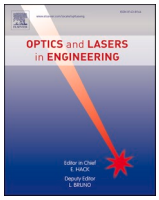
Optics and Lasers in Engineering 182 (2024) 108455

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng



Influence of lock-in thermography set-up parameters on the capability of a temporal convolutional neural network to characterize defects in a CFRP

Tiziana Matarrese [a,*], Roberto Marani [b], Davide Palumbo [a], Tiziana D'Orazio [b], Umberto Galietti [a]

[a] Department of Mechanics, Mathematics and Management (DMMM), Politecnico di Bari, Via Orabona, 4, 70125, Bari, Italy
[b] National Research Council of Italy, Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, via Amendola 122D/O, 70126 Bari, Italy

ARTICLE INFO

Keywords:
Lock-in thermography
Deep learning
CFRP
1-D CNN
Defects detection

ABSTRACT

Lock-in thermography is a well-established non-destructive technique for detecting damage in composite materials. The success of the lock-in technique in industrial applications depends on several test parameters, such as the excitation frequency, the number of frames per cycle, and the number of excitation cycles that need to be correctly set to reduce testing and processing time. However, quantitative analysis to characterize defects using lock-in thermography is still a challenging task. Machine and deep learning algorithms can be useful to automatize the classification of defects in terms of size and depth. In this regard, the aim of this work is to investigate the influence of lock-in thermography set-up parameters on the capability of a temporal convolutional neural network to characterize defects in a carbon fiber-reinforced polymer specimen. The performance of the proposed neural network in identifying and classifying the depth of defects was evaluated as a function of the number of cycles, the frames per cycle, and the random frames lost during the acquisition of the thermal sequence by an infrared camera. The achieved results have been critically discussed through qualitative and quantitative analyses.

1. Introduction

Nowadays, fiber-reinforced polymer (FRP) composite materials are widely used in many industrial fields such as aerospace, automotive, civil, and so on. The growing success of these materials is due to their constituents, resin, and fiber, which have the task of directing and withstanding mechanical loads. However, due to the heterogeneous nature of the manufactured components, during the production phase or operative conditions, internal defects like bridging, fiber breakage, and matrix cracks could significantly decrease the strength of components or structures, leading to serious safety problems. In this regard, non-destructive tests (NDT) are necessary for characterizing defects and estimating the residual strength. Currently, several NDT methods can be used to control composite materials, such as ultrasound [1,2], X-ray [1], shearography [3], and thermography techniques [4–6].

Active thermography (AT) techniques are based on the deposition of a thermal external stimulus on the surface of the test piece through halogen/infrared/flash lamps or laser [7]. According to the signal shape and duration of the thermal excitation, the thermographic techniques can be classified as pulse (PT) [8–10], step (ST) [11,12], and lock-in thermography (LI) [9,13–20]. The first two techniques adopt a square wave as an external stimulus, a Dirac impulse for PT, or a long pulse for ST. Instead, the lock-in consists of modulating the heat source with sine, square, or chip form [9,15]. Independently of the modality to excite the surface of the test piece, the physics thermal principle is the same. The thermal wave propagates into the body of the test piece following Fourier's conduction law. The presence of a defect beneath the surface, which has different thermophysical properties with respect to the surrounding area, alters heat propagation. This physics phenomenon can be appreciated on the surface of the test piece as a thermal signature that can help the operator localize the defects.

In recent years, AT has been successfully used for controlling composite components since it is a contactless and full-field technique. This means that a significant component area can be inspected in a few minutes [4]. Many authors focused on improving the signal-to-noise ratio of thermal data by employing specific post-processing algorithms [5] to help an expert operator identify defects.

It is important to highlight that these algorithms allow the operator

* Corresponding author.
E-mail address: tiziana.matarrese@poliba.it (T. Matarrese).

https://doi.org/10.1016/j.optlaseng.2024.108455
Received 6 February 2024; Received in revised form 23 June 2024; Accepted 23 July 2024
Available online 5 August 2024

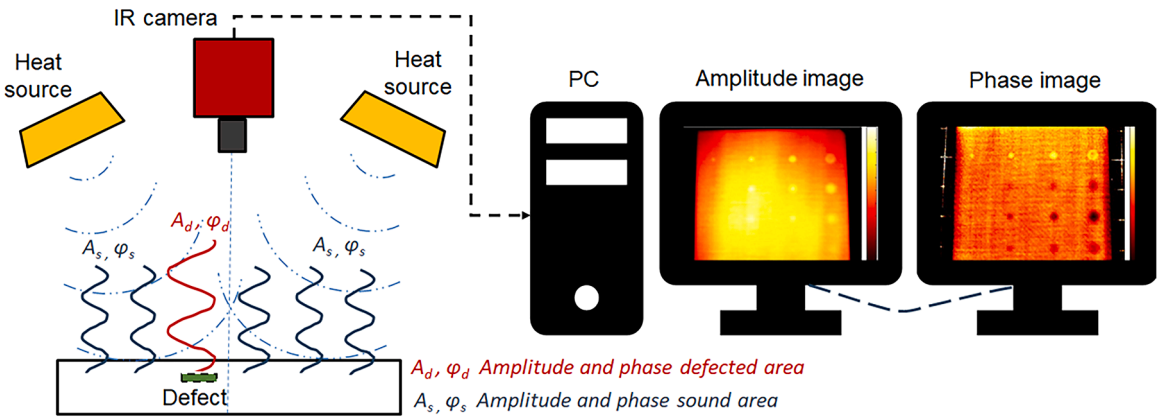


Fig. 1. A schematic representation of the lock-in set-up configuration in reflection mode.

to perform only qualitative analyses without characterizing defects in terms of shape and size. The quantitative approaches remain a challenging task that the researchers are still investigating.

Currently, the approaches used in literature to get quantitative data from LI are mainly based on analytical approximate solutions [21,22], numerical finite element analyses [23], and experimental tests to characterize defects [24]. An analytical one-dimensional model, based on a 1D hypothesis [21], has been used in [22] to estimate the thickness of a glass-FRP (GFRP) laminate. However, this model provides reliable results only for large defects.

Another way to characterize the defects is using hybrid methods based on numerical and analytical analyses. In this regard, in [25] an iterative approach that minimizes the cost function to get information on the defects in a carbon-FRP (CFRP) specimen has been adopted. The cost function is the difference between phase lag given by the least square algorithm applied on the acquired thermal wave and the numerical phase values obtained by performing a finite element simulation. However, this procedure is time-consuming and not suitable for industrial applications.

In recent years, due to the growing success of CNNs in computer vision, image recognition, and feature segmentation [26], many researchers have adopted this kind of approach in NDT to characterize defects [27–30]. The CNN networks were initially developed to elaborate large datasets of images (2-D data). Therefore, a large amount of data is required to train the networks correctly and create general models to avoid overfitting. In order to apply CNNs in the NDT context, many experimental tests on master specimens with different shapes and size defects are necessary, significantly increasing the costs of tests [29,31]. Numerical models can reduce the time for CNN training, but experimental tests are always required to validate and verify the models [32].

A feasible way to bridge the gap between the lack of a huge dataset and the use of neural networks is to consider 1-D convolution layers [33–37] to train the network with thermal temporal signals. In this way, from a thermographic sequence of a single specimen, it is possible to extract a large number of signals of both defects and sound regions. In the literature, some works successfully train the neural networks with lock-in thermal temporal signals to label the defects on CFRP. In [34], a 1-D residual attention network (1-DRAN) has been trained with thermal temporal signals obtained with lock-in and long-pulsed tests both performed on a CFRP. The lock-in test was carried out at 0.5 Hz with two excitation cycles, and the IR camera was set to 50 frames/s. The authors found that long-pulse signals, unlike lock-in, showed an enhanced sensitivity in detecting deeper defects. The capability of segmenting defects in a CFRP test piece using a convolutional residual network trained with temporal thermal data was also studied in the work presented in [38]. In particular, the surface of the test piece was excited with a lock-in square wave at a fixed period of 45 s for three excitation cycles, and in each cycle, 856 frames were acquired. The goal of this work was to show the capability of the proposed neural network in segmenting the defects independently of the configuration of the excitation sources.

Considering the just exposed works, deep learning algorithms have been studied to automatically identify and classify the defects with specific lock-in test parameters. However, some aspects must be further analyzed to understand their applicability in actual industrial contexts. To our knowledge, no one has evaluated the effect of lock-in set-up test parameters in classifying defects using a neural network.

The aim of this work is to evaluate how the variation of some lock-in test parameters can affect the performance of a temporal convolutive neural network (1-D CNN) in identifying and classifying defects inside a CFRP specimen. In particular, two test parameters usually imposed by the operator, the number of cycles and the number of frames/cycle, in an offline process have been varied.

The performance of the investigated 1-D CNN has also been evaluated as a function of dropped frames, i.e., the number of random frames lost by the infrared camera (IR camera) during the acquisition. In addition, a parametric study has been performed by varying some hyperparameters of the net to choose the most suitable net architecture. So, the main novelty of the work is to investigate the limits and advantages of using a neural network to characterize defects in CFRP as a function of the experimental lock-in parameters to reduce testing and processing time.

The paper is organized as follows: Section 2 presents the Lock-in thermography theory; Section 3 describes the specimen and the experimental set-up; Section 4 explains the methodology adopted; Section 5 describes the different architectures that have been tested in the experimental campaign; Section 6 reports the experimental results, and finally Section 7 discusses the conclusions.

2. Theory: lock-in thermography

Lock-in thermography is based on a periodic deposition of modulated thermal heat on the surface of the specimen [7,9,13], a schematic representation of a typical lock-in set-up in reflection mode is shown in Fig. 1. Halogen lamps are usually employed to heat the surface of the specimen homogeneously, avoiding localized thermal gradients that could hide or be confused as defects. Simultaneously, on the same side of the specimen, the thermal material response is remotely recorded by an IR camera. After that, the acquired thermal sequence is processed by a suitable lock-in algorithm [13,14] that gives amplitude and phase maps as an output.

As shown in Fig. 1, the phase map is mainly considered since it is less affected by non-homogeneous heating. The phase map is related to thermal diffusion length according to this equation:

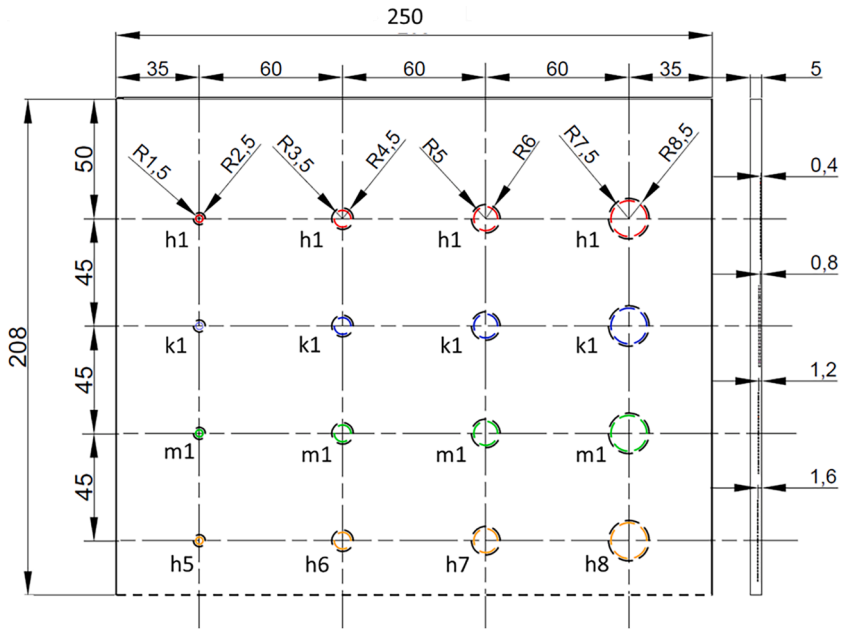


Fig. 2. Investigated CFRP specimen made of artificial defects having different materials (halar, kevlar, metal), placed at different depths and with different diameters (in black are represented the dimensions of flash-bakers).

$$\varphi(z) = \frac{z}{\mu} \tag{1}$$

The thermal flow propagates into the body of the specimen and the reached depth (z) is ruled by the thermal diffusion length μ. This parameter μ strictly depends on the thermophysical properties of the material and the modulation period T [6,9,15,16]:

$$\mu = \sqrt{\frac{2k}{\rho c_p \omega}} = \sqrt{\frac{2\alpha}{\omega}} = \sqrt{\frac{\alpha T}{\pi}} \tag{2}$$

where k is the thermal conductivity, ρ is the density, c_p is the specific heat at constant pressure, ω is the angular frequency, and α is the thermal diffusivity. The thermal diffusion length represents the distance at which the temperature reaches 37 % of the surface temperature value. Below this value, detecting defects becomes very hard.

