

DETECCIÓN TEMPRANA DE FALLAS EN REDES DE ENERGÍA ELÉCTRICA
A PARTIR DE VARIABLES METEOROLÓGICAS

Sebastian Ruiz Ruiz



Escuela de Ingeniería Eléctrica y Electrónica
Facultad de Ingeniería
Programa de Pregrado en Ingeniería Electrónica
Junio
2025

DETECCIÓN TEMPRANA DE FALLAS EN REDES DE ENERGÍA ELÉCTRICA A PARTIR DE VARIABLES METEOROLÓGICAS

Sebastian Ruiz Ruiz

Documento presentado al
Programa de Pregrado en Ingeniería Electrónica
como requisito para optar por el título de
Ingeniero Electrónico

Director: Wilfredo Alfonso-Morales, PhD.

Codirector: Cristhian Eduardo Fernández, MSc.



Universidad del Valle

Escuela de Ingeniería Eléctrica y Electrónica
Facultad de Ingeniería
Programa de Pregrado en Ingeniería Electrónica
Junio
2025

Aceptación

Nota de aceptación:

Firma del Jurado

Firma del Jurado

Wilfredo Alfonso-Morales, PhD., Director

Cristhian Eduardo Fernández, MSc., Codirector

Cali, Junio de 2025

Dedicatoria

Dedico este trabajo a mi familia, por su amor incondicional y por creer siempre en mí.

Gracias por acompañarme en cada etapa de este proceso, por brindarme su apoyo constante y por ser un pilar fundamental en mi formación como profesional y como persona. Su ejemplo y enseñanzas me han mostrado el verdadero valor del esfuerzo y la perseverancia.

A mis padres:

Marta Elena Ruiz

Eutimio Ruiz

A mis hermanos:

Harbey Ruiz Ruiz

Noralba Ruiz Ruiz

Luis Alberto Ruiz Ruiz

Agradecimientos

Agradezco sinceramente a mi director, Wilfredo Alfonso, y a mi codirector, Cristhian Fernández, por su invaluable guía a lo largo de este proceso. Su apoyo, dedicación y enseñanzas fueron fundamentales para mi formación y crecimiento académico. Ha sido un verdadero privilegio compartir este camino de aprendizaje en torno a temas apasionantes, los cuales me motivaron y fortalecieron en la culminación de mi carrera profesional.

También extendo mi agradecimiento a Wilmar A. Torres por compartir su conocimiento y facilitar mi comprensión del tema de investigación, contribuyendo de manera significativa al desarrollo de este trabajo.

A Dayana Quiñones, quien ha sido un pilar fundamental en mi vida. Gracias por tu apoyo incondicional en cada etapa de este camino, por animarme en los momentos difíciles y por no dejar que me rindiera. Agradezco profundamente haber compartido este trayecto a tu lado, y por ser parte esencial en la culminación de mi carrera profesional.

Finalmente, deseo expresar mi agradecimiento a mis amigos Edier Alexis Murillo y Alejandro Escobar por su constante apoyo, por compartir conmigo sus conocimientos a lo largo de toda esta carrera, así como por su confianza y amistad. También extendo mi gratitud a todas las personas con las que tuve la oportunidad de relacionarme durante mi estancia en la universidad. De cada una de ellas aprendí algo valioso, encontré apoyo incondicional y dejaron una huella significativa en mi formación académica y personal.

Glosario

- ACC** Accuracy. 9, 41, 42, 52
- ANN** Artificial Neural Network. 10, 13, 14, 16, 17
- API** Application Programming Interface. VII, 20, 26–30, 56
- AUC** Area Under the Curve. 9, 10, 13
- BNN** Bayesian Neural Networks. 9, 10, 13
- CNN** Conventional neural networks. 8
- CREG** Comisión de Regulación de Energía y Gas. 2
- DES** Duración Equivalente de las Interrupciones del Servicio. 2
- DL** *Deep Learning*. x, xi, 6, 12, 15, 16, 19
- DNN** Deep Neural Networks. 10
- ECMWF** European Medium-Term for Medium-Range Forecasts. 27
- FCNN** Fully Connected Neural Network. 9
- FES** Frecuencia Equivalente de las Interrupciones del Servicio. 2
- IA** Inteligencia Artificial. VII, 3, 6, 12, 15
- IFS** Integrated Forecasting System. 27
- KNN** k-Nearest Neighbors. 9, 10, 14
- LGBM** LightGBM. 9
- LOF** Local Outlier Factor. 10
- LSTM** Long Short-Term Memory. VI, VII, x, xi, 8, 10, 13, 14, 18, 19, 34–36
- MCC** Matthews Correlation Coefficient. 9, 13
- ML** *Machine Learning*. 6, 10, 12, 15, 16
- N.E.** Not Specified. 13, 14
- NaN** Not a Number. 22
- NERC** North American Electric Reliability Corporation. 9
- OR** Operador de Red. 2
- Recall** Sensibilidad. x, xi, 10, 13, 40–42, 52, 53
- RF** Random Forest. 9, 10, 13
- RMSE** Error cuadrático medio. 8, 13
- RNN** Recurrent Neural Network. VII, 19, 36
- ROC** Receiver Operating Characteristic. 9, 13, 52
- SAIDI** *System Average Interruption Duration Index*. 2
- SAIFI** *System Average Interruption Frequency Index*. 2
- SDL** Sistema de Distribución Local. 2
- SSPD** Superintendencia de Servicios Públicos Domiciliarios. 2
- SVM** *Support Vector Machines*. 3, 9, 10, 14
- UTC** Universal Time Coordinated. 29
- WandB** Weights & Biases. x, xi, 43, 44, 49, 51, 55, 57
- XGB** XGBoost. 9

Tabla de Contenido

	Pág.
Aceptación	I
Dedicatoria.....	II
Agradecimientos	III
Glosario	IV
Lista de Figuras	VII
Lista de Tablas	IX
Resumen	X
Abstract	XI
1. Introducción.....	1
1.1. Objetivos	4
1.1.1. Objetivo General	4
1.1.2. Objetivos Específicos	4
1.2. Estructura del Documento	5
2. Detección temprana de fallas en redes eléctricas	6
2.1. Redes Eléctricas.....	6
2.2. Detección Temprana de Fallas	7
2.3. “Marco conceptual”	15
2.4. Resumen.....	18
3. Análisis de Información y Definición del Modelo de clasificación	20
3.1. Zona de Estudio	20
3.2. Fuente de Datos.....	21
3.3. Limpieza y Selección de Datos	23
3.4. Variables Meteorológicas	26
3.4.1. Base de Datos Unificada	29
3.5. Análisis Exploratorio	31

3.6.	Selección y Especificación del Modelo	34
3.6.1.	Long Short-Term Memory (LSTM)	34
3.7.	Selección y Especificación del Modelo	36
3.7.1.	Long Short-Term Memory (LSTM)	36
3.7.2.	Preparación y Estructuración de la Base de Datos.....	36
3.8.	Métrica para la Evaluación del Desempeño	40
3.9.	Resumen.....	42
4.	Evaluación y Análisis de Resultados	43
4.1.	Metodología de Entrenamiento	44
4.2.	Resultados y Análisis	46
4.2.1.	Modelos para la Detección Temprana	47
4.2.1.1.	Ventana de observación	47
4.2.1.2.	Modelos de barrido amplio	49
4.2.1.3.	Modelos con Hiperparámetros Sintonizados	51
4.2.2.	Comparación de los Modelos Propuestos y el Consultado en la Literatura	54
4.3.	Resumen.....	55
5.	Conclusiones	56
5.1.	Conclusiones Generales	56
5.2.	Trabajos Futuros.....	57
	Referencias	59

Lista de Figuras

FIGURA	Pág.
1.1. Flujo de trabajo de la solución propuesta.	3
2.1. Distribución de la energía eléctrica.	7
2.2. Jerarquía de la inteligencia artificial (IA).	15
2.3. Estructura de una Long Short-Term Memory (LSTM).	18
2.4. Diagrama de metodología propuesta.	19
3.1. Ubicación y Componentes de North West England.	21
3.2. Estructura de la base de datos en formato <i>DataFrame</i>	22
3.3. Distribución de categorías de las fallas.	22
3.4. Fallas de ambiente y clima en el noroeste de Inglaterra.	24
3.5. Fallas depuradas por condiciones meteorológicas en el noroeste de Inglaterra.	26
3.6. Cantidad de datos a solicitar a la API.	30
3.7. Datos con variables meteorológicas.	31
3.8. Matriz de correlación de variables meteorológicas.	32
3.9. Relación de Variables Meteorológicas entre Fallas (1) y No Fallas (0).	34
3.10. Arquitectura del modelo LSTM para la predicción anticipada de fallas. ..	35
3.11. Tipos de arquitecturas de RNN.	36
3.12. Ventana de 4 datos antes de la falla 1.	37
3.13. Ventana de 4 datos antes de la no falla 1.	38
3.14. Etiquetado de fallas.	39

3.15. Diagrama de las variaciones de los datos a evaluar.	40
4.1. Esquema general de la metodología de entrenamiento.	44
4.2. Distribución de parámetros e hiperparámetros.	45
4.3. Entrenamiento de todos los modelos para $t=0$. El mejor modelo se marca con un recuadro.	48
4.4. Mejores modelos en el barrido amplio para cada escenario (resaltados en un recuadro).	49
4.5. Matriz de confusión de los mejores modelos en la sintonización de parámetros amplios.	50
4.6. Mejores modelos en el barrido final para cada escenario (resaltados en un recuadro).	51
4.7. Matriz de confusión de los mejores modelos en la sintonización de parámetros final.	52
4.8. Evaluación con ventanas deslizantes en el modelo con anticipación $t=-1$ alrededor de una falla.	54

Lista de Tablas

TABLA	Pág.
2.1. Referencias relevantes utilizadas en el desarrollo de los antecedentes	13
3.1. Subcategoría de fallas por condiciones meteorológicas.	25
3.2. Variables meteorológicas utilizadas en el análisis.	28
3.3. Códigos de interpretación meteorológica de la OMM (WW).	28
3.4. Interpretación del coeficiente de correlación.....	33
3.5. Matriz de confusión para clasificación binaria.	41
4.1. Especificaciones del equipo de cómputo utilizado.....	43
4.2. Distribución de los datos en las etapas de entrenamiento, validación y prueba para cada escenario.....	47
4.3. Métricas de desempeño en los datos de prueba para cada escenario.	50
4.4. Hiperparámetros óptimos seleccionados para cada escenario.	51
4.5. Métricas de desempeño para los mejores modelos en los datos de prueba para cada escenario.	52
4.6. Métricas de desempeño de los modelos propuestos y el de la literatura....	55

Resumen

En este trabajo de grado se exploran e implementan modelos de inteligencia artificial, específicamente técnicas de *Deep Learning*, con el propósito de desarrollar un sistema de detección temprana de fallas en redes eléctricas, con **resolución horaria**, utilizando variables meteorológicas en el noroeste de Inglaterra. Se plantea una metodología para el desarrollo e implementación de un modelo para la detección de fallas basado en clasificadores con **variables exógenas**. Inicialmente, se realizó una revisión bibliográfica de técnicas y modelos relevantes, a partir de la cual se definió la arquitectura del modelo y el tratamiento de los datos. Este último incluyó la recopilación de variables meteorológicas horarias asociadas a fallas, el preprocesamiento de datos y la visualización de correlaciones entre variables. Se definió una arquitectura basada en redes LSTM para procesar las series temporales y detectar fallas ocasionadas por eventos meteorológicos con una resolución temporal en horas. Con el fin de evaluar su desempeño en la predicción anticipada, se estableció un esquema de etiquetado que desplaza las etiquetas de falla una y dos horas antes del evento real, generando así tres escenarios de clasificación: $t = 0$ (momento de la falla), $t - 1$ (una hora antes) y $t - 2$ (dos horas antes). La base de datos se dividió en conjuntos de entrenamiento, validación y prueba. Asimismo, se utilizó la herramienta WandB para la optimización de hiperparámetros y el seguimiento del rendimiento de los modelos. Los resultados mostraron un desempeño consistente: para $t = 0$, se obtuvo una exactitud de 0.8338, F1-score de 0.8342, precisión de 0.8357 y Recall de 0.8327; para $t = -1$, una exactitud de 0.8320, F1-score de 0.8377, precisión de 0.8132 y Recall de 0.8638; y para $t = -2$, una exactitud de 0.8261, F1-score de 0.8343, precisión de 0.7996 y Recall de 0.8721. Para contextualizar el desempeño del modelo propuesto, se compararon sus resultados con los reportados en [1], aplicado en un contexto similar y que alcanzó valores de 0.929, 0.865, 0.843 y 0.889 respectivamente. Los resultados obtenidos en este trabajo reflejan las particularidades del modelo propuesto. Estas particularidades incluyen un enfoque centrado en la detección anticipada de fallas hasta dos horas antes del evento real, el uso de un esquema de etiquetado desplazado y una configuración que prioriza la sensibilidad del modelo, aun a costa de una mayor tasa de falsos positivos.

Palabras clave— Detección temprana de fallas, redes eléctricas, variables meteorológicas, aprendizaje profundo, LSTM, series temporales, inteligencia artificial.

Abstract

This thesis explores and implements artificial intelligence models, specifically *Deep Learning* techniques, with the purpose of developing an early fault detection system for power grids using **hourly resolution** meteorological variables in Northwest England. A methodology is proposed for the development and implementation of a fault detection model based on classifiers with **exogenous variables**. Initially, a literature review of relevant techniques and models was carried out, from which the model architecture and data processing strategy were defined. The latter included the collection of hourly meteorological variables associated with faults, data preprocessing, and visualization of correlations among variables. An architecture based on LSTM networks was defined to process time series and detect failures caused by meteorological events with an hourly temporal resolution. To evaluate its performance in early prediction, a labeling scheme was established that shifts fault labels one and two hours before the actual event, thus generating three classification scenarios: $t = 0$ (failure time), $t - 1$ (one hour before), and $t - 2$ (two hours before). The dataset was divided into training, validation, and test sets. In addition, the WandB tool was used for hyperparameter optimization and model performance tracking. The results showed consistent performance: for $t = 0$, an accuracy of 0.8338, F1-score of 0.8342, precision of 0.8357, and Recall of 0.8327 were obtained; for $t = -1$, an accuracy of 0.8320, F1-score of 0.8377, precision of 0.8132, and Recall of 0.8638; and for $t = -2$, an accuracy of 0.8261, F1-score of 0.8343, precision of 0.7996, and Recall of 0.8721. To contextualize the performance of the proposed model, its results were compared with those reported in [1], applied in a similar context and which achieved values of 0.929, 0.865, 0.843, and 0.889, respectively. The results obtained in this work reflect the particularities of the proposed model. These include a focus on early fault detection up to two hours before the actual event, the use of a shifted labeling scheme, and a configuration that prioritizes model sensitivity, even at the cost of a higher false positive rate.

Keywords— Early fault detection, power grids, meteorological variables, deep learning, LSTM, time series, artificial intelligence.

Capítulo 1

Introducción

El suministro de energía eléctrica es fundamental para la vida moderna, dado el alto grado de dependencia de los diferentes sectores sociales, económicos e industriales hacia este recurso. La importancia de garantizar la confiabilidad del suministro de energía eléctrica y de las redes eléctricas no puede subestimarse, ya que cualquier interrupción o falla en su funcionamiento podría tener repercusiones directas en los hogares, en los sectores industriales y en sectores críticos como los hospitales, las plantas de tratamiento de agua, las instalaciones de seguridad pública y los centros de datos que dependen de un suministro eléctrico continuo para operar eficazmente [2]. Las interrupciones pueden ocasionar consecuencias severas, como afectaciones a la salud pública, interrupciones de servicios esenciales y pérdidas económicas significativas para las empresas proveedoras, comprometiendo incluso la seguridad nacional. Por esta razón, las empresas de distribución de energía están prestando más atención a la confiabilidad, en respuesta al gran aumento de las expectativas de los clientes y los reguladores [3].

En este contexto, resulta clave comprender las causas principales de las interrupciones, ya que las fallas en el sistema de distribución eléctrica pueden originarse por mantenimiento deficiente, errores humanos, fallas técnicas o condiciones climáticas adversas. Estas últimas, en particular, representan una amenaza creciente debido a su naturaleza aleatoria, destructiva y cada vez más frecuente. Fenómenos como huracanes o tormentas pueden afectar el entorno, derribando árboles y afectando las redes de suministro eléctrico, ya sea de manera directa o indirecta [3, 1], con un alto impacto en los costos operativos y de mantenimiento. En general, estas fallas pueden producir diversas consecuencias importantes en la red, afectando no solo a la misma, sino también a otros elementos circundantes, debido a daños, disparos, sobrecargas o perturbaciones [4, 1]. Por ejemplo, los generadores, transformadores, disyuntores y líneas de transmisión pueden sufrir daños, dependiendo de la gravedad y

el tiempo de la falla, se pueden presentar daños en los equipos de los clientes [4].

Lo anterior se ve reflejado en evidencia reciente. Por ejemplo, reportes de fallas en redes eléctricas de EE.UU. entre 2000 y 2021 indican que aproximadamente el 83 % de los grandes apagones fueron provocados por eventos meteorológicos, evidenciando un incremento significativo frente al periodo 2000–2010. Durante estos años, se registraron 1.542 cortes de energía relacionados con el clima, donde el 58 % fueron provocados por tormentas severas, el 22 % por condiciones invernales como nieve y hielo, el 15 % por tormentas tropicales y huracanes, y el 5 % por calor extremo e incendios forestales. Estos datos reflejan la creciente vulnerabilidad de la infraestructura eléctrica ante fenómenos meteorológicos cada vez más frecuentes e intensos [5].

Ante este panorama, la capacidad de respuesta de los operadores toma especial relevancia. Es por eso que actualmente los operadores de la red de energía eléctrica se enfrentan a la necesidad de realizar una rápida identificación y localización de las fallas, para desplazar al sitio los recursos técnicos y humanos a la mayor brevedad, y reparar la falla en el menor tiempo posible [4]. Sin embargo, la detección tardía y el posterior restablecimiento de fallas contradicen las políticas de calidad y confiabilidad que se establecen a nivel nacional e internacional. A nivel nacional, los parámetros Frecuencia Equivalente de las Interrupciones del Servicio (FES) y Duración Equivalente de las Interrupciones del Servicio (DES), requeridos por la Comisión de Regulación de Energía y Gas (CREG), son fundamentales para garantizar la prestación adecuada del servicio de energía eléctrica [1]. A nivel internacional, los índices de calidad *System Average Interruption Frequency Index* (SAIFI) y *System Average Interruption Duration Index* (SAIDI), que miden la frecuencia y duración promedio de las interrupciones por usuario de los eventos sucedidos en el Sistema de Distribución Local (SDL) del Operador de Red (OR), y el estándar IEEE 1366-1998, también son importantes para evaluar la calidad del servicio eléctrico [3]. Bajo estos parámetros, las empresas de transmisión y distribución de energía eléctrica pueden ser penalizadas por no prestar el servicio de manera eficaz y confiable, según los lineamientos impuestos por los entes de regulación como la CREG y la Superintendencia de Servicios Públicos Domiciliarios (SSPD), donde las sanciones están relacionadas directamente a la cantidad de usuarios afectados y el tiempo de afectación.

