POLITECNICO DI TORINO
Repository ISTITUZIONALE

A physics-informed framework for feature extraction and defect segmentation in pulsed infrared thermography

*Original*
A physics-informed framework for feature extraction and defect segmentation in pulsed infrared thermography / Santoro, L., Sesana, R.. - In: ENGINEERING FAILURE ANALYSIS. - ISSN 1350-6307. - 175:(2025). [10.1016/j.engfailanal.2025.109542]

*Availability:*
This version is available at: 11583/2998636 since: 2025-06-08T12:58:17Z

*Publisher:*
Elsevier

*Published*
DOI:10.1016/j.engfailanal.2025.109542

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

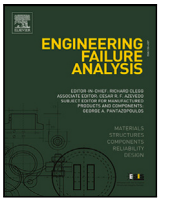
16 November 2026

Engineering Failure Analysis 175 (2025) 109542

Contents lists available at ScienceDirect

Engineering Failure Analysis

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



# A physics-informed framework for feature extraction and defect segmentation in pulsed infrared thermography

Luca Santoro *, Raffaella Sesana

*Politecnico di Torino, Corso Duca degli Abruzzi, 24, Turin, 10129, Italy*

ARTICLE INFO

*Keywords:*
Infrared thermography
Defect detection
Nondestructive testing
Physics-based feature engineering
Random forest
Engineering failure analysis
Spatial augmentation
PVC

ABSTRACT

This paper presents a robust and interpretable methodology for defect detection in active infrared thermography data applied to polyvinyl chloride (PVC) specimens. Our approach integrates a physics-based cooling model to describe the transient thermal response of each pixel, from which five primary temporal features are extracted via least-squares fitting. These features are then enriched with local spatial statistics through neighborhood-based computations, resulting in a 15-dimensional descriptor per pixel. The resulting feature set is used to train a random forest classifier, which achieves high overall accuracy (99.3%), competitive intersection-over-union (0.705), and an outstanding ROC AUC (0.998). In contrast to deep encoder–decoder networks that require extensive computational resources and large annotated datasets, the proposed pipeline offers enhanced interpretability and significantly reduced computational overhead. Comparative analysis with state-of-the-art deep learning models, such as those reported in Wei et al., (2023), demonstrates that our approach achieves similar performance while providing a transparent insight into the contribution of each feature. The proposed method is especially suitable for engineering failure analysis where model transparency, rapid evaluation, and integration into existing inspection protocols are critical. Future work will extend the framework to accommodate a broader range of defect types and material systems, aiming to further enhance industrial applicability and diagnostic reliability.

## 1. Introduction

Nondestructive testing (NDT) has become an indispensable tool in ensuring the reliability and safety of critical engineering components by detecting subsurface and internal defects without compromising structural integrity [1–3]. Over the past decades, a variety of NDT techniques have been developed—including ultrasonic testing [4,5], X-ray computed tomography [6,7], and infrared thermography (IRT) [8–12]—each offering distinct advantages in terms of speed, cost, and inspection versatility [13,14].

Among these techniques, IRT has emerged as a particularly attractive method due to its rapid, contact-free, and cost-effective nature [15–17]. Recent research has demonstrated that IRT can be effectively combined with advanced data analysis techniques to enhance defect detectability in a range of materials—from weld joints and composite laminates to complex cultural heritage artifacts [18,19]. In parallel, deep learning approaches have gained prominence in NDT, providing automated solutions that reduce reliance on manual interpretation while increasing detection accuracy [7,20–23].

Convolutional neural networks (CNNs) and their variants, such as U-Net, Mask R-CNN, and two-stream architectures, have been successfully applied to defect segmentation and classification in IRT and other imaging modalities [9,24,25]. Complementary to

* Corresponding author.
*E-mail address:* luca.santoro@polito.it (L. Santoro).
*URL:* https://lucasantoro97.github.io/cv/ (L. Santoro).

https://doi.org/10.1016/j.engfailanal.2025.109542
Received 15 February 2025; Received in revised form 15 March 2025; Accepted 17 March 2025
Available online 25 March 2025
1350-6307/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

these methods, generative adversarial networks (GANs) have been employed to augment data and enhance the quality of thermal images, thereby addressing challenges posed by limited datasets and noisy backgrounds [16,26]. In addition, recent studies have integrated transfer learning and ensemble-based strategies to further improve defect detection performance in both industrial and aerospace applications [18,27].

Beyond purely data-driven approaches, there is an emerging trend toward incorporating physical principles into machine learning models. Physics-informed neural networks (PINNs) have been proposed to embed heat diffusion laws into the learning process, thereby enhancing model robustness and interpretability in thermographic inspections [28–33]. At the same time, hybrid methods which combine handcrafted features with deep learning representations have proven effective in mitigating issues related to nonuniform heating and background disturbances [34,35].

The challenge of limited and imbalanced training data has motivated the use of synthetic data generation and advanced data augmentation techniques. Finite element method (FEM) based simulations have been leveraged to generate virtual thermographic datasets which closely mimic real-world conditions [36,37], while transfer learning strategies have allowed models pretrained on large datasets to be adapted for NDT applications with limited sample sizes [38,39]. Moreover, multimodal fusion approaches that combine IRT with complementary imaging modalities, such as CT and ultrasonic testing, have further enriched the defect characterization process [40,41].

Advances in network architectures have also contributed significantly to the field. Recent work has introduced self-attention mechanisms and Transformer-based models to capture both spatial and temporal features in thermographic data [42,43], while efficient deep learning models based on DenseNet, ResNet, and attention-enhanced U-Nets have achieved remarkable performance in classifying defects across diverse materials [44–48]. In parallel, studies on defect quantification and the reconstruction of defect shapes have shown promising results by integrating segmentation networks with regression and LSTM modules [49,50].

Other complementary NDT approaches have been explored in the literature. For instance, methods based on magnetic flux leakage (MFL) [27,51], microwave imaging [5], and resonant acoustic testing [40] have further expanded the scope of automated defect detection. Furthermore, recent investigations into the application of deep learning for defect detection in cultural heritage [52,53] and infrastructure inspection [54,55] underscore the versatility of these methods across various sectors.

Despite these advances, several challenges remain. Many state-of-the-art methods demand substantial computational resources and large annotated datasets, and some struggle with the interpretability of their predictions [56,57]. In addition, nonuniform background conditions, limited training samples, and noise in thermographic data continue to impede optimal defect detection performance [58–60]. Recent efforts to enhance defect detection through adaptive background suppression and signal enhancement techniques have shown promise but often require further refinement to achieve industrial-level robustness [61].

In this context, our work proposes an alternative strategy which combines physics-based feature extraction with efficient machine learning classification. Specifically, we introduce a novel methodology that models the cooling behavior of thermographic sequences using a parametric function, extracts five primary temporal features, and augments these with local spatial statistics to form a 15-dimensional descriptor for each pixel. These descriptors are subsequently employed to train a random forest classifier, which achieves competitive accuracy and a near-perfect ROC AUC while offering enhanced interpretability and reduced computational overhead compared to deep encoder–decoder networks [26,62,63].

The aim of our paper is to present a robust, transparent, and computationally efficient approach for defect detection in infrared thermography data, specifically applied to PVC specimens. By integrating a physics-based cooling model with spatial augmentation and leveraging classical machine learning techniques, our method provides an effective alternative for industrial defect detection and engineering failure analysis. We demonstrate that our approach not only achieves high classification performance but also offers significant advantages in terms of model interpretability and ease of deployment, thereby contributing to the advancement of practical NDT methodologies for critical engineering applications. (see Table 1).

## 2. Materials and methods

This work uses a thermal imaging dataset from [64] containing nineteen distinct polyvinyl chloride (PVC) specimens, each featuring subsurface defects in the form of cylindrical blind holes of varying depths and diameters. For each specimen, an infrared camera was used to record a sequence of thermal frames over time, yielding three-dimensional data of size $H \times W \times N_t$, where $H$ and $W$ denote the height and width in pixels, respectively, and $N_t$ is the number of frames in the sequence. The temporal sampling was 10 Hz, yielding approximately 181 seconds of cooling data per specimen immediately following a short heating pulse provided by flash lamps. The thermal camera used is a FLIR SC5000 with a spectral range of $3 - 5\mu m$, the heat source are two Broncolor lamps with a power of 6.4 KJ. The camera captured the infrared video with a resolution of $320 \times 256$ and recorded it for 181 s at 10 Hz, resulting in a video with a dimension of $320 \times 256 \times 1810$ for each measurement.

Let us denote the measured temperature at pixel coordinates $(j, k)$ and discrete time $t_f$ by

$$T(j, k, t_f), \quad 1 \le j \le H, \quad 1 \le k \le W, \quad 1 \le f \le N_t. \tag{1}$$

We store these measurements in a three-dimensional array $\mathbf{Z} \in \mathbb{R}^{H \times W \times N_t}$. Additionally, we define a corresponding ground-truth mask $\mathbf{M} \in \{0, 1\}^{H \times W}$, where

$$\mathbf{M}(j, k) = \begin{cases} 1, & \text{if pixel } (j, k) \text{ belongs to a defect,} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

