

Análisis de Fallas Mecánicas Internas en Motores AC Utilizando Imágenes Termográficas

Banda Moreno, Carlos
Departamento de Automatización
Universidad de Piura
Piura, Perú
carlos.banda@alum.udep.edu.pe

Chunga Shimokawa, Diego
Departamento de Automatización
Universidad de Piura
Piura, Perú
diego.chunga.s@alum.udep.edu.pe

Farias Guerrero, Marcelo
Departamento de Automatización
Universidad de Piura
Piura, Perú
marcelo.farias@alum.udep.edu.pe

Imán Miranda, Anthony
Departamento de Automatización
Universidad de Piura
Piura, Perú
anthony.iman@alum.udep.edu.pe

Serrepe Castillo, Joan
Departamento de Automatización
Universidad de Piura
Piura, Perú
joan.serrepe@alum.udep.edu.pe

Abstract—This paper explores various techniques for the identification of mechanical failures in an alternating current motor using thermographic imaging. The methodology focuses on the thermal analysis of components to identify abnormal temperature patterns that suggest indications of failure. To do this, a set of images from public and private sources were obtained and preprocessed using techniques such as segmentation, feature extraction, contrast enhancers and filtering to improve the quality of the image set. Subsequently, the images are divided into a training and validation set in order to train different classification models such as convolutional neural networks, VGG16, ResNet50, ViT, ERT, SVM and K-NN. Preprocessing techniques such as CLAHE, DWT and other contrast enhancement techniques are also used. In this context, the model which is a combination of HSI, DWT and SVM has proven to be particularly effective achieving outstanding results in terms of accuracy and speed of classification.

Index Terms—CNN, thermographic images, Visual Transformers, induction motor, sorting algorithms and transformers.

I. INTRODUCCIÓN

Los motores de corriente alterna (ACM) son elementos fundamentales en una amplia variedad de aplicaciones industriales, sin embargo, no están exentos de sufrir fallas que puedan afectar su rendimiento y, en casos más graves puedan generar interrupciones en las operaciones en las cuales estén involucrados, lo que conlleva a pérdidas económicas. Las fallas mecánicas internas en los ACM pueden manifestarse de diversas maneras como sobrecalentamiento, vibraciones y ruidos anómalos y/o pérdidas de potencia, si estas fallas no se detectan y corrigen a tiempo pueden derivar a pérdidas económicas como se mencionó anteriormente, es por todo esto que la capacidad de identificar tempranamente estos problemas es de suma importancia no solo para asegurar la continuidad de operación, sino también para garantizar la máxima eficiencia operativa del proceso que se esté llevando a cabo. Una de las técnicas emergentes para la detección de fallas mecánicas en ACM es el uso de imágenes termográficas, estas imágenes muestran la distribución de temperaturas en la superficie del

motor y al analizarla es posible identificar patrones de calor anómalos que pueden ser indicios de fallas internas. Existen muchos tipos de fallas internas en un ACM, una de las clasificaciones más usadas, es según su origen, tales como fallas térmicas, electromagnéticas, mecánicas, etc. Siento la falla más común las del tipo mecánico como desequilibrio entre piezas, desalineamiento de los ejes, fallas en los rodamientos, en las barras del rotor, etc. El presente trabajo, tiene como objetivo principal explorar la aplicación de la termografía infrarroja para la detección de fallas en motores de corriente alterna. Dentro del trabajo, se buscará identificar las fallas más comunes, analizar las imágenes obtenidas y evaluar la eficacia de esta técnica como herramienta de diagnóstico en el entorno industrial.

II. MARCO TEÓRICO

Los motores de inducción son un tipo de ACM ampliamente usados en la industria, es por eso que en muchas ocasiones funcionan bajo condiciones no favorables, lo que ocasiona en muchos casos la aparición de fallas internas, para evitar estos problemas, la implementación de programas de mantenimiento es esencial, dentro de estos, la detección pronta de fallas en los ACM es uno de los factores clave con miras a optimizar recursos y disminuir costos. En este contexto, los métodos empleados más comunes han sido el análisis de vibraciones, la firma de corriente del motor e inspecciones visuales y/o auditivas. Aunque los métodos antes mencionado han presentado una gran efectividad, también tienen algunas limitaciones, como que pueden resultar invasivos y que además pueden ser susceptibles a interferencias externas, por lo que se surge como alternativa no invasiva, el uso de imágenes termográficas, a través de las cuales es posible analizar las distribuciones de temperatura superficial mediante el uso de redes neuronales entre otras técnicas, y se puede inferir el estado del motor. Este proceso conjunto (análisis termográfico + clasificador) nos permitirá la detección de fallas basadas en

patrones térmicos anómalos desarrollados en presencia de un problema mecánico y eléctrico, e inclusive distinguir entre las mismas; para la obtención y buena ejecución de esta proceso serán necesarias técnicas de procesamiento de imágenes antes y después de ingresarlas al modelo clasificador, las cuales serán explicadas en los siguientes puntos del presente apartado. Por ello, para el adecuado entendimiento de este documento, es necesaria la previa definición de ciertos términos y palabras consideradas fundamentales en cuanto al tema de motores y todo lo relacionado.

A. Redes Neuronales

Una red neuronal es una técnica de inteligencia artificial, que tiene como objetivo enseñar a las computadoras a procesar datos de una manera similar a como lo haría el cerebro humano, utiliza nodos o neuronas interconectadas en una estructura organizada en capas. [3]

B. Motores de corriente alterna

Dispositivos electromecánicos que convierten energía eléctrica en mecánica, son ampliamente usados en industrias y en variedad de aplicaciones comerciales como bombas, ventiladores, compresores y sistemas de transporte. Dentro de esta clasificación se encuentran los motores de inducción de jaula de ardilla y de rotor bobinado, así como los motores síncronos.

C. Fallas

Defectos, averías o condiciones anómalas que afectan el funcionamiento óptimo de un activo, en el contexto de los motores AC, estas pueden ser de naturaleza mecánica o eléctrica.

1) *Fallas Mecánicas (MF)*: Son causadas por problemas físicos en los componentes mecánicos del motor, entre estas se incluyen:

- **Desgaste en los rodamientos (BD)**: Cuya principal causa es la fricción y la fatiga del material, provoca vibraciones excesivas que en cierto punto afectan notablemente la eficiencia del motor.
- **Desalineamiento (MSL)**: Ocurre si las partes móviles del motor no están alienadas adecuadamente, también pueden causar vibraciones excesivas, aceleramiento en el desgaste de los componentes, etc.

2) *Fallas Eléctricas (EF)*: Ocurren por anomalías en los circuitos eléctricos o componentes electromagnéticos del motor, entre estas se incluyen:

- **Cortocircuito en las bobinas del estator (ECF)**: Causado por el desgaste excesivo del aislamiento entre bobinas, permitiendo el contacto entre dos conductores, puede provocar un flujo de corriente no deseado, capaz de dañar el motor.
- **Rotura de barras del rotor (BRB)**: Este tipo de fallas es común en motores de jaula de ardilla, se refiere a la ruptura de una o más barras conductoras del rotor, es generalmente causada por fatiga o vibraciones anómalas, generan una pérdida de torque que puede llegar a ser significativa.

D. Radiación

La radiación es la energía emitida por la materia en forma de ondas electromagnéticas (o fotones) como resultado de los cambios en las configuraciones electrónicas de los átomos o moléculas, no requiere la presencia de un medio interventor. [5]

E. Termografía por infrarrojos (TI)

Es la ciencia que estudia el uso de dispositivos opto-electrónicos para detectar y medir la radiación, para posteriormente obtener la temperatura de la superficie bajo análisis. Los dispositivos que nos ayudan a la obtención de estas imágenes son las cámaras termográficas. [9]

F. Preprocesamiento de imágenes (IPP)

Conjunto de técnicas que se aplican a las imágenes antes de ejecutar cualquier análisis o procesamiento complejo. Su objetivo es mejorar la calidad, eliminación de ruido y/o resaltar características.

1) *Filtros*: La aplicación de filtros es una de las principales técnicas en el procesamiento de imágenes digitales, se pueden aplicar en el dominio del espacio o en el dominio de la frecuencia, tienen como objetivo suavizar la imagen, eliminar ruido, detectar bordes, etc. [25]

2) *Contrast Limited Adaptive Histogram Equalization (CLAHE)*: Método que construye un histograma de intensidad por pixel, para luego aumentar la longitud de cada valor, de esta manera logra aumentar el contraste en toda la imagen.

3) *Transformada de Wavelet Discreta (DWT)*: Es una herramienta matemática utilizada para la compresión tanto de imágenes como de audio, es una descomposición de señales en frecuencias. Permite analizar imágenes en diferentes escalas y frecuencias. [4].

4) *Análisis de Componentes Independientes (ICA)*: Es un método computacional utilizado para separar una señal multivariante en subcomponentes independientes constituyentes, se utiliza en una amplia gama de aplicaciones, como el procesamiento de señales, reconocimiento de patrones, procesamiento de imágenes, etc. [1]

5) *Segmentación de imágenes (IS)*: Es una técnica de visión por computadora que separa una imagen digital en distintos grupos de píxeles (segmentos) para facilitar la detección de objetos, entre otras tareas [12]. Existen numerosas técnicas de segmentación:

- **Umbralización**: Clasifican píxeles en función a si su intensidad está por encima o por debajo de un determinado valor umbral. Uno de los métodos más usados para determinar el valor umbral es el método de Otsu.
- **Histogramas**: Representan la frecuencia de los valores de píxeles en la imagen, se usan para determinar umbrales.
- **Detección de bordes**: Identifican límites de objetos o clases a través de la existencia de discontinuidades o contrastes.
- **Algoritmos de watershed**: Usan las imágenes convertidas a escala de grises y generan un mapa topográfico en base a la luminosidad de los píxeles.

