

Universidad Peruana Cayetano Heredia



FACULTAD DE CIENCIAS Y FILOSOFÍA

Construcción de modelos de aprendizaje automático químicamente interpretativos en la clasificación de moléculas de colorantes naturales de celdas solares sensibilizadas por tintes basados en relaciones estructura-propiedad

Proyecto de tesis

Autor:

Bach. Dan Santivañez Gutarra

Asesora:

Dra. María Quintana Caceda

Co-Asesores:

Dr. Anthony Huamán Aguirre

Lima - Perú

2022

Índice

1. Introducción	3
1.1. Contexto	3
1.2. Celdas solares	4
1.3. Celdas solares sensibilizadas por tintes (CSPT)	5
1.4. Medidas del rendimiento de CSPT	7
1.5. Optimización de las CSPT	8
1.6. Tintes para CSPT	9
1.7. Problemática de los tintes naturales para CSPT	9
1.8. Enfoque de investigación direccionado por datos o <i>Data-Driven</i>	9
1.9. Aprendizaje automático aplicado a la química	10
1.10. Limitaciones del uso del aprendizaje automático en el desarrollo de materiales	11
1.11. Problemática	11
1.12. Justificación	12
1.13. Antecedentes	12
1.13.1. Internacionales	13
1.13.2. Nacionales	13
2. Hipótesis de investigación y objetivos	13
2.1. Hipótesis de investigación	13
2.2. Objetivos	13
3. Materiales y métodos	13
3.1. Análisis estadístico	13
3.1.1. Pruebas de normalidad	13
3.1.2. Extracción de datos	13
3.1.3. Preprocesamiento de datos	13
3.1.4. Descubrimiento y análisis de características	13
3.1.5. Entrenamiento y evaluación del modelo 1	13
3.1.6. Entrenamiento y evaluación del modelo 2	13
3.1.7. Entrenamiento y evaluación del modelo 3	13
3.1.8. Comparación de modelos	13
4. Referencias bibliográficas	13
5. Cronograma de trabajo	16
6. Presupuesto y financiamiento	16

Resumen

Las celdas solares son dispositivos que convierten energía lumínica en energía eléctrica útil a través de fenómenos electroquímicos. El desarrollo de materiales sostenibles y con mayor eficiencia es un reto de investigación y también ambiental, ya que resolverlo conlleva explorar miles de millones de compuestos y hallar dicho tipos de materiales favorecerían el consumo de energías limpias. Esta exploración requiere recursos humanos y materiales que muchos investigadores no pueden proporcionarse, es entonces que las simulaciones computacionales y el enfoque dirigido por datos cobran importancia. El uso de experimentos *in silico* han ...

1. Introducción

1.1. Contexto

La dependencia de los combustibles fósiles traería consigo problemas como la inseguridad energética y el incremento de la temperatura global a causa del aumento de los gases de efecto invernadero. Las sociedades que concentran la producción de los combustibles fósiles podrían controlar geopolítica y económicamente a otras [1], esta inseguridad energética se evidenció en la crisis del petróleo de 1973 [2] y en la disputa entre Rusia y la Unión Europea por el suministro de gas en este año [3]. En contraste, el problema del incremento de la temperatura global es un problema mediáticamente más discreto pero no menos alarmante. Tanto así que, en el 2018, la *Intergovernmental Panel on Climate Change* (IPCC) emitió un informe sobre los impactos que causaría dicho incremento en 1.5°C con respecto a los niveles preindustriales para el 2040, que en resumidas cuentas, se prevé un detrimento crítico y sin retorno de la civilización y la biósfera [4]. A fin de hacer frente a estos problemas, la transición hacia una matriz energética mundial donde predominen fuentes energéticas menos contaminantes y descentralizadas es la solución.

Las fuentes renovables reúnen dichas características, por lo que muchos gobiernos y organizaciones han tomado acciones para aprovecharlas. En ese marco, la *International Renewable Energy Agency* (IRENA) realizó un análisis multisectorial en el 2020 donde propuso una hoja de ruta para que las energías renovables generen el 86 % de la electricidad global [5]. En ese documento también se menciona que esta cifra no se alcanzaría sin la investigación ni el desarrollo de tecnologías que aprovechen dichas fuentes para hacerlas sostenibles y comercialmente viables.

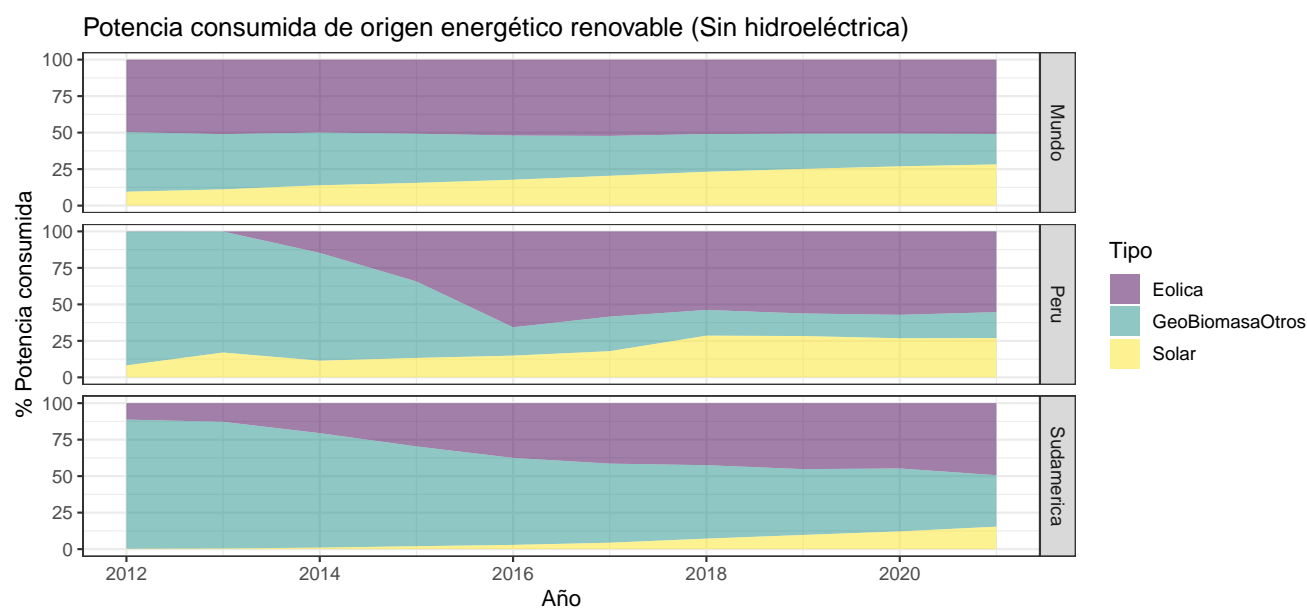


Figura 1: Comparación del porcentaje de la cantidad de potencia consumida por tipo de energía renovable sin considerar la hidroeléctrica. Fuente: Our World in Data [6]

