
Peer reviewed version

License (if available):
CC BY-NC-ND

Link to published version (if available):
10.1016/j.compositesa.2021.106527

Link to publication record in Explore Bristol Research
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via Elsevier at https://doi.org/10.1016/j.compositesa.2021.106527. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/
**Phase Segmentation of Uncured Prepreg X-Ray CT Micrographs**

Pedro Galvez-Hernandez, Karolina Gaska, James Kratz

Bristol Composites Institute (ACCIS), Department of Aerospace Engineering

University of Bristol, University Walk, Bristol UK BS8 1TR

1Corresponding author: james.kratz@bristol.ac.uk

**Abstract**

This paper explores methods to investigate voids and dry fibre areas in pre-impregnated aligned carbon fibre reinforced epoxy lay-ups. A deep learning segmentation approach was compared to conventional thresholding techniques to characterise the interlaminar voids (entrapped air) and dry areas (unsaturated fibre bed) phases obtained by micro-CT scanning of samples from uncured laminates. The performance of both approaches was quantitatively assessed in three regions of interest having different levels of porosity, ranging from a low 1% to a high 25%. Deep learning consistently outperformed thresholding in the segmentation of both interlaminar voidage and dry areas. Furthermore, deep learning improved the ability to detect small voids and was able to accurately segment voids in volumes with less than 2% voidage, whereas thresholding techniques fail in this task. Finally, the application of deep learning to the segmentation of dry areas in micro-CT scans provided sharper results than thresholding, without needing filtering.

Key words: A. Prepreg; B. Porosity; D. CT analysis; Machine learning
Graphical abstract
1 Introduction

Considerable work has been done studying porosity in advanced composites. A recent review summarising nearly five decades of research on voids in composite laminates emphasised the need to minimise porosity in cured laminates [1]. Generally speaking, the review showed that a 1% increase in void content leads to a 10% decrease in interlaminar shear strength in thermoset resin laminates made with layers of aligned unidirectional fibres [1]. To minimise the risk of porosity in high-performance composite parts, the autoclave process is preferred over lower pressure options such as vacuum-bag-only cure of thermoset prepregs or in-situ consolidation of thermoplastic prepregs.

Investigating the initial morphology of prepreg laminates before manufacturing provides valuable information about the initial porosity needing to be filled and prevent further void nucleation and growth [2]. Porosity can be separated into two categories, as shown in Figure 1. Firstly, there are voids: the spaces between plies that mechanically trap air during deposition due to the surface properties of prepreg, such as roughness and surface energy. There is another potential source of voids – air/volatiles entrapped within the resin film itself. Secondly, there are unsaturated dry areas: the reinforcing spread fibre tows that do not contain polymer after the pre-impregnation process. In most prepreg materials, the volume of voids and dry tows in the uncured laminate is above 10%. This initial porosity will influence evacuation of entrapped air and other volatiles during the early stages of processing before resin gelation [3] and have a strong influencing the final cured laminate voidage [4].

![Figure 1: Schematic of interlaminar voids and dry areas in uncured prepreg laminates.](image)

A promising method to detect porosity in uncured laminates is X-ray computed tomography (XCT). It is often used for studying the microstructure and morphology of materials in a wide range of
applications, including medicine, aerospace, automotive and defence sectors [5-7]. This technique is especially useful in applications where it follows the evolution of the microstructure at different stages of the manufacturing process [8, 9], observing changes in critical features [10, 11], and structural integrity under applied loads [12, 13]. When applied to composite materials, XCT could be one of the most accurate techniques available for 3D evaluation of microscopic features [14]. However, challenges remain when choosing optimal machine settings during image acquisition and pre-processing the acquired data [15].

The output from a CT scan is presented as groups of voxels (3D pixels). Each voxel has a grey level strictly connected to the effective linear attenuation coefficient for X-rays of the incident beam [16] interacting with the sample. The linear attenuation coefficient is function of the material density and the atomic number [5]. Therefore, materials with a different density will be displayed at different grey levels. In the case of uncured composites, three phases can be identified in a CT image [17]. Firstly, the air phase is represented by voxels with lower grey levels due to its low attenuation coefficient. Secondly, the dry areas containing both fibres and air, have a grey level depending on the weighted sum of the attenuation coefficient considering the fibre fraction within the volume of the voxel [16]. Finally, the composite phase, containing resin-only and resin-impregnated fibre areas, is represented with higher intensity grey values due to the higher density of its constituents. However, the presence of noise, artefacts and low contrast might hinder a clear identification of these phases in the corresponding grey scale histogram, requiring of additional pre-processing steps to facilitate a subsequent segmentation.

Segmentation methods are proposed in the literature to distinguish between constituents of composites, such as the polymer matrix, carbon or other reinforcing fibres, porosity etc. Pre-processing before segmentation may improve image quality. Usually, filters to enhance contrast or smooth the image are found in software, such as VGStudio Max® [18], Avizo® [19] or Fiji [20]. After pre-processing, segmentation is applied to assign each voxel the “air” or “composite” phase [16]. Segmentation is a critical step since the segmented images are used to evaluate the porosity level in the selected volume under investigation.
A common technique for segmenting the different phases in a CT image is the selection of a grey value threshold in the greyscale histogram of the input scan. The grey values below the threshold are assigned a new value of 255 (white), and the grey values above the threshold are assigned a new value of 0 (black), in an 8-bit image. The threshold can be chosen either after visually analysing the histogram shape [21] or it can be determined based on the histogram properties. Otsu’s segmentation method [22], based on minimising the intra-class variance of the histogram, the ISO-50% method [23], where the middle grey value between two consecutive peaks in the scan histogram, or the grey value with the lowest counts between two peaks, i.e., the local minimum [24] are often selected as the grey value threshold. A comprehensive review of different thresholding segmentation techniques can be found in the paper by Tretiak and Smith [16].