- **Segmentación basada en clústeres:** Se dividen los datos visuales en grupos de píxeles con valores similares, una de las variantes más comunes es la técnica K-means.

6) *Bag of Visual Words (BoVW)*: Es una técnica usada en trabajos de visión por computadora y aprendizaje automático, representa una imagen como un conjunto de palabras visuales que extraen características notables en la imagen (bordes, esquinas, texturas, etc.). [22]

7) *Speeded-UP- Robust Features (SURF)*: Es un algoritmo de detección de características en imágenes, tiene como objetivo identificar características significativas que puedan ser usadas en tareas de comparación, coincidencia, etc. Algunas de sus principales ventajas son su rapidez y robustez.

8) *Hue Saturation Intensity (HSI)*: Modelo descriptivo de color similar al modelo RGB, donde se describen valores de los mismos, pero esta vez en base a las características de Intensidad, saturación y tono.

9) *Linear Discriminant Analysis (LDA)*: Técnica para filtrar características de forma que las variables de ingreso sean reducidas en una dimensión que representa mucho mejor la dinámica de un proceso o imagen.

10) *Principal Component Analysis (PCA)*: Es un método que mediante el análisis de una matriz de tamaño definido que almacena varios valores correspondientes a las características extraídas de una o más imágenes, calcula la matriz de covarianza, de la cual, mediante el cálculo de autovalores y vectores propios, tras lo cual puede determinar los grupos de características más representativas de la matriz original.

G. Clasificadores en base a características (IFE)

Es la aplicación de técnicas con el objetivo de reducción de la complejidad y dimensionalidad de una imagen gracias a la extracción de características relevantes. [2]

1) *Redes neuronales convolucionales (CNN)*: Son arquitecturas de red para Deep Learning (DL) eficaces para la detección de patrones en imágenes, con el fin de reconocer objetos, clases y categorías, poseen capas de convolución en su estructura, estas capas pueden aplicar filtros convolucionales a las imágenes, para activar diferentes características de estas [17], entre estas arquitecturas encontramos:

- **VGG16:** Desarrollada por Visual Geometry Group (VGG), presentada en 2014, se basa en 13 capas convolucionales y 3 capas completamente conectadas.
- **ResNet (Residual Networks):** Desarrollada por ImageNet en 2015, su principal característica es el uso de bloques residuales, que permitan el armado de redes con muchas capas, existiendo ResNet-50, ResNet-101 o ResNet-152, con 50, 101 y 152 capas respectivamente.
- **Inception (Google Net):** Desarrollada por Google, introduce el uso de bloques inception, permitiendo usar filtros de múltiples tamaños al mismo tiempo.
- **MobileNet:** Arquitectura ligera, pensada para equipo con recursos limitados o móviles, utiliza convoluciones separables, para reducir el número de parámetros.

2) *K-Nearest Neighbors (K-NN)*: Es un clasificador de aprendizaje supervisado no paramétrico, que realiza clasificaciones o predicciones sobre la agrupación de un conjunto de datos, en base a la proximidad.

3) *Support Vector Machine (SVM)*: Es un algoritmo de aprendizaje supervisado que se usa en problemas de clasificación y regresión, su objetivo es la obtención de un hiperplano que separe dos clases de puntos de datos.

4) *Random Tree (RT)*: Es un modelo de aprendizaje automático utilizado para predicción y clasificación, que se caracteriza por emplear un enfoque aleatorio en su construcción, seleccionando de manera aleatoria un subconjunto de datos y características en cada fase del proceso. Su principal ventaja es que ayuda a prevenir el sobreajuste del modelo a datos específicos, mejorando así su capacidad para hacer predicciones más generales y precisas.

5) *Visual Transform (ViT)*: Modelo que adapta la arquitectura de Transforms para poder recepcionar imágenes, mediante el agregado de un paso previo donde la imagen es segmentada en un grupo finito de partes a los cuales se les conoce como batches y se le adapta de tal forma que pueda ser ingresada al Transformer Encoder. Para este método en particular y su correcta comprensión es importante tener en cuenta los siguientes conceptos claros:

- **Multilayer Perceptron (MLP):** Tecnología basada en el funcionamiento de una neurona, conforma la base de todos los modelos basados en Deep Learning.
- **Feed Forward:** Metodología de aprendizaje que permite que los MLP aprendan por cada "epoch" entrenada.
- **Patch Division:** División de una imagen en patches o imágenes más pequeñas, los cuales contienen un número menor de información.
- **Positional Encoding:** Matriz que guarda el posicionamiento de cada Patch en una sola matriz.
- **Encoder Stack:** Herramienta encargada de aplanar un input cualquiera y convertirlo en un vector representativo.
- **Decoder Stack:** Herramienta que convierte un vector representativo en un output o vector probabilístico que sirva como predicción de alguna clase.
- **Auto Attention:** Función encargada de relacionar N inputs con M outputs de tal forma que para una predicción $M + 1$ discierne que N input es más importante para realizar la predicción.
- **Multi Heading Self Attention:** Múltiples funciones Auto Attention conectadas en paralelo.

III. ESTADO DEL ARTE

En los últimos años, se ha profundizado en la detección de fallas en motores mediante el uso de las imágenes termográficas y de técnicas avanzadas de procesamiento de imágenes y algoritmos de aprendizaje automático. En esta sección, se presentará la literatura más relevante y reciente relacionada a la detección de fallas en los ACM mediante el uso de imágenes termográficas, analizando la metodología propuesta por los mismos, resaltando que técnicas de proce-

samiento y algoritmos de clasificación han sido los más efectivos.

En [14], describen como la inteligencia artificial y el aprendizaje profundo se están empleando para mejorar las técnicas convencionales de diagnóstico de fallas en motores de inducción a medida que estos se vuelven más comunes. Resaltan el gran potencial para la exploración en estudios futuros y la dificultad de detectar fallas en tiempo real actualmente por los métodos laboriosos de descomposición de señales.

Con respecto al procesamiento de imágenes térmicas, el modelo de color HSV considera la región de tono (HUE) del modelo HSV debido a la utilidad para evaluación y medición del color de un objeto como la rojez o amarilleo predominantes en la imagen térmica, además de técnicas para detección de bordes como Roberts, Prewitt y Otsu con el fin de segmentar las imágenes térmicas de equipos eléctricos defectuosos. La comparativa de rendimiento de umbralización de estos métodos se realizó calculando el error cuadrático medio y la relación señal ruido de pico siendo el método de Otsu el más efectivo para determinar las regiones más calientes en este tipo de imágenes como se observa en [10].

Si bien es cierto, trabajar con el color de las imágenes térmicas resaltando las zonas de calor es una buena técnica de preprocesamiento, no se toma en cuenta el ruido visual de la imagen, entonces, la transformada discreta de Wavelet en dos dimensiones (2D-DWT) mejora la relación señal-ruido para luego reducir la dimensionalidad de características a extraer identificando las más relevantes. Después, calcula la MD (Distancia de Mahalanobis) para clasificar las características según su relevancia descartándose las menos relevantes. Todos estos datos son usados para entrenar un CDT (Complex Decision Tree) para predecir el estado de los rodamientos basándose en un esquema de validación cruzada, obteniéndose una precisión de clasificación de 99.85% [13].

En contraste, en [27] diseñan un método híbrido para la detección de fallas en los rodamientos de motores de inducción cuyo preprocesamiento se basa en el uso de CLAHE logrando una mejor visualización de los detalles en las imágenes térmicas para resaltar los gradientes de calor y anomalías de temperatura para luego ser usadas en Inception V3 con un módulo SE integrado en su arquitectura para potenciar la extracción de características y la última capa de clasificación de Inception V3 se reemplaza por un algoritmo (SVM) mejorando la capacidad de representación y generalización del modelo siendo este capaz de clasificar hasta 11 tipos diferentes de fallos de motor con una gran precisión acelerando el proceso de aprendizaje y estabilizando la pérdida durante el entrenamiento.

En [16] se presenta una nueva metodología para la clasificación de fallas en motores de inducción trifásicos. Como técnica de preprocesamiento de imágenes proponen el uso de BoVW (Bag of Visual Words) y SURF (Speeded-UP-Robust Features) para la extracción de características visuales. Siendo el modelo de clasificación empleado ERT (Extremely Randomized Tree) resultando una gran precisión y estabilidad en la detección y clasificación de fallas superando a métodos

tradicionales como KNN, SVM y DT.

Terminado el preprocesamiento de las imágenes, se debe considerar un clasificador entre las fallas de los motores y el estado óptimo de operación de la máquina, entre ellos, los métodos basados en aprendizaje profundo utilizando imágenes térmicas pueden clasificar de manera efectiva las condiciones de los motores de inducción con una alta precisión y baja tasa de error. VGG-16 y ResNet-16, dos arquitecturas de redes neuronales convolucionales surgen como las alternativas más fiables en este campo siendo la primera superior a la otra debido a su gran capacidad de aprendizaje profundo y manejo efectivo del problema de la desaparición del gradiente.