De acuerdo con lo expuesto anteriormente, la detección temprana de fallas en redes

eléctricas es esencial para las empresas de servicios eléctricos. Algunos estudios, como el de Zhou [3], han analizado datos históricos de la red y han comparado diferentes técnicas tanto tradicionales como aquellas basadas en Inteligencia Artificial (IA). Las técnicas tradicionales, que se apoyan en la analítica matemática como las ecuaciones diferenciales o la estadística, han mostrado una confiabilidad histórica, pero no siempre han logrado resultados satisfactorios. Por otro lado, las técnicas de inteligencia artificial, que utilizan algoritmos entrenados con datos históricos y con información exógena, han avanzado significativamente en los últimos años, ofreciendo modelos de predicción más precisos. Por ejemplo, las máquinas de soporte vectorial (por sus siglas en inglés SVM) se han utilizado para prever la duración de los cortes y la capacidad del equipo para enfrentar condiciones extremas, lo que ha representado una mejora en la gestión de los sistemas de energía [1].

En este sentido, y considerando los avances recientes en técnicas de IA, se propone el desarrollo de un modelo computacional para la **detección anticipada de fallas**, utilizando **variables meteorológicas como entrada exógena**. La Figura 1.1 presenta la solución propuesta mediante un diagrama de bloques, en el cual se estructuran las etapas clave del proceso. Inicialmente, se dispone de un bloque de base de datos, que integra tanto el registro de fallas como las variables meteorológicas asociadas. Para este fin, se emplea una base de datos histórica de fallas eléctricas suministrada por una empresa operadora de redes del noroeste de Inglaterra. Posteriormente, se define la metodología y el modelo a emplear para el análisis de los datos. Finalmente, el proceso termina con la implementación de un modelo de detección temprana de fallas, el cual permite identificar posibles incidentes a partir de patrones extraídos de los datos meteorológicos.



Figura 1.1: Flujo de trabajo de la solución propuesta.

En este contexto, contar con un modelo que permita la detección temprana de fallas en redes eléctricas constituye una herramienta clave para prever y mitigar las potenciales consecuencias derivadas de las fallas por eventos, en la medida que se puedan identificar y corregir y así, evitar mayores complicaciones [6]. Adicionalmente,

los modelos para predecir fallas en redes de energía eléctrica están experimentando un crecimiento significativo, pero pocos han considerado las series de tiempo para realizar una detección temprana a los eventos. Por ello, este trabajo se enfoca en las fallas de las redes eléctricas causadas por condiciones climáticas adversas utilizando variables meteorológicas, tal que respondiera al interrogante:

¿Cómo detectar oportunamente fallas eléctricas relacionadas con eventos meteorológicos, dentro de una serie temporal, que ayuden a mejorar la respuesta ante interrupciones prolongadas en el suministro de energía eléctrica?

1.1. Objetivos

1.1.1. Objetivo General

Desarrollar un modelo computacional para la detección temprana de fallas de una red eléctrica usando técnicas de inteligencia artificial a partir de variables meteorológicas.

1.1.2. Objetivos Específicos

- Definir el marco de investigación para desarrollar un modelo de detección temprana de fallas en redes eléctricas con una revisión bibliográfica.
- Identificar las características relevantes para el desarrollo de un modelo de detección temprana de fallas en redes eléctricas a partir de un análisis de las series de tiempo de variables meteorológicas.
- Implementar el modelo computacional para la detección temprana de fallas en redes eléctricas considerando técnicas de inteligencia artificial.
- Validar el modelo computacional a través de pruebas funcionales y usando métricas de desempeño.

1.2. Estructura del Documento

Este trabajo de grado está conformado por 5 capítulos, que buscan responder los objetivos específicos planteados, dichos capítulos se describen a continuación:

- En el Capítulo 1 se expone el problema de investigación, destacando la importancia de la energía eléctrica, los retos ante fallas en redes de distribución y la necesidad de anticiparlas mediante variables climáticas. Además, se exponen los métodos de detección temprana y modelos existentes, y se define el alcance del trabajo mediante una pregunta problema y sus objetivos.
- El Capítulo 2 aborda la revisión del estado del arte, presentando los trabajos más relevantes que han sido considerados como referencia para esta investigación. Asimismo, se desarrollan los conceptos teóricos fundamentales que sustentan el estudio, tales como las series de tiempo, las variables meteorológicas y los principios básicos de la inteligencia artificial. También se introducen los diferentes modelos de aprendizaje utilizados y los mecanismos empleados para evaluar su desempeño.
- El Capítulo 3 expone detalladamente la solución metodológica desarrollada en este trabajo. Se inicia con una descripción de la base de datos utilizada, seguida del tratamiento y preparación de los datos para su posterior análisis. Posteriormente, se explica el proceso de extracción de las variables meteorológicas necesarias para la implementación del modelo. Finalmente, se presenta la arquitectura propuesta para el sistema de detección temprana de fallas.
- El Capítulo 4 presenta los resultados obtenidos a partir de la implementación de los diferentes modelos propuestos para cada uno de los escenarios definidos. Se recopilan y analizan los mejores parámetros de desempeño alcanzados, con el fin de realizar una comparación objetiva frente a un modelo de referencia encontrado en la literatura, el cual aborda una metodología similar.
- Finalmente, en el Capítulo 5 se exponen las conclusiones de la investigación, en las que se sintetizan los principales hallazgos obtenidos, se reconocen las limitaciones del estudio y se proponen posibles líneas de trabajo futuro, a partir de los aportes realizados en este trabajo.

Capítulo 2

Detección temprana de fallas en redes eléctricas

Este capítulo presenta el marco de referencia para la detección temprana de fallas en redes de energía eléctrica a partir de una revisión bibliográfica. El enfoque de esta revisión hace un énfasis particular en las variables meteorológicas como un factor clave en la ocurrencia de fallas. De los diversos trabajos desarrollados en esta temática se examinan los diferentes métodos de IA que se han utilizado, abarcando tanto técnicas de *Machine Learning* (ML) como de *Deep Learning* (DL).

La revisión consideró trabajos en otras áreas además de las redes eléctricas, donde se consideraron variables meteorológicas u orientaron sus aportes hacia la detección temprana de fallas. En general, se realizaron consultas en fuentes como IEEE y Scopus, abarcando información desde el 2010, utilizando palabras clave como: *forecasting, prediction, early detection, artificial intelligence, electrical networks*.

2.1. Redes Eléctricas

Las redes eléctricas constituyen un componente fundamental para garantizar el suministro de energía, canalizando este recurso desde las centrales generadoras hasta los consumidores finales. Este proceso se inicia mediante conexiones con la red de transmisión, que abarca desde las subestaciones de alta tensión hasta las redes de media tensión. La energía llega a los usuarios a través de centros de transformación, donde se realiza la conversión de media tensión a baja tensión [7]. Una representación gráfica de lo que compone el sistema eléctrico de potencia se presenta en la Figura 2.1. En general, las redes de distribución corresponden principalmente a líneas aéreas, las cuales se encuentran expuestas a las condiciones meteorológicas a lo largo de su

recorrido.

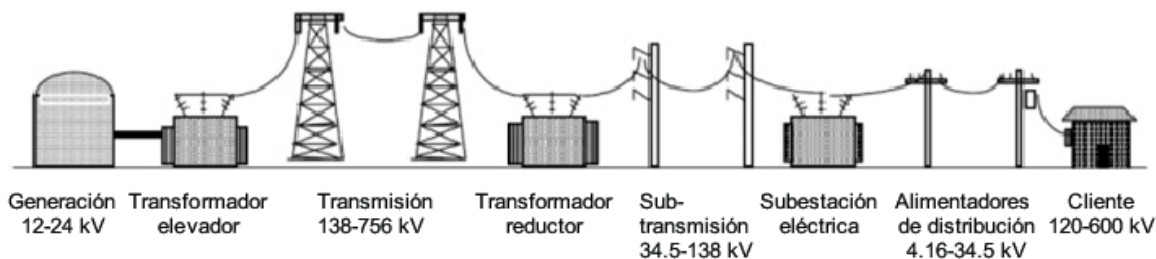


Figura 2.1: Distribución de la energía eléctrica. Adaptado de [8]

En el contexto de las fallas en las redes eléctricas, las variables exógenas son aquellas condiciones que ejercen una influencia indirecta sobre el sistema, modificando la probabilidad de ocurrencia de interrupciones. Por ejemplo, las condiciones climáticas, como la precipitación, la humedad relativa o la presión atmosférica, pueden tener un impacto significativo en el desempeño de la red. Una alta pluviosidad puede incrementar la probabilidad de cortocircuitos en líneas aéreas, mientras que una elevada humedad relativa favorece procesos de corrosión en los equipos y estructuras, aumentando así la probabilidad de falla en el suministro eléctrico [9].

Entre estas variables exógenas, las condiciones meteorológicas adversas no solo afectan la demanda de energía, sino que también pueden ser una causa directa de fallas en la infraestructura del sistema eléctrico. Fenómenos como tormentas eléctricas, vientos fuertes, nieve, hielo o lluvias intensas interactúan con las líneas eléctricas, provocando fallos momentáneos en los equipos o incluso la interrupción total del suministro eléctrico [6].

2.2. Detección Temprana de Fallas

Esta sección analiza investigaciones enfocadas en la detección temprana de fallas, abarcando tanto el sector eléctrico como otras áreas, mediante la aplicación de modelos de pronóstico basados en inteligencia artificial. Se examinan distintas metodologías utilizadas para la predicción de fallos, junto con las métricas empleadas para evaluar su desempeño, permitiendo así comprender las estrategias más efectivas y su impacto en la fiabilidad de los sistemas.

En un trabajo desarrollado por Ruijun Guo et al. [10] se resalta la importancia de detectar tempranamente fallas en la caja de cambios de turbinas eólicas. Para esta tarea, se emplearon redes neuronales de memoria larga a corto plazo (LSTM) optimizadas mediante el algoritmo *Adam*. Las LSTM resultan óptimas para capturar de manera efectiva la información en series temporales, posibilitando una precisa predicción de variables que experimentan cambios graduales y aprovechando la correlación de datos a largo plazo, aspecto fundamental en la detección de fallas.

Para este análisis, se seleccionaron ocho parámetros como variables de entrada, utilizando datos operativos previamente recolectados de turbinas eólicas en condiciones normales. Los valores históricos de estas variables se emplearon como entrada para la red LSTM, mientras que la temperatura actual del aceite de la caja de cambios se tomó como la variable a predecir. Para caracterizar condiciones normales y anormales, se calcularon los residuos entre los valores predichos por el modelo LSTM y las mediciones reales de la temperatura del aceite de la caja de cambios, generando así un indicador de diagnóstico de fallas. El umbral de alarma de fallas se estableció en función del error residual previsto del modelo LSTM en condiciones de operación normal.

Los resultados de este estudio son alentadores, ya que el método propuesto permite detectar anomalías en la temperatura del aceite de la caja de cambios con **hasta una hora de anticipación**. Además, se destaca la eficacia de las LSTM optimizadas con Adam en el manejo de datos, superando a los métodos tradicionales de monitoreo y detección de fallas. En comparación con las redes neuronales convolucionales (CNN), las LSTM resultan especialmente aptas para capturar correlaciones a largo plazo en series temporales. Finalmente, se empleó el error cuadrático medio (RMSE) como métrica de desempeño para evaluar la precisión del modelo desarrollado, calculando la diferencia entre los valores reales y los valores predichos de la temperatura del aceite durante condiciones normales de funcionamiento.

En otro estudio llevado a cabo por Mauro Atrigna et al. [11], se destaca la aplicación de diversas técnicas de aprendizaje automático para predecir fallas en redes de distribución de energía durante olas de calor. El enfoque del estudio integra información meteorológica, datos de infraestructura y mediciones del sistema para anticipar posibles fallas en la red de distribución. Inicialmente, se analizan estos datos para identificar correlaciones entre eventos de fallas previos y series temporales

asociadas a las condiciones climáticas y las mediciones del sistema. Posteriormente, se emplea un clasificador para evaluar datos en tiempo real y prever fallas futuras en la red.

Se implementaron varios algoritmos de aprendizaje automático, incluyendo Random Forest (RF), LightGBM (LGBM), XGBoost (XGB), *Support Vector Machines* (SVM), Fully Connected Neural Network (FCNN) y k-k-Nearest Neighbors (KNN). Para evaluar el rendimiento de estos algoritmos, se utilizaron métricas como el Accuracy, el coeficiente de correlación de Matthews (MCC) y el área bajo la curva característica operativa del receptor (ROC-AUC). Los resultados revelaron que el algoritmo RF obtuvo el mejor desempeño, con un MCC de 0.45 y un ROC-AUC de 0.80. Le siguieron LGBM y XGB, con valores similares de MCC de 0.43 y ROC-AUC de 0.79, respectivamente. En contraste, FCNN y KNN mostraron el rendimiento más bajo en la evaluación.

En un reciente estudio realizado por T. Alquthami et al. [4], se desarrolló un modelo basado en redes neuronales artificiales para predecir interrupciones en líneas de transmisión de alta tensión. Para este propósito, se utilizó una base de datos con eventos históricos proporcionada por la *North American Electric Reliability Corporation (NERC)*. El modelo fue entrenado con datos de diversas fallas, incluyendo aquellas causadas por condiciones climáticas extremas, fallos de equipo y operaciones de mantenimiento. El análisis incorporó factores como las condiciones climáticas, las características de la red y el historial de mantenimiento. Se emplearon técnicas de retropropagación para optimizar el modelo, logrando una alta precisión predictiva. La investigación demostró la capacidad del modelo para anticipar interrupciones y mejorar la gestión de la red eléctrica. Los resultados obtenidos reflejan la eficacia del modelo en la reducción del impacto de las fallas en el sistema eléctrico, asegurando una operación continua y segura del sistema.

En otro trabajo desarrollado por Ying Du et al. [1] Los autores diseñaron un modelo de predicción de fallas en un sistema de distribución eléctrica en una ciudad del noroeste de China donde se utilizó una base de datos de la empresa de servicios públicos de la ciudad, que incluye detalles sobre fechas, ubicaciones, tipos de fallas y condiciones climáticas, abarcando de enero de 2011 a octubre de 2018. La técnica principal fue una red neuronal bayesiana (BNN, por sus siglas en inglés), comparada

con otros cuatro modelos. Se evaluaron métricas de desempeño como precisión, sensibilidad (Recall), puntuación F1 y exactitud. Las variables exógenas incluyeron parámetros meteorológicos obtenidos localmente de una estación meteorológica. El modelo BNN se destacó por su capacidad para estimar la incertidumbre mediante intervalos de **confianza al 95 %**, lo que aporta valor en la gestión de riesgos y en la toma de decisiones operativas en empresas de servicios públicos.

El estudio realizado por Weining Lu, et al. [12], se aborda una técnica similar a la de Ruijun Guo, pero, utilizando redes neuronales LSTM. Este trabajo adopta un enfoque diferente al emplear un modelo de extracción de características mediante una red neuronal profunda (DNN, por sus siglas en inglés). **Este enfoque se basa en conjuntos de datos** que incluyen rodamientos defectuosos y rodamientos en funcionamiento hasta su fallo, abarcando cuatro categorías distintas de fallas. El modelo se entrena con 300 muestras de datos históricos y se comparó con diversos métodos de detección temprana de fallas, como KNN, Local Outlier Factor (LOF) y SVM. Los resultados indican que el método propuesto puede detectar las fallas con aproximadamente 6 horas de antelación.

En otro documento desarrollado por Richard Tituana et al. [13], se lleva a cabo una evaluación detallada de varios algoritmos de predicción de fallas en redes eléctricas, con un enfoque cuantitativo. Se inicia con la recopilación y limpieza de datos **históricos de operaciones y fallas eléctricas**, incluyendo variables como **voltaje, corriente y temperatura**. Posteriormente, se seleccionan algoritmos de *Machine Learning* (ML), como redes neuronales artificiales (ANN, por sus siglas en inglés), SVM, árboles de decisión, entre otros. La evaluación de estos modelos se realiza mediante métricas como precisión, sensibilidad, especificidad y AUC. Los resultados muestran que las ANN y RF alcanzan los mejores rendimientos, con una precisión del 92% y 89% respectivamente. Las ANN destacan en sensibilidad, mientras que RF y SVM muestran una mayor especificidad.

En otro trabajo desarrollado por Pradeep KV et al. [14], se emplean técnicas de aprendizaje automático para predecir fallas en convertidores de potencia, principalmente en circuitos abiertos y cerrados. Se utilizan datos de mediciones de voltaje y corriente, junto con imágenes procesadas de los convertidores. Mediante un algoritmo y preprocesamiento, se obtienen datos puros de cada convertidor. Las fallas

se detectan utilizando un algoritmo genético y técnicas de optimización. Posteriormente, se aplican técnicas de minería de datos, incluyendo clasificación, regresión y agrupamiento, lo que permite detectar fallas en el sistema eléctrico en una etapa temprana, mejorando así la eficiencia y el rendimiento general.

En el ámbito de las técnicas estadísticas, se destaca un trabajo realizado por Yan Zhang GU et al. [15], aunque el enfoque principal de este proyecto se centra en técnicas de inteligencia artificial, resulta útil debido a su detallada información sobre el uso de datos climatológicos. El estudio examina la detección temprana de fallas en la operación de redes de distribución mediante un *gateway* inteligente. Este sistema emite alertas tempranas a través del análisis de datos históricos y en tiempo real para evaluar las condiciones operativas.

El análisis utiliza datos tanto eléctricos, como voltaje, corriente y nivel de aislamiento, como no eléctricos, incluyendo temperatura, humedad y velocidad del viento. La integración de estos dos tipos de datos permite identificar correlaciones y detectar posibles fallas que podrían surgir bajo condiciones climáticas extremas. Las técnicas empleadas en el análisis incluyen el coeficiente de correlación de producto-momento de Pearson, la prueba de causalidad de Granger y un modelo de mapa de probabilidad. Para emitir alertas tempranas, se establece un umbral de 0.75 en el coeficiente de alerta temprana.

El desempeño del modelo se evalúa cuantificando el número de alertas correctas y falsas alarmas. Este enfoque facilita una evaluación precisa y efectiva de la operatividad de la red de distribución, mejorando la capacidad de respuesta ante condiciones adversas y reduciendo el riesgo de fallas en el sistema. El trabajo subraya la importancia de integrar datos heterogéneos y aplicar técnicas avanzadas de análisis para mejorar la confiabilidad y seguridad en la operación de redes de distribución.

En resumen, los distintos artículos y trabajos revisados destacan la inclusión de variables exógenas a los datos históricos como una tendencia relevante. Un ejemplo de ello es el trabajo de T. Alquthami. [4] en el que se incorporan variables meteorológicas del día de la falla como factores clave. Este enfoque es también uno de los objetivos de nuestro estudio, donde se pretende utilizar las condiciones meteorológicas presentes en el momento de la falla para facilitar su detección temprana. Asimismo, se observa

una amplia gama de modelos de IA, tanto de ML como de DL, además de modelos estadísticos. Sin embargo, es evidente que la elección del modelo más adecuado depende de las variables disponibles, la granularidad de los datos históricos, y, como se mencionó anteriormente, las variables exógenas.

La Tabla 2.1 presenta un **resumen de los antecedentes** revisados, destacando sus principales aportes. Entre las características consideradas se encuentran:

- **Título del documento/artículo:** título de la investigación.
- **Ref:** referencia del documento/artículo.
- **Ubicación:** país donde se desarrolló la investigación.
- **Año:** año de publicación de la investigación.
- **Temática:** Principales aportes y enfoques abordados en la investigación.
- **Variables Exógenas:** descripción de las variables exógenas usadas en el modelo de pronóstico.

