0.20	0.13 ± 0.04	0.08 ± 0.02
	β	0.68 ± 0.14	0.17 ± 0.04	0.12 ± 0.03
<i>Elastic Net</i>	α	0.76 ± 0.18	0.12 ± 0.04	0.08 ± 0.02
	β	0.71 ± 0.10	0.16 ± 0.03	0.12 ± 0.02
SVR (lineal)	α	0.79 ± 0.20	0.11 ± 0.05	0.06 ± 0.01
	β	0.82 ± 0.09	0.13 ± 0.05	0.07 ± 0.02
SVR (Nu)	α	0.79 ± 0.23	0.11 ± 0.05	0.06 ± 0.01
	β	0.82 ± 0.10	0.13 ± 0.05	0.06 ± 0.02
Árbol de decisión	α	0.77 ± 0.20	0.12 ± 0.04	0.06 ± 0.02
	β	0.78 ± 0.12	0.14 ± 0.04	0.08 ± 0.02
Bosque aleatorio	α	0.78 ± 0.16	0.12 ± 0.04	0.07 ± 0.01
	β	0.80 ± 0.07	0.14 ± 0.05	0.08 ± 0.03
Árboles-extra	α	0.79 ± 0.20	0.11 ± 0.05	0.06 ± 0.01
	β	0.80 ± 0.10	0.14 ± 0.05	0.07 ± 0.02
Red Neuronal (MLP)	α	0.76 ± 0.21	0.12 ± 0.05	0.08 ± 0.02
	β	0.79 ± 0.08	0.14 ± 0.04	0.09 ± 0.02
<i>AdaBoost</i> (árbol)	α	0.80 ± 0.20	0.11 ± 0.05	0.06 ± 0.02
AdaBoost(SVR)	β	0.85 ± 0.08	0.12 ± 0.05	0.07 ± 0.02
Gradient Boosting	α	0.81 ± 0.21	0.10 ± 0.05	0.06 ± 0.01
<i>Gradient Boosting</i>	β	0.75 ± 0.11	0.15 ± 0.04	0.09 ± 0.03
<i>Bagging</i> (NuSVR)	α	0.77 ± 0.24	0.11 ± 0.05	0.06 ± 0.01
	β	0.83 ± 0.09	0.13 ± 0.05	0.07 ± 0.02

3.4. Modelos en producción

Una vez ajustados y entrenados los modelos de ML para cada tipo de borde, se realizó la predicción de la EFS para el resto de nanotriángulos no calculados con DFT, comprendiendo 998 estructuras para cada geometría de borde (en total 1996 nanotriángulos), α y β . Para estas predicciones, se consideraron todos los metales de transición (MT) de los grupos III al XII y de los periodos 4 a 6 (excepto molibdeno y tungsteno). Se incluyeron tamaños desde $n = 4$ hasta $n = 10$ para los MT que no formaron parte de la base de datos original (véase Figura 5), y tamaños de $n = 5$ a $n = 10$ para aquellos MT incluidos inicialmente. Como casos particulares de interés que no fueron integrados en la BD de ejemplo, se evaluaron los nanotriángulos de $MoSO$ en geometría α con composiciones $Mo_{10}S_{18}O_{18}$ y $Mo_{55}S_{75}O_{75}$, y en geometría β con composiciones $Mo_{10}S_{15}O_{15}$ y $Mo_{55}S_{66}O_{66}$. Los resultados indican que la mayoría de los nanotriángulos son estables según la EFS predicha por los modelos de ML para ambas geometrías (véase la Figura 15). Casi todas las EFS resultan negativas, lo que señala viabilidad química, es decir,

son estables. Concretamente, el 95 % de las composiciones $M_aX_bY_c$ en el modelo α presentan $EFS < 0$ (estables) y el 5 % restante muestran $EFS \geq 0$ (no estables). Para el modelo β , el 91% de las composiciones son estables y el 8,3 % son no estables.

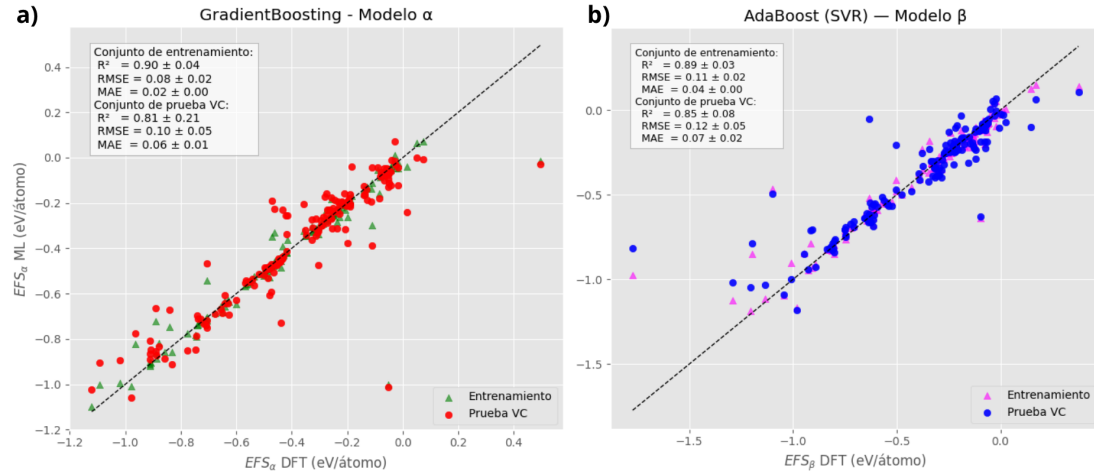


Figura 13. Diagramas de dispersión entre los valores reales de la energía de formación de superficie (EFS) y las predicciones de los modelos de aprendizaje automático para nanotriángulos tipo Janus. El panel (a) muestra el desempeño de *Gradient Boosting* para el borde α ; el panel (b), el del ensamble *AdaBoost* con SVR como estimador base para el borde β . En cada panel, los triángulos representan las predicciones sobre el conjunto de entrenamiento y los círculos las de la validación cruzada ($k = 5$); la línea discontinua negra indica la tendencia ideal $y = x$. En los recuadros internos se reportan las medias y desviaciones estándar de R^2 , $RMSE$ y MAE , tanto para entrenamiento como para validación cruzada.

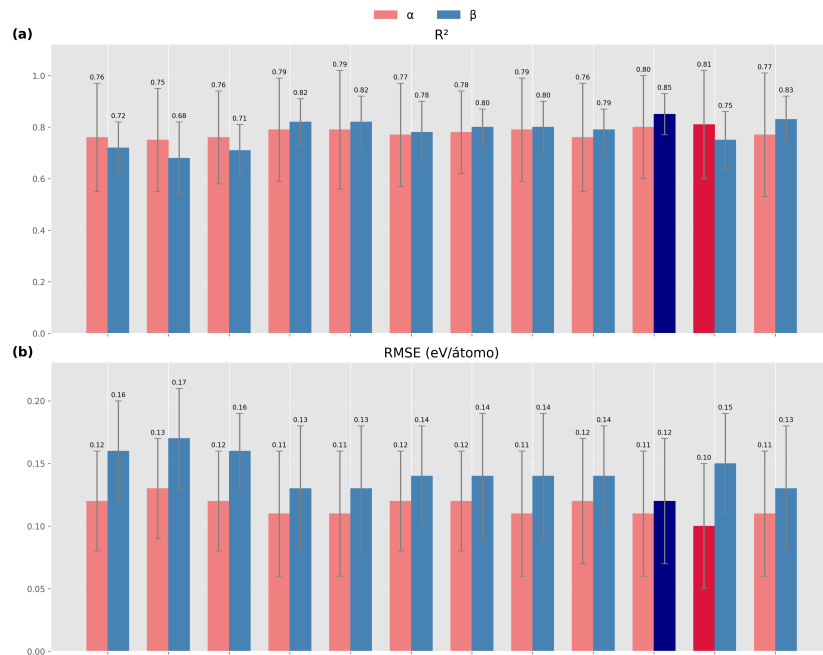


Figura 14. Desempeño comparado de los modelos de borde α (barras rojas) y β (barras azules) para la predicción de la energía de formación de superficie. (a) Coeficiente de determinación R^2 : el modelo *Bagging* (NuSVR) alcanza el mayor R^2 en α , mientras que *AdaBoost* (NuSVR) lo maximiza en β . (b) Error cuadrático medio (RMSE): *Bagging* (NuSVR) consigue el menor RMSE en α , y *AdaBoost* (NuSVR) en β .

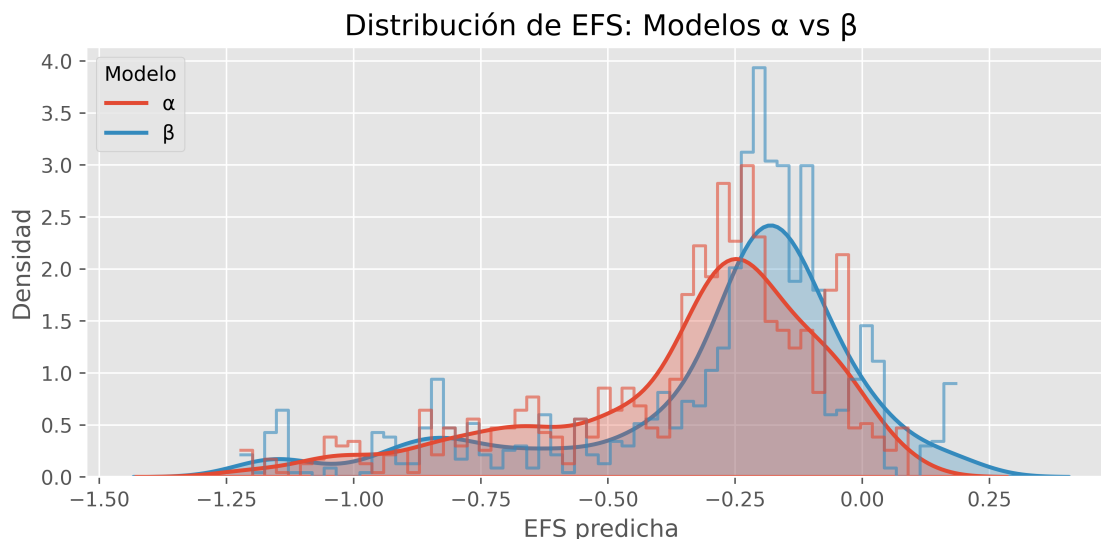


Figura 15. Distribución de valores predichos de *EFS* por los modelos de ML para los modelos de borde α en color rojo y β en color azul, se puede observar que de acuerdo a los modelos de ML la mayoría de valores de *EFS* son negativos lo que indica que la estabilidad química es favorable de acuerdo a esta cantidad.