Developments in machine learning might overcome the drawbacks of segmenting and thresholding porosity in composite materials. Machine learning models have been applied as a framework to address: discovering new chemical compounds or molecules with designed physical properties [25-29], automated feature detection [30-32], and even predicting properties from the microstructure [33]. Specific examples of where machine learning has been applied to CT images was done by Sinchuk et al. [34], who proposed and evaluated two different approaches for solving the segmentation problem of very-low-contrast X-ray CT images of carbon epoxy woven composites by means of variational and deep-learning-based segmentation methods. Ali et al. [35] trained a deep learning model for the generation of virtual specimens of fibrous reinforcements via the segmentation of CT images of real specimens. Similarly, Badran et al. [36] used a deep learning-based procedure for automatic segmentation of 3D tomography images from fibre reinforced ceramic composites.

In this paper, we explore the benefit of using machine learning to quantify the interlaminar voids and dry areas in micro-CT scans of uncured composite laminates. Convolutional Neural Networks (CNN) that have recently gained interest among material scientists are used here to train a deep learning model based on the convolution of a sequence of filters using only raw, unprocessed X-ray CT images. The performance of the deep learning model was found to consistently exceed the thresholding
approach when compared to the ground truth. This new method to investigate phases in composite materials is remarkably better at quantifying low porosity levels in high-performance composites.

2 Materials and methods

2.1 Sample preparation

An out-of-autoclave prepreg material was selected for this study. A Hexcel HexPly® M56 epoxy matrix with 35% resin content and reinforced with IM7 unidirectional fibers having an areal weight of 268 g/m² was used. Initially a 30-ply 0° laminate with dimensions of 100 mm × 100 mm was laid-up by hand and consolidated under vacuum for 10 min at ambient temperature (20 ± 1 °C). Thereafter, three rectangular test samples of 6 mm × 100 mm were cut from the centre of the laminate with the 0° fibres parallel to the long edge. The test samples remained uncured throughout the study.

The three samples are referenced as follows: Training Sample 1, Training Sample 2, and Validation Sample. The CT scans from Training Samples 1–2 were used to train the deep learning model. Validation Sample was used to validate the deep learning model and compare it to the standard thresholding methods as shown in Figure 2.

![Flowchart](image)

Figure 2: Flowchart of the image pre-processing, segmentation, and analysis to define the voids and dry areas of the uncured prepreg. The steps associated to the deep learning and thresholding segmentations are represented in blue and red, respectively.

2.2 X-Ray CT Setup

A lab-based Nikon XTH-320 CT Scanner was used to investigate the laminate. A voltage of 103 kV and a power of 9.5 W was applied to all three scans. The exposure time was set at 500 ms/projection and 4 frames/projection, resulting in a scan time of 1 hour and 6 minutes. The final resolution was 8.24 µm/voxel, which is in the typical range for lab-based CT scanners, but greater than
the fibre diameter of 5 \( \mu \text{m} \). Therefore, individual fibres cannot be separated from the polymer matrix within a voxel and features below the size of a voxel are unlikely to be detected [37]. The resulting scans were reconstructed with the Nikon CT Pro software. No filter was applied during reconstruction.

2.3 Segmentation methods

An image pre-processing flow was implemented in Fiji (ImageJ) [20] and applied to each scan (see Figure 2). Firstly, the air surrounding each sample in the scan was cropped and converted to 8-bit (256 grey values). Secondly, the phase contrast of the scan was enhanced by stretching the scan histogram (voxel counts for each grey value). This operation assigns a value of 0 to the minimum and 255 to the maximum grey values initially present in the scan. The remaining grey values were linearly mapped to this new range. To prevent voxels with abnormally low or high values negatively affecting the histogram normalisation, a set of voxels (0.2\% of the total voxels) with extreme grey values were automatically assigned 0 and 255 grey values, i.e., saturated values.

2.3.1 Thresholding

A conventional thresholding approach was applied to segment voids and dry areas in the uncured composite samples. Before selecting the thresholds (one is needed for each phase), the Validation Sample sub-volume was denoised with a 3D-Median filter of size 2 using Fiji. This filter preserves the sharpness of the image while reducing the noise and artefacts in the original scan by replacing the grey value of a given voxel by the median grey value of the voxels surrounding it. To illustrate the effect of the intraphase homogenisation achieved with this operation, five patches of dimensions 5 \( \times \) 15 pixels were selected within each of the three phases (air, dry areas and resin-saturated areas) and their greyscale histograms were generated as shown in Figure 3. The average and standard deviation grey value for each phase in the original and filtered scans are, respectively: 62 ± 12 and 62 ± 8 for the interlaminar voids, 132 ± 17 and 132 ± 12 for the dry areas and 162 ± 8 and 161 ± 4 for the resin-saturated areas. The reduction in the dispersion indicates a higher intraphase homogenisation that ultimately eases a further selection of the thresholds to separate the three different phases. Following the filtering, the contrast was enhanced and the histogram was generated. The peaks and minimum grey
values corresponding to the different phases are identified as shown in Figure 4 and summarised in Table 1.

Three separate peaks representing each phase of the laminate become visible in the scan histograms shown in Figure 3e and Figure 4. The first peak represents air. The second and third peaks represent unsaturated dry fibre areas and resin-saturated areas (resin-only and resin-impregnated fibre areas), respectively. Since the scans were performed at a higher resolution than the diameter of a single fibre, it was not possible to explicitly separate the peaks corresponding to the fibres and resin-saturated areas in the resulting histogram.

