

Análisis comparativo entre un modelo de redes neuronales artificiales (RNA) y la metodología de superficie de respuesta (MSR) para la predicción de fallas en equipos industriales.

Comparative analysis between an artificial neural network (ANN) model and the response surface methodology (RSM) for predicting failures in industrial equipment.

A. Vélez Velázquez ^{a,*}, J. Medina Marín ^a, J. C. Seck Tuoh Mora ^a

^a Área Académica de Ingeniería y Arquitectura, Universidad Autónoma del Estado de Hidalgo, 42184, Pachuca, Hidalgo, México.

Resumen

Actualmente, el uso de tecnologías de la industria 4.0 se han convertido en una herramienta clave para la toma de decisiones en sectores industriales complejos, ya que permiten predecir fallas con mayor exactitud, optimizar el mantenimiento predictivo y reducir paros inesperados. En este trabajo se compara el rendimiento de dos modelos entrenados en el software MATLAB 2020®, utilizando algoritmos de aprendizaje automático para predecir fallas en equipos industriales: redes neuronales artificiales (RNA) y la metodología de superficie de respuesta (MSR). Se analizaron 7672 datos históricos obtenidos de sensores instalados directamente en turbinas, compresores y bombas, considerando variables como temperatura, presión, vibración, humedad, ubicación y equipo. El modelo de la MSR presenta un r^2 de 57.59% y un RMSE de 0.1637, por otro lado, la RNA con mayor rendimiento muestra un r^2 de 88.81% y un RMSE de 0.0958. La RNA presenta una precisión superior a la MSR del 41.5 %, lo que significa que se ajusta mejor a los datos.

Palabras Clave:

Industria 4.0, Mantenimiento predictivo, Predicción de fallas, Aprendizaje automático, Red neuronal artificial, Metodología de superficie de respuesta

Abstract

Currently, the use of Industry 4.0 technologies has become a key tool for decision-making in complex industrial sectors, as they allow for more accurate failure prediction, optimized predictive maintenance, and reduced unexpected downtime. This work compares the performance of two models trained in MATLAB 2020® software, using machine learning algorithms to predict failures in industrial equipment: artificial neural networks (ANN) and the response surface methodology (RSM). A total of 7,672 historical data points obtained from sensors installed directly on turbines, compressors, and pumps were analyzed, considering variables such as temperature, pressure, vibration, humidity, location, and equipment. The RSM model presents an r^2 of 57.59% and an RMSE of 0.1637, while the ANN with the highest performance shows an r^2 of 88.81% and an RMSE of 0.0958. The ANN has a higher accuracy than the MSR of 41.5%, which means it fits the data better.

Keywords:

Industry 4.0, Predictive maintenance, Failure prediction, Machine learning, Artificial neural network, Response surface methodology

1. Introducción

En el entorno industrial actual donde la eficiencia y confiabilidad operativa de los equipos son significativas, la predicción de fallas y el mantenimiento predictivo en equipos industriales es una estrategia clave. Estas técnicas permiten detectar y solucionar problemas antes de que ocurran; este tipo de mantenimiento permite una intervención planificada antes de que se produzcan paradas inesperadas por fallas generando costos altos para la organización (Carrera, 2024).

Uno de los mayores problemas a nivel industrial es la aplicación del mantenimiento correctivo, debido al impacto económico que conlleva en sus elevados costos de reparación y producción. Por otra parte, el mantenimiento preventivo si bien es

planificado por medio de una agenda para realizar un paro programado, no siempre se adapta a las condiciones reales del equipo, ya que no se considera si el dispositivo industrial aún puede trabajar o si sus componentes siguen en condiciones óptimas; el mantenimiento debe realizarse porque así fue programado (Martínez, 2021).

Los tipos de mantenimiento preventivo y correctivo permiten evitar y corregir fallas, ambos enfoques son útiles pero presentan limitaciones ya que dejan de lado el pronóstico hacia un futuro cercano o inmediato, es decir, se desconoce el tiempo o ciclo aproximado donde una parte de la maquinaria fallará o también cuando la falla está en la etapa inicial. En respuesta a estos acontecimientos surge el mantenimiento predictivo cuya función principal es monitorear continuamente el estado de los equipos

*Autor para la correspondencia: ve440672@uaeh.edu.mx

Correo electrónico: ve440672@uaeh.edu.mx (Allison Vélez Velázquez), jmedina@uaeh.edu.mx (Joselito Medina Marín), jseck@uaeh.edu.mx (Juan Carlos Seck Tuoh Mora)

industriales rotativos como turbinas, compresores, bombas, ventiladores, agitadores y mezcladoras. A través de sensores el mantenimiento predictivo registra parámetros operativos como amplitud, frecuencia y velocidad, con el objetivo de poder identificar patrones que indiquen el inicio de una falla y así estimar una "fecha aproximada" de intervención (Falla & Ortega, 2022).

En este sentido, el mantenimiento predictivo puede definirse como un conjunto de técnicas utilizadas para lograr pronosticar el punto de falla de un equipo o componente industrial. Gracias a la inteligencia artificial y el aprendizaje automático por medio de (Redes neuronales artificiales, RNA) y la estadística aplicada a la optimización de procesos (Metodología de Superficie de respuesta, MSR) hoy en día es posible predecir con el mayor porcentaje de efectividad y un margen de error mínimo el estado futuro de los sistemas industriales.

Gracias a la incorporación de tecnologías propias de la Industria 4.0, el mantenimiento predictivo ha alcanzado un nuevo nivel de evolución, conocido como PdM 4.0 (mantenimiento predictivo 4.0). Este enfoque permite la predicción de fallas futuras a partir de un análisis exhaustivo de diversas variables como el estado técnico del equipo, su historial de uso y mantenimiento, el entorno en el que se encuentra operando, así como datos comparativos con equipos similares presentes en otros lugares. La integración y análisis de estas variables facilita correlaciones más precisas respecto al rendimiento de los activos industriales (Arencibia, Anías, & Hernández, 2024).

Entre las principales ventajas de implementar el PdM 4.0 en las empresas se encuentra la obligatoriedad de realizar la programación de las acciones de mantenimiento antes de que ocurran serias averías, ya que al ser predictivo, el mantenimiento se realizará con más calidad y, además, permitirá que los fallos se detecten con suficiente antelación (Arencibia, Anías, & Hernández, 2024).

En el artículo "Prediction of Machine Failure in Industry 4.0: A Hybrid CNN-LSTM Framework" se propuso un modelo híbrido para PdM, evaluando tres enfoques: redes neuronales convolucionales (CNN), redes neuronales de memoria a largo plazo (LSTM) y un modelo híbrido CNN-LSTM. Las redes CNN forman parte de las redes neuronales artificiales (RNAs), su objetivo es identificar patrones espaciales dentro de los datos; al combinarse con modelos LSTM que capturan relaciones temporales permite reconocer tanto la estructura como la evolución en el tiempo de los datos, mejorando la capacidad predictiva de los modelos en entornos complejos. De los modelos evaluados, el híbrido CNN-LSTM fue el que obtuvo mejores resultados, mostrando un alto nivel de precisión y los errores más bajos en las métricas utilizadas (Wahid, Breslin, & Intizar, 2022). Los autores concluyen que la combinación de métodos híbridos dentro del marco de las RNAs puede mejorar significativamente la predicción de fallas en escenarios reales de la Industria 4.0.

En su estudio (Villegas & Pérez, 2023) señalan que los turbocompresores son fundamentales para la producción de gas natural en la industria petrolera, dichos equipos presentan fallas y baja disponibilidad debido a la ausencia del mantenimiento que permita aumentar su vida útil. Por esta razón, el autor decide implementar inteligencia artificial con el diseño y entrenamiento de un modelo de aprendizaje automático supervisado, utilizando el algoritmo de clasificación a partir de redes neuronales artificiales para predecir tres tipos de fallas (NORMAL, ALARMA y CRITICO). La arquitectura de la RNA toma en cuenta 7 variables de entrada, la capa oculta con 5 entradas y 3 variables de salida. Como resultado se obtiene una precisión del 99.965% en el modelo, y con ello afirma que la técnica resultó favorable ya que ayudó a implementar soluciones en la toma de

decisiones respecto a fallas detectadas y planificación del mantenimiento predictivo en los equipos turbocompresores.

Por otra parte (Younes, Álvarez, & Vázquez, 2021) describe que los álabes son el principal origen de fallas y averías de las turbinas de vapor, a partir de lo anterior, desarrolló un modelo de redes neuronales artificiales (RNA) utilizando 8 variables de entrada, 40 neuronas y una variable de salida para predecir la vida útil de los álabes. La RNA se entrenó con el algoritmo Backpropagation en el software Matlab a partir de la herramienta neural networks toolbox (RNA). El autor dio a conocer un resultado positivo al aplicar una prueba estadística de regresión lineal obteniendo un nivel de confianza del 99%; (Younes, Álvarez, & Vázquez, 2021) agregaron que el entrenamiento y uso de redes neuronales son capaces de predecir y simular eficazmente el análisis de fallas de ingeniería y estimar el ciclo de vida de álabes en turbinas de vapor al monitorear variables de comportamiento.

