Composites Part B 264 (2023) 110907

Contents lists available at ScienceDirect

Composites Part B

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

Barely visible impact damage detection in composite structures using deep learning networks with varying complexities

Ali Tabatabaeian [a,*], Bruno Jerkovic [b], Philip Harrison [a], Elena Marchiori [b], Mohammad Fotouhi [a,c,**]

[a] James Watt School of Engineering, University of Glasgow, Glasgow, UK
[b] Institute of Computing and Information Sciences, Radboud University, Nijmegen, the Netherlands
[c] Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands

ARTICLE INFO

Handling Editor: Prof. Ole Thomsen

Keywords:
Structural health monitoring
Deep learning
Hybrid composite sensors
Barely visible impact damage

ABSTRACT

Visual inspection is one of the most common non-destructive testing (NDT) methods that offers a fast evaluation of surface damage in aerospace composite structures. However, it is highly dependent on human-related factors and may not detect barely visible impact damage (BVID). In this research, low velocity impact tests with different energy levels are conducted on two groups of composite panels, namely 'reference' and 'sensor-integrated' samples. Then, the results of impact tests, together with C-scan and visual inspection images, are analysed to define the BVID range and create an original image dataset. Next, four different deep learning models are trained, validated and tested to capture the BVID only from the images of the impacted and non-impacted surfaces. The results show that all four networks can learn and detect BVID quite well, and the sensor-integrated samples reduce the training time and improve the accuracy of deep learning models. ResNet outperforms other networks with the highest accuracy of 96.2% and 98.36% on the back-face of reference and sensor-integrated samples, respectively. The proposed damage recognition method can act as a fast, inexpensive and accurate structural health monitoring tool for composite structures in real-life applications.

1. Introduction

Artificial Intelligence (AI)-based techniques for detecting impact damage in composite materials can be broadly categorised into the following primary types: (i) image-based, (ii) vibration-based, and (iii) acoustic-based techniques. Image-based methods utilise computer vision algorithms to scrutinise images of the surface of composite panels both before and after an impact event. This technique detects any changes in surface topography, such as cracks and fibre breakage. Vibration-based methods, on the other hand, focus on measuring and analysing the vibrational response of the composite structure to detect any changes in mechanical properties, including stiffness and damping, caused by impact damage. Acoustic-based techniques involve using acoustic sensors to detect changes in the acoustic emission signals generated by impact damage [1–4]. Several studies have explored the application of AI-based methods for detecting impact-induced damage in polymer composite materials [5–11]. In addition to impact damage detection, AI-based techniques can be applied for damage classification, damage quantification, and remaining useful life prediction [12]. For example, Zargar and Yuan [9] demonstrated the possibility of using a Deep Learning (DL) model to characterise impact damage by analysing the evolution of impact-generated propagating waves. The DL model used there was not merely data-driven, but the physics of the wave propagation phenomenon was embedded into the model's architecture to make it a physics-inspired DL model. Fotouhi et al. [13] proposed the exploitation of DL for quantitative assessment of visual detection of both microscale damage (matrix cracking) and macroscale damage (impact and erosion damages). The damage severity was identified on images of the impacted and non-impacted sides with a validation accuracy of 96% and 87%, respectively. Their results indicated the promising performance of DL to automate visual inspection; however, it highlighted the need for an improved dataset library and customised classifiers for DL training. Wei et al. [10] used infrared thermography data of impacted curved Carbon Fibre Reinforced Polymer (CFRP) composites to train two

* Corresponding author.
** Corresponding author. James Watt School of Engineering, University of Glasgow, Glasgow, UK.
E-mail addresses: 2611578T@student.gla.ac.uk (A. Tabatabaeian), m.fotouhi-1@tudelft.nl (M. Fotouhi).

https://doi.org/10.1016/j.compositesb.2023.110907
Received 24 May 2023; Received in revised form 12 July 2023; Accepted 25 July 2023
Available online 26 July 2023

different DL models. Both models could identify impact damage and predict the damaged location with an F1-score of 92.74% on mid-wave infrared data and an F1-score of 87.39% on long-wave infrared data. Hasebe et al. [8] applied three Machine Learning (ML) models on a dataset extracted from low-velocity impact (LVI) tests on composites with particular attention to three influential factors, namely stacking sequence, impactor shape, and impact energy. Their results indicated that local volume, the gradient of the dent surface, and the pure dent depth could be used for characterising internal damage in CFRP laminates.

