Modeling and Classification of Defects in CFRP Laminates by Thermal Non-Destructive Testing

R. Marani[1,*], D. Palumbo[2], V. Renò[1], U. Galietti[2], E. Stella[1], T. D'Orazio[1]

[1]Institute of Intelligent Systems for Automation, Italian National Research Council, via Amendola 122 D/O, Bari, Italy

[2]Department of Mechanics, Mathematics and Management, Politecnico di Bari, viale Japigia 182, Bari, Italy

*Corresponding Author: marani@ba.issia.cnr.it

Abstract

Pulsed thermography has been used for many years to investigate the presence of subsurface defects in composite materials for aeronautics. Several methods have been proposed but only few of them include a complete automated approach for the effective defect characterization. This paper presents a novel method which approximates the thermal decays on the laminate surface, induced by a short heat pulse, by means of an exponential model in three unknowns (model parameters), estimated in the least squares sense. These parameters are discriminant and noise-insensitive features used to feed several classifiers, which are trained to label possible defects according to their depths. Experimental tests have been performed on a carbon-fiber reinforced polymer (CFRP) laminate having four inclusions of known properties. The comparative analysis of the proposed classifiers has demonstrated that the best results are achieved by a decision forest made of 30 trees. In this case the mean values of standard and balanced accuracies reach 99.47% and 86.9%, whereas precision and recall are 89.87% and 73.67%, respectively.

Keywords

Defect classification; feature extraction; exponential model; CFRP; pulsed thermography; decision forest.

1. Introduction

Carbon fiber reinforced plastic (CFRP) is a composite material widely used in many applications in the automotive and aerospace fields. As clearly demonstrated in the literature [1], the presence of defects can dramatically change the strength of the CFRP material and so, non-destructive

techniques (NDT) play a key role for estimating the residual strength of structures or components. In this regard, many NDT techniques are currently used for defect detection in composites such as X-ray [2],[3], ultrasound [4]-[8], shearography [9], vibration testing [10] and electrical potential technique [11] and in many applications a combination of them is necessary to identify different kinds of defects to quantify.

Infrared thermography - Non Destructive Technique (IR-NDT) presents peculiarities suitable for investigation of large areas [1],[12],[13] since i) it does not require the coupling with the component, ii) it is easily automatable and iii) the testing time is relatively shorter with respect to other well-established NDT techniques. With regards of testing setup and data processing, the most diffused thermographic techniques for the non-destructive evaluation of composite materials are: Lock-in Thermography (LT) and Pulsed/Stepped Thermography (PT) [14]. Further techniques use infrared thermography in combination with acoustic emission (AE) to identify damage evolution in carbon fiber reinforced composites [15]. In any case, all techniques use an external heat source (mechanical [12],[13], electromagnetic [1],[16], and optical [8]) to stimulate with thermal waves the material in order to induce a heat flux in the material and analyse the temperature behaviour on the surface of the component.

Among IR-NDT, Pulsed Thermography (PT) is largely used for inspecting composite materials. In particular, PT employs a short thermal stimulation to produce a thermal perturbation within the material. The presence of a defect is revealed by monitoring the surface temperature decay of the specimen. The defect appears as a limited area of different temperature with respect to a surrounding sound area: the subsurface defect produces an abnormal behavior of the temperature decay curve. In the last years, different algorithms associated with PT technique have been developed with the aim to extract information about defects [17]. Among them, the Difference Temperature Signal [14], the Thermographic Signal Reconstruction [18],[19] and the Polynomial Fitting [20],[21] can enhance thermal signals, thus allowing the detection of deep defects, even with high noise levels. On the other hand, most of them are limited to the mere segmentation of defects and are not able to infer their complete properties (defect characterization).

Since the last decades, researchers have investigated the possibility of using machine learning approaches for both the detection and the classification of internal defects, inspected by NDT techniques. Artificial Neural Networks (ANNs), have been proposed for the automatic analysis of results of inspections performed with ultrasonic techniques in [22]-[24] and active infrared thermography in [25],[26]. In these cases, supervised techniques require a learning phase during which a number of examples of the observable phenomenon are used to build the model of the

different defects that has to be recognized. Different kinds of impact damages are recognized in [27] by active thermography. Distinguishing features are extracted from the thermographic signals of damaged parts in the image and related to the impact energy with an ANN in order to model the entity of the impact. An unsupervised approach has been used in [28] to evaluate the reliability of sound regions by a clustering technique which aggregates similar thermal data belonging to the same kind of regions. In this case, the number of target classes is provided by the user to group surface regions which share the same probability of being defective.

According to our knowledge, thermal maps and videos are typically treated as standard images by classical machine methods and algorithms. As an example, image processing tools, such as two-dimensional median filters, histogram equalization, contrast enhancement, standard feature spaces (e.g. Scale Invariant Feature Transform, SIFT [29], or Speeded Up Robust Features, SURF [30]), are often used to enhance images and highlight defects. However, thermal imaging follows well-known physical laws, whose understanding can further improve results: feature representations with physical meanings can be much more informative than standard processing. Besides, a comparative analysis of different machine learning methodologies, able to explain the relation between features, classifiers and performances, has not been applied before in this context.

This paper presents a novel approach for automatic defect detection and classification in composite materials by pulsed thermography imaging. The main novelties of the proposed approach are summarized in the following points.

- First of all, new features are extracted from the thermographic signals, in order to model the differences among different defects and pristine regions. In particular, following the principles of pulsed thermography, local transient temperature decays, collected by a thermocamera, are normalized and compared with those expected in pristine regions. Resulting temperature contrasts are modeled by robust and compact features whose analytical meaning is directly related to the presence of defects buried in the composite.
- Then, several machine learning algorithms have been selected. The extracted features are thus used to train the classifiers to detect defects depending on their depths.
- A comparative analysis has been performed on a known specimen made of carbon fiber, pre-preg epoxy resin, which shows few inclusions of foreign objects of known geometry and material, used to simulate interlaminar delaminations. Results demonstrate that the use of the proposed exponential-model-based features, together with a decision forest of 30 trees, can correctly detect and classify defects in accordance with their depths within the sample under testing.

The paper is organized as follows: Section 2 briefly describes the experimental method used to provide the thermal data, the feature extraction pipeline with reference to the analytical model, and the classifiers which will be used to characterize defects; Section 3 deals with the experiments performed on an actual sample, whereas final conclusions and remarks are in Section 4.

2. Methodology

The proposed method aims to classify CFRP laminates depending on the presence of defects and to characterize them in terms of their properties. This purpose is achieved starting from experimental analyses performed by pulsed thermography. Then robust and discriminative features are defined to describe exhaustively the response of the composite under testing, thus enabling the classification of any possible defect within the medium. The next subsections will describe the whole pipeline to reach the complete classification of flaws, according to their depths.

2.1. Pulsed thermography

The investigation of CFRP is performed by using active pulsed thermography [31]. This technique is based on the analysis of the thermal behavior of the target specimen in response to a short (few milliseconds) thermal pulse. Pristine semi-infinite and isotropic media produce a transient decay, after a Dirac pulse, which follows the well-known equation:

$$T(x,t) = \frac{Q}{\sqrt{\pi\rho ckt}} \exp\left(-\frac{x^2}{4\alpha t}\right) \tag{1}$$

where x represents the depth coordinate, t is time and T is the temperature value. ρ [kg/m^3], c [J/kg K] and k [W/m K] are properties of the specific material under testing, respectively equal to mass density, heat capacity and thermal conductivity, whereas the term α [m^2/s] is the thermal diffusivity, equal to the ratio k/ρc. Q [J/m^2] takes into account the source energy transferred to the specimen surface.

Pulsed thermography applied for the analysis of finite-thickness materials can be exploited in two different configurations: reflection or transmission configuration. With reference to the sketch in Fig. 1, the proposed method will be defined in case of reflection configuration, which requires that both input lamps and receiving thermocamera are positioned on the same side with respect to the specimen. Here, d_L and d_T are the specimen-lamps and specimen-thermocamera distances, respectively, whereas β is the angle enclosed by the camera optical axis and the dominant light direction. Although this configuration shows lower resolution for deeper defects, as heat has to

travel within the specimen for a round trip, it is the most suitable for in-line testing, since a single system can both heat up the surface and scan the response.

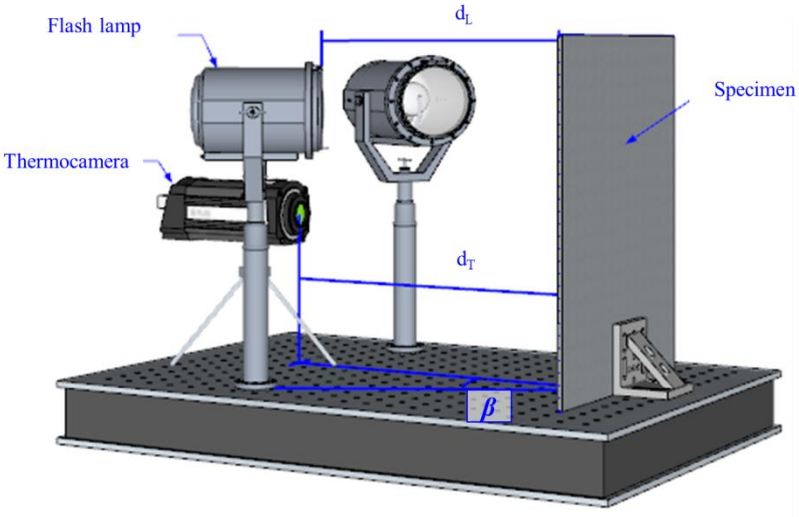

Fig. 1. Sketch of a setup for pulsed thermography.

In this case the analytical model in Eq. (1) can be reduced to the one of Eq. (2), since thermocamera can map the top surface of the specimen, producing an image of size I×J.

$$T(0,t) = \frac{Q}{\sqrt{\pi\rho ckt}} \tag{2}$$

Following the simplified models in [32],[33], the presence of an interlaminar delamination within the medium alters the transient temporal decay of temperature, depending on the difference of effusivities $\varepsilon = \sqrt{\rho ck}$ [$Wm^{-2}K^{-1}s^{1/2}$] between the defect and the background material. Specifically, the collected thermal decay is [32]:

$$T(0,t) = \frac{Q}{\sqrt{\pi\rho ckt}}\left[1 + 2\sum_{n=1}^{\infty} R^n \exp\left(-n^2 \frac{d^2}{\alpha t}\right)\right] \tag{3}$$

where d [m] is the depth of the discontinuity due to the presence of the defect (e.g. delaminations or foreign objects), R is the reflection coefficient related to the difference of effusivities of the two involved media, and α refers to the sample material. Eq. (3) shows a summation term due to the presence of multiple internal reflections of the travelling heat wave impinging on the interfaces between the different media of the whole air-composite-defect arrangement. Moreover, the inspection of Eq. (3) reveals that the heat wave propagates faster in regions where defects are at lower depths d (or of higher diffusivity α) than the case of defects at higher depths (or in regions of

lower diffusivity). This phenomenon is of great importance for the feature extraction process which will follow in the next sections.