3.5. Validación de primeros principios del conjunto *MoSO*

Para ampliar el estudio de la estabilidad, se decidió validar mediante DFT las predicciones de los modelos de ML sobre la estabilidad de los nanotriángulos de *MoSO* (dos tamaños que no se incluyeron en la BD de ejemplo) en geometrías α y β . Calculando la *EFS* mediante DFT y, adicionalmente, con simulaciones AIMD, se evaluó la estabilidad a temperatura ambiente. Además, se obtuvieron las propiedades electrónicas mediante el análisis de las isosuperficies de potencial electrostático (ISPE). El análisis se centró en los conjuntos formados por el metal de transición molibdeno (Mo). En esta sección se reportan las dinámicas moleculares AIMD y las ISPE para contrastar las diferencias entre tamaños y modelos de borde. En la Figura 16 se muestran los nanotriángulos optimizados con el algoritmo de optimización de VASP hasta alcanzar la configuración de mínima energía.

3.5.1. Estabilidad química

Se calculó la *EFS* de los nanotriángulos con composiciones $Mo_{10}S_{18}O_{18}$ ($n = 4$) y $Mo_{55}S_{75}O_{75}$ ($n = 10$) del modelo de borde α . Y del modelo de borde β con composiciones $Mo_{10}S_{15}O_{15}$ ($n = 4$) y $Mo_{55}S_{66}O_{66}$ ($n = 10$) usando DFT y la ecuación 5 para comparar con las predicciones hechas por

los modelos de ML. De acuerdo a los modelos entrenados las EFS_{α} de los nanotriángulos α para $n = 4$ y $n = 10$ se predicen como estables coincidiendo en tendencias con el formalismo de DFT que también predice como estables, aunque con valores ligeramente diferentes como se reporta en la Tabla 3. Logrando capturar la tendencia en las EFS de los nanotriángulos. Para los modelos β los valores de EFS_{β} calculados con ML y DFT mantienen las mismas tendencias de estabilidad química.

Tabla 3. Comparación de las energías de formación superficie predichas por ML y calculadas mediante DFT, con la diferencia absoluta.

Composición	EFS_{ML}	EFS_{DFT}	$ \Delta EFS $
$Mo_{10}S_{18}O_{18}$	-0.405154	-0.425341	0.020187
$Mo_{55}S_{75}O_{75}$	-0.448917	-0.451005	0.002088
$Mo_{10}S_{15}O_{15}$	-0.524784	-0.577446	0.052662
$Mo_{55}S_{66}O_{66}$	-0.560680	-0.534952	0.025728

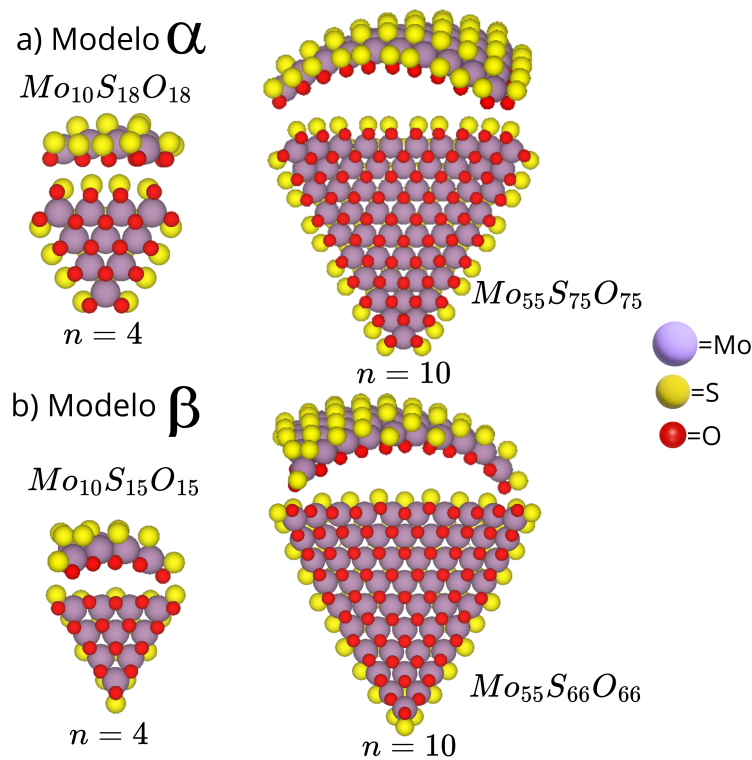


Figura 16. Nanoestructuras optimizadas mediante DFT de nanotriángulos Mo–S–O en geometrías de borde α y β . (a) Modelo α : composiciones $Mo_{10}S_{18}O_{18}$ ($n = 4$) y $Mo_{55}S_{75}O_{75}$ ($n = 10$). (b) Modelo β : composiciones $Mo_{10}S_{15}O_{15}$ ($n = 4$) y $Mo_{55}S_{66}O_{66}$ ($n = 10$). Las esferas moradas representan átomos de molibdeno, las amarillas de azufre y las rojas de oxígeno.

3.5.2. Estabilidad termodinámica

Con el fin de evaluar la estabilidad térmica de los nanotriángulos, se realizaron simulaciones de AIMD empleando VASP bajo un ensamble NVT a 300 K, durante 5 ps para cada caso. Se seleccionaron los sistemas *MoSO* con tamaños $n = 4$ y $n = 10$ para ambos modelos de borde, de modo que se pudiera comparar el comportamiento de estructuras pequeñas y grandes sin exceder el costo computacional de estas simulaciones.

En la Figura 17 (panel a) se muestra la evolución de la temperatura (línea roja) y la energía total (línea azul) para $\text{Mo}_{10}\text{S}_{18}\text{O}_{18}$ en el modelo α . Desde etapas tempranas, a partir de 0.5 ps, ambos valores fluctúan alrededor de un valor medio estable, con oscilaciones en el rango de 250–350 K para la temperatura y variaciones de energía de 0.5 eV alrededor de -310 eV. En el panel b de la misma figura, correspondiente a $\text{Mo}_{55}\text{S}_{75}\text{O}_{75}$ ($n = 10$) en el modelo α , las fluctuaciones de temperatura y energía también oscilan alrededor de 300 K y de -1508 eV, respectivamente, pero con amplitudes ligeramente menores que en $n = 4$. Este comportamiento sugiere que, al aumentar el tamaño del nanotriángulo, la energía se estabiliza más rápidamente y con oscilaciones menos pronunciadas, lo que refuerza la idea de que las estructuras grandes presentan mayor estabilidad térmica y química. De manera similar, en la Figura 24 se reportan los resultados para el modelo β . En el panel a, $\text{Mo}_{10}\text{S}_{15}\text{O}_{15}$ ($n = 4$) exhibe fluctuaciones de temperatura cercanas a 300 K y oscilaciones de energía alrededor de -282 eV, estabilizándose después del primer pico (alrededor de 0.5 ps). En el panel b, el sistema $\text{Mo}_{55}\text{S}_{66}\text{O}_{66}$ ($n = 10$) muestra una curva de energía más suave y una temperatura con variaciones más pequeñas, indicando una estabilización aún más rápida en comparación con $n = 4$. En ambos modelos (α y β), las configuraciones inicial y final (mostradas a la izquierda y a la derecha de cada gráfico) demuestran que la geometría del nanotriángulo se mantiene razonablemente intacta tras 5 ps a 300 K. No se observan reordenamientos significativos ni desprendimientos de átomos en los planos basales o en los bordes, lo cual confirma la estabilidad estructural de estas nanoestructuras Janus bajo condiciones de temperatura ambiente. En conjunto, estos resultados de AIMD respaldan la viabilidad termodinámica de los nanotriángulos Janus basados en TMDs y su posible uso en aplicaciones que requieran estabilidad térmica.

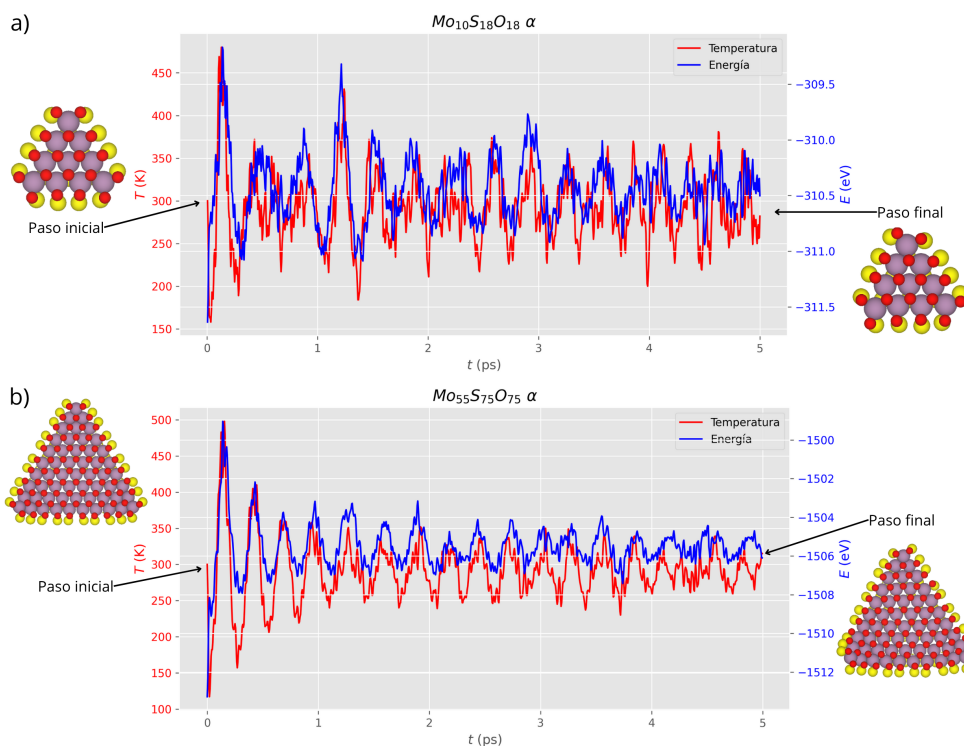


Figura 17. Resultados de las simulaciones AIMD en ensamble NVT a 300 K para nanotriángulos Janus MoSO en el modelo de borde α . Se presentan dos tamaños distintos: $n = 4$ (panel superior) y $n = 10$ (panel inferior). En cada panel, la curva roja muestra la evolución de la temperatura (eje izquierdo) y la curva azul la evolución de la energía total (eje derecho) durante 5 ps de simulación. A la derecha e izquierda de cada gráfico se muestran las configuraciones atómicas en el paso inicial y en el paso final respectivamente, evidenciando que, a pesar de las fluctuaciones, la estructura mantiene su forma triangular y converge hacia valores medios estables de temperatura y energía.

