



Modelo de aprendizaje automático para la detección de anomalías en registros de temperatura de sensores meteorológicos in-situ en el Valle de Aburrá.

Ronald Antonio Vásquez Peña

Daissy Milenys Herrera Posada

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor

Julián David Arias Londoño, PhD.

Universidad de Antioquia

Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2026

Cita	(Vásquez Peña & Herrera Posada, 2026)
Referencia	Vásquez Peña, R., & Herrera Posada, D. (2026). <i>Modelo de aprendizaje automático para la detección de anomalías en registros de temperatura de sensores meteorológicos in-situ en el Valle de Aburrá</i> . [Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia.
Estilo APA 7 (2020)	



Grupo de Investigación Intelligent Information Systems Lab In2Lab

Especialización en Analítica y Ciencia de Datos, Cohorte IX.

Centro de Investigación Ambientales y de Ingeniería (CIA).



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: Héctor Iván García García.

Decano: Elkin Libardo Ríos Ortiz.

Jefe departamento: Danny Alejandro Múnera Ramírez

Coordinadora del Programa: Maria Bernarda Salazar Sánchez

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Dedicatoria

A nuestras familias, por su apoyo incondicional durante este proceso de formación académica y profesional.

Agradecimientos

Los autores expresan su gratitud al Profesor Julián David Arias Londoño, PhD., por su orientación académica y apoyo constante durante el desarrollo de esta investigación. Igualmente, se agradece al Sistema de Alerta Temprana del Valle de Aburrá (SIATA) por facilitar el acceso a los datos de la red de sensores meteorológicos, insumo fundamental para la realización de este trabajo. Se reconoce también el aporte de la comunidad científica de código abierto, cuyos desarrollos en arquitecturas de aprendizaje profundo fueron referencia clave para el diseño metodológico adoptado.

Tabla de contenido

1. Introducción	9
2. Materiales y Métodos	10
2.1 Descripción de los datos	10
2.2 Preprocesamiento y limpieza de datos	13
2.3 Métodos de análisis y modelado	13
2.3.1 Auto-codificadores recurrentes	14
2.3.2 USAD – UnSupervised Anomaly Detection con aprendizaje por transferencia	15
2.3.3 Anomaly-Transformer con aprendizaje por transferencia	16
2.4 Evaluación de los modelos	16
2.5 Diseño experimental	17
2.5.1 Auto-codificadores recurrentes	17
2.5.2 USAD	18
2.5.3 Anomaly-Transformer con aprendizaje por transferencia	19
3. Resultados y Discusión	20
3.1 Modelos	20
3.1.1 Autoencoder recurrente	20
3.1.2 USAD con aprendizaje por transferencia	25
3.1.3 Anomaly-Transformer con aprendizaje por transferencia	27
3.2 Análisis e interpretación de resultados	29
3.3 Limitaciones	30
4. Conclusiones	30

Lista de Figuras

Figura 1. Ubicación de las estaciones meteorológicas empleadas.....	11
Figura 2. Serie temporal de temperatura de la estación 203, donde se resaltan los registros normales (azul) y anómalos (rojo).....	12
Figura 3. A) Ciclo diurno y B) Autocorrelación de la temperatura para la estación 203, estimados utilizando los datos evaluados.....	12
Figura 4. Arquitectura del autoencoder recurrente.....	15
Figura 5. Arquitectura USAD: codificador compartido (E), decodificador D_1 (AE1) y decodificador D_2 (AE2) con esquema de entrenamiento adversarial. Adaptado de Audibert et al. (2020).....	16
Figura 6. Curva de entrenamiento para los auto-codificadores recurrentes: A) Simple RNN. B) LSTM y C) GRU.....	18
Figura 7. Curvas de pérdida (entrenamiento y validación) del modelo USAD durante el ajuste fino.....	19
Figura 8. Curvas de pérdida durante el preentrenamiento del modelo Anomaly-Transformer.....	20
Figura 9. Curvas de pérdida durante el ajuste fino del modelo Anomaly-Transformer.....	20
Figura 10. A) Curva ROC y criterio de Youden J para la selección del umbral $\theta : 0,2965$, empleando los datos de validación. B) Error en la reconstrucción. C) Curva ROC para los datos de prueba. D) Comparación serie original y reconstruida — Simple RNN.....	22
Figura 11. A) Curva ROC y criterio de Youden J para la selección del umbral $\theta : 0,0835$, empleando los datos de validación. B) Error en la reconstrucción. C) Curva ROC para los datos de prueba. D) Comparación serie original y reconstruida — LSTM.....	23
Figura 12. A) Curva ROC y criterio de Youden J para la selección del umbral $\theta : 0,0920$, empleando los datos de validación. B) Error en la reconstrucción. C) Curva ROC para los datos de prueba. D) Comparación serie original y reconstruida — GRU.....	24
Figura 13. Matriz de confusión normalizada para los modelos auto-decodificadores recurrentes en el conjunto de prueba A) Simple RNN, B) LSTM y C) GRU.....	24
Figura 14. A) Curva ROC (AUC = 0,5580) B) criterio de Youden J para la selección del umbral $\theta = 0,308279$ — USAD — validación.....	25
Figura 15. Error de reconstrucción USAD.....	25
Figura 16. Curva ROC sobre el conjunto de prueba (AUC = 0,5826) — USAD.....	26
Figura 17. Señal reconstruida USAD vs. señal original.....	26
Figura 18. Matriz de confusión normalizada USAD.....	27
Figura 19. A) Curva ROC (AUC = 0,6944) B) criterio de Youden J para la selección del umbral $\theta = -0,000121$ — Anomaly-Transformer.....	28
Figura 20. Error de reconstrucción Anomaly-Transformer.....	28
Figura 21. Curva ROC sobre el conjunto de prueba (AUC = 0,6044) — Anomaly-Transformer. El umbral θ fue definido exclusivamente en validación.....	28
Figura 22. Señal reconstruida Anomaly-Transformer vs. señal original.....	29
Figura 23. Matriz de confusión normalizada Anomaly-Transformer	29

Lista de Tablas

Tabla 1. Estadísticas básicas para los datos de los sensores 203 en el periodo de estudio...	12
Tabla 2. Mejores hiper-parámetros encontrados para los auto-codificadores recurrentes.	17
Tabla 3. Resultados comparativos de los modelos auto-codificadores recurrentes en el conjunto de prueba.....	24

Siglas, acrónimos y abreviaturas

APA	American Psychological Association
AELM	Autoencoder LSTM Multivariante
AUC	Area Under the Curve (Área bajo la curva)
BA	Balance Accuracy
AUC	Area under the curve
EDA	Exploratory Data Analysis (Análisis Exploratorio de Datos)
F1	Medida F1 (media armónica de precisión y recall)
GRU	Gated Recurrent Unit (Unidad Recurrente con Compuertas)
ICLR	International Conference on Learning Representations
IJCAI	International Joint Conference on Artificial Intelligence
KDD	Knowledge Discovery in Databases
LSTM	Long Short-Term Memory (Memoria de Largo y Corto Plazo)
ML	Machine Learning (Aprendizaje Automático)
MSE	Mean Squared Error (Error Cuadrático Medio)
NAB	Numenta Anomaly Benchmark
ROC	Receiver Operating Characteristic
RNN	Recurrent Neural Network (Red Neuronal Recurrente)
SIATA	Sistema de Alerta Temprana del Valle de Aburrá
TL	Transfer Learning (Aprendizaje por Transferencia)
UdeA	Universidad de Antioquia
USAD	UnSupervised Anomaly Detection

Resumen

El monitoreo continuo de variables atmosféricas en redes de sensores meteorológicos enfrenta el desafío persistente de registrar datos erróneos o atípicos, derivados de fallas técnicas, interferencias ambientales o degradación de los equipos, lo que compromete la calidad de la información que sustenta la toma de decisiones. En este contexto, se evalúan modelos de aprendizaje automático no supervisado para la detección automática de anomalías en registros de temperatura de sensores in situ en el Valle de Aburrá, empleando los datos de la estación 203 (Universidad Nacional de Colombia) de la red SIATA para el período enero–junio de 2023. Se implementaron y compararon tres enfoques: (1) USAD (Unsupervised Anomaly Detection) con aprendizaje por transferencia de dominio; (2) Anomaly-Transformer con entrenamiento progresivo en dos fases; y (3) ensambles de auto-codificadores recurrentes (RNN, LSTM y GRU). La detección de las anomalías se realiza a partir del error de reconstrucción y el umbral definido mediante el criterio de Youden J sobre el conjunto de validación. Para los datos de prueba, el experimento USAD obtuvo $F1 = 0,3015$, el Anomaly-Transformer-TL alcanzó $F1 = 0,4298$ y de los auto-codificadores recurrentes, se tiene que RNN simple posee un $F1 = 0,5768$, GRU alcanzó un $F1 = 0,5621$ y LSTM obtiene $F1 = 0,6120$. Donde la arquitectura LSTM logra el mejor desempeño global entre los modelos comparados.

Palabras clave: *aprendizaje por transferencia, auto-codificadores recurrentes, detección de anomalías, sensores meteorológicos, series de tiempo.*

1. Introducción

El Valle de Aburrá concentra aproximadamente 3,9 millones de habitantes (Área Metropolitana del Valle de Aburrá [AMVA], 2019) y se caracteriza por una topografía compleja que favorece la acumulación de contaminantes atmosféricos y la ocurrencia de eventos hidrometeorológicos extremos. Para hacer frente a estas condiciones, el Sistema de Alerta Temprana de Medellín y el Valle de Aburrá (SIATA) opera una extensa red de sensores in situ que registran, entre otras variables, la temperatura del aire con una resolución temporal de un minuto. La disponibilidad de estos datos en tiempo real es fundamental para la calibración de modelos numéricos de predicción atmosférica, la generación de alertas por precipitaciones intensas y el seguimiento de los indicadores de la calidad del aire (Alcaldía de Medellín & AMVA, 2020).

La operación continua de sensores en ambientes expuestos puede introducir registros anómalos, los cuales comprenden valores físicamente improbables, variaciones abruptas y lecturas constantes por congelamiento de la señal. Si estas anomalías no se detectan y corrigen oportunamente, el error se propaga hacia otros análisis. Aunque existen métodos clásicos no supervisados ampliamente documentados —como el factor de anomalía local (Breunig et al., 2000) y el bosque de aislamiento (Liu et al., 2008)—, su aplicación directa a series de alta densidad temporal presenta limitaciones ante las dependencias temporales y el volumen de datos (Chandola et al., 2009).

Como respuesta a este desafío, el aprendizaje automático no supervisado (*unsupervised machine learning*) se posiciona como una alternativa viable para la detección automática de anomalías en series temporales, dado que omite la necesidad de contar con etiquetas manuales durante la fase de entrenamiento de los modelos (Audibert et al., 2020; Kieu et al., 2019; Xu et al., 2022). Específicamente, los enfoques fundamentados en auto-codificadores recurrentes (*recurrent autoencoders*) —con antecedentes relevantes en el trabajo de Malhotra et al. (2016) para series temporales— aprovechan la capacidad de las redes neuronales recurrentes para modelar dependencias temporales. Estos modelos emplean el error de reconstrucción como indicador de anomalía: los patrones normales se reconstruyen con un error mínimo, mientras que las observaciones anómalas, al no estar representadas en los datos de entrenamiento, generan errores de reconstrucción significativamente mayores. Un enfoque complementario lo constituyen las arquitecturas basadas en Transformers (Xu et al., 2022), las cuales reemplazan el mecanismo de reconstrucción por el análisis de la distribución de asociación de la autoatención: su hipótesis central plantea que los puntos normales establecen asociaciones de largo alcance en la serie, mientras que los puntos anómalos, dada su rareza, quedan restringidos a conexiones locales, lo que permite discriminarlos a través de la discrepancia entre ambas distribuciones de atención.

