Automatic detection of defective crankshafts by image analysis and supervised classification

Beatriz Remeseiro · Javier Tarrío-
Saavedra · Mario Francisco-Fernández ·
Manuel G. Penedo · Salvador Naya ·
Ricardo Cao

Abstract A crankshaft is a mechanical component of an engine that performs a conversion of an alternative movement of a piston in a rotational motion of a shaft. It is a critical part and one of the most expensive of an engine. Defects in crankshafts may imply serious failures and, consequently, possible injuries and high costs. Therefore, the manufacture quality is of primordial importance for security and economic reasons. Nowadays, the quality control of crankshafts manufactured by forging in the automotive industry consists, among others, in inspecting them at the final process, using a magnetic particle procedure. This slow and highly stressful technique depends on operators and consumes many human resources, time, and space. This paper presents a methodology to automatically detect defective crankshafts. The proposed procedure is based on digital image analysis techniques, to extract a set of representative features from crankshaft images. Statistical techniques for supervised classification are used to classify the images into defective or not. The experimental results demonstrated the good performance of the proposed method with a classification accuracy over 99%, a 10% higher than the one obtained by manual inspection. Therefore, working time and personnel required for this task can be reduced when using this automated procedure.

Beatriz Remeseiro · Manuel G. Penedo
Research Group VARPA, CITIC, Departamento de Computación, Universidade da Coruña.
Campus de Elviña s/n, 15071 A Coruña, Spain.
Present address (Beatriz Remeseiro): Department of Computer Science, Universidad de Oviedo. Campus de Gijón s/n, 33203 Gijón, Spain.
E-mail: bremeseiro@uniovi.es, mgpenedo@udc.es

Javier Tarrío-Saavedra · Mario Francisco-Fernández · Salvador Naya · Ricardo Cao
Research Group MODES, CITIC and ITMATI, Departamento de Matemáticas, Universidade da Coruña. Campus de Elviña s/n, 15071 A Coruña, Spain.
E-mail: jtarrio@udc.es, mariofr@udc.es, salva@udc.es, rcao@udc.es

1 Introduction

Quality assurance of crankshafts used in the automotive industry is the principal aim of this work. A crankshaft (see Figure 1) is a shaft composed of throws with webs and crankpins to transform the rectilinear reciprocating motion of pistons within the cylinders into rotary movement of crankpins through the connecting rods, which is also transmitted to the wheels and other elements.
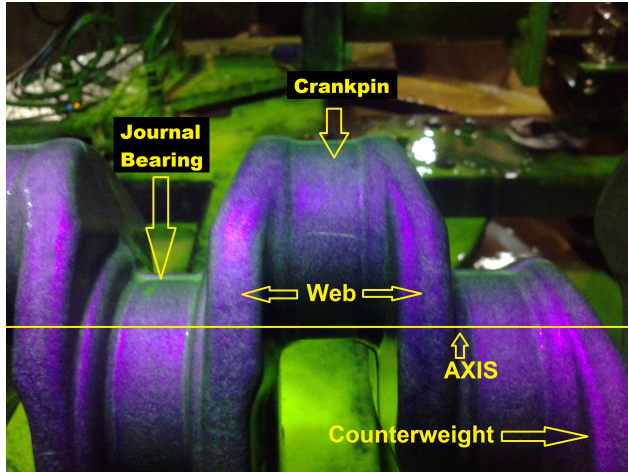
Fig. 1 Parts of a crankshaft.

It is important to emphasize that during operation, the crankshaft is subjected to violent efforts, caused by explosions and reactions due to the acceleration of bodies moved by the reciprocating motion (pistons and connecting rods). Thus, this type of engine part requires special mechanical properties to provide a correct performance and prevent breakdowns and accidents. The crankshaft is the most expensive component of the engine and requires a comprehensive and reliable quality control procedure.

Figure 2 illustrates the procedure to obtain a forging crankshaft. As can be seen, this procedure is divided in five steps: reception of raw material (microalloyed steel), ingots preprocessing, forge procedure, post-processing of forging cranks, and quality assurance, including magnetic particles inspection of defects and dimensional control [22,11,19,3]. The receiving material are ingots of microalloyed steel of high impact, fatigue and mechanical resistance, codified by the Aircraft Material Specification [22,3]. In the preprocessing step, the ingots are cut according to standardized dimensions, taking into account that the posterior forging process reduces their initial size.

Crankshafts can be also produced by casting and machining (depending of the engine requirements), but this work is only focused on the identification of crank defects produced in forging factories by closed-die forging [10]. Broadly speaking, forging consists in heating a steel ingot to about 1200ºC, where
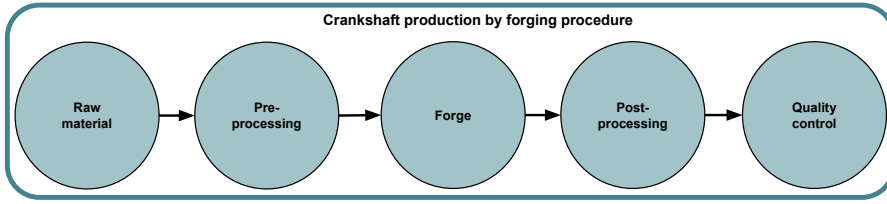
Fig. 2 Steps to obtain a forging crankshaft.

the steel becomes to be malleable without melting, while it is compressed by pressing or hammering into a matrix that reproduces the shape of the final crankshaft. This pressing procedure increases the mechanical resistance of crankshafts, and the alignment and strengthening of the steel grain. Once the crankshaft is forged, the burr formed between the two matrix pieces is cut in the post-processing step by machining procedures. Moreover, the stress produced by the machining is relieved by annealing between 200ºC and 400ºC.

Once the crankshaft is forged and the burr is cut, the application of an exhaustive quality control analysis is absolutely necessary to detect surface defects that could result in the crank failing during operation [22,11,3]. Crankshaft failure is not very common but, when it occurs, catastrophic accidents and significant economic losses are caused [19], apart from a worsening of the brand image. The crankshaft is not only a critical piece of automotive engines, but also the piece of the engine with the largest cost [1] and, if damaged, it is very difficult to access and repair.

Nowadays, quality control of crankshafts is manually performed by trained and experienced operators, using a magnetic particle inspection procedure [19, 26]. This inspection is a non-destructive test applied to steel pieces and structures that consists in magnetizing the inspected items, thus the magnetic field lines point out the flaws or defects. These lines are denser in the crack surroundings and so, if the piece is sprayed with small magnetic particles, they will be concentrated in the neighborhood of the flaw allowing to observe it by the naked eye. Summarizing, crankshafts go through a manual inspection procedure (see Figure 3) in which the ferromagnetic pieces are sequentially magnetized, sprayed with magnetic particles, inspected and demagnetized. Notice that this manual inspection procedure is very time consuming (workers spend about 45 seconds with each single piece), and dependent on the operator. This is the manual procedure (using the commercial version MAGNAFLUX^{TM}) employed by CIE Galfor S.A.[1], the company that has motivated this research to automate the inspection at the end of the production line.

