Analysis of Internal Mechanical Failures in AC Motors Using Thermographic Imaging

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. INTRODUCTION

Alternating current motors (ACM) are essential elements in a wide variety of industrial applications, however, they are not exempt from suffering failures that can affect their performance and, in more serious cases, can cause interruptions in the operations in which they are involved, leading to economic losses. Internal mechanical failures in ACMs can manifest themselves in various ways such as overheating, vibrations and abnormal noises and/or power losses, if these failures are not detected and corrected in time they can lead to economic losses as mentioned above, which is why the ability to identify these problems early is of utmost importance not only to ensure the continuity of operation, but also to ensure maximum operational efficiency of the process being carried out. One of the emerging techniques for the detection of mechanical failures in ACMs is the use of thermographic images, these images show the distribution of temperatures on the surface of the engine and by analysing it it is possible to identify anomalous heat patterns that may be indicative of internal failures. There are many types of internal faults in an ACM, one of the most commonly used classifications is according to their origin, such as thermal, electromagnetic, mechanical, etc. I feel that the most common faults are of the mechanical type, such as unbalance between parts, misalignment of shafts, bearing failures, rotor bar failures, etc. The main objective of the present work is to explore the application of infrared thermography for the detection of faults in alternating current motors. The aim of the work is to identify the most common faults, analyse the images obtained and evaluate the effectiveness of this technique as a diagnostic tool in the industrial environment.

II. THEORETICAL FRAMEWORK

Induction motors are a type of ACM widely used in industry, which is why they often operate under unfavourable conditions, which in many cases causes the appearance of internal faults. To avoid these problems, the implementation of maintenance programmes is essential, and within these, the early detection of faults in ACMs is one of the key factors with a view to optimising resources and reducing costs. In this context, the most common methods employed have been vibration analysis, motor current signature and visual and/or auditory inspections. Although the aforementioned methods have been very effective, they also have some limitations, such as the fact that they can be invasive and that they can also be susceptible to external interferences, which is why the use of thermographic images has emerged as a non-invasive alternative, through which it is possible to analyse surface temperature distributions by using neural networks among other techniques, and the state of the engine can be inferred. This joint process (thermographic analysis + classifier) will allow us to detect faults based on anomalous thermal patterns developed in the presence of a mechanical and electrical problem, and even distinguish between them; in order to obtain and carry out this process correctly, image processing techniques will be necessary before and after entering them into the classifier model, which will be explained in the following points of this section. Therefore, for the proper understanding of this document, it is necessary to first define certain terms and words that are considered fundamental in terms of the subject of engines and everything related to them.

A. Redes neuronales

A neural network is an artificial intelligence technique, which aims to teach computers to process data in a similar way to the human brain, using interconnected nodes or neurons in a layered structure. [3]

B. AC motors

Electromechanical devices that convert electrical energy into mechanical energy, they are widely used in industries and in a variety of commercial applications such as pumps, fans, compressors and conveyor systems. Within this classification are squirrel cage and wound rotor induction motors, as well as synchronous motors.

C. Faults

Defects, malfunctions or abnormal conditions that affect the optimal functioning of an asset, in the context of AC motors, these can be mechanical or electrical in nature.

1) Mechanical Failures (MF): They are caused by physical problems in the mechanical components of the engine, these include:

- **Wear on bearings (BD):** Mainly caused by friction and material fatigue, it causes excessive vibrations which at a certain point significantly affect the efficiency of the engine.
- **Misalignment (MSL):** Occurs if the moving parts of the engine are not properly aligned, can also cause excessive vibrations, accelerated wear of components, etc.

2) Electrical Faults (EF): Occur due to anomalies in the electrical circuits or components. electromagnetic components of the motor, these include:

- **Short circuit in the stator coils (ECF):** Caused by excessive wear of the insulation between coils, allowing contact between two conductors, it can lead to unwanted current flow, capable of damaging the motor.
- **Rotor bar breakage (BRB):** This type of failure is common in squirrel cage motors and refers to the breakage of one or more rotor busbars, usually caused by fatigue or abnormal vibrations, resulting in a loss of torque that can be significant.

D. Radiation

Radiation is energy emitted by matter in the form of electromagnetic waves (or photons) as a result of changes in the electronic configurations of atoms or molecules; it does not require the presence of an intervening medium. [5].

E. Infrared thermography (IT)

The science of using optoelectronic devices to detect and measure radiation to obtain the temperature of the surface under analysis. The devices that help us to obtain these images are thermographic cameras. [8]

F. Image pre-processing (IPP)

A set of techniques that are applied to images before any complex analysis or processing is performed. Their aim is to improve quality, remove noise and/or highlight features.

1) Filters: The application of filters is one of the main techniques in the They can be applied in the space domain or in the frequency domain, and are aimed at smoothing the image, removing noise, detecting edges, etc. [22]

2) Contrast Limited Adaptive Histogram Equalization (CLAHE): A method that constructs a histogram of intensity per pixel, and then increases the length of each value, thereby increasing the contrast throughout the image.

3) Transformada Wavelet discreta (DWT): A mathematical tool used for both image and audio compression, it is a decomposition of signals into frequencies. It allows to analyse images at different scales and frequencies. [4].

4) Independent Component Analysis (ICA): A computational method used to separate a multivariate signal into independent constituent subcomponents, used in a wide range of applications, such as signal processing, pattern recognition, image processing, etc. [1]

5) Image segmentation (IS): A computer vision technique that separates a digital image into distinct groups of pixels (segments) to facilitate object detection, among other tasks [10]. There are numerous segmentation techniques:

- **Thresholding:** They classify pixels according to whether their intensity is above or below a certain threshold value. One of the most commonly used methods to determine the threshold value is the Otsu method.
- **Histograms:** They represent the frequency of pixel values in the image, they are used to determine thresholds.
- **Edge detection:** They identify boundaries of objects or classes through the existence of discontinuities or contrasts.
- **Watershed algorithms:** They use the images converted to greyscale and generate a topographic map based on pixel luminosity.
- **Cluster-based segmentation:** The visual data is divided into groups of pixels with similar values, one of the most common variants is the K-means technique.

6) Bag of Visual Words (BoVW): A technique used in computer vision and machine learning work, it represents an image as a set of visual words that extract notable features in the image (edges, corners, textures, etc.). [19]

7) Speeded-UP- Robust Features (SURF): It is an image feature detection algorithm, which aims to identify significant features that can be used in comparison, matching, etc. tasks. Some of its main advantages are its speed and robustness.

