

Comparative Assessment of Baseline LSTM and GA-Tuned LSTM for RUL Estimation in Diesel Engines with Synthetic OBD-II Data

Rolando Chavez Artica¹; Alert Mendoza Acosta²

^{1,2} Universidad Tecnológica del Perú, Perú, U19309183@utp.edu.pe, E08141@utp.edu.pe

Abstract— *Unplanned diesel-engine failures generate significant operational and economic impacts in heavy-transport systems. Predictive maintenance based on remaining useful life (RUL) estimation can mitigate downtime; however, the availability of labeled degradation datasets remains limited. This work presents a simulation-based and reproducible framework for generating multivariate OBD-II-like signals and evaluating RUL prediction using a Long Short-Term Memory (LSTM) network. A Genetic Algorithm (GA) is further employed for hyperparameter optimization, using validation performance as the fitness criterion. The system is evaluated as a regression problem using R^2 , RMSE, MAE, and relative RMSE under an asset-based split, where a complete unseen engine is reserved for testing. Results show that both models meet the predefined acceptance criteria ($R^2 \geq 0.85$ and relative RMSE $\leq 15\%$). However, the baseline LSTM achieves lower prediction error on the test set compared to the GA-optimized model, indicating superior generalization performance. This finding highlights that improved validation fitness does not necessarily translate into better performance on unseen assets. The proposed framework provides a controlled proof of concept for reproducible evaluation of predictive models and supports future extensions using real OBD-II data and more robust validation schemes.*

Keywords— *Predictive maintenance, Remaining Useful Life (RUL), Long Short-Term Memory (LSTM), Genetic Algorithm (GA), OBD-II data.*

Evaluación comparativa de una LSTM base y una LSTM con optimización por Algoritmos Genéticos para la estimación de la Vida Útil Remanente en motores diésel usando datos sintéticos OBD-II

Rolando Chavez Artica¹; Alert Mendoza Acosta²

^{1,2} Universidad Tecnológica del Perú, Perú, U19309183@utp.edu.pe, E08141@utp.edu.pe

Resumen— Las fallas no planificadas en motores diésel generan impactos operativos y económicos significativos en el transporte pesado. El mantenimiento predictivo basado en la estimación de la vida útil remanente (RUL) permite reducir paradas no previstas; sin embargo, la disponibilidad de datos reales etiquetados con degradación es limitada. Este estudio propone un framework reproducible basado en simulación para la generación de señales multivariadas tipo OBD-II y la evaluación de la estimación de RUL mediante una red Long Short-Term Memory (LSTM). Adicionalmente, se emplea un Algoritmo Genético (AG) para la optimización de hiperparámetros, utilizando el desempeño en validación como función fitness. La evaluación se plantea como un problema de regresión utilizando R^2 , RMSE, MAE y RMSE relativo, bajo un esquema de partición por activos que reserva un motor no visto para prueba. Los resultados muestran que ambos modelos cumplen los criterios de aceptación definidos ($R^2 \geq 0.85$ y RMSE relativo $\leq 15\%$). Sin embargo, el modelo LSTM baseline presenta un menor error en el conjunto de prueba respecto al modelo optimizado por AG, evidenciando una mejor capacidad de generalización. Este hallazgo demuestra que una mejora en la función fitness durante la validación no garantiza necesariamente un mejor desempeño en escenarios no observados. El framework propuesto constituye un proof of concept controlado para la evaluación reproducible de modelos predictivos y sienta las bases para futuras investigaciones con datos reales OBD-II y esquemas de validación más robustos.

Palabras clave— Mantenimiento predictivo, Vida útil remanente (RUL), Long Short-Term Memory (LSTM), Algoritmo genético (AG), Datos OBD-II.

I. INTRODUCCIÓN

Las fallas no planificadas en motores diésel generan impactos operativos y económicos significativos en el sector transporte, debido a paradas imprevistas, retrasos logísticos y costos asociados al mantenimiento correctivo. En contextos donde la operación depende de flotas de vehículos de carga, la disponibilidad del motor se convierte en un factor crítico para garantizar la continuidad del servicio. En este escenario, el mantenimiento predictivo ha adquirido un rol estratégico al permitir anticipar fallas antes de su ocurrencia mediante el

análisis de señales operativas y patrones de degradación, en concordancia con las tendencias de la industria 4.0 [1], [2].

No obstante, uno de los principales desafíos para el desarrollo de sistemas predictivos robustos es la escasez de conjuntos de datos reales etiquetados, particularmente en tareas de estimación de la Vida Útil Remanente (Remaining Useful Life, RUL). En aplicaciones reales, los registros históricos suelen ser incompletos, no presentan trayectorias claras de degradación progresiva o carecen de etiquetas confiables asociadas a eventos de falla. Esta limitación afecta directamente el entrenamiento supervisado de modelos basados en *deep learning* y restringe su capacidad de generalización [3].

