UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

FACULTAD DE CONTADURÍA, ADMINISTRACIÓN E INFORMÁTICA

MAESTRÍA EN OPTIMIZACIÓN Y CÓMPUTO APLICADO

APLICACIÓN DE APRENDIZAJE AUTOMÁTICO PARA LA CLASIFICACIÓN DEL DESEMPEÑO DE BATERÍAS DE IONES DE LITIO A PARTIR DE LA EFICIENCIA COULÓMBICA Y LA CONDUCTIVIDAD DEL ELECTROLITO

T E S I S

QUE PARA OBTENER EL GRADO DE

MAESTRÍA EN OPTIMIZACIÓN Y CÓMPUTO APLICADO

PRESENTA

SERGIO RUBÉN OCAMPO PÉREZ

DIRECTOR DE TESIS

DR. OUTMANE OUBRAM

CO-DIRECTOR

DR. NOUREDDINE LAKOUARI

REVISORES:

DR. OUTMANE OUBRAM
DR. NOUREDDINE LAKOUARI
DR. FEDERICO ALONSO PECINA
DRA. LORENA DÍAZ GONZÁLEZ
DR. JOSÉ ALBERTO HERNÁNDEZ AGUILAR

Facultad de Contaduría, Administración e Informática

CUERNAVACA, MORELOS. AGOSTO,2026

Resumen

Este trabajo propone una metodología basada en aprendizaje automático para clasificar dos parámetros esenciales de las baterías de iones de litio: la eficiencia coulómbica, que se estudia como un indicador del rendimiento electroquímico a nivel celular, y la conductividad iónica, que se analiza como una propiedad intrínseca del electrolito para el transporte. Se desarrolló un proceso de curación de la base de datos, basado en datos experimentales que se encuentran en la literatura. Este proceso abarcó el reconocimiento de compuestos a través del sistema SMILES y la determinación de propiedades fisicoquímicas esenciales, produciendo 13 descriptores atómicos para medir la eficiencia coulómbica y 17 descriptores moleculares y termodinámicos para estudiar la conductividad iónica.

Luego, se llevó a cabo un análisis exploratorio de datos en un conjunto curado, que incluye 283 datos para la eficiencia coulómbica y 10196 para la conductividad iónica. La regresión LASSO (Least Absolute Shrinkage and Selection Operator) fue la técnica utilizada para seleccionar las características. Se definieron límites de rendimiento rigurosos para analizar las formulaciones de manera sistemática: la eficiencia coulómbica fue dividida en dos categorías (bajo rendimiento CE $<$ 99 % y alto rendimiento CE $\geq$ 99 %), mientras que la conductividad iónica se clasificó en tres grupos: baja ($\sigma < 1$ mS/cm), media ($1 \leq \sigma \leq 10$ mS/cm) y alta ($\sigma > 10$ mS/cm).

Se entrenaron varios métodos de aprendizaje automático, entre los que se incluyen K-Vecinos más cercanos, Regresión Logística, eXtreme Gradient Boosting (XGBoost), Bosques Aleatorios y Light Gradient Boosting Machine (LightGBM). La validación cruzada y el entorno de Optuna se usaron para optimizar estos modelos. Además, para determinar la relevancia y el efecto de cada variable en las predicciones, se utilizaron los valores SHAP (SHapley Additive exPlanations), lo que otorgó interpretabilidad a los modelos creados.

Los resultados demuestran que el uso de modelos de aprendizaje automático puede mejorar la clasificación de electrolitos de alto y bajo rendimiento en baterías de iones de litio, y establecer una base sólida para el diseño de nuevos electrolitos en investigaciones futuras.

Abstract

This work proposes a machine learning-based methodology to classify two essential parameters of lithium-ion batteries: coulombic efficiency, studied as an indicator of electrochemical performance at the cell level, and ionic conductivity, analyzed as an intrinsic electrolyte property related to ion transport. A data curation process was developed based on experimental data reported in the literature. This process included compound identification through the SMILES system and the calculation of key physicochemical properties, resulting in 13 atomic descriptors for coulombic efficiency assessment and 17 molecular and thermodynamic descriptors for ionic conductivity analysis.

Subsequently, an exploratory data analysis was performed on the curated dataset, which comprised 283 samples for coulombic efficiency and 10,196 samples for ionic conductivity. Least Absolute Shrinkage and Selection Operator (LASSO) regression was employed for feature selection. Strict performance thresholds were established to systematically analyze the formulations: coulombic efficiency was divided into two categories (low performance, CE $< 99\,\%$, and high performance, CE $\geq 99\,\%$), while ionic conductivity was classified into three groups: low ($\sigma < 1$ mS/cm), medium ($1 \leq \sigma \leq 10$ mS/cm), and high ($\sigma > 10$ mS/cm).

Several machine learning algorithms were trained, including K-Nearest Neighbors, Logistic Regression, eXtreme Gradient Boosting (XGBoost), Random Forest, and Light Gradient Boosting Machine (LightGBM). Cross-validation and the Optuna framework were used to optimize these models. In addition, SHAP (SHapley Additive exPlanations) values were employed to determine the relevance and contribution of each variable to the predictions, providing interpretability to the developed models.

The results demonstrate that machine learning models can effectively improve the classification of high- and low-performance electrolytes for lithium-ion batteries, while establishing a solid foundation for the design of novel electrolytes in future research.

Agradecimientos

Agradezco a mi madre Anita Pérez Díaz, quien siempre me apoyó en todo momento e hizo lo imposible por mantener a flote a 3 hijos sin importar lo fuerte de la marea.

A mis hermanos, Jesús Enrique Ocampo Pérez y Víctor Uriel Ocampo Pérez, les agradezco sinceramente por estar siempre a mi lado, brindándome apoyo, motivación y confianza en cada etapa de este camino. Su acompañamiento ha sido un pilar constante durante mi formación.

Deseo expresar mi más sincero agradecimiento al Dr. Noureddine Lakouari por su guía académica, su constante disposición y sus valiosas observaciones a lo largo del desarrollo de este trabajo, las cuales fueron fundamentales para fortalecer su calidad científica y rigor metodológico.

De igual manera, agradezco profundamente al Dr. Outmane Oubram por su acompañamiento, orientación y apoyo durante todo el proceso de investigación, así como por la generosidad al compartir su conocimiento y experiencia, contribuyendo significativamente al enriquecimiento de este proyecto.

Agradezco también a cada uno de mis asesores por el tiempo dedicado, las sugerencias aportadas y la retroalimentación proporcionada, las cuales enriquecieron el desarrollo de esta tesis y ampliaron mi perspectiva académica.

Reconozco y agradezco el apoyo otorgado por la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), a través de la beca concedida, la cual fue fundamental para la realización de este trabajo de investigación y para el fortalecimiento de mi formación académica y profesional.

Finalmente, agradezco a la Facultad de Contabilidad, Administración e Informática (FCAeI) por facilitar el uso de sus instalaciones y recursos, los cuales fueron esenciales para el desarrollo adecuado de esta investigación.

Índice general

1. Introducción **10**

1.1. Contexto de la investigación . . . 10

1.2. Planteamiento del problema . . . 11

1.3. Justificación . . . 12

1.4. Hipótesis . . . 13

1.5. Objetivos . . . 13

1.5.1. Objetivo General. . . . 13

1.5.2. Objetivos Específicos. . . . 13

1.6. Conceptos clave . . . 14

2. Marco Teórico **16**

2.1. Fundamentos de baterías de iones de Litio . . . 16

2.1.1. Composición y Naturaleza de los Electrolitos Líquidos . . . 19

2.1.2. Eficiencia Coulómbica . . . 20

2.1.3. Conductividad Iónica . . . 23

2.2. Aprendizaje Automático . . . 25

2.2.1. Algoritmos de Aprendizaje Supervisado . . . 25

2.2.2. Algoritmos de Clasificación . . . 26

2.2.3. Optimización de hiperparámetros de modelos de clasificación . . . 28

2.2.4. Métricas de evaluación de modelos de clasificación . . . 29

2.2.5. Interpretabilidad de modelos de aprendizaje automático . . . 32

2.2.6. Aprendizaje Automático Aplicado a Baterías . . . 33

3. Metodología **40**

3.1. Descripción de los Datasets 40

3.1.1. Dataset de Eficiencia Coulómbica 40

3.1.2. Dataset de Conductividad Iónica 42

3.2. Definición de Clases y Umbrales 43

3.3. Preprocesamiento e Ingeniería de Características 47

3.3.1. Expansión y Selección de Variables para la Eficiencia Coulómbica 47

3.3.2. Extracción de Descriptores Fisicoquímicos para la Conductividad Iónica . . . 48

3.4. Diseño del Experimento de Cómputo 51

3.4.1. Selección de Modelos 51

3.4.2. Optimización de Hiperparámetros con Optuna 52

3.5. Interpretabilidad del Modelo con SHAP 53

4. Resultados y Discusión **54**

4.1. Análisis de la Clasificación de Eficiencia Coulómbica 54

4.1.1. Capacidad de Detección de Electrolitos de Alto Desempeño 57

4.1.2. Interpretabilidad SHAP para Eficiencia Coulómbica 59

4.2. Análisis de la Clasificación de Conductividad Iónica 62

4.2.1. Capacidad de Detección de Electrolitos de Alta Conductividad 62

4.2.2. Interpretabilidad SHAP para la Conductividad Iónica 65

5. Conclusiones **68**

Bibliografía **70**

Índice de figuras

2.1. Representación esquemática de una celda de iones de litio y el flujo de partículas durante la descarga. 17

2.2. Cuantificación de la composición del litio electroquímicamente inactivo. En la gráfica las barras grises representan la cantidad de litio metálico muerto, y las barras de color verde azulado representan el porcentaje de porosidad. Adaptado de Hobold et al. (2021). 21

2.3. Diagrama de flujo del proceso de aprendizaje supervisado. El esquema ilustra cómo los datos etiquetados se utilizan en la fase de entrenamiento para construir un modelo matemático, el cual posteriormente genera predicciones en la fase de inferencia. Adaptado de Mobarak et al. (2023) 26

2.4. Modelo de ensamble de árboles. La predicción final se determina a partir de la suma de las contribuciones realizadas por cada árbol que compone el modelo. Adaptado de Chen y Guestrin (2016) 27

2.5. En Optuna, cada trabajador ejecuta de manera independiente una instancia de la función objetivo dentro de un estudio. A través de las API del sistema, las pruebas pueden consultar y actualizar un almacenamiento compartido que contiene la información de evaluaciones anteriores. Este esquema permite que múltiples trabajadores colaboren en un mismo proceso de optimización, compartiendo continuamente el avance y los resultados obtenidos. Adaptado de Akiba et al. (2019) 29

2.6. Diagrama común de una matriz de confusión 31

2.7. Diagrama común de una curva ROC: El gráfico muestra la relación dinámica entre la Tasa de Verdaderos Positivos y la Tasa de Falsos Positivos. La línea azul sólida representa el clasificador propuesto, mientras que la línea punteada negra define el umbral de una clasificación puramente aleatoria (AUC = 0,5). 32

3.1. Distribución de los valores de la eficiencia coulombica y balace de las clases resultantes de la aplicacion del umbral 45
3.2. Distribución de los valores de conductividad y balance de clases resultantes tras la aplicación del umbral. 46

4.1. Comparación de áreas bajo la curva ROC (AUC-ROC) de modelos de clasificación de eficiencia coulombica. 56
4.2. Matrices de confusión de los modelos mediante ajuste de hiperparámetros con Optuna. 59
4.3. Interpretabilidad SHAP para la alta eficiencia coulombica. 60
4.4. Análisis de la dependencia de las 2 variables con mayor impacto en la eficiencia coulombica. 61
4.5. Comparación de area en curva Roc, en clase alta conductividad 63
4.6. Matrices de confusión de los modelos optimizados mediante ajuste de hiperparámetros con optuna para la conductividad. 64
4.7. Interpretabilidad SHAP para la alta conductividad 65
4.8. Valores shap para rangos de temperatura 66

Índice de cuadros

3.1. Variables descriptoras del conjunto de datos de electrolitos. 42

3.2. Descripción detallada de los campos originales en la base de datos. 43

3.3. Variables seleccionadas tras la aplicación del método de regularización L1 y su interpretación. 48

3.4. Variables finales y descripción del conjunto de datos procesado de conductividad iónica. 50

4.1. Comparación de métricas globales en entrenamiento y prueba para los modelos evaluados para la eficiencia coulombica. 55

4.2. Hiperparámetros optimizados del modelo XGBoost mediante el uso de Optuna para la eficiencia coulombica. 57

4.3. Desempeño en la clasificación de la Clase 1 ($CE \geq 99\,\%$) en eficiencia coulombica. . 58

4.4. Comparación de métricas globales en entrenamiento y prueba para los modelos evaluados para la conductividad iónica. 62

4.5. Desempeño en la clasificación de la clase Alta para los modelos evaluados. 63

Capítulo 1

Introducción

En este capítulo se presenta la fundamentación general de la investigación. Se introduce el contexto de las baterías de iones de litio y el papel crítico de los electrolitos en su desempeño dentro de la transición energética. A partir de ello, se plantea el problema y se justifica el uso de enfoques basados en aprendizaje automático para superar las limitaciones del diseño empírico de materiales. Asimismo, se definen la hipótesis y los objetivos del estudio, y se establecen los conceptos clave necesarios para comprender los fenómenos fisicoquímicos y computacionales abordados en el trabajo.

1.1. Contexto de la investigación

En la actualidad, las baterías de litio han alcanzado una posición elevada en el ámbito del almacenamiento de la energía, intensificándose, especialmente, por su alta densidad energética y su alta robustez en términos de las cargas de descarga. En el contexto de la transición energética actual, el litio se sitúa en el centro de los recursos minerales estratégicos bajo el imperativo de la industria de las baterías recargables, que, según U.S. Geological Survey (2026), es responsable del 88 % de su consumo final a nivel global. Este crecimiento en la demanda, hacia vehículos eléctricos y sistemas de almacenamiento de las energías, llevó hasta el consumo de 263,000 toneladas en 2025. Con respecto a la disponibilidad, los recursos que se han podido detectar en el ámbito internacional son de unas 150 millones de toneladas, encontrando importantes concentraciones, incluso en la zona latinoamericana: Argentina tiene unos 28 millones, Bolivia 23, Chile 13 y México tiene unos recursos de 1.7 millones de toneladas métricas.

Sobre todo, resulta necesario atravesar las limitaciones en cuanto a rendimiento y durabilidad de las baterías existentes para lograr que la implementación a gran escala de las tecnologías a partir de los recursos existentes sea una realidad. Y los materiales que las constituyen son los que determinan por sí mismos esta optimización. En el funcionamiento de estas celdas, el electrolito cumple un papel relevante dado que posibilita el transporte de los iones de litio entre el ánodo y el cátodo. El rendimiento electroquímico está supeditado a dos aspectos claves: la eficiencia coulómbica, que de acuerdo a Kim et al. (2023) se define formalmente como la relación entre la capacidad de descarga y la capacidad de carga de la celda, además de que tiene una incidencia derivada de la resistencia interna y el transporte iónico. Frente a la necesidad de mejorar el rendimiento de estos sistemas, la investigación en ciencia de materiales deja atrás el tradicional enfoque empírico de ensayo y error y se encuentra ahora en un nuevo paradigma que es guiado por datos. En este contexto, el aprendizaje automático juega un rol importante para realizar un descubrimiento y optimización de componentes de modo más rápido, porque procesa y analiza un gran volumen de datos experimentales en un tiempo mucho más corto del habitual (Mobarak et al., 2023). El uso de estas herramientas matemáticas resulta crucial en el diseño de electrolitos, pues la conductividad iónica y la eficiencia coulómbica son dependencias fisicoquímicas no lineales que hacen que la optimización simultánea de las propiedades sea muy difícil (Deng et al., 2026). Para tratar de superar este obstáculo, la presente investigación propone un enfoque de aprendizaje automatizado supervisado que está enfocado en clasificar el comportamiento de formulaciones líquidas. De un modo concreto, el trabajo divide la eficiencia coulómbica en regímenes de alta y baja eficiencia, un planteamiento de la metodología que da adecuadamente cuenta de los diferentes mecanismos de la degradación interfacial que evitan la reversible. De modo complementario, la conductividad iónica también se divide en umbrales operativos en los que es posible evaluar mejor el transporte de carga.

1.2. Planteamiento del problema

Existe una gran dificultad para identificar materiales que sean óptimos para electrolitos. El método experimental tradicional es un procedimiento lento que muchas veces se apoya en procesos de estandarización y en el acceso a bases de datos. Otro aspecto destacado lo constituye la gran inversión que requiere la validación experimental de materiales; para la medición de la conductividad iónica es suficiente medir únicamente el electrolito, pero la determinación de la eficiencia coulómbica es mucho

más laboriosa. Esto es porque requiere llevar a cabo la construcción de una celda electroquímica completa, con la participación del ánodo, cátodo y separador, y que el proceso de carga y descarga se controle durante mucho tiempo para lograr obtener datos fiables. Puesto que este protocolo no solo conlleva un incremento de costes de operación, sino que también genera cuellos de botella temporales en la investigación de nuevos materiales, entonces debe dar respuesta a la siguiente pregunta: ¿Cómo desarrollar un modelo de clasificación basado en algoritmos de aprendizaje automático y descriptores fisicoquímicos que permita predecir el desempeño de nuevos electrolitos a partir de su conductividad iónica y su eficiencia coulómbica?

1.3. Justificación

El aprendizaje automático proporciona herramientas de análisis de predicción, que potencian la extracción de patrones de conjuntos de datos complejos. Modelos de ensamble como Random Forest demostraron en la literatura reciente una alta precisión en la predicción de propiedades fisicoquímicas y capacidades de descarga para realizaciones de selección de materiales de acuerdo con la previsión para futuras formulaciones (Wang et al., 2021). Este paradigma basado en datos permite el manejo de grandes volúmenes de información y potencia un descubrimiento más rápido de formulaciones que superen las limitaciones de los métodos empíricos ortodoxos (Deng et al., 2026).

Esta metodología representa una forma viable de optimización de costes en la investigación, contando con una posibilidad técnica, ya que los modelos de aprendizaje automático, al predecir el desempeño de un electrolito antes de su síntesis física, funcionan como una criba inicial que elimina formulaciones subóptimas y conserva únicamente los candidatos más prometedores para la validación de las medidas experimentales. De esta forma, disminuye drásticamente el coste operativo relacionado con el consumo de reactivos y las largas horas de experimentación manual que suelen acompañar a las aproximaciones de prueba y error. Así, el proceso se torna en un diseño guiado por datos que potencia la eficiencia de los recursos invertidos. Incluyendo también a esta investigación la corriente del objetivo global de sostenibilidad y la transición energética donde las baterías de iones de litio son el pilar para la integración de fuentes renovables y la electrificación del transporte, factores clave para la reducción de las emisiones de dióxido de carbono (Ren y Du, 2023). Por consiguiente, mejorar la eficiencia coulómbica y prolongar la vida útil de las celdas mediante modelos predictivos

disminuye la necesidad de reemplazos prematuros, reduciendo el impacto ambiental asociado a la generación de residuos tecnológicos.

1.4. Hipótesis

H1 (Hipótesis de trabajo): La implementación de modelos de clasificación basados en aprendizaje automático, utilizando descriptores moleculares y fisicoquímicos como variables de entrada, permite clasificar con alta precisión el desempeño de los electrolitos para baterías de iones de litio en función de su conductividad iónica y su eficiencia coulómbica.

H0 (Hipótesis nula): La implementación de modelos de clasificación basados en aprendizaje automático, utilizando descriptores moleculares y fisicoquímicos como variables de entrada, no permite clasificar con una precisión significativamente superior al azar o a un modelo base el desempeño de los electrolitos para baterías de iones de litio en función de su conductividad iónica y su eficiencia coulómbica.

1.5. Objetivos

1.5.1. Objetivo General.

Desarrollar y aplicar modelos de aprendizaje automático capaces de clasificar electrolitos en categorías de alto y bajo desempeño, utilizando como criterios su eficiencia coulómbica y su conductividad iónica.

1.5.2. Objetivos Específicos.

- Recolectar y estructurar una base de datos robusta de electrolitos líquidos tanto de eficiencia coulómbica como de conductividad.
- Analizar y seleccionar los atributos clave de los materiales para el entrenamiento de los modelos.
- Desarrollar y entrenar un modelos de aprendizaje automático adaptado a las características de los datos.
- Evaluar y ajustar el modelo utilizando métricas de desempeño como Accuracy y AUC.

- Validar la interpretabilidad de los modelos de clasificación evaluando la contribución de las variables de entrada y su correlación con el comportamiento experimental de las formulaciones electrolíticas.

1.6. Conceptos clave

Para facilitar la comprensión del presente trabajo, se definen conceptos técnicos de la investigación:

- **Eficiencia Coulómbica (CE):** Representa la relación cuantitativa entre la capacidad de energía extraída durante la descarga y la capacidad suministrada en la carga previa. Se define matemáticamente como:

$$CE = \frac{Q_{\text{descarga}}}{Q_{\text{carga}}} \times 100\,\% \tag{1.1}$$

  $Q_{descarga}$ es la cantidad de carga extraída de la batería durante la descarga.

  Q_{carga} es la cantidad de carga suministrada a la batería durante el proceso de carga.

- **Interfase sólido-electrolito:** Capa protectora, iónicamente conductora pero electrónicamente aislante, que se forma sobre la superficie del ánodo debido a la reducción del electrolito. Su estabilidad es el factor crítico que determina la eficiencia coulómbica y la reversibilidad de la batería (Hobold et al., 2021).

- **SMILES:** Sistema de notación estándar que representa la estructura y topología de las moléculas (como solventes y sales) mediante cadenas de texto.

- **Conductividad Iónica (σ):** Propiedad que influye directamente en la capacidad de transporte de iones litio dentro del electrolito. Una baja conductividad incrementa la resistencia interna, mientras que una alta mejora el transporte iónico.

- **Aprendizaje Automático:** Subárea de la inteligencia artificial centrada en algoritmos que permiten a las máquinas aprender automáticamente a partir de datos.

- **Baterías de Iones de Litio:** Dispositivo electroquímico que almacena energía mediante el movimiento reversible de iones de litio entre un ánodo y un cátodo.

- **Sobreajuste:** Sucede cuando un modelo de aprendizaje automático se ajusta con excesiva precisión a los datos de entrenamiento, incluyendo el ruido estadístico. Por lo tanto, el modelo pierde su habilidad de generalizar y no logra hacer predicciones confiables con datos nuevos o que no ha visto antes.

Capítulo 2

Marco Teórico

En este capítulo se presentan las bases teóricas que sustentan la investigación, integrando principios electroquímicos y herramientas de aprendizaje automático. Se tratan las bases de las baterías de iones de litio, con énfasis en los electrolitos líquidos y en mediciones fundamentales como la conductividad iónica y la eficiencia coulómbica. Más adelante, se describen conceptos fundamentales de aprendizaje automático, como la optimización de hiperparámetros, las métricas de evaluación, los algoritmos de clasificación y la interpretabilidad. Por último, se muestra el estado del arte de su utilización en baterías, resaltando la capacidad de estos métodos para convertir descriptores fisicoquímicos en modelos predictivos que posibilitan analizar y mejorar el rendimiento de los electrolitos.

2.1. Fundamentos de baterías de iones de Litio

Las baterías de iones de litio han revolucionado el almacenamiento de energía gracias a su altísima densidad de energía y a su elevado rendimiento energético. La función operativa de una celda electroquímica se basa en tres partes fundamentales: el ánodo, el cátodo y el electrolito, el cual cuenta con un separador microporoso. El ánodo o electrodo negativo consiste, normalmente, en materiales grafíticos, que permiten de esta forma la intercalación de los iones litio. En el caso del cátodo (electrodo positivo) está normalmente formado por un óxido de metal. Al realizar la descarga, se convierte en el ánodo en el que los átomos de litio se oxidan, generándose electrones al circuito externo que fluyen a través del electrolito hacia el cátodo.