In this context, several artificial vision procedures have been proposed to identify and/or characterize defects in forged crankshafts: Iborra et al. [11] proposed an automatic visual identification procedure based on image segmentation and syntactic-neural network machine learning; a classifier based on neural networks to evaluate the fatigue failure risk was applied in [1] to a

---

[1] http://www.cieautomotive.com/-/cie-galfor
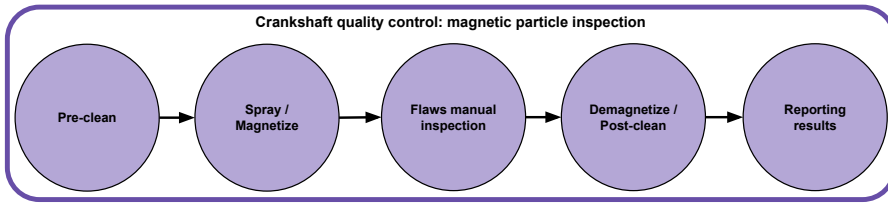
Fig. 3 Steps of the most common, manually crankshaft quality assessment procedure based on the magnetic particles inspection.

large crankshaft for a diesel engine for power plants; and a scattering technique to measure and check the surface quality of cranks was provided in [8]. However, there is still a lack of automated procedures applied to the special case of identification of defects in crankshafts, with all the steps comprehensively described to provide a useful tool to practitioners and companies requiring an automatic solution.

In this paper, a way to automate the quality control in the crankshaft manufacturing process, right at the end of the production line, is described. Applying the proposed methodology, surface defects in crankshafts can be identified and, thus, those units that are not in perfect conditions can be automatically discarded. The main defects occurring in forged crankshafts (and that can be detected by the new approach) are under-filling, overlap, foreign bodies and scale pits [22,3]. Note that about 85% of all crank defects come from these types of problems, as suggested in [3].

In the next sections, the proposed classification procedure based on color and texture analysis is described, applied and discussed to solve the industrial problem of defective crankshaft identification. Notice that color and texture properties have been successfully used in other industrial applications [4,16].

2 Material and methods

This section describes the procedure for data collection, and the images used for validation. Additionally, it includes all the methods used in the proposed methodology to automatically detect defective crankshafts.

2.1 Data collection

A total of 88 representative crankshaft images were used to validate the proposed methodology. In particular, 37 of them correspond to crankshafts that were manually rejected, and the other 51 images correspond to correct crankshafts. All the images have been stored at a spatial resolution of 2048×2048 pixels in the RGB color space.

Regarding the acquisition procedure, all the images were obtained using an industrial camera provided by Infaimon S.L. company[2]. The camera was placed in a fixed position to take images in an ad-hoc pilot unit located in the crankshaft manufacturing company, CIE Galfor S.A. This pilot unit reproduces the common manual process of magnetic particle inspection (see Figure 4): the crankshafts arrive to a special dark place where they are impregnated (one by one) in a liquid solution with magnetic particles (MAGNAFLUX), then they are magnetized (applying a magnetic field intensity), illuminated, and finally the images are taken.
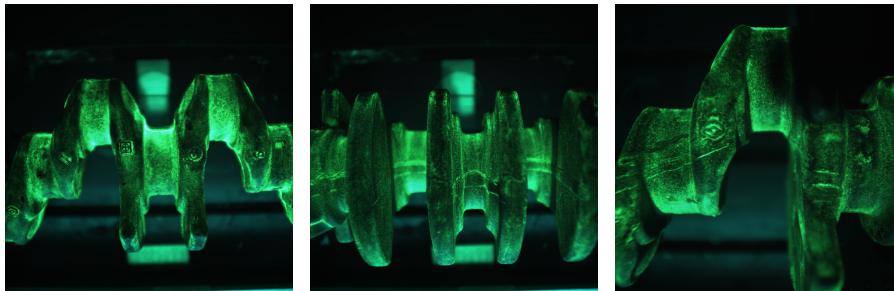
Fig. 4 Representative images of magnetized and illuminated crankshafts when they are manually inspected.

Once the crankshaft images were obtained, the next step was to mark those areas with defects in order to find out whether our approach is able to detect them. It was a manual process based on the observation of the images, in which the defective areas were marked by means of GIMP[3], an image manipulation program. It is important to stress that the defects were identified with the supervision of trained personnel of the crankshaft factory, and it took 30 seconds each on average. Figure 5 shows a representative crankshaft image and, as can be observed, defective areas are characterized as thin lines of a more intense green color than the rest of the crankshaft. A mask for each annotated image was created to delimit the defective areas and make easier the further analysis.

In a defective crankshaft image, there are both correct and defective areas. For this reason, a local level analysis of the images is needed in order to discriminate defective areas from those which are not. In this way, not only it is possible to determine if a crankshaft is defective or not, but also to estimate the location of the defects on the image.

Local image patches are defined as small groups of nearby pixels, and they are used in local analysis. Thus, once the 88 crankshaft images were annotated and the defective areas were delimited, the next step was to extract patches from them. Taking into account the resolution of the images and the

---

[2] http://www.infaimon.com/es/

[3] http://www.gimp.org
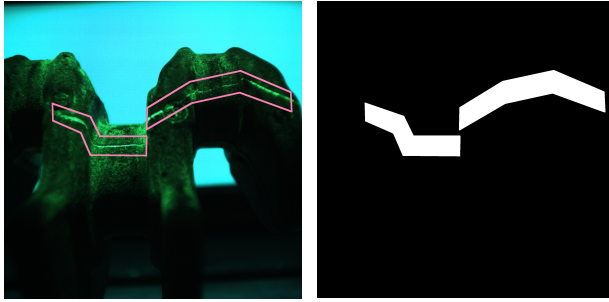
Fig. 5 Representative crankshaft image with the defective areas manually delimited, and the mask created from it to facilitate the further analysis.

average size of the defective areas, we have considered to use square patches of 128 pixels. And, instead of extracting all the possible patches from the images, we have used a margin of separation of four pixels between patches. This process was automatically performed by using the masks previously generated (see Figure 5). In this manner, over 100,000 patches per image were obtained. As expected, most of them correspond to non-defective areas. In order to avoid the common problems which occur with imbalanced datasets [14], we have created a well-balanced dataset to be used in the experimentation that is composed of 16,778 patches, half of which correspond to defective patches. Notice that the non-defective patches included in the final dataset were randomly selected among all the non-defective patches extracted from the images. Figure 6 shows a representative image of the database with two patches included in the dataset.
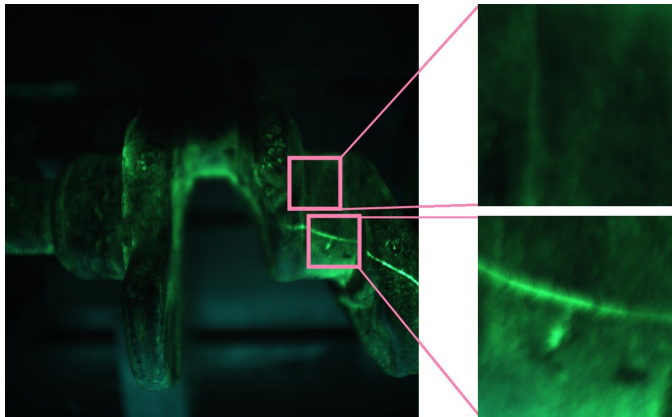
Fig. 6 Representative image of the database with two patches, from top to bottom: non-defective patch, and defective patch.

