Electronic Letters on Computer Vision and Image Analysis 22(1):35-51; 2022

Infrared thermography for seal defects detection on packaged products: unbalanced machine learning classification with iterative digital image restoration

Victor Guillot*

*Thimonnier, 11 avenue de la Paix, 69650 Saint Germain au Mont-d'Or, France

Received 9th of April, 2022; accepted 10th of May 2023

Abstract

Non-destructive and online defect detection on seals is increasingly being deployed in packaging processes, especially for food and pharmaceutical products. It is a key control step in these processes as it curtails the costs of these defects.

To address this cause, this paper highlights a combination of two cost-effective methods, namely machine learning algorithms and infrared thermography. Expectations can, however, be restricted when the training data is small, unbalanced, and subject to optical imperfections.

This paper proposes a classification method that tackles these limitations. Its accuracy exceeds 93% with two small training sets, including 2.5 to 10 times fewer negatives. Its algorithm has a low computational cost, and does not need any prior statistical studies on defects characterization.

Key words: Control, seal, machine learning, small and unbalanced training set, thermography, iterative image restoration

1 Introduction

To maintain an optimum throughput, the packaging industry has increasing demands for quality control. One of the most critical stages in the process of such industries is the sealing of materials (Figure 1).

Package integrity is usually checked on small test samples, and offline. A small defect in sealing material can induce significant losses if not detected immediately. Such losses include leakage of raw materials used before and after sealing as well as a perturbation on the production line. A minor leakage forces to remove the faulty package out of line, clean the soiled area, and sometimes cancel the post sealing operations like over-wrapping on several batches. Leaks are not the only defects; particle contamination, wrinkles, bubbles, and over-sealing also affect the strength and tightness of the seal, leading to potential issues during the package's life after sealing.

Correspondence to: victorguillot_eu@protonmail.com

Recommended for acceptance by Angel D. Sappa
https://doi.org/10.5565/rev/elcvia.1567
ELCVIA ISSN: 1577-5097
Published by Computer Vision Center / Universitat Autonoma de Barcelona, Barcelona, Spain

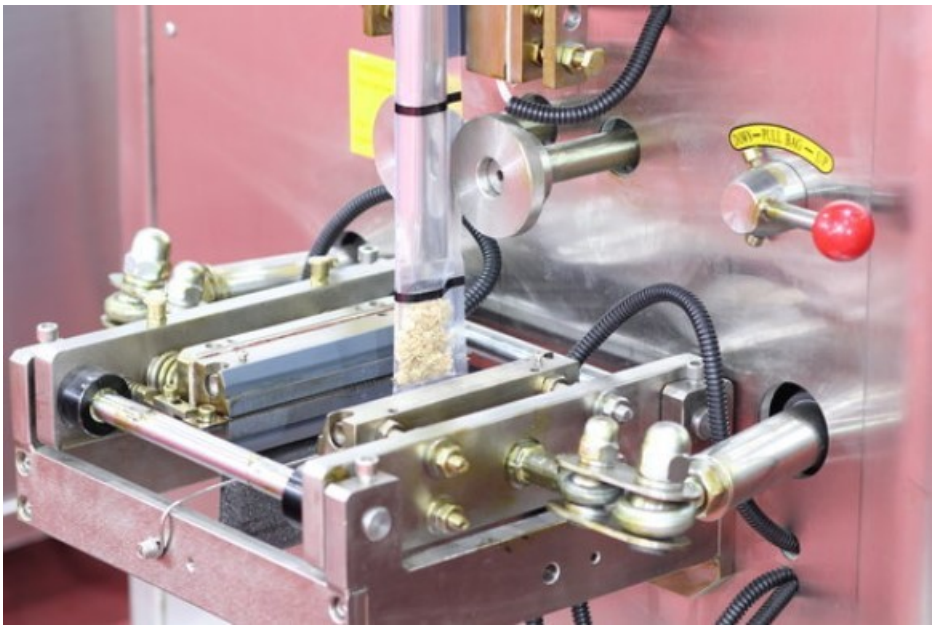

Figure 1 Horizontal sealing on a Vertical Form-Fill-Seal (VFFS) machine

Various non-destructive and online tests have been deployed for sealing. Some of these tests rely on 1D variables control like the voltage and current waveforms, as in high-frequency dielectric sealing [1], or the energy transmitted, as in ultrasonic sealing [2]. These methods are easy and quick, but their mono-dimensional characteristic cannot provide a real 2D integrity image of the seal.

On the other hand, several non-destructive 2D sealing control methods also exist. Infrared (IR) thermography is one of the most attractive methods among them. It showcases various advantages like the absence of electrical and mechanical stress and low cost when used with uncooled microbolometers. It can be applied to a variety of materials, including opaque ones in the visible band. Nevertheless, it has certain limitations like low resolution due to a relatively high wavelength, loss of accuracy on moving objects (Figure 2), and compromise of pixel resolution and thermal sensitivity.

Some commercial products offer a sealing control with IR like Qipcam from Qipack, 1420 Braine-l'Alleud, Belgium. Their performances and image processing methods are not made available in publications. Some of the recent implementations rely on statistical tests and discontinuity detection [3], whose formulations depend strongly on the human experience. Machine learning provides a solution that can mitigate such pre-requisites. In this paper, various learning models are evaluated with small and unbalanced datasets, i.e., having a small fraction of IR images of bad sealings (called negatives) compared to good ones (called positives). This consideration enables a practical and quick implementation of the control whereas the large and balanced datasets require time-consuming runs and tests. However, it represents a technical challenge as the traditional machine learning models for binary classification are built for large and balanced datasets. Another challenge comes from optical limitations. Putting aside the loss of resolution due to the spatial frequency bandwidth and the quantization of thermal captures, optical blurring limits the spatial resolution to values higher than those related to the pixel size. This paper proposes a method to address this limitation.

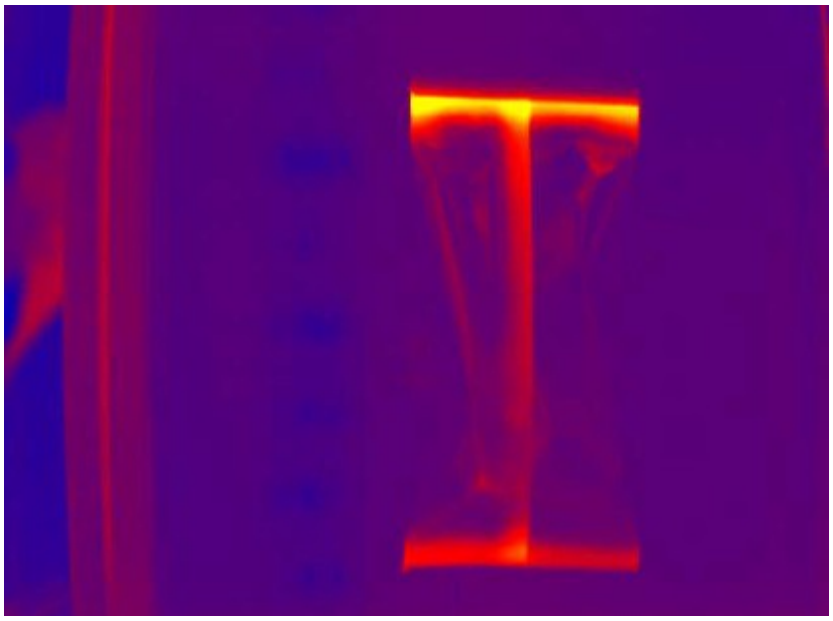

Figure 2 Thermal image of a package on a conveyor belt with motion blur (two horizontal seals and one longitudinal seal are highlighted)

2 Related work

Since the early 2000s, IR thermography has been examined for sealing inspection. Umezaki et al. [4] noticed that cavities and insufficient adherence in plastic film seals could be visualized with IR pictures but not by visible light or micrographs. In the year 1994, Mitsubishi Heavy Industries, Ltd [5] patented the first-ever application of IR thermography for sealing control.