Para que la celda opere de manera eficiente, los colectores de corriente (típicamente láminas de cobre para el ánodo y aluminio para el cátodo) actúan como soporte mecánico de los materiales

activos y proveen la conducción electrónica hacia el circuito externo (Nasajpour-Esfahani et al., 2024). Internamente, el electrolito sirve como el medio de transporte iónico. Simultáneamente, el separador garantiza la integridad del sistema al funcionar como una barrera física que impide el contacto directo entre los electrodos, evitando así cortocircuitos internos y permitiendo de manera exclusiva el flujo selectivo de los iones.

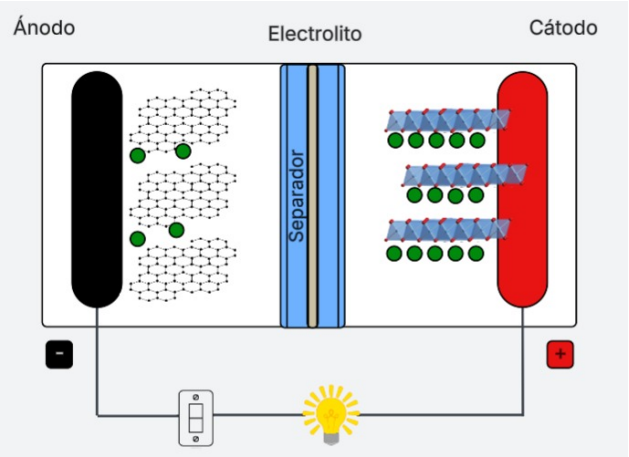


Figura 2.1: Representación esquemática de una celda de iones de litio y el flujo de partículas durante la descarga.

El funcionamiento de la celda se basa en la transferencia coordinada de partículas como se ilustra en la Figura 2.1. Mientras que los iones de litio se desplazan internamente a través del electrolito y el separador, los electrones son forzados a circular a través de un circuito externo.

Este fenómeno ocurre porque el separador actúa como un aislante electrónico, impidiendo el paso directo de electrones entre los electrodos. Durante la descarga, el ánodo se oxida y libera electrones que viajan por el circuito externo hacia el cátodo para compensar la carga de los iones que llegaron por el interior. Este flujo de electrones es, precisamente, la corriente eléctrica que alimenta los dispositivos.

En la Figura 2.1 se puede apreciar visualmente esta arquitectura: el ánodo (negativo) a la izquierda, el cátodo (positivo) a la derecha, y el sistema de interruptor y carga (bombilla) que representa el camino externo por donde circulan los electrones.

Manthiram (2017) nos provee una revisión de los avances tecnológicos en las baterías de iones de litio, señalando puntos relevantes como lo son la seguridad del almacenaje de la energía y la optimización de los materiales utilizados en su fabricación. En ella el autor recalca la importancia de una buena estrategia para estabilizar la interfaz electrodo-electrolito así como la búsqueda de sistemas de reacción de conversión con alta capacidad, teniendo como sus principales resultados

la identificación de la estabilidad de la SEI como la clave principal para la viabilidad de baterías futuras sumando a esta la búsqueda de rebasar las limitaciones de las estructuras como por ejemplo cambios masivos en el volumen en ánodos de conversión de silicio de hasta un 400 %.

Pese a que la arquitectura más novedosa presenta una madurez comercial suficientemente amplia, las exigencias de mayores densidades de energía han mostrado sus carencias de los mecanismos de intercalación convencional. La definición de la frontera de investigación frente a esta limitación se dirige hacia sistemas posteriores al litio con materiales activos de elevada capacidad a partir de reacciones de aleación y conversión. No obstante, tal y como analizan Choi y Aurbach (2016), la inclusión de estas nuevas químicas se encuentra notablemente condicionada por el bajo grado de reversibilidad electroquímica que presentan. Los materiales para el ciclado en alternancia encabezados por este tipo de reacciones experimentan transiciones de fase y cambios volumétricos muy intensos que desestabilizan el contacto con el electrolito y previenen la formación de láminas de pasivación que sean lo suficientemente estables, condicionando un contacto más propenso a toponos de otros eventos paralelos. De ahí que la viabilidad de estas baterías de nueva generación dependa en buena medida del incremento de dicha inestabilidad; un reto que otorga al diseño del electrolito un papel fundamental en la ecuación. A fin de gestionar dicho reto de forma sistemática y conseguir la degradación de la interfaz, se impone la comprensión de los límites de primer orden que regulan las interacciones moleculares dentro de la celda. Desde un punto de vista termodinámico, Goodenough y Park (2013) apuntan que la inestabilidad de los elementos que componen a una batería es fijada por las energías de los orbitales moleculares del electrolito en función de los potenciales galvánicos de los electrodos. Ellos afirmaban que el voltaje de la celda electrolítica se hallaba limitado por la llamada ventana del electrolito, caracterizada por la diferencia de energía de los orbitales moleculares entre el más bajo desocupado (LUMO) y el más alto ocupado (HOMO). Cuando el potencial del ánodo se encuentra por encima del nivel de energía del LUMO del electrolito como sucede en sistemas con litio metálico, se produce la reducción termodinámica del solvente y la interfase sólido-electrolito (SEI) es generada. La pasivación de la misma resulta necesaria para evitar la reacción entre el electrodo y el solvente continua, pero el coste electroquímico de la misma es alto al despojar de litio irreversible (menor eficiencia coulómbica) y generar la impedancia interfacial que limita el transporte iónico.

2.1.1. Composición y Naturaleza de los Electrolitos Líquidos

Función y requisitos generales

El electrolito forma parte fundamental de una batería de iones de litio, situada entre el ánodo y cátodo, su principal funciones es permitir el paso de los iones de un lado a otro dentro de la celda (dependiendo si la batería está almacenando o liberando energía), sin que los electrones lo hagan ya que estos se obligan a fluir unicamente por el campo exterior, realizando así la función de una aislante y así garantizar la movilidad iónica manteniendo la seguridad operativa de la celda. Para atender a estas exigencias funcionales, este debe cumplir con estrictos requerimientos cinéticos y termodinámicos. Según Deng et al. (2026), para reducir la polarización por concentración y mantener las exigencias de carga rápida, es esencial que el líquido tenga una conductividad iónica alta en un rango extenso de temperaturas. Además, su investigación enfatiza que para evitar la descomposición oxidativa y reductiva en los electrodos, el sistema necesita una extensa ventana de estabilidad electroquímica. Como Goodenough y Park (2013) establecieron, esta tolerancia al voltaje se basa en la brecha energética entre el orbital molecular más bajo desocupado (LUMO) y el orbital molecular más alto ocupado (HOMO).

Matriz de solventes orgánicos

La matriz de solventes orgánicos funciona como el medio líquido del electrolito; su objetivo principal es disolver la sal de litio y permitir que los iones se muevan libremente de un electrodo al otro. Para lograr eso de manera efectiva, los solventes deben mantener una relación equilibrada entre dos características físicas opuestas: por un lado, deben poseer capacidad de disociar la sal en iones (alta permitividad dieléctrica) y, por otro, deben ser relativamente fluidos (baja viscosidad) para facilitar el viaje de los iones a través de la celda. Basando en este equilibrio se puede definir como las moléculas provenientes del solvente y que rodean el ión de litio supone una relación entre la velocidad del transporte de iones y la durabilidad de la batería.

De Blasio et al. (2024) en su investigación, señalan cómo los electrolitos comerciales tienden a emplear combinaciones de disolventes que se basan en carbonatos (por ejemplo, DMC, DEC o EC) para optimizar la conductividad iónica y equilibrar las propiedades mencionadas. Nasajpour-Esfahani et al. (2024) Menciona que, a pesar de que estos carbonatos toleran bien el voltaje, su reactividad química disminuye la vida útil de la batería. Por este motivo, Hobold et al. (2021) subrayan que el uso de solventes a base de éteres es termodinámicamente más estable frente al litio,

lo cual permite incrementar la eficiencia coulómbica al reducir las reacciones no deseadas.

Sales de Litio

Las sales de litio constituyen la fuente mayoritaria de portadores de carga en el electrolito y posibilitan que fluya la corriente necesaria para que la batería funcione electroquímicamente. Para asegurar un transporte eficaz, es necesario que la sal se disocie por completo en la matriz orgánica, lo que impide que se formen pares iónicos o agrupaciones que eleven la resistencia interna. La naturaleza química del anión, al margen de su impacto en la conductividad iónica, tiene un efecto directo sobre la formación y estabilidad de las capas pasivantes que se encuentran en los electrodos. Entender la reactividad y la dinámica de disociación de estas sales posibilita explicar por qué los descriptores estructurales y la concentración del soluto son variables predictivas fundamentales para el diseño y desarrollo de los modelos de aprendizaje automático que se presentan en este estudio.

El hexafluorofosfato de litio (LiPF6) ha permanecido como el soluto estándar en aplicaciones comerciales por su balance funcional entre estabilidad y conductividad (Yan et al., 2025). Pero observando más profundamente el panorama, la investigación actual se ha centrado más en sales alternativas con imidas, como el LiFSI y el LiTFSI. Trabajos de Weng et al. (2024) y Xue et al. (2020) informan que el anión FSI^- se descompone con preferencia para que se forme una interfase sólida rica en LiF, la cual es mecánicamente resistente, evita el crecimiento de dendritas y mejora la eficiencia coulómbica de la celda.

2.1.2. Eficiencia Coulómbica

La eficiencia coulómbica (CE) se utiliza como un evaluador de la celda en términos de reversibilidad, pudiendo tener con esta un indicador tanto termodinámico como cinético de la celda. Dicha métrica se puede expresar como el porcentaje obtenido de la capacidad de descarga dividido entre la capacidad de carga completado un ciclo de la celda 1.1.

En el caso particular de las baterías de litio metálico, valores elevados de eficiencia coulómbica están asociados con una reducción en la formación de litio irreversible y con una menor degradación del electrolito. Alcanzar valores próximos al 100 % es esencial para asegurar una larga vida útil de la batería, puesto que incluso pérdidas mínimas repetidas a lo largo de los ciclos conducen a un consumo acelerado del litio activo y a la degradación del electrolito (Hobold et al., 2021).

Para comprender estas limitaciones y superar la barrera técnica del 99.9 % de CE, es indispensable cuantificar microscópicamente el origen de esta pérdida de litio activo. Como lo mencionan

Hobold et al. (2021), mediante técnicas analíticas como la cromatografía de gases de titulación, la literatura ha establecido que en electrolitos de baja eficiencia (<95 %), la pérdida de capacidad está dominada por litio metálico encapsulado y desconectado electrónicamente. Por el contrario, en regímenes de alta eficiencia, la ineficiencia se deriva predominantemente de los iones de litio atrapados durante la formación de la SEI. Como se ilustra en la Figura 2.2, esta transición de un régimen dominado por litio muerto a uno dominado por la SEI se correlaciona directamente con la morfología del depósito. Las reconstrucciones tridimensionales obtenidas mediante haz de iones enfocados criogénico revelan que los electrolitos de baja eficiencia generan depósitos con alta porosidad (barras color verde azulado que representan hasta un 17 %), mientras que, al transitar a regímenes de alta eficiencia, la porosidad se reduce drásticamente a menos del 1 %, dando lugar a depósitos de litio densamente empaquetados. Empíricamente, la optimización física de estas variables ha permitido resultados relevantes, como el uso de electrolitos de gas licuado que han logrado reportar de forma excepcional un 99.992 % de CE.

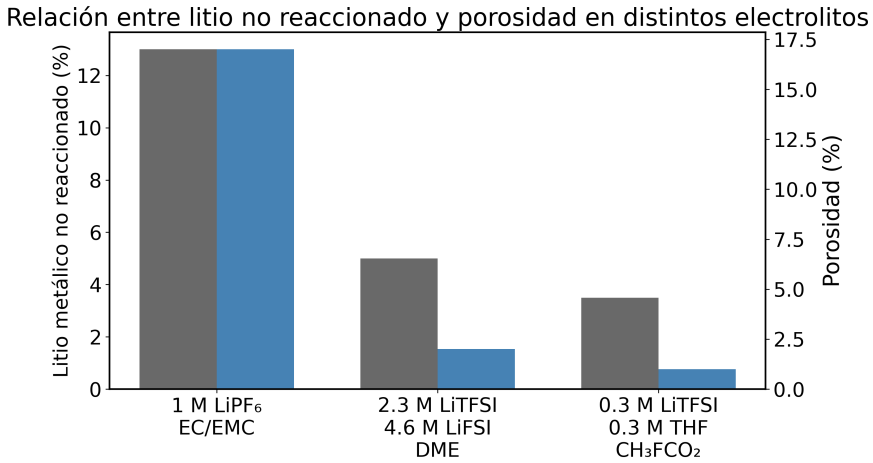


Figura 2.2: Cuantificación de la composición del litio electroquímicamente inactivo. En la gráfica las barras grises representan la cantidad de litio metálico muerto, y las barras de color verde azulado representan el porcentaje de porosidad. Adaptado de Hobold et al. (2021).

Para mitigar la pérdida de litio activo y estabilizar la reversibilidad de las celdas, la investigación experimental se ha enfocado en el diseño químico de los electrolitos. Desde un enfoque físico-químico, se ha intentado contrarrestar la inestabilidad intrínseca del sistema mediante el desarrollo de interfases artificiales protectoras que regulen el depósito del metal. Como documentan Xu et al. (2019) en su revisión sobre el tema, estrategias híbridas como el uso de capas compuestas por PVDF-HFP y LiF han logrado estabilizar la deposición y reportar eficiencias coulómbicas superiores al 99.2 %. En este mismo sentido, la implementación de interfases avanzadas de conductores mixtos ha permitido

sostener de manera excepcional eficiencias del 99.5 % a lo largo de 500 ciclos.

Los electrolitos de alta concentración esperan estabilizar la reversibilidad del litio, dado que modifican la red de solvatación y la cinética de pasivación, la cual genera decrementos en la cantidad de litio activo. En la modalidad experimental, esta forma de proceder iónico puede controlar la tasa de las velocidades de reacción y suprimir las reacciones parásitas. Según el estudio de Qian et al. (2015), un sistema de 4 M de LiFSI en DME obtiene un depósito libre de dendritas gracias a la fuerte coordinación solvente-catión. Los resultados del trabajo de Qian son sobresalientes bajo regímenes en los que la exigencia es elevada: se registran unas celdas asimétricas de Cu‖Li a 4 mA cm^{-2} que superan los 1000 ciclos con un promedio de 98.4 % de eficiencia coulómbica, al igual que se experimentan celdas simétricas de Li‖Li que son capaces de operar en 10 mA cm^{-2} hasta los 6000 ciclos sin presentar cortocircuitos. Estos resultados apuntan a que la química de solvatación es el determinante del límite de eficiencia de la celda, estableciendo el fundamento teórico que hace necesario recurrir a métodos computacionales para estudiar estas formulaciones de forma sistemática, del modo menos dependiente que sea posible para la validación empírica.

El diseño molecular de nuevos solventes orientado a la optimización de la pasivación interficial invita a considerar el uso de descriptores moleculares como predictores en los modelos desarrollados. En la misma dirección del diseño molecular, Xue et al. (2020) se propusieron el diseño de un çompleto fluorosulfonilo"partiendo del disolvente fluoruro de N,N-dimetilsulfamoílo, que es la estructura que se deriva del anión FSI y cuya investigación computacional fue realizada a partir de la Teoría del Funcional de la Densidad (DFT), demostrando una sobresaliente estabilización de la deposición — en su estudio se informa de una eficiencia coulómbica de un 91 % en el estado inicial, alcanzando el 99 % tras 10 ciclos y manteniendo un promedio de 99.03 % durante 400 ciclos a densidad de corriente de 0.5 mA cm^{-2}.

Gracias al diseño de electrolitos avanzados y con arquitecturas estructurales, se permite mejorar la estabilidad del litio bajo condiciones operativas extremas. Desde una óptica orientada a la estabilidad térmica y a la tolerancia de voltajes extremos, Weng et al. (2024) dieron cuenta de que los líquidos iónicos de solvente único con un anión dual (LiFSI y EMIMTFSI) eran capaces de estabilizar la deposición de litio en condiciones extremas. En ese estudio se reportó una conductividad iónica de 2.8 mS cm^{-1} a 25 °C y una eficiencia coulómbica (CE) media del 99 % tras 300 ciclos en celdas Li‖LiFePO4, marcando así umbrales termodinámicos como referencia para la validación de la clasificación de la conductividad, marcando la alta capacidad de almacenamiento.

De forma complementaria a la química del electrolito, el diseño físico de los electrodos proporciona las soluciones mecánicas idóneas para controlar la morfología del depósito de litio. De hecho, Yang et al. (2022) realizaron una película híbrida multidimensional (MD-HFF) que mediante el dopaje *in situ* con yodo favoreció a la formación de LiI que minimiza la barrera de migración de los iones de litio. Experimentalmente, dicha modificación favoreció la estabilización de la deposición (CE del 99 % durante así mas de 230 ciclos a 1 $\mathrm{mA\,cm^{-2}}$ y 1 $\mathrm{mAh\,cm^{-2}}$). La parametrización de éstas variaciones del transporte iónico resulta clave para que los algoritmos realicen una evaluación matemática de cómo se mitiga la expansión volumétrica y la degradación interfacial.

Por último, en el objetivo de tolerar condiciones cinéticas aún más exigentes para aplicaciones de carga rápida, Gong et al. (2022) pusieron en práctica anfitriones de carbono duro presentando nanoestructuras porosas en forma de flor. Su aportación permitió operar ánodos híbridos de ion-litio/litio metálico con una eficiencia coulómbica del 99.5 % bajo densidades de corriente extremas -de $\mathrm{16mA\,cm^{-2}}$. Es así como estos indicadores revelan que la estabilidad del medio conductor y la deposición uniforme en estructuras huésped, son rutas eficaces para sostener la reversibilidad cinética; a su vez, refuerzan la base de datos de que se extraen características estructurales que los modelos de clasificación pueden utilizar para predecir fallos en condiciones de alta tasa de corriente.

La validación experimental de la eficiencia coulómbica requiere protocolos de medida estrictos para suprimir sesgos analíticos, y certificar la accesibilidad termodinámica de los materiales. Gracias a la implementación de procedimientos de macrociclado galvanostático sobre ánodos de magnesio metálico, Attias et al. (2022) pusieron de relieve que la estabilidad a largo plazo sólo se puede definir eliminando ambigüedades operativas de técnicas tradicionales, y lograron reportar eficiencias sostenidas del 99.59 %.

2.1.3. Conductividad Iónica

Así como la eficiencia coulómbica, la conductividad iónica controla la capacidad cinética del electrolito estableciendo en este qué tan rápido puede una batería cargarse y descargarse, y la única manera para poder estimar por medio del aprendizaje automático esta transferencia de iones, es entendiendo la física que lo rige, es por ello que comprender estas exigencias es primordial para obtener descriptores cuantitativos exactos que posibiliten a los algoritmos aprender a clasificar correctamente nuevas formulaciones no exploradas experimentalmente.

Algo importante a mencionar en la búsqueda de la maximización del nivel de conductividad de

un electrolito es la necesidad de que la sal de litio se disuelva por completo. De acuerdo con lo que De Blasio et al. (2024) han expuesto, este proceso requiere que la matriz de solventes orgánicos ofrezca un ambiente con elevada permitividad eléctrica, característica esencial para disolver la sal y dividir los pares iónicos. Al mismo tiempo, el medio debe tener una viscosidad baja para que la resistencia mecánica disminuya y los iones libres puedan moverse por todo el sistema. Esta conexión inherente entre las características del solvente y la movilidad de los iones establece los fundamentos teóricos que explican la elección de las variables reológicas y estructurales al crear modelos predictivos.

Para seguir comprendiendo las propiedades de transporte, es esencial examinar la influencia termodinámica que tiene la saturación de sal en la movilidad iónica. A pesar de que uno podría pensar que un aumento en la cantidad de portadores de carga optimizaría el flujo de corriente, los electrolitos presentan una dinámica molecular extremadamente no lineal que contradice las normas del diseño empírico convencional. Qian et al. (2015) comprobaron empíricamente esta restricción mediante la evaluación de sistemas basados en LiFSI en dimetoxietano (DME). Su estudio muestra que, cuando la concentración de sal aumenta de 1 M a 4 M, la viscosidad del medio se incrementa rápidamente, lo que provoca una disminución significativa de la conductividad iónica, pasando de 16,9 mS cm^{-1} a 5,7 mS cm^{-1}. Este fenómeno se explica porque la modificación de la red de solvatación favorece la formación de pares iónicos y agregados moleculares que limitan la libertad de movimiento de los iones.

Investigar líquidos iónicos ha aparecido como una vía prometedora para modificar la química de coordinación del electrolito con el fin de sobrepasar las estrictas limitaciones cinéticas provocadas por la elevada concentración de sales convencionales. Aunque estos compuestos son conductores excepcionales cuando están en su estado puro, la manera en que se transportan se torna extremadamente complicada al interactuar directamente con los portadores de carga. Como explican Appetecchi et al. (2012), la incorporación de sales de litio a los líquidos iónicos produce una disminución notable en la conductividad intrínseca del medio. Esto sucede porque se establecen vínculos intermoleculares sólidos entre el catión litio y los aniones existentes, formando conglomerados que aumentan la viscosidad y limitan el flujo. Con el objetivo de contrarrestar este fenómeno y promover la movilidad de iones, la estrategia física ha buscado optimizar las mezclas diluyéndolas con solventes de permitividad alta. En esta línea, Ferasso et al. (2025) documentan en su revisión crítica que la adaptación meticulosa de la fracción molar de líquidos iónicos posibilita la ruptura de las

interacciones coulómbicas dañinas y se pueden obtener conductividades excepcionales, llegando a 40 mS cm^{-1} en condiciones térmicas favorables.

La temperatura tiene un rol esencial en las características de transporte de los electrolitos, ya que cambia la viscosidad del medio y la movilidad iónica. Su aumento brinda energía extra al sistema, lo que disminuye la resistencia al flujo y promueve el desplazamiento de los portadores de carga. En un estudio experimental realizado por Weng et al. (2024), se demostró esta influencia en líquidos iónicos de solvente único; hallaron que la conductividad iónica pasó de 2.8 mS cm^{-1} a 25 °C a 9.0 mS cm^{-1} a 80 °C, lo cual fue consecuencia de la reducción de la resistencia reológica. Estos hallazgos demuestran que la energía térmica altera la red de coordinación iónica y promueve el transporte de carga, lo cual hace que la temperatura sea un factor crucial para el funcionamiento electroquímico de los electrolitos.

2.2. Aprendizaje Automático

2.2.1. Algoritmos de Aprendizaje Supervisado

El aprendizaje supervisado, que se basa en el entrenamiento de modelos con conjuntos de datos etiquetados anteriormente para prever la salida ante nuevas entradas, es la modalidad más común del aprendizaje automático (Valizadeh y Amirhosseini, 2024).En la fase inicial, o de entrenamiento, el algoritmo examina un conjunto de datos que incluye muestras de entrada y sus etiquetas asociadas para detectar patrones subyacentes, convirtiendo dicha información en un modelo matemático. Luego, en la etapa de inferencia, se utiliza este modelo para hacer predicciones exactas sobre datos nuevos que no se analizaron durante el proceso de entrenamiento. El objetivo matemático de este proceso es aprender la función de mapeo exacta entre las características de entrada y sus salidas correspondientes. Las tareas analíticas más comunes dentro de este enfoque son la regresión y la clasificación. En los problemas de clasificación, por ejemplo, el algoritmo aproxima el comportamiento de una función capaz de asignar un vector de datos a una de varias clases predefinidas. Entre las técnicas e implementaciones algorítmicas más utilizadas bajo este esquema de aprendizaje en la ciencia de materiales se encuentran la Regresión Lineal, la Regresión Logística, los Árboles de Decisión, los Bosques Aleatorios (Random Forests), las Máquinas de Vectores de Soporte (SVM) y las Redes Neuronales (Mobarak et al., 2023). El flujo metodológico de este paradigma se puede representar mediante un proceso secuencial donde los datos de entrenamiento estructuran el algorit-

mo para crear un modelo clasificador capaz de evaluar datos no observados previamente, tal como se esquematiza en la Figura 2.3.



