A Deep Learning Solution for Quality Control in a Die Casting Process

Paula Mielgo[1], Anibal Bregon[2], Carlos J. Alonso-González[3], Daniel López[4], Miguel A. Martínez-Prieto[5] and Belarmino Pulido[6]

[1,2,3,5,6] *Department of Computer Science, University of Valladolid, Valladolid, Spain*
paula.mielgo@uva.es, anibal.bregon@uva.es, calonso@uva.es, miguelamp@uva.es, b.pulido@uva.es

[4] *Factoría de Motores, HORSE Spain, Valladolid, Spain*
daniel.g.lopez@horse.tech

ABSTRACT

Industry 4.0 aims for a digital transformation of manufacturing and production systems, producing what is known as smart factories, where information coming from Cyber-Physical Systems (core elements in Industry 4.0) will be used in all the manufacturing stages to improve productivity. Cyber-physical systems through their control and sensor systems, provide a global view of the process, and generate large amounts of data that can be used for instance to produce data-driven models of the processes. However, having data is not enough, we must be able to store, visualize and analyze them, and to integrate induced knowledge in the whole production process. In this work, we present a solution to automate the quality control process of manufactured parts through image analysis. In particular, we present a Deep Learning solution to detect defects in manufactured parts from thermographic images of a die casting machine at an aluminum foundry.

1. INTRODUCTION

In 2015, the foundational definition and main design principles of Industry 4.0 were presented by Hermann, Pentek, and Otto (Hermann, Pentek, & Otto, 2016) as a guide for implementing Industry 4.0. This definition encompasses the four key components of Industry 4.0: Cyber-Physical Systems (CPS), the Internet of Things (IoT), the Internet of Services (IoS), and Smart Factories. The objective of Industry 4.0 is the digital transformation of manufacturing and production industries. Cyber-physical systems represent the foundation of Industry 4.0, which frequently involve control systems, embedded software, and a substantial array of data coming from sensors and actuators. These systems generate a vast quantity of data, which must be integrated and analysed in order to achieve the designation of "smart factories".

The concept of Smart Manufacturing was introduced in the United States to facilitate the deployment of emerging technologies in manufacturing, including the Industrial Internet of Things (IIoT) and Artificial Intelligence (AI). Smart manufacturing, also known as intelligent manufacturing, focus on the adoption of these advanced information and manufacturing technologies to optimize the production (Zhong, Xu, Klotz, & Newman, 2017). The main focus of this methodology is to enhance the quality, traceability, and efficiency of the production process. Each industrial revolution has been accompanied by an increase in productivity, which has been attributed to the introduction of new technologies, including the steam engine, electricity, and digital technology. For the fourth industrial revolution, the primary factor driving productivity enhancement is the far-reaching impact of these vast quantities of data, which influence not only production but also other sectors, particularly engineering processes. This allows for more effective decision-making processes. However, having data is not enough; it must be stored, visualised and analysed, and the resulting knowledge must be integrated into the entire production process. This can be achieved, for instance, by producing data-driven models that can subsequently be employed in a digital twin (which represents another crucial component in the smart factory framework). This is of particular importance in those smart factories where there are few, if any, analytical models available, due to the nature or complexity of the processes.

Artificial intelligence in general and Machine Learning (ML) in particular play a pivotal role in this contemporary development of smart manufacturing characterised by the production of data-driven models. The integration of ML with the production process facilitates the reduction of production time, improvement of quality and the elimination of unnecessary

waste. Some literature reviews demonstrate the use of ML techniques in industrial environments to enhance planning and control procedures (Usuga Cadavid, Lamouri, Grabot, Pellerin, & Fortin, 2020), as well as specific applications in quality control tasks (Peres, Barata, Leitao, & Garcia, 2019) (Patel & Jokhakar, 2016). Moreover, Deep Learning (DL) has contributed to a significant developement of computer vision. Architectures such as Convolutional Neural Networks (CNNs) or Vision Transformers (ViTs) have made significant contributions in many fields, including manufacturing. These models are able to process images and videos to ensure product quality (EL Ghadoui, Mouchtachi, & Majdoul, 2023) (Villalba-Diez et al., 2019) (Cumbajin et al., 2023) and also to monitor manufacturing processes (Alfaro-Viquez, Zamora-Hernandez, Benavent-Lledo, Garcia-Rodriguez, & Azorín-López, 2022).

In this paper we present a proposal to address the automation of the quality control process through image analysis. The reasons behind our proposal are twofold. On the one hand, productivity will increase if we can predict, on the early stages of the manufacturing process, a subset of manufactured parts that will pass, with total safety, all the quality control checks. This will enable the skipping of checks that increase the costs of the manufacturing process and limit the capacity of manufactured parts per shift. On the other hand we can perform an early detection of defective manufactured parts, that will allow their removal from the production chain. To achieve this objective, we will use CNNs (LeCun, Bottou, Bengio, & Haffner, 1998) and ViTs (Dosovitskiy et al., 2020) to detect, from thermographic images, defects in manufactured parts. In order to achieve this objective, a case study of an aluminum die casting plant at a car engine manufacturing plant will be employed.

The remaining of the paper is organized as follows. Section 2 presents the case study of the die casting plant. Section 3 briefly introduces the preprocessing/DL techniques used in this work. Section 4 presents the proposed architecture for quality control at the die casting plant by the analysis of the thermographic images of the mold. Section 5 introduces the experimental setup and the experiments we carried out, and discusses the results obtained with the proposed solution against a subset of state-of-the-art computer vision architectures. Finally, Section 6 draws the main conclusions and future directions of this work.

2. THE ALUMINUM DIE CASTING PLANT

In a casting plant, the first process is the aluminum melting. The aluminum arrives at the facilities in the form of ingots that are melted at a temperature of about 720 degrees Celsius in the different melting towers. This molten aluminum is transferred to holding furnaces that are located next to the die casting machine. A small ladle takes the molten aluminum and pours it into a shot chamber where it is ready to be injected into a steel mold, known as die. The molten aluminum is forced into the die with a clamping force of about 22kN. Right after injection, the machine introduces water and air into the different cooling circuits of the mold to reduce the mold temperature and to solidify the part. The high pressure holds the metal in the die until it solidifies. Afterwards, the injector opens the die, and a robot extracts the part with a series of leftovers that are later removed. Right after the part is removed, thermographic cameras check the mold for temperature problems capturing two thermographic images, one for each part of the mold (Figure 1 illustrates one of these images). The next step is to pour a release agent into the mold to prevent the part from sticking to it, and right after that, two additional thermographies are taken.