Hace un año, la energía hidroeléctrica generó más de 4 mil tera-vatios por hora (TWh) en el mundo, lo que representó más del 50 % de la generación de todas las renovables [7]. No obstante, no es posible extender la construcción de centrales hidroeléctricas dado que muchas comunidades

no cuentan cuencas hidrográficas o caídas de agua que puedan aprovecharse. Sin contar esta fuente, las energías eólica y solar son las que encabezan la producción energética. Con el propósito de contextualizar el crecimiento del uso de estas fuentes renovables geográfica y temporalmente se muestra la Figura 1. En ella se visualizan los porcentajes de potencia consumida provenientes de tres grupos de fuentes renovables no hidroeléctricas: La eólica, solar y otras (*GeoBiomasaOtros*). Estos porcentajes se distribuyen desde el 2012 hasta el 2021 y se comparan las tendencias en el Perú, Sudamérica y el mundo. Se aprecia el aumento del consumo energético de la eólica y solar sobre las demás, y de entre estas dos, la eólica es mayor. No obstante, también es evidente que la energía solar ha tenido un crecimiento sostenido en el tiempo y esto se debe, entre otras razones, a la disponibilidad de nuevas tecnologías más comercial y ambientalmente más viables que aprovechan la luz solar.

Existen dos tipos de tecnologías solares: La termosolar y la fotovoltaica [8]. En la primera se usa la energía térmica de un fluido calentado por concentradores solares para generar electricidad por medio del movimiento de turbinas conectadas a generadores electromagnéticos. Pese a los esfuerzos por reducir los costos de construcción y mantenimiento de las centrales termosolares, solo países con alta demanda energética industrial y la geografía apta pueden hacer rentables estas tecnologías [9].

En contraste a la termosolar, en la fotovoltaica se aprovecha directamente la radiación solar mediante fenómenos fotoeléctricos y de transporte de cargas, mimetizando la fotosíntesis. Si las plantas tienen a las células con clorofila en sus hojas como sus unidades generadoras de energía, en los sistemas fotovoltaicos estas serían las células o celdas solares. La estructura de una celda solar depende de los mecanismos de generación eléctrica y transporte de carga, y estos a su vez, de la naturaleza de los materiales donde suceden estos mecanismos. El interés por desarrollar tecnologías fotovoltaicas se da por dos razones: Por su versatilidad, es decir que pueden ser usadas para fines domésticos o industriales, y porque reducirían notablemente las emisiones de CuO_2/kWh , por ejemplo esta reducción sería cercana al 90 % para sistemas que sustituyan al gas natural [10].

1.2. Celdas solares

Las celdas solares llevan en el mercado más de 60 años, desde su descubrimiento en los Laboratorios Bell en 1954 [11] y su primer uso comercial en el satélite Vanguard [12] en 1958, estos se han diversificado en su composición, por ende, también en su eficiencia y estabilidad. Según la clasificación del *National Renewable Energy Laboratories* (NREL) los tipos de celdas se pueden agrupar en cinco familias: Cristalinas de silicio, de unión simple de galio (Ga) y arsénico (As), de unión múltiple, basadas en películas delgadas y emergentes [13]. Los puntos en común de estas familias son la similitud de sus componentes, arquitecturas y/o coetaneidad [14]. Cabe mencionar que en adelante solo se tomarán en cuenta tecnologías fotovoltaicas que no tengan concentradores.

En primer lugar se encuentra la familia de las celdas basadas en cristales de silicio que congrega a las celdas de monocristal [15], multicristal [16], con heteroestructuras de silicio [17] y con películas delgadas de silicio cristalino [18]. Segundo, las celdas de unión simple (Ga-As) en donde están las que se basan en cristales y las de películas delgadas [19]. Tercero, las celdas de unión múltiple que tienen celdas con arquitecturas con dos, tres y cuatro a más uniones [20, 21]. Cuarto, en las celdas basadas en películas delgadas están las celdas de cobre-indio-galio-selenio (CIGS) [22], cadmio-teluro (CdTe) [23] y silicio amorfo hidrogenado (a-Si:H) [24]. Y en quinto lugar, las celdas emergentes, que se caracteriza por agrupar las tecnologías que llevan menos tiempo siendo investigadas tales como las celdas sensibilizadas por tintes (CSPT) [25], las inorgánicas basadas en kesterita $\text{Cu}_2\text{ZnSn}(\text{S}, \text{Se})_4$ (CZTSSe) [26], las celdas orgánicas

[27], orgánicas tipo *sándwich* [28], las celdas perovskitas [29], perovskitas-CIGS monolítica tipo *sándwich* [30], perovskitas-SI monolítica tipo *sándwich* [30] y celdas con *quantum-dots* [31].