The thresholds chosen for segmenting the air voids and dry areas are based on two methods:

- **Th 1**: The thresholds are selected based on the ISO-50% technique [23]. The mid-grey value was calculated for two consecutive peaks and chosen as the threshold for each phase. The range of grey values covering each phase was as follows:
  - Interlaminar voids: [0 - 111]
  - Dry Areas: [112 – 195]

- **Th 2**: The thresholds correspond to the local minimum, excluding grey values with no counts, between two consecutive peaks in the greyscale histogram [24]. This method provided the following grey value ranges:
  - Interlaminar voids: [0 – 72]
  - Dry Areas: [73 – 185]

Finally, the thresholds were applied to the enhanced contrast scans, resulting in four different binary 3D images (one binary image per phase and per thresholding method). The voxels with grey values falling into the range of interest are assigned a value of 255 (white), leaving the rest of the voxels with a value of 0 (black).
Figure 3: Original and filtered scans (a and d) together with their associated greyscale histograms (b and e) and the histogram of the specific patches covering each phase (c and f). The effect of the median filter in removing noise and providing higher intraphase homogeneity is a reduced grey value spread for each phase.

Figure 4: X-ray micrograph after enhancing the contrast of the filtered scan (a) and grey values of interest in its associated histogram (b). Full label definition summarised in Table 1.
Table 1: Grey values of interest from the histogram in Figure 4b.

<table>
<thead>
<tr>
<th>Histogram Label</th>
<th>Description</th>
<th>Grey Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Air Peak</td>
<td>49</td>
</tr>
<tr>
<td>B</td>
<td>Air – Dry Areas Minimum</td>
<td>72</td>
</tr>
<tr>
<td>C</td>
<td>ISO-50% Air – Dry Areas</td>
<td>111</td>
</tr>
<tr>
<td>D</td>
<td>Dry Areas Peak</td>
<td>173</td>
</tr>
<tr>
<td>E</td>
<td>Dry Areas – Resin-saturated areas Minimum</td>
<td>185</td>
</tr>
<tr>
<td>F</td>
<td>ISO-50% Dry Areas – Resin-saturated areas</td>
<td>195</td>
</tr>
<tr>
<td>G</td>
<td>Resin-saturated areas Peak</td>
<td>217</td>
</tr>
</tbody>
</table>

2.3.2 Deep Learning

Model Definition

The deep learning approach used in this study is based on the U-Net architecture developed by Ronnenberger et al. [38]. U-Net is a convolution neural network (CNN) that has been successfully applied to phase segmentation in a wide range of fields, such as civil engineering [39], biology [40] and additive manufacturing [41]. A major advantage of U-Net is in providing satisfactory results despite noise, artefacts or overall low quality of the input data set. In our study, the U-Net architecture was implemented in Python 3.6 and Tensorflow 2.1 [42].

The visual representation of the U-Net architecture is shown in Figure 5. Starting from the left-hand side of the U-shaped model, an input 2D image of size 128 × 128 pixels is taken through a series of blocks that generate a set of feature maps containing weights that store information about contextual and intrinsic properties of each pixel in the input image. The first part of the model is called “contracting path” in Blocks 1-5. Each of the blocks contains two 3 × 3 convolutional layers with a rectified linear unit (ReLU) [43] as an activation function. Blocks 2-5 are preceded by a 2 × 2 maximum pooling layer that halves the feature maps in the x-y dimension. Furthermore, at each block, the number of feature
maps are doubled from the previous block value. Following the contracting path and after reducing the feature maps in the x-y dimension by sixteen, the “expanding path” begins (Blocks 6-10). This path is symmetrical to the contracting path, but the feature maps in the x-y dimensions are first doubled (2 × 2 upsampling followed by a 2 × 2 convolution). The resulting feature map is concatenated with the feature maps from the equivalent block in the contracting path and passed through two additional convolutional layers. In the expanding path, the number of feature maps is halved at each block. Finally, a classification layer based on a 1 × 1 convolution and the soft-max activation function (Eq. 1), is applied to the final set of feature maps to provide a 3 pixel-wise probability map, one for each class (background, voids, and dry areas):

\[ p_i(z) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \]  

(1)

where \( p_i(z) \) corresponds to the \( i^{th} \) class probability calculated at each pixel given the output values \( z \) from the previous 1 × 1 convolutional layer for a set of \( C \) classes.

Figure 5: U-Net architecture with the number of feature maps (green) and feature map x-y dimensions (red) for each convolutional layer. PlotNeuralNet was used to create this schematic [44].
Model Training

In order to train the model, a total of one hundred 2D images of size 128 × 128 pixels were cropped out from the Sample 1 and Sample 2 scans. From each image, a ground truth mask for voids and dry areas was manually annotated using the following annotation software: VGG Image Annotator [45] and Pixel Annotation Tool [46], respectively. An additional mask is automatically generated for the background class. The ground truth masks are binary images that contain two types of labels. Firstly, positive labels, i.e., pixels that belong to the class of interest (either voids or dry areas) and are displayed in white, as shown in Figure 6. Secondly, negative labels, i.e., pixels that do not belong to the class of interest and are represented in black. For example, in the case of the voids ground truth masks, the pixels representing a void will receive a positive label, i.e., will be displayed as white pixels, and the pixels corresponding to the background and dry areas will be displayed in black.

