Naturaleza y TECNOLOGÍA
Número Especial "Verano de la Investigación DCNE 2026"
ISSN 2007-672X
Universidad de Guanajuato

REDES NEURONALES PROFUNDAS COMO MODELOS SUSTITUTOS EN INGENIERÍA DE PROCESOS: CONTRASTE ENTRE UNA SECUENCIA DE DESTILACIÓN Y UN INTERCAMBIADOR DE CALOR

Sebastián Ayala Palomares [a], Ángel Ramón Chávez Ortíz [a], Jesús Manuel Núñez López [a,*]

[a] Universidad de Guanajuato, Campus Guanajuato, División de Ciencias Naturales y Exactas, Departamento de Ingeniería Química, Noria Alta s/n, 36050, Guanajuato, Gto., México.
jm.nunez@ugto.mx

Resumen

El diseño riguroso de equipos de separación y de transferencia de calor mediante simuladores de proceso y algoritmos de optimización híbridos (determinísticos-metaheurísticos) es computacionalmente costoso, lo que limita su utilización en aplicaciones que requieren evaluaciones repetidas o respuestas rápidas. Este trabajo evalúa el desempeño de redes neuronales profundas (DNN) frente a modelos de regresión lineal múltiple como modelos sustitutos en dos casos de estudio de distinta naturaleza: 1) una secuencia directa de dos columnas de destilación (sistema benceno-etilbenceno-dietilbenceno; 6,000 diseños, 10 variables de entrada, y dos respuestas: costo total anual y Eco-Indicador 99) y 2) un intercambiador de calor (5,000 casos, 5 variables de entrada, y una respuesta: coeficiente global de transferencia de calor). En cada caso de estudio, los modelos se compararon utilizando particiones de datos y criterios de evaluaciones consistentes, mientras que las DNN conservaron una arquitectura común de tres capas ocultas (64-32-16). Para destilación, 30 entrenamientos independientes de la DNN produjeron, para el costo total, un MAE de 27.52 ± 8.28 y un R^2 de 0.9977 ± 0.0016, frente a 52.47 y 0.9941 para la regresión lineal. En contraste, para el Eco-Indicador 99 la regresión lineal presentó menor error que la DNN (MAE de 13.28 frente a 27.07 ± 7.47), mostrando que una mayor complejidad del modelo no garantiza una mejora predictiva. Para el intercambiador de calor, la DNN redujo el MAE de 29.90 a 4.88 W/m²K y aumentó el R^2 de 0.8993 a 0.9978 respecto a la regresión lineal, mientras que un Bosque Aleatorio alcanzó un desempeño comparable R^2=0.9979. Los resultados muestran que la conveniencia de utilizar modelos de mayor complejidad depende de la estructura de las relaciones entrada–salida y respaldan la evaluación de modelos de referencia simples y de múltiples entrenamientos independientes antes de seleccionar una DNN como modelo sustituto.

Palabras clave: Modelos Sustitutos; Redes Neuronales Profundas; Regresión Lineal; Ingeniería de Procesos.

DEEP NEURAL NETWORKS AS SURROGATE MODELS IN PROCESS ENGINEERING: A CONTRAST BETWEEN A DISTILLATION SEQUENCE AND A HEAT EXCHANGER

Abstract

Rigorous design of separation and heat-transfer equipment using process simulators and optimization algorithms can involve substantial computational cost, limiting its use in applications requiring repeated evaluations or rapid responses. This work evaluates the performance of deep neural networks (DNNs) against multiple linear regression models as surrogate models in two case studies of different nature: a direct sequence of two distillation columns (6,000 designs, 10 input variables, and two responses: total cost and Eco-Indicator 99) and a heat exchanger (5,000 cases, five input variables, and one response: overall heat-transfer coefficient). Within each case study, the models were compared using consistent data partitions and evaluation criteria, while the DNNs retained a common three-hidden-layer architecture (64–32–16). For distillation, 30 independent DNN training runs yielded a cost MAE of 27.52 ± 8.28 and an R^2 of 0.9977 ± 0.0016, compared with 52.47 and 0.9941 for linear regression. In contrast, linear regression produced a lower error for Eco-Indicator 99 than the DNN (MAE of 13.28 versus 27.07 ± 7.47), demonstrating that greater model complexity does not necessarily translate into improved predictive performance. For the heat exchanger, the DNN reduced the MAE from 29.90 to 4.88 W/m²K and increased R^2 from 0.8993 to 0.9978 relative to linear regression, while a Random Forest achieved comparable performance R^2=0.9979. The results demonstrate that the benefit of using more complex surrogate models depends on the structure of the input–output relationships and support evaluating simple baseline models and multiple independent training run before selecting a DNN as a surrogate model.

Keywords: Surrogate Models; Deep Neural Networks; Linear Regression; Process Engineering.

Número Especial "Verano de la Investigación DCNE 2026"
ISSN 2007-672X
Universidad de Guanajuato

1. Introducción

El diseño de equipos de proceso —columnas de destilación, intercambiadores de calor, reactores— exige resolver simultáneamente balances de materia y energía, correlaciones termodinámicas no lineales y restricciones de operación, lo que en la práctica se resuelve mediante simuladores rigurosos (p. ej. Aspen Plus) usando algoritmos de optimización determinística o metaheurística. Cuando el diseño de los equipos se optimiza junto con la distribución de recursos de una cadena de suministro, el costo computacional puede volverse prohibitivo: Núñez-López y col. (2025) reporta un tiempo de cómputo de aproximadamente 22 horas para una sola corrida con un método híbrido de optimización determinística-metaheurística aplicado al diseño de una cadena de suministro de poliestireno con columnas de destilación integradas.

Frente a ese costo, los *modelos sustitutos* (surrogate models) —aproximaciones estadísticas o de aprendizaje automático entrenadas con un número limitado de evaluaciones del modelo— permiten reemplazar la simulación completa por una función mucho más barata de evaluar, conservando una fracción relevante de su exactitud (Bhosekar y Ierapetritou, 2018). Dentro de esta familia, las redes neuronales profundas se han posicionado como la técnica dominante gracias a su capacidad demostrada, de aproximar cualquier función continua con un número suficiente de neuronas ocultas (Hornik y col., 1989), y a los avances algorítmicos y de cómputo que han hecho práctico su entrenamiento a gran escala (LeCun y col., 2015; Goodfellow y col., 2016). En ingeniería química, Venkatasubramanian (2019) describe esta tendencia como parte de un "cuarta paradigma" de la ingeniería de procesos, impulsado por la disponibilidad de datos y cómputo; entonces, la adopción de gemelos digitales y manufactura inteligente en la industria química —los pilares de la Industria 4.0— depende en gran parte de que estos modelos sustitutos sean suficientemente rápidos y confiables, de manera que permitan el control y optimización en tiempo real (Pal y col., 2025).