2.2 Analysis and interpretation

In order to characterize defective areas in crankshaft images, and so to automatize the quality control process, a four-step methodology was applied as illustrated in Figure 7. Apart from employing specific image analysis software, it is important to stress that the free statistical software R [23] has been used to perform the corresponding statistical analysis. Specifically, the following R packages were employed in this research: `pls` (for dimensional reduction), `MASS`, `e1071` (for classification purposes), and `cvtools` (for validation tasks).
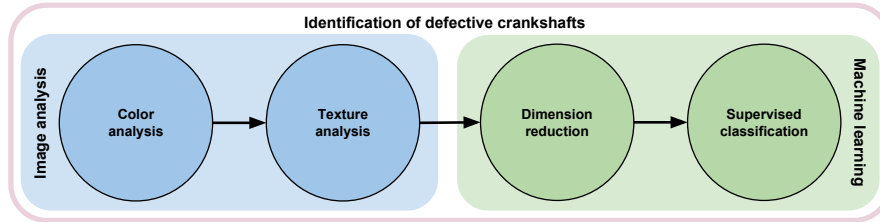
Fig. 7 Steps of the proposed research methodology to identify defective crankshafts.

In the following sections, the proposed procedure is described in depth. Color and texture information were initially extracted from an input image in order to create a feature vector, whose dimension was subsequently reduced and, finally, the image is classified into one of the two categories considered (defective and non-defective areas).

2.3 Color analysis

Color seems to be one of the discriminant features of defective areas, since they can be observed in green color. For this reason, crankshaft images were analyzed by means of the Lab color space, a chromatic model based on the color perception of the human brain. Grayscale images were also considered in order to verify the adequacy of using color information. Therefore, this stage consists in transforming the input images in RGB to grayscale images and to the Lab color space, according to the procedures subsequently defined.

2.3.1 Grayscale images

A grayscale image is one in which the only color is gray, represented by different levels from black to white. Therefore, only a single intensity value is specified for pixel (see Figure 8). In order to generate a grayscale image, the three channels of the RGB image (R, G and B) have to be converted into only one gray channel (Gr), according to the following expression [2]:

$$Gr = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B$$
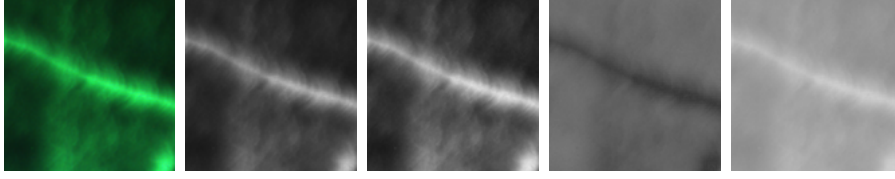
Fig. 8 One representative defective patch in different color spaces, from left to right: original image in RGB, grayscale image, and the three components of Lab (L, a and b).

2.3.2 The Lab Color Space

The CIE 1976 L*a*b [20] is a chromatic color space that describes all the colors that the human eye can perceive. It is a 3D model whose coordinates are: the luminance of the color L, its position between magenta and green a, and its position between yellow and blue b (see Figure 8). This color space is perceptually uniform, that is, a change of a certain amount in a color value produces a change of the same visual relevance. The significance of this property lies in the fact that operators' visual perception is trying to be imitated.

Using the Lab color space entails converting the three components of the image in RGB to the three channels of Lab. This transformation was done by using the CIE XYZ color space and its three components X, Y and Z [2]:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124563 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

and then,

$$X = X/0.950456$$
$$Z = Z/1.088754$$

Next, the Lab channels are calculated according to:

$$L = \begin{cases} 116 \cdot Y^{1/3} - 16 & \text{if } Y > 0.008856 \\ 903.3 \cdot Y & \text{if } Y \leq 0.008856 \end{cases}$$

$$a = 500(f(X) - f(Y)) + 128$$
$$b = 200(f(Y) - f(Z)) + 128$$

where:

$$f(t) = \begin{cases} t^{1/3} & \text{if } t > 0.008856 \\ 7.787t + 16/116 & \text{if } t \leq 0.008856 \end{cases}$$

2.4 Texture analysis

Texture is used to characterize the defective areas, in addition to the color information, since the distribution of the intensity values in the images is different depending on if there is a defect or not. Several techniques can be found in the literature to extract texture information from images and, in this research, two of the most common ones were selected: the analysis of histograms, and the discrete wavelet transform. Thus, this step consists in extracting the texture information from each color component: one channel (Gr) when using grayscale images, and three channels (L, a and b) when using the Lab color space. We next explain the texture extraction procedure.

2.4.1 Uniform histograms

An image histogram represents the distribution of the intensity values of its pixels. Instead of analyzing the traditional fixed bin histograms, uniform histograms were considered. They allow a better representation of the intensity distribution in the problem at hand. Uniform histograms are histograms with equiprobable bins, and the process to obtain them is described as follows: given all the input images, the limits of the histogram are defined so that each bin contains $\frac{N}{N_{bins}}$ intensity values, where N is the number of pixels in the image, and N_{bins} the number of histogram bins.

For this problem, and after some preliminary experiments, 16-bin histograms were considered and, consequently, the quantitative vector of an input image is composed of 16 features (frequency distribution) for each color component. That is, the vector computed for a single patch is composed of 16 features when using grayscale images (see Table 1), whilst it is composed of $16 \times 3 = 48$ features when using the Lab color space.

f1: 0.0561	f2: 0.0612	f3: 0.0466	f4: 0.0793	f5: 0.0518	f6: 0.0729
f7: 0.0676	f8: 0.0605	f9: 0.0504	f10: 0.0657	f11: 0.0687	f12: 0.0670
f13: 0.0582	f14: 0.0657	f15: 0.0640	f16: 0.0634		

Table 1 Sample vector obtained when using uniform histograms: numerical values of the 16 features (f) computed from the grayscale image illustrated in Figure 8.

2.4.2 Discrete wavelet transform

The discrete wavelet transform (DWT) [18] generates a set of wavelets by scaling and translating a *mother wavelet*, which is a function defined in the spatial and frequency domain, represented in 2D as:

$$\phi^{a,b}(x,y) = \frac{1}{\sqrt{a_x a_y}} \phi\left(\frac{x-b_x}{a_x}, \frac{y-b_y}{a_y}\right),$$

where x and y are the horizontal and vertical pixel coordinates, respectively, $a = (a_x, a_y)$ governs the scale, and $b = (b_x, b_y)$ the translation of the function. The values of a and b control the band-pass of the filter in order to generate high-pass or low-pass filters.

The wavelet decomposition of an image consists in applying those wavelets in the two dimensions, i.e. horizontally and vertically. Taking into account the high-pass (H) and low-pass (L) filter for every dimension, four sub-images are generated (LL, HL, LH and HH). Every one of these four images is then sub-sampled by a factor of two, i.e. only every second pixel is retained.

Some statistical measures were used in order to generate the quantitative vector from an input image: mean, absolute average deviation and energy; which are respectively defined as:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} p(i),$$

$$AAD = \frac{1}{N} \sum_{i=1}^{N} |p(i) - \mu|$$

and

$$e = \frac{1}{N^2} \sum_{i=1}^{N} p(i)^2,$$

where $p(i)$ is the i-th entry in the image, sorted row-wise, and N represents its number of pixels.

Finally, the feature vector is constructed from the μ and the AAD of the input and LL images, and the e of the LH, HL and HH images. Thus, the quantitative vector of an input image is composed of 7 features per color component. Consequently, the vector computed for a single patch is composed of 7 features when using grayscale images (see Table 2), whilst it is composed of $7 \times 3 = 21$ features when using Lab. Note that a representative type of basis for wavelets, known as Daubechies [6], was considered as the mother wavelet.