At this stage, it is important to notice that the one-dimensional simplified model of Eq. (3) well fits only the behavior of isotropic media and, consequently, it can lead to misleading interpretations of the defects depth buried in finite actual samples. Nevertheless, this model gives a realistic approximation of thermal propagation if defects are larger than their depths. These conditions are verified for the analysis of targets made for aeronautics. In this context, actual mechanical tolerances impose the detection and classification of flaws having sizes of few millimeters, in panels having thickness which rarely exceeds 3 mm.

In addition, this hypothesis further justifies the use of pulsed thermography to provide the best results [34], since the finite thickness of these panels meets the capability achievable in reflective configuration. Here, thickness values are within the limits of penetration of the input thermal wave.

2.2. Analytical representation

As a result of the application of Eq. (3), the surface of a defective region shows a temporal decay of the temperature which is directly linked to the constitutive material of the defect and to the specific depth at which it arises. Specifically, a defect beneath the surface alters the standard linear profile in natural logarithmic scale (with slope -0.5), leading to the thermal decays in Fig. 2(a), which depend on the ratio d^2/α. In these examples, the constant ratio $Q/\sqrt{\pi\rho ck}$ has been kept to unity for simplicity and R is equal to 0.5.

As shown previously, the discrepancy between the expected linear profile and the one due to the presence of the defect is linked to the properties of the inhomogeneity, in terms of position in depth and thermal effusivities of both materials. The aim of testing laminates is the detection of defects and their classification, dividing target sets, namely defects, in several categories which differ in geometrical properties. To achieve this task, input thermal behaviors, captured from the top surface of the specimen by a thermocamera, are treated to further enhance alterations with respect to the reference profile. Specifically, at this stage input enhancement is obtained by normalizing the profiles, i.e. enclosing their dynamics in the range between 0 and 1. This transformation can be easily obtained as:

$$T_{norm}(0,t) = \frac{T(0,t) - T_0}{T(0,t_\delta) - T_0} \tag{4}$$

where T_0 is the minimum value of temperature, corresponding to the last observation of the target at the time instant t_{tot}, and $T(0,t_\delta)$ is the temperature at the end of the heat pulse excitation, namely the maximum temperature level reached by the top surface of the composite. The result of this normalization is displayed in Fig. 2(b) where the normalized profiles are gained starting from the same input thermal decays of Fig. 2(a).

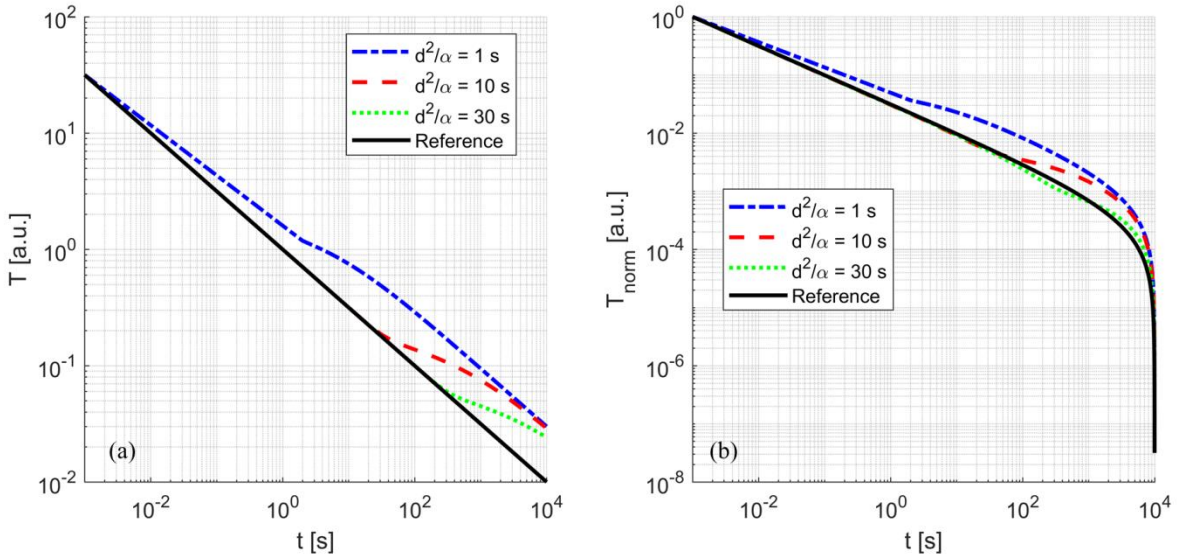

Fig. 2. (a) Transient temporal decay of the surface temperature of a semi-infinite medium heated by a pulsed-source and (b) corresponding normalized temperature curves. Profiles are obtained by changing the ratio d^2/α, linked to the presence of a reflecting defect at depth d. The reference decay is $T = t^{-1/2}$.

The use of the temperature normalization has a twofold advantage:

1. Increase discrimination: the normalized temperature decays are varied so that their levels can be over or below the reference at specific time instants: the presence of a defect not only changes how fast the material is locally heated up, but also alters the normalized temperature level over or below the reference. In particular, deeper defects can apparently "cool down", in terms of normalized temperature, the laminate faster than the reference if normalized behaviors are considered (see the green dotted curve of Fig. 2(b)). It is worth noticing that separation among profiles can be solved with different resolution depending on the extension of the time range of observation [t_δ, t_{tot}]: the shorter time is considered, the higher resolution is achieved. On the other hand, reducing the observation range can delete information on the presence of deeper inhomogeneity;
2. Independence of the properties of the media: normalization adjusts temperature dynamics to a common scale, thus making the following steps insensitive to the specific properties of the considered material. As a consequence, there is no need to exactly characterize the media, since the response of the composite will be bounded within known limits.

Normalized temperatures are then processed to highlight defect contrast ΔT, which is the result of the difference between the responses of defective and non-defective regions. By subtracting the normalized reference to the normalized temperature profile, after simple algebra, it is possible to obtain the following:

$$\Delta T\left(0,t\right)=Q\sqrt{\frac{t_\delta}{t}}\frac{\Sigma_t\left(Q-T_0\varepsilon\sqrt{\pi t_\delta}\right)-\Sigma_{t_\delta}\left(Q-T_0\varepsilon\sqrt{\pi t}\right)}{\left(Q-T_0\varepsilon\sqrt{\pi t_\delta}\right)\left(Q\left(1+\Sigma_{t_\delta}\right)-T_0\varepsilon\sqrt{\pi t_\delta}\right)} \tag{5}$$

where $\Sigma_t=2\sum_{n=1}^{\infty}R^n\exp\left(-n^2\frac{d^2}{\alpha t}\right)$ and $\Sigma_{t_\delta}=2\sum_{n=1}^{\infty}R^n\exp\left(-n^2\frac{d^2}{\alpha t_\delta}\right)$. Fig. 3(a) presents the thermal contrasts corresponding to temperature transient profiles in Fig. 2(b).

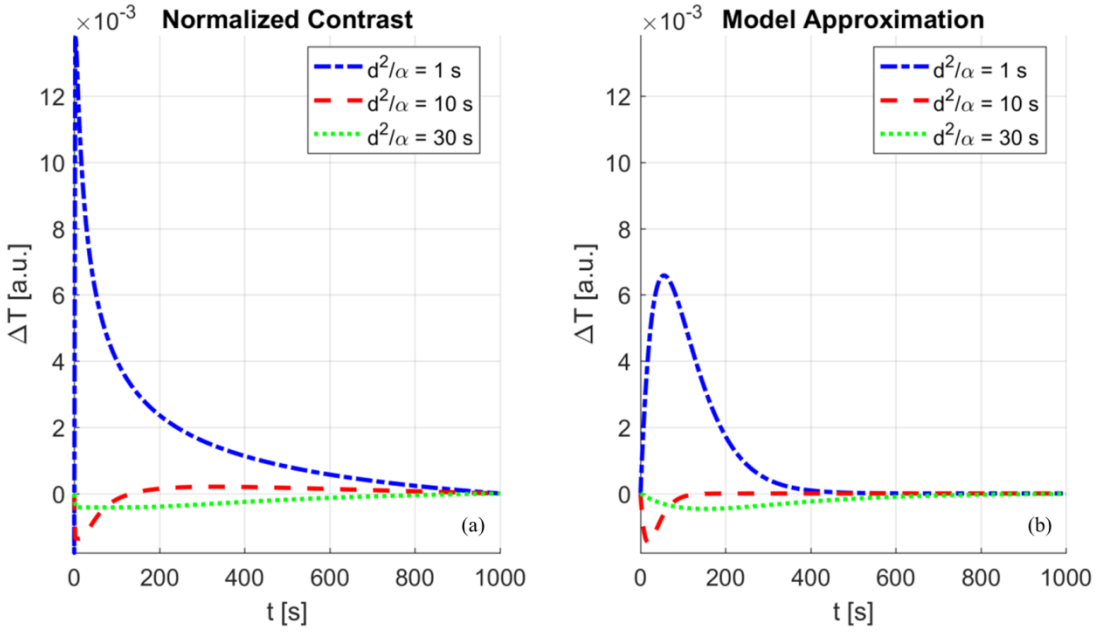

Fig. 3. (a) Thermal contrast behavior related to the normalized temperature profiles after the source pulse as a function of d^2/α. (b) Corresponding model approximation for feature extraction.

As shown previously, the use of temperature normalization enhances discrimination between the different contrasts which can be positive or negative, depending on the value of the ratio d^2/α.

The proposed methodology for defect classification tries to group transient contrasts by looking at their time evolutions, which are directly related to the depth of the possible defect beneath the surface of the laminate. For this reason, a simple exponential model is developed to approximate all thermal contrasts, including the trivial one, which corresponds to homogeneity. Specifically, the analytical model approximates the thermal contrast as:

$$\Delta T_{approx}\left(0,t\right)=\left(a_1t^2+a_2t\right)\exp\left(-a_3t\right) \tag{6}$$

being $a_1 \geq 0$ and $a_3 \geq 0$. The estimation of the three parameters in the least squares sense leads to the plots in Fig. 3(b), which approximate the corresponding thermal contrasts in Fig. 3(a).

As shown in Fig. 3, this simplified model is able to reproduce the whole profile of the normalized thermal contrast by using only three unknowns, which include the contributions of all parameters in Eq. (5). As a consequence, the possible presence of subsurface defects can be easily estimated by the inspection of the values assumed by the three parameters a_1, a_2 and a_3, computed in every point captured by the thermocamera. The three unknowns a_1, a_2 and a_3 are the starting discriminant features for effective material assessment.