La destilación ha sido uno de los procesos donde más se ha explorado esta sustitución. Gutiérrez-Antonio y Briones-Ramírez (2015), en uno de los primeros trabajos mexicanos sobre el tema, entrenaron redes neuronales como sustitutos de las funciones objetivo y restricciones de columnas

Número Especial "Verano de la Investigación DCNE 2026"
ISSN 2007-672X
Universidad de Guanajuato

divididas (dividing-wall), acoplándolas a un algoritmo genético multiobjetivo para alcanzar el frente de Pareto con una reducción sustancial en el número de evaluaciones del simulador riguroso. Trabajos más recientes han extendido este enfoque: Ye y col. (2023) combinaron una red neuronal tipo backpropagation (GA-BP) con un algoritmo genético para minimizar el costo total anualizado y el impacto ambiental de una columna de propileno, y Bubel y col. (2026) han propuesto arquitecturas de sustitutos reutilizables entre distintas configuraciones de columna. No obstante, Streb y col. (2022) advierten que el desempeño reportado para redes neuronales sustitutas suele evaluarse sobre rangos de operación estrechos alrededor de un punto de diseño, lo que puede inflar métricas como el coeficiente de determinación sin que eso implique una ventaja real sobre modelos más simples.

Esta línea de trabajo tiene una comunidad de investigación activa en el ámbito hispanohablante: Ramírez Rodríguez (2025) desarrolló e implementó modelos de aprendizaje profundo como sensores virtuales (soft sensors) para una columna despropanizadora industrial real, Gómez Jurado (2025) documenta, en una revisión sistemática, la adopción creciente de técnicas de inteligencia artificial en el control, la seguridad y la optimización de procesos químicos en la región.

Entonces surge una pregunta que rara vez se responde de forma explícita en la literatura aplicada: *¿cuándo se justifica, en términos de exactitud predictiva, el uso de una red neuronal profunda en lugar de un modelo lineal considerablemente más simple y barato de entrenar?* Evidencia reciente de este caso sugiere que la respuesta depende más de la estructura de los datos que de la sofisticación del algoritmo: Schulz y col. (2020), en un estudio a gran escala con datos de neuroimagen, muestran que los modelos lineales son iguales a las redes profundas en tareas de baja dimensionalidad efectiva o de curvatura débil, y que la ventaja del aprendizaje profundo solo se manifiesta y crece con el tamaño de muestra, cuando la relación implícita es genuinamente no lineal. En ingeniería de procesos esta pregunta se traduce en que: dado que entrenar una red neuronal exige más decisiones de diseño (arquitectura, regularización, inicialización) y ofrece menor interpretabilidad que una regresión lineal, su uso como sustituto debería estar justificado por evidencia de no linealidad en el sistema, y no asumirse por defecto.

Este trabajo aborda esa pregunta de manera empírica, comparando una red neuronal profunda contra una regresión lineal múltiple entrenadas bajo condiciones idénticas sobre dos casos de estudio de ingeniería química con estructuras físicas distintas: una secuencia directa de dos columnas de destilación diseñada siguiendo la metodología híbrida determinística-metaheurística de Núñez-López y col. (2025), y un intercambiador de calor cuyo coeficiente global de transferencia de calor depende de forma notoriamente no lineal de los flujos másicos de operación (Çengel y Ghajar, 2020; Incropera y col., 2017; Kakaç y col., 2012). El análisis de ambos casos permite examinar si la ventaja de la red neuronal es consistente entre sistemas o si, como sugieren Schulz y col. (2020) y Streb y col. (2022), depende de la curvatura real de la función que se busca sustituir. El objetivo del trabajo es doble: (i) cuantificar el desempeño predictivo de ambos modelos —incluyendo un análisis de la variabilidad entre corridas independientes de la red neuronal, un aspecto frecuentemente omitido en estudios similares— y (ii) proponer un criterio, basado en el análisis de sensibilidad de las variables de entrada, para decidir cuándo la complejidad adicional de una red neuronal profunda está justificada en el diseño de modelos sustitutos para ingeniería de procesos.

2. Marco teórico

2.1. Modelos sustitutos y diseño híbrido determinístico-metaheurístico.

Un modelo sustituto reemplaza una función costosa f(x) —en este caso, una simulación rigurosa de proceso seguida de una evaluación económica y ambiental— por una aproximación $f(x)$ entrenada sobre un conjunto finito de pares entrada-salida generados previamente (Bhosekar y Ierapetritou, 2018). La estrategia de diseño híbrido de Núñez-López y col. (2025), que sirve de base metodológica para el caso de destilación de este trabajo, ejemplifica el origen de esos pares: el diseño riguroso de cada columna (número de etapas, alimentación, relación de reflujo, diámetro y flujo de destilado) se resuelve mediante Evolución Diferencial con Lista Tabú en Aspen Plus, mientras que la distribución de recursos de la cadena de suministro se resuelve de forma determinística en GAMS; ambas etapas se comunican de manera iterativa hasta converger. Bajo ese esquema, cada evaluación del diseño y la cadena de suministro es costosa, lo que motiva sustituir

Número Especial "Verano de la Investigación DCNE 2026"
ISSN 2007-672X
Universidad de Guanajuato

el mapeo variable de diseño → costo/impacto ambiental por un modelo entrenado una sola vez sobre un conjunto de diseños previamente evaluados.

2.2. Redes neuronales profundas.

Una red neuronal artificial tipo perceptrón multicapa (MLP) transforma un vector de entrada $x \in \mathbb{R}^d$ en una salida $\hat{y} \in \mathbb{R}^m$ mediante una composición de L capas (Eq. 1):

$$h_l = \sigma(W_l h_{l-1} + b_l), \quad \ell = 1, \ldots, L-1 \qquad (1)$$

$$\hat{y} = W_l h_{l-1} + b_l \qquad (2)$$

donde W_l y b_l son los pesos y sesgos de la capa ℓ, $h_0 = x$, y σ es una función de activación no lineal —en este trabajo, la unidad lineal rectificada, ReLU(z) = max(0, z)— aplicada en las capas ocultas; la capa de salida usa activación lineal, adecuada para tareas de regresión. Hornik y col. (1989) demostraron que una red con una sola capa oculta y suficientes neuronas puede aproximar cualquier función continua sobre un dominio compacto con un error arbitrariamente pequeño (teorema de aproximación universal); en la práctica, sin embargo, la existencia de tal aproximación no garantiza que el algoritmo de entrenamiento la encuentre con los datos y el tiempo de cómputo disponibles. Los pesos se ajustan minimizando el error cuadrático medio entre $\hat{y}$ y el valor real y (Eq. 2), mediante retropropagación del gradiente y el optimizador Adam (Kingma y Ba, 2015), que adapta la tasa de aprendizaje de cada parámetro combinando estimaciones de primer y segundo momento del gradiente.