$f1$: 64.1641	$f2$: 23.9111	$f3$: 128,328	$f4$: 47.7307	$f5$: 0.0013	$f6$: 0.0043
$f7$: 0.0003					

Table 2 Sample vector obtained when using the discrete wavelet transform: numerical values of the 7 features (f) computed from the grayscale image illustrated in Figure 8.

2.5 Dimensionality reduction

The complexity of any classification process depends on the number of input attributes, apart from the difficulty of the corresponding classifier. This

determines both memory and computing time, and also the necessary number of samples to train the classifier. In this context, different dimensionality reduction techniques can be applied to reduce the computational (memory and time) requirements without compromising the classification performance. Two of the most important techniques for dimensionality reduction, principal component analysis (PCA) and partial least squares (PLS), have been here considered. The application of these techniques is useful in a wide range of domains such as classification, regression or multivariate quality control [28].

Given the number of observations, or what is the same, crankshaft patches (cases) to be classified per image, there would be no problem in applying classification techniques directly to the set of features extracted from them. However, the application of PCA or PLS results in a new data matrix, projecting the primal one, with a number of desirable properties for the subsequent application of classification methods:

- The new features are independent, orthogonal to each other, very important for some classification methods to ensure their optimality.
- The features are sorted from the highest to the lowest explanation of the total variability of the data, i.e. from the most to the least important. This fact allows us to carry out the classification from a few relevant characteristics, thus not losing efficiency and saving computational requirements.

2.5.1 Principal component analysis

Principal component analysis (PCA) [13] is a feature transformation technique widely used for dimensionality reduction. It is an unsupervised method that aims at maximizing the variance of the data projection.

In order to calculate the new PCA components, the singular value decomposition algorithm [29] has been used. It allows us to decompose the $n \times p$ data matrix $\mathbf{X}$, once centered and scaled, in three new matrices: $\mathbf{X} = \mathbf{UDV}'$, where $\mathbf{U}$ is an orthonormal matrix ($n \times a$), $\mathbf{D}$ is a diagonal matrix ($a \times a$) which contains the singular values, and $\mathbf{V}$ is an orthonormal matrix ($p \times a$). The product $\mathbf{UD}$ results in the projection matrix with a new coordinate axis, where the first axis indicates the maximum variability of the data (i.e. the most informative component), the second axis is perpendicular to the first one and indicates the second largest direction in terms of variance, and so forth.

2.5.2 Partial least squares

Partial least squares (PLS) [30] is a statistical method used for dimensionality reduction which has some relation to principal component regression. However, instead of finding hyperplanes of minimum variance between the response and the independent variables, it finds a linear regression model by projecting the predicted variables and the observable ones into a new space. This linear regression is fitted using an initial complete sample, called the training sample. The PLS method has the advantage of taking into account the relationship

between the response variable (class) and the feature vector of each test sample to form the new matrix of scores.

To apply the PLS procedure, the so-called kernel algorithm [17] has been used, which is faster and more stable than other alternatives. Each new test sample is first centered and scaled with respect to the mean and standard deviation vector of the training sample, and then the new scores of the test sample are calculated by multiplying the original vectors by the PLS projection matrix obtained from the training sample. Next, the classification model is created by using the PLS scores of the training sample, and then the class of the test samples can be predicted.

2.6 Supervised classification

Supervised machine learning is one of the tasks most frequently carried out by the so-called intelligent systems [21]. Its application is absolutely necessary to develop an automated crankshaft anomaly detection method. In this case, the goal of the supervised learning application is to construct a classifier than can correctly predict if a new crankshaft contains defective areas or not, given training samples obtained from old crankshafts.

A large number of techniques have been developed for supervised learning, based on statistics and artificial intelligence [15]. Among them, traditional algorithms have been selected due to their fast computing time, simplicity and optimality under certain conditions. More innovative and flexible techniques have been also considered, although they require more computational resources. Next sections describe the methods considered in this research.

2.6.1 Multivariate statistical methods

Two of the most popular supervised classification techniques are linear discriminant analysis (LDA) [7] and Naive Bayes (NBC) [12]. They are here used due to their fast computing time and good performance when classes are linearly separable. In addition, LDA is applied taking into account its optimality when multivariate Gaussian assumption of feature vector is met. NBC is based on Bayes' theorem, particularly suitable when the dimension of the feature vectors is high compared with the number of observations. It requires small training datasets to estimate the classification parameters, assuming that each feature distribution can be independently estimated apart from the others. Even if predictors are not independent, it has performed well in many cases [9,25], thus its application is advisable in the present case. Another method widely used when the LDA assumptions are not fulfilled is the logistic regression model (LR). It is a generalized linear regression model (GLM) [24], in which the response variable is dichotomous. Therefore, LR application is also recommended in the present case.

2.6.2 Machine learning algorithms

The performance of the previous multivariate classification methods are compared with machine learning approaches. The methods of this type are very flexible and powerful, and they have provided very good solutions to complex classification problems where the classes are not linearly separable, as it could be the case of our problem. One of the most popular and useful machine learning techniques is the so-called support vector machine (SVM) [27], an advanced classifier that performs supervised classification by constructing hyperplanes in a multidimensional space, and so they separate samples which belong to different classes. Vapnik [27] found that, if there is a wide margin between regions with distributions from different classes, the mathematical model obtained by an optimal hyperplane will possess excellent predictive characteristics, even if the dimension of the feature vector is very high.

The most important task in the SVM classification process is the dimensional elevation, using a kernel function which assigns points of the input space into a higher dimensional space by means of a non-linear transformation. In these high-dimensional spaces, the points which belong to the initial sample, not linearly separable, become linearly separable points with a wide margin in the new feature space. For the construction of an optimal hyperplane, SVM uses an iterative training algorithm with the main goal of minimizing an error function. Some of the most common types of kernels are linear, Gaussian, and sigmoid. The SVM algorithm here considered uses a C parameter (in this case, set to 1) that controls the trade-off between training errors and rigid margins, creating a soft margin which permits some misclassification.

3 Results and discussion

The experimentation was carried out with a dataset composed of 16,778 patches obtained from 88 crankshaft images, as described in Section 2.1. Therefore, all the performance results reported in this section refer to this dataset.

The validation procedure of the defect identification in forging crankshafts is summarized as follows:

1. Apply the two color models and the two texture analysis methods to the dataset composed of 16,778 image patches. As a result, four feature matrices are available. Their sizes are $16,778 \times 16$, $16,778 \times 7$, $16,778 \times 48$, and $16,778 \times 21$, when the following methods are respectively applied (see Figure 9): grayscale and histograms, grayscale and DWT, Lab and histograms, and Lab and DWT.
2. Apply the two dimensionality reduction methods (PCA and PLS) to the four feature matrices.
3. Train the supervised classifiers (see Figure 10) using different configurations. Note that a 10-fold cross-validation [29] was used, so the average error across all 10 trials was computed.