Al-Habaibeh et al. [6] qualified the possibility to identify abnormal laser sealing of film on food containers using IR thermography. Accordingly, if the number of white pixels after segmentation and transformation under the variance operator is higher than a set threshold, the seal is identified as bad.

Morris [7] demonstrated the method to detect interrupted seals on Polyethylene Terephthalate/Aluminium/Poly Propylene laminated films with active IR thermography and an erosion/centroid-count algorithm. In this method, activation was done by employing a 4W laser source.

D'huys et al. [8] compared the size detection limit and the influence of the recording start time of six image processing methods to detect defects on Oriented Poly Amide/Poly Ethylene laminated films. IR images were taken on a static seal in a VFFS after thermal sealing. The evaluation was done on the basis of total overlapped bad pixels between different processed thermal images and a digital image of the defective seal.

The lowest detection limit of 0.6 mm with a probability of 95 %, along with a mean of falsely-detected particle percentage among the lowest was obtained by a method based on fitting of the cooling profile over a sequence of thermal images. A method based on a single frame led to comparable results, provided the capture time after sealing is the lowest possible (less than 5s). Its advantage over other methods based on multiple frames was its processing time. This result motivated the author of this work to focus on single frame methods captured immediately after the sealing.

After 2010, IR control methods of seals started commercializing. However, their performance and image processing methods are not yet published.

Some current implementations rely on statistical tests on random variables characterizing defects [6] and discontinuities detection [3]. To determine suitable thresholds, especially for new products, the statistical tests require substantial experience increasing time and cost. The automatic detection of thermal discontinuities is less dependent on experience but not exhaustive.

To avoid such expensive and lengthy processes, Machine Learning (ML) proves to be an attractive

solution. The current subject is identified as a binary classification problem in the ML environment. To resolve it, a classifier is built with a training set of positives and negatives and then applied to infer the state of new images. A classification problem is said unbalanced when one class is in minority in the training set. Negative thermal images are much more time consuming to produce than positive ones. Hence, practical consideration often prompts unbalanced classification problem with negatives forming the minority class.

Since the end of the 1990s, ML models for this problem and multidimensional data like images are an active research area. Recently, deep learning approaches aroused a multiplication of such algorithms.

Early application of ML to IR thermography dates back to the 2000s [9]. Since then, the deep learning approach has been adopted for numerous cases of material control and medical diagnosis. Cruz et al [10] introduced an automated scheme of controlling bottle sealings using a thermal camera. A small and unbalanced dataset was used for the application. Augmentation and segmentation were done artificially. The sum of white pixels for each row and column of a segmented image defined its signature. An Artificial Neural Network (ANN) was trained with these signatures. Its net contained two hidden layers with 100 and 10 neurons, respectively. It achieved a significant precision of 98.96% and a recall of 100% on the test set.

This paper investigates various ML approaches on two small and unbalanced datasets, including neural networks and less computational models. The selection of the prior dimensionality reduction is justified through this approach. The limitations of these ML approaches, as used in early publications [10], are exposed in the presented case, and a digital optical correction is eventually provided to address these limitations.

3 Test setup and datasets

The evaluation of ML models requires a dataset of positive as well as negative IR images. Using a semi-automatic sealing machine, seal samples are prepared (model variant of UV D3 Inox, Thimonnier, Saint Germain au Mont d'Or, Rhône-Alpes, France).

Between two heating jaws kept at a controlled temperature above the Poly Ethylene (PE) fusion temperature, an unsealed package composed of two laminates Polyethylene Terephthalate (PET)/ PE is inserted. The thermal regulation is ensured with the help of a digital controller (E5CN, Omron, Kyoto, Japan). Constant pressure is applied on the laminates for 2 s using jaws driven by pneumatic actuators. The heated package is then laid down on a horizontal aluminum plate (Figure 3).

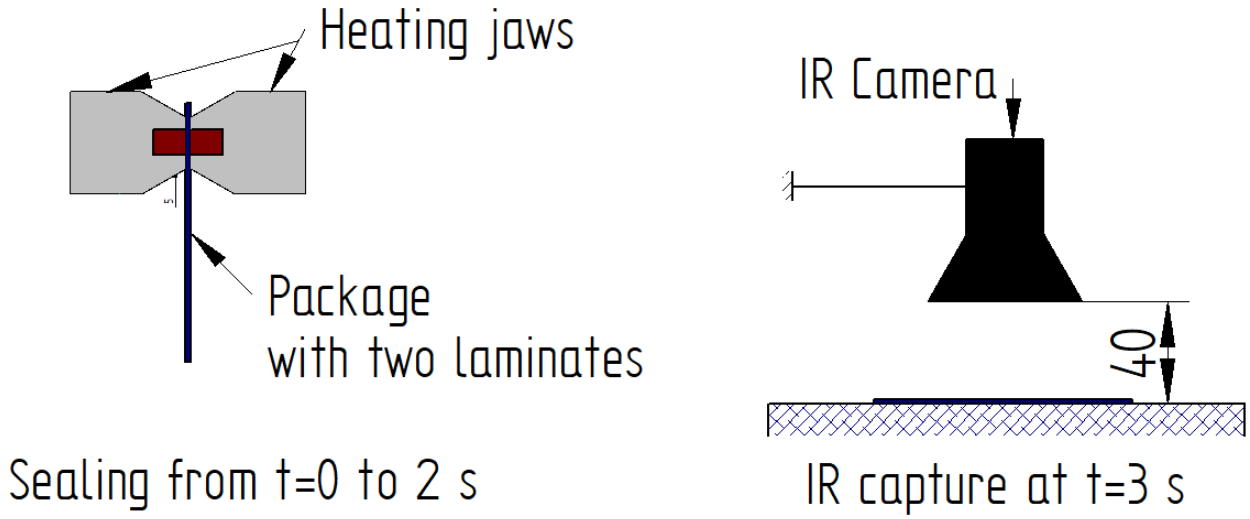

Figure 3 Scheme for dataset building

One second after the release, an IR image is captured on the sealing area at a distance of 0.4 m.

The IR camera (FLIR E50, FLIR Systems Inc., Wilsonville, OR, USA) has an IR resolution of 240 × 180 pixels, an angular spatial resolution of 1.82 mRad, a thermal sensitivity below 0.05 °C at 30 °C,

and a spectral range between 7.5 and 13 um. The deduced spatial resolution is 0.728 mm per pixel and the thermal resolution is about 3 times fold or 2.184 mm when dispersion and aberrations are considered [11].