8) Hue Saturation Intensity (HSI): Descriptive colour model similar to the RGB model, where colour values are described, but this time based on the characteristics of intensity, saturation and hue.

9) Linear Discriminant Analysis (LDA): Technique for filtering features so that the input variables are reduced in a dimension that much better represents the dynamics of a process or image.

10) Principal Component Análisis (PCA): It is a method that by analysing a matrix of defined size that stores several values corresponding to the features extracted from one or more images, calculates the covariance matrix, from which, by calculating eigenvalues and eigenvectors, it can determine the most representative groups of features of the original matrix.

G. Feature-based classifiers (IFE)

It is the application of techniques with the objective of reducing the complexity and dimensionality of an image by extracting relevant features. [2]

1) Convolutional Neural Networks (CNN): They are network architectures for Deep Learning (DL) effective for the detection of patterns in images, in order to recognise objects, classes and categories, they have convolution layers in their structure, these layers can apply convolutional filters to the images, to activate different characteristics of these [14], among these architectures we find:

- **VGG16:** Developed by Visual Geometry Group (VGG), presented in 2014, it is based on 13 convolutional layers and 3 fully connected layers.
- **ResNet (Residual Networks):** Developed by ImageNet in 2015, its main feature is the use of residual blocks, which allow the assembly of networks with many layers, with ResNet-50, ResNet-101 or ResNet-152, with 50, 101 and 152 layers respectively.
- **Inception (Google Net):** Developed by Google, it introduces the use of inception blocks, allowing filters of multiple sizes to be used at the same time.
- **MobileNet:** Lightweight architecture, designed for resource-constrained or mobile equipment, uses separable convolutions to reduce the number of parameters.

2) K-Nearest Neighbors (K-NN): It is a non-parametric supervised learning classifier, which makes classifications or predictions about the clustering of a data set, based on proximity.

3) Support Vector Machine (SVM): A supervised learning algorithm used in classification and regression problems, its objective is to obtain a hyperplane that separates two classes of data points.

4) Random Tree (RT): A machine learning model used for prediction and classification, characterised by the use of a random approach in its construction, randomly selecting a subset of data and features at each stage of the process. Its main advantage is that it helps prevent overfitting of the model to specific data, thus improving its ability to make more general and accurate predictions.

5) Visual Transform (ViT): Model that adapts the architecture of Transforms to be able to receive images, by adding a previous step where the image is segmented in a finite group of parts which are known as batches and is adapted in such a way that it can be entered into the Transformer Encoder. For this particular method and its correct understanding it is important to keep in mind the following clear concepts:

- **Multilayer Perceptron (MLP):** Technology based on the functioning of a neuron forms the basis of all models based on Deep Learning.
- **Feed Foward:** Learning methodology that allows MLPs to learn for each "epoch" trained.
- **Patch Division:** Division of an image into smaller patches or images, which contain a smaller amount of information.
- **Positional Encoding:** Array that stores the positioning of each Patch in a single array.
- **Encoder Stack:** Tool for flattening any input into a representative vector.
- **Decodar Stack:** A tool that converts a representative vector into a probabilistic output or vector that serves as a prediction of some kind.
- **Auto Attention:** Function in charge of relating inputs to outputs in such a way that for a prediction $M+1$ discerns which N input is more important to make the prediction.
- **Multi Heading Self Attention:** Multiple Auto Attention functions connected in parallel.

III. State of the Art

In recent years, engine failure detection has been further developed through the use of thermographic imaging and advanced image processing techniques and machine learning algorithms. In this section, the most relevant and recent literature related to the detection of faults in ACMs using thermographic images will be presented, analysing the methodology proposed by them, highlighting which processing techniques and classification algorithms have been the most effective.

In [11], they describe how artificial intelligence and deep learning are being employed to improve conventional induction motor fault diagnosis techniques as induction motors become more common. They highlight the great potential for exploration in future studies and the difficulty of detecting faults in real time at present due to laborious signal decomposition methods.

With respect to thermal image processing, the HSV colour model considers the hue region (HUE) of the HSV model due to the usefulness for evaluation and measurement of the colour of an object such as predominant redness or yellowness in the thermal image, in addition to edge detection techniques such as Roberts, Prewitt and Otsu in order to segment thermal images of faulty electrical equipment. The thresholding performance comparison of these methods was performed by calculating the mean square error and the peak signal to noise ratio with the Otsu method being the most effective in determining the hottest regions in these types of images as seen in [9].

While it is true that working with the colour of thermal images highlighting the hot spots is a good preprocessing technique, it does not take into account the visual noise of the image, so the two-dimensional discrete Wavelet transform (2D-DWT) improves the signal-to-noise ratio and then reduces the dimensionality of features to be extracted by identifying the most relevant ones. Then, it calculates the MD (Mahalanobis Distance) to classify the features according to their relevance by discarding the less relevant ones. All this data is used to train a CDT (Complex Decission Tree) to predict the state of the bearings based on a cross-validation scheme, obtaining a classification accuracy of 99.85%.

In contrast, in [24] design a hybrid method for the detection of faults in induction motor bearings whose preprocessing is based on the use of CLAHE achieving a better visualisation of the details in the thermal images to highlight heat gradients and temperature anomalies to be used in Inception V3 with an SE module integrated in its architecture to enhance feature extraction and the last classification layer of Inception V3 is replaced by an algorithm (SVM) improving the representation and generalisation capability of the model being able to classify up to 11 different types of engine faults with a high accuracy accelerating the learning process and stabilising the loss during training.

In [13] a new methodology for fault classification in three-phase induction motors is presented. As an image preprocessing technique, they propose the use of BoVW (Bag of Visual Words) and SURF (Speeded-UP- Robust Features) for the extraction of visual features. The classification model used is ERT (Extremely Randomized Tree) resulting in a high accuracy and stability in the detection and classification of faults, surpassing traditional methods such as KNN, SVM and DT.

After image preprocessing is finished, a classifier between motor faults and the optimal machine operation state should be considered, among them, deep learning based methods using thermal imaging can effectively classify the conditions of induction motors with high accuracy and low error rate. VGG-16 and ResNet-16, two convolutional neural network architectures emerge as the most reliable alternatives in this field with the former being superior to the other due to its high deep learning capability and effective handling of the gradient vanishing problem.