In addition, it is worth noticing that the use of least square fitting of the model in Eq. (6) on actual input decays also ensures the reduction of the effects of measurement noise, which can alter significantly the results of surface labeling. As a consequence, defect classifiers can perform with increasing robustness with respect to the negative contributions of actual noise.

Further and deeper discussions on feature extraction processing are in the next subsection.

2.3. Feature processing

As shown previously exponential-model-based features are directly linked to the values of the three parameters a_1, a_2 and a_3 in Eq. (6). These values are obtained as the approximation of the normalized thermal contrast, which in turn is related to the actual transient decay of temperature and to the reference profile in Eq. (2). The former term is the output temperature measured by the camera on the top surface of the specimen, whereas the latter has to be found by a preliminary processing on the acquired data. Since no *a priori* assumption can be stated on the presence of defects, which is the actual unknown within this lines, it is not possible to select a trust area of homogeneity to be used to study the reference response of the constitutive medium. For this reason, the temperature reference $T_{ref}(i,j,t)$ of a pixel of coordinates (i,j) at time t is determined as the median value of temperature measurements belonging to a patch $P_{i,j}$, having custom shape:

$$T_{ref}(i,j,t) = \underset{(i',j')\in P_{i,j}}{median}\left(T(i',j',t)\right) \tag{7}$$

It is possible to take two assumptions:

1. the number of defective pixels within the patch P is much lower than the number of non-defective ones;

2. the source pulse illuminates homogeneously the whole surface of the specimen, or equivalently it is possible to recognize the direction of the thermal gradient produced by the source wave.

The first hypothesis states that the normalized procedures will work only if the defected area is significantly smaller than the patch $P_{i,j}$. In this way the median value of surface temperature can be reliably attributed to sound material. For this reason, the patch $P_{i,j}$ must be larger than the size of the expected defects. Alternatively, the size of P defines a limit for delaminations that can be detected. In this case, wide defects can generate a false positive as the median temperature would be close to the one of the delamination and not to that of the sound material.

Moreover, it is important to notice that the choice of the patch $P_{i,j}$ strictly depends on the experimental setup and can be tuned in agreement with the lamp-sample arrangement. As an example, if no gradients are present along horizontal lines, i.e. the input thermal excitation is homogeneous along the rows of the thermal map, it is possible to choose a patch element $P_{i,j}$ of a flat, linear and horizontal shape. Consequently, the reference thermal profiles are computed along each row of the frames of the thermal video sequence. The normalized reference temperatures can be then arranged in a 2D matrix $T_{ref}(i,t)$ of median decays, having size equal to $I \times t_{tot}$. In particular, the rows of $T_{ref}(i,t)$ are the median profiles related to each row of the video frames at a specific time instant t. The columns of $T_{ref}(i,t)$ report the time development of the normalized reference temperatures for the specific i-th row, till the final observation instant t_{tot}.

Once the reference is defined, it is possible to compute the normalized contrast for every pixel of the thermal images. Then the feature extraction processing can follow the development outlined in Fig. 4.

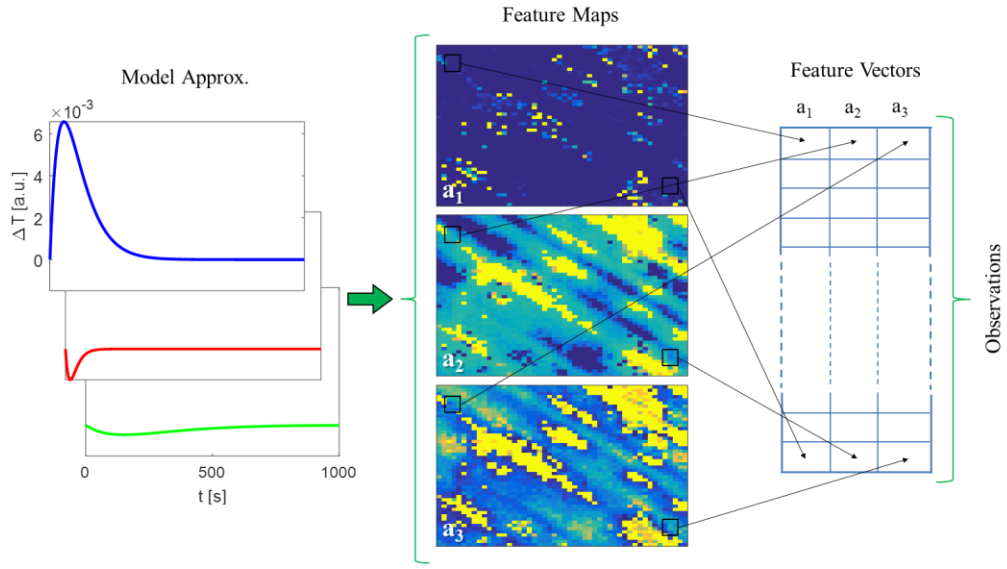

Fig. 4. Feature vectors arrangement from the model approximation through the generation of feature maps. Each line of the final matrix corresponds to a transient temporal decay collected by a pixel of the thermocamera (observation).

Every pixel of the image produces the entries of three feature maps, filled by the values of a_1, a_2 and a_3. Feature maps can be directly used for material characterization or can be further processed to enhance the accuracy and the reliability of defect detection. As discussed previously, thermal videos are mainly affected by measurement noise, which is intrinsically limited by the application of the least square fitting of the model of Eq. (6) on the actual normalized contrasts. Although this process of noise reduction, noise can still alter significantly the results of the model approximation, introducing sparsely occurring impulse of noise within the feature maps (salt and pepper noise). This noise can be effectively contrasted by the application of a simple median filter over a 5×5 neighborhood [35]. Parameters can now be referred in the (i,j)-th pixel of the video frames as $a_1^F(i,j)$, $a_2^F(i,j)$ and $a_3^F(i,j)$.

These results are then rearranged in vectors of features $F(i,j) = [A_1(i,j), A_2(i,j), A_3(i,j)]$. Here $A_1(i,j)$, $A_2(i,j)$ and $A_3(i,j)$ can be equal to the scalar values of the corresponding parameters, either filtered or directly equal to the input, or can be made of blocks of values, having size N×N, obtained by windowing feature maps around the (i,j)-th coordinates. Each feature vector has a length equal to $3 \cdot N^2$. As an example, Fig. 5 shows the process of reshaping feature maps in case of 5×5 neighborhood. The example produces the feature vector F(i,j), referred to the (i,j)-th coordinate of the feature maps.

The use of extended windows of exponential-model-based features can introduce better results since the analysis of local contributions add more information than single point values. In this case classifiers, which will be then responsible for deciding whether a pixel of the map belongs to defect areas, can receive more comprehensive information of the thermal behavior of the whole specimen

surface. As a consequence, defect characterization can be more reliable against sparse noise or at edges, where transitions between homogeneities and defects are estimated considering feature changes.

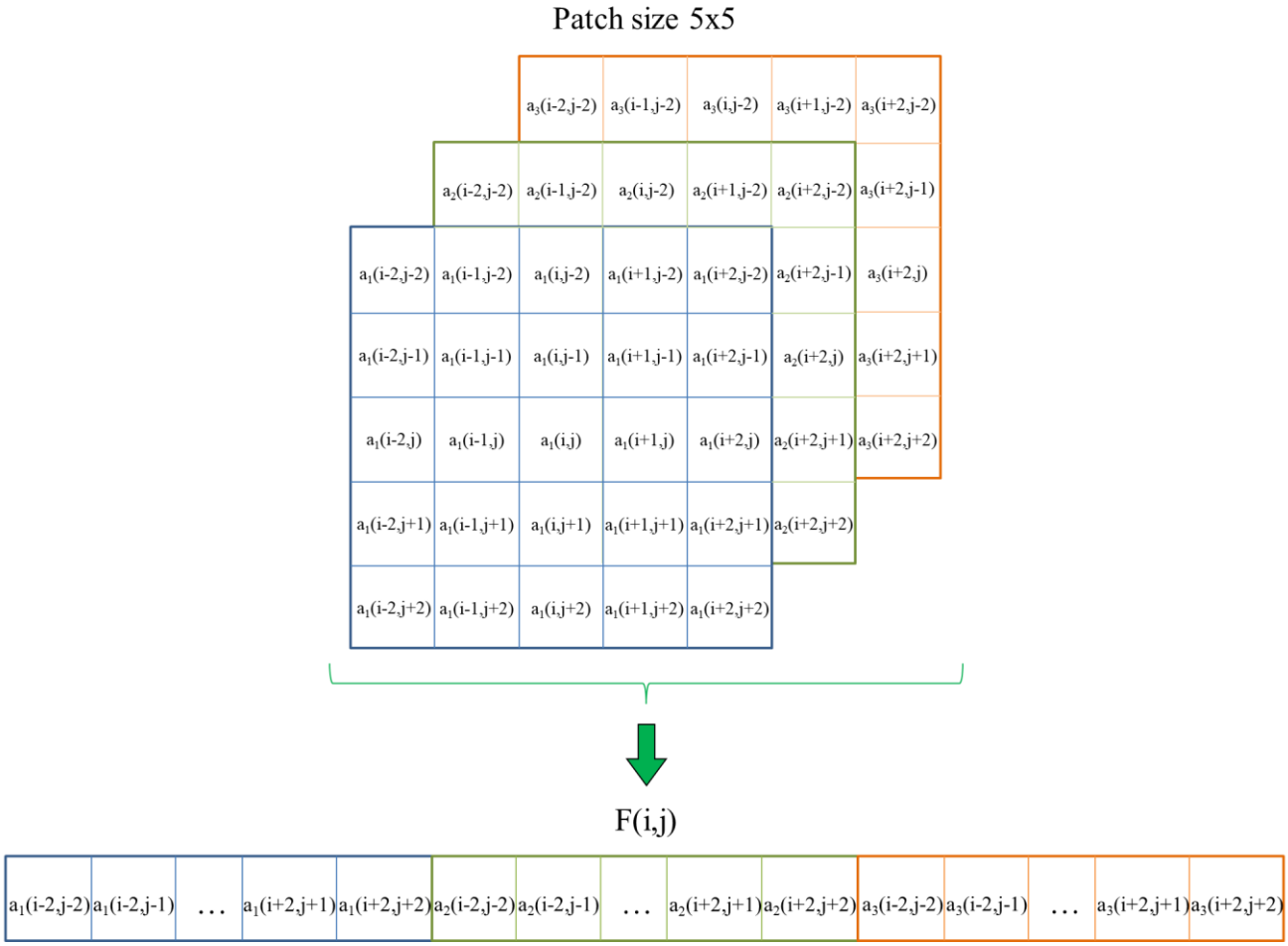

Fig. 5. Single feature vector formation in case of extended patch size. This example substitutes a generic line of the final feature matrix. Neighboring observations can share some entries.