A pesar de estas limitaciones, diversos estudios han demostrado que las redes neuronales recurrentes, en particular las arquitecturas *Long Short-Term Memory* (LSTM), son adecuadas para modelar series temporales complejas y capturar dependencias de largo plazo en señales multivariadas, lo que las convierte en una alternativa viable para la estimación de RUL [4], [5]. Asimismo, se ha explorado el uso de técnicas de optimización evolutiva, como los Algoritmos Genéticos (AG), para el ajuste de hiperparámetros de modelos de aprendizaje profundo. Sin embargo, la literatura también señala que dichas estrategias no garantizan mejoras consistentes en la capacidad de generalización, especialmente en escenarios con datos limitados o alta variabilidad entre unidades evaluadas [6].

En el contexto del transporte y los motores diésel, la dependencia de vehículos pesados refuerza la necesidad de desarrollar enfoques predictivos aplicables a flotas, orientados a reducir fallas inesperadas y mejorar la planificación del mantenimiento. A pesar de ello, en la práctica aún predominan estrategias de mantenimiento programado o correctivo, debido a barreras asociadas a la instrumentación, la disponibilidad de datos y los costos de implementación de sistemas predictivos avanzados [7].

Una alternativa viable para la construcción de modelos predictivos en vehículos consiste en aprovechar fuentes de información ya disponibles, como los sistemas de diagnóstico a

bordo. El estándar OBD-II permite acceder a múltiples variables operativas del motor mediante modos definidos en normas técnicas como SAE J1979, facilitando la adquisición de parámetros relevantes para diagnóstico y monitoreo sin requerir instrumentación adicional [8]. Diversos estudios han reportado el uso de datos OBD-II en combinación con técnicas de *machine learning* como un mecanismo accesible para el desarrollo de soluciones de diagnóstico y mantenimiento predictivo en sistemas vehiculares [9], [10].

Desde una perspectiva de ingeniería de sistemas, el desarrollo de soluciones predictivas integradas requiere un marco metodológico que asegure trazabilidad entre las fases de diseño, implementación y validación. La metodología VDI 2206 propone un enfoque estructurado para el desarrollo de sistemas mecatrónicos, integrando modelado, diseño y verificación como elementos centrales del ciclo de ingeniería, lo que resulta adecuado para guiar propuestas que combinan adquisición de datos, modelado predictivo e integración de sistemas [11].

En respuesta a la problemática de la falta de *datasets* reales etiquetados con degradación progresiva, diversos trabajos han explorado el uso de datos sintéticos y enfoques basados en simulación como mecanismos para construir entornos controlados de evaluación y *proofs of concept* reproducibles para la estimación de RUL [12], [13]. Estos enfoques permiten analizar el comportamiento de modelos predictivos bajo condiciones definidas, sin afirmar una correspondencia directa con sistemas reales, pero facilitando comparaciones metodológicas consistentes. Asimismo, estudios recientes en diagnóstico y pronóstico de motores diésel refuerzan la relevancia de este tipo de aproximaciones en aplicaciones vehiculares [14], [15].

En este contexto, la principal contribución de este trabajo es el desarrollo de un framework reproducible basado en simulación para la estimación de la Vida Útil Remanente (RUL) en motores diésel a partir de datos sintéticos tipo OBD-II. A diferencia de estudios orientados a proponer nuevas arquitecturas o a reportar mejoras de desempeño mediante técnicas de optimización, esta investigación presenta una evaluación comparativa controlada entre un modelo LSTM baseline y un modelo LSTM con optimización de hiperparámetros mediante Algoritmos Genéticos, empleando un esquema de validación basado en activos. El framework propuesto permite analizar de forma sistemática la capacidad de generalización de los modelos sobre motores no vistos, evidenciar las limitaciones de la optimización evolutiva en escenarios con datos limitados y establecer un *proof of concept* metodológico que puede extenderse a futuras investigaciones con datos reales OBD-II y esquemas de validación más robustos.

II. METODOLOGÍA

Este trabajo desarrolla un sistema de mantenimiento predictivo orientado a estimar la Vida Útil Remanente (RUL) en motores diésel empleando datos sintéticos generados con base en variables asociadas a la operación del motor. La propuesta considera la construcción de un dataset sintético con degradación progresiva y fallas complejas, el entrenamiento de un modelo predictivo basado en redes neuronales LSTM y la optimización de hiperparámetros mediante Algoritmo Genético (AG). La evaluación del sistema se realiza como un problema de regresión, considerando como criterios de validación un desempeño mínimo de $R^2 \geq 0.85$ y RMSE relativo $\leq 15\%$.

II.A Variables seleccionadas

El sistema se basa en cuatro variables principales para la simulación: temperatura del refrigerante ($^{\circ}\text{C}$), presión del riel de combustible (bar), carga del motor (%) y revoluciones por minuto (RPM). Para cada variable se consideraron rangos típicos de operación y patrones de falla simulados asociados al comportamiento degradativo del motor.