Figura 2.3: Diagrama de flujo del proceso de aprendizaje supervisado. El esquema ilustra cómo los datos etiquetados se utilizan en la fase de entrenamiento para construir un modelo matemático, el cual posteriormente genera predicciones en la fase de inferencia. Adaptado de Mobarak et al. (2023)

2.2.2. Algoritmos de Clasificación

Un algoritmo de clasificación consiste en un procedimiento de aprendizaje supervisado que distribuye datos con múltiples dimensiones a categorías discretas. Según indica Macwan (2024), estos modelos posibilitan que se reconozcan patrones complejos y se respalde la toma de decisiones. En la investigación de electrolitos para baterías de iones de litio, las métricas que miden la eficiencia coulómbica y la conductividad iónica tienen relaciones no lineales, lo cual restringe el rendimiento de métodos tradicionales como K-Nearest Neighbors o la Regresión Logística. Para tratar esta complejidad, se utilizan métodos de aprendizaje de conjunto, que consisten en la combinación de varios clasificadores para aumentar la capacidad predictiva. Los métodos de Bagging, que buscan disminuir la varianza, y los de Boosting, que se centran en corregir errores secuencialmente para obtener una representación más sólida del espacio de materiales, son algunos de ellos. Esta arquitectura predictiva, como se muestra en la Figura 2.4, funciona según el principio de funciones aditivas. El resultado final para un caso específico es la suma o el promedio de las puntuaciones que cada árbol independiente calcula Chen y Guestrin (2016). Esta suma matemática en el análisis de electrolitos es lo que posibilita que el modelo otorgue un peso particular a varios descriptores atómicos sin depender de un solo árbol, que podría estar sujeto a errores

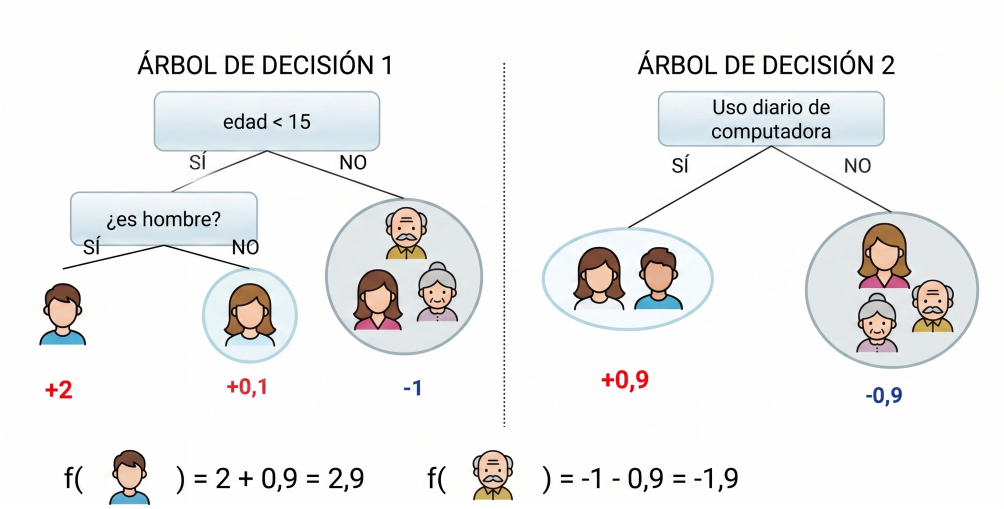


Figura 2.4: Modelo de ensamble de árboles. La predicción final se determina a partir de la suma de las contribuciones realizadas por cada árbol que compone el modelo. Adaptado de Chen y Guestrin (2016)

En el estado del arte en aprendizaje de conjunto aplicado a datos tabulares, la familia Bagging tiene un perfecto ejemplo en la arquitectura Random Forest, que reduce el sobreajuste al entrenar por separado árboles de decisión en subespacios aleatorios. En su estudio comparativo, Bentéjac et al. (2021) mostraron que este método es extraordinariamente estable y no necesita ajustes de hiperparámetros, promediando sus predicciones espaciales para lograr una gran capacidad de generalización. No obstante, su investigación también estableció que, con respecto a las dinámicas no lineales más complejas, los procedimientos de Gradient Boosting consiguen una discriminación predictiva superior al edificar árboles secuencialmente para reducir de manera iterativa los errores residuales.

Stumpfe y Shongwe (2026) informaron que XGBoost es especialmente buena en entornos experimentales ruidosos y funciones objetivos complejas, lo que puede favorecer a la mejora de esta propuesta matemática. Los autores enfatizan que este algoritmo incluye una formal penalización en la función de pérdida, lo que permite tener un previo control metodológico sobre una complejidad estructural del árbol para combatir la memorización de este tipo de datos anómalos que provienen del laboratorio. La evolución de esta tipología algorítmica nos lleva a alcanzar un rendimiento computacional incomparable con LightGBM. Ke et al. (2017) mencionan que esta arquitectura realiza una transformación del entrenamiento practicando una estrategia de crecimiento foliar vertical (leaf-wise) que combina con el Muestreo Unilateral Basado en Gradientes (Gradient-Based One-Side Sampling [GOSS]), un mecanismo creado para poder preservar solamente ejemplos informativos y

agrupar descriptores dispersos. En el marco del presente trabajo, esta evolución teórica difunde que el uso de ensamblajes avanzados es la metodología más sobrada para poder identificar descriptores asociados a una conductividad iónica ideal y una alta eficacia coulómbica, con lo que poder superar las limitaciones de los modelos predictivos tradicionales.

2.2.3. Optimización de hiperparámetros de modelos de clasificación

La evaluación de la conductividad y la eficiencia coulúmbica por medio del electrolito usando modelos de aprendizaje automático requiere elecciones apropiadas de algoritmos que puedan reproducir adecuadamente estas interacciones no lineales. Dadas estas condiciones, las configuraciones por defecto tienden a dar lugar a sobreajuste o, en su defecto, a una escasa generalización a nuevas formulaciones. Por eso, la optimización de hiperparámetros ha pasado de ser empírica a constituirse en una metodología más eficiente para la exploración del espacio de búsqueda. Este proceso permite encontrar configuraciones que puedan reflejar las relaciones más necesidades entre las variables del sistema, lo que permite mejorar la precisión de las predicciones garantizando que los modelos obtenidos sean estadísticamente pequeños y sean la correcta representación de los fenómenos que condicionan el output de las baterías.

La optimización de hiperparámetros en el aprendizaje automático ha ido progresando desde técnicas exhaustivas hasta procedimientos que resultan más eficaces. De acuerdo con Subaşı (2024), una de las técnicas más usadas, llamada Grid Search, examina de manera sistemática todas las combinaciones que existen de los parámetros; sin embargo, su costo computacional se incrementa rápidamente a medida que la dimensionalidad del problema crece. La búsqueda aleatoria (Random Search), por otra parte, disminuye este costo al examinar el espacio de forma probabilística, pero sin utilizar la información adquirida en evaluaciones anteriores. Rom et al. (2025) proponen la optimización bayesiana como una de las metodologías más efectivas para sobrepasar estas restricciones, porque a partir de un modelo probabilístico de la función objetivo, emplea ese conocimiento para orientar la búsqueda. Este método permite hallar configuraciones óptimas con un número limitado de evaluaciones caras, equilibrando la explotación y la exploración.

En este modelo probabilístico, Optuna es la herramienta que logra la mayor madurez analítica del ecosistema de herramientas. Akiba et al. (2019) evidenciaron que este marco de trabajo de última tecnología transforma el mapeo paramétrico a través de una arquitectura de diseño por ejecución, que posibilita la estructuración nativa de espacios de búsqueda dinámicos. La figura 2.5 reseña tal

arquitectura eficiente; detalla el procedimiento que los distintos procesos de cómputo pueden seguir para explorar instancias de la función objetivo en una estrategia asíncrona y paralela; la forma de compartir el historico de los resultados a través de una memoria central. Este trabajo resaltó la aplicación del TPE (Tree-structured Parzen Estimator) con técnicas de poda secuencial que permiten descartar previamente configuraciones no prometedoras acelerando así la búsqueda. En Optuna, este concepto es adaptado para optimizar eficientemente los hiperparámetros de modelos complejos, eligiendo el mejor compromiso entre la mayor precisión y un coste computacional soportable. Esto resulta en la posibilidad de identificar configuraciones con capacidad de captar relaciones no lineales, mejorando el rendimiento predictivo sin la necesidad de tener un régimen de exploración exhaustiva del espacio de búsqueda.

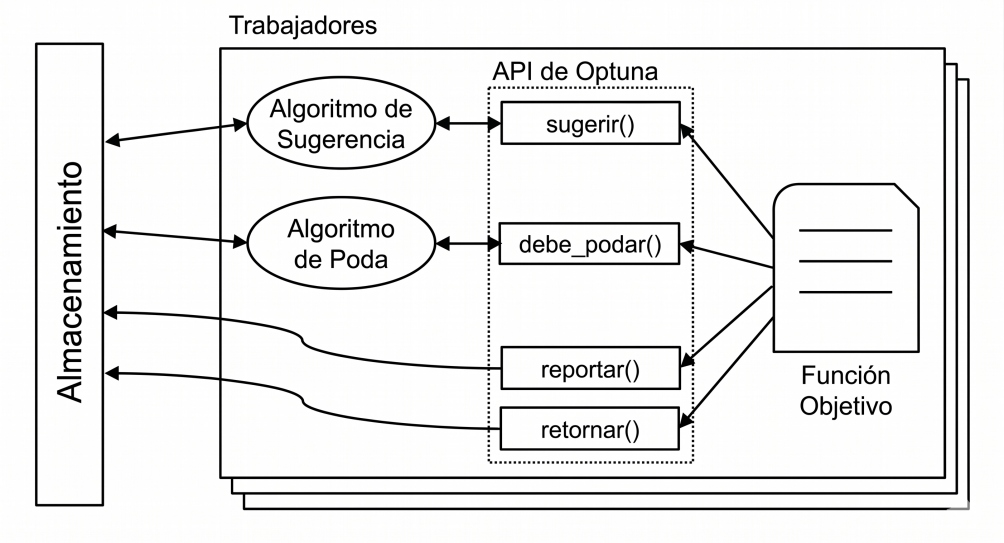


Figura 2.5: En Optuna, cada trabajador ejecuta de manera independiente una instancia de la función objetivo dentro de un estudio. A través de las API del sistema, las pruebas pueden consultar y actualizar un almacenamiento compartido que contiene la información de evaluaciones anteriores. Este esquema permite que múltiples trabajadores colaboren en un mismo proceso de optimización, compartiendo continuamente el avance y los resultados obtenidos. Adaptado de Akiba et al. (2019)

2.2.4. Métricas de evaluación de modelos de clasificación

Según Chicco y Jurman (2020), el rendimiento de un algoritmo de clasificación se organiza y evalúa mediante una matriz de confusión (Figura 2.6), que segmenta los resultados en cuatro grupos básicos: verdaderos positivos (TP), falsos negativos (FN), verdaderos negativos (TN) y falsos positivos (FP). De estas categorías, se obtienen métricas estadísticas concretas para medir cuantitativamente la calidad de las predicciones:

- **Exactitud (Accuracy):** Se refiere a la proporción de casos clasificados correctamente en comparación con el total de muestras. Matemáticamente, se establece como:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precisión (Precision):** Determina la proporción de predicciones positivas que son verdaderamente correctas. Se calcula mediante:

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Sensibilidad (Recall o Tasa de Verdaderos Positivos):** Mide el porcentaje de casos positivos reales que el modelo identifica correctamente, calculado como:

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1-Score:** Corresponde al promedio armónico entre la precisión y la sensibilidad. Se define mediante:

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN}$$

donde su valor se encuentra en el intervalo $[0, 1]$. Aunque es una métrica ampliamente utilizada, no considera los verdaderos negativos y no es simétrica frente a la permutación de clases.

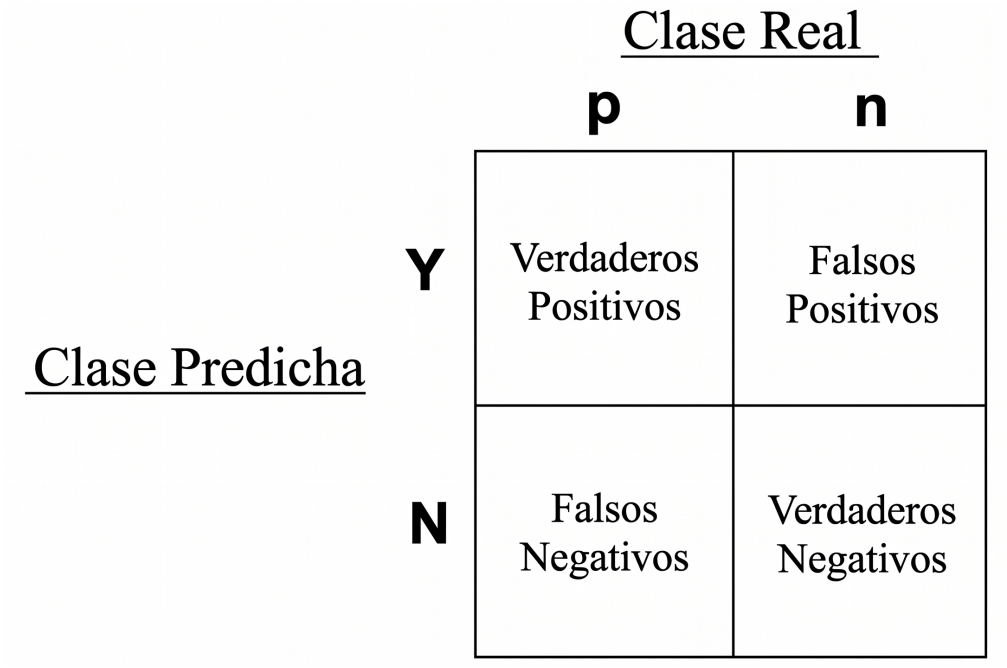


Figura 2.6: Diagrama común de una matriz de confusión

Curvas ROC (Receiver Operating Characteristics): se emplean para analizar el desempeño del modelo en diversos niveles de decisión. Según Fawcett (2006), la curva ROC es un diagrama en dos dimensiones que representa el compromiso entre los costos y las ventajas de un clasificador (ver Figura 2.7); se dibuja la tasa de verdaderos positivos a lo largo del eje Y y la tasa de falsos positivos a lo largo del eje X. Para comparar varios modelos utilizando un único valor escalar, se determina el Área Bajo la Curva ROC (AUC). El AUC oscila entre 0 y 1.0; si es igual a 0.5, significa que se está utilizando una táctica de adivinación al azar. El AUC, en términos estadísticos, es la probabilidad de que una muestra positiva seleccionada aleatoriamente tenga una calificación más alta que una negativa también elegida al azar.

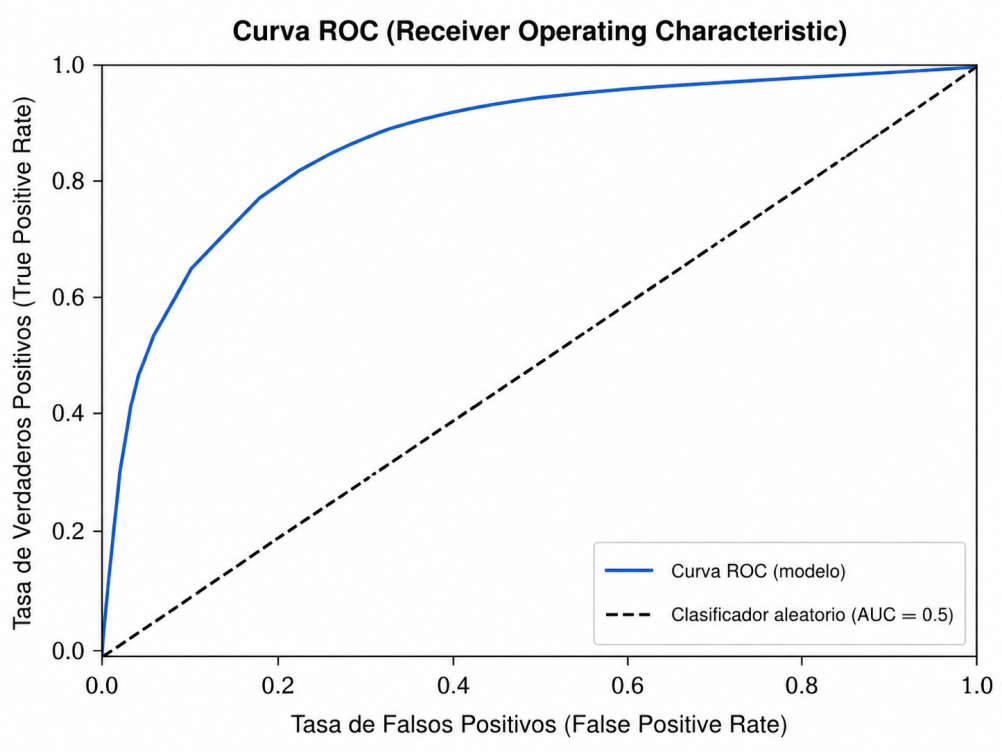


Figura 2.7: Diagrama común de una curva ROC: El gráfico muestra la relación dinámica entre la Tasa de Verdaderos Positivos y la Tasa de Falsos Positivos. La línea azul sólida representa el clasificador propuesto, mientras que la línea punteada negra define el umbral de una clasificación puramente aleatoria (AUC = 0,5).

2.2.5. Interpretabilidad de modelos de aprendizaje automático

En el diseño de baterías de iones de litio y en la ciencia de materiales, un modelo predictivo tiene que explicar los principios químicos y físicos que respaldan sus decisiones además de proporcionar una precisión elevada. No obstante, las arquitecturas de aprendizaje de conjunto más sofisticadas, empleadas para modelar las complejas relaciones que definen la conductividad iónica y la eficiencia coulómbica, tienden a actuar como sistemas de caja negra. Los métodos post-hoc crean explicaciones acerca de modelos previamente entrenados sin que su rendimiento se vea comprometido, a diferencia de los modelos intrínsecamente interpretables, que intercambian la capacidad para predecir por simplicidad. Así, se pueden determinar los descriptores fisicoquímicos más importantes, verificar el comportamiento del modelo y transformar sus pronósticos en información valiosa para la creación de nuevos electrolitos.

Entre las principales perspectivas que se encuentran en la literatura acerca de las explicaciones independientes del modelo (model-agnostic) sobresalen dos. Por una parte, Ribeiro et al. (2016) sugirieron LIME (Local Interpretable Model-agnostic Explanations), un método que se aproxima a la

conducta de un clasificador complejo alterando datos en torno a una instancia particular y generando un modelo lineal alternativo que es útil para examinar formulaciones individuales de electrolitos. Sin embargo, Lundberg y Lee (2017) presentaron SHAP (SHapley Additive exPlanations), que unifica metodologías anteriores bajo los fundamentos de la teoría de juegos cooperativos, con el objetivo de superar las debilidades en matemáticas y la inestabilidad de estas aproximaciones puramente locales. Su estudio preliminar mostró que SHAP proporciona propiedades teóricas fundamentales como la consistencia y la eficacia cuando permite obtener la correcta evaluación marginal de cada variable predictiva. De acuerdo con Muia et al. (2025), esta exactitud en el procedimiento de calidad convierte SHAP en una herramienta más desarrollada para entregar índices confiables cuando varía el espacio de datos. También es destacable como S. Saliu et al. (2025) dejan claro que la implementación de esto se traduce en el cumplimiento del punto de transparencia necesario para validar arquitecturas de gran nivel de opacidad. La aplicación de la estructura SHAP en el ámbito de la presente investigación ofrece la herramienta para abrir la caja negra de ensamblajes como LightGBM o XGBoost y así poder ofrecer una adecuada interpretación de cómo las fracciones atómicas o las variables controlan tanto la estabilidad de la interfase como la movilidad de los iones de litio.

2.2.6. Aprendizaje Automático Aplicado a Baterías

La investigación científica ha sido revolucionada por la transición de los métodos de ensayo y error hacia el descubrimiento de materiales basado en datos. La complejidad del espacio químico hace que en los sistemas de almacenamiento de energía sea difícil crear nuevos materiales por medio de métodos convencionales. En estas circunstancias, los algoritmos de aprendizaje automático hacen posible el reconocimiento de relaciones y patrones no lineales. Esto ayuda a evaluar y clasificar el rendimiento de los electrolitos mediante sus propiedades moleculares, así como a prever métricas importantes como la conductividad iónica y la eficiencia coulómbica. Mobarak et al. (2023) describen la investigación basada en datos como el cuarto paradigma de la ciencia, enfatizando que el aprendizaje automático acelera la identificación de materiales al descubrir patrones estadísticos en datos con alta dimensionalidad. No obstante, su investigación también alerta acerca de las restricciones en cuanto a interpretabilidad que son intrínsecas a los modelos complejos de **caja negra**, lo que plantea dudas sobre su habilidad para generalizarse a partir de datos limitados. Esta observación es crucial desde el punto de vista metodológico para esta tesis, pues justifica la necesidad ineludible de emplear instrumentos de interpretabilidad algorítmica no solo con el fin de clasificar correctamente

las variables del electrolito, sino también para obtener una comprensión clara desde el punto de vista físico-químico acerca de los mecanismos que dirigen la predicción.

En particular, en el ámbito de los sistemas de iones de litio, Valizadeh y Amirhosseini (2024) indagaron sobre la influencia del aprendizaje automático durante todo el ciclo vital de las baterías. Muestran cómo técnicas como Bosques Aleatorios y Máquinas de soporte vectorial logran superar limitantes de escalabilidad, además de reducir costos computacionales a la hora de buscar electrolitos y materiales óptimos, apoyando de esta manera la aplicación de estos algoritmos en la predicción de clasificación del desempeño de estas soluciones líquidas y evitando de esta manera experimentación exhaustiva costosa, siendo una opción más eficaz.

Esta forma de ver la química se aplica tal cual a la formulación química de la celda, Donde Deng et al. (2026) afirman y sostienen que el diseño empírico clásico no es suficiente para optimizar unas características en oposición como la estabilidad interfacial o la conductividad iónica por la gran cantidad de espacio químico disponible. Los autores proponen que el diseño basado en datos constituyen una forma revolucionaria de poner solución a las relaciones no lineales que existen entre los descriptores de los componentes y el rendimiento del dispositivo. Por lo tanto, este método muestra la capacidad de los procedimientos fundamentados en datos para establecer una relación entre la composición de los electrolitos y su rendimiento electroquímico, lo que permite detectar formulaciones con posibilidades sin tener que apoyarse únicamente en experimentos exhaustivos.