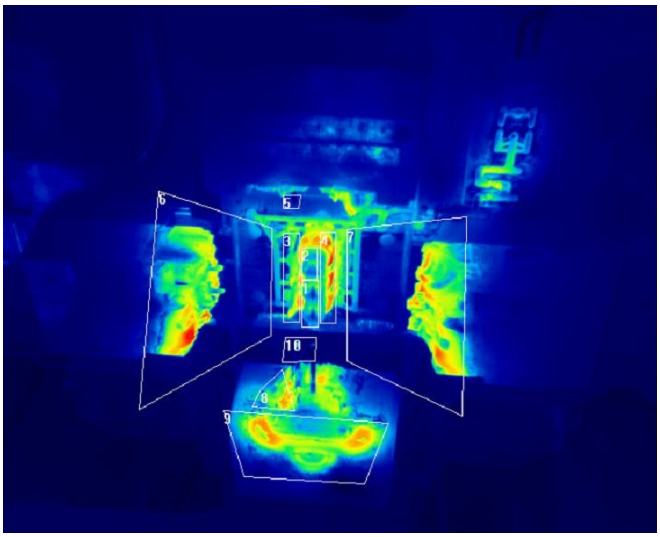


Figure 1. Thermographic image

The process continues with a first visual inspection of the manufactured part and then a thermal treatment process is carried out to release any stresses in the material. To complete the manufacturing process in this plant, a series of machining operations are carried out on the parts. Next, in order to test the quality of the injection process a leak test is performed. Finally, all parts are inspected by an operator in a second visual inspection, looking for parts that do not comply with quality standards. Figure 2 schematically shows this production process.

For each manufactured part, the thermographic cameras generate three files, a source file, where all the data is found, and it is only exploitable with a specific proprietary software of the supplier. A csv file, where the image search areas (ROI) and the maximum temperature of that region are found. Finally, there are the four thermographic images of the mold. These thermographic images are used to detect degradation in the molds, but not to check the quality of the parts. As a consequence, these images are randomly inspected by human

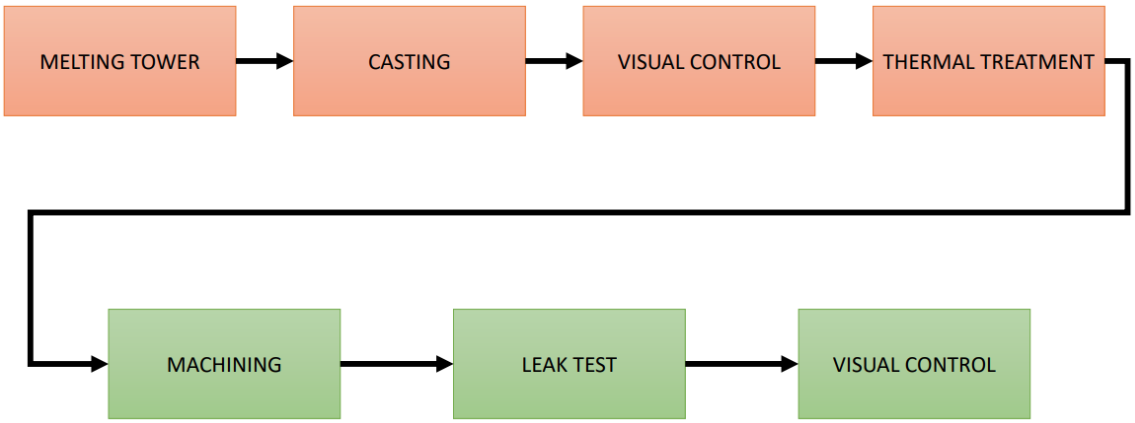


Figure 2. Process diagram

experts, who do not have any automatic tool to support them in this task.

2.1. Problem Formulation

As we just mentioned, with the current available tools it is not possible to detect non-conforming parts as soon as they are manufactured by means of these thermographic images. The parts follow the manufacturing flow until they reach the leaking inspection. Those parts with evident defects are rejected either during the casting process or in the first visual inspection. However, a small percentage of the parts can exhibit internal defects not evident to the human eye. Consequently, to guarantee the good quality of all the manufactured parts, all the pieces that fulfill the visual control requirements must also pass through the leak test. Our aim in this work is to use the thermographic images of the molds created by the casting machine to identify a subset of the parts that can skip this inspection, thus improving the performance of the machinery on the production lines.

This task can be done either by identifying a part as defective (NK) and sending it for melting down, or, more importantly, since most of the manufactured parts are non-defective (OK), by determining, with absolute certainty, that a part is correct. In both cases the result is the same, no additional quality controls are required, and the bottleneck will be reduced. However, this is a difficult task since, on the one hand, defects that reach the leak test are not evident to the human eye and we only have thermographic images of the molds, not from the manufactured parts; and on the other hand, because quality requirements in the automotive industry are quite strict and we must guarantee that defective parts are not considered as non-defective.

3. BACKGROUND

This section provides a brief overview of preprocessing techniques, CNNs and ViTs, which are the main techniques used in this paper.

3.1. Preprocessing techniques

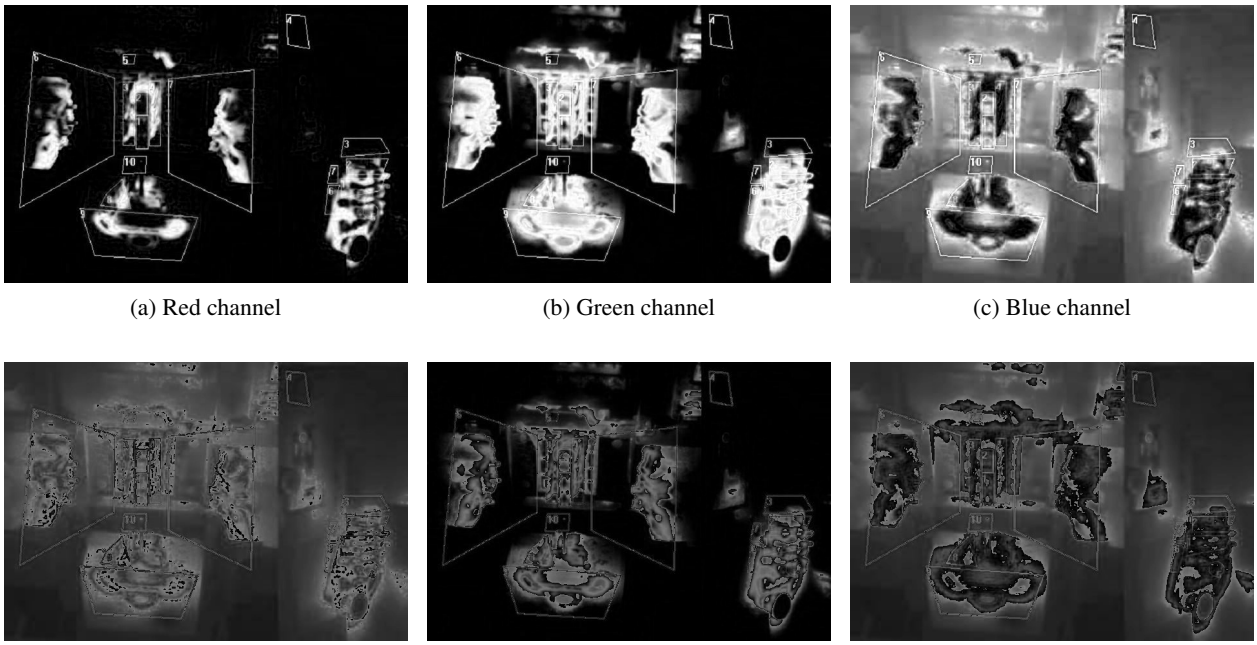
Classical image transformations are frequently used as an initial step prior to training a model with the objective of enhancing the classification results.