Tabla 2.1: Referencias relevantes utilizadas en el desarrollo de los antecedentes

Título	Ubicación	Año	Temática	Variables
Early Fault Detection of Wind Turbine Gearbox Based on Adam-trained LSTM [10]	China	2021	Utilizando un modelo de red neuronal LSTM entrenado con el algoritmo Adam, se ha logrado detectar de manera temprana fallas en la caja de cambios de una turbina eólica, basándose en características relacionadas con la temperatura del aceite. Este enfoque permite anticipar las fallas con hasta una hora de antelación, además se empleó métricas de desempeño como el RMSE para evaluar el modelo desarrollado.	Temperatura del aceite
A Machine Learning Approach to Fault Prediction of Power Distribution Grids Under Heatwaves [11]	Italia	2023	Se están utilizando diversas técnicas de aprendizaje automático para predecir fallas en redes de distribución de energía eólica, tomando en cuenta tanto datos internos como externos del sistema que puedan influir en su desempeño. Para ello, se han analizado datos históricos con el objetivo de identificar correlaciones entre fallas previas y series temporales de condiciones climáticas . Como resultado, el algoritmo RF ha mostrado el mejor rendimiento, evaluado mediante las métricas de MCC y ROC-AUC.	Flujos de energía operativos y Variables meteorológicas
Forecasting Transmission System Outages Using Artificial Neural Networks [4]	Egipto	2021	Se implementa un algoritmo de inteligencia artificial diseñado para el pronóstico de interrupciones en un sistema de transmisión de CA de voltaje extra-alto. El modelo se desarrolla a partir del análisis de registros históricos de interrupciones causadas por diferentes tipos de eventos. Se emplea una ANN, entrenada utilizando técnicas de retropropagación, y se aplican diversas métricas para evaluar su desempeño, obteniendo resultados favorables.	N.E.
Predicting Weather-Related Failure Risk in Distribution Systems Using Bayesian Neural Network [1]	China	2020	Debido a la alta probabilidad de fallas en los sistemas de distribución causadas por condiciones climáticas , se propuso un modelo basado en una BNN para pronosticar interrupciones relacionadas con el viento, la lluvia y los rayos. Para ello, se utilizaron datos históricos de fallas en la región del noroeste de China, junto con información meteorológica de una estación local. El modelo se evaluó utilizando diversas métricas, incluyendo Recall, obteniendo resultados satisfactorios.	Variables meteorológicas

Continúa en la siguiente página

Tabla 2.1 – Continúa de la página anterior

Título	Ubicación	Año	Temática	Variables
Early Fault Detection Approach With Deep Architectures [12]	China	2018	El estudio se enfoca en la detección de fallas en rodamientos, utilizando datos históricos de rodamientos defectuosos y en funcionamiento hasta su fallo. Se emplearon 300 muestras para entrenar una red neuronal del tipo LSTM. Los resultados obtenidos se compararon con otros métodos, como KNN y SVM, y demostraron que las LSTM pueden predecir fallas con hasta 6 horas de antelación.	Señales operativas transformadas en un indicador de desviación
Application of Machine Learning Algorithms in the Prediction of Failures in Electrical Networks [13]	Ecuador	2024	Se evalúan y comparan diversos algoritmos de inteligencia artificial, tanto supervisados como no supervisados, para la predicción temprana de fallas en redes eléctricas. Se analizan algoritmos como ANN y SVM, entre otros, utilizando datos históricos de operación y fallas eléctricas. Los algoritmos se evalúan mediante diferentes técnicas, y los resultados indican que ANN es el modelo más destacado.	N.E.
Detection of Early Fault in Power Electronic Converters through Machine Learning and Data Mining Techniques [14]	India	2023	Se emplean técnicas de aprendizaje automático para predecir fallas en convertidores de potencia utilizando mediciones de voltaje , corriente y otros parámetros relevantes. Se incorpora un algoritmo genético combinado con técnicas de optimización para mejorar la precisión de las predicciones. Finalmente, se aplican técnicas de minería de datos para detectar fallas en el sistema eléctrico en una etapa temprana.	Señales de voltaje y corriente
Research on Intelligent Early Warning Algorithm for Distribution Network Considering Extreme Climate Conditions [15]	China	2020	Se emplean técnicas estadísticas para la detección temprana de fallas en redes de distribución a través de un gateway inteligente, utilizando tanto datos eléctricos como climáticos, como temperatura , humedad y velocidad del viento . Se establece un umbral de 0.75 para activar alertas tempranas ante una posible falla.	Variables meteorológicas

2.3. “Marco conceptual”

Inteligencia Artificial (IA): Existen diversas definiciones de inteligencia artificial, pero en general, se puede entender como "la habilidad de los ordenadores para realizar actividades que normalmente requieren inteligencia humana". En términos amplios, la IA se define como la capacidad de imitar el comportamiento humano mediante el uso de algoritmos, aprendiendo de los datos y aplicando ese conocimiento en la toma de decisiones de manera similar a un ser humano [16, 17].

La inteligencia artificial abarca varios tipos y áreas de estudio, destacando dos principales: el ML y el DL. En la Figura 2.2 se ilustra la jerarquía de estos enfoques, que incluyen técnicas como el aprendizaje supervisado, el aprendizaje no supervisado y el aprendizaje por refuerzo. Estas metodologías se emplean para entrenar sistemas según los objetivos específicos que se desean alcanzar [16, 17].

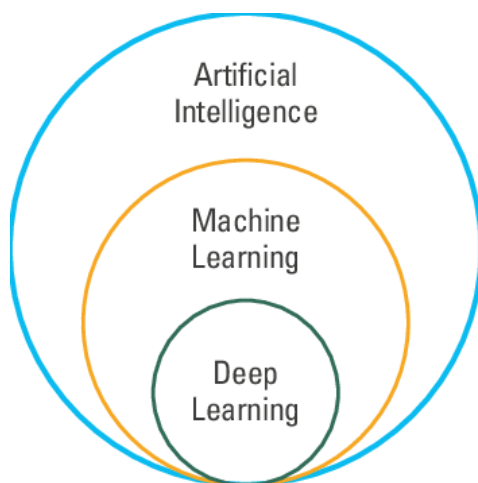


Figura 2.2: Jerarquía de la inteligencia artificial (IA). Tomado de [18]

***Machine Learning* (ML):** Es uno de los enfoques principales de la Inteligencia Artificial. Se centra en permitir que los **algoritmos identifiquen patrones recurrentes** en conjuntos de datos para realizar tareas específicas. Una vez entrenados, los algoritmos pueden reconocer patrones en nuevos conjuntos de datos, realizar predicciones o tomar decisiones basadas en esos patrones. En esencia, el *Machine Learning* permite que las máquinas aprendan de los datos sin estar explícitamente programadas para ello, un ejemplo típico de su aplicación serían las sugerencias o predicciones en situaciones particulares [13, 16].

En el ámbito del *Machine Learning*, existen tres técnicas principales que se utilizan con mayor frecuencia para entrenar los algoritmos, estas se describen a continuación:

- **Aprendizaje supervisado:** Este tipo de aprendizaje emplea datos que han sido etiquetados u organizados previamente para guiar la categorización de nueva información. Este método requiere la intervención humana para ofrecer retroalimentación y ajustar el modelo. Un ejemplo claro de este enfoque es el uso de detectores de los correos no deseados, que son entrenados con ejemplos de correos etiquetados como correo no deseado o correo deseado, con el propósito de identificar y filtrar automáticamente mensajes similares en el futuro [16].
- **Aprendizaje no supervisado:** Los algoritmos trabajan sin datos previamente etiquetados u organizados que les indiquen cómo clasificar la nueva información. En lugar de recibir directrices externas, los algoritmos deben identificar patrones y estructuras por sí mismos. Por lo tanto, este enfoque no requiere intervención humana para la categorización o clasificación de los datos [16].
- **Aprendizaje reforzado:** En el aprendizaje por refuerzo, los algoritmos aprenden a través de experiencias pasadas, evaluando los resultados obtenidos mediante recompensas o penalizaciones. A medida que el modelo interactúa con el entorno, ajusta su comportamiento para maximizar las recompensas y minimizar las penalizaciones, optimizando así su desempeño en tareas específicas [16].

Deep Learning (DL): Es una rama del aprendizaje automático enfocada en resolver problemas que demandan el procesamiento de grandes volúmenes de datos. Este enfoque utiliza redes neuronales profundas, estructuradas en múltiples capas, para identificar y aprender relaciones y patrones complejos en los datos. Debido a la sofisticación de estos modelos, es necesaria una significativa capacidad de procesamiento. El aprendizaje profundo se emplea en una variedad de campos, incluyendo el reconocimiento de voz, la visión artificial, entre otros [16].

Artificial Neural Network (ANN): Es un modelo matemático diseñado para emular el funcionamiento del cerebro humano, que se alimenta de datos para

identificar relaciones o patrones complejos y resolver problemas a gran escala. Las ANN se componen de tres capas principales: la capa de entrada, donde se ingresan los datos; las capas ocultas, que realizan el procesamiento de los datos; y la capa de salida, que proporciona el resultado de la red. Estas redes son ideales para tareas como clasificación, regresión y reconocimiento de patrones. La expresión matemática típica para una ANN incluye el cálculo de la salida en función de las entradas y los pesos ajustados durante el entrenamiento [19, 20]. A continuación, se presenta la expresión matemática que define una ANN:

$$y_k = f_o \left[\sum_{j=1}^n W_{kj} f_h \left(\sum_{i=1}^m W_{ji} x_i + W_{jb} \right) + W_{kb} \right] \quad (2.1)$$

Donde, y_k es la salida, x_i es la i -ésima entrada; w_{ji} son los pesos sinápticos que conectan la capa de entrada y la capa oculta; w_{kj} son los pesos sinápticos que conectan las neuronas de las capas oculta y de salida; f_o y f_h son las funciones de activación en las capas de salida y oculta, respectivamente; n y m son el número de capas de salida y capas ocultas, respectivamente. w_{jb} y w_{kb} son el *bias* (sesgo) para las capas oculta y de salida [21].

Para el entrenamiento de las redes, se utiliza un algoritmo de aprendizaje que itera para encontrar la mejor combinación de pesos sinápticos que minimice una función de costo [21]. La actualización de estos pesos se realiza mediante el algoritmo de retropropagación (Backpropagation), el cual propaga el error a través de la red utilizando el gradiente de la función de costo, basada en el error cuadrático medio entre la salida real y la salida calculada. La expresión matemática correspondiente se presenta en la Ecuación 2.2:

$$E = \sum_{p=1}^P E_p = \sum_{p=1}^P \sum_{k=1}^n (y_{pk} - \hat{y}_{pk})^2 \quad (2.2)$$

Donde, E es el error total, P es el número de datos de entrada, y E_p es el error de la diferencia al cuadrado entre las salidas reales y_{pk} y las salidas previstas $\hat{y}_{pk}$ para el

patrón p [22].

Long Short-Term Memory (LSTM): Son un tipo especial de redes neuronales, creadas específicamente para abordar el problema del **desvanecimiento del gradiente durante el entrenamiento de las redes neuronales tradicionales**. Estas redes están diseñadas con celdas de memoria que permiten mantener y preservar información a lo largo del tiempo. Además, las LSTM se conciben como una solución para retener la activación durante un periodo más extenso. Una unidad básica de LSTM consta de una celda de memoria, una puerta de entrada y una puerta de salida, que juntas controlan el flujo de información y permiten la persistencia de datos a largo plazo [23].

La estructura de las redes con celdas LSTM se compone de cinco elementos principales: la celda de memoria (**C**), que almacena información relevante para futuras predicciones, y el estado oculto (**h**), que conserva la información más reciente procesada. Además, incluye tres puertas que regulan el flujo de información: la puerta de olvido (**f**), que decide qué información descartar; la puerta de entrada (**i**), que incorpora nueva información a la memoria; y la puerta de salida (**O**), que determina el valor de salida y actualiza el estado oculto, como se muestra en la Figura 2.3 [24, 25].

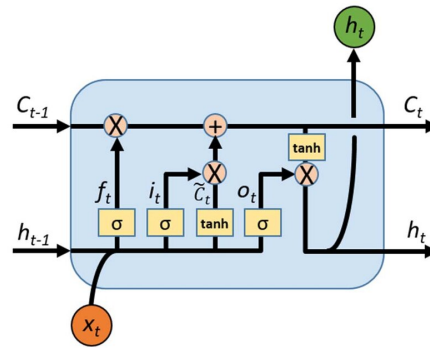


Figura 2.3: Estructura de una Long Short-Term Memory (LSTM). Tomado de [24]

2.4. Resumen

Teniendo en cuenta la retroalimentación de la literatura mencionada anteriormente, junto con los resultados y características relevantes de cada artículo revisado, este proyecto propone una solución para desarrollar un modelo que permita

la detección temprana de fallas en la red eléctrica ocasionadas por variables meteorológicas. En el esquema presentado en la Figura 2.4, se evidencia todo el proceso metodológico propuesto. Para alcanzar este objetivo, se plantea el uso de técnicas de DL, específicamente *Recurrent Neural Network (RNN)*, debido a su capacidad para procesar grandes volúmenes de datos secuenciales, detectar patrones ocultos en series temporales y extraer características de manera automática. Dentro de este enfoque, se utilizará la arquitectura LSTM, que ofrece ventajas clave para el análisis de datos temporales, tales como la capacidad de capturar dependencias a largo plazo, permitiendo al modelo aprender relaciones temporales relevantes en las variables meteorológicas previas a una falla, y la habilidad para procesar datos secuenciales sin la pérdida de contexto temporal.

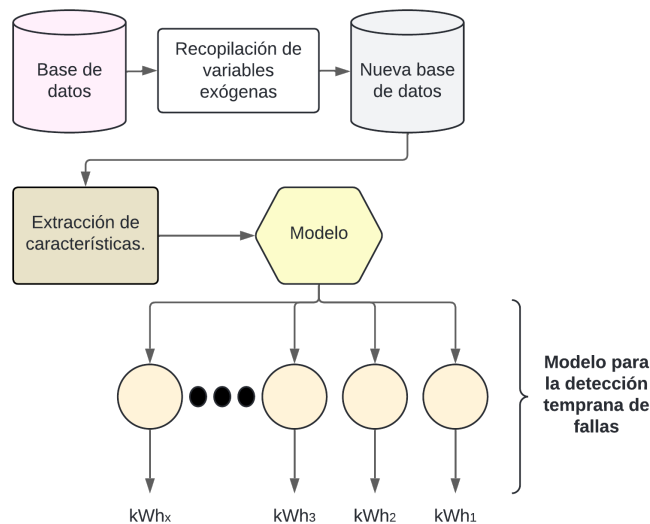


Figura 2.4: Diagrama de metodología propuesta. Adaptado de [25].

Capítulo 3

Análisis de Información y Definición del Modelo de clasificación

En este capítulo se describe el proceso de análisis y construcción del modelo propuesto para la detección temprana de fallas en redes eléctricas causadas por eventos meteorológicos, siguiendo las etapas mostradas en la Figura 2.4. El enfoque inicia con una contextualización geográfica de la base de datos, la cual tiene su origen en el Reino Unido, permitiendo así comprender las características particulares del entorno donde fueron recolectados los datos. Posteriormente, se realiza un proceso de limpieza y selección de datos con el objetivo de eliminar inconsistencias, reducir el ruido y asegurar la calidad del conjunto de entrenamiento. A continuación, se define la API meteorológica utilizada para obtener y recopilar variables meteorológicas relevantes cuya variación esta asociada con la generación de fallas en infraestructuras eléctricas. Luego, se presenta la arquitectura del modelo de clasificación, diseñado para aprender y generalizar la relación entre las variables meteorológicas y la ocurrencia de las fallas, con el objetivo de mejorar la capacidad predictiva del sistema. Finalmente, se establecen las métricas de evaluación y los experimentos realizados para medir el desempeño de los modelos.

3.1. Zona de Estudio

La zona de estudio corresponde al *North West England* que es una región ubicada al norte de *West Midlands* y al oeste de *Yorkshire* y *North East England*, con límites menores con *Gales del Norte*, *East Midlands* y el suroeste de Escocia. Su geografía es diversa, con un sur predominantemente urbano, con ciudades como Liverpool y Mánchester, mientras que el norte alberga impresionantes paisajes naturales, como el Distrito de los Lagos y los Peninos de Lancashire. La región está compuesta por los

condados de Lancashire, Cumbria y Cheshire, además de las áreas metropolitanas de Gran Mánchester y Merseyside; mientras que Mánchester, Liverpool y Warrington son zonas urbanizadas, Lancashire y Cumbria presentan un carácter más rural, con algunas ciudades grandes, y Cheshire se distingue por sus llanuras agrícolas. El clima de *North West England* es templado y húmedo, con frecuentes lluvias durante todo el año; en la costa, los vientos pueden ser fríos, mientras que en verano las temperaturas suelen oscilar entre 20 y 25°C, con ocasionales picos de más de 30°C en el interior [26].

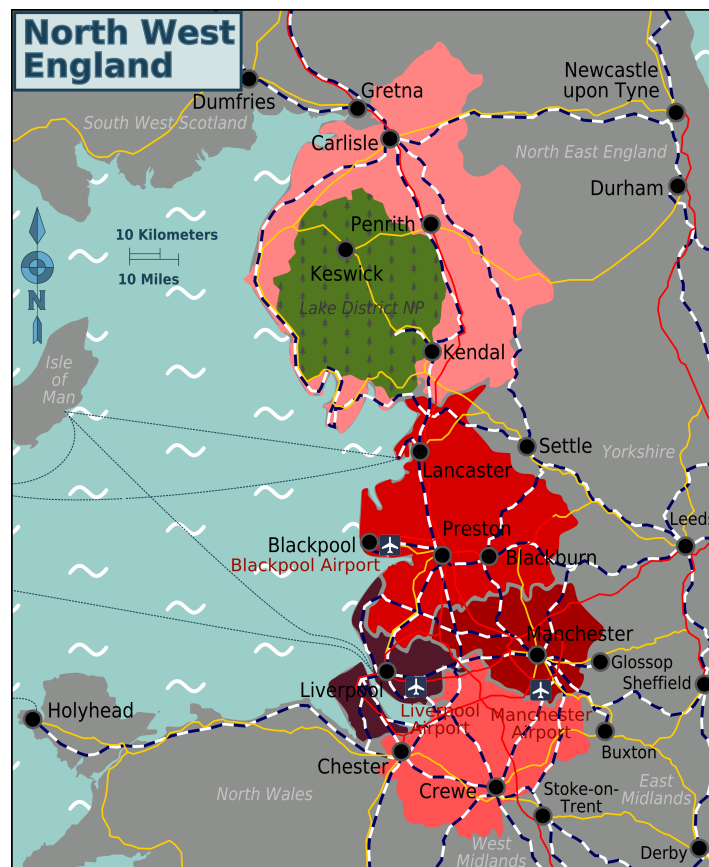


Figura 3.1: Ubicación y Componentes de North West England. Tomado de [26]

3.2. Fuente de Datos

La base de datos utilizada para el desarrollo del modelo de detección temprana de fallas a partir de variables meteorológicas proviene de la empresa *Electricity North West* ubicada en el noroeste de Inglaterra y ha sido obtenida mediante acceso libre. Esta base de datos recopila todas las fallas ocurridas entre los años 2000 y 2023, clasificadas por categoría y subcategoría. Cada registro incluye información detallada, como el ID de

la falla, las coordenadas geográficas del incidente, el código y nombre del distrito, y el tiempo requerido para la resolución del problema. Además, la base de datos contiene categorías especiales que describen el componente afectado en cada falla, en la Figura 3.2 se pueden observar las columnas principales.