II.B Generación del dataset sintético

El dataset sintético fue generado para emular señales OBD-II multivariadas con degradación progresiva, con el objetivo de disponer de un conjunto de entrenamiento reproducible ante la escasez de registros reales etiquetados para RUL. Para asegurar trazabilidad experimental, el proceso de simulación se diseñó bajo tres principios: reproducibilidad, variabilidad inter-motor y realismo operativo. En primer lugar, se fijó una semilla global (`np.random.seed(42)`) para garantizar que los experimentos puedan reproducirse bajo las mismas condiciones. En segundo lugar, se incorporó variabilidad entre motores simulando diferentes vidas útiles mediante una distribución normal $\text{total_ciclos} \sim N(\mu=217, \sigma=15)$, lo cual permite generar historiales no idénticos y representar dispersión esperable entre unidades reales. Finalmente, se modeló un patrón de operación discontinuo (2 h de operación y 4 h de inactividad) con la finalidad de distinguir entre timestamp y horas operativas efectivas, aproximando condiciones de uso industrial.

Una vez establecidos estos elementos base, la degradación se implementó como un proceso no lineal que desplaza gradualmente las variables desde rangos normales hacia valores críticos conforme progresa el ciclo de vida del motor, empleando interpolación temporal mediante `np.interp`. Para reflejar la pérdida de estabilidad asociada al envejecimiento, se añadió ruido creciente, incrementando la variabilidad de las mediciones conforme avanza el proceso degradativo. Además, se incorporaron comportamientos de falla más realistas mediante fallas intermitentes (eventos transitorios de

baja probabilidad con recuperación temporal) y fallas terminales con interdependencia multivariable activadas cuando $RUL < 40$ horas. Por último, se introdujeron imperfecciones en los datos mediante huecos temporales por paradas de mantenimiento y truncamiento de historiales en ciertos casos, simulando retiros preventivos antes de alcanzar la falla completa.

II.C Preprocesamiento y segmentación temporal

El módulo de preprocesamiento implementa un pipeline de transformaciones programado en Python cuyo objetivo es convertir el dataset crudo generado en un formato estructurado y normalizado, apto para el entrenamiento de una Red Neuronal Recurrente (LSTM). Este módulo aborda tres desafíos principales: validación de series temporales, normalización de variables en diferentes escalas y segmentación de datos en secuencias.

En primer lugar, se realiza una limpieza y depuración de los registros, que contempla la identificación y eliminación de lecturas inconsistentes, valores atípicos (outliers) y observaciones incompletas. En caso de detectarse valores faltantes aislados, se aplica interpolación temporal lineal para preservar la continuidad de la serie sin alterar su tendencia original, manteniendo la integridad de las dependencias temporales requeridas para modelos secuenciales.

Posteriormente, se aplica una normalización Min-Max normalizando las variables críticas a un rango entre 0 y 1, definida por:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Donde X_{min} y X_{max} corresponden a los valores mínimo y máximo observados dentro del conjunto de entrenamiento. Esta normalización se aplica a las cuatro variables de interés (temperatura del refrigerante, presión del riel de combustible, carga del motor y RPM), permitiendo una convergencia más estable durante el entrenamiento y evitando la dominancia entre variables por diferencias de escala. Para prevenir fuga de datos (data leakage), el escalador se ajusta únicamente sobre los datos de entrenamiento.

Para la división de datos se emplea una estrategia de División Temporal basada en Activos (Asset-based Split), evitando particionados temporales simples o aleatorios debido al riesgo de fuga de datos. En este protocolo, los historiales completos de los motores 1, 2, 3 y 4 se destinan al conjunto de entrenamiento y validación (80% del total de activos), mientras que el historial completo del motor 5 se reserva como conjunto de prueba (20% del total de activos), evaluando así el modelo sobre un activo no visto durante entrenamiento.

Finalmente, la segmentación temporal se realiza mediante la técnica de ventanas deslizantes (Sliding Window), definiendo un hiperparámetro lookback de 20 pasos de tiempo (SEQUENCE_LENGTH=20). Cada ventana representa los últimos 20 registros de las 4 variables de entrada y se utiliza para predecir el valor de RUL(hours) al final de dicha ventana. Este procedimiento transforma la estructura de datos 2D en un tensor 3D, (muestras, pasos-de-tiempo, variables) con forma (N,20,4). La segmentación se aplica por separado a cada motor para evitar que las ventanas crucen entre motores o a través de huecos de mantenimiento simulados

II.D Modelo predictivo LSTM

El modelo predictivo se implementó mediante una red neuronal recurrente Long Short-Term Memory (LSTM) apilada para estimación de RUL_{Horas} . La arquitectura se definió con una dimensión de entrada (SEQUENCE_LENGTH, N_FEATURES) = (20,4). El modelo contiene dos capas LSTM con 50 y 25 unidades respectivamente, seguidas de una capa Dropout de 0.2 para reducir sobreajuste y una capa densa final de una neurona con activación lineal para la predicción del valor de RUL(hours). El modelo se compiló utilizando el optimizador Adam y la función de pérdida Error Cuadrático Medio (MSE).