It is seen that in all similar research works, collecting a dataset needs advanced knowledge and measurement facilities such as Pulsed Thermography equipment [5], 3D measurement systems for evaluating impact dent depth [6,8], or signal processing to convert impact signals into input image data [11]. In addition, these methods may not be widely feasible. For example, when measuring the dent depth, the relaxation in dent depth over time is often neglected [8]. Moreover, the dent depth of impact-induced damage depends on impactor size and cannot be determined based on visual observation alone [14]. In the case of using thermal images as input data, there is a challenge of light reflection and not enough emissivity, as most composite structures have a reflective surface [15]. Accordingly, the method cannot be widely used to collect thermal image datasets of various composite materials and structures. As for further research, the authors of [8] suggested developing new Structural Health Monitoring (SHM) methods to directly use the surface profiles as features without a reduction in raw data (surface profile) to human-designed features (depth, volume, etc.). Also, they suggested studying whether AI models are effective even if the target contains paint or other features which may be found in real structures, but not in laboratory-level research. This is a source of inspiration for this research, to use only surface images as inputs to the AI system. An improvement of this work over other research in the literature is the

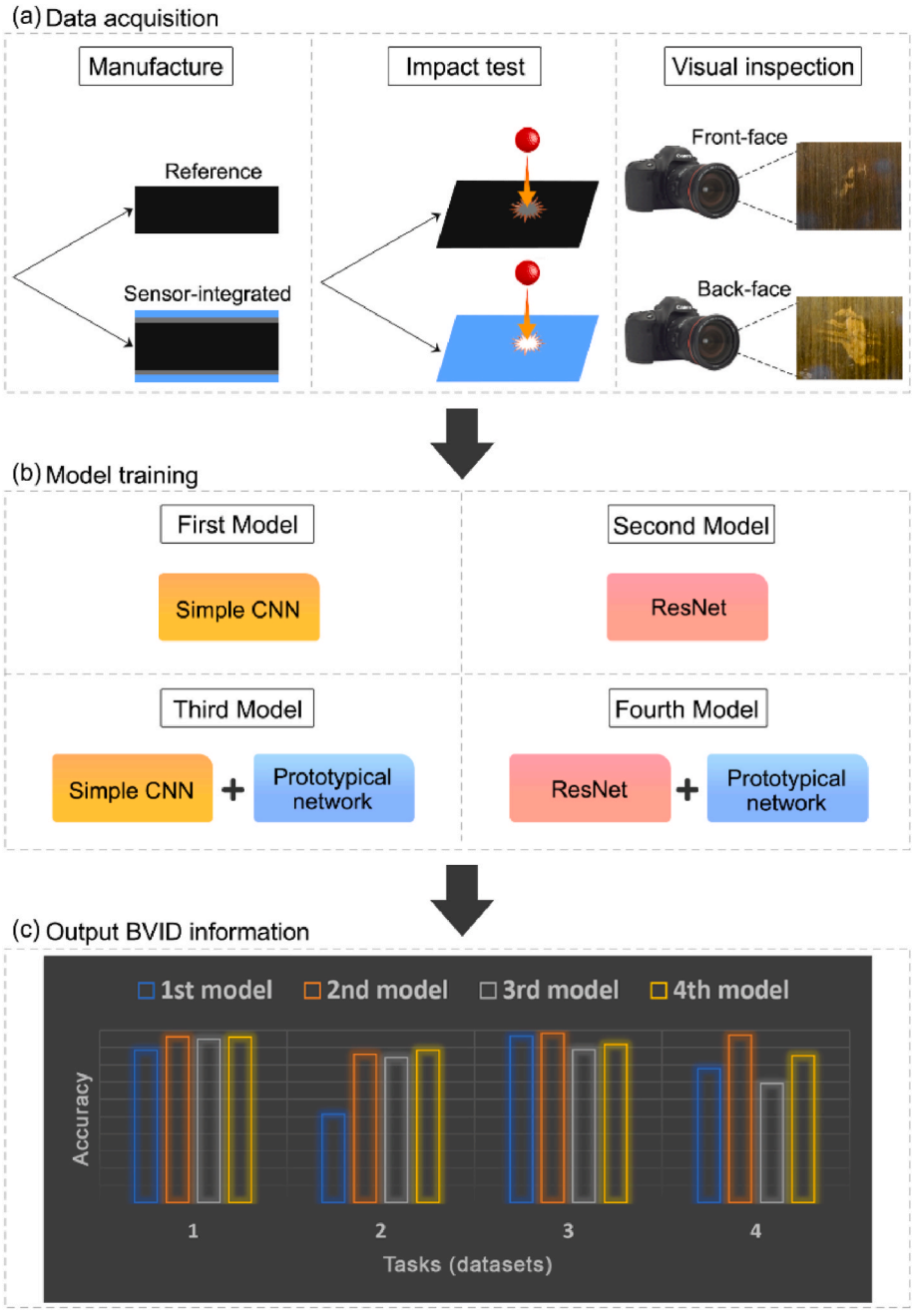

Fig. 1. The procedure followed in this study: a) data acquisition, b) model training, c) output BVID information, including accuracy, F-1 score, precision, recall and training time of each of the four DL models (shown with different colours) for four different tasks. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

successful implementation of four different DL models, from a simple two-layer Convolutional Neural Network (CNN) to an advanced CNN composed of a Residual Network (ResNet) and a Prototypical network, to predict the Barely Visible Impact Damage (BVID) only from very simple surface images. The data collection procedure in this research is very simple, and there is potential to add a remote photography system to make the SHM process fully autonomous. In addition, the idea of applying AI while adding self-reporting hybrid glass/carbon coatings [16] on the surface is explored and shows how this sensing technology can improve the autonomous recognition of BVID using DL models. A detailed description of different DL models used in this research is presented in Section 2. Fig. 1 shows the procedure followed in this research.