	Exceptional Event ID	Direct Cause Category	Direct Cause	District Code	District Name	Incident Date-time	Restoration Date-time	geometry	
	0	NaN	Third Party	by Local Building Authorities or their Contrac...	16	Manchester	1998-01-09T08:09:00+00:00	1998-01-09T08:09:00+00:00	53.47331296940958, -2.2312190158439638
	1	NaN	Unclassified or Unknown	Faulty Manufacturing, Design, Assembly or Mate...	41	Preston	1997-07-19T09:12:00+01:00	1997-07-19T09:12:00+01:00	53.61392948503613, -2.7044979634646418
	2	NaN	Third Party	Willful Damage, Interference (not including Met...	16	Manchester	2000-04-03T00:43:00+01:00	2000-04-03T00:43:00+01:00	53.48307488356014, -2.244229572164601
	285204	NaN	Weather and Environment	Corrosion	61	Kendal	2024-02-14T16:20:00+00:00	NaN	54.22856701429348, -2.769090744341674
	285205	NaN	Companies	Deterioration due to Ageing or Wear (excluding...	61	Kendal	2024-02-13T11:26:00+00:00	NaN	54.36999155356873, -2.911797283335996
	285206	NaN	Companies	Deterioration due to Ageing or Wear (excluding...	63	Carlisle	2024-02-14T18:40:00+00:00	NaN	54.8859518584598, -2.966797640792277

Figura 3.2: Estructura de la base de datos en formato *DataFrame*.

En la Figura 3.3 se muestra la distribución de fallas según cuatro categorías del RIIO-ED1: Anexo F (Ofgem). *Companies* agrupa fallas atribuibles a la empresa distribuidora (mantenimiento, errores operativos, defectos en equipos); *Unclassified or Unknown* incluye fallas sin causa determinada por falta de datos o naturaleza del evento; *Weather and Environment*, de especial interés en este estudio, abarca condiciones meteorológicas adversas o factores ambientales como erosión o vegetación; y *Third Party* refiere a daños causados por agentes externos como contratistas o accidentes. Se priorizó la categoría de *Weather and Environment* y se aplicó un proceso de limpieza de datos, dada la presencia de múltiples valores no numéricos (NaN) que dificultaban el análisis.

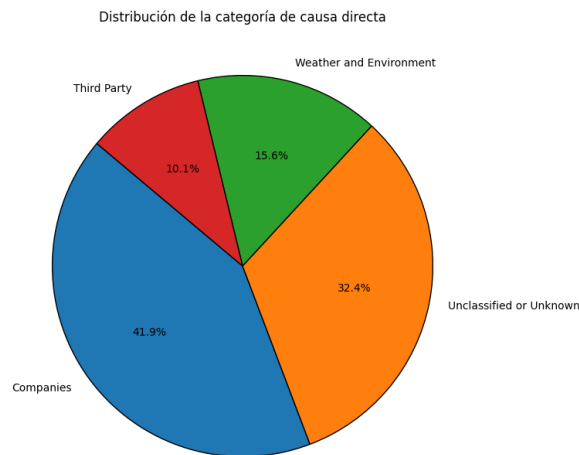


Figura 3.3: Distribución de categorías de las fallas.

3.3. Limpieza y Selección de Datos

Para este proyecto, se trabajó exclusivamente con las fallas clasificadas en la categoría de ambiente y clima. Como primera aproximación, se realizó la limpieza de la base de datos resultante, eliminando aquellos registros que contengan valores faltantes en la columna de coordenadas, ya que esta información es fundamental para el estudio y análisis. Del mismo modo, se eliminaron los registros con datos nulos en las columnas de fecha y hora, al ser esenciales para la temporalidad del análisis. Tras la limpieza de los datos, se llevó a cabo un análisis exploratorio inicial, utilizando GeoDataFrame y shapefile¹ para representar las ubicaciones geográficas de las fallas en un mapa (Ver Figura 3.4). El objetivo de esta etapa fue identificar bases de datos históricas de variables meteorológicas relacionadas con los puntos donde se registraron las fallas. Para ello, se seleccionó la región del Reino Unido, específicamente el noroeste de Inglaterra, dado que es la zona donde se encuentran localizados los eventos de falla.

Para la visualización de las fallas en un mapa del noroeste de Inglaterra, se implementó el siguiente algoritmo:

```
Inicio
  Inicializar variables
  Importar librerías
  Copiar los datos originales
  Eliminar valores faltantes en la columna coordenadas
  Convertir coordenadas en puntos geográficos
  Crear un GeoDataFrame con los puntos y definir su sistema de coordenadas
  Cargar el shapefile del mapa
  Filtrar la región North West de "UKD"
  Mostrar información del shapefile
  Si las proyecciones no coinciden:
    Ajustar sistema de coordenadas del GeoDataFrame
  FinSi
  Graficar el mapa con la región "UKD" y los puntos en rojo
  Mostrar el mapa
Fin
```

¹Un GeoDataFrame es una extensión de un DataFrame que integra datos espaciales mediante geometrías asociadas a atributos tabulares, mientras que un shapefile es un formato clásico para almacenar datos vectoriales como puntos, líneas o polígonos junto a sus descripciones [27].

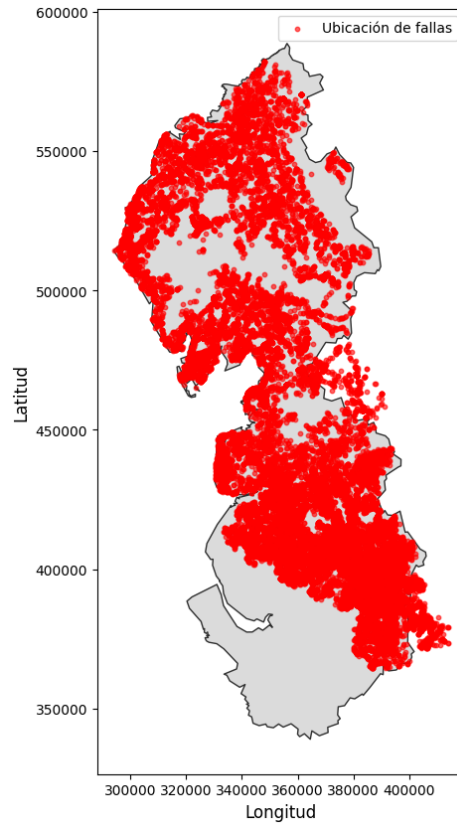


Figura 3.4: Fallas de ambiente y clima en el noroeste de Inglaterra.

Teniendo en cuenta lo mencionado en [11], el uso de variables climáticas para el pronóstico de fallas resulta una estrategia útil. Sin embargo, a diferencia de dicho estudio, este trabajo **se enfocó exclusivamente en variables meteorológicas, mientras que ellos incorporan también variables internas del sistema.** Durante una inspección más detallada de la base de datos, se identificó que muchas de las fallas clasificadas dentro de la categoría de ambiente y clima no eran óptimas para este tipo de análisis. Por ejemplo, causas como **corrosión, condensación o Árboles en crecimiento** pueden estar influenciadas por múltiples factores no necesariamente climáticos, como el diseño de la infraestructura, deficiencias en el mantenimiento o condiciones particulares del entorno. Por esta razón, se decidió aplicar un filtro adicional, seleccionando únicamente aquellas fallas que pertenecían a una subcategoría específica, las cuales se encuentran listadas en la Tabla 3.1. Esta selección permitió garantizar que los datos utilizados fueran representativos y adecuados para el estudio de la influencia de las condiciones meteorológicas en la ocurrencia de fallas.

Tabla 3.1: Subcategoría de fallas por condiciones meteorológicas.

Categoría
Viento y vendaval
Materiales transportados por el viento
Nieve, aguanieve y ventisca
Inundación
Incendio no debido a averías
Lluvia
Hielo
Niebla y escarcha

Un aspecto fundamental en el desarrollo de este proyecto fue asegurar la precisión de las coordenadas geográficas de las fallas, específicamente la longitud y latitud. De acuerdo con [28], la georreferenciación debe reflejar la incertidumbre real de los datos, por lo que se recomienda un formato de entre 4 y 6 decimales. En este caso, se optó por redondear las coordenadas a 5 decimales, buscando un equilibrio entre precisión y eficiencia en el procesamiento. Para facilitar la búsqueda de variables meteorológicas, el *DataFrame* se estructuró con cuatro columnas clave: longitud y latitud (extraídas de una columna combinada), fecha y hora del evento, e ID de la falla, lo que permite contextualizar temporalmente cada registro. Como se muestra en la Figura 3.5, también se realizó una depuración de datos duplicados, garantizando una base de datos organizada y lista para el análisis. La organización se llevó a cabo mediante el siguiente código:

```

Inicio
  Importar librerías
  Cargar la base de datos
  Filtrar las fallas de la categoría "Weather and Environment"
  Seleccionar solo las fallas que pertenecen a subcategorías
  Eliminar registros sin coordenadas ("geometry")
  Convertir las coordenadas de texto a tuplas (latitud, longitud)
  Redondear las coordenadas a 5 decimales
  Convertir la fecha a formato datetime
  Separar latitud y longitud en columnas individuales
  Seleccionar solo columnas relevantes (latitud, longitud, fecha)
  Eliminar registros duplicados para obtener datos únicos
Fin

```

	id	latitude	longitude	date
0	1	53.35261	-1.92755	2000-05-09 18:00:00+00:00
1	2	53.84827	-2.75814	2000-04-04 13:00:00+00:00
2	3	53.91820	-2.37710	2000-08-21 17:00:00+00:00
3	4	54.13923	-2.61192	2000-04-21 15:00:00+00:00
4	5	54.09821	-2.43470	2000-09-11 23:00:00+00:00
...	...	...	...	...
13894	13895	54.69711	-2.67191	2024-01-23 20:00:00+00:00
13895	13896	54.68908	-2.83192	2024-01-22 05:00:00+00:00
13896	13897	54.81570	-3.21231	2024-01-21 18:00:00+00:00
13897	13898	54.58042	-3.18501	2024-01-25 16:00:00+00:00
13898	13899	54.48219	-2.42163	2024-01-29 15:00:00+00:00

Figura 3.5: Fallas depuradas por condiciones meteorológicas en el noroeste de Inglaterra.

3.4. Variables Meteorológicas

Después de organizar y analizar los datos sobre fallas causadas por condiciones meteorológicas, se identificó que los datos cubren un rango temporal que va desde el año 2000 hasta el año 2023. Debido a que el modelo a desarrollar requiere analizar la evolución de las variables meteorológicas en las horas previas a la ocurrencia de cada falla, es fundamental disponer de información climática no solo en el momento exacto del evento, sino también en los períodos anteriores. Se realizó una búsqueda de bases de datos meteorológicas que permitieran obtener información histórica acorde con la ubicación geográfica de cada falla [3]. Dentro de las fuentes exploradas, Met Office [29] se identificó como una alternativa relevante para la obtención de datos meteorológicos del Reino Unido. No obstante, su información se encuentra disponible principalmente en formato diario, lo que resulta insuficiente para este estudio, en el que se requiere una resolución mayor, idealmente con registros con una frecuencia de horas.

Ante la falta de bases de datos con la resolución temporal requerida, se procedió a explorar el acceso a información meteorológica histórica mediante APIs meteorológicas. Durante este proceso, se identificaron diversas restricciones en los servicios evaluados. Algunas APIs presentaban limitaciones en los planes gratuitos, restringiendo el número de consultas diarias o mensuales, mientras que otras no contenían todas las variables meteorológicas necesarias para el análisis. Finalmente, tras una evaluación comparativa

de las opciones disponibles, se seleccionó la API Open-Meteo, la cual proporciona datos meteorológicos históricos con una **resolución mínima de 1 hora**, permitiendo así la recopilación de información climática con el nivel de detalle requerido para el desarrollo del modelo predictivo de fallas.

La API de Open-Meteo proporciona tanto datos de pronóstico como datos históricos de variables meteorológicas. Los datos históricos abarcan desde 1940 hasta dos días antes de la fecha de consulta. La API permite obtener información con diferentes resoluciones temporales: diaria, horaria y de 15 minutos para los pronósticos; sin embargo, para los datos históricos, la resolución mínima disponible es de 1 hora. Los conjuntos de datos meteorológicos en Open-Meteo se construyen a partir de una combinación de observaciones provenientes de estaciones meteorológicas, aeronaves, boyas, radares y satélites. La integración de estas múltiples fuentes permite generar un registro histórico continuo y detallado de las condiciones climáticas. Además, mediante el uso de modelos matemáticos avanzados, se pueden interpolar y estimar valores en regiones con datos incompletos o sin cobertura directa, como áreas rurales o en océano abierto [30].

Para mejorar la precisión en la representación de las condiciones meteorológicas pasadas, los modelos empleados en estos conjuntos de datos operan con una **resolución espacial de 9 km**, lo que permite capturar detalles más finos en zonas costeras y terrenos montañosos complejos. En términos generales, una mayor resolución espacial proporciona un mayor nivel de detalle y mejora la representación de los fenómenos climáticos a escalas más pequeñas. Uno de los conjuntos de datos empleados por Open-Meteo es el *Integrated Forecasting System (IFS)* del *European Medium-Term for Medium-Range Forecasts (ECMWF)*, el cual ha sido procesado meticulosamente a través de simulaciones de reanálisis realizadas dos veces al día (00:00 UTC y 12:00 UTC), utilizando la versión más actualizada del modelo IFS. Este conjunto de datos, al estar basado en simulaciones que integran observaciones atmosféricas reales, ofrece una de las mayores resoluciones y precisiones para representar condiciones meteorológicas históricas [30]. Además, en la documentación oficial de Open-Meteo, se recomienda el uso del conjunto de datos ERA5 para garantizar la coherencia de los registros meteorológicos a lo largo del tiempo. ERA5 es un conjunto de datos de reanálisis que proporciona información con alta resolución y precisión [31].

Para el análisis de fallas, se seleccionaron variables meteorológicas con base en las subcategorías descritas en la Tabla 3.1. Las variables empleadas, mostradas en la Tabla 3.2, se obtuvieron mediante la API de Open-Meteo.

Tabla 3.2: Variables meteorológicas utilizadas en el análisis.

Variable	Unidad	Descripción
Temperatura 2m	°C	Temperatura del aire a 2 metros sobre el suelo.
Humedad relativa 2m	%	Humedad relativa a 2 metros sobre el suelo.
Velocidad del viento 10m	km/h	Velocidad del viento a 10 metros sobre el suelo.
Dirección del viento 10m	–	Dirección del viento a 10 o 100 metros sobre el suelo.
Código meteorológico	Código	Código numérico de condiciones meteorológicas
Lluvia	mm	Cantidad de lluvia acumulada en un periodo de tiempo.
Cubierta de nube	%	Nubosidad total como fracción de área.

Es importante comprender el significado de cada código climatológico presentado en la Tabla 3.2, ya que estos representan diferentes condiciones meteorológicas relevantes para el análisis. La tabla 3.3 proporciona una descripción detallada de cada código, permitiendo interpretar adecuadamente la información meteorológica asociada a los eventos de interés.

Tabla 3.3: Códigos de interpretación meteorológica de la OMM (WW).

Código	Descripción
0	Cielo despejado
1, 2, 3	Mayormente despejado, parcialmente nublado y cubierto
45, 48	Niebla y niebla con escarcha
51, 53, 55	Llovizna: Intensidad ligera, moderada y densa
56, 57	Llovizna helada: Intensidad ligera y densa
61, 63, 65	Lluvia: Intensidad ligera, moderada y fuerte
66, 67	Lluvia helada: Intensidad ligera y fuerte
71, 73, 75	Nevada: Intensidad ligera, moderada y fuerte
77	Granos de nieve
80, 81, 82	Chubascos de lluvia: Intensidad ligera, moderada y violenta
85, 86	Chubascos de nieve: Intensidad ligera y fuerte
95 *	Tormenta: Intensidad ligera o moderada
96, 99 *	Tormenta con granizo: Intensidad ligera y fuerte

3.4.1. Base de Datos Unificada

Con el propósito de consolidar una base de datos unificada fue necesario disponer de un conjunto de variables meteorológicas que coincidieran tanto temporal como espacialmente con las fallas previamente identificadas. Para ello, se construyó una base de datos que integra las coordenadas geográficas de cada falla, así como la fecha y hora exactas de su ocurrencia, con el fin de solicitar los datos meteorológicos correspondientes a través de la API. Adicionalmente, se realizó un ajuste en la zona horaria, alineando las fechas de la base de datos bajo el formato Universal Time Coordinated (UTC), de modo que se garantizara la correcta correspondencia temporal entre los eventos de falla y las mediciones meteorológicas. Sin embargo, dado que la base de datos meteorológica disponía únicamente de una resolución temporal horaria, fue necesario aplicar un ajuste adicional a las horas de ocurrencia de las fallas. Este ajuste consistió en redondear los registros horarios a la hora entera más cercana. Por ejemplo, si el valor registrado en la columna *date* correspondía a las 17:39:00, tras su conversión a formato UTC, se redondeaba a 18:00:00+00:00. Esta estandarización horaria resultó fundamental para aproximar el momento de las fallas y asegurar la coherencia en la integración de la información meteorológica.

De igual forma, fue importante considerar el intervalo de tiempo antes y después de la falla, entendido como un día completo de datos, teniendo en cuenta que algunos eventos podrían presentarse al final o al inicio de un día. Tras revisar la columna *date*, se concluyó que, para obtener una serie temporal sólida y relevante para el modelo, se debían incluir tanto las variables meteorológicas correspondientes al día de la falla como las del día anterior y posterior a la misma. Esto permitió disponer de un conjunto de datos más extenso y detallado para analizar el comportamiento de las condiciones meteorológicas en relación con las fallas [1].

Para ilustrar de manera clara la cantidad de datos que se solicitaron a la base de datos histórica, una representación gráfica se presenta en la Figura 3.6, que facilita la comprensión de este proceso.

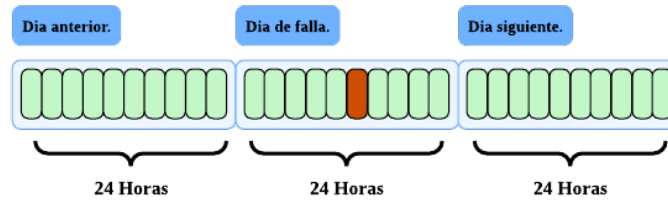


Figura 3.6: Cantidad de datos a solicitar a la API.

Para la obtención de los datos meteorológicos, se utilizó el conjunto de datos ERA5, en conformidad con las recomendaciones proporcionadas por la plataforma Open-Meteo. Dado que la base de datos de fallas puede contener un número elevado de registros, se optó por realizar las solicitudes a la API en bloques de 1.400 fallas por iteración. Esta estrategia se implementó con el objetivo de evitar la pérdida de información relevante y minimizar el riesgo que la API no responda correctamente debido a solicitudes de gran volumen.