Therefore, the achieved depth can be obtained [14,18]:

$$z = C_1 \sqrt{\frac{\alpha T}{\pi}} = C_1 \mu \tag{3}$$

where C_1 is a constant of the material.

To detect the defect, the phase contrast is evaluated as follows:

$$\Delta\varphi = \varphi_d - \varphi_s \tag{4}$$

where φ_d and φ_s are the phase signals on the defect and sound regions, respectively. For some frequencies (named blind frequencies) that depend on the size, shape, and depth of the defects, the phase signals of the defect and sound regions can be equal. In this way, the defect disappears and seems to merge with the sound region, even if the thermal diffusion length μ is greater than the depth of the defect. Therefore, the experimental tests must be performed considering multiple sine wave lock-in periods to ensure the detection of the defects. This constraint considerably increases the testing time. Currently, in literature, an attempt to overcome this problem is to excite the surface of the specimen with a modulated square wave [9,15].

3. Specimen and experimental set-up

The lock-in tests have been performed on a CFRP specimen composed of 25 layers arranged in a cross-ply configuration [(0/90°)]s to achieve a nominal thickness of 5 mm. The specimen has dimensions of 250 × 500 mm^2. The prepreg is MTM228 having a volume fraction of 42%. This epoxy matrix resin was cured in a dry room for 60 min at 120 °C with a pressure of 2 bar in the autoclave. To simulate the delaminations that can occur between two plies of the specimen, each defect has been manufactured with two patches of the same material (to enclose air), that have been put inside a flash-baker to fix them. As shown in Fig. 2, a limited area of dimensions 250 × 208 mm^2 has been inspected. In this area, sixteen defects have been placed. These defects can be divided into four groups according to their different depths and materials. Starting from the first row (on the top of the specimen), the first group consists of four defects h1, h2, h3, and h4. Each of them has been obtained by inserting two circular patches of halar materials with varying diameters, placed at the same depth of 0.4 mm. Moving down to the specimen, the depths of defects increase. In the second row, the group of defects k1, k2, k3, k4 are manufactured with kevlar and placed

Fig. 3. Experimental set-up. The experimental tests have been carried out by adopting a reflection configuration mode.

Table 1
Set up parameters used during the experimental phase.

Excitation period [s]	Sampling rate [Hz]	N° of frames/ cycle	N° of cycles	N° tot frames	N° of repetitions
50	10	500	3	1500	3

at 0.8 mm. In the third and fourth rows, the groups of defects are in metal and halar material and are placed at a depth of 1.2 mm and 1.6 mm, respectively. By observing the specimen from left to right, each defect has an increasing diameter ranging between 5 mm to 17 mm (considering the flash-bakers represented in black in Fig. 2)

This specimen has been inspected adopting a reflection configuration mode, as shown in Fig. 3. The data acquisition system is performed by the Multides systems (DES Diagnostic Engineering Solutions S.r.l.), which has a proprietary software namely IRTA able to synchronize the halogen lamps and the IR camera. Two halogen lamps both of 500 W have been modulated to excite sinusoidally (for three cycles, Table 1) the surface of the test piece, and simultaneously the thermal material response has been recorded with a microbolometer IR camera (FLIR A655sc) working in the longwave range of 7 – 14 μm, with a focal plane array of 640 × 480 elements and NETD < 30 mK. The obtained geometrical resolution has been 0.48 mm/pixel.

According to Table 1, the set parameters of the IR camera have been imposed to acquire 500 frames in each cycle, by collecting a total of 1500 frames. Different excitation frequencies have been investigated, and three repetitions for each frequency have been performed. In this work, only the thermal sequences acquired at 50 s of the excitation period have been taken into account since this period ensures that almost all the defects are well-detected in the phase map [14]. Moreover, only two of the three repetitions have been considered, as explained in detail in the following sections.

4. Methods

The adopted methodology is described in the following subsections: the pre-processing procedure of thermal data, the lock-in test parameters considered for the analysis, and the investigated neural network architectures.

4.1. Signal pre-processing

It is difficult to obtain homogeneous heating due to the convective heat exchanges that occur at the edges of the specimen. Then, it is necessary to pre-process the acquired thermal sequences. In Fig. 4 the flowchart of the adopted procedure is shown.

On the left, there is a frame extracted (4(a)) from the raw thermal sequence on which two spatial profiles, both in correspondence with kapton defects and the adjacent sound region, have been represented (black and red dashed lines, respectively). Both spatial profiles are affected by non-homogeneous heating. By focusing on the black profile, the three peaks correspond to the defects (*k2, k3, k4*), while defect *k1* is too small to be detected. In addition, the thermal temporal behaviors of two different points belonging to defect and sound regions appear quite comparable. Therefore, a pre-processing technique is necessary to remove the effects of non-homogeneous heating and increase the signal-to-noise ratio. All the frames of the thermal sequence have been normalized (4 (b)), and a median spatial filter (4 (c)) with a large kernel size [121,121] has been applied to each of them. Fig. (4 (d)) reports a frame extracted by the output sequence obtained as a difference between the normalized and median filter sequences. The two resulting spatial profiles are less affected by non-homogeneous heating. In the same way, the two thermal temporal profiles present a better signal-to-noise ratio.

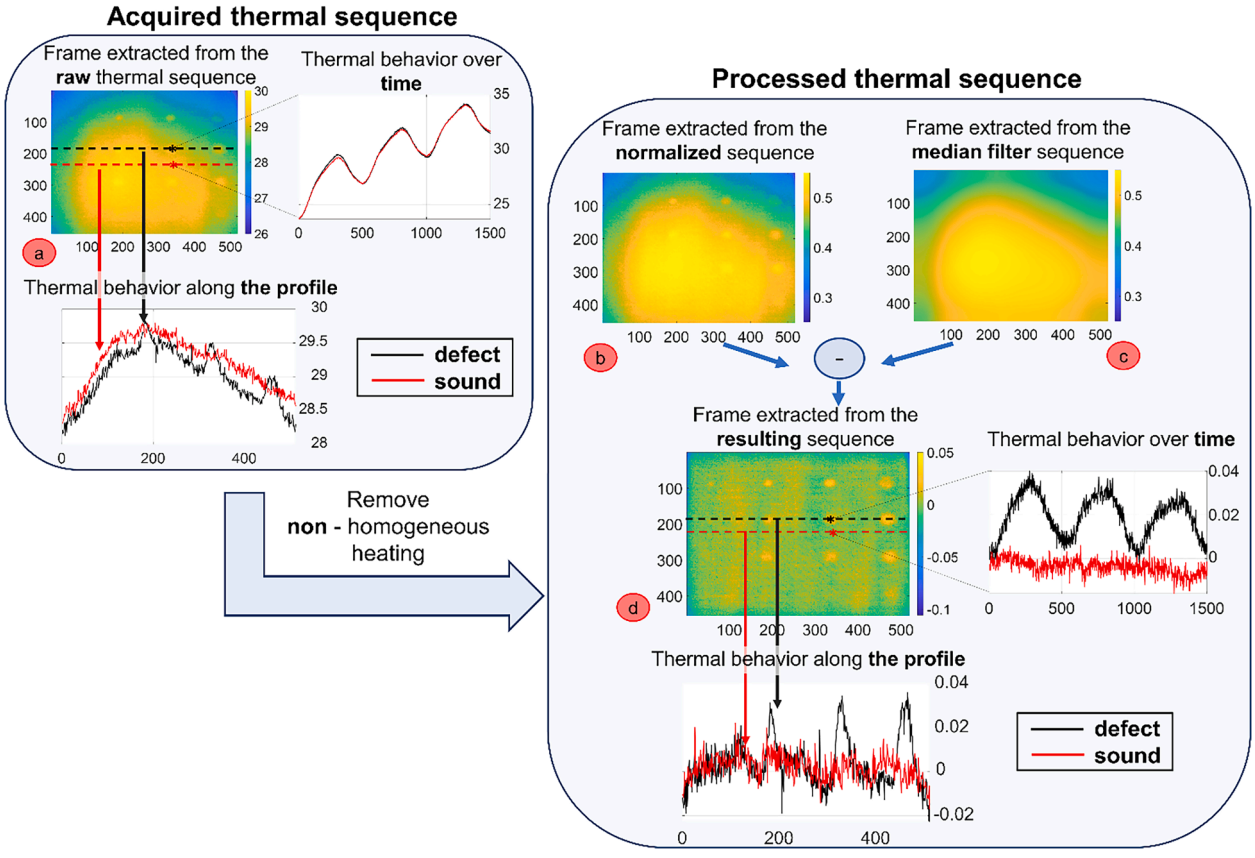


Fig. 4. Flowchart of the adopted signal pre-processing procedure to remove non-homogeneous heating and increase the signal-to-noise ratio.

Table 2
The investigated offline parameters with their relative input signal lengths reported in parentheses and in bold.

Excitation period [s]	50
N° of cycles (dataset length)	1 **(500)** – 2 **(1000)** – 3 **(1500)**
N° of frames/cycle (dataset length)	25 **(75)** – 50 **(150)** – 100 **(300)** – 125 **(375)** – 250 **(750)** – 500 **(1500)**
Dropped frames [%]	0 – 2 – 5 –10 –15 – 20 – 25 – 30

4.2. Lock-in test parameters

After the pre-processing of the thermal sequences, some offline analyses have been performed as reported in Table 2.

Different signal lengths have been achieved for both the number of cycles and the number of frames per cycle, as represented in bold inside the parentheses. For example, at one excitation cycle, the signal length is equal to 500 frames, at two cycles is 1000, and so on. The same thing can be said for the frames/cycle. It is important to underline that for convolution neural networks is not possible to use different dataset lengths for the training and testing phase. Therefore, for each investigated case, for the number of cycles and the number of frames per cycle, the 1-D CNN has been trained every time.

A different approach has been used for the dropped frames. The dropped frames were expressed as a percentage value of the total number of imposed frames, as listed in Table 2. This analysis simulates an uncontrolled event that could occur during the acquisition of the thermal sequence, in which some frames are missing. The direct consequence of this event is that the total number of acquired frames is less than the expected one, providing different signal lengths. To have the same data length (1500 frames and 3 cycles), the dropped frames were substituted with null frames (frames for which all temperature values are 0). Then, the robustness and sensitivity of the network architectures were evaluated as a function of the percentual values of dropped frames in a range from 0 to 30 %.

4.3. 1- D CNN neural network

The proposed methodology must be able to classify the thermal input signal as belonging to the sound or defective region. In this case, five classes have been considered: one is related to the background/sound region, and the other ones are the four defective groups, corresponding to defects placed at four different depths, as described in Fig. 1. The input thermal signals have been pre-processed as described above, and after a learning phase, the output is a class number corresponding to the range [0–4]. Starting from the results obtained in [33], a temporal convolutional neural network (1-D CNN) architecture was selected, as shown in Fig 5.

The proposed 1-D CNN architecture can be divided into four macro-blocks: the input layer, the feature extraction layer, the classification layer, and the decision layer. The number of input thermal data provided to the 1-D CNN is defined by the batch size parameter (in this work, it is set to 32).

For instance, on the left, a thermal normalized and filtered input signal of 1500 frames is graphically depicted as a series of cubes. This input signal gets into three consecutive convolutional layers, which have the task of feature extraction. Each convolutive layer is further divided into three sub-blocks:

- The input thermal signal is convolved with N_F different 1D filters (each has a kernel size $1 \times N_K$). Each filter slides one frame over the thermal signal, and another convolutional operation is performed. This operation is repeated till the whole input temporal signal is analyzed. As a result, N_F thermal output signals that are smaller but have an increasing significance level with respect to the input signal, are generated. It is important to note that the N_F value must be properly set for each convolutional layer.
- The Rectified Linear Unit (ReLU) adds non-linearity to the neural network to increase its complexity. This operation is easily performed: all the negative values are brought to null; instead, all positive values are unchanged.
- The Max Polling performs a downsampling operation. A rectangular window of dimension 1×2 moves with a stride equal to 1×2 on the thermal data and selects each time the maximum value. This operation is performed on all N_F thermal signals given by the convolutional filter. The results are new N_F output signals that are half with respect to the input one but with a further increasing level of significance.