Más aún (Acosta, Ruiz, Prieto, Orozco, & Valdés, 2023) probaron distintas técnicas de aprendizaje automático para detectar fallas intermitentes en un entorno industrial simulado con Factory I/O. Dentro de los modelos que evaluaron, utilizaron redes neuronales multicapa entrenadas con datos de sensores y motores. Aunque estas redes no superaron en precisión a modelos como RandomForest o Árboles de Decisión, ofrecieron buenos resultados en escenarios con varios tipos de falla, alcanzando una precisión del 97.57%. Los autores señalan que, si bien otros métodos pueden ofrecer un mejor desempeño en ciertos casos, las redes neuronales continúan siendo una opción confiable para el mantenimiento predictivo, especialmente cuando se afinan sus parámetros y se eliminan variables irrelevantes.

Por otro lado (Onokwai, y otros, 2024), realizaron un estudio basado en datos para la predicción de fallas en máquinas de la industria de bebidas mediante la Metodología de Superficie de Respuesta (MSR) y el Sistema de Inferencia Neurodifusa Adaptativa (SNAFIS). Generaron un diseño experimental con la metodología de Superficie de Respuesta (MSR) a partir del diseño compuesto central (DCC) con el software estadístico Design-Expert versión 13.0. Con lo anterior, desarrollaron un modelo matemático a partir de la de MSR, con el cual analizaron parámetros operativos como: presión, voltaje, corriente, vibración y temperatura con el propósito de identificar cuales afectan la incidencia de fallas. (Onokwai, y otros, 2024) Concluyen que el modelo tiene alta precisión al obtener ($R^2=98.22\%$) y una significancia estadística (valor $p<0.05$), demostrando así confiabilidad en la predicción de fallas y optimizando efectivamente las relaciones de parámetros a través de modelos de MSR y SNAFIS.

Con respecto a la toma de decisiones, esta acción representa la etapa final del proceso dentro de la estrategia para la implementación del mantenimiento predictivo. El autor (Pinzón, 2022) numera una gran variedad de métodos con capacidad para establecer criterios de mantenimiento y los clasifica en cuatro categorías según el tipo de investigación a realizarse para su cálculo o análisis correspondiente en trabajos futuros: sistemas basados en la experiencia, sistemas basados en datos, sistemas basados en la física del proceso y sistemas híbridos.

De acuerdo con la clasificación de (Pinzón, 2022), este trabajo utiliza la categoría número dos: *sistemas basados en datos*, la cual se enfoca en analizar datos numéricos recolectados por sensores (routers) instalados en el sistema de los equipos. El análisis de los datos tiene como objetivo apoyar la toma de decisiones en torno al mantenimiento predictivo de equipos industriales, con base en la predicción de fallas. Por lo tanto, se consideran variables físicas como la temperatura, presión, vibración y humedad, registradas en turbinas, compresores y bombas instalados en cinco ciudades de Estados Unidos: Atlanta, Chicago, San Francisco, Houston y

Nueva York. Además del comportamiento de los equipos, también se considera la posible influencia de la variable geográfica y ambiental (ubicación de los equipos), así como la variable categórica (tipo de equipo), en la predicción de fallas, con el fin de mejorar la planificación del mantenimiento predictivo.

En este artículo se propone entrenar dos modelos uno de redes neuronales artificiales (RNA) y otro a partir de la metodología de superficie de respuesta (MSR) para determinar cuál de los dos enfoques ofrece mejores resultados en términos de optimización del mantenimiento predictivo y predicción de fallas en equipos rotativos industriales, se consideran 7672 datos recopilados del conjunto **Industrial Equipment Monitoring Dataset** (Samudrala, 2024) disponible en la plataforma Kaggle, sitio web de ciencia de datos e inteligencia artificial.

A pesar de que ya se han realizado diversos estudios en los que se aplican tanto modelos de Redes neuronales artificiales (RNA) así como métodos estadísticos, como la Metodología de superficie de respuesta (MSR), para la predicción de fallas en equipos industriales, son muy pocos los que realmente comparan ambos enfoques bajo las mismas condiciones. En consecuencia, se deja abierta una brecha importante al momento de establecer qué modelo resulta más útil o eficiente para planificar el mantenimiento predictivo, el cual depende en gran medida de una predicción de fallas precisa. Por lo tanto, este trabajo propone un análisis comparativo entre ambas técnicas, utilizando el mismo conjunto de datos para su formulación y entrenamiento y los mismos criterios de evaluación, con el objetivo de identificar cuál ofrece mejores resultados en escenarios reales. A partir de los resultados obtenidos, se espera brindar una referencia clara que contribuya a mejorar la toma de decisiones en torno a la implementación de estrategias más efectivas dentro del mantenimiento predictivo en la industria.

2. Conceptos básicos

2.1. Fundamentos del mantenimiento industrial

2.1.1 Concepto e importancia del mantenimiento industrial

El concepto de mantenimiento industrial comprende una serie de acciones que deben realizar los usuarios responsables del departamento o área, con el objetivo de que los equipos, máquinas, componentes e instalaciones involucradas en un proceso industrial se encuentren en las condiciones de funcionamiento requeridas para lo cual fueron diseñados, construidos, instalados y puestos en operación (Arcos, Lizarzaburu, Marín, & Arcos, 2023).

Así mismo, la importancia del mantenimiento industrial reside en que para la industria, es fundamental mantener el equipo, la maquinaria y los componentes operando de forma continua y eficiente, por lo tanto, el pronosticar el tiempo de la falla en un equipo puede ayudar a programar de forma más eficiente el mantenimiento de equipos industriales (Villanueva, González, Blanco, Buenabad, & Carrera, 2024).

2.1.2 Ingeniería de mantenimiento

De acuerdo con (Hernández, 2005) la ingeniería de mantenimiento es fundamental en la industria, ya que tiene la responsabilidad de establecer procedimientos, métodos, técnicas y líneas de investigación para la aplicación de planes de mantenimiento, tomando en cuenta oportunidades del presente y futuro a partir de estrategias de actualización, innovación y mejora continua en áreas de mantenimiento. Además, se enfoca en el análisis estadístico y la evaluación del mismo con el propósito de anticipar fallas mediante técnicas predictivas.

El objetivo final es aumentar la confiabilidad a través de estrategias y tendencias tecnológicas, elaboración de procedimientos de mantenimiento, formulación y evaluación de proyectos que deberán ejecutarse.

2.1.3 Tipos de mantenimiento

Como lo establece (Garrido, 2003) al día de hoy no es posible tener una planta sujeta a un tipo de mantenimiento debido a que cada equipo tiene características propias que lo hacen diferente del resto, incluso de otros equipos similares. El mantenimiento principalmente se clasifica en tres tipos: preventivo, correctivo y predictivo, como se muestra en la tabla 1.

Tabla 1. Tipos de mantenimiento

Tipo de mantenimiento	Descripción
Mantenimiento correctivo	Conjunto de tareas propuestas para corregir los defectos que se van presentando en los equipos industriales. Su objetivo es mantener un buen nivel de servicio en los equipos, programando las correcciones de puntos vulnerables en el momento oportuno.
Mantenimiento preventivo	Es el más tecnológico, persigue conocer e informar permanentemente el estado y operatividad de los equipos a partir variables físicas cuya variación sea indicativa de problemas que puedan estar apareciendo en el equipo.

Fuente: (Garrido, 2003)

2.2 Industria 4.0 y mantenimiento predictivo

2.2.1 Introducción a la industria 4.0

Según (Luna Pérez, Badillo Piña, & Vázquez Álvarez, 2020) la industria 4.0 es considerada la cuarta revolución industrial al combinar técnicas avanzadas de producción y operaciones utilizando tecnologías inteligentes que se integran en las organizaciones empresariales. Hace referencia a una industria donde las máquinas, equipos y elementos de producción están digitalizados e interconectados, permitiendo el manejo de cantidades inmensas de datos conocidos como Big Data.

Con la llegada de la cuarta revolución industrial se tiene en mente la perspectiva de la fábrica del futuro o la fábrica inteligente. Esta visión se construye a través de transformación digital y a la incorporación de tecnologías avanzadas como las que se muestran en la figura 1, permitiendo a las empresas adaptarse, crecer y responder mejor a las necesidades del presente y del futuro.