El objetivo general de la presente investigación consiste en desarrollar y evaluar modelos de aprendizaje automático para la detección de anomalías en los registros de temperatura provenientes de sensores meteorológicos in situ en el Valle de Aburrá,

mediante el uso de arquitecturas de auto-codificadores recurrentes, estrategias de aprendizaje por transferencia (transfer learning) y arquitecturas basadas en Transformers. Para alcanzar este propósito, se implementaron tres enfoques complementarios sobre los datos del sensor 203 (enero–junio 2023), el modelo USAD (*Unsupervised Anomaly Detection*) con aprendizaje por transferencia de dominio desde el conjunto SWaT (51 sensores industriales); la arquitectura Anomaly-Transformer con entrenamiento progresivo en dos fases; y auto-codificadores recurrentes en tres variantes arquitectónicas —RNN simple, LSTM y GRU—.

Para la identificación de las anomalías, se define un umbral de decisión que permite clasificar la serie en función del error de reconstrucción, obtenido por los modelos. El criterio de selección del umbral θ , se obtiene maximizando el índice de Youden J en el conjunto de validación. Empleando las anomalías predichas y las reales, se cuantificó el rendimiento de los modelos a través de las siguientes métricas: exactitud balanceada (*balanced accuracy*), precisión (*precision*), exhaustividad (*recall*) y el valor F1 (*F1-score*), complementadas con el análisis de matrices de confusión.

El presente documento se estructura de la siguiente manera: la sección de Materiales y Métodos describe la procedencia de los datos, las técnicas de preprocesamiento, la configuración arquitectónica de los modelos y las métricas de evaluación; posteriormente, la sección de Resultados y Discusión expone los hallazgos derivados de la implementación experimental de los enfoques; finalmente, la sección de Conclusiones sintetiza los resultados más relevantes y plantea posibles líneas de trabajo futuro.

2. Materiales y Métodos

2.1 Descripción de los datos

Los datos empleados en el presente estudio proceden de la red de sensores meteorológicos del SIATA. Específicamente, se utilizaron registros de temperatura del aire del sensor 203 ubicado en la Universidad Nacional de Colombia (Figura 1), para el período enero–junio de 2023 con una resolución temporal de un minuto.

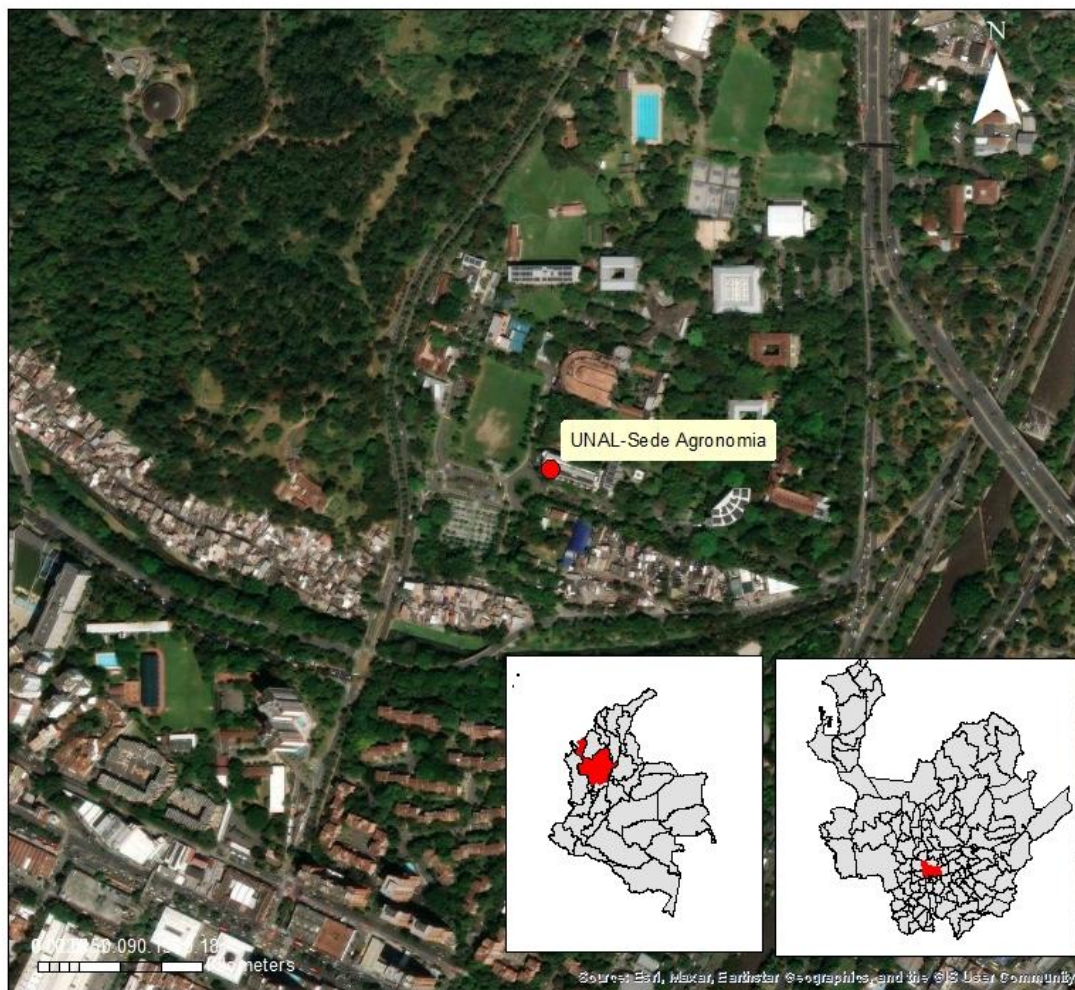


Figura 1. Ubicación de las estaciones meteorológicas empleadas.

La estación y periodo seleccionado se debe a que, de la cantidad total de datos para la estación 203 en el periodo de enero-junio del 2023, el 1,7% son anomalías y estas corresponden principalmente a tres categorías: (1) valores físicamente improbables (temperaturas fuera del rango operativo del sensor); (2) variaciones abruptas no atribuibles al régimen meteorológico local y (3) lecturas constantes por congelamiento de la señal durante periodos prolongados. La Figura 2 presenta las anomalías identificadas en las series de temperatura.

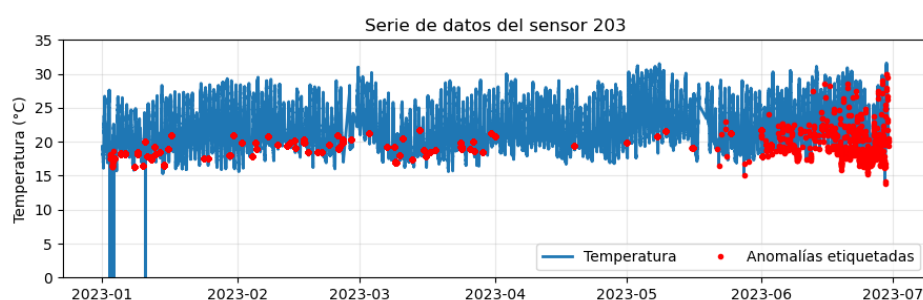


Figura 2. Serie temporal de temperatura de la estación 203, donde se resaltan los registros normales (azul) y anómalos (rojo).

El análisis exploratorio permitió caracterizar el comportamiento estadístico y temporal de la serie. La temperatura en el conjunto de datos cuenta con media de 21,83 °C y desviación estándar de 3,28 °C (Tabla 1). En la estación se detectaron patrones periódicos diurnos marcados, con temperaturas mínimas registradas entre las 05:00 y las 07:00 horas, y máximas entre las 13:00 y las 15:00 horas (Figura 3A). La función de autocorrelación (Figura 3B), muestra una correlación elevada para retardos cortos que decrece de forma gradual al aumentar el desfase temporal, lo que evidencia una marcada persistencia temporal. En conjunto, el ciclo diurno y la función de autocorrelación confirman la presencia de un patrón periódico sostenido y dependencia temporal, modulado principalmente por los procesos radiativos diarios que determinan la dinámica térmica local.

Tabla 1. Estadísticas básicas para los datos de los sensores 203 en el periodo de estudio.

	Sensor 203
Cantidad	244.428
Promedio	21,83
Desviación estándar	3,28
Mínimo	15,3
25%	19,3
50%	21,2
75%	24,2
Máximo	31,6

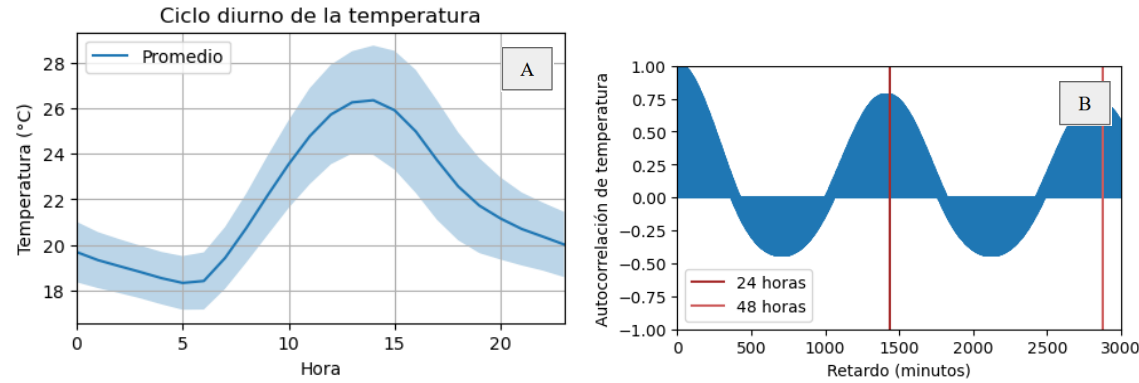


Figura 3. A) Ciclo diurno y B) Autocorrelación de la temperatura para la estación 203, estimados utilizando los datos evaluados.

2.2 Preprocesamiento y limpieza de datos

El acondicionamiento de los datos se estructuró en cuatro etapas:

1. Partición secuencial. Los datos se dividieron preservando el orden temporal de la serie: el 70 % inicial conformó el conjunto de entrenamiento, el 15 % siguiente el de validación y el 15 % final el de prueba.
2. Normalización z-score. Los tres modelos —auto-codificadores recurrentes, USAD y Anomaly-Transformer— emplearon estandarización z-score. La media y la desviación estándar se calcularon exclusivamente sobre las observaciones válidas del conjunto de entrenamiento y se aplicaron a los conjuntos de validación y prueba.
3. Tratamiento de anomalías en el entrenamiento. El tratamiento de las observaciones anómalas estuvo condicionado por las características estructurales y las posibilidades de implementación propias de cada arquitectura. En el caso del autoencoder recurrente, al tratarse de una arquitectura desarrollada específicamente para este estudio, fue posible incorporar un mecanismo de enmascaramiento directo mediante el uso del valor centinela -2000 , evitando que los registros previamente identificados como anómalos contribuyeran al proceso de reconstrucción. En contraste, USAD y Anomaly-Transformer corresponden a arquitecturas previamente definidas, cuya formulación y estructura de entrenamiento no contemplan de manera directa la aplicación de máscaras sobre las entradas. Por esta razón, en USAD se adopta una estrategia de exclusión a nivel de ventana, eliminando del *DataLoader* aquellas ventanas que contenían al menos una observación anómala mediante enmascaramiento por índice. Por su parte, Anomaly-Transformer no requirió una exclusión explícita de dichas observaciones, dado que su mecanismo de discrepancia de asociación está diseñado para diferenciar patrones normales y anómalos durante el propio proceso de optimización.
4. Segmentación por ventanas deslizantes. La serie se divide en ventanas de longitud fija con desplazamiento (*stride*) constante, se emplea ventanas de $C = 30$ pasos con *stride* = 10. La selección de cada tamaño se realizó considerando la estructura temporal de la serie: el análisis de correlación (Figura 3B) evidenció codependencia temporal significativa, mientras que el ciclo diurno (Figura 3A) sugiere variaciones relevantes en períodos de corta duración. Ventanas extensas favorecen una reconstrucción global pero atenúan la amplitud local de la señal —reduciendo la sensibilidad ante anomalías puntuales—; ventanas reducidas incrementan la resolución temporal a costa de mayor costo computacional.