4. Evaluate the effectiveness of the different methods in terms of the percentage of samples correctly classified.
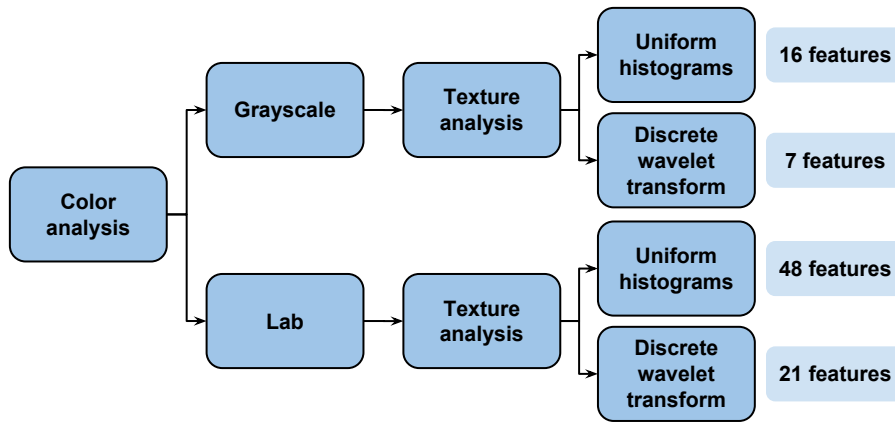
Fig. 9 Feature sets obtained with the proposed color and texture analysis techniques.
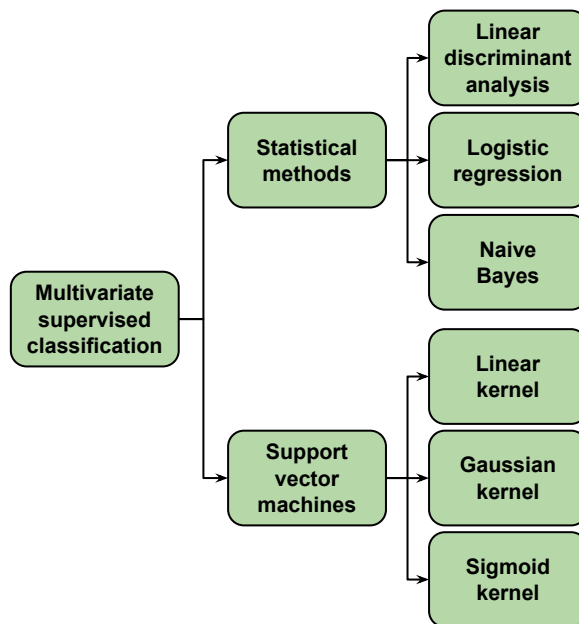
Fig. 10 Multivariate classification methods applied to the feature sets (Figure 9).

One of the goals of the proposed methodology is to find the image analysis approach and the classification method that produce an optimal trade-off be-

tween performance-classification and low demand of computational resources. Since on-line classification is aimed, the time consumption of the computational method is an important aspect. Four different feature sets are obtained from the different possible combinations of color an texture analysis techniques (see Figure 9), each one defined by a different number of features. Then, PLS and PCA components are obtained from these feature sets. A variable number of PLS and PCA components were used to implement the classification methods, depending on the original feature set (Figure 9): between 3 and 16 (from grayscale and histogram set), 3 and 7 (from grayscale and DWT set), 3 and 48 (from Lab and histogram set), and 3 and 21 (from Lab and DWT set). Overall, if we take into account the different dimensions of PLS and PCA projection matrices, more than 160 sets have been tested in order to obtain an optimal combination between good classification performance and short computing time. In the following sections, the results of the application of the different classifiers are summarized, presenting the misclassification proportions obtained by the application of LDA, LR, NBC, and SVM (see Figure 10) to the feature sets obtained from the PLS and PCA projections. It is important to note that traditional classification methods, although less flexible and powerful than SVM, are rather more efficient in terms of runtime and computing resources demand, and they present less risk of overfitting, thus justifying their application to obtain reliable results.

Table 3 shows the best results (i.e. the lowest misclassification proportions) obtained by each classification method applied to the PLS/PCA sets. These PCA or PLS components were obtained projecting the features extracted from image patches by different combinations of color and texture analysis techniques. Considering the high number of the alternatives, this summary intends to highlight the most effective techniques in terms of misclassification errors. Additional results are shown in Figures 11-14, and subsequently discussed.

Projection method	Classification method	Histograms		DWT	
		Grayscale	Lab	Grayscale	Lab
PCA	LDA	0.252(14)	—	0.194(6)	**0.088(16)**
	LR	0.251(14)	0.35(48)	0.167(6)	**0.006(16)**
	NBC	0.387(14)	0.38(13)	0.240(6)	0.240(16)
	SVM^L	0.237(14)	0.334(48)	0.174(6)	**0.015(16)**
	SVM^G	**0.010(14)**	**0.050(37)**	**0.081(6)**	**0.009(19)**
	SVM^S	0.591(14)	—	0.328(7)	0.190(19)
PLS	LDA	0.246(6)	0.410(39)	0.183(5)	**0.072(8)**
	LR	0.244(6)	0.350(48)	0.162(5)	**0.00006(16)**
	NBC	0.273(7)	0.410(17)	0.233(5)	0.136(14)
	SVM^L	0.230(6)	0.250(15)	0.173(5)	**0.0017(17)**
	SVM^G	**0.008(14)**	**0.040(37)**	**0.081(6)**	**0.0005(14)**
	SVM^S	0.503(3)	—	0.332(7)	0.145(16)

Table 3 Lowest mean misclassification errors and number of PLS/PCA components (in brackets), obtained by a 10-fold cross-validation. SVM methods were applied using different kernels, labeled as: ^{L}linear kernel, ^{G}Gaussian kernel, and ^{S}sigmoidal kernel.

As a final remark, we would like to highlight that we have also evaluated the proposed methodology using only the green (G) channel of the input images, since defective areas can be observed in green color. In this sense, the same texture analysis methods (uniform histograms and DWT), PCA and PLS reduction dimension procedures and classification algorithms were applied in this case. The obtained results are presented in Table 4 and, as can be observed, they are relatively competitive. For example, in the case of DWT and using only the G channel, better results are obtained compared to those achieved with a grayscale color analysis. However, in this scenario, the G channel provides worse misclassification errors than those obtained when using the Lab color space. In the case of uniform histograms, the proportions of misclassification errors obtained analyzing only the G channel tend to be generally higher than those obtained for the grayscale and Lab approaches. Note that, when analyzing the G channel, the SVM with sigmoid kernel has not been applied taking into account the poor results obtained with this method in the previous experiments, in terms of classification performance and computing runtime. Therefore, in view of this experimentation, we conclude that in this case it is worthy to use the information provided by all the color components and not only the G channel.

| Projection method	Classification method	Green channel	
		Histograms	DWT
PCA	LDA	—	0.158(6)
	LR	0.365(16)	0.136(6)
	NBC	0.429(9)	0.173(6)
	SVM^L	0.398(16)	0.138(6)
	SVM^G	**0.109(16)**	**0.069(16)**
PLS	LDA	0.368(16)	0.152(5)
	LR	0.365(16)	0.136(6)
	NBC	—	—
	SVM^L	0.376(16)	0.138(6)
	SVM^G	**0.114(15)**	**0.069(6)**

Table 4 Lowest mean misclassification errors and number of PLS/PCA components (in brackets), obtained by a 10-fold cross-validation procedure. SVM methods were applied using different kernels, labeled as: ^{L}linear kernel and ^{G}Gaussian kernel.