2.3. Regresión lineal múltiple como modelo de referencia.

La regresión lineal múltiple asume una relación aditiva entre las variables de entrada estandarizadas y cada salida (Eq. 3):

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \qquad (3)$$

con coeficientes β_j estimados por mínimos cuadrados ordinarios. Se emplea aquí no solo como línea base de comparación frente a la red neuronal, sino como herramienta de diagnóstico: al entrenarse sobre variables estandarizadas (media cero, varianza unitaria), la magnitud de cada coeficiente β_j es directamente comparable entre variables y permite identificar cuáles dominan cada salida —información que una red neuronal, por su naturaleza de caja negra, no ofrece de forma directa.

2.4. Métricas de desempeño.

El desempeño de ambos modelos se evaluó, para cada variable de salida por separado, con el error absoluto medio (MAE) (Eq. 4), la raíz del error cuadrático medio (RMSE) (Eq. 5) y el coeficiente de determinación (R^2) (Eq. 6):

$$MAE = (1/n)\ \Sigma_i\ |y_i - \hat{y}_i| \quad (4)$$

$$RMSE = \sqrt{[(1/n)\ \Sigma_i\ (y_i - \hat{y}_i)^2]} \quad (5)$$

$$R^2 = 1 - [\Sigma_i(y_i - \hat{y}_i)^2\ /\ \Sigma_i(y_i - \bar{y})^2] \quad (6)$$

Las tres métricas se calcularon sobre el conjunto de prueba. Deliberadamente no se reporta una métrica de error agregada entre los dos objetivos del caso de destilación (costo y Eco-Indicador 99), dado que ambos se expresan en unidades distintas (USD/año frente a puntos/año); promediar sus errores absolutos produciría una cifra sin significado físico y podría sesgar la comparación entre modelos hacia la variable de mayor magnitud numérica.

3. Casos de estudio

Para evaluar el desempeño de los modelos sustitutos se consideraron dos casos de estudio con características físicas y operativas diferentes: una secuencia de destilación directa compuesta por dos columnas y un intercambiador de calor de tubos y coraza. La Figura 1 presenta una representación esquemática de ambos sistemas, permitiendo identificar sus principales características y las variables consideradas en cada caso.

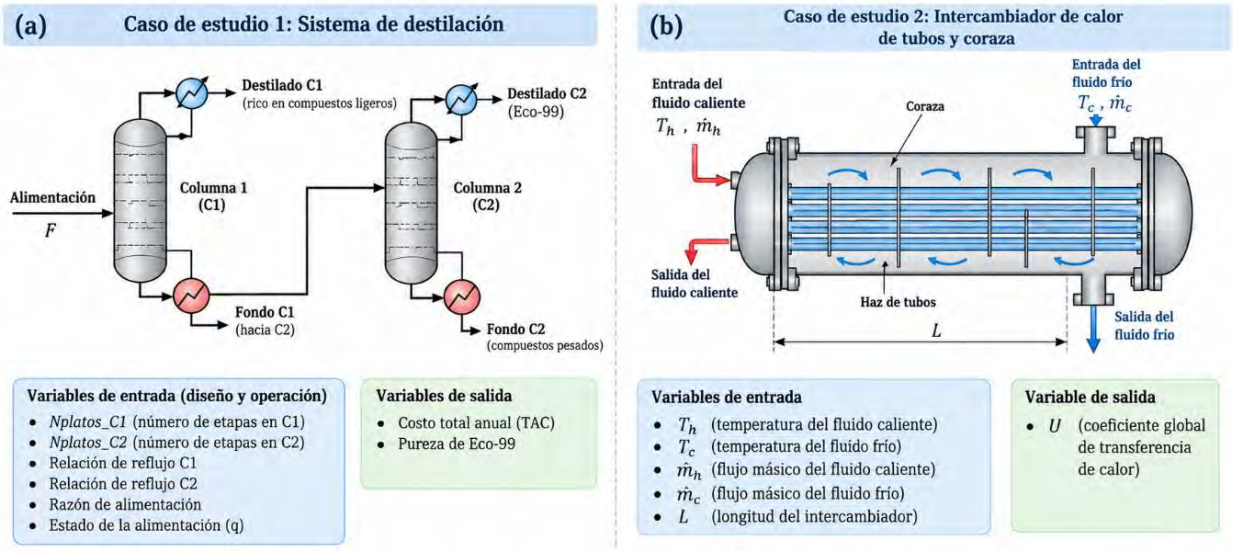


Figura 1. Representación esquemática de los casos de estudio: (a) sistema de destilación de dos columnas y (b) intercambiador de calor de tubos y coraza.

Número Especial "Verano de la Investigación DCNE 2026"
ISSN 2007-672X
Universidad de Guanajuato

3.1. Secuencia de destilación directa.

El primer caso de estudio reproduce la etapa de diseño de equipo de la metodología híbrida de Núñez-López y col. (2025): una secuencia directa de dos columnas de destilación convencionales que separan una mezcla ternaria de benceno (B), etilbenceno (EB) y dietilbenceno (DEB) —subproducto de la síntesis de etilbenceno a partir de benceno y etileno— en la primera etapa de una cadena de producción de poliestireno. En la columna C1, el benceno se recupera por el domo; los fondos, una mezcla de EB y DEB, alimentan la columna C2, donde el EB se recupera como destilado y el DEB como fondos (Figura 2).

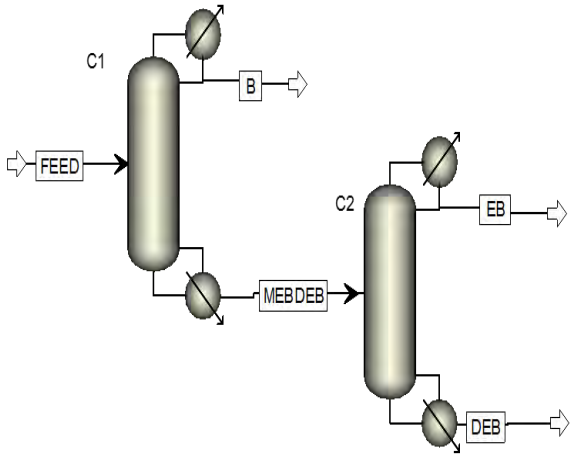


Figura 2. Secuencia directa de destilación: la columna C1 separa benceno (B) por el domo; la columna C2 recupera etilbenceno (EB) como destilado y dietilbenceno (DEB) como fondos. Adaptado de Núñez-López y col. (2025).

Cada columna se caracteriza por cinco variables de diseño: número de etapas, flujo de alimentación, relación de reflujo, flujo de destilado y diámetro; el conjunto de datos utilizado contiene 6,000 combinaciones de estas diez variables (cinco por columna), junto con el costo total anual del proceso (equipo, operación y distribución) y el Eco-Indicador 99, una métrica de impacto ambiental de ciclo de vida que pondera acero, vapor y electricidad consumidos por categoría de daño (Goedkoop y Spriensma, 2001), como variables de salida. La Tabla 1 resume el intervalo de valores observado para cada variable en el conjunto completo.