2.3 Métodos de análisis y modelado

Se implementaron y evaluaron tres enfoques algorítmicos para la detección de anomalías, orientados a explorar distintos paradigmas dentro de la arquitectura de auto-codificadores:

2.3.1 Auto-codificadores recurrentes

Basado en el trabajo de Kieu et al. (2019), se evaluó una variante del enfoque S-RNN. Se implementó un modelo univariado de autoencoder recurrente basado en tres arquitecturas base: RNN simple, LSTM y GRU. El *encoder-decoder* se orienta a la reconstrucción de secuencias temporales. El objetivo de los modelos es aprender una representación latente de los patrones normales de la serie y utilizar el error de reconstrucción como indicador de anomalía.

La elección de arquitecturas recurrentes se sustenta en la fuerte dependencia temporal presente en las series de temperatura. Dichas arquitecturas son capaces de modelar dependencias secuenciales y representar patrones temporales complejos. En este contexto, se evalúa el impacto de las capacidades de memoria a largo plazo de la LSTM (Hochreiter & Schmidhuber, 1997) y la GRU (Chung et al., 2014), en contraste con la memoria más limitada de la RNN simple, con el fin de analizar su desempeño en la detección de anomalías en series de tiempo.

En la Figura 4, se muestra el esquema de la arquitectura implementada, donde el *encoder* (Bloque azul) recibe como entrada secuencias univariadas. Previamente, se aplica una capa de enmascaramiento (*Masking*) que ignora los valores correspondientes al valor (-2000), evitando que datos previamente identificados como anómalos influyan en el aprendizaje del modelo. Posteriormente, una capa recurrente procesa la secuencia y condensa la información temporal en un único vector de estado oculto, que representa la codificación latente de la secuencia de entrada. El *decoder* (Bloque verde) toma este vector latente (h_c) y lo replica C veces mediante una capa *RepeatVector*, generando una secuencia de entrada para la etapa de reconstrucción. Esta secuencia es procesada por una segunda capa recurrente con salida secuencial, permitiendo reconstruir la dinámica temporal original. Finalmente, una capa *TimeDistributed Dense* proyecta cada paso temporal a la dimensión de salida, produciendo la secuencia reconstruida.

Durante el entrenamiento, el modelo se optimiza utilizando el algoritmo Adam, minimizando una función de pérdida basada en el error cuadrático medio enmascarado (*masked MSE*), la cual excluye del cálculo aquellos valores asociados al valor -2000. De esta manera, el modelo aprende únicamente a partir de observaciones consideradas válidas.

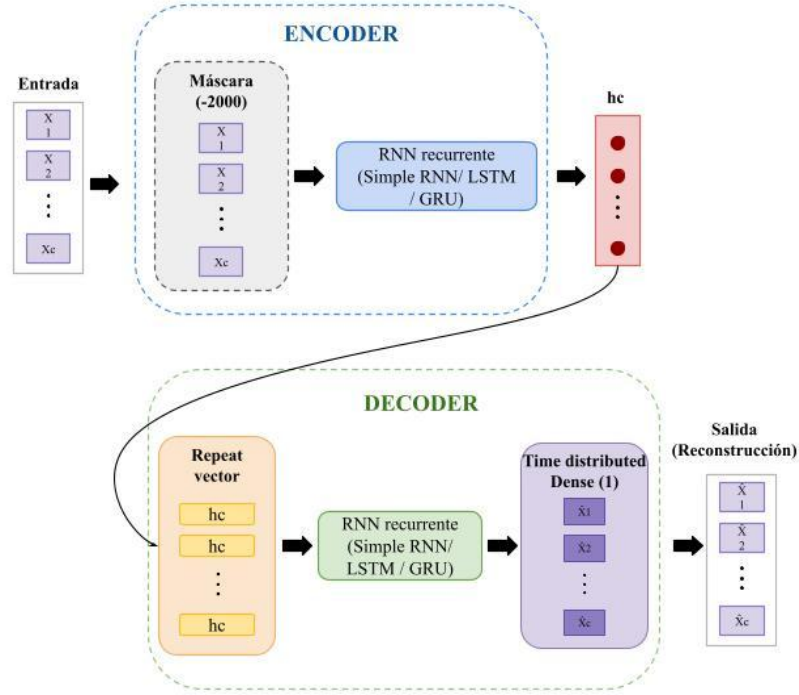


Figura 4. Arquitectura del autoencoder recurrente.

2.3.2 USAD – UnSupervised Anomaly Detection con aprendizaje por transferencia

El modelo USAD (Audibert et al., 2020) es un auto-codificador no supervisado diseñado para series temporales multivariantes. Su arquitectura (Figura 5) acopla un auto-codificador inicial (AE1, bloques naranjas) y un segundo auto-codificador (AE2, bloques morados). AE1 reconstruye la entrada original a partir de la representación latente generada por el codificador compartido (E, bloque verde), mientras que AE2 reconstruye la salida generada por AE1. Esta configuración establece una cadena de reconstrucción diseñada para amplificar el error frente a instancias anómalas. La puntuación de anomalía (anomaly score) se calcula según la Ecuación 1:

$$\text{Score}(x) = (1-\alpha) \cdot \|x - \text{AE1}(x)\|^2 + \alpha \cdot \|x - \text{AE2}(\text{AE1}(x))\|^2 \quad (1)$$

Donde $\alpha \in [0, 1]$ es un hiperparámetro que regula el compromiso entre la sensibilidad y la especificidad. Dado que el entrenamiento es no supervisado, el parámetro α no puede calibrarse durante la fase de entrenamiento. Por esta razón, α se optimizó empíricamente sobre el conjunto de validación (con las etiquetas disponibles), evaluando el índice de Youden J para cada valor candidato de α . Un valor de α cercano a 1 prioriza la detección exhaustiva de anomalías (incrementando la exhaustividad o recall), mientras que un valor próximo a 0 minimiza la tasa de falsos positivos (mejorando la precisión).

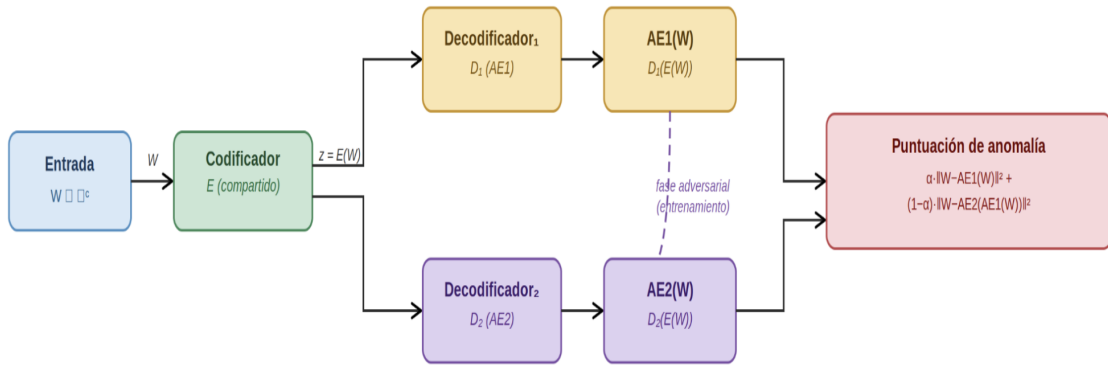


Figura 5. Arquitectura USAD: codificador compartido (E), decodificador D_1 (AE1) y decodificador D_2 (AE2) con esquema de entrenamiento adversarial. Adaptado de Audibert et al. (2020).

2.3.3 Anomaly-Transformer con aprendizaje por transferencia

La arquitectura Anomaly-Transformer (Xu et al., 2022) introduce un enfoque basado en la distribución de asociación de los mecanismos de atención para la detección de anomalías en series temporales. Su hipótesis central postula que las observaciones anómalas, dada su baja frecuencia de ocurrencia, carecen de la capacidad para establecer asociaciones de largo alcance con otros puntos de la serie temporal. En consecuencia, sus distribuciones de atención se encuentran dominadas por conexiones locales (prior-association). Por el contrario, las observaciones normales desarrollan patrones de atención con dependencias a largo alcance que capturan las periodicidades y las tendencias globales subyacentes en la serie, lo que le confiere una mayor eficiencia matemática para capturar dependencias de largo alcance en series temporales extensas.

El modelo formaliza esta divergencia mediante el mecanismo de *Anomaly Attention*, el cual computa simultáneamente la asociación a priori (*prior-association*) —una distribución gaussiana centrada en cada instante de tiempo que modela la propensión local de las anomalías— y la asociación de la serie (*series-association*), correspondiente a la distribución de atención aprendida empíricamente a partir de los datos. La puntuación de anomalía se calcula maximizando la divergencia de Kullback-Leibler (KL) entre ambas distribuciones mediante un esquema de optimización Minimax. De este modo, las anomalías exhiben la mayor discrepancia matemática entre sus distribuciones prior y series-association (distribución de asociación de la serie aprendida empíricamente).

2.4 Evaluación de los modelos

Los modelos evaluados corresponden a enfoques no supervisados, en los cuales la detección de anomalías se realiza a partir del error de reconstrucción. Empleando los datos de validación, se define un umbral de decisión (θ) que permite discriminar entre observaciones normales y anómalas. Las anomalías identificadas a partir de este procedimiento se comparan posteriormente con las etiquetas de datos anómalos

conocidas, lo que permite el cálculo de métricas de evaluación.

El criterio de selección del umbral θ para todo los modelos evaluados, es aquel donde se maximice el índice de Youden ($J = \text{sensibilidad} + \text{especificidad} - 1$), ya que permite identificar el punto donde la curva ROC posee la mayor capacidad discriminativa global.

Debido al pronunciado desbalance de clases presente en los datos, se seleccionaron métricas de evaluación robustas ante este fenómeno. Las métricas reportadas para el conjunto de prueba incluyen: la exactitud balanceada (*balanced accuracy*), la cual calcula el promedio entre la sensibilidad y la especificidad; la precisión (*precision*), que cuantifica la proporción de detecciones positivas que corresponden a anomalías reales; la exhaustividad (*recall* o sensibilidad), que mide la fracción de anomalías reales correctamente identificadas; y el valor F1, definido como la media armónica entre la precisión y la exhaustividad. De manera complementaria, se reporta la matriz de confusión detallada para cada modelo evaluado.