3.1 Classification performance using grayscale and uniform histograms

Figure 11 shows the boxplots of misclassification proportions obtained by performing a 10-fold cross-validation procedure using SVM classification model to different datasets. These datasets are composed of a variable number of PCA or PLS components (between 3 and 16, according to the maximum number of extracted features). As can be seen, the larger the number of components, the lower the misclassification errors. The same trend can be observed in Figures 12, 13 and 14, which show the misclassification proportions obtained from the

application of SVM^G and LR methods to the PCA and PLS components corresponding to Lab color space and uniform histograms, grayscale and DWT, and Lab color space and DWT, respectively.
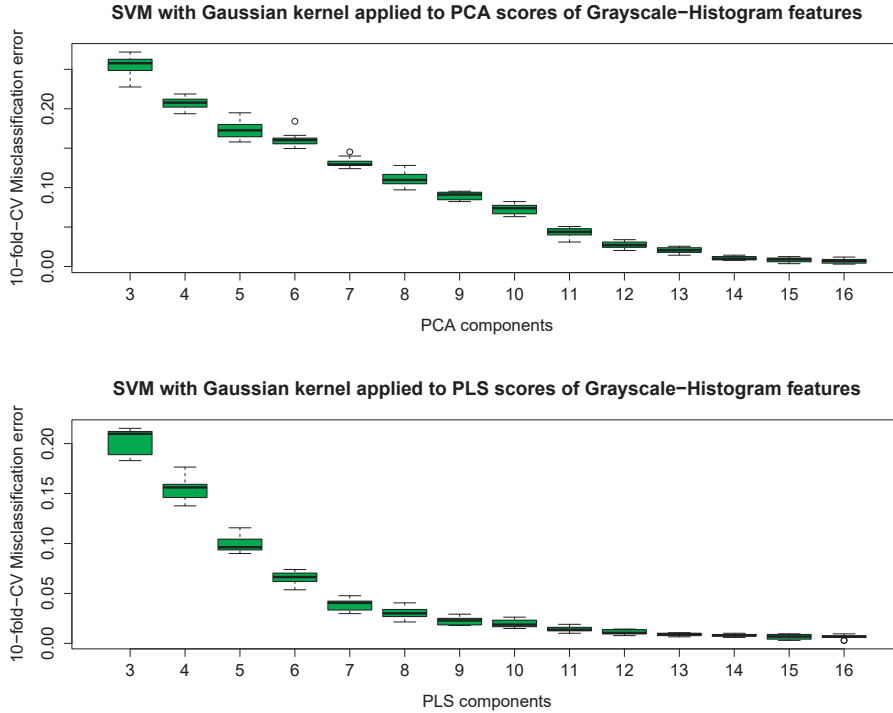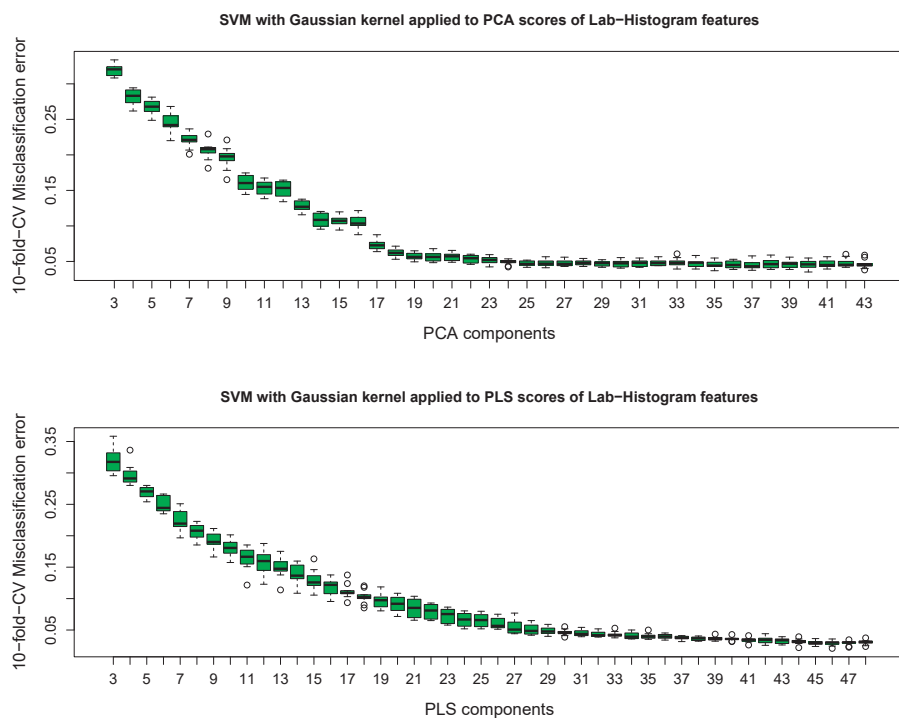

Fig. 11 Boxplots of the misclassification proportions using SVM applied to the PCA and PLS components (obtained from the features extracted using grayscale and uniform histograms). Different scales were used in the plots for the sake of legibility and interpretation.

As observed in Figure 11, misclassification errors of less than 5% were achieved when SVM with a Gaussian kernel were applied, at least, to the first 11 PCA components, and to the first 7 PLS components. As expected, when the PLS components were used for classification purposes, lower misclassification errors were obtained with a smaller number of components (i.e. leading to a shorter computing time). In this line, when using 14 or more PLS components, more than the 99% of the crankshaft images were correctly classified. The good performance of this approach can be also observed considering the small dispersion of the corresponding boxplots. The misclassification errors obtained by the remaining classification methods were higher, thus only SVM with a Gaussian kernel provided an adequate solution for the requirements.

3.2 Classification performance using Lab and uniform histograms

When the Lab color space and uniform histograms are applied to the crankshaft images, a set composed of 48 features is obtained. Table 3 shows that the best results are obtained by the application of SVM with a Gaussian kernel. In this line, Figure 12 presents the boxplots of the 10-fold cross-validation misclassification errors, obtained by the implementation of SVM classification method to a different number of PLS and PCA components.
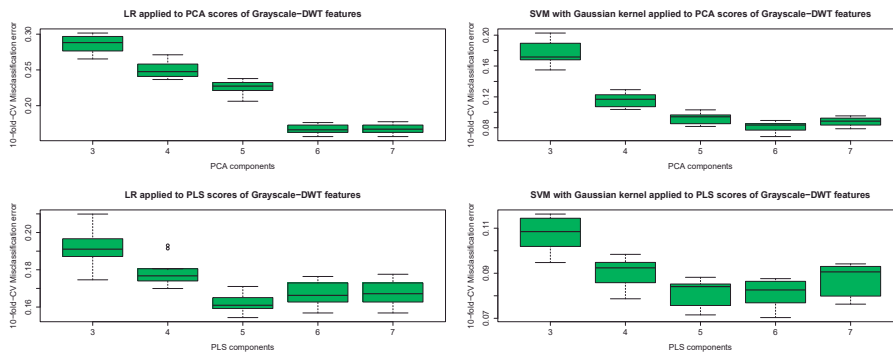

Fig. 12 Boxplots of the misclassification proportions using SVM applied to the PCA and PLS components (obtained from the features extracted using Lab and uniform histograms). Different scales were used in the plots for the sake of legibility and interpretation.