RGB plane decomposition. An RGB image is composed of three primary colour channels: red, green and blue. It may be interesting to process either of these channels separately in order to remove noise or to simplify the computation process. This decomposition has already been done with other formats such as HSV, CIEL*a*b* or YCbCr, improving the results compared to the original format (Sachin, Sowmya, Govind, & Soman, 2018). In an RGB image, a primary channel is derived directly from the original image, whereas a complementary channel is computed by averaging two primary channels. Let x be a pixel, and let x_i be its value in the channel i, the complementary channels (cyan, magenta and yellow) are computed as follows.

$$x_{cyan} = \frac{x_{green} + x_{blue}}{2}$$
$$x_{magenta} = \frac{x_{blue} + x_{red}}{2}$$
$$x_{yellow} = \frac{x_{red} + x_{green}}{2}$$

Consequently, it is possible to decompose not only primary channels but also complementary ones. Figure 3 illustrates an example of the RGB plane decomposition technique applied to an image of the dataset.

Grayscale transformation. Some papers in the literature



(a) Red channel (b) Green channel (c) Blue channel

(d) Magenta channel (e) Yellow channel (f) Cyan channel

Figure 3. Plane decomposition applied to thermographic image

have employed grayscale transformation as a preprocessing step prior to CNN classification (Y. Xie & Richmond, 2018) (Hsu et al., 2021). This not only improves the results, but also reduces the computation time. There are several methods for transforming an RGB image to grayscale. The most popular is the National Television System Committee (NTSC) formula (Jack, 2011). Let x be a pixel, and let x_i its value in the channel i. Then, the value of x in the grayscale will be:

$$x_{gray} = 0.299 \cdot x_{red} + 0.587 \cdot x_{green} + 0.114 \cdot x_{blue}$$

Figure 4a illustrates the grayscale transformation applied to a thermographic image.

Gamma correction. This technique was initially developed to correct the power-low transformations applied by some image capture, printing or display devices (Gonzalez & Woods, 2008). However, it can also be used to adjust the contrast of the image, making it a preprocessing technique. For instance, it has been used in the process of binarising thermographic images (Wang, Zhang, Ni, & Ren, 2021). The gamma correction has the following equation for a pixel x:

$$g(x) = \left(\left(\frac{x}{255} \right)^{\left(\frac{1}{\gamma}\right)} \right) \times 255$$

where γ is a parameter.

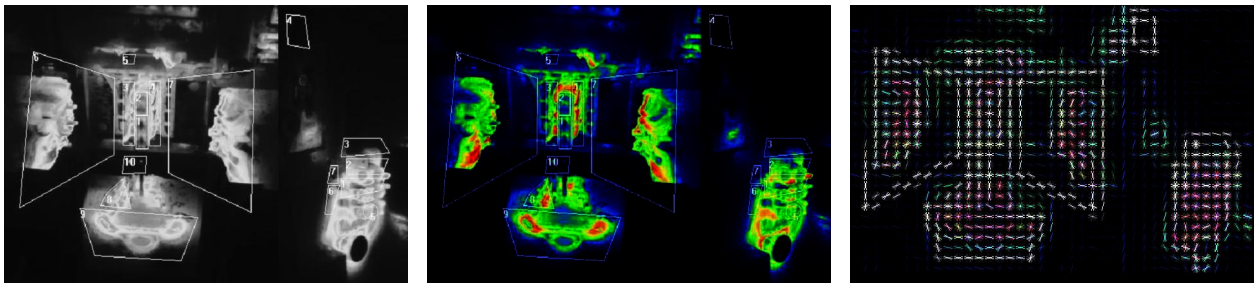
Figure 4b provides an example of gamma correction.

RGB-HOG. The Histogram of Oriented Gradients (HOG) is a well-established technique used for image feature extraction. It was introduced in (Dalal & Triggs, 2005) for pedestrian detection but has subsequently been used for other applications such as face recognition (Déniz, Bueno, Salido, & De la Torre, 2011). The fundamental concept is the representation of small image cells by accumulating the 1-D histogram of gradient directions of their pixels, which allows for the characterisation of local object appearance in the image. The application of HOG descriptors has traditionally been limited to grayscale images (with only one channel). However, it may be important to retain the color information when dealing with three-channel images (Lahmyed, El Ansari, & Ellahyani, 2019). One possible approach is the use of RGB-HOG, which calculates the HOG characteristic on each RGB channel and then stacks them to form the final descriptor. Figure 4c illustrates an example of the RGB-HOG technique.

3.2. Deep learning architectures

DL models are able to exploit the underlying information in large volumes of data in an efficient way. Consequently, in recent years there has been a significant impact of DL in a range of technological fields, including Speech Recognition (Nassif, Shahin, Attili, Azzeh, & Shaalan, 2019), Natural Language Processing (NLP) (Deng & Liu, 2018) and Computer Vision (Esteva et al., 2021). For image processing, CNNs and ViTs are the most widely used architectures.

Convolutional Neural Networks. CNNs are a type of Deep Neural Network that attempts to emulate the visual cortex. In



(a) Grayscale (b) Gamma correction 0.2 (c) RGB-HOG

Figure 4. Preprocessing techniques applied to thermographic image

recent years, a number of proposed works have demonstrated their ability to extract patterns in images and videos (Li et al., 2014) (Dyrmann, Karstoft, & Midtiby, 2016). This is possible thanks to their hierarchical layer structure. In essence, the initial layers identify fundamental elements such as curves or lines, which are subsequently combined to form objects in deeper layers, such as silhouettes or faces. The main components of a CNN are (Figure 5): convolutional layers, pooling layers and the fully connected layer.