Its parameters (ambient temperature, humidity, reflected apparent temperature and material emissivity) are set according to the camera's standard procedure.

The temperature measurement of an IR camera makes assumptions of reflection, transmission as well as absorption factors of the material and the environment. Some of these assumptions were not verified. For instance, it assumes constant transmission and absorption factors of the material within the spectral range, which is not the actual case.

The best way to reduce temperature errors would be to use a narrow spectral range within which the laminate foils exhibit constant optical factors. This narrow band could be chosen to obtain a highly transmissive upper foil (PET) and a highly reflective inner sealing foil (PE). In this way, the IR capture of the sealing area approximates an IR capture of the sealing foil alone, which reveals some of the critical defects like leaks and incomplete sealing.

However, to compare different seals, the control method does not need an absolute temperature measurement but only a repetitive process. Also, some defects can be revealed only by the upper foil (for example, surface holes or wrinkles). Hence, having an IR capture with an upper foil partially transmissive is desired.

Four groups of captures are prepared: two groups for positives captured on 15th June 2017 and 18th July 2017; two groups of negatives captured on the same days (Table 1).

Seals are qualified as good ones only after a 50 N shear stress and visual checking.

Various defects are produced: insertion of coffee powder/hear/bread grain between plastic laminates, hole piercing through one laminate of diameter 1 to 3 mm, and placing piece(s) of Teflon^{TM} adhesive on the surface. Each type of defect is enumerated in Table 1.Some raw images are shown in Figure 4.

Date	Good seals (Positives)	Bad seals (Negatives)	List of defects
15/06/2017	30	24	3 (coffee powder), 2 (1 mm hole),2 (2 mm hole),2 (2 mm hole), 5 (Teflon^{TM} adhesive bands), 5 (hair), 5 (bread grain)
18/07/2017	15	19	3 (coffee powder), 2 (1 mm hole),2 (2 mm hole),2 (2 mm hole), 4 (Teflon^{TM} adhesive bands), 3 (hair), 3 (bread grain)

Table 1 Datasets of June and July 2017

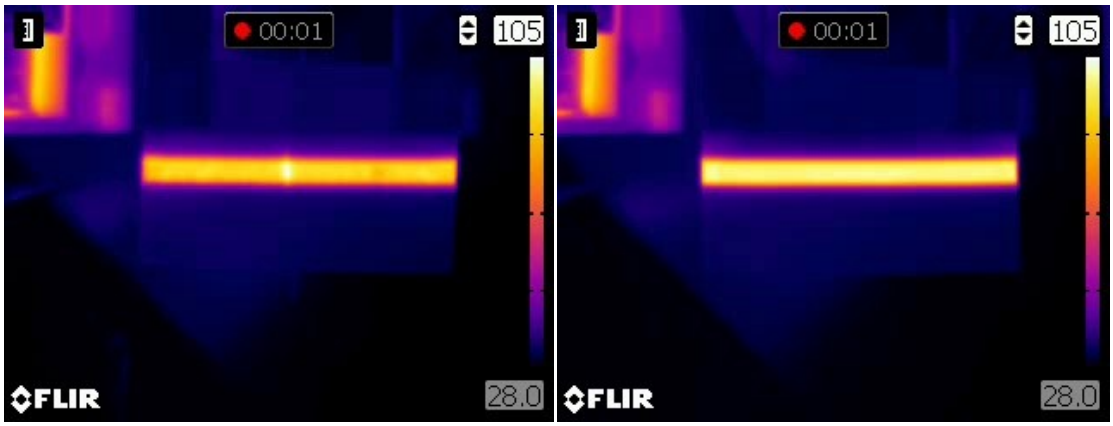

Figure 4 Some raw images from 2017's dataset (bad seal on left, good seal on right)

4 Prior treatments and dimensionality reduction analysis

Prior treatments are applied to the dataset for dimensionality reduction and metrics (Figure 5). Following successive treatments are performed:

- Initially, a Region of Interest (ROI) is selected to ignore the objects lying outside the sealing area.
- A 5×5 Gaussian kernel is applied to smooth the sharp edges and reduce the noise prior to segmentation and Principal Component Analysis (PCA).
- Filtered ROI is then converted into a grayscale image.
- Global thresholding is done to remove the cold pixels and the outer region of sealing. Since the ROI has a bimodal histogram, Otsu's method fits here. However, this thresholding could delete some useful information, as cold pixels within a hot area could exist in case of a defect. Thus, the distorted metrics which are valuable for classification like energy, entropy, mean, and variance are calculated on the ROI before thresholding and saved for later use.
- To avoid distortion on pictures, no resizing is done to a fixed size as they have different ratios of height over width.