2.4. Classifiers

Exponential-model-based features are input predictors able to characterize the area framed by a pixel (or observation) of the thermocamera. This information is used by classifiers to establish whether there is a defect or not. Furthermore, if a defect is detected, classifiers can also recognize the kind of discontinuity, thus grouping features together. For the next development, three classes will be considered:

- Class 0: Regions of homogeneity;
- Class 1: Surface defects, i.e. defects placed within the medium but close to the top surface, which is illuminated by the thermal pulse;
- Class 2: Internal defects, i.e. buried defects deeper than those of class 1.

For each of these classes an input set of examples, made of couples of known inputs (feature vectors) and expected labels, will be provided to train the classifiers. In this manuscript, the following classifiers will be implemented to manage the input features [36]:

1. Decision tree classifier: Decision trees are graph-based classifiers which label input observations by splitting them at different levels through a threshold mechanism. This process can be traced explicitly as a sequence of decisions, starting from the most significant till the output node, where classification is achieved. As a drawback, small alterations of the input predictors can lead to wrong labeling, since thresholds can be closely related to the specific training set;
2. Ensemble of decision trees: Ensemble of classifiers can provide better results in the decision process. With reference to the case of ensembles of decision trees, also known as decision forests, several trees are trained to produce different labeling on the same input. This approach takes advantage of the instability of the trees in response to slightly altered input predictors, thus reaching a more flexible structure, preventing overfitting;
3. k – Nearest Neighbor (k-NN) classifier: this classifier defines several representations (prototypes), which act as a reference of the three classes of interest. Every input observation is compared with the closer k prototypes and then labeled accordingly with a voting mechanism. Two different variants will be considered in the following experiments:
    a. Standard k-NN: The number of the nearest neighbors is set to k and all prototypes have the same weight in the voting process;
    b. Weighed k-NN: The contribution of the k nearest prototypes is weighted in the voting phase by the inverse of the Euclidean distance between the input observation and the corresponding prototype.

As a result, the trained classifiers will be then able to make a prediction on the behavior of the inspected region of the specimen, thus estimating the presence and the type of defect, accordingly with the preliminary training phase.

3. Experimental results

The proposed methodology has been tested for the inspection of an actual CFRP laminate, in which several foreign objects of known geometries have been placed at specific positions. Exponential-model-based features have been used as input for four classifiers in order to compare results. The next subsections will introduce the experiments and the sample, and a detailed discussion on final results.

3.1. Experiments

3.1.1. Setup description

Fig. 6 shows the experimental setup adopted for thermographic tests. Two flash lamps with a power of 1500 W were used and controlled by a flash power control (Bowens Quadx®) with a max stored energy of 3000 Ws and an approximate flash duration of 1/1430 s. The lamps were placed about to 120 mm from the specimen surface (d_L in Fig. 1). The two lamps are arranged with an angle β equal to 15°.

The FLIR X6540sc [37] thermal image camera with a indium antimonide (InSb) cooled detector (640×512 pixels, NETD < 25 mK) has been used for acquiring thermal sequences with a framerate of 100 Hz. This framerate is achieved by downsizing the sensor area to a region of interest of 384×452 pixels. The IR camera is placed between the flash lamps at a distance d_T = 1500 mm from the specimen in order to obtain a geometrical resolution of about 0.25 mm/pixel over a 100-mm-wide region.

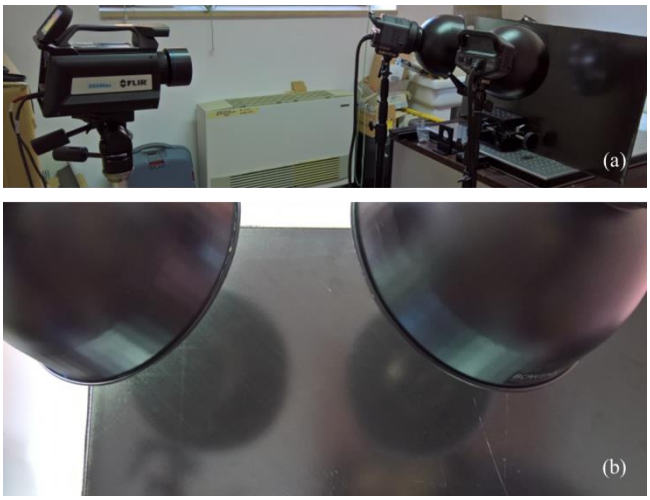

Fig. 6. (a) Picture of the actual setup used for experiments. (b) Camera point of view focused on the specimen.

3.1.2. Sample specification

Defect classification has been tested on a reference sample made of a vertical arrangement of eight layers P_k, k = 1, …, 8, of CFRP taper, consisting of unidirectional carbon fiber impregnated with an epoxy resin (NTA 62470, form 1, Type 35, Class 3, Grade a90), having width and thickness equal to 150 mm and 0.1 mm, respectively. Each layer is made of several plies, whose in-plane orientation is defined accordingly with those reported in Fig. 7(a), in order to obtain a balanced and symmetric laminate with thickness of 0.8 mm. During the specimen manufacture with the vacuum bag process, several foreign objects (red blocks in the sketch of Fig. 7(a)) have been placed between the layers P_4-P_5 and P_7-P_8. These inclusions are different in size and material. Specifically, 0.05-

mm-thick fluorinated ethylene propylene (FEP) and 0.08-mm-thick white paper have been used within these experiments. Further details on the target defects are listed in Tab. 1.

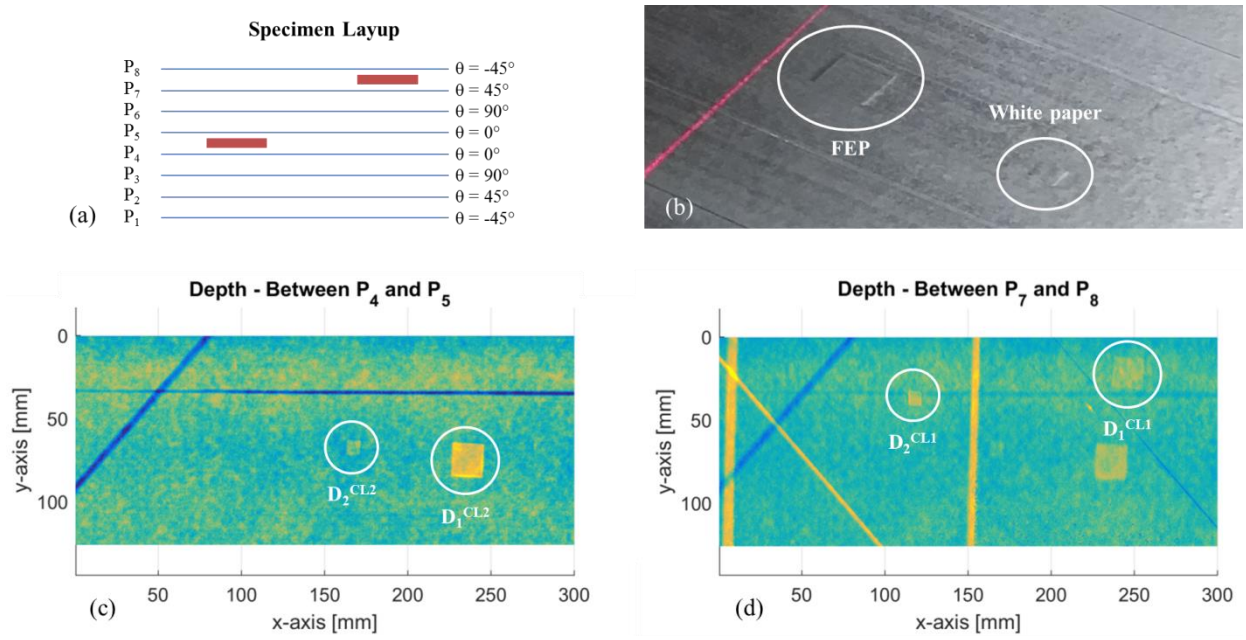

Fig. 7. (a) Detailed layup of the specimen having foreign object inclusions within layers P_4-P_5 (internal defect, class 2) and P_7-P_8 (surface defect, class 1). Also ply orientations (θ) are reported in the scheme. (b) Surface of the specimen before polymerization. Circles are centered on top defects made of FEP and white paper. (c) Depth map acquired after the deposition of P_5 over P_4 which sandwiches two defects of class 2. (d) Depth map resulting by scanning the top layer. Two new defects of class 1 are highlighted.

Tab. 1. List of defects within the specimen and corresponding properties.

Name	Size [mm]	Depth	Material	Class
D_1^{Cl2}	25×25	P_4-P_5	White paper	2
D_2^{Cl2}	12.5×12.5	P_4-P_5	FEP	2
D_1^{Cl1}	25×25	P_7-P_8	FEP	1
D_2^{Cl1}	12.5×12.5	P_7-P_8	White paper	1

Furthermore, the specimen has been preliminary scanned by means of a custom laser profilometer in order to gain information about the shape and the position of every defects beneath its surface. Acquisitions have been performed by using a custom triangulation setup [38],[39] made by a laser source able to generate a line over the surface of the specimen (see the red line in Fig. 7(b)). The laser line is revealed on the sensor plane of a standard camera, which complete the optical probe of the system. As known, a rigid translation of the target, or equivalently every corrugation of the surface, is translated in a shift of the reveled laser spot onto the camera plane. After a first calibration phase it is possible to define exactly the shape of the sample surface, i.e. a three-dimensional model of the surface, with high resolution. In this case, the optical receiver is a Dalsa

Falcon 4M60 [40]. The whole probe is set to reach the resolution of 0.05 mm in depth measurements. 3D models have been acquired after the deposition of every single layer before polymerization. As a result, measurements are able to characterize with high precision all defects, as shown in Figs. 7(c)-(d) which display the depth maps computed after the deposition of P_5 and P_8, respectively. With more details, blue regions of depth maps in Figs. 7(c)-(d) correspond to deeper points, whereas yellow areas are related to points over the surface level. Circles in Figs. 7(c)-(d) enclose the target defects.

It is important to notice that the specimen has a specific texture, due to further artificial defects added during manufacturing. This texture is underlined by the presence of further in-plane valleys, such as the blue stripes in Fig. 7(c), or sharp ridges, such as yellow stripes in Fig. 7(d)). These structures belong to the 0 class of homogeneity, which actually includes all regions other than those of inclusions, labeled in class 1 and 2.

As discussed previously, the thermal setup is able to frame a limited area of the specimen. For this reason the analysis of the whole specimen has been divided in three different steps, namely one used for training the algorithms and the rest for testing them (Test 1 and Test 2).

Once the specimen has been completely scanned and thus modeled in three dimensions, all geometrical properties about each element beneath the surface are thoroughly known. Hence it is possible to build a reference ground truth having exact information about the in-plane position of both classes of defects (see the maps in Fig. 8). This information will be then used for both training and testing phases for the proposed classifiers. In this case the training dataset corresponds to a square region placed at the highest values of the x-axis in Fig. 7(d), since this area covers both classes of defects, thus producing a significant input for the learning phase.