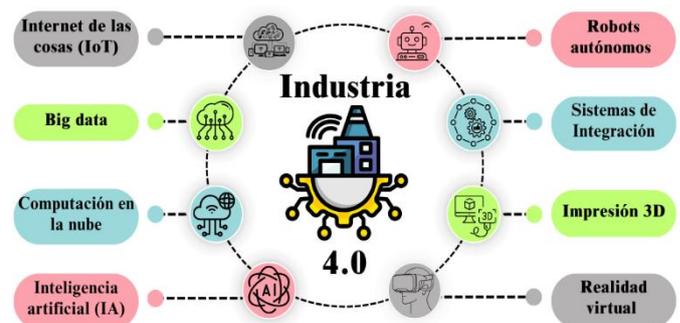


Figura 1. Tecnologías habilitadoras de la industria 4.0

Fuente: Elaboración propia

2.2.2 Tecnologías de la industria 4.0 habilitadas al PdM

La industria 4.0 reúne una serie de avances tecnológicos, según la clasificación realizada por (Vázquez, 2023) y mostrada en la tabla 2, estos avances son ideales o propios del enfoque de PdM 4.0. Abren la puerta a nuevas formas de trabajo más eficientes y proactivas. Son herramientas clave para desarrollar un modelo que facilite la predicción de fallas y la planificación del mantenimiento de manera anticipada, ayudando así a prevenir problemas antes de que ocurran.

Tabla 2. Tecnologías de la industria 4.0

Tecnología	Descripción
Internet de las cosas (IOT)	Incorporación de sensores para la recolección de información en cualquiera de las etapas de producción.
Big data	Análisis de datos en grandes volúmenes, para la mejora de procesos, productos, predicción de fallas y anticipación de operaciones de mantenimiento.
Inteligencia artificial	Desarrollo de aplicaciones a partir de algoritmos de aprendizaje automático (machine learning), utilizando grandes cantidades de datos como insumo, y los de aprendizaje profundo (deep learning), que jerarquizan los algoritmos en redes neuronales para ser utilizados en distintos contextos.

Fuente: (Vázquez, 2023)

2.2.3 Mantenimiento predictivo 4.0

En el contexto de la industria 4.0, (Villanueva, González, Blanco, Buenabad, & Carrera, 2024) definen al PdM como el conjunto de acciones y técnicas aplicadas en el sector industrial, haciendo uso de sensores IoT, inteligencia artificial y big data a partir del monitoreo y análisis de datos en tiempo real cuyo objetivo es predecir fallas y obtener la reducción del tiempo de inactividad no planificada. La estrategia de PdM permite a los ingenieros detectar y corregir la falla en su etapa inicial para evitar que los equipos se averíen y se pueda prolongar su vital útil.

2.2.4 Predicción de fallas como base del PdM

Como señala (Yan, 2021), la predicción de fallas considera enfoques de aislamiento e identificación; el aislamiento de fallas se refiere a la localización del componente defectuoso; y la identificación de fallas determina la naturaleza de la falla cuando se detecta. Gracias a los avances tecnológicos en áreas como la minería (DM), transmisión y base de datos es posible realizar pronósticos de fallas, los cuales se encargan de predecir fallas antes de que ocurran, lo anterior, también permite indicar si algo está mal en el sistema monitoreado. Por lo tanto, la predicción de fallas es una labor para determinar si una falla es inminente y estimar qué tan pronto y qué tan probable es que pueda ocurrir. Este enfoque es fundamental para el PdM ya que ayuda a anticiparse a posibles fallas y permite programar acciones de mantenimiento en el momento preciso. De esta manera, se asegura que los equipos sigan funcionando correctamente, se evitan paros inesperados y se aprovechan mejor los recursos disponibles.

2.3 Técnicas de inteligencia artificial aplicadas al PdM

2.3.1 Aprendizaje automático (Machine Learning)

Como lo explica (Guartazaca, 2024) el aprendizaje automático, conocido como machine learning (ML) es parte de la inteligencia artificial, su objetivo es dar a las máquinas la capacidad de aprender de la experiencia creando algoritmos con datos de entrenamiento, los cuales se comprendan y optimicen de forma autónoma utilizando redes neuronales artificiales y aprendizaje profundo. El ML toma parámetros de entrada, para convertirlos en salida, como se muestra en la figura 2, lo anterior, con el propósito de realizar predicción de fallas y tomar decisiones respecto al PdM sin necesidad de una programación minuciosa.

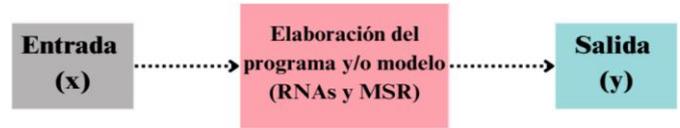


Figura 2. Programación tradicional del machine learning

Fuente: Elaboración propia

2.3.2 Redes neuronales artificiales

Las Redes Neuronales Artificiales (RNA) son modelos informáticos que toman inspiración del funcionamiento del cerebro humano y se emplean para el aprendizaje automático y el reconocimiento de patrones en un conjunto de datos (Echeverría & Nivelá, 2024).

De acuerdo con (Yandar & Moreno, 2019) las RNAs son construidas a partir de elementos llamados nodos, unidades, o neuronas y se componen de capas de entrada (x) (input layers), capas ocultas (hidden layers) y capas de salida (y) (output layers) como se observa en la figura 3. Las capas de entrada son las encargadas de la recepción de datos para alimentar la red neuronal y estos se transfieren a las capas ocultas que tienen diferentes tipos de operaciones matemáticas que procesan los datos, por último, las capas de salida poseen generalmente una función de activación que ayuda a determinar el tipo de dato que se está procesando.

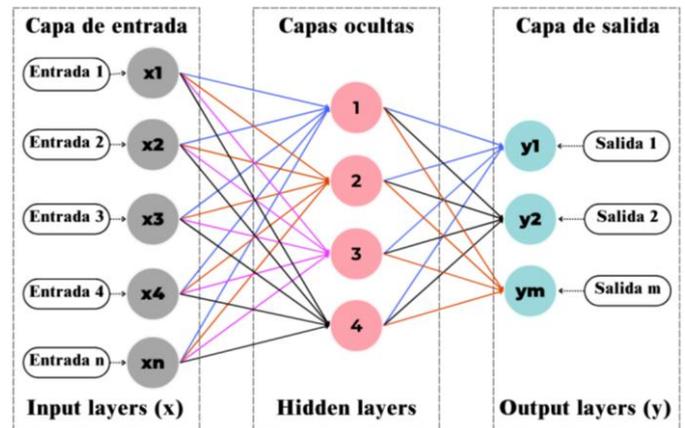


Figura 3. Modelo de red neuronal artificial

Fuente: Elaboración propia

2.3.2.1 Predicción de fallas con RNAs

Para (Rangel, 2024) el uso de redes neuronales surge como una herramienta eficaz, ofreciendo mayor precisión en la detección de comportamientos anómalos a través del análisis de datos históricos y en tiempo real, en la actualidad, es indispensable detectar anomalías o fallas tempranamente en equipos industriales. Existen algunos métodos como los de monitoreo y mantenimiento empleados habitualmente, estos resultan efectivos, pero, presentan ciertas limitaciones en la identificación de patrones complejos o sutiles que puedan anteceder a una falla.

2.3.2.2 RNAs y mantenimiento predictivo

Como señala (Rangel, 2024) los algoritmos de inteligencia artificial y machine learning tienen relevancia en avances tecnológicos y científicos, de igual forma, estas herramientas son aplicadas en educación, medicina, ingeniería, etc. como solución o caso de estudio a problemas en diferentes áreas del sector industrial. En este caso específico, resaltar la importancia de las redes neuronales en el campo del mantenimiento predictivo es primordial, debido a su capacidad para manejar un gran volumen de datos también llamados “Big Data”, basándose en el comportamiento del cerebro humano, permitiendo así la detección de fallas en equipos, máquinas y elementos industriales con el propósito de optimizar recursos y mejorar la toma de decisiones.

2.3.3 Metodología de superficie de respuesta

La metodología de superficie de respuesta (MSR) es una técnica estadística que surge como un recurso sofisticado para la optimización y el modelado de sistemas complejos a partir de un conjunto de datos o registros numéricos. Este método destaca por su utilidad en la experimentación con múltiples variables y niveles, lo que permite comprender detalladamente el impacto de estos factores en la predicción de fallas en equipos industriales y múltiples escenarios de aplicación (Toapanta, Paredes, Meneses, & Salinas, 2024).

Según (Alkharisi & Dahish, 2025) la MSR es capaz de admitir datos históricos y modelos definidos por el usuario. El tipo de datos y el número de variables determinan qué modelo se debe utilizar. Este método se puede utilizar para desarrollar modelos y analizar una matriz de experimentos preparada previamente mediante un enfoque definido por el usuario. Con los modelos creados con la MSR en algunos casos, se obtienen gráficos de respuesta tridimensionales como resultado del entrenamiento como el de la figura 4, los cuales pueden discutirse e interpretarse de acuerdo al problema y/o modelo planteado.