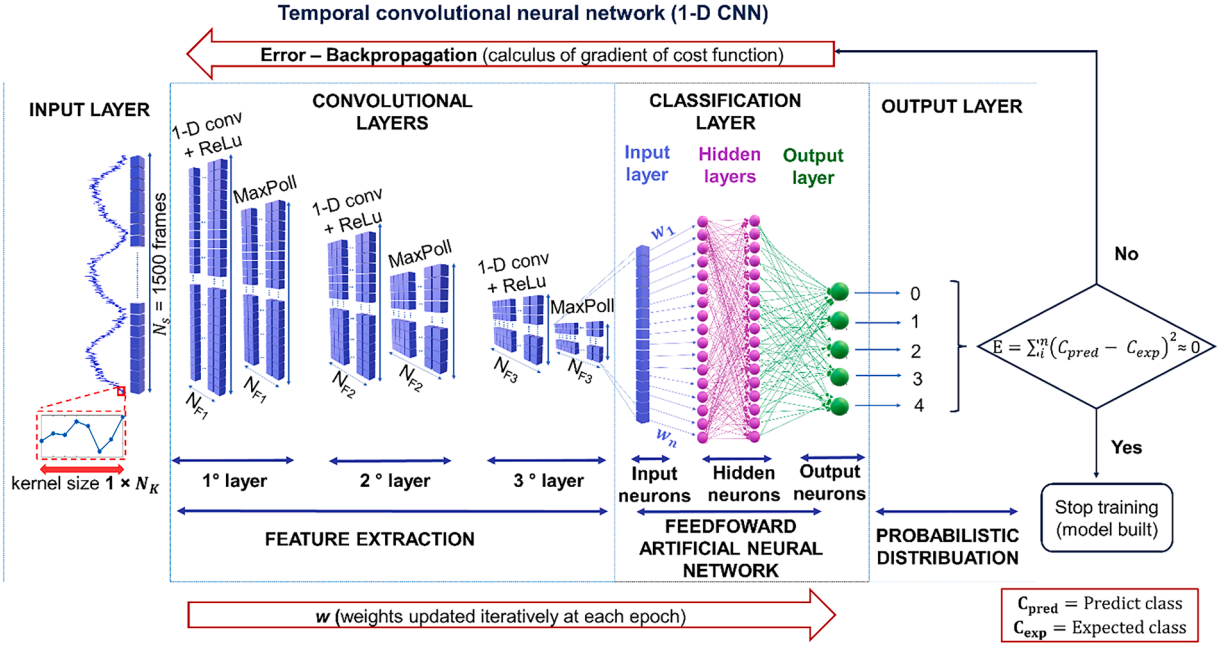


Fig. 5. Schematics representation of the adopted net architecture trained with the thermal signal. N_s is the normalized and filtered thermal signal. N_k is the length of the kernel size, N_{Fi}, with i = 1,2,3, is the number of output 1 D convolution filters of each layer.

The above-cited operations are repeated for each convolutional layer. Then, the thermal signals output at the end of the feature extraction layer depends on the N_F value imposed for the third convolutional layer (N_{F3}). Each of these N_{F3} signals contains the most important information about the input thermal signal; in other words, the network has N_F different representations of shorter lengths of the input thermal signal. This information helps carry out the classification tasks performed by the fully connected network, following the three steps:

- The N_{F3} thermal signals are flattened into a vector that represents the input layer of the fully connected network.
- The softmax function assigns a probability value to the input thermal signal to belong to one of the N_c investigated classes.
- The cross-entropy loss compares the output value given by the build model with the actual value; if these values are comparable, the cross-entropy provides a low value. Therefore, during the training, the goal is to minimize the cross entropy-loss value as much as possible.

In the proposed 1-D CNN architecture, three convolutional layers have been considered. It is easy to understand that the deeper the neural network, the more information is caught. Nevertheless, it is not recommended to significantly increase the network complexity when unnecessary, since a few layers can create a general model, preventing the overfitting problem. Moreover, it is necessary to ensure that the thermal signals do not become null after pooling operations. In this work, since different analyses with different lengths of the thermal signal will be provided as input to the 1-D CNN, a net architecture with three convolutional layers has been considered. This choice represents a proper trade-off between the network complexity and the informative content of the thermal signals at the end of the feature extraction phase.

5. Experimental campaign

The proposed methodology has been implemented using the Pyhorch framework, which runs with an Intel Core i7 8700K (clock frequency 3.7 GHz) supported by 16 GB of RAM, and an nVidia GPU, namely a Geforce RTX350 having memory of 8 GB GDDR6 and 2560 CUDA cores).

5.1. Metrics

To evaluate the performance of the investigated model, some typically used metrics in the deep learning approach have been considered. It is essential to highlight that in multiclassification problems, the performance of the net architecture is assessed with a one-versus-all approach. It consists of reducing the multiclassification problem to a binarization one in which only two classes exist, namely the c-th class, against all the remaining ones. In the following list, the three investigated metrics are described.

- The positive prediction value (PPV_c), also namely Precision.

$$PPV_c = \frac{TP_c}{TP_c + FP_c} \tag{5}$$

where TP_c and FP_c are the True Positives and False Positives, respectively.

This metric represents the number of correct positive predictions for all the times the model forecasts that the event will occur. As can be seen, this metric does not consider the FN_c (False Negatives).

It could occur that although the model has a high precision value, it might not detect many defective pixels, so it does not give the operator the necessary alert.

Table 3
Parameters of the adopted temporal Convolutive Neural Network (1-D CNN) architecture.

Parameters	Description	Value
N_s	Length of the normalized thermal signal	1500
N_c	Number of classes	5
N_K	Length of kernels of 1D convolutional filters	1×8
N_F	Number of 1D convolutional filters	Experimentally evaluated

- The true positive rate (TPR_c) or recall defines the sensitivity of the model.

$$TPR_c = \frac{TP_c}{TP_c + FN_c} \tag{6}$$

This metric represents the number of correct positive predictions over all the actual positive occurrences. Contrary to the precision metric, the recall does not give any information about the number of FP_c, that could be significant. It could happen that the model incorrectly assigned to some pixel of the sound region as belonging to a defective one, thereby giving unnecessary alerts to the operator.

- The Balanced Accuracy metric is widely used when the dataset is strongly unbalanced, as it weights the standard accuracy by the number of occurrences of the c-th label:

$$BACC_c = \frac{\frac{TP_c}{TP_c+FN_c} + \frac{TN_c}{TN_c+FP_c}}{2} \tag{7}$$

It is the mean arithmetic between the sensitivity (first term) and the specificity (second term). The specificity is the negative correct prediction TN_c (True Negative) with respect to all the negative events that occur. Of course, having all the metrics close to unity should be preferable. However, these conditions are not always possible, so a trade-off must be reached. So, in all the fields where safety is mandatory, it is better to get in the worst case. In other words, having a significant value of FP_c with respect to having a high value of FN_c is preferable. In these cases, models that have good performance in all the metrics but maintain, in particular, high recall values represent the best choice.

5.2. 1- D CNN parameters

Table 3 shows the parameters used to define the architecture of the proposed neural network. The input/output constraints, such as the length of the thermal signal N_s and the number of classes N_c determine some of these parameters. The remaining ones are explained in this section.

The lengths of the kernels of 1D convolutional filters N_K must be defined, and generally, it depends on the specific application. More specifically, a low hyperparameter value means that the network will be able to appreciate the quick variations of the input thermal signal, but at the same time, it will be more affected by the noise. On the contrary, a high N_K hyperparameter value means that the network is less affected by noise but can appreciate only the slow variations, thus neglecting the fast variations. In this work, as the noise contribution has been removed thanks to the pre-processing phase, the hyperparameter N_K value has been empirically fixed to 8.

The parameter N_F represents the number of 1D convolutional filters for each of the three considered convolutional layers. In this regard, a parametric study has been carried out in which the N_F value varies between 32 and 64. These values have been chosen since they, being

Table 4
The eight investigated configuration net architecture with the relative number of epochs and time employed to achieve convergence.

Net Architecture	Layer 1 (N_{F1})	Layer 2 (N_{F2})	Layer 3 (N_{F3})	Epochs	Time [min]
Net 1	32	32	32	229	16
Net 2	32	32	64	148	**12**
Net 3	32	64	32	230	18
Net 4	64	32	32	203	**12**
Net 5	64	64	64	206	20
Net 6	64	64	32	211	18
Net 7	64	32	64	184	16
Net 8	32	64	64	176	16

powers of two, ensure an efficient computational approach, allowing the neural network to reach convergence in less time. Table 4 shows the investigated configurations with the relative number of epochs and time to achieve convergence.

As Table 4 reports, the net architectures Net 2 and Net 4 employed less time to achieve the convergence (the model is built in 12 min, in bold). To choose the suitable architecture, all the investigated net architecture has been accurately evaluated and compared by studying the training and test results in qualitative and quantitative ways.

5.3. Training and testing

After pre-processing, each signal has been classified as belonging to one of the five classes. In this work, to have the map of the defects, the acquired thermal sequence has been processed with the signal reconstruction (SR) algorithm [14,19,22], as shown in Fig. 6. The labeling of defects has been manually performed on the phase map to obtain the ground truth (GT) map. Then, the partition in training and test has been performed.

The training dataset is composed of the thermal signals extracted from the defective areas belonging to the columns of size, 5 mm, 12 mm, and 17 mm. The second column (defects diameter equal to 9 mm) has been considered to extract the signals for the test phase.

It is essential to highlight that two different acquisitions of the same specimen were considered to evaluate the actual robustness of the considered 1-D CNN. From the first sequence, the three training regions were extracted; from the second one, the testing region was selected.

5.4. Balanced dataset

One of the potential limits that can affect the convergence of the neural network during the training phase is the quality of the data provided as input. Specifically, some signal values of the sound region can have completely different values with respect to the other ones or can be similar to data belonging to the defect region; this is due to, i.e., edge effects, inhomogeneity deposition of the resin, or something like that. About defected regions, since the defects have finite dimensions, the lateral diffusion effect significantly affects the signal at the edge of defects with a consequent reduction of the thermal signal with respect to the center. Therefore, as shown in Fig. 7, the examples were limited inside the circular red areas in correspondence with the defective regions. In contrast, in the sound region, the examples were chosen randomly inside the two central limited areas in the red rectangular boxes [34]. As can be seen, although the sound region has been delimited, the dataset is still imbalanced since the training examples related to class 0 are still much more with respect to the defective areas. The under-sampling method, among the different approaches present in the literature, has been used to balance the dataset [39]. It consists of randomly removing examples from the majority dataset. The main

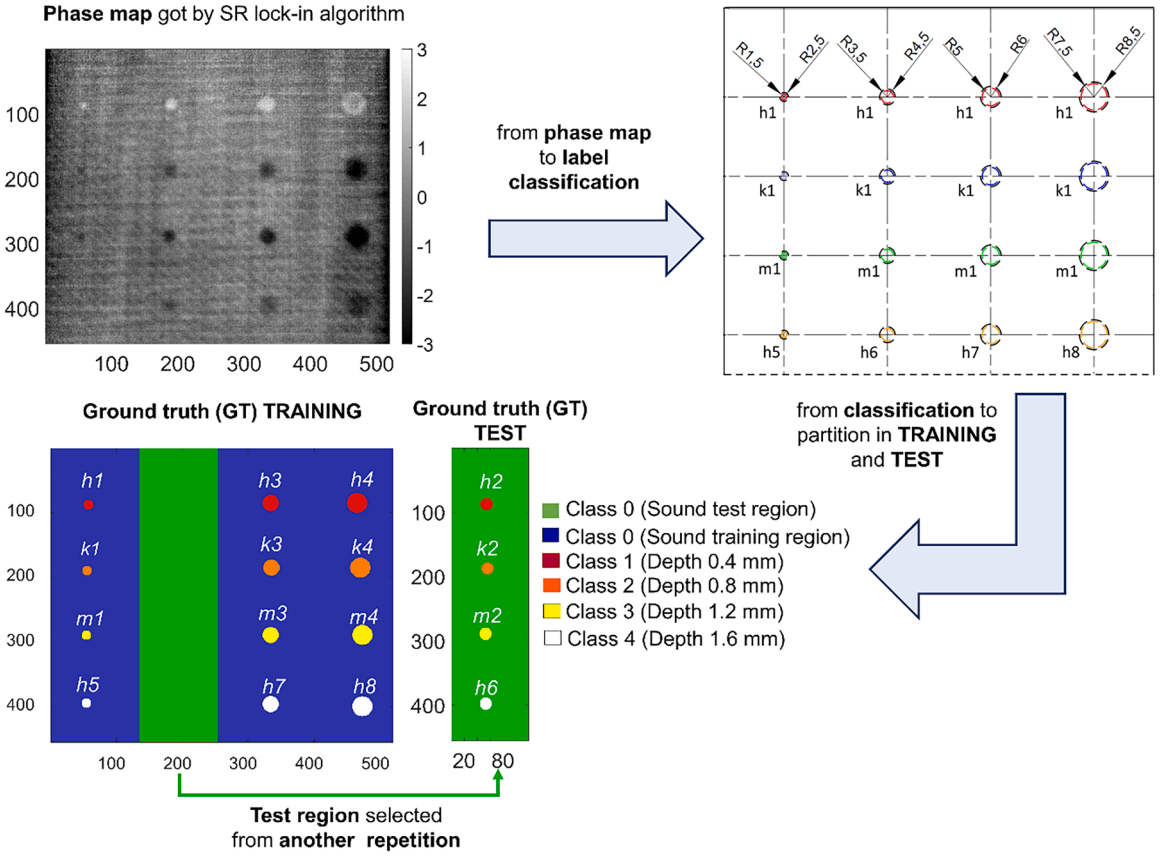


Fig. 6. Classification of defects, on the phase map obtained by the SR lock-in algorithm. The dataset has been split into training (blue sound region) and test regions (green sound region). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