![Figure 6: Example of a patch (a) used for training and its associated ground truth masks for interlaminar voids (b) and dry areas (c). Positive segmentation labels are represented in white and negative segmentation labels in black.](image)

The training images together with their respective masks are fed into the model in batches of a user-defined number of images. The model then goes through the batch and generates a prediction for the voids and dry areas depending on the value of the convolutional layers weights. The predicted mask was compared to the ground truth mask for each class and an error was computed using the Categorical Cross-Entropy Loss function:

\[
Loss = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{C} [y_{ij} \log(p_{ij})]
\]  

(2)
where $N$ represents the total number of pixels contained in the batch for a given iteration, $C$ is the total number of classes, and $p_{ji}$ and $y_{ji}$ are the pixel score provided by the softmax function (Eq. 1) in the classification layer and the ground truth label respectively.

The computed loss was then backpropagated through the network and the weights of the convolutional layers are adjusted using the Adam optimizer [47] with a learning rate of 0.0001. Each complete training cycle, comprising the prediction of a training set, the loss computation, and the adjustment of the weights, is called an “epoch”. In this study, the number of epochs was set to 75 and the batch size was 1. The training set loss evolution is shown in Figure 7a. The model training and the Validation Sample scan prediction were performed on an Intel® Xeon® Gold 5220 CPU and took 17 hours to complete.

![Graphs showing training and control set loss evolution](image)

*Figure 7: Training set loss a) and control set loss evolution b) during model training for twenty-five, fifty, and one hundred 2D images.*

To decide when the model was trained and to prevent overfitting, i.e., the case in which the model successfully predicts the classes in the training set images but fails to predict unknown data, a control set of twenty 2D images together with their void, dry areas and background masks, was also prepared. The control set was unknown to the training weights and allowed the measurement of the model performance on unknown data, also known as “generalisation”. The evolution of the control set loss was computed and tracked as shown in Figure 7b. The set of model weights that minimise the control set loss at a given epoch was saved and used for future predictions.
Figure 7b shows that training was completed after fourteen, twenty-three, and twenty-two epochs and a minimum loss of 0.18, 0.16 and 0.14 for DL-25, DL-50, and DL-100, respectively. Model overfitting was observed after the control set loss reached a minimum. At this point, the reduction of the training set loss does not translate into a reduction of the control set loss, therefore this indicated that the model performance on unknown data has decreased.

2.3.3 Entire Scan Characterisation

The main goal is to segment voids and dry areas within the volumes obtained via CT-scanning. However, the proposed deep learning model uses 2D patches of 128 × 128 pixels whereas the 3D scan consisted of one thousand 2D-slices having 700 × 700 pixels. The strategy to overcome this obstacle was based on the “Overlap-tile” principle, also proposed by Ronnenberger et al. [38]. In this approach, each 2D slice was divided into multiple overlapping tiles of 128 × 128 pixels. Each tile was passed to the network, and the voids and dry areas contained in the tile were predicted. Since the tiles have an overlap of fourteen pixels, only an inner tile of size 100 × 100 was kept after prediction. This procedure was applied over the entire set of 2D-slices for a given scan.

The prediction set for the deep learning method consists of three probability maps, where each voxel is assigned a probability of belonging to each of the three classes. The sum of the class probabilities for a given voxel is 100%. The probability maps are then converted into binary masks after applying a cut-off probability $p_t$. This means that a voxel was assigned to the interlaminar voids or dry areas classes only if the probability associated to either class is equal or higher than $p_t$, otherwise it is assigned to the background class. In this study, choosing $p_t = 55\%$ allowed the segmentation of a wide range of interlaminar voids and dry areas. Consequently, the binary masks contain positive and negative labels, depending on whether the pixel was assigned a given class or not, in the same fashion as the ground truth masks were generated. Void and dry areas percentages, average volumes and counts are calculated with Fiji plugin BoneJ [48].

2.3.4 Performance Evaluation

The performance of the deep learning approach for voids and dry areas segmentation was evaluated against the thresholding method. For this task, three Regions of Interest (ROIs) of size 128 ×
128 × 20 voxels (1 mm × 1 mm × 0.16 mm) were selected as shown in Figure 8a. The ROIs are located in the central sub-volume of the scan, spanning slices 491–510. Each ROI was chosen because of its different void distribution and content through visual inspection. This selection allows the measurement of the performance of the segmentation techniques to address different porosity levels as well as the definition of the optimum training set that best predicts voids and dry areas in a 3D-image.

The evaluation of the segmentation performance consists of four steps:

1. Generate the voids and dry areas ground truths for each ROI (Figure 8b). These ground truths will serve as the benchmark for the segmented images.

2. Crop out the ROIs from the scans segmented by deep learning and thresholding.

3. Definition of a confusion matrix containing the four categories that can be assigned to a voxel depending on its class prediction. For each class (voids and dry areas), the categories are:
   - True Positive (TP): The voxel is segmented as positive and has a positive value in the ground truth mask,
   - False Positive (FP): The voxel is segmented as positive, but it is negative in the ground truth mask,
   - True Negative (TN): The voxel segmentation is negative, and it has a negative value in the ground truth mask, and
   - False Negative (FN): The voxel segmentation is negative, but the ground truth value is positive.

4. Based on the category of the voxel (TP, TN, FP and FN), the following metrics are generated to evaluate the performance of each segmentation method:

\[
Precision = \frac{TP}{TP + FP} \quad (3)
\]

\[
Recall = \frac{TP}{TP + FN} \quad (4)
\]
The precision for a given class (either voids or dry areas) provides information on the level of noise or false positives. A high precision value will indicate that most of the positive predictions match the positive ground truth labels. The Recall informs the ability of the predictor to capture positive ground truth labels. Therefore, a high Recall will indicate a low amount of false negatives. Finally, the MCC (Mathews Correlation Coefficient) \[49\] provides information on the overall performance of the segmentation technique integrating the four voxels categories (TP, TN, FP and FN) into a unique formula. As a result, MCC returns a high score only if the segmentation approach produces a good estimation of all categories \[50\]. Precision and Recall are defined in the range \([0-1]\) and the MCC in \([-1, 1]\).