**Table 1**
Nomenclature.

| Symbol | Description |
|---|---|
| $T(j,k,t_f)$ | Measured temperature at pixel $(j,k)$ and time $t_f$ |
| $H$ | Image height (number of pixels) |
| $W$ | Image width (number of pixels) |
| $N_t$ | Number of frames in the thermographic sequence |
| $t_f$ | Discrete time instance corresponding to frame $f$ |
| $\mathbf{Z} \in \mathbb{R}^{H\times W\times N_t}$ | 3D temperature data array |
| $\mathbf{M} \in \{0,1\}^{H\times W}$ | Ground-truth mask (1: defect, 0: healthy) |
| $a$ | Baseline parameter in the cooling model |
| $b$ | Cooling coefficient in the cooling model |
| $\widehat{T}_f$ | Estimated temperature at time $t_f$, given by $a + \frac{b}{\sqrt{t_f}}$ |
| $T_f$ | Measured temperature at time $t_f$ for a given pixel |
| $r_f$ | Residual: $r_f = T_f - \widehat{T}_f$ |
| RMSE | $\sqrt{\frac{1}{N_t}\sum_{f=1}^{N_t} r_f^2}$ |
| MaxAbsDeriv | $\max_{1\le f<N_t} \left\lvert T_{f+1} - T_f \right\rvert$ |
| Var($T$) | $\frac{1}{N_t}\sum_{f=1}^{N_t}\left(T_f - \overline{T}\right)^2$ |
| $\overline{T}$ | $\frac{1}{N_t}\sum_{f=1}^{N_t} T_f$ |
| $\mathcal{F}(j,k)$ | Generic feature map at pixel $(j,k)$ |
| $\Omega$ | Spatial neighborhood for local feature computation |
| $\lvert\Omega\rvert$ | Number of pixels in $\Omega$ |
| $\mu_\Omega(j,k)$ | Local mean over $\Omega$ at $(j,k)$ |
| $\sigma_\Omega(j,k)$ | Local standard deviation over $\Omega$ at $(j,k)$ |
| $\mathbf{x}_i \in \mathbb{R}^{15}$ | 15-dimensional feature vector for pixel $i$ |
| $y_i \in \{0,1\}$ | Class label for pixel $i$ (0: healthy, 1: defect) |
| $N$ | Total number of pixel samples |
| $T^{(m)}(\mathbf{x})$ | Prediction of the $m$th decision tree |
| $M$ | Total number of trees in the random forest |
| $\hat{y}(\mathbf{x})$ | Predicted class (by majority vote) |
| $p^{(m)}(\mathbf{x})$ | Defect probability from the $m$th tree |
| $p(\mathbf{x})$ | Averaged probability: $\frac{1}{M}\sum_{m=1}^{M} p^{(m)}(\mathbf{x})$ |
| $H(S)$ | Entropy: $-\sum_{c\in\{0,1\}} p_c \log_2(p_c)$ |
| $p_c$ | Fraction of samples in $S$ of class $c$ |
| $\mathcal{D}_{\text{train}}$ | Training dataset |
| $\mathcal{D}_{\text{test}}$ | Test dataset |
| $p_i$ | Predicted probability for pixel $i$ |
| $H(p_i)$ | $-\left[p_i \ln(p_i) + (1-p_i)\ln(1-p_i)\right]$ |
| $\hat{p}(j,k)$ | Raw probability at pixel $(j,k)$ |
| $\widetilde{p}(j,k)$ | Smoothed probability after median filtering |
| $\Omega_m$ | Neighborhood for median filtering |
| $\tau$ | Threshold for segmentation (e.g., 0.5) |
| Seg$(j,k)$ | Binary segmentation map |
| $\mathcal{B}$ | Structuring element for morphological operations |
| IoU | Intersection over Union |
| Accuracy | Overall classification accuracy |
| Precision | Ratio of true positives to predicted positives |
| Recall | Ratio of true positives to actual positives |
| F1 score | Harmonic mean of precision and recall |
| ROC AUC | Area under the ROC curve |

These masks were obtained from prior knowledge of the cylindrical holes manufactured into the PVC specimens and confirmed through reference images.

Thermal transients under adiabatic or near-adiabatic assumptions often exhibit a $t^{-1/2}$ cooling behavior immediately after a uniform heat pulse. We thus employ a simplified parametric model at each pixel $(j,k)$:

$$T(t) \approx a + \frac{b}{\sqrt{t}}, \tag{3}$$

where $a$ and $b$ are scalar parameters. Denoting discrete times by $\{t_1, t_2, \ldots, t_{N_t}\}$, we perform a least-squares fit to the measured data

$$\min_{a,b} \sum_{f=1}^{N_t} \left( T_f - \left( a + \frac{b}{\sqrt{t_f}} \right) \right)^2, \tag{4}$$

with

$$T_f = T(j,k,t_f).$$

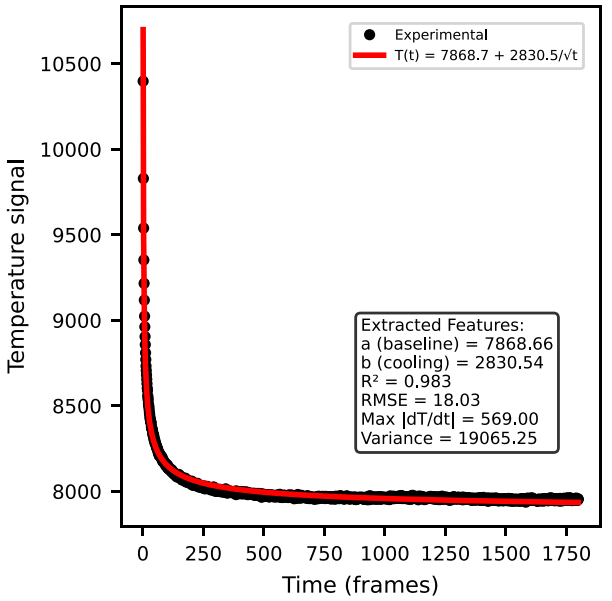



**Fig. 1.** Fitting of the temperature decay curve to experimental data.

Once $(a, b)$ is estimated, we compute additional temporal features. Let

$$\widehat{T}_f = a + \frac{b}{\sqrt{t_f}} \quad \text{and} \quad r_f = T_f - \widehat{T}_f. \tag{5}$$

We define the root-mean-square error

$$\text{RMSE} = \sqrt{\frac{1}{N_t}\sum_{f=1}^{N_t} r_f^2}, \tag{6}$$

the maximum absolute derivative

$$\text{MaxAbsDeriv} = \max_{1\le f<N_t} \left|T_{f+1} - T_f\right|, \tag{7}$$

and the temporal variance

$$\text{Var}(T) = \frac{1}{N_t}\sum_{f=1}^{N_t}\left(T_f - \overline{T}\right)^2, \quad \text{where} \quad \overline{T} = \frac{1}{N_t}\sum_{f=1}^{N_t} T_f. \tag{8}$$

Fig. 1 shows an example of fitting the experimental cooling curve data.

Thus, for each pixel, a feature vector

$$\left[a,\ b,\ \text{RMSE},\ \text{MaxAbsDeriv},\ \text{Var}(T)\right] \tag{9}$$

encodes its local cooling characteristics.

Thermographic data can exhibit spatial gradients and local texture variations due to defects, edges, or nonuniform heating. To incorporate local context, we convolve each feature map (viewed as a two-dimensional scalar field) with a small spatial kernel of size $p \times p$. For a generic feature map $\mathcal{F}(j, k)$, the local mean is defined as

$$\mu_\Omega(j,k) = \frac{1}{|\Omega|}\sum_{(u,v)\in\Omega} \mathcal{F}(j+u, k+v), \tag{10}$$

where $\Omega$ is the neighborhood around $(j, k)$ and $|\Omega|$ is its cardinality. Similarly, the local standard deviation is computed by

$$\sigma_\Omega(j,k) = \sqrt{\mu_\Omega(j,k;\mathcal{F}^2) - \left(\mu_\Omega(j,k;\mathcal{F})\right)^2}, \tag{11}$$

with $\mu_\Omega(j, k; \mathcal{F}^2)$ representing the local mean of $\mathcal{F}^2$. Concatenating these local means and standard deviations with the original features captures spatial regularities that enhance subsequent classification.

To clarify the final dimensionality, each of the five base features (baseline $a$, cooling coefficient $b$, RMSE, MaxAbsDeriv, and $\text{Var}(T)$) has an associated local mean and local standard deviation (see Fig. 2). This yields a total of 15 features per pixel (5 base features + 5 local means + 5 local standard deviations), as outlined in Table 2.

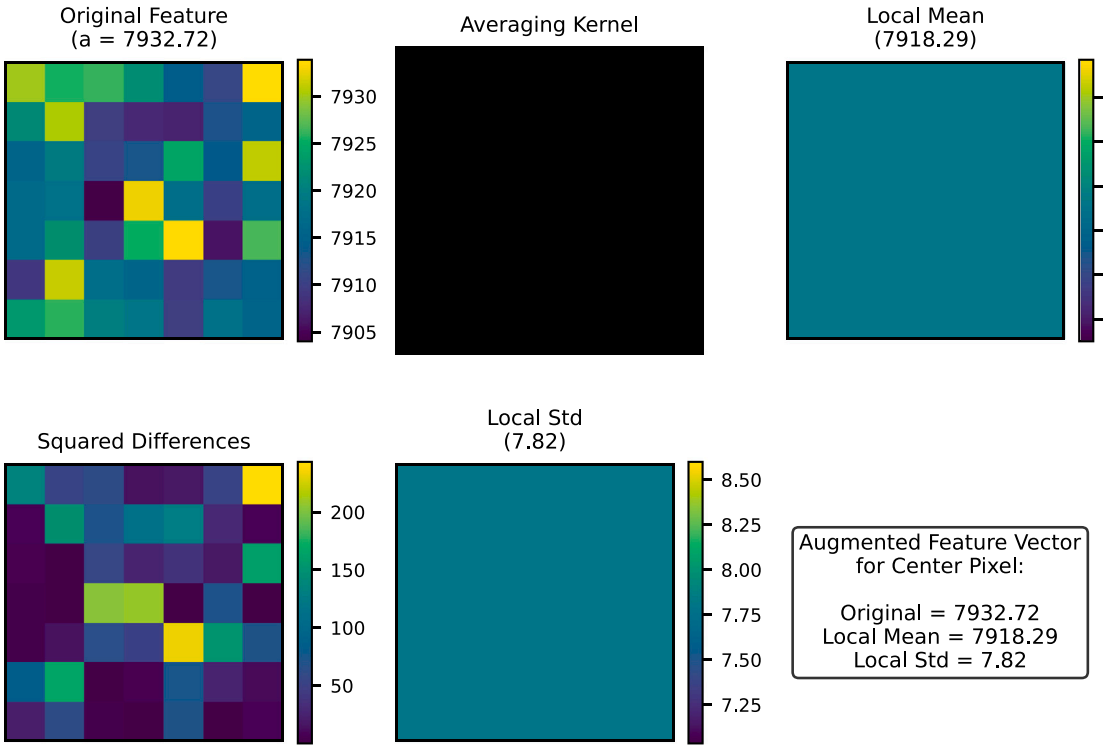



**Fig. 2.** Local context feature extraction process.

**Table 2**
Overview of the 15 extracted features per pixel.

| Feature index | Description |
|---|---|
| 1 | Baseline term $a$ |
| 2 | Cooling coefficient $b$ |
| 3 | RMSE of the $\sqrt{t}$-based fit |
| 4 | Maximum absolute derivative |
| 5 | Temporal variance of the pixel sequence |
| 6–10 | Local means of features 1 to 5 |
| 11–15 | Local standard deviations of features 1 to 5 |

Once all pixelwise features have been extracted for every specimen, we label each pixel according to the ground-truth mask $\mathbf{M}$. Let $\mathbf{x}_i \in \mathbb{R}^d$ denote the $d$-dimensional feature vector for pixel $i$, and let $y_i \in \{0, 1\}$ be its corresponding label (healthy or defect). The aggregated dataset is $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$. A random forest is an ensemble of decision trees, each trained on a bootstrapped subset of the data. Denote by $T^{(m)}(\mathbf{x})$ the $m$th decision tree in an ensemble of $M$ trees. The random forest prediction is obtained by majority vote:

$$\hat{y}(\mathbf{x}) = \text{Mode}\Big(T^{(1)}(\mathbf{x}),\, T^{(2)}(\mathbf{x}),\, \ldots,\, T^{(M)}(\mathbf{x})\Big), \tag{12}$$

or, in a probabilistic sense, by averaging the probability estimates

$$p(\mathbf{x}) = \frac{1}{M}\sum_{m=1}^{M} p^{(m)}(\mathbf{x}), \tag{13}$$

where $p^{(m)}(\mathbf{x})$ is the output probability (of being a defect) from the $m$th tree. In training each tree, the optimal split at any node is often determined by minimizing an impurity measure. For instance, when using information entropy, the split is chosen to minimize

$$H(S) = -\sum_{c\in\{0,1\}} p_c \log_2(p_c), \tag{14}$$

where $p_c$ is the fraction of samples in set $S$ belonging to class $c$. By aggregating many such trees with randomized selections of training data and feature subsets, a robust classifier is obtained.