Un aspecto notable de la Tabla 1, retomado en la discusión, es que tres de las diez variables de entrada —la relación de reflujo y el flujo de destilado de la columna C1, y el flujo de destilado de la columna C2— varían menos de 0.05% de su media en las 6,000 combinaciones disponibles, mientras que el resto varía entre 10% y 20%. El costo total anual varía apenas 0.42% respecto a su media, y el Eco-Indicador 99 solo 0.09%. Aunque el número de etapas de una columna de destilación es físicamente una variable discreta, $Nplatos_{C1}$ y $Nplatos_{C2}$ se manejaron inicialmente como variables continuas durante la generación de las configuraciones

Número Especial "Verano de la Investigación DCNE 2026"
ISSN 2007-672X
Universidad de Guanajuato

de diseño. Antes de implementar cada configuración en el simulador riguroso, estos valores se redondearon hacia arriba al entero inmediato superior, garantizando que todas las simulaciones se realizaran con un número entero y físicamente realizable de etapas. En consecuencia, los datos utilizados para el desarrollo de los modelos sustitutos corresponden a configuraciones discretas, aun cuando estas variables fueron tratadas inicialmente como continuas.

Tabla 1. Intervalo de las variables de entrada y salida del conjunto de datos de destilación (n = 6,000). C1/C2: columnas 1 y 2.

Variable	Unidad	Mín.	Media	Máx.
$N_{platos,C1}$	—	50.0	54.9	60.0
F^{alim}_{C1}	kmol/h	30.0	35.0	40.0
$Reflujo_{C1}$	—	0.656	0.657	0.659
$Destilado_{C1}$	kmol/h	973.28	973.40	973.52
$Diámetro_{C1}$	m	0.55	0.60	0.65
$N_{platos,C2}$	—	15.0	17.5	20.0
F^{alim}_{C2}	kmol/h	5.0	8.7	12.5
$Reflujo_{C2}$	—	10.05	10.55	11.05
$Destilado_{C2}$	kmol/h	694.34	694.49	694.64
$Diámetro_{C2}$	m	0.45	0.50	0.55
Costo total anual	kUSD/año	1,731.8	1,732.9	1,739.1
Eco-Indicador 99	10^7 pts/año	1.98	1.98	1.99

3.2. Intercambiador de calor.

El segundo caso de estudio corresponde a un intercambiador de calor de tubos y coraza, en el cual se evaluó la influencia de las condiciones de operación y de una característica geométrica sobre el coeficiente global de transferencia de calor (U). Las variables de entrada consideradas fueron las temperaturas de los fluidos caliente (T_h) y frío (T_c), sus respectivos flujos másicos (m_h y m_c) y la longitud del intercambiador (L). Esta última se relaciona directamente con el área disponible para la transferencia de calor (Tabla 2).

Número Especial "Verano de la Investigación DCNE 2026"
ISSN 2007-672X
Universidad de Guanajuato

Tabla 2. Intervalo de las variables de entrada y salida del conjunto de datos del intercambiador de calor (n = 5,000). T_{hin} (°C), T_{cin} (°C): temperaturas de entrada, caliente/fría; m_h (kg/s), m_c (kg/s): flujos másicos; L (m): longitud; U (W/m^2K): coeficiente global (salida).

Variable	Unidad	Mín.	Media	Máx.
T_{hin}	°C	120.00	159.75	199.98
T_{cin}	°C	20.00	49.49	79.97
m_h	kg/s	1.001	4.51	7.999
m_c	kg/s	1.002	4.55	8.000
L	M	2.000	6.07	9.999
U	W/m^2K	165.71	438.55	717.38

A diferencia del caso de destilación, aquí las cinco variables cubren un intervalo amplio y comparable entre sí (coeficientes de variación entre 14% y 51%), lo que permite que un análisis de sensibilidad posterior (Sección 5.4) sea representativo del comportamiento del sistema en todo el dominio muestreado, y no solo en un entorno estrecho alrededor de un punto de diseño nominal.

4. Metodología

4.1. Preparación y partición de datos.

4.1.1. Para destilación.

Para el caso de destilación, los 6,000 registros disponibles se dividieron mediante una partición reproducible en tres subconjuntos: 3,840 observaciones (64%) para entrenamiento, 960 (16%) para validación y 1,200 (20%) para prueba. Inicialmente, el 20% de los datos se reservó como conjunto de prueba mediante una partición aleatoria con random_state = 42; posteriormente, el 20% del conjunto restante se destinó a validación utilizando una partición fija. Las diez variables de entrada y las dos variables de salida se estandarizaron mediante StandardScaler. Los parámetros de escalamiento se estimaron exclusivamente a partir del conjunto de entrenamiento y posteriormente se aplicaron, sin reajuste, a los conjuntos de validación y prueba. Las predicciones se transformaron a las unidades originales antes del cálculo de las métricas de error.

4.1.2. Para el intercambiador.

Para el caso del intercambiador de calor se empleó una partición de 70%, 15%, 15% para entrenamiento, validación y prueba, respectivamente. Las variables de entrada se

Número Especial "Verano de la Investigación DCNE 2026"
ISSN 2007-672X
Universidad de Guanajuato

estandarizaron mediante StandardScaler, mientras que la variable de salida (U) se conservó en sus unidades originales durante el entrenamiento y la evaluación. Los subconjuntos se mantuvieron fijos para la comparación de los diferentes modelos dentro de este caso de estudio.

4.2. Arquitectura y entrenamiento de los modelos.

Para ambos casos se implementó una red neuronal densa de tres capas ocultas con 64, 32 y 16 neuronas, activación ReLU, y una capa de salida lineal con tantas neuronas como variables de respuesta. La arquitectura 64–32–16 se seleccionó como una configuración común de capacidad progresivamente decreciente, con el propósito de proporcionar suficiente flexibilidad para representar relaciones no lineales sin incrementar innecesariamente el número de parámetros entrenables. Mantener la misma estructura de capas ocultas permitió además evitar que la arquitectura constituyera una fuente adicional de variación entre los casos analizados. No se realizó una búsqueda exhaustiva de hiperparámetros; por tanto, esta estructura debe interpretarse como una configuración de referencia y no como la arquitectura óptima para cada sistema.

La DNN del caso de destilación presenta una arquitectura 10–64–32–16–2 y contiene 3,346 parámetros entrenables, mientras que las dos regresiones lineales independientes requieren en conjunto 22 parámetros. Para el intercambiador de calor, la arquitectura 5–64–32–16–1 contiene 3,009 parámetros entrenables, frente a seis parámetros de la regresión lineal. En consecuencia, las DNN contienen aproximadamente 152 y 502 veces más parámetros ajustables que los correspondientes modelos lineales para destilación e intercambio de calor, respectivamente.