Del mismo modo, el uso de una red neuronal convolucional puede demandar un alto costo computacional, así [23] introducen un enfoque de aprendizaje utilizando SqueezeNet modificada, un modelo de CNN de solo 7 capas con 16 veces menos parámetros que la original reduciendo significativamente el tiempo de entrenamiento y uso de recursos computacionales haciéndolo ideal para dispositivos móviles o sistemas embebidos. Del mismo modo, se diseñó una red prototípica aplicando Few-Shot Learning para que el modelo aprenda a clasificar fallos eficazmente con pocos ejemplos. En este estudio se llegó a la conclusión que la implementación de modelos más ligeros no necesariamente implica pérdidas significativas en la eficiencia en la detección de fallas en motores eléctricos usando imágenes térmicas.

Estos resultados han sugerido que los métodos basados en DL pueden predecir eficientemente las condiciones del motor incluso cuando se entrenan con datos pequeños y desbalanceados. [19]

Pese a ello, algunas arquitecturas con pesos pre-entrenados pueden presentar problemas. Entonces, se puede preparar el Dataset con el fin de forzar la falla de manera extrema y que esta se observe de manera evidente en la cámara infrarroja logrando una precisión de 95% empleando técnicas de aumento de datos para simular escenarios reales, la presencia de ruido y la precisión del modelo. [24]

Aparte de la metodología mencionada anteriormente, existen algoritmos de clasificación que usan las características recopiladas en la etapa de preprocesamiento. Por ejemplo, la extracción de FOS (First Order Statistics) como datos extraídos de las imágenes. En [21], realizan una comparativa entre clasificadores como redes neuronales, KNN (K-Nearest Neighbors), LR (Logistics Regressions) y SVM (Support Vector Machine). Llegando a la conclusión de que para el uso de FOS, se destaca RF (Random Forest) como mejor clasificador debido a su equilibrio entre precisión y estabilidad.

Por otro lado, en [20] basan su modelo en tres niveles de procesamiento incluyendo la mayoría de los puntos resaltados por los anteriores autores. En el nivel bajo normalizan la data recortando, rotando y ampliando las imágenes y utilizando ecualización de histograma para mejorar el contraste de las imágenes. En el nivel medio se emplea una red de segmentación semántica con SegNet pre-entrenada con pesos de VGG-16. Para esto, se etiquetan y clasifican las imágenes en 3 clases: fondo, bobinado y falla. Posterior a ello, se

entrena y valida la red hasta alcanzar una precisión del 99.29% destacando regiones de interés en las áreas calientes de los devanados. Finalmente, en el procesamiento de alto nivel se establecen tres variables de temperatura como entradas a un sistema de interferencia difusa clasificando el motor en tres niveles: regular, alerta y alarma.

En la búsqueda de solucionar el problema de la pérdida de información en las RNN durante el entrenamiento, en ese contexto surgió la función Attention como una forma de relacionar los inputs y outputs, de esta forma la pérdida de información es mínima y la eficiencia del modelo aumenta. Este nuevo método llamado Transformer permite a los elementos analizados, llamados tokens, “elegir” a que otros tokens prestar atención, logrando de esta forma acelerar el proceso y no perder relación entre patches y aun mejor combinado con el uso de las funciones Attention en paralelo (Multi-Head Attention). [26]

Es por ello que en [8] se explora la posibilidad de usar la arquitectura propia de una Transformer en el ámbito de computer vision para tareas de clasificación mediante el uso de una división en patches de la imagen original con el objetivo de reducir el peso computacional del modelo y agilizar las operaciones; asimismo usa un Position Embedding que sirva de índice para cada patch, con estos 2 valores y tras ser sometidos a un aplanamiento, se pueden aplicar un Multi-Head Self Attention (MSA) que permite predecir con criterios de relación, es decir aprende a saber que valores previos son los que deben influir más en la predicción futura específica. Este método es ampliamente usado para Datasets muy grandes dada su gran eficiencia además de robustez, y es este paper en conjunto con [7] que demostró que su empleo es extrapolable a usos que se pensaban imposibles por su costo computacional.

Sin embargo, los Transformers destacan de gran manera cuando se les entrena con Datasets grandes, esto es con respecto a otros modelos como por ejemplo las CNN, mientras que en Datasets pequeños no se nota una gran diferencia de rendimiento, es por ello que [15] usa 2 cosas: un tipo de partición de data en Patches especial denominado Shifted Patch Tokenization (SPT) donde mediante un conjunto de pasos establecidos amplía el número de patches, un ajuste de la función Self-Attention para hacer que se enfoque solo en los tokens relevantes (no nulls), eliminando la posibilidad de un token de construir relaciones con dichos tokens nulls.

A pesar de que las ViT según la bibliografía consultada se desempeñan mejor que las CNN, por ejemplo, en el caso de contar con datasets grandes, no dándose el mismo caso para datasets pequeños, por lo cual en [11] se presenta el concepto de Compact Convolutional Transformer (CCP), el cual se basa en no usar una simple división de la imagen para la separación en patches, sino en aplicar sobre estos patches una convolución que ayude a captar características locales similares a como lo hace una CNN, además de conservar la información espacial de la imagen.

Otros enfoques válidos para afrontar esta situación es la que se plantea en [6] donde se centran en una doble tokenization a escalas diferentes, es decir una imagen, la tokeniza o la

divide en patches grandes y además hace el mismo proceso pero para patches más pequeños, es decir obtiene 2 tokens por imagen, de esta forma captura no solo la información global de la imagen, sino también la información local de los patches pequeños, ambos tokens son ingresados a un encoder propio el cual es conocido como cross-attention, y según los resultados observados, se notó un aumento en la eficiencia del modelo con este nuevo enfoque.

IV. METODOLOGÍA

En esta sección se detallará el procedimiento llevado a cabo, así mismo los materiales tanto de hardware como de software que se usaran para llevar a cabo la experimentación. Cabe señalar, que el presente trabajo es de carácter investigativo-experimental, por lo que tras identificar una problemática y encontrar una solución novedosa, se propone revisar las herramientas actuales, experimentar con las mismas, con el fin de determinar la combinación adecuada de herramientas y modelos para obtener el mejor resultado. Es por ello que la metodología se ha dividido en 3 etapas: Investigación, experimentación, validación.

A. Investigación

En esta etapa inicial, se recopiló literatura relacionada con el tema en cuestión, con el fin de determinar cuales serán los objetivos, alcances y posibles técnicas a usar, para ello se consultaron un conjunto de portales web de talla nacional e internacional relacionados con la investigación científica, tales como: IEEEExplore, Scopus, ResearchGate, etc. En esta etapa, se incluye la búsqueda del repositorio de imágenes térmicas usadas en el desarrollo tanto públicas (repositorio online) como propias (tomadas en el campus de la Universidad de Piura (UDEP)), la definición de técnicas de preprocesamiento y algoritmos de clasificación.

B. Experimentación

Esta etapa inicia una vez concluida la investigación, aquí se llevarán a cabo experimentos con el objetivo de obtener una combinación de preprocesamientos y algoritmos de clasificación óptima, todo esto basado en la literatura antes consultada.

C. Validación

En esta última parte, se realiza un análisis de validación, para ello se evalúan las estadísticas que representan el buen o mal desempeño de las combinaciones antes planteadas, para ellos nos apoyamos de criterios y parámetros como: precisión, matriz de confusión, pérdida, peso computacional y rapidez de entrenamiento.

V. DESARROLLO

Una vez que ya se ha explicado de manera general cuáles serán los pasos y criterios por seguir, podemos pasar a resumir de manera específica que Datasets, métodos de preprocesamiento, clasificadores, criterios de evaluación se han usado y que tal ha sido su desempeño en el transcurso del trabajo.

A. Datasets Revisados

Con respecto a los Datasets consultados y obtenidos podemos enumerarlos mediante la siguiente tabla:

TABLA I
DATASET REVISADOS

Dataset	Fuente	Img	Clases	Observación
Data 1	Eksploracja i Niezawodność Journal	10007	105	Las fallas se han evaluado para distintos rangos de corriente de funcionamiento y se duplican, pues se posee 2 fuentes fotográficas para una misma falla.
Data 2	Babol Noshirvani University of Technology	369	11	Las imágenes son obtenidas de un motor en funcionamiento libre, es decir sin carga alguna.
Data 3	Universidad Autónoma de Querétaro	80	4	Las fallas inducidas son muy poco realistas.
Data 4	Universidad de Piura	479	1	Solo imagenes de motores sanos.

Con respecto a los Datasets es necesario aclarar que de todos los antes mencionados, solo se han usado las Data 1 y 2 para el entrenamiento de los modelos clasificadores, puesto que es el que se ha considerado cumple todos los requerimientos antes planteados, a su vez también se debe esclarecer que para solucionar el problema del exceso de clases, se ha considera plausible la unificación de varias clases y el descarte de otras muchas en la Data 1, siguiendo los 2 siguientes criterios: Unir todas las imágenes sin distintivo de corriente de funcionamiento para el caso de una misma falla en el primer caso (unificación) y descarte de las imágenes de 1 de las cámaras, pues la otra presenta mejor resolución (descarte) en el segundo caso. Habiendo entendido este contexto, se presenta a continuación la estructura de la data objeto de estudio en la figura 1.