Figure 8: a) Location of the Regions of Interest in slice 491 of Validation Sample. ROI 1 outlined in blue, ROI 2 in yellow, and ROI 3 in red. b) Ground truth masks generated for each of the Regions of Interest and each class illustrating positive voxel labels (white) and negative voxel label (black).
3 Results

3.1 Entire Scan Characterisation

The quantitative results from the thresholding and deep learning segmentation are summarised in Figure 9. At first glance, the global results look similar for total sample porosity, however deep learning was found to identify porosity with greater precision than thresholding. Overall, both segmentation methods are able to detect voids and unsaturated dry areas in the uncured composite. However, deep learning had a lower count and larger average volume in the dry areas. As a result, deep learning segmentation was able to identify the true void and dry area phases, as described in the following sections.

Figure 9: Interlaminar voids and dry areas quantitative assessment for each segmentation method.

For each feature, their respective percentage in the scan together with their counts and the average volume is plotted for each segmentation approach.
3.1.1 Voids

Both segmentation approaches were able to segment medium and large interlaminar voids, leading to the similar percentages shown in Figure 9. However, the segmentation of small interlaminar voids is where major differences can be found between deep learning and thresholding. On the one hand, deep learning based techniques captured small interlaminar voids with the overall segmentation accuracy slightly increasing after doubling the training set size from twenty-five (DL-25) to fifty (DL-50) images. A further doubling in the size of the training set to one hundred images did not result in a significant change in the segmentation performance. On the other hand, thresholding-based methods failed to segment the small interlaminar voids as their grey values fall outside of the grey value range chosen for each thresholding option described in Section 2.3.1. Furthermore, dry areas containing darker grey values are also incorrectly segmented as voids as the context of such voxels was not considered during thresholding segmentation unlike the results based on deep learning segmentation. This behaviour can be visually assessed in Figure 10. DL-50 captures finer details and fewer incorrectly labelled voxels than Th 1, which is the thresholding-based technique with a higher MCC score (see Table 2). Th 2 provided the lowest porosity estimation (4.30%) whereas DL-25 produced the highest porosity (5.58%). Th 2 also resulted in the lowest void counts at 6696 and highest average void volume at 1.7E6 µm³, indicating a low performance for capturing small interlaminar voids.

![Image](image_url)

*Figure 10: Interlaminar voids segmentation using Th 1(a) and DL-50 (b). ROIs locations are represented in blue (ROI 1), yellow (ROI 2) and red (ROI 3). Positive segmentation labels are represented in white and negative segmentation labels in black. Presented images correspond to the slice 491 of the scan.*
3.1.2 Dry Areas

Unlike voids, the segmentation of dry areas represents a big challenge for thresholding-based methods, which fail to provide an accurate prediction of their volume. As a consequence, dry areas cannot be correctly captured by these methods. Figure 11a shows a noisy segmentation because of a high number of false positives and unfilled dry area volumes due to more false negatives. However, since deep learning methods are not solely based on the voxel grey value but also consider the context of each of them, the accuracy of the dry area segmentation increases significantly, providing sharper and cleaner visual results as shown in Figure 11b.

![Figure 11: Dry Areas segmentation using Th 1(a) and DL-50 (b). ROIs locations are represented in blue (ROI 1), yellow (ROI 2) and red (ROI 3). Positive segmentation labels are represented in white and negative segmentation labels in black. Presented images correspond to the slice 491 of the entire scan.](image)

Quantitative results of the dry areas segmentation are also given in Figure 9. The dry areas percentage is estimated at the same level by all the approaches. However, notable discrepancies exist in the dry areas counts and average volumes from each technique. Deep learning methods result in a significant decrease in the dry areas counts and higher average dry area volumes with respect to the thresholding technique. In other words, deep learning models captured a lower number of false positives and false negatives than thresholding. Consequently, deep learning models provided a more homogeneous and consistent segmentation of dry areas than thresholding.
3.1.3 3D reconstruction

A 3D reconstruction of voids and dry areas for DL-50 and Th-1 is shown in Figure 12. When using thresholding to reconstruct the interlaminar voids and dry area phases the result is noisy. On the other hand, void and dry area volumes reconstructed from deep learning segmentation showed higher continuity in the 3D space. These two approaches might exhibit similar voids and dry areas percentages. However, DL-50 provides a cleaner and better defined volume reconstruction. The reconstruction was done with Avizo® 2019.1.

![Figure 12: 3D reconstruction of interlaminar voids (white) and dry areas (green) for Th 1 (a) and DL-50 (b). The edge view of the volume from the 1st slice of the scan (left-hand side) and a perspective view (right-hand side) are shown.](image-url)
3.2 Performance Evaluation

3.2.1 Voids

**ROI 1**

This ROI had the smallest size and volume of voids chosen from Figure 8a, with a ground-truth void percentage of 1.07% (Table 2). The interlaminar voidage was underestimated by Th 1 and Th 2 because most of the voids were represented by grey values that fell outside of the range defined for both thresholding techniques. Th 1 estimates a voidage of 0.53% and Th 2 0.15%, which are 50% and 86% lower than the ground truth, respectively. This underestimation is corroborated by their respective recall values, both of which fall below 0.5. The recall metric indicates that thresholding fails in segmenting positive voids labels present in the ground truth volume. Furthermore, the fact that the precision is 1.0 for Th 2 indicates that all the segmented air voxels are present in the ground truth, even though there are few as indicated by the low recall.