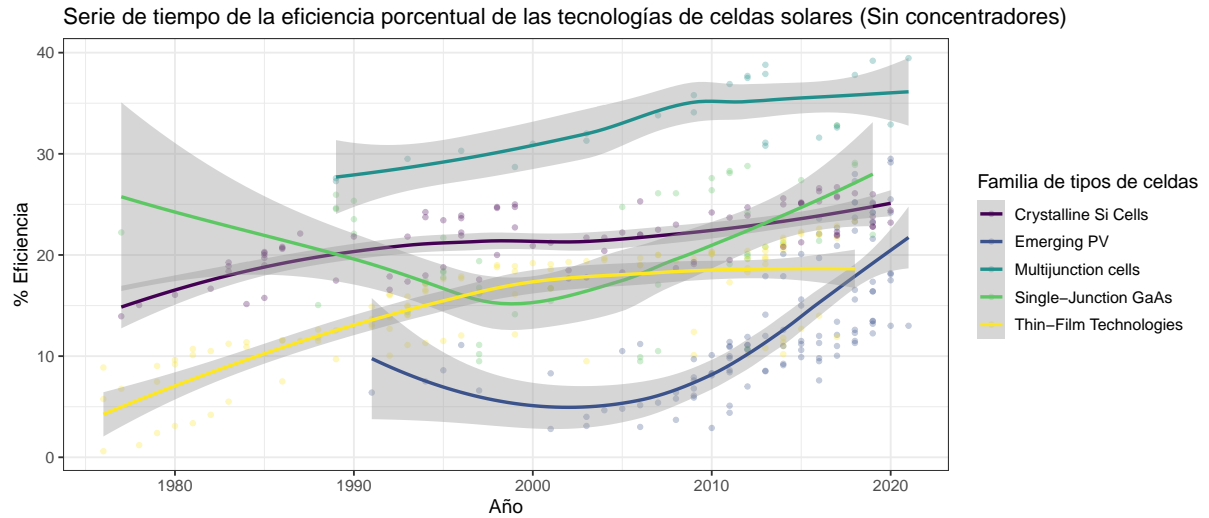


Figura 2: Serie de tiempo de las mejores eficiencias obtenidas en la investigación de las celdas solares según la agrupación dada por NREL. Fuente: *National Renewable Energy Laboratory* [13]

Para contrastar dichas familias de celdas se usaron las mejores eficiencias de conversión energética de estos dispositivos obtenidos en el laboratorio que fueron recopilados por la NREL entre los años de 1980 y 2021 (Figura 2). En esa gráfica se observa que las tecnologías con mayor rendimiento han sido las celdas de unión múltiple, esto se debe al efecto sinérgico entre uniones de las multicapas, no obstante, estas serían más costosas [32]. También se aprecia que hay menos celdas de silicio cristalino y las películas delgadas siendo investigadas. Este hecho puede deberse a que las tecnologías emergentes las están igualando, en términos de su performance en el laboratorio. Es por ello que es importante seguir investigándolas gracias a la sostenibilidad ambiental de sus materiales.

Con los datos del NREL se construyó una serie de tiempo de las mejores eficiencias que muestran en el laboratorio todos los tipos de celdas que pertenecen a las tecnologías emergentes (Figura 3). En ella se observa que las celdas basadas en perovskitas son las más recientes y las que mayor rendimiento tienen. La perovskita está compuesto por átomos de titanio y calcio dispuestos ortorrómbicamente cuando es cristalino, tiene un *bandgap* de 1.5 eV, un alta constante dieléctrica, capacidad de separar rápidamente la carga y de ser funcionalizados con CIGS, Si, Sn y/o Pb para extender el rango de absorción del espectro solar (300-800 nm) y aumentar el coeficiente de absorción ($10^5 cm^{-1}$) [33]. A pesar de este gran número de ventajas, el alto costo de fabricación y la toxicidad de sus metales de dopamiento [34], limita extender su uso. Otras celdas que presentan el mismo problema son: Las que se basan en *quantum dots* [35] o compuestos inorgánicos CZTSSe [26].

1.3. Celdas solares sensibilizadas por tintes (CSPT)

Contrario a las anteriores, las celdas orgánicas y las sensibilizadas por tintes (CSPT) son más sostenibles, ya que los polímeros que los componen pueden ser degradados por el ambiente.

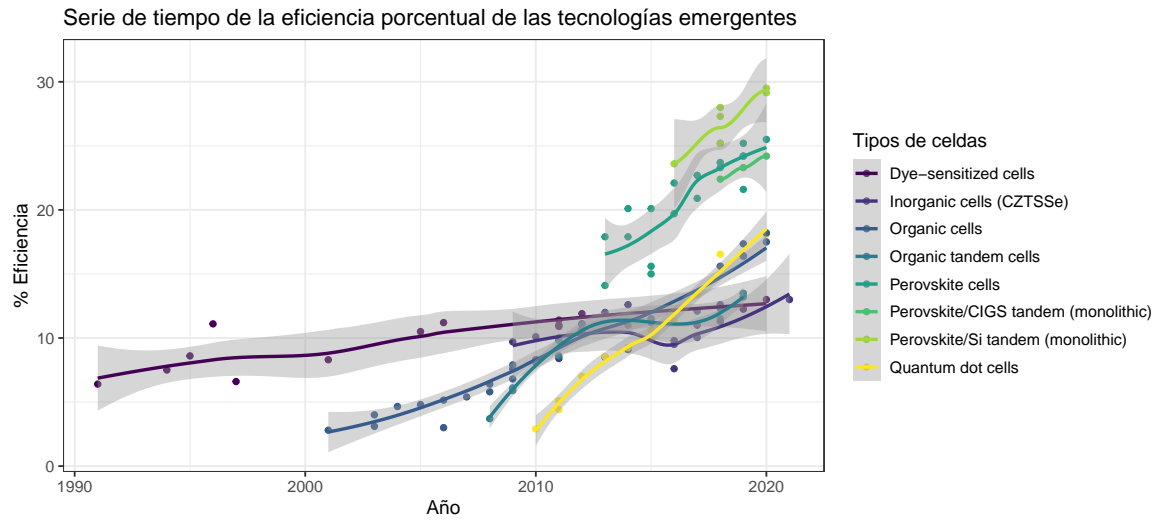


Figura 3: Serie de tiempo de las mejores eficiencias obtenidas en la investigación de las tecnologías emergentes de celdas solares. Fuente: *National Renewable Energy Laboratory* [13]

Además, pueden ser contruidos por módulos con diversos materiales orgánicos y metal-orgánicos, por lo que sus costos de fabricación serían comercialmente más rentables; aparte de ser más flexibles y resistentes. Si bien las CSPT han sido descubiertas antes que las orgánicas, estas últimas las han superado en términos de eficiencia, véase imagen 3; debido al descubrimiento de moléculas aceptoras y donadoras de alto rendimiento que generan y transportan carga [36]. Sin embargo, estas moléculas especializadas son más costosas que los tintes de las CSPT. Sumado a la brecha de conocimiento de las propiedades fotoactivas de muchos productos naturales que pueden ser usados como tintes sensibilizantes, el presente proyecto se concentrará en el estudio de las CSPT.