After extracting features and concatenating all pixelwise samples, we split the dataset into a training partition $\mathcal{D}_{\text{train}}$ (e.g., 80% of the samples) and a hold-out test partition $\mathcal{D}_{\text{test}}$ (e.g., 20%). The classifier is trained using only $\mathcal{D}_{\text{train}}$, and its performance is

later evaluated on $\mathcal{D}_{\text{test}}$. To quantify predictive ability, we consider the confusion matrix, from which classical metrics (accuracy, precision, recall, etc.) are derived. For example, precision is given by

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{15}$$

and recall (or sensitivity) by

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{16}$$

where TP (true positives), FP (false positives), and FN (false negatives) are computed with respect to the defect class. We further compute the receiver operating characteristic (ROC) curve and the area under the curve (AUC). For the probability outputs $p(\mathbf{x}_i)$, the ROC curve is generated by varying a threshold $\theta$: pixels with $p(\mathbf{x}_i) > \theta$ are predicted as defects, and those with $p(\mathbf{x}_i) \leq \theta$ as healthy.

In this work, each pixel's defect probability is estimated as the average probability from all trees in the random forest. A useful measure of prediction uncertainty is given by the Shannon entropy of the Bernoulli distribution associated with $p(\mathbf{x}_i)$. Concretely, let

$$p_i = p(\mathbf{x}_i), \qquad H(p_i) = -\big[p_i \ln(p_i) + (1 - p_i)\ln(1 - p_i)\big]. \tag{17}$$

This scalar $H(p_i)$ is larger near $p_i = 0.5$ (reflecting uncertainty) and minimal near 0 or 1 (reflecting confidence). Mapping $H(p_i)$ across all pixels yields a spatial representation of classification confidence.

To reduce granular noise or salt-and-pepper artifacts in the segmentation, spatial smoothing (e.g., via a median filter or morphological operators) is applied to the probability map $p(j, k)$. Let $\hat{p}(j, k)$ be the raw probability assigned to pixel $(j, k)$. A median filter of size $m \times m$ yields the smoothed probability

$$\widetilde{p}(j, k) = \text{median}\Big\{\hat{p}(j + u, k + v) \mid (u, v) \in \Omega_m\Big\}, \tag{18}$$

where $\Omega_m$ defines the local neighborhood. This spatial regularization consolidates adjacent pixels with similar probabilities and removes isolated outliers. A binary segmentation map is then generated by thresholding $\widetilde{p}(j, k)$ at a suitable level $\tau$ (e.g., $\tau = 0.5$):

$$\text{Seg}(j, k) = \begin{cases} 1, & \text{if } \widetilde{p}(j, k) > \tau, \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

Further morphological operations, such as a closing step,

$$\text{Seg} \leftarrow \text{Seg} \bullet \mathcal{B}, \tag{20}$$

(where $\bullet$ denotes the morphological closing operator and $\mathcal{B}$ is a structuring element, typically circular or elliptical), ensure that the final segmentation map is spatially coherent and fills small holes within defect regions.

## 3. Results

In this study, a total of 389120 pixel samples were obtained by extracting 15 features per pixel from the nineteen thermographic video sequences of PVC specimens. These features included parameters from the cooling model, various temporal descriptors, and local spatial statistics. The ground truth for each pixel was based on known defect masks. A random forest classifier, trained on 80% of the pixel-level data, was evaluated on the remaining 20% of the data, yielding the following overall metrics on the test partition: an IoU of 0.705, accuracy of 0.993, precision of 0.943, recall of 0.736, F1 score of 0.827, and a ROC AUC of 0.998. Although the accuracy is extremely high, the slightly lower recall indicates that a fraction of actual defects were missed by the classification process, which is reflected in the IoU measure. Nonetheless, the overall performance suggests that the methodology captures the thermographic signatures of subsurface defects effectively.

The detailed classification report shows that healthy pixels achieved near-perfect precision and recall, whereas defect pixels exhibited a precision of 0.94 and a recall of 0.74. This high precision means that when a pixel was predicted to be defective, it was correct most of the time. However, certain small or subtle defect regions were occasionally missed. In scenarios where the cost of failing to detect a defect is high, it may be advantageous to use a lower decision threshold or apply further morphological expansions, though this might increase false positives. Additional explorations could include domain-specific adjustments to the threshold or tuning the spatial neighborhood size used in the feature augmentation step.

Fig. 3 presents the confusion matrix of the random forest classifier, displaying predicted labels along the horizontal axis and true labels along the vertical axis. The matrix is strongly dominated by the large diagonal entries, where 76 094 healthy pixels are correctly identified (top-left cell) and 1219 defect pixels are correctly classified (bottom-right cell). This aligns with the overall high accuracy of the model. The off-diagonal cells, which indicate misclassifications, are comparatively small: only 73 healthy pixels are predicted as defect (top-right cell), and 438 defect pixels are incorrectly labeled as healthy (bottom-left cell). Although the low count of false positives (73) highlights a model conservative in defect calls, the false negatives (438) remain a noteworthy concern, particularly if undetected flaws pose serious structural risks. In many practical nondestructive testing scenarios, such false negatives can be mitigated by adopting a lower probability threshold for defect identification or by applying spatial post-processing to group borderline classifications. It has to be said that the dataset contains the specimen, the gripping system and the background. This might introduce noise and lower the quality of data. However, the presence of external objects in the view cannot be exclude for

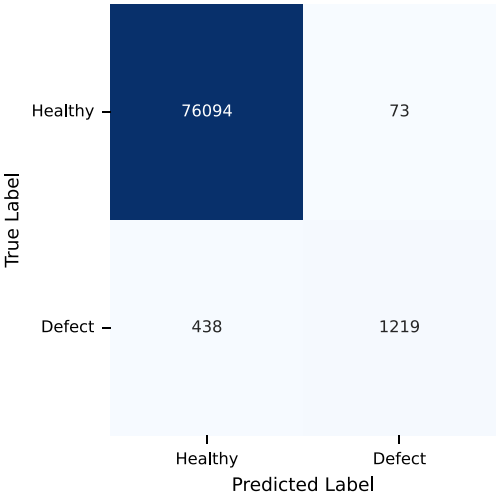



**Fig. 3.** Confusion matrix of predicted versus true classes.

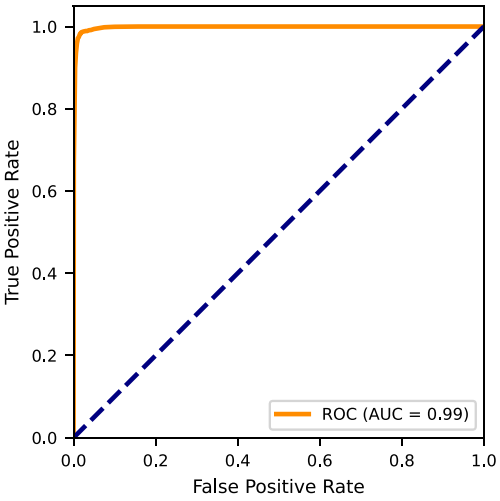



**Fig. 4.** Receiver operating characteristic (ROC) curve. AUC=0.998.

common use application. Nevertheless, the external component in the view were never detected as defective as will be showed later in the results. Overall, the confusion matrix confirms that the proposed feature-based approach is robust, with a strong emphasis on correctly identifying healthy regions while retaining a high rate of true defect detection.

Fig. 4 presents the ROC curve, which approaches the top-left corner of the plot, yielding a near-perfect area under the curve of 0.998. This high AUC indicates that the defect pixels are generally well separated from healthy pixels in the feature space. In practical terms, this suggests that if one adjusts the classification threshold, it is possible to trade precision for recall (or vice versa) while still maintaining an overall strong discriminative capability. Nevertheless, with highly skewed data in some specimens, care must be taken when selecting operating points for industrial scenarios.

Fig. 5 compares the predicted defect probability from the random forest classifier (blue solid line) with the empirical fraction of positive (defect) samples, as partitioned into bins of similar predicted probability. The dashed diagonal represents ideal calibration, where the classifier's predicted probabilities perfectly match the observed frequencies of defect pixels. Overall, the curve lies close to this diagonal, indicating that when the model assigns, for example, an average of 60% probability, the true fraction of defects in that bin is also near 60%. However, mild deviations can be observed in the mid-probability range (around 0.3–0.5), where the classifier slightly underestimates the defect likelihood. This region suggests that certain borderline defect pixels are assigned lower probabilities than they actually warrant. Nevertheless, at high probability levels ($\geq 0.8$), the curve approaches or slightly exceeds the diagonal, implying that when the model strongly predicts a defect, the positive rate is indeed very high. Such calibration quality is beneficial in real-world nondestructive testing scenarios, as reliable probability estimates enable users to make more informed decisions regarding inspection thresholds and maintenance actions.