Una vez establecido el diseño basado en datos como un paradigma en la ciencia de materiales, el reto siguiente es anticipar cómo evolucionarán las celdas bajo condiciones operativas. La medición directa de los procesos internos que determinan la degradación y el rendimiento de las baterías es dificultosa debido a su complejidad estructural y a la influencia de factores externos. Los métodos fundamentados en datos han aparecido en este contexto como un instrumento efectivo para calcular estados internos y anticipar cómo se desarrollará el sistema sin necesidad de hacer mediciones invasivas. Según Nasajpour-Esfahani et al. (2024), el ruido exterior y la temperatura del ambiente aumentan la complejidad de estas estimaciones; por ello, se han utilizado modelos de aprendizaje profundo para determinar el estado y la vida útil de las baterías. Estos progresos muestran la capacidad de la inteligencia artificial para examinar sistemas electroquímicos que son muy complejos y no lineales.

Ren y Du (2023) revisaron la implementación de algoritmos fundamentados en datos para esti-

mar el estado de salud (SOH) y el estado de carga (SOC) dentro del marco de la supervisión del rendimiento de las baterías. Su análisis revela que, en diversas condiciones operativas, arquitecturas de aprendizaje profundo como LSTM y Unidades Recurrentes Cerradas (GRU) tienen la capacidad de captar el carácter no lineal del deterioro y llegar a MAE por debajo del 1.5 %. Asimismo, enfatizan que no solamente la arquitectura utilizada, sino también la calidad de los datos de entrenamiento y una correcta sintonización de hiperparámetros determinan el rendimiento de estos modelos, subrayando así lo crucial que es disponer de bases de datos estandarizadas y confiables.

Severson et al. (2019) probaron que es posible prever la vida útil de las baterías a través de modelos fundamentados en datos, sin necesidad de tener un entendimiento explícito de los mecanismos que provocan su degradación. Utilizando redes elásticas y atributos obtenidos de las curvas de voltaje en los ciclos iniciales, fueron capaces de calcular la vida útil con un margen de error (MAPE) del 9.1 %. Asimismo, un modelo de clasificación detectó si las celdas sobrepasarían los 550 ciclos con un error de prueba del 4.9 % solamente, empleando datos de los cinco ciclos iniciales. Estos hallazgos demuestran que las señales de operación tempranas tienen la información necesaria para prever el rendimiento futuro de una batería.

Además, Kauwe et al. (2019) utilizaron SVM y Bosques Aleatorios para predecir la capacidad de descarga específica de materiales para baterías a partir de descriptores composicionales. Su modelo de bosque aleatorio, en el conjunto de prueba, logró un coeficiente de determinación (R^2) de 0.48 y un error absoluto medio de 15.7 mAh g^{-1}. A pesar de que los investigadores indican que la falta de estandarización en las informaciones experimentales afecta la exactitud de los modelos, sus hallazgos muestran el potencial de los algoritmos de ensamble para hallar materiales y subrayan lo relevante que es utilizar descriptores fisicoquímicos definidos con claridad y bases de datos organizadas.

La estabilidad y el rendimiento de las baterías se ven influidos directamente por la composición química del electrolito y la concentración de los grupos funcionales que lo componen, más aún en sistemas de alta tensión en los que las reacciones interfaciales aceleran su degradación. Yan et al. (2025) construyeron un modelo basado en datos para identificar electrolitos de alto voltaje en celdas de litio metálico, usando como variables de entrada la molaridad de los grupos químicos que aparecen en los disolventes (enlaces carbonato y éter). En su comparación de diferentes arquitecturas de aprendizaje automático, mostraron que los algoritmos de ensamble son muy buenos para describir la relación entre el rendimiento y la composición. En su caso concreto, el modelo de Bosque Aleatorio obtuvo un error cuadrático medio (MSE) de 0.0074.

Conforme los modelos fundamentados en datos se vuelven más precisos prediciendo el desempeño y la degradación de las baterías, surge el reto de entender los mecanismos que respaldan sus predicciones. Es crucial determinar qué interacciones y variables fisicoquímicas afectan el comportamiento del sistema, además de lograr una precisión elevada. Por lo tanto, estudios recientes han incluido modelos dirigidos por principios físicos y métodos de interpretabilidad que posibilitan entender la lógica interna de los algoritmos y obtener información valiosa para crear nuevos materiales. Al integrar principios físicos en los modelos de aprendizaje automático, es posible describir fenómenos complejos con menos dependencia de grandes cantidades de datos. Gu et al. (2022) evidenciaron esta perspectiva a través de las Redes Neuronales Informadas por la Física (PINNs), que incorporan las ecuaciones diferenciales del sistema en el proceso de entrenamiento. Mediante una cantidad reducida de datos, su modelo logró pronosticar la conducta de materiales con errores relativos en el rango de 10^{-4}. De acuerdo con estos hallazgos, la inclusión de limitaciones físicas es una táctica efectiva para aumentar la exactitud de los modelos en situaciones donde no hay suficiente información experimental.

La teoría de juegos cooperativos ha surgido como una posible solución para evaluar la contribución individual de cada variable, con el propósito de encarar de manera directa el problema de la interpretabilidad algorítmica en baterías. Wang et al. (2021) llevaron a cabo varias implementaciones de algoritmos no lineales con el objetivo de estimar la capacidad de descarga en cátodos dopados de tipo NCM, subrayando la eficiencia de la Máquina de Potenciación del Gradiente (GBM), que logró un coeficiente de determinación (R^2) igual a 0.76 y un error cuadrático medio (RMSE) en prueba equivalente a 16.66 mAh g^{-1}. La utilización innovadora de los valores SHAP (SHapley Additive Explanations) para evaluar la relevancia de las variables fue la contribución más importante de su estudio, mostrando con números que una menor cantidad del dopante y una electronegatividad más baja del mismo favorecen que haya mayor retención de capacidad.

En última instancia, se ha corroborado que los valores SHAP tienen la capacidad de convertir un modelo predictivo en una herramienta para diseñar materiales en varios campos de la fisicoquímica. Para pronosticar la temperatura de Curie basándose en proporciones atómicas, Liu et al. (2024) usaron algoritmos de ensamble durante el diseño de aleaciones amorfas que tienen como base el hierro. En el conjunto de prueba, su modelo ideal, basado en un Árbol de Potenciación del Gradiente (GBT), alcanzó un coeficiente de determinación (R^2) de 0.92 y un mínimo error porcentual absoluto medio (MAPE) del 5.21 %. Al igual que Wang et al. (2021), los autores integraron SHAP para

establecer una clasificación de importancia de las características, identificando los valores umbrales críticos de cada fracción elemental y diseñando con éxito nuevas aleaciones con propiedades térmicas específicas, evidenciando así, como los métodos de interpretabilidad nos permiten facilitar entender qué factores o propiedades complejas determinan el desempeño de materiales

Como último paso en el diseño de materiales basado en datos, y después de haber identificado que la aplicación de algoritmos robustos y de marcos de interpretabilidad que permiten superar el obstáculo de caja negra de los modelos de aprendizaje automático, es necesario modelar las variables que controlan el rendimiento de la celda, y es aquí que la eficiencia coulómbica y la conductividad iónica son los indicadores del transporte de masa y de la reversibilidad interfacial, por este orden, La consideración de modelos predictivos sobre estas variables específicas permite sustituir el clásico estudio empírico por una interesante categorización computacional.

La combinación de descriptores estructurales y químicos con algoritmos de ensamble ha mostrado un alto potencial para hacer predicciones en el campo de la conductividad iónica. Para pronosticar la energía de enlace y el transporte de carga en electrolitos haluros dopados (Li_2ZrCl_6), El Massafi et al. (2026) utilizaron modelos como LightGBM y XGBoost. Para la conductividad iónica, su modelo basado en XGBoost obtuvo un coeficiente de determinación (R^2) de 0.8062 y un error absoluto medio (MAE) de 0.6012. A pesar de que la investigación está enfocada en electrolitos en estado sólido, sus hallazgos muestran que una táctica efectiva para predecir propiedades de transporte en sistemas electroquímicos y captar relaciones no lineales es el uso combinado de descriptores elementales con algoritmos que potencian el gradiente.

Kumar et al. (2024) compilaron una de las bases de datos experimentales más completas sobre conductividad iónica empleando LightGBM y Redes Neuronales Gráficas para crear modelos predictivos, abarcando únicamente electrolitos líquidos. El mejor resultado que lograrón fue de un MAE de 0.09 log mS cm^{-1}, lo que equivale a 1.23 mS cm^{-1}. Un añadido fundamental es que dicho modelo mostró una dependencia no lineal intensa entre la movilidad iónica, la temperatura y la concentración de sal, y así demostrando cómo, a partir de descriptores físicos de los materiales, se puede predecir de manera confiable la conductividad en los electrolitos.

La reversibilidad en la interfaz entre el electrolito y el electrodo, además del transporte ionico, determinará si la batería es de larga duración por lo que es crucial evaluarla parametrizando dicha situación. Kim et al. (2023) abordaron este reto mediante el diseño de modelos supervisados de

Machine Learning usando como variable de entrada la composición elemental con el objeto de estimar en celdas de litio metálico la eficiencia coulómbica. Los autores cambiaron la métrica hacia una escala logarítmica (LCE):

$$LCE = -\log_{10}(1 - CE) \tag{2.1}$$

con el objetivo de sensibilizar al algoritmo en regímenes de alto rendimiento, lo que posibilitó que su modelo de Bosque Aleatorio (Random Forest) tuviera un error cuadrático medio (MSE) de prueba de solo 0.333. El análisis de la importancia de las características determinó que el elemento limitante más significativo es la proporción de oxígeno en el disolvente. Siguiendo este modelo, los autores desarrollaron electrolitos sin flúor que mostraron eficiencias coulómbicas experimentales de hasta el 99.70 %. Estos resultados muestran que los descriptores fundamentados en fracciones moleculares y los algoritmos de ensamble tienen la capacidad de vincular la composición química con el rendimiento electroquímico y guiar el diseño lógico de electrolitos novedosos.

Conforme las formulaciones de electrolitos logran niveles más altos de rendimiento, los indicadores convencionales de diseño disminuyen su capacidad explicativa. En estas circunstancias, categorizar los datos en grupos de alta y baja eficiencia es una estrategia eficaz para determinar cuáles son los factores fisicoquímicos que restringen el rendimiento en diferentes regímenes operativos. Esta perspectiva hace posible desvelar patrones y mecanismos subyacentes que no se pueden identificar fácilmente por medio de métodos empíricos o a través de la predicción de una sola variable continua.

Oyakhire et al. (2024) diseñaron modelos fundamentados en Bosques Aleatorios (Random Forest) para analizar electrolitos en celdas de litio metálico, utilizando una estrategia de segmentación del rendimiento. Las 189 formulaciones examinadas se separaron en dos grupos: una con alta eficiencia coulómbica (más del 98 %) y otra con baja eficiencia. La relación entre flúor y oxígeno (F/O) y la proporción de oxígeno en el solvente fueron determinantes en la clase de alta eficiencia, según el análisis de importancia de características, que reveló que los factores predominantes son distintos entre los dos regímenes, identificando la corrosión galvánica como un factor limitante para la eficiencia coulómbica.

Una vez demostrada la posibilidad de predicción de la eficiencia coulómbica, el diseño de electrolitos se convierte en un proceso que debe optimizar varias propiedades de forma simultánea. En especial, la estabilidad interfacial a menudo es antagónica con la conductividad iónica, lo que deriva en un compromiso entre la reversibilidad y el transporte de carga. La inteligencia artificial permite

abordar las dos métricas al mismo tiempo, así como analizar eficientemente grandes espacios químicos. En consecuencia, el enfoque de múltiples objetivos favorece la respuesta de formulaciones que puedan redundar en la estabilidad y la conductividad de los sistemas electroquímicos de vanguardia.

Zhang et al. (2025) constataron un marco de optimización bayesiana, junto con un análisis de frente de Pareto, para concebir electrolitos localizados de elevada concentración (LHCE), además de abordar la optimización de varias propiedades. Usando como mediante la regresión de proceso gaussiano como modelo sustituto, donde se procura simultáneamente el sobrepotencial cinético y la eficiencia coulómbica, utilizando un método iterativo, encontraron la formulación que, en baterías sin ánodo, muestra una capacidad de retención del 55 % tras 100 ciclos, lo que supera el 45 % que obtenía el electrolito de referencia, al mismo tiempo que se obtiene una eficiencia coulómbica del 99.39 %.

Para analizar de manera conjunta la conductividad iónica y la eficiencia coulómbica, se necesitan métricas unificadas que puedan clasificar espacios químicos extensos de forma coherente. Este método no solo ayuda a identificar formulaciones con potencial, sino que también promueve el diseño inverso de materiales, lo que posibilita la creación de nuevas estructuras moleculares dirigidas hacia metas específicas de rendimiento.

Kumar et al. (2024) idearon un método fundamentado en macrodatos para encarar la evaluación simultánea de varias propiedades, que combina la estabilidad oxidativa, la conductividad iónica y la eficiencia coulómbica a través de una métrica unificada conocida como eScore. Los autores lograron clasificar y filtrar disolventes previamente inexplorados utilizando este marco en repositorios que contienen más de 24 millones de compuestos. El fluoruro de etanosulfonilo (ESF) fue uno de los candidatos que se identificaron y sobresalió al obtener un rendimiento excepcional con una eficiencia coulómbica del 99.5 %. Estos hallazgos demuestran que los modelos de múltiples objetivos tienen la capacidad de acelerar el descubrimiento de electrolitos, más allá del método tradicional de ensayo y error.

Yang et al. (2026) establecieron un marco integrado que combina modelos generativos de difusión condicional con redes neuronales predictivas de grafos (GNN), llevando la capacidad analítica a la frontera del diseño computacional. En base a este sistema, sus electrolitos formulados alcanzaron valores de eficiencia coulúmbica de más del 96 % además de conductividades iónicas por arriba de los 15 mS cm^{-1}, valores considerables de alto desempeño y que brindan un referente para la clasificación de estos materiales.

Capítulo 3

Metodología

En este tercer capítulo se describe el flujo de trabajo computacional y analítico para la clasificación del desempeño de electrolitos. Se presenta la recolección y curación de datos experimentales, así como la definición de clases y umbrales de rendimiento. Posteriormente, se detalla el preprocesamiento e ingeniería de características, incluyendo la extracción de descriptores fisicoquímicos a partir de estructuras moleculares. Finalmente, se expone el diseño experimental, abarcando la selección de algoritmos de clasificación, la optimización de hiperparámetros y la aplicación del marco SHAP para la interpretabilidad de las predicciones.

3.1. Descripción de los Datasets

Las bases de datos utilizadas para cada uno de los objetivos, como lo son la eficiencia coulómbica y la conductividad iónica, se detallan a continuación, incluidos su composición y su preprocesamiento de curación.

3.1.1. Dataset de Eficiencia Coulómbica

El conjunto de datos para la evaluación de la eficiencia coulómbica ha sido creado mediante la combinación de las dos principales bases de datos que se encuentran en la literatura existente en baterías de litio metálico: Kim et al. (2023) y Kumar et al. (2024). La factibilidad y la adecuada compatibilidad de las metodologías en las que se basa la fusión de las dos bases de datos, se debe a que ambos estudios toman sus métricas sobre el mismo tipo de configuraciones experimentales, donde los dos estudios han reportado mediciones de eficiencia coulómbica tomando medidas unicamente con

configuraciones de celdas eléctricas Litio-Cobre (Li‖Cu), las que deben ser realizadas a temperatura ambiente y normalmente utilizando intensidades de corriente en el rango 0.4-1 $\mathrm{mA\,cm^{-2}}$. Al utilizar el mismo diseño de celda para la prueba, los valores a través de datos médicos corresponden al mismo proceso de reversibilidad electroquímica y de pérdida de litio, permitiendo así su fusión, sin que el diseño del experimento interfiera en los resultados de la fusión final. Para asegurar la explotación de la base de datos unificada, se utilizó el marco de descriptores químicos de Kim et al. (2023), donde en su propuesta se introducen 13 variables descriptivas que son aquellas que se encuentran basadas en las fracciones molares (porcentajes) de los componentes químicos que forman el electrolito, que abarca tanto el solvente como la sal y la composición total en sí, proporciones que están alineadas junto a su correspondiente eficiencia coulómbica en porcentajes del tipo lineal. Teniendo en cuenta que la base de datos extraída de Kumar et al. (2024), la cual está formada por 142 registros, utilizaba la representación SMILES para explicar la estructura y composición química de las sales y de los electrolitos, fue necesario realizar un proceso de curado y armonización. En primer lugar, se identifican los propios materiales que forman el electrolito y se obtiene su propiedades fisicoquímicas (por ejemplo la densidad y el peso molecular) a través del uso de repositorios químicos tales como PubChem y ChemSpider. Una vez caracterizada la composición, se aplicó un método volumétrico-atómico. Este procedimiento consistió en calcular la fracción molar dividiendo el número de átomos de cada elemento entre el número total de átomos presentes en un litro de solución. De esta manera, se lograron realizar los conteos atómicos exactos para derivar con precisión las 13 variables planteadas por el enfoque de Kim et al. (2023).

Para brindar mayor claridad sobre los descriptores químicos adoptados, la Tabla 3.1 detalla las 13 variables predictivas generadas. Estas características se centran en la composición elemental del electrolito y en las relaciones molares entre el solvente y el anión, abordando parámetros como las fracciones de oxígeno, carbono y flúor, así como la proporción inorgánica-orgánica. La elección de estas variables es esencial para el modelado, ya que representan factores críticos que influyen directamente en la estructura de solvatación de los iones y en la formación de la interfase electrolito-sólido.

Cuadro 3.1: Variables descriptoras del conjunto de datos de electrolitos.

Variable	**Descripción**
O %, C %, F %	Fracción molar global de Oxígeno, Carbono y Flúor en la composición del electrolito.
sO, sC, sF	Fracción molar de Oxígeno, Carbono y Flúor específicamente en los componentes solventes.
aO, aC, aF	Fracción molar de Oxígeno, Carbono y Flúor en las especies aniónicas.
F/O, F/C, O/C	Razones elementales entre Flúor, Oxígeno y Carbono, que describen relaciones composicionales.
InOr	Relación entre componentes inorgánicos y orgánicos en el sistema de electrolito.

Para asegurar la integridad del dataset y evitar sesgos en el aprendizaje de los modelos, se aplicó un filtro de detección de valores atípicos. Se eliminaron aquellos registros con una eficiencia coulómbica inferior al 80 % (por ejemplo valores de CE = 9 %), ya que representaban ruido de una medición y se consideran celdas disfuncionales, resultando en un conjunto de datos final consolidado de 283 registros aptos para el entrenamiento.

3.1.2. Dataset de Conductividad Iónica

Para el estudio y modelado del transporte de carga en los electrolitos, se empleó la base de datos experimental de Kumar et al. (2024), recopilada y curada por ellos en su estudio. Esta base de datos representa una de las recopilaciones empíricas más extensas reportadas en la literatura sobre electrolitos líquidos apróticos de litio, abarcando un total de 10,196 observaciones obtenidas de diversas fuentes y evaluadas bajo múltiples condiciones térmicas y de concentración. En su estado original, es decir, previo a cualquier etapa de preprocesamiento o ingeniería de características, el conjunto de datos se compone de 13 variables principales que describen de forma precisa la formulación de la mezcla multicomponente y sus condiciones de operación. La arquitectura del conjunto de datos original hace posible la representación de la formulación de sistemas complejos de un máximo de cuatro solventes diferentes y una sal de litio. Cada uno de estos compuestos químicos se devuelve mediante su notación SMILES – subyacente a esta identidad química, cuya

cadena es capaz de codificar la topología de la estructura molecular por medio de cadenas de texto absolutamente legibles por ordenador – además de la proporción de los solventes que componen la mezcla y las precisas condiciones físicas que se aplicaron para poder medir el rendimiento. A continuación, la Tabla 3.2 se ocupará de describir en detalle las variables que configuran el dataset en su forma cruda, las cuales se convierten en el punto de inicio para después calcular los descriptores moleculares.

Cuadro 3.2: Descripción detallada de los campos originales en la base de datos.

Variable original	**Concepto Físico**	**Descripción**	**Unidades / Formato**
`solv_1_sm` a `solv_4_sm`	Identidad de los solventes	Estructura química de hasta cuatro posibles solventes que componen la mezcla del electrolito.	Texto (SMILES)
`solv_ratio_1` a `solv_ratio_4`	Proporción de solventes	Fracción volumétrica o molar de cada solvente respectivo dentro de la mezcla formulada.	Numérico (0 a 1)
`salt_sm`	Identidad de la sal	Estructura química de la sal de litio disuelta en el sistema.	Texto (SMILES)
`conc_salt`	Concentración de la sal	Concentración de la sal de litio presente en el electrolito.	Molaridad (M)
`temperature`	Temperatura de operación	Temperatura a la cual se realizó la medición experimental de la celda.	°C
`conductivity`	Conductividad iónica	Variable objetivo original. Medición experimental de la conductividad iónica del electrolito.	mS/cm

3.2. Definición de Clases y Umbrales

Para transformar el problema de regresión en uno de clasificación binaria, se establecieron umbrales críticos basados en criterios de desempeño electroquímico:

- **Eficiencia Coulómbica**: Se eligió un límite crítico que permitiera distinguir de forma objetiva entre electrolitos con bajo rendimiento y aquellos con alto rendimiento. Por lo tanto, se designó

como Clase 1 (alto rendimiento) a los electrolitos que tienen una Eficiencia Coulómbica de al menos el 99 %, y como Clase 0 (bajo rendimiento) a los que están por debajo. La elección de un umbral del 99 % no es caprichosa, sino que se ajusta rigurosamente a las exigencias de reversibilidad electroquímica necesarias para el desempeño de baterías en la vida real. Según la investigación de Attias et al. (2022), para que una batería de metal recargable tenga viabilidad en aplicaciones prácticas y pueda resistir cientos de ciclos estables de carga y descarga, es obligatorio que su eficiencia coulómbica supere el 99 %. Bajo condiciones de aplicación práctica, los ánodos de litio deben someterse a altas densidades de corriente y capacidades de ciclado, escenarios donde los cambios drásticos de volumen y el crecimiento incontrolado de litio son fenómenos perjudiciales altamente anticipados. En este sentido, el estudio de Xu et al. (2019), aduce que la CE del litio protegido debe ser superior al 99 % para mantener la ventaja teórica de alta densidad de energía y asegurar la longeva vida de la célula. Yang et al. (2022), mostraron experimentalmente que una CE del 99 % y su mantenimiento durante ciclos largos representa un umbral de prueba de alta reversibilidad electroquímica, deposición uniforme de litio y reducción de cambios de volumen. Establecer este límite del 99 % va en la línea de perseguir escenarios de corto plazo viables y al mismo tiempo, es capaz de sentar las bases para escalar hacia un futuro con parámetros más exigentes. Como advierten Hobold et al. (2021), la adopción de vehículos eléctricos de forma masiva exigirá, eventualmente, mayores eficiencias que supere el 99.9 % o 99.95 %, para poder garantizar una retención de la capacidad de un 80-90 % tras 1,000 ciclos de uso. Por consiguiente, clasificar a los electrolitos con una CE inferior al 99 % como Clase 0 se justifica en que estos sistemas implican una pérdida acumulativa de litio activo en cada ciclo, resultando incompatibles con los estándares de viabilidad comercial a largo plazo.

Como resultado de la aplicación del umbral del 99 %, el conjunto de datos consolidado de 283 registros experimentó una segmentación binaria. La Figura 3.1 ilustra la distribución continua de la Eficiencia Coulómbica y el balance de clases resultante tras aplicar el criterio de separación.