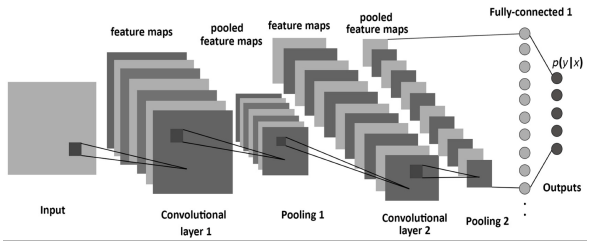


Figure 5. CNN architecture. Extracted from (Albelwi & Mahmood, 2017)

The purpose of the convolutional layer is to extract a feature map from an input pixel matrix. This is achieved through the use of kernels that weight the input values. Then, it is necessary to apply an activation function, which introduces the non-linearity property and restricts the values to be passed to the next layer.

The pooling layer reduces the dimensionality of the feature map obtained in the convolutional layer. This approach helps to reduce the computational complexity of the training process while maintaining the key information extracted by kernels. For this purpose, a pooling kernel is used to perform mathematical operations on sub-matrices of the feature maps.

Convolutional and pooling layers are employed iteratively to generate multiple feature maps that accurately reflect the information contained in the image. Finally, the fully connected layer is used to transform the three-dimensional feature matrix into a one-dimensional array. In this process, all the features extracted in the preceding layers are efficiently combined.

In recent years, CNNs have been demonstrated to be an effective solution for anomaly detection in industrial environments. This is exemplified in (Weimer, Scholz-Reiter, & Shpitalni, 2016), where a CNN was evaluated against other manual feature extraction methods using a textured surface dataset for defect detection.

Vision Transformers. Transformers (Vaswani et al., 2017) are a type of DL architecture that was developed for use in Natural Language Processing. Due to their success, some researchers attempted to adapt them to the field of Computer Vision. This resulted in the creation of ViTs (Dosovitskiy et al., 2020).

The fundamental concept of transformers is the use of attention mechanisms to determine and weight the relationships between the input network elements (tokens). This is achieved by constructing a matrix in which all the tokens are related, assigning them a value between -1 and 1 according to the importance of their relationship. To make this process more efficient, transformers use Multihead Attention, which divides the dimension of the token space to enable each of the attention mechanisms (referred to as attention heads) to process one of the subspaces. Although the objective of Transformers is to identify relationships between all elements, this is not feasible in ViTs due to the high dimensionality of the images. To overcome this challenge, prior to entering the network, a division into small cells, known as patches, is carried out in order to study the relationships between them. In addition, a position embedding is added to patches in order to preserve information of their position. Then, the network incorporates encoder blocks. Each block alternates Multi-Head Attention blocks with Multi Layer Perceptron (MLP) blocks, adding a Normalisation Layer before each one and a Residual Connection behind. The final part of the network is a classification layer, formed by an MLP with a single linear layer. Figure 6 provides an illustration of the complete structure.

ViT has also been employed in the detection of anomalies in industrial environments. In (Smith, Du, & Kurien, 2023), a conventional and a modified version of ViT were used on a leather object detection dataset.

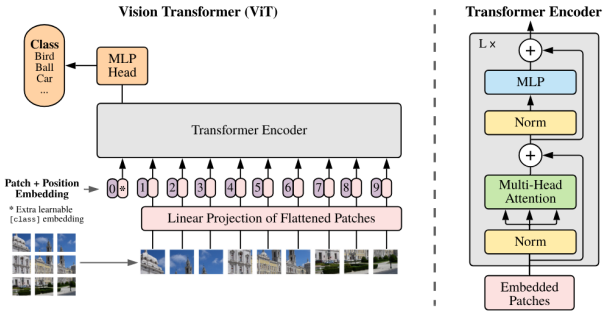


Figure 6. ViT architecture. Extracted from (Dosovitskiy et al., 2020)

4. ARCHITECTURE PROPOSAL

This section presents the proposed architecture. An overview of such architecture can be found in Figure 7. As shown in the figure, two distinct parts are distinguished. First, the construction, which includes the fitting process and the selection of two models, the best for detecting non-defective parts and the best for determining that a part is defective. Second, the inference, which enables the classification of new thermographic images.

As previously stated in section 2, there are four thermographic images of the mold for each part, two of them taken before the application of the release agent and two after its application. It was determined that the images taken prior the application of the release agent would be the ones to use. Furthermore, due to the lack of individual labels for each of the images of a part, the two images of the mold were combined by placing one next to the other without the color scales (as shown in Figures 3 and 4).

4.1. Model construction

The initial stage of the process involves the combination of preprocessing techniques with pretrained models over the experimental set to obtain fitted models. After that, a search is conducted in order to select the best model for identifying non-defective parts and, on the other hand, the best for identifying the defective ones. In addition, for each of these models, a minimum confidence value, known as threshold, is set to ensure the quality of predictions. The output of this stage will be, on the one hand, the best model for classifying OK images, together with its associated preprocessing technique and the OK threshold; and on the other hand, the best model for classifying NK, in conjunction with its associated preprocessing technique and the NK threshold. All this process is detailed next.

First, preprocessing techniques described in subsection 3.1 were applied to the experimental set (training and validation subsets). In particular, for each image, the colour planes (red, green, blue, cyan, magenta and yellow) are decomposed, a transformation to grayscale is performed, an RGB-HOG is applied and a gamma correction is conducted with $\gamma \in \{0.1, 0.2, 0.5, 1.5, 2.0, 3.0, 4.0, 5.0\}$. All techniques are applied separately and we also consider the non-preprocessed version.

Second, each of the image sets obtained from the previous phase, and also the original set, are employed to train CNNs and ViTs. In order to facilitate the training process, pretrained models are used for the experiments. The selected CNNs are ResNet-50 (He, Zhang, Ren, & Sun, 2016), SqueezeNet (Iandola et al., 2016), EfficientNet (Tan & Le, 2019), ResNeXt-50 (S. Xie, Girshick, Dollár, Tu, & He, 2017), ConvNeXt-S and ConvNeXt-L (Liu et al., 2022). The selected ViTs are the original ViT, a hybrid ResNet - ViT (Dosovitskiy et al., 2020), EVA-02 (Fang et al., 2023) and EfficientViT (Cai, Li, Hu, Gan, & Han, 2022). In all cases, an EarlyStopping regularization function is applied based on the validation loss. Additionally, the best model is selected based on its F1-score value in the validation set. The output of this stage is a set of fitted models that combine the preprocessing techniques and the original images with the selected models.