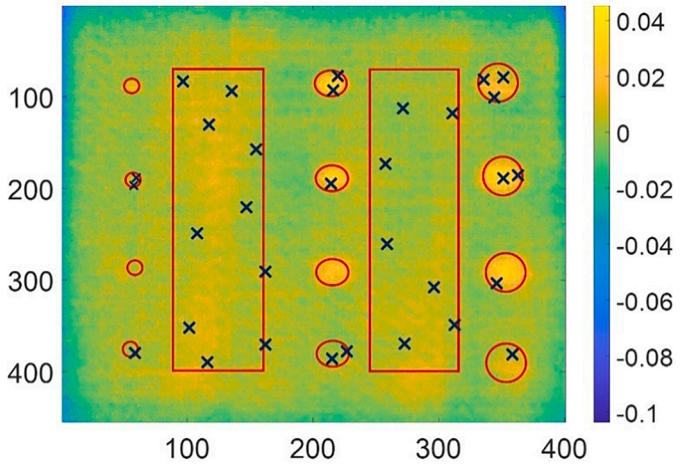


Fig. 7. Examples over which the 1-D CNN learns the model delimited by two red rectangular areas and by circular areas for sound and defects regions, respectively. The number of these examples, chosen randomly at each epoch, depends on the value imposed on the batch size.

drawback is to significantly reduce the number of training examples and induce overfitting, which means that the network cannot build a general model. For our purpose, it has been imposed that each defective class has the same number of examples, and that class 0 is five times more than each defective class.

In addition, for robust training, the achieved dataset has been split into training and validation, 75 % and 25 %, respectively. During training, the objective cost function was optimized by the stochastic gradient descent method (SGDM), with momentum and learning rate set to 0.9 and 0.001, respectively. The training phase stopped when the accuracy of the validation set did not improve for several consecutive epochs (patience parameter). In this work, the patience parameter was set to 8.

6. Results and discussion

In this section, the qualitative and quantitative results obtained by each investigated net architecture will be shown in order to choose the best one. Subsequently, the effect of the number of cycles, number of frames/cycle, and dropped frames on the performance of the adopted 1-D CNN will be illustrated.

6.1. *Parametric study: training phase*

In Fig. 8, the training maps obtained by each net architecture are presented.

As can be seen, the defects *h3,* and *h4* of class 1, are correctly identified and classified by each architecture. In this case, the lateral diffusion effect that affects the edge of the defects is quite negligible since these defects are shallower than the others [8]. Focusing on the defects of classes 2 (*k3, k4*), the lateral diffusion effect becomes evident: the defect edges are not correctly classified. The same happens for defects of class 3 (*m3, m4*) and class 4 (*h7, h8*).

More in detail, it is worth highlighting that no architecture can

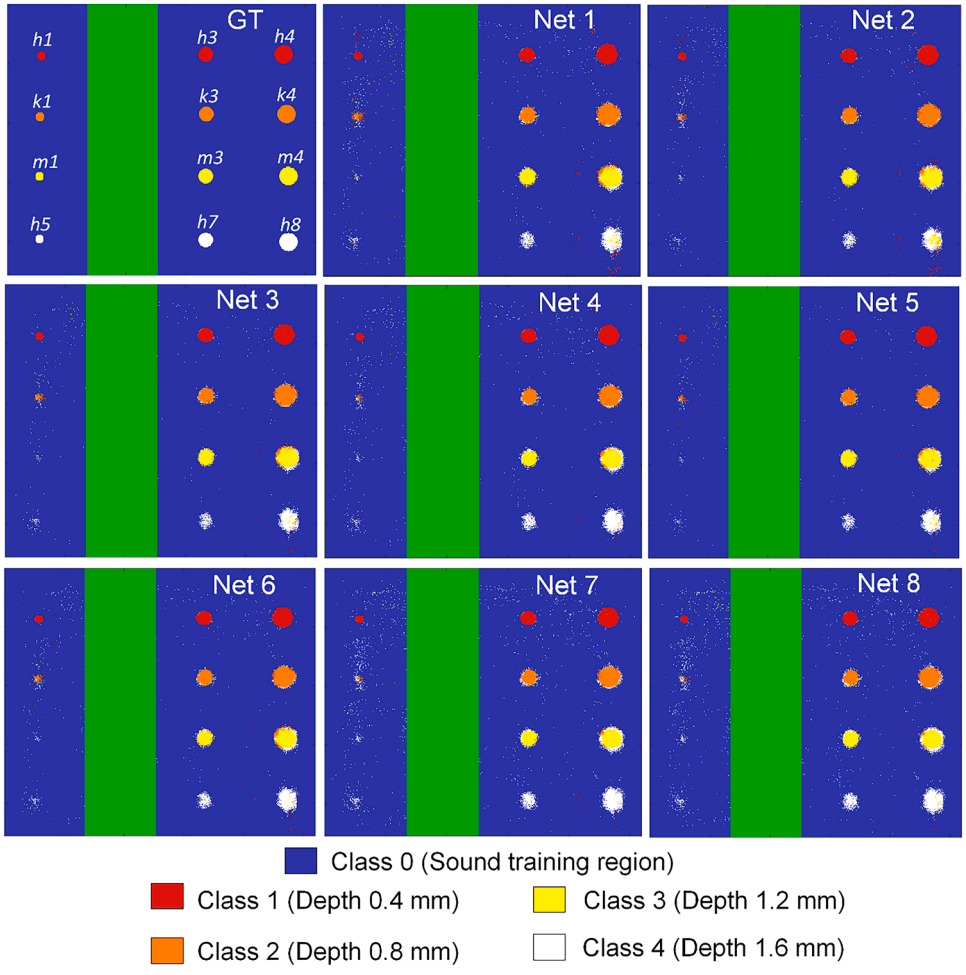


Fig. 8. Classification results of the training phase for each proposed 1- D CNN architecture.

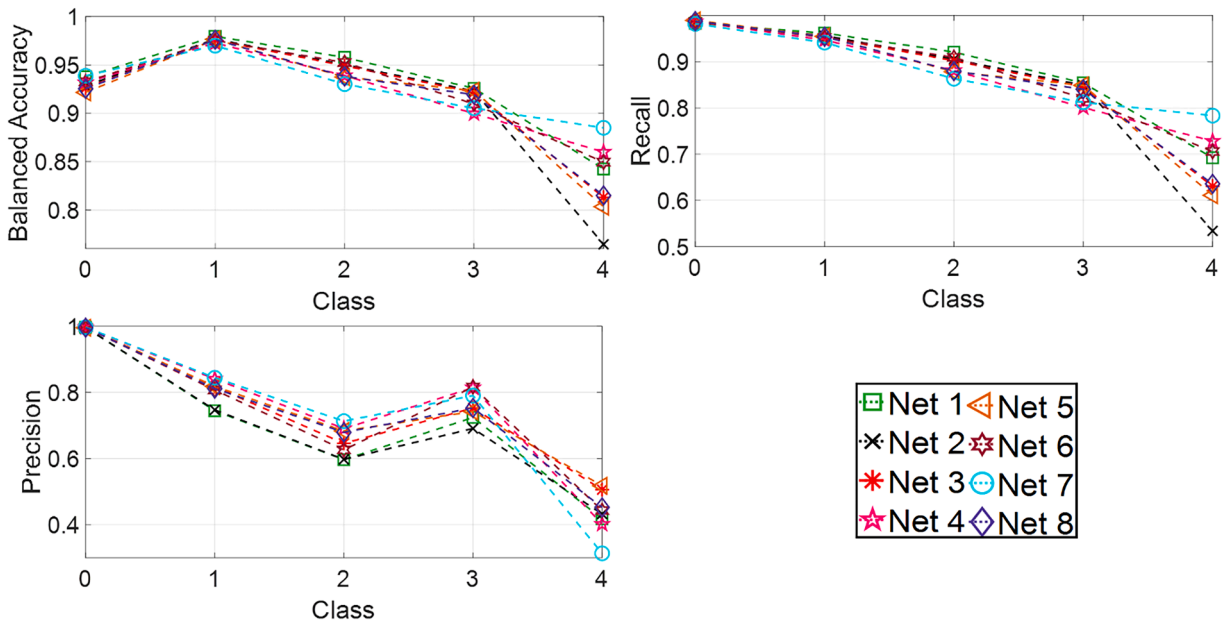


Fig. 9. Comparison among quantitative results obtained in the training phase by each investigated net architecture for the three studied metrics.

Table 5
The mean value of all investigated classes has been calculated after the training phase for the balanced accuracy, recall, and precision metrics.

Net Architecture	BACC	Recall	Precision
Net 1	0.9287	0.8823	**0.6958**
Net 2	0.9085	0.8460	**0.6926**
Net 3	0.9171	0.8640	0.7433
Net 4	0.9205	0.8686	0.7485
Net 5	0.9124	0.8564	0.7513
Net 6	0.9239	0.8748	0.7377
Net 7	0.9258	0.8763	0.7314
Net 8	0.9147	0.8595	0.7382

accurately determine the shape of the defects of class 4. In this case, the thermal signal value in correspondence of pixels on the edges is too low. Therefore, these pixels are erroneously labeled as class 0. The outcome is that these defects seem smaller than the actual dimensions. By observing the defects in the first column (*h1, k1, m1, h5*), all the models provide detections that are not as precise as for the others belonging to the same class but with larger diameters. These defects are challenging to detect since they are smaller and deeper than the others. In this regard, as seen in Fig. 8, the defects *h1* and *k1* of class 1 and class 2 are always correctly well-segmented by each configuration, although the defect *k1* seems smaller in size than the actual one.

Instead, with increasing depth, detecting the *m1* and *h5* defects of classes 3 and 4 becomes critical. As can be seen, all the architectures provide a very small cluster of thermal data grouped in correspondence with these defects. However, the segmentation could be more reliable and accurate, and it cannot be concluded that these pixels belong to a defective region. So, no one architecture can correctly segment them.

The quantitative results are shown in Fig. 9, where all the metrics, i.e., balanced accuracy, recall, and precision, are graphically depicted as a function of the investigated classes.

Starting from class 0 until class 3, for all the investigated metrics, all the architectures are comparable since they have more or less the same value. As expected, for the detection of defects belonging to class 4, the architectures have lower performance. In Table 5, the overall behaviors of the eight models are summarized by showing the mean value of all

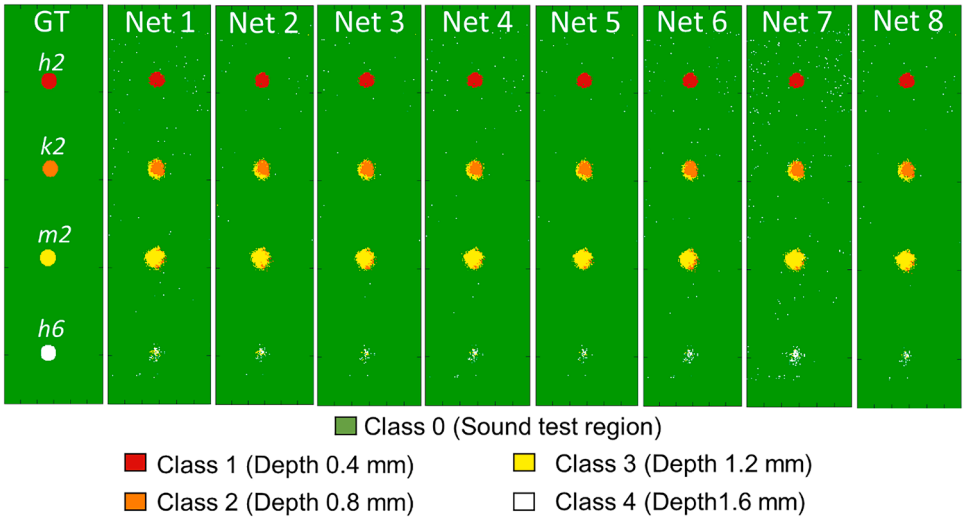


Fig. 10. Classification results of the test phase for each proposed 1-D CNN architecture.