II.E Optimización de hiperparámetros mediante Algoritmo Genético (AG-LSTM)

Con el objetivo de optimizar hiperparámetros del modelo LSTM, se integró un Algoritmo Genético empleando la librería DEAP, definiendo una estructura evolutiva orientada a minimizar el fitness. La evaluación de cada individuo se ejecutó mediante un wrapper AG-Keras, el cual construye un modelo LSTM bajo una configuración de hiperparámetros determinada, lo entrena y retorna su desempeño en validación. El fitness de cada individuo se definió como el valor más bajo de val_loss (Error Cuadrático Medio) alcanzado durante el entrenamiento del modelo correspondiente.

II.F Protocolo de entrenamiento y criterios de validación

El entrenamiento del modelo consideró una separación automática de validación mediante validation_split=0.2. La evaluación del sistema se realizó como un problema de regresión, considerando como criterios de validación un desempeño mínimo de $R^2 \geq 0.85$ y un $RMSE_{rel} \leq 15\%$. No obstante, la incorporación de estrategias adicionales de validación, como validación cruzada temporal o evaluación sobre múltiples activos no vistos, podría fortalecer la robustez del protocolo experimental.

III. RESULTADOS

En esta sección se presentan los resultados obtenidos durante la validación del sistema propuesto para estimación de Vida Útil Remanente (RUL(hours)) en motores diésel empleando datos sintéticos OBD-II. La evaluación se realizó bajo una estrategia de partición por activos, reservando un motor completo no visto durante entrenamiento como conjunto de prueba. Los resultados se organizan en tres partes: (i) comparación cuantitativa del desempeño predictivo, (ii) análisis gráfico del seguimiento temporal del RUL, y (iii) comportamiento del proceso de optimización mediante Algoritmo Genético.

III.A Comparación cuantitativa del desempeño predictivo

El desempeño del sistema se evaluó como un problema de regresión mediante métricas estándar: coeficiente de determinación (R^2), raíz del error cuadrático medio (RMSE), error absoluto medio (MAE) y RMSE relativo. La Tabla I resume los valores obtenidos para el modelo baseline LSTM y para el modelo LSTM optimizado mediante Algoritmo Genético (AG-LSTM) en el conjunto de prueba (motor reservado).

Como se observa en la Tabla I, el modelo baseline LSTM alcanzó un desempeño de $R^2 = 0.986$, con $RMSE = 14.42h$, $MAE = 11.02h$ y un $RMSE_{rel} = 6.70\%$. Por su parte, el modelo AG-LSTM obtuvo $R^2 = 0.982$, con $RMSE = 16.46h$, $MAE = 13.01h$ y un $RMSE_{rel} = 7.6\%$. En términos generales, estos valores evidencian que ambos modelos cumplen los criterios globales de validación definidos para el sistema, ya que presentan un nivel elevado de ajuste ($R^2 \geq 0.85$) y un error relativo dentro del umbral permitido ($RMSE_{rel} \leq 15\%$), tal como se exige en la Tabla II.

No obstante, al comparar específicamente la métrica RMSE en la Tabla I, se aprecia que el modelo baseline generaliza mejor en el conjunto de prueba, al presentar un error menor que el modelo optimizado por AG. En consecuencia, no se cumple el criterio RNF-3 (mejora mínima del 10% sobre baseline), definido en la Tabla II, debido a que el modelo AG-LSTM incrementa el RMSE respecto al modelo base. Este resultado sugiere que, si bien la optimización evolutiva permite explorar configuraciones alternativas de hiperparámetros orientadas a minimizar val_loss , el desempeño final depende de la estabilidad del conjunto de validación y de la capacidad de generalización sobre activos no vistos, por lo que una menor pérdida en validación no necesariamente se traduce en una mejora del error final en prueba.

TABLA I

RESUMEN DE CUMPLIMIENTO DE CRITERIOS DE ÉXITO (RNF)

Métricas	LSTM Base	LSTM (AG-optimizado)	Resultado
R^2	0.986	0.982	Alcanzado
RMSE	14.42	16.46	No alcanzado
MAE	11.02	13.01	Alcanzado
RMSE relativo	6.70%	7.6%	Alcanzado

TABLA II

CRITERIOS DE ACEPTACIÓN CUANTITATIVOS PARA LA VALIDACIÓN DEL SISTEMA

Criterio	Métrica	Umbral de éxito	Justificación
RNF-1	Coeficiente de Determinación (R^2)	≥ 0.85	El modelo debe explicar al menos el 85% de la varianza del RUL.
RNF-2	RMSE Relativo	$\leq 15\%$	El error promedio de predicción debe ser inferior al 15% del rango del RUL.
RNF-3	Mejora sobre Baseline (RMSE)	$\geq 10\%$	El AG debe proporcionar una mejora de rendimiento de al menos 10% vs. el modelo base.
RNF-4	Análisis de Residuales	Aleatorio	Los errores no deben mostrar patrones sistemáticos (evaluación cualitativa).
RNF-5	Reproducibilidad	100%	El pipeline debe ejecutarse de forma consistente