Similarly, the use of a convolutional neural network can demand a high computational cost, so [20] introduce a learning approach using modified SqueezNet, a 7-layer CNN model with 16 times fewer parameters than the original, significantly reducing training time and computational resource usage making it ideal for mobile devices or embedded systems. CNN model with only 7 layers and 16 times fewer parameters than the original one, significantly reducing training time and computational resource usage, making it ideal for mobile devices or embedded systems. Similarly, a prototypical network was designed by applying Few-Shot Learning so that the model learns to classify faults efficiently with few examples. This study concluded that the implementation of lighter models does not necessarily imply significant losses in the efficiency of electric motor fault detection using thermal imaging.

These results have suggested that DL-based methods can efficiently predict motor conditions even when trained on small and unbalanced data. [16]

However, some architectures with pre-trained weights may present problems. Then, the Dataset can be prepared in order to force the fault in an extreme way so that it is clearly observed in the infrared camera and an accuracy of 95% can be achieved by using data augmentation techniques to simulate real scenarios, the presence of noise and the accuracy of the model. [21]

Apart from the methodology mentioned above, there are also classification algorithms that use the features classification algorithms that use the features collected in the pre-processing stage. For example, the extraction of FOS (First Order Stadistics) as data extracted from the images. In [18], they make a comparison between classifiers such as neural networks, KNN (K-Nearest Neighbors), LR (Logistics Regressions) and SVM (Support Vector Machine). They conclude that for the use of FOS, RF (Random Forest) stands out as the best classifier due to its balance between accuracy and stability.

On the other hand, in [17] they base their model on three levels of processing including most of the points highlighted by the previous authors. At the low level they normalise the data by cropping, rotating and enlarging the images and using histogram equalisation to improve the contrast of the images. At the medium level, a semantic segmentation network is employed with SegNet pre-trained with VGG-16 weights. For this, the images are labelled and classified into 3 classes: background, winding and fault. Subsequently, the network is trained and validated until an accuracy of 99.29% is achieved by highlighting regions of interest in the hot areas of the windings. Finally, in high-level processing, three temperature variables are set as inputs to a fuzzy interference system classifying the motor into three levels: regular, alert and alarm.

In the search for a solution to the problem of the loss of information in the RNN during training, the Attention function emerged as a way of relating inputs and outputs, thus minimising the loss of information and increasing the efficiency of the model. This new method called Transformer allows the analysed elements, called tokens, to "choose" which other tokens to pay attention to, thus speeding up the process and not losing the relationship between patches and even better combined with the use of the Attention functions in parallel (Multi-Head Attention). [23]

This is why in ¨citedosovitskiy2020image we explore the possibility of using the architecture of a Transformer in the field of computer vision for classification tasks by using a division into patches of the original image with the aim of reducing the computational weight of the model and speeding up the operations; It also uses a Position Embedding that serves as an index for each patch, with these 2 values and after being subjected to a flattening, a Multi-Head Self Attention (MSA) can be applied that allows predicting with relation criteria, that is to say, it learns to know which previous values are the ones that should influence more in the specific future prediction. This method is widely used for very large Datasets given its great efficiency as well as robustness, and it is this paper in conjunction with [7] that demonstrated that its use is extrapolable to uses that were thought impossible due to its computational cost.

However, Transformers stand out in a great way when trained on large Datasets, this is with respect to other models such as CNNs, while on small Datasets there is not a big difference in performance, which is why [12] uses 2 things: a special type of data partitioning in Patches called Shifted Patch Tokenization (SPT) where by a set of set steps it enlarges the number of patches, a setting of the Self-Attention function to make it focus only on relevant tokens (not nulls), eliminating the possibility of a token to build relationships with those null tokens.

Although ViTs, according to the literature, perform better than CNNs, for example, in the case of large datasets, the same is not true for small datasets, which is why the concept of Compact Convolutional Transformer (CCP) was presented in ¨citehassani2021escaping, which is based on not using a simple division of the image for the separation into patches, but applying on these patches a convolution that helps to capture local characteristics similar to those of a CNN, as well as preserving the spatial information of the image. Other valid approaches to deal with this situation is the one proposed in [6] where they focus on a double tokenisation at different scales, i.e. one image, tokenise it or split it into large patches and then do the same process but for smaller patches, i.e. get 2 tokens per image, In this way it captures not only the global information of the image, but also the local information of the small patches, both tokens are input to a proprietary encoder which is known as cross-attention, and according to the observed results, an increase in the efficiency of the model was noticed with this new approach.

IV. Methodology

This section will detail the procedure carried out, as well as the hardware and software materials used to carry out the experimentation. It should be noted that this work is of a research-experimental nature, so after identifying a problem and finding a novel solution, it is proposed to review the current tools, experiment with them, in order to determine the appropriate combination of tools and models to obtain the best result. This is why the methodology has been divided into 3 stages: research, experimentation, validation.

A. Research

In this initial stage, literature related to the topic in question was compiled in order to determine the objectives, scope and possible techniques to be used, for which a series of national and international web portals related to scientific research were consulted, such as: IEEExplore, Scopus, ResearchGate, etc. This stage includes the search for the repository of thermal images used in the development, both public (online repository) and own (taken at the campus of the University of Piura (UDEP)), the definition of preprocessing techniques and classification algorithms.

B. Experimentation

This stage starts once the research has been completed, here experiments will be carried out with the aim of obtaining a combination of preprocessing and optimal classification algorithms, all based on the literature previously consulted.

C. Validation

In this last part, a validation analysis is carried out, for which we evaluate the statistics that represent the good or bad performance of the combinations proposed above, using criteria and parameters such as: precision, confusion matrix, loss, computational weight and training speed.

V. Development

Now that the steps and criteria to be followed have been explained in general terms, we can go on to summarise specifically which datasets, pre-processing methods, classifiers, evaluation criteria have been used and how they have performed in the course of the work.

A. Revised Datasets

With regard to the Datasets consulted and obtained, we can list them in the following table:

TABLE I
Revised Dataset

Dataset	Source	Img	Classes	Observation
Data 1	Eksploatacja 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 Univeristy 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.