2.5 Diseño experimental

En esta sección se describe el diseño experimental implementado para la evaluación de los modelos de detección de anomalías propuestos. Se presentan las estrategias empleadas para el entrenamiento y validación de los modelos, así como los criterios utilizados para la selección de hiperparámetros (en los casos que aplique) y la evaluación del desempeño.

2.5.1 Auto-codificadores recurrentes

La selección de los hiperparámetros de los modelos (RNN simple/ LSTM /GRU) se realizó mediante un proceso de optimización bayesiana, empleando el desempeño sobre el conjunto de validación como criterio objetivo. Para cada arquitectura se definió un espacio de búsqueda compuesto por el número de unidades ocultas (16, 32, 64, 128), la función de activación (ReLU, Tanh), la tasa de aprendizaje (0,001; 0,003; 0,01) y el *recurrent dropout* (0; 0,1; 0,2). Durante el proceso de optimización, cada configuración propuesta fue entrenada y evaluada automáticamente, permitiendo identificar la combinación de hiperparámetros con mejor capacidad de generalización, las cuales se muestran a continuación:

Tabla 2. Mejores hiper-parámetros encontrados para los auto-codificadores recurrentes.

Arquitectura	Unidades ocultas	Activación	Tasa de aprendizaje (LR)	Recurrent dropout
RNN Simple	32	ReLU	0,01	0,0
LSTM	128	Tanh	0,01	0,2
GRU	128	Tanh	0,001	0,1

Una vez seleccionada la configuración óptima, se realizó el entrenamiento final de cada modelo y se analizaron las curvas de pérdida para los datos de entrenamiento y validación con el propósito de mostrar la evolución del aprendizaje. Estas curvas permiten evaluar la estabilidad del proceso de entrenamiento, donde se evidencia la convergencia de la función de pérdida y la presencia de posibles fenómenos de sobreajuste o subajuste. La Figura 6, muestran las curvas de aprendizaje, donde se puede observar la convergencia de la función de pérdida y no se aprecia sobre ajuste o subajuste en el proceso.

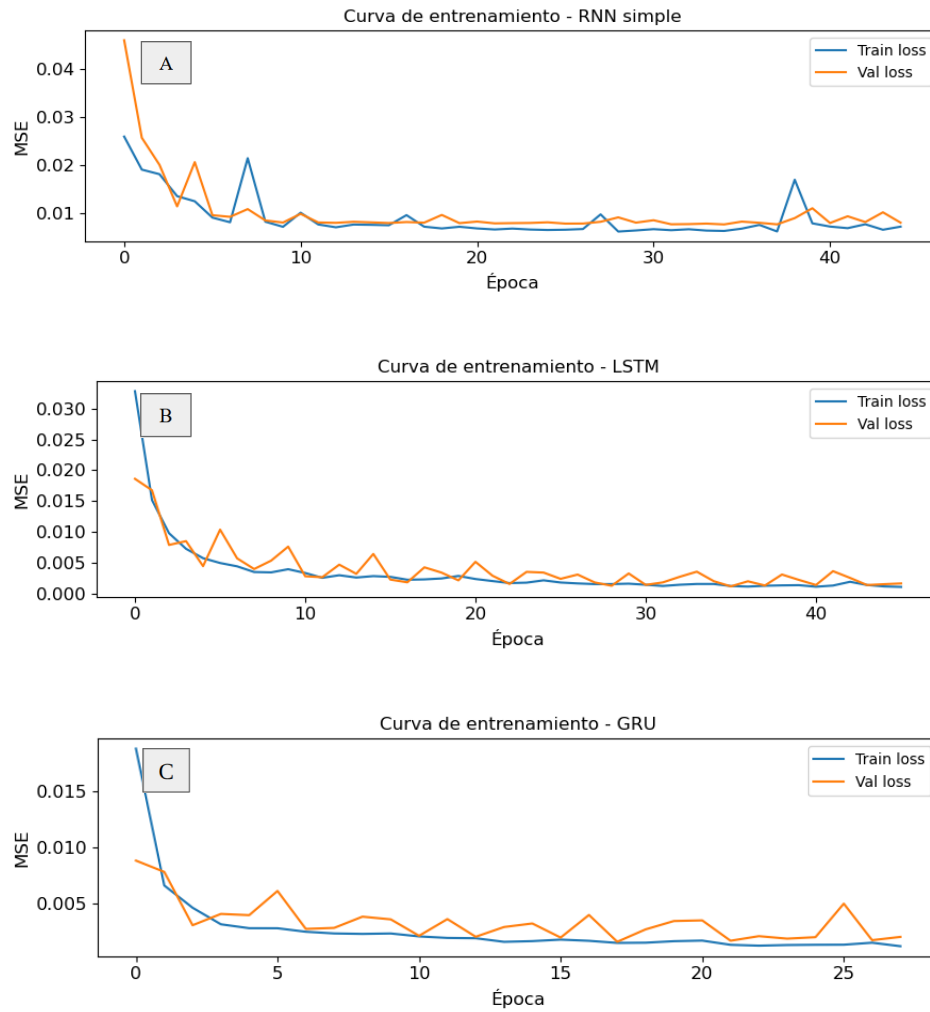


Figura 6. Curva de entrenamiento para los auto-codificadores recurrentes: A) Simple RNN. B) LSTM y C) GRU.

2.5.2 USAD

El diseño experimental del aprendizaje por transferencia de dominio consistió en las siguientes etapas: (1) carga del modelo USAD pre-entrenado en el conjunto SWaT (H=306, 51 variables industriales), extrayendo la submatriz del codificador adaptada al canal de temperatura (columnas [:, :12]); (2) fase de transferencia TL-1 con las capas internas congeladas sobre el conjunto de datos de entrenamiento, $lr=5 \times 10^{-4}$, hasta 30 épocas con parada temprana; (3) fase de ajuste fino TL-2 con todos los parámetros descongelados, $lr=1 \times 10^{-4}$, hasta 60 épocas con parada temprana.

La Figura 7 presenta las curvas de pérdida durante las dos fases del ajuste fino (TL-1 y TL-2) del modelo USAD, que evidencian convergencia del proceso de optimización: la pérdida de validación (Val L1) descendió de 0,427 a 0,424 y se mantuvo estable, mientras que la pérdida de entrenamiento L1 se estabilizó en torno a 0,504, sin signos de sobreajuste entre ambas fases.

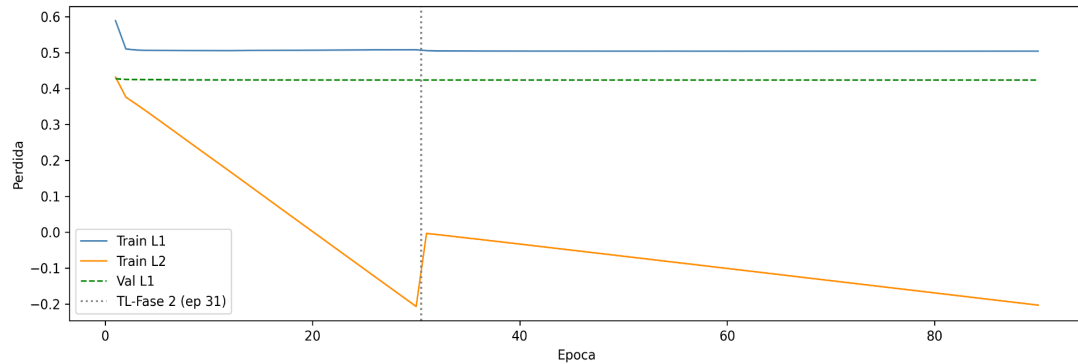


Figura 7. Curvas de pérdida (entrenamiento y validación) del modelo USAD durante el ajuste fino.

2.5.3 Anomaly-Transformer con aprendizaje por transferencia

El entrenamiento se estructuró en dos rondas sucesivas sobre el mismo 70 % del conjunto de entrenamiento: (1) pre-entrenamiento (PT) con encoder congelado (PT-P1: 10 épocas, $lr=1 \times 10^{-4}$) seguido de ajuste fino completo (PT-P2: máx. 40 épocas, $lr=5 \times 10^{-5}$); y (2) fase de transferencia (TL) con la misma estrategia de descongelamiento progresivo (TL-P1: 10 épocas; TL-P2: máx. 30 épocas, $lr=5 \times 10^{-5}$). Ambas rondas utilizan el mismo conjunto de entrenamiento, transfiriendo los pesos de PT como punto de partida para TL. En las fases de ajuste fino (PT-P2 y TL-P2) se empleó parada temprana con criterio de mínima pérdida de reconstrucción sobre el conjunto de validación (paciencia de 10 y 8 épocas, respectivamente), restaurando al final de cada ronda el estado del modelo con mejor desempeño registrado.

El criterio de puntuación de anomalía adoptó el enfoque combinado propuesto por Xu et al. (2022). Durante la inferencia, el modelo produce, para cada ventana, tanto la señal reconstruida como la discrepancia de asociación —la divergencia entre la distribución de atención de la serie y el prior gaussiano aprendido—. Aplicar la función softmax a la discrepancia negada convierte estos valores en pesos normalizados sobre los pasos de tiempo de la ventana: los instantes cuya atención se aleja del prior gaussiano (comportamiento atípico) reciben pesos menores, mientras que los instantes cuya atención es coherente con el prior (comportamiento normal) reciben pesos mayores. Este mecanismo actúa como una ponderación estructural que regula la contribución relativa de cada paso de tiempo al *score* final, amplificando el contraste entre regiones normales y anómalas de la ventana.

Las Figuras 8 y 9 presentan las curvas de pérdida durante el pre-entrenamiento (PT) y la fase de transferencia (TL), respectivamente, que evidencian la convergencia del proceso de aprendizaje en cada ronda: la pérdida de validación descendió de Val L1 $\approx 1,40$

al inicio del PT hasta $\text{Val L1} \approx 0,008$ al restaurar el mejor estado, mientras que en la fase TL convergió desde $\text{Val L1} \approx 0,009$ hasta $\text{Val L1} \approx 0,004$, sin signos de sobreajuste entre las fases congelada y descongelada.

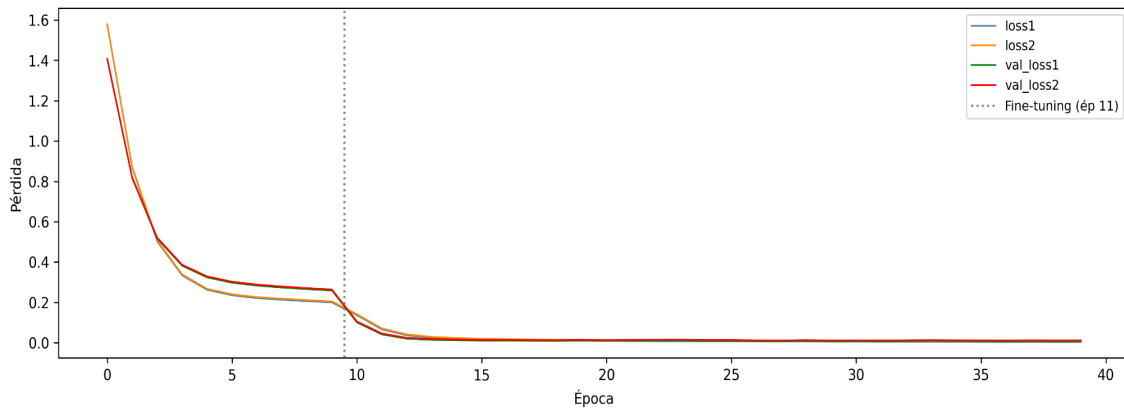


Figura 8. Curvas de pérdida durante el preentrenamiento del modelo Anomaly-Transformer.