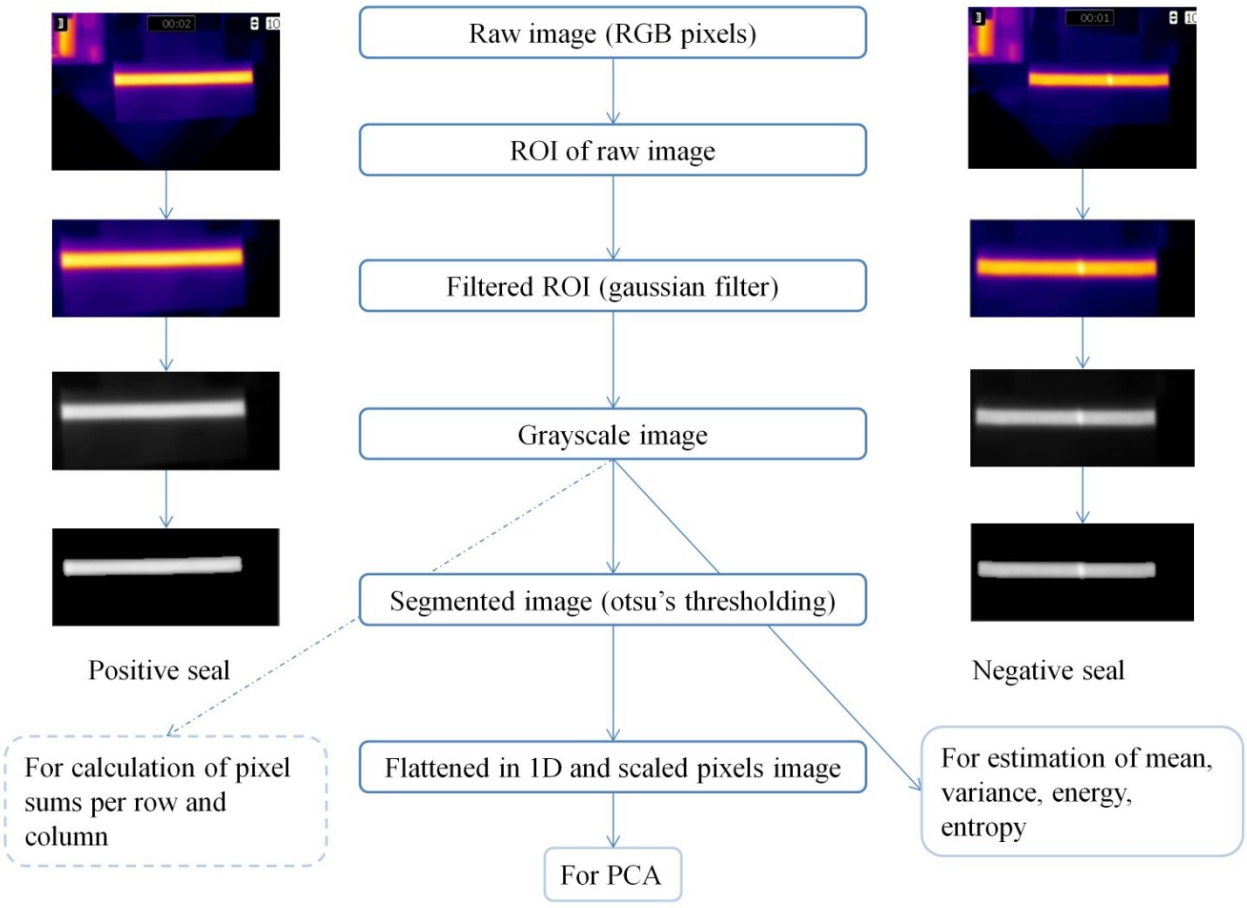

Figure 5 Prior treatments on IR images

An extensive comparison of good and bad seals has been made using classical metrics: mean, variance, and energy. Through Boxplots of these metrics, it can be inferred that none of them can be independently used to distinguish positives from negatives (Figure 6).

However, a representation of these two classes along three metrics: mean, variance, and entropy (or energy), improves the separation (Figure 7).

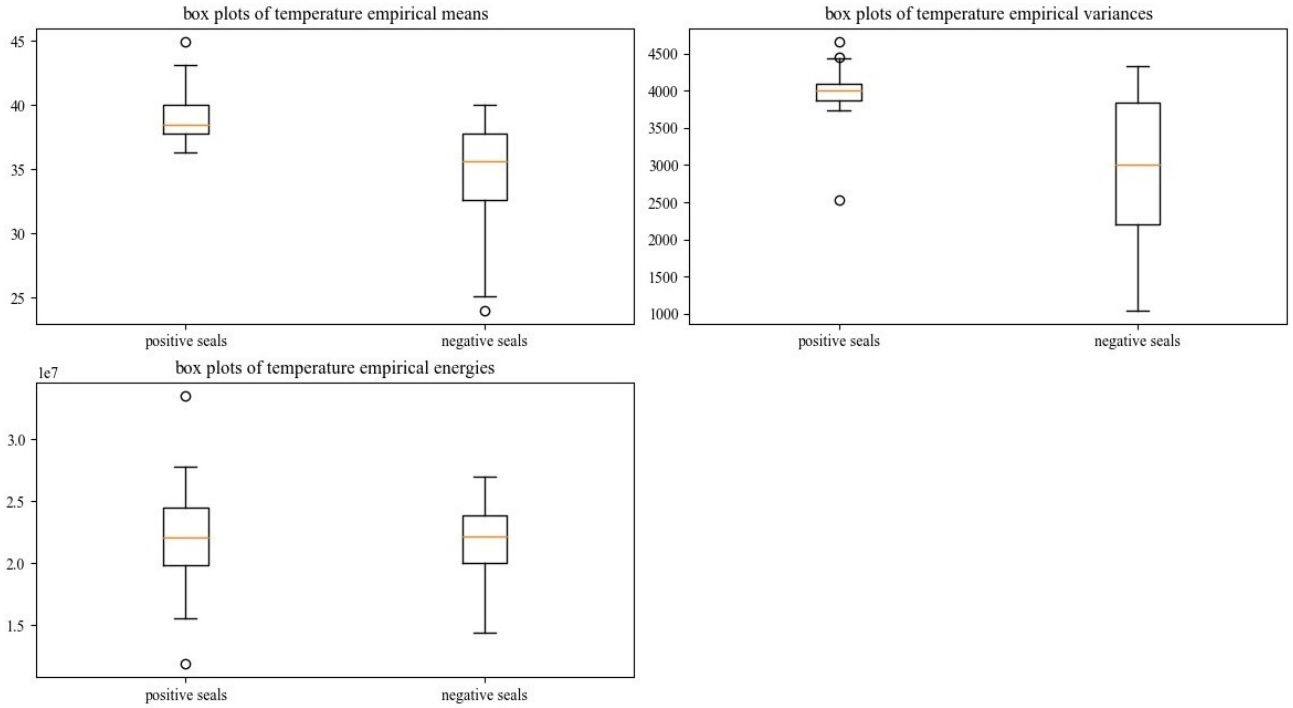

Figure 6 Boxplots of empirical means, variances and energies, for the dataset of 15th June