Third, the fitted models can classify any image directly. However, it is important to considerate the confidence of the prediction. To this end, models return a score value in the $[0, 1]$ interval, together with the predicted label, which represents the likelihood that the image belongs to each class. Binary classification usually provides a single score value that represents the likelihood for the positive class, which is the defective one in our case. Consequently, if the associated score for the defective class is $S_{defective}$, the non-defective score will be $S_{non-defective} = 1 - S_{defective}$. These values may be used to make more accurate predictions by setting thresholds that must be exceeded by the scores. In particular, two values are selected. One to detect if a part is non-defective, the OK threshold, which implies that the $S_{non-defective}$ score must be over that value to classify an image as OK, and another to determine whether a part is defective, the NK threshold, which must be exceeded by the $S_{defective}$ score to classify a part as NK.

Finally, the best model for classifying OK images and the best model for classifying NK images are selected based on the concepts of scoring and threshold. The objective is to identify the most effective model for detecting non-defective parts, associated with the first threshold, and the most effective model for determining defective parts, related to the second threshold. For that end, once the models have been fitted, a search is conducted to identify both models. In the first case, the objective is to identify the model that maximizes the number of OK images classified correctly before classifying an NK as non-defective. In the second one, the aim is the opposite. In both cases this is achieved by following the steps

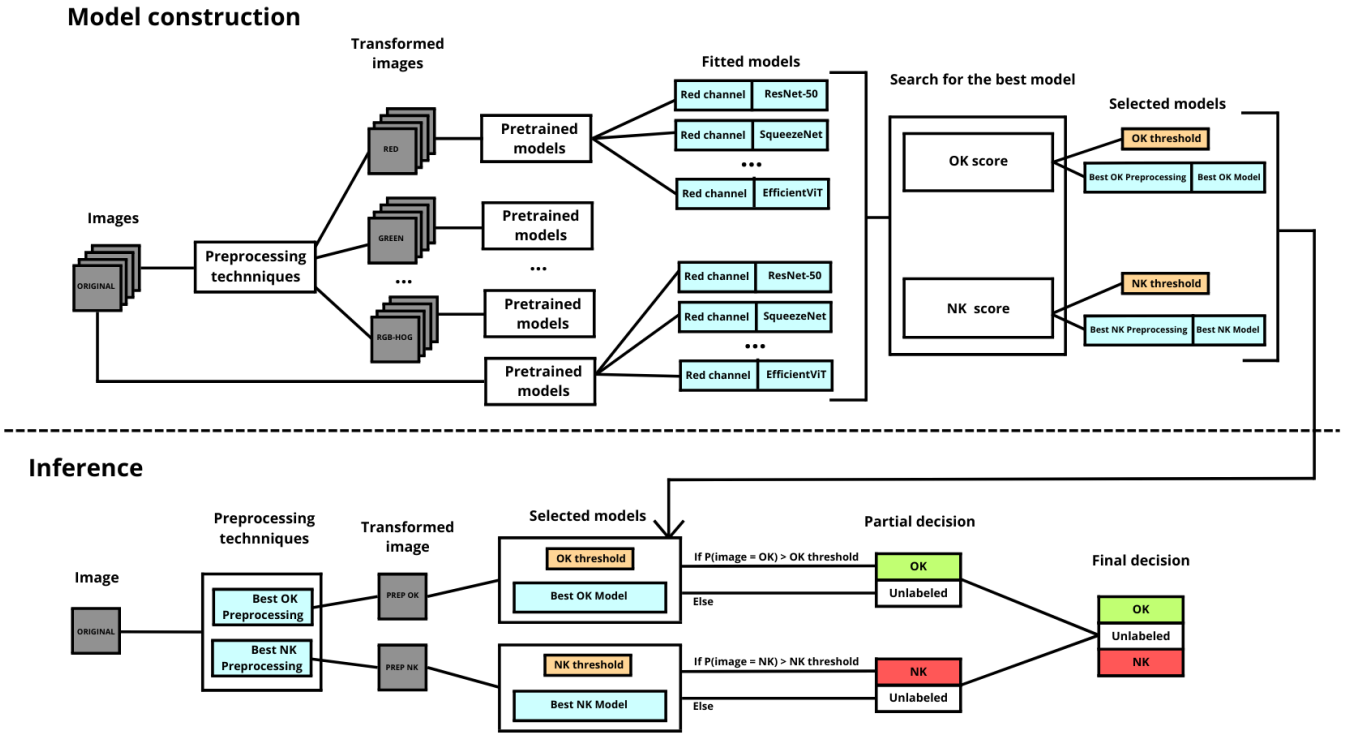


Figure 7. Proposed architecture

of the Algorithm 1. The input of the algorithm is composed of five parameters, namely N, VI, $scores$, $labels$ and $label$. Let $M_{m \times n}$ be a matrix with m rows and n columns, these five parameters are defined as follows:

- N is the number of fitted models.
- VI is the number of images on the validation set,
- $scores \in M_{N \times VI}$ represents the $S_{defective}$ scores provided by fitted models for images on the validation set.
- $labels \in M_{N \times VI}$ contains the labels predicted by fitted models for images on the validation set.
- $label \in \{'OK', 'NK'\}$ includes the search objective. $'OK'$ indicates that the best model for classifying non-defective parts must be selected and the associated OK threshold must be defined. $'NK'$ means the opposite.

The output of the algorithm is the threshold value ($threshold$) and the index of the selected model ($best\_model$).

Algorithm 1 Search for the best model

function FIND_BEST_MODELS_THRESHOLDS(N, VI, $scores$, $labels$, $label$)
  $best\_count \leftarrow 0$
  $threshold \leftarrow 0$
  $best\_model \leftarrow 0$
  if $label == 'OK'$ **then**
    $scores \leftarrow 1 - scores$
  end if
  for $i \leftarrow 1$ **to** num_models **do**
    $indices \leftarrow sort\_indices\_descend(scores[i])$
    $sorted\_scores \leftarrow scores[i][indices]$
    $sorted\_labels \leftarrow labels[i][indices]$
    $j \leftarrow 1$
    while $(sorted\_labels[i][j] == label) \& (j <= VI)$ **do**
      $j \leftarrow j + 1$
    end while
    if $j > best\_count$ **then**
      $best\_count \leftarrow j$
      $threshold \leftarrow sorted\_scores[j]$
      $best\_model \leftarrow i$
    end if
  end for
  return $threshold, best\_model$
end function

4.2. Inference

Once again, as shown in Figure 7, the initial step is to preprocess the data. This is achieved by applying the two transformations associated with the models selected during the architecture construction. This results in two images that will be processed independently.