A mean correct classification proportion of 96% was achieved when this method was applied to, at least, the first 37 PLS components. Very competitive results (around the 94% of correct classification) were also obtained from the first 18 PCA components, from the 30 first PLS components (over 95% of correct classifications), and using the 45 first PLS components (about the 97.5%). The remaining methods did not offer a good classification solution, and so they would not be applicable to the identification of defective crankshafts.

3.3 Classification performance using grayscale and DWT

A smaller set composed of 7 features is obtained by the application of grayscale and DWT for color and texture analysis, respectively. Figure 13 shows the boxplots of the misclassification errors obtained for LR and SVM applied to a different number of PLS and PCA components. Figure 13 and Table 3 show a mean misclassification proportion of around 8% (i.e. a 92% proportion of correct classification) reached by the application of SVM with a Gaussian kernel to, at least, the first 6 PLS components. Focusing on the traditional methods, a mean misclassification proportion of 16% was obtained when applying LR to the first 5 PLS components.
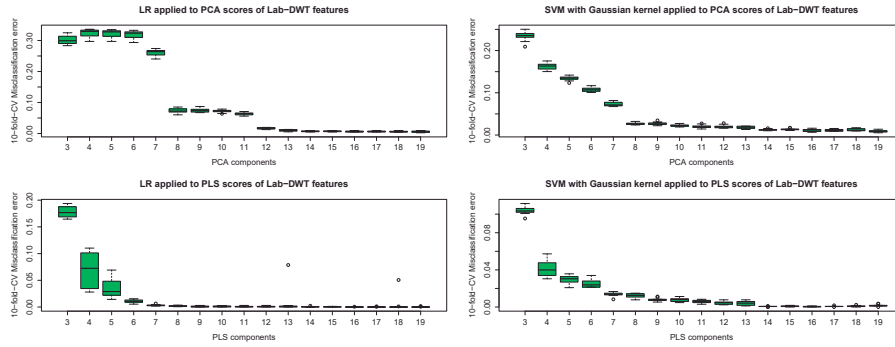

Fig. 13 Boxplots of the misclassification proportions using LR and SVM applied to the PCA and PLS components (obtained from the features extracted using grayscale and DWT). Different scales were used in the plots for the sake of legibility and interpretation.

3.4 Classification performance using Lab and DWT

If the Lab color space and DWT for color-texture analysis are applied to the crankshaft image patches, a set composed of 21 features is obtained. Table 3 shows that very low mean misclassification errors are reached by almost all the classification methods and datasets. NBC and SVM with a sigmoid kernel present a worse performance, but even their results could be competitive compared to the manual inspection of defective crankshafts.

Figure 14 shows the boxplots of the misclassification errors obtained for LR and SVM applied to a different number of PLS and PCA components. They were under 5% when SVM with a Gaussian kernel was applied to, at least, the first 4 PLS components. This is a great result regarding both computer runtime and classification performance. The time to make a prediction is 0.87 seconds if LR method is applied, and 8.73 seconds if SVM^G is used, rather shorter than manual inspection. If we consider that the simpler LR provided very accurate results significantly faster (see Table 6), this method should be the first option

in order to distinguish defective from non-defective areas in crankshaft images. In fact, Figure 14 shows that misclassification errors under 5% are obtained when applying LR to, at least, the first 5 PLS components. Once again, the PLS projections provided better results with a less number of components than the PCA projections. Furthermore, an almost perfect classification was obtained by applying LR to the first 16 PLS components. In this case, a misclassification error of 0.006% was reached (Table 3), which means that the 99.994% of image patches were correctly classified into defective and non-defective. Similar results were obtained when using SVM with a Gaussian kernel, which is the classification method that, in general, provided better results in this research.

Fig. 14 Boxplots of the misclassification proportions using LR and SVM applied to the PCA and PLS components (obtained from the features extracted using Lab and DWT). Different scales were used in the plots for the sake of legibility and interpretation.

Table 5 shows the confusion matrices corresponding to the application of LR and SVM^G to the 10 cross-validation partitions. Almost all the image patches are correctly classified as (0) non-defective or (1) defective areas. Notice that the dataset used in this study is balanced, and thus there is no risk to obtain spuriously low misclassification proportions due to the ability to correct classify a specific class. In fact, the values of kappa index [5] (alternative to misclassification proportion or accuracy) are in accordance with correct classification proportions: $\kappa_{LR} = 0.999$ and $\kappa_{SVM^G} = 0.996$, showing a perfect concordance between real classes and predictions.

In order to illustrate how much time each classification alternative takes, Table 6 shows the computing running times (measured in seconds) corresponding to the complete 10-fold cross-validation process applied to PCA and PLS sets composed of 3 to 19 components. It is important to note that these execution times refer to the complete comparison procedure, including 10-fold cross-validation and several sets. In a real application, the running time to assign a class to each feature vector would be considerably lower. They could be estimated dividing the numbers in Table 6 by 285,260 (17 sets × 10 subsets related to 10-fold cross-validation × 1,678 feature vectors in each subset).

			Estimated									
			0	1	0	1	0	1	0	1	0	1
LR	Real	0	819	1	839	1	820	1	805	2	769	42
		1	0	857	1	836	0	856	0	871	90	777
		0	810	2	868	3	826	0	823	0	831	0
		1	2	864	0	807	0	852	0	855	0	847
SVM^G	Real	0	818	2	833	7	815	6	801	6	806	5
		1	1	856	3	834	2	854	3	868	5	862
		0	811	1	870	1	820	6	817	6	828	3
		1	2	864	1	806	0	852	7	848	3	844

Table 5 Confusion matrices corresponding to the application of LR and SVM^G to the 10 cross-validation partitions of data matrix corresponding to the first 16 PLS components of the original Lab-DWT feature matrix. In each partition of about 1678 individuals, there are a variable number of defective and non-defective sections of crankshafts.

	LDA	NBC	LR	SVM^L	SVM^G	SVM^S
PCA	19.32	129.62	22.64	2,063.13	4,197.10	6,142.31
PLS	55.28	154.97	143.57	883.45	2,266.41	2,703.66

Table 6 Execution times (seconds) of the complete 10-fold cross-validation process, applying each classification method to a number of PLS or PCA components ranging from 3 to 19. The PCA and PLS components are obtained from the feature set corresponding to the Lab color space and DWT.

As a final remark, the automatic classification between defective and non-defective areas of crankshafts is highly reliable when applying the LR method to, at least, the first 5 PLS components obtained from the features extracted by the Lab color space and the DWT technique. Note that the Lab-DWT set provided better results for the identification of crankshaft defects than the remaining feature sets. Thus, our recommendation is to use this set to be analyzed in the production line of crankshafts inside forging factories.

4 Conclusions

Several approaches can be found in the literature to identify and/or characterize defects in forged crankshafts, which include image segmentation methods, neural networks and scattering techniques. However, there is still a lack of automatic procedures for the identification of defective crankshafts, with all the steps comprehensively described to provide companies requiring automatic solutions with a useful tool. Thus, we propose a way to automate the quality control in the crankshaft manufacturing process, right at the end of the production line. Applying the proposed method, surface defects on crankshafts can be identified and, thus, defective units can be automatically discarded. Therefore, our approach can be used as part of any reliable quality control procedure in the automotive industry, playing an important role in the digitalization of the manufacturing process.