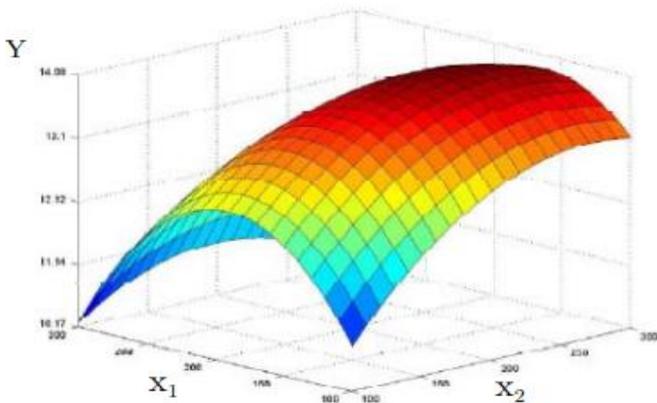


Figura 4. Superficie de respuesta tridimensional
Fuente: (Alkharisi & Dahish, 2025)

Los autores (Alkharisi & Dahish, 2025) afirman que la MSR es utilizada para diseñar, simular, evaluar, desarrollar y mejorar procesos de optimización a partir de enfoques estadísticos y matemáticos avanzados, por lo tanto, es considerada como un método de optimización multiobjetivo para situaciones que buscan un objetivo de respuesta y variable.

2.3.3.1 Predicción de fallas con MSR

En palabras de (Onokwai, y otros, 2024) la MSR es una técnica estadística y matemática utilizada para modelar y optimizar

procesos complejos ya que ha sido aplicada con éxito en industrias manufactureras para analizar el impacto de parámetros operativos críticos en la ocurrencia de fallas. Lo anterior permite comprender que al crear un modelo matemático, la MSR permite investigar las relaciones entre múltiples variables y su impacto en la frecuencia de fallas.

Por lo tanto, se concluye que el objetivo principal de aplicar MSR en el contexto industrial es identificar las mejores condiciones operativas que minimicen la probabilidad de fallas a partir de una planificación minuciosa de los experimentos, debido a que la MSR permite evaluar los efectos individuales e interactivos de diversos equipos industriales.

2.3.3.2 MSR y mantenimiento predictivo

Como lo explican (Onokwai, y otros, 2024) dentro del mantenimiento predictivo se utiliza el aprendizaje automático para crear un modelo a partir de datos históricos y poder evaluar patrones de fallas con datos en tiempo real. En este contexto, el generar un diseño experimental a partir de la metodología de superficie de respuesta (MSR) es una herramienta clave de modelado matemático ya que permite ajustar de manera precisa las variables operativas (x) y su efecto sobre las variables de respuesta (y) dentro de un entorno de modelado.

Por lo tanto, el PdM con aprendizaje automático busca optimizar el uso de los recursos y predecir fallas antes de que ocurran, ya que las prácticas conservadoras desperdician recursos. Por lo tanto, el mantenimiento predictivo busca mantener un equilibrio entre la prevención excesiva y la inacción.

2.4 Desarrollo experimental

La capacidad de la inteligencia artificial para analizar grandes conjuntos de datos y generar modelos predictivos precisos a partir del aprendizaje automático se manifiesta como un recurso valioso dentro de la industria.

En esta investigación, se hará uso del software de aprendizaje automático MATLAB 2020®, ya que posee la flexibilidad de implementar algoritmos de inteligencia artificial de manera eficiente como RNA y MSR, con el objetivo de optimizar los cálculos y la solución de problemas complejos dentro de la industria y otros campos disciplinarios. En la figura 5 se muestra el diagrama de fases y nombre de cada una de ellas utilizado en esta investigación para la elaboración y entrenamiento de ambos modelos de predicción.

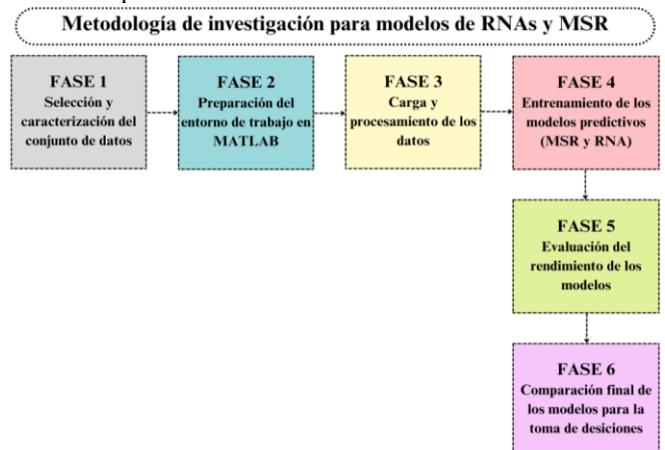


Figura 5. Diagrama de fases para la elaboración y entrenamiento de los modelos de RNA y MSR

Fuente: Elaboración propia

2.4.1 Primera fase

A partir de una investigación exhaustiva en el sitio web Kaggle, se empleó el conjunto “Industrial Equipment Monitoring Dataset” (Samudrala, 2024), disponible en el siguiente enlace:

<https://www.kaggle.com/datasets/dnkumars/industrial-equipment-monitoring-dataset/data>, el cual contuvo un total de 7672 registros históricos correspondientes a seis variables independientes (x) y una variable dependiente (y); cuya clasificación se muestra en la tabla 3. Las variables mencionadas se utilizaron para la elaboración y entrenamiento de los modelos de RNA y MSR, por lo tanto, fue indispensable identificar las variables de entrada (x) y variables de salida (y).

Tabla 3. Tipo de variables experimentales del caso de estudio

Variables independientes o de entrada		
x_1	Temperatura	Lectura de temperatura en el momento de la observación (en °C).
x_2	Presión	Lectura de presión en el momento de la observación (en bar).
x_3	Vibración	Lectura del nivel de vibración (unidades normalizadas).
x_4	Humedad	Porcentaje de humedad registrado en la ubicación del equipo.
x_5	Equipo	Tipo de equipo industrial que se está monitoreando (1. Turbina, 2. Compresor y 3. Bomba).
x_6	Ubicación	Ubicación del equipo en cinco ciudades de Estados Unidos (1. Atlanta, 2. Chicago, 3. San Francisco, 4. Houston y 5. New York).
Variable dependiente o de salida		
y	Estado del equipo	Indicador binario (0 = No defectuoso, 1 = defectuoso) para especificar si el equipo está funcionando correctamente o requiere mantenimiento.

Fuente: Elaboración propia, tomando como base el conjunto de datos “Industrial Equipment Monitoring Dataset” del sitio Kaggle (Samudrala, 2024).

Las variables independientes descritas en la tabla 3 intervinieron en la predicción de fallas de tres equipos rotativos industriales: turbinas, compresores y bombas; por lo tanto, el entrenamiento de los modelos tuvo como objetivo determinar el estado final de los equipos a partir del comportamiento de los valores experimentales. Resulta indispensable señalar que las variables x_5 y x_6 presentaban texto, por lo que fueron transformadas en valores numéricos o indicadores binarios, con la finalidad de que el software MATLAB 2020® pudiera procesarlas de manera correcta durante el entrenamiento de ambos modelos. La tabla 4 muestra algunos de los datos experimentales que se utilizaron para el entrenamiento de los modelos de RNA y MSR.

Tabla 4. Datos experimentales para la detección de fallas

No.	x_1	x_2	x_3	x_4	x_5	x_6	y
1	58.18	25.03	0.61	45.69	1	1	0
2	75.74	22.95	2.34	41.87	2	2	0
3	71.36	27.28	1.39	58.95	1	3	0
4	71.62	32.24	1.77	40.57	3	1	0
5	66.51	45.20	0.35	43.25	3	5	0
6	49.51	76.43	4.81	20.27	2	5	1
7	66.90	45.53	1.55	64.80	2	4	0
...							

7670	72.03	34.76	1.71	49.97	3	1	0
7671	107.0	23.75	1.14	23.97	2	1	1
7672	28.47	66.59	4.68	79.67	1	2	1

Fuente: Elaboración propia, tomando como base el conjunto de datos “Industrial Equipment Monitoring Dataset” del sitio Kaggle (Samudrala, 2024).

2.4.2 Segunda fase

Se estableció el método de importación de la base de datos al software MATLAB 2020® para gestionar las variables de entrada y salida en el espacio de trabajo, posteriormente, se determinó la librería y caja de herramientas a utilizarse en cada uno de los modelos a evaluar. Para el caso del modelo de redes neuronales artificiales (RNA), se aplicó la librería *Deep Learning Toolbox*, la cual contiene funciones para el desarrollo y entrenamiento de redes neuronales profundas y con la metodología de superficie de respuesta (MSR), se utilizó la librería *Statistics and Machine Learning Toolbox*, compuesta por herramientas para análisis estadístico, modelado predictivo y aprendizaje automático.

2.4.3 Tercera fase

Para entrenar la base de datos se cargó la base de datos al software MATLAB 2020® en formato de Microsoft Excel (.xlsx) dejando de lado palabras, encabezados, etc., presentes en la tabla o libro de Excel, haciendo uso solo de datos numéricos adecuados a los algoritmos de ML (machine learning) o aprendizaje automático.