Las CSPT aplican mecanismos diferentes a los otros dispositivos fotovoltaicos. En contraste a los sistemas convencionales que utilizan semiconductores que integran los mecanismos de absorción de luz y el transporte de carga, los CSPT tienen estos dos mecanismos separados. En las CSPT el tinte sensibilizante se ancla al semiconductor para absorber la luz, mientras que el transporte de carga se da en la interfaz del sensibilizante y la banda de conducción del semiconductor por medio de inyección fotoinducida de electrones y creación de huecos. Estas características les confieren a las CSPT las ventajas de ser diseñadas e implementadas modularmente [37] así como de operar establemente, incluso en condiciones de temperatura, irradiación y ángulo de incidencia solar no estándares. Adicionalmente, la construcción de las CSPT no requiere altas temperaturas o sistemas de vacío. Gracias a las ventajas de las CSPT, existen esfuerzos para mejorar sus propiedades así como darle usos comerciales. Un ejemplo de esto último son los primeros CSPT comerciales usados como sensores y electrónica domótica ofrecidos en el 2009 por la empresa israelí 3G Solar Photovoltaics Ltd. (<https://www.3gsolar.com/>.)

El electrolito puede ser líquido, gel o sólido y se une al semiconductor como fase de contacto, las otras celdas se unen por contacto de fases sólidas. El corazón del sistema Grätzel es el fotoánodo que consiste en una monocapa de tinte sensibilizante absorbido por una capa de un semiconductor mesoporoso, de aproximadamente $10\ \mu\text{m}$ de grosor. El semiconductor usualmente es TiO_2 , pero también puede ser ZnO , SnO_2 , Nb_2O_5 , etc. El electrolito que permeabiliza el

semiconductor y que dona electrones a las moléculas oxidadas del tinte es comúnmente el par redox yoduro/yodo. Bajo la radiación lumínica el tinte promueve electrones (e^-) del HOMO al LUMO. El LUMO debe tener mayor energía que la banda de conducción del semiconductor para que la inyección suceda, posteriormente los electrones excitados difunden por el semiconductor y son colectados por el óxido conductor del vidrio FTO/ITO. Estos electrones salen del fotoánodo hacia el cátodo y en su paso pueden ser aprovechados como potencial eléctrico (V). Al reentrar a la celda por medio del contra electrodo (típicamente a construido a base de Pt-s/TCO), los electrones reducen los transportadores de huecos de electrones ($I_3^- \longrightarrow I^-$). Por último el agente reductor del electrolito difunde hacia el semiconductor para regenerar el tinte oxidado en cuestión de micro segundos, de tal forma que cierre el ciclo operativo de óxido-reducción de la celda sin afectar teóricamente las estructuras químicas de los componentes.

Las CSPT son dispositivos que aprovechan la absorción luminosa de los tintes para generar energía fotovoltaica. Sus arquitecturas se componen de 4 materiales: El semiconductor, el tinte o sensibilizador, el electrolito y los vidrios conductores (Figura 4). Las CSPT modernas derivan de la celda creada por Brian O'Rigan y Michael Gratzel en 1991 [38]. En dicha celda, se usó óxido de estaño dopado con flúor y partículas de platino (Pt-FTO) como vidrio conductor, el cual soportaba el óxido de titanio TiO_2 nanocrystalino dopado a su vez por nanopartículas de complejos triméricos de rutenio(II) bipiridina (Rutenio-bipiridina) como semiconductor y tinte sensibilizante, respectivamente. Para la regeneración del tinte se utilizó yoduro/triyoduro como electrolito redox. El fotovoltaje (ΔV) de celda corresponde al potencial teórico de trabajo útil y puede aproximarse a la diferencia entre el nivel de energía de Fermi del semiconductor bajo iluminación y el potencial de Nernst del par redox del electrolito.

1.4. Medidas del rendimiento de CSPT

Los componentes influyen directamente al rendimiento de la celda, y esta puede medirse por medio de dos ecuaciones: La eficiencia de corriente monocromática ($\eta_i(\lambda)$) y la eficiencia del potencial de conversión (PCE). La ecuación 1 describe la capacidad del dispositivo para mantener el flujo eléctrico de celda generado por fenómenos fotoeléctricos. Esta capacidad se calcula como el producto de la eficiencia de recolección de luz de longitud λ ($LHE(\lambda)$) que mide la fracción de fotones incidentes que son absorbidos por el tinte, el rendimiento cuántico de inyección de carga (ϕ_{inj}) que mide el número de huecos inyectados hacia la capa de valencia o electrones a la banda de conducción del semiconductor sensibilizado por número de einstein absorbidos, y la eficiencia de colección de carga de retorno (η_e) que mide la fracción de electrones que llegan al contraelectrodo versus los que salieron del electrodo de trabajo en un circuito cerrado. Como esta medición sólo toma en cuenta valores de parámetros para una longitud de onda, es más práctico tomar como medida de referencia la segunda ecuación.

$$\eta_i(\lambda) = LHE(\lambda) \times \phi_{inj} \times \eta_e \quad (1)$$

Por otra parte, la ecuación 2 describe a la eficiencia del potencial de conversión como el porcentaje del potencial lumínico del haz incidente aprovechado en forma de potencial eléctrico de salida. Para calcularlo, la CSPT es sometida a densidades de corriente y voltaje (J/V) controlados dentro de un simulador de luz solar. Los parámetros extraídos de estos experimentos son la densidad de corriente de corto circuito (J_{SC}), el voltaje del circuito abierto (V_{OC}), el factor de llenado (FF) y la potencia incidente (P_{in}). El J_{SC} permite conocer las características de absorción de la capa fotoactiva, la generación y el transporte de carga; debido a que estas la afectan. Mientras que el V_{OC} permite obtener la información de los niveles de energía de los materiales fotoactivos, el estado del funcionamiento de los materiales del electrodo de trabajo y la velocidad de recombinación del portador de carga. Por la parte del FF , que se define como