Además, el código ha sido adaptado de [32] donde se implementaron mecanismos de manejo de errores para asegurar la estabilidad y robustez del proceso. En caso de presentarse algún problema en la solicitud, como errores de conexión o códigos de estado HTTP inesperados, el sistema está diseñado para gestionar dichas situaciones de manera adecuada. Las solicitudes a la API se realizan a través del protocolo HTTP GET, garantizando así una comunicación eficiente y adecuada con el servicio de datos meteorológicos. A continuación se puede observar el algoritmo utilizado, el cual es una adaptación [32, 31]

```

Iniciar lista vacía 'dataframes'
Especificar url de la api: "https://archive-api.open-meteo.com/v1/era5"
Para cada 1.400 filas en 'data_fallas':
    Obtener latitud, longitud y fecha
    Definir parámetros de solicitud con la información de la fila
    Realizar solicitud http get
        Un día antes de la fecha.
        La fecha actual.
        Un día después de la fecha.
    Si la respuesta es exitosa:
        extraer los datos json
        crear un dataframe con los datos meteorológicos
        añadir el dataframe a 'dataframes'
    Sino:
        imprimir error con el índice de la fila
Concatenar los dataframes en 'dataframes'
Guardar el dataframe final en archivo csv.

```

Una vez se realizadas todas las solicitudes y verificando que se habían obtenido los datos correspondientes a las **72 horas por cada falla**, los *DataFrames* generados se organizaron y concatenaron en un único conjunto de datos. De este modo, se obtuvo una base de datos compacta, lista para realizar las verificaciones necesarias. Esta base de datos, que incluye las variables meteorológicas relevantes, constituye la base de datos total sobre la cual se podrían obtener los datos necesarios para entrenar el modelo de detección temprana de fallas. Como se muestra en la Figura 3.7, la base de datos contiene las variables meteorológicas necesarias para llevar a cabo este análisis.

	latitude	longitude	date	temperature_2m	relative_humidity_2m	wind_speed_10m	wind_direction_10m	precipitation	cloudcover	weather_code
0	53.35261	-1.92755	2000-05-08 00:00:00+00:00	8.488999	81.136310	18.504139	52.907085	0.0	19.800001	0.0
1	53.35261	-1.92755	2000-05-08 01:00:00+00:00	8.139000	81.938890	17.283749	54.323517	0.0	22.200000	1.0
2	53.35261	-1.92755	2000-05-08 02:00:00+00:00	7.839000	83.336900	16.489416	53.880566	0.0	21.900002	1.0
3	53.35261	-1.92755	2000-05-08 03:00:00+00:00	7.539000	85.057990	16.135872	51.340170	0.0	27.600000	1.0
4	53.35261	-1.92755	2000-05-08 04:00:00+00:00	7.389000	86.833630	15.629971	51.546265	0.0	35.400000	1.0
...	...	...	...	...	...	...	...	...	...	...
1010875	54.48219	-2.42163	2024-01-30 19:00:00+00:00	3.136000	95.837760	6.696387	216.253920	0.0	59.399998	2.0
1010876	54.48219	-2.42163	2024-01-30 20:00:00+00:00	2.936000	96.172550	7.993297	215.837740	0.0	44.700000	1.0
1010877	54.48219	-2.42163	2024-01-30 21:00:00+00:00	2.786000	93.797646	10.195057	227.862470	0.0	35.400000	1.0
1010878	54.48219	-2.42163	2024-01-30 22:00:00+00:00	2.686000	92.792070	10.296251	216.469330	0.0	43.500000	1.0
1010879	54.48219	-2.42163	2024-01-30 23:00:00+00:00	3.186000	91.832810	14.904173	217.146770	0.0	84.900000	3.0

942264 rows × 10 columns

Figura 3.7: Datos con variables meteorológicas.

3.5. Análisis Exploratorio

Con el objetivo de comprender el comportamiento de las variables meteorológicas en relación a la incidencia de fallas, se planteó realizar un análisis de correlación. Este método estadístico cuantifica la relación entre dos variables mediante un coeficiente que oscila entre -1 y 1, donde los valores extremos indican una correlación máxima. En este contexto, un coeficiente cercano a 1 refleja una fuerte correlación positiva, mientras que uno cercano a -1 indica una fuerte correlación negativa; por el contrario, valores próximos a 0 sugieren una relación lineal débil o nula. La aplicación de este método permite identificar y cuantificar las tendencias existentes entre las variables meteorológicas, aportando información valiosa para entender su impacto en la ocurrencia de fallas [33].

El análisis de correlación puede llevarse a cabo utilizando diferentes coeficientes. El coeficiente de correlación de Pearson es muy común, especialmente con variables de distribución normal. Por otro lado, el coeficiente de Spearman se utiliza cuando las variables no siguen una distribución normal. Así, antes de aplicar una prueba de correlación, es fundamental conocer el comportamiento de **la distribución de las variables** para seleccionar el coeficiente adecuado y lograr resultados más precisos [33].

Para analizar la relación entre las variables meteorológicas asociadas a las fallas, se utilizó el coeficiente de correlación de Spearman, dado que los datos no presentan una distribución normal. La Figura 3.8 muestra las correlaciones obtenidas entre las variables evaluadas.



Figura 3.8: Matriz de correlación de variables meteorológicas.

Los resultados fueron interpretados conforme a los rangos establecidos en la Tabla 3.4. Se observó que **la correlación entre la precipitación y la temperatura fue prácticamente nula**, clasificándose como sin correlación. De manera similar, **la**

relación entre la precipitación y la dirección del viento presentó un coeficiente cercano a cero, lo que indica la ausencia de una asociación significativa. En cuanto a la velocidad del viento, esta mostró una correlación débil tanto con la precipitación como con la dirección del viento. Por otro lado, se destacó la relación entre la humedad relativa y la nubosidad, cuya correlación se aproximó a un nivel moderado, al igual que la existente entre la nubosidad y la precipitación. No obstante, no se identificaron correlaciones buenas ni fuertes entre las variables meteorológicas analizadas, por lo que no se consideró necesario descartar ni combinar ninguna de ellas en esta etapa del análisis.

Tabla 3.4: Interpretación del coeficiente de correlación. Adaptado de [33]

Coeficiente	Interpretación*
0	Sin correlación
± 0.20	Correlación débil
± 0.50	Correlación moderada
± 0.80	Correlación buena
± 1.00	Correlación perfecta

*Parámetros solo de referencia, no deben ser considerados como estrictos puntos de corte. Estos valores son afectados por el tamaño de muestra.

Otro de los análisis exploratorios realizados consistió en examinar la relación entre las variables meteorológicas durante un evento de falla y las mismas variables en momentos en los que no ocurrió dicha falla. Este enfoque fue esencial para visualizar cómo se comportan los datos en términos de similitud [34]. Con base en estos resultados, se avanzó en la selección del modelo adecuado. Como se puede observar en la Figura 3.9, la observación de una **alta similitud entre los datos correspondientes a eventos de falla y no falla** sugiere que las técnicas de clasificación convencionales, que suelen basarse en supuestos de separabilidad lineal, podrían ser inadecuadas para discriminar eficazmente entre ambas clases. Esta superposición en el espacio de características indica la necesidad de emplear métodos más sofisticados capaces de capturar relaciones complejas y no lineales presentes en

los datos [35].

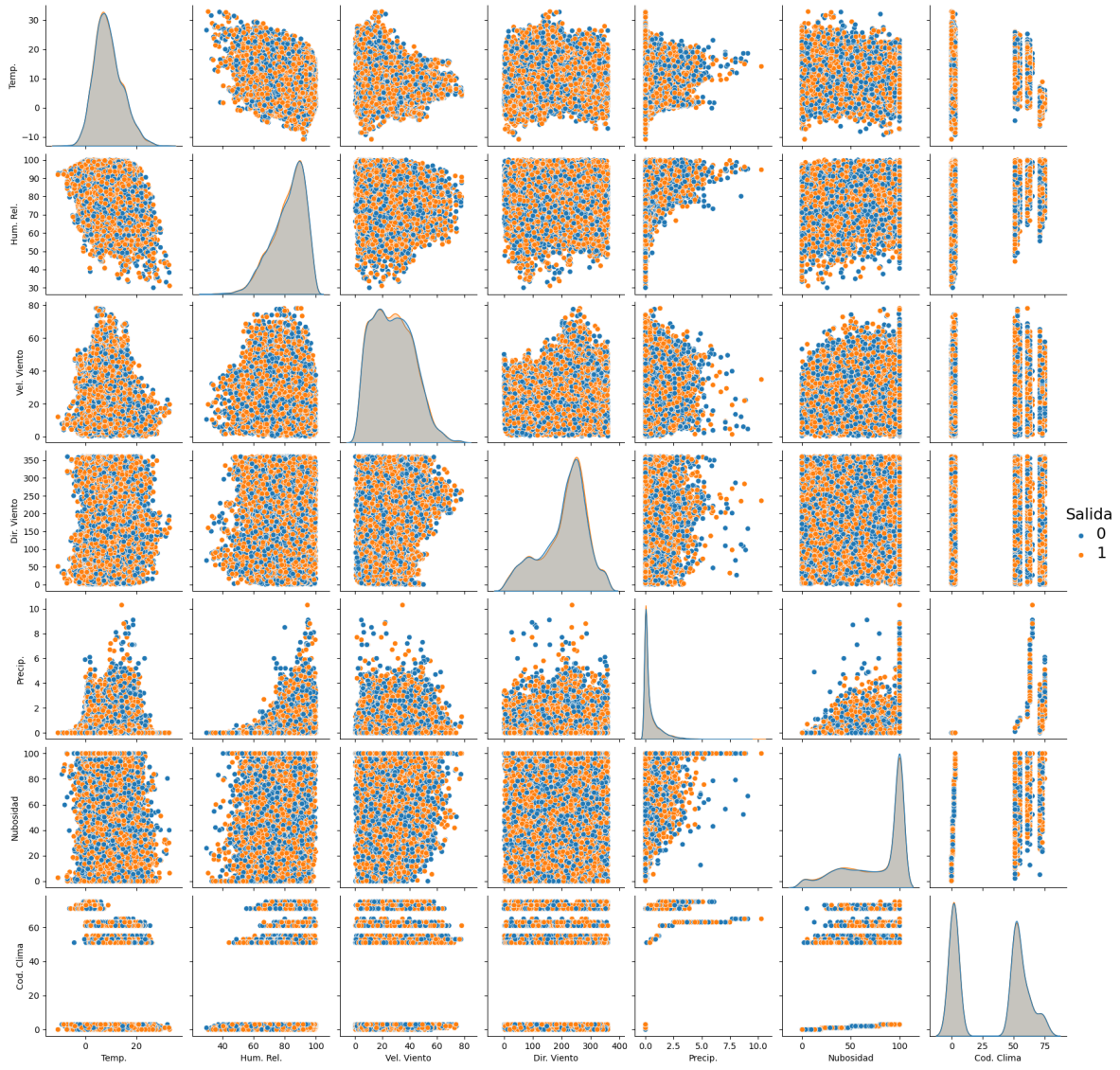


Figura 3.9: Relación de Variables Meteorológicas entre Fallas (1) y No Fallas (0).

3.6. Selección y Especificación del Modelo

3.6.1. Long Short-Term Memory (LSTM)

Para el momento en que se pretendía detectar la falla, era posible utilizar modelos de clasificación similares a los empleados en [13]. Sin embargo, a través del análisis de

correlación, se observó que la relación entre las variables meteorológicas en condiciones de falla y no falla era similar. Esto implica que modelos como los utilizados en [11, 13] presentaron una baja eficiencia cuando se aplicaron a datos donde no existían diferencias significativas entre ambas clases. Por esta razón, se determinó que el modelo para la detección temprana debía implementarse utilizando las técnicas mencionadas en [12]. En este contexto, se seleccionó una arquitectura basada en redes neuronales LSTM, debido a su capacidad para identificar relaciones temporales complejas y capturar patrones secuenciales que podían anticipar eventos futuros.

El enfoque propuesto se centró en predecir la ocurrencia de fallas a partir del análisis temporal de variables meteorológicas. Para lograr una detección temprana efectiva, **la etiqueta de falla fue desplazada una y dos horas antes del evento real, permitiendo anticipar su aparición y facilitar la toma de decisiones oportunas**. La arquitectura del modelo, representada en la 3.10, está compuesta por dos capas LSTM que capturan patrones secuenciales relevantes. Posteriormente, se aplica una normalización por lotes que estabiliza el entrenamiento, seguida de capas densas que transforman las representaciones extraídas. El uso de una capa de *dropout* mejora la capacidad de generalización del modelo. La salida final es una probabilidad que indica la posible ocurrencia de una falla con antelación.

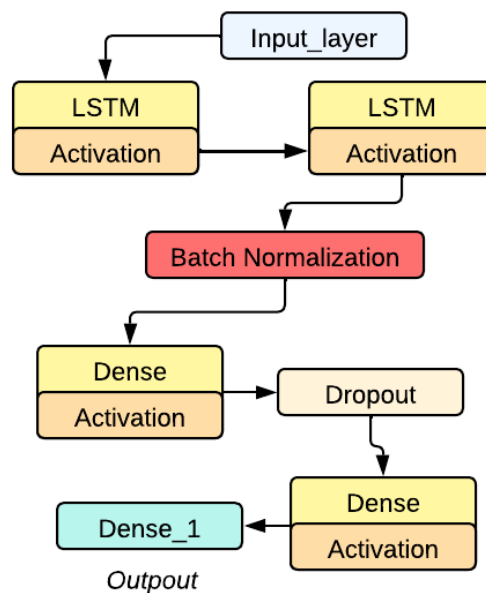


Figura 3.10: Arquitectura del modelo LSTM para la predicción anticipada de fallas.

3.7. Selección y Especificación del Modelo

3.7.1. Long Short-Term Memory (LSTM)

Como se mencionó en [36, 37], existen diferentes tipos de arquitecturas, como se muestra en la Figura 3.11. Sin embargo, para este modelo, la opción más adecuada fue *Many-to-One*, ya que permite analizar secuencias de datos históricos y generar una única predicción basada en los patrones detectados a lo largo del tiempo. Esta estructura fue ideal para la detección de fallas, pues facilitó la identificación de patrones previos al evento crítico, proporcionando un margen de respuesta para la toma de decisiones preventivas.

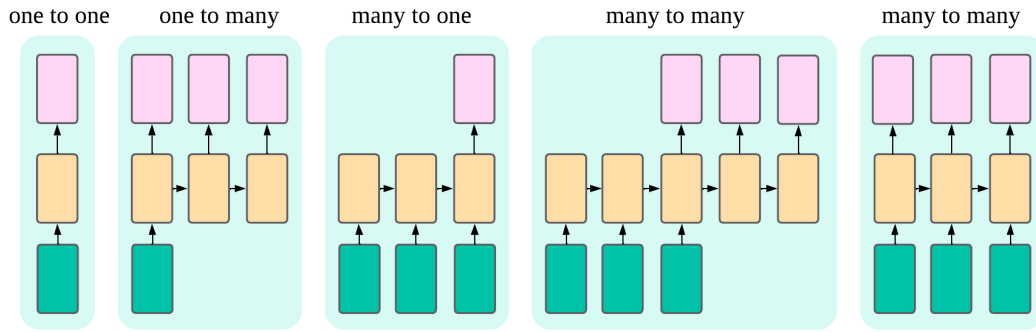


Figura 3.11: Tipos de arquitecturas de RNN. Adaptado de [37]

En este contexto, el modelo propuesto opera como un **clasificador *Many-to-One***, donde la salida es binaria: donde una probabilidad de falla por encima del 50 %, se clasifica como falla; de lo contrario, se considera no falla.

3.7.2. Preparación y Estructuración de la Base de Datos

Para garantizar un procesamiento adecuado de los datos en el modelo, se llevó a cabo una **normalización** de las variables meteorológicas, ya que algunas de ellas presentaban escalas considerablemente mayores que otras. Esta disparidad en las magnitudes puede afectar tanto la estabilidad numérica como la eficiencia del entrenamiento del modelo. En este sentido, se estableció como un hiperparámetro a evaluar la mejor opción entre dos técnicas de normalización: la primera era el *StandardScaler*, que transformaba los

datos para que tuvieran una media de 0 y una desviación estándar de 1; y la segunda era la normalización mediante escalado de máximos y mínimos, que ajustaba los valores a un rango entre 0 y 1 en todos los datos procesados [38].

Otro aspecto fundamental en la configuración del modelo fue la selección del **enventanado**, un hiperparámetro clave que definió la cantidad de datos previos que el modelo LSTM utilizaría para la detección de fallas (Figura 3.12). Este proceso consistió en la recopilación de n registros previos dentro de un intervalo de tiempo ($t - n$), los cuales sirvieron como referencia para que el modelo pudiera estimar la ocurrencia de una falla [39]. La correcta definición del tipo de enventanado fue determinante para el desempeño del modelo, ya que influyó directamente en su capacidad de aprendizaje y generalización.

	id	latitude	longitude	date	temperature_2m	relative_humidity_2m	wind_speed_10m	wind_direction_10m	precipitation	cloudcover	weather_code	salida
0	1	53.35261	-1.92755	2000-05-09 14:00:00+00:00	18.639000	65.903980	1.138420	251.56496	0.1	53.1	51.0	0
1	1	53.35261	-1.92755	2000-05-09 15:00:00+00:00	18.939000	66.619934	1.800000	270.00000	0.0	44.7	1.0	0
2	1	53.35261	-1.92755	2000-05-09 16:00:00+00:00	18.889000	67.709404	1.800000	306.86996	0.0	47.7	1.0	0
3	1	53.35261	-1.92755	2000-05-09 17:00:00+00:00	18.338999	71.002655	2.902413	330.25522	0.0	48.3	1.0	0
4	1	53.35261	-1.92755	2000-05-09 18:00:00+00:00	17.588999	75.164116	4.334974	355.23645	1.5	47.7	61.0	1

Figura 3.12: Ventana de 4 datos antes de la falla 1.

Otro aspecto fundamental en la construcción del modelo fue el **etiquetado de fallas y no fallas**, como se muestra en la Figura 3.12. En este esquema, un valor de 1 en la columna de salida indicaba el instante en el que ocurría la falla, mientras que los registros previos se etiquetaban con 0, ya que en esos momentos no se detectaba ninguna anomalía. Asimismo, en la misma figura se presentó una ventana de observación compuesta por los cuatro registros anteriores a la falla, lo que permitió capturar patrones previos al evento crítico. Este hiperparámetro se evaluó utilizando ventanas de **4, 9 y 14 registros previos**, seleccionadas para representar diferentes escalas temporales: eventos recientes (4), intermedios (9) y tendencias más prolongadas (14), con el fin de identificar cuál ofrecía la mejor representación para la detección de fallas.

Para garantizar un entrenamiento equilibrado, fue fundamental contar con un conjunto de datos balanceado que incluyera tanto momentos en los que se presentaban fallas como aquellos en los que no. En la Figura 3.13, se presentó un ejemplo del mismo conjunto de datos (coordenadas) de la primera falla mostrada en la Figura

3.12, pero en una condición normal, donde no se registró una falla. En este caso, todos los valores de la columna de salida estaban etiquetados como 0, asegurando así la representatividad de los estados normales del sistema dentro del conjunto de datos.

id	latitude	longitude	date	temperature_2m	relative_humidity_2m	wind_speed_10m	wind_direction_10m	precipitation	cloudcover	weather_code	salida
13900	53.35261	-1.92755	2000-05-08 00:00:00+00:00	8.488999	81.13631	18.504139	52.907085	0.0	19.800001	0.0	0
13900	53.35261	-1.92755	2000-05-08 01:00:00+00:00	8.139000	81.93889	17.283749	54.323517	0.0	22.200000	1.0	0
13900	53.35261	-1.92755	2000-05-08 02:00:00+00:00	7.839000	83.33690	16.489416	53.880566	0.0	21.900002	1.0	0
13900	53.35261	-1.92755	2000-05-08 03:00:00+00:00	7.539000	85.05799	16.135872	51.340170	0.0	27.600000	1.0	0
13900	53.35261	-1.92755	2000-05-08 04:00:00+00:00	7.389000	86.83363	15.629971	51.546265	0.0	35.400000	1.0	0

Figura 3.13: Ventana de 4 datos antes de la no falla 1.