The image obtained after applying the OK transformation is processed with the Best OK Model. Then, the architecture evaluates the score of the model indicating whether the image is OK. If the value is greater than the OK threshold, it is determined that the image is OK (non-defective). Otherwise, it remains unlabeled.

The same process is conducted with the image obtained after applying the NK transformation. In this case, the score that the image is NK is observed. If the NK threshold is exceeded, the image is classified as NK (defective). Otherwise, it remains unlabeled.

Finally, the partial decisions are integrated to make a final decision, as illustrated in Table 1.

Table 1. Final decision.

Partial decision OK	Partial decision NK	Final decision
OK	Unlabeled	OK
OK	NK	Unlabeled
Unlabeled	NK	NK
Unlabeled	Unlabeled	Unlabeled

5. RESULTS

This section begins with a brief overview of the experimental setup. Then, the results obtained with CNNs and ViTs are shown as a baseline. Finally, the results obtained with the proposed architecture are presented.

5.1. Experimental setup

The experiments were conducted on an Intel Xeon Silver 4310 CPU at 2.10 GHz with 32 GB RAM. Additionally, GPU acceleration was employed with a 48 GB NVIDIA A40. Some of the selected libraries in the execution environment were Python 3.9.12, Pytorch 1.13.1 and Fastai 2.7.13.

The dataset was balanced by randomly removing some non-defective images until a ratio of 80% - 20% is achieved. After that, this set was divided into three subsets: training, with 312 images (50%), validation, with 125 images (20%) and test, with 188 images (30%). This partition was random and stratified.

The models were evaluated using metrics derived from the confusion matrix, which in the case of binary classification includes:

- True positive (TP). The actual and predicted values are both positive.
- False negatives (FN). The actual value is positive, while the predicted is negative.
- True negative (TN). The actual and predicted values are both negative.
- False positives (FP). The actual value is negative, while the predicted is positive.

Given the significant imbalance between classes and the importance of false positives and false negatives, the selected metric for evaluating the models was the F1-score.

$$F1{-}score = 2{\cdot}\frac{Precision \cdot Recall}{Precision + Recall} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

The minority class, which represents the defective set, was selected as the positive class.

Moreover, additional metrics are employed to enhance the comprehension of the results.

$$TruePositiveRate = \frac{TP}{TP + FN}$$
$$FalsePositiveRate = \frac{FP}{FP + TN}$$
$$TrueNegativeRate = \frac{TN}{TN + FP}$$
$$FalseNegativeRate = \frac{FN}{FN + TP}$$

5.2. Baseline

The results obtained after fitting the pretrained models with the training data were used as a baseline. This can be consulted in Table 2. In particular, the ten best models are enumerated, in conjunction with their preprocessing technique. For a more comprehensive understanding, in addition to displaying the F1-score results, the elements of the confusion matrix are also included.

Table 2. Classification test results for the baseline models

Preprocessing	Model	F1	TP	FN	FP	TN
Cyan ch.	ResNet-ViT	0.457	24	9	48	107
Gamma 0.5	ConvNeXt-S	0.453	17	16	25	130
Red ch.	ResNet-50	0.442	17	16	27	128
Original	ConvNeXt-S	0.422	19	14	38	117
RGB-HOG	ResNeXt-50	0.407	12	21	14	141
Original	ViT	0.407	24	9	61	94
Blue ch.	EfficientViT	0.400	11	22	11	144
Magenta ch.	SqueezeNet	0.400	15	18	27	128
Green ch.	SqueezeNet	0.395	17	16	36	119
Magenta ch.	EfficientNet	0.395	16	17	32	123

The color plane decomposition appears to be an effective approach, as evidenced by the fact that more than half of the models presented employ this technique. Although the table includes more CNN models, the hybrid Resnet-ViT architecture is the one with the best results. However, all the results are quite poor. The considerable number of false negatives is unacceptable in a real manufacturing environment,

as they would result in defective parts that continue in the assembly line. Furthermore, a third part of the non-defective parts would be sent to melting down unnecessarily.

5.3. Architecture results

First, the models selected during the architecture construction are presented.

- In the case of OK score, the selected model was EVA-02 with the grayscale transformation, which was able to correctly classify 23 OK instances before the appearance of an NK, as can be observed in Figure 8. The threshold was determined to be 0.898.

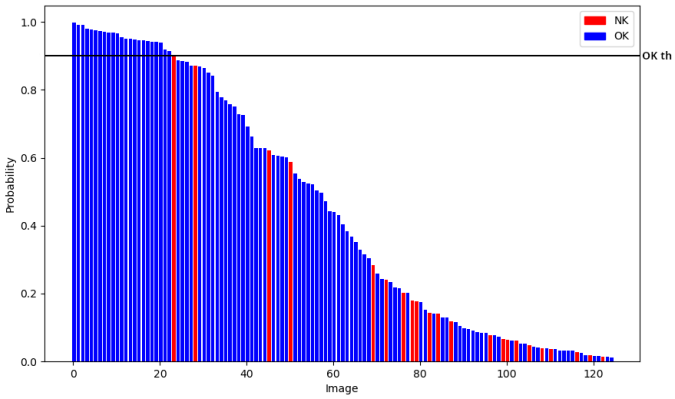


Figure 8. Best OK Model

- In the case of NK score, the best model was ConvNeXt-L with original images. This model was able to classify 4 NK instances before the appearance of the first OK, as illustrated in Figure 9. The threshold was determined to be 0.794.

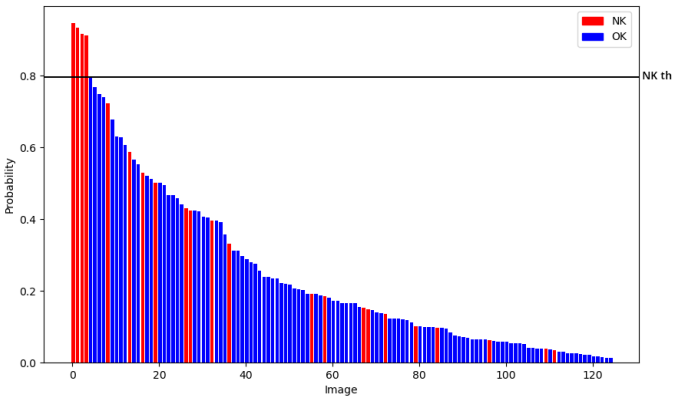


Figure 9. Best NK Model

The results obtained on the test set using the selected models and thresholds can be found in Table 3.