III.B Análisis de entrenamiento y convergencia del proceso AG-LSTM

Para complementar el análisis cuantitativo presentado en la Tabla I, se evaluó el comportamiento del entrenamiento del modelo y la convergencia del proceso evolutivo mediante inspección gráfica de curvas de error y fitness. En primer lugar, la Fig. 1 muestra la curva de MSE del modelo baseline LSTM, evidenciando una reducción progresiva del error conforme avanzan las épocas de entrenamiento, lo que confirma estabilidad y convergencia del modelo durante el aprendizaje. Adicionalmente, esta figura permite observar la influencia del

mecanismo de EarlyStopping, el cual detiene el entrenamiento una vez que la pérdida de validación deja de mejorar.

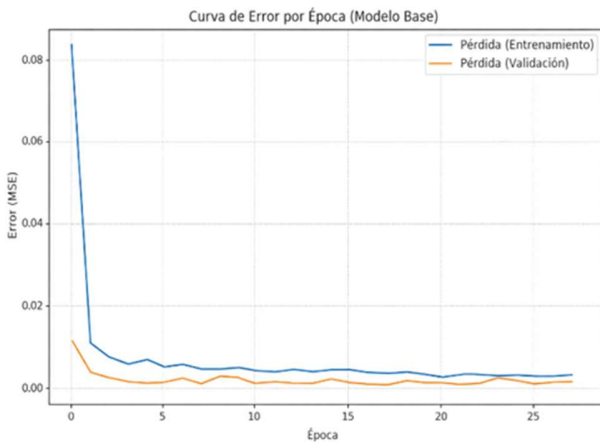


Fig. 1 Curva MSE por época – modelo LSTM base

Asimismo, con el objetivo de evaluar el desempeño del Algoritmo Genético en la búsqueda de hiperparámetros, la Fig. 2 presenta la convergencia del fitness en función de las generaciones, incluyendo la evolución del error promedio y mínimo de la población. Este comportamiento evidencia que el proceso evolutivo logra explorar el espacio de búsqueda y aproximarse hacia configuraciones con menor valor de *val loss*, validando la operatividad del esquema de optimización AG–LSTM.

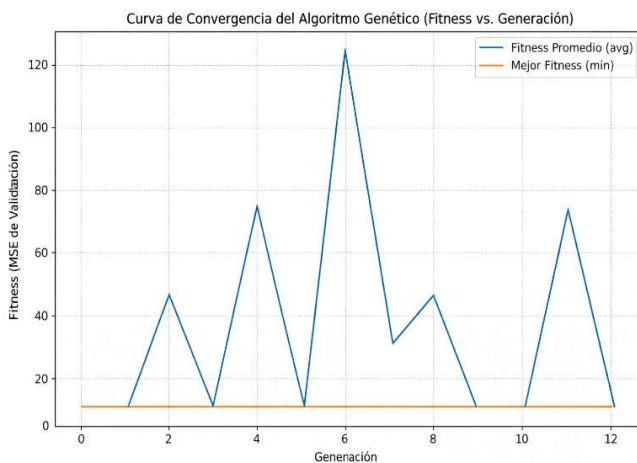


Fig. 2 Convergencia del fitness del Algoritmo Genético

Finalmente, la Fig. 3 muestra la curva de MSE del modelo LSTM entrenado con los hiperparámetros optimizados mediante AG, observándose también una tendencia decreciente del error durante entrenamiento y validación, lo cual confirma

que el modelo resultante mantiene estabilidad de aprendizaje bajo la configuración seleccionada por el algoritmo evolutivo.

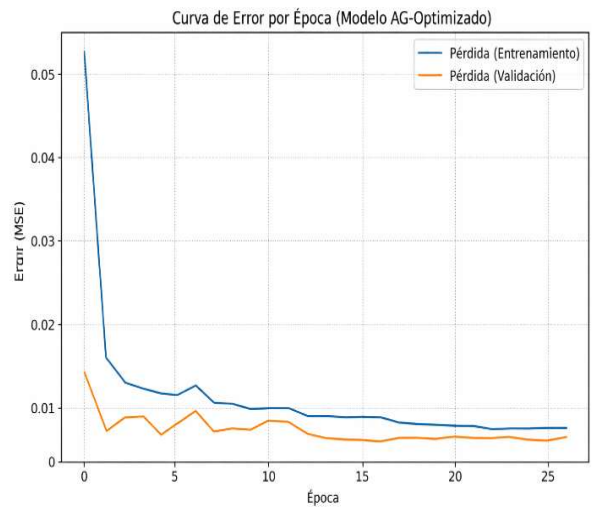


Fig. 3 Curva MSE por época – modelo LSTM optimizado

En conjunto, las Fig. 1–3 respaldan que tanto el modelo baseline como el modelo optimizado alcanzan convergencia durante entrenamiento. Sin embargo, estos resultados también sugieren que una mejora en la pérdida de validación (fitness) durante la optimización genética no garantiza necesariamente una reducción del error final sobre el conjunto de prueba, lo cual es consistente con la comparación cuantitativa mostrada previamente en la Tabla I.