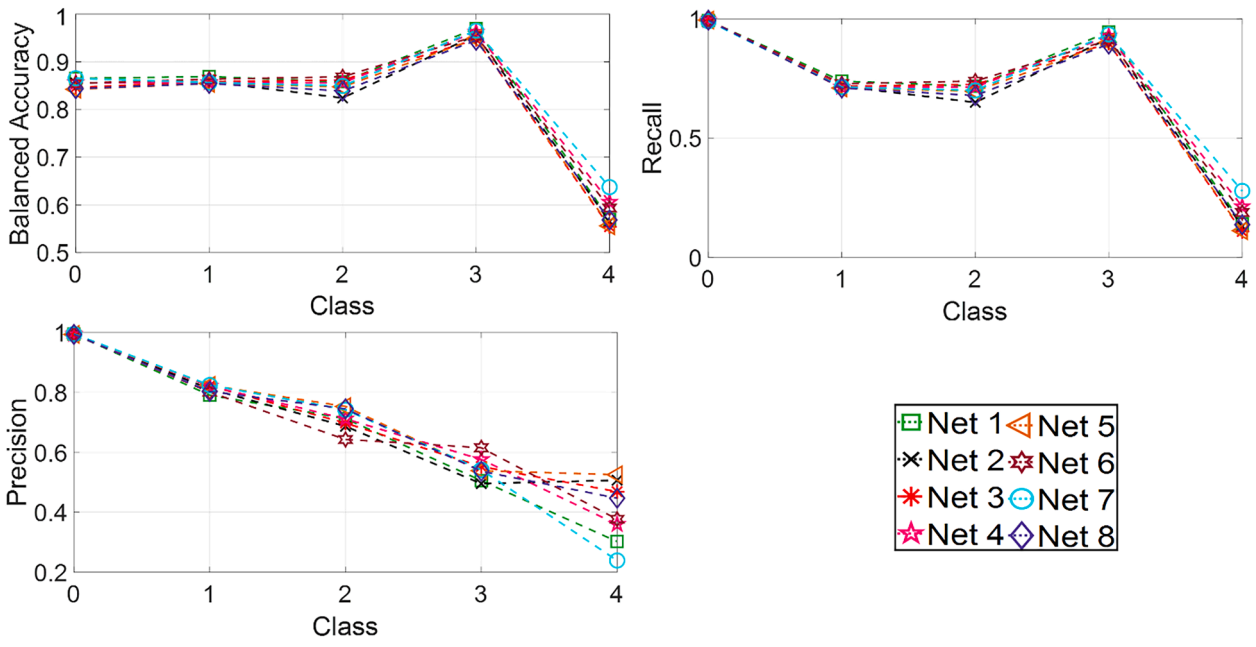


Fig. 11. Comparison among quantitative results obtained in the test phase by each investigated net architecture for the three studied metrics.

Table 6
The mean value of all investigated classes has been calculated for the balanced accuracy, recall, and precision metrics.

Net architecture	BACC	Recall	Precision
Net 1	0.8265	0.7073	0.6611
Net 2	0.8119	0.6820	0.6989
Net 3	0.8134	0.689	0.7068
Net 4	**0.8275**	**0.7127**	**0.6925**
Net 5	0.8098	0.6830	0.7268
Net 6	0.8270	0.7122	0.6859
Net 7	**0.8348**	**0.7237**	**0.6684**
Net 8	0.8100	0.6830	0.7052

investigated classes for the balanced accuracy, recall, and precision metrics.

Considering the results of Table 5, Net1, and Net2 although have good values of balanced accuracy and recall, both have a low value of precision (in bold), therefore they are not considered in the following evaluations. The other architectures are quite comparable. The only parameter that greatly varies among these remaining models is the processing time (see Table 4). For this reason, the Net 4 is the one that after the training phase presents good values of the metrics, and also reaches the convergence in less time.

6.2. *Parametric study: testing phase*

A test dataset extracted from the test region depicted in Fig. 6, has been provided as input to all the architectures. In Fig. 10, the obtained qualitative results are shown.

As can be seen, each architecture recognizes the defect *h2* belonging to class 1, and the lateral diffusion effect is quite negligible. Instead, starting from class 2 up to class 4, the depth of defects increases, and the thermal contrast decreases, so detecting the defects is more difficult.

In particular, the defect of class 4 is detected with less accuracy. In Fig. 11, the quantitative results of the test phase for each proposed architecture are compared to critically analyze the performance of the considered net architectures.

Considering both BACC and recall metrics, all the investigated architectures perform similarly in segmenting the defects from class 0 to class 3. At the same time, they show a different behavior only in correspondence with class 4. Table 6 summarizes the results for the test phase by showing the mean value of all investigated classes for the balanced accuracy, recall, and precision metrics.

From this analysis, Net 4 and Net 7 have balanced accuracy and recall metrics higher with respect to the other net architectures (in bold). Net 4 has a precision value higher than the Net 7. Therefore, in light of these considerations (both taking into account the training and testing results) and the less convergence time, Net 4 was chosen for our further analyses.

6.3. *Effect of the number of cycles*

As mentioned earlier in Section 2, detecting the defects with the lock-in thermography requires correctly choosing the excitation period since it is strictly related to the penetration depth of the thermal flaw into the body of the specimen. Moreover, it is necessary to consider several excitation cycles to collect the suited amount of data for the processing phase. Therefore, it is interesting to assess the minimum number of cycles that ensure a significant thermal contrast to recognize the defects and, at the same time, speed up the experimental testing time. Moreover, reducing the number of cycles can resolve the problem related to the amount of data that needs to be stored for each acquisition. In this work, the experimental tests were carried out by heating the surface of the specimen for three excitation cycles, as depicted in Table 1. The acquired thermal sequence was offline post-processed to achieve one and two excitation cycles, respectively. By doing that, three thermal sequences of different lengths have been obtained, and each of them has been provided as input to the neural network to train the related models.

The qualitative results obtained after the training phase are shown in Fig. 12. As can be seen, the capability of the models to label the defects of class 1 is not significantly affected by the number of cycles. The same consideration applies to classes 2 and 3, which are only less reliable in detecting the smallest defects (*k1* and *m1*). The models perform differently in labeling the deeper defects of class 4: the models of the first two cycles do not detect the defects *h5* and *h7*. In contrast, the model trained with three cycles partially detects some areas of the defects belonging to four classes.

The quantitative results regarding BACC, recall, and precision metrics are depicted in Fig. 13.

The BACC, recall, and precision metrics for the classes 0–3 seem

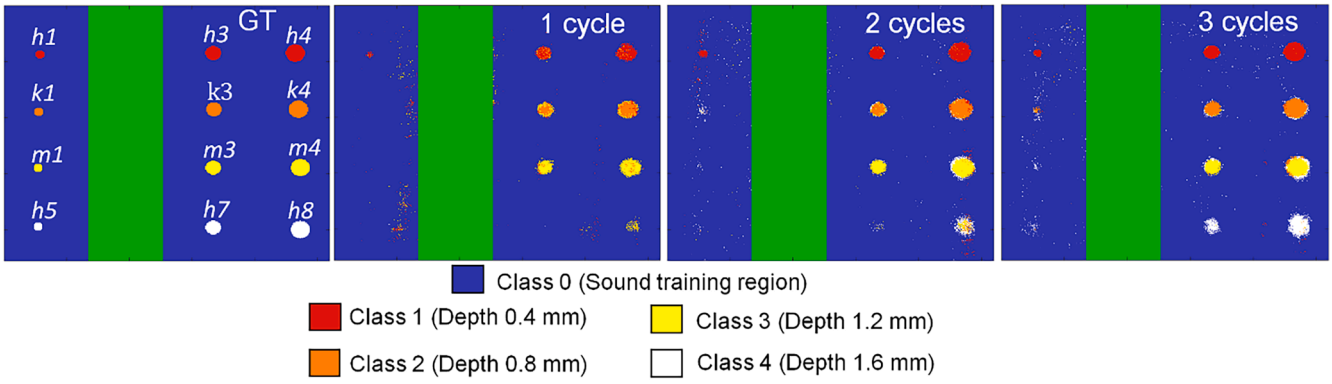


Fig. 12. Classification results of the training phase as a function of the excitation cycles.

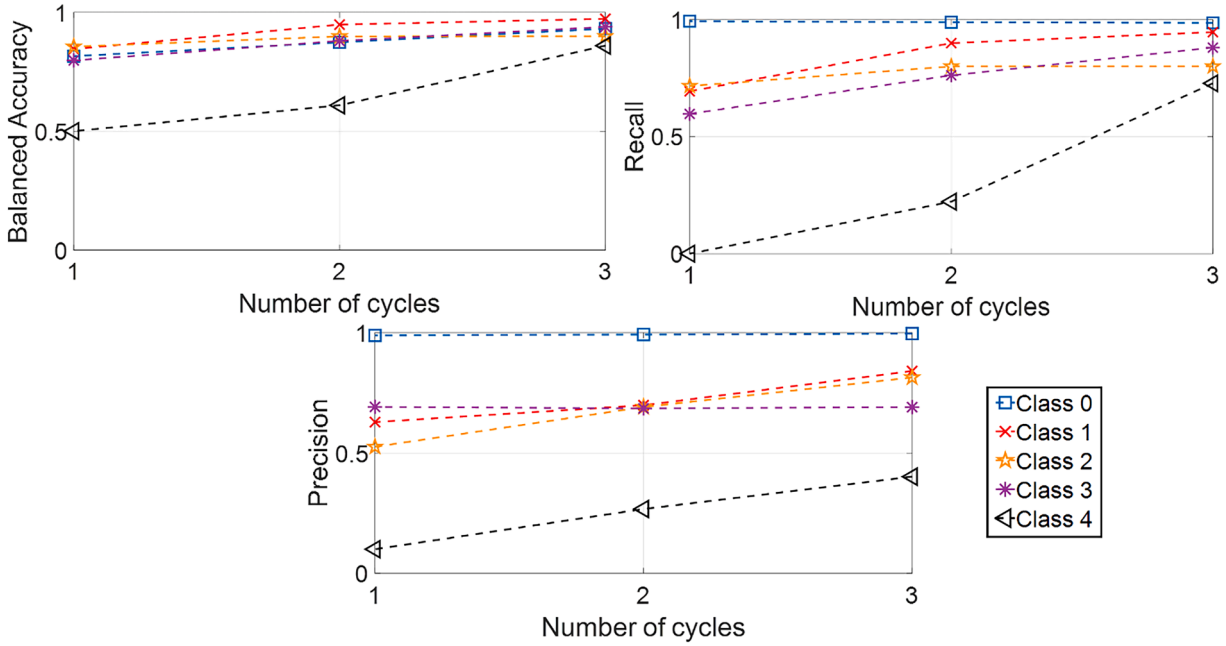


Fig. 13. Quantitative results of the investigated metrics as a function of the excitation cycles obtained in the training phase.

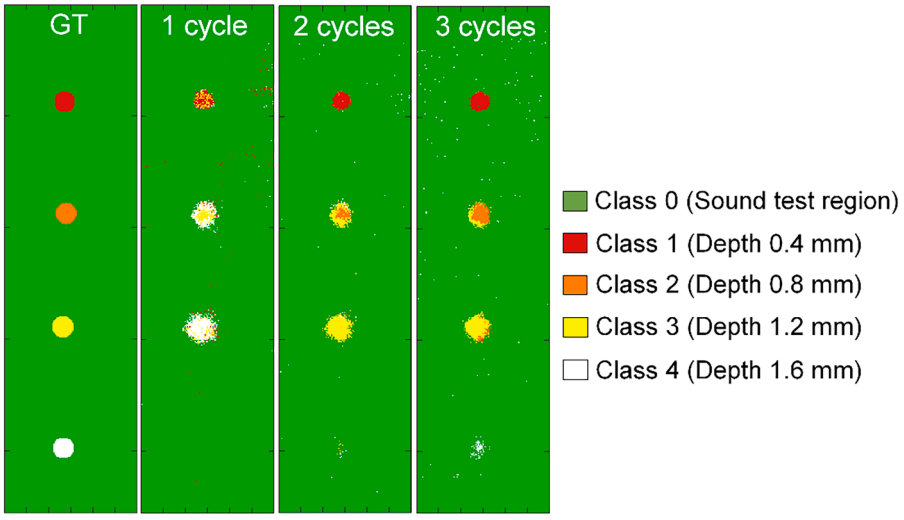


Fig. 14. Classification results of the test phase as a function of the excitation cycles.