3.5.3. Propiedades electrónicas

Para estudiar las propiedades electrónicas de los nanotriángulos de composición $MoSO$ con tamaños $n = 4$ y $n = 10$ de ambos modelos de borde (α y β), se calcularon las ISPE de las nanoestructuras optimizadas. Este formalismo permite calcular los valores de potencial electrostático en la superficie del material y visualizar la distribución espacial de potencial, identificando regiones con exceso (color rojo) o déficit (color azul) de electrones. Estas zonas son relevantes porque corresponden a los sitios más reactivos del nanotriángulo, donde pueden iniciarse procesos de catálisis o funcionalización superficial, según la naturaleza electrofílica o nucleofílica de las especies reactivas. Por convención, las zonas rojas se asocian a potencial electrostático negativo (exceso de electrones), mientras que las azules indican potencial positivo (déficit electrónico). Así, regiones azules serán susceptibles a ataques nucleofílicos, y regiones rojas a ataques electrofílicos. Además las ISPE permiten anticipar sitios donde es más viable eliminar átomos o iniciar reacciones químicas. En nanotriángulos prístinos de MoS_2 , las zonas más

activas se localizan en los bordes y vértices, mientras que las superficies basales permanecen inertes, en concordancia con reportes previos hechos por Bollinger et al. (2003); Helveg et al. (2000); Lauritsen et al. (2007); Tuxen et al. (2010). Esto se observa también en las ISPE del MoS_2 (Figura 25), donde los bordes presentan déficit electrónico (potencial positivo) y el basal se mantiene con potencial negativo. La menor coordinación de los átomos en los bordes facilita su extracción o funcionalización. Al intercambiar un plano de átomos de azufre por uno de oxígeno, se induce una curvatura en la estructura (ver Figura 16), lo que produce una redistribución significativa de la densidad electrónica. Para el modelo de borde α , las ISPE muestran un comportamiento similar al prístino: los bordes siguen siendo los sitios más activos, con potencial positivo y déficit electrónico, mientras que la cara del azufre se muestra con una menor concentración de carga negativa y con una tendencia más neutra en contraste con la cara del oxígeno que muestra un gran exceso de electrones mucho mayor que la cara del azufre mostrando una gran concentración de carga negativa. En contraste, el modelo de borde β exhibe un patrón más complejo: la alternancia entre zonas de exceso y déficit electrónico se extiende no sólo en los bordes, sino también hacia el centro del triángulo, especialmente conforme crece el tamaño ($n = 10$). Esta mayor diversidad de sitios activos es consecuencia de la distinta coordinación de los átomos metálicos: en el modelo β , los átomos de borde están coordinados a dos dímeros S–O, mientras que en α sólo a uno. Así, β favorece una redistribución de la carga más pronunciada, generando sitios con diferente reactividad tanto en los bordes como en el centro. Estos resultados sugieren que los nanotriángulos Janus de borde β pueden ser más versátiles en catálisis o funcionalización superficial, al disponer de múltiples tipos de sitios activos. En ambos modelos, la presencia de oxígeno y la curvatura inducida por la asimetría de los calcógenos modulan el potencial electrostático superficial, lo que podría aprovecharse para diseñar estrategias de síntesis selectiva o aplicaciones en catálisis. En resumen, el análisis de las ISPE evidencia que el tipo de borde y la composición química influyen de manera directa en la distribución de carga y, por ende, en la reactividad superficial de los nanotriángulos Janus de $MoSO$.

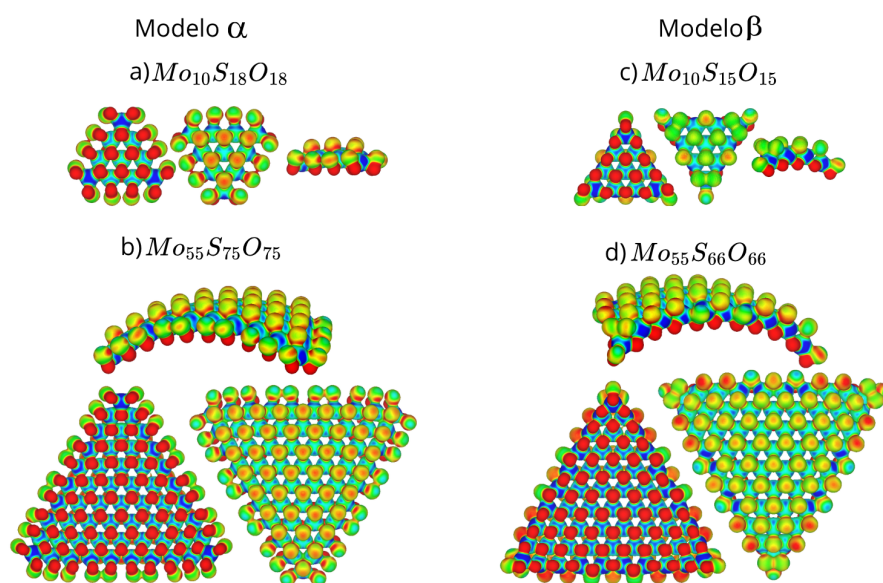


Figura 18. ISPE (isovalor = 20) para nanotriángulos de $MoSO$: modelo α de tamaño (a) $n = 4$ y (b) $n = 10$; modelo β de tamaño (c) $n = 4$ y (d) $n = 10$. En la escala de colores, el rojo corresponde a valores de potencial negativos y el azul a valores de potencial positivos.

Capítulo 4. Discusión

Una vez obtenida la BD de ejemplo mediante cálculos de DFT para los modelos de borde α y β , que abarcan las composiciones y tamaños resumidos en la Tabla 4, y cuya distribución de valores es mayoritariamente negativa (véase la Figura 11), estos cálculos predicen que la mayoría de las composiciones son químicamente estables respecto a la *EFS*. La fase de referencia empleada fue el nanotriángulo prístino con el calcógeno menos electronegativo. Sólo algunos nanotriángulos pequeños resultan químicamente inviables de formarse ($EFS > 0$), como se muestra en la Tabla 5. La BD incluye composiciones variadas de metales de transición (Figura 5) y tamaños desde $n = 4$ hasta $n = 10$, con predominio de $n = 4$ en la mayoría de las composiciones. Con esta información se construyeron dos modelos de ML independientes para cada borde (α y β) que, a partir de una representación puramente composicional $M_aX_bY_c$, permiten predecir la *EFS* con un costo computacional muy bajo y aplicable a todos los datos de la base.

En cuanto a la capacidad de generalización, ambos modelos *Gradient Boosting* para α y *AdaBoost-SVR* para β alcanzan coeficientes de determinación mayores al 90 % en entrenamiento y 80 % en validación, lo que indica que capturan la mayor parte de la variabilidad de la *EFS* tanto en el conjunto de entrenamiento como, crucialmente, en datos no vistos. Las métricas de error *RMSE* y *MAE* se sitúan en torno a 0.1 eV/átomo (desviaciones estándar $\sim 10^{-2}$). En la BD, la mínima diferencia absoluta de energía entre pares de nanotriángulos es $\sim 10^{-5}$ eV/átomo para el modelo α y $\sim 10^{-6}$ eV/átomo para el modelo β , mientras que la mayoría de las diferencias cae en el intervalo $[10^{-1}, 10^0]$ eV/átomo (Figura 19). Dado que los errores típicos de los modelos ($\sim 10^{-1}$ eV/átomo) son del mismo orden que las variaciones más frecuentes, las predicciones resultan lo suficientemente precisas para discriminar entre diferentes nanotriángulos. Sin embargo, el enfoque podría no capturar detalles finos; su fortaleza radica en la identificación rápida y escalable. En efecto, el tiempo de predicción de la *EFS* se reduce de días a unos segundos (Figura 23). Así, con únicamente descriptores basados en composición (número y tipo de metales de transición, etc.), el modelo aprende la dependencia de la *EFS* con la química del nanotriángulo. En la práctica, sin recurrir a nuevos cálculos DFT, el modelo distingue y prioriza candidatos prometedores, aunque la ausencia de información estructural puede limitar la capacidad predictiva. Los modelos en producción predicen que, para el resto de los metales de transición y tamaños no considerados inicialmente, más del 90 % de los nanotriángulos son químicamente estables, según se observa en la Figura 15. Las principales excepciones corresponden a estructuras con plata (Ag) y paladio (Pd), las cuales se predicen inestables en comparación con sus nanotriángulos prístinos. Esto refuerza la viabilidad química de la síntesis, aunque se requieren estudios más robustos para confirmar la estabilidad termodinámica y dinámica de estas nanoestructuras.

Para validar la extrapolación, se estudiaron específicamente las composiciones *MoSO* de tamaños $n = 4$ y $n = 10$, ausentes en la BD inicial. Sus *EFS* fueron predichas con los modelos de ML (*Adaboost* y *Gradient Boosting*) y corroboradas mediante DFT, coincidiendo en su estabilidad (Tabla 3). Además, las simulaciones AIMD confirman la estabilidad termodinámica a temperatura ambiente, pues las nanoestructuras conservan su integridad sin ruptura de enlaces ni deformación significativa. Finalmente, los cálculos de ISPE revelan una asimetría en la densidad de carga de los nanotriángulos Janus (Figura 18) en contraste con los prístinos (Figura 25), formándose un dipolo eléctrico que apunta fuera del plano, tal como se ha reportado para monocapas de *MoSO* en el artículo de Yagmurcukardes & Peeters (2020). Asimismo, los nanotriángulos Janus exhiben un incremento de posibles sitios activos en el centro, originado por la geometría particular del borde, la alta electronegatividad del oxígeno y la curvatura espontánea observada en las fases oxidadas (Figura 16). Tanto las optimizaciones estructurales como las simulaciones AIMD y los análisis ISPE confirman que las propiedades de los nanotriángulos varían con el tamaño, la geometría del borde y la composición. A medida que el tamaño aumenta, la curvatura se incrementa debido a la diferencia de electronegatividad entre las especies X e Y . Además, los nanotriángulos de mayor tamaño muestran una mayor estabilidad, coherente con los valores de *EFS* y con los resultados de las AIMD. Por último, el incremento de tamaño también influye en la aparición de posibles sitios activos: de acuerdo con las ISPE, cuanto mayor es el nanotriángulo, mayor es el número de sitios catalíticamente activos. Finalmente, la extracción basada en la composición ha reducido el tiempo de cálculo de los nanotriángulos de varios días a solo segundos (véase la Tabla 23), con un grado de certidumbre superior al 80 % respecto a los cálculos teóricos, lo que permite acelerar la identificación de un gran conjunto de composiciones. Las predicciones están basadas en descriptores composicionales promedio y no incorporan detalles geométricos precisos, como la relajación estructural o la coordinación local, que pueden tener un impacto significativo en las propiedades electrónicas predichas.

Capítulo 5. Conclusiones

Combinando cálculos de DFT y modelos de ML se ha logrado identificar la estabilidad química de nanotriángulos tipo Janus de composición $M_aX_bY_c$ prediciendo que la mayoría de composiciones α y β son químicamente estables de acuerdo a las *EFs* para cada geometría de borde excepto para un grupo muy puntual de composiciones pequeñas que sin embargo para tamaños mayores son estables. Esto quiere decir que es viable la formación de los nanotriángulos Janus. Los modelos de ML lograron disminuir el tiempo de cálculo de días a sólo algunos segundos permitiendo identificar de forma rápida y escalable de todas las combinaciones posibles de nanotriángulos considerando las dos geometrías (α y β), los 29 MT, 4 calcógenos y los 7 tamaños diferentes de nanotriángulos. Observando una tendencia estructural interesante del conjunto de ejemplo a medida que se aumenta el tamaño del nanotriángulo aumenta la presencia de una curvatura espontánea que es más intensa en las fases oxidadas. Esta metodología está basada puramente en una representación basada en la composición lo que no requiere de cálculos extras de DFT para obtener el vector característico de nuevos datos y una capacidad predictiva tolerable para una identificación rápida y escalable de nuevas posibles combinaciones. La selección de características y la correlación entre *EFs* y los atributos elementales reafirma el hecho crucial de la selección del tipo de átomos y el efecto de la geometría del borde, de acuerdo a esto átomos más pequeños y con mayor electronegatividad como el oxígeno minimizan la *EFs* y presentan una mayor curvatura espontánea.