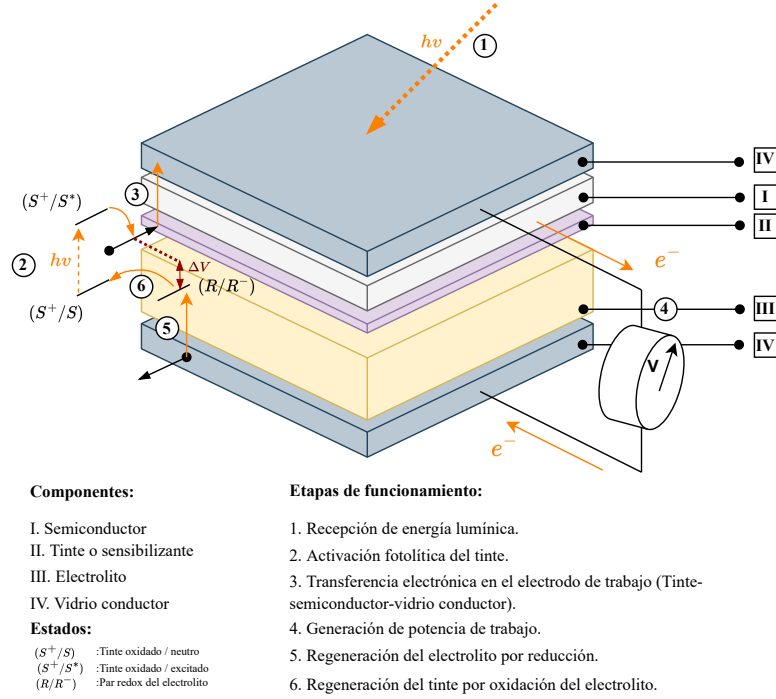


Figura 4: Estructura y funcionamiento de una celda solar sensibilizada por tinte (CSPT). Fuente: O'Rigan y Grätzel [38]

la relación entre la potencia real de la celda y la potencia sin la presencia de resistencia; es posible coleccionar información del estado de los electrodos y la estabilidad de los materiales en un determinado tiempo. Cabe señalar que el FF idealmente debe ser 1, lo cual significa que no existen fenómenos de desgaste que consuman potencial de trabajo. [Añadir Referencias](#)

$$PCE = \frac{P_{out}}{P_{in}} = \frac{J_{SC} \times V_{OC} \times FF}{P_{in}} \times 100\% \quad (2)$$

1.5. Optimización de las CSPT

Si bien la eficiencia de las celdas depende del rango de longitudes del espectro de absorción, la inyección electrónica y las reacciones redox, fenómenos que suceden en la interacción de los componentes de celda; el tinte juega un rol importante en estos. Es por ello que el presente trabajo se centrará en mejorarlos.

No obstante, existen reacciones que merman el rendimiento, las cuales son: El decaimiento del estado excitado del tinte, la migración de electrolitos de la banda conductora del semiconductor al tinte o al electrolito, también conocidas como reacciones de retroceso y oscuras, respectivamente.

Teniendo en cuenta lo anterior, se infiere que la optimización de las celdas CSPT recae en

contrarrestar al máximo dichas reacciones. Una manera de hacerlo es diseñando rutas para que la inyección electrónica y la regeneración del tinte por medio del electrolito sean más rápidas que el decaimiento del tinte excitado y las reacciones de retroceso, respectivamente.

1.6. Tintes para CSPT

Los tintes sensibilizantes ideales deben absorber un amplio rango de longitudes de onda, en especial cubrir la región visible y cercana al infrarrojo. Además de mostrar alta foto estabilidad, una absorptividad molar grande y debe ser capaz de inyectar los electrones fotogenerados rápidamente desde su estado excitado a la banda de conducción del semiconductor. Esto se logra con un fuerte anclaje a la superficie del óxido conductor y con un potencial redox propicio que permita su rápida regeneración gracias al electrolito. No obstante, las celdas tienen un tiempo de vida a causa de los procesos de enfriamiento o *quenching* y otros generados por el medio que degradan al tinte o al electrolito.

(G. Calogero, A. Bartolotta, G. Di Marco, A. Di Carlo, F. Bonaccorso, Vegetablebased dye sensitized solar cells, Chem. Soc. Rev. 44 (2015) 3244e3294.)

Hasta la fecha se han clasificado los tintes sensibilizantes en tres clases: Basados en complejos metálicos, en estructuras orgánicas libres de metales y de fuentes naturales.

Aquellos basados en complejos metálicos tienen dos tipos de ligantes: De anclaje y auxiliares. Los ligantes de anclaje son responsables de absorber el tinte a la superficie del semiconductor. Por otro lado, los ligantes auxiliares permiten la modulación de las propiedades del complejo. Algunos tintes de este tipo son los derivados del bipyridil de rutenio como el N3, N719 y el N749 o tinte negro. A pesar que estos tintes tienen altas eficiencia de conversión (η), sus altos costos de fabricación, limitados coeficientes de extinción y potenciales efectos contaminantes derivados de los metales dificultan la sostenibilidad de las CSTP y por ende la investigación basada en estos materiales.

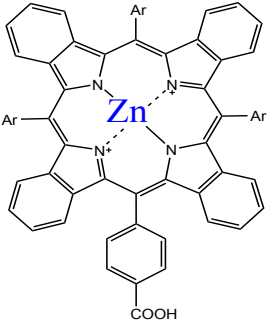
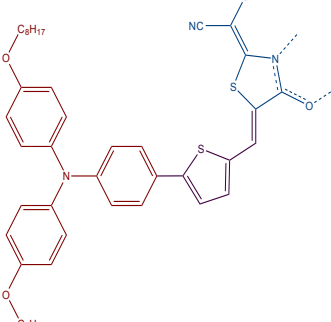
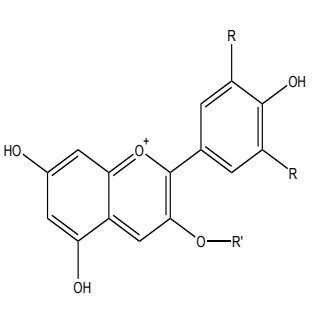
En segundo lugar, los tintes orgánicos libre de metales resuelven las limitaciones anteriormente mencionadas, ya que pueden ser diseñadas molecularmente a fin de tener propiedades que aumenten sus rendimientos. Generalmente, las estructuras de estos tintes se configuran químicamente en regiones donadoras y aceptoras de electrones, y puentes conjugados que unen y espacian dichas regiones ($D - \pi - A$). Las configuraciones donadoras de electrones pueden ser cumarinas, fenilaminas, indolinas, triarilaminas, carbazoles, entre otras. Mientras que las estructuras aceptoras pueden ser ácido cianoacrílico, rodaminas o piridinas. Por otra parte las configuraciones puentes pueden ser derivados de tiofenos, polienos o benzotiadiazoles.

