

# Detección de fallas de motores en turbinas de gas

## Informe del Estado del Arte

Adrián Hernández Castellanos (C412)

Laura Martir Beltrán (C411)

Ariadna Velázquez Rey (C411)

Lía Stephanie López Rosales (C412)

Yesenia Valdés Rodríguez (C411)

Olivia Ibañez Mustelier (C411)

**Tutora:** Deborah Famadas Rodríguez

## Resumen

La detección temprana de fallos en turbinas de gas es un desafío crítico para la operación segura y eficiente de instalaciones energéticas. Este informe sintetiza el estado del arte relacionado con técnicas de detección y diagnóstico de fallos en turbinas y componentes asociados (p. ej. rodamientos, rotores, generadores). Se emplea como referencia la tabla de estado del arte (archivo **State of the Art.csv** / **State of the Art.xlsx**) que acompaña este documento y contiene la extracción y comparación de atributos clave de trabajos relevantes. Debido a la limitada disponibilidad de datos y metainformación en el proyecto original, se ha decidido utilizar el *Gas Turbine Engine Fault Detection Dataset* (Kaggle) como fuente principal de datos, aplicando técnicas de aumento de información para robustecer el entrenamiento de modelos de aprendizaje automático.

# Índice

1. Qué estamos investigando	3
2. Título y justificación	3
3. Dataset	3
4. Cómo interpretar la tabla (State of the Art)	3
5. Qué se representa en la tabla	4
6. Estado del arte – síntesis y hallazgos principales	4
7. Metodología propuesta	4
7.1. Ingesta y limpieza . . . . .	5
7.2. Preprocesado y extracción de características . . . . .	5
7.3. Aumentación de datos (Data Augmentation) . . . . .	5
7.4. Modelado . . . . .	5
7.5. Evaluación y métricas . . . . .	5
8. Conclusiones provisionales	5

# 1. Qué estamos investigando

El objetivo principal del proyecto es **desarrollar y evaluar modelos de machine learning** capaces de detectar y clasificar fallos en motores de turbinas de gas a partir de señales y registros de sensores. Los objetivos específicos incluyen:

- Revisar y sintetizar investigaciones previas (estado del arte) en diagnóstico de fallas para turbinas y equipos rotativos.
- Preparar y enriquecer el dataset seleccionado para entrenamiento y validación.
- Diseñar pipelines de preprocesado, extracción de características y modelos (ML/DL) adecuados.
- Evaluar modelos con métricas robustas y analizar interpretabilidad.

# 2. Título y justificación

**Título (propuesto):** *Detección de fallas de motores en turbinas de gas*

Nota sobre cambio de título: el título anterior (*Diagnóstico Dinámico de Turbogrupos Rotativos de Generación Eléctrica*) se ha sustituido por uno más acorde al dataset disponible y al objetivo real del proyecto: la detección de fallos en turbinas. Este cambio se justifica por la falta de datos y metadatos suficientes para continuar con el enfoque original; por ello se adoptó el dataset de Kaggle como recurso principal y se aplicarán técnicas de aumentación y síntesis de datos cuando sea necesario.

# 3. Dataset

**Nombre:** Gas Turbine Engine Fault Detection Dataset.

**Motivación de su uso** El dataset de Kaggle se adoptó por ofrecer registros suficientes (sensores operacionales) y por su accesibilidad. Dado que el corpus original del proyecto carecía de ejemplos etiquetados y de metadatos suficientes para entrenar modelos con generalización adecuada (contaba con menos de 30 entradas), se decidió usar este dataset como fuente principal y aplicar técnicas de aumentación y síntesis para compensar la escasez y el desbalance de clases.

# 4. Cómo interpretar la tabla (State of the Art)

La tabla del estado del arte que acompaña este informe tiene la forma de una matriz donde:

- **Cada columna** representa un artículo / trabajo revisado (título del paper en la cabecera).
- **Cada fila** (campo *Característica (Fila)*) representa un atributo extraído de los trabajos: autor, año, algoritmos usados, tipo de objeto diagnosticado (rodamiento, rotor, generador), señales analizadas (v. gr., vibración) y otras observaciones concretas extraídas de cada trabajo.
- **Celdas vacías** indican que el trabajo no reportó explícitamente esa característica o que no aplica.

## 5. Qué se representa en la tabla

La tabla sintetiza, para cada trabajo revisado, las características que se pudieron extraer de las fuentes consultadas. Dado que los artículos no reportan uniformemente la misma información, la tabla combina filas normalizadas (atributos estandarizados) y notas específicas por paper. En particular:

- **Identificación bibliográfica:** título, autor(es) y año (estos campos aparecen de forma consistente).
- **Técnicas y algoritmos:** se registran categorías generales como redes neuronales (ANN/DL), modelos de ensamble/árboles, SVM, etc.; la especificación más fina (p. ej. CNN vs LSTM) no aparece de forma sistemática en todas las fichas.
- **Objeto de diagnóstico:** entradas como turbo-grupo/generador y componentes rotativos (p. ej. rodamientos/rotor) están presentes en la tabla.
- **Señales y análisis:** la tabla incluye explícitamente *vibración* como tipo de señal en una fila normalizada; otros tipos de señal (corriente, temperatura, velocidad) aparecen, si acaso, de forma puntual en celdas individuales pero no como filas estandarizadas.
- **Preprocesado y mejora:** aparecen filas normalizadas para elementos como reducción de atributos, fusión de dominios y referencias a optimizadores (p. ej. PSO). Otras técnicas de preprocesado pueden estar descritas solo dentro de celdas de papers concretos.
- **Interpretabilidad y observaciones:** existe una fila relativa a enfoques de interpretabilidad y varias celdas contienen notas sobre interpretabilidad o recomendaciones; sin embargo, limitaciones y recomendaciones no están normalizadas en filas comunes.