Aunque ambas DNN conservaron la misma estructura de capas ocultas, sus protocolos de entrenamiento no fueron completamente idénticos. En destilación se utilizó el optimizador Adam con tasa de aprendizaje de 0.001, MSE como función de pérdida, tamaño de lote de 16 y 300 épocas. Para el intercambiador de calor se utilizó Adam, MSE, tamaño de lote de 32 y detención temprana con paciencia de 20 épocas, restaurando los mejores pesos observados; el entrenamiento finalizó en la época 153 de un máximo de 300. En consecuencia, la comparación directa entre DNN y regresión lineal se realiza principalmente dentro de cada caso de estudio. (Kingma y Ba, 2015)

Número Especial "Verano de la Investigación DCNE 2026"
ISSN 2007-672X
Universidad de Guanajuato

Como modelo no lineal adicional para el intercambiador de calor se utilizó un Bosque Aleatorio implementado mediante RandomForestRegressor de scikit-learn. El modelo estuvo compuesto por 200 árboles de decisión, con profundidad máxima de 15 y semilla aleatoria de 42; los parámetros restantes conservaron los valores predeterminados de la implementación utilizada. El modelo se ajustó utilizando el conjunto de entrenamiento y se evaluó sobre el conjunto de prueba. La importancia relativa de las variables se obtuvo mediante feature_importances_, correspondiente a una medida basada en la disminución media de impureza (MDI), y se utilizó como medida de importancia predictiva, no de causalidad física.

4.3. Verificación independiente de resultados.

Con el propósito de cuantificar la variabilidad asociada con la naturaleza estocástica del entrenamiento de la DNN, el caso de destilación se entrenó de manera independiente 30 veces. En todas las ejecuciones se mantuvieron constantes las particiones de entrenamiento, validación y prueba, el procedimiento de estandarización, la arquitectura 10–64–32–16–2 y los hiperparámetros de entrenamiento. La modificación controlada entre ejecuciones correspondió a la semilla aleatoria, utilizando valores de 0 a 29.

Para cada ejecución se calcularon MAE, RMSE y R^2 sobre el mismo conjunto de prueba de 1,200 observaciones. El desempeño se caracterizó mediante la media, desviación estándar, mínimo y máximo de cada métrica. De esta manera, la evaluación del caso de destilación no depende de una única inicialización de la DNN y permite cuantificar explícitamente la variabilidad asociada con el entrenamiento.

5. Resultados

5.1. Desempeño predictivo: secuencia de destilación.

La Tabla 3 resume el desempeño de ambos modelos sobre el conjunto de prueba (1,200 diseños). Para el costo total, la DNN alcanzó un MAE promedio de 27.52 ± 8.28, un RMSE de 48.24 ± 15.06 y un R^2 de 0.9977 ± 0.0016. La regresión lineal presentó valores de 52.47, 80.92 y 0.9941, respectivamente. En términos del desempeño promedio, la DNN redujo el MAE aproximadamente un 47.6% y el RMSE un 40.4%, confirmando la ventaja del modelo no lineal para esta respuesta.

Para el Eco-Indicador 99 se observó el comportamiento contrario. La regresión lineal alcanzó un MAE de 13.28, un RMSE de 18.77 y un R^2 de 0.9999, mientras que la DNN presentó 27.07 ± 7.47, 40.38 ± 9.99 y 0.9996 ± 0.0002, respectivamente. Por tanto, la mayor complejidad de la DNN no produjo una mejora predictiva para esta respuesta.

Tabla 3. Desempeño predictivo de los modelos en el conjunto de prueba para el caso de destilación.

Modelo	Respuesta	MAE	RMSE	R^2
DNN	Costo Total	27.52±8.28	48.24±15.06	0.9977±0.0016
	Eco-Indicador 99	27.07±7.47	40.38±9.99	0.9996±0.0002
Regresión Lineal	Costo Total	52.47	80.92	0.9940
	Eco-Indicador 99	13.28	18.77	0.9999

La Figura 3 muestra las gráficas de paridad (valor predicho frente a valor real) para ambos modelos y ambos objetivos sobre el conjunto de prueba de esta réplica; en los cuatro paneles los puntos se ajustan estrechamente a la diagonal identidad, consistente con los coeficientes de determinación superiores a 0.99 reportados en la Tabla 3.

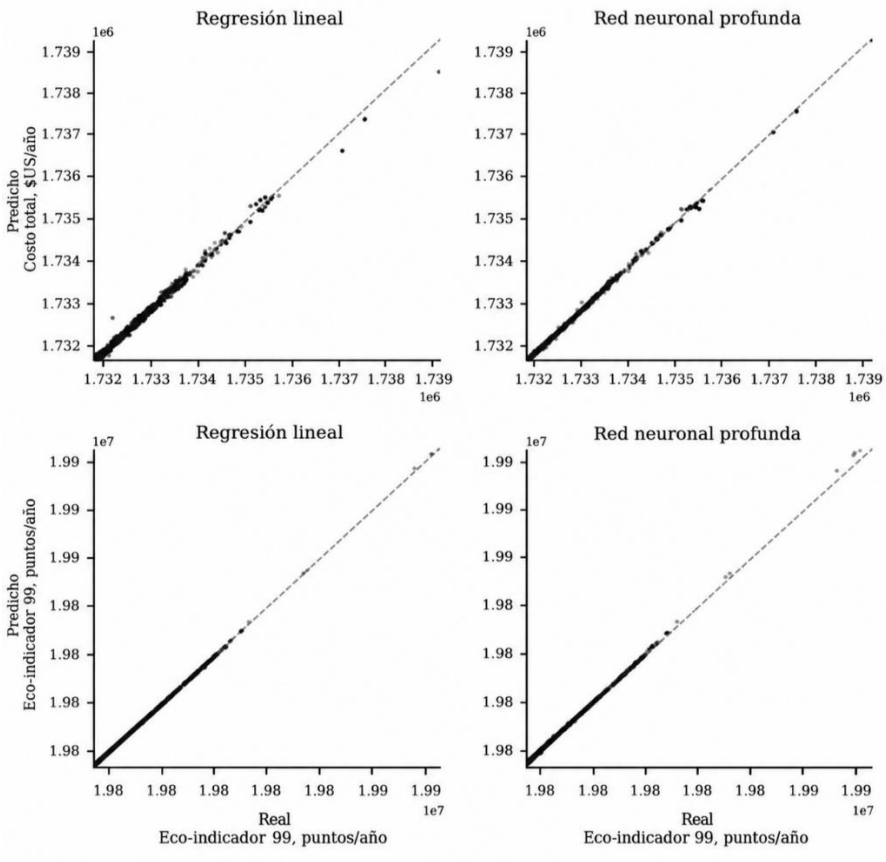


Figura 3. Gráficas de paridad (valor predicho vs. valor real) sobre el conjunto de prueba de la réplica independiente (Sección 4.3), para el costo total anual (arriba) y el Eco-Indicador 99 (abajo), con la regresión lineal (izquierda) y la red neuronal (derecha). La línea discontinua indica identidad perfecta.