Table 3. Proposed architecture test results

	Predicted		
Actual	**OK**	**NK**	**Unlabeled**
OK	30	5	120
NK	1	5	27

This approach enhances the baseline results, not only in terms of the F1 score, which has increased from 0.457 to 0.625, but also in the confusion matrix values. Only one defective image is classified as OK, and only five correct parts would be sent to melting down. Furthermore, approximately 22% of the parts would be exempt from the leak test.

The True Positive Rate and True Negative Rate are 0.833 and 0.857 respectively. In contrast, the False Negative Rate has a value of 0.167 and the False Positive Rate is 0.143. Another particularly significant value is the one that tells us what percentage of parts classified as OK are really OK, which would be obtained by calculating $\frac{TN}{TN+FN} = 0.968$.

Moreover, although in this work the threshold has been set according to the validation subset results, the proposal has the sufficient flexibility to be adapted to more or less strict requirements. Consequently, if the use case requires it, higher thresholds could be employed to minimize the False Positive and False Negative rates, or more relaxed thresholds can be used to reduce the unlabeled percentage.

6. CONCLUSIONS AND FUTURE WORK

The application of classical preprocessing techniques in association with CNNs and ViTs (the baseline) has demonstrated that this approach does not provide sufficient quality guarantees. Production lines requires that all parts undergo rigorous quality control checks. However, this can result in bottlenecks and constraints on the production capacity of the line.

In this work we have proposed an architecture that combines several DL techniques, which effectively prioritizes the number of parts which can avoid these tests while ensuring strict quality requirements. The proposed architecture combines the models that maximize the number of correctly classified instances before the first classification error according to their OK score (Best OK Model) and to their NK score (Best NK Model).

The proposed architecture results in a 22% reduction in the number of parts that would not require additional tests. Consequently, it can be concluded that the proposed techniques enhance the quality control tasks in a die-casting process, resulting in an overall improvement of the productivity system. Since these thermographic images are proprietary data, it is not possible to compare the results with other publications in the literature. However, experiments have been conducted

with the main state-of-the-art techniques to demonstrate differences in performance.

Considering these results, it may be advisable to use alternative image data for quality control purposes. Although thermographic imaging of the manufactured parts is not a feasible option, tomography provides a potential solution for obtaining a more detailed representation of the parts. However, the extended capture process would represent a bottleneck, preventing the current production rate. Therefore, as the objective of the proposal is to enhance the production process capacity, it is essential to preserve the use of thermographic images of the molds.

In order to improve these results, future work will focus on the use of more advanced DL techniques and architectures such as ensemble methods. Furthermore, due to the considerable complexity of the image, along with the possibility that thermographic images may be taken from different angles and distances, it is planned to process it by regions of interest, studying each of the portions individually. Additionally, it is also intended to use thermographic images of different molds and injectors.

ACKNOWLEDGMENT

Paula Mielgo's work has been funded by UVa 2023 predoctoral contracts, co-financed by Banco Santander. This work has been funded by the Spanish Ministerio de Ciencia e Innovación under grant PID2021-126659OB-I00. We acknowledge HORSE Powertrain for granting us access to the thermographic images dataset and Javier Moral Blanco for the help and support.

REFERENCES

Albelwi, S., & Mahmood, A. (2017). A framework for designing the architectures of deep convolutional neural networks. *Entropy*, *19*(6), 242.

Alfaro-Viquez, D., Zamora-Hernandez, M.-A., Benavent-Lledo, M., Garcia-Rodriguez, J., & Azorín-López, J. (2022). Monitoring human performance through deep learning and computer vision in industry 4.0. In *International workshop on soft computing models in industrial and environmental applications* (pp. 309–318).

Cai, H., Li, J., Hu, M., Gan, C., & Han, S. (2022). Efficientvit: Multi-scale linear attention for high-resolution dense prediction. *arXiv preprint arXiv:2205.14756*.

Cumbajin, E., Rodrigues, N., Costa, P., Miragaia, R., Frazão, L., Costa, N., ... Pereira, A. (2023). A real-time automated defect detection system for ceramic pieces manufacturing process based on computer vision with deep learning. *Sensors*, *24*(1), 232.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 ieee computer society conference on computer vision and pattern recognition* (Vol. 1, p. 886-893).

Deng, L., & Liu, Y. (2018). *Deep learning in natural language processing*. Springer.

Déniz, O., Bueno, G., Salido, J., & De la Torre, F. (2011). Face recognition using histograms of oriented gradients. *Pattern recognition letters*, *32*(12), 1598-1603.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... others (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Dyrmann, M., Karstoft, H., & Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosystems engineering*, *151*, 72-80.

EL Ghadoui, M., Mouchtachi, A., & Majdoul, R. (2023). Intelligent surface roughness measurement using deep learning and computer vision: a promising approach for manufacturing quality control. *The International Journal of Advanced Manufacturing Technology*, *129*(7), 3261–3268.

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., ... Socher, R. (2021). Deep learning-enabled medical computer vision. *NPJ digital medicine*, *4*(1), 5.

Fang, Y., Sun, Q., Wang, X., Huang, T., Wang, X., & Cao, Y. (2023). Eva-02: A visual representation for neon genesis. *arXiv preprint arXiv:2303.11331*.

Gonzalez, R. C., & Woods, R. E. (2008). *Digital image processing*. Pearson Education.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (p. 770-778).

Hermann, M., Pentek, T., & Otto, B. (2016). Design principles for industrie 4.0 scenarios. In *2016 49th hawaii international conference on system sciences (hicss)* (pp. 3928–3937).

Hsu, C.-M., Hsu, C.-C., Hsu, Z.-M., Shih, F.-Y., Chang, M.-L., & Chen, T.-H. (2021). Colorectal polyp image detection and classification through grayscale images and deep learning. *Sensors*, *21*(18), 5995.

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size. *arXiv:1602.07360*.

Jack, K. (2011). *Video demystified: a handbook for the digital engineer*. Elsevier.

Lahmyed, R., El Ansari, M., & Ellahyani, A. (2019). A new thermal infrared and visible spectrum images-based pedestrian detection system. *Multimedia Tools and Applications*, *78*, 15861-15885.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recogni-

tion. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D. D., & Chen, M. (2014). Medical image classification with convolutional neural network. In *2014 13th international conference on control automation robotics & vision* (p. 844-848).

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (p. 11976-11986).