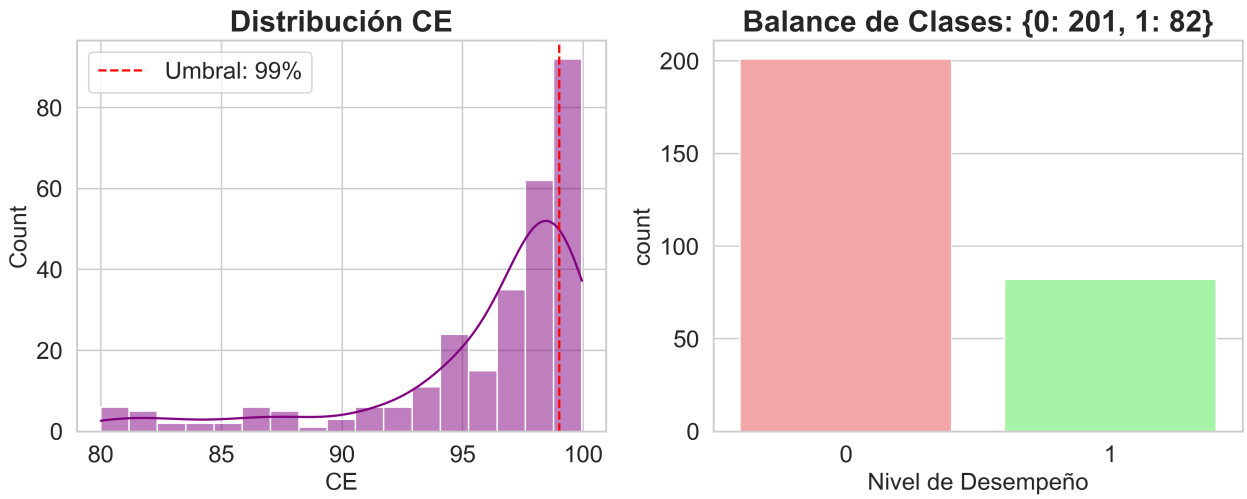


Figura 3.1: Distribución de los valores de la eficiencia coulombica y balace de las clases resultantes de la aplicacion del umbral

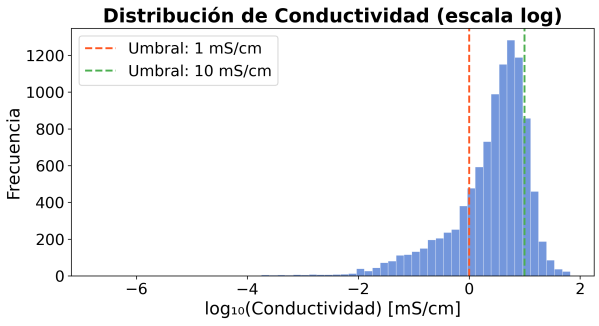
Como se puede apreciar, el conjunto de datos final tiene un desequilibrio de clases que es natural, ya que está compuesto por 201 muestras de la Clase 0 y 82 de la Clase 1. Es relevante enfatizar esta distribución, porque una mayoría de electrolitos de bajo rendimiento muestra la dificultad inherente en términos físicos y químicos de formular sistemas que logren eficiencias adecuadas para ser comercializables.

- **Conductividad Iónica**: Para modelar el transporte iónico, la variable continua de conductividad fue discretizada para plantear un problema de clasificación de múltiples clases. Establecer fronteras operativas claras ayuda al algoritmo a diferenciar entre formulaciones que limitan el flujo de carga y aquellas que mejoran la cinética de la celda. Por lo tanto, el conjunto de datos se dividió en tres categorías de rendimiento basadas en parámetros estrictos de viabilidad física y comercial.

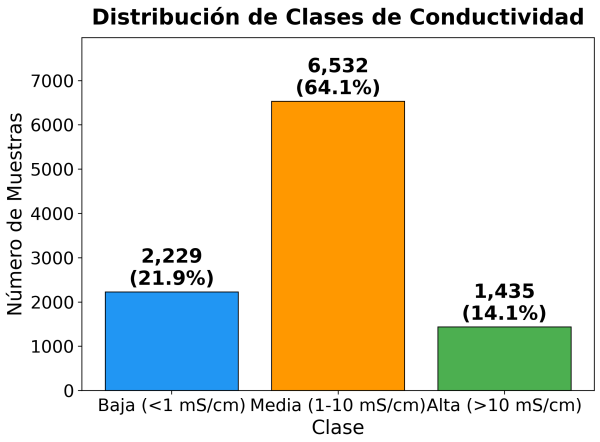
  Clase 0 (conductividad baja, < 1 mS/cm): Indica un sistema de transporte ineficaz. En su metodología, Kumar et al. (2024) establecieron que el límite mínimo necesario para disminuir la caída óhmica en la celda es de 1.0 mS/cm. Appetecchi et al. (2012) concluyeron que no es posible garantizar propiedades de transporte viables si no se alcanza un rango de 10^{-3} S/cm, en línea con esta limitación. Además, Li et al. (2023) informaron que es indispensable sobrepasar la barrera de 1 mS/cm para asegurar que el electrolito funcione en dispositivos reales. En consecuencia, se descartan las formulaciones de esta categoría para aplicaciones prácticas, ya que presentan alta resistencia interna.

Clase 1 (Conductividad media, de 1 a 10 mS/cm): Establece el régimen operativo funcional. Las formulaciones en esta clase han sobrepasado la barrera crítica de caída óhmica y tienen la movilidad iónica requerida para garantizar el transporte en aplicaciones estándar.Ferasso et al. (2025) demostraron en su revisión que un amplio espectro de electrolitos seguros y funcionales, como los basados en líquidos iónicos, presentan conductividades operativas entre 10^{-3} y 10^{-2} S/cm (es decir, entre 1 y 10 mS/cm), lo cual justifica este intervalo. Esto indica que estas mezclas son factibles para el ciclado, pero aún no alcanzan el límite superior requerido por tecnologías de alta potencia o carga rápida.

Clase 2 (Conductividad alta, >10 mS/cm): Representa el régimen de alto rendimiento y de transporte de carga mayor. Según el análisis de electrolitos comerciales que Kumar et al. (2024) han recopilado, el estándar operativo de la industria, por ejemplo el LiPF6 en combinaciones de carbonatos (EC/DMC), ha mostrado un promedio registrado de 10.11 mS/cm. En el marco de esta tesis, incluir un electrolito en esta categoría quiere decir que su movilidad iónica tiene una competencia directa o la supera, en comparación con el estado del arte de la industria, lo cual lo hace adecuado para resistir altas exigencias de transmisión de energía. En la Figura 3.2a se pueden observar la distribución de valores de conductividad de manera logarítmica con los umbrales definidos, y en la Figura 3.2b el tamaño de las clases resultantes.



(a) Distribución de la conductividad en escala logarítmica.



(b) Balance de clases después de aplicar el umbral.

Figura 3.2: Distribución de los valores de conductividad y balance de clases resultantes tras la aplicación del umbral.

3.3. Preprocesamiento e Ingeniería de Características

3.3.1. Expansión y Selección de Variables para la Eficiencia Coulómbica

Las 13 variables descriptivas elementales originales no son suficientes para captar la elevada no linealidad de los procesos electroquímicos, y en particular, la reacción de formación y estabilización de la interfase electrolito-sólido (SEI); para solventar la limitación expositiva mencionada y aumentar la capacidad predictiva de los algoritmos, se llevó a cabo una estrategia de pre-procesado para la reducción de dimensionalidad estructurada en dos etapas: i) una expansión polinómica de las características y ii) la selección regularizada de las características. En la primera etapa, se aplicó una transformación de polinomios de 2do grado al espacio de las variables de entrada. Este paso por fin dará lugar a términos de interacción ($x_i \cdot x_j$) y términos cuadráticos (x_i^2) exhaustivamente. Incorporar esos nuevos atributos es muy relevante desde una perspectiva química, puesto que permite modelar matemáticamente los efectos sinérgicos de los diferentes constituyentes del electrolito y es capaz de capturar relaciones individuales no lineales derivadas de fenómenos de saturación en las interficies. Esta ampliación del espacio de características incrementó en gran medida la dimensionalidad del conjunto de datos: dado que el tamaño de la muestra era muy pequeño, este incremento sorpresivo suponía un alto peligro de sobreajuste, donde el modelo podría .ªprender.el ruido apenas presente en el conjunto de entrenamiento; por el contrario, se hizo una segunda etapa de reducción de dimensionalidad aplicando regularización L1 (Lasso), la cual fue llevada a cabo algorítmicamente a través de una función SelectFromModel, utilizando un estimator de Regresión Logística con penalización L1. Un aspecto del que fui muy consciente y que fue relevante en esta etapa fue el de garantizar la exigencia metodológica y evitar la fuga de información: para ello, el flujo de preprocesamiento, es decir, estándar previa de la magnitud de las variables con *StandardScaler*, la transformación polinómica, la selección con licenciatura regularizada L1 y el modelo clasificador, fue encapsulado secuencialmente en un *Pipeline* acoplado a la validación cruzada, la cual garantizaría que, en cada iteración de la validación, los parámetros de escalado (intento de evitar el riesgo de fuga de información en el mismo conjunto de muestra) y los coeficientes de Lasso se ajustaran exclusivamente con los datos del pliegue (*fold*) de entrenamiento activo, aislando en ese sentido cualquier información del pliegue de validación. Como resultado de ese proceso estricto de regularización, los coeficientes matemáticos de las características redundantes o que no aportaban valor predictivo fueron penalizados, culminando en coeficientes que fueron reducidos a cero y

etc., reduciendo así el espacio de entrada a un vector denso de 9 características que determinan el comportamiento del electrolito que en la tabla 3.3 se clasifica y describe y que se agrupan según su naturaleza matemática y su interpretación física del rendimiento de este electrolito.

Cuadro 3.3: Variables seleccionadas tras la aplicación del método de regularización L1 y su interpretación.

Tipo de variable	**Variables seleccionadas**	**Interpretación**
Interacciones ($x_i \cdot x_j$)	$sO \cdot C$, $FO \cdot aO$, $FO \cdot C$, $sF \cdot aF$, $aF \cdot C$, $aO \cdot aC$, $FO \cdot sC$	Efectos sinérgicos entre los componentes del sistema (ej. interacciones entre las fracciones elementales del solvente, el anión y el total).
Cuadráticas (x_i^2)	$(O/C)^2$	Efectos no lineales asociados a fenómenos de saturación.
Lineales (x_i)	C	Efecto directo de la composición elemental global.

3.3.2. Extracción de Descriptores Fisicoquímicos para la Conductividad Iónica

Para que los algoritmos de aprendizaje automático puedan procesar e interpretar la información química contenida en la base de datos, es indispensable transformar las representaciones topológicas (cadenas de texto SMILES) en vectores numéricos continuos. Para llevar a cabo esta ingeniería de características, se empleó la librería de quimioinformática de código abierto RDKit, un estándar en la literatura para la generación de descriptores moleculares de relaciones estructura-actividad. Para garantizar la integridad y consistencia de la representación molecular, al igual que Kumar et al. (2024) todos los identificadores SMILES extraídos de la literatura fueron validados y estandarizados mediante la librería quimioinformática RDKit. Las secuencias se convirtieron a su forma canónica (Canonical SMILES) para eliminar representaciones redundantes y unificar el formato de entrada. Dado que el análisis se basó en descriptores topológicos y fisicoquímicos bidimensionales (1D/2D), las diferencias estereoquímicas fueron homogeneizadas sin afectar el cálculo de propiedades. Para poder integrar los datos correctamente a los modelos de aprendizaje automático, se verificó que no hubiera información faltante o valores nulos en la base de datos; una vez confirmado esto, se prosiguió a la generación de los descriptores. A partir de cada cadena SMILES, el algoritmo RDKit extrajo

cinco descriptores fisicoquímicos fundamentales que impactan la movilidad iónica y la solvatación: el peso molecular (MolWt), el coeficiente de partición octanol-agua (LogP), el número de donadores de puentes de hidrógeno (HDonors), el número de aceptores de puentes de hidrógeno (HAcceptors) y el Área de superficie polar topológica (TPSA). En la medida en que los electrolitos actúan como sistemas termodinámicos multicomponente, no es suficiente tomar las propiedades individuales de los solventes de forma aislada, de ahí que se haya calculado un descriptor global para el sistema solvente; exactamente como lo sería la sumatoria del producto de la propiedad individual calculada por RDKit para cada molécula por su respectiva fracción en el electrolito, de esta manera podemos brindar al modelo una manera cuantitativa de ponderar la distribución de cada componente de la mezcla. El conjunto de datos definitivo para el entrenamiento de los modelos quedó estructurado por un total de 17 variables. Mediante este flujo de preprocesamiento, la base de datos cruda se transformó en una matriz de características numéricas robusta y dotada de un profundo significado fisicoquímico. En la Tabla 3.4 se presenta la descripción detallada de estas variables resultantes, agrupadas según su naturaleza operativa, descriptores ponderados, descriptores de la sal y la variable objetivo, las cuales alimentaron directamente a los algoritmos de clasificación multiclase.

Cuadro 3.4: Variables finales y descripción del conjunto de datos procesado de conductividad iónica.

Variable	**Descripción**
`solv_ratio_1`– `solv_ratio_4`	Fracción individual de cada uno de los cuatro posibles solventes presentes en la mezcla.
`conc_salt`	Concentración de la sal de litio en el sistema, medida en molaridad.
`temperature`	Temperatura experimental en la que se realizó la medición de conductividad.
`mix_MolWt`	Peso molecular promedio ponderado de los solventes de la mezcla.
`mix_LogP`	Coeficiente de partición octanol-agua promedio ponderado de la mezcla.
`mix_HDonors`	Número de átomos donadores de enlaces de hidrógeno promedio ponderado de la mezcla.
`mix_HAcceptors`	Número de átomos aceptores de enlaces de hidrógeno promedio ponderado de la mezcla.
`mix_TPSA`	Área de superficie polar topológica promedio ponderada de los solventes.
`salt_MolWt`	Peso molecular calculado de la sal de litio utilizada.
`salt_LogP`	Coeficiente de partición octanol-agua calculado para la sal de litio.
`salt_HDonors`	Número de átomos donadores de enlaces de hidrógeno en la estructura de la sal.
`salt_HAcceptors`	Número de átomos aceptores de enlaces de hidrógeno en la estructura de la sal.
`salt_TPSA`	Área de superficie polar topológica calculada de la sal de litio.
`cond_class`	**Variable objetivo.** Categoría de conductividad iónica: 0 (Baja), 1 (Media) o 2 (Alta).

3.4. Diseño del Experimento de Cómputo

3.4.1. Selección de Modelos

Para clasificar el rendimiento de los electrolitos, tanto en lo que respecta a su conductividad iónica (problema multiclase) como a su eficiencia coulómbica (problema binario), se eligió una variedad de algoritmos de aprendizaje automático. Cada modelo tiene fortalezas particulares para enfrentar los retos inherentes a los datos, como el desequilibrio natural de las clases y la no linealidad de las interacciones químicas.

La regresión logística fue empleada como modelo básico predictivo en primer lugar por ser un procedimiento estadístico muy adecuado para modelar variables categóricas de respuesta. Al calcular probabilidades a través de una función sigmoide, este algoritmo ofrece una interpretación muy buena, lo que posibilita la evaluación directa y clara del efecto de las covariables (Becerra Correa y Leguizamón Páez, 2024). Se implementaron enfoques algorítmicos específicos, dependiendo de la naturaleza y del tamaño de cada base de datos.

Se eligió el algoritmo K-Vecinos Más Cercanos (K-NN) para la eficiencia coulómbica, que tiene un conjunto de datos de menor tamaño. Este método es no paramétrico y se basa en instancias; examina la cercanía geométrica entre los datos sin necesitar suposiciones previas acerca de la distribución subyacente (Vahedifar et al., 2023). Sin embargo, se decidió reemplazar K-NN con LightGBM debido a la gran cantidad de datos iónicos (más de 10,000 observaciones experimentales). Este algoritmo utiliza un marco de optimización fundamentado en histogramas que mejora notablemente la velocidad del entrenamiento y disminuye el uso de memoria, manteniendo al mismo tiempo una precisión elevada. Por lo tanto, es muy eficaz y apropiado para procesar las extensas matrices de características que provienen de la conductividad (El Massafi et al., 2026).

Para captar las dinámicas fisicoquímicas sumamente complejas en las dos propiedades, se utilizaron dos técnicas de aprendizaje en ensamble de vanguardia: XGBoost y Random Forest. Random Forest funciona con el principio de bagging, que crea varios árboles de decisión en subespacios aleatorios de datos; esta estructura compensa las carencias de los árboles individuales y disminuye la varianza, ofreciendo un modelo sólido ante el sobreajuste o overfitting (Fatima et al., 2023). XGBoost (Extreme Gradient Boosting), es un sistema escalable que representa el estado del arte por su excepcional velocidad y precisión (Chen y Guestrin, 2016). Su elección es crítica para todo el estudio debido a su destacada capacidad para lidiar con conjuntos de datos desbalanceados (como la

proporción de 201 muestras de baja eficiencia frente a 82 de alta, o la prevalencia de bajas conductividades en la base multiclase); mediante su enfoque iterativo, el algoritmo ajusta proactivamente los pesos (class weight), priorizando a las clases subrepresentadas y minimizando de forma eficaz los errores de predicción en los dominios no lineales de los electrolitos (Fatima et al., 2023).

3.4.2. Optimización de Hiperparámetros con Optuna

El buen desempeño predictivo de los algoritmos de aprendizaje automático depende de forma crítica de la configuración interna de los mismos. Para realizar un ajuste fino de los hiperparámetros de los modelos seleccionados, se empleó una estrategia de búsqueda automatizada utilizando el entorno de trabajo Optuna. Esta herramienta implementa el algoritmo de Estimador de Parzen Estructurado en Árbol (TPE, por sus siglas en inglés) para realizar una optimización bayesiana. A diferencia de una búsqueda de cuadrícula exhaustiva convencional, el muestreador TPE es mucho más eficiente computacionalmente, ya que modela y aprende de las distribuciones de los hiperparámetros que otorgan buenos y malos resultados, maximizando el ratio de probabilidad para concentrar la exploración estrictamente en las regiones más prometedoras del espacio de búsqueda. Dado que las bases de datos de esta investigación presentan arquitecturas distintas, la optimización se adaptó a las particularidades de cada problema:

Optimización para la Eficiencia Coulómbica (Clasificación Binaria). Para los modelos que predicen la reversibilidad del litio, una dificultad esencial fue el desequilibrio de clases intrínseco al conjunto de datos, este se forma por 201 muestras para la Clase 0 y únicamente 82 para la Clase 1. Se empleó un método de aprendizaje sensible a los costos para impedir que los modelos generaran una inclinación predictiva hacia la clase mayoritaria. En el caso del modelo de XGBoost, el hiperparámetro scale pos weight se ajustó a 2.45, garantizando que el algoritmo penalice matemáticamente con mayor fuerza los errores cometidos sobre la clase minoritaria de alto rendimiento. Se realizó la calibración mediante un esquema de validación cruzada de 5 iteraciones (5-Fold CV). La función objetivo utilizada por Optuna se enfocó en maximizar el Área Bajo la Curva ROC (AUC-ROC), utilizando el F1-Score como criterio secundario para asegurar un balance riguroso en la precisión y sensibilidad del modelo.

Optimización para la Conductividad Iónica (Clasificación Multiclase). Para la predicción del transporte iónico, el problema planteado involucra la clasificación en tres categorías (Baja, Media

y Alta) con un volumen masivo de datos. Para cada modelo de aprendizaje automático (Regresión Logística, Random Forest, XGBoost y LightGBM), Optuna se configuró de forma que se ejecutaran 120 iteraciones de búsquedas de hiperparámetros para cada uno. La manera más apropiada de manejar el desbalance de clases sin sesgar las predicciones fue incorporar el hiperparámetro class weight='balanced' directamente en los algoritmos, a excepción de la regresión logística, que no cuenta con este parámetro, otorgando pesos inversamente proporcionales a las frecuencias de cada clase operativa. Se utilizó la validación cruzada para mejorar la estabilidad de los resultados de entrenamiento, manejando 6 particiones (KFolds) de los datos. La función objetivo se definió para maximizar la métrica AUC OvR (One-vs-Rest), ya que con esta el problema queda binarizado, evaluando cada clase contra todo el resto y midiendo de igual manera la métrica F1-Score Macro, pudiendo así determinar la capacidad de discriminación entre las diferentes clases.

3.5. Interpretabilidad del Modelo con SHAP

Las técnicas de aprendizaje por ensemblaje, tales como XGBoost y Random Forest, brindan precisión predictiva superior para captar la no linealidad de los electrolitos. Sin embargo, su naturaleza algorítmica (caja negra) complica la comprensión de cómo interactúan las variables para obtener una predicción. Esta limitación la solventamos integrando una fuente de conocimiento fisicoquímico que permite su interpretación mediante la metodología SHAP (Lundberg y Lee, 2017), ya que este método es, desde el punto de vista matemático, un marco unificado y aditivo que incorpora los principios de la teoría de juegos cooperativos y otorga a cada atributo un valor de importancia calculando valores de Shapley que permiten evidenciar la contribución marginal de una variable promediada en todas las combinaciones (o subconjuntos) posibles de los atributos del modelo. La decisión conceptual de adoptar SHAP en lugar de otros métodos de interpretación radica en que, teóricamente, ha sido el único medio aditivo que ha demostrado ser aquel que cumple con tres propiedades matemáticas deseadas: ausencia (las características ausentes en el caso de entrada no tienen algún impacto atribuido), consistencia y precisión local (la suma de las atribuciones de las características se corresponde a la perfección con la salida del modelo para la predicción de ciertos casos) (Lundberg y Lee, 2017), permitiendo así un cálculo preciso y seguro de qué descriptores químicos y estructurales determinan la clasificación de las celdas.

Capítulo 4

Resultados y Discusión

Durante este Cuarto Capítulo se detallan los resultados obtenidos, contrastando entre cada uno de los modelos sus métricas ópticas para cada una de las cuestiones a resolver, seguido de un análisis detallado usando las gráficas de curva ROC y matrices de confusión, así como la interpretación de estos resultados desde un punto de vista físico-químico para poder así comprender las decisiones tomadas por los modelos.

4.1. Análisis de la Clasificación de Eficiencia Coulómbica

Al concluir el procedimiento de entrenamiento y conseguir la optimización de hiperparámetros mediante la herramienta Optuna, se puso a prueba el rendimiento predictivo de las cuatro arquitecturas elegidas (Regresión Logística, K-NN, Random Forest y XGBoost). Para comprobar la robustez de los modelos, se comprobaron sus métricas globales a partir del conjunto de prueba independiente (n=82), que actuó como suelo correspondiente a un proceso de validación. Tal y como reflejan los resultados que aparecen en la tabla 4.1, el algoritmo XGBoost mostró una capacidad predictiva superior al resto del grupo, manifestada en un área bajo la curva (AUC) de 0.9449, una exactitud (Accuracy) de 0.8947 y un F1-Score de 0.8333.

Cuadro 4.1: Comparación de métricas globales en entrenamiento y prueba para los modelos evaluados para la eficiencia coulombica.

Modelo	**AUC (Entrenamiento)**	**AUC (Prueba)**	**Accuracy (Prueba)**	**F1-Score (Prueba)**
XGBoost	0.9541	**0.9449**	**0.8947**	**0.8333**
Random Forest	0.9272	0.9084	0.8070	0.7027
Regresión Logística	0.9094	0.9316	0.8596	0.7895
KNN	0.9369	0.8853	0.8421	0.7273

La Figura 4.1 presenta de mejor manera la comparativa de las curvas ROC de los modelos evaluados, dejando en claro como XGBoost fue el modelo que mejor capacidad tuvo para distinguir entre la tasa de verdaderos positivos y la tasa de falsos positivos. Por su parte, la regresión logística (con una penalización L1 de C=2.015) se posicionó como la segunda mejor alternativa y, de este modo, se dio lugar a un modelo con un buen nivel de generalización, alcanzando un AUC de 0.9316, una accuracy de 0.8596 y un F1-Score de 0.7895. Por otro lado, las arquitecturas Random Forest y K-NN mostraron un peor rendimiento predictivo, siendo Random Forest con un AUC de 0.9096 y una accuracy de 0.8070 y K-NN el que peor rendimiento obtuvo, con un AUC de 0.8853 y un F1-Score de 0.7273.