5.2. Desempeño predictivo: intercambiador de calor.

En el caso del intercambiador de calor, la brecha entre modelos es mucho más marcada (Tabla 4). La red neuronal reduce el error absoluto medio en 84% respecto a la regresión lineal (4.88 frente a 29.90 W/m²K) y el coeficiente de determinación aumenta de 0.899 a 0.998. Como referencia adicional, un modelo de bosque aleatorio (300 árboles) entrenado sobre los mismos datos alcanzó un desempeño estadísticamente equivalente al de la red neuronal (MAE = 4.72 W/m²K, R^2 = 0.9979), lo que sugiere que la ganancia observada proviene de la capacidad de ambos modelos no lineales para capturar la forma funcional de U, y no de una característica particular de la arquitectura de la red neuronal profunda. Los resultados evidencian que la ventaja de la DNN depende de la estructura de la relación entrada–salida. Mientras que para el costo total el modelo no lineal reduce

Número Especial "Verano de la Investigación DCNE 2026"
ISSN 2007-672X
Universidad de Guanajuato

sustancialmente los errores, para el Eco-Indicador 99 la regresión lineal proporciona un desempeño superior. Este resultado muestra que incrementar la complejidad del modelo sustituto no garantiza una mejora predictiva y que un modelo de menor complejidad puede resultar suficiente cuando la estructura de la respuesta puede representarse adecuadamente mediante relaciones aproximadamente lineales.

Tabla 4. Desempeño de los modelos en el conjunto de prueba, caso del intercambiador de calor (n = 750), sobre U. MAE y RMSE en W/m²K.

Modelo	MAE	RMSE	R^2
Regresión lineal	29.90	40.99	0.8993
Red neuronal	4.88	6.02	0.9978
Bosque aleatorio	4.72	5.93	0.9979

5.3. Sensibilidad de la regresión lineal en destilación.

Al estar entrenados sobre variables estandarizadas, los coeficientes de la regresión lineal son directamente comparables entre sí y permiten identificar qué variables dominan cada objetivo (Tabla 5). El costo total anual está dominado por el número de etapas de la columna C1 (β = 0.671) y, en menor medida, por el número de etapas de C2 (β = 0.276); el Eco-Indicador 99, en cambio, está dominado casi por completo por la relación de reflujo de C1 (β = 0.797) y, en segundo lugar, por el flujo de destilado de C1 (β = 0.267) —coherente con que el mayor componente del Eco-Indicador 99 en columnas de destilación suele ser la carga energética del rehervidor, gobernada por la relación de reflujo. El resto de las variables tiene coeficientes estandarizados menores a 0.11 en valor absoluto para ambos objetivos.

Tabla 5. Coeficientes estandarizados (β) de la regresión lineal múltiple, caso de destilación. Magnitudes mayores indican mayor influencia relativa sobre el objetivo correspondiente.

Variable	β Costo	β Eco-99
$N_{platos,C1}$	0.685	0.013
F^{alim}_{C1}	0.090	−0.006
$Reflujo_{C1}$	0.095	0.793
$Destilado_{C1}$	0.031	0.270
$Diámetro_{C1}$	0.030	0.005
$N_{platos,C2}$	0.281	0.003
F^{alim}_{C2}	0.003	−0.001
$Reflujo_{C2}$	−0.006	−0.0001
$Destilado_{C2}$	−0.001	−0.002
$Diámetro_{C2}$	0.078	0.001

Para el costo total, la mayor magnitud corresponde al número de etapas de C1 ($\beta = 0.685$), seguido por el número de etapas de C2 ($\beta = 0.281$). Para el Eco-Indicador 99 se observa una estructura más concentrada: $Reflujo_{C1}$ presenta un coeficiente de $\beta =$

0.793, seguido por Destilado_{C1} con $\beta = 0.270$, mientras que las variables restantes presentan contribuciones considerablemente menores. Este comportamiento ayuda a explicar el elevado desempeño alcanzado por la regresión lineal para esta respuesta.

5.4. Importancia de variables en el intercambiador de calor.

Figura 4 muestra la importancia relativa de cada variable de entrada según el modelo de bosque aleatorio del intercambiador de calor. Los flujos másicos del fluido caliente y del fluido frío concentran, en conjunto, 99.9% de la importancia total (50.2% y 49.7%, respectivamente), mientras que las dos temperaturas de entrada y la longitud del equipo contribuyen, cada una, menos de 0.04%. Este patrón es consistente con correlaciones convectivas de tipo potencia, como la de Dittus-Boelter ($Nu = 0.023\ Re^{0.8}\ Pr^{n}$), en las que el número de Nusselt —y, por lo tanto, el coeficiente de película y U— depende fuertemente del número de Reynolds, proporcional al flujo másico, y solo débilmente de la temperatura a través de las propiedades del fluido (Çengel y Ghajar, 2020; Incropera y col., 2017). Dado que Re entra a esa correlación elevado a una potencia de 0.8 y no de forma lineal, un modelo lineal simple tiene una desventaja estructural para capturar esta relación, lo que explica la magnitud de la brecha observada en la Tabla 4. Esta ventaja de los modelos no lineales sobre correlaciones o ajustes lineales para predecir el desempeño térmico de intercambiadores de tubos ha sido reportada previamente en la literatura (Thanikodi y col., 2020), lo que refuerza que la brecha observada en este trabajo no es un artefacto del conjunto de datos particular empleado, sino un rasgo esperable de este tipo de sistemas.

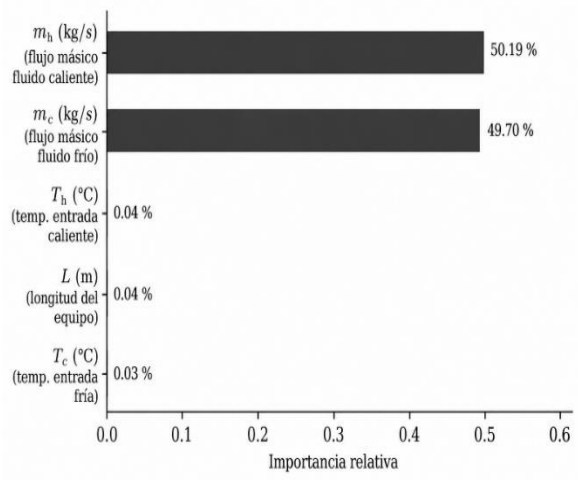


Figura 4. Importancia relativa de las variables de entrada obtenida mediante el Bosque Aleatorio del caso de estudio del intercambiador de calor. Las importancias corresponden a la disminución media de impureza (MDI) del modelo compuesto por 200 árboles, con profundidad máxima de 15, entrenado sobre el conjunto de entrenamiento, y se presentan normalizadas respecto a la importancia total.