Deep learning-based methods provided a better performance across all three parameters. The interlaminar void percentage estimation is higher than the thresholding-based methods, even though is still below the ground truth value. DL-25 returned a voidage of 0.86%, DL-50 was 0.91%, and DL-100 was 0.81% (Table 2). DL-50 was the closest to the ground truth value. Precision remained above 0.97 for the three models and the recall increased only by 5% after doubling the training set size from 25 to 50 images. No further improvement in recall was achieved for DL-100.

The MCC values for Th 1 (0.43) and Th 2 (0.37) were significantly lower than the deep learning model values (>0.86), in agreement with the precision and recall results previously discussed.

**ROI 2**

This region contains medium-sized interlaminar voids (Figure 8a), and its ground truth interlaminar voidage was 6.09%. The difference between the values obtained by the thresholding-based methods and the ground-truth is slightly less acute in ROI 2 than for ROI 1. Th 1 and Th 2 provide a percentage of interlaminar voids of 4.88% (20% lower) and 3.63% (40% lower). Precision remains >0.98 for both thresholding methods and recall for Th 2 is 0.59, which notably improves ROI 1 recall. Recall of Th 1 also improves (0.78).
Similarly, to the behaviour observed in ROI 1, deep learning-based methods provide interlaminar void percentage values closer to ground truth, with $DL_{-25} = 6.19\%$ (1.6\% higher), $DL_{-50} = 5.78\%$ (5\% lower) and $DL_{-100} = 5.41\%$ (11\% lower) $DL_{-25}$ provides the highest porosity estimation due to the combination of more false positives and fewer false negatives (lower precision and higher recall) than in $DL_{-50}$ and $DL_{-100}$. These methods also achieve high precision (>0.93) and high recall (>0.87). Recall in ROI 2 is improved with respect to ROI 1 values, and therefore a higher amount of true air voxels are now captured.

Thresholding-based methods improved their overall performance, given by the MCC score. MCC for $Th_{1}$ and $Th_{2}$ was roughly doubled compared to ROI 1.

$Table 2$: Interlaminar voids segmentation results in each ROI.

<table>
<thead>
<tr>
<th>ROI Id</th>
<th>Segmentation Method</th>
<th>(%)</th>
<th>Precision</th>
<th>Recall</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROI 1</td>
<td>GT</td>
<td>1.07</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Th 1</td>
<td>0.53</td>
<td>0.62</td>
<td>0.31</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Th 2</td>
<td>0.15</td>
<td>1.00</td>
<td>0.14</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>DL-25</td>
<td>0.86</td>
<td>0.97</td>
<td>0.78</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>DL-50</td>
<td>0.91</td>
<td>0.97</td>
<td>0.82</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>DL-100</td>
<td>0.81</td>
<td>0.99</td>
<td>0.75</td>
<td>0.86</td>
</tr>
<tr>
<td>ROI 2</td>
<td>GT</td>
<td>6.09</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Th 1</td>
<td>4.88</td>
<td>0.98</td>
<td>0.78</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>Th 2</td>
<td>3.63</td>
<td>1.00</td>
<td>0.59</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>DL-25</td>
<td>6.19</td>
<td>0.93</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>DL-50</td>
<td>5.78</td>
<td>0.96</td>
<td>0.91</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>DL-100</td>
<td>5.41</td>
<td>0.98</td>
<td>0.87</td>
<td>0.92</td>
</tr>
<tr>
<td>ROI 3</td>
<td>GT</td>
<td>26.4</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Th 1</td>
<td>25.4</td>
<td>0.99</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>Th 2</td>
<td>24.2</td>
<td>1.00</td>
<td>0.91</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>DL-25</td>
<td>25.6</td>
<td>0.99</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>DL-50</td>
<td>25.4</td>
<td>1.00</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>DL-100</td>
<td>24.9</td>
<td>1.00</td>
<td>0.94</td>
<td>0.96</td>
</tr>
</tbody>
</table>

$ROI 3$

As shown in Figure 8a, large voids dominate this sub-volume, and the ground truth percentage of interlaminar voids reaches 26.4\%. Both thresholding and deep learning segmentation methods provided void values close to the ground truth. Precision was >0.99 and recall >0.91 for all segmentation
methods. Furthermore, a high value in the MCC score confirms the high segmentation performance for all techniques.

3.2.2 Dry Areas

The percentage of dry areas in all three ROIs is similar upon visual inspection of Figure 8a and was confirmed by the dry area average ground truth value (18.9 ± 1.6%) for the three ROIs.

Considering precision, recall and MCC, for example in ROI 1, thresholding-based method Th 1 achieves the lowest precision (0.87) whereas Th 2 records the lowest recall (0.58) and MCC (0.67). Th 2 is also the methodology capturing the lowest percentage of dry areas (Table 3). The deep learning-based approaches were more successful across all three metrics (for DL-50 precision = 0.89, recall = 0.92, MCC = 0.88). A similar trend was observed for ROI 2 and 3 (see Table 3).

Table 3: Dry areas segmentation results in each ROI.