Llegados a este punto, es importante resaltar que para las pruebas de algunos de los modelos que se han llevado a cabo se consideró que el acoplamiento Coupling 1 no tenía el suficiente número de clases por lo que se trabajó únicamente con las imágenes del Coupling2, mientras que en otros modelos si se usaron, pues nos puede servir como segundo valor de validación, ya que a pesar de pertenecer al mismo dataset, presentara una dinámica diferente de interacción que las imágenes del “Coupling2”, posee una clase de “Missalingment”, por lo que la correcta o no identificación de

esta clase con el modelo creado a partir de las otras imágenes puede ayudar a su mejora continua.

B. Métodos de preprocesamientos

De la misma forma, en las tablas II y III se describe todos los métodos de preprocesamiento revisados y testeados, así como cuál fue la apreciación que se concluyó al usarlos.

TABLA II
PREPROCESAMIENTOS (PARTE I)

Preprocesamiento	Tipo	Observación
Otsu	Segmentación	Eficiente pero solo solo puede “Diferenciar entre 2 clases”.
FFT	Diferenciación	Permite acrecentar contrastes y de esta forma acrecentar las diferencias por clase.
YOLOV5	Segmentación	Muy útil, preciso y ademas logra identificar ACM en imágenes termográficas con una muy buena precisión.
CLAHE	Diferenciación	Logro acrecentar contrastes en la imagen, sin embargo, depende un filtro pasa alto previo para mejorar dicho proceso.
DWT	Extracción de Características	Mejora el rendimiento del clasificador SVM, puesto que le proporciona un data input mucho más adecuado.
SIFT-BoVW	Extracción de Características	Permite realizar una segmentación, pues identifica muy bien limites, sin embargo, es en conjunto con su método clasificador el que peor precisión ha obtenido.
K-means	Segmentación	Menor rendimiento que Otsu, pero permite “Diferenciar entre más de 2 clases”.
FOS	Extracción de Características	Sirve para identificar patrones de intensidad por clase, también sirve para un posible análisis termográfico simple, pues permite convertir los valores de intensidad de pixeles en temperatura.
GLCM	Extracción de Características	Conjunto de información más completa y de interés para la tarea de clasificación.

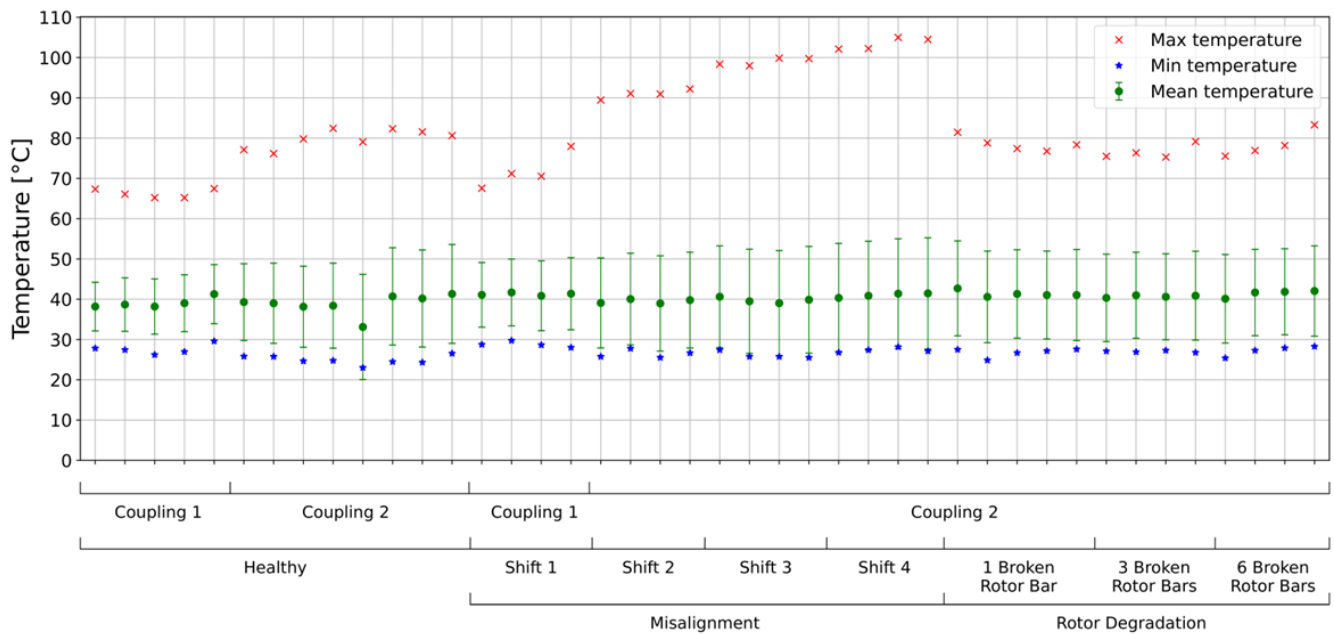


Fig. 1. Distribución de temperatura para la Data1 [18]

TABLA III
PREPROCESAMIENTOS (PARTE 2)

Preprocesamiento	Tipo	Observación
Data Augmentation	Aumento de Data	Técnica ampliamente conocida en el ámbito, ha dado buenos resultados evitando el overfitting.
LDA	Diferenciación	No significo una mejora sustancial en la precisión del modelo clasificador.
PCA	Diferenciación	No se observó una mejora significativa en los modelos evaluados.

C. Métodos de Clasificación

Tras realizar la evaluación de los métodos de preprocesamiento vistos en las tablas II y III, en las tablas IV, V y VI se menciona todos los métodos de clasificación que se han considerado, así como sus características.

TABLA IV
ENFOQUES DE CLASIFICACIÓN (PARTE I)

Método de clasificación	Estructura	Observaciones
VGG16	Concolutional neural network	Es de estructura simple y fácil de comprender.

TABLA V
ENFOQUES DE CLASIFICACIÓN (PARTE 2)

Método de clasificación	Estructura	Observaciones
ResNet50	Convolutional neural network	Ampliamente usada para labores de clasificación.
InceptionV3		Permite la reducción de convoluciones mediante factorización.
EfficientNetB2		Modelo más pesado que las otras CNN pero que no muestra una mejora comparativa con respecto a otras CNN.
EfficientNetB5		
EfficientNetB7		
ERT	Ensemble supervised machine learning method	Método de clasificación basado en la prueba y error.
SVM	Supervised learning algorithm	Es un método que solo admite entradas 1D
KNN	Non-parametric, supervised learning classifier	Dio peores resultados que cuando se usó SVM como clasificador

TABLA VI
ENFOQUES DE CLASIFICACIÓN (PARTE 2)

Método de clasificación	Estructura	Observaciones
ViT	Transformer	Excelentes resultados, pero gran peso computacional en comparación con SVM
ViTFSD	Transformer	Mejoro los resultados obtenidos en comparación con las CNN y ViT, pero demoro más en entrenar
CCT	Transformer	Buenos resultados, pero hubo una época en la que el resultado fue pésimo, pero es mucho más liviana que la ViTFSD
CrossViT	Transformer	El modelo más pesado de todos y no represento una mejora sustancial

D. Datasets de Entrenamiento y Testeo

Culminada la etapa donde se ha detallado los tipos de métodos de clasificación probables a usar, debemos establecer los protocolos de testeo y entrenamiento, pues este es un paso de igual de importante que los antes repasados, sobre todo para corroborar errores de overfitting, de esta forma tendremos los siguientes Datasets de training y testing presentados en las tablas VII, VIII, IX y X.

TABLA VII
DATASETS DE ENTRENAMIENTO (PARTE 1)

Dataset Train	Fuente	Datos	Clases
Train 1	Data 1	939 img (80%) + metadata (4GLCM x 939 img)	Healthy
	Coupling2	1480 img (80%) + metadata (4GLCM x 1480 img)	Missalingment1
Train 2	Data 1	939 img (80%) + 1878 img (Data augmentation) + metadata (4GLCM x 2817 img)	Healthy
	Coupling2	1480 img (80%) + 1480 img (Data augmentation) + metadata (4GLCM x 2960 img)	Missalingment1

TABLA VIII
DATASETS DE ENTRENAMIENTO (PARTE 2)

Dataset Train	Fuente	Datos	Clases
Train 3	Data 1	1009 img (80%) + 4 sub-bandas por imagen (DWT)	Healthy
	Coupling2	1439 img (80%) + 4 sub-bandas por imagen (DWT)	Missalingment1
Train 4	Data 1 Coupling 1 y 2	1511 img (80%)	Healthy
		3538 img (80%)	Missalingment1
Train5	Data1 Coupling2 gris	1196 img (80%)	Healthy
		720 img (80%)	Missalingment2
		888 img (80%)	Missalingment3
		942 img (80%)	Missalingment4
		148 img (80%)	RotorBar1
		363 img (80%)	RotorBar3
		390 img (80%)	RotorBar6
Train6	Data1 + Data 4	1782 img (20%) + 4 sub-bandas por imagen (DWT) + metadata (2GLCM x 1782 img)	Healthy
		3530 img (20%) + 4 sub-bandas por imagen (DWT) + metadata (2GLCM x 3530 img)	Missalingment
		901 img (20%) + 4 sub-bandas por imagen (DWT) + metadata (2GLCM x 901 img)	Missalingment1

TABLA IX
DATASETS DE TESTEO (PARTE 1)