III.C Validación de robustez y estabilidad en entorno simulado.

Con el fin de complementar el análisis cuantitativo y de convergencia presentado en las secciones anteriores, se realizó una evaluación cualitativa del desempeño del sistema sobre el motor reservado para prueba, el cual no fue utilizado durante el proceso de entrenamiento ni validación. La Fig. 4 muestra la comparación entre el valor real de la Vida Útil Remanente (*RUL_Horas*) y el valor estimado por el modelo LSTM optimizado mediante Algoritmo Genético, a lo largo del horizonte temporal evaluado.

IV. DISCUSIÓN

Los resultados obtenidos en este trabajo confirman que los modelos basados en redes neuronales recurrentes LSTM son capaces de estimar la Vida Útil Remanente (RUL) de motores diésel a partir de series temporales multivariadas, incluso cuando los datos provienen de entornos simulados. El desempeño alcanzado por ambos modelos evaluados —LSTM baseline y LSTM optimizado mediante Algoritmo Genético (AG-LSTM)— cumple con los criterios globales de validación definidos ($R^2 \geq 0.85$ y RMSE relativo $\leq 15\%$), lo que evidencia la viabilidad del enfoque como un sistema predictivo orientado a mantenimiento basado en condición. Es importante señalar que el uso exclusivo de datos sintéticos constituye una limitación en términos de aplicabilidad directa a escenarios reales. Si bien el framework propuesto permite experimentación controlada y reproducibilidad, el comportamiento de señales OBD-II reales puede presentar variabilidad adicional que no está completamente representada en el entorno simulado. Por ello, resulta necesario realizar validaciones futuras con datos reales para confirmar la robustez del modelo en condiciones operativas reales.

No obstante, el análisis comparativo muestra que el modelo LSTM baseline presentó un menor error de predicción en el conjunto de prueba que el modelo AG-LSTM. Este resultado indica que la optimización evolutiva de hiperparámetros, si bien puede mejorar el desempeño durante la fase de validación, no garantiza necesariamente una mejor capacidad de generalización sobre activos no vistos. Este comportamiento es consistente con estudios previos que reportan que los métodos automáticos de búsqueda de hiperparámetros pueden conducir a configuraciones sobreajustadas al conjunto de validación, sin traducirse en mejoras consistentes en escenarios de prueba independientes [16].

Desde una perspectiva metodológica, este comportamiento puede explicarse por la sensibilidad del proceso de optimización al conjunto de validación empleado como función *fitness*, así como por la variabilidad intrínseca entre unidades evaluadas. La literatura en mantenimiento predictivo señala que, en problemas de estimación de RUL, la capacidad de generalización del modelo depende fuertemente de la representatividad de las trayectorias de degradación incluidas en el entrenamiento y del esquema de partición de datos utilizado [17]. En este contexto, la evaluación sobre un activo completamente no visto, aunque adecuada para evitar fuga de datos, puede amplificar diferencias de desempeño entre modelos. Adicionalmente, el número reducido de motores simulados limita la generalización estadística de los resultados. Aunque el esquema de validación basado en activos permite evaluar el desempeño en un motor no visto, incrementar la cantidad de datos fortalecería la robustez de las conclusiones obtenidas.

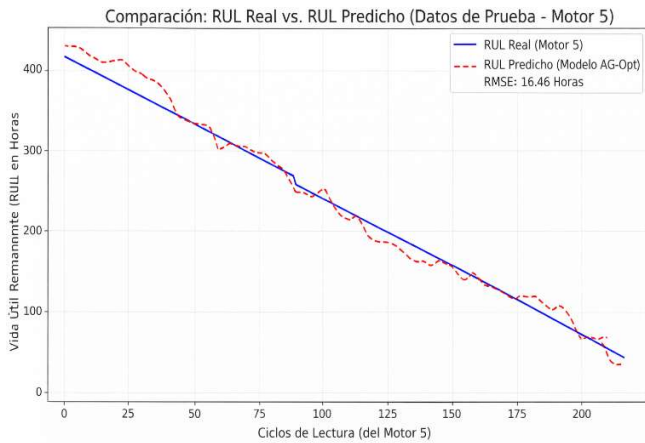


Fig. 4 Comparación de RUL Real vs. RUL Predicho (Modelo AG-Optimizado sobre Motor 5).