En tercer lugar, la clase de tintes naturales son aquellos que se extraen de seres vivos, mayoritariamente de fuentes vegetales como frutas, flores, hojas, entre otras. Muchos de los pigmentos vegetales extraídos han sido investigados como sensibilizantes para CSTP y presentaron rendimientos fotovoltaicos significativos, en especial colorantes como derivados de clorofila, betalainas, antocianicas, carotenoides, etc. [1](#)

1.7. Problemática de los tintes naturales para CSPT

1.8. Enfoque de investigación direccionado por datos o *Data-Driven*

La investigación de tintes para CSTP de alto rendimiento puede realizarse por medio del camino clásico de ensayo y error guiado por intuición química y por métodos computacionales que sirven para filtrar los candidatos con características estructurales y funcionales teóricamente óptimas. Ambos enfoques deben ir de la mano a fin de ahorrar recursos en la búsqueda de materiales y validar experimentalmente los materiales candidatos. Sin embargo, los recursos

		
Metalorgánico	Libre de metales	Naturales

Cuadro 1: Tipos de tintes para CSPT

computacionales son limitados debido a la complejidad de los cálculos teóricos-computacionales en el diseño químico y el costo de las licencias de los softwares que los realizan.

A fin de expandir el uso de modelos computacionales para el tamizaje de materiales, se han creado metodologías que aprovechan datos experimentales y computacionales para extraer y extrapolar información en dichos datos. Estas metodologías se engloban dentro de un enfoque llamado investigación dirigida por datos (*Data-driven* del inglés). La creación de este enfoque es consecuencia del desarrollo de hardware para el almacenamiento y procesamiento informáticas, y los movimientos *Open-Source* de datos y software. Los cuales son los soportes de las bases de datos y herramientas informáticas para el procesamiento de datos. Una de estas herramientas, y es la que más presencia tiene en las investigaciones para el desarrollo de materiales, es el aprendizaje automático [39].

1.9. Aprendizaje automático aplicado a la química

El aprendizaje automático es un subcampo de la inteligencia artificial que estudia modelos computacionales que emulen procesos de aprendizaje basados en la experiencia para resolver problemas. Esta experiencia se obtiene entrenándolos con un volumen considerable de datos. Son muchos los campos de aplicación del aprendizaje automático en la química, tales como en el descubrimiento de nuevos medicamentos [40], la síntesis orgánica [41], las reacciones catalíticas [42], el diseño de baterías [43], supercapacitores [44] o también el desarrollo de las celdas solares [45].

La manera de abordar un problema con algoritmos de aprendizaje automático depende de la complejidad de las tareas a resolver y el volumen de datos disponibles. Si las tareas consisten en predecir valores numéricos o clasificar datos a partir del entrenamiento de modelos con un conjunto de variables independientes y dependientes, se podrían usar algoritmos de aprendizaje supervisado; tales como las regresiones polinomiales, árboles de decisión o máquinas de soporte vectorial. En caso solo se conocieran las variables independientes, las tareas recaen en la identificación de asociaciones y para ello se usarían algoritmos de aprendizaje no supervisado; como es el caso de modelos mixtos gaussianos, K-medias o de *clustering* jerárquico.

Con respecto al volumen de datos, estos tienen una relación inversa al nivel de complejidad del problema. Para ilustrar, Li utilizó miles de cálculos de dinámica molecular retroalimentar modelos de aprendizaje automático capaces de predecir fuerzas atómicas [46] de moléculas simples. En cambio, De Sandip se limitó a usar menos de cien descriptores atómicos y moleculares para predecir y comparar microestructuras cristalinas de moléculas pequeñas [47], dado que los cálculos

tomaban más tiempo computacional. En conclusión, mientras más interacciones existan entre unidades estructurales químicas (átomos, moléculas o materiales) más complejo es el espacio químico del problema, por ende, disminuye el número de fuentes de datos que describen dicho espacio.

Un espacio químico, se puede entender a primera mano como un espacio vectorial donde cada vector representa un descriptor físico o químico de una molécula. En tal sentido, es importante delimitar el objetivo principal de la investigación dentro del espacio químico del problema, ya que no es igual a construir modelos que predigan propiedades o estructuras exactas a que clasifiquen materiales en función de una serie de características dadas.

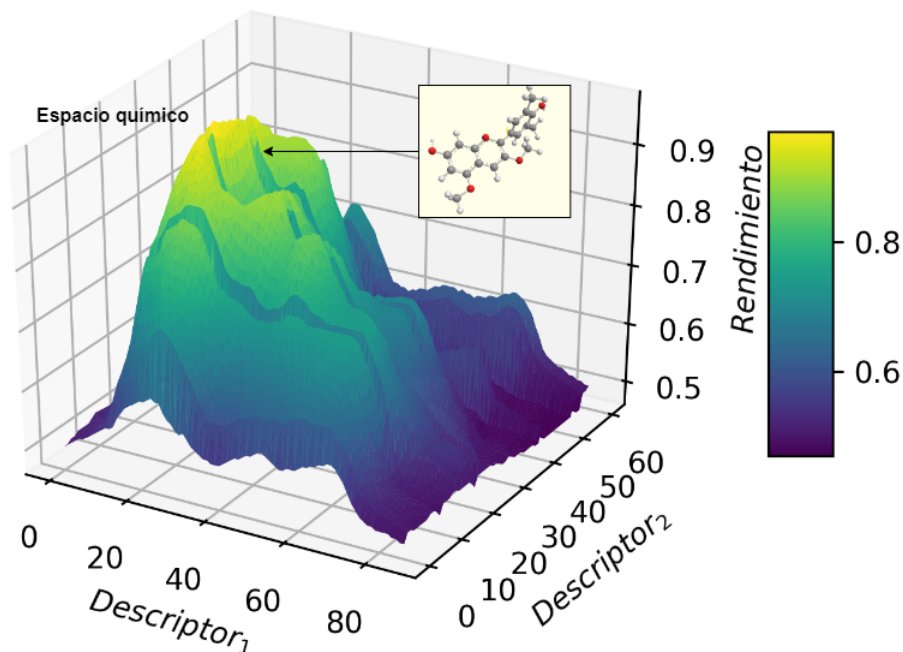


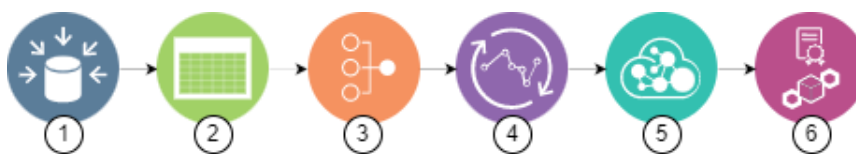
Figura 5: Espacio químico de investigación

1.10. Limitaciones del uso del aprendizaje automático en el desarrollo de materiales

1.11. Problemática

En las investigaciones de la aplicación del aprendizaje automático en las CSTP se hace uso de las relaciones cuantitativas estructura-propiedad de las moléculas (*QSAR* del inglés) para predecir el rendimiento de la celda a partir de las características estructurales de los tintes. Los algoritmos utilizan las

¿Es posible la construcción de modelos de aprendizaje automático químicamente interpretativos para la clasificación de moléculas de colorantes naturales según el posible rendimiento que puedan tener como sensibilizantes de las CSPT a partir de datos de relaciones cuantitativos de estructura y propiedad?