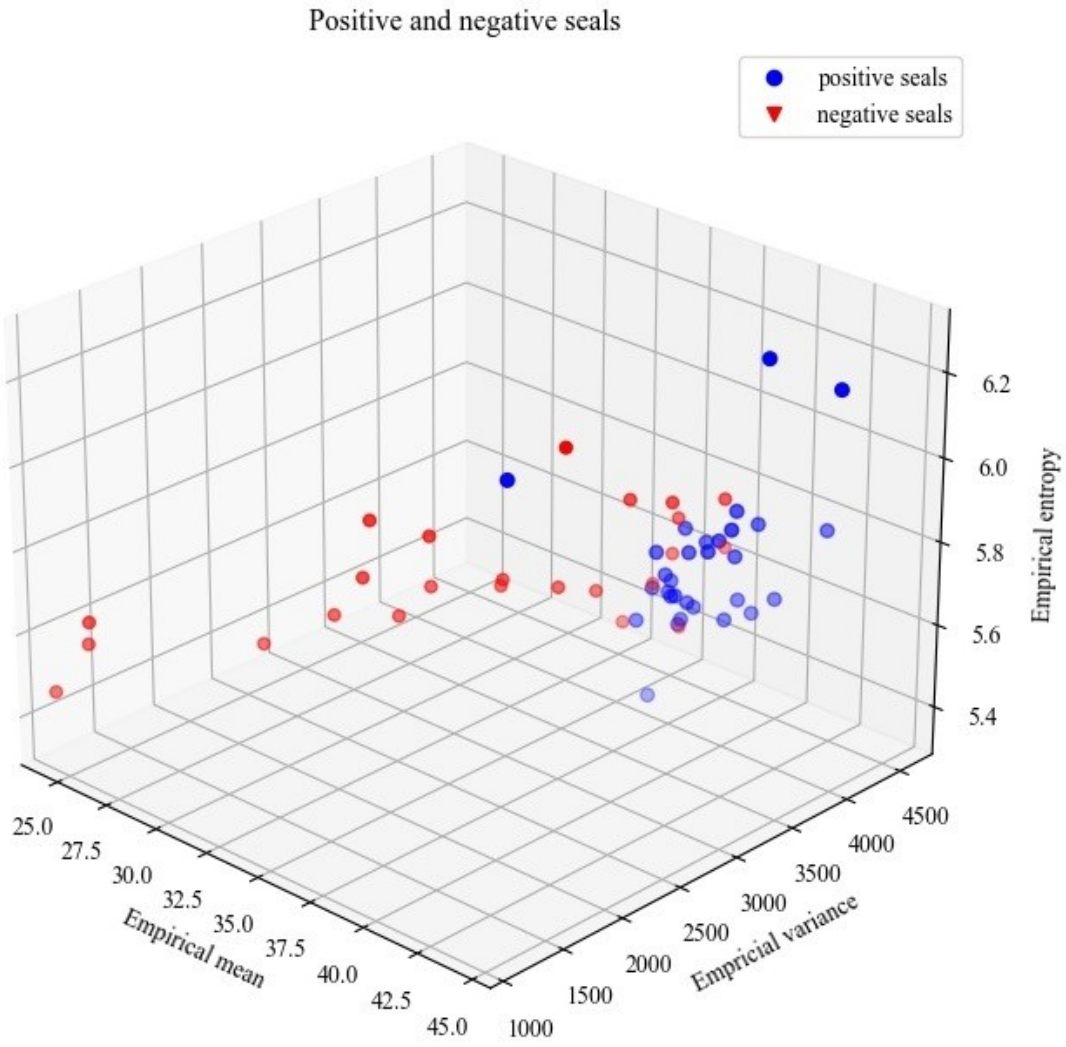

Figure 7 Spatial representation of images along 3 metrics, for 15th June dataset

A PCA is performed to improve further separation. The principal axis are determined with part of the dataset of positives in the training set (see below for further details), followed by a calculation of Principal Components (PC) of every positive and negative along these axes. Before PCA, each positive is flattened to 1D, standardized, and added as a line to a matrix of positives. A similar process is done for negatives. PCA is exclusively conducted on these two matrices.

With a variance of 0.99, PCA leads to 27 principal components for the dataset of 15th June and 4 components for the dataset of 18th July dataset.

The distance between two classes is computed for various linear decompositions (Table 2).

Linear decomposition of pictures	Minimum Euclidian distance between classes	Mean distance between classes	Ward distance
15th June dataset			
3 D (mean, variance, energy)	13	17	237
27 D (PCA with a variance of 0.99)	1337	3647	7680
30 D (above PCA with a variance of 0.99 + mean, variance, energy)	1337	3647	7684
372 D (sums of pixels per column and row)	66	221	967
18th July dataset			
3 D (mean, variance, energy)	38	43	170
4 D (PCA with a variance of 0.99)	993	2186	1667
7 D (above PCA with a variance of 0.99 + mean, variance, energy)	994	2187	1675
372 D (sums of pixels per column and row)	180	834	1579

Table 2 Distances between classes for different linear decompositions

A calculation is done on three different types of distances. This calculation proves that as compared to a simple 3D decomposition, the decomposition of each picture along its PC increases the distance between the classes. This separation can be marginally improved by adding three other metrics to the PC.

The decomposition in 372 vectors corresponds to a similar prior treatment used by Cruz et al. [10]. This prior treatment resulting to these 372 dimensions consists of:

- ROI extraction
- Gaussian Filter
- Grayscale conversion
- Binary conversion (Thresholding)
- A calculation of sums of each row and each column to obtain a vector of 102 rows and 270 columns (372 components).

This decomposition exhibits lower distances between classes than the PCA decomposition, and its higher dimension induces higher computational costs whatever the ML model used after. This outcome holds if the Gaussian filter is suppressed and the grayscale conversion is substituted by a Hue Saturation Value (HSV) conversion as in [10].This has led to discard this decomposition in the presented work.

5 Anomaly detection

The practical distribution of images across the known classes is skewed: negatives are in the minority as compared to positives. This problem is identified as anomaly detection.

5.1 Workflow

Dataset is split into two sets: one for building the classifier ("training data") and the other for its testing ("test data"). For ML algorithms with few parameters and relatively low complexity like Support Vector Machine (SVM), the training data is partitioned into two subgroups: one for training and one for the pre-test. k combinations of subgroups of training and pre-test are evaluated in this paper ("k-fold cross validation"). The classifier is evaluated on the training data as an average of rates over the k combinations for each set of model parameters ("hyperparameters"). These rates refer to an evaluation metric like balanced accuracy or $f_{0.5}$ score to quote a few, chosen to get the best performance on test data.

After a training phase, the classifier is evaluated on the test set with the following metrics: False Positive Rate (FPR), False Negative Rate (FNR), and Area Under the Receiver Operating Characteristic (ROC) Curve (AUC). To achieve minimum variance and acceptable bias within the acceptable calculation costs, the hyperparameters and the evaluation metrics of a model are iteratively updated.

The training data used here is unbalanced as the ratio of negatives over positives is less than one. At least two configurations of training data are tested for each model: inclusion of 70% or 80% of all positives, inclusion of 40% of all negatives.The first five models listed below can accept only one class in the training set, avoiding the production of faulty seals. As such, they are also tested without any negatives in the training data.

Seven models of unbalanced classifiers are compared in the June 2017 dataset:

- One Class Classification – Support Vector Machine (OCC-SVM)
- Isolation forest
- Local outlier factor
- Robust covariance matrix estimation
- Deep Support Vector Data Description (Deep SVDD)
- Unbalanced Convolutional Neural Network (CNN)
- Unbalanced SVM

Each model, except Deep SVDD, is fed with the PC, mean, variance, and energy of pre-processed images as described earlier.

5.2 Results and discussion

An interesting observation can be inferred from three models: OCC-SVM, Deep SVDD and Unbalanced SVM. Other models give less satisfactory results and are not discussed here.

The Table 3 sums up the results:

Model	False Positive Rate (FPR)	False Negative Rate (FNR)
OCC-SVM w/o data augmentation	0	>50
Deep SVDD w/o data augmentation	<15	>30
Unbalanced SVM	6.67 to 100 (details in Table 4)	0 to 2.56 (details in Table 4)

Table 3 FPR and FNR of three models

OCC-SVM, from Schölkopf et al [12], is trained by minimizing the FNR. It leads to an FPR of 0% and a high FNR on test data (>50%), regardless of the configuration. With an increase in the quantity of trained data, its FNR tends to decrease. Data augmentation methods are evaluated, but they do not compensate for the lack of data.

Deep SVDD, from Ruff et al [13], is trained by minimizing the mean distance of each training data to a hypersphere center in a feature space. The feature space is an output of a CNN whose weights are optimized for the above-mentioned minimization. LeNet is the chosen architecture of CNN. For all the configurations, AUC exceeds 87%. AUC has also been evaluated for an artificial data augmentation of 100 more images for each class. It is observed to reach 97% in case of 80% of positives in training data and inclusion of 0% to 40% of negatives (Figure 8). Though these scores are high, the false and positive-negative rates depend strongly on the set value of the threshold. Deep SVDD outputs for each test data the squared difference between distance from this test data to the hypersphere center and its radius. An outlier should have a positive score, whereas an inlier or positive should have a negative score. Accordingly, the logical threshold should be zero. However, the best threshold found is non-zero and is a function of the configuration of the test, the parameters of the model, and the run. This makes the data classification impossible with the optimum performance metrics observed on the ROC curve.