Número Especial "Verano de la Investigación DCNE 2026"
ISSN 2007-672X
Universidad de Guanajuato

5.5. Identificación de un escenario de compromiso costo-ambiente.

Con el modelo de red neuronal ya entrenado, se buscó, dentro de los 1,200 diseños del conjunto de prueba, aquel que ofrece el mejor compromiso conjunto entre costo y Eco-Indicador 99. Es importante precisar el alcance de este ejercicio: no se trata de una optimización sobre el espacio continuo de variables de diseño, sino de una búsqueda exhaustiva sobre las predicciones de la red para los 1,200 puntos ya existentes en el conjunto de prueba, seleccionando el que minimiza la suma de ambos objetivos expresados en unidades estandarizadas. El diseño así identificado —Tabla 6— alcanza un costo total anual predicho de 1,732,505.5 USD/año y un Eco-Indicador 99 predicho de 19,839,884 puntos/año, ambos dentro del intervalo reportado en la Tabla 1. Este resultado debe interpretarse como una demostración de que el modelo entrenado puede utilizarse para explorar rápidamente los diseños ya evaluados —una tarea que tomaría fracciones de segundo frente a las horas de cómputo que exige el método híbrido original— y no como evidencia de que la red haya descubierto un diseño nuevo fuera del conjunto de datos proporcionado.

Tabla 6. Escenario candidato de compromiso costo–ambiente identificado entre los 1,200 diseños del conjunto de prueba mediante las predicciones promedio del ensamble de 30 DNN.

Objetivo	Valor de referencia	Predicción media ± DE	Error relativo (%)
Costo total (USD/año)	1,732,494.4	1,732,499.9 ± 11.7	0.000318
Eco-indicador 99 (puntos/año)	19,839,875.0	19,839,874.0 ± 21.4	0.000005

5.6. Robustez entre corridas independientes.

La Tabla 7 compara las métricas de la red neuronal de destilación reportadas en el análisis original con las obtenidas en la réplica independiente descrita en la Sección 4.3, ambas sobre el mismo conjunto de prueba. El coeficiente de determinación se mantuvo estable y superior a 0.998 para ambos objetivos en las dos corridas. El error absoluto medio del costo también fue consistente (21.97 frente a 20.15 miles de USD/año), pero el del Eco-Indicador 99 casi

Número Especial "Verano de la Investigación DCNE 2026"
ISSN 2007-672X
Universidad de Guanajuato

se duplicó entre corridas (23.73 frente a 43.37 puntos/año), pese a usar exactamente los mismos datos, partición, arquitectura e hiperparámetros. Esta variación es atribuible únicamente a la inicialización aleatoria de los pesos de la red, que no se controló mediante una semilla en ninguna de las dos corridas, y constituye evidencia directa de que una sola ejecución de una red neuronal —como la reportada de forma habitual en trabajos aplicados— puede no ser representativa de su desempeño típico, particularmente para el objetivo con menor variación absoluta en el conjunto de datos (Sección 3.1).

Tabla 7. Variabilidad del desempeño predictivo de la DNN para el caso de destilación obtenida a partir de 30 entrenamientos independientes.

Respuesta	Métrica	Media ± DE	Mín.	Máx.
Costo	MAE	27.52 ± 8.28	16.12	49.20
	RMSE	48.24 ± 15.06	30.77	91.19
	R^2	0.99770 ± 0.00162	0.99248	0.99914
Eco-99	MAE	27.07 ± 7.47	19.19	54.48
	RMSE	40.38 ± 9.99	28.46	73.69
	R^2	0.99964 ± 0.00021	0.99888	0.99983

6. Discusión

6.1. La ventaja de la red neuronal no es universal.

En la secuencia de destilación, la ventaja de la DNN se concentra en la predicción del costo total y se revierte para el Eco-indicador 99. Para esta última respuesta, la relación de reflujo de C1 presenta el coeficiente estandarizado de mayor magnitud (β=0.793, Tabla 5), considerablemente superior al de la mayoría de las variables restantes. Este predominio, junto con el elevado desempeño alcanzado por la regresión lineal, indica que la estructura entrada–salida del Eco-Indicador 99 puede representarse adecuadamente mediante el modelo lineal dentro del dominio analizado.

6.2. Un criterio práctico: diagnosticar antes de complejizar.

De lo anterior se desprende un criterio operativo para la selección de modelos sustitutos en ingeniería de procesos: antes de comprometerse con una red neuronal profunda —que exige más decisiones de diseño, mayor costo de entrenamiento y ofrece menor interpretabilidad—, conviene ajustar primero un modelo lineal y, opcionalmente, un modelo de árboles de bajo

Número Especial "Verano de la Investigación DCNE 2026"
ISSN 2007-672X
Universidad de Guanajuato

costo computacional (p. ej., un bosque aleatorio), y comparar su desempeño. Si ambos coinciden y el modelo lineal ya alcanza un error aceptable para la aplicación prevista, la complejidad adicional de una red neuronal puede no traducirse en beneficio práctico y sí en menor interpretabilidad y mayor sensibilidad a la inicialización (Sección 6.3). Si, por el contrario, el modelo de árboles supera claramente al lineal —como ocurrió en el intercambiador de calor—, existe evidencia de no linealidad que justifica invertir en una red neuronal, con la ventaja adicional de que las importancias de variables del modelo de árboles orientan qué relaciones físicas está aprendiendo el sustituto, útil tanto para depurar el modelo como para comunicar sus resultados a especialistas del proceso.

6.3. Un criterio práctico: diagnosticar antes de complejizar.

Los 30 entrenamientos independientes muestran que la naturaleza estocástica del entrenamiento de una DNN puede producir variaciones apreciables en las métricas de error aun cuando se mantienen constantes la arquitectura, los hiperparámetros y las particiones de datos. Sin embargo, los valores de R^2 permanecieron consistentemente elevados en todas las ejecuciones.

Estos resultados resaltan la importancia de no caracterizar el desempeño de una red neuronal a partir de una única ejecución. Cuando se emplean DNN como modelos sustitutos, particularmente para comparar modelos con diferencias relativamente pequeñas de desempeño, resulta recomendable realizar múltiples entrenamientos independientes y reportar medidas de tendencia central y dispersión.

6.4. Limitaciones.

Este trabajo presenta algunas limitaciones que acotan el alcance de sus conclusiones. En primer lugar, los conjuntos de datos fueron proporcionados previamente y la información disponible no permite reconstruir completamente su procedimiento de generación ni, en el caso del intercambiador de calor, todos los detalles físicos del sistema. Asimismo, varias variables del caso de destilación presentan intervalos relativamente estrechos, por lo que el elevado desempeño predictivo observado debe interpretarse dentro del dominio representado por los datos y no como evidencia de capacidad de extrapolación.

Número Especial "Verano de la Investigación DCNE 2026"
ISSN 2007-672X
Universidad de Guanajuato

Por otra parte, la arquitectura 64–32–16 se utilizó como una configuración común de referencia y no fue seleccionada mediante una búsqueda sistemática de hiperparámetros. Además, aunque ambas DNN conservaron esta arquitectura, sus protocolos de entrenamiento no fueron completamente idénticos, particularmente en el tamaño de lote y el uso de detención temprana. Finalmente, el análisis de variabilidad mediante 30 entrenamientos independientes se realizó únicamente para el caso de destilación. Estas consideraciones deberán abordarse en estudios posteriores mediante conjuntos de datos con mayor caracterización, protocolos de entrenamiento estandarizados y análisis de reproducibilidad para diferentes sistemas.