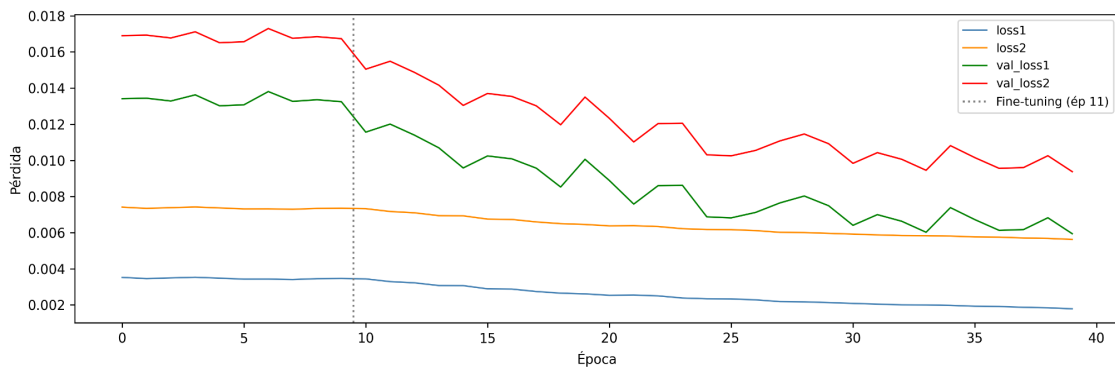


Figura 9. Curvas de pérdida durante el ajuste fino del modelo Anomaly-Transformer.

3. Resultados y Discusión

3.1 Modelos

3.1.1 Autoencoder recurrente

A continuación se presentan los resultados de la evaluación de los modelos auto-codificadores recurrentes (Simple RNN, LSTM y GRU) para la detección de anomalías en la serie de temperatura. Las Figuras 10,11,12-A presentan las curvas ROC y criterio de Youden J para la selección del umbral de cada arquitectura evaluada sobre el conjunto de validación. Las Figuras 10,11,12-B ilustran la distribución del error de reconstrucción sobre el conjunto de prueba, donde se aprecia los eventos anómalos que superan el umbral establecido y los que no. Adicionalmente, las Figuras 10,11,12-C representan la curva ROC obtenida a partir del error de reconstrucción de los datos de prueba. Finalmente, las Figuras 10,11,12-D muestran la superposición entre la serie original y la reconstruida, junto con las anomalías reales y las detectadas por cada

modelo, evidenciando que las diferentes arquitecturas construyen adecuadamente la serie original y son capaces de detectar las anomalías.

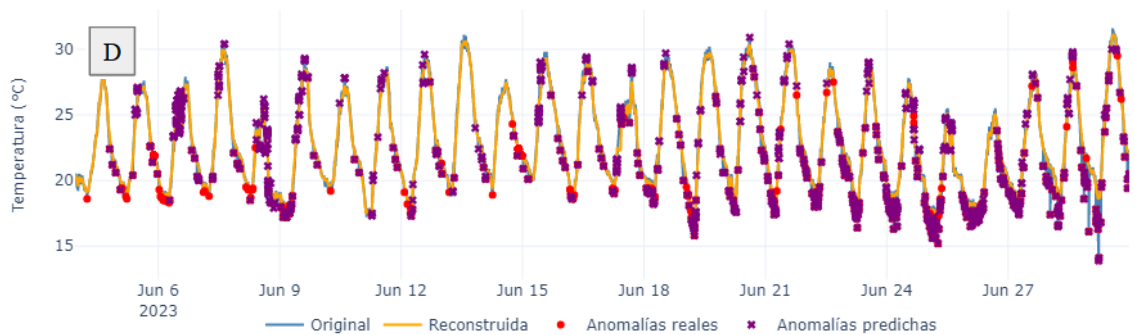
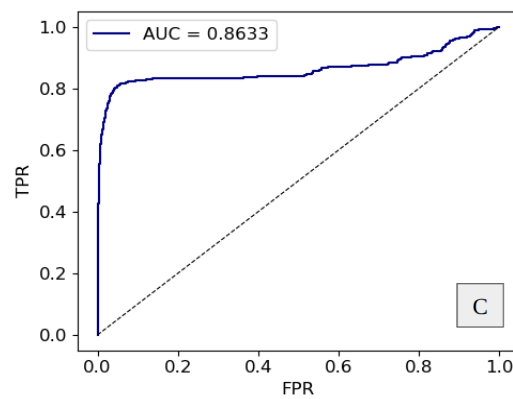
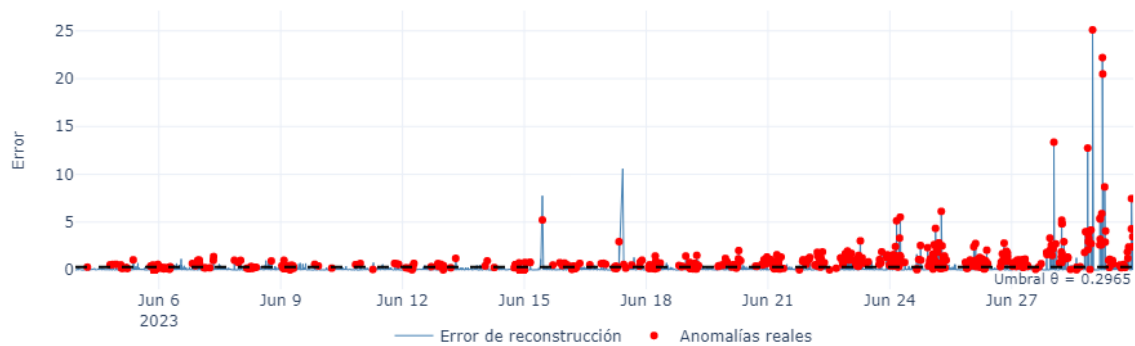
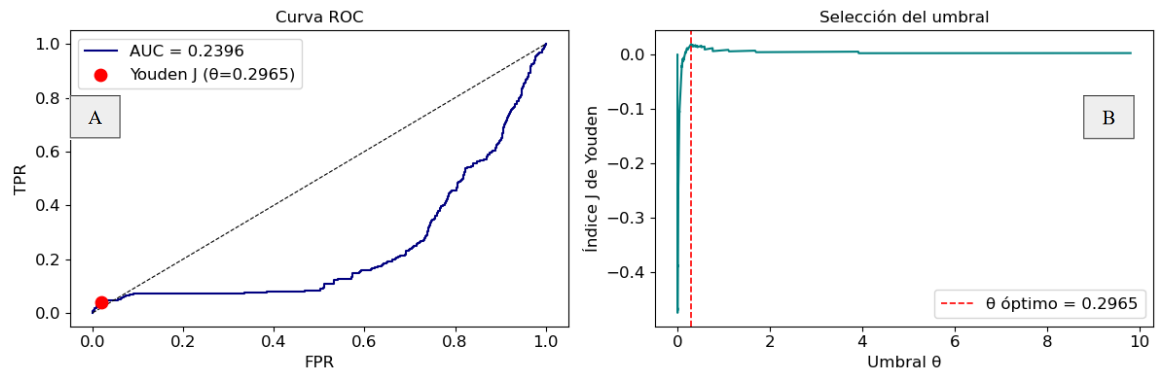


Figura 10. A) Curva ROC y criterio de Youden J para la selección del umbral $\theta : 0,2965$, empleando los datos de validación. B) Error en la reconstrucción. C) Curva ROC para los datos de prueba. D) Comparación serie original y reconstruida — Simple RNN.

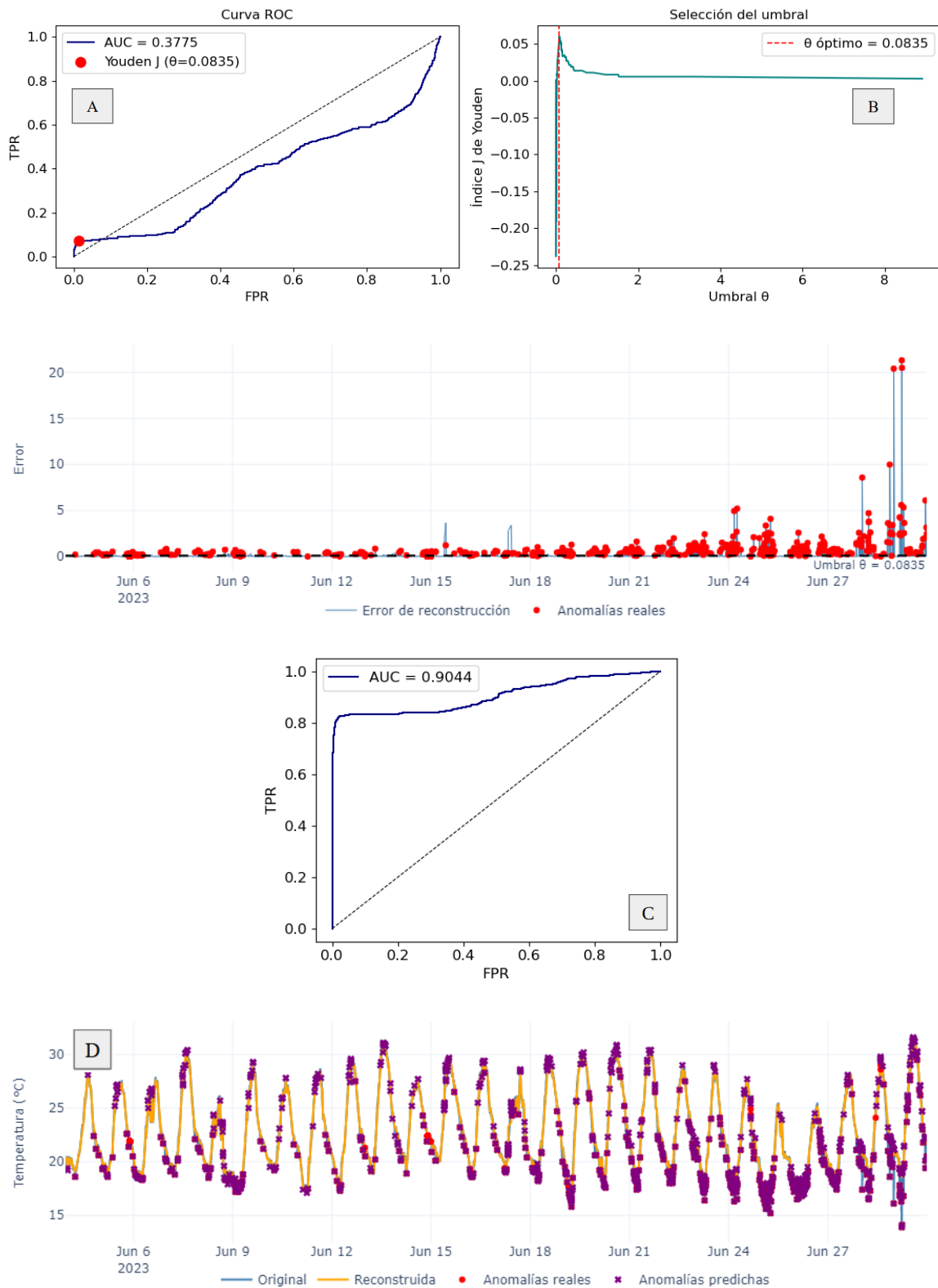


Figura 11. A) Curva ROC y criterio de Youden J para la selección del umbral θ : 0,0835, empleando los datos de validación. B) Error en la reconstrucción. C) Curva ROC para los datos de prueba. D) Comparación serie original y reconstruida — LSTM.