## 6. Estado del arte – síntesis y hallazgos principales

A partir de la lectura y comparación de los trabajos (tabla adjunta), se extraen las siguientes observaciones, especificando qué aspectos están realmente reflejados en la tabla:

- **Tendencia hacia métodos basados en aprendizaje profundo:** la tabla recoge entradas que mencionan redes neuronales y enfoques de aprendizaje profundo (ANN/DL y referencias a arquitecturas profundas).
- **Preprocesado y extracción de características:** aunque la literatura general tiende a usar transformaciones tiempo–frecuencia y extracción estadística, esas técnicas *no* aparecen como filas normalizadas en la tabla.
- **Enfoques híbridos y fusión de dominios :** la tabla incluye al menos un trabajo que presenta explícitamente fusión tiempo–frecuencia / fusión de dominios.
- **Interpretabilidad:** existen varias entradas en la tabla que hacen referencia a interpretabilidad y a enfoques XAI (por ejemplo, revisiones y trabajos que discuten XAI).

## 7. Metodología propuesta

A modo de plan técnico, podemos basarnos en el siguiente plan de acción para el desarrollo del proyecto:

## 7.1. Ingesta y limpieza

- Carga del dataset Kaggle y verificación de integridad.
- Tratamiento de valores faltantes y normalización por estación de operación.
- Segmentación temporal (ventanas/frames) adecuada al régimen de la máquina.

## 7.2. Preprocesado y extracción de características

- Filtrado y atenuación de ruido.
- Extracción de features en dominio temporal y frecuencial (cuando estén disponibles o sean relevantes).
- Uso de librerías: `numpy`, `scipy.signal` para extracción y procesamiento.

## 7.3. Aumentación de datos (Data Augmentation)

Dada la escasez y desbalance, existen varias propuestas que podemos utilizar:

- **Jittering**: añadir ruido gaussiano pequeño a las señales.
- **Scaling**: variación de magnitud para simular condiciones de medición distintas.
- **Time warping / time stretching**: ligero escalado temporal para simular variaciones de régimen.
- **Window slicing / permutation**: cortar y recombinar segmentos para crear nuevas ventanas.
- **Generación sintética**: uso de SMOTE para características tabulares; uso de GANs o modelos generativos para series temporales si se justifica.
- **Fusión de datos**: combinar ejemplos reales con sintéticos y validar con hold-out estratificado.

## 7.4. Modelado

- Modelos candidatos: Random Forest, XGBoost, SVM (baseline), CNN 1D, LSTM/GRU, arquitecturas híbridas (CNN+LSTM).
- Selección y optimización de hiperparámetros (GridSearch / RandomSearch / Bayesian Opt).
- Entrenamiento con validación cruzada estratificada y validación temporal cuando aplique.

## 7.5. Evaluación y métricas

- Métricas recomendadas: precisión, recall, F1-score, matriz de confusión, tiempo de detección promedio...
- Evaluar robustez frente a ruido y a variaciones operativas.
- Uso de técnicas de interpretabilidad para explicar decisiones del modelo.

## 8. Conclusiones provisionales

- La literatura muestra presencia de soluciones basadas en DL y de trabajos que discuten interpretabilidad.

- La disponibilidad de datos etiquetados y la consistencia en los protocolos de evaluación permanecen como retos para comparar enfoques.
- La estrategia de emplear el dataset de Kaggle y complementar con aumentación y generación sintética es razonable y práctica para el objetivo del proyecto (esta estrategia es una recomendación metodológica; la tabla contiene evidencia parcial sobre la adopción de ciertos enfoques, pero no documenta de forma completa el uso de aumentación).

## Anexo – Enlaces y recursos

- Gas Turbine Engine Fault Detection Dataset (Kaggle): <https://www.kaggle.com/datasets/ziya07/gas-turbine-engine-fault-detection-dataset>
- Herramientas recomendadas: Python (pandas, numpy, scipy), scikit-learn, Tensor-Flow/PyTorch, herramientas vistas en clases prácticas.

## Referencias

- [1] Zhang (2010). *Turbo-Generator Vibration Fault Diagnosis Based on PSO-BP Neural*.
- [2] Sabani (2024). *Bearing Faults Diagnosis of a Turbogenerator System Using Machine Learning Training Functions*.
- [3] Leite (2024). *Fault Detection and Diagnosis in Industry 4.0: A Review on Challenges and Opportunities*.
- [4] Xing (2025). *Time-Frequency-Domain Fusion Cross-Attention Fault Diagnosis Method Based on Dynamic Modeling of Bearing Rotor System*.
- [5] Maged (2024). *Explainable Artificial Intelligence Techniques for Accurate Fault Detection and Diagnosis – A Review*.
- [6] Tran (2024). *The Use of eXplainable Artificial Intelligence and Machine Learning Operation Principles to Support the Continuous Development of Machine Learning-Based Solutions in Fault Detection and Identification*.
- [7] Bechara (2024). *Review of Artificial Intelligence Methods for Faults Monitoring, Diagnosis, and Prognosis in Hydroelectric Synchronous Generators*.
- [8] *Investigación de modelos.pdf*. (Resumen de investigación realizada basada en modelos de predicciones utilizados en este campo).