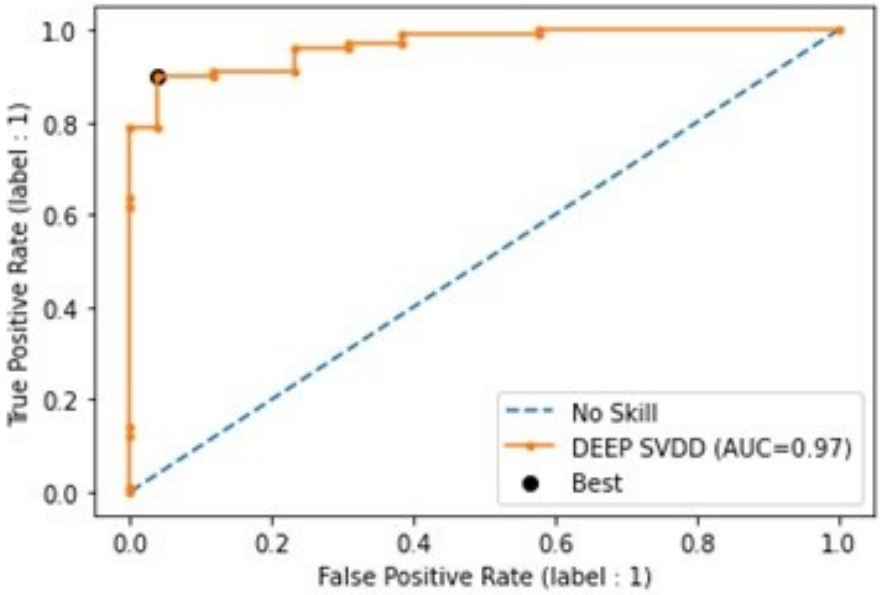

Figure 8 ROC curve and AUC score in deep SVDD (batch size = 3, number of epochs = 150, training data = 80% of positives augmented)

The SVM in which the minority class (negatives here) is given more importance or more weight over the majority class for the definition of the decision function is identified as the unbalanced SVM. Introduced by Yang et al. [14], the penalty term of negatives is fixed higher than that for the positives in the training set. To maximize the evaluation metrics during the classifier building, different class weightings are tuned along an exhaustive search. In this case, the metric does not have much influence on the results: $f_{0.5}$ score, f_0 score, balanced accuracy, and AUC lead to almost similar results.

It is evaluated over five iterations of k-fold cross-validation, where k ranges between 2 to 5 depending on the size of the training set.

Two main types of augmentation were further tested: augmentation of all datasets, augmentation of the minority class of the training set with Synthetic Minority Oversampling Technique (SMOTE), or variants like Borderline-SMOTE, Borderline-SMOTE SVM, and Adaptive Synthetic Sampling (ADASYN).

SMOTE has two arguments tuned for maximization: number of the nearest neighbors of negatives taken to increase the minority class size, and data augmentation ratio.

Some results are listed in Table 4.

Composition of the training set	False Positive Rate (FPR)	False Negative Rate (FNR)	AUC
Without augmentation			
70% positives + 10% negatives	100	0	0.96
70% positives + 40% negatives	6.67	0	0.99
80% positives + 10% negatives	100	0	0.96
80% positives + 40% negatives	6.67	0	0.99
With overall data augmentation and Borderline-SMOTE SVM			
70% positives + 10% negatives	61.8	2.56	>0.95
70% positives + 40% negatives	18.92	2.56	>0.95
80% positives + 10% negatives	60.9	0	>0.95
80% positives + 40% negatives	20.27	0	>0.95
With ADASYN			
70% positives + 10% negatives	Not enough negatives		
70% positives + 40% negatives	10	0	0.97
80% positives + 10% negatives	Not enough negatives		
80% positives + 40% negatives	10	0	0.97 (Figure 9)

Table 4 Performance of Unbalanced SVM after PCA with different training sets and data augmentations

In SVM, the inclusion of negatives in training data is mandatory. In the case of 40% of negatives in the training set (Figure 9), the best results are achieved without data augmentation or with ADASYN. Performance starts to degrade when the ratio of positives over negatives increases in the training set. Out of all seven models, the unbalanced SVM, with its relatively low complexity compared to the neural networks used in a former publication [10], offers the best performance on the dataset.

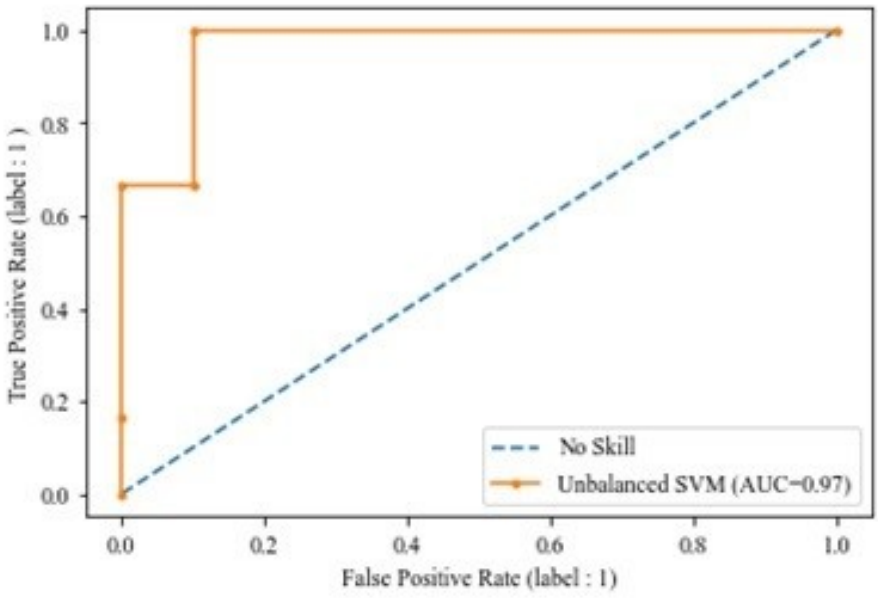

Figure 9 ROC curve and AUC score in Unbalanced SVM (training data = 80% of positives, 40% of negatives augmented with ADASYN)

6 Alleviating the effects of blur

6.1 Inclusion of an iterative deconvolution in workflow

In the above-mentioned ML runs, some negatives remain misclassified for all the models. These negatives are seals with small defects like the inclusion of coffee powder or 1 mm holes.

Possible reasons for this include hardware and numerical limits of the thermal spatial resolution.

Numerically, image resolution is significantly affected by the quantization and all post-processing steps. Raw data in compressed format (jpeg), 8-bit grayscale conversion, and low pass filtering are the key factors that affect the resolution.

On the hardware front, the camera has a number of pixels and a Field of View (FOV), limiting the optical resolution to 0.728 mm. According to this consideration and without taking diffusion into account, a defect of size less than 0.728 mm cannot be detected (see chapter 3). Considering the diffusion, the spreading's area of a defect expands with time making it possible to detect the defects smaller than the optical resolution. This encourages capturing the images after a diffusion time, sufficient for a defect to exceed the thermal sensitivity of the camera. Thus, diffusion helps for some defects smaller than the optical resolution. The initial image can be restored with its defects and without diffusion by a reverse scattering algorithm [15]. Reverse scattering can also deduce heat power sources [16] from the 2D conduction equation and successive thermal frames. This can help if the capture time is not optimal. However, its numerical cost can prevent it from being used for quick control.