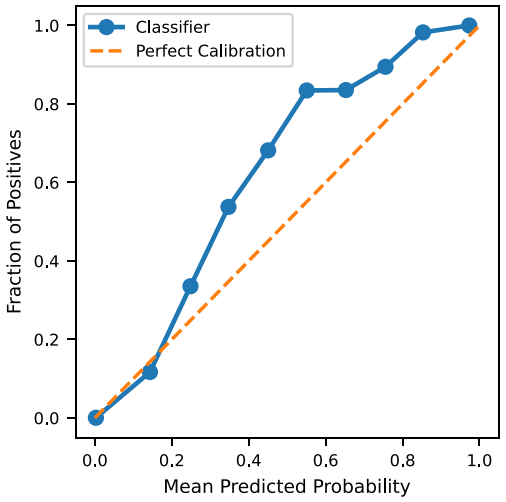



**Fig. 5.** Calibration curve for the random forest classifier.

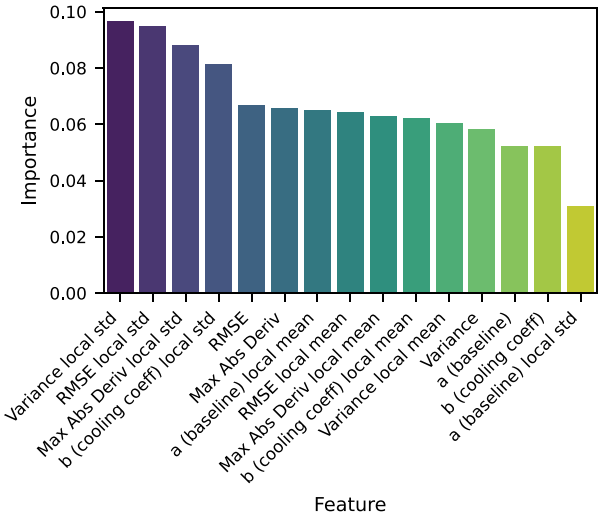



**Fig. 6.** Feature importance plot.

Fig. 6 ranks the fifteen features by their mean decrease in impurity (MDI) within the random forest ensemble. Notably, local standard deviation measures dominate the top positions: Variance local std emerges as the single most influential feature, followed closely by RMSE local std, Max Abs Deriv local std, and b (cooling coeff) local std. This pattern highlights that localized fluctuations around each pixel (as captured by the standard deviation in a small neighborhood) provide a highly discriminative cue for identifying subsurface defects. Even when the absolute value of a thermal parameter might not differ drastically between healthy and defect regions, spatial variation in the same parameter often reveals the presence of anomalies.

In the midrange, the raw features RMSE and Max Abs Deriv are comparably significant, reflecting their direct linkage to pixelwise deviations from the $t^{-\frac{1}{2}}$ cooling model and abrupt temperature changes, respectively. Additionally, the local mean versions of a (baseline), RMSE, Max Abs Deriv, b (cooling coeff), and Variance fill out the middle tier. These metrics reinforce that averaging over a neighborhood helps capture smooth thermal trends—complementary to the more sensitive local standard deviations—thereby contributing substantively to defect segmentation.

In contrast, the lower-ranked features include the global (i.e., single-pixel) Variance, a (baseline), b (cooling coeff), and especially a (baseline) local std, which shows the smallest feature importance. Although baseline and cooling coefficients describe the fundamental shape of the thermal decay, the moderate importance of their global values implies that other features—particularly those incorporating higher-order or spatial information—play a more crucial role in discriminating subtle defects. Overall, these findings highlight the synergistic interplay between raw, per-pixel thermal descriptors and neighborhood-based statistics, with local standard deviations emerging as a particularly robust indicator of anomalous heat-flow patterns.

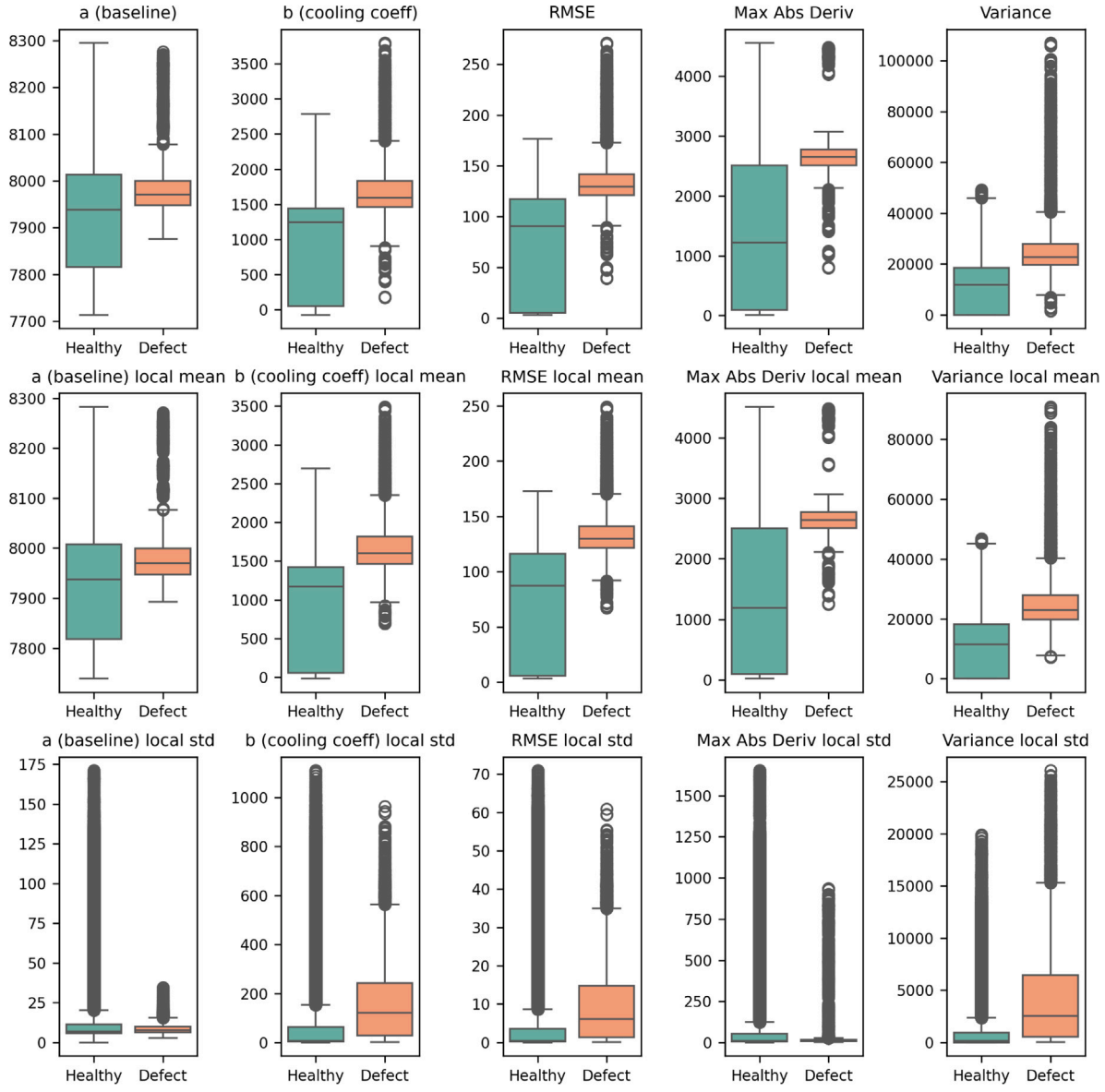
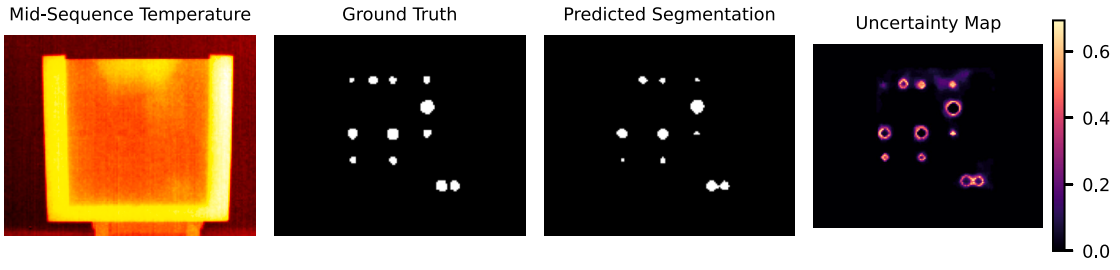


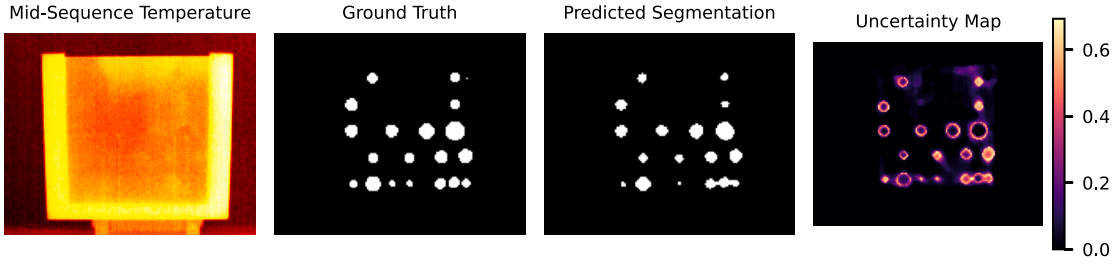

**Fig. 7.** Box plots for multiple features comparing defect and healthy classes.

Fig. 7 presents fifteen box plots for the extracted features, partitioned into three rows of five subplots each. In the *top row*, we show the raw, per-pixel features derived from the fitting procedure and temporal statistics: (1) baseline $a$, (2) cooling coefficient $b$, (3) RMSE, (4) maximum absolute derivative, and (5) variance. In almost every case, the defect class displays either a shifted median or a broader range compared to healthy pixels. Specifically, the baseline $a$ is typically higher for defective regions, underscoring a residual heat accumulation effect. Conversely, the cooling coefficient $b$ is lower for defects, reflecting a disrupted thermal dissipation pattern. The RMSE plots also reveal larger residual errors in defective areas, indicating that the simple $t^{-\frac{1}{2}}$ cooling model struggles to capture local heat flow distortions introduced by subsurface voids. Both maximum absolute derivative and variance exhibit generally higher median values for defect pixels, pointing to more pronounced temperature gradients and fluctuations around flaws.