Dada la relevancia del inventariado en la clasificación de fallas, fue crucial establecer una metodología adecuada para etiquetar las horas en las que ocurrieron las fallas. Inicialmente, se asignó un valor de 1 a la hora exacta de la falla, mientras que las horas anteriores a la misma se etiquetaron con un valor de 0. Este enfoque permitió clasificar la falla en el momento preciso de su ocurrencia, proporcionando un punto de referencia para evaluar el desempeño del modelo.

No obstante, este proyecto tiene como objetivo la **detección temprana de fallas**, por lo que fue necesario ajustar la estrategia de etiquetado para anticipar la ocurrencia del evento con varias horas de antelación. Para ello, se modificó la asignación de etiquetas de la siguiente manera: se marcó con un valor de 1 en $t = -1$ (una hora antes de la falla) y, adicionalmente, se generó una **segunda versión** en la que la etiqueta se asignó en $t = -2$ (dos horas antes). En ambos casos, el objetivo es permitir que el modelo aprenda a reconocer patrones previos a la falla, facilitando su detección anticipada.

Asimismo, se incorporaron las horas anteriores a la etiqueta como contexto de entrada para el modelo, según el tamaño de ventana definido. Esta estrategia se ilustra en la Figura 3.14, donde se observa cómo se adelanta la etiqueta en el tiempo. Se recomienda prestar especial atención a la columna *date*, ya que en ella se evidencia claramente la progresión temporal. En total, se construyeron tres modelos, con el propósito de mejorar la capacidad de clasificación de fallas.

id	latitude	longitude	date	temperature_2m	relative_humidity_2m	wind_speed_10m	wind_direction_10m	precipitation	cloudcover	weather_code	salida
1	53.35261	-1.92755	2000-05-09 14:00:00+00:00	18.639000	65.903980	1.138420	251.56496	0.1	53.1	51.0	0
1	53.35261	-1.92755	2000-05-09 15:00:00+00:00	18.939000	66.619934	1.800000	270.00000	0.0	44.7	1.0	0
1	53.35261	-1.92755	2000-05-09 16:00:00+00:00	18.889000	67.709404	1.800000	306.86996	0.0	47.7	1.0	0
1	53.35261	-1.92755	2000-05-09 17:00:00+00:00	18.338999	71.002655	2.902413	330.25522	0.0	48.3	1.0	0
1	53.35261	-1.92755	2000-05-09 18:00:00+00:00	17.588999	75.164116	4.334974	355.23645	1.5	47.7	61.0	1

Etiquetado de la falla en t0 (momento exacto de la falla)

id	latitude	longitude	date	temperature_2m	relative_humidity_2m	wind_speed_10m	wind_direction_10m	precipitation	cloudcover	weather_code	salida
1	53.35261	-1.92755	2000-05-09 13:00:00+00:00	18.489000	66.744770	0.720000	180.00000	0.1	77.399994	51.0	0
1	53.35261	-1.92755	2000-05-09 14:00:00+00:00	18.639000	65.903980	1.138420	251.56496	0.1	53.100000	51.0	0
1	53.35261	-1.92755	2000-05-09 15:00:00+00:00	18.939000	66.619934	1.800000	270.00000	0.0	44.700000	1.0	0
1	53.35261	-1.92755	2000-05-09 16:00:00+00:00	18.889000	67.709404	1.800000	306.86996	0.0	47.700000	1.0	0
1	53.35261	-1.92755	2000-05-09 17:00:00+00:00	18.338999	71.002655	2.902413	330.25522	0.0	48.300000	1.0	1

Etiquetado de la falla en t-1 (una hora antes de la falla)

id	latitude	longitude	date	temperature_2m	relative_humidity_2m	wind_speed_10m	wind_direction_10m	precipitation	cloudcover	weather_code	salida
1	53.35261	-1.92755	2000-05-09 12:00:00+00:00	18.039	68.430490	1.08000	90.00000	0.0	89.700000	3.0	0
1	53.35261	-1.92755	2000-05-09 13:00:00+00:00	18.489	66.744770	0.72000	180.00000	0.1	77.399994	51.0	0
1	53.35261	-1.92755	2000-05-09 14:00:00+00:00	18.639	65.903980	1.13842	251.56496	0.1	53.100000	51.0	0
1	53.35261	-1.92755	2000-05-09 15:00:00+00:00	18.939	66.619934	1.80000	270.00000	0.0	44.700000	1.0	0
1	53.35261	-1.92755	2000-05-09 16:00:00+00:00	18.889	67.709404	1.80000	306.86996	0.0	47.700000	1.0	1

Etiquetado de la falla en t-2 (dos horas antes de la falla)

Figura 3.14: Etiquetado de fallas.

El primer modelo se enfocó en la pronosticación del momento exacto de la falla, utilizando un vector de entrada de dimensiones $(n, x, 7)$, donde n correspondía a la cantidad total de datos, tanto de momentos de falla como de no falla, x era el parámetro del inventariado con los datos asociados a la falla, y 7 representaba el número de características o variables meteorológicas consideradas. Este modelo se dedicó a la clasificación de la falla en su momento preciso. El segundo modelo buscó pronosticar la falla una hora antes, es decir, en el instante $t - 1$, y el tercer modelo se orientó a prever la falla con una antelación dos horas, en el instante $t - 2$, con las mismas configuraciones de los vectores de entrada. Para este proyecto, el valor de n resultó ser 27.798, con 13.899 momentos de falla y 13.899 momentos de no falla, con el objetivo de mantener un conjunto de datos equilibrado. Para una comprensión más detallada de las variaciones y el proceso de etiquetado, se realizó el proceso como el mostrado en la Figura 3.15.

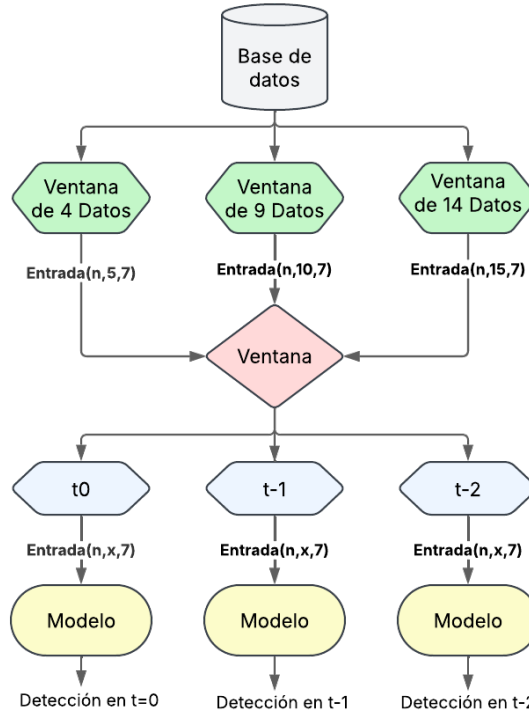


Figura 3.15: Diagrama de las variaciones de los datos a evaluar. Adaptado de [25]

3.8. Métrica para la Evaluación del Desempeño

La selección de una métrica de evaluación de desempeño para los modelos de clasificación es crucial, ya que permite comparar directamente los resultados obtenidos con otros modelos. Estas métricas buscan medir la desviación entre los datos reales y los pronosticados. Generalmente, la métrica empleada para evaluar el desempeño de los modelos esta vinculada a la problemática a resolver. Por lo tanto, es fundamental comprender el contexto del problema para seleccionar adecuadamente la métrica [25]. En el caso de los modelos de clasificación binaria, se sugieren siete métricas que se destacan por su capacidad para evaluar la eficacia del modelo. Según [40], para un clasificador binario son útiles técnicas como la matriz de confusión, el accuracy, el Recall, la precisión, el F1-score, la curva ROC y el AUC; además se menciona que las dos primeras medidas son muy eficientes en el caso de trabajar con datos balanceados; sin embargo, pierden confiabilidad al trabajar con datos desbalanceados.

La métrica de la matriz de confusión tiene como objetivo evaluar el desempeño del

modelo al indicar tanto las predicciones correctas como los errores. En esta matriz, las filas representan la clase predicha, mientras que las columnas corresponden a la clase real, mostrando el número de predicciones realizadas para cada categoría de manera mutuamente excluyente [40]. La Tabla 3.5 ilustra un caso de clasificación binaria, donde los verdaderos positivos y verdaderos negativos representaban las instancias correctamente clasificadas en sus respectivas clases. En contraste, los falsos positivos y falsos negativos corresponden a los casos en los que el modelo asignó incorrectamente una instancia a una categoría distinta de la real.

Tabla 3.5: Matriz de confusión para clasificación binaria. Adaptado de [40]

	Clase 1: Positivo	Clase 2: Negativo
Clase 1: Positivo	Verdadero positivo	Falso negativo
Clase 2: Negativo	Falso positivo	Verdadero negativo

Dado que este proyecto trabaja con clases balanceadas, la métrica principal a considerar fue el *Recall*. En la detección de fallas es fundamental anticiparlas, ya que el costo de clasificar erróneamente una falla como no falla es mucho mayor que el de predecir una falla cuando en realidad no lo es. Esta métrica, también conocida como *sensibilidad* o tasa de verdaderos positivos, mide la capacidad del modelo para identificar correctamente las instancias positivas; esta se calcula como la proporción de verdaderos positivos respecto a la suma de verdaderos positivos y falsos negativos. Un *Recall* alto indica que el modelo es eficiente en la detección de fallas, minimizando los falsos negativos [40, 41]. Por esta razón, es una métrica clave que debe maximizarse, asegurando al mismo tiempo que la precisión del modelo no se vea comprometida significativamente; la forma de calcular la sensibilidad se encontró en la Ecuación 3.1.

$$\text{Recall} = \frac{\text{V.P}}{\text{V.P} + \text{F.N}} \quad (3.1)$$

Sin embargo, además de la *Recall*, es conveniente incorporar otras métricas de desempeño que aseguran un equilibrio adecuado en la clasificación; en este sentido, se emplea el *Accuracy*, que representa la proporción de predicciones correctas sobre el total de eventos (Ecuación 3.2). También se utiliza el *F1-Score*, que constituye la

media armónica entre la *Recall* y el *Accuracy* (Ecuación 3.4), complementadas por la *Precision*, que evaluó la exactitud del modelo en la predicción de los eventos positivos (Ecuación 3.3) [40].

$$\text{Accuracy} = \frac{\text{V.P} + \text{V.N}}{\text{V.P} + \text{V.N} + \text{F.P} + \text{F.N}} \quad (3.2)$$

$$\text{Precision} = \frac{\text{V.P}}{\text{V.P} + \text{F.P}} \quad (3.3)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.4)$$

3.9. Resumen

En este capítulo se desarrolló la metodología para la limpieza, organización y procesamiento de los datos, con el objetivo de extraer variables meteorológicas relevantes que fueron utilizadas como características en el modelo. En primer lugar, se realizó un acercamiento inicial a la base de datos, organizando y depurando los registros para garantizar su calidad. Posteriormente, se llevó a cabo la segmentación de los datos con el fin de extraer las variables meteorológicas necesarias para alimentar el modelo. A continuación, se procedió a la integración y concatenación de los datos, consolidando una única base de datos que incluye las fechas y horas de las fallas, sus coordenadas geográficas y las variables meteorológicas asociadas. Como parte del análisis exploratorio, se efectuaron pruebas de correlación entre las variables meteorológicas, utilizando el coeficiente de Pearson, y se llevó a cabo un análisis de similitud entre las condiciones meteorológicas en eventos de falla y no falla. Con base en estos análisis, se avanzó en la definición de la arquitectura del modelo. Finalmente, se llevó a cabo la estructuración de los datos para su posterior procesamiento en el modelo de predicción. Este procedimiento contempló la normalización de las variables, la segmentación en ventanas temporales y la organización de los datos de manera que permitiera la detección de fallas tanto en el momento en que ocurren como en instantes previos. Además, se definieron las métricas de desempeño para un clasificador binario que se utilizarían para evaluar el rendimiento de los modelos.

Capítulo 4

Evaluación y Análisis de Resultados

Para el desarrollo de este proyecto en sus diferentes etapas, desde la obtención de variables meteorológicas hasta la organización de los datos, se utilizó Google Colaboratory como entorno de programación, una herramienta basada en la nube que permitió la ejecución de código en Python sin la necesidad de contar con un equipo de alto rendimiento a nivel local. Su funcionalidad basada en *notebooks* interactivos facilitó la integración de código, documentación y análisis de datos en un mismo entorno, lo que lo hizo especialmente útil en aplicaciones de inteligencia artificial y análisis de datos. Inicialmente, la organización y procesamiento de las variables meteorológicas se llevaron a cabo íntegramente en Google Colaboratory. Sin embargo, a medida que la complejidad del análisis aumentó, se identificaron limitaciones en el tiempo de ejecución y en el uso de recursos computacionales. Debido a estas restricciones, se tomó la decisión de migrar la ejecución del proyecto a un equipo de cómputo local con especificaciones detalladas en la Tabla 4.1.

Tabla 4.1: Especificaciones del equipo de cómputo utilizado.

Elemento	Especificación
Procesador	Intel Core i5 7200
RAM	16GB
Sistema Operativo	Windows 10
Almacenamiento	SSD 480

Adicionalmente, se utilizó Weights & Biases (WandB) ¹ como plataforma para optimizar y gestionar el entrenamiento de modelos en Python, esta herramienta permite realizar un seguimiento detallado de los experimentos, visualizar métricas,

¹ WandB, es un panel de control central que permite realizar un seguimiento de los hiperparámetros, las métricas del sistema y las predicciones para comparar modelos de ML en tiempo real y compartir los resultados. <https://wandb.ai/site>

comparar modelos y gestionar hiperparámetros de manera eficiente.

4.1. Metodología de Entrenamiento

En la Figura 4.1 se ilustra el esquema general de la metodología utilizada para el entrenamiento de los modelos, la cual sigue un enfoque basado en agentes. En este enfoque, **un agente representa una unidad de procesamiento que está vinculada a un coordinador**. La función del coordinador es distribuir los parámetros e hiperparámetros de entrenamiento a cada agente conectado. Cada agente es responsable de entrenar una cantidad específica de modelos y posteriormente enviarlos al coordinador, quien se encarga de su almacenamiento remoto. Para llevar a cabo el entrenamiento, los datos deben estar centralizados y ser accesibles para su consulta o descarga por parte de los agentes. Implementar una metodología de entrenamiento basada en agentes ofrece diversas ventajas, entre las cuales se destacan:

- Información y artefactos disponibles de forma centralizada, esto evita errores locales y la carga manual de datos.
- El coordinador tiene la capacidad de realizar diferentes barridos de parámetros e hiperparámetros para encontrar el mejor modelo.
- Computación distribuida, los agentes permiten tener diferentes máquinas entrenando modelos de forma paralela y reportando sus resultados al coordinador.

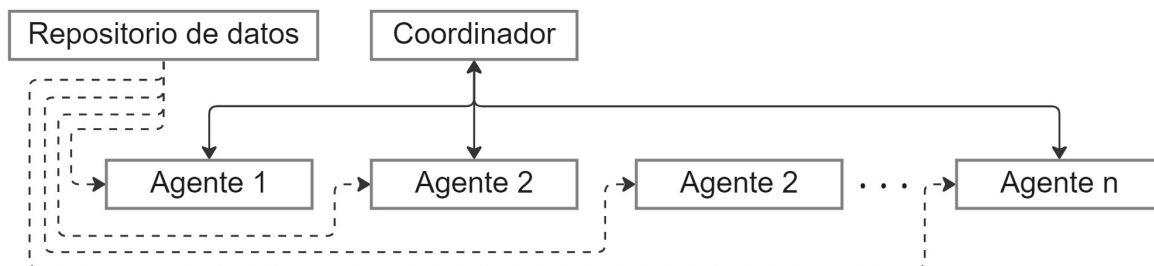


Figura 4.1: Esquema general de la metodología de entrenamiento. Tomado de [42]

La herramienta usada para configurar el coordinador y los agentes es WandB esta herramienta permite configurar diversos parámetros e hiperparámetros utilizando

enfoques aleatorios o adaptativos. El coordinador asigna un conjunto de parámetros a cada agente, el cual entrena los modelos con estos valores y posteriormente envía los artefactos generados junto con los registros del entrenamiento al coordinador.

En los diferentes *sweeps* realizados para cada modelo, se configuraron **variaciones en múltiples hiperparámetros** con el objetivo de optimizar el rendimiento de la red neuronal. Posteriormente, los agentes asociados a estos *sweeps* llevaron a cabo el proceso de entrenamiento. Las principales variaciones incluyeron **ajustes en el tamaño del lote de datos, la ventana de observación, el horizonte temporal** para la detección de fallas, la configuración de la **primera capa (bidireccional o unidireccional)**, el **método de normalización** de datos, la **tasa de aprendizaje**, el **número de unidades por capa** y la **tasa de dropout** aplicada después de la **capa densa**. El proceso de exploración de hiperparámetros se estructuró en dos etapas: inicialmente, se tomó un modelo base y se exploraron valores dentro de un rango amplio. Posteriormente, tras analizar la distribución de los parámetros e hiperparámetros obtenidos (ver Figura 4.2) y evaluar su convergencia, se acotaron los rangos de variación para los *sweeps*, permitiendo una optimización más fina del modelo.

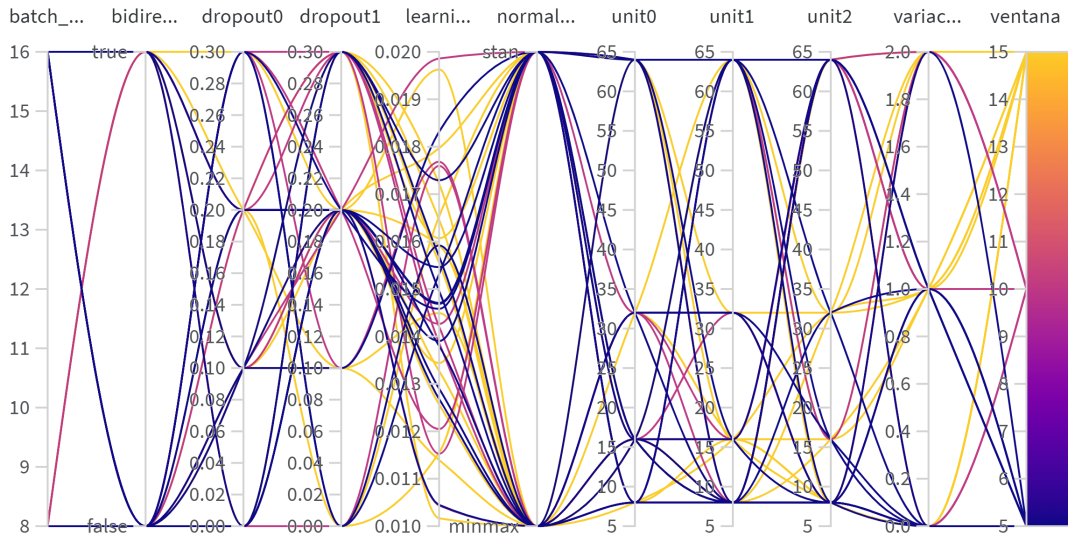


Figura 4.2: Distribución de parámetros e hiperparámetros.