Dataset Train	Fuente	Datos	Clases
Test 1	Data 1	322 img (20%) + metadata (4GLCM x 322 img)	Healthy
	Coupling2	319 img (20%) + metadata (4GLCM x 319 img)	Missalingment1
Test 2	Data 1	322 img (20%) + 644 img (Data augmentation) + metadata (4GLCM x966 img)	Healthy
	Coupling 2	319 img (20%) +638 img (Data augmentation) + metadata (4GLCM x 957 img)	Missalingment1
Test 3	Data 1	252 img (20%) + 4 sub-bandas por imagen (DWT)	Healthy
	Coupling2	360 img (20%) + 4 sub-bandas por imagen (DWT)	Missalingment1
Test 4	Data 1 Coupling 1 y 2	378 img (20%)	Healthy
		885 img (20%)	Missalingment1
Test 5	Data1 Coupling2 gris	299 img (20%)	Healthy
		180 img (20%)	Missalingment2
		222 img (20%)	Missalingment3
		236 img (20%)	Missalingment4
		37 img (20%)	RotorBar1
		91 img (20%)	RotorBar3
		98 img (20%)	RotorBar6

TABLA X
DATASETS DE TESTEO (PARTE 2)

Dataset Train	Fuente	Datos	Clases
Test 6	Data1 + Data 4	445 img (20%) + 4 sub-bandas por imagen (DWT) + metadata (2GLCM x 445 img)	Healthy
		883 img (20%) + 4 sub-bandas por imagen (DWT) + metadata (2GLCM x 883 img)	Missalingment
		225 img (20%) + 4 sub-bandas por imagen (DWT) + metadata (2GLCM x 225 img)	Missalingment1

En las tablas VII VIII, IX y X se aprecian que los conjuntos de entrenamiento y prueba siguen una proporción de 80-20%, sin embargo, sería interesante probar en futuros proyecto alguna otra distribución como 70-30% para observar el comportamiento de nuevos modelos, aunque tal y como se verá más adelante, no se espera una mejora significativa.

E. Modelos Ensayados

Para el ensamblado de todos los modelos se ha usado un conjunto de técnicas de preprocesamiento, así como un método de clasificación, es importante resaltar que inclusive dentro de un mismo tipo de clasificador existen variaciones, es por eso que existe un índice que lo denote.

TABLA XI
MODELOS EVALUADOS

Modelo	Preprocesamiento	Clasificador	Data
Modelo1	YOLOV5+Otsu+FFT+CLAHE+GLCM	VGG16-1	Train-test1
Modelo2	YOLOV5+Otsu+FFT+CLAHE+GLCM + Data Augmentation	ResNet50	Train-test2
Modelo3	YOLOV5+Otsu+FFT+CLAHE	SVM-1	Train-test3
Modelo4		EfficientNetB2	Train-test4
Modelo5		EfficientNetB5	

TABLA XII
MODELOS EVALUADOS

Modelo	Preprocesamiento	Clasificador	Data
Modelo6	YOLOV5+Otsu+FFT+CLAHE	EfficientNetB7	Train-test4
Modelo7	YOLOV5+Otsu+FFT+CLAHE	VGG16-2	Train-test4
Modelo8		InceptionV3-1	
Modelo9		SVM-2	
Modelo10	YOLOV5+Otsu+FFT+CLAHE	KNN	Train-test4
Modelo11		InceptionV3-2	
Modelo12	YOLOV5+SIFT-BoVW	ERT	Train-test5
Modelo13	YOLOV5+Otsu+FFT+CLAHE	VIT	Train-test5
Modelo14		VITFSD	Train-test6
Modelo15		CCT	Train-test6
Modelo16		CrossViT	Train-test6
Modelo17	YOLOV5m	ViT	Train-test6

F. Resultados

La evaluación se basará en tres métricas referenciales y en una matriz de confusión, estas proporcionan información sobre el desempeño del modelo.

TABLA XIII
RESULTADOS DE LOS ENSAYOS (PARTE 1)

Modelo	Precisión		Recall		F1-score	
	Healthy	Failed	Healthy	Failed	Healthy	Failed
Modelo 1	1	0.78	0.72	1	0.84	0.87
Modelo 2	0.98	0.79	0.74	0.99	0.85	0.88

TABLA XIV
RESULTADOS DE LOS ENSAYOS (PARTE 2)

Modelo	Precisión		Recall		F1-score	
	Healthy	Failed	Healthy	Failed	Healthy	Failed
Modelo 3	0.99	0.992	1	0.97	0.993	0.995
Modelo 4	1	0.78	0.84	1	0.91	0.88
Modelo 5	0.99	0.8	0.85	0.99	0.92	0.89
Modelo 6	1	0.9	0.89	0.89	0.91	0.9
Modelo 7	1	0.8	0.82	0.84	0.9	0.92
Modelo 8	0.98	1	1	0.99	0.99	0.99
Modelo 9	0.84	0.82	0.68	0.92	0.75	0.87
Modelo 10	0.88	0.87	0.78	0.93	0.83	0.9
Modelo 11	0.83	0.9	0.84	0.89	0.84	0.9
Modelo 12	1	0.998	0.95	0.996	0.97	0.997
Modelo 13	0.99	0.96	0.943	0.999	0.97	0.979
Modelo 14	0.995	0.997	0.995	1	0.995	0.998
Modelo 15	1	0.999	0.998	1	0.999	0.999
Modelo 16	0.997	0.998	0.997	1	0.997	0.999
Modelo 17	0.998	0.955	0.936	0.999	0.966	0.976

VI. ANÁLISIS RESULTADOS

En esta sección se repasarán los resultados obtenidos, se comentará lo aprendido del uso de cada herramienta y cuáles se han determinado como útiles y cuales no, asimismo se analizarán los resultados de los métodos clasificadores y su rendimiento en base a los criterios del capítulo previo, para finalmente establecer, en base a lo antes comentado un modelo (preprocesamiento + clasificador). Finalmente se comentarán las pruebas de validación a la que se le someterá a este último modelo y cuáles son sus limitaciones.

A. Análisis de preprocesamientos

En el preprocesamiento, se usaron 2 tipos de segmentación, una principal para detectar y recortar el ACM y una secundaria para calcular el ROI. Se uso YOLOv5 para la segmentación principal y Otsu o K-means para la segmentación secundaria, sin embargo, es posible usar YOLOv5 para ambas tareas o inclusive usar Otsu Thresholding como único método de segmentación. Pero, se planteo el uso de un método de identificación primario y después un secundario, pues de esta forma se evitaba caer en el error de calcular un ROI sin haber confirmado antes que el ACM se encontraba en la imagen alcanzada. Para la identificación primaria se usó YOLOv5m, el modelo mediano de 5ta generación de YOLO, debido a que daba los mejores resultados para los datasets usados. Para la segmentación se evaluaron 2 métodos: OTSU y K-means, ambos ofreciendo buenos resultados en un inicio, sin embargo, para tareas de segmentación más complejas se usó OTSU, ya que presentó un desempeño similar a k-means, pero es un método mucho mas simple y de bajo coste computacional como se detalla en las ecuaciones 1, 2, 3 y 4.

$$w_{1(T)} = \sum_{i=0}^T P(i) \quad (1)$$

$$w_{2(T)} = \sum_{i=T+1}^{L-1} P(i) \quad (2)$$

$$\mu_{1(T)} = \frac{\sum_{i=0}^T i \cdot P(i)}{w_{1(T)}} \quad (3)$$

$$\mu_{1(T)} = \frac{\sum_{i=T+1}^{L-1} i \cdot P(i)}{(w_{2(T)})} \quad (4)$$

Donde w_1 representa la probabilidad de que un píxel pertenezca a la clase 1, w_2 la probabilidad de que pertenezca a la clase 2. Además, μ_1 es la media de la clase 1, habiendo establecido un umbral T de tal forma que los píxeles de intensidad inferior a dicho umbral sean de la clase 1 y los de intensidad superior pertenezcan a la clase 2. Tras esto, se calcula la varianza intra-clase σ_w^2 y la varianza inter-clase σ_b^2 .

$$\sigma_w^2 = w_1\sigma_1^2 + w_2\sigma_2^2 \quad (5)$$

$$\sigma_b^2 = w_1w_2 \cdot (\mu_1 - \mu_2)^2 \quad (6)$$

Para finalmente optimizarse de tal forma que se establezca un nuevo umbral T que maximice la varianza inter-clase y minimice la varianza intra-clase.

Por otro lado, el método K-means usa una función de optimización o también conocida como función de costo (J), la cual debe ser minimizada y se representa en la ecuación 7.

$$J = \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - \mu_j\|^2 \quad (7)$$

Siendo k el número de clusters o agrupamientos, x_i el vector que indica la intensidad de cada pixel y μ_j el centroide del j -esimo cluster, que se calcula de la forma representada en la ecuación 8.

$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \quad (8)$$

Donde C_j es una agrupación de píxeles, en caso de que sean 2 clusters como en nuestro caso, solo existirán C_1 y C_2 .

Tras armar 2 clústeres iniciales en base a puntos aleatorios, de tal forma que los demás puntos o píxeles de la imagen más cercanos a un clúster u otro le pertenezcan a dicho clúster. Posterior a ello, se calcula y recalcula la función de costo hasta alcanzar su valor mínimo de tal forma que, una vez encontrada la función mínima, se posee 2 clústeres nuevos de tal forma que ahora los píxeles en efecto se guardan en el clúster cuya “distancia de color” es la mínima, lo cual quiere decir que se agrupa en 2 clases a las que cada grupo de pixel posee una relación, en este caso por intensidad y contraste.