In the *middle row*, each of the five base features is augmented with its local mean. This spatial averaging leverages neighboring pixels to amplify characteristic zones of thermal contrast. Here, defective pixels again manifest consistently distinct medians, often surpassing the healthy class for $a$ (local mean), maximum absolute derivative, and variance. Meanwhile, the local mean of the cooling coefficient $b$ remains lower for defective areas, mirroring the raw $b$ feature's class separation but in a more spatially smoothed form. These trends confirm that neighborhood-based statistics help reveal localized irregularities in the underlying thermal field.




(a) Specimen 004




(b) Specimen 012

**Fig. 8.** Illustration of the predicted segmentation compared to ground truth and the associated uncertainty map for PVC specimens.

Finally, the *bottom row* shows the local standard deviations for the same five base features. Interestingly, healthy regions exhibit a higher spread for most features (e.g., baseline $a$ and maximum absolute derivative), suggesting that intact areas can sustain more pixel-to-pixel variability. In contrast, defected regions tend to have smaller local standard deviations, indicating a more spatially clustered disruption—possibly because the subsurface flaw imposes a relatively uniform thermal anomaly in its vicinity. Notwithstanding a few overlapping whiskers between classes, each box plot demonstrates a meaningful shift in central tendency or dispersion for defective pixels, thereby reinforcing the discriminative power of combining raw and spatially derived attributes. Overall, these insights illustrate that integrating local neighborhood statistics with the core physical features of cooling behavior significantly enhances the separation between healthy and defective regions in pulsed thermography data.

Finally, the qualitative performance is demonstrated by segmentation overlays and uncertainty maps in Figs. 8(a) and 8(b), which show representative specimens from the dataset. In each four-panel layout, the first image corresponds to a reference thermal map (such as mid-sequence temperature), the second image is the ground-truth mask, the third image is the predicted defect segmentation (with spatial regularization), and the fourth image is the associated uncertainty map derived from the Shannon entropy of the predicted probability. Darker tones in the uncertainty map indicate that the classifier is more confident in its label assignment, whereas lighter tones indicate intermediate probabilities. It is visible that inside and outside the defect the uncertainty is low, while in the bounding areas the uncertainty is higher. This can be explained by the fact that in transition zones, such as boundaries, the thermal behavior is not fully correspondent to the expected one for a defect nor for a healthy region. This highlights a potential workflow in which human inspectors or automated systems could focus on regions of high uncertainty to gather additional information or apply targeted inspection.

Overall, these results highlight both the strong accuracy and the remaining challenges in capturing certain subtle defects. The near-perfect ROC curve and high precision affirm that the extracted features and ensemble classification approach are suitable for robust identification of deeper or more pronounced subsurface anomalies. The need for enhanced recall, however, indicates that some advanced or smaller-scale defects could require additional refinement of features or further post-processing. This holistic analysis thus provides a thorough assessment of the pipeline's capability and points to directions for future enhancements in thermographic nondestructive testing.

Compared to the work of [64], which reported an IoU of up to 0.722 when using a deep encoder–decoder network (U-Net) with a VGG11 backbone on the same PVC-Infrared dataset, our IoU of 0.705 is close to that range while relying on a different methodological paradigm. Whereas [64] leverages deep neural architectures that may demand substantial training time and specialized hardware, our approach with carefully engineered features, a random forest model, and relatively straightforward spatial augmentation strategies attains similarly strong IoU values, with a very high ROC AUC of 0.998 and a notably transparent interpretation of how each feature contributes to defect detection. This high interpretability can be particularly advantageous for industrial applications where model transparency and ease of deployment are valuable. The reduced computational overhead associated with the presented pipeline also underscores its practical utility, making it more accessible in settings where deep learning infrastructure or large-scale training resources may be limited.

**Table 3**
Comparison of selected defect detection methods in IRT. The proposed approach addresses the gap between interpretability and high performance, while maintaining moderate data and resource requirements.

| Method | Computational cost | Interpretability | Data requirements | Performance (IoU/AUC) |
|---|---|---|---|---|
| Deep U-Net [64] | High | Low | High | 0.72 IoU |
| Proposed (RF + physics features) | Low-Moderate | High | Moderate | 0.70–0.71 IoU/0.998 AUC |

## 4. Conclusions

In this study, we introduced a robust and interpretable pipeline for defect detection in thermographic data obtained from PVC specimens. The proposed methodology leverages a physics-based cooling model to extract five temporal features which, when augmented with local spatial statistics, yield a 15-dimensional descriptor for each pixel. The subsequent application of a random forest classifier resulted in high overall accuracy, a near-perfect ROC AUC of 0.998, and competitive IoU and precision values. These results affirm that the engineered feature set is highly discriminative for identifying subsurface defects, while the classifier's high confidence levels and well-calibrated probability estimates facilitate a clear interpretation of the failure mechanisms.

Our approach demonstrates several strengths that are particularly relevant for the field of Engineering Failure Analysis. Unlike complex deep learning models, such as the U-Net with a VGG11 backbone, which require extensive computational resources and can be difficult to interpret, the proposed pipeline offers transparency in feature contribution and a significantly reduced computational overhead. This is of critical importance in engineering applications where understanding the underlying physical basis of failure and making rapid, reliable assessments are paramount.

Furthermore, the ability to generate spatial uncertainty maps provides a practical tool for targeted inspection and risk-based decision making. By highlighting areas of higher uncertainty, the method enables engineers to focus additional diagnostic efforts on regions that may harbor subtle or early-stage defects. This aspect not only improves the reliability of defect detection but also contributes to a more efficient allocation of maintenance resources.

Table 3 provides a high-level comparison of related approach along with ours. We focus on key aspects such as computational cost, interpretability, data requirements, and typical performance metrics. This overview highlights a recurring trade-off: more complex neural network models may yield high accuracy but demand extensive training data and computational power, whereas simpler or more classical approaches can be faster and more interpretable, though sometimes less accurate.

The comparative analysis with the results reported in [64] underscores the efficacy of our method. While deep encoder–decoder networks have achieved slightly higher IoU values on the same dataset, our results, obtained with a straightforward machine learning framework, are competitive and offer enhanced interpretability and ease of deployment. These advantages align well with the aims of Engineering Failure Analysis, where practical considerations such as model transparency, computational efficiency, and ease of integration with existing inspection protocols are essential.

Future work will aim to extend the current framework to accommodate a wider variety of defect shapes, depths, and material properties, and to integrate additional sensing modalities for a more comprehensive evaluation. Further investigation into adaptive thresholding and dynamic spatial post-processing may also help in further increasing recall without compromising precision.

This experiment where performed using an high-end thermal camera, future works might explore the use of cheaper instruments that might introduce more noise in the signal.

In summary, the integration of physics-based feature engineering with robust machine learning techniques as presented herein offers a viable and effective alternative for defect detection in thermographic data. This approach not only meets the stringent demands of engineering failure analysis but also lays the groundwork for further advancements in reliable, interpretable nondestructive evaluation methods.

## CRediT authorship contribution statement

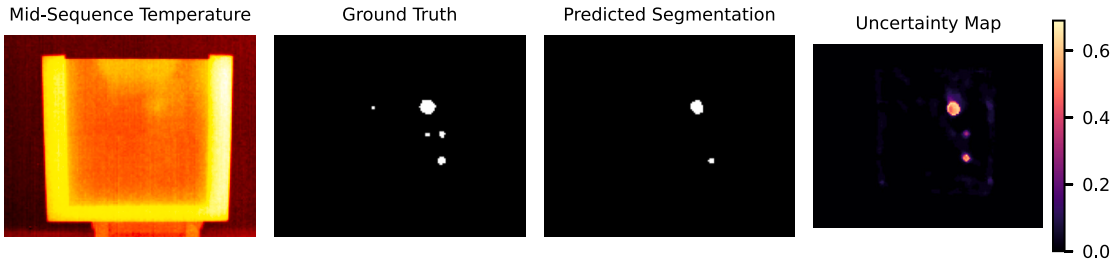
**Luca Santoro:** Writing – original draft, Visualization, Software, Methodology, Conceptualization. **Raffaella Sesana:** Writing – review & editing, Validation, Supervision.

## Declaration of competing interest

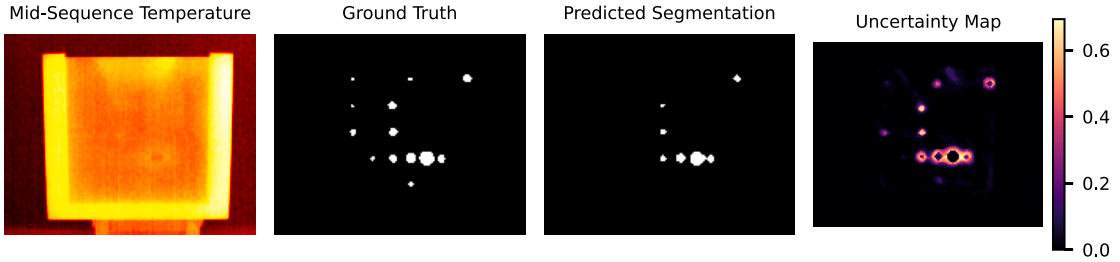
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix. Figure appendix

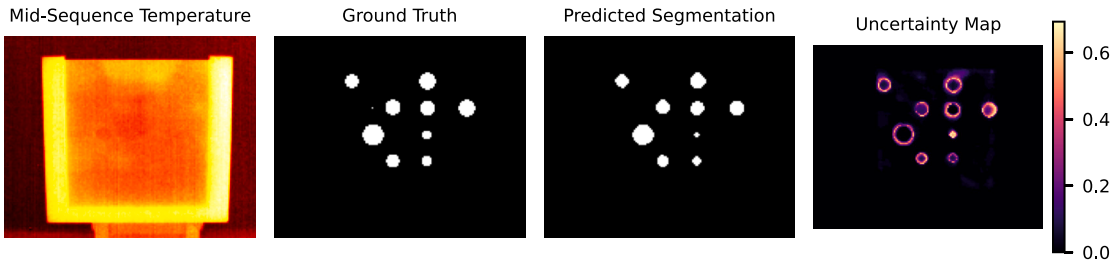
See Fig. 9.