La configuración de los *sweeps* para los modelos se definió de la siguiente forma.

```
sweep_config = {
  "metric":{"goal": "minimize","name": "val_loss"},
  "method": "bayes",
  "parameters": {
    "bidirectional":{"values": [True, False]},
    "variacion_t": {"values": [0,1,2]},
    "ventana": {"values": [5, 10, 15]},
    "normalizador": {"values": ["stan", "minmax"]},
    "batch_size": {"values": [8, 16]},
    "learning_rate": {"min": 0.01, "max": 0.02},
    "unit0": {"values": [8, 16, 32, 64]},
    "unit1": {"values": [8, 16, 32, 64]},
    "unit2": {"values": [8, 16, 32, 64]},
    "dropout": {"values": [0.0, 0.1, 0.2, 0.3]},
  },
}
```

El siguiente paso fue optimizar los modelos y sus hiperparámetros, comenzando por la selección de la ventana de observación, que define el intervalo previo a la falla utilizado para su clasificación. Para ello, se tomo como referencia $t = 0$ (momento de la falla) y se evaluó el desempeño de los modelos desplazando la etiqueta del evento una o dos horas antes, permitiendo analizar su precisión en distintos horizontes temporales de predicción.

4.2. Resultados y Análisis

En esta sección se presentan las características y estructuras de los modelos LSTM implementados, destacando aquellos que han obtenido un mejor rendimiento en la detección temprana de fallas en redes eléctricas. Se analizan los resultados obtenidos en las etapas de entrenamiento, validación y prueba, con el propósito de evaluar la capacidad de generalización de cada modelo. La Tabla 4.2 muestra la distribución del conjunto de datos en estas etapas, donde n representa la ventana de observación, la cual será definida posteriormente como un hiperparámetro clave en el diseño del modelo.

Tabla 4.2: Distribución de los datos en las etapas de entrenamiento, validación y prueba para cada escenario.

Modelo	Etapas	Datos de Entrada	Datos de Salida
t=0	Entrenamiento	(17.790, n, 7)	(17.790)
	Validación	(4.448, n, 7)	(4.448)
	Prueba	(5.560, n, 7)	(5.560)
t=-1	Entrenamiento	(17.790, n, 7)	(17.790)
	Validación	(4.448, n, 7)	(4.448)
	Prueba	(5.560, n, 7)	(5.560)
t=-2	Entrenamiento	(17.790, n, 7)	(17.790)
	Validación	(4.448, n, 7)	(4.448)
	Prueba	(5.560, n, 7)	(5.560)

4.2.1. Modelos para la Detección Temprana

4.2.1.1. Ventana de observación

El primer parámetro a definir en la configuración de los hiperparámetros fue la ventana de observación, que determina cuántas horas antes de la falla son relevantes para predecir la ocurrencia del evento. En el primer barrido de hiperparámetros, realizado en $t = 0$ (momento del evento), se identificó que el mejor modelo emplea una ventana de **15 horas**, es decir, considera las variables meteorológicas de las 14 horas previas y la hora en que ocurre la falla (ver Figura 4.3). Sin embargo, otros modelos con ventanas de **10 y 5 horas** también mostraron un desempeño competitivo. Dado que una ventana más corta implica una reducción en el costo computacional, se evaluó el impacto de reducir la ventana de 15 a 10 horas en términos de precisión y eficiencia. Para ello, **se aplicó una prueba ANOVA, comparando el mejor modelo de cada ventana en $t = 0$.**

El procedimiento consiste en encontrar el mejor modelo para cada ventana de observación y evaluarlos 10 veces, recopilando los resultados para analizarlos estadísticamente si existen diferencias significativas en su desempeño. Esto permitió

determinar si la reducción de la ventana de observación ofrece beneficios sin comprometer la precisión en la detección de fallas.

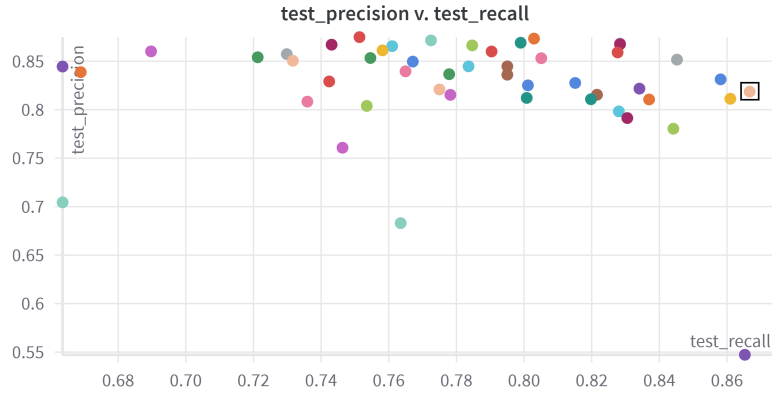


Figura 4.3: Entrenamiento de todos los modelos para $t=0$. El mejor modelo se marca con un recuadro.

Para evaluar si existen diferencias significativas entre los modelos entrenados con una ventana de 10 horas y 15 horas, se realizó un análisis de varianza (ANOVA), considerando la interpretación del p -value para determinar la hipótesis a aceptar. Las hipótesis establecidas fueron:

- **Hipótesis nula** (H_0): la media del residual es la misma en los diferentes modelos, los modelos son similares.
- **Hipótesis alternativa** (H_a): al menos dos medias difieren de forma significativa, algún modelo no es similar.

Para un p -value menor o igual al nivel de significancia de 0.05 ($p \leq 0.05$) se debe considerar la H_a y para un p -value mayor al nivel de significancia de 0.05 ($p > 0.05$) se debe considerar H_0 . La finalidad de entrenar los modelos con una ventana de menos datos, implica mejor costo computacional, en la evaluación para la métrica precisión, el p -value fue de 0.0651, lo que indica que no hay diferencias significativas entre los modelos de ventana 10 y 15. Dado que el p -value es mayor al umbral de 0.05, se acepta la hipótesis nula H_0 , lo que implica que **ambos modelos presentan un rendimiento estadísticamente similar en esta métrica.**

4.2.1.2. Modelos de barrido amplio

Una vez determinada la ventana de observación óptima, se procedió a realizar un barrido de hiperparámetros utilizando la técnica de optimización bayesiana en WandB. Este método permitió identificar progresivamente los mejores hiperparámetros basándose en los resultados de entrenamientos previos, optimizando la búsqueda de configuraciones más eficientes. Para guiar este proceso, se definió una métrica de evaluación junto con un objetivo. En este caso, al tratarse de un clasificador binario, se estableció el costo en la validación como métrica a minimizar.

Siguiendo esta estrategia, se entrenaron 20 modelos para cada uno de los escenarios: $t=0$ (momento exacto de la falla), $t=-1$ (una hora antes de la falla), $t=-2$ (dos horas antes de la falla). El propósito de esta exploración fue identificar hiperparámetros comunes entre los modelos, lo que permitió realizar un barrido más preciso en iteraciones posteriores. De esta manera, se logro refinar progresivamente la búsqueda del modelo óptimo para cada escenario, asegurando un rendimiento óptimo tanto en la detección de fallas en $t=0$, como en los pronósticos anticipados antes de que la falla ocurriese. Las Figuras 4.4 y 4.5 muestran los mejores modelos del barrido amplio y sus matrices de confusión, destacando la similitud en los hiperparámetros seleccionados.

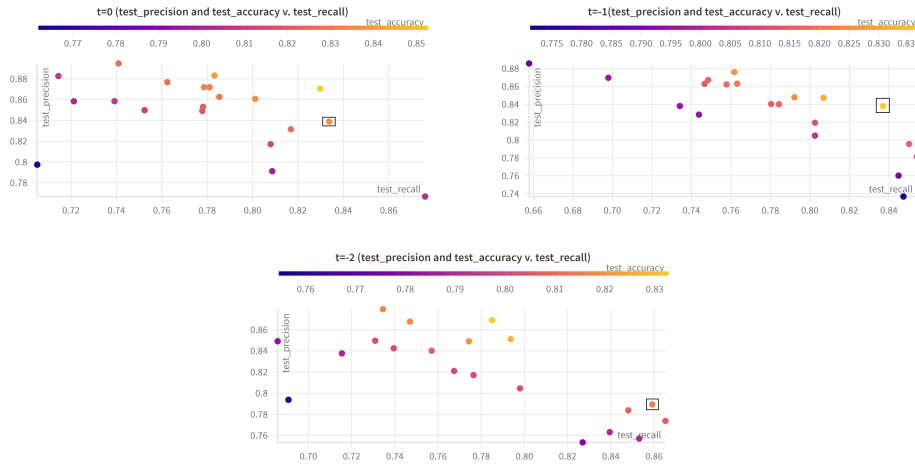


Figura 4.4: Mejores modelos en el barrido amplio para cada escenario (resaltados en un recuadro).

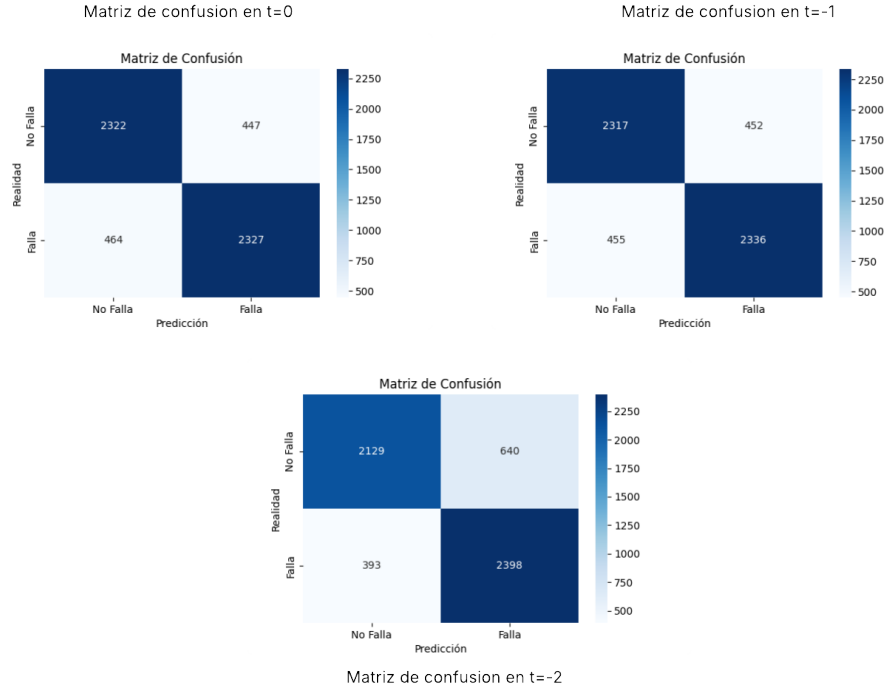


Figura 4.5: Matriz de confusión de los mejores modelos en la sintonización de parámetros amplios.

Como se muestra en la Figura 4.4, se identificaron los modelos con mejor desempeño para cada escenario tras un barrido amplio de hiperparámetros. A partir de estos resultados, se analizaron tres métricas que evidenciaron la convergencia de ciertos hiperparámetros entre los modelos óptimos, con el fin de refinar su selección y lograr un ajuste más preciso. En este análisis se observó que, en los tres escenarios evaluados, algunos hiperparámetros mostraron una tendencia consistente: el **batch size converge a 16**, la **primera capa LSTM se torna unidireccional**, el método de normalización más efectivo fue el **StandardScaler**, y el valor óptimo de **dropout** fue 0.2. Las métricas de desempeño obtenidas en el conjunto de prueba para los modelos seleccionados se presentan en la Tabla 4.3.

Tabla 4.3: Métricas de desempeño en los datos de prueba para cada escenario.

Escenario	Accuracy	F1-score	Loss	Precision	Recall
t=0	0.8362	0.8363	0.3813	0.8389	0.8338
t=-1	0.8369	0.8374	0.4000	0.8379	0.8370
t=-2	0.8142	0.8228	0.4177	0.7893	0.8592

4.2.1.3. Modelos con Hiperparámetros Sintonizados

Para la sintonización óptima de los hiperparámetros en cada escenario, se realizó un ajuste fino considerando únicamente aquellos hiperparámetros que, en la primera exploración amplia, no mostraron una tendencia clara en su impacto sobre el desempeño del modelo. Para ello, se continúa utilizando el método de optimización bayesiana mediante WandB, entrenando **50 modelos por escenario**. Este proceso garantizó la identificación del mejor modelo para cada caso, ya sea en la detección de fallas en el momento exacto del evento o en la detección temprana de posibles fallas. La selección de los modelos óptimos se basó en las mismas métricas de desempeño utilizadas en la Figura 4.4. En la Tabla 4.4 se presentan los hiperparámetros óptimos seleccionados para cada escenario.

Tabla 4.4: Hiperparámetros óptimos seleccionados para cada escenario.

Esc.	Batch	Bi-dir	Dropout	LR	Norm.	Unit0	Unit1	Unit2
t=0	16	No	0.2	0.01085	Stan	64	32	8
t=-1	16	No	0.2	0.01007	Stan	64	8	64
t=-2	16	No	0.2	0.01365	Stan	64	16	64

Con base en los hiperparámetros seleccionados, las Figuras (4.6, 4.7) muestran los resultados obtenidos para cada escenario.

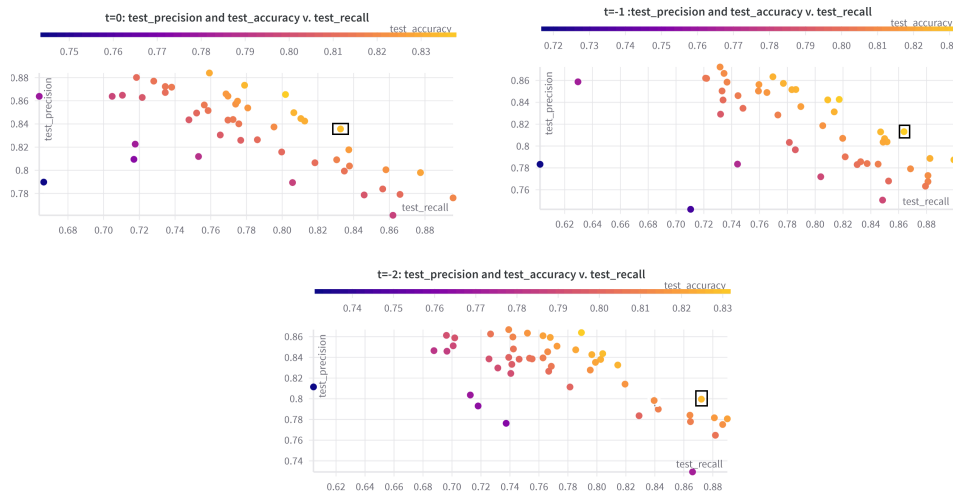


Figura 4.6: Mejores modelos en el barrido final para cada escenario (resaltados en un recuadro).

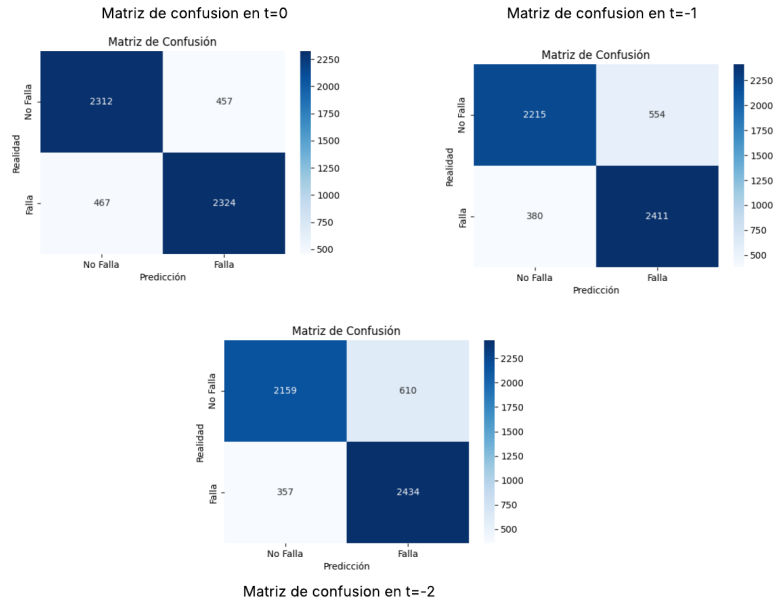


Figura 4.7: Matriz de confusión de los mejores modelos en la sintonización de parámetros final.

Para seleccionar el mejor modelo en cada escenario, ilustrado en la Figura 4.6, se priorizaron tres métricas clave en la evaluación de clasificadores binarios: *Recall*, *Accuracy* y *precision*. Esta decisión se basó en que enfocarse en una o dos métricas generaba desequilibrios significativos en el desempeño. Por ejemplo, un Recall alto podía acompañarse de valores bajos en las demás métricas, afectando la capacidad de generalización del modelo. Adicionalmente, se incorporó el área bajo la curva ROC, que permitió evaluar la capacidad del modelo para distinguir entre clases mediante la relación entre la tasa de verdaderos positivos y falsos positivos. Los resultados de estas métricas se presentan en la Tabla 4.5.

Tabla 4.5: Métricas de desempeño para los mejores modelos en los datos de prueba para cada escenario.

Escenario	Accuracy	F1-score	Loss	Precision	Recall	ROC
t=0	0.8338	0.8342	0.3873	0.8357	0.8327	0.92
t=-1	0.8320	0.8377	0.4003	0.8132	0.8638	0.91
t=-2	0.8261	0.8343	0.3997	0.7996	0.8721	0.91

En los modelos optimizados para cada escenario, se observaron mejoras

significativas en la detección temprana de fallas en comparación con los modelos iniciales. Sin embargo, el modelo encargado de la clasificación en el momento exacto de la falla ($t=0$) no presentó mejoras sustanciales, lo que indica que ya era robusto en su configuración original. Se destacó que el modelo es capaz de predecir una falla con hasta dos horas de anticipación ($t=-2$), lo que demuestra su efectividad en escenarios de detección temprana. No obstante, **al priorizar el *Recall*, se observó un incremento en la tasa de falsos positivos**, un comportamiento esperado al aumentar el horizonte de predicción. Otro aspecto relevante es que en los tres escenarios analizados, la métrica ROC se mantuvo en valores elevados en los modelos finales, lo que confirma su capacidad para distinguir de manera efectiva entre fallas y no fallas. Finalmente, al evaluar la función de pérdida (loss), se concluyó que **el modelo de detección en $t=0$ presenta la menor pérdida, lo que sugiere una mayor estabilidad y precisión en sus predicciones**.

Con el objetivo de evaluar la capacidad del modelo para anticipar eventos críticos, se seleccionó un caso representativo en el que la falla fue correctamente pronosticada. A partir de este caso, se construyó una ventana deslizante centrada en el instante de anticipación, definido como una hora antes del evento real. Se generaron un total de 11 muestras, cada una compuesta por las nueve horas previas más la hora correspondiente al instante que se desea clasificar. Estas muestras se distribuyeron temporalmente desde cinco horas antes hasta cinco horas después del punto de anticipación. Este enfoque permitió analizar cómo varía la respuesta del modelo frente a distintos escenarios temporales cercanos al evento crítico. Tal como se muestra en la Figura 4.8, se observa que, **a medida que las ventanas se aproximan al instante de anticipación, la probabilidad de falla predicha por el modelo aumenta de forma sostenida**. Este comportamiento genera un escalón pronunciado en la salida del modelo, el cual se mantiene, reflejando adecuadamente su capacidad de detección temprana.