2.4.4 Cuarta fase

Los 7672 datos históricos fueron entrenados en el software MATLAB 2020®, tomando en cuenta dos modelos de predicción, RNA y MSR con el objetivo de identificar cuál de los dos se ajusta mejor a los datos experimentales y ofrece predicciones más confiables.

2.4.4.1 Metodología de superficie de respuesta

Los 7672 datos históricos, obtenidos a partir del monitoreo en tiempo real de tres tipos de equipos industriales, recopilaron parámetros como temperatura, presión, vibración y humedad en diversas ciudades de Estados Unidos. Esta información se empleó para construir un modelo basado en la MSR, con el propósito de analizar el rendimiento de los equipos e identificar posibles fallas, a partir del comportamiento de las seis variables independientes o de entrada.

2.4.4.1.1 Construcción y validación del modelo

Los datos se ajustaron y entrenaron en el software MATLAB 2020®, utilizando el comando o caja de herramientas (`rstool`). La variable de respuesta, que representa el estado del equipo o la predicción de falla, se modeló como una ecuación polinomial, como se muestra en la ecuación 1, basada en la propuesta de (Serna, Tellez, Seck-Tuoh-Mora, Hernández, & Medina, 2024). Este modelo es de tipo cuadrático, ya que incluye términos lineales, términos de segundo orden (cuadráticos) y términos de interacción entre variables independientes, por lo tanto, permitió describir el comportamiento de cada variable en sus respectivos puntos experimentales y predecir resultados con base en los patrones observados.

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{j=2}^k \sum_{i=1}^{j-1} \beta_{ij} x_i x_j + \sum_{i=1}^k \beta_{ii} x_i^2 + e \quad (1)$$

Donde y representa el estado del equipo y/o predicción de falla, β_0 es el término constante de la ecuación, β_i el coeficiente del término lineal, β_{ij} el coeficiente del término de interacción entre las variables, β_{ii} el coeficiente del término cuadrático; x_i y x_j son los índices que recorren las variables independientes, de 1 a k , e establece el error observado y k es el número total de variables independientes, en nuestro caso $k = 6$.

Para validar el ajuste del modelo obtenido mediante la MSR, cuyo propósito fue establecer el estado del equipo y/o predicción de falla, se determinó a partir del análisis de varianza (ANOVA). Esta técnica permitió evaluar la adecuación y confiabilidad del modelo. Para determinar su significancia estadística, se empleó el coeficiente de determinación (r^2), el cual indica que tan bien se ajustan los datos al modelo. Además, para la interacción entre las variables independientes y dependiente se representó a través de una gráfica de regresión generada por el modelo de MSR, entrenado en MATLAB 2020®.

2.4.4.2 Redes neuronales artificiales

Utilizando los 7672 datos históricos obtenidos del monitoreo en tiempo real de tres tipos de equipos industriales, se construyó un modelo de RNA con el objetivo de replicar la interacción entre seis variables independientes que influyeron directamente en el estado del equipo y la predicción de posibles fallas.

2.4.4.2.1 Arquitectura y entrenamiento de la red

La estructura de la red neuronal se constituyó por una capa de entrada con seis neuronas, representando a las variables independientes, y una capa de salida con una sola neurona, estableciendo la variable dependiente, es decir, el estado del equipo o la predicción de una posible falla. El modelo fue entrenado y evaluado con el software MATLAB 2020® utilizando el comando o caja de herramientas (`nftool`). Para facilitar el entrenamiento de la red, se empleó la función de activación *Sigmoide tangente hiperbólica* y en las capas ocultas y la capa de salida se utilizó la función de *transferencia lineal*. Esta configuración permitió capturar patrones no lineales complejos y generó salidas continuas útiles para la predicción del comportamiento de los equipos.

Para determinar la significancia y que tan bien explica los datos del modelo de RNA se recurrió al coeficiente de determinación (r^2) al igual que en la MSR.

2.4.5 Quinta fase

En esta fase se recopilaron los resultados numéricos y gráficos relacionados con el error cuadrático medio (**RMSE**) y el coeficiente de determinación (r^2). Estos indicadores permitieron tener una visión clara del comportamiento y precisión de cada uno de los modelos desarrollados (RNA y MSR) y con ello se estableció qué tan bien logran ajustarse los datos analizados a cada uno de los modelos.

2.4.6 Sexta fase

A partir de los resultados obtenidos en la fase anterior, se llevó a cabo una comparación entre los modelos de redes neuronales artificiales (RNA) y la metodología de superficie de respuesta

(MSR). Este análisis, apoyado de los valores del **RMSE** y r^2 , tuvo como propósito identificar cuál de los dos enfoques ofrece un mejor ajuste a los datos y una mayor precisión para su aplicación en la predicción de fallas, facilitando así una toma de decisiones más eficiente dentro del mantenimiento predictivo.

3. Resultados y discusión

3.1 Resultados experimentales

Antes de dar inicio al procesamiento de los datos y el entrenamiento de los modelos en el software de aprendizaje automático MATLAB 2020®, se llevó a cabo un caso de estudio utilizando el conjunto de datos "Industrial Equipment Monitoring Dataset" disponible en el sitio web Kaggle. Esta etapa permitió observar y comprender las variables que intervienen en la detección de fallas en equipos industriales y por lo tanto, su relación con el diagnóstico de estos sistemas. Con base en seis variables independientes relacionadas en la detección de fallas y el estado final de los equipos industriales, se analizaron 7672 registros históricos (ver Tablas 3 y 4). De estos datos obtenidos mediante monitoreo en tiempo real, se identificó que 6905 equipos se encuentran no defectuosos, mientras que 767 si se encuentran defectuosos. Esta distribución demuestra la influencia de las variables independientes sobre el comportamiento de la variable dependiente o de respuesta.

Más allá de la distribución descriptiva, se analizaron los rangos máximos y mínimos registrados para cada métrica física. Se observó que, cuando alguna variable física excede los niveles esperados, por ejemplo, una temperatura o vibración fuera del rango operativo nominal, tiende a presentarse una mayor probabilidad de fallo. Asimismo, se consideró que la ubicación geográfica de los equipos influye significativamente: no es lo mismo operar en una ciudad con condiciones ambientales distintas o infraestructura diferente respecto a otra, y en nuestro caso se consideraron cinco. También el tipo de equipo tiene características propias (como tolerancias, fabricación o uso), lo que establece que su funcionamiento normal o posible fallo se determinará a través del comportamiento físico y geográfico de las variables independientes que intervienen. Desde una perspectiva ingenieril, resulta lógico suponer que estas seis variables, al combinar condiciones físicas, factores geográficos y el tipo de maquinaria, actúan de manera sinérgica y afectan el estado operativo de los equipos. Aunque en la reflexión inicial no se demuestra de forma estadística la interacción de las variables independientes con la variable dependiente, la decisión de incluir estas variables en los modelos predictivos se justifica a partir de una perspectiva ingenieril y en el análisis previo de los datos históricos contenidos en la base de datos.

La validación formal del impacto de las variables se aborda en las secciones siguientes, mediante el entrenamiento y comparación de los modelos de MSR y RNA, utilizando métricas como el coeficiente de determinación (r^2) y el error cuadrático medio (RMSE). Este análisis será la clave para cumplir con el objetivo del estudio y respaldar la justificación inicial, ya que de ello depende la comprensión y cierre de la brecha principal planteada en este artículo.

3.2 Metodología de superficie de respuesta

El modelo de la metodología de superficie de respuesta se desarrolla utilizando un modelo cuadrático que permite describir como las variables de entrada influyen en una variable de salida, en nuestro caso particular, el modelo cuadrático se compone de seis ejes para las variables del proceso. La ecuación 2 se obtuvo para predecir el comportamiento de los equipos industriales (y).

$$\begin{aligned}
 y = & 1.8055 - 2.19 \times 10^{-2}x_1 - 2.11 \times 10^{-2}x_2 - 4.52 \times \\
 & 10^{-2}x_3 - 2.67 \times 10^{-2}x_4 - 5.92 \times 10^{-2}x_5 + 1.80 \times 10^{-2}x_6 - \\
 & 1.4 \times 10^{-5}x_1x_2 - 7.35 \times 10^{-4}x_1x_3 + 1 \times 10^{-6}x_1x_4 + \\
 & 5.79 \times 10^{-4}x_1x_5 + 1.27 \times 10^{-4}x_1x_6 - 1.8 \times 10^{-3}x_2x_3 + \\
 & 1.1 \times 10^{-5}x_2x_4 - 1.94 \times 10^{-4}x_2x_5 + 3.1 \times 10^{-5}x_2x_6 - \\
 & 5.7 \times 10^{-5}x_3x_4 + 4.53 \times 10^{-3}x_3x_5 - 5.10 \times 10^{-3}x_3x_6 - \\
 & 6.2 \times 10^{-5}x_4x_5 - 1.6 \times 10^{-4}x_4x_6 + 1.66 \times 10^{-3}x_5x_6 + \\
 & 1.55 \times 10^{-4}x_1^2 + 3.58 \times 10^{-4}x_2^2 + 0.06.15 \times 10^{-2}x_3^2 + \\
 & 2.68 \times 10^{-4}x_4^2 + 4.39 \times 10^{-3}x_5^2 - 3.92 \times \\
 & 10^{-3}x_6^2 \quad (2)
 \end{aligned}$$

El modelo cuadrático empleado (ver Ecuación 2) fue construido con base en la metodología de superficie de respuesta y considera un total de 28 términos. El modelo incorpora un valor constante (intercepto) y el efecto individual de cada una de las seis variables del proceso (x_1 a x_6), cómo estas se combinan entre sí (15 interacciones bilaterales) y sus efectos de forma curvilínea (6 términos cuadráticos). Es importante aclarar que, en un posible caso hipotético donde las seis variables independientes del modelo se igualaran a cero, el resultado sería $y = 1.8055$, correspondiente al valor del intercepto. No obstante, este valor representaría únicamente una referencia matemática, ya que en la práctica no es permisible que todas las condiciones del proceso estén en cero al mismo tiempo. Por consiguiente, este valor no debe interpretarse como un diagnóstico real del equipo.