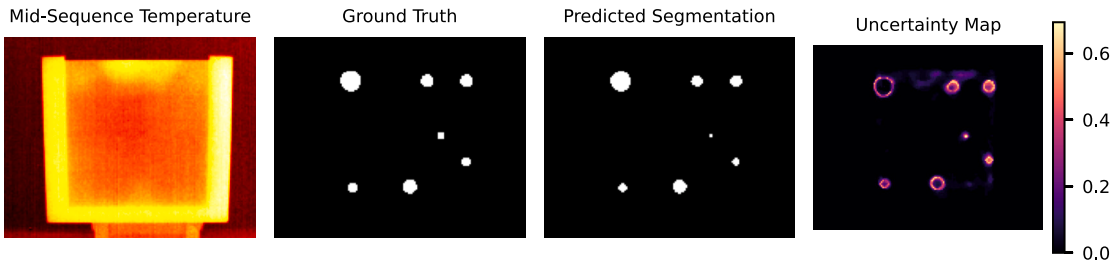

(a) Specimen 002



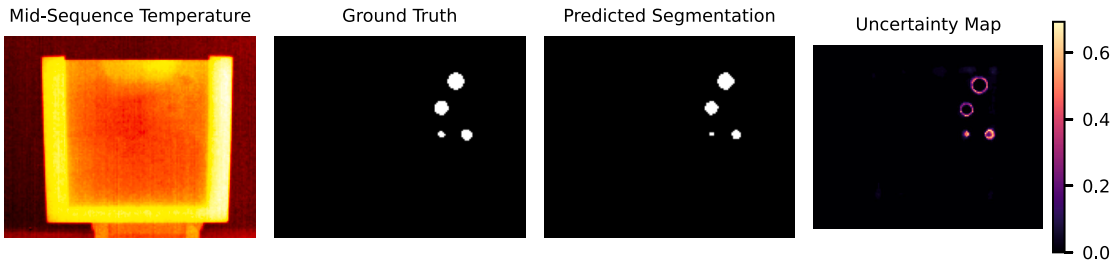

(b) Specimen 003




(c) Specimen 005

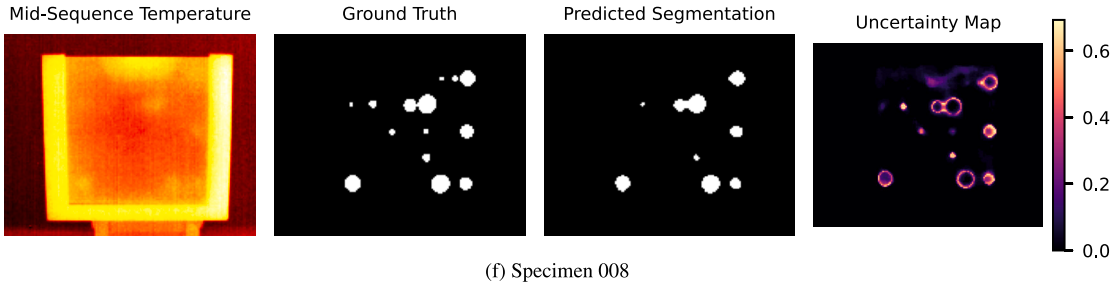



(d) Specimen 006

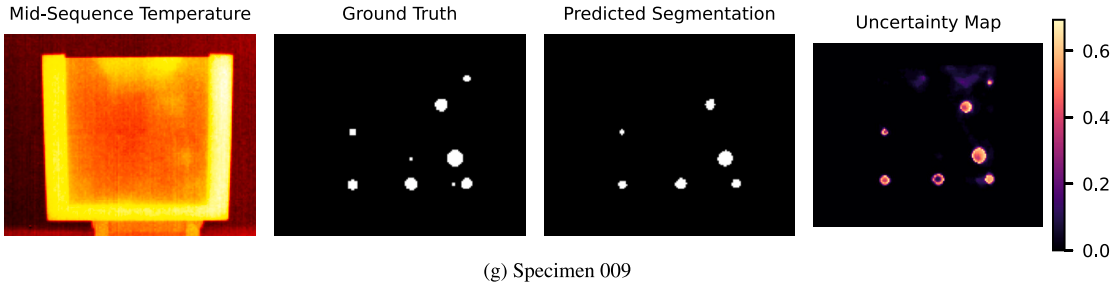



(e) Specimen 007

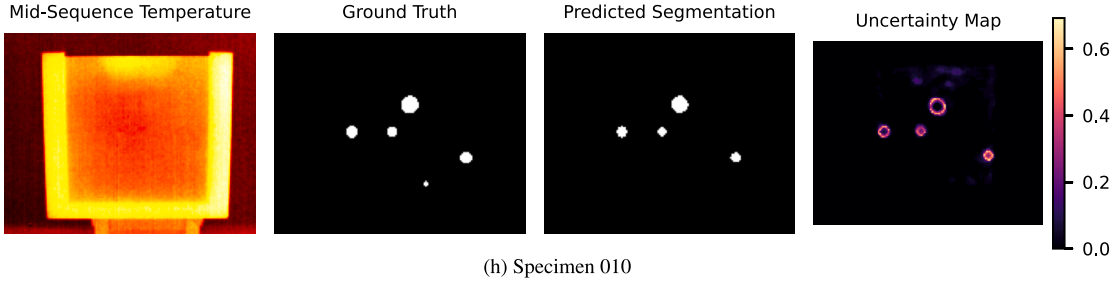
**Fig. 9.** Advanced segmentation results for various specimens.



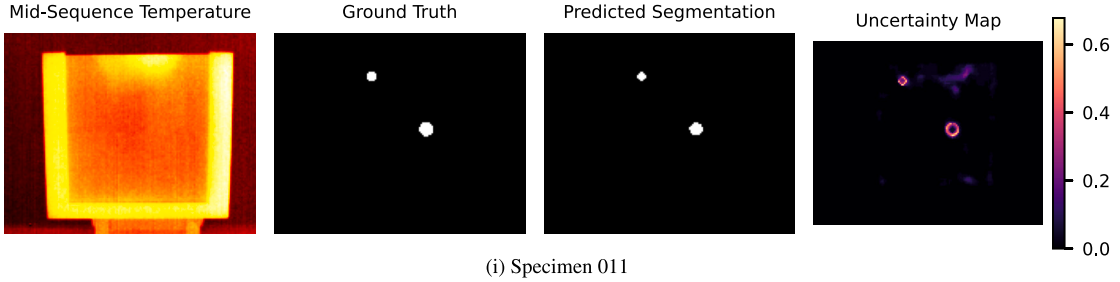

(f) Specimen 008



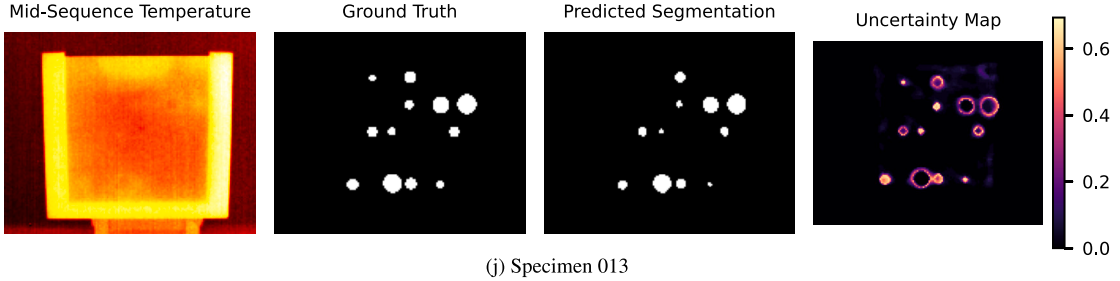

(g) Specimen 009




(h) Specimen 010

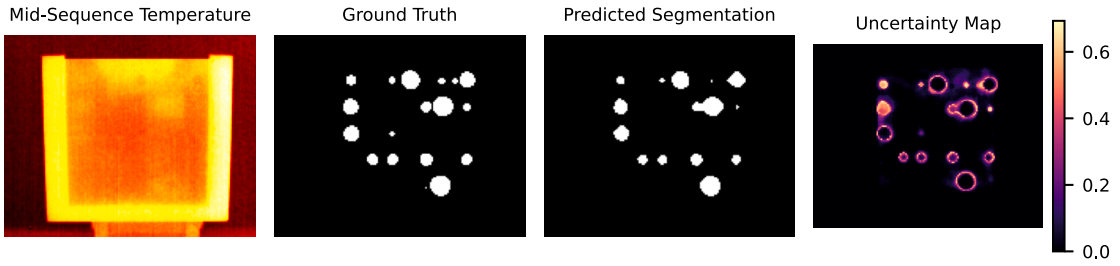



(i) Specimen 011

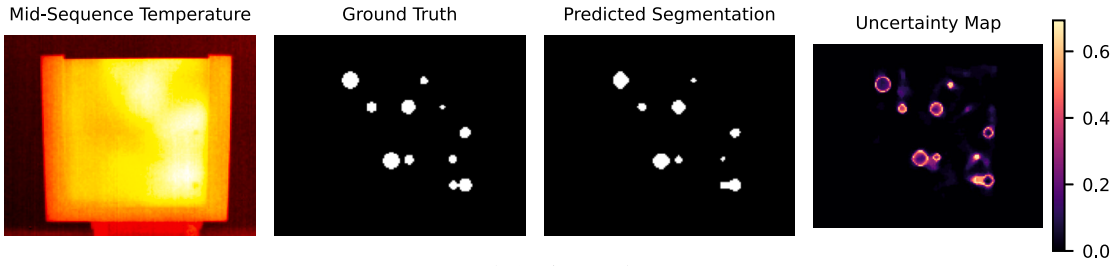



(j) Specimen 013

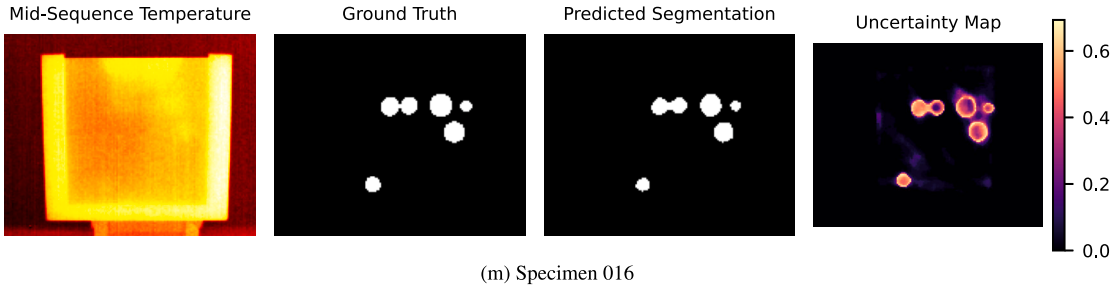
**Fig. 9.** (*continued*).