Como se aprecia en la Fig. 4, el modelo logra seguir de manera adecuada la tendencia decreciente del RUL asociada al proceso de degradación progresiva del motor, manteniendo una proximidad consistente entre la curva real y la curva predicha. Este comportamiento confirma la capacidad del modelo para capturar las dependencias temporales presentes en las señales multivariadas simuladas y para representar la dinámica global del deterioro, incluso bajo condiciones de ruido creciente y presencia de fallas complejas.

Adicionalmente, el análisis visual de los errores de predicción permite verificar el cumplimiento del criterio RNF-4 definido en la Tabla II, el cual establece que los residuales no deben presentar patrones sistemáticos. La distribución de las desviaciones observadas no evidencia tendencias estructurales ni acumulaciones persistentes, lo que sugiere un comportamiento aproximadamente aleatorio del error y refuerza la validez del modelo desde un punto de vista cualitativo.

En conjunto, la evaluación gráfica sobre el motor de prueba respalda los resultados numéricos presentados en la Tabla I, demostrando que el sistema propuesto es capaz de estimar el RUL de manera coherente y estable bajo un esquema de validación basado en activos no vistos, lo que constituye un indicador clave de generalización para aplicaciones de mantenimiento predictivo.

En relación con el estado del arte, diversas investigaciones han reportado resultados favorables al emplear arquitecturas LSTM para la predicción de RUL, destacando su capacidad para capturar dependencias temporales de largo plazo y relaciones no lineales en series temporales multivariadas [18]. Sin embargo, estos trabajos también enfatizan que el desempeño final está condicionado por la calidad de los datos disponibles, el número de activos considerados y la estrategia de validación aplicada. En este sentido, los resultados obtenidos en este estudio son coherentes con la literatura al demostrar un alto desempeño predictivo, pero también reflejan las limitaciones inherentes a escenarios con disponibilidad restringida de datos. Asimismo, este estudio no incluye comparaciones con otros modelos del estado del arte, como XGBoost, LSTM bidireccional o arquitecturas basadas en Transformers, los cuales han sido reportados en la literatura para tareas de estimación de RUL. Incluir este tipo de comparaciones permitiría ampliar el análisis comparativo y constituye una línea relevante de trabajo futuro.

El uso de datos OBD-II como fuente de información para mantenimiento predictivo constituye otro punto de convergencia con investigaciones recientes. Revisiones del estado del arte indican que las aplicaciones basadas en OBD-II combinadas con técnicas de *machine learning* representan una alternativa viable y de bajo costo para el monitoreo de condición en vehículos, aunque su adopción a gran escala continúa limitada por la escasez de *datasets* etiquetados y la heterogeneidad de las señales disponibles. En este contexto, la estrategia de generación de datos sintéticos con degradación progresiva empleada en este trabajo se alinea con enfoques *simulation-based* utilizados para construir *proofs of concept* reproducibles cuando los datos reales no están disponibles.

Es importante reconocer las limitaciones del presente estudio. En primer lugar, la validación se realizó exclusivamente sobre datos sintéticos, lo que impide extrapolar directamente los resultados a escenarios reales de operación sin una etapa adicional de validación experimental. En segundo lugar, el número de motores simulados es reducido, lo que limita el análisis estadístico de la generalización del modelo. Asimismo, la optimización mediante Algoritmo Genético se basó en una única métrica de desempeño, lo cual puede restringir la exploración de configuraciones más robustas.

A pesar de estas limitaciones, el trabajo aporta implicaciones relevantes tanto a nivel práctico como metodológico. Desde el punto de vista aplicado, demuestra que es posible construir un sistema de mantenimiento predictivo reproducible utilizando únicamente variables accesibles vía OBD-II, sin requerir instrumentación adicional. Desde una perspectiva científica, los resultados evidencian que la optimización de hiperparámetros no debe asumirse como una mejora garantizada del desempeño, reforzando la necesidad de evaluar modelos bajo esquemas de validación estrictos y con

activos no vistos, en concordancia con las tendencias actuales del estado del arte en mantenimiento predictivo y estimación de RUL.

Este hallazgo es relevante, ya que evidencia que la optimización de hiperparámetros no garantiza mejoras en la generalización, lo que resalta la importancia de utilizar esquemas de validación más rigurosos en problemas de estimación de RUL.

V. CONCLUSIONES

En este trabajo se desarrolló y evaluó un sistema de mantenimiento predictivo orientado a la estimación de la Vida Útil Remanente (RUL) en motores diésel, empleando datos sintéticos multivariados basados en variables accesibles vía OBD-II. La propuesta integró un modelo predictivo basado en redes neuronales LSTM y una estrategia de optimización de hiperparámetros mediante Algoritmo Genético, siguiendo un enfoque metodológico estructurado conforme a la metodología VDI 2206.

Los resultados obtenidos demuestran que tanto el modelo LSTM baseline como el modelo optimizado mediante Algoritmo Genético alcanzan un desempeño satisfactorio en términos de precisión, cumpliendo los criterios globales de validación definidos ($R^2 \geq 0.85$ y RMSE relativo $\leq 15\%$). Esto confirma la viabilidad del uso de arquitecturas LSTM para la estimación de RUL a partir de series temporales multivariadas, incluso en escenarios donde los datos provienen de entornos simulados.