La validación del conjunto *MoSO* para $n = 4$ y $n = 10$ mediante cálculos de DFT demuestra que estas nanoestructuras oxidadas son estables químicamente de acuerdo a los modelos de ML y los cálculos de AIMD a temperatura ambiente, confirmando la estabilidad termodinámica. Las ISPE confirman la aparición de una asimetría en la densidad de carga por las especies atómicas del oxígeno y del azufre, la carga se concentra en el oxígeno mostrando un exceso de electrones en la cara del oxígeno lo que potencialmente indica su capacidad de interactuar con moléculas o especies electrofílicas. Mientras que la cara del azufre muestra déficit de electrones y bordes que intercalan entre potencial positivo y negativo indicando que estos sitios pueden interactuar con moléculas nucleofílicas y siendo sitios de interés para iniciar reacciones químicas ya que al exponerse el plano basal de la capa de átomos de azufre aumenta la aparición de posibles sitios activos sobre todo en la fase β , como sugieren las ISPE de los nanotriángulos. Estos podrían ser muy buenas plataformas para la generación de nanomateriales en catalis sobre todo en procesos de hidrodesulfurización rompiendo así con el paradigma de que en las nanoestructuras basadas en TMDs sólo son activos los vértices y los ejes externos del nanomaterial en concordancia con el reporte de Paez-Ornelas et al. (2021).

Finalmente, la combinación de metodologías basadas en ML y ciencia computacional de los materiales

en el paradigma de DFT permite acelerar el descubrimiento de nuevos nanomateriales con importancia en nanotecnología. Parte del trabajo a futuro en esta línea de investigación podría ser generar una representación basada en estructura de los nanotriángulos para utilizar algoritmos de ML más complejos como las *Graph Neural Networks* (Wei et al., 2019) que utilizan una representación basada en un gráfo de la estructura lo que podría capturar la complejidad de los bordes para predecir otras propiedades físicas de interés.

Literatura citada

- Aftab, S., Iqbal, M. Z., & Rim, Y. S. (2023). Recent advances in rolling 2d tmds nanosheets into 1d tmds nanotubes/nanoscrolls. *Small*, 19(1). <https://doi.org/10.1002/smll.202205418>.
- Aguilar, G. (2002). *El hombre y los materiales*, (3a ed.). Colección La Ciencia para Todos. Fondo de Cultura Económica (FCE), SEP, CONACyT, México.
- An, Y., Fan, X., Liu, H., & Luo, Z. (2019). Improved catalytic performance of monolayer nano-triangles ws2 and mos2 on her by 3d metals doping. *Computational Materials Science*, 159, 333–340. <https://doi.org/https://doi.org/10.1016/j.commatsci.2018.12.032>.
- Asghari, J. & Naderi, M. (2024). Artificial intelligence in nanomaterials studies: A statistical overview. *Nanoscale and Advanced Materials*, 1(2), 73–86. <https://doi.org/10.22034/nsam.2024.04.01>.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer Science.
- Blöchl, P. E. (1994). Projector augmented-wave method. *Physical Review B*, 50, 17953–17979. <https://doi.org/10.1103/PhysRevB.50.17953>.
- Bollinger, M. V., Jacobsen, K. W., & Nørskov, J. K. (2003). Atomic and electronic structure of mos2 nanoparticles. *Physical Review B*, 67, 085410. <https://doi.org/10.1103/PhysRevB.67.085410>.
- Bollinger, M. V., Lauritsen, J. V., Jacobsen, K. W., Nørskov, J. K., Helveg, S., & Besenbacher, F. (2001). One-dimensional metallic edge states in mos2. *Physical Review Letters*. <https://doi.org/10.1103/PhysRevLett.87.196803>.
- Chhowalla, M., Shin, H. S., Eda, G., Li, L.-J., Loh, K. P., & Zhang, H. (2013). The chemistry of two-dimensional layered transition metal dichalcogenide nanosheets. *Nature Chemistry*, 5, 263–275. <https://doi.org/https://doi.org/10.1038/nchem.1589>.
- Choudhary, K., Garrity, K. F., Reid, A. C. E., DeCost, B., Biacchi, A. J., Trautt, Z., Wisinger, M., Clark, E., Joshi, K., Kumar, A., Chowdhury, S., Talapatra, A., & Tavazza, F. (2020). The joint automated repository for various integrated simulations (jarvis) for data-driven materials design. *npj Computational Materials*, 6, 173. <https://doi.org/10.1038/s41524-020-00440-1>.
- Damewood, J., Karaguesian, J., Lunger, J. R., Tan, A. R., Xie, M., Peng, J., & Gómez-Bombarelli, R. (2023). Representations of materials for machine learning. *Annual Review of Materials Research*, 53, 399–426. <https://doi.org/10.1146/annurev-matsci-080921-085947>.
- Dickinson, R. G. & Pauling, L. (1923). The crystal structure of molybdenite. *Journal of the American Chemical Society*, 45(6), 1466–1471. <https://doi.org/10.1021/ja01659a020>.
- Dumaz, M., Romero-Bohórquez, C., Adjero, D., & Romero, A. H. (2023). Topic modeling in density functional theory on citations of condensed matter electronic structure packages. *Scientific Reports*, 13, 11881. <https://doi.org/10.1038/s41598-023-3855>.
- G. Kresse, J. F. (1996). Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B*, 54, 11169–11186. <https://doi.org/D0I:https://doi.org/10.1103/PhysRevB.54.11169>.
- Grimme, S., Antony, J., Ehrlich, S., & Krieg, H. (1999). From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59, 1758–1775. <https://doi.org/https://doi.org/10.1103/PhysRevB.59.1758>.

- Haastrup, S., Strange, M., Pandey, M., Deilmann, T., Schmidt, P. S., Hinsche, N. F., Gjerding, M. N., Torelli, D., Larsen, P. M., Riis-Jensen, A. C., Gath, J., Jacobsen, K. W., Mortensen, J. J., Olsen, T., & Thygesen, K. S. (2018). The computational 2d materials database: high-throughput modeling and discovery of atomically thin crystals. *2D Materials*, 5(4), 042002. Topical Review, <https://doi.org/10.1088/2053-1583/aacfc1>.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*, (3a ed.). Elsevier Inc.
- Helveg, S., Lauritsen, J. V., Lægsgaard, E., Stensgaard, I., Nørskov, J. K., Clausen, B. S., Topsøe, H., & Besenbacher, F. (2000). Atomic-scale structure of single-layer mos2 nanoclusters. *Physical Review Letters*, 84, 951–954. <https://doi.org/10.1103/PhysRevLett.84.951>.
- Hohenberg, P. & Kohn, W. (1964). Inhomogeneous electron gas. *Physical Review Journals Archive*, 136, 864–871. <https://doi.org/https://doi.org/10.1103/PhysRev.136.B864>.
- Hoover, W. G. (1985). Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31, 1695. <https://doi.org/10.1103/PhysRevA.31.1695>.
- Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., & Persson, K. A. (2013). Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1), 011002. <https://doi.org/10.1063/1.4812323>.
- Kalidindi, S. R. & De Graef, M. (2015). Materials data science: Current status and future outlook. *Annual Review of Materials Research*, 45, 171–193. <https://doi.org/10.1146/annurev-matsci-070214-020844>.
- Kim, H. J., Chong, M., Rhee, T. G., Khim, Y. G., Jung, M.-H., Kim, Y. M., Jeong, H. Y., & Chang, B. K. C. . Y. J. (2023). Machine-learning-assisted analysis of transition metal dichalcogenide thin-film growth. *Nano Convergence*, 10, 10. <https://doi.org/10.1186/s40580-023-00359-5>.
- Kresse, G. & Furthmüller, J. (1996). Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6, 15–50. [https://doi.org/https://doi.org/10.1016/0927-0256\(96\)00008-0](https://doi.org/https://doi.org/10.1016/0927-0256(96)00008-0).
- Kresse, G. & Hafner, J. (1993). Ab initio molecular dynamics for liquid metals. *Physical Review B*, 47, 558–561. <https://doi.org/https://doi.org/10.1103/PhysRevB.47.558>.
- Kresse, G. & Joubert, D. (1999). From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59, 1758–1775. <https://doi.org/10.1103/PhysRevB.59.1758>.
- Kumar, A. & Ahluwalia, P. K. (2012). Electronic structure of transition metal dichalcogenides monolayers 1h-mx₂ (m = mo, w; x = s, se, te) from ab-initio theory: new direct band gap semiconductors. *The European Physical Journal B*, 85. <https://doi.org/https://doi.org/10.1140/epjb/e2012-30070-x>.
- Kumar, P., Sharma, V., Shirodkar, S. N., & Dev, P. (2022a). Predicting phase preferences of two-dimensional transition metal dichalcogenides using machine learning. *Phys. Rev. Mater.*, 6, 094007. <https://doi.org/10.1103/PhysRevMaterials.6.094007>.
- Kumar, U., Mishra, K. A., Kushwaha, A. K., & Cho, S. B. (2022b). Bandgap analysis of transition-metal dichalcogenide and oxide via machine learning approach. *Journal of Physics and Chemistry of Solids*, 171, 110973. <https://doi.org/https://doi.org/10.1016/j.jpcs.2022.110973>.