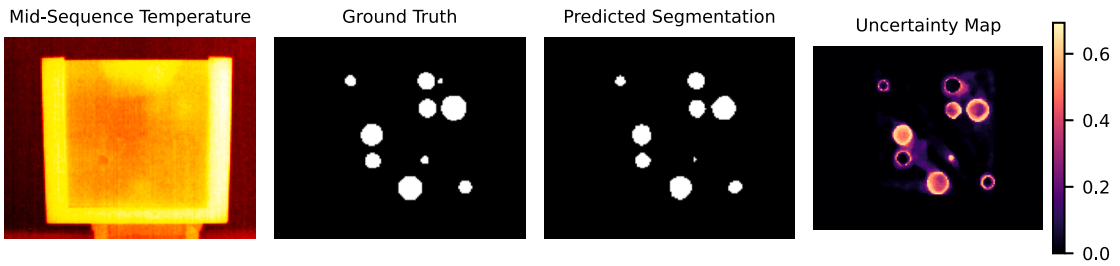

(k) Specimen 014



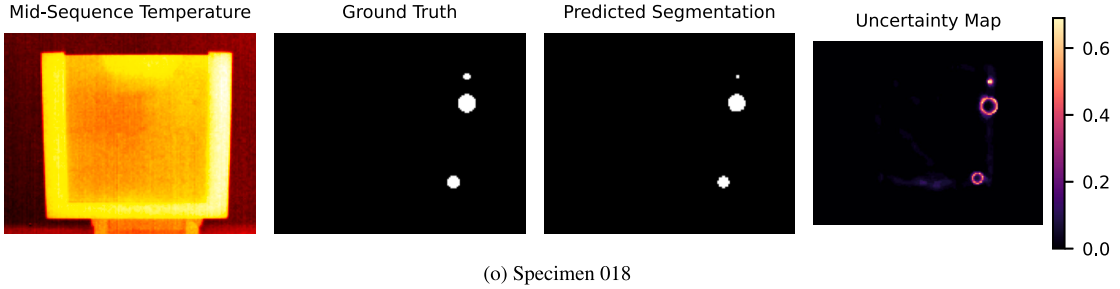

(l) Specimen 015




(m) Specimen 016

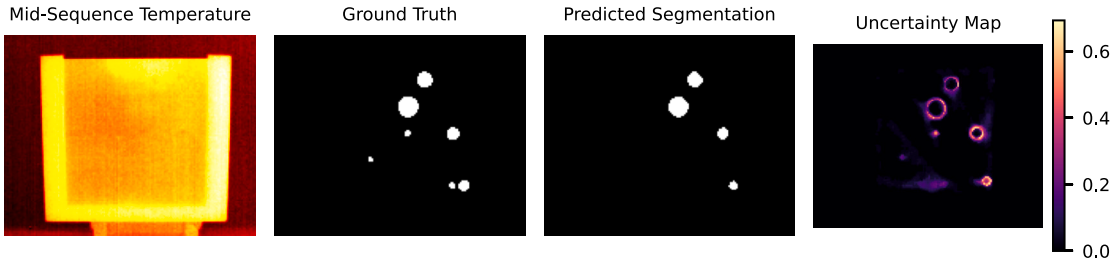



(n) Specimen 017

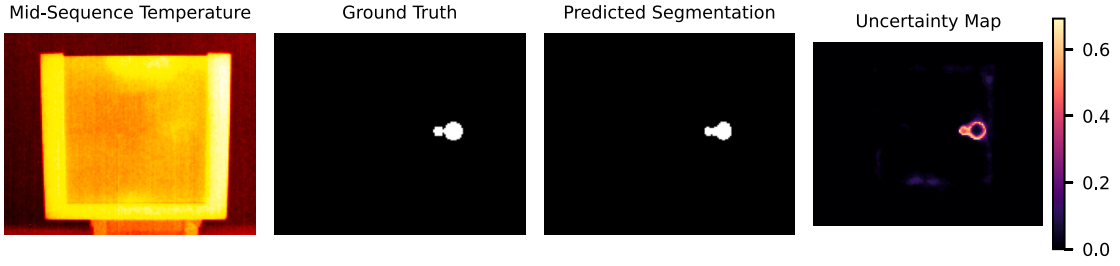



(o) Specimen 018

**Fig. 9.** (*continued*).




(p) Specimen 019




(q) Specimen 020

**Fig. 9.** (*continued*).

## Data availability

Dataset is referred in the text.

## References

[1] M. Amarnath, N. Sudharshan, P. Srinivas, Automatic detection of defects in welding using deep learning - a systematic review, Mater. Today: Proc. (2023).
[2] A.R. Diogo, B. Moreira, C.A.J. Gouveia, J.M.R.S. Tavares, A review of signal processing techniques for ultrasonic guided wave testing, Metals 12 (6) (2022).
[3] V. Borghese, L. Santoro, S. Santini, R. Sesana, Correlation between thermal and density properties of chestnuts: preliminary results of experimental non-destructive testing, Arch. Civ. Mech. Eng. 24 (2024) 1–15.
[4] A.R. Dakak, V. Kaftandjian, P. Duvauchelle, P. Bouvet, Deep learning-based defect detection in industrial CT volumes of castings, Insight: Non- Destr. Test. Cond. Monit. 64 (11) (2022) 647–658.
[5] N.H.M.M. Shrifan, G.N. Jawad, N.A.M. Isa, M.F. Akbar, Microwave nondestructive testing for defect detection in composites based on K-means clustering algorithm, IEEE Access 9 (2021) 4820–4828.
[6] C. Xu, G. Wei, Y. Guan, S. Zhang, H. Wang, X. Chen, F. Wang, H. Liu, High-performance deep learning segmentation for non-destructive testing of X-ray tomography, J. Manuf. Process. 128 (2024) 98–110.
[7] X. Dong, C.J. Taylor, T.F. Cootes, Defect detection and classification by training a generic convolutional neural network encoder, IEEE Trans. Signal Process. 68 (2020) 6055–6069.
[8] D. Buongiorno, M. Prunella, S. Grossi, S.M. Hussain, A. Rennola, N. Longo, G. Di Stefano, V. Bevilacqua, A. Brunetti, Inline defective laser weld identification by processing thermal image sequences with machine and deep learning techniques, Appl. Sci. (Switzerland) 12 (13) (2022).
[9] Y. Cao, Y. Dong, Y. Cao, J. Yang, M.Y. Yang, Two-stream convolutional neural network for non-destructive subsurface defect detection via similarity comparison of lock-in thermography signals, NDT E Int. 112 (2020).
[10] M. Quercio, L. Santoro, E. Poskovic, R. Sesana, A. Canova, Magnetic and thermal characterization of Fe2.9wt.%Si for magnetic shielding applications, IEEE Access (2024) 1.
[11] L. Santoro, R. Sesana, R.M. Nardo, F. Curá, Infrared in-line monitoring of flaws in steel welded joints: a preliminary approach with SMAW and GMAW processes, Int. J. Adv. Manuf. Technol. 128 (2023) 2655–2670.
[12] L. Santoro, V. Razza, M.D. Maddis, Frequency-based analysis of active laser thermography for spot weld quality assessment, Int. J. Adv. Manuf. Technol. 130 (2024) 3017–3029.
[13] R. Dvořák, L. Jakubka, L. Topolář, M. Rabenda, A. Wirowski, J. Puchýř, I. Kusák, L. Pazdera, Non-destructive characterization of cured-in-place pipe defects, Materials 16 (24) (2023).
[14] V. Guillot, Infrared thermography for seal defects detection on packaged products: unbalanced machine learning classification with iterative digital image restoration, Electron. Lett. Comput. Vis. Image Anal. 22 (1) (2023) 35–51.
[15] L. Cheng, Z. Tong, S. Xie, M. Kersemans, IRT-GAN: A generative adversarial network with a multi-headed fusion strategy for automated defect detection in composites using infrared thermography, Compos. Struct. 290 (2022).
[16] Q. Fang, C. Ibarra-castanedo, X. Maldague, Automatic defects segmentation and identification by deep learning algorithm with pulsed thermography: Synthetic and experimental data, Big Data Cogn. Comput. 5 (1) (2021) 1–21.
[17] L. Santoro, R. Sesana, J. Diller, C. Radlbeck, M. Mensinger, Dissipative and thermal aspects in cyclic loading of additive manufactured AISI 316L, Eng. Fail. Anal. 163 (2024) 108446.
[18] J. Cui, B. Zhang, X. Wang, J. Wu, J. Liu, Y. Li, X. Zhi, W. Zhang, X. Yu, Impact of annotation quality on model performance of welding defect detection using deep learning, Weld. the World 68 (4) (2024) 855–865.