With regard to the Datasets, it is necessary to clarify that of all those mentioned above, only Data 1 and 2 have been used for the training of the classifier models, since this is the one that has been considered to meet all the requirements set out above. It should also be clarified that to solve the problem of excess classes, it has been considered plausible to unify several classes and discard many others in Data 1, following the following 2 criteria: Unifying all the images without distinctive operating current for the case of the same fault in the first case (unification) and discarding the images of 1 of the cameras, because the other one presents better resolution (discarding) in the second case. Having understood this context, the structure of the data under study is presented below in figure 1 1.

At this point, it is important to highlight that for the tests of some of the models that have been carried out, it was considered that the Coupling 1 coupling did not have a sufficient number of classes, so we worked only with the images of Coupling 2, while in other models they were used, as they can serve as a second validation value, Since despite belonging to the same dataset, it will present a different interaction dynamics than the Coupling2 images, it has a ”Missalingment” class, so the correct or not identification of this class with the model created from the other images can help to its continuous improvement.

B. Methods of pre-posting

Similarly, tables ¨reftab2 and ¨reftab3 describe all the pre-processing methods reviewed and tested, as well as what was concluded by using them.

TABLE II
PRE-PROCESSING (PART 1)

Pre-processing	**Type**	**Observation**
Otsu	Segmentation	Efficient but can only ”Differentiate between 2 classes”.
FFT	Differentiation	It allows to increase contrasts and thus to increase class differences.
YOLOV5	Segmentation	Very useful, accurate and can identify ACMs in thermographic images with very good accuracy.
CLAHE	Differentiation	I manage to increase contrasts in the image, however, it depends on a previous high-pass filter to improve this process.
DWT	Feature Extraction	It improves the performance of the SVM classifier by providing it with much better data input.

TABLE III
PRE-PROCESSING (PART 2)

Pre-processing	**Type**	**Observation**
SIFT-BoVW	Feature extraction	It allows for segmentation, as it identifies boundaries very well, but together with its classifier method, it is the one with the worst accuracy.
K-means	Segmentation	Lower performance than Otsu, but allows ”Differentiate between more than 2 classes”.
FOS	Feature extraction	It is used to identify intensity patterns by class, it is also useful for a possible simple thermographic analysis, as it allows the conversion of pixel intensity values into temperature.
GLCM	Feature extraction	A more complete set of information relevant to the classification task.
Data Aumentation	Data aumentation	This technique, which is widely known in the field, has been successful in preventing overfitting.
LDA	Differentiation	It did not mean a substantial improvement in the accuracy of the classifier model.
PCA	Differentiation	No significant improvement was observed in the models evaluated.

C. Classification Methods

After the evaluation of the pre-processing methods seen in tables II and III, tables IV, V and VI mention all the classification methods that have been considered, as well as their characteristics.

TABLE IV
CLASSIFICATION APPROACHES (PART 1)

Classification method	**Structure**	**Remarks**
VGG16	Concolutional neural network	It is simple in structure and easy to understand.

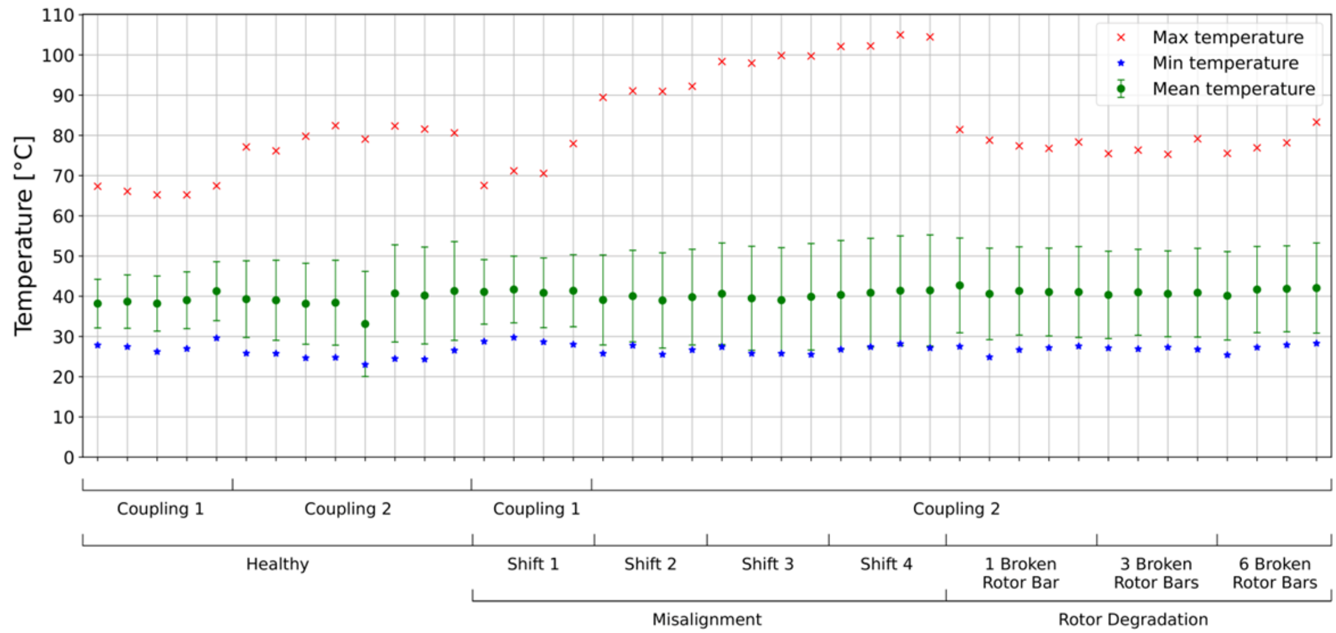

Fig. 1. Distribución de temperatura para la Data1 [15]

TABLE V
CLASSIFICATION APPROACHES (PART 2)

Classification method	Structure	Remarks
ResNet50	Convolutional neural network	Widely used for sorting work.
InceptionV3		Allows the reduction of convolutions by factorisation.
EfficientNetB2		Heavier model than the other CNNs but which does not show a comparative improvement with compared to other CNNs.
EfficientNetB5		
EfficientNetB7		
ERT	Ensemble supervised machine learning method	Classification method based on trial and error.
SVM	Supervised learning algorithm	It is a method that only supports 1D inputs.
KNN	Non-parametric, supervised learning classifier	It gave worse results than when SVM was used as a classifier.