- Lauritsen, J., Kibsgaard, J., Helveg, S., Topsøe, H., Clausen, B. S., Lægsgaard, E., & Besenbacher, F. (2007). Size-dependent structure of MoS₂ nanocrystals. *Nature Nanotechnology*, 2, 53–58. <https://doi.org/10.1038/nnano.2006.171>.
- Lens, T. (2025). Search and analysis results for transition metal dichalcogenides. <https://www.lens.org/lens/search/scholar/analysis?q=Transition%20metal%20dichalcogenides&preview=true>. Consultado el 21 de marzo de 2025.
- Li, X., Chen, L., Cui, D., Jiang, W., Han, L., & Niu, N. (2022). Preparation and application of janus nanoparticles: Recent development and prospects. *Coordination Chemistry Reviews*, 454, 214318. <https://doi.org/https://doi.org/10.1016/j.ccr.2021.214318>.
- Li, X. & Zhu, H. (2015). Two-dimensional mos2: Properties, preparation, and applications. *Journal of Materiomics*, 1(1), 33–44. <https://doi.org/https://doi.org/10.1016/j.jmat.2015.03.003>.
- Lu, A.-Y., Xiao, H. Z. J., Chuu, C.-P., Han, Y., Chiu, M.-H., Cheng, C.-C., Yang, C.-W., Wei, K.-H., Yang, Y., Wang, Y., Sokaras, D., Nordlund, D., Yang, P., Muller, D. A., Chou, M.-Y., Zhang, X., & Li, L.-J. (2017). Janus monolayers of transition metal dichalcogenides. *Nature Nanotechnology*, 12, 744–749. <https://doi.org/https://doi.org/10.1038/nnano.2017.100>.
- Lu, G., Hu, Y., Fan, S., Liu, Z., Liu, C., Xu, J., Zheng, W., Zhang, J., & Liu, X. (2023). Synthesis of wse2 concentric nanotriangles for fully recoverable photoelectric gas sensors. *Sensors and Actuators B: Chemical*, 386, 133682. <https://doi.org/https://doi.org/10.1016/j.snb.2023.133682>.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA.
- Nguyen, D., Guerrero, J., & Hoat, D. (2023). Hfxo (x=s and se) janus monolayers as promising two-dimensional platforms for optoelectronic and spintronic applications. *Journal of Material Research*, 38, 2600–2612. <https://doi.org/https://doi.org/10.1557/s43578-023-00989-9>.
- Nosé, S. (1984). A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81, 511–519. <https://doi.org/10.1063/1.447334>.
- Novoselov, K. S., Geim, A. K., Morozov, S. V., Jiang, D., Zhang, Y., Dubonos, S. V., Grigorieva, I. V., & Firsov, A. A. (2004). Electric field effect in atomically thin carbon films. *Science*, 306(5696), 666–669. <https://doi.org/10.1126/science.1102896>.
- Ohno, K., Esfarjani, K., & Kawazoe, Y. (2018). *Computational Materials Science: From Ab Initio to Monte Carlo Methods*, (2a ed.). Springer.
- On, V. V., Nguyen, D. K., Guerrero-Sanchez, J., & Hoat, D. M. (2021). Exploring the electronic band gap of janus moseo and wseo monolayers and their heterostructures. *New J. Chem.*, 45, 20776–20786. <https://doi.org/https://doi.org/10.1039/C4RA17320A>.
- Paez-Ornelas, J. I., Ponce-Pérez, R., Fernández-Escamilla, H. N., Hoat, D. M., Murillo-Bracamontes, E. A., Moreno-Armenta, M. G., Galván, D. H., & Guerrero-Sánchez, J. (2021). The effect of shape and size in the stability of triangular janus mosse quantum dots. *Scientific Reports*, 11. <https://doi.org/https://doi.org/10.1038/s41598-021-00287-6>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://doi.org/https://doi.org/10.48550/arXiv.1201.0490>.

- Perdew, J. P., Burke, K., & Ernzerhof, M. (1997). Generalized gradient approximation made simple. *Physical Review Letters*, 78, 3865–3868. <https://doi.org/https://doi.org/10.1103/PhysRevLett.77.3865>.
- Perdew, J. P., Ernzerhof, M., & Burke, K. (1996). Rationale for mixing exact exchange with density functional approximations. *The Journal of Chemical Physics*, 105, 9982–9985. <https://doi.org/https://doi.org/10.1063/1.472933>.
- Rajan, A. C., Mishra, A., Satsangi, S., Vaish, R., Mizuseki, H., Lee, K.-R., & Singh, A. K. (2018). Machine-learning-assisted accurate band gap predictions of functionalized mxene. *Chemistry of Materials*, 30(12), 4031–4038. <https://doi.org/10.1021/acs.chemmater.8b00686>.
- Rajan, K. (2005). Materials informatics. *Materials Today*, 8(10), 38–45. [https://doi.org/https://doi.org/10.1016/S1369-7021\(05\)71123-8](https://doi.org/https://doi.org/10.1016/S1369-7021(05)71123-8).
- Ramos, M., Nogan, J., Ortíz-Díaz, M., Enriquez-Carrejo, J. L., Rodríguez-González, C. A., Mireles-Jr-García, J., Ambrosio-Lazáro, R. C., Ornelas, C., Hurtado-Macias, A., Boll, T., Chassaing, D., & Heilmaier, M. (2019). Mos2 thin films for photo-voltaic applications. *IntechOpen*, 6. <https://doi.org/10.5772/intechopen.83512>.
- Russell, S. J. & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*, (3a ed.). Pearson Education, Inc.
- Schaefer, H. E. (2010). *Nanoscience: The science of the small in physics, engineering, Chemistry, Biology and Medicine*, (2a ed.). Springer-Verlag.
- Sham, L. J. & Kohn, W. (1965). Self-consistent equations including exchange and correlation effects. *Physical Review*, 140, 1133–1138. <https://doi.org/https://doi.org/10.1103/PhysRev.140.A1133>.
- Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, (2a ed.). Dover Publications, Inc.
- Thejo Kalyani, N., Dhoble, S. J., Michalska-Domańska, M., Vengadaesvaran, B., Nagabhushana, H., & Arof, A. K. (2023). Index. In *Quantum Dots*, Woodhead Publishing Series in Electronic and Optical Materials, (pp. 581–590). Woodhead Publishing.
- Tong, X., Ashalley, E., Lin, F., Li, H., & Wang, Z. M. (2015). Advances in mos2-based field effect transistors (fets). *Nano-Micro Letters*. <https://doi.org/https://doi.org/10.1007/s40820-015-0034-8>.
- Tuxen, A., Kibsgaard, J., Gøbel, H., Lægsgaard, E., Topsøe, H., Lauritsen, J. V., & Besenbacher, F. (2010). Size threshold in the dibenzothiophene adsorption on mos2 nanoclusters. *ACS NANO*. <https://doi.org/10.1021/nn1011013>.
- van Mourik, T., Bühl, M., & Gageot, M.-P. (2014). Density functional theory across chemistry, physics and biology. *Philosophical Transactions of the Royal Society A*, 372, 20120488. <https://doi.org/10.1098/rsta.2012.0488>.
- VanderPlas, J. (2022). *Python Data Science Handbook*, (2a ed.). O'Reilly Media, Inc.
- Voiry, D., Yang, J., & Chhowalla, M. (2016). Recent strategies for improving the catalytic activity of 2d tmd nanosheets toward the hydrogen evolution. *Advanced Materials*, 28, 6197–6206. <https://doi.org/https://doi.org/10.1002/adma.201505597>.

- Waheed, H. S., Ullah, H., Iqbal, M. W., & Shin, Y.-H. (2022). Optoelectronic and photocatalytic properties of mo-based janus monolayers for solar cell applications. *Optik*, 271. <https://doi.org/https://doi.org/10.1016/j.ijleo.2022.170071>.
- Wang, H.-C., Schmidt, J., Marques, M. A. L., Wirtz, L., & Romero, A. H. (2023). Symmetry-based computational search for novel binary and ternary 2d materials. *2D Materials*, 10(3), 035007. Open access, <https://doi.org/10.1088/2053-1583/accc43>.
- Wang, Q., Zhang, Z., & Lee, Y.-W. (2018). Transition metal dichalcogenides: Synthesis, properties, and applications. *Nature Materials*, 15, 842–850. <https://doi.org/10.1038/nmat4749>.
- Wang, T., Su, M., Jin, H., Li, J., Wan, L., & Wei, Y. (2020). Optical, electronic, and contact properties of janus-moso/mos2 heterojunction. *The Journal of Physical Chemistry C*, 124, 15988–15994. <https://doi.org/https://doi.org/10.1021/acs.jpcc.0c04042>.
- Ward, L., Agrawal, A., Choudhary, A., & Wolverton, C. (2016). A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials*, 2, 16028. <https://doi.org/10.1038/npjcompumats.2016.28>.
- Ward, L., Dunn, A., Faghaninia, A., Zimmermann, N. E., Bajaj, S., Wang, Q., Montoya, J., Chen, J., Bystrom, K., Dylla, M., Chard, K., Asta, M., Persson, K. A., Snyder, G. J., Foster, I., & Jain, A. (2018). Matminer: An open source toolkit for materials data mining. *Computational Materials Science*, 152, 60–69. <https://doi.org/https://doi.org/10.1016/j.commatsci.2018.05.018>.
- Wegner, K. D. & Resch-Genger, U. (2024). The 2023 nobel prize in chemistry: Quantum dots. *Analytical and Bioanalytical Chemistry*, 416, 3283,3293. <https://doi.org/https://doi.org/10.1007/s00216-024-05225-9>.
- Wei, J., Chu, X., Sun, X.-Y., Xu, K., Deng, H.-X., Chen, J., Wei, Z., & Lei, M. (2019). Machine learning in materials science. *InfoMat*, 1(3), 338–358. <https://doi.org/https://doi.org/10.1002/inf2.12028>.
- Wilson, J. A. & Yoffe, A. D. (1969). The transition metal dichalcogenides: discussion and interpretation of the observed optical, electrical and structural properties. *Advances in Physics*, 18(73), 193–335. <https://doi.org/10.1080/00018736900101307>.
- Wua, Z., Li, L., Liao, T., Chena, X., Jianga, W., Luoa, W., Yanga, J., & Sun, Z. (2018). Janus nanoarchitectures: From structural design to catalytic applications. *Nano Today*, 22, 62–82. <https://doi.org/https://doi.org/10.1016/j.nantod.2018.08.009>.
- Yagmurcukardes, M. & Peeters, F. M. (2020). Stable single layer of janus moso: Strong out-of-plane piezoelectricity. *Physical Review B*, 101, 155205. <https://doi.org/10.1103/PhysRevB.101.155205>.
- Zhang, J., Jia, S., Kholmanov, I., Dong, L., Er, D., Chen, W., Guo, H., Jin, Z., Shenoy, V. B., Shi, L., & Lou, J. (2017). Janus monolayer transition-metal dichalcogenides. *ACS Nano*, 11(8), 8192–8198. <https://doi.org/10.1021/acsnano.7b03186>.
- Zhang, L. & Sun, X. (2019). Janus transition metal dichalcogenides: Properties and potential applications. *Advanced Functional Materials*, 29(13), 1808380. <https://doi.org/10.1002/adfm.201808380>.
- Zhang, X., Teng, S. Y., Loy, A. C. M., How, B. S., Leong, W. D., & Tao, X. (2020). Transition metal dichalcogenides for the application of pollution reduction: A review. *Nanomaterials*, 10(6). <https://doi.org/10.3390/nano10061012>.

Anexos

Tabla 4. Nanotriángulos Janus simulados que formaron la BD de ejemplo en total se obtuvieron 172 nanotriángulos del modelo α y 172 para el modelo β . Para cada fórmula general MX_Y se indica el rango de n .