[19] L. Deng, H. Zuo, W. Wang, C. Xiang, H. Chu, Internal defect detection of structures based on infrared thermography and deep learning, KSCE J. Civ. Eng. 27 (3) (2023) 1136–1149.
[20] D. Yang, Y. Cui, Z. Yu, H. Yuan, Deep learning based steel pipe weld defect detection, Appl. Artif. Intell. 35 (15) (2021) 1237–1249.
[21] L. Santoro, V. Razza, M.D. Maddis, Nugget and corona bond size measurement through active thermography and transfer learning model, Int. J. Adv. Manuf. Technol. 133 (2024) 5883–5896.
[22] L. Santoro, M. Quercio, A. Canova, R. Sesana, Measuring thermal and electrical performances of additively manufactured magnetic shielding material: an active thermography approach, Nondestruct. Test. Eval. (2024) 1–24.
[23] R. Sesana, L. Santoro, F. Curà, R.M. Nardo, P. Pagano, Assessing thermal properties of multipass weld beads using active thermography: microstructural variations and anisotropy analysis, Int. J. Adv. Manuf. Technol. 128 (2023) 2525–2536.
[24] S. Kumaresan, K.S.J. Aultrin, S. Kumar, M.D. Anand, Deep learning-based weld defect classification using VGG16 transfer learning adaptive fine-tuning, Int. J. Interact. Des. Manuf. 17 (6) (2023) 2999–3010.
[25] S. Deng, R. Gao, Y. Wang, W. Mao, W. Zheng, Structure of a semantic segmentation-based defect detection network for laser cladding infrared images, Meas. Sci. Technol. 34 (8) (2023).
[26] K. Liu, Y. Li, J. Yang, Y. Liu, Y. Yao, Generative principal component thermography for enhanced defect detection and analysis, IEEE Trans. Instrum. Meas. 69 (10) (2020) 8261–8269.
[27] V. Yuksel, Y.E. Tetik, M.O. Basturk, O. Recepoglu, K. Gokce, M.A. Cimen, A novel cascaded deep learning model for the detection and quantification of defects in pipelines via magnetic flux leakage signals, IEEE Trans. Instrum. Meas. 72 (2023).
[28] W.H. Lim, S. Sfarra, Y. Yao, A physics-informed neural network method for defect identification in polymer composites based on pulsed thermography †, Eng. Proc. 8 (1) (2021).
[29] W. Hng Lim, S. Sfarra, T.-Y. Hsiao, Y. Yao, Physics-informed neural networks for defect detection and thermal diffusivity evaluation in carbon fiber-reinforced polymer using pulsed thermography, IEEE Trans. Instrum. Meas. 74 (2025).
[30] M. Quercio, E. Poskovic, F. Franchini, E. Fracchia, L. Ferraris, A. Canova, A. Tenconi, H. Tiismus, A. Kallaste, Application of active thermography for the study of losses in components produced by laser powder bed fusion, J. Magn. Magn. Mater. 592 (2024) Cited by: 2.
[31] R. Asghar, F.R. Fulginei, M. Quercio, L. Sabino, F. Crescimbini, M. Abusara, A hybrid LSTM-CNN model for short-term photovoltaic power forecasting in Italy, Dig. Tech. Pap. - IEEE Int. Conf. Consum. Electron. (2024) 67–73, Cited by: 0.
[32] M. Quercio, L. Sabino, G.M. Lozito, R. Asghar, M. Parise, F.R. Fulginei, Prediction of DC/DC boost converter switching power losses using a backpropagation algorithm neural network, in: 8th IEEE International Forum on Research and Technologies for Society and Industry Innovation, RTSI 2024 - Proceeding, 2024, pp. 449–452, Cited by: 1.
[33] R. Asghar, F.R. Fulginei, M. Quercio, A. Mahrouch, Artificial neural networks for photovoltaic power forecasting: A review of five promising models, IEEE Access 12 (2024) 90461–90485, Cited by: 6; All Open Access, Gold Open Access.
[34] A. Moskovchenko, M. Svantner, Thermographic data processing and feature extraction approaches for machine learning-based defect detection †, Eng. Proc. 51 (1) (2023).
[35] G. Vesala, V. Ghali, A.V. Lakshmi, R. Naik, Deep and handcrafted feature fusion for automatic defect detection in quadratic frequency modulated thermal wave imaging, Russ. J. Nondestruct. Test. 57 (6) (2021) 476–485.
[36] K.A. Pareek, D. May, P. Meszmer, M.A. Ras, B. Wunderle, Synthetic data generation using finite element method to pre-train an image segmentation model for defect detection using infrared thermography, J. Intell. Manuf. (2024).
[37] Q. Li, H. Zhang, J. Hu, S. Sfarra, M. Mostacci, D. Yang, M. Georges, V.P. Vavilov, X.P.V. Maldague, Using the unsupervised mixture of Gaussian models for multispectral non-destructive evaluation of the replica of botticelli's "the birth of venus", J. Nondestruct. Eval. 42 (2) (2023).
[38] Q. Fang, C. Ibarra-Castanedo, X. Maldgue, University laval infrared thermography databases for deep learning multiple types of defect detections training †, Eng. Proc. 2 (1) (2021).
[39] C. Dunderdale, W. Brettenny, C. Clohessy, E.E. van Dyk, Photovoltaic defect classification through thermal infrared imaging using a machine learning approach, Prog. Photovolt., Res. Appl. 28 (3) (2020) 177–188.
[40] A.-F. Obaton, Y. Wang, B. Butsch, Q. Huang, A non-destructive resonant acoustic testing and defect classification of additively manufactured lattice structures, Weld. the World 65 (3) (2021) 361–371.
[41] D. Müller, U. Netzelmann, B. Valeske, Defect shape detection and defect reconstruction in active thermography by means of two-dimensional convolutional neural network as well as spatiotemporal convolutional LSTM network, Quant. InfraRed Thermogr. J. 19 (2) (2022) 126–144.
[42] Z. Tong, L. Cheng, S. Xie, M. Kersemans, A flexible deep learning framework for thermographic inspection of composites, NDT E Int. 139 (2023).
[43] H. Wu, K. Zheng, S. Sfarra, Y. Liu, Y. Yao, Multiview learning for subsurface defect detection in composite products: A challenge on thermographic data analysis, IEEE Trans. Ind. Informatics 16 (9) (2020) 5996–6003.
[44] G. Liu, W. Gao, W. Liu, Y. Chen, T. Wang, Y. Xie, W. Bai, Z. Li, Automatic defect classification for infrared thermography in CFRP based on deep learning dense convolutional neural network, J. Nondestruct. Eval. 43 (3) (2024).
[45] G. Zhou, Z. Zhang, W. Yin, H. Chen, L. Wang, D. Wang, H. Ma, Surface defect detection of CFRP materials based on infrared thermography and attention U-net algorithm, Nondestruct. Test. Eval. 39 (2) (2024) 238–257.
[46] A. Gulsen, B. Kolukisa, A.T. Ozdemir, B. Bakir-Gungor, V.C. Gungor, Defect classification of composite materials using transfer learning methods, Nondestruct. Test. Eval. (2024).
[47] J. Cao, G. Yang, X. Yang, A pixel-level segmentation convolutional neural network based on deep feature fusion for surface defect detection, IEEE Trans. Instrum. Meas. 70 (2021).
[48] D.M. Tsai, S.K.S. Fan, Y.H. Chou, Auto-annotated deep segmentation for surface defect detection, IEEE Trans. Instrum. Meas. 70 (2021).
[49] R. Wu, H. Wei, C. Lu, Y. Liu, Automatic and accurate determination of defect size in shearography using U-net deep learning network, J. Nondestruct. Eval. 44 (1) (2025).
[50] Q. Li, X. Peng, X. Zhong, X. Xiao, H. Wang, C. Zhao, K. Zhou, Quantitative identification of debonding defects in building façades based on UAV-thermography using a two-stage network integrating dual attention mechanism, Infrared Phys. Technol. 138 (2024).
[51] S. Liu, X. Hua, Y. Liu, L. Shan, D. Wang, Q. Wang, Y. Sun, Accurate wire rope defect MFL detection using improved Hilbert transform and LSTM neural network, Nondestruct. Test. Eval. (2024).
[52] Y. Ding, G. Russo, R.K. Tshiangomba, E. Pellegrino, A. Cicone, S. Sfarra, H. Zhang, Stabilization system for solar loading thermography applied on cultural heritage objects exposed outdoors: the contribution of advanced algorithms and dual-branch U-net, J. Therm. Anal. Calorim. (2024).
[53] Z. Zhang, H. Zhang, J. Hu, S. Sfarra, M. Mostacci, Y. Wang, D. Yang, X. Maldague, D. Niu, Y. Duan, Defect detection: An improved YOLOX network applied to a replica of "the birth of venus" by botticelli, J. Cult. Herit. 62 (2023) 404–411.
[54] S. Pozzer, M.P.V.d. Souza, B. Hena, R.K. Rezayiye, S. Hesam, F. Lopez, X. Maldague, Defect segmentation in concrete structures combining registered infrared and visible images: A comparative experimental study †, Eng. Proc. 8 (1) (2021).
[55] S. Li, S. Sun, Y. Liu, W. Qi, N. Jiang, C. Cui, P. Zheng, Real-time lightweight YOLO model for grouting defect detection in external post-tensioned ducts via infrared thermography, Autom. Constr. 168 (2024).
[56] N. Shipway, P. Huthwaite, M. Lowe, T. Barden, Using ResNets to perform automated defect detection for fluorescent penetrant inspection, NDT E Int. 119 (2021).

[57] E. Westphal, H. Seitz, A machine learning method for defect detection and visualization in selective laser sintering based on convolutional neural networks, Addit. Manuf. 41 (2021).
[58] K. Liu, Y. Tang, W. Lou, Y. Liu, J. Yang, Y. Yao, A thermographic data augmentation and signal separation method for defect detection, Meas. Sci. Technol. 32 (4) (2021).
[59] X. Li, X. Ying, W. Zhu, W. Liu, B. Hou, L. Zhou, Nondestructive detection and analysis based on data enhanced thermography, Meas. Sci. Technol. 33 (6) (2022).
[60] P. Shen, Z. Luo, S. Wang, F. Mao, Z. Su, H. Zhang, Feature detection of GFRP subsurface defects using fast randomized sparse principal component thermography, Int. J. Thermophys. 43 (10) (2022).
[61] F. Wang, Z. Jiang, Y. Liu, C. Ibarra-Castanedo, H. Zhang, K. Cao, X. Maldague, S. Sfarra, Y. Yao, Enhancing defect detection in active infrared thermography using adaptive background suppression techniques, J. Therm. Anal. Calorim. (2024).
[62] L. Zeng, S. Zhang, P. Wang, Z. Li, Y. Hu, T. Xie, Defect detection algorithm for magnetic particle inspection of aviation ferromagnetic parts based on improved DeepLabv3+, Meas. Sci. Technol. 34 (6) (2023).
[63] K.A. Pareek, D. May, P. Meszmer, M.A. Ras, B. Wunderle, Synthetic data generation using finite element method to pre-train an image segmentation model for defect detection using infrared thermography, J. Intell. Manuf. (2024).
[64] Z. Wei, A. Osman, B. Valeske, X. Maldague, Pulsed thermography dataset for training deep learning models, Appl. Sci. 13 (5) (2023).

**Luca Santoro** is a dedicated researcher at the Politecnico di Torino, within the Department of Mechanical and Aerospace Engineering. His expertise lies in non-destructive testing techniques and the mechanical fatigue of materials. He specializes in active thermography and innovative ultrasonic applications, with a recent focus on using thermography for monitoring welding processes. His work has led to significant industrial applications and the development of patents, contributing to advancements in material testing and engineering practices.

**Raffaella Sesana** is M.Sc. in Mechanical Engineering, Ph.D. in Machine Design. Today Associate Professor in Department of Mechanic and Aerospace Engineering, Politecnico di Torino. The main research topics are HCF, non-destructive fatigue limit and damage assessment, LCF and TMF damage models, characterization of materials and components, constitutive models for cyclic plastic behavior of metallic, polymeric materials, foams, thermographic NDT.