<table>
<thead>
<tr>
<th>ROI Id</th>
<th>Segmentation Method</th>
<th>(%)</th>
<th>Precision</th>
<th>Recall</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROI 1</td>
<td>GT</td>
<td>20.7</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Th 1</td>
<td>16.0</td>
<td>0.87</td>
<td>0.67</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>Th 2</td>
<td>13.3</td>
<td>0.90</td>
<td>0.58</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>DL-25</td>
<td>21.9</td>
<td>0.86</td>
<td>0.91</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>DL-50</td>
<td>21.3</td>
<td>0.89</td>
<td>0.92</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>DL-100</td>
<td>19.3</td>
<td>0.93</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>ROI 2</td>
<td>GT</td>
<td>18.1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Th 1</td>
<td>17.5</td>
<td>0.75</td>
<td>0.73</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>Th 2</td>
<td>15.8</td>
<td>0.72</td>
<td>0.63</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>DL-25</td>
<td>19.3</td>
<td>0.85</td>
<td>0.91</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>DL-50</td>
<td>18.4</td>
<td>0.88</td>
<td>0.90</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>DL-100</td>
<td>16.2</td>
<td>0.94</td>
<td>0.84</td>
<td>0.86</td>
</tr>
<tr>
<td>ROI 3</td>
<td>GT</td>
<td>17.9</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Th 1</td>
<td>15.4</td>
<td>0.83</td>
<td>0.71</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>Th 2</td>
<td>14.4</td>
<td>0.79</td>
<td>0.64</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>DL-25</td>
<td>19.0</td>
<td>0.87</td>
<td>0.93</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>DL-50</td>
<td>19.2</td>
<td>0.88</td>
<td>0.94</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>DL-100</td>
<td>17.0</td>
<td>0.93</td>
<td>0.89</td>
<td>0.89</td>
</tr>
</tbody>
</table>
4 Discussion

Deep learning techniques to segment the phases of micro-CT scans of uncured composite laminates yielded higher accuracy void and dry areas detection when compared to standard segmentation techniques based on thresholding. Furthermore, the deep learning approach captured a wider range of porosity levels that are usually present in uncured laminates, which thresholding-methods failed to fully recognise.

Deep Learning showed a robust behaviour against noise, artefacts or intra-phase heterogeneity present in the scan and provided an accurate segmentation of interlaminar voids and dry areas. Thresholding, unlike deep learning, required a pre-processing step involving the scan denoising by convolving the scan with a median filter to obtain higher intra-phase homogeneity and remove the effect of abnormally high/low grey values. This process results in the identification of three peaks in the scan histogram and facilitates the threshold selection and the void and dry areas segmentation. However, it produces an unavoidable blurring of the scan that, even if small, produces a loss of information.

Deep learning methods provided higher MCC values for both interlaminar voids and dry areas. Of particular interest, it has been proven that deep learning approaches outperform thresholding in volumes containing low interlaminar voidage, as shown in Figure 13. The best thresholding method (Th 1) achieved a score of 0.43, whereas deep learning counterpart was found to be twice as good, reaching 0.89. Low porosity identification is especially relevant for the study of the mechanical properties of high-performance composite parts, where 2% porosity is usually the maximum acceptable level. As the size of the interlaminar voids increases, such as in ROI 2 and ROI 3, the grey values associated with the voids become darker, entering the segmentation range of the thresholding-based methods. Therefore, a higher amount of true positives were found after a ground truth percentage of interlaminar voids exceeded 5%, where the segmentation performance of thresholding and deep learning converge; no significant differences in the MCC score were observed above 5% voidage.
The effect of false positives and false negatives on the phase segmentation is shown in Figure 14 for the thresholding and deep learning methodologies. Overall, the deep learning segmentation approaches are better at identifying true interlaminar voids and is particularly accurate for lower voidage levels.

Deep learning continued to outperform thresholding when classifying the dry areas. Even though the percentage of the dry areas was estimated to be similar regardless of the segmentation approach used (see Figure 9b), the segmentation noise was higher with thresholding. The noise manifests in a higher dry count (see Figure 9d) and lower average volumes (see Figure 9f) for thresholding. As a result, thresholding-based methods captured a high proportion of false positives and failed to capture a large number of true positives. In contrast, the deep learning models predict a higher number of true positives and lower proportion of false positives. Figure 14 shows sharper and cleaner results from deep learning, and a better representation the real dry areas content in the material. This effect can also be observed in Figure 12, where the 3D volume generated from the deep learning method is characterised by a well-organised distribution of dry areas and interlaminar voids.

Figure 13: MCC dependency on the ground truth (GT) interlaminar voids and dry areas percentage.
Figure 14: Visual performance of DL-50 and Th-1 for the segmentation of interlaminar voids and dry areas. Pixel-wise classification is illustrated as follows: TP (green), TN (white), FP (red) and FN (blue).

Even though the deep learning approach to phase segmentation of uncured prepregs has been shown to outperform thresholding, an advantage to thresholding is quasi-instant generation of the segmented 3D image from CT data. Training a deep learning model took three and a half hours. This time can be reduced by using a GPU (Graphical Processor Unit) instead of the computer CPU (Central Processing Unit), as GPUs are optimized to perform matrix convolution, which is the predominant mathematical concept in CNN.

A training set of fifty images was found to provide the best performance overall for both voids and dry areas regardless of the ROI. However, a model trained with a set of twenty-five images, and therefore requiring half of the annotation and training effort, accounts for a decrease of the MCC value of only 2% in the segmentation of interlaminar voids in ROI 1, and no impact in the other ROIs. Similarly, the dry areas MCC only decreases by an average of 2% in all ROIs when using a training set.
of twenty-five images. No significant performance improvement was observed when further doubling the training set size to one hundred images.