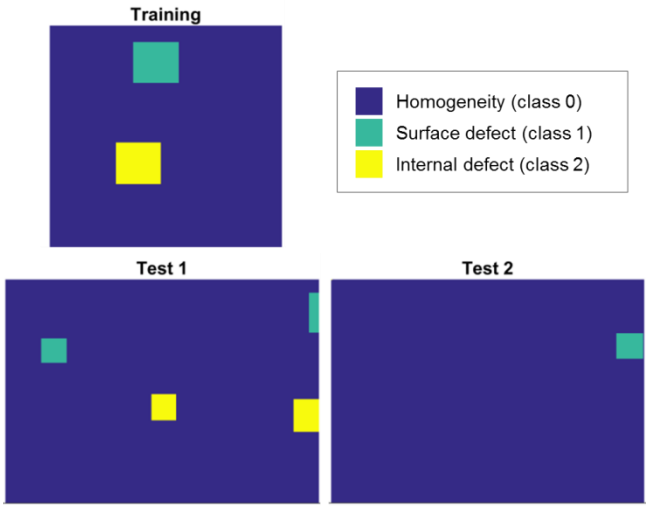

Fig. 8. Ground truth corresponding to the three regions considered for training classifiers and test predictions.

3.1.3. Evaluation metrics

Ground truths are the basis to define the evaluation metrics, which is useful for the comparison of final results. At the end of the classification task, it is possible to compare predictions, namely class assignments, with expected ones, defined by the maps in Fig. 8. Specifically, quantitative tests can be performed by evaluating the following measurements:

$$\begin{cases} BACC_c = \dfrac{\dfrac{TP_c}{TP_c + FN_c} + \dfrac{TN_c}{TN_c + FP_c}}{2} \\ PPV_c = \dfrac{TP_c}{TP_c + FP_c} \\ TPR_c = \dfrac{TP_c}{TP_c + FN_c} \end{cases} \tag{8}$$

where $BACC_c$ is the balanced accuracy, PPV_c refers to the positive prediction value (or precision) and TPR_c is the true positive rate (or recall). In these formulations TP_c, TN_c, FP_c and FN_c are the numbers of True Positive, True Negative, False Positive and False Negative, respectively. All evaluations are obtained for binary classification, i.e. they explain whether the method can correctly label inputs belonging to the c-th class against the others.

With more details, the balanced accuracy estimates the number of correct predictions, weighted by the number of occurrence of the class of interest. Precision is the ratio of correct assignments over all labels of that class, i.e. it shows if all defect detections are true. Then recall is related to the number of missed labelling (FN_c) of observations of the c-th class. In this case it states the ability of the classifier to detect a defect beneath the specimen surface.

It is important to notice that balanced accuracy is preferred with respect to standard accuracy, namely the ratio between true labels over the total number of observations, since datasets are heavily unbalanced. In this case the number of observations of class 0 is much higher than the one of the other classes, since defective regions are much smaller in extension than pristine ones. Using the standard formulation for accuracy might lead to ambiguous results, numerically close to the unity, as effect of the structure of the given dataset. On the contrary, balanced accuracy always produces informative results, since the comparison of output labels with the expected ones has the same weight, regardless the number of occurrences of the class c under evaluation.

3.2. Results and discussion

The pulse-source excitation produces a transient decay of surface temperature, described by the equations discussed previously. The response of the actual specimen has been tested by the

inspection of thermal sequences, as the one reported in Fig. 9, which shows several frames acquired by the thermocamera after the generation of the heat pulse. In this case the heat pulse illuminates the specimen surface at t = 0 s. All images are related to the inspection of the region of Test 1, where all defects are placed.

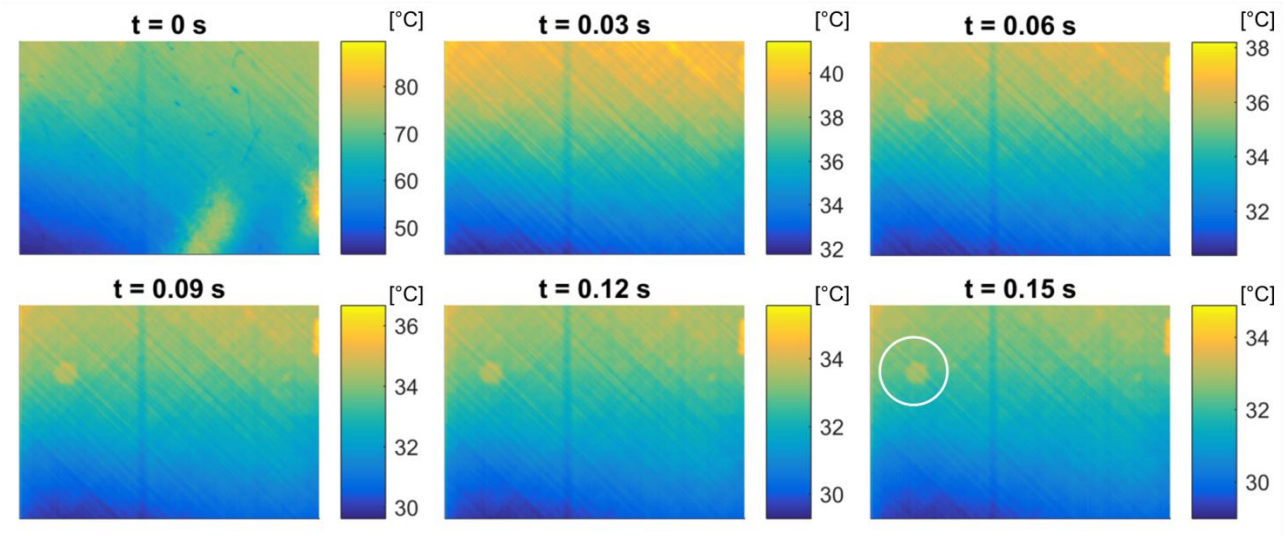

Fig. 9. Thermal sequence samples after the heat pulse generation (t = 0 s). Frames are captured from region Test 1. The white circle at t = 0.15 s encloses the defect D_2^{CL1}

The inspection of the maps of surface temperatures reveals that a thermal gradient is generated along the vertical direction. This result is actually due to the specific excitation used in these experiments. Flash lamps do not produce a thermal wave of infinite extension, able to illuminate orthogonally the whole specimen surface. Consequently heat is not radiated homogeneously from areas without defects. Nevertheless a specific regularity can be sensed during the cooling phase, i.e. regions without defects produce a response with an almost vertical in-plane gradient of temperature. This consideration enables the computation of the reference temperature following the approach discussed in Eq. (7), where the kernel patch P_{ij} is assumed to be equal to the whole i-th row of each frame of the thermal video.

On the contrary, the vertical direction of the in-plane thermal gradient is altered in regions where a defect arises. As an example, it is possible to notice this alteration, enclosed by the circle at t = 0.15 s, which corresponds to the presence of D_2^{CL1}. This different behavior of defective regions is highlighted in Figs. 10(a), where temperature profiles, corresponding to points which belong to the three classes, are displayed.

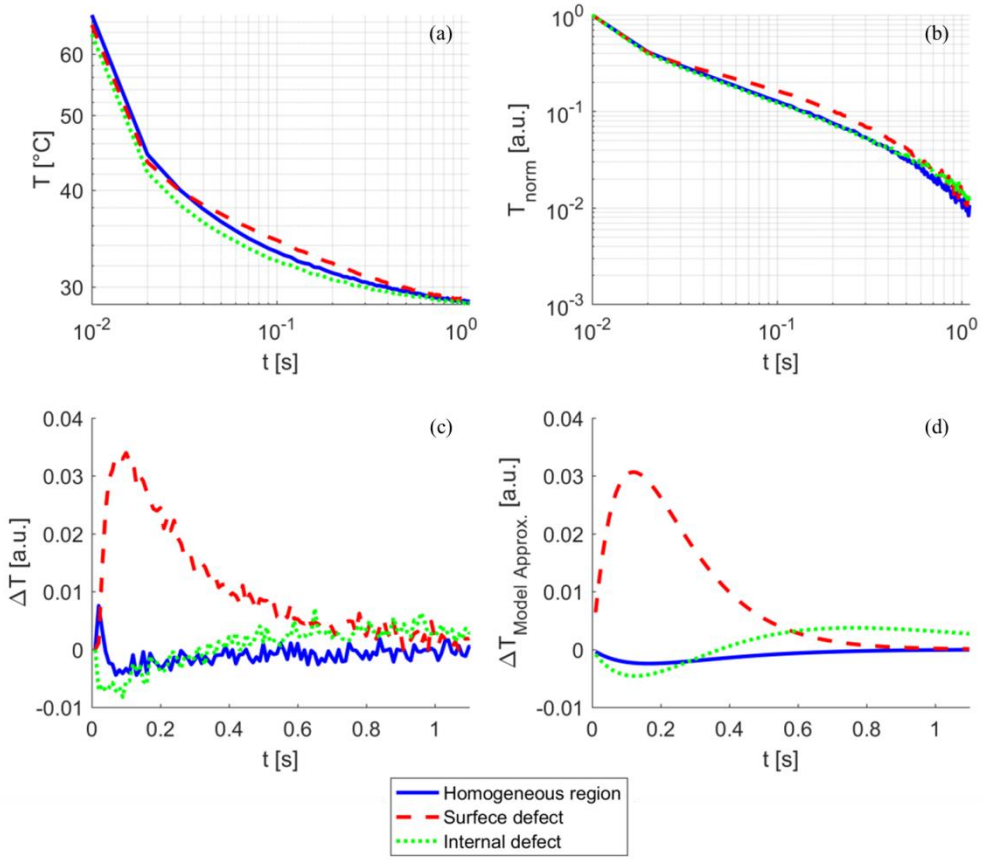

Fig. 10. (a) Actual decays of surface temperature and (b) corresponding normalized behaviors. (c) Thermal contrast obtained by the normalized temperature profiles and (d) corresponding model approximations. Curves are extracted from a flawless region (blue solid) and defective regions of class 1 (red dashed) and class 2 (green dotted).

Experimental results in Fig. 10(a) clearly prove that the temperature decay in logarithmic scale no longer follows the linear profile as an effect of the finite thickness of the laminate, also in the case of inspection of a non-defective area (solid blue curve). As a consequence, every algorithm of feature extraction working on the analysis of the deviations of temperature decays from the linear model will work with low accuracy, since the reference profile is actually far from the expected one. For this reason, the proposed methodology goes through the normalization of temperature decays, whose results are in Fig. 10(b). In this case the process of temperature normalization is performed on a time interval of 1.2 s, which is enough extended to enhance discrimination between the two classes of defects of interest.