Se puede observar que la herramienta K-means aparte de ser un proceso iterativo cuyo funcionamiento es mucho más complejo de entender que el de Otsu, constituye una herramienta cuyo objetivo no se limita a la separación de 2 clases, como si lo hace Otsu. Pues al contrario que este último, K-means permite la separación de una imagen en más de 2 clases, característica que en este caso no se requiere, pues solo se desea separar el ROI (ACM) del fondo. Por todo lo antes detallado, se considera adecuado la elección de Otsu como método de segmentación secundario.

Para extracción de características destacan 2 grupos principales los cuales son extracción de características simples y extracción de características complejas tal y como se muestra en la tabla XV.

TABLA XV
TIPOS DE PREPROCESAMIENTO PARA EXTRACCIÓN DE CARACTERÍSTICAS

Métodos de Preprocesamiento Extracción de Características	
Características Simples	Características Complejas
FOS	DWT
GLCM	CNN

Esta distinción se debe a que tanto FOS como GLCM extraen características simples de la imagen, datos como: Media, varianza, asimetría (FOS) y Contraste, Energía, Homogeneidad (GLCM), mientras que, por otro lado, las herramientas DWT y CNN analizan y extraen características de capas más profundas, en el caso de las DWT analiza la imagen a diversas frecuencias, haciendo una separación de patrones en base a esto y a la dirección de los píxeles (horizontal, vertical y diagonal), mientras que las CNN calculan y asignan pesos a características inobservables de las imágenes que han pasado por múltiples procesos de convolución a fin de reducir la dimensionalidad de sus componentes.

Para poder determinar que método es el adecuado, se estudió el alcance y desempeño que tiene cada uno de estos. Para la extracción de características simples se tomó GLCM debido a que analiza parámetros similares a los FOS, pero tomando en cuenta la influencia de píxeles vecinos, lo que puede resultar beneficioso para analizar cambios de temperatura en el motor. Sin embargo, para la extracción de características complejas se tuvo en cuenta el rendimiento y el tiempo de computo de cada método, resultando la DWT el método elegido, debido que superó por mucho el tiempo de ejecución a las convoluciones.

Con respecto a aumentadores/diferenciadores de contraste, CLAHE ha destacado como una gran herramienta a diferencia de Data Augmentation. También se usó HSI como método directo para acrecentar el contraste, ello es de manera previa a CLAHE, para que de esta manera no se pierda la dinámica original de la imagen.

Finalmente, con respecto a los métodos de preprocesamiento de filtrado es menester comentar que, el filtro FFT a pesar de que lleva a cabo una labor de filtrado, lo hace con el objetivo de acrecentar contraste en la imagen y por lo explicado antes, CLAHE ya realiza esta acción de una manera mucho más efectiva, por lo que sería redundante mantener estos 2 preprocesamientos y, por ende, solo se mantuvo CLAHE para futuros modelos.

B. Análisis clasificadores

Los métodos de clasificación se agruparon en tres categorías: CNN, algoritmos de aprendizaje automático y Vision Transformers (ViT).

- **CNN:** Se evaluaron seis arquitecturas. En algunos modelos, las CNN se usaron para entrenamiento y predicción (como en los modelos 1 y 2), mientras que en otros (como los modelos 4, 6 y 11) solo como algoritmo de extracción de características. Aunque el rendimiento fue positivo, se observó una ligera mejora en precisión al reemplazar la última capa de las CNN por un clasificador SVM.
- **Algoritmos de Aprendizaje Automático:** Se usaron SVM, KNN y ERT. Los mejores resultados en esta categoría provinieron del modelo 3 basado en SVM, destacando por su rapidez y precisión. KNN mostró resultados menos precisos y ERT, usado con histogramas generados por BoVW, ofreció buenos resultados, aunque con un modelo más pesado que otras opciones de precisión similar.

- **Vision Transformers (ViT):** Se evaluaron cuatro modelos de ViT. Se obtuvieron resultados de predicción casi perfectos y tiempos de entrenamiento menores que los de las CNN.

C. Análisis de datasets

En cuanto a los datasets es relevante analizar las 2 principales fuentes de datos de las cuales se nutre este trabajo, en primer lugar, tenemos a la data obtenida en [18], donde a pesar de contar con una gran cantidad de imágenes y un número considerable de clases, se observa que el tipo de máquina objeto de estudio es la misma, además de que el ángulo de captura de la imagen se mantiene para todo el dataset, lo que significa una pérdida de información del ACM.

Esto explicaría el motivo de porque la mayoría de los modelos entrenados, incluyendo los detallados en el presente trabajo dan tan buenos resultados, pues los patrones de falla al pertenecer a una misma máquina son ampliamente perceptibles para un modelo clasificador como los evaluados.

Esto sin embargo no es extrapolable a industria, pues en dicho contexto, no existe un solo tipo de motor por analizar y menos aún en una misma sala o con un mismo fondo y/o carga de funcionamiento, por lo cual los resultados observados si bien son altamente prometedores, deberían ser tomados como una oportunidad de extrapolar este tipo de investigaciones y promover la obtención de data independiente.

Concatenando lo anteriormente mencionado respecto a la problemática de la data, es que surgió la idea de obtener un nuevo dataset en base a motores de diversa índole, uso, carga, contexto de funcionamiento, etc. Es por ello que, en coordinación con múltiples entes institucionales de la Universidad de Piura, se coordinó la visita a múltiples lugares y áreas de funcionamiento donde se podrían encontrar este tipo de máquinas. Finalmente se puede resumir la estructura de esta data recopilada en la tabla XVI.

TABLA XVI
ESTRUCTURA DE LA DATA 4 (PARTE 1)

Ubicación	Nº imágenes	Modelo
Laboratorio de Sistemas Automáticos de Control	144	El motor llegó a una elevada temperatura de manera rápida
Estación de bombeo de agua para la Facultad de Comunicación	25	Se observa una concentración de temperatura en el eje del motor
Estación de bombeo de agua para el Aulario UDEP	63	Motores colocados de manera vertical
Estación de bombeo de residuos del Edificio 80	80	Mayor calentamiento debido a la mayor demanda de potencia de la carga

TABLA XVII
ESTRUCTURA DE LA DATA 4 (PARTE 2)

Ubicación	Nº imágenes	Modelo
Laboratorio de Electrotecnia	167	El motor en vacío presentaba fugas de ventilación y el motor del módulo de tanques fue variando su potencia

D. Mejor modelo

El modelo final recolectó una combinación de los mejores algoritmos de preprocesamiento, los cuales se detallan a continuación:

- **Detección del motor con YOLOv5m:** Se cambió de YOLOv5s a YOLOv5m para mejorar la detección del motor en imágenes de bajo contraste. Tras entrenarlo, el modelo detecta el área de interés, recorta las imágenes y elimina elementos irrelevantes.

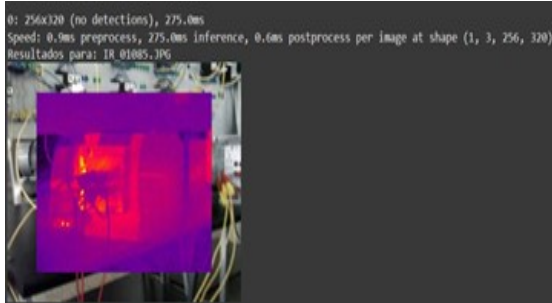


Fig. 2. Detección fallida con YOLOv5s con la nueva data

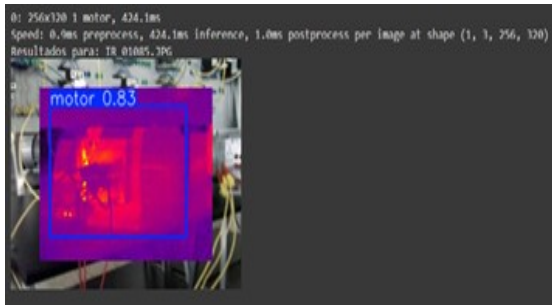


Fig. 3. Detección correcta con YOLOv5m con la nueva data

- **Segmentación con Otsu** Segmenta el motor recortado, identificando regiones relevantes mediante un umbral óptimo en escala de grises.
- **Conversión a los canales de color HSI:** En este paso se cambian los canales de las imágenes de RGB a HSI, de los cuales solo se emplea el canal I (intensidad), al ser el que mejor representa la temperatura de una imagen termográfica.
- **Redimensionamiento y mejora de contraste con CLAHE:** Las imágenes se redimensionan a un tamaño

de 224 x 224 píxeles, ya que este es un tamaño bastante usado en algoritmos de procesamiento de imágenes, posteriormente se aplica CLAHE, con el objetivo de mejorar el contraste y hacer más visibles los pequeños detalles.

- **Extracción de características con GLCM:** Se extraen las características de contraste y homogeneidad de las imágenes resultantes del paso anterior, estas son almacenadas en un vector de longitud 2 para su posterior uso.
- **Transformada discreta de Wavelet (DWT):** A través de esta herramienta se busca extraer 4 subbandas de frecuencia: cA (baja frecuencia) y cH, cV y cD (estas 3 últimas de alta frecuencia), que se aplanan y combinan en un vector de características. Cada una de las subbandas corresponde a una imagen de 112 x 112 píxeles, al ser aplanadas y concatenadas, se obtiene un vector de longitud 50176.
- **Preparación de datos:** Las características combinadas dan un total de 50178, cada una de las 9071 imágenes que se tenían inicialmente posee un vector de longitud 50178.
- **Reducción de dimensionalidad con PCA:** Con el objetivo de hacer mucho más ligero el modelo, se optó por usar PCA para la reducción de componentes, estudios previos avalaron la reducción de 50178 a 50 componentes, sin sacrificar eficacia del modelo.
- **Optimización de hiperparámetros:** Se usó "Grid Search" para seleccionar los mejores valores del parámetro de regularización, el factor gamma y el kernel para SVM, los parámetros identificados como óptimos fueron un valor de regularización de 1, un valor de gamma de 0.01 y un el uso del kernel "linear".
- **Entrenamiento y evaluación:** Con los parámetros óptimos y una distribución de datos del 80% para el entrenamiento y el 20% restante para la validación, se entrenó el modelo SVM, alcanzando un 99.72% de precisión, confirmando el buen desempeño de los preprocesamientos antes seleccionados.
- **Matriz de confusión y resultados:** La matriz muestra excelente precisión en la clasificación de "healthy", "misalignment" y "broken rotor bars".