Nassif, A. B., Shahin, I., Attili, I., Azzeh, M., & Shaalan, K. (2019). Speech recognition using deep neural networks: A systematic review. *IEEE access*, *7*, 19143-19165.

Patel, S., & Jokhakar, V. N. (2016). A random forest based machine learning approach for mild steel defect diagnosis. In *2016 ieee international conference on computational intelligence and computing research (iccic)* (p. 1-8).

Peres, R. S., Barata, J., Leitao, P., & Garcia, G. (2019). Multistage quality control using machine learning in the automotive industry. *IEEE Access*, *7*, 79908-79916.

Sachin, R., Sowmya, V., Govind, D., & Soman, K. (2018). Dependency of various color and intensity planes on cnn based image classification. In *Advances in signal processing and intelligent recognition systems: Proceedings of third international symposium on signal processing and intelligent recognition systems* (p. 167-177).

Smith, A. D., Du, S., & Kurien, A. (2023). Vision transformers for anomaly detection and localisation in leather surface defect classification based on low-resolution images and a small dataset. *Applied Sciences*, *13*(15), 8716.

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (p. 6105-6114).

Usuga Cadavid, J. P., Lamouri, S., Grabot, B., Pellerin, R., & Fortin, A. (2020). Machine learning applied in production planning and control: a state-of-the-art in the era of industry 4.0. *Journal of Intelligent Manufacturing*, *31*, 1531-1558.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Villalba-Diez, J., Schmidt, D., Gevers, R., Ordieres-Meré, J., Buchwitz, M., & Wellbrock, W. (2019). Deep learning for industrial computer vision quality control in the printing industry 4.0. *Sensors*, *19*(18), 3987.

Wang, K., Zhang, J., Ni, H., & Ren, F. (2021). Thermal defect detection for substation equipment based on infrared image using convolutional neural network. *Electronics*, *10*(16), 1986.

Weimer, D., Scholz-Reiter, B., & Shpitalni, M. (2016). Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. *CIRP annals*, *65*(1), 417-420.

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (p. 1492-1500).

Xie, Y., & Richmond, D. (2018). Pre-training on grayscale imagenet improves medical image classification. In *Proceedings of the european conference on computer vision workshops.*

Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent manufacturing in the context of industry 4.0: a review. *Engineering*, *3*(5), 616–630.

BIOGRAPHIES

Paula Mielgo is a Ph.D. Student in the Computer Science Department at the University of Valladolid, Spain. Mielgo completed her two Bachelor's Degrees in Mathematics and Computer Science Engineering in 2022. Furthermore, she obtained a Master's Degree in Computer Science Engineering in 2023. During her studies, she received several academic awards, including the award for the best academic results in both of her CS programs and a regional award for her final degree thesis. Later that year, she started her Ph.D. thesis in Computer Science.

Anibal Bregon received his B.Sc., M.Sc. and Ph.D. degrees in Computer Science from the University of Valladolid (Spain) in 2005, 2007 and 2010, respectively. He joined the Department of Computer Science at the University of Valladolid in 2011, where he is Associate Professor since February 2018. He has carried out both basic and applied research in the areas of fault diagnosis and prognosis for aerospace and industrial systems, and has co-authored more than 85 journal and conference papers. He is currently leading a national funded project on advanced learning for smart manufacturing and several technology transfer contracts on Deep Learning, and has also participated as researcher on several funded projects, networks and contracts on fault diagnosis and prognosis topics, on Big Data analytics and on Deep Learning. He has been guest researcher with the Intelligent Systems Division at NASA Ames Research Center and the Institute for Software Integrated Systems at Vanderbilt University, among others. His current research interests include model-based reasoning for diagnosis and prognosis, health-management, Big Data, Industry 4.0 and Deep Learning. Among various other professional activities, he has held different chair positions at the PHM and PHME conferences, has been co-administrator of several courses and summer schools on diagnosis, prognosis, and artificial intelligence, and has been the Local Chair of the 2016 European Conference of the Prognostics and Health Management Society.

Carlos J. Alonso-González received the B.S. and Ph.D. degrees in physics from the University of Valladolid, in 1985 and 1990, respectively. After a brief stay in private companies and the Public University of Navarra, he joined the University of Valladolid, where he is currently an Associate Professor with the Department of Computer Science. He is also the Head of the Intelligent Systems Group, Department of Computer Science. He has worked on different national and European-funded projects related to the monitoring and diagnosis of continuous industrial environments and dynamic hybrid systems. He is also involved in projects related to the application of deep learning and casual and explainable AI to Industry 4.0, both for manufacturing and continuous processes. His current research interests include knowledge, model, and data-based systems for health management of dynamic systems, model and data-based diagnosis and prognosis of complex physical systems, and machine learning. He has been a member of the Board of Trustees of the Sugar Technology Center, being responsible for projects related to the application of artificial intelligence to online production supervision.

Daniel López is a graduate in Industrial and Automatic Electronics with more than 8 years of experience in the automotive sector. He graduated from the School of Industrial Engineering at the University of Valladolid in 2016 and subsequently completed his studies with a Masters in Industrial Electronics and Automation at the UVa and another in Project Management at the UEMC. He continues to work on improving his knowledge of the world of data.

Miguel A. Martínez-Prieto is an Associate Professor and Researcher in Computer Science with the Department of Computer Science, University of Valladolid, Spain. He received his B.Sc., M.Sc., and Ph.D. degrees in Computer Science from the University of Valladolid, Valladolid, Spain, in 2005, 2007, and 2010, respectively. He held a Postdoctoral position with the Department of Computer Science, University of Chile, from 2010 to 2012. His research has been in the area of data management, mainly in data compression and indexing of semantic, text and biological data, and the resolution of specific queries in each of these scenarios. His current research interests focus on data science, with applications in air traffic management and Industry 4.0. He has co-authored more than 90 peer-reviewed papers on these topics and has been involved in several European and national funded projects, as well as several transfer contracts with companies and government institutions.

Belarmino Pulido received his Licenciate degree, M.Sc. degree, and Ph.D. degree in Computer Science from the University of Valladolid, Valladolid, Spain, in 1992, 1995, and 2001 respectively. In 1994 he joined the Department of Computer Science at the University of Valladolid, where he is Associate Professor since 2002. His main research interests are in Systems Health Management using different techniques such as model-based reasoning, knowledge-based systems, and data-driven models (using both classical and advanced machine-learning). He has worked in different national and european funded projects related to Supervision and Diagnosis. He is the coordinator of the Spanish Network on Supervision and Diagnosis of Complex Systems since 2005.