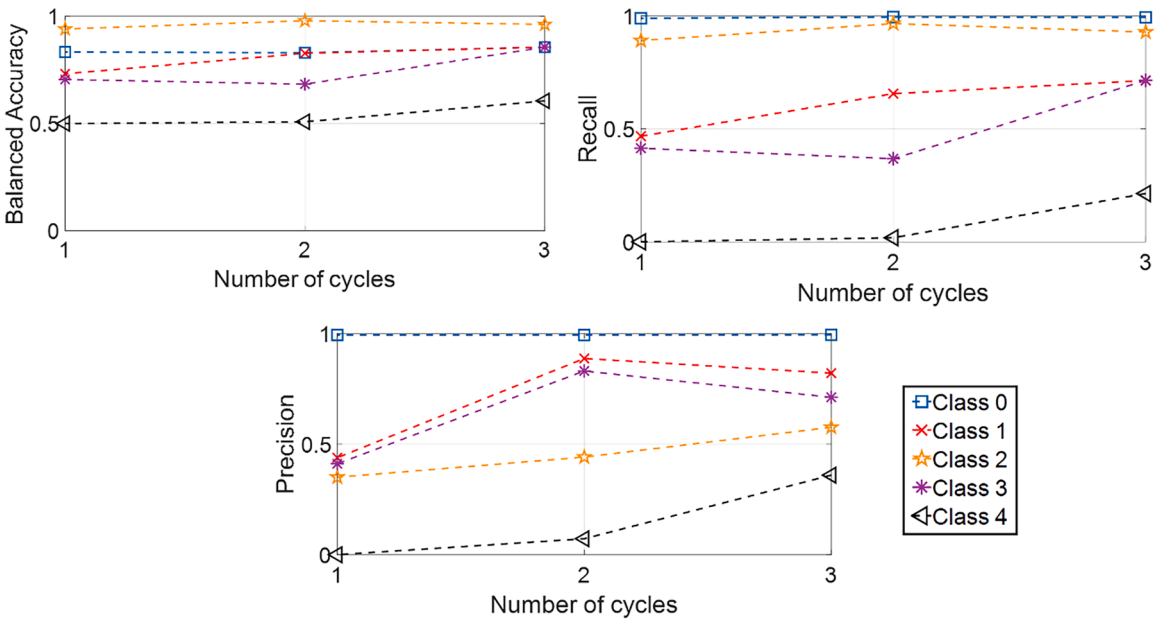


Fig. 15. Quantitative results of the investigated metrics as a function of the excitation cycles obtained in the test phase.

Table 7
Balance accuracy, true positive, and positive predictive values obtained by testing the adopted 1-D CNN as a function of the number of cycles.

	Class	Number of cycles		
		1 cycle	2 cycles	3 cycles
BACCc	Class 0 vs All	0.833	0.830	0.856
	Class 1 vs All	0.732	0.823	0.856
	Class 2 vs All	0.941	0.980	0.962
	Class 3 vs All	0.706	0.684	0.856
	Class 4 vs All	0.500	0.509	0.606
	Average	**0.742**	**0.765**	**0.827**
TPR_C	Class 0 vs All	0.989	0.994	0.993
	Class 1 vs All	0.467	0.656	0.714
	Class 2 vs All	0.891	0.966	0.929
	Class 3 vs All	0.414	0.368	0.714
	Class 4 vs All	n.a.	n.a.	0.214
	Average	**0.552**	**0.597**	**0.713**
PPV_c	Class 0 vs All	0.993	0.993	0.994
	Class 1 vs All	0.439	0.887	0.821
	Class 2 vs All	0.350	0.442	0.576
	Class 3 vs All	0.411	0.831	0.712
	Class 4 vs All	n.a.	0.073	0.359
	Average	**0.439**	**0.645**	**0.693**

almost independent of the number of cycles, showing only a slight improvement with increasing cycles. The effect of the number of cycles on the degradation of the metrics becomes significant as the depth of defects increases. As can be seen, for class 4 the values of the metrics are lower for one cycle while they considerably increase at three cycles.

Fig. 14 shows the qualitative results of the test phase. The same considerations of the learning phase can be extended in the test phase.

Table 8
Number of epochs and time employed by the net architecture as a function of the number of frames per cycle.

Frames/cycle	Epochs	Time [min]
25	401	19
50	341	16
100	183	8
125	226	11
250	202	9
500	203	12

The main difference can be appreciated by looking at one cycle results. In particular, the model identifies the defect *h2*, while the defects *k2* and *m2* are detected but belong to the wrong class. This behavior is undoubtedly due to the low signal-to-noise ratio. The defect *h6* with only one excitation cycle is not detected at all. This defect *h6* does not even detect with two excitation cycles, while with three cycles, only some points are detected.

The quantitative results in Fig. 15 confirm the above qualitative considerations. One excitation cycle is enough to correctly identify the defects belonging to the first three classes. Performing three cycles, the defect *h6* (class 4) is partially detected, with lower values of the metrics with respect to the other classes.

The quantitative results are summarized in Table 7. As can be seen, all the matrics increase as a function of the number of cycles. In order to emphasize this consideration, in Table 7, also the average values are reported. As expected, the average values increase as a function of the investigated parameter.

6.4. *Effect of the number of frames per cycle*

The number of frames per cycle is another lock-in user-controllable parameter that can be set by the operator and can influence the quality of the results. As reported in Table 1, 500 frames have been acquired in each cycle. However, since the composite thermal material response has a low thermal diffusivity, acquiring a huge number of frames per cycle could not be necessary. Indeed, in the works of [14], it has been proven that 100 frames/cycle is a good trade-off value in terms of signal-to-noise ratio and reduction of storage requirements of the thermal data.

To simulate different acquisition frequencies, the original thermal sequences have been subsampled to integer numbers, as shown in Table 2. Also, in this case, each analysis consists of a sequence with a proper length of the thermal data. Therefore, the 1-D CNN has been retrained for each case.

Table 8 shows the training time values for each number of frames per cycle. It is worth noting that the neural network needs more time to build the model with 25 and 50 frames/cycle. This is due that the neural network must learn the model on a dataset made of few examples. For this reason, the cost function optimization needs more time to achieve convergence.

All the metrics are graphically depicted in Fig. 16 as a function of the

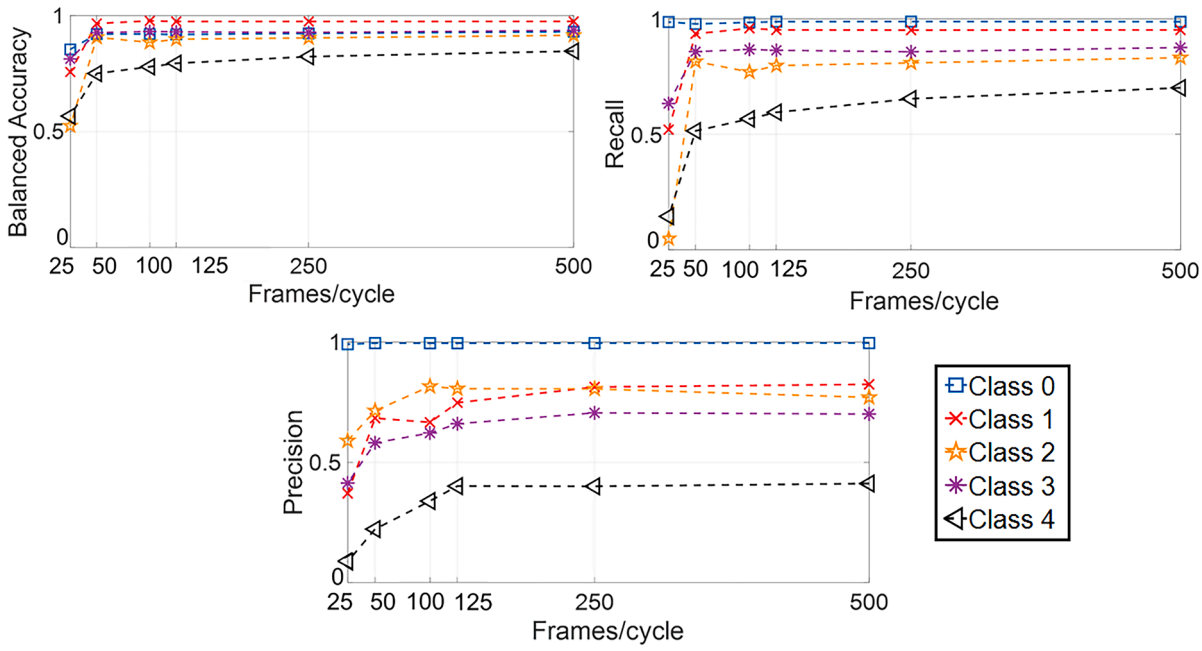


Fig. 16. Quantitative results of the investigated metrics as a function of the number of frames obtained in the training phase.

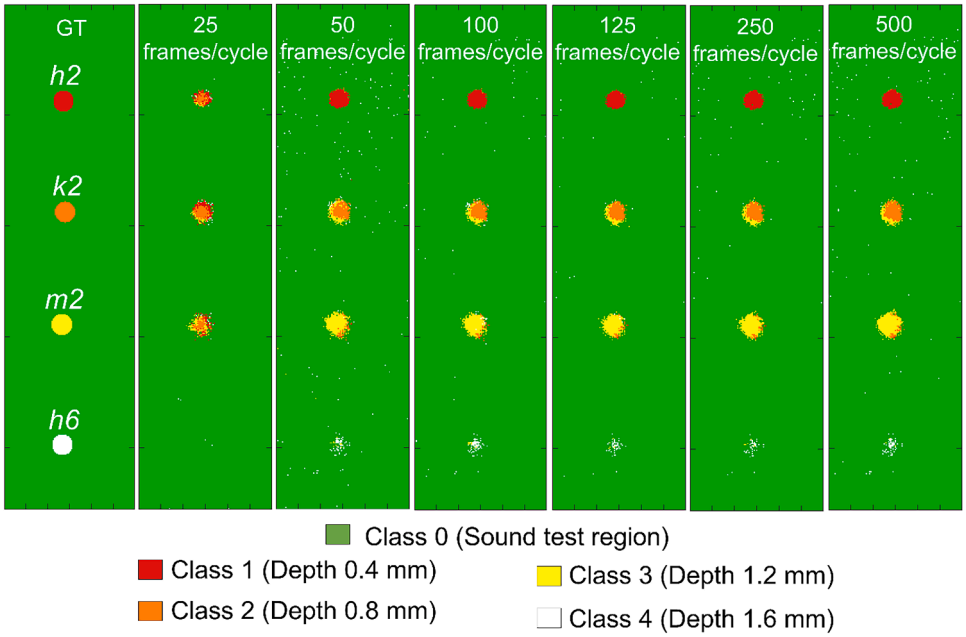


Fig. 17. Classification results of the test phase as a function of the number of frames/cycle.

number of frames per cycle. The BACC values of all investigated classes abruptly decrease for values lower than 100 frames/cycle, and the same considerations can be extended to both recall and precision metrics. Specifically, in the range between 100 and 500 frames, a slight increment in the performance of all investigated metrics can be seen for the defects of class 4. However, from a practical point of view, this analysis shows that 100 frames/cycle is a reasonable threshold value above which the sensitivity of the 1-D CNN in labeling all the defects does not significantly change.

Fig. 17 shows the qualitative results obtained by the test phase. Focusing on the detections obtained between 100 and 500 frames, the defects are similarly labeled. Instead at 25 and 50 frames/cycle, the classification of defects is less accurate and reliable.

These results are confirmed by the quantitative results of the test phase reported in Fig. 18. So, it can be concluded that training the neural network with more than 100 frames/cycles does not significantly increase the value of the three investigated metrics.

The quantitative results obtained in the tests phase are summarized in Table 9. As can be seen, all the metrics increase as a function of the number of frames/cycle.