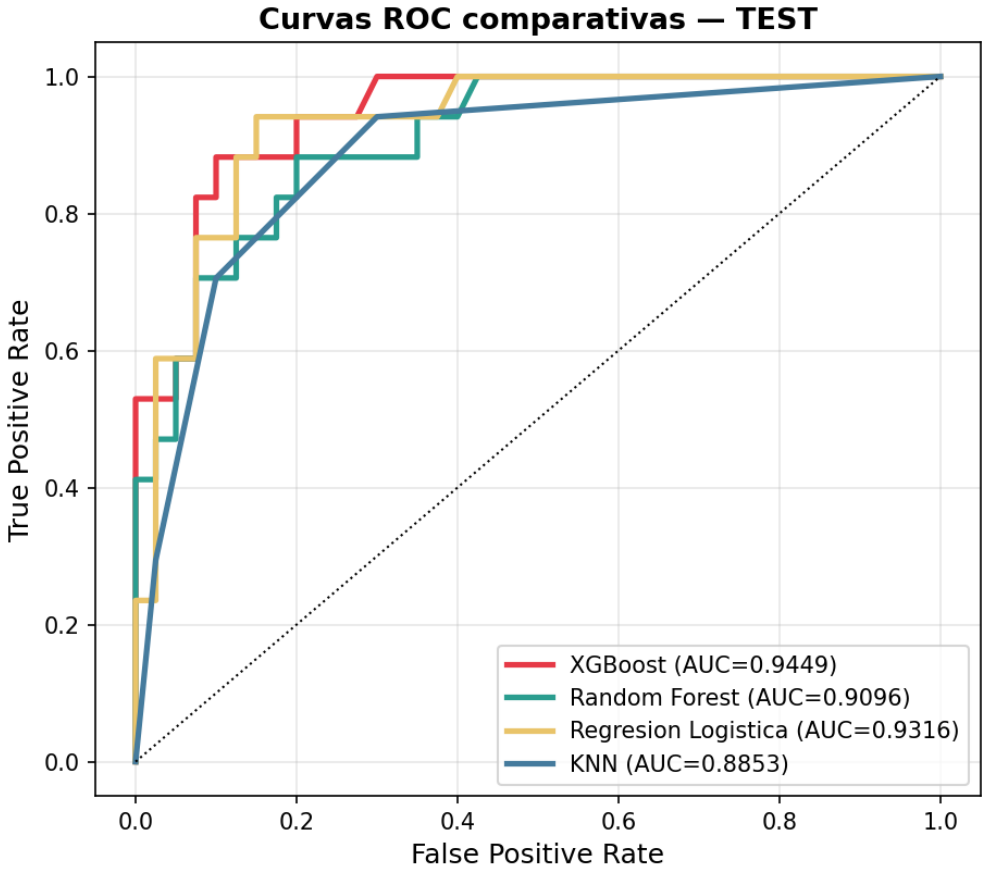


Figura 4.1: Comparación de áreas bajo la curva ROC (AUC-ROC) de modelos de clasificación de eficiencia coulombica.

El hecho de que XGBoost haya conseguido un rendimiento superior nos da a entender cómo pudo de manera matemática distinguir la fisicoquímica no lineal de los electrolitos, siendo altamente eficiente a la hora de medir su eficiencia coulómbica. Esta eficacia radica en su capacidad de capturar estos patrones subyacentes sin sobre ajustarse a pesar de la expansión polinomial. En la Tabla 4.2 se detallan los hiperparámetros óptimos por Optuna que permitieron a XGBoost alcanzar este desempeño, destacando una configuración de 874 estimadores, una profundidad máxima de 11 y ajustes específicos de submuestreo (subsample=0.977).

Cuadro 4.2: Hiperparámetros optimizados del modelo XGBoost mediante el uso de Optuna para la eficiencia coulombica.

Hiperparámetro	**Valor**	**Descripción**
n_estimators	874	Número de iteraciones de boosting.
max_depth	11	Profundidad máxima de los árboles para controlar la complejidad.
learning_rate	0.13587	Tasa de aprendizaje que reduce el impacto de cada árbol para evitar sobreajuste.
subsample	0.97717	Fracción de muestras utilizadas en cada árbol.
colsample_bytree	0.60369	Fracción de variables utilizadas en cada árbol.
gamma	2.67414	Reducción mínima de pérdida requerida para realizar una partición.
scale_pos_weight	2.45	Factor de ponderación para manejar el desbalance de clases (201/82).

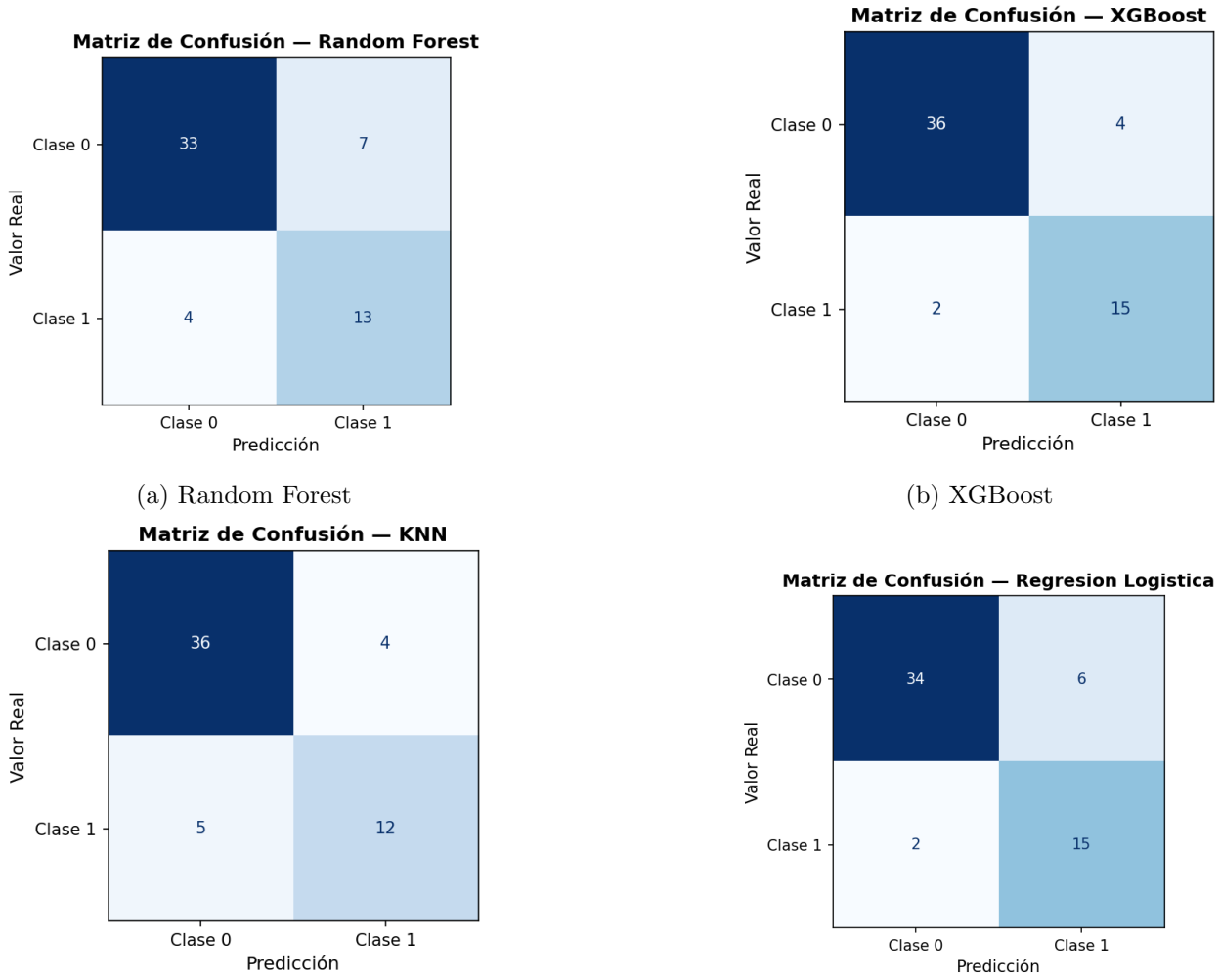
4.1.1. Capacidad de Detección de Electrolitos de Alto Desempeño

Aunque las métricas globales ofrecen una visión general de la calidad del modelo, el verdadero rendimiento del marco propuesto como instrumento de cribado virtual está determinado de manera específica por su desempeño en la clase minoritaria (Clase 1: CE $\geq$ 99 %). En el campo del descubrimiento acelerado de materiales, la prioridad cambia de la precisión general a maximizar la sensibilidad (recall) con el fin de reducir los falsos negativos. No incluir a un candidato de alto rendimiento conlleva un costo considerable en términos de innovación perdida, mientras que un falso positivo únicamente representa la validación experimental en laboratorio de una formulación subóptima. En el Cuadro 4.3, el modelo XGBoost demostró mayor efectividad, logrando detectar correctamente el 88 % (Recall = 0.88) de los electrolitos con alto rendimiento en el conjunto de prueba, incluyendo en esto un valor de 0.83 de F1-Score en dicha clase y mostrando así una viabilidad beneficiosa para el método experimental tradicional.

Cuadro 4.3: Desempeño en la clasificación de la Clase 1 ($CE \geq 99\,\%$) en eficiencia coulombica.

Modelo	**Precisión**	**Recall**	**F1-Score (Clase 1)**
XGBoost	**0.79**	**0.88**	**0.83**
Random Forest	0.65	0.76	0.70
Regresión Logística	0.71	**0.88**	0.79
KNN	0.75	0.71	0.73

El rendimiento en el ámbito de instancias individuales está ilustrado mediante las matrices de confusión en la Figura 4.2. El algoritmo XGBoost logró clasificar a 36 electrolitos como verdaderos negativos y a 15 electrolitos como verdaderos positivos, teniendo errores solo en 2 falsos negativos y 4 falsos positivos. La Regresión Logística obtuvo la misma sensibilidad que en el modelo XGBoost al clasificar 15 electrolitos como verdaderos positivos y 2 como falsos negativos (Recall = 0.88), pero sufrió una caída en la precisión (0.71) producida por los 6 falsos positivos y los 34 verdaderos negativos. Los modelos Random Forest y K-NN presentan tasas de error que limitan su uso como herramientas de filtrado. Random Forest fue el modelo con menor precisión en la clase positiva (precision = 0.65) y un recall de 0.76 y 13 verdaderos positivos, 33 verdaderos negativos, 7 falsos positivos y 4 falsos negativos, y K-NN obtuvo el peor rendimiento en la recuperación de la clase positiva al tener un recall de 0.71. La matriz de confusión revela que este modelo clasificó correctamente a 36 verdaderos negativos y 12 verdaderos positivos, pero también mostró la tasa más alta de falsos negativos de toda clase (5), así como 4 falsos positivos. El empleo de XGBoost asegura la recuperación de la mayoría de los electrolitos prometedores y reduce mucho el espacio de búsqueda, mejorando notablemente la inversión de tiempo y recursos de diseño experimental.



(a) Random Forest

(b) XGBoost

(c) k-NN

(d) Regresión Logística

Figura 4.2: Matrices de confusión de los modelos mediante ajuste de hiperparámetros con Optuna.

4.1.2. Interpretabilidad SHAP para Eficiencia Coulómbica

Para interpretar mejor las decisiones matemáticas de las predicciones realizadas por XGBoost y relacionarlas con la fisicoquímica de las mezclas, se utilizó el modelo de interpretabilidad SHAP.

Análisis Global de Características: En el gráfico de resumen de la Figura 4.3, la evaluación global revela que la predicción hacia la clase de alto rendimiento (CE $\geq 99\,\%$) está gobernada por la interacción sO C, seguida por la interacción FO aO y el término lineal de la concentración global de carbono (C).

La distribución espacial de los valores SHAP nos relata un comportamiento direccional serio para aquella característica que más peso tiene (sO C). Hay una fuerte correlación inversa: los valores bajos para esa variable (puntos azules) hacen que la predicción sea positiva, ayudando así a

que, por culpa de este hecho, la clasificación sea electrolito con una alta eficiencia. Por contra, los valores altos de sO C (puntos rojos) imponen una severa penalización algoritmia para el resultado, haciéndolo caer en clasificación baja. En una segunda línea, la FO aO tiene la relación inversa y complementaria; altos valores de FO aO (puntos magenta/rojos) tienen una dinámica de impulso positivo en el desempeño. Este orden de importancia llega a corroborar que las relaciones sinérgicas que tienen las fracciones de oxígeno en el solvente, los aniones y el carbono total y, tal y como matemáticamente se concluye, dictan la formación de una interfase electrolito-sólido robusta.

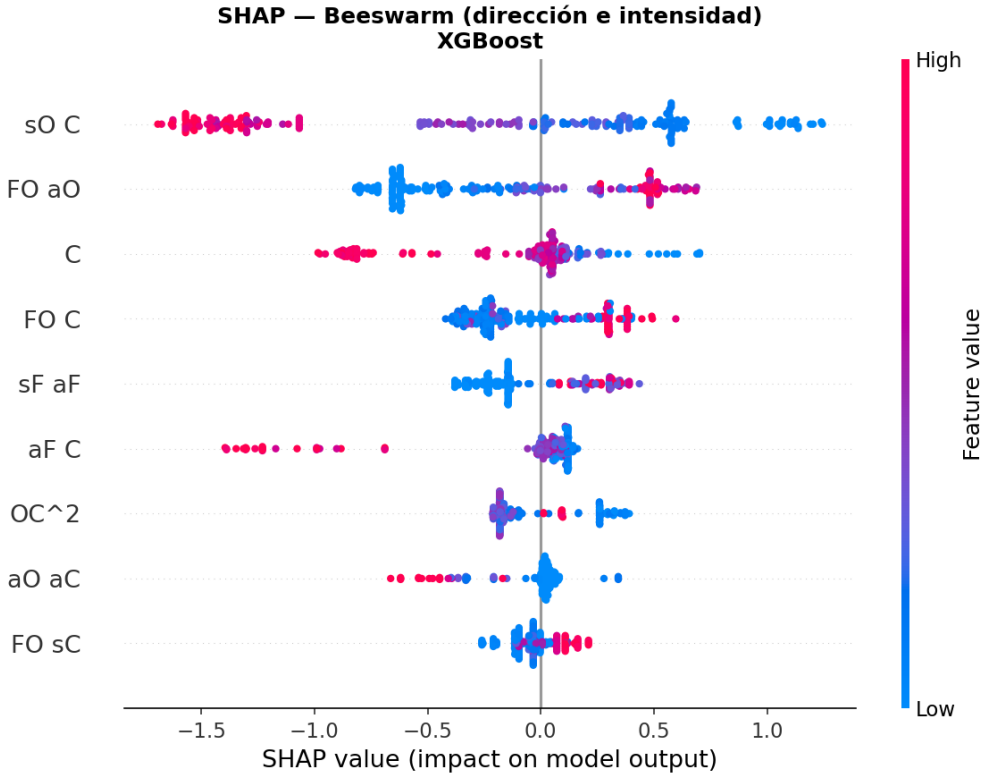


Figura 4.3: Interpretabilidad SHAP para la alta eficiencia coulombica.

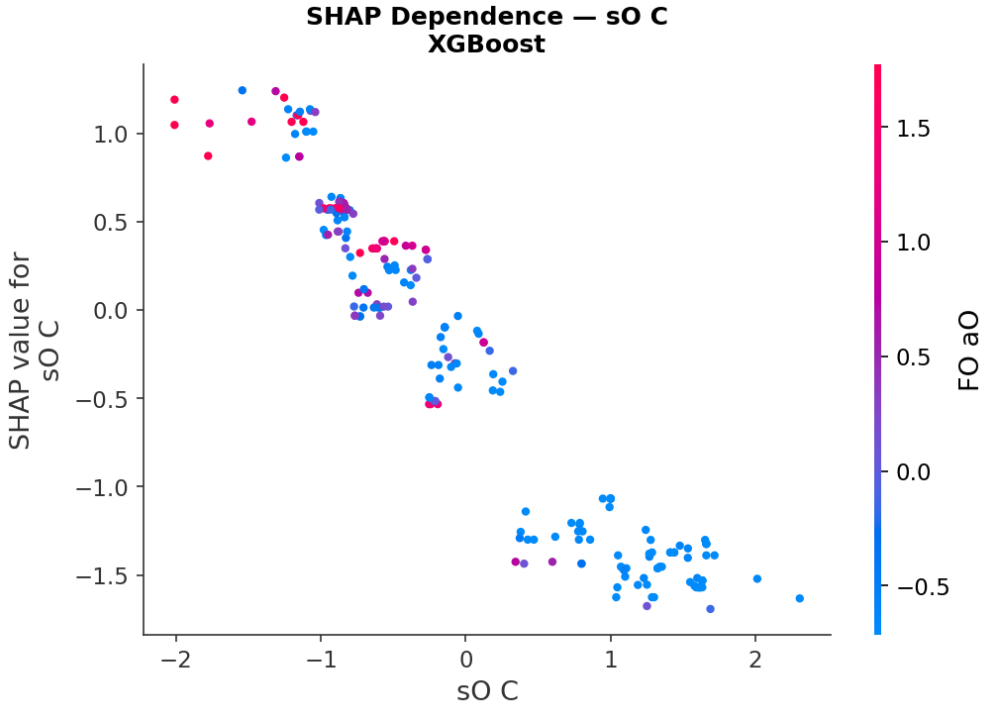


Figura 4.4: Análisis de la dependencia de las 2 variables con mayor impacto en la eficiencia coulombica.

Análisis de Dependencia e Interacciones: Para poder estudiar el comportamiento no lineal de la variable principal y observar los efectos conjuntos, la Figura 4.4 corresponde con el gráfico de dependencia referida a sO C. El eje horizontal de la figura pone de manifiesto la relación negativa descrita en el análisis global: el valor de sO C crece (es decir, se avanza a la derecha en el eje X) con respecto a su contribución (valor SHAP en el eje Y), todo ello disminuyendo de una forma bien monotónica, realizando una transición de una posición muy positiva (valores superiores a 1.0) a una penalización restrictiva (valores inferiores a -1.5). Por si fuera poco, el mapeo por color que superpone la variable FO aO expone una interacción fisicoquímica muy importante. Se observa que la propensión del modelo a clasificar un electrolito como uno de alta eficiencia coulómbica se maximiza considerablemente cuando cohabitan dos hechos: valores de sO C nulos o próximos a cero / valores altos de FO aO (puntos magenta que se agrupan en la parte superior izquierda del gráfico). Por otro lado, en la zona de alta penalización (valores altos de sO C), los valores de FO aO son bajos (puntos azules), cosa que complica notablemente la pérdida de reversibilidad. Esta sinergia extraída define una regla de diseño paramétrico directa para la formulación empírica de nuevos electrolitos.

4.2. Análisis de la Clasificación de Conductividad Iónica

Tras concluir la optimización de hiperparámetros, se evaluó el rendimiento predictivo de los cuatro modelos seleccionados (Regresión Logística, Random Forest, XGBoost y LightGBM) para el problema multiclase de conductividad iónica. La evaluación consideró las métricas de exactitud (Accuracy), F1-Score macro y el Área Bajo la Curva en esquema One-vs-Rest (AUC OvR) tanto en los conjuntos de entrenamiento como en los de prueba

Cuadro 4.4: Comparación de métricas globales en entrenamiento y prueba para los modelos evaluados para la conductividad iónica.

Modelo	**AUC (Ent.)**	**AUC (Pr.)**	**Acc. (Ent.)**	**Acc. (Pr.)**	**F1 (Ent.)**	**F1 (Pr.)**
XGBoost	0.9988	**0.9882**	0.9779	**0.9309**	0.9733	**0.9152**
LightGBM	0.9989	0.9878	0.9700	0.9284	0.9647	0.9138
Random Forest	0.9939	0.9793	0.9332	0.8980	0.9214	0.8791
Regresión Logística	0.8497	0.8406	0.6437	0.6446	0.6291	0.6244

En la Tabla 4.4 se presentan estos resultados globales, los cuales demuestran de manera contundente la superioridad de los métodos de ensamble frente al modelo lineal base. Como se observa en los datos recopilados, XGBoost y LightGBM alcanzaron una exactitud en el conjunto de prueba de 0.9308 y 0.9284, respectivamente, acompañados de un AUC OvR superior a 0.987 . Estas métricas contrastan marcadamente con la Regresión Logística, que obtuvo una exactitud de apenas 0.6446 y un AUC OvR de 0.8406 en la fase de prueba. Por este hecho podemos confirmar que la conductividad iónica está rígida por relaciones altamente no lineales entre sus características químicas y dando por hecho el uso necesario de modelos de inteligencia artificial capaces de clasificar correctamente estas no linealidades.

4.2.1. Capacidad de Detección de Electrolitos de Alta Conductividad

En el contexto del diseño computacional de materiales, el objetivo primordial es identificar formulaciones sobresalientes y, simultáneamente, minimizar los falsos positivos que derivarían en experimentos de laboratorio infructíferos. Por ello, el análisis de rendimiento se centra de manera específica en la viabilidad predictiva sobre la clase de alto desempeño, correspondiente a conductividades superiores a 10 mS/cm.

Cuadro 4.5: Desempeño en la clasificación de la clase Alta para los modelos evaluados.

Modelo	Precisión	Recall	F1-Score
XGBoost	**0.9068**	0.8815	**0.8940**
LightGBM	0.8370	**0.9303**	0.8812
Random Forest	0.7479	0.9303	0.8292
Regresión Logística	0.4798	0.8293	0.6079

Las métricas de precisión, sensibilidad (recall) y puntuación F1 de cada uno de los algoritmos quedan reflejadas en la Tabla 4.5 acompañadas con el análisis explicativo. La Figura 4.5 muestra las curvas ROC para la clase de alta conductividad, lo cual confirmar la capacidad predictiva de los algoritmos avanzados ya que el AUC para el XGBoost alcanza un 0.993 y el lightGBM un 0.992. Como se puede constatar en la visualización de la figura, el algoritmo XGBoost predijo a los candidatos de alta conductividad con valores de precisión de 0.9068 y sensibilidad de 0.8815. En el caso de LightGBM y Random Forest, predominó la sensibilidad, ya que estos métodos obtuvieron un recall de 0.9303, aunque la precisión también fue muy inferior: 0.8369 y 0.7478, respectivamente, y la regresión logística demostró un rendimiento muy inferior con un área bajo la curva ROC del 0.912 y un valor de precisión realmente poco fiable para el cribado virtual de tan solo 0.4798.