Etapas:

1. Recopilación de datos en fuentes bibliográficas o simulaciones computacionales.
2. Limpieza, tratamiento e ingeniería de datos
3. Análisis estadístico de las variables características y objetivo.
4. Entrenamiento y evaluación de los modelos de aprendizaje automático.
5. Análisis de interpretabilidad química q de los modelos de aprendizaje automático.
6. Extracción del modelo de aprendizaje automático interpretativo con mejor desempeño.

Figura 6: Etapas de investigación basada en datos

1.12. Justificación

La referencia es citada en [48]

Los procesos en el descubrimiento de celdas solares con mayor eficiencia de producción energética y sostenibilidad tales como el diseño, la prospección, la construcción son lentos, costosos y poco eficientes. Y el tiempo que toma que estas nuevas celdas solares tengan rentabilidad comercial para las partes que financiaron la investigación es reducido [49].

La ciencia de materiales dirigida por datos

- Los modelos de aprendizaje automáticos son activos académicos y comerciales para una institución que desarrolla materiales.
- Es importante solucionar los problemas de interpretabilidad de los modelos de aprendizaje automático a fin de que los expertos en la materia pero que no desarrollaron los modelos puedan incorporar sus conocimientos en estos.
- Estos modelos pueden ser contruidos con herramientas de acceso abierto y con datos extraídos tanto de simulaciones computacionales, experimentos y la bibliografía.
- Estos modelos ahorraran tiempo de procesamiento de la estimación del rendimiento de una CSTP genérico a partir de las propiedades químicas de un tinte natural en un tiempo menor a lo que lo haría una simulación computacional. Esto sirve para filtrar de moléculas candidatas antes de la fase experimental.
- Investigar tintes

1.13. Antecedentes

En los antecedentes se

1.13.1. Internacionales

1.13.2. Nacionales

2. Hipótesis de investigación y objetivos

2.1. Hipótesis de investigación

2.2. Objetivos

3. Materiales y métodos

3.1. Análisis estadístico

3.1.1. Pruebas de normalidad

La verificación de la normalidad de la distribución de un conjunto de datos se puede obtener a través de técnicas gráficas y por pruebas.

En primer lugar, aquellas basadas en gráficas tienen como

Data

La cien Actualmente, cerca del 50 % de las nuevas publicaciones de artículos son de acceso abierto y estimaciones hechas por (Ref 33 Y 34) indican que estás casi en su totalidad lo serán también para el 2040.

Herramientas de libre acceso

El software de acceso abierto es el conjunto de herramientas computacionales que permiten las operaciones de la ciencia de datos basada en datos. Este tipo de software adquiere importancia desde la aparición pública del internet y el sistema operativo LINUX en 1990, cuyos desarrollos fueron netamente colaborativos. Adicionalmente, el nacimiento de herramientas de control de versiones y lenguajes de programación de alto nivel permitieron el nacimiento de repositorios proyectos científicos (46), notebooks electrónicos (47), paquetes de simulación y experimentos(49) y librerías de aprendizaje automático y ciencia de datos (50).

3.1.2. Extracción de datos

3.1.3. Preprocesamiento de datos

3.1.4. Descubrimiento y análisis de características

3.1.5. Entrenamiento y evaluación del modelo 1

3.1.6. Entrenamiento y evaluación del modelo 2

3.1.7. Entrenamiento y evaluación del modelo 3

3.1.8. Comparación de modelos

4. Referencias bibliográficas

Referencias

1. Mayer A. Fossil fuel dependence and energy insecurity. Energy, Sustainability and Society 2022; 12:1-13
2. Vernon R. Oil Crisis. 1976

3. Rodriguez-Fernandez L, Carvajal ABF y Tejada VF de. Improving the concept of energy security in an energy transition environment: Application to the gas sector in the European Union. *The Extractive Industries and Society* 2022; 9:101045
4. Guilyardi E, Lescarmonnier L, Matthews R, Point SP, Rumjaun AB, Schlüpmann J y Wilgenbus D. IPCC Special Report “Global Warming of 1.5° C”: Summary for Teachers. 2018
5. Asmelash E, Prakash G, Gorini R y Gielen D. Role of IRENA for global transition to 100 % renewable energy. *Accelerating the transition to a 100 % renewable energy era*. Springer, 2020 :51-71
6. Ritchie H, Roser M y Rosado P. Energy. Our World in Data 2020. <https://ourworldindata.org/energy>
7. IRENA. Renewable Energy Statistics 2022. Renewable Energy Target Setting, Abu Dhabi, UAE 2022
8. Hammarström L. Overview: capturing the sun for energy production. *Ambio* 2012; 41:103-7
9. Xu Y, Pei J, Yuan J y Zhao G. Concentrated solar power: technology, economy analysis, and policy implications in China. *Environmental Science and Pollution Research* 2022; 29:1324-37
10. Tawalbeh M, Al-Othman A, Kafiah F, Abdelsalam E, Almomani F y Alkasrawi M. Environmental impacts of solar photovoltaic systems: A critical review of recent progress and future outlook. *Science of The Total Environment* 2021; 759:143528
11. Green MA. The path to 25 % silicon solar cell efficiency: History of silicon cell evolution. *Progress in photovoltaics: research and applications* 2009; 17:183-9
12. Singh GK. Solar power generation by PV (photovoltaic) technology: A review. *Energy* 2013; 53:1-13
13. NREL. National Renewable Energy Laboratory Best Research-Cell Efficiency Chart. Available from: <https://www.nrel.gov/pv/cell-efficiency.html> [Accessed on: 2022 Sep 26]
14. Blakers A, Zin N, McIntosh KR y Fong K. High efficiency silicon solar cells. *Energy Procedia* 2013; 33:1-10
15. Gul M, Kotak Y y Muneer T. Review on recent trend of solar photovoltaic technology. *Energy Exploration & Exploitation* 2016; 34:485-526
16. Möller HJ, Funke C, Rinio M y Scholz S. Multicrystalline silicon for solar cells. *Thin Solid Films* 2005; 487:179-87
17. Sark Wv, Korte L y Roca F. Introduction–Physics and Technology of Amorphous-Crystalline Heterostructure Silicon Solar Cells. *Physics and technology of amorphous-crystalline heterostructure silicon solar cells*. Springer, 2012 :1-12
18. Brendel R. Thin-film crystalline silicon solar cells: physics and technology. John Wiley & Sons, 2011
19. Konagai M, Sugimoto M y Takahashi K. High efficiency GaAs thin film solar cells by peeled film technology. *Journal of crystal growth* 1978; 45:277-80
20. Dimroth F y Kurtz S. High-efficiency multijunction solar cells. *MRS bulletin* 2007; 32:230-5
21. Philipps SP, Dimroth F y Bett AW. High-efficiency III–V multijunction solar cells. *McEvoy’s handbook of photovoltaics*. Elsevier, 2018 :439-72
22. Wada T, Hashimoto Y, Nishiwaki S, Satoh T, Hayashi S, Negami T y Miyake H. High-efficiency CIGS solar cells with modified CIGS surface. *Solar Energy Materials and Solar Cells* 2001; 67:305-10