En la Tabla 5 se muestran los coeficientes de cada uno de los términos del modelo, en conjunto con otros datos como el error estándar, el estadístico t y el valor p. La información de la tabla permite comprender la significancia de cada variable, su interacción y término cuadrático dentro del modelo, para poder identificar los factores que tienen mayor impacto en el comportamiento de los equipos industriales.

Tabla 5. Resultados del modelo de regresión lineal

Término	Coefficiente Estimado	Error Estándar	t-Statistic	Valor-p
Intercepto	1.8055	0.0658	27.43	< 0.001
x_1	-0.0219	0.00076	-28.92	< 0.001
x_2	-0.0212	0.00119	-17.83	< 0.001
x_3	-0.0452	0.0163	-2.77	0.0056
x_4	-0.0267	0.00123	-21.63	< 0.001
x_5	-0.0592	0.0232	-2.55	0.0109
x_6	0.0181	0.0119	1.51	0.1306
$x_1 : x_2$	-1.39E-05	6.39E-06	-2.17	0.0297
$x_1 : x_3$	-0.00073	9.09E-05	-8.09	< 0.001
$x_1 : x_4$	1.03E-06	5.86E-06	0.18	0.8604
$x_1 : x_5$	0.00058	0.00015	3.99	< 0.001
$x_1 : x_6$	0.00013	8.30E-05	1.53	0.1266
$x_2 : x_3$	-0.0018	0.00017	-10.92	< 0.001
$x_2 : x_4$	1.15E-05	1.07E-05	1.07	0.2832
$x_2 : x_5$	-0.00019	0.00022	-0.87	0.3866
$x_2 : x_6$	3.07E-05	0.00013	0.24	0.8116
$x_3 : x_4$	-5.74E-05	0.00015	-0.39	0.6946
$x_3 : x_5$	0.00453	0.00323	1.4	0.1604
$x_3 : x_6$	-0.0051	0.00185	-2.76	0.0059
$x_4 : x_5$	-6.24E-05	0.00019	-0.32	0.749
$x_4 : x_6$	-1.61E-05	0.00011	-0.14	0.8856
$x_5 : x_6$	0.00167	0.00162	1.03	0.3038
x_1^2	0.00015	3.35E-06	46.26	< 0.001
x_2^2	0.00036	1.01E-05	35.49	< 0.001

x_3^2	0.0615	0.00188	32.66	< 0.001
x_4^2	0.00027	9.08E-06	29.48	< 0.001
x_5^2	0.0044	0.00397	1.11	0.2682
x_6^2	-0.00393	0.00112	-3.5	0.0005

Fuente: Elaboración propia, tomando como base los resultados obtenidos del entrenamiento del modelo de MSR en el software MATLAB 2020®.

La influencia de las variables independientes sobre la variable de respuesta, se muestran en la tabla 6, la cual representa el análisis de varianza (ANOVA). El modelo cuadrático de la metodología de superficie de respuesta es altamente significativo ya que presenta un valor de ($p = < 0.001$), lo que confirma su capacidad para predecir fallas en equipos industriales; el modelo explica la mayor parte de la variabilidad total (**484.8** de **690.3**), destacando los efectos lineales, se obtuvo un valor F de **1015.1**, lo que indica una fuerte influencia individual de las variables. También los efectos no lineales (interacciones y términos cuadrados) son significativos, aportando a la precisión del modelo con un valor F de **568.3**. El error residual es bajo, lo que respalda la solidez del ajuste.

Tabla 6. ANOVA del modelo cuadrático de la MSR

	SC	GL	CM	Valor-F	Valor-p
Total	690.3	7671	0.090		
Modelo	484.8	27	17.95	667.6	< 0.001
Lineal	163.8	6	27.29	1015.1	< 0.001
No lineal	321.0	21	15.28	568.3	< 0.001
Residual	205.6	7644	0.027		

Fuente: Elaboración propia, tomando como base los resultados del ANOVA obtenidos a partir del entrenamiento del modelo de MSR en el software MATLAB 2020®.

Para concluir el análisis de resultados de la metodología de superficie de respuesta, se evaluó la capacidad predictiva del modelo a partir del coeficiente de determinación (r^2), el cual alcanzó un valor de **0.5759**. El resultado indica que el **57.59%** de la variabilidad observada en los datos experimentales puede explicarse a partir de las predicciones generadas por el modelo entrenado en MATLAB 2020®, como se muestra en la figura 6.

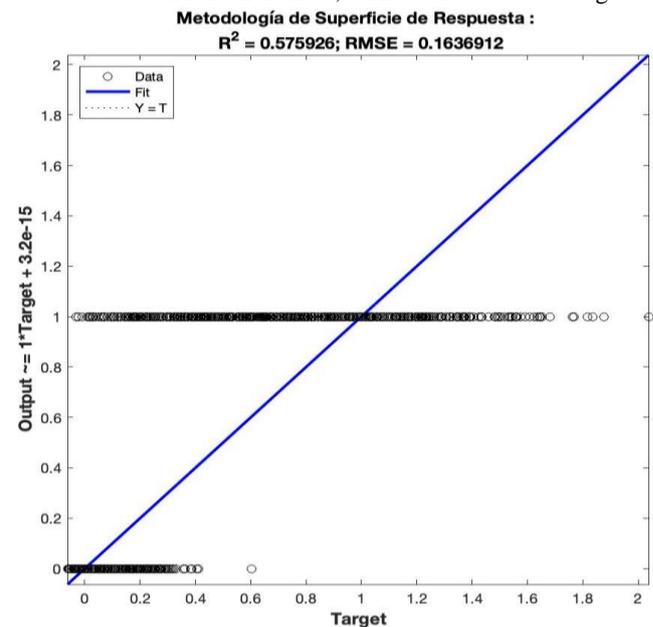


Figura 6. Ajuste de regresión para el modelo de MSR.

Fuente: MATLAB 2020®

3.3 Redes neuronales artificiales

El entrenamiento de la red neuronal artificial con arquitectura multicapa se diseñó con el propósito de predecir fallas en equipos industriales utilizando seis variables independientes. Este proceso se llevó a cabo en el software MATLAB 2020®, empleando el algoritmo de retropropagación *Levenberg–Marquardt*, destacado por su eficacia en problemas de regresión. Para encontrar la configuración más adecuada de la red neuronal, se desarrolló un *script* que analizó distintas combinaciones posibles.

La referencia fue un modelo de tres capas. La primera, conocida como capa de entrada, incluyó 6 neuronas, correspondientes a cada variable independiente; después, se añadieron tres capas ocultas, ajustando el número de neuronas entre una y diez para cada una de ellas; finalmente, la capa de salida se mantuvo con una sola neurona, encargada de predecir el estado final del equipo, es decir, si presenta o no una falla.

En total, se generaron 1,110 combinaciones posibles, para cada una de ellas se obtuvieron valores numéricos de r^2 y $RMSE$ con el propósito de encontrar la que mejor se ajustara a los datos reales del estudio.

La red con mayor rendimiento se muestra en la Figura 7, compuesta por una capa de entrada con 6 neuronas, tres capas ocultas con 10, 9 y 8 neuronas respectivamente, y una capa de salida con una neurona.

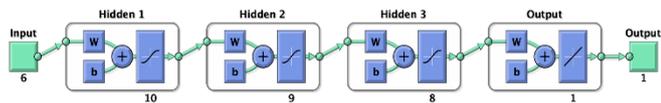


Figura 7. RNA con mayor rendimiento

Fuente: MATLAB 2020®

Para evaluar el desempeño del modelo entre los datos experimentales y predichos, se utilizó como métrica el valor estadístico del coeficiente de determinación (r^2), el cual alcanzó un valor de **0.8881**, indicando una capacidad explicativa del **88.81%** respecto a los datos reales (Figura 8).