The other hardware limit is due to the optical imperfections: diffraction and aberrations. Diffraction causes an airy disk of diameter between 24 and 42 um, depending on the wavelength. This effect is seen on the focal plane array with the chosen optics and configuration (f-number of 1.3, focal length of 18 mm, diameter aperture of 47 mm, and distance image to the lens of 400 mm). Similar to diffusion, the initial image without diffraction can be restored by reversing the Fresnel approximation of the Huygens-Fresnel model, which is valid in this case. However, if the pixel size exceeds 21 um (42/2 = 21 um), diffraction is transparent on the picture according to the Rayleigh criterion. Inverse diffraction does nothing more than distorting the image. Also, the method of inversion should bother the inherent noise [17].

Aberrations models are more complex. A way to model the aberrations (and diffraction) is the use of the impulse response of the optics, known as the Point Spread Function (PSF). Output image after the optics is defined as the convolution of the input image and the PSF. However, this definition is valid only when the PSF is a function of the difference between input and output plane coordinates. This assumption is known as "shift-invariance". Over certain regions of the input plane, high-quality optical systems are often considered as "shift-invariant ". Input image restoration can be deduced directly by applying a 2D inverse Fourier transform of the PSF to the output. However, this process amplifies the noise at high frequencies, which makes the inversion unreadable. Deconvolution of the PSF is instead made with models smoothing the effect of noise, like the Wiener filter [18] or the Richardson-Lucy algorithm [19]. Some models need to identify the PSF and eventually the noise. Others, known as blind deconvolutions [20], cope without the knowledge of PSF.

In this work, to enhance the spatial resolution, the Gaussian low pass filter is canceled and a PSF deconvolution is undergone on all the datasets with the Richardson-Lucy algorithm.

PSF is not measured; instead, the expression of the PSF is estimated iteratively to maximize the performance of the classifier on the training data. Once the optimal PSF is identified, its deconvolution is applied to the test data prior to its classification (Figure 10). The shape of PSF is considered Gaussian. Optimization consists then of finding its size and standard deviation.

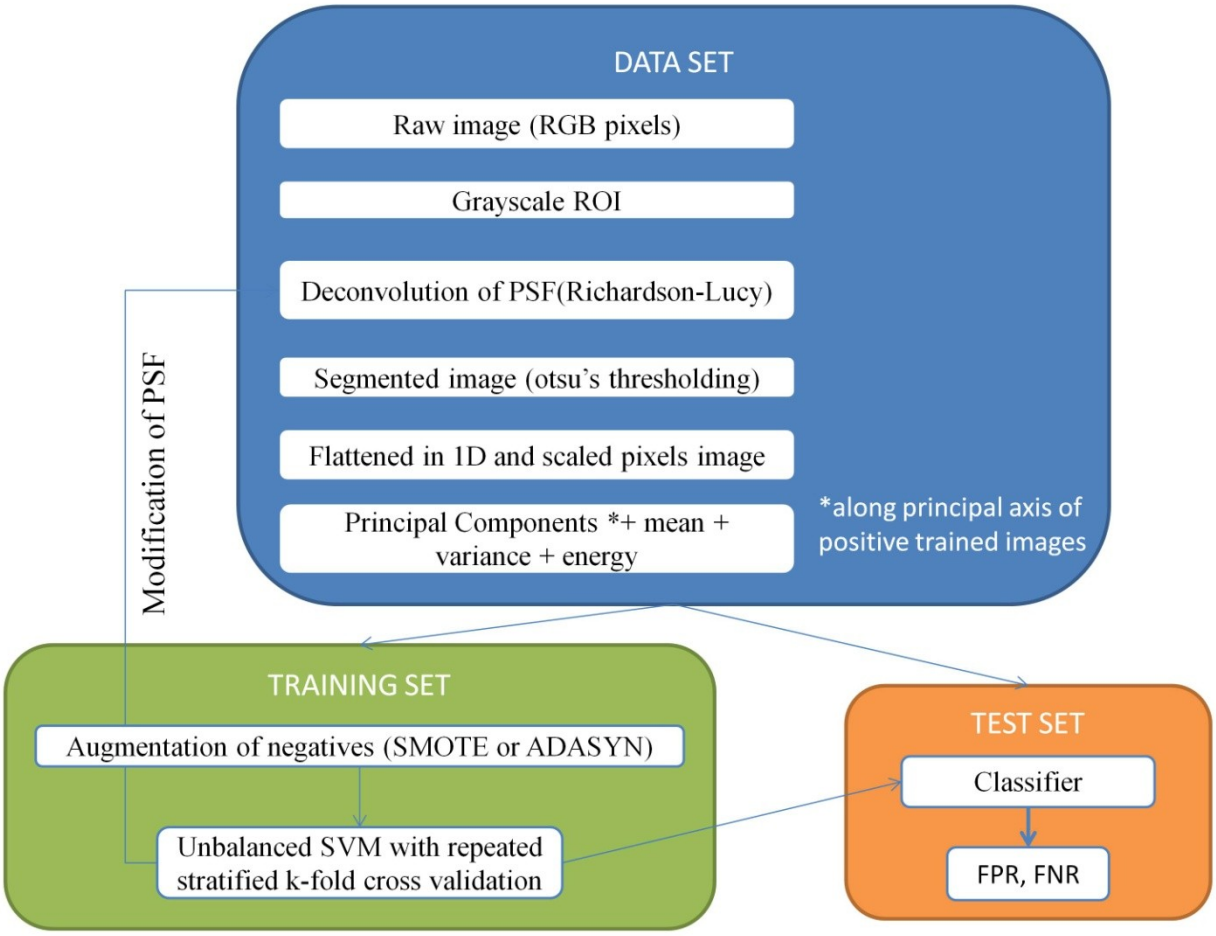

Figure 10 Pipeline of classification with iterative PSF deconvolution

6.2 Results and discussion

With the above-mentioned pipeline, the test sets of June and July 2017 datasets are classified. FPR and FNR are null in both cases, with 30% or 40% of all negatives and 80% of all positives in the training set (Table 5). FPR is a bit higher, 5 to 9%, in the case of the June 2017 dataset and inclusion of 10 or 20% of negatives. With deconvolution, the results are overall improved compared to previous ones without.

Model	Composition of the training set	False Positive Rate (FPR)	False Negative Rate (FNR)	Accuracy
Dataset of June 2017 (test setup described in chapter 3)				
Unbalanced SVM + SMOTE	80% positives + 40% negatives (ratio = 2.5)	0	0	1
Unbalanced SVM + SMOTE	80% positives + 30% negatives	0	0	1
Unbalanced SVM + SMOTE	80% positives + 20% negatives(ratio = 5)	5	0	0.96
Unbalanced SVM + SMOTE	80% positives + 10% negatives (ratio = 10)	9.1	0	0.93
Dataset of July 2017(test setup described in chapter 3)				
Unbalanced SVM + SMOTE	80% positives + 40% negatives	0	0	1
Unbalanced SVM + SMOTE	80% positives + 30% negatives	0	0	1
Unbalanced SVM + SMOTE	80% positives + 20% negatives (ratio = 3.1)	0	0	1
Unbalanced SVM + SMOTE	80% positives + 10% negatives (ratio = 6.3)	0	0	1

Table 5 Performance of the SVM classifier with PSF deconvolution. Ratio = number of positives per number of negatives in training data

By averaging the test times of ten random images, the time needed to classify one single image is evaluated to be 57 ms. It considers:

- the second training configuration as listed in Table 5.
- the inclusion of all prior treatments of Figure 10 (grayscale conversion, deconvolution, PCA).
- execution with a 1.9 GHz processor Intel® Core™ i7 running on Windows 10 operating system.