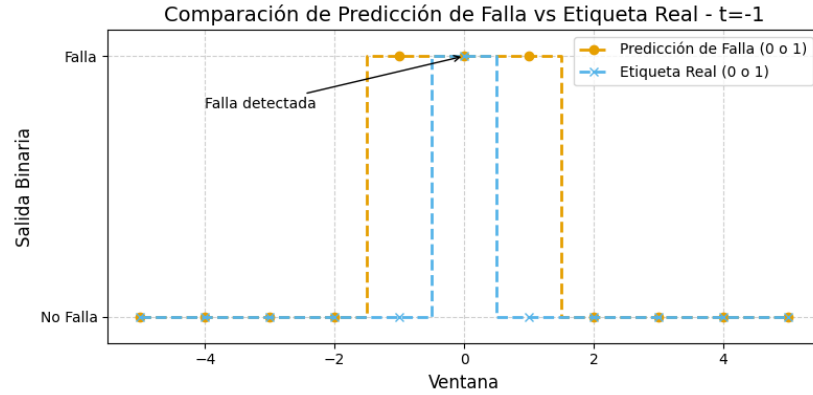


Figura 4.8: Evaluación con ventanas deslizantes en el modelo con anticipación $t=-1$ alrededor de una falla.

4.2.2. Comparación de los Modelos Propuestos y el Consultado en la Literatura

En la Tabla 4.6 se presentan los resultados comparativos entre los modelos LSTM propuestos para cada escenario y el modelo de referencia encontrado en la literatura [1]. Este último corresponde a un enfoque basado en redes neuronales bayesianas (BNN), acompañado de otros modelos utilizados para contrastar su desempeño. El estudio de referencia emplea una base de datos proveniente de una ciudad de china y consideró cinco variables meteorológicas para la clasificación de fallas en tres niveles: 0, 1 y 2. Es importante destacar que el modelo de referencia realiza predicciones de fallas con una periodicidad semanal, mientras que los modelos propuestos en este trabajo operan con datos en una resolución temporal distinta. A pesar de esta diferencia metodológica, se considera relevante la comparación, ya que permite evaluar el desempeño relativo de los enfoques utilizados.

Los resultados muestran que el modelo de referencia presenta un mejor desempeño en todas las métricas evaluadas. Sin embargo, los modelos basados en LSTM desarrollados en este estudio presentan valores competitivos y no se encuentran significativamente alejados de los obtenidos por la red neuronal bayesiana. Además, es importante destacar que el modelo de referencia emplea variables meteorológicas obtenidas de una estación meteorológica local, lo que garantiza **mediciones directas** y precisas en un punto geográfico específico. En contraste, el presente estudio utiliza datos meteorológicos extraídos de una API, los cuales provienen de modelos de

reanálisis y estaciones distribuidas, permitiendo una mayor cobertura espacial pero con posibles diferencias en la resolución y precisión de los datos.

Tabla 4.6: Métricas de desempeño de los modelos propuestos y el de la literatura.

Escenario	Accuracy	F1-score	Precision	Recall
t=0	0.8338	0.8342	0.8357	0.8327
t=-1	0.8320	0.8377	0.8132	0.8638
t=-2	0.8261	0.8343	0.7996	0.8721
Referencia [1]	0.929	0.865	0.843	0.889

4.3. Resumen

En este capítulo se presentan los resultados obtenidos a partir de los clasificadores desarrollados para la detección de fallas en redes eléctricas en tres escenarios: $t=0$ (momento del evento), $t=-1$ (una hora antes) y $t=-2$ (dos horas antes), utilizando una arquitectura *many-to-one* en una red LSTM y utilizando los datos analizados en el Capítulo 3. Inicialmente, se describe el entorno de desarrollo, incluyendo los programas, el lenguaje de programación y el hardware empleado (Tabla 4.1), seguido por la presentación de la plataforma de sintonización de hiperparámetros WandB (Figura 4.2) y los conjuntos de datos para entrenamiento, validación y prueba (Tabla 4.2). Se establece la ventana de observación como el hiperparámetro principal, cuya justificación se realiza mediante ANOVA (Figura 4.3), tras lo cual se lleva a cabo una sintonización amplia de hiperparámetros con el fin de identificar patrones en los modelos más efectivos para cada escenario (Figura 4.5), permitiendo detectar hiperparámetros recurrentes. Posteriormente, se realiza una segunda fase de ajuste más detallado, fijando los parámetros previamente identificados y logrando mejoras significativas, especialmente en los modelos de detección temprana. Finalmente, se comparó el desempeño de los mejores modelos obtenidos con un modelo reportado en la literatura para un problema similar en un contexto comparable, evidenciando que, aunque las métricas obtenidas no son idénticas (Tabla 4.6), los modelos propuestos presentan una estructura robusta y resultan altamente eficaces para la detección temprana de fallas en redes eléctrica usando variables meteorológicas.

Capítulo 5

Conclusiones

5.1. Conclusiones Generales

La revisión detallada de artículos, tesis, resoluciones y otros documentos técnicos relacionados con la problemática abordada facilitó la extracción de elementos clave para el desarrollo de modelos de clasificación destinados a la detección temprana de fallas en redes eléctricas, a partir del análisis de variables meteorológicas en un operador del Reino Unido. Este proceso permitió identificar métodos, estrategias para el manejo de bases de datos, consulta de variables meteorológicas y diversas metodologías, que contribuyeran significativamente para la construcción de clasificadores.

En el Capítulo 2, dedicado a la revisión de antecedentes que contribuyeron al desarrollo de este trabajo, se identificó que en Colombia existe una limitada investigación sobre la detección temprana de fallas eléctricas a partir de variables exógenas. Asimismo, se encontró que la combinación de datos de la red con variables meteorológicas podía mejorar la robustez de los modelos para la detección de fallas en redes eléctricas, lo que resultaba una **estrategia prometedora** para el pronóstico y la detección de fallas en redes eléctricas.

En el Capítulo 3, correspondiente al análisis exploratorio de los datos, se identificó un gran potencial en el uso de las APIs meteorológicas, ya que, en comparación con las bases de datos históricas, fue difícil encontrar variables meteorológicas con resoluciones horarias adecuadas para el estudio, por lo que uno de los principales parámetros en este trabajo fue la obtención de dichas variables a partir de la API de Open-Meteo. Además, se evidenció una correlación moderada entre algunas variables meteorológicas, encontrándose, de acuerdo con la Tabla 3.4, una correlación de 0.44 entre la cobertura nubosa y la humedad relativa, así como una correlación de 0.35 entre la precipitación

y la cobertura nubosa, tal como se muestra en la Figura 3.8. Finalmente, los hallazgos de este capítulo resaltan la importancia de la preparación y organización de los datos de entrada, así como la adecuada selección de la arquitectura del modelo, factores fundamentales para garantizar la obtención de resultados precisos y confiables en el estudio.

En el Capítulo 4, donde se presentan los resultados de los modelos desarrollados, se resalta la importancia y utilidad de herramientas como WandB, las cuales permiten monitorear el entrenamiento de los modelos en tiempo real y generar reportes detallados para cada ajuste de hiperparámetros. Además, **la optimización mediante la búsqueda de hiperparámetros permitió obtener modelos con mejor desempeño**, particularmente en la detección temprana de fallas. Asimismo, se observó que los modelos presentan robustez ante cambios en el escenario de predicción, dado que la variabilidad en el desempeño es mínima al detectar fallas con una o dos horas de anticipación. Sin embargo, a pesar de que las métricas de desempeño de los modelos son similares, el costo computacional varía significativamente. En particular, el entrenamiento de 20 modelos para la detección en $t = 0$ tomó aproximadamente 3 horas, mientras que para $t = -1$ el tiempo de entrenamiento se incrementó a 10 horas y para $t = -2$ fue de 6 horas.

5.2. Trabajos Futuros

- Incluir datos históricos propios del sistema eléctrico, como registros provenientes de sistemas de medición avanzada (Advanced Metering Infrastructure - AMI), con el objetivo de evaluar el impacto de las condiciones operativas de la red en la predicción de fallas. Este análisis permitirá comparar el desempeño del modelo cuando se trabaja exclusivamente con variables meteorológicas exógenas frente a un enfoque híbrido que combine información climática y datos del sistema eléctrico.
- Explorar diferentes configuraciones en la ventana temporal de observación de las variables meteorológicas, ampliando el rango de horas previas al evento de falla. Este enfoque permitirá analizar si el desempeño del modelo mejora o se ve afectado al considerar períodos de tiempo más largos. No obstante, es importante tener en

cuenta que el aumento en la cantidad de datos históricos utilizados conlleva un incremento en el costo computacional.

- Explorar el desempeño de diversas arquitecturas de aprendizaje automático y procesamiento de datos, incluyendo el uso de redes neuronales convolucionales para la extracción automática de características relevantes en datos meteorológicos temporales. Además, se podría considerar la implementación de estas arquitecturas en Colombia, con el objetivo de satisfacer la demanda de calidad de energía del país, cumpliendo con las normativas locales.
- Implementar datos meteorológicos en tiempo real en el sistema de detección temprana de fallas. Esta implementación permitiría monitorear continuamente la red eléctrica y generar alertas ante condiciones climáticas adversas, facilitando la toma de **decisiones oportunas** por parte de los operadores del sistema eléctrico.

Referencias

- [1] Y. Du, Y. Liu, X. Wang, J. Fang, G. Sheng, and X. Jiang, “Predicting weather-related failure risk in distribution systems using bayesian neural network,” *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 350–360, 2020.
- [2] J. E. C. González, D. V. O. Villavicencio, M. del Cisne Carrión González, A. F. V. León, J. C. C. Tinitana, and R. N. C. Pérez, *Fiabilidad de Redes Eléctricas*. Ediloja, 2024. [Online]. Available: <https://www.ediloja.com.ec>
- [3] Y. Zhou, A. Pahwa, and S.-S. Yang, “Modeling weather-related failures of overhead distribution lines,” *IEEE Transactions on power systems*, vol. 21, no. 4, pp. 1683–1690, 2006.
- [4] T. Alquthami, M. K. Alghamdi, B. S. Almajnuni, and O. M. Alarbidi, “Forecasting transmission system outages using artificial neural networks,” in *2021 22nd International Middle East Power Systems Conference (MEPCON)*. IEEE, 2021, pp. 654–658.
- [5] C. Central, “Aumento de los cortes de energía y cambio climático,” Climate Matters, 2023, <https://www.climatecentral.org/climate-matters/surging-weather-related-power-outages?lang=es>.
- [6] V. Sultan and B. Hilton, “A spatial analytics framework to investigate electric power-failure events and their causes,” *ISPRS International journal of geo-information*, vol. 9, no. 1, p. 54, 2020.
- [7] J. A. B. Contreras and E. G. S. Villegas, “Diseño de redes eléctricas en medio voltaje, bajo voltaje y alumbrado público para urbanizaciones,” *Polo del Conocimiento: Revista científico-profesional*, vol. 7, no. 8, pp. 1527–1547, 2022.
- [8] M. Mikati, M. Santos, and C. Armenta-Deu, “Modelado y simulación de un sistema conjunto de energía solar y eólica para analizar su dependencia de la red eléctrica,” *Revista Iberoamericana de Automática e Informática Industrial RIAI*, vol. 9, p. 267–281, 07 2012.

- [9] J. F. J. García, “Técnicas de predicción aplicada a la demanda desagregada con variables exógenas,” Proyecto Fin de Máster, Universidad de Sevilla, Escuela Técnica Superior de Ingeniería, Sevilla, España, 2023, máster en Sistemas de Energía Eléctrica.
- [10] R. Guo, G. Zhang, Q. Zhang, L. Zhou, H. Yu, and P. Yun, “Early fault detection of wind turbine gearbox based on adam-trained lstm,” in *2021 6th International Conference on Power and Renewable Energy (ICPRE)*. IEEE, 2021, pp. 285–289.
- [11] M. Atrigna, A. Buonanno, R. Carli, G. Cavone, P. Scarabaggio, M. Valenti, G. Graditi, and M. Dotoli, “A machine learning approach to fault prediction of power distribution grids under heatwaves,” *IEEE Transactions on Industry Applications*, vol. 59, no. 4, pp. 4835–4845, 2023.
- [12] W. Lu, Y. Li, Y. Cheng, D. Meng, B. Liang, and P. Zhou, “Early fault detection approach with deep architectures,” *IEEE Transactions on instrumentation and measurement*, vol. 67, no. 7, pp. 1679–1689, 2018.
- [13] R. M. R. Tituana, “Aplicación de algoritmos de aprendizaje automático en la predicción de fallas en redes eléctricas,” *INNOVA CIENTIFICA*, vol. 3, no. 1, 2024.
- [14] K. Pradeep, K. Gowrishankar, E. Sivanantham, K. S. Rao, N. Kiran, and A. Vimal, “Detection of early fault in power electronic converters through machine learning and data mining techniques,” in *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*. IEEE, 2023, pp. 1438–1442.
- [15] Y. Z. Gu, Z. R. Wu, J. G. Zhao, L. Q. Han, L. L. Yuan, and W. T. Huang, “Research on intelligent early warning algorithm for distribution network considering extreme climate conditions,” in *Proceedings of the 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2020, pp. 412–417.
- [16] L. Rouhiainen, “Inteligencia artificial,” *Madrid: Alienta Editorial*, pp. 20–21, 2018.
- [17] Á. F. R. Torres, F. S. R. Alvear, D. R. C. Lapo, and J. C. R. Alvear, “Diferencias y aplicaciones de big data, inteligencia artificial, machine learning y deep learning,” *Dominio de las Ciencias*, vol. 10, no. 3, pp. 960–982, 2024.

- [18] C. P. E. Yáñez and G. A. N. Cedillo, “Detección de pistas aéreas ilegales en imágenes digitales empleando técnicas de inteligencia artificial,” *Ciencia y Poder Aéreo*, vol. 18, no. 1, pp. 47–56, 2023.
- [19] P. J. Esquivel-Saenz, R. Ortiz-Gómez, M. Zavala, and R. S. Flowers-Cano, “Artificial neural networks for drought forecasting in the central region of the state of zacatecas, mexico,” *Climate*, vol. 12, no. 9, p. 131, Aug. 2024. [Online]. Available: <https://www.mdpi.com/2225-1154/12/9/131>
- [20] J. Vorobioff, S. Cerrotta, E. Morel, and A. Amadio, *Inteligencia Artificial y Redes Neuronales: Fundamentos, Ejercicios y Aplicaciones*. —, March 2022.
- [21] B. Martín del Brío and C. Serrano Cinca, “Fundamentos de redes neuronales artificiales: hardware y software,” *Scire: representación y organización del conocimiento*, no. June, pp. 103–125, 1995.
- [22] N. Khalili, S. R. Khodashenas, K. Davary, M. M. Baygi, and F. Karimaldini, “Prediction of rainfall using artificial neural networks for synoptic station of Mashhad: a case study,” *Arabian Journal of Geosciences*, vol. 9, no. 13, 2016. [Online]. Available: <http://dx.doi.org/10.1007/s12517-016-2633-1>
- [23] M. R. B. de Quiros, “Análisis de líneas de costa con redes neuronales lstm,” *Universitat Oberta de Catalunya (UOC)*, pp. 7–10, 2020, publicación académica interna.
- [24] Q. Sun, M. V. Jankovic, L. Bally, and S. G. Mougiakakou, “Predicting blood glucose with an lstm and bi-lstm based deep neural network,” in *2018 14th Symposium on Neural Networks and Applications (NEUREL)*, 2018, pp. 1–5.
- [25] J. David and R. Gomez, “Pronóstico de la demanda de energía eléctrica activa a largo plazo para un operador de red colombiano mediante técnicas de aprendizaje profundo,” Tesis de Pregrado, Universidad del Valle, Cali, Colombia, November 2023.
- [26] Colaboradores de Wikivoyage, “Noroeste de inglaterra,” Wikivoyage, March 2025, https://en.wikivoyage.org/wiki/North_West_England.
- [27] B. P. McClain, *Geospatial Data Science with Python*. Sebastopol, CA, USA: O’Reilly Media, 2022.

- [28] J. Wiecek, Q. Guo, and R. Hijmans, “The point-radius method for georeferencing locality descriptions and calculating associated uncertainty,” *International journal of geographical information science*, vol. 18, no. 8, pp. 745–767, 2004.
- [29] Met Office, “Historical weather and climate data,” Website, 2024, <https://www.metoffice.gov.uk/research/climate/maps-and-data/historic-station-data>.
- [30] P. Zippenfenig, “Open-meteo.com weather api,” Software, 2023, <https://open-meteo.com/>, doi: <https://doi.org/10.5281/zenodo.7970649>, Licencia: CC-BY-4.0.
- [31] H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. M. Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, and J.-N. Thépaut, “Era5 hourly data on single levels from 1940 to present,” Dataset, Copernicus Climate Data Store (CDS), 2023, <https://cds.climate.copernicus.eu/doi/10.24381/cds.adbb2d47>, doi: 10.24381/cds.adbb2d47.
- [32] isakimfv, “Como_usar_open_meteo.ipynb,” GitHub, n.d., https://github.com/isakimfv/TutorialesPython/blob/main/Como_usar_Open_Meteo.ipynb.
- [33] I. Roy-García, R. Rivas-Ruiz, M. Pérez-Rodríguez, and L. Palacios-Cruz, “Correlation: Not all correlation entails causality,” *Revista alergia Mexico (Tecamachalco, Puebla, Mexico: 1993)*, vol. 66, no. 3, p. 354, 2019.
- [34] M. L. Waskom, “seaborn: statistical data visualization,” *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021. [Online]. Available: <https://doi.org/10.21105/joss.03021>
- [35] E. López-Iñesta, F. Grimaldo, and M. Arevalillo-Herráez, “Classification similarity learning using feature-based and distance-based representations: A comparative study,” *Applied Artificial Intelligence*, vol. 29, pp. 445–458, 05 2015.
- [36] A. Karpathy, “The unreasonable effectiveness of recurrent neural networks,” Karpathy Blog, 2015, <https://karpathy.github.io/2015/05/21/rnn-effectiveness/>.
- [37] A. Thakur, “Lstm rnn in keras: Examples of one-to-many, many-to-one & many-to-many,” W&B Reports, 2023, <https://tinyurl.com/3mdxxc62>.

- [38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [39] J. S. Caicedo, “Pronóstico a corto plazo de la demanda de energía eléctrica activa para un operador de red colombiano,” Ph.D. dissertation, Universidad del Valle, Cali, Colombia, June 2023.
- [40] R. Borja-Robalino, A. Monleon-Getino, and J. Benedé, “Estandarización de métricas de rendimiento para clasificadores machine y deep learning,” in *Proceedings of an Academic Conference*, –, February 2020, pp.–.
- [41] M. Hossin and S. M.N, “A review on evaluation metrics for data classification evaluations,” *International Journal of Data Mining Knowledge Management Process*, vol. 5, pp. 01–11, 03 2015.
- [42] C. E. F. Álvarez, “Pronóstico del indice estandarizado de precipitación (spi) en la región de nariño utilizando técnicas de inteligencia artificial,” Master’s thesis, Universidad del Valle, Facultad de Ingeniería, Escuela de Ingeniería Eléctrica y Electrónica, Cali, Colombia, December 2022.