El análisis comparativo evidenció que la optimización evolutiva de hiperparámetros no garantizó una mejora en la capacidad de generalización sobre el activo no visto, ya que el modelo baseline presentó un menor error en el conjunto de prueba. Este hallazgo constituye un resultado relevante, al mostrar que la optimización automática puede mejorar el desempeño en validación sin traducirse necesariamente en una reducción del error en escenarios reales o no observados durante el entrenamiento, reforzando la importancia de esquemas de validación basados en activos.

Asimismo, el uso de datos OBD-II como fuente principal de información confirma su potencial como alternativa accesible y de bajo costo para el desarrollo de sistemas de mantenimiento predictivo en vehículos diésel, sin requerir instrumentación adicional. La estrategia de generación de datos sintéticos con degradación progresiva permitió construir un entorno reproducible para la evaluación de modelos predictivos, mitigando la limitación asociada a la escasez de *datasets* reales etiquetados para RUL.

No obstante, el estudio presenta limitaciones que deben ser consideradas. La validación se realizó exclusivamente sobre

datos sintéticos y con un número reducido de activos, lo que restringe la extrapolación directa de los resultados a escenarios reales de operación. Además, la optimización mediante Algoritmo Genético se basó en una única métrica de desempeño, lo que limita la exploración de configuraciones potencialmente más robustas bajo criterios multiobjetivo. Es importante destacar que los resultados obtenidos se limitan a un entorno controlado basado en simulación, por lo que su extrapolación directa a escenarios reales debe realizarse con cautela. El principal aporte de este trabajo radica en el desarrollo de un framework reproducible y controlado para la evaluación de modelos, más que en la implementación de una solución completamente validada en entorno industrial.

Como trabajo futuro, se plantea la incorporación de datos reales OBD-II provenientes de flotas operativas, la ampliación del número de activos para fortalecer la evaluación estadística de la generalización, y la exploración de estrategias de optimización más robustas, incluyendo enfoques multiobjetivo o híbridos. En conjunto, este trabajo aporta un framework reproducible y validable para la estimación de RUL en motores diésel, alineado con las tendencias actuales del estado del arte en mantenimiento predictivo.

REFERENCIAS

- [1] R. K. Mobley, *An Introduction to Predictive Maintenance*, 2nd ed. Oxford, UK: Butterworth-Heinemann, 2002.
- [2] J. Lee et al., "Prognostics and health management design for rotary machinery systems," *Mechanical Systems and Signal Processing*, 2014.
- [3] N. K. Singh et al., "Machine learning techniques for predictive maintenance in Industry 4.0," *Journal of Manufacturing Systems*, 2021.
- [4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, 1997.
- [5] I. Goodfellow et al., *Deep Learning*. MIT Press, 2016.
- [6] J. Serradilla et al., "Deep learning architectures for predictive maintenance," *Applied Sciences*, 2020.
- [7] Asociación Automotriz del Perú (AAP), *Reporte del mercado de camiones y tractocamiones en el Perú*, Lima, Perú, 2024.
- [8] SAE International, *SAE J1979: E/E Diagnostic Test Modes*, Society of Automotive Engineers, 2019.
- [9] A. Michailidis, E. Tryfonas, and S. B. Cooper, "Vehicle diagnostics and prognostics using OBD-II data," *IEEE Access*, vol. 7, pp. 78145–78156, 2019.
- [10] S. Sazli, M. R. Ahmad, and N. A. Rahman, "Diesel engine fault diagnosis using OBD-II data and machine learning techniques," *Sensors*, vol. 24, no. 3, 2024.
- [11] Verein Deutscher Ingenieure (VDI), *VDI 2206 – Design Methodology for Mechatronic Systems*, Düsseldorf, Germany, 2004.
- [12] Y. Wang, "Remaining useful life prediction using deep learning: A review," *Journal of Manufacturing Systems*, vol. 54, pp. 1–20, 2020.
- [13] H. Zheng et al., "Deep learning-based remaining useful life estimation: A review," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 1–14, 2021.
- [14] M. Alshammari, M. Alharbi, and K. Saleh, "Diesel engine fault diagnosis using multisensor data fusion and machine learning," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 6, pp. 4763–4772, 2020.
- [15] C. Huang et al., "A review of data-driven prognostics and health management for predictive maintenance," *International Journal of Production Research*, 2024.
- [16] A. Theissler, "Predictive maintenance enabled by machine learning," *European Journal of Operational Research*, vol. 298, pp. 1–15, 2021.
- [17] J. B. Coble and J. W. Hines, "Applying the general path model to estimation of remaining useful life," *International Journal of Prognostics and Health Management*, vol. 1, no. 1, 2010.
- [18] E. T. Michailidis et al., "A review of OBD-II-based machine learning applications," *Sensors*, vol. 25, no. 13, 2025.