Reference temperature decays at each row are finally used to obtain the normalized thermal contrasts, which are able to highlight alterations of transient thermal behaviors, as shown in Fig. 10(c). In this case, the homogeneous region produces a flat contrast, which evolves around zero, whereas regions with surface or internal defects have representative responses to the heat pulse. The use of the exponential approximation in Eq. (6) is able to model with three parameters the temporal evolution of all thermal contrasts, as demonstrated by the inspection of Fig. 10(d). The analysis of

results also confirms that the model approximation through least square fitting can filter measurement noise, denoted by the fluctuations of thermal contrasts in Fig. 10(c), adding more reliability to the estimation of the feature vectors.

The parameters (a_1, a_2, a_3) of the fitting model are finally reported in the feature maps of Fig. 11 (first row), which is related to the analysis of the training region. The application of the median filter produces the maps of (a_1^F, a_2^F, a_3^F) reported in the second row of Fig. 11.

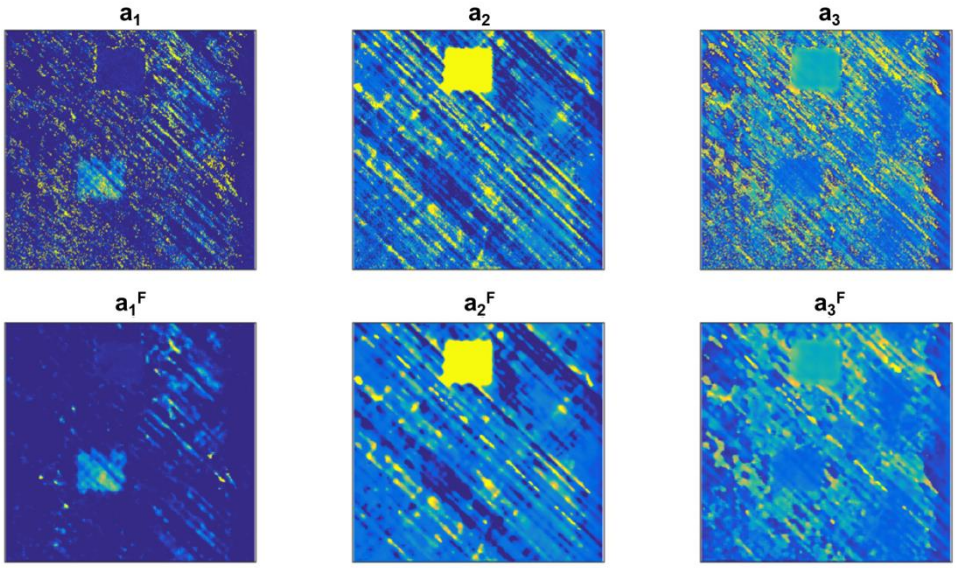

Fig. 11. Feature maps obtained by processing thermal images of the training region. First row: input data; Second row: filtered data.

At a glance, it can be noticed that these two sets of maps contain the information needed to detect the defective areas of the training region, whose ground truth is in Fig. 8. This suggests that the proposed process can give accurate results in solving the defect detection task. Accordingly, feature values in each pixel of the maps have to be arranged in vectors F(i,j), whose size depends on the considered extension of the neighborhood.

In these experiments, four neighborhood sizes have been considered, namely 1×1 (single scalar input), 3×3, 5×5 and 7×7, for the two sets of features (a_1, a_2, a_3) and (a_1^F, a_2^F, a_3^F). Predictors are arranged following the processing of Fig. 5 which results in input feature vectors of increasing length, equal to 3, 27, 75 and 147, respectively, which will clearly increase the computational cost. At the same time, four classifiers are used, namely a decision tree (CTree), a decision forest made of 30 trees (BagTree), a k-nearest neighbor classifier defined over a neighborhood of 10 prototypes (KNN10) and a k-nearest neighbor classifier with an inverse-distance voting function (WeiKNN). Numerical experiments are performed on 32 predictors, resulting from the application of four

classifiers receiving four different feature sizes out of two sets of inputs. The response to testing predictors can be compared by means of the metrics in Eq. (8).

Each classifier has been trained by using a set of examples made of couples of feature vectors out of the training region and the corresponding known class, given by the ground truth. Training is performed following the k-fold cross-validation approach (k = 10) [41]. Once classifiers are correctly trained, they are used to predict the labels, i.e. the class, of every observation of the two test regions. The outcomes can be then compared with the expected ones, thus producing the balanced accuracy results in Tab. 2, which refers to the use of input features, and Tab. 3, which exploits filtered data.

Tab. 2. Balanced accuracy measurement obtained by testing the four classifiers considering the c-th class against the remaining. Results are obtained by the application of input feature values (a_1,a_2,a_3). Best results for the c-th class are highlighted in bold.

Classifier	Patch Size	Class 0 vs All	Class 1 vs All	Class 2 vs All
CTree	1×1	80.1187	77.1856	83.8733
	3×3	77.9718	72.5782	84.5683
	5×5	79.6243	75.3876	84.7479
	7×7	78.5763	73.4712	84.6012
BagTree	1×1	82.4784	79.4222	86.4161
	3×3	**83.9361**	80.3682	**88.1210**
	5×5	81.1887	76.3457	86.7372
	7×7	82.4784	79.4222	86.4161
KNN10	1×1	82.8589	**81.8610**	84.3473
	3×3	78.4377	80.2449	76.4500
	5×5	71.7832	79.0738	63.6606
	7×7	69.7845	80.4208	58.0958
WeiKNN	1×1	82.3086	81.0790	84.3413
	3×3	79.9985	80.8842	79.1083
	5×5	73.3447	79.7357	66.2910
	7×7	70.9883	80.9987	60.0410

Tab. 3. Balanced accuracy measurement obtained by testing the four classifiers considering the c-th class against the remaining. Results are obtained in case of filtered input data (a_1^F, a_2^F, a_3^F). Best results for the c-th class are highlighted in bold.

Classifier	Patch Size	Class 0 vs All	Class 1 vs All	Class 2 vs All
CTree	1×1	85.3218	81.9400	89.4274
	3×3	84.8390	81.7495	88.5487
	5×5	86.8784	**85.2614**	88.8876
	7×7	86.0653	85.1633	87.3013
BagTree	1×1	86.1391	83.0930	**89.9358**
	3×3	**86.7301**	84.0869	89.8968
	5×5	85.1996	81.4178	89.5964
	7×7	86.5600	84.2786	89.2063
KNN10	1×1	85.6396	83.4455	88.4794
	3×3	80.6140	74.4382	88.0342
	5×5	81.2919	75.6974	87.9962
	7×7	83.1736	78.9857	88.2649
WeiKNN	1×1	85.5247	83.3362	88.4411
	3×3	80.9267	74.6763	88.4815
	5×5	81.6792	76.0688	88.4396
	7×7	83.3858	79.1780	88.5799

The first analysis of results shows that balanced accuracies approach satisfying values, often higher than 80%. It proves the robustness of the proposed exponential-model-based features, which always perform accurately despite the classifier used for predictions. On the other hand, although all classifiers can properly label observations, results in both Tabs. 2 and 3 demonstrate that decision trees and decision forests produce better results than the k-NN-based classifiers. At the same time the comparison of results given by decision trees and decision forests proves that ensembles work better than single predictors, taking advantages of the improved ability in producing accurate responses to noisy inputs.

In addition, increasing the number of the considered neighbors to a 3×3 patch can give better results with respect to the scalar single input. On the other hand, the patch extension has to be limited to

this value in case of decision forest classifiers, since adding more neighbors can introduce ambiguities to the classifier.

Further analyses on the results of classification are shown in Fig. 12, where precision versus recall is plotted in scatter diagrams. Parametric investigations are performed as a function of the exponential-model-based features, both input and filtered (blue and red marks, respectively), and of the size of the patch used to rearrange data in feature vectors.

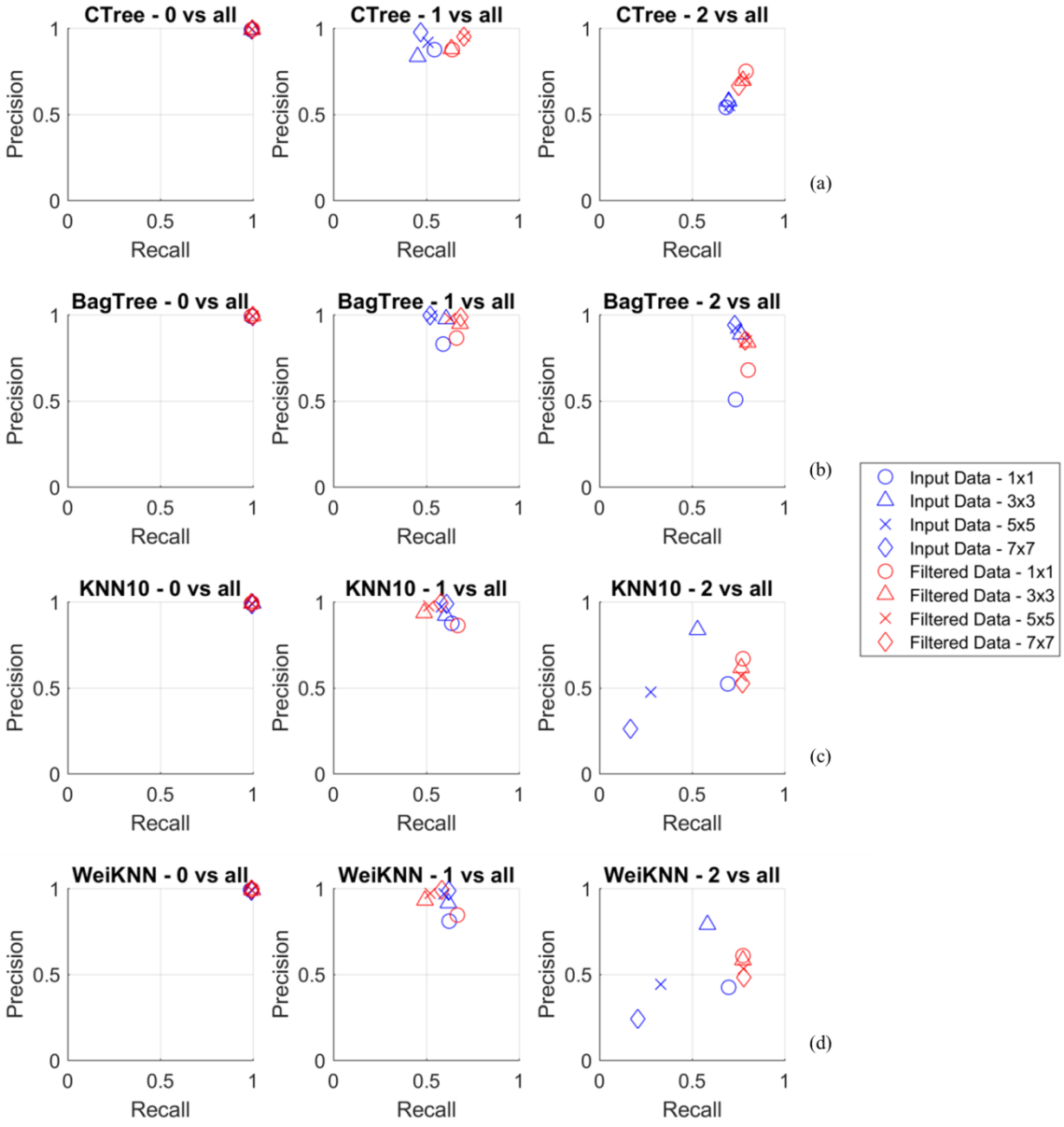

Fig. 12. Classification results in terms of precision *versus* recall as a function of the set of input predictors: (a) Decision tree; (b) Decision forest made of 30 trees; (c) KNN with 10 neighbors; (d) Weighted KNN with squared inverse distance weights. Each scatter plot describes the performance in classifying a class against the others.