Fórmula	n	Fórmula	n
MoSO	5–9	WSO	4–10
MoSeO	4–10	WSeO	4–10
MoSSe	4–10	WSSe	4–10
MoTeO	4–10	WTeO	4–10
MoSeTe	4–10	WSTe	4–10
MoSTe	4–10	WSeTe	4–10
NbSO	4	TiSO	4
NbSeO	4	TiSeO	4
NbTeO	4	TiTeO	4
NbSSe	4	TiSSe	4
NbSTe	4	TiSTe	4
NbSeTe	4	TiSeTe	4
TaSO	4	CrSO	4
TaSeO	4	CrSeO	4
TaTeO	4	CrTeO	4
TaSSe	4	CrSSe	4
TaSTe	4	CrSTe	4
TaSeTe	4	CrSeTe	4
HfSO	4	RuSO	4
HfSeO	4	RuSeO	4
HfTeO	4	RuTeO	4
HfSSe	4	RuSSe	4
HfSTe	4	RuSTe	4
HfSeTe	4	RuSeTe	4
PdSO	4	AgSO	4
PdSeO	4	AgSeO	4
PdTeO	4	AgTeO	4
PdSSe	4	AgSSe	4
PdSTe	4	AgSTe	4
PdSeTe	4	AgSeTe	4
ZrSO	4	VSO	4
ZrSeO	4	VSeO	4
ZrTeO	4	VTeO	4
ZrSSe	4	VSSe	4
ZrSTe	4	VSTe	4
ZrSeTe	4	VSeTe	4
ScSO	4	MnSO	4
ScSeO	4	MnSeO	4
ScTeO	4	MnTeO	4
ScSSe	4	MnSSe	4
ScSTe	4	MnSTe	4
ScSeTe	4	MnSeTe	4

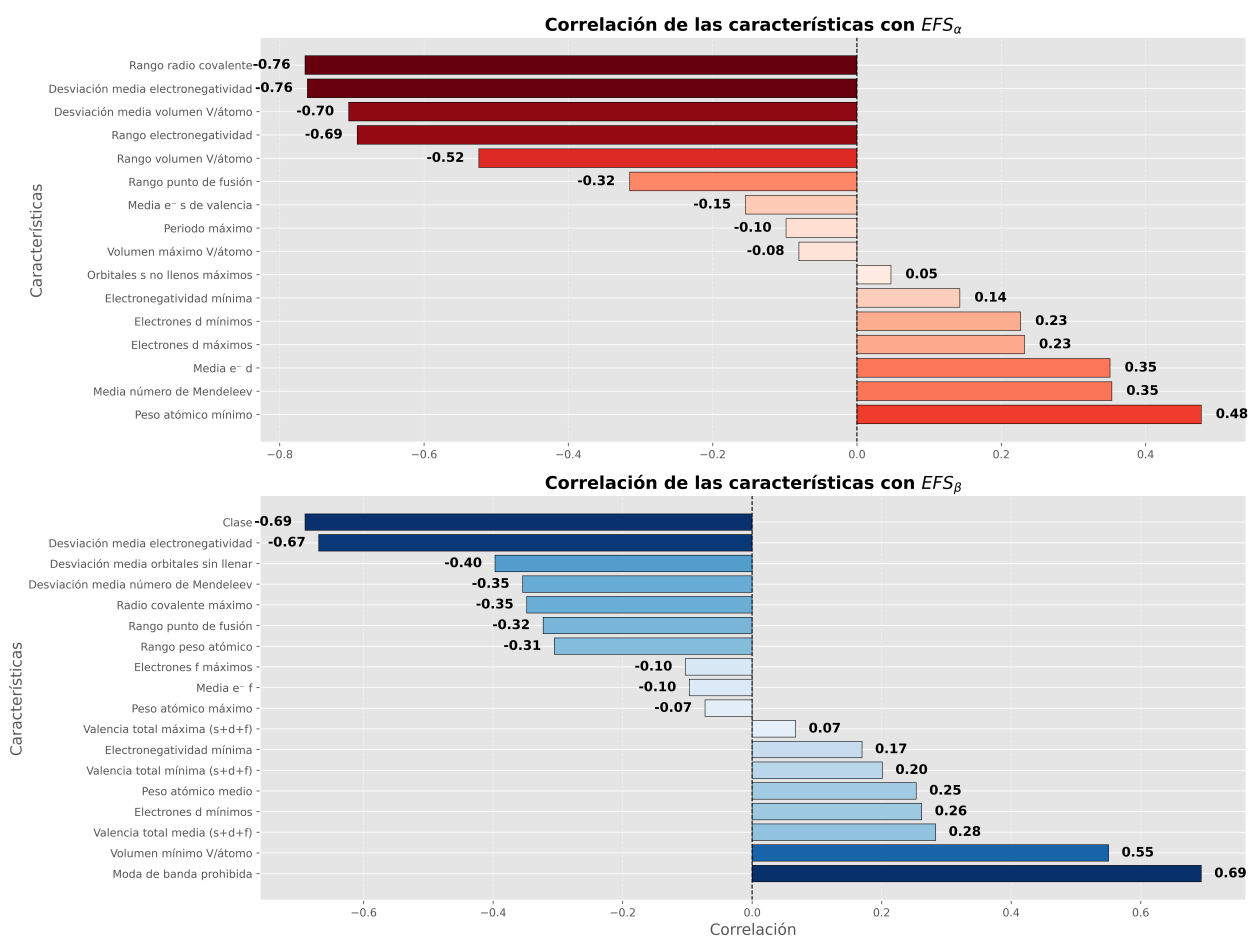


Figura 22. Coeficientes de correlación de Pearson entre las características seleccionadas y la EFS para los modelos de borde α (escala de rojos) y β (escala de azules). En el modelo α la mayor correlación positiva corresponde al mínimo peso atómico ($r = 0.48$) y la mayor correlación negativa al rango del radio covalente ($r = -0.76$). En el panel b la característica con correlación más alta es la moda de la banda prohibida ($r = 0.69$), mientras que la variable «Clase» (óxido vs. no óxido) presenta la asociación negativa más fuerte ($r = -0.69$).

Tabla 8. Hiperparámetros optimizados para cada algoritmo de ML de los modelos de borde α y β de acuerdo a la nomenclatura usada en la librería *scikit-learn* desarrollada por Pedregosa et al. (2011)

Algoritmo	Modelo α	Modelo β
Ridge	alpha = 2.25	alpha = 0.5
Lasso	alpha = 0.001, tol = 0.1	alpha = 0.001, tol = 0.1
Elastic Net	alpha = 0.1, l1_ratio = 0.01, tol = 0.1	alpha = 0.001, l1_ratio = 0.5, tol = 0.1
SVR (RBF)	epsilon = 0.03, gamma = 0.01, kernel = rbf, tol = 0.1	epsilon = 0.02, gamma = 0.05, kernel = rbf, tol = 0.0001
Nu-SVR	gamma = 0.01, kernel = rbf, nu = 1, tol = 1e-6	gamma = scale, kernel = rbf, nu = 0.5, tol = 0.001
Árbol de decisión	max_depth = 9, max_features = 0.75, min_samples_leaf = 1, min_samples_split = 4, splitter = random	max_depth = 8, max_features = 0.34, min_samples_leaf = 1, min_samples_split = 8, splitter = random
Random Forest	criterion = absolute_error, max_depth = 11, max_features = sqrt, min_samples_split = 11, n_estimators = 6	criterion = friedman_mse, max_depth = 8, max_features = 0.45, min_samples_split = 10, n_estimators = 2
Extra Trees	criterion = absolute_error, max_depth = none, max_features = sqrt, min_samples_leaf = 1, min_samples_split = 8, n_estimators = 40	criterion = absolute_error, max_depth = 10, max_features = sqrt, min_samples_leaf = 1, min_samples_split = 10, n_estimators = 58
MLP	activation = identity, alpha = 0.12, hidden_layer_sizes = (41,), learning_rate = constant, learning_rate_init = 1e-4, solver = lbfgs	activation = tanh, alpha = 0.158, hidden_layer_sizes = (21,), learning_rate = constant, learning_rate_init = 1e-3, solver = adam
AdaBoost	base_estimator = DecisionTree, learning_rate = 0.001, loss = square, n_estimators = 69, random_state = 42	base_estimator = SVR, learning_rate = 0.004, loss = linear, n_estimators = 4, random_state = 42
Gradient Boosting	criterion = friedman_mse, learning_rate = 0.25, loss = absolute_error, min_samples_split = 0.325, n_estimators = 400	criterion = friedman_mse, learning_rate = 0.5, loss = absolute_error, min_samples_split = 0.775, n_estimators = 100

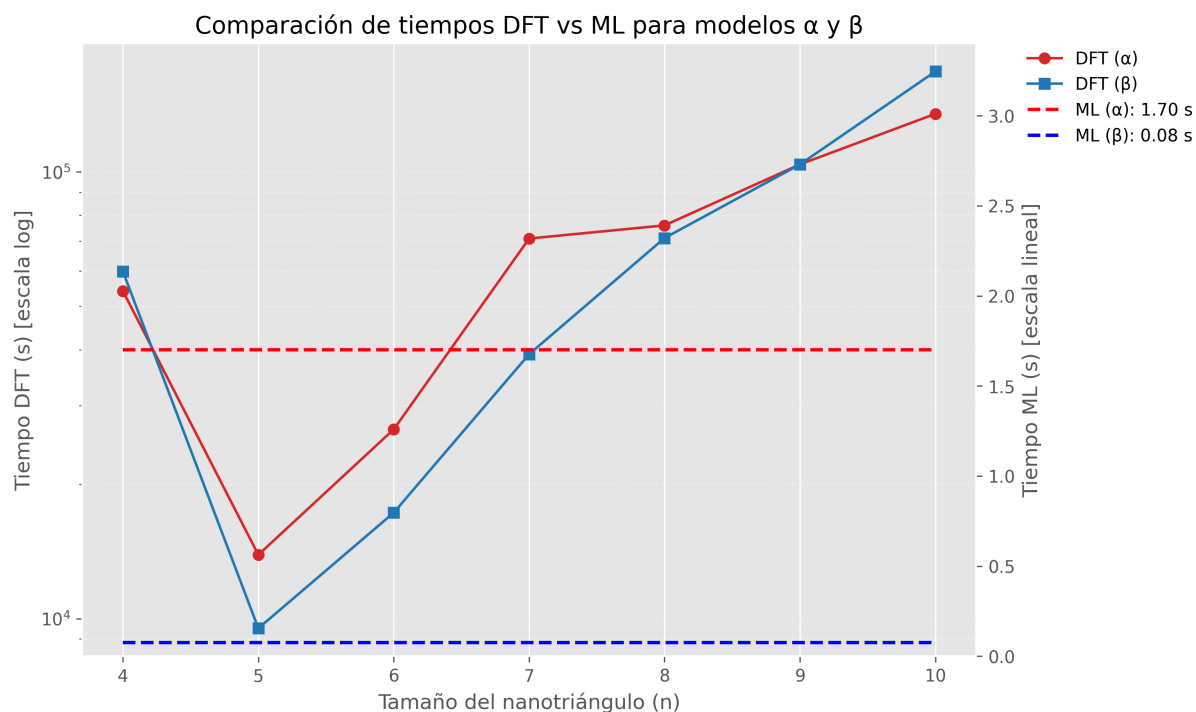


Figura 23. Comparación de los tiempos promedio de simulación DFT frente al coste computacional de los modelos de ML para los nanotriángulos Janus. Cada marcador (círculo rojo para α , cuadrado azul para β) indica el tiempo medio de un cálculo DFT en función de n , el número de átomos de metales de transición en el borde exterior del nanotriángulo. Las líneas rojas y azules continuas (escala logarítmica, eje izquierdo) muestran cómo crece el tiempo de DFT al aumentar n . Las líneas punteadas horizontales en rojo (α) y azul marino (β), trazadas en escala lineal (eje derecho), representan el tiempo total de entrenar y evaluar cada modelo de ML (validación cruzada 5-fold) para α y β , respectivamente, evidenciando la drástica reducción de tiempo al usar ML en lugar de DFT.