TABLE VI
CLASSIFICATION APPROACHES (PART 2)

Classification method	Structure	Remarks
ViT	Transformer	Excellent results, but computational burden compared to SVM
ViTFSD	Transformer	I improve the results obtained in compared to CNNs and ViTs, but it takes longer to train than CNNs and ViTs.
CCT	Transformer	Good results, but there was a time when the result was lousy, but it is much lighter than ViTFSD.
CrossViT	Transformer	The heaviest model of all and not a substantial improvement.

D. Training and Testing Datasets

Having completed the stage where we have detailed the types of classification methods likely to be used, we must establish the testing and training protocols, as this is an equally important step as those reviewed above, especially to corroborate overfitting errors, so we will have the following training and testing Datasets presented in tables VII, VIII, IX and X.

TABLE VII
TRAINING DATASETS (PART 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 x2817 img)	Healthy
	Coupling2	1480 img (80%) + 1480 img (Data augmentation) + metadata (4GLCM x 2960 img)	Missalingment1
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

TABLE VIII
DATASETS DE ENTRENAMIENTO (PARTE 2)

Dataset Train	Fuente	Datos	Clases
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

TABLE 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

TABLE X
DATASETS DE TESTEO (PARTE 2)

Dataset Train	Fuente	Datos	Clases
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
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

In the tables VII, VIII, IX and X it can be seen that the training and test sets follow a ratio of 80-20%, however, it would be interesting to test in future projects some other distribution such as 70-30% to observe the behaviour of new models, although as we will see later on, a significant improvement is not expected.

E. Tested Models

For the assembly of all the models, a set of pre-processing techniques has been used, as well as a classification method. It is important to note that even within the same type of classifier there are variations, which is why there is an index to denote it.

F. Results

The evaluation will be based on three benchmark metrics and a confusion matrix, which provide information on the performance of the model.

TABLE XI
TESTED MODELS

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	
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

TABLE XII
TEST RESULTS (PART 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
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. Analysis Results

This section will review the results obtained, comment on what has been learned from the use of each tool and which have been determined to be useful and which have not, and will also analyse the results of the classifier methods and their performance based on the criteria of the previous chapter, in order to finally establish a model (pre-processing + classifier) based on the aforementioned. Finally, the validation tests to which the latter model will be subjected and its limitations will be discussed.

A. Pre-processing analysis

In the pre-processing, 2 types of segmentation were used, a main one to detect and trim the ACM and a secondary one to calculate the ROI. YOLOv5 was used for the primary segmentation and Otsu or K-means for the secondary segmentation, however, it is possible to use YOLOv5 for both tasks or even use Otsu Thresholding as the only segmentation method. However, the use of a primary and then a secondary identification method was considered, as this avoided making the mistake of calculating an ROI without first confirming that the ACM was in the achieved image. For the primary identification, YOLOV5m, YOLO's 5th generation median model, was used because it gave the best results for the datasets used. For segmentation, 2 methods were evaluated: OTSU and K-means, both offering good results initially, however, for more complex segmentation tasks OTSU was used, as it presented a similar performance to k-means, but it is a much simpler method and of low computational cost as detailed in equations 1, 2, 3 and 4.

$$w_{1_{(T)}} = \sum_{i=0}^{T} P_{(i)} \tag{1}$$

$$w_{2_{(T)}} = \sum_{i=T+1}^{L-1} P_{(i)} \tag{2}$$

$$\mu_{1_{(T)}} = \frac{\sum_{i=0}^{T} i \cdot P_{(i)}}{w_{1_{(T)}}} \tag{3}$$

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

Where w_1 represents the probability that a pixel belongs to class 1, w_2 the probability that it belongs to class 2. Furthermore, μ_1 is the mean of class 1, having established a threshold T such that pixels of intensity below this threshold are of class 1 and those of higher intensity belong to class 2. After this, the intra-class variance σ_w^2 and the inter-class variance σ_b^2 are calculated.

$$\sigma_w^2 = w_1\sigma_1^2 + w_2\sigma_2^2 \tag{5}$$

$$\sigma_b^2 = w_1 w_2 \cdot (\mu_1 - \mu_2)^2 \tag{6}$$

Finally, it is optimised in such a way as to establish a new threshold T that maximises the inter-class variance and minimises the intra-class variance.

On the other hand, the K-means method uses an optimisation function or also known as a cost function (J), which must be minimised and is represented in the equation 7.

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

Where k is the number of clusters, x_i the vector indicating the intensity of each pixel and μ_j the centroid of the jth cluster, which is calculated as represented in the equation 8.

$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \tag{8}$$

Where C_j is a grouping of pixels, in case there are 2 clusters as in our case, only C_1 and C_2 will exist.

After assembling 2 initial clusters based on random points, in such a way that the other points or pixels of the image closest to one cluster or the other belong to that cluster. After that, the cost function is calculated and recalculated until it reaches its minimum value so that, once the minimum function has been found, 2 new clusters are obtained so that now the pixels are actually stored in the cluster whose "colour distance" is the minimum, which means that it is grouped into 2 classes to which each group of pixels has a relationship, in this case by intensity and contrast.

It can be seen that the K-means tool, apart from being an iterative process whose operation is much more complex to understand than that of Otsu, constitutes a tool whose objective is not limited to the separation of 2 classes, as Otsu does. Unlike Otsu, K-means allows the separation of an image into more than 2 classes, a feature that is not required in this case, as it is only desired to separate the ROI (ACM) from the background. For all of the above, the choice of Otsu as a secondary segmentation method is considered appropriate.

For feature extraction there are 2 main groups which are simple feature extraction and complex feature extraction as shown in the table below.

TABLE XIII
TIPOS DE PREPROCESAMIENTO PARA EXTRACCIÓN DE CARACTERÍSTICAS

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

This distinction is due to the fact that both FOS and GLCM extract simple image features, data such as: Mean, Variance, Skewness (FOS) and Contrast, Energy, Homogeneity (GLCM), while, on the other hand, DWT and CNN tools analyse and extract deeper layered features, in the case of DWT analyses the image at various frequencies, in the case of DWT it analyses the image at various frequencies, making a pattern separation based on this and the direction of the pixels (horizontal, vertical and diagonal), while CNN calculates and assigns weights to unobservable features of the images that have gone through multiple convolution processes in order to reduce the dimensionality of their components.