With reference to Fig. 12, the best results are obtained for scatter measures in the top-right corner of the recall-precision plane, which corresponds to the lowest value of missed predictions, both false positives and negatives. In this case all classifiers perform properly in detecting the 0-class, which

corresponds to homogeneity. On the contrary, results can differ significantly if detection of class 1 or 2 is compared with the corresponding remaining classes. For the sake of simplicity, results of classification of defective classes are averaged in Fig. 13, where marks refer to the average precision $\overline{PPV}$ and the average recall $\overline{TPR}$. These measurements are the sum of precision and recall of class 1 and 2, weighted by the probability of that value, i.e. the number of occurrence of class 1 and 2 (N_1 and N_2, respectively):

$$\begin{cases} \overline{PPV} = \dfrac{N_1 PPV_1 + N_2 PPV_2}{N_1 + N_2} \\ \overline{TPR} = \dfrac{N_1 TPR_1 + N_2 TPR_2}{N_1 + N_2} \end{cases} \tag{9}$$

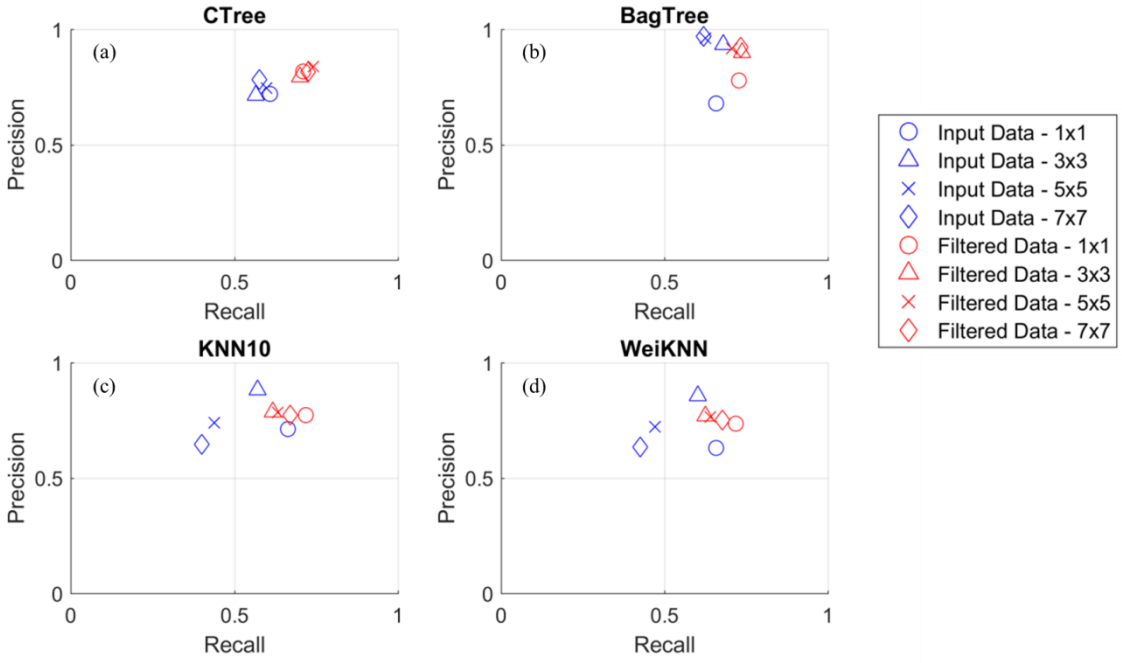

Fig. 13. Average precision $\overline{PPV}$ and recall $\overline{TPR}$ for estimating classification performances in detecting defects against regions of homogeneity: (a) Decision tree; (b) Decision forest; (c) KNN; (d) Weighted KNN.

The inspection of Fig. 13 clearly indicates that the best results in terms of average precision and recall are obtained with the decision forest classifier made of 30 trees having filtered input data with patch size equal to 3×3 or 7×7. In these cases $\overline{PPV}$ and $\overline{TPR}$ are equal to 89.87% and 73.67% for a 3×3 neighborhood, and 92.21% and 73.28% for a 7×7 neighborhood, respectively. High values of average precision $\overline{PPV}$ are of great importance in the analysis of defects beneath the surface of the laminate since it shows that in case of positive labeling that region has a defect of that specific class with high probability.

Furthermore, the insight on the achieved values suggests that the use of a 7×7 neighborhood allows better predictions than the case of using a 3×3 patch for feature extraction. Nevertheless, this improvement is paid in terms of an increase of computational cost, since the number of input predictors rises from 27 to 147 entries. For this reason, downsizing the patch is much more efficient for in-line inspections, at the expense of slightly alterations of results in terms of precision and recall. Final results of classifications are described in Fig. 14, where input thermal responses are treated by the proposed methodology to classify filtered input data from 3×3 patches with the trained decision forest.

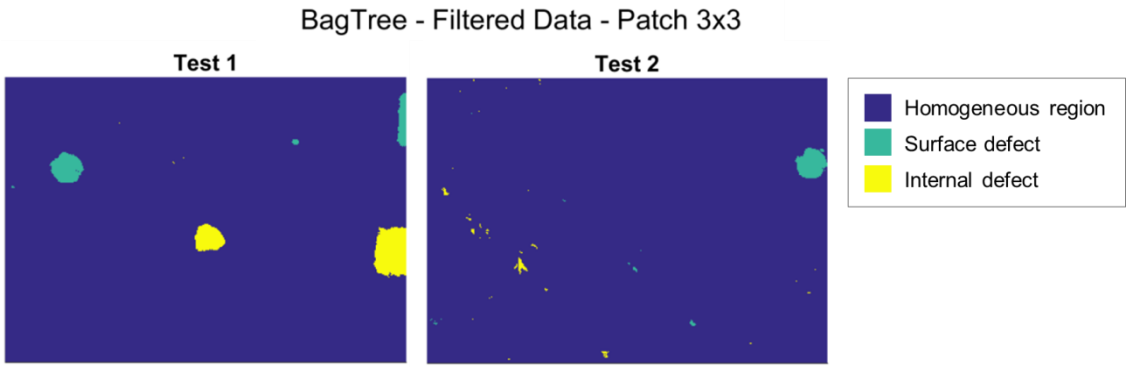

Fig. 14. Final predictions resulting by the application of the decision forest having filtered input features with 3×3 patch size.

Final predictions in Fig. 14 are in good agreement with the expected results defined by ground truths in Fig. 8. In this case, the mean balanced accuracy, namely the average value of the results of Tab. 3 corresponding to this specific set of features and classifier, is equal to 86.9%. On the other hand, standard accuracy values, i.e. the ratios between the exact labels of the c-th class (TP_c and TN_c) and the total number of observations (number of pixels of the c-th class of the maps in Fig. 14), can be averaged among classes to define the mean accuracy, which here reaches 99.47%. As a consequence, the two mean values of balanced and standard accuracy show that exponential-model-based features are enough discriminant to allow the classification of defects, as clearly stated by the maps in Fig. 14.

A deeper analysis of the resulting maps in Fig. 14 reveals that defects are labeled with more rounded edges in comparison with the expected ones. This result is more evident for the smallest defects D_2^{CL1} and D_2^{CL2}, whose shapes are closer to circles than squares. Actually this behavior can be an effect of lateral heat flow which can reduce thermal contrast near the defects corners. In addition, it can be also due to further alterations of the specimen during polymerization. These possible changes are not captured in the definition of the ground truths, since they are obtained by scanning the specimen before its polymerization. The actual shapes of defects can be different from the one of the ground truth. Therefore, the overall accuracy and recall of class 1 and 2 are implicitly

limited below the unity as missing detections at edges introduce false negatives in the estimation of both classes.

Besides, false positives are mainly due to the specific layup of the specimen. As an example, results of Test 2 display sparse points labeled as internal defects, which actually correspond to the sharp ridges, highlighted by the tilted yellow stripe in Fig. 7(d). The presence of these points can be easily reduced improving the training phase, by adding more examples of class 0, which includes homogeneity and all possible structures different from the defects of class 1 and 2. In addition, sparse labels representative of defective regions can be filtered out by means of proper morphological filters working on connected components. As an example, components having areas, namely the number of pixels, lower than a threshold can be erased from the image. The threshold value has to be set in accordance with the mechanical tolerances of the specific production process. Further investigations on image filtering are out of the scope of this paper and will be studied in future works.

The understanding of results of Fig. 14 further proves that the approximation due to the use of the analytical model of Eq. (3), even more amplified by the proposed model in three unknowns, is suitable for this analysis. It is mainly due to the size of the defects, which is larger than their depth, and it is ascribable to the significant difference between the depths of the two classes of defects at comparison (first interface against middle interface). On the contrary, smaller and/or deeper defects, and shorter depth differences among delaminations may bring forward the limitations of the analytic model. These aspects will be object of future validations through further experiments.

4. Conclusions

In this paper, a complete approach for the automatic detection and classification of defects in composite materials for aeronautics has been developed and tested. Starting from the thermal inspection of the response of a laminate surface to the application of a heat pulse, the proposed processing approximates the temperature decay with an exponential model made of three parameters. These unknowns are able to characterize the behavior of the specimen, dividing homogeneous regions from defective ones and classifying the latter depending of their depths. Experiments have proven the robustness of the proposed exponential-model-based features in reducing measurement noise, which is further limited by the use of suitable filters on extended feature maps. As a result, a decision forest made of 30 trees trained on the input image is able to classify three classes of surface areas (pristine, surface defects and internal defects) with high values of both standard accuracy (99.47%) and balanced accuracy (86.9%). In addition, results are

almost insensitive to the specific classifier used in labeling, thus proving that exponential-model-based features are enough informative to always achieve defect classification. Finally, further contributions will be dedicated to additional processing of the input thermal measurements in order to aid the extraction of the parameters used by the method, thus improving its performances in terms of precision and recall.

Acknowledgements

This work is part of the Italian MIUR Project DITECO (Ref. PAC-PON03PE_00067_2). The authors would like to thank Mr. Giuseppe Bono and Mr. Michele Attolico for the fruitful and high-level discussions and Mr. Pasquale Cavallo for the precious help in experimental activity performed in this work.