6.5. *Effect of dropped frames*

Differently from the two previously discussed lock-in parameters, the dropped frame parameter is not user-controllable since it is related to an online event that can occur when the IR camera during the acquisition

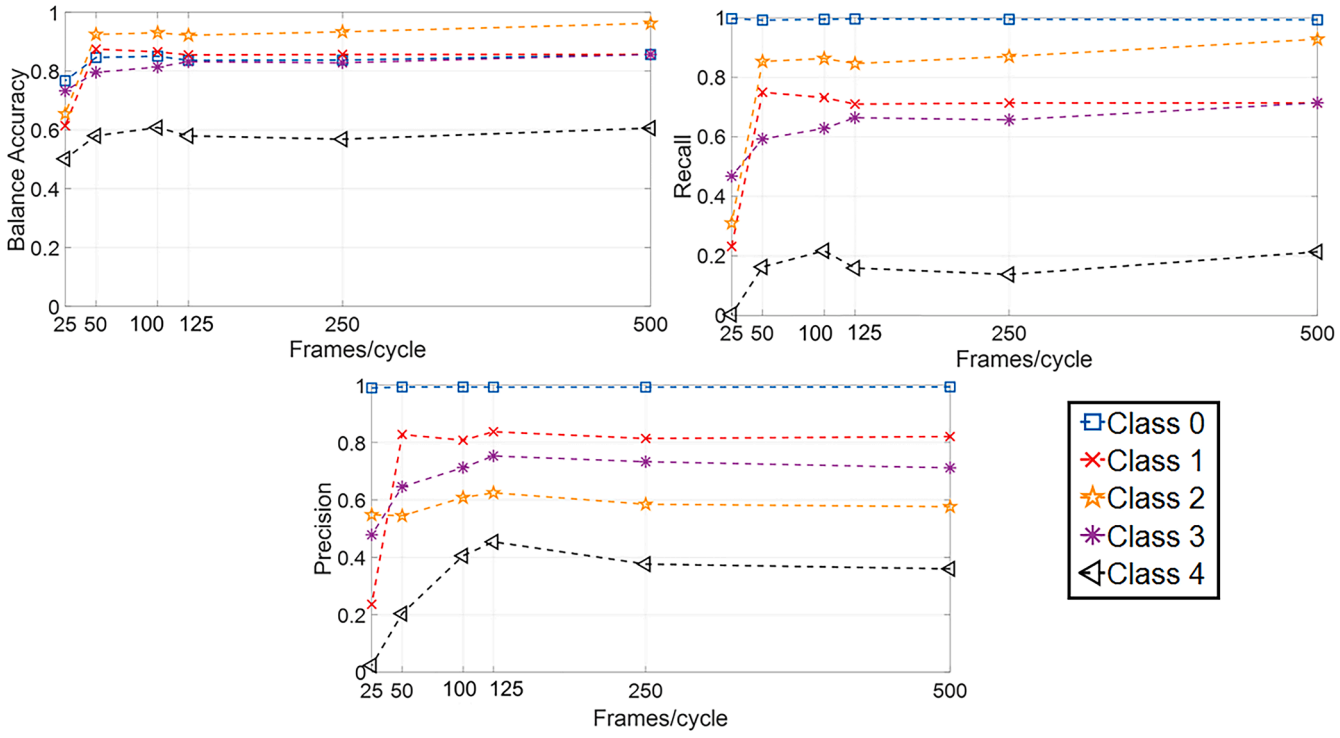


Fig. 18. Quantitative results of the investigated metrics as a function of the number of frames obtained in the test phase.

Table 9
Balance accuracy, true positive, and positive predictive values obtained by testing the adopted 1-D CNN as a function of the number of frames/cycle.

	Class	Number of frames/cycle					
		25	50	100	125	250	500
$BACC_C$	Class 0 vs All	0.767	0.846	0.850	0.8363	0.837	0.856
	Class 1 vs All	0.614	0.875	0.866	0.8547	0.857	0.857
	Class 2 vs All	0.733	0.796	0.814	0.8316	0.828	0.962
	Class 3 vs All	0.655	0.925	0.930	0.9218	0.934	0.857
	Class 4 vs All	0.501	0.579	0.608	0.5793	0.568	0.606
	Average	**0.654**	**0.804**	**0.814**	**0.8047**	**0.805**	**0.828**
TPR_C	Class 0 vs All	0.997	0.992	0.995	0.996	0.995	0.993
	Class 1 vs All	0.232	0.750	0.732	0.710	0.714	0.714
	Class 2 vs All	0.311	0.853	0.864	0.846	0.870	0.928
	Class 3 vs All	0.468	0.593	0.629	0.664	0.657	0.714
	Class 4 vs All	n.a.	0.163	0.217	0.159	0.138	0.214
	Average	**0.402**	**0.670**	**0.687**	**0.675**	**0.675**	**0.713**
PPV_C	Class 0 vs All	0.990	0.994	0.994	0.993	0.993	0.994
	Class 1 vs All	0.236	0.828	0.808	0.821	0.814	0.821
	Class 2 vs All	0.548	0.555	0.608	0.605	0.585	0.576
	Class 3 vs All	0.478	0.646	0.713	0.739	0.733	0.712
	Class 4 vs All	0.233	0.204	0.405	0.416	0.376	0.359
	Average	**0.497**	**0.645**	**0.706**	**0.715**	**0.700**	**0.693**

phase randomly loses some frames. In the following experiments, the model previously trained at 1500 frames for three excitation cycles has been considered. The effort is to evaluate the robustness and sensitivity of this model in labeling the defects when a thermal sequence that has lost some information is provided as an input.

To simulate the online lost events, some frames of the sequences are dropped and substituted with null frames. Several percentages of dropped frames are considered starting from a low value of 2 % up to 30 %. Fig. 19 presents the results obtained in the test phase.

As expected, the classification provides worse results as the dropped frames increase. Comparing the results, the size of all defects seems gradually smaller as the number of dropped frames increases. This behavior is significant above all for the defect of class 4, where the 1-D CNN is able to classify the defect up to 10% of the total frames.

The quantitative results are shown in Fig. 20. The values of the three considered metrics for all the investigated classes when the percentage of dropped frames varies in the range of 0–5 % are slightly affected by the lost frames.

The quantitative results obtained in the tests phase for the dropped frames are summarized in Table 10.

7. Conclusions

This work evaluates the performance of a convolutive neural network as a function of some lock-in test parameters for defect classification in a CFRP specimen. This analysis is fundamental to understanding the applicability of 1-D CNN techniques in a real industrial context. In particular, the number of cycles and frames per cycle has been investigated. Moreover, the performance of the 1-D CNN when frames are dropped (which is a not user-controllable lock-in parameter) has been studied. A neural network with three convolutive layers has been considered. In addition, a parametric study by varying some hyperparameters of this adopted convolutive network has also been done to select its best configuration.

The main results can be summarized as follows:

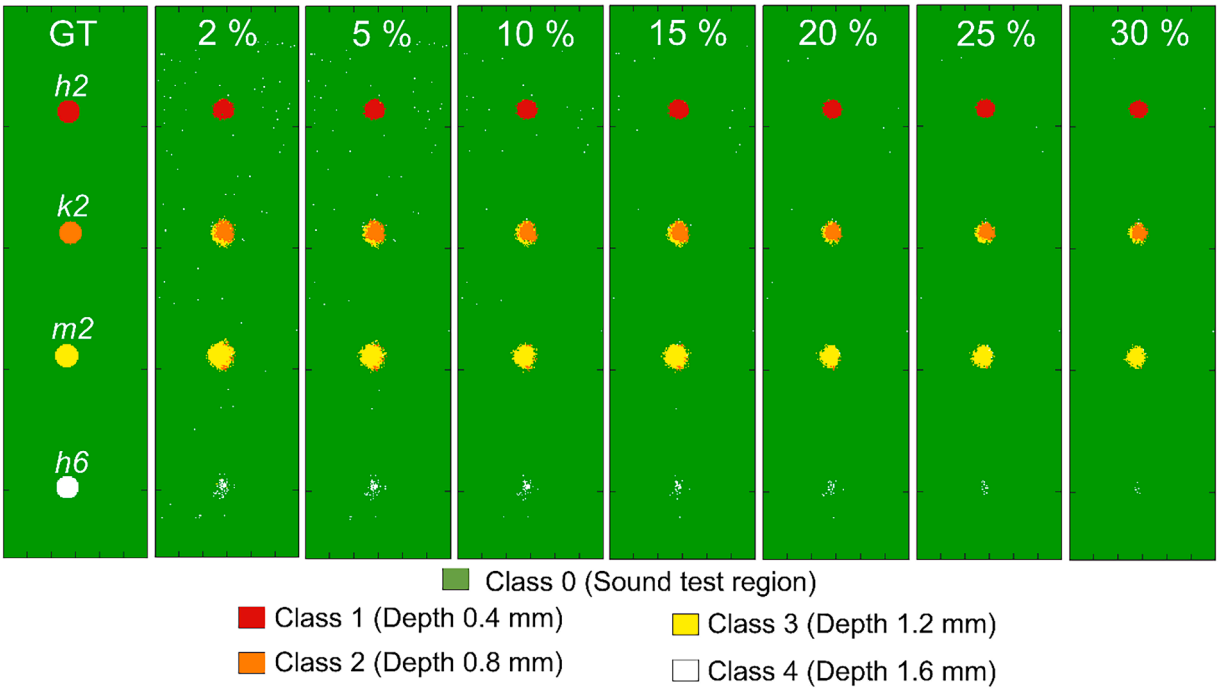


Fig. 19. Classification results of the test phase as a function of the number of dropped frames.

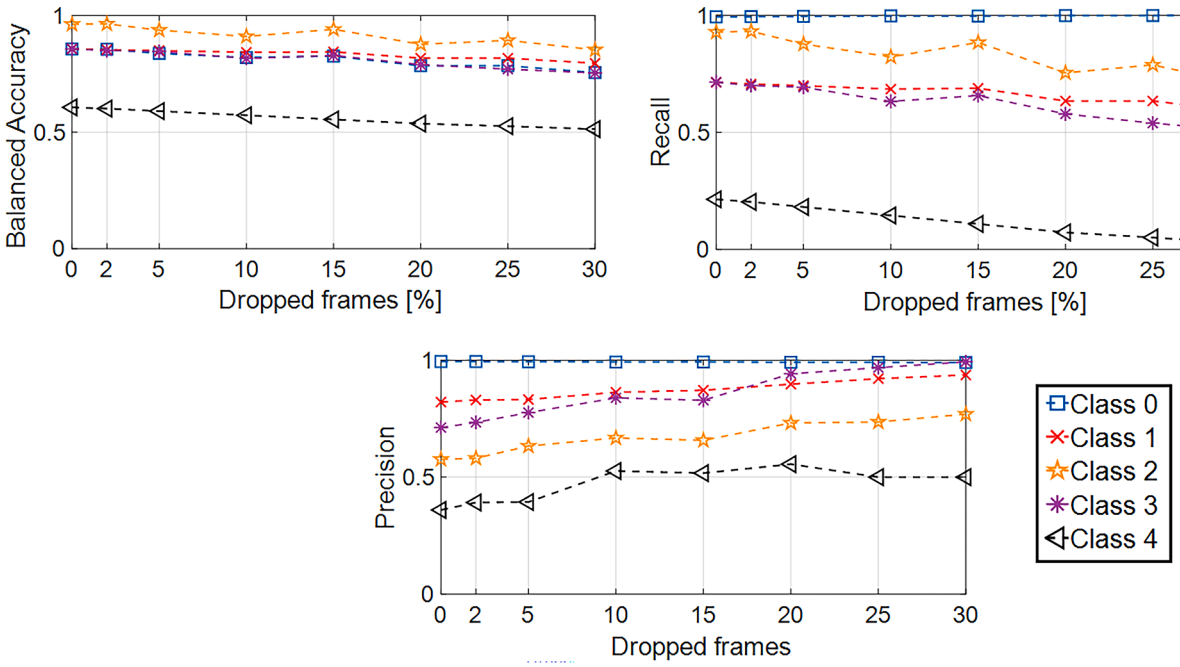


Fig. 20. Quantitative results of the investigated metrics as a function of the dropped frames obtained in the test phase.

- Considering just one excitation cycle (that can reduce the duration of experiments), the adopted 1-D CNN can identify the first three classes of defects both in training and in the test phase. The deepest defects can be detected only with three cycles, even if with a reduced precision than the actual one.
- The investigated metrics do not increase significantly as the number of frames per cycle rises above 100. Acquiring more than 100 frames/cycle to detect and classify defects is unnecessary. This analysis is helpful to reduce the amount of data storage.
- The defect detection performances of the proposed 1-D CNN architecture remain acceptable for up to 5% of dropped frames.

As previously said, the training phase requires some time to build the model that depends on the performance of the personal computer.

However, during the test phase, only a few seconds are required to get quantitative information about the defects. The possibility of getting results very quickly is one of the main advantages of the proposed approach thereby allowing to speed up and automatize the defect detection process. The main limitation of the proposed approach is due to the pre-processing phase which needs the application of both the normalization and median filter. In particular, a median filter with a large kernel size [121,121], useful to remove the non-homogeneous heating, has been chosen. It is worth mentioning that this filter takes

Table 10
Balance accuracy, true positive, and positive predictive values obtained by testing the adopted 1-D CNN as a function of the percentage of dropped frames.