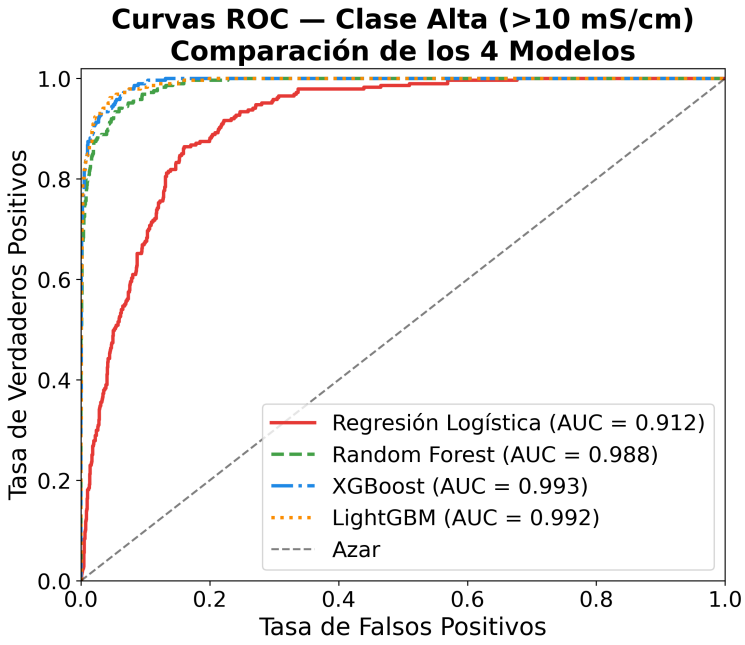
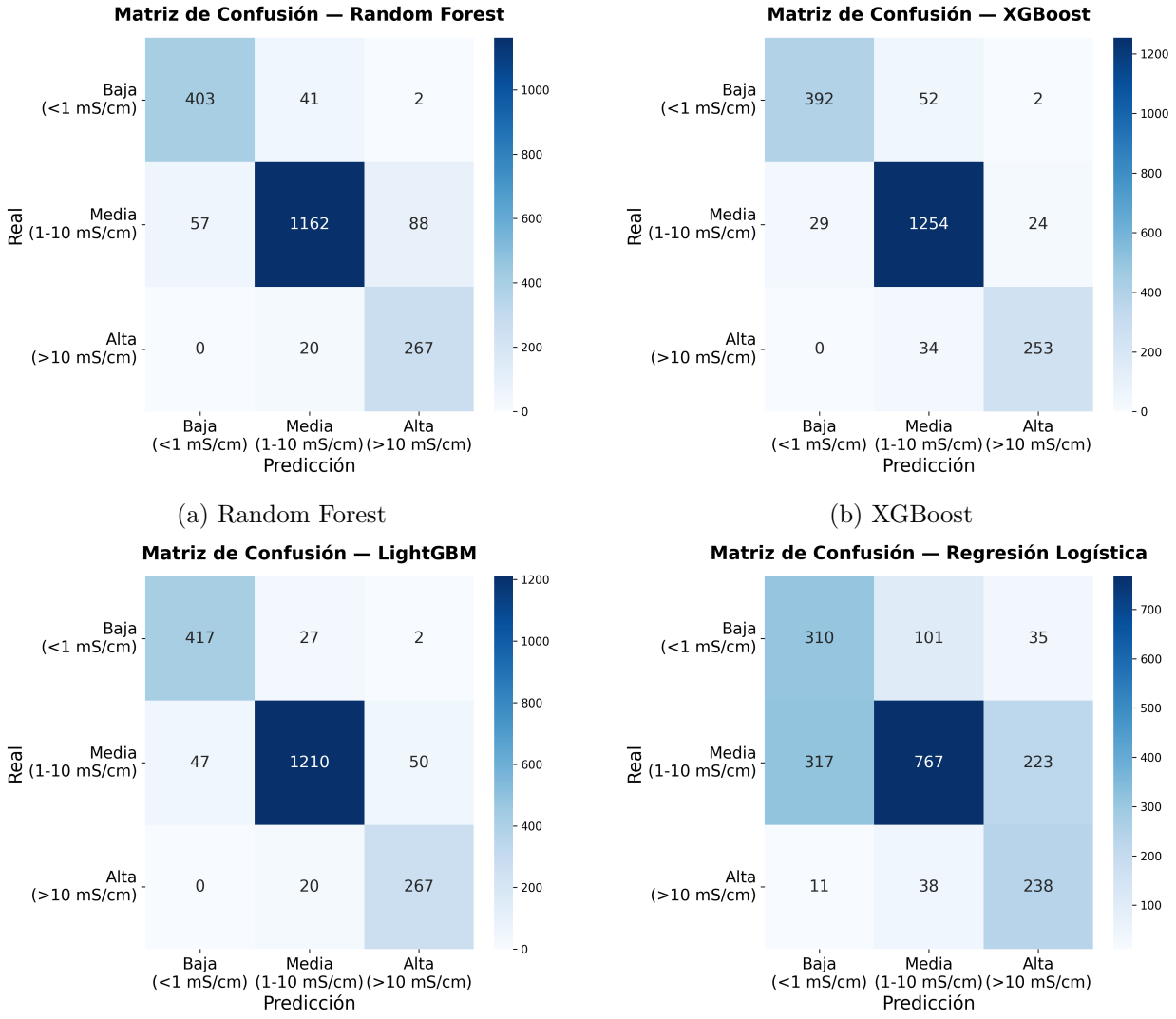


Figura 4.5: Comparación de area en curva Roc, en clase alta conductividad

El análisis del comportamiento predictivo a nivel de instancias individuales respalda la viabilidad operativa de estos modelos de ensamble. Como se evidencia en las matrices de confusión presentadas

en la Figura 4.6, XGBoost logró clasificar correctamente 253 muestras de alta conductividad como verdaderos positivos, incurriendo únicamente en 2 falsos positivos provenientes de la clase de baja conductividad y 24 provenientes de la clase media. LightGBM también exhibió una alta tasa de detección con 267 verdaderos positivos, aunque registró un mayor número de falsos positivos procedentes de la clase media (50 muestras). Por el contrario, la matriz correspondiente a la Regresión Logística revela una clasificación transversal errónea de 35 electrolitos de baja conductividad y 223 de conductividad media, etiquetándolos incorrectamente como candidatos de alto rendimiento.



(a) Random Forest

(b) XGBoost

(c) LightGBM

(d) Regresión Logística

Figura 4.6: Matrices de confusión de los modelos optimizados mediante ajuste de hiperparámetros con optuna para la conductividad.

Basados en los valores obtenidos por los modelos de aprendizaje automático, encontramos que

modelos basados en árboles logran una mejor clasificación de electrolitos, disminuyendo su error y creando así un filtro de cribado funcional y efectivo para la búsqueda de mezclas viables para su experimentación, quitando aquellas formulaciones no aptas para el alto transporte iónico.

4.2.2. Interpretabilidad SHAP para la Conductividad Iónica

La mejor manera de comprender las decisiones en las que el modelo XGBoost se basa para clasificar los electrolitos es implementar el análisis de explicabilidad SHAP. Observar exclusivamente la interpretación algorítmica de los factores que favorecen la clase de alta conductividad (>10 mS/cm) es primordial, por lo que la manera más precisa de hacerlo es desglosar el impacto global y su comportamiento segmentado por rangos térmicos.

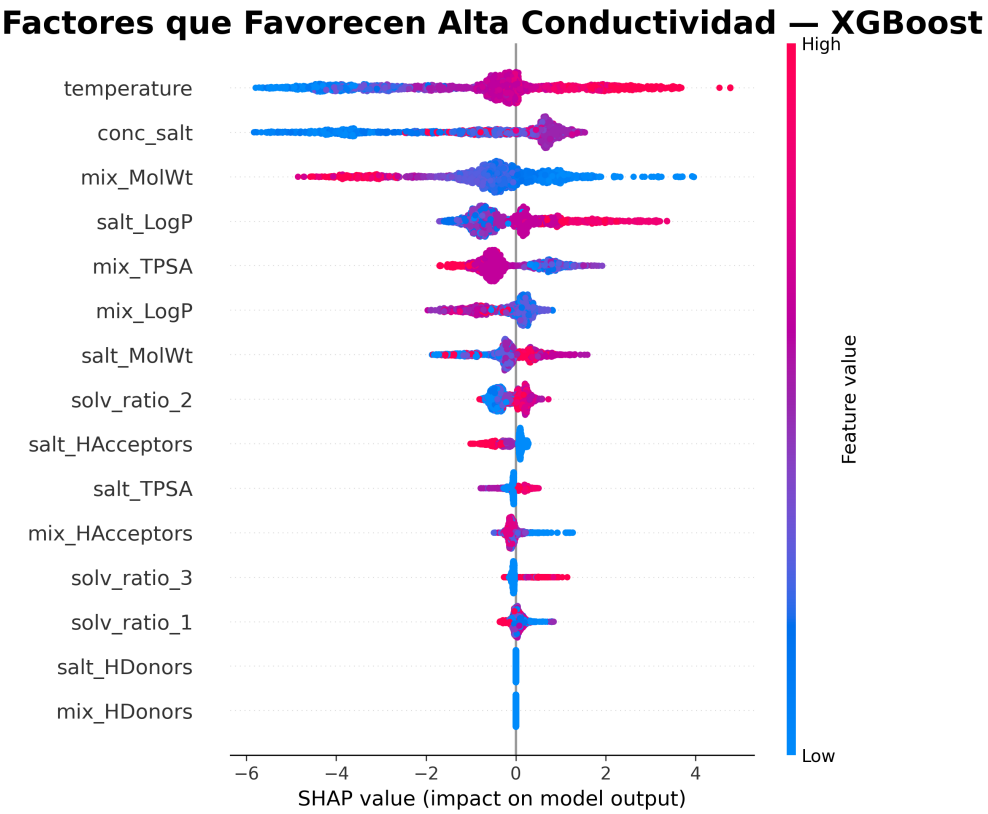
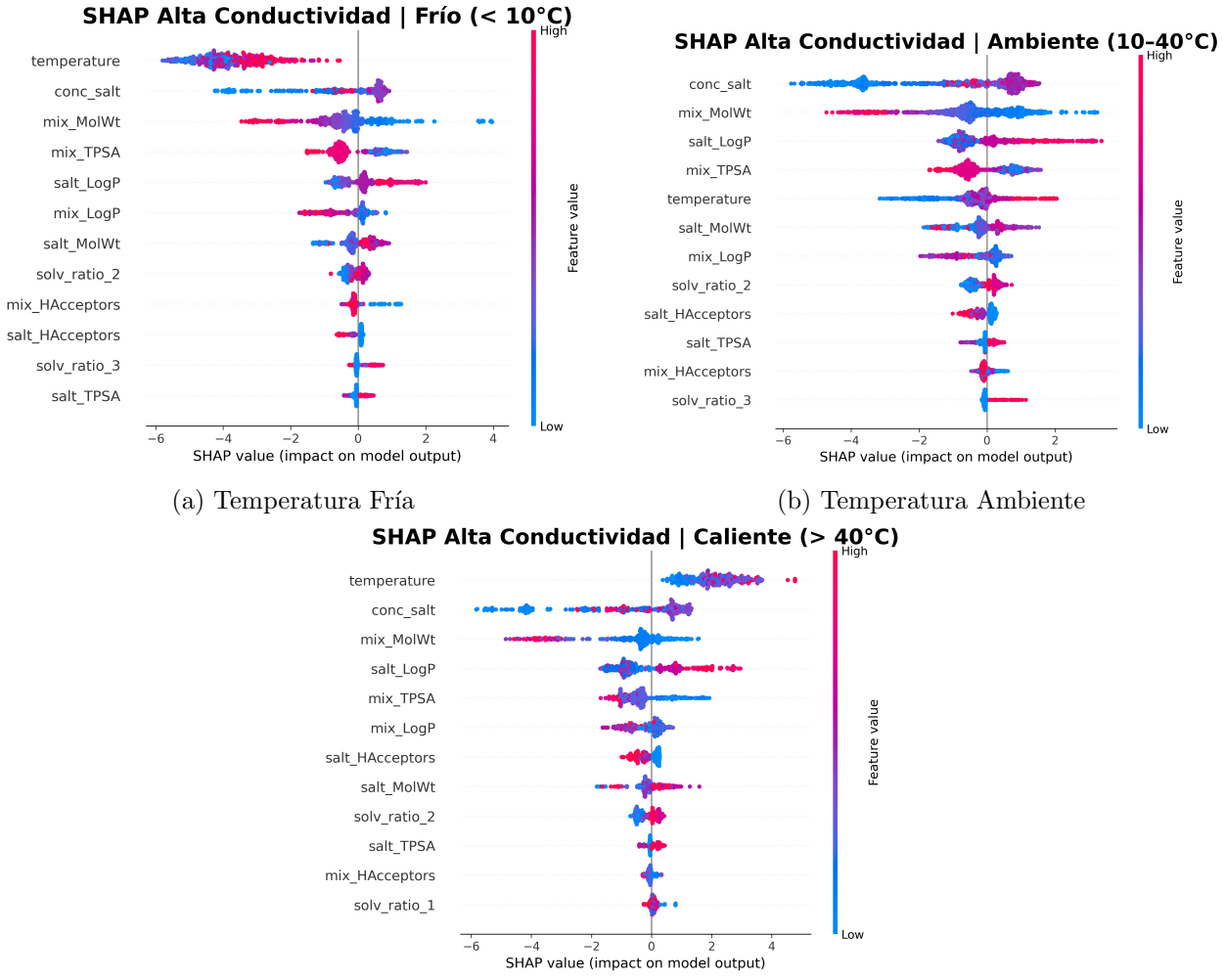


Figura 4.7: Interpretabilidad SHAP para la alta conductividad

Análisis Global de Alta Conductividad: La gráfica resumen de la Figura 4.7 nos muestra que tres descriptores paramétricos son los que más contribuyen a predecir alta conductividad, estos son peso molecular medio del conjunto de solventes (mix MolWt), concentración de sal (conc salt) y temperatura de operación (temperature). Los valores SHAP muestran que la temperatura tiene fuerte correlación positiva: los puntos rojos (los de valores más altos) desplazan la predicción hacia

arriba, mientras que los puntos azules (las temperaturas más bajas) hacen lo contrario y, por tanto, penalizan la predicción. El peso molecular de la mezcla tiene correlación inversa estricta: los solventes ligeros (puntos azules) son considerados impulsores de transporte de iones. La concentración de sal presenta un comportamiento muy particular: distancia muy no lineal, donde las concentraciones intermedias maximizan el valor SHAP positivo (puntos púrpuras) y las concentraciones extremas la minimizan (las concentraciones muy altas o muy bajas).

Análisis dividido en rangos térmicos: Como la temperatura es la variable de mayor importancia, la figura 4.8 muestra el análisis dividido en rangos térmicos para mostrar cómo cambian las reglas operativas según las condiciones ambientales.



(a) Temperatura Fría

(b) Temperatura Ambiente

(c) Temperatura cálida

Figura 4.8: Valores shap para rangos de temperatura

Régimen Frío (< 10°C): Podemos ver en la Figura 4.8a, la interpretabilidad SHAP de rango catalogado como temperatura fría, la temperatura sigue siendo el factor limitante principal. Las

temperaturas más bajas dentro de este rango generan penalizaciones (con valores SHAP cercanos a -6). Aunque la concentración de sal y el peso molecular mantienen su orden de importancia relativa, su capacidad para impulsar la conductividad está acotada por la barrera térmica, como se muestra en dicho gráfico.

Régimen Ambiente (10 - 40°C): entonces, al eliminar la condición térmica extrema, la jerarquía algorítmica se transforma. Como ejemplifica la Figura 4.8b, la sal se convierte en el controlador predictivo. La tendencia no lineal es imprescindible: facilitar la formulación a concentraciones intermedias (púrpura) es determinante para lograr alta conductividad ambientalmente y utilizar solventes de bajo peso molecular se convierte en el segundo parámetro dominante. Régimen Caloroso ($>$ 40°C): finalmente, la Figura 4.8c sostiene que cuando se trabaja a temperaturas elevadas, la temperatura vuelve a posicionarse en primer lugar en factor a la predicción. En otras palabras, que aun en estas circunstancias los altos gradientes térmicos conducen a una mayor optimización de la conductividad. La concentración media de sal y la baja masa molecular siguen facilitando el transporte de la carga, pero la penalización que imponen los altos pesos moleculares (puntos rojos) se torna menos restrictiva que en las situaciones medioambientales, lo cual sugiere que existe una mayor tolerancia del sistema al aumento de temperatura.

Capítulo 5

Conclusiones

Este trabajo de investigación ha evidenciado que la transición hacia un diseño de materiales basado en datos es una estrategia sumamente eficaz para mejorar las baterías de iones de litio. Se estableció un marco predictivo sólido, basado en el aprendizaje automático, para clasificar el desempeño de las formulaciones electrolíticas debido a la dificultad de equilibrar propiedades fisicoquímicas con dependencias cruzadas y no lineales. Esto exigió la creación de extensas bases de datos y el uso de algoritmos para obtener descriptores desde la representación molecular SMILES. El modelo, al emplear métodos de regularización como la regresión LASSO, se logró disminuir el extenso espacio químico a un grupo de variables realmente decisivas, lo que le otorgó una profundidad física al entrenamiento y le permitió sobrepasar las restricciones operativas propias de los enfoques empíricos basados en ensayo y error.

Los resultados de la eficiencia coulómbica corroboran que las técnicas de aprendizaje ensambladas son más eficaces, ya que el algoritmo XGBoost sobresalió al alcanzar una sensibilidad del 88 % y un AUC de 0.9449 a la hora de encontrar mezclas de electrolitos optimas. La herramienta SHAP logró mostrar que el principal factor limitante es la interacción cruzada entre el oxígeno del solvente y el carbono total de la mezcla. El modelo aprendió por medio de los propios datos que las formulaciones con poco o ningún oxígeno aumentan significativamente la reversibilidad, lo que proporciona un soporte a los fenómenos reportados en la literatura como lo demuestran Kim et al. (2023) y valida que el modelo puede separar matemáticamente las razones por las cuales se pierde litio activo.

La evaluación de la conductividad iónica consolidó el alcance del modelo para aproximar la conducta termodinámica del sistema. Las arquitecturas basadas en árboles evidenciaron un com-

portamiento superior puesto que se alcanzaron precisiones globales superiores al 93 %. De acuerdo con los hallazgos interfaciales, el análisis SHAP transformó el proceso de razonamiento del modelo en un conocimiento cinético aplicable, demostrando que la concentración de la sal —a temperatura ambiente— es una combinación que conlleva un foco de predicción absoluto bajo una dinámica fuertemente no lineal. Igualmente, el algoritmo demostró una sensible característica de penalizar de forma consistente a los disolventes de alto peso molecular, a la par que se deducen los fundamentos físicos que indican que una menor viscosidad aminora la ventaja del transporte de masa. Por otro lado, este algoritmo realizó una jerarquía dinámica para definir estas reglas de diseño, según diferentes rangos de temperatura.

La hipótesis de trabajo planteada se validó a partir de los resultados obtenidos, puesto que con el uso de extractores de variables fisicoquímicas, aunados a técnicas de aprendizaje automático, logró clasificar de manera efectiva el desempeño de electrolitos líquidos haciendo que sea posible crear un filtro antes de la validación experiencial para prever mezclas que mantendrán tanto una alta conductividad iónica como una alta eficiencia coulombica, reduciendo drásticamente los recursos usados en los experimentos, es así que como a medida que los modelos de aprendizaje automático vayan evolucionando y los datos experimentales vayan en aumento las predicciones de estos serán mucho mas superiores y explorar estas combinaciones químicas permitan diseñar baterías próximas mas optimas.

Bibliografía

Gustavo M. Hobold, Jeffrey Lopez, Rui Guo, Nicolò Minafra, Abhik Banerjee, Y. Shirley Meng, Yang Shao-Horn, y Betar M. Gallant. Moving beyond 99.9 % Coulombic efficiency for lithium anodes in liquid electrolytes. *Nature Energy*, 6(10):951–960, October 2021. ISSN 2058-7546. doi: 10.1038/s41560-021-00910-w.

Md Hosne Mobarak, Mariam Akter Mimona, Md. Aminul Islam, Nayem Hossain, Fatema Tuz Zohura, Ibnul Imtiaz, y Md Israfil Hossain Rimon. Scope of machine learning in materials research—A review. *Applied Surface Science Advances*, 18:100523, December 2023. ISSN 26665239. doi: 10.1016/j.apsadv.2023.100523.

Tianqi Chen y Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, San Francisco California USA, August 2016. ACM. ISBN 978-1-4503-4232-2. doi: 10.1145/2939672.2939785.

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, y Masanori Koyama. Optuna: A Next-generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2623–2631, Anchorage AK USA, July 2019. ACM. ISBN 978-1-4503-6201-6. doi: 10.1145/3292500.3330701.

U.S. Geological Survey. Mineral Commodity Summaries 2026. Technical report, U.S. Geological Survey, Reston, VA, February 2026.

Sang Cheol Kim, Solomon T. Oyakhire, Constantine Athanitis, Jingyang Wang, Zewen Zhang, Wenbo Zhang, David T. Boyle, Mun Sek Kim, Zhiao Yu, Xin Gao, Tomi Sogade, Esther Wu, Jian Qin, Zhenan Bao, Stacey F. Bent, y Yi Cui. Data-driven electrolyte design for lithium metal

anodes. *Proceedings of the National Academy of Sciences*, 120(10):e2214357120, March 2023. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2214357120.

Jiaojiao Deng, Xiaozhen Chen, Yu Bai, Jinhan Mo, y Xiaoliang Yu. Data-Driven Electrolyte Design for Advanced Rechargeable Lithium Batteries†. *Chinese Journal of Chemistry*, 44(3):403–418, February 2026. ISSN 1001-604X, 1614-7065. doi: 10.1002/cjoc.70321.

Guanyu Wang, Tom Fearn, Tengyao Wang, y Kwang-Leong Choy. Machine-Learning Approach for Predicting the Discharging Capacities of Doped Lithium Nickel–Cobalt–Manganese Cathode Materials in Li-Ion Batteries. *ACS Central Science*, 7(9):1551–1560, September 2021. ISSN 2374-7943, 2374-7951. doi: 10.1021/acscentsci.1c00611.

Zhong Ren y Changqing Du. A review of machine learning state-of-charge and state-of-health estimation algorithms for lithium-ion batteries. *Energy Reports*, 9:2993–3021, December 2023. ISSN 23524847. doi: 10.1016/j.egyr.2023.01.108.

Navid Nasajpour-Esfahani, Hamid Garmestani, Mohsen Bagheritabar, Dheyaa J. Jasim, D. Toghraie, Shohreh Dadkhah, y Hooman Firoozeh. Comprehensive review of lithium-ion battery materials and development challenges. *Renewable and Sustainable Energy Reviews*, 203:114783, October 2024. ISSN 13640321. doi: 10.1016/j.rser.2024.114783.

Arumugam Manthiram. An Outlook on Lithium Ion Battery Technology. *ACS Central Science*, 3 (10):1063–1069, October 2017. ISSN 2374-7943, 2374-7951. doi: 10.1021/acscentsci.7b00288.

Jang Wook Choi y Doron Aurbach. Promise and reality of post-lithium-ion batteries with high energy densities. *Nature Reviews Materials*, 1(4):16013, March 2016. ISSN 2058-8437. doi: 10.1038/natrevmats.2016.13.

John B. Goodenough y Kyu-Sung Park. The Li-Ion Rechargeable Battery: A Perspective. *Journal of the American Chemical Society*, 135(4):1167–1176, January 2013. ISSN 0002-7863, 1520-5126. doi: 10.1021/ja3091438.

Paolo De Blasio, Jonas Elsborg, Tejs Vegge, Eibar Flores, y Arghya Bhowmik. CALiSol-23: Experimental electrolyte conductivity data for various Li-salts and solvent combinations. *Scientific Data*, 11(1):750, July 2024. ISSN 2052-4463. doi: 10.1038/s41597-024-03575-8.

Yizhe Yan, Feng Hai, Bin Wang, Wenrui Cao, Mingtao Li, Chaohui Wang, Naipeng Li, y Dan Zhao. Machine learning accelerates high-voltage electrolyte discovery for lithium metal batteries. *Energy Storage Materials*, 79:104312, June 2025. ISSN 24058297. doi: 10.1016/j.ensm.2025.104312.

Chaocang Weng, Liang Ma, Bingfang Wang, Fanyue Meng, Jiaqi Yang, Yingying Ji, Botian Liu, Wenjie Mai, Sumei Huang, Likun Pan, y Jinliang Li. Single-solvent ionic liquid strategy achieving wide-temperature and ultra-high cut-off voltage for lithium metal batteries. *Energy Storage Materials*, 71:103584, August 2024. ISSN 24058297. doi: 10.1016/j.ensm.2024.103584.