Reference

[1] Palumbo D, Ancona F, Galietti U. Quantitative damage evaluation of composite materials with microwave thermographic technique: feasibility and new data analysis. Meccanica 2015;50:443-460.

[2] Nikishkov Y, Airoldi L, Makeev A. Measurement of voids in composites by X-ray Computed Tomography. Composites Science and Technology 2013;89:89-97.

[3] Hassen AA, Taheri H, Vaidya UK. Non-destructive investigation of thermoplastic reinforced composites, Composites Part B: Engineering 2016;97:244-254

[4] Hosur MV, Murthy CRL, Ramamurthy TS, Shet A. Estimation of impact-induced damage in CFRP laminates through ultrasonic imaging. NDT&E International 1998;31(5):359-374.

[5] Poudel A, Shrestha SS, Sandhu JS, Chu TP, Pergantis CG. Comparison and analysis of Acoustography with other NDE techniques for foreign object inclusion detection in graphite epoxy composites. Composites Part B: Engineering 2015;78:86-94.

[6] Castellano A, Foti P, Fraddosio A, Marzano S, Piccioni MD. The ultrasonic C-Scan technique for damage evaluation of GFRP composite materials. International Journal of Mechanics 2016;10:206-212.

[7] Castellano A, Foti P, Fraddosio A, Marzano S, Paparella F, Piccioni MD. Monitoring applied and residual stress in materials and structures by non-destructive acoustoelastic techniques. In

IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS 2016) 2016; ISBN: 978-150902369, DOI:10.1109/EESMS.2016.7504830.

[8] Tamborrino R, Palumbo D, Galietti U, Aversa P, Chiozzi S, Luprano VAM. Assessment of the effect of defects on mechanical properties of adhesive bonded joints by using non destructive methods. Composites Part B 2016;91:337-345.

[9] De Angelis G, Meo M, Almond DP, Pickering SG, Angioni SL. A new technique to detect defect size and depth in composite structures using digital shearography and unconstrained optimization. NDT&E International 2012;45:91-96.

[10] Pérez MA, Gil L, Oller S. Impact damage identification in composite laminates using vibration testing. Composite Structures 2014;108: 267-276.

[11] Angelidis N, Irving PE. Detection of impact damage in CFRP laminates by means of electrical potential techniques. Composites Science and Technology 2007;67:594-604.

[12] Palumbo D, De Finis R, Demelio GP, Galietti U. A new rapid thermographic method to assess the fatigue limit in GFRP composites. Composite Part B 2016;103:60-67.

[13] Palumbo D, De Finis R, Demelio GP, Galietti U. Study of damage evolution in composite materials based on the Thermoelastic Phase Analysis (TPA) method. Composite Part B 2017;117:49-60.

[14] Maldague XPV. Theory and practice of infrared technology of non-destructive testing. John Wiley & Sons; 2001.

[15] Munoz V, Valès B, Perrin M, Pastor ML, Welemane H, Cantarel A, Karama M. Damage detection in CFRP by coupling acoustic emission and infrared thermography, Composites Part B: Engineering 2016;85:68-75

[16] Cheng L, Tian GY. Comparison of Nondestructive Testing Methods on Detection of Delaminations in Composites. Journal of Sensors 2012;2012:408437.

[17] Vavilov VP, Burleigh DD. Review of pulsed thermal NDT: Physical principles, theory and data processing. NDT&E International 2015;73:28-52.

[18] Shepard S. Advances in pulsed thermography. Proceedings of SPIE, Thermosense-XXIII, 2001;4360:511.

[19] Balageas D, Roche JM. Common tools for quantitative time resolved pulse and step-heating thermography – part I: theoretical basis. QIRT Journal 2014;11:10-28.

[20] Palumbo D, Galietti U. Damage investigation in composite materials by means of new thermal data processing procedures. Strain 2016;52(4):276-285.

[21] Alvarez-Restrepo CA, Benitez-Restrepo HD, Tobon LE. Characterization of defects of pulsed thermography inspections by orthogonal polynomial decomposition. NDT & E International 2017;91:9-21.

[22] D'Orazio T, Leo M, Distante A, Guaragnella C, Pianese V, Cavaccini G. Automatic ultrasonic inspection for internal defect detection in composite materials. NDT & E International 2008;41(2):145-154.

[23] Leo M, Looney D, D'Orazio T, Mandic DP. Identification of defective areas in composite materials by bivariate EMD analysis of ultrasound. IEEE Transactions on Instrumentation and Measurement 2012;61(1):221-232.

[24] Farley SJ, Durodola JF, Fellows NA, Hernandez-Gomez LH. High resolution non-destructive evaluation of defects using artificial neural networks and wavelets. NDT&E International 2012;52:69-75.

[25] Darabi A, Maldague X. Neural network based defect detection and depth estimation in TNDE. NDT & E International 2002;35(3):165-175.

[26] D'Orazio T, Guaragnella C, Leo M, Spagnolo P. Defect detection in aircraft composites by using a neural approach in the analysis of thermographic images. NDT & E International 2005;38(8):665-673.

[27] Usamentiaga R, Venegas P, Guerediaga J, Vega L, López I. Feature extraction and analysis for automatic characterization of impact damage in carbon fiber composites using active thermography. NDT & E International 2013;54:123-132.

[28] Marani R, Palumbo D, Galietti U, Stella E, D'Orazio T. Automatic detection of subsurface defects in composite materials using thermography and unsupervised machine learning. In IEEE 8th International Conference on Intelligent Systems (IS) 2016;516-521.

[29] Lowe DG. Object recognition from local scale-invariant features. In IEEE International Conference on Computer Vision (ICCV), 1999;2:1150-1157.

[30] Bay H, Ess A, Tuytelaars T, van Gool L. Surf: Speeded up robust features. Computer Vision and Image Understanding 2008;110(3):346-359.

[31] Parker WJ, Jenkins RJ, Butler CP, Abbott GL. Flash method of determining thermal diffusivity, heat capacity, and thermal conductivity. Journal of applied physics 1961;32(9):1679-1684.

[32] Almond DP, Pravin P. Photothermal science and techniques. Springer Science & Business Media 10; 1996.

[33] Lau SK, Almond DP, Milne JM. A quantitative analysis of pulsed video thermography. NDT & E International 1991;24(4):195-202.

[34] Maierhofer C, Röllig M, Ehrig K, Meinel D, Céspedes-Gonzales G. Validation of flash thermography using computed tomography for characterizing inhomogeneities and defects in CFRP structures. Composites Part B: Engineering 2014;64:175-186.

[35] Gonzalez RC, Woods RE. Digital image processing. Upper Saddle River, New Jersey: Prentice-Hall; 2001.

[36] Kuncheva LI. Combining pattern classifiers: methods and algorithms. Hoboken, New Jersey: John Wiley & Sons; 2004.

[37] Online available: http://www.flir.com/science/blog/details/?ID=74531

[38] Marani R, Nitti M, Cicirelli G, D'Orazio T, Stella E. High-Resolution Laser Scanning for Three-Dimensional Inspection of Drilling Tools. Advances in Mechanical Engineering. 2013;2013:620786.

[39] Patruno C, Marani R, Nitti M, D'Orazio T, Stella E. An Embedded Vision System for Real-Time Autonomous Localization Using Laser Profilometry. IEEE Transactions on Intelligent Transportation Systems, 2015;16(6):3482-3495.

[40] Available Online: www.teledynedalsa.com.

[41] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. International Joint Conference on Artificial Intelligence, 1995;14(2):1137-1145.

Figure caption

Fig. 1. Sketch of a setup for pulsed thermography.

Fig. 2. (a) Transient temporal decay of the surface temperature of a semi-infinite medium heated by a pulsed-source and (b) corresponding normalized temperature curves. Profiles are obtained by changing the ratio d^2/α, linked to the presence of a reflecting defect at depth d. The reference decay is $T = t^{-1/2}$.

Fig. 3. (a) Thermal contrast behavior related to the normalized temperature profiles after the source pulse as a function of d^2/α. (b) Corresponding model approximation for feature extraction.

Fig. 4. Feature vectors arrangement from the model approximation through the generation of feature maps. Each line of the final matrix corresponds to a transient temporal decay collected by a pixel of the thermocamera (observation).

Fig. 5. Single feature vector formation in case of extended patch size. This example substitutes a generic line of the final feature matrix. Neighboring observations can share some entries.

Fig. 6. (a) Picture of the actual setup used for experiments. (b) Camera point of view focused on the specimen.

Fig. 7. (a) Detailed layup of the specimen having foreign object inclusions within layers P_4-P_5 (internal defect, class 2) and P_7-P_8 (surface defect, class 1). Also ply orientations (θ) are reported in the scheme. (b) Surface of the specimen before polymerization. Circles are centered on top defects made of FEP and white paper. (c) Depth map acquired after the deposition of P_5 over P_4 which sandwiches two defects of class 2. (d) Depth map resulting by scanning the top layer. Two new defects of class 1 are highlighted.

Fig. 8. Ground truth corresponding to the three region considered for training classifiers and test predictions.

Fig. 9. Thermal sequence samples after the heat pulse generation (t = 0 s). Frames are captured from region Test 1. The white circle at t = 0.15 s encloses the defect D_2^{CL1}

Fig. 10. (a) Actual decays of surface temperature and (b) corresponding normalized behaviors. (c) Thermal contrast obtained by the normalized temperature profiles and (d) corresponding model approximations. Curves are extracted from a flawless region (blue solid) and defective regions of class 1 (red dashed) and class 2 (green dotted).

Fig. 11. Feature maps obtained by processing thermal images of the training region. First row: input data; Second row: filtered data.

Fig. 12. Classification results in terms of precision *versus* recall as a function of the set of input predictors: (a) Decision tree; (b) Decision forest made of 30 trees; (c) KNN with 10 neighbors; (d) Weighted KNN with squared inverse distance weights. Each scatter plot describes the performance in classifying a class against the others.

Fig. 13. Average precision $\overline{PPV}$ and recall $\overline{TPR}$ for estimating classification performances in detecting defects against regions of homogeneity: (a) Decision tree; (b) Decision forest; (c) KNN; (d) Weighted KNN.

Fig. 14. Final predictions resulting by the application of the decision forest having filtered input features with 3x3 patch size.

Table caption

Tab. 1. List of defects within the specimen and corresponding properties.

Tab. 2. Balanced accuracy measurement obtained by testing the four classifiers considering the c-th class against the remaining. Results are obtained by the application of input feature values (a_1,a_2,a_3). Best results for the c-th class are highlighted in bold.

Tab. 3. Balanced accuracy measurement obtained by testing the four classifiers considering the c-th class against the remaining. Results are obtained in case of filtered input data (a_1^F,a_2^F,a_3^F). Best results for the c-th class are highlighted in bold.