In order to determine which method is suitable, the scope and performance of each method was studied. For the extraction of simple features, GLCM was used because it analyses similar parameters to FOS, but taking into account the influence of neighbouring pixels, which can be beneficial for analysing temperature changes in the engine. However, for the extraction of complex features, the performance and computational time of each method was taken into account, resulting in DWT being the method of choice, as it outperformed convolutions by far.

With respect to contrast enhancers/differentiators, CLAHE has stood out as a great tool in contrast to Data Augmentation. HSI was also used as a direct method for contrast enhancement, this is prior to CLAHE, so that the original dynamics of the image are not lost.

Finally, with respect to the filtering preprocessing methods, it is necessary to comment that, although the FFT filter performs a filtering task, it does it with the objective of increasing contrast in the image and, as explained before, CLAHE already performs this action in a much more effective way, so it would be redundant to maintain these 2 preprocessing methods and, therefore, only CLAHE was kept for future models.

B. Classifier analysis

The classification methods were grouped into three categories: CNN, machine learning algorithms and Vision Transformers (ViT).

- **CNN:** Six architectures were evaluated. In some models, CNNs were used for training and prediction (as in models 1 and 2), while in others (such as models 4, 6 and 11) only as a feature extraction algorithm. Although the performance was positive, a slight improvement in accuracy was observed when replacing the last layer of CNNs with an SVM classifier.
- **Machine Learning Algorithms:** SVM, KNN and ERT were used. The best results in this category came from SVM-based model 3, which stood out for its speed and accuracy. KNN showed less accurate results and ERT, used with BoVW-generated histograms, gave good results, although with a heavier model than other options of similar accuracy.
- **Vision Transformers (ViT):** Four ViT models were evaluated. Near-perfect prediction results and shorter training times than CNNs were obtained.

C. Analysis of datasets

Regarding the datasets, it is relevant to analyse the two main sources of data from which this work is based. Firstly, we have the data obtained in [15], where, despite having a large number of images and a considerable number of classes, it is observed that the type of machine under study is the same, in addition to the fact that the angle of capture of the image is maintained for the entire dataset, which means a loss of information from the ACM.

This would explain why most of the trained models, including those detailed in the present work, give such good results, as the fault patterns belonging to the same machine are widely perceptible for a classifier model such as the ones evaluated.

However, this cannot be extrapolated to industry, since in this context, there is no single type of engine to be analysed and even less so in the same room or with the same background and/or operating load, so the results observed, although highly promising, should be taken as an opportunity to extrapolate this type of research and promote the collection of independent data.

In conjunction with the aforementioned data issues, the idea of obtaining a new dataset based on engines of different types, use, load, operating context, etc. arose. This is why, in coordination with multiple institutional bodies of the University of Piura, a visit to multiple places and areas of operation where these types of machines could be found was coordinated. Finally, the structure of the data collected can be summarised in table XIV.

TABLE XIV
ESTRUCTURA DE LA DATA 4 (PARTE 1)

Ubicación	N° imágenes	Modelo
Laboratorio de Sistemas Automáticos de Control	144	El motor llego 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

D. Best model

The final model collected a combination of the best preprocessing algorithms, which are detailed below:

- **Engine detection with YOLOv5m:** Changed from YOLOv5s to YOLOv5m to improve engine detection in low contrast images. After training, the model detects the area of interest, crops the images and removes irrelevant elements.

TABLE XV
ESTRUCTURA DE LA DATA 4 (PARTE 2)

Ubicación	N° 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

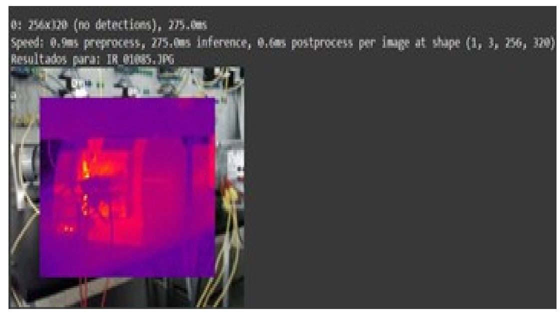

Fig. 2. Failed detection with YOLOv5s with new data

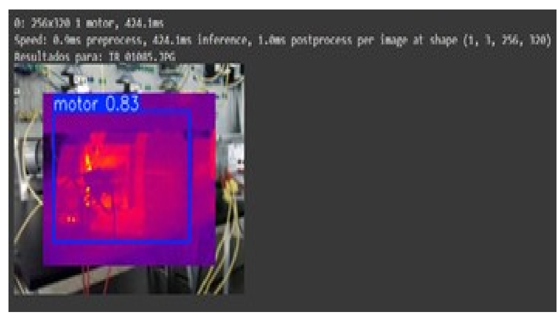

Fig. 3. Detección correcta con YOLOv5m con la nueva data

- **Segmentation with Otsu** It segments the trimmed engine, identifying relevant regions by means of an optimal greyscale threshold.
- **Conversion to HSI colour channels:** In this step, the image channels are changed from RGB to HSI, of which only channel I (intensity) is used, as it best represents the temperature of a thermographic image.
- **Resizing and contrast enhancement with CLAHE:** The images are resized to a size of 224 x 224 pixels, as this is a size widely used in image processing algorithms, then CLAHE is applied, in order to improve contrast and make small details more visible.
- **Feature extraction with GLCM:** Contrast and homogeneity features are extracted from the images resulting from the previous step and stored in a vector of length 2 for later use.

- **Transformada wavelet discreta (DWT):** This tool is used to extract 4 frequency subbands: cA (low frequency) and cH, cV and cD (the latter 3 high frequency), which are flattened and combined into a feature vector. Each of the subbands corresponds to an image of 112 x 112 pixels, when flattened and concatenated, a vector of length 50176 is obtained.
- **Data preparation:** The combined features give a total of 50178, each of the initial 9071 images has a vector of length 50178.
- **Dimensionality reduction with PCA:** In order to make the model much lighter, it was decided to use PCA for the reduction of components, previous studies supported the reduction of 50178 to 50 components, without sacrificing the efficiency of the model.
- **Hyperparameter optimisation:** Grid Search was used to select the best values for the regularisation parameter, the gamma factor and the kernel for SVM, the parameters identified as optimal were a regularisation value of 1, a gamma value of 0.01 and the use of the "linear" kernel.
- **Training and evaluation:** With the optimal parameters and a data distribution of 80% for training and the remaining 20% for validation, the SVM model was trained, achieving 99.72% accuracy, confirming the good performance of the previously selected preprocessing.
- **Confusion matrix and results:** The matrix shows excellent accuracy in the classification of "healthy", "misalignment" and "broken rotor bars".