Integration in a packaging machine should account for the additional time required to capture, amplify, serialize, digitize, compress an image, and transfer it to the hard drive where the classification is done. For an uncooled camera with a USB 2.0 transmission interface and a compression to jpeg format, the frequency of all these operations can reach 20 to 30 Hz, resulting in a maximum of 50 ms per frame.

The overall time of control of seals defects results in 107 ms, enabling integration in the packaging process with throughputs until 33,500 packages/h. This fits the majority of the current applications. It must be noted that the time required to physically accept or reject a package on an actual production line is not counted in the throughput calculation as these operations are performed after the control step on the same line. Also, the overall time is conservative; it can be significantly boosted by code optimization and up-to-date hardware (fast multi-GPUs, high-speed camera).

7 Conclusion

To support the quality assurance of industrial and critical processes, seal defect detection is an essential utility. The Infrared non-destructive test is a low-cost and practical solution to integrate on a production line. For obtaining reliable results, conventional methods based on statistics need a thorough experience on defects characterization. On the contrary, machine learning methods avoid this condition, and their performances have been proven to specific cases but at the cost of high computational expenses. In this paper, different machine learning approaches are evaluated. Weighted SVM combined with dimensionality reduction and minority data augmentation methods is observed to give correct results for two practical cases of small and unbalanced training sets containing a minimum ratio of negatives. Improved performance is finally achieved by inserting an iterative PSF deconvolution of images in the ML pipeline. Accuracy is over 93% for both datasets, achieving 100% with sufficient negatives in the training sets. This approach has a low computational cost compared to deep learning models, thereby providing faster anomaly detection.

8 Acknowledgment

This paper and the research involved in it would not have been possible without the initial support of Fredéric Roumanet. The author is also grateful to the company Thimmonier that has provided financial support to build the dataset and conduct initial studies on the subject.

9 Data availability statement

The image datasets and code that support the findings of this study are available upon request from Thimonnier. Nevertheless, certain restrictions apply to the availability of some licensed data used for this study.

References

[1] S. Benamou, “RF energy console including method for vessel sealing,” Oct. 20, 2015 [Online]. Available: https://patents.google.com/patent/US9161813B2/en

[2] D. A. Peter, “Ultrasonic welding control,” May 24, 1988 [Online].Available: https://patents.google.com/patent/US4746051A/en

[3] G. Rusch and K. N. Murphy, “Systems and methods for detecting syringe seal defects,” Mar. 26, 2019 [Online]. Available: https://patents.google.com/patent/US10242437B2/en

[4] E. Umezaki, H. Yamaguchi , K. Futase, H. Aoki and Y. Kamata “Evaluation of Heat-Seal Quality of Liquid Packaging Bags Made of Plastic Films by Infrared Thermography”, *www.ndt.net*. https://www.ndt.net/apcndt2001/papers/902/902.htm

[5] N. Kenji, N. Masaru, and M. Noriaki, “Method and apparatus for inspection of sealed part with container-opening sealing material,” Jul. 16, 1996 [Online]. Available: https://patents.google.com/patent/JPH08184571A/en

[6] A. Al-Habaibeh, F. Shi, N. Brown, D. Kerr, M. Jackson, and R. M. Parkin, “A novel approach for quality control system using sensor fusion of infrared and visual image processing for laser sealing of food containers,” *Measurement Science and Technology*, vol. 15, no. 10, pp. 1995–2000, Aug. 2004, doi: 10.1088/0957-0233/15/10/008.

[7] S. A. Morris, “Detection and Characterization of Package Defects and Integrity Failure using Dynamic Scanning Infrared Thermography (DSIRT),” *Journal of Food Science*, vol. 81, no. 2, pp. E388–E395, Dec. 2015, doi: 10.1111/1750-3841.13178.

[8] K. D’huys, W. Saeys, and B. De Ketelaere, “Active Infrared Thermography for Seal Contamination Detection in Heat-Sealed Food Packaging,” *Journal of Imaging*, vol. 2, no. 4, p. 33, Nov. 2016, doi: 10.3390/jimaging2040033.

[9] B. J. Papež, M. Palfy, and Z. Turk, “Infrared Thermography Based on Artificial Intelligence for Carpal Tunnel Syndrome Diagnosis,” *Journal of International Medical Research*, vol. 36, no. 6, pp. 1363–1370, Dec. 2008, doi: 10.1177/147323000803600625.

[10] S. Cruz, A. Paulino, J. Duraes, and M. Mendes, “Real-Time Quality Control of Heat Sealed Bottles Using Thermal Images and Artificial Neural Network,” *Journal of Imaging*, vol. 7, no. 2, p. 24, Feb. 2021, doi: 10.3390/jimaging7020024.

[11] “Pyromètre ou caméra thermique ?,” *www.flir.fr*. https://www.flir.fr/discover/rd-science/temperature-guns-versus-thermal-imaging-technology/ (accessed Jan. 20, 2021).

[12] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, “Estimating the Support of a High-Dimensional Distribution,” *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001, doi: 10.1162/089976601750264965.

[13] L. Ruff *et al.*, “Deep One-Class Classification,” *proceedings.mlr.press*, Jul. 03, 2018. http://proceedings.mlr.press/v80/ruff18a.html

[14] X. Yang, Q. Song, and A. Cao, “Weighted support vector machine for data classification,” *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, doi: 10.1109/ijcnn.2005.1555965.

[15] D. J. Crowther, L. D. Favro, P. K. Kuo, and R. L. Thomas, “Inverse scattering algorithm applied to infrared thermal wave images,” *Journal of Applied Physics*, vol. 74, no. 9, pp. 5828–5834, Nov. 1993, doi: 10.1063/1.354202.

[16] S. D. Holland and J. Renshaw, “Physics-based image enhancement for infrared thermography,” *NDT & E International*, vol. 43, no. 5, pp. 440–445, Jul. 2010, doi: 10.1016/j.ndteint.2010.04.004.

[17] D. N. Ghosh Roy, J. Roberts, M. Schabel, and S. J. Norton, “Noise propagation in linear and nonlinear inverse scattering,” *The Journal of the Acoustical Society of America*, vol. 121, no. 5, pp. 2743–2749, May 2007, doi: 10.1121/1.2713671.

[18] N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time series : with engineering applications*. Mansfield Centre, Ct: Martino Publishing, 2013.

[19] W. H. Richardson, “Bayesian-Based Iterative Method of Image Restoration,” *JOSA*, vol. 62, no. 1, pp. 55–59, Jan. 1972, doi: 10.1364/JOSA.62.000055.

[20] E. Y. Lam and J. W. Goodman, “Iterative statistical approach to blind image deconvolution,” *Journal of the Optical Society of America A*, vol. 17, no. 7, p. 1177, Jul. 2000, doi: 10.1364/josaa.17.001177.