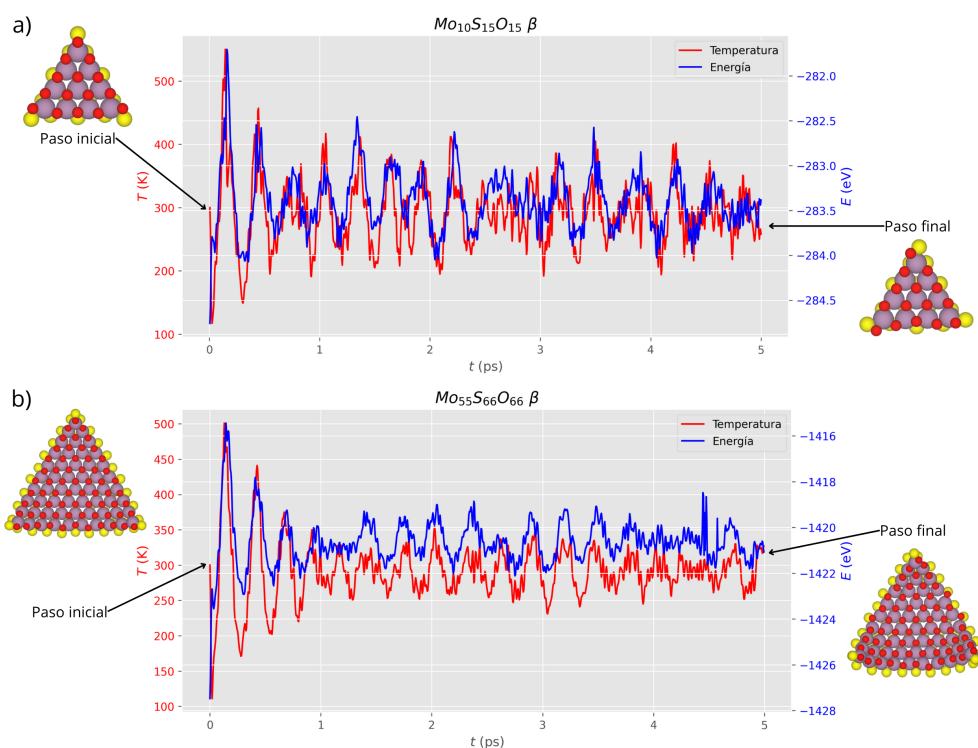


Figura 24. Resultados de las simulaciones AIMD en ensamble NVT a 300 K para nanotriángulos Janus MoSO en el modelo de borde β . Se muestran dos tamaños distintos: $n = 4$ (panel superior) y $n = 10$ (panel inferior). En cada panel, la curva roja indica la evolución de la temperatura (eje izquierdo) y la curva azul la evolución de la energía total (eje derecho) durante 5 ps de simulación. A la derecha e izquierda de cada gráfico aparecen las configuraciones atómicas inicial y final respectivamente, demostrando que, pese a las oscilaciones térmicas y energéticas, la forma triangular se conserva y las magnitudes oscilan alrededor de valores medios estables.

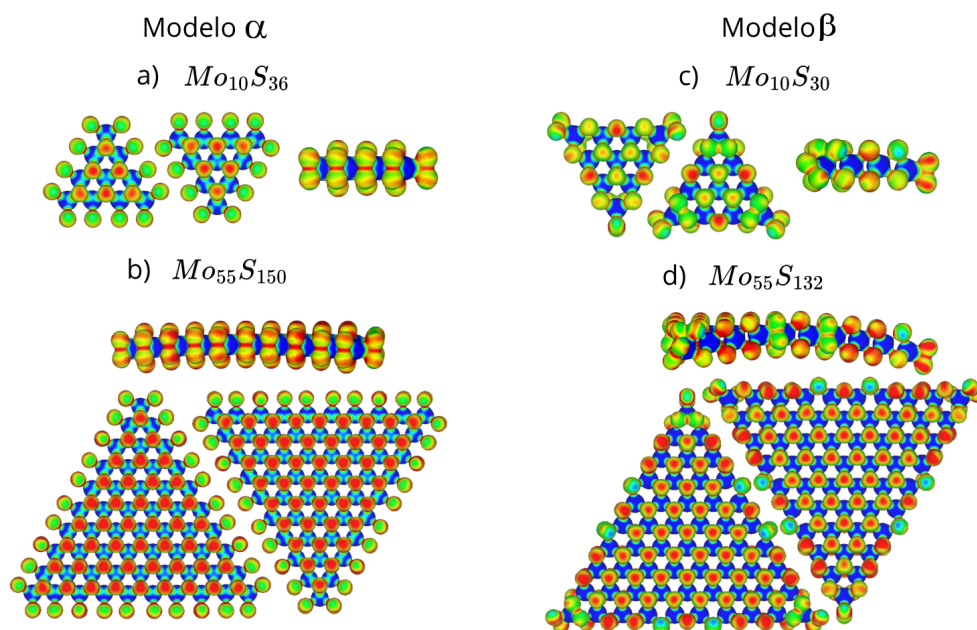


Figura 25. Isosuperficies de potencial electrostático para nanotriángulos prístinos de MoS_2 en dos modelos de borde: α (paneles a y b) y β (paneles c y d), para tamaños $n = 4$ y $n = 10$. El código de colores indica potencial negativo (exceso de electrones, en rojo) y potencial positivo (déficit de electrones, en azul). En ambos tamaños, el modelo α muestra una clara localización de potencial positivo en los bordes, lo que indica que estos sitios son los más activos y, por tanto, los más susceptibles a procesos como arranque de átomos o catálisis, mientras que el centro (plano basal) exhibe una mayor acumulación de carga negativa (potencial negativo). Para la geometría β , se observa una distribución más uniforme del potencial, con una mayor presencia de exceso de electrones en toda la superficie y menos diferenciación en los bordes, sugiriendo una menor actividad relativa en comparación con el modelo α . El aumento de tamaño del nanotriángulo refuerza estos patrones, intensificando la diferenciación entre bordes y superficies basales en ambos modelos.

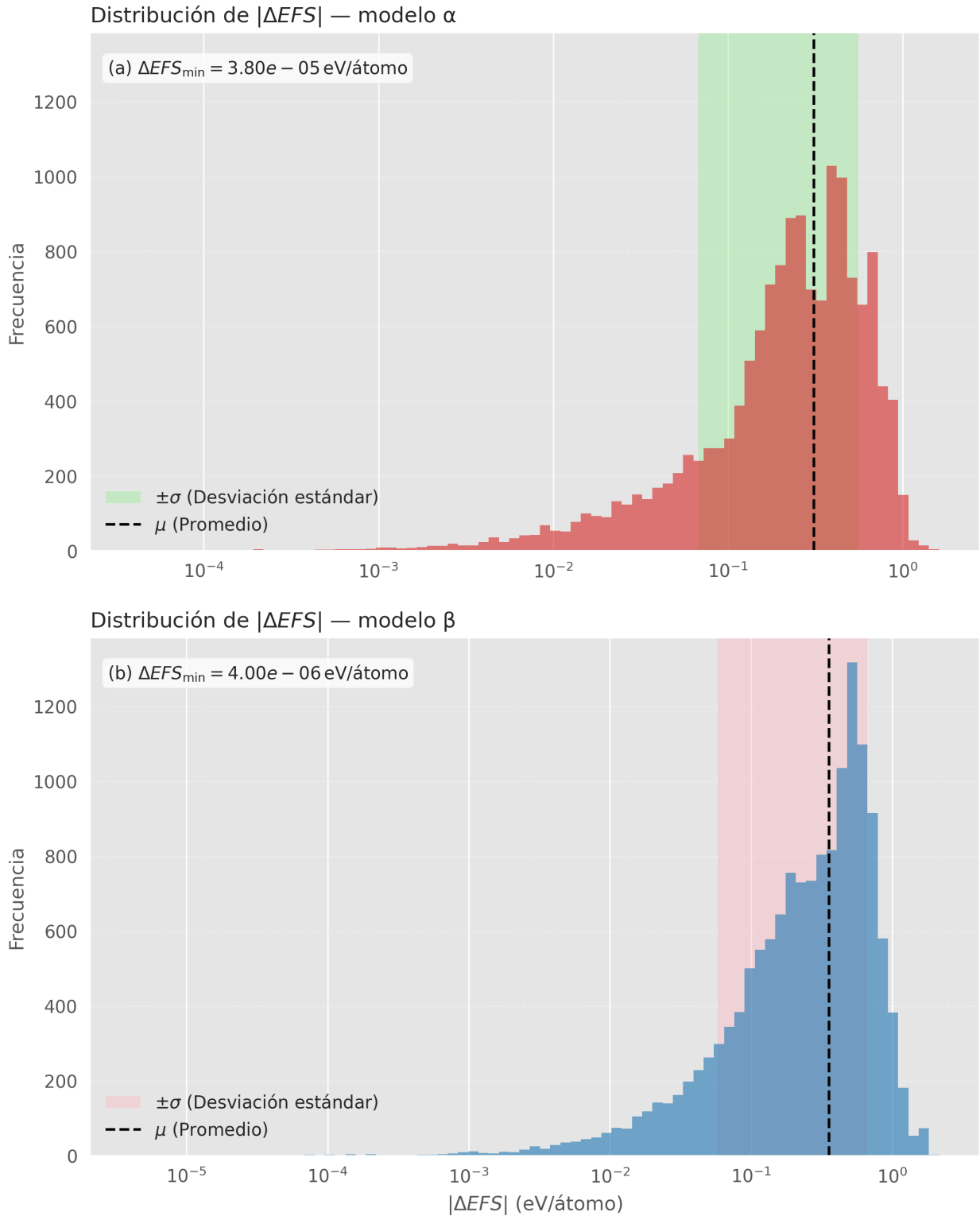


Figura 19. Distribución de los valores absolutos de las diferencias en la energía de formación de superficie (ΔEFS) entre pares de predicciones para los modelos de borde α (panel superior, color rojo) y β (panel inferior, color azul). Ambas distribuciones se presentan en escala logarítmica; las franjas sombreadas muestran $\pm\sigma$ (desviación estándar) y las líneas verticales discontinuas señalan la media μ . La mayoría de las diferencias se concentra en el rango de 10^{-1} a 10^0 eV/átomo. Los valores mínimos de diferencia son $\Delta EFS_{\min}^{\alpha} = 3.80 \times 10^{-5}$ eV/átomo y $\Delta EFS_{\min}^{\beta} = 4.00 \times 10^{-6}$ eV/átomo.

Tabla 5. Nanotriángulos con energías de formación de superficie positivas (formación no favorable según los cálculos de DFT), estos datos pertenecen al conjunto de ejemplos para entrenar los algoritmos de ML.