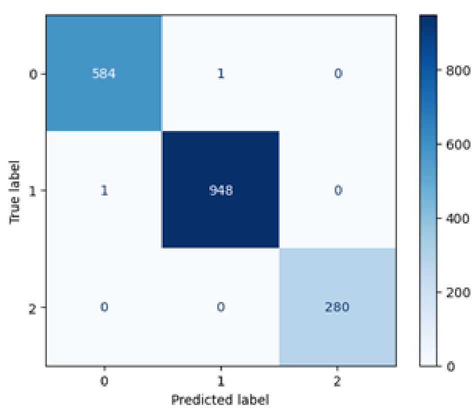

Fig. 4. Matriz de confusión

- **Subclasses of faults:** According to the data studied, it was possible to appreciate within the misalignment faults and broken rotor bars, different types of grades for each one, which is why, with the aim of making a much more complete and detailed identification, a hierarchical identification system was established, where the first SVM model (the one mentioned above) recognises whether or not there is a fault in the engine (misalignment or broken bars), if there is, there are two additional SVM models (one for each fault) that classify the corresponding fault in each of its subclasses, the models were trained with parameters similar to the general model and only with data unique to each fault, obtaining the following results:

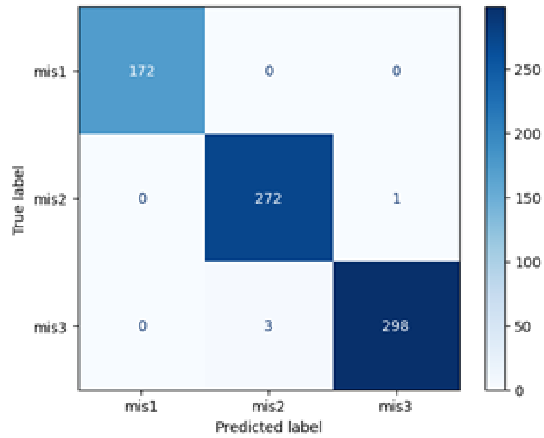

Fig. 5. Matriz de confusión de desalineamiento

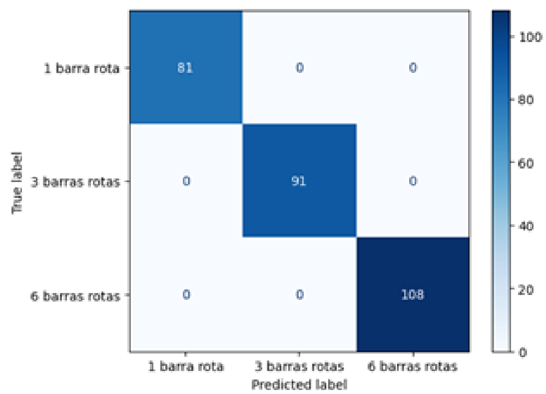

Fig. 6. Matriz de confusión de barras rotas

- **Validation test:** As has been detailed in previous chapters, it is suspected that the models may be suffering from overfitting, which is why it has been decided to carry out a test with a validation dataset composed only of engines in a good state (Healthy) but of different sizes, positioning, power, location, load, etc. In other words, Data4, is used to test the robustness of this model before using them as training data. This test was carried out using only the YOLOV5m trained with these images, so that it can recognise engines and, based on the engines found, and after applying the methods described above, the result shown in figure 7 was obtained.

```
[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. Interface

As a final product, a user interface (GUI) was developed, with the objective of making it easier and more user-friendly for the user to detect engine failures in real applications. The pyQT5 library was used to create the GUI, which offers a wide series of customisable widgets and a specialised application for the design, which is QtDesigner, as shown in figure 8.

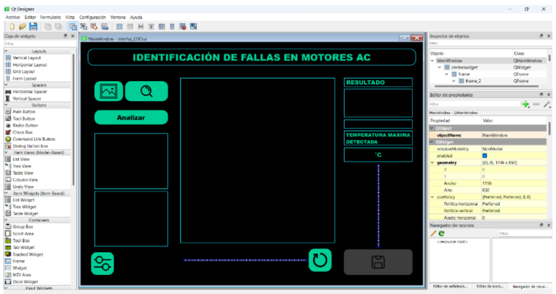

Fig. 8. Diseño de la interfaz dentro de QtDesigner

With the button shown in the following image, a dialogue box will open where you can choose the image from the local files of the device.

Fig. 9. Botón para seleccionar la imagen

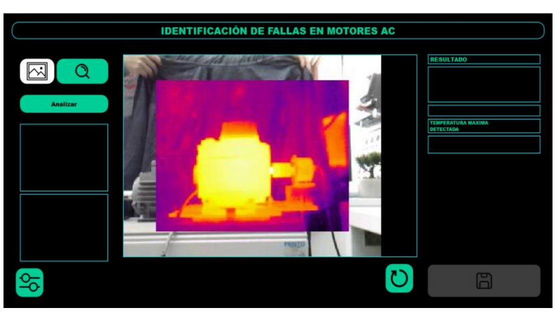

Fig. 10. Interfaz una vez seleccionada la imagen

With the following button, the YOLO model will identify the engine in the entered image, in case it does not identify it correctly, it will give the option to select it on its own.

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

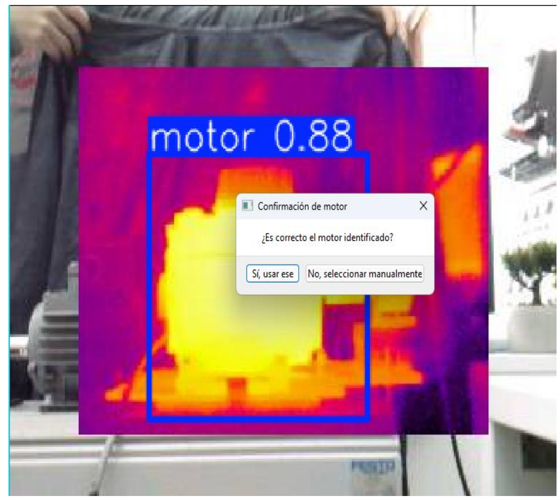

Fig. 12. Motor identificado de manera correcta

Once the engine has been identified, either manually or automatically, the SVM model proceeds to the analysis, using the ”analyse” button. The results are displayed on the right-hand side of the application as shown in figure 1 13.

It should be noted that the upper and lower temperature limits can be set using the settings button on the bottom left, as shown in Figure 14.

Finally, the application offers the possibility to generate a pdf report of the analysed image, which contains the results obtained: the class predicted by the SVM, the radar graph with the probabilities of each class and the maximum temperature estimated in the image, as shown in figure 15.

VII. Conclusions