Even though X-ray micro-CT considers the whole specimen volume and provides a more realistic estimate of void level than optical microscopy, the technique is only accurate to within one voxel [37]. As a consequence, voids smaller than one voxel are neglected from the void level calculation. A hypothetical increase in scan resolution was considered in Appendix A. The analysis in Figure 15 shows that the influence is most pronounced in materials having low void volumes, but a significant number of previously undetected one-voxel voids are needed to increase the porosity. The actual number of undetected voids remains an open question and motivates further studies into X-ray CT scan resolution of low void fraction composite materials. When improvements in scan resolution arrive, Figure 15 suggests that phase segmentation by deep learning is better positioned than thresholding to detect small voids.

5 Conclusions

Machine learning was explored as a segmentation technique for the three main phases that are present in uncured prepreg materials. A laminate of unidirectional out-of-autoclave prepreg was laid-up by hand and lightly consolidated under a vacuum bag before X-ray CT data was collected. The total volume of pores in the laminate was segmented by a deep learning convolutional neural network and conventional thresholding to identify the interlaminar void spaces and unsaturated dry areas of reinforcing fibres. Both types of porosity need to be filled with resin during the manufacturing process to produce high-performance composite structures.

Deep learning was proven to be a suitable technique to segment voids and dry areas from CT images of uncured composite laminates. Specifically, deep learning models more accurately captured the void and dry area percentages, as well as its distribution. Overall, the results of segmentation by deep learning are superior to histogram thresholding and are particularly well-suited to detect small voids. For example, when the total volume porosity was below 2%, deep learning outperforms histogram thresholding of the ground truth by a factor of two. In addition, the segmentation based on deep learning was able to provide better performance than thresholding, despite the noise and ring
artefacts present in the raw CT data. Therefore, an added benefit of deep learning segmentation is that it does not require the application of a filter in the pre-processing stage, unlike thresholding, to improve the phase identification, preventing any information loss or image deformation.

Potential applications of phase segmentation of X-ray micrographs by deep learning involve: the study of the initial state of an uncured composite laminate and the evolution of its different phases during the manufacturing process, and microstructural characterisation of complex geometry multi-axial laminates, such as corners, changes in thickness, and sandwich structures. Before reaching its full potential, new questions will need to be answered. For instance, the influence of annotation on phase segmentation, application of trained models to different materials, CT scan resolution and noise, and any consequence of polymer cross-link density increases due to curing on the CT scan data and model training.

Data Access Statement
The raw CT data underlying this article are not available by agreement with our industrial partners to protect their commercial confidentiality.

Acknowledgements
The authors would like to acknowledge the Engineering and Physical Sciences Research Council (EPSRC) for their support of this research through Investigation of Fine-Scale Flows in Composites Processing [EP/S016996/1]. A PhD studentship for P. Galvez-Hernandez was supported through the Rolls-Royce Composites University Technology Centre at the University of Bristol.

Appendix A. CT-Scan Sensitivity Study
The CT-Scan resolution used in this study was 8.24 µm/voxel, therefore, any features having a size smaller than the resolution are likely to be missed. For this reason, quantifying the impact on the total voidage assessment of those interply voids that have not been captured by the CT-Scan remains an important aspect to consider. Therefore, a sensitivity study was performed as follows:

The current CT-Scan settings estimate an initial value for the interply porosity ($P_i$), void counts ($C_i$), number of void voxels ($V_{vl}$) and the total number of voxels ($V_t$) for a given ROI:
\[ P_i = \frac{V_{vi}}{V_t} \times 100 \]  

(A.1)

The detectability has been defined as the number of voids existing in the sample that are captured by the CT-Scan. Therefore, increasing the void detectability by Y\% will also increase the void counts detected by the CT-Scan. The new number of total voids counts \( C_u \) is calculated as:

\[ C_u = C_i \times \frac{Y}{100} + C_i \]  

(A.2)

The additional voids that are captured as a result of the increase in detectability have a size of one voxel. The updated number of void voxels \( V_{vu} \) becomes:

\[ V_{vu} = V_{vi} + (C_u - C_i) \]  

(A.3)

An updated value for the interply porosity \( P_u \) is then calculated considering the new number of voids and their size:

\[ P_u = \frac{V_{vu}}{V_t} \times 100 \]  

(A.4)

Finally, the interply porosity increase is given by:

\[ \text{Interply Porosity Increase} = \frac{P_u - P_i}{P_i} \times 100 \]  

(A.5)

The increase between the updated interply porosity \( P_u \) and the initial interply porosity \( P_i \) depending on the increase of the detectability is calculated for all three ROIs (from Figure 8) and is shown in Figure 15.
Figure 15: Interply porosity increase with void detectability. Note the percentage is a relative increase and not an absolute increase in void level (i.e. a 1.16% interply porosity increase in Th1 causes a rise in void level from 0.52% to 0.53% in the material volume).

The following observations can be made from the results presented in Figure 15:

- The effect that an increase in the detectability has on the porosity estimation is more evident in low porosity volumes, where any addition in the void counts has a more pronounced impact than in high porosity volumes due to their relative contributions to the total voidage.
- The highest increase in the porosity estimation is provided by Th1 in ROI 1 for a detectability increase of 100%. This value will lead to a void counts of 40 (voxel counts = 1744) compared to the initial value of 20 (voxel counts = 1724).
- The lower porosity increase takes place in ROI 3 for the ground truth segmentation (0.007%), due to the presence of voids with a size greater than 60,000 voxels. Therefore, the contribution of the additional one-voxel voids to the total porosity remains negligible.

Overall, the effect of the interply porosity due to voids with a size below the chosen resolution is an increase of less than 1% in volumes with a ground truth voidage level of 1%. As the ground truth voidage level increases, the impact of the non-detected voids have on the total voidage becomes negligible.
References