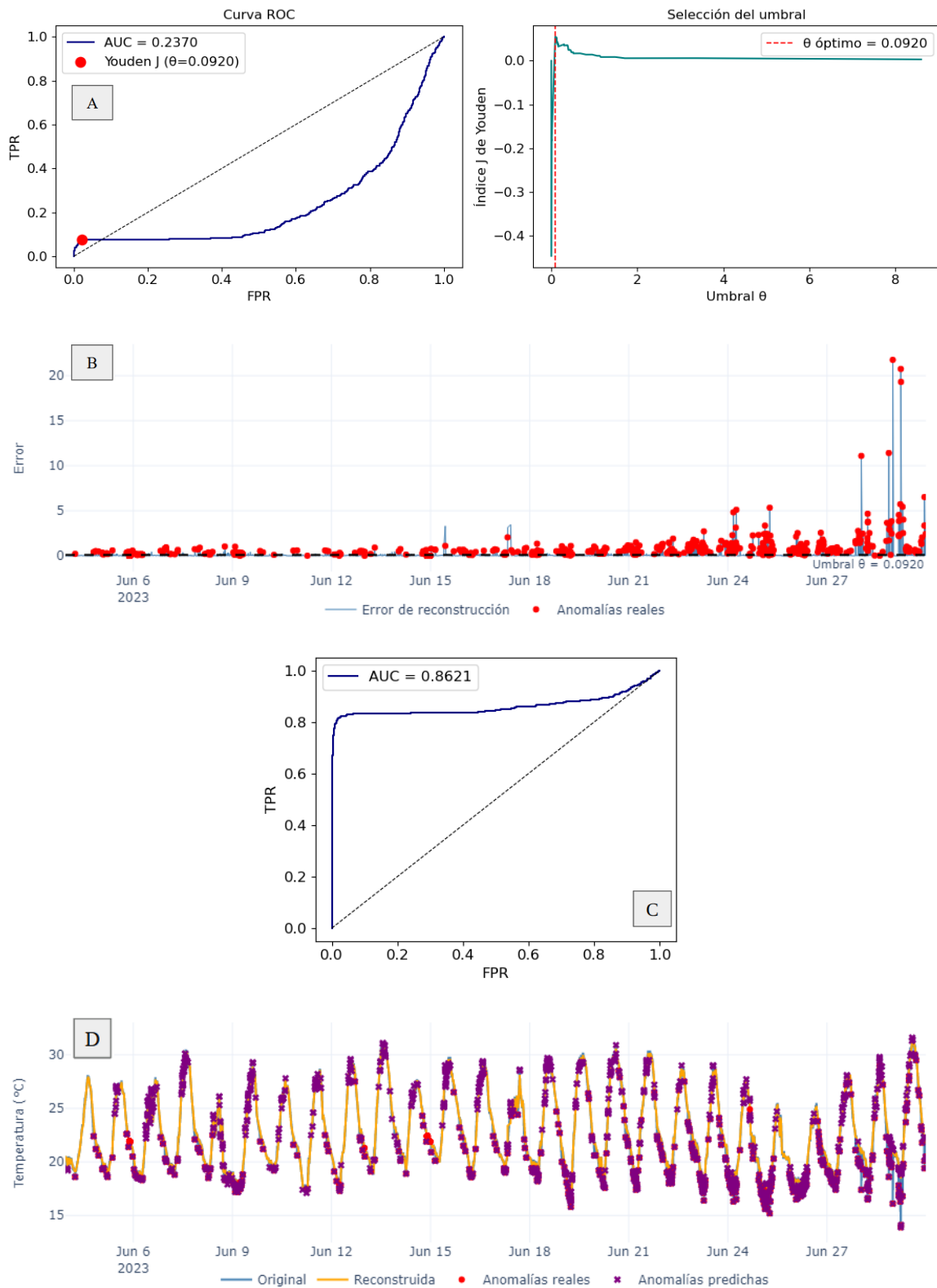


Figura 12. A) Curva ROC y criterio de Youden J para la selección del umbral $\theta : 0,0920$, empleando los datos de validación. B) Error en la reconstrucción. C) Curva ROC para los datos de prueba. D) Comparación serie original y reconstruida — GRU.

Para conocer la capacidad de detección de anomalías en los tres modelos evaluados, se muestran los resultados de la matriz de confusión normalizada (Figura 13) y las diferentes métricas de evaluación (Tabla 3)

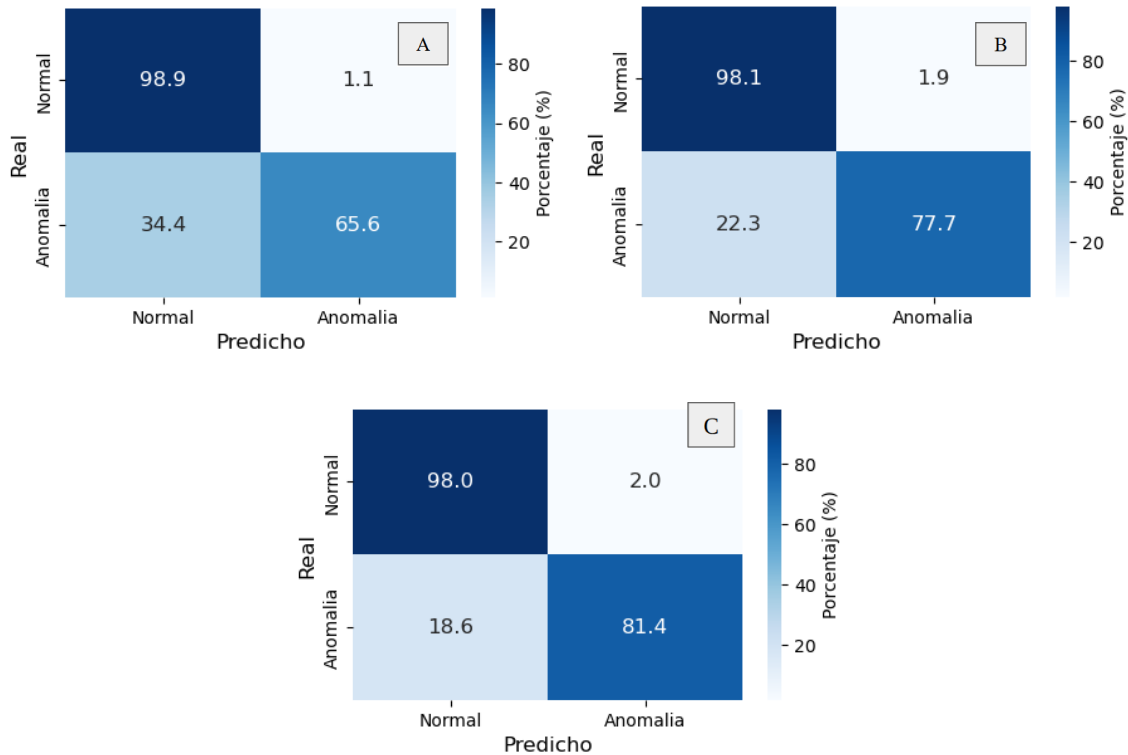


Figura 13. Matriz de confusión normalizada para los modelos auto-decodificadores recurrentes en el conjunto de prueba A) Simple RNN, B) LSTM y C) GRU.

Tabla 3. Resultados comparativos de los modelos auto-codificadores recurrentes en el conjunto de prueba.

Modelo	Exactitud Balanceada	Precisión	Exhaustividad (Recall)	Valor F1	Umbral
RNN Simple	0,8224	0,5145	0,6562	0,5768	0,2965
LSTM	0,8978	0,4914	0,8110	0,6120	0,0835
GRU	0,8970	0,4293	0,8139	0,5621	0,0920

Las métricas revelan un comportamiento consistente: la arquitectura LSTM alcanzó el mejor desempeño global, registrando un valor F1 de 0,612 y una exactitud balanceada de 0,8978.

Un aspecto relevante es la variación del umbral de decisión óptimo entre arquitecturas. La RNN simple presenta un umbral de 0,2965, notablemente superior al de

LSTM y GRU (0,035 y 0,0920, respectivamente). Esta diferencia refleja una reconstrucción menos precisa de las observaciones normales: la RNN simple genera un error más elevado durante períodos sin anomalías, lo que desplaza el umbral hacia valores más altos para evitar un exceso de falsos positivos. Esta limitación se manifiesta directamente en el recall del modelo (0,6562), inferior al de LSTM (0,811) y GRU (0,8139).

3.1.2 USAD con aprendizaje por transferencia

Dentro de los resultados del modelo USAD, empleando los datos de validación se obtuvo un umbral de detección de $\theta = 0,3083$, con un AUC = 0,5580 (Figura 14).

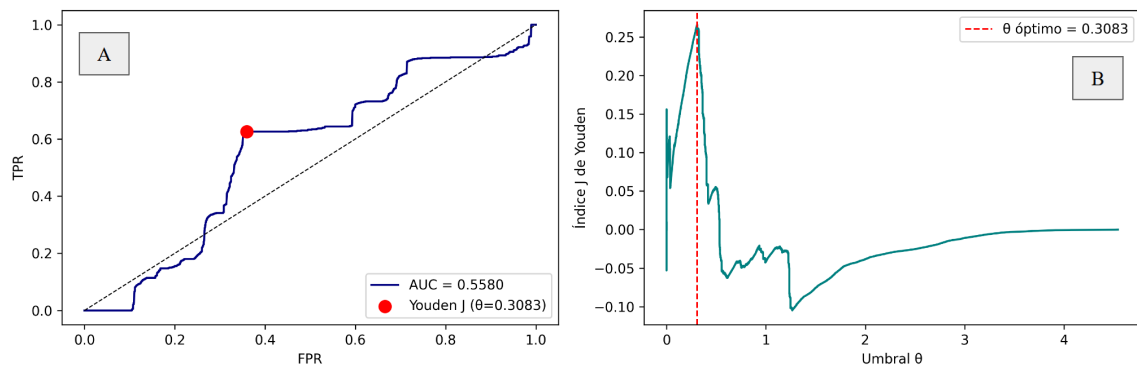


Figura 14. A) Curva ROC (AUC = 0,5580) B) criterio de Youden J para la selección del umbral $\theta = 0,308279$ — USAD — validación.

La Figura 15 y 16 muestra, para los datos de prueba, la distribución del error de reconstrucción del modelo USAD y la curva ROC, donde este alcanzó un AUC = 0,5826.

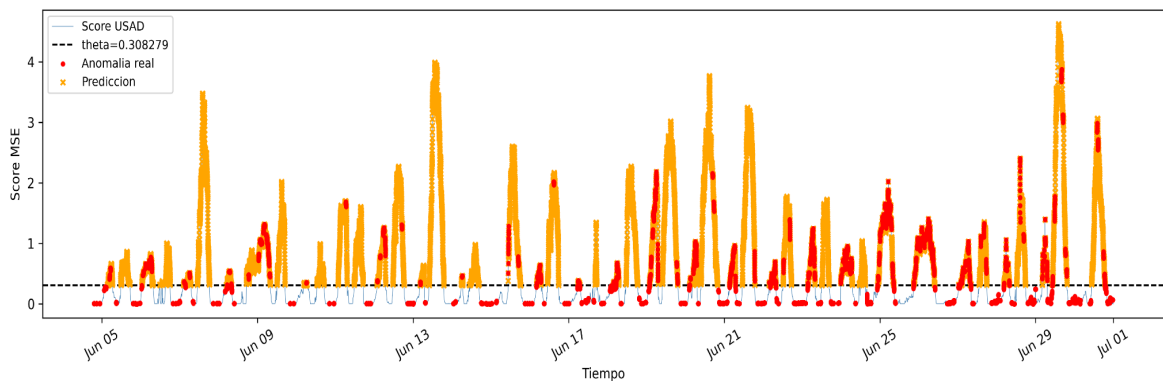


Figura 15. Error de reconstrucción USAD.