Nanotriángulo	Tamaño
<i>Modelo α</i>	
PdSTe	$n = 4$
PdSeTe	$n = 4$
MnSSe	$n = 4$
AgSSe	$n = 4$
<i>Modelo β</i>	
NiSSe	$n = 4$
CoSSe	$n = 4$
RuSSe	$n = 4$
AgSSe	$n = 4$
AgSTe	$n = 4$
AgSeTe	$n = 4$

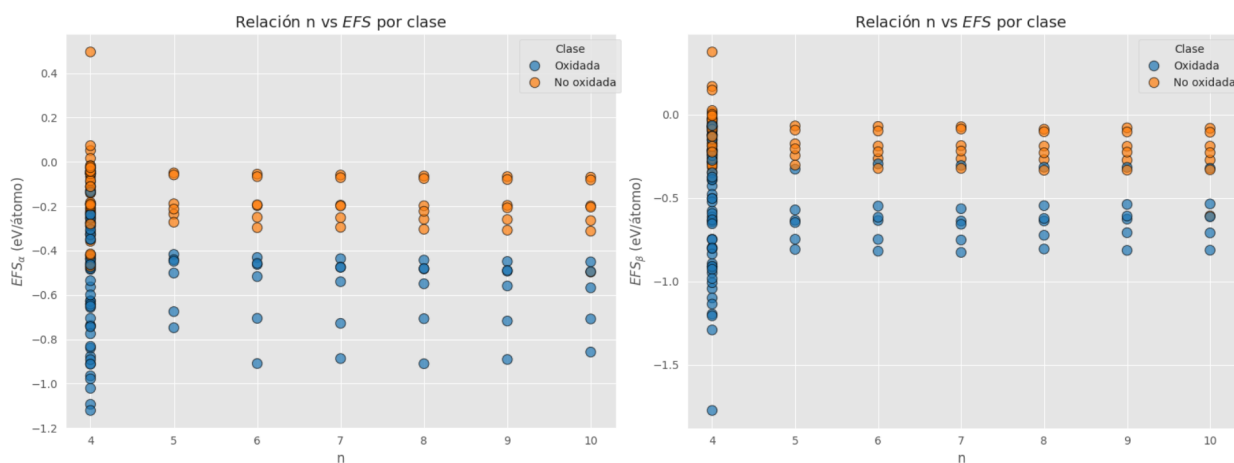


Figura 20. Relación entre el número de átomos de metal de transición en el borde (n) y la EFS para los modelos α (izquierda) y β (derecha). Los puntos están coloreados según la clase: oxidados (azul) y no oxidados (naranja). Al aumentar n , las EFS tienden a valores más negativos (mayor estabilidad), especialmente en las estructuras oxidadas.

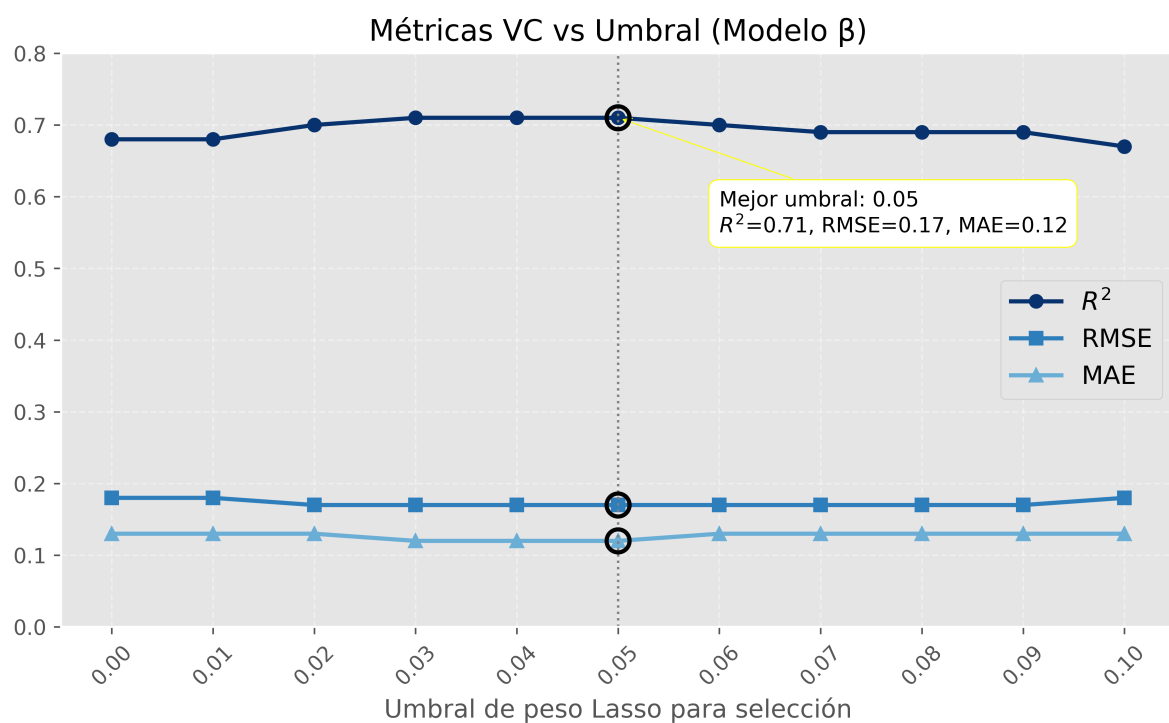
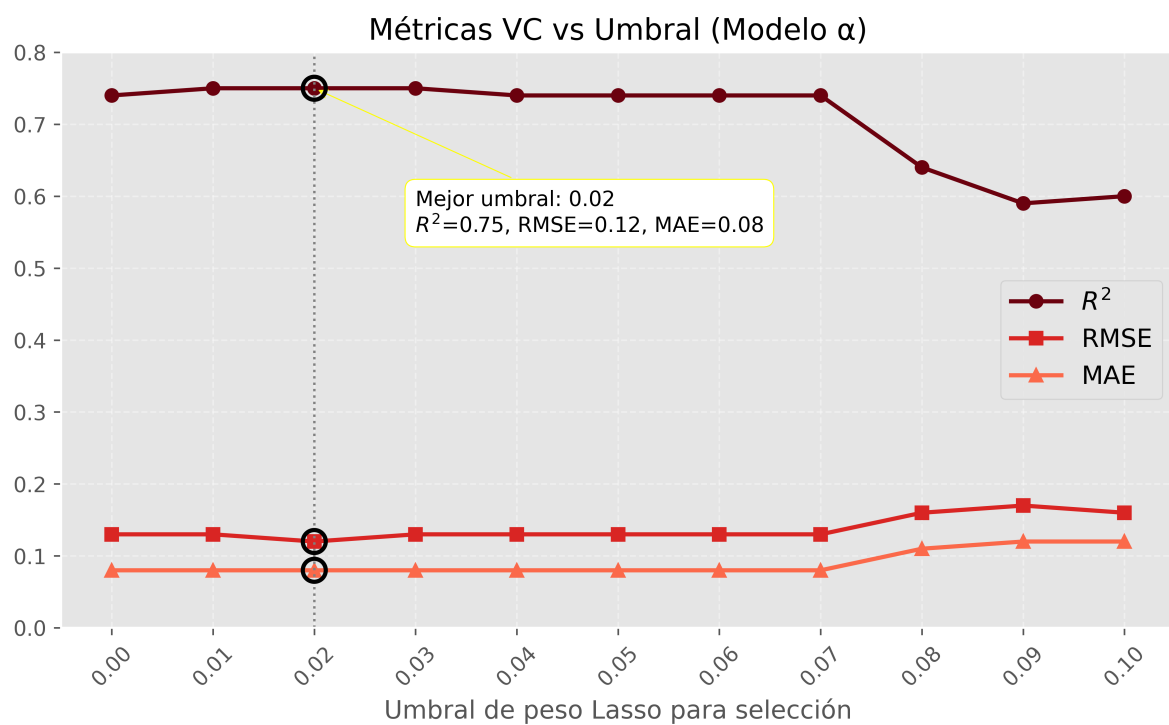


Figura 21. Comparativa de las métricas de validación cruzada (R^2 , RMSE y MAE) promedio de 5 pliegues para distintos umbrales de selección con Lasso en los modelos α y β . Los umbrales óptimos se destacan con anotaciones y líneas verticales.

Tabla 6. Descripción de las características utilizadas en el modelo α .

Característica	Descripción
<i>Media número de Mendeleev</i>	Media del número de Mendeleev (posición en la tabla periódica) de los elementos que componen el nanotriángulo.
<i>Peso atómico mínimo</i>	Peso atómico mínimo entre los elementos de la composición.
<i>Rango punto de fusión</i>	Rango (diferencia entre máximo y mínimo) de los puntos de fusión de los elementos.
<i>Periodo máximo</i>	Número de periodo (fila de la tabla periódica) más alto entre los elementos.
<i>Rango radio covalente</i>	Diferencia entre los radios covalentes máximo y mínimo.
<i>Electronegatividad mínima</i>	Electronegatividad de Pauling mínima entre los elementos.
<i>Rango electronegatividad</i>	Rango de electronegatividades (máxima – mínima).
<i>Desviación media electronegatividad</i>	Desviación media de las electronegatividades.
<i>Media electrones s de valencia</i>	Valor medio del número de electrones de valencia en orbitales <i>s</i> .
<i>Electrones d mínimos</i>	Número mínimo de electrones de valencia en orbitales <i>d</i> .
<i>Electrones d máximos</i>	Número máximo de electrones de valencia en orbitales <i>d</i> .
<i>Media electrones d</i>	Valor medio de electrones de valencia <i>d</i> .
<i>Orbitales s no llenos máximos</i>	Número máximo de orbitales <i>s</i> no llenos entre los elementos.
<i>Volumen máximo V/átomo</i>	Volumen de celda unidad en estado base por átomo (valor máximo).
<i>Rango volumen V/átomo</i>	Rango de volúmenes en estado base por átomo (máximo–mínimo).
<i>Desviación media volumen V/átomo</i>	Desviación media de los volúmenes en estado base por átomo.

Tabla 7. Descripción de las características utilizadas en el modelo β .

Característica	Descripción
<i>Desviación media número de Mendeleev</i>	Desviación media del número de Mendeleev de los elementos.
<i>Peso atómico máximo</i>	Peso atómico máximo entre los elementos de la composición.
<i>Rango peso atómico</i>	Rango de pesos atómicos (máximo – mínimo).
<i>Peso atómico medio</i>	Peso atómico medio de los elementos.
<i>Rango punto de fusión</i>	Rango de puntos de fusión (máximo – mínimo).
<i>Radio covalente máximo</i>	Radio covalente máximo entre los elementos.
<i>Electronegatividad mínima</i>	Electronegatividad mínima de Pauling.
<i>Desviación media electronegatividad</i>	Desviación media de las electronegatividades.
<i>Electrones d mínimos</i>	Número mínimo de electrones de valencia <i>d</i> .
<i>Electrones f máximos</i>	Número máximo de electrones de valencia <i>f</i> .
<i>Media electrones f</i>	Valor medio de electrones de valencia <i>f</i> .
<i>Valencia total mínima ($s+d+f$)</i>	Número mínimo total de electrones de valencia ($s + d + f$).
<i>Valencia total máxima ($s+d+f$)</i>	Número máximo total de electrones de valencia ($s + d + f$).
<i>Valencia total media ($s+d+f$)</i>	Valor medio total de electrones de valencia ($s + d + f$).
<i>Desviación media orbitales sin llenar</i>	Desviación media del número de orbitales no llenos.
<i>Volumen base V/átomo</i>	Volumen en estado base por átomo (valor mínimo).
<i>Moda de banda prohibida</i>	Moda (valor más frecuente) de la banda prohibida en estado base de los elementos.