Regarding the proposed methodology, the identification of defective forged crankshafts has been tackled using image analysis and supervised classification

approaches. Once crankshaft images were obtained in a pilot plant, different color and texture analysis techniques were applied in order to extract feature vectors from them. Then, different well-known classification methods were applied to 160 feature sets composed of the PLS and PCA projections derived from the matrices of original features, and subsequently validated by a 10-fold cross-validation procedure.

The main conclusion is that the problem of identifying defective crankshafts can be solved by applying the proposed methodology, since a very reliable defective crankshaft classification was achieved. Significantly low misclassification errors were obtained using all the feature sets composed of different image properties. Thus, these features extracted by the proposed color and texture analysis methods provided representative information about the presence or absence of surface defects in forging crankshafts. The proposed methodology was compared with manual inspection in a pilot plant specifically built for this purpose. The proportion of correct classification is about 10% higher than the one obtained by manual inspection, according to the technical staff of the company who participate in this project (personal communication).

Additionally, the proposed methodology prevents discarding false defective crankshafts that is a current problem with the existing manual inspection system. In terms of efficiency and economic impact, the demand of personnel and execution time in supervision tasks can be noticeably reduced by using this methodology, and the efficiency can be increased due to the elimination of manual errors.

With respect to the best classification results, the highest proportion of defects correctly identified were obtained when applying LR or SVM with a Gaussian kernel to the PLS component set derived from the Lab and DWT features. These results are summarized as follows:

- Over 99% of correct classification was obtained when applying LR to the first 16 PLS components.
- Correct classification proportions over 95% were reached when applying the LR method to, at least, the first 5 PLS components. Similar proportions were also estimated when applying SVM with a Gaussian kernel to, at least, the first 4 components.
- Despite its good performance, the SVM model is more time-consuming than LR. Thus, for crankshaft classification purposes, the authors recommend the use of the LR method applied to, at least, the first 5 PLS components obtained from the combination of Lab and DWT.

It is important to stress that, in general, SVM with Gaussian kernel provided a high classification performance regardless of the dataset:

- A mean correct classification proportion of 99% was obtained when using the first 14 PLS components from the features extracted with grayscale as color model and uniform histograms as texture analysis method.
- A mean correct classification proportion of 96% was achieved when using the first 37 PLS components from the features extracted with the Lab color space and uniform histograms.

– A mean correct classification proportion of 92% was reached when using the first 6 PLS components from the features extracted with grayscale as color model and DWT as texture analysis method.

Finally, note that highest proportions of correct classification in defective and non-defective areas in crankshaft images were obtained when using PLS components instead of PCA.

Acknowledgments

References

1. Becerra-Villanueva, J.A., Jiménez-Espadafor, F., Cruz-Peragon, M., Torres-García, M.: A methodology for cracks identification in large crankshafts. Mechanical Systems and Signal Processing **25**(8), 3168–3185 (2011)
2. Bradski, G.: The OpenCV library. Dr. Dobb's Journal of Software Tools **25**, 120–126 (2000)
3. Chandna, P., Chandra, A.: Quality tools to reduce crankshaft forging defects: an industrial case study. Journal of Industrial and Systems Engineering **3**(1), 27–37 (2009)
4. Chiou, Y.C., Lin, C.S., Chiou, B.C.: The feature extraction and analysis of flaw detection and classification in BGA gold-plating areas. Expert Systems with Applications **35**(4), 1771–1779 (2008)
5. Cook, R.J.: Kappa. Wiley Encyclopedia of Clinical Trials pp. 1–7 (2008)
6. Daubechies, I.: Ten Lectures on Wavelets, vol. 61. SIAM, CBMS series (1992)
7. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Annals of genetics **7**(2), 179–188 (1936)
8. Fontani, D., Francini, F., Longobardi, G., Sansoni, P.: Control of crankshaft finish by scattering technique. Optics and Lasers in Engineering **35**(6), 387–396 (2001)
9. Francisco-Fernández, M., Tarrío-Saavedra, J., Mallik, A., Naya, S.: A comprehensive classification of wood from thermogravimetric curves. Chemometrics and Intelligent Laboratory Systems **118**, 159–172 (2012)
10. Grote, K.H., Antonsson, E.K.: Springer Handbook of Mechanical Engineering, vol. 10. Springer Science+Business Media (2009)
11. Iborra, A., Alvarez, B., Jiménez, C., Fernández-Merono, J.M., Fernández, C., Suardíaz, J.: Automated Visual Inspection system (AVI) for crankshaft production processes. European Journal of Mechanical and Enviromental Engineering **45**(1), 29–34 (2000)

---

[4] http://vigotec.com

12. Jensen, F.V.: An Introduction to Bayesian Networks, vol. 210. Springer-Verlag (1996)
13. Jolliffe, I.T.: Principal Component Analysis and Factor Analysis. In: Principal Component Analysis, pp. 115–128. Springer-Verlag Berlin Heidelberg (1986)
14. Kotsiantis, S., Kanellopoulos, D., Pintelas, P.: Handling imbalanced datasets: A review. GESTS International Transactions on Computer Science and Engineering **30**(1), 25–36 (2006)
15. Kotsiantis, S.B., Zaharakis, I., Pintelas, P.: Supervised machine learning: a review of classification techniques. Emerging Artificial Intelligence applications in Computer Engineering **160**, 3–24 (2007)
16. Li, T.S.: Applying wavelets transform, rough set theory and support vector machine for copper clad laminate defects classification. Expert Systems with Applications **36**(3), 5822–5829 (2009)
17. Lindgren, F., Geladi, P., Wold, S.: The kernel algorithm for PLS. Journal of Chemometrics **7**(1), 45–59 (1993)
18. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **11**(7), 674–693 (1989)
19. McEvily, A.J.: Failures in inspection procedures: case studies. Engineering Failure Analysis **11**(2), 167–176 (2004)
20. McLaren, K.: XIII The development of the CIE 1976 (L* a* b*) uniform colour space and colour-difference formula. Coloration Technology **92**(9), 338–341 (1976)
21. Mitchell, T.M.: Machine Learning. McGraw-Hill (1997)
22. Pater, B.V., Thakkar, H.R., Mehta, S.B.: Review of Analysis on Forging Defects for Quality Improvement in Forging Industries. Journal of Emerging Technologies and Innovative Research **1**(7), 871–876 (2014)
23. R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (2018). Available at `http://www.R-project.org`
24. Seber, G.A.F., Lee, A.J.: Linear Regression Analysis, vol. 329. John Wiley & Sons (2012)
25. Tarrío-Saavedra, J., Francisco-Fernández, M., Naya, S., López-Beceiro, J., Gracia-Fernández, C., Artiaga, R.: Wood identification using pressure DSC data. Journal of Chemometrics **27**(12), 475–487 (2013)
26. Thottungal, A.P., Sijo, M.T.: Controlling Measures to Reduce Rejection Rate due to Forging Defects. International Journal of Scientific and Research Publications **3**(3), 238–243 (2013)
27. Vapnik, V.: The Nature of Statistical Learning Theory, 2nd edn. Springer Science+Business Media (2000)
28. Wang, K., Jiang, W.: High-dimensional process monitoring and fault isolation via variable selection. Journal of Quality Technology **41**(3), 247–258 (2009)
29. Wehrens, R.: Chemometrics with R. Multivariate Data Analysis in the Natural Sciences and Life Sciences. Springer-Verlag Berlin Heidelberg (2011)
30. Wold, H.: Partial Least Squares. Encyclopedia of Statistical Sciences **9**, 1 (2006)