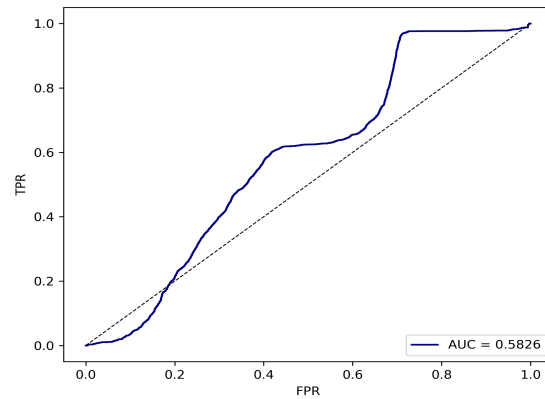


Figura 16. Curva ROC sobre el conjunto de prueba (AUC = 0,5826) — USAD

Por último, La Figura 17 representa la comparación de la serie original y reconstruida del conjunto de datos de prueba; en esta se señalan las anomalías reales y las detectadas por el modelo.

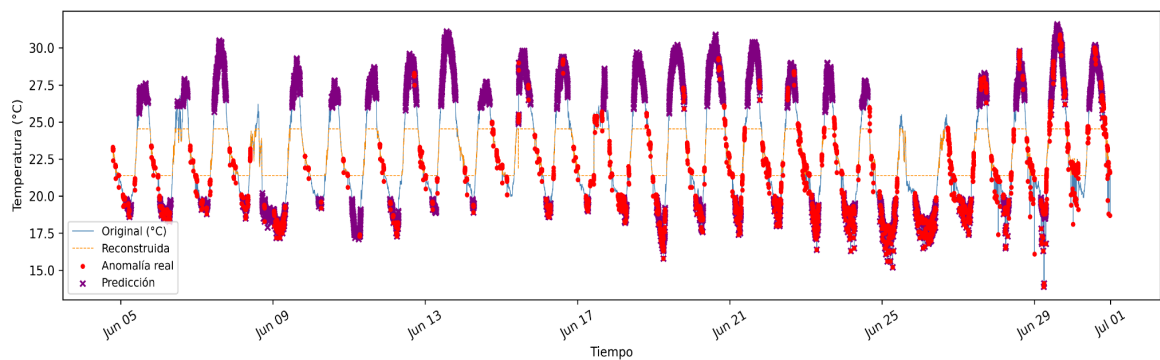


Figura 17. Señal reconstruida USAD vs. señal original.

En el conjunto de prueba, el modelo alcanzó exactitud balanceada = 0,5788, F1 = 0,3015, Exhaustividad (Recall) = 0,5414 y Precisión = 0,2089. La Figura 18 muestra la matriz de confusión normalizada por fila y en porcentaje.

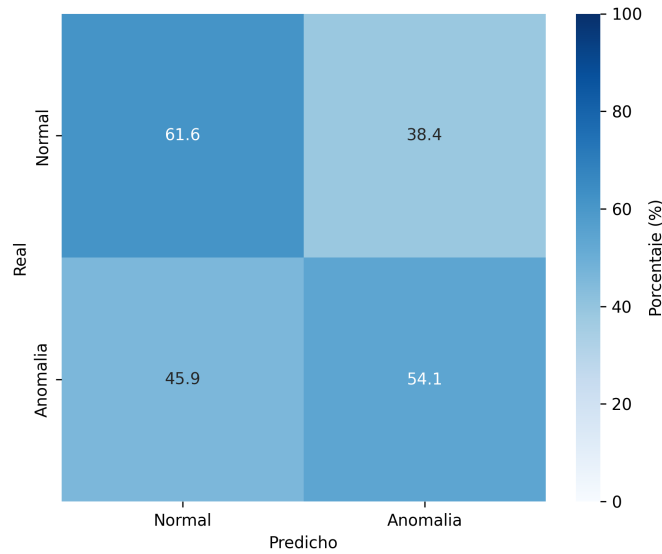


Figura 18. Matriz de confusión normalizada USAD.

En cuanto a la reconstrucción de la señal, el modelo logra aproximar la tendencia central, pero presenta una subestimación sistemática de la amplitud de la serie temporal. Los picos de temperatura y los valles no son adecuadamente reconstruidos, lo que genera errores de reconstrucción elevados. Esto queda reflejado en las métricas de evaluación: si bien el modelo alcanza una exhaustividad (Recall) del 54,14 %, lo que indica que detecta poco más de la mitad de las anomalías reales, la precisión se reduce a apenas 20,89 %, confirmando que una proporción significativa de las alertas generadas corresponden a falsas alarmas. Como resultado, el F1-score se sitúa en 0,3015 y la exactitud balanceada en 0,5788, valores que evidencian un desempeño con sesgo hacia la sobredetección. La matriz de confusión refuerza este diagnóstico: el 38,4 % de los puntos normales son clasificados erróneamente como anomalías, mientras que el 45,9 % de las anomalías reales no son detectadas.

3.1.3 Anomaly-Transformer con aprendizaje por transferencia

Para el Anomaly-Transformer, empleando los datos de validación se obtuvo un umbral de detección $\theta = -0,000121$, con un AUC = 0,6944 (Figura 19); donde el signo negativo del umbral es gracias al *score* de anomalías dado por el modelo.

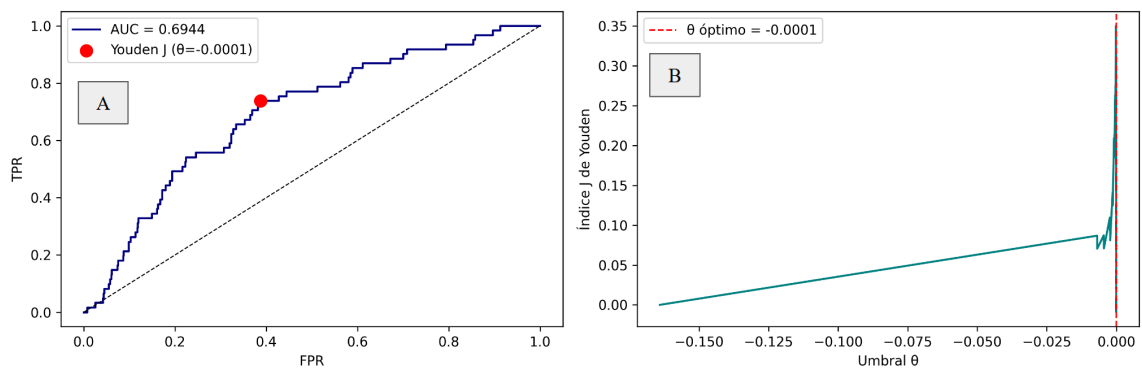


Figura 19. A) Curva ROC (AUC = 0,6944) B) criterio de Youden J para la selección del umbral $\theta = -0,000121$ — Anomaly-Transformer.

La Figura 20 y 21 muestra, para los datos de prueba, el *score* de anomalía del modelo y la curva ROC, donde el modelo alcanzó un AUC = 0,6044.

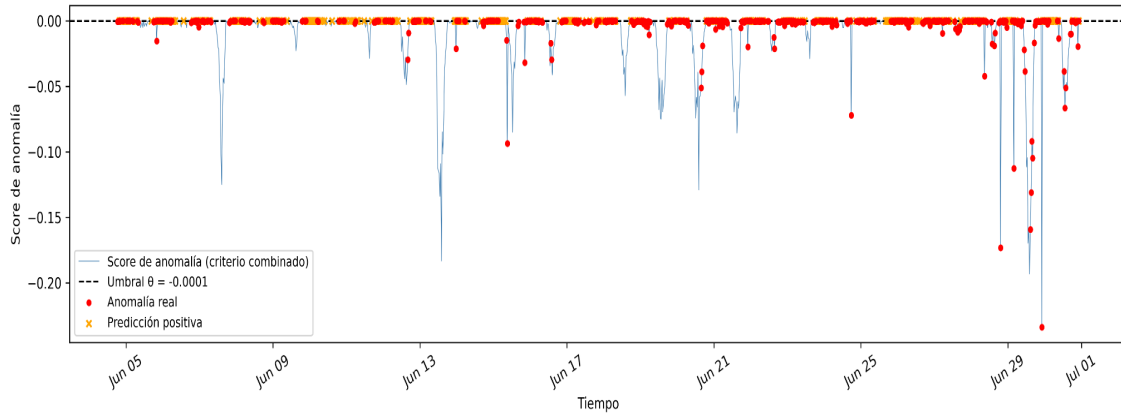


Figura 20. Error de reconstrucción Anomaly-Transformer.

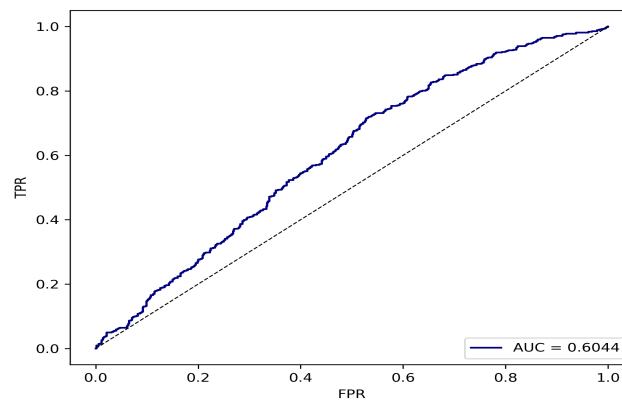


Figura 21. Curva ROC sobre el conjunto de prueba (AUC = 0,6044) — Anomaly-Transformer. El umbral θ fue definido exclusivamente en validación.

La Figura 22, representa la señal reconstruida frente a la original sobre el conjunto de prueba y resalta las anomalías reales y predichas por el modelo.

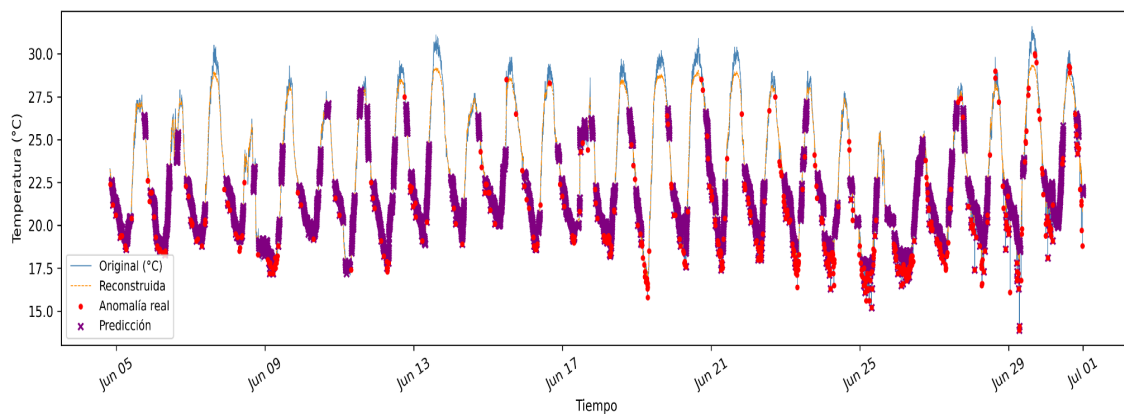


Figura 22. Señal reconstruida Anomaly-Transformer vs. señal original.