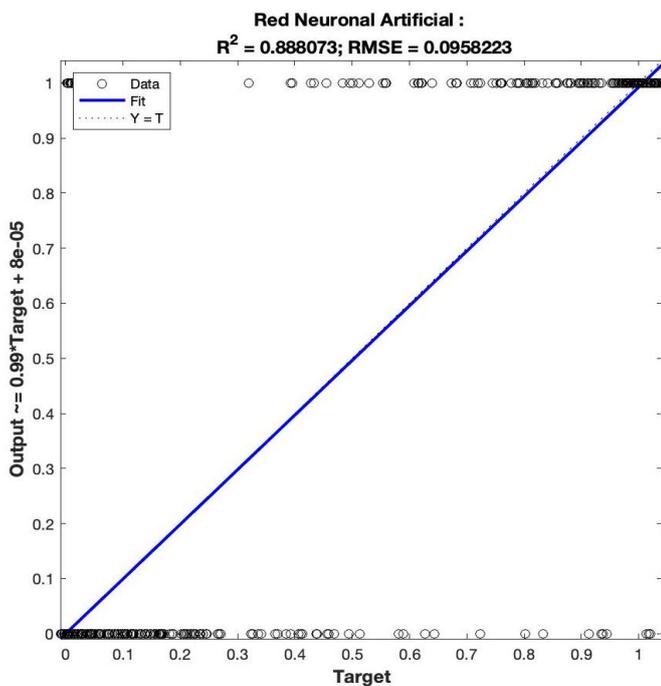


Figura 8. Modelo de regresión para la RNA con mayor rendimiento.

Fuente: MATLAB 2020®

El buen nivel de ajuste (coeficiente de determinación) muestra que existe una fuerte relación entre lo que predice la red neuronal y lo que realmente ocurre en los datos. Por lo que se concluye que el modelo logró aprender de forma efectiva los patrones complejos que existen entre las variables monitoreadas y el estado final de los equipos.

3.4 Comparación entre los modelos a partir de r^2 y $RMSE$

Para establecer cuál de los dos modelos predice mejor las fallas, se recurrió a dos indicadores estadísticos importantes: el coeficiente de determinación (r^2) y la raíz del error cuadrático medio ($RMSE$). Estos valores permiten tener una idea clara de qué tan bien se ajusta cada modelo a los datos reales observados.

3.4.1 Coeficiente de determinación (r^2)

De acuerdo con (Palafox, Medina, Seck Tuoh Mora, Serna, & Hernández, 2023) el r^2 nos ayuda a entender qué tanto un modelo es capaz de explicar lo que ocurre en los datos reales. En otras palabras, indica qué tan bien el modelo logra predecir el comportamiento de la variable que estamos estudiando en función de las variables que utilizamos como entrada, lo anterior, se establece por la ecuación número 3, la cual fue adaptada de (Serna, Tellez, Seck-Tuoh-Mora, Hernández, & Medina, 2024).

$$r^2 = \frac{(\sum_{i=1}^n (A_i - \bar{A}_i) (F_i - \bar{F}_i)^2)}{\sum_{i=1}^n (A_i - \bar{A}_i)^2 (F_i - \bar{F}_i)^2} \quad (3)$$

Donde A_i es el valor real observado de las variables independientes, \bar{A}_i es el promedio de los valores reales observados, F_i es el valor pronosticado por el modelo, \bar{F}_i es el promedio de los valores pronosticados y n es el número total de observaciones desde $i = 1$ hasta n .

3.4.2 Raíz del error cuadrático medio (RMSE)

Por otra parte (Palafox, Medina, Seck Tuoh Mora, Serna, & Hernández, 2023) establecen que el $RMSE$ permite evaluar qué tan precisa es la predicción del modelo en comparación con los datos reales; un valor bajo de este indicador refleja un buen ajuste del modelo, lo que indica que las predicciones se acercan bastante a lo que realmente ocurre. De acuerdo con la formulación propuesta por (Serna, Tellez, Seck-Tuoh-Mora, Hernández, & Medina, 2024) la medida estadística del RMSE se calcula a partir de la ecuación número 4.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (F_i - A_i)^2} \quad (4)$$

Donde F_i es el valor pronosticado por el modelo, A_i es el valor real observado de la variable independiente, n es el número total de observaciones y \sum la suma total de los errores al cuadrado.

En la Tabla 7 se muestran los resultados de r^2 y $RMSE$ obtenidos a partir de las fallas que cada modelo logró predecir, ya sea mediante redes neuronales artificiales o con la metodología de superficie de respuesta. Estos resultados fueron indispensables para comprender qué tan cerca estuvo cada enfoque de lo que realmente ocurrió y qué tan útil puede ser para anticiparse a posibles fallas en los equipos industriales.

Tabla 7. Comparativa de métricas estadísticas entre el modelo de MSR y RNA

	Metodología de superficie de respuesta	Redes neuronales artificiales
r^2	0.575926	0.888073
RMSE	0.1636912	0.0958223

Fuente: Elaboración propia, tomando como base los resultados del entrenamiento del modelo de MSR y RNA en el software MATLAB 2020®.

En los resultados anteriores (ver Tabla 7) se observan los valores numéricos del coeficiente de determinación (r^2) alcanzando un valor del **88.81 %** para el modelo de RNA, mientras que el modelo basado en MSR obtuvo **57.59 %**. Esto significa que las RNAs lograron explicar con mayor precisión la variabilidad en los datos reales, frente a la MSR.

En cuanto al error cuadrático medio (**RMSE**), el modelo de RNA registró un valor de **0.0958**, mientras que el de MSR fue de **0.1637**, a partir de lo anterior queda establecido que las predicciones realizadas por la red neuronal tienen un margen de error menor del **41.5 %** en comparación a las obtenidas con la metodología de superficie de respuesta, lo que refuerza aún más la precisión del modelo de RNA.

3.4.3 Análisis final

De acuerdo con los resultados obtenidos queda expuesta una diferencia significativa entre ambos modelos. Mientras que la metodología de superficie de respuesta genera una función matemática explícita para modelar la relación entre variables independientes y dependiente. A diferencia de, la red neuronal artificial (RNA) comprende dicha relación a través del entrenamiento del modelo, en lugar de utilizar una fórmula definida, la RNA ajusta internamente los valores de sus conexiones (pesos) entre la capa de entrada, capas ocultas y capa de salida, para poder representar de forma precisa los patrones existentes en los datos. Ambos enfoques resultan útiles cuando el comportamiento del sistema es complejo y no puede describirse fácilmente mediante una ecuación analítica.

Para finalizar el análisis comparativo de los modelos, el resultado de las métricas estadísticas pone en evidencia que el modelo basado en RNA presenta un mejor ajuste a los datos reales y posee mayor precisión al predecir fallas en equipos industriales. A partir de lo anterior, se establece que el modelo de RNA puede posicionarse como una herramienta más eficiente para implementar estrategias de mantenimiento predictivo, contribuyendo a reducir tiempos de inactividad, prevenir daños mayores en los equipos industriales y optimizar los recursos en entornos de producción.

4. Conclusiones

A partir de los resultados obtenidos en este estudio se confirma que el uso de modelos de aprendizaje automático como las redes neuronales artificiales (RNA) supera de manera significativa a la metodología de superficie de respuesta (MSR) en la predicción de fallas en equipos rotativos industriales. Al analizar 7672 registros históricos y modelar seis variables físicas y geográficas, el modelo de RNA alcanzó un ajuste del 88.81 % (r^2), en comparación con el 57.59 % obtenido por el modelo de MSR. De igual forma, el **RMSE** fue considerablemente menor en las RNA (0.0958), lo que indica una mayor capacidad para replicar y anticipar el comportamiento real de los equipos. Estos hallazgos son relevantes para el fortalecimiento de estrategias en la predicción de fallas como base del mantenimiento predictivo, ya que

permiten a los diferentes sectores industriales intervenir de manera oportuna, reducir costos de operación y evitar paros no planificados.

Al comparar el rendimiento de ambos modelos, las RNA presentan ventajas superiores para ser utilizadas en sectores industriales en comparación con la MSR, debido a que presentan mayor precisión y ofrecen a los sectores industriales la posibilidad de anticiparse con exactitud a posibles fallas, optimizando así la planificación del mantenimiento predictivo, reduciendo costos operativos y tiempos de inactividad no programados. En el contexto de la Industria 4.0, la integración de RNAs se posiciona como una herramienta esencial del aprendizaje automático para el monitoreo inteligente, la toma de decisiones basada en datos y la mejora continua de procesos industriales. En conclusión, las RNA no solo representan una solución técnica eficaz, sino también una estrategia clave para fortalecer la competitividad industrial a largo plazo.