Weijiang Xue, Zhe Shi, Mingjun Huang, Shuting Feng, Chao Wang, Fei Wang, Jeffrey Lopez, Bo Qiao, Guiyin Xu, Wenxu Zhang, Yanhao Dong, Rui Gao, Yang Shao-Horn, Jeremiah A. Johnson, y Ju Li. FSI-inspired solvent and "full fluorosulfonyl" electrolyte for 4 V class lithium-metal batteries. *Energy & Environmental Science*, 13(1):212–220, 2020. ISSN 1754-5692, 1754-5706. doi: 10.1039/C9EE02538C.

Rui Xu, Xin-Bing Cheng, Chong Yan, Xue-Qiang Zhang, Ye Xiao, Chen-Zi Zhao, Jia-Qi Huang, y Qiang Zhang. Artificial Interphases for Highly Stable Lithium Metal Anode. *Matter*, 1(2): 317–344, August 2019. ISSN 25902385. doi: 10.1016/j.matt.2019.05.016.

Jiangfeng Qian, Wesley A. Henderson, Wu Xu, Priyanka Bhattacharya, Mark Engelhard, Oleg Borodin, y Ji-Guang Zhang. High rate and stable cycling of lithium metal anode. *Nature Communications*, 6(1):6362, February 2015. ISSN 2041-1723. doi: 10.1038/ncomms7362.

Jian Yang, Tingting Feng, Junming Hou, Xinran Li, Boyu Chen, Cheng Chen, Zhi Chen, Yaochen Song, y Mengqiang Wu. Multi-dimensional hybrid flexible films promote uniform lithium deposition and mitigate volume change as lithium metal anodes. *Journal of Energy Chemistry*, 65: 583–591, February 2022. ISSN 20954956. doi: 10.1016/j.jechem.2021.07.001.

Huaxin Gong, Yuelang Chen, Shucheng Chen, Chengyi Xu, Yufei Yang, Yusheng Ye, Zhuojun Huang, Rui Ning, Yi Cui, y Zhenan Bao. Fast-Charging of Hybrid Lithium-Ion/Lithium-Metal Anodes by Nanostructured Hard Carbon Host. *ACS Energy Letters*, 7(12):4417–4426, December 2022. ISSN 2380-8195, 2380-8195. doi: 10.1021/acsenergylett.2c02130.

Ran Attias, Ben Dlugatch, Omer Blumen, Keren Shwartsman, Michal Salama, Netanel Shpigel, y Daniel Sharon. Determination of Average Coulombic Efficiency for Rechargeable Magnesium

Metal Anodes in Prospective Electrolyte Solutions. *ACS Applied Materials & Interfaces*, 14(27): 30952–30961, July 2022. ISSN 1944-8244, 1944-8252. doi: 10.1021/acsami.2c08008.

G. B. Appetecchi, M. Montanino, y S. Passerini. Ionic Liquid-Based Electrolytes for High Energy, Safer Lithium Batteries. In Ann E. Visser, Nicholas J. Bridges, y Robin D. Rogers, editors, *ACS Symposium Series*, volume 1117, pages 67–128. American Chemical Society, Washington, DC, January 2012. ISBN 978-0-8412-2763-7 978-0-8412-2764-4. doi: 10.1021/bk-2012-1117.ch004.

Clauber Andre Ferasso, Cesar Pandolfi, y Flavia Manica Siviero. O uso de líquidos iônicos para o aumento da condutividade elétrica em baterias: uma revisão crítica. *Brazilian Journal of Development*, 11(8):e81521, August 2025. ISSN 2525-8761. doi: 10.34117/bjdv11n8-018.

Alireza Valizadeh y Mohammad Hossein Amirhosseini. Machine Learning in Lithium-Ion Battery: Applications, Challenges, and Future Trends. *SN Computer Science*, 5(6):717, July 2024. ISSN 2661-8907. doi: 10.1007/s42979-024-03046-2.

Kevin Macwan. Improved Decision Support System for Personal Loan Eligibility Using Artificial Neural Networks. *International Journal of Computer Trends and Technology*, 72(7):138–155, July 2024. ISSN 22312803. doi: 10.14445/22312803/IJCTT-V72I7P119.

Candice Bentéjac, Anna Csörgő, y Gonzalo Martínez-Muñoz. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3):1937–1967, March 2021. ISSN 0269-2821, 1573-7462. doi: 10.1007/s10462-020-09896-5.

Siegfried Fredrich Stumpfe y Sandile Charles Shongwe. Comparative Analysis and Optimisation of Machine Learning Models for Regression and Classification on Structured Tabular Datasets. *Mathematics*, 14(3):473, January 2026. ISSN 2227-7390. doi: 10.3390/math14030473.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, y Tie-Yan Liu. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Neural Information Processing Systems*, Long Beach, CA, USA, 2017.

Nadir Subaşı. Comprehensive Analysis of Grid and Randomized Search on Dataset Performance. *European Journal of Engineering and Applied Sciences*, 7(2):77–83, December 2024. ISSN 2651-3412. doi: 10.55581/ejeas.1581494.

Abdul Rahman Mohamad Rom, Shafaf Ibrahim, Ahmad Firdaus Ahmad Fadzil, Nur Nabilah Abu Mangshor, y Nor Azura Md Ghani. A Review of Hyperparameter Tuning Methods in Machine Learning. In *2025 6th International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, pages 1–6, West Java, Indonesia, September 2025. IEEE. ISBN 979-8-3315-8603-4. doi: 10.1109/AiDAS67696.2025.11213530.

Davide Chicco y Giuseppe Jurman. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1):6, December 2020. ISSN 1471-2164. doi: 10.1186/s12864-019-6413-7.

Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, June 2006. ISSN 01678655. doi: 10.1016/j.patrec.2005.10.010.

Marco Tulio Ribeiro, Sameer Singh, y Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144, San Francisco California USA, August 2016. ACM. ISBN 978-1-4503-4232-2. doi: 10.1145/2939672.2939778.

Scott M Lundberg y Su-In Lee. A Unified Approach to Interpreting Model Predictions. *Curran Associates Inc.*, abs/1705.07874, 2017. ISSN 9781510860964.

Muia et al. Explainable Artificial Intelligence: A Comprehensive Review of Techniques, Applications, and Emerging Trends. *International Journal of Scientific Research in Computer Science and Engineering*, 2025. ISSN 23207639. doi: 10.26438/ijsrcse.v13i4.740.

Ayodeji S. Saliu, Lucky Anthony Osayuki, Oliseamaka N Chiedu, John Cherechim Nwaigbo, Airat Aderoju Aroyewun, y Afeez Olamilekan Isiaka. The Rise of Explainable AI: Enhancing Transparency and Trust in Machine Learning Models. *Global Journal of Engineering and Technology Advances*, 25(3):080–090, December 2025. ISSN 25825003. doi: 10.30574/gjeta.2025.25.3.0343.

Kristen A. Severson, Peter M. Attia, Norman Jin, Nicholas Perkins, Benben Jiang, Zi Yang, Michael H. Chen, Muratahan Aykol, Patrick K. Herring, Dimitrios Fraggedakis, Martin Z. Bazant, Stephen J. Harris, William C. Chueh, y Richard D. Braatz. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(5):383–391, March 2019. ISSN 2058-7546. doi: 10.1038/s41560-019-0356-8.

Steven K. Kauwe, Trevor David Rhone, y Taylor D. Sparks. Data-Driven Studies of Li-Ion-Battery Materials. *Crystals*, 9(1):54, January 2019. ISSN 2073-4352. doi: 10.3390/cryst9010054.

Yan Gu, Chuanzeng Zhang, y Mikhail V. Golub. Physics-informed neural networks for analysis of 2D thin-walled structures. *Engineering Analysis with Boundary Elements*, 145:161–172, December 2022. ISSN 09557997. doi: 10.1016/j.enganabound.2022.09.024.

Chengcheng Liu, Yongchao Lu, Jianfa Feng, Weidong Cai, y Hang Su. Prediction and customized design of Curie temperature of Fe-based amorphous alloys based on interpretable machine learning. *Materials Today Communications*, 38:107667, March 2024. ISSN 23524928. doi: 10.1016/j.mtcomm.2023.107667.

Nisryne El Massafi, Othman El Kssiri, Kawtar Zerhouni, Mohamed Naji, Abdessamad Faik, y Anass Sibari. Machine Learning Prediction of Ionic Conductivity and Binding Energy in Doped Li_2ZrCl_6 Halides for Solid-State Electrolyte. *ACS Applied Materials & Interfaces*, 18(2):3991–4003, January 2026. ISSN 1944-8244, 1944-8252. doi: 10.1021/acsami.5c18028.

Ritesh Kumar, Minh Canh Vu, Peiyuan Ma, y Chibueze Amanchukwu. Electrolytomics: A unified big data approach for electrolyte design and discovery. https://chemrxiv.org/doi/full/10.26434/chemrxiv-2024-vqtc7, April 2024.

Solomon T. Oyakhire, Sang Cheol Kim, Wenbo Zhang, Sanzeeda Baig Shuchi, Yi Cui, y Stacey F. Bent. Galvanic Corrosion Underlies Coulombic Efficiency Differences in High-Performing Lithium Metal Battery Electrolytes. https://chemrxiv.org/doi/full/10.26434/chemrxiv-2024-6g0h2, November 2024.

Jiayi Zhang, Debanjan Sarker, Matthew Beltran, Yu Xiang, Cormac Toher, y Laisuo Su. Dual-objective optimization of lithium metal battery electrolytes via machine learning. *Materials Today Energy*, 51:101909, July 2025. ISSN 24686069. doi: 10.1016/j.mtener.2025.101909.

Zhenze Yang, Yifan Wu, Xu Han, Ziqing Zhang, Haoen Lai, Zhenliang Mu, Tianze Zheng, Siyuan Liu, Zhichen Pu, Zhi Wang, Zhiao Yu, Sheng Gong, y Wen Yan. A unified predictive and generative solution for liquid electrolyte formulation. *Nature Machine Intelligence*, 8(2):186–196, January 2026. ISSN 2522-5839. doi: 10.1038/s42256-025-01173-w.

Kai Li, Jifeng Wang, Yuanyuan Song, y Ying Wang. Machine learning-guided discovery of ionic polymer electrolytes for lithium metal batteries. *Nature Communications*, 14(1):2789, May 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-38493-7.

Nelson Becerra Correa y Miguel Angel Leguizamón Páez. Regresión Logística Técnica de Machine Learning para predicciones académicas. *XIKUA Boletín Científico de la Escuela Superior de Tlahuelilpan*, 12:71–80, July 2024. ISSN 2007-4948. doi: 10.29057/xikua.v12iEspecial.12746.

Mohammad Ali Vahedifar, Azim Akhtarshenas, Mohammad Mohammadi Rafatpanah, y Maryam Sabbaghian. Shapley-Based Data Valuation with Mutual Information: A Key to Modified K-Nearest Neighbors, 2023.

Sana Fatima, Ayan Hussain, Sohaib Bin Amir, Syed Haseeb Ahmed, y Syed Muhammad Huzaifa Aslam. XGBoost and Random Forest Algorithms: An in Depth Analysis. *Pakistan Journal of Scientific Research*, 3(1):26–31, October 2023. ISSN 0552-9050, 0552-9050. doi: 10.57041/pjosr.v3i1.946.

Cuernavaca, Morelos a 9 de Noviembre del 2026.

DR. MIGUEL ÁNGEL GARCÍA GARNICA
DIRECTOR DE LA FACULTAD DE CONTADURÍA,
ADMINISTRACIÓN E INFORMÁTICA.
PRESENTE

En mi carácter de revisor de tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado de la Maestría en Optimización y Cómputo Aplicado, del estudiante Sergio Rubén Ocampo Pérez, con matrícula 10025065, con el título **APLICACIÓN DE APRENDIZAJE AUTOMÁTICO PARA LA CLASIFICACIÓN DEL DESEMPEÑO DE BATERÍAS DE IONES DE LITIO A PARTIR DE LA EFICIENCIA COULÓMBICA Y LA CONDUCTIVIDAD DEL ELECTROLITO**, por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que el estudiante continúe con los trámites que la Universidad Autónoma del Estado de Morelos tenga establecidos para obtener el grado mencionado.

Atentamente
Por una humanidad culta

Dr. Outmane Oubram
Profesor- investigador
Facultad de Contaduría, Administración e Informática

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

Se expide el presente documento con firma electrónica UAEM, soportada por el certificado vigente a la fecha de su elaboración y con efectos plenos de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS PUBLICADOS en el ÓRGANÓ INFORMATIVO UNIVERSITARIO "ADOLFO MENÉNDEZ SAMARÁ" número 117 de fecha 20 de abril de 2021.

Sello electrónico

OUTMANE OUBRAM | Fecha:2026-11-09 14:17:42 | FIRMANTE

V3P9ESxVKjh23tjxF5nxbHOMpdki+yueM+9CDpSDoyy4wKkTJ+idT+YoxXFOaU3Qi2xAehthwZaRuD4/Ao+QPE759y5M4FqbgO9XJkgaYh2jJ+HB/8Cwpf45L408C21wrQ2buh8KI
/fH1zndbvgYtg/CY8RbOnAcezt6d2DznRwma368M/l8Lp70d8T/FhKF9OWpm0koATV2qYKKX6HAKlEIoMfUbTyeX/F49vrZmzB0IOcD1wl9lyos+YxIrcup49t/s+TDJmF6sN69AwR/
Q/XepuktPjkc4wr8ySHyoSpiJXJfWZTV+4OlqJU79elZioe2scESY9CCmwhZlQBAqg==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:

dQDZUos1b

https://efirma.uaem.mx/noRepudio/UpbCIlk9tvBXlh3ucDDZoVa7vcGaqk5Z

Cuernavaca, Morelos a 10 de Noviembre del 2026.

DR. MIGUEL ÁNGEL GARCÍA GARNICA
DIRECTOR DE LA FACULTAD DE CONTADURÍA,
ADMINISTRACIÓN E INFORMÁTICA.
PRESENTE

En mi carácter de revisor de tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado de la Maestría en Optimización y Cómputo Aplicado, del estudiante Sergio Rubén Ocampo Pérez, con matrícula 10025065, con el título **APLICACIÓN DE APRENDIZAJE AUTOMÁTICO PARA LA CLASIFICACIÓN DEL DESEMPEÑO DE BATERÍAS DE IONES DE LITIO A PARTIR DE LA EFICIENCIA COULÓMBICA Y LA CONDUCTIVIDAD DEL ELECTROLITO**, por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que el estudiante continúe con los trámites que la Universidad Autónoma del Estado de Morelos tenga establecidos para obtener el grado mencionado.

Atentamente
Por una humanidad culta

Dr. Noureddine Lakouari
Investigador por México
Instituto nacional de astrofísica óptica y electrónica

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

Se expide el presente documento con firma electrónica UAEM, soportada por el certificado vigente a la fecha de su elaboración y con efectos plenos de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS PUBLICADOS en el ÓRGANÓ INFORMATIVO UNIVERSITARIO "ADOLFO MENÉNDEZ SAMARÁ" número 117 de fecha 20 de abril de 2021.

Sello electrónico

NOUREDDINE LAKOUARI | Fecha:2026-11-10 13:00:20 | FIRMANTE

o0HSywPuXL9byO3/RuGCV9o3AXRYfQ/8edSlvpCZs2rlaUR4rn1fFEuXfjPoK9MkMSqi/JM6aJHgnK9KplBcSwlADbMFEul7v5vZA7CnPVk1C8esP1T2NENF35VUESxZtQ0gpH2uELUNvf8cSRT3DYNbvyuURX3IkUU+3aicuVo1WIl+i3ICDaYvecprlhFdDQG3rnNoQ+oiuG/MJArnbqQ+BS96DdTCOmPYw07Bw1HH2aHrBabUwVECdsz0aNBjsrtend0/giiM0kVFJ2yqY0mpXAQx1/X4VTDR0UglrL4NNbpYs/D+3QWannlPSeQrdhRoKqhKkXcXzgEdLD8aKw==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:

JvWXSnDHA

https://efirma.uaem.mx/noRepudio/aobTEqXNCx48ZWVQyyEUz6EKo9Mdwbxw

Cuernavaca, Morelos a 9 de Noviembre del 2026.

DR. MIGUEL ÁNGEL GARCÍA GARNICA
DIRECTOR DE LA FACULTAD DE CONTADURÍA,
ADMINISTRACIÓN E INFORMÁTICA.
PRESENTE

En mi carácter de revisor de tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado de la Maestría en Optimización y Cómputo Aplicado, del estudiante Sergio Rubén Ocampo Pérez, con matrícula 10025065, con el título **APLICACIÓN DE APRENDIZAJE AUTOMÁTICO PARA LA CLASIFICACIÓN DEL DESEMPEÑO DE BATERÍAS DE IONES DE LITIO A PARTIR DE LA EFICIENCIA COULÓMBICA Y LA CONDUCTIVIDAD DEL ELECTROLITO**, por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que el estudiante continúe con los trámites que la Universidad Autónoma del Estado de Morelos tenga establecidos para obtener el grado mencionado.

Atentamente
Por una humanidad culta

Dr. Federico Alonso Pecina
Profesor- investigador
Facultad de Contaduría, Administración e Informática

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

Se expide el presente documento con firma electrónica UAEM, soportada por el certificado vigente a la fecha de su elaboración y con efectos plenos de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS PUBLICADOS en el ÓRGANÓ INFORMATIVO UNIVERSITARIO "ADOLFO MENÉNDEZ SAMARÁ" número 117 de fecha 20 de abril de 2021.

Sello electrónico

FEDERICO ALONSO PECINA | Fecha:2026-11-09 16:43:20 | FIRMANTE

R6zEXvxWCt6HIwhr6kgclsQR34GafxDhGzS6RcSuOmKtALfsqtwnG5mV7jCrJcJsaGcbo0C9IfytmXzf0tm/ZNsooFtaM5spnyGwseaODpqELGLP8F5l4C+MprjUpqjJS1qo4jHebpLa4bmzqRyN8d3TjoC2HPbJ6WBmwoheCRpwbNRI2pUJZh413Az2OeKV6NxS2NSqVlWDUng+0A/wjWkpCljXO2Q48/SMUdGSyL7mhgrbVK9AK+2rnfLrK/1t3NN8CnYCHjna+jm/pfFrx5fCO+vMrkrNpTp9y4aC/AevVgQtlpQLxfyXp5j3NprxpzItutKL0rGItPHI2+XcqQ==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:

qZyctbK87

https://efirma.uaem.mx/noRepudio/EfHTqVeIjETUq05ppAZM7G9ZdB0zB6RZ

Facultad de Contaduría, Administración e Informática

Cuernavaca, Morelos a 9 de Noviembre del 2026.

DR. MIGUEL ÁNGEL GARCÍA GARNICA
DIRECTOR DE LA FACULTAD DE CONTADURÍA,
ADMINISTRACIÓN E INFORMÁTICA.
PRESENTE

En mi carácter de revisor de tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado de la Maestría en Optimización y Cómputo Aplicado, del estudiante Sergio Rubén Ocampo Pérez, con matrícula 10025065, con el título **APLICACIÓN DE APRENDIZAJE AUTOMÁTICO PARA LA CLASIFICACIÓN DEL DESEMPEÑO DE BATERÍAS DE IONES DE LITIO A PARTIR DE LA EFICIENCIA COULÓMBICA Y LA CONDUCTIVIDAD DEL ELECTROLITO**, por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que el estudiante continúe con los trámites que la Universidad Autónoma del Estado de Morelos tenga establecidos para obtener el grado mencionado.

Atentamente
Por una humanidad culta

Dr José Alberto Hernández Aguilar
Profesor- investigador
Facultad de Contaduría, Administración e Informática

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

Se expide el presente documento con firma electrónica UAEM, soportada por el certificado vigente a la fecha de su elaboración y con efectos plenos de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS PUBLICADOS en el ÓRGANÓ INFORMATIVO UNIVERSITARIO "ADOLFO MENÉNDEZ SAMARÁ" número 117 de fecha 20 de abril de 2021.

Sello electrónico

JOSE ALBERTO HERNANDEZ AGUILAR | Fecha:2026-11-09 19:08:23 | FIRMANTE

Zvi7qoQnMC48aTtyhuapK/oZGl4j0VvHUj89JKlodWHgy66LVWPalTWQVPfdR4H54rhShyx9MTaR1BTaUWWAeuM0QhuV0DaXUiJCVkOTriXN6xm/UeDqPi+4zCizykWpy2RLsqpS3l/178/IgmaB2V1rOi1zkjvrCT+jpmgvPTyTzb/u3BtBJFX67nl+Izrjgn7ymTUPT1qHLUu9r/l3vqD2mDqL4yiRBEBP2+yaONVt37Sc4gJoDtFzfWpmg5/8jZpwxjL+nalsjfSUSARD/uK4rqfWujJl3wTv6Y3GUVy2uYKXGprMkmruIHv4XMNjYICj/ll3jct/ldb0jsqJig==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:

ONsma5Jxr

https://efirma.uaem.mx/noRepudio/j6flDmLeHTUrBspuENQfLfzy2cCQrfcZ

Cuernavaca, Morelos a 9 de Noviembre del 2026.

DR. MIGUEL ÁNGEL GARCÍA GARNICA
DIRECTOR DE LA FACULTAD DE CONTADURÍA,
ADMINISTRACIÓN E INFORMÁTICA.
PRESENTE

En mi carácter de revisora de tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado de la Maestría en Optimización y Cómputo Aplicado, del estudiante Sergio Rubén Ocampo Pérez, con matrícula 10025065, con el título **APLICACIÓN DE APRENDIZAJE AUTOMÁTICO PARA LA CLASIFICACIÓN DEL DESEMPEÑO DE BATERÍAS DE IONES DE LITIO A PARTIR DE LA EFICIENCIA COULÓMBICA Y LA CONDUCTIVIDAD DEL ELECTROLITO**, por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que el estudiante continúe con los trámites que la Universidad Autónoma del Estado de Morelos tenga establecidos para obtener el grado mencionado.

Atentamente
Por una humanidad culta

Dra. Lorena Díaz González
Profesora- investigadora
Centro de Investigación en Ciencias UAEM

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

Se expide el presente documento con firma electrónica UAEM, soportada por el certificado vigente a la fecha de su elaboración y con efectos plenos de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS PUBLICADOS en el ÓRGANÓ INFORMATIVO UNIVERSITARIO "ADOLFO MENÉNDEZ SAMARÁ" número 117 de fecha 20 de abril de 2021.

Sello electrónico

LORENA DIAZ GONZALEZ | Fecha:2026-11-09 21:50:04 | FIRMANTE

H3ITc7BY0qAKR8HT+TmeYdZlUVFU88hWs+xpqRJuTVZ03icxQ2njpP6r+RgKHfprJ3hdY3SEF49fN/HQtOh/BNwhL05s8Ccw1eJMYd4p6J0OK8vRKEQ0jkDu68GgEDvVp+msU7IBwJgFFZO851/bdL18LU9Ho/4a4DwJk1oheP2W3UsLKAJKw6WQl1HaIbtAW8Ju14+WayGb1Kr/IIIsg9bSXBazJdRpIy1QpG26qktwyAydgEkQMxEd5FbVTJGovvremjZTOiutI54uPm6dS13ehCu9ybLcYxV7QTBNyNmF/YENffZENR3Zd/fHCVvyDkXjlk2pqJ8fNFmHKEGI+Q==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:

jkWd6obFN

https://efirma.uaem.mx/noRepudio/g38HaYHliKjM9OKWuUbgQKgK8Azaa3xK