23. Ferekides C, Balasubramanian U, Mamazza R, Viswanathan V, Zhao H y Morel D. CdTe thin film solar cells: device and technology issues. *Solar energy* 2004; 77:823-30
24. Krč J, Smole F y Topič M. Analysis of light scattering in amorphous Si: H solar cells by a one-dimensional semi-coherent optical model. *Progress in photovoltaics: Research and Applications* 2003; 11:15-26
25. Sharma K, Sharma V y Sharma S. Dye-sensitized solar cells: fundamentals and current status. *Nanoscale research letters* 2018; 13:1-46
26. Suryawanshi M, Agawane G, Bhosale S, Shin SW, Patil P, Kim JH y Moholkar A. CZTS based thin film solar cells: a status review. *Materials Technology* 2013; 28:98-109
27. Hoppe H y Sariciftci NS. Organic solar cells: An overview. *Journal of materials research* 2004; 19:1924-45
28. Ameri T, Dennler G, Lungenschmied C y Brabec CJ. Organic tandem solar cells: A review. *Energy & Environmental Science* 2009; 2:347-63
29. Jung HS y Park NG. Perovskite solar cells: from materials to devices. *small* 2015; 11:10-25
30. Li H y Zhang W. Perovskite tandem solar cells: from fundamentals to commercial deployment. *Chemical Reviews* 2020; 120:9835-950
31. Kim M, Jeong J, Lu H, Lee TK, Eickemeyer FT, Liu Y, Choi IW, Choi SJ, Jo Y, Kim HB y col. Conformal quantum dot-SnO₂ layers as electron transporters for efficient perovskite solar cells. *Science* 2022; 375:302-6
32. Chowdhury MS, Rahman KS, Chowdhury T, Nuthammachot N, Techato K, Akhtaruzzaman M, Tiong SK, Sopian K y Amin N. An overview of solar photovoltaic panels' end-of-life material recycling. *Energy Strategy Reviews* 2020; 27:100431
33. Velilla E, Ramirez D, Uribe JI, Montoya JF y Jaramillo F. Outdoor performance of perovskite solar technology: Silicon comparison and competitive advantages at different irradiances. *Solar Energy Materials and Solar Cells* 2019; 191:15-20
34. Wang X, Dong B, Feng M, Xue DJ y Wang SM. Sustainable management of lead in perovskite solar cells. *Journal of Materials Chemistry A* 2022; 10:15861-4
35. Pan Z, Mora-Seró I, Shen Q, Zhang H, Li Y, Zhao K, Wang J, Zhong X y Bisquert J. High-efficiency "green" quantum dot solar cells. *Journal of the American Chemical Society* 2014; 136:9203-10
36. Li Y, Huang W, Zhao D, Wang L, Jiao Z, Huang Q, Wang P, Sun M y Yuan G. Recent Progress in Organic Solar Cells: A Review on Materials from Acceptor to Donor. *Molecules* 2022; 27:1800
37. Hashmi G, Miettunen K, Peltola T, Halme J, Asghar I, Aitola K, Toivola M y Lund P. Review of materials and manufacturing options for large area flexible dye solar cells. *Renewable and Sustainable Energy Reviews* 2011; 15:3717-32
38. O'regan B y Grätzel M. A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO₂ films. *nature* 1991; 353:737-40
39. Himanen L, Geurts A, Foster AS y Rinke P. Data-driven materials science: status, challenges, and perspectives. *Advanced Science* 2019; 6:1900808
40. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, Li B, Madabhushi A, Shah P, Spitzer M y col. Applications of machine learning in drug discovery and development. *Nature reviews Drug discovery* 2019; 18:463-77

41. Maryasin B, Marquetand P y Maulide N. Machine learning for organic synthesis: are robots replacing chemists? *Angewandte Chemie International Edition* 2018; 57:6978-80
42. Kitchin JR. Machine learning in catalysis. *Nature Catalysis* 2018; 1:230-2
43. Liu Y, Guo B, Zou X, Li Y y Shi S. Machine learning assisted materials design and discovery for rechargeable batteries. *Energy Storage Materials* 2020; 31:434-50
44. Zhu S, Li J, Ma L, He C, Liu E, He F, Shi C y Zhao N. Artificial neural network enabled capacitance prediction for carbon-based supercapacitors. *Materials Letters* 2018; 233:294-7
45. Mahmood A y Wang JL. Machine learning for high performance organic solar cells: current scenario and future prospects. *Energy & environmental science* 2021; 14:90-105
46. Li Z, Kermode JR y De Vita A. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Physical review letters* 2015; 114:096405
47. De S, Bartók AP, Csányi G y Ceriotti M. Comparing molecules and solids across structural and alchemical space. *Physical Chemistry Chemical Physics* 2016; 18:13754-69
48. Wen Y, Fu L, Li G, Ma J y Ma H. Accelerated Discovery of Potential Organic Dyes for Dye-Sensitized Solar Cells by Interpretable Machine Learning Models and Virtual Screening. *Solar RRL* 2020; 4:2000110
49. Eagar TW. Bringing new materials to Market. en. *ABI/INFORM Global* 1995; 98:42

5. Cronograma de trabajo

6. Presupuesto y financiamiento