El modelo Anomaly-Transformer alcanzó $BA = 0,5477$, $F1 = 0,4298$, Precisión = $0,5088$ y Recall = $0,3720$. La Figura 23 visualiza la matriz de confusión resultante.

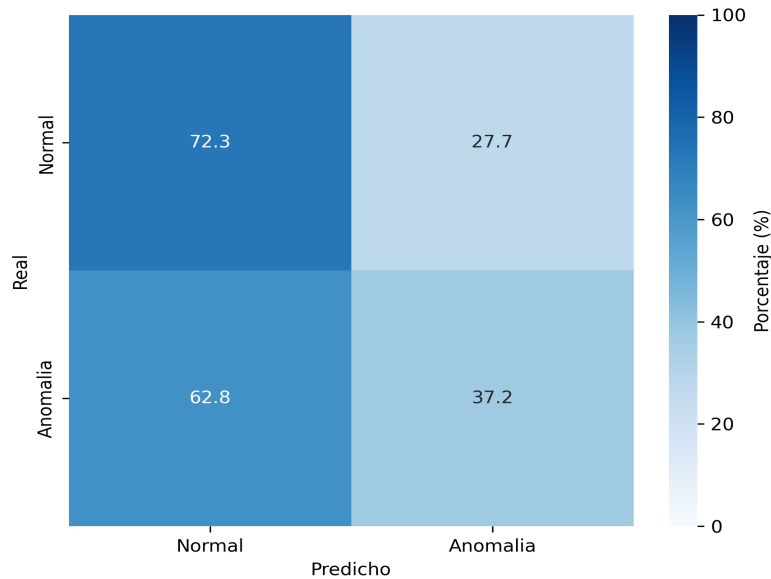


Figura 23. Matriz de confusión normalizada Anomaly-Transformer .

De lo obtenido, se encuentra que el modelo Anomaly-Transformer logra discriminar entre observaciones normales y anómalas, con una especificidad notable: el 72,3 % de las observaciones normales son clasificadas correctamente, mientras que solo el 27,7% son etiquetadas erróneamente como anomalías. Sin embargo, el modelo exhibe un sesgo hacia la infra detección: únicamente el 37,2 % de las anomalías reales son detectadas, lo que implica que el 62,8 % de las anomalías no son identificadas como tales. Este comportamiento se refleja directamente en las métricas: la precisión de $0,5088$ indica que más de la mitad de las alertas generadas corresponden a anomalías reales, pero el recall reducido de $0,3720$ limita la capacidad del modelo para capturar la totalidad del fenómeno anómalo.

3.2 Análisis e interpretación de resultados

En el modelo USAD, la señal reconstruida no logra capturar la amplitud de los ciclos diurnos, manteniéndose en un rango estrecho alrededor de la media, lo que genera errores de reconstrucción elevados en los picos y valles. Como consecuencia, el modelo opera con una precisión de apenas $0,2089$, confirmando que la mayoría de sus alertas corresponden a falsos positivos derivados de la variabilidad normal de la serie. Si bien su *recall* de $0,5414$ indica que detecta algo más de la mitad de las anomalías reales, el *F1-score* de $0,3015$ y la exactitud balanceada de $0,5788$ evidencian un desempeño global insuficiente para un sistema de alerta confiable.

El modelo Anomaly-Transformer, por su parte, resuelve el problema de reconstrucción de manera sobresaliente, siguiendo tanto los picos como los valles de la

serie original. El criterio combinado de puntuación permite al modelo discriminar entre observaciones normales y anómalas, alcanzando una precisión de 0,5088 y una exactitud balanceada de 0,5477. No obstante, el recall de 0,3720 indica que el modelo detecta apenas el 37,2 % de las anomalías reales, exhibiendo un sesgo hacia la infra detección.

Los modelos de autoencoder recurrente —RNN Simple, LSTM y GRU— superan a los dos anteriores en términos de desempeño global. Los tres logran reconstruir la dinámica general de la serie y sus métricas reflejan un comportamiento discriminativo entre las observaciones normales y anómalas. Entre ellos, el modelo LSTM se posiciona como la arquitectura de mejor desempeño, alcanzando una exactitud balanceada de 0.8978, un recall de 0.8110 y un F1-score de 0.6120, lo que lo convierte en el modelo con mayor capacidad para detectar anomalías reales manteniendo un nivel razonable de precisión (0.4914). El GRU ofrece resultados comparables aunque ligeramente inferiores, mientras que el RNN Simple, pese a obtener la precisión más alta de los auto-codificadores recurrentes (0.5145), sacrifica exhaustividad al generar un error más elevado durante períodos sin anomalías.

3.3 Limitaciones

El presente estudio advierte las siguientes limitaciones metodológicas. En primer lugar, la evaluación cuantitativa de los modelos se ejecuta exclusivamente sobre el sensor 203 de enero-junio del 2023. Por consiguiente, los resultados no son directamente extrapolables a la totalidad de la red SIATA, se requiere hacer más experimentos para conocer el rendimiento del modelo frente a los datos de otros sensores meteorológicos con variaciones de temperatura similares y en otros periodos de tiempo. En segundo lugar, la dependencia del rendimiento de los algoritmos respecto a la calidad del conjunto de etiquetas de supervisión introduce un margen de incertidumbre.

4. Conclusiones

El presente trabajo evaluó tres enfoques de detección de anomalías basadas en aprendizaje profundo —USAD, Anomaly-Transformer y auto-decodificadores recurrentes (RNN Simple/ LSTM / GRU)— sobre una serie temporal de temperatura del aire del sensor 203, con el objetivo de identificar el modelo más adecuado para este tipo de datos en términos de precisión, exhaustividad y desempeño global.

Los resultados demuestran que la calidad de la reconstrucción de la señal y la calibración del umbral de decisión son los dos factores determinantes del éxito de un detector de anomalías basado en error de reconstrucción. El modelo USAD evidenció una reconstrucción deficiente en los picos y valles de la serie lo que conduce inevitablemente a una alta tasa de falsos positivos. Por otro lado, el Anomaly-Transformer presenta un desempeño conservador, caracterizado por una adecuada identificación de patrones normales, pero con una baja tasa de detección de anomalías lo que indica que una proporción importante de eventos anómalos permanece sin identificar.

Los modelos de auto-codificadores recurrentes demostraron ser las arquitecturas más robustas para el problema abordado. Entre ellos, el modelo LSTM se consolidó como la mejor alternativa, alcanzando una exactitud balanceada de 0.8978, un recall de 0.8110 y un F1-score de 0.6120, valores que reflejan un equilibrio entre sensibilidad y especificidad. El modelo GRU ofreció resultados comparables, posicionándose como una alternativa eficiente, mientras que el RNN Simple, aunque más preciso, sacrificó exhaustividad al operar con un umbral más conservador.

Estos hallazgos señalan que la elección de la arquitectura no es el único factor relevante en el diseño de sistemas de detección de anomalías: la estrategia de preprocesamiento, el manejo de la señal y la metodología de calibración del umbral son elementos igualmente críticos.

Finalmente, las líneas de investigación futuras prioritarias son: (1) extender la validación a múltiples sensores de la red SIATA para verificar la generalización espacial de los modelos; (2) implementar el marco de ensamble completo S-RNN conforme a la propuesta de Kieu et al. (2019); (3) explorar estrategias de aprendizaje continuo (*online learning*) para adaptación ante derivas de los datos (*data drift*); y (4) desarrollar una interfaz de visualización integrada con la infraestructura operacional del SIATA para detección en tiempo real de anomalías en la totalidad de la red térmica.

5. Referencias

- Abdulaal, A., Liu, L., & Benesty, M. (2021). Practical approach to asynchronous multivariate time series anomaly detection and localization. En Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (pp. 2485–2494). <https://doi.org/10.1145/3447548.3467174>
- Alcaldía de Medellín, & Área Metropolitana del Valle de Aburrá. (2020). Sistema de Alerta Temprana del Valle de Aburrá (SIATA): Descripción general y cobertura de la red de monitoreo. Área Metropolitana del Valle de Aburrá.
- Área Metropolitana del Valle de Aburrá. (2019). Plan de Gestión Integral de Residuos Sólidos del Valle de Aburrá y diagnóstico ambiental regional. AMVA. https://www.medellin.gov.co/normograma/docs/astrea/docs/A_AMVA_0004_2020.htm
- Audibert, J., Michiardi, P., Guyard, F., Marti, S., & Zuluaga, M. A. (2020). USAD: UnSupervised Anomaly Detection on multivariate time series. En Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 3395–3404). <https://doi.org/10.1145/3394486.3403392>
- Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. En Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (pp. 93–104). <https://doi.org/10.1145/342009.335388>
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58. <https://doi.org/10.1145/1541880.1541882>

- Chen, J., Sathe, S., Aggarwal, C. C., & Turaga, D. S. (2017). Outlier detection with autoencoder ensembles. En *Proceedings of the 2017 SIAM International Conference on Data Mining* (pp. 90–98). <https://doi.org/10.1137/1.9781611974973.11>
- Chung, J., Gülçehre, Ç., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*. <https://arxiv.org/abs/1412.3555>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Soderstrom, T. (2018). Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. En *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 387–395). <https://doi.org/10.1145/3219819.3219845>
- Instituto de Hidrología, Meteorología y Estudios Ambientales. (2015). *Atlas Climatológico de Colombia*. IDEAM.
- Intergovernmental Panel on Climate Change. (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (V. Masson-Delmotte et al., Eds.). Cambridge University Press. <https://doi.org/10.1017/9781009157896>
- Kieu, T., Yang, B., Guo, C., & Jensen, C. S. (2019). Outlier detection for time series with recurrent autoencoder ensembles. En *Proceedings of the 28th International Joint Conference on Artificial Intelligence* (pp. 2725–2732). <https://doi.org/10.24963/ijcai.2019/378>
- Lavin, A., & Ahmad, S. (2015). Evaluating real-time anomaly detection algorithms: The Numenta Anomaly Benchmark. En *Proceedings of the 14th IEEE International Conference on Machine Learning and Applications* (pp. 38–44). <https://doi.org/10.1109/ICMLA.2015.141>
- Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. En *Proceedings of the 8th IEEE International Conference on Data Mining* (pp. 413–422). <https://doi.org/10.1109/ICDM.2008.17>
- Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., & Shroff, G. (2016). LSTM-based encoder-decoder for multi-sensor anomaly detection. *arXiv preprint arXiv:1607.00148*. <https://arxiv.org/abs/1607.00148>
- Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., & Pei, D. (2019). Robust anomaly detection for multivariate time series through stochastic recurrent neural network. En *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2828–2837). <https://doi.org/10.1145/3292500.3330672>
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. En *Advances in Neural Information Processing Systems* (Vol. 27, pp. 3104–3112). Curran Associates.

https://proceedings.neurips.cc/paper_files/paper/2014/hash/a14ac55a4f27472c5d894ec1c3c743d2-Abstract.html

Xu, J., Wu, H., Wang, J., & Long, M. (2022). Anomaly Transformer: Time series anomaly detection with association discrepancy. En Proceedings of the 10th International Conference on Learning Representations.
<https://openreview.net/forum?id=LzQQ89U1qm>