Como trabajo futuro se recomienda probar el modelo en entornos industriales diferentes al caso de estudio actual. Lo anterior permitirá evaluar de manera amplia la viabilidad y efectividad de los modelos de aprendizaje automático para generar información clave que apoye a las organizaciones industriales en la toma de decisiones.

Abreviaciones y Acrónimos

RNA: Redes neuronales artificiales o red neuronal artificial.
MSR: Metodología de superficie de respuesta.
PdM 4.0: Mantenimiento predictivo 4.0
PdM: Mantenimiento predictivo
IOT: Internet de las cosas.
ML: Aprendizaje automático o machine learning.

Agradecimientos

Expreso mi más sincero agradecimiento a la Universidad Autónoma del Estado de Hidalgo por brindarme la oportunidad de formarme profesionalmente en la Licenciatura en Ingeniería Industrial. A mi asesor, el Dr. Joselito Medina Marín, por su paciencia, orientación y apoyo constante durante el desarrollo y redacción de este trabajo. A Dios, a mi familia y a mis amigos, por su motivación, cariño y respaldo incondicional, que han sido pilares fundamentales en este camino. De manera muy especial, a mi mamá, mi mayor inspiración para superarme profesionalmente, gracias por tu amor, ejemplo y compañía en cada etapa de este proceso.

Conflicto de intereses

Los autores declaran que no existen conflictos de intereses.

Referencias

- Acosta, S., Ruiz, E., Prieto, J. A., Orozco, J. L., & Valdés, J. O. (2023). Diagnóstico De Fallas Intermitentes En Sistemas Automáticos De Manufactura Usando Machine Learning. *Pistas Educativas*, 14,11,15. Obtenido de <https://pistaseducativas.celaya.tecnm.mx/index.php/pistas/article/view/2888>
- Alkharisi, M. K., & Dahish, H. A. (2025). The Application of Response Surface Methodology and Machine Learning for Predicting the Compressive Strength of Recycled Aggregate Concrete Containing Polypropylene Fibers and Supplementary Cementitious Materials. *MDPI*, 8. DOI:<https://doi.org/10.3390/su17072913>

- Arcos, J. A., Lizarzaburu, A. S., Marín, B. E., & Arcos, Á. P. (2023). ¿Cómo Es La Gestión De Mantenimiento De Una Empresa Metalmeccánica? *Ingeniar*, 2,7. Obtenido de <https://journalingeniar.org/index.php/ingeniar/article/view/131>
- Arencibia, N. T., Anías, C., & Hernández, F. (2024). Procedimiento para implementar el mantenimiento predictivo basado en la industria 4.0. *Telemática*, 4. Obtenido de <https://revistatelematica.cujae.edu.cu/index.php/tele/article/view/635>
- Carrera, E. A. (2024). Análisis de vibraciones como método de mantenimiento predictivo en máquinas rotativas. *Uleam*, 1. Obtenido de <https://repositorio.uleam.edu.ec/handle/123456789/5778>
- Echeverría, S. V., & Nivelá, M. A. (2024). Análisis exhaustivo de las tecnologías clave en la Industria 4.0. *Código Científico Revista de Investigación*, 18. Obtenido de <https://revistacodigocientifico.itslosandes.net/index.php/1/article/view/632>
- Falla, B. F., & Ortega, D. A. (2022). Evaluación de algoritmos de inteligencia artificial aplicados al mantenimiento predictivo. *AUNAR*, 1-2. Obtenido de <http://repositorio.aunar.edu.co:8080/xmlui/handle/20.500.1276/1258>
- Garrido, S. G. (2003). *Organización y gestión integral de mantenimiento*. Madrid, España: Díaz de Santos, S. A. Recuperado el 16 de Abril de 2025
- Guartazaca, J. S. (Julio de 2024). *Repositorio Institucional de la Universidad Politécnica Salesiana*. Recuperado el 8 de Mayo de 2025, de Predicción de fallas mediante un modelo de aprendizaje automático en motores de combustión interna: <https://dspace.ups.edu.ec/handle/123456789/28661>
- Hernández, T. F. (2005). *Planificación y gestión del mantenimiento industrial*. Caracas, Venezuela: Panapo. Recuperado el 12 de Abril de 2025
- Luna Pérez, M. A., Badillo Piña, I., & Vázquez Álvarez, G. (2020). Diseño de una metodología de mantenimiento predictivo para asegurar procesos de producción de la industria 4.0. *South Florida Journal of Development*, 1,2. DOI:<https://doi.org/10.46932/sfjdv2n1-074>
- Martínez, A. M. (12 de Septiembre de 2021). *Medium*. Recuperado el 15 de Abril de 2025, de Predicción de fallas futuras en motores de turboventiladores por medio de LSTM: <https://medium.com/saturdays-ai/prediccion-de-fallas-futuras-en-motores-de-turboventiladores-por-medio-de-lstm-48d44f331128>
- Onokwai, A. O., Olusanya, O. O., Onifade, M. K., Anyaegbuna, B. E., Onoriode, A., & Agholor, D. (2024). Comparative analysis of response surface methodology and adaptive neuro-fuzzy inference system for predictive fault detection and optimization in beverage industry. *Frontiers in Mechanical Engineering*, 1,7,13,19. DOI:<https://doi.org/10.3389/fmech.2024.1428717>
- Palafox, D., Medina, J., Seck Tuoh Mora, J. C., Serna, M. G., & Hernández, N. (2023). Modelo de pronostico de cadena de suministro mediante redes neuronales. *Padi*, 5. DOI:<https://doi.org/10.29057/icbi.v11iEspecial3.11482>
- Pinzón, C. (2022). Modelos matemáticos aplicados al mantenimiento predictivo. *CMMSHere*, 6-7. Obtenido de <https://cmmshere.com/wp-content/uploads/2022/11/art-CMMSHere-mantenimiento-predictivo.pdf>
- Rangel, L. D. (19 de Diciembre de 2024). *Universidad De Piura*. Recuperado el 1 de Mayo de 2025, de Diseño de modelo basado en redes neuronales para la detección de anomalías en motorreductor de molino de secado en empresas agroindustriales: <https://pirhua.udep.edu.pe/backend/api/core/bitstreams/971de da3-c630-4e40-ba6d-4825b429312e/content>
- Samudrala, D. N. (7 de Diciembre de 2024). *Kaggle*. Obtenido de Industrial Equipmen Monitoring Dataset: <https://www.kaggle.com/datasets/dnkumars/industrial-equipment-monitoring-dataset/data>
- Serna, M. G., Tellez, A., Seck-Tuoh-Mora, J. C., Hernández, N., & Medina, J. (2024). Modeling Conidiospore Production of *Trichoderma harzianum* Using Artificial Neural Networks and Response Surface Methodology. *MDPI*, 3-9. DOI:<https://doi.org/10.3390/app14125323>
- Toapanta, O. G., Paredes, J., Meneses, M., & Salinas, M. (2024). Validation of DOE Factorial/Taguchi/Surface Response Models of Mechanical Properties of Synthetic and Natural Fiber Reinforced Epoxy Matrix Hybrid Material. *MDPI*, 6. DOI: <https://doi.org/10.3390/polym16142051>
- Vázquez, R. D. (2023). De las tic y la digitalización a la industria 4.0 y la transformación digital. *CONICET*, 13,14. Obtenido de <http://hdl.handle.net/11336/247843>
- Villanueva, J., González, J., Blanco, A., Buenabad, H., & Carrera, E. d. (2024). Mantenimiento predictivo de motores de corriente directa empleando redes neuronales artificiales. *Research in Computing Science* 153(9), 2. Obtenido de https://rcs.cic.ipn.mx/2024_153_9
- Villegas, L., & Pérez, R. (2023). Inteligencia Artificial para Toma de Decisiones en el Mantenimiento Predictivo a equipos Turbocompresores en la Industria Petrolera. *Ieras JIP Jornadas de Investigación de Postgrado*, 2, 8, 10, 11, 12, 13 y 14. Obtenido de <https://www.researchgate.net/publication/371008558>
- Wahid, A., Breslin, J. G., & Intizar, M. A. (2022). Prediction of Machine Failure in Industry 4.0: A Hybrid CNN-LSTM Framework. *MDPI*, 3,6,14, 15. DOI:<https://doi.org/10.3390/app12094221>
- Yan, J. (4 de Octubre de 2021). *Predictiva21*. Obtenido de Gestión de mantenimiento orientada al pronóstico y pronóstico de maquinaria: <https://predictiva21.com/gestion-mantenimiento-orientada-pronostico-maquinaria>
- Yandar, M. A., & Moreno, J. M. (2019). *La industria 4.0 desde la perspectiva organizacional*. Quito, Ecuador: Artes y Letras S.A.S.

Younes, E. H., Álvarez, J. A., & Vázquez, J. L. (2021). Estimación de la Vida Útil de la Turbina de Vapor de 110MW Utilizando Redes Neuronales Artificiales. *ECORFAN®*, 82, 85, 105. DOI:10.35429/H.2021.16.81.96