- During the tests, the discrete wavelet transform proved to be the preprocessing that provided the most information, reaching an accuracy of 99.18% in its evaluation with SVM, however, due to the nature of the Dataset, there is always the risk of overtraining the model, this is evident when analysing the model’s separation margin, which is why techniques such as crossvalidation were applied, and parameters such as the number of components analysed or the regularisation parameter were varied. All these

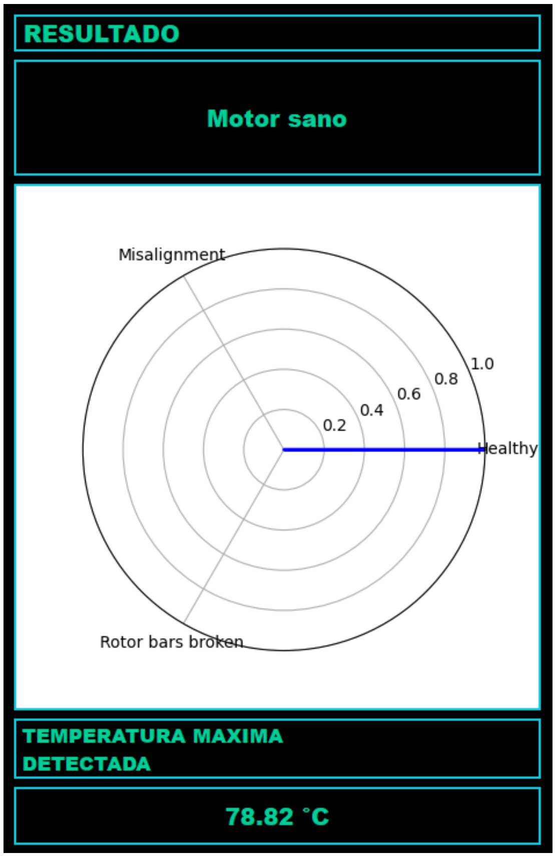

Fig. 13. Resultados obtenidos

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

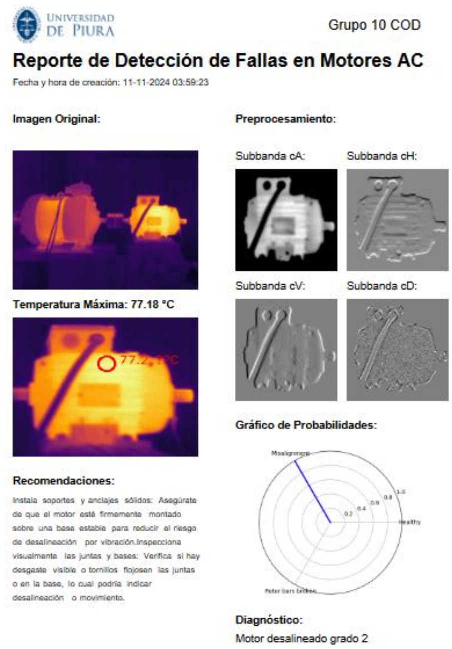

Fig. 15. Reporte final de la interfaz

experiments obtained equally high accuracy percentages, as well as much higher separation margins, corroborating the validity of the model and making it much more robust.

- Within the analysis performed with PCA (Principal Component Analysis) in model 3, it was detected that within the 50176 features obtained from the 4 sub-bands by DWT, 95% of the variance was represented by only the first 593 components of each image, this means that only this number of components can be used without sacrificing relevant information from the sub-bands.
- Models that have used deep CNNs such as EfficientNetB2, B5 and B7 have not proven to be good prospects, as compared to other models, despite achieving very good accuracy, the training time represents by far the biggest problem in terms of industrial implementation.
- The SVM-1 and SVM-2 classification method which has employed an SVM as the primary classification tool has demonstrated better performance in this task than the KNN tool.
- The model that uses FOS as a feature extractor, given its accuracy and training speed, is a plausible model to implement in case what is sought is a fast response system, as the computational load it requires is the lowest of all the models evaluated, obviously also taking into account that, in turn, they have been the least accurate models among them (Model 11 and 12).
- In case you want to apply a predictive control system or simply a predictive maintenance protocol using ther-

mographic analysis in conjunction with neural networks, given the particularity of the latter, the most appropriate would be to provide at least a set of thermographic images of the healthy engine in the context in which it will work, so that the classifier method can at least learn the patterns of a healthy engine, in this way we will always ensure that at least the task of identifying faults in an ACM will be fulfilled efficiently.

REFERENCES

[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] Fluke Corporation and The Snell Group. Introducción a los principios de la termografía, 2009. Accedido: 11 de noviembre de 2024.

[9] 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.

[10] IBM. ¿qué es la segmentación de imágenes? Accedido: 11 de noviembre de 2024.

[11] 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.

[12] Seung Hoon Lee, Seunghyun Lee, and Byung Cheol Song. Vision transformer for small-size datasets. *arXiv preprint arXiv:2112.13492*, 2021.

[13] 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.

[14] MathWorks. Convolutional neural network. Accedido: 11 de noviembre de 2024.

[15] Mohamad Najafi, Yasser Baleghi, and Seyyed Mehdi Mirimani. Thermal image of equipment (induction motor), 2020.

[16] 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.

[17] Ashraf Navaid Sabah and Zainul Abdin Jaffery. Fault detection of induction motor using thermal imaging. In *2022 IEEE IAS Global Conference on Emerging Technologies (GlobConET)*, pages 84–90. IEEE, 2022.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] Universidad de Sevilla. Programa de la asignatura pid-filtros digitales. Accedido: 11 de noviembre de 2024.

[23] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

[24] 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.