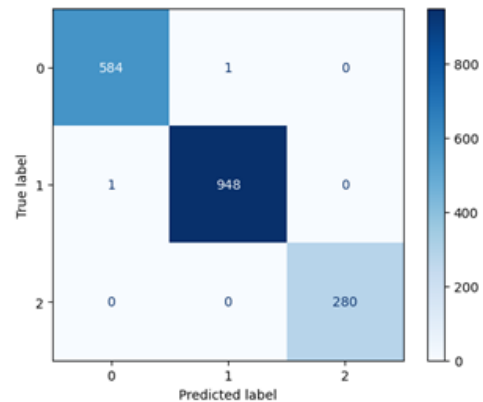


Fig. 4. Matriz de confusión

- **Subclases de fallas:** De acuerdo a la data estudiada, se pudieron apreciar dentro de las fallas de desalineamiento y barras de rotor rotas, diferentes tipos de grados para cada una, es por eso que, con el objetivo de realizar una identificación mucho más completa y detallada, se estableció un sistema de identificación jerárquica, donde el primer modelo SVM (el antes mencionado) reconoce si existe o no falla en el motor (desalineamiento o barras rotas), en caso de existir, existen dos modelos adicionales de SVM (uno para cada falla) que clasifican la falla correspondiente en cada una de sus subclases, los modelos fueron entrenados con parámetros similares al modelo general y solo con data exclusiva de cada falla, obteniendo los siguientes resultados:

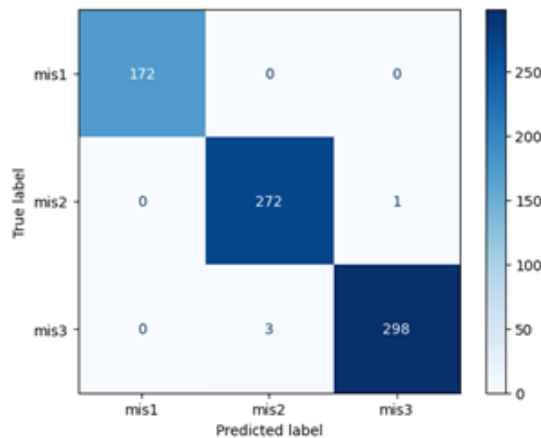


Fig. 5. Matriz de confusión de desalineamiento

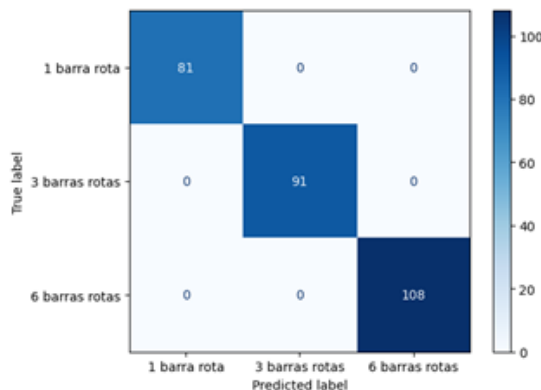


Fig. 6. Matriz de confusión de barras rotas

- **Prueba de validación:** Como se ha venido detallando en capítulos previos, se sospecha que los modelos pueden estar sufriendo de Overfitting, motivo por el cual se ha decidido realizar una prueba donde un dataset de validación compuesto únicamente por motores en un buen estado (Healthy) pero de diverso tamaño, posicionamiento, potencia, ubicación, carga, etc. Es decir, Data4, sirva para probar la robustez de este modelo, ello es antes de usarlas

como data de entrenamiento, dicha prueba se llevó a cabo teniendo únicamente el YOLOV5m entrenado con dichas imágenes, a fin de que, si pueda reconocer motores y en base a los motores encontrados, y tras aplicar los métodos antes detallados se obtuvo el resultado mostrado en la figura 7.

```
[32] 1 from sklearn.metrics import accuracy_score
      2 accuracy = accuracy_score(y, predicciones)
      3 print("Precisión:", accuracy*100)
```

Precisión: 100.0

Fig. 7. Prueba de validación con Data4 al modelo 18

E. Interfaz

Como producto final, se desarrolló una interfaz de usuario (GUI), con el objetivo de hacer más sencilla y amigable para el usuario la detección de fallas en motores en aplicaciones reales, se utilizó la librería PyQt5 para la creación de la GUI, la cual ofrece una amplia serie de widgets personalizables y una aplicación especializada para el diseño, la cual es QtDesigner como se observa en la figura 8.



Fig. 8. Diseño de la interfaz dentro de QtDesigner

Con el botón presente en la siguiente imagen, se abrirá un cuadro de diálogo donde se podrá elegir la imagen desde los archivos locales del dispositivo.



Fig. 9. Botón para seleccionar la imagen

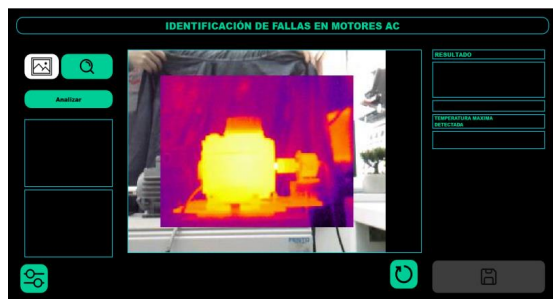


Fig. 10. Interfaz una vez seleccionada la imagen

Con el siguiente botón, el modelo YOLO identificará el motor en la imagen ingresada, en caso no lo identifique de manera correcta, se dará la opción de seleccionarlo por cuenta propia.



Fig. 11. Botón para seleccionar la imagen de manera manual

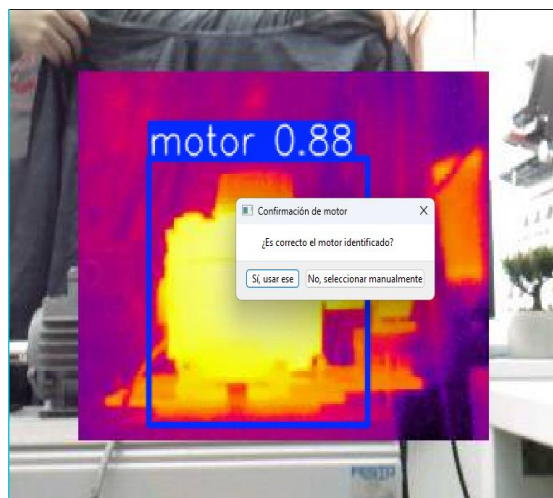


Fig. 12. Motor identificado de manera correcta

Ya que se tiene identificado el motor, ya sea de manera manual o automáticamente, se procede al análisis por parte del modelo SVM, mediante el uso del botón “analizar”. Los resultados son mostrados en la parte derecha de la aplicación como se observa en la figura 13.

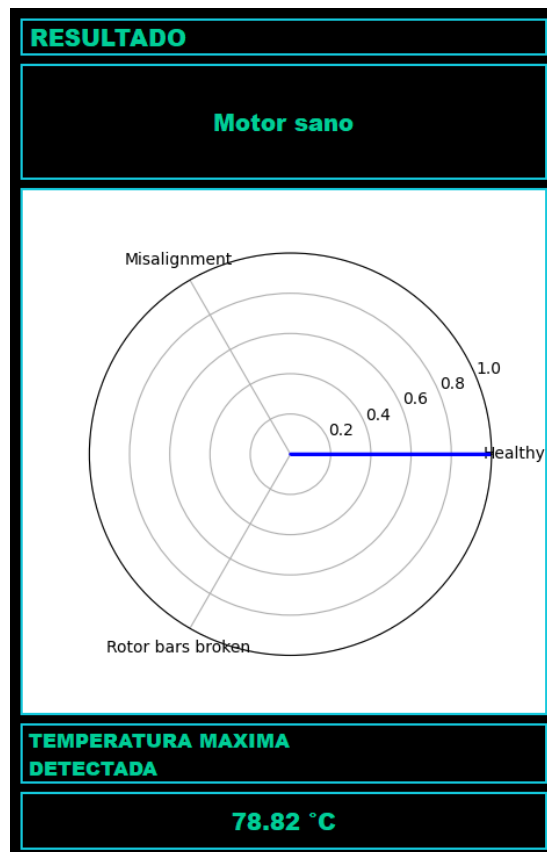


Fig. 13. Resultados obtenidos

Cabe resaltar que se pueden establecer los límites inferior y superior de temperatura, a través del botón configuraciones, ubicado en la parte inferior izquierda, tal y como se demuestra en la figura 14.