7. Conclusiones

Los resultados muestran que la conveniencia de utilizar una DNN como modelo sustituto depende de la estructura de las relaciones entrada–salida y no únicamente de su mayor complejidad. En la secuencia de destilación, la DNN redujo aproximadamente un 47.6% el MAE promedio para el costo total respecto a la regresión lineal (27.52 ± 8.28 frente a 52.47); sin embargo, para el Eco-Indicador 99 la regresión lineal presentó menor error (13.28 frente a 27.07 ± 7.47). En contraste, para el intercambiador de calor los modelos no lineales mostraron una ventaja clara: la DNN alcanzó un R^2 de 0.9978 y el Bosque Aleatorio de 0.9979, frente a 0.8993 para la regresión lineal.

Los 30 entrenamientos independientes realizados para el caso de destilación mostraron además que las métricas de una DNN pueden variar entre ejecuciones aun cuando se mantienen constantes las particiones, la arquitectura y los hiperparámetros, lo que respalda el uso de múltiples entrenamientos y el reporte de medidas de tendencia central y dispersión. Asimismo, las predicciones promedio de las 30 DNN permitieron identificar un escenario candidato de compromiso costo–ambiente entre los diseños previamente evaluados, sin que este resultado deba interpretarse como una nueva optimización continua.

En conjunto, los resultados respaldan el uso de modelos simples e interpretables como referencia inicial y la incorporación de modelos más flexibles cuando la estructura de los datos justifique su mayor complejidad. Como trabajo futuro, se recomienda ampliar la comparación a otros modelos y arquitecturas, estandarizar los protocolos de

entrenamiento y validar mediante modelos rigurosos de proceso los diseños seleccionados.

Declaración sobre el uso de herramientas de inteligencia artificial

Durante la preparación de este manuscrito, los autores emplearon el asistente de inteligencia artificial Claude (Anthropic) como herramienta de apoyo para la redacción y edición del texto borrador, la generación del código utilizado para reproducir de forma independiente el análisis de la Sección 4.3 y elaborar las Figuras 1 y 3, así como para la localización y verificación de las referencias bibliográficas citadas. Todo el contenido generado mediante esta herramienta fue revisado, verificado y editado por los autores, quienes asumen plena responsabilidad por la exactitud de los datos, el análisis e interpretación de los resultados y las fuentes citadas en el presente trabajo.

Contribución de los autores

Sebastián Ayala Palomares y Ángel Ramón Chávez Ortíz: investigación, análisis de datos, desarrollo de software, redacción del borrador original. Jesús Manuel Núñez López: conceptualización, supervisión, administración del proyecto, redacción — revisión y edición.

Referencias bibliográficas

Bhosekar, A., y Ierapetritou, M. (2018). Advances in surrogate based modeling, feasibility analysis, and optimization: A review. Computers & Chemical Engineering, 108, 250-267. https://doi.org/10.1016/j.compchemeng.2017.09.017

Bubel, M., Seidel, T., y Bortz, M. (2026). Reusable surrogate models for distillation columns. Computers & Chemical Engineering, 206, 109523. https://doi.org/10.1016/j.compchemeng.2025.109523

Çengel, Y. A., y Ghajar, A. J. (2020). Heat and mass transfer: Fundamentals and applications (6a ed.). McGraw-Hill.

Goedkoop, M., y Spriensma, R. (2001). The Eco-indicator 99: A damage oriented method for life cycle impact assessment (3a ed.). PRé Consultants.

Gómez Jurado, E. A. (2025). Técnicas de inteligencia artificial usadas en el control, seguridad y optimización de procesos químicos: una revisión sistemática de la

literatura [Trabajo de grado, Universidad del Valle]. https://hdl.handle.net/10893/36124

Goodfellow, I., Bengio, Y., y Courville, A. (2016). Deep learning. MIT Press.

Gutiérrez-Antonio, C., y Briones-Ramírez, A. (2015). Multiobjective stochastic optimization of dividing-wall distillation columns using a surrogate model based on neural networks. Chemical and Biochemical Engineering Quarterly, 29(4), 491-504. https://doi.org/10.15255/CABEQ.2014.2132

Hornik, K., Stinchcombe, M., y White, H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks, 2(5), 359-366. https://doi.org/10.1016/0893-6080(89)90020-8

Incropera, F. P., DeWitt, D. P., Bergman, T. L., y Lavine, A. S. (2017). Fundamentals of heat and mass transfer (8a ed.). Wiley.

Kakaç, S., Liu, H., y Pramuanjaroenkij, A. (2012). Heat exchangers: Selection, rating, and thermal design (3a ed.). CRC Press.

Kingma, D. P., y Ba, J. (2015). Adam: A method for stochastic optimization. En Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015). https://arxiv.org/abs/1412.6980

LeCun, Y., Bengio, Y., y Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539

Núñez-López, J. M., Segovia-Hernández, J. G., Sánchez-Ramírez, E., y Ponce-Ortega, J. M. (2025). Integrating metaheuristic methods and deterministic strategies for optimizing supply chain equipment design in process engineering. Chemical Engineering Research and Design, 214, 93-104. https://doi.org/10.1016/j.cherd.2024.12.021

Pal, P. K., Hens, A., Behera, N., y Lahiri, S. K. (2025). Digital twins: Transforming the chemical process industry—A review. The Canadian Journal of Chemical Engineering, 103(8), 3611-3636. https://doi.org/10.1002/cjce.25611

Ramírez Rodríguez, P. (2025). Desarrollo de modelos de deep learning para la implementación de soft sensor de variables en procesos industriales [Tesis de maestría, Universidad Politécnica de Madrid]. Archivo Digital UPM. https://oa.upm.es/89953

Schulz, M.-A., Yeo, B. T. T., Vogelstein, J. T., Mourao-Miranda, J., Kather, J. N., Kording, K., Richards, B., y Bzdok, D.

(2020). Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets. Nature Communications, 11, 4238. https://doi.org/10.1038/s41467-020-18037-z

Streb, A., y Mazzotti, M. (2022). Performance limits of neural networks for optimizing an adsorption process for hydrogen purification and CO2 capture. Computers & Chemical Engineering, 165, 107974. https://doi.org/10.1016/j.compchemeng.2022.107974

Thanikodi, S., Singaravelu, D. K., Devarajan, C., Venkatraman, V., y Rathinavelu, V. (2020). Teaching learning optimization and neural network for the effective prediction of heat transfer rates in tube heat exchangers. Thermal Science, 24(1B), 575-581. https://doi.org/10.2298/TSCI190714438T

Venkatasubramanian, V. (2019). The promise of artificial intelligence in chemical engineering: Is it here, finally? AIChE Journal, 65(2), 466-478. https://doi.org/10.1002/aic.16489 1

Ye, L., Zhang, N., Li, G., Gu, D., Lu, J., y Lou, Y. (2023). Intelligent optimization design of distillation columns using surrogate models based on GA-BP. Processes, 11(8), 2386. https://doi.org/10.3390/pr11082386