	Class	Percentage of the dropped frames							
		0	2	5	10	15	20	25	30
$BACC_C$	Class 0 vs All	0.856	0.856	0.849	0.847	0.849	0.846	0.839	0.846
	Class 1 vs All	0.857	0.856	0.849	0.850	0.849	0.851	0.849	0.854
	Class 2 vs All	0.962	0.967	0.959	0.947	0.969	0.937	0.939	0.935
	Class 3 vs All	0.857	0.857	0.854	0.842	0.848	0.846	0.848	0.845
	Class 4 vs All	0.606	0.595	0.589	0.576	0.560	0.548	0.516	0.502
	Average	**0.828**	**0.826**	**0.820**	**0.812**	**0.815**	**0.806**	**0.798**	**0.796**
TPR_C	Class 0 vs All	0.993	0.995	0.997	0.998	0.999	0.999	0.999	0.999
	Class 1 vs All	0.714	0.713	0.713	0.713	0.712	0.698	0.698	0.682
	Class 2 vs All	0.928	0.932	0.894	0.883	0.907	0.858	0.877	0.855
	Class 3 vs All	0.714	0.713	0.713	0.692	0.701	0.615	0.553	0.503
	Class 4 vs All	0.214	0.212	0.184	0.176	0.164	0.158	0.104	n.a.
	Average	**0.713**	**0.713**	**0.700**	**0.692**	**0.696**	**0.665**	**0.646**	**0.608**
PPV_C	Class 0 vs All	0.994	0.996	0.995	0.995	0.994	0.993	0.993	0.993
	Class 1 vs All	0.821	0.829	0.832	0.853	0.861	0.866	0.873	0.893
	Class 2 vs All	0.576	0.581	0.625	0.649	0.647	0.722	0.734	0.786
	Class 3 vs All	0.712	0.732	0.797	0.849	0.845	0.882	0.923	0.993
	Class 4 vs All	0.359	0.374	0.373	0.521	0.504	0.529	0.497	0.507
	Average	**0.693**	**0.703**	**0.724**	**0.773**	**0.770**	**0.799**	**0.804**	**0.834**

about 3 h to give the results using a MATLAB (version 2023b) software code. Another limitation is that the proposed 1-D CNN could only partially classify small and deeper defects. The thermographic signals for these small and deep defective regions contain non-significant information to separate them from sound areas. To overcome this problem, further studies will be carried out considering a multifrequency approach in which more excitation periods will be provided as input to the 1-D CNN.

CRediT authorship contribution statement

Tiziana Matarrese: Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Roberto Marani:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Davide Palumbo:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Tiziana D'Orazio:** Writing – review & editing, Supervision. **Umberto Galietti:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors would like to thank DES S.r.l. for providing the specimen and for the support during the experimental activity performed within the "National Operational Programme 2014–2020, PhD programmes for sustainability and innovation-related subjects", Ministerial Decree n. 1061.

Part of the work has been Financed by the European Union - Next-GenerationEU (National Sustainable Mobility Center CN00000023, Italian Ministry of University and Research Decree n. 1033 - 17/06/2022, Spoke 11 - Innovative Materials & Lightweighting). The opinions expressed are those of the authors only and should not be considered as representative of the European Union or the European Commission's official position. Neither the European Union nor the European Commission can be held responsible for them.

References

[1] Evans EE, Brooks RA, Liu J, Hall ZEC, Liu H, Lowe TJE, et al. Comparison of X-ray computed tomography and ultrasonic C-scan techniques and numerical modelling of impact damage in a CFRP composite laminate. Appl Compos Mater 2023. https://doi.org/10.1007/s10443-023-10171-3.
[2] Huang J, Wei Q, Zhuo L, Zhu J, Li C, Wang Z. Detection and quantification of artificial delaminations in CFRP composites using ultrasonic thermography. Infrared Phys Technol 2023;130. https://doi.org/10.1016/j.infrared.2023.104579.
[3] Guo B, Zheng X, Gerini-Romagnoli M, Yang L. Digital shearography for NDT: determination and demonstration of the size and the depth of the smallest detectable defect. NDT E Int 2023;139. https://doi.org/10.1016/j.ndteint.2023.102927.
[4] Maldague XP V. Nondestructive evaluation of materials by infrared thermography. London: Springer London; 1993. https://doi.org/10.1007/978-1-4471-1995-1.
[5] Breitenstein O, Warta W, Langenkamp M. Lock-in thermography. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. https://doi.org/10.1007/978-3-642-02417-7. vol. 10.
[6] Meola C, Boccardi S, Carlomagno GM, Boffa ND, Monaco E, Ricci F. Nondestructive evaluation of carbon fibre reinforced composites with infrared thermography and ultrasonics. Compos Struct 2015;134:845–53. https://doi.org/10.1016/j.compstruct.2015.08.119.
[7] EN 17119. Non-destructive testing – thermographic testing – active thermography, 2018, p. 15.
[8] Almond DP, Pickering SG. An analytical study of the pulsed thermography defect detection limit. J Appl Phys 2012;111. https://doi.org/10.1063/1.4704684.
[9] Chatterjee K, Tuli S, Pickering SG, Almond DP. A comparison of the pulsed, lock-in and frequency modulated thermography nondestructive evaluation techniques. NDT E Int 2011;44:655–67. https://doi.org/10.1016/j.ndteint.2011.06.008.
[10] Pickering S, Almond D. Matched excitation energy comparison of the pulse and lock-in thermography NDE techniques. NDT E Int 2008;41:501–9. https://doi.org/10.1016/j.ndteint.2008.05.007.
[11] Almond DP, Angioni SL, Pickering SG. Long pulse excitation thermographic non-destructive evaluation. NDT E Int 2017;87:7–14. https://doi.org/10.1016/j.ndteint.2017.01.003.
[12] Shepard S.M. Advances in pulsed thermography. In: Rozlosnik A.E., Dinwiddie R. B., editors., 2001, p. 511–5. doi:10.1117/12.421032.
[13] Pitarresi G. Lock-in signal post-processing techniques in infra-red thermography for materials structural evaluation. Exp Mech 2015;55:667–80. https://doi.org/10.1007/s11340-013-9827-1.
[14] Matarrese T, Palumbo D, Galietti U. Comparison in the transient regime of four lock-in thermography algorithms by means of synthetic and experimental data on CFRP. NDT E Int 2023;139. https://doi.org/10.1016/j.ndteint.2023.102925.
[15] Mulaveesala R, Tuli S. Theory of frequency modulated thermal wave imaging for nondestructive subsurface defect detection. Appl Phys Lett 2006;89. https://doi.org/10.1063/1.2382738.
[16] Maierhofer C, Röllig M, Krankenhagen R, Myrach P. Comparison of quantitative defect characterization using pulse-phase and lock-in thermography. Appl Opt 2016;55:D76. https://doi.org/10.1364/AO.55.000D76.
[17] Delanthabettu S, Menaka M, Venkatraman B, Raj B. Defect depth quantification using lock-in thermography. Quant Infrared Thermogr J 2015;12:37–52. https://doi.org/10.1080/17686733.2015.1013663.
[18] Montanini R. Quantitative determination of subsurface defects in a reference specimen made of Plexiglas by means of lock-in and pulse phase infrared

thermography. Infrared Phys Technol 2010;53:363–71. https://doi.org/10.1016/j.infrared.2010.07.002.
[19] Matarrese T, Palumbo D, Santonicola G, Ancona F, Galietti U. Comparison among four lock-in algorithms in transient regime on CFRP. In: 2023 IEEE 10th International Workshop on Metrology for AeroSpace (MetroAeroSpace). IEEE; 2023. p. 677–82. https://doi.org/10.1109/MetroAeroSpace57412.2023.10189933.
[20] Busse G, Wu D, Karpen W. Thermal wave imaging with phase sensitive modulated thermography. J Appl Phys 1992;71:3962–5. https://doi.org/10.1063/1.351366.
[21] Bennett CA, Patty RR. Thermal wave interferometry: a potential application of the photoacoustic effect. Appl Opt 1982;21:49. https://doi.org/10.1364/AO.21.000049.
[22] Palumbo D. On the thickness quantification of composite materials by using lock-in thermography. Materials 2019;12. https://doi.org/10.3390/ma12071185.
[23] Salazar A, Sagarduy-Marcos D, Rodríguez-Aseguinolaza J, Mendioroz A, Celorrio R. Characterization of semi-infinite delaminations using lock-in thermography: theory and numerical experiments. NDT E Int 2023;138. https://doi.org/10.1016/j.ndteint.2023.102883.
[24] Ekanayake S, Gurram S, Schmitt RH. Depth determination of defects in CFRP-structures using lock-in thermography. Compos B 2018;147:128–34. https://doi.org/10.1016/j.compositesb.2018.04.032.
[25] Jinlong G, Junyan L, Fei W, Yang W. Inverse heat transfer approach for nondestructive estimation the size and depth of subsurface defects of CFRP composite using lock-in thermography. Infrared Phys Technol 2015;71:439–47. https://doi.org/10.1016/j.infrared.2015.06.005.
[26] Khan S, Rahmani H, Shah SAA, Bennamoun M. A guide to convolutional neural networks for computer vision. Cham: Springer International Publishing; 2018. https://doi.org/10.1007/978-3-031-01821-3.
[27] Gholizadeh S. Deep learning and machine learning techniques in advanced non-destructive testing. J Harbin Eng Univ 2023;44:977–93.
[28] He Y, Deng B, Wang H, Cheng L, Zhou K, Cai S, et al. Infrared machine vision and infrared thermography with deep learning: a review. Infrared Phys Technol 2021;116. https://doi.org/10.1016/j.infrared.2021.103754.
[29] Dong Y, Zhao B, Yang J, Cao Y, Cao Y. Two-stage convolutional neural network for joint removal of sensor noise and background interference in lock-in thermography. NDT E Int 2023;137. https://doi.org/10.1016/j.ndteint.2023.102816.
[30] Gao Y, Gao M, Wang F, Liu J, Yang F, Sheng J, et al. Convolution neural network fusion high power halogen lamp induced thermal-wave diffusion multidimensional features: an intelligent defect identification approach for aviation honeycomb sandwich composites. Infrared Phys Technol 2023;133:104772. https://doi.org/10.1016/j.infrared.2023.104772.
[31] Saeed N, King N, Said Z, Omar MA. Automatic defects detection in CFRP thermograms, using convolutional neural networks and transfer learning. Infrared Phys Technol 2019;102. https://doi.org/10.1016/j.infrared.2019.103048.
[32] Fang Q, Ibarra-Castanedo C, Maldague X. Automatic defects segmentation and identification by deep learning algorithm with pulsed thermography: synthetic and experimental data. Big Data Cognit Comput 2021;5:9. https://doi.org/10.3390/bdcc5010009.
[33] Marani R, Palumbo D, Galietti U, D'Orazio T. Deep learning for defect characterization in composite laminates inspected by step-heating thermography. Opt Lasers Eng 2021;145. https://doi.org/10.1016/j.optlaseng.2021.106679.
[34] Cheng X, Chen P, Wu Z, Cech M, Ying Z, Hu X. Automatic detection of CFRP subsurface defects via thermal signals in long pulse and lock-in thermography. IEEE Trans Instrum Meas 2023;72. https://doi.org/10.1109/TIM.2023.3277996.
[35] Luo Q, Gao B, Woo WL, Yang Y. Temporal and spatial deep learning network for infrared thermal defect detection. NDT E Int 2019;108:102164. https://doi.org/10.1016/j.ndteint.2019.102164.
[36] Schmid S, Reinhardt J, Grosse CU. Spatial and temporal deep learning for defect detection with lock-in thermography. NDT E Int 2024:103063. https://doi.org/10.1016/j.ndteint.2024.103063.
[37] Dong Y, Xia C, Yang J, Cao Y, Cao Y, Li X. Spatio-temporal 3-D residual networks for simultaneous detection and depth estimation of CFRP subsurface defects in lock-in thermography. IEEE Trans Industr Inform 2022;18:2571–81. https://doi.org/10.1109/TII.2021.3103019.
[38] Morelli D, Marani R, D'Accardi E, Palumbo D, Galietti U, D'Orazio T. A convolution residual network for heating-invariant defect segmentation in composite materials inspected by lock-in thermography. IEEE Trans Instrum Meas 2021;70. https://doi.org/10.1109/TIM.2021.3116300.
[39] Yen SJ, Lee YS. Cluster-based under-sampling approaches for imbalanced data distributions. Expert Syst Appl 2009;36:5718–27. https://doi.org/10.1016/j.eswa.2008.06.108.