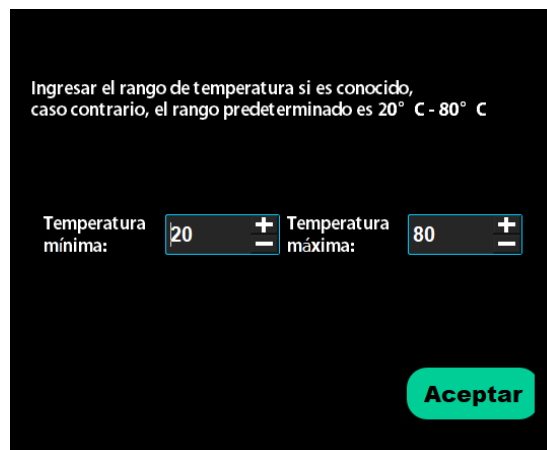


Fig. 14. Botón configuraciones para establecer los límites inferior y superior de temperatura

Finalmente , la aplicación nos ofrece la posibilidad de generar un reporte en pdf de la imagen analizada, el cual contiene los resultados obtenidos los cuales son: la clase

predicha por el SVM, el gráfico de radar con las probabilidades de cada clase y la temperatura máxima estimada en la imagen, esto es apreciable en la figura 15.

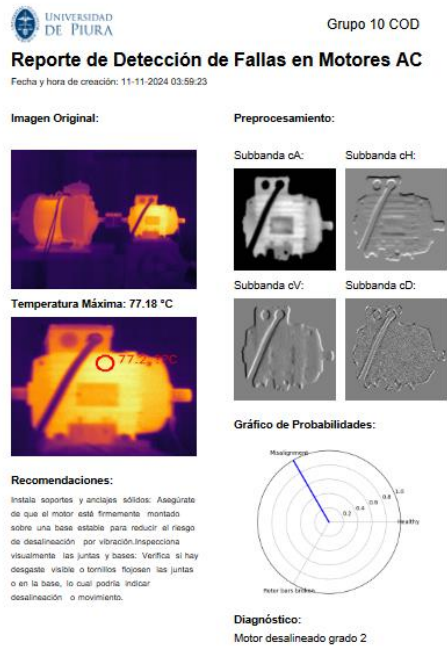


Fig. 15. Reporte final de la interfaz

VII. CONCLUSIONES

- Durante las pruebas, la transformada discreta de wavelet demostró ser el preprocesamiento que más información brindaba, llegando a alcanzar una precisión de 99.18% en su evaluación con SVM, sin embargo, debido a la naturaleza del Dataset, siempre está el riesgo de sobreentrenamiento del modelo, esto se evidencia al analizar el margen de separación del modelo, es por eso que se aplicaron técnicas como la crossvalidation, y se variaron parámetros como la cantidad de componentes analizadas o el parámetro de regularización. Todos estos experimentos obtuvieron porcentajes de precisión igualmente bastante altos, además de márgenes de separación mucho más altos, corroborando la validez del modelo y haciéndolo mucho más robusto.
- Dentro del análisis realizado con PCA (Principal Component Analysis) en el modelo 3, se detectó que dentro de las 50176 características obtenidas de las 4 sub bandas mediante DWT, el 95% de la varianza era representada por solamente los primeros 593 componentes de cada imagen, esto significa que se puede usar solo este número de componentes sin sacrificar información relevante de las subbandas.

- Los modelos que han usado CNN de gran profundidad como lo han sido EfficientNetB2, B5 y B7 no han demostrado ser buenos prospectos, pues comparativamente con otros modelos, a pesar de conseguir una muy buena precisión, el tiempo de entrenamiento representa por mucho un problema mayor en cuanto a su implementación industrial.
- El método de clasificación SVM-1 y SVM-2 el cual ha empleado un SVM como herramienta primordial de clasificación, ha demostrado un mejor desempeño en esta tarea que la herramienta de KNN.
- El modelo que emplea FOS como extractor de características, dada su precisión y velocidad de entrenamiento, un modelo plausible a implementar en caso lo que se busque sea un sistema de respuesta rápido, pues la carga computacional que exige es la menor de todos los modelos evaluados, obviamente tomando en cuenta también que, a su vez, han sido los modelos menos precisos entre los mismos (Modelo 11 y 12).
- En caso se desee aplicar un sistema de control predictivo o simplemente un protocolo de mantenimiento predictivo usando análisis termográfico en conjunto con de redes neuronales, dada la particularidad de estas últimas, lo más adecuado sería suministrar al menos un conjunto de imágenes termográficas del motor sano en el contexto en el que trabajara, para que el método clasificador pueda al menos aprender los patrones propios de un motor sano, de esta forma nos aseguraremos siempre de que al menos la tarea de identificación de fallas en un ACM la cumplirá de manera eficiente.

REFERENCIAS

- [1] Aalto University. Independent component analysis. Accedido: 11 de noviembre de 2024.
- [2] Enrique Alegre, Gonzalo Pajares, and Arturo de la Escalera. Conceptos y métodos en visión por computador. *Conceptos y métodos en visión por computador*, 2016.
- [3] Amazon Web Services. ¿qué es una red neuronal?, 2024. Accedido: 11 de noviembre de 2024.
- [4] Dora María Ballesteros Larrotta. Aplicación de la transformada wavelet discreta en el filtrado de señales bioeléctricas. *Umbral Científico*, 2004.
- [5] Y.A. Cengel and J.H.P. Castellanos. *Transferencia de calor y masa: un enfoque práctico*. Elibro Catedra. McGraw-Hill, 2007.
- [6] Chun-Fu Richard Chen, Quanfu Fan, and Rameswar Panda. Crossvit: Cross-attention multi-scale vision transformer for image classification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 357–366, 2021.
- [7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Pre-training of deep bidirectional transformers for language understanding. *arxiv. arXiv preprint arXiv:1810.04805*, 2018.
- [8] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [9] Fluke Corporation and The Snell Group. Introducción a los principios de la termografía, 2009. Accedido: 11 de noviembre de 2024.
- [10] Mohammad Haider, Amit Doegar, and Ram Kumar Verma. Fault identification in electrical equipment using thermal image processing. In *2018 International conference on computing, power and communication technologies (GUCON)*, pages 853–858. IEEE, 2018.
- [11] Ali Hassani, Steven Walton, Nikhil Shah, Abulikemu Abuduweili, Jiachen Li, and Humphrey Shi. Escaping the big data paradigm with compact transformers. *arXiv preprint arXiv:2104.05704*, 2021.
- [12] IBM. ¿qué es la segmentación de imágenes? Accedido: 11 de noviembre de 2024.

- [13] S Padma Kala, R Palanivel, Murlidhar Prasad Singh, G Arulkumaran, and Ammar Hameed Shnain. Infrared thermography-based induction motor bearing fault detection using hybrid method. In *2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, pages 1–4. IEEE, 2024.
- [14] Gurbhej Singh Kalyan and Poonam Syal. Recent advancements of thermal imaging in induction motor: A review. In *2023 5th International Conference on Energy, Power and Environment: Towards Flexible Green Energy Technologies (ICEPE)*, pages 1–7. IEEE, 2023.
- [15] Seung Hoon Lee, Seunghyun Lee, and Byung Cheol Song. Vision transformer for small-size datasets. *arXiv preprint arXiv:2112.13492*, 2021.
- [16] Amine Mahami, Chemseddine Rahmoune, Toufik Bettahar, and Djamel Benazzouz. Induction motor condition monitoring using infrared thermography imaging and ensemble learning techniques. *Advances in Mechanical Engineering*, 13(11):16878140211060956, 2021.
- [17] MathWorks. Convolutional neural network. Accedido: 11 de noviembre de 2024.
- [18] Mohamad Najafi, Yasser Baleghi, and Seyyed Mehdi Mirimani. Thermal image of equipment (induction motor), 2020.
- [19] Saira Parveen, Tanweer Hussain, Dileep Kumar, and Bhawani Chowdhry. Induction motor fault detection and classification using thermal images and deep learning. 6:16–31, 11 2021.
- [20] Ashraf Navaid Sabah and Zainul Abidin Jaffery. Fault detection of induction motor using thermal imaging. In *2022 IEEE IAS Global Conference on Emerging Technologies (GlobConET)*, pages 84–90. IEEE, 2022.
- [21] Gönül Sakalli and Hasan Koyuncu. Discrimination of electrical motor faults in thermal images by using first-order statistics and classifiers. In *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pages 1–5. IEEE, 2022.
- [22] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Accedido: 11 de noviembre de 2024.
- [23] Farhan Md Siraj, Syed Tasnimul Karim Ayon, Md Abdus Samad, Jia Uddin, and Kwonhue Choi. Few-shot lightweight squeezenet architecture for induction motor fault diagnosis using limited thermal image dataset. *IEEE Access*, 2024.
- [24] Omar Trejo-Chavez, Irving A Cruz-Albarran, Emmanuel Resendiz-Ochoa, Alejandro Salinas-Aguilar, Luis A Morales-Hernandez, Jesus A Basurto-Hurtado, and Carlos A Perez-Ramirez. A cnn-based methodology for identifying mechanical faults in induction motors using thermography. *Machines*, 11(7):752, 2023.
- [25] Universidad de Sevilla. Programa de la asignatura pid-filtros digitales. Accedido: 11 de noviembre de 2024.
- [26] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.
- [27] Lifu Xu, Soo Siang Teoh, and Haidi Ibrahim. A deep learning approach for electric motor fault diagnosis based on modified inceptionv3. *Scientific Reports*, 14(1):12344, 2024.