2. Methodology

2.1. Experiments

Two groups of samples were manufactured, including CFRP composites without a sensor (reference) and CFRP composites with a sensor (sensor-integrated). In the former, 32 plies of T800 carbon prepregs were laid up with the [45/0/90/-45]4s configuration. In the latter, a sensor composed of a layer of YS-90 carbon and a layer of S-glass (both with ply degree of 90°) was added on each side of the panel, where the 32-ply substrate and sensors on two sides were co-cured according to supplier's instructions (Fig. 2) [17].

Quasi-static indentation tests were first used to understand the behaviour of the investigated samples and to choose appropriate energy levels for the impact tests, as described in a previous paper [17]. This is because quasi-static indentation and LVI tests can provide comparable results according to several research in the literature [18–20]. LVI tests with different energy levels, from 3J to 128J, were then carried out using an Instron Dynatup 9250 HV drop weight impact tower and according to the ASTM D7136 standard (see Fig. 3(a)). A single accelerometer inside the tup measured the impact load and deflection, and the measured data was processed by a 4 kHz filter of the console software to reduce the noise and oscillations. Fig. 3(b) shows different types of impact damage with respect to the impact energy level. At the first stage, the dent size is so small, and damage happens in the form of matrix cracks. As impact energy increases, damage appears in the form of delamination, and dent size becomes larger. In the second stage, all three damage types, matrix cracks, delamination, and fibre breakage, can occur, and visual inspection becomes easier due to a larger dent size (dent depth and diameter). The fibre breakage in this stage can help achieve better inspection results. Nevertheless, this can have a detrimental effect on the residual strength after impact, which explains the complexity of the interaction of different damage types during an impact event. In the third stage, the damage is visible to the naked eye due to perforation. Therefore, the second stage, which deals with BVID and the interaction of different damage types, is of great interest and is the focus of this paper.

After completing LVI tests, two Non-Destructive Evaluation (NDE) methods, including C-scan and visual inspection were conducted. In the first approach, a10 MHz transducer was used to scan the samples in a water tank. The damage area was measured in the software and the results were then recorded [17]. In the second approach, a Nikon D5600 camera was used to take images of the front-face and back-face of all samples. Multiple images of each sample with viewing angle of 0 ° and viewing distance of 40 cm were taken. Effort was made to set a lighting level which is intense enough to achieve quality images but avoid the reflection as much as possible. Accordingly, a detailed internal (C-scan) and surface (visual inspection) damage dataset was collected.

2.2. An overview of AI algorithms

The field of AI, especifically ML and computer vision, offers a way to detect defects in materials in seconds [2,22]. Regardless of the task at hand, certain parts of a ML algorithm's pipeline are almost always prevalent. The general idea is that data-driven algorithms are not completely hard-coded by the programmer, but rather learn from the data itself. In this way, ML algorithms can be tailored to specific tasks. A ML algorithm is defined by a hypothesis, a loss function and an optimisation procedure. A hypothesis is a function that receives data as input and outputs a prediction. It is parametrised by its trainable parameters and hyperparameters. The data automatically determines the former, while the ML engineer pre-sets the latter. The loss function is a function that outputs the error made by the model while using the output of the hypothesis as input. In the case of classifying if an image contains damage or not, since it is a binary classification, the common approach is to use a Binary Cross Entropy (BCE) loss function. This measures the difference between the classification predicted by the model's hypothesis and the true classification of the datapoint. Another popular loss function for classifying is a Categorical Cross Entropy (CCE). This is a more generic form of the BCE, meaning that a datapoint can be classified into more than only classes for damaged and undamaged [23,24]. An optimisation procedure is an algorithm that defines how the trainable parameters of the ML model are iteratively changed to improve its accuracy, which is achieved by minimising the error of the model's prediction tasks. These errors are defined by the loss function. When it comes to the construction of the dataset, the data is split into three parts: the training set, the validation set and the testing set. The training set is used to determine the trainable parameters of the model via an optimisation process. The validation set determines the model's hyperparameters by trying out different values. Finally, the testing set estimates the model's performance on unseen data.

ML can be used to automate damage detection in composite materials. However, more complicated ML architectures are required when the dataset is a collection of images. A ML architecture that mimics the workings of the human brain is known as an Artificial Neural Network (ANN). The simplest neural network is a Multi-Layer Perceptron (MLP). It contains multiple functions where the output of a function is used as an input to the next one. These functions are called the layers of the model. The simplest layers of a MLP are Fully Connected (FC) layers [25]. When using a neural network with multiple layers as an ML hypothesis, it falls into the realm of DL. Rautela et al. [26] applied two ML and one DL algorithms in three benchmark experimental datasets to develop a SHM tool for delamination prediction in composite structures. It was clarified that the DL approach could generate better reconstructions with lower mean squared error and provide higher accuracy on all the datasets compared to ML models. When dealing with an image-based dataset, a more complex family of neural networks, such as Convolutional Neural Networks (CNNs) might be needed [3,4]. CNNs differ from other neural network architectures because of their convolutional layers [27]. There might be some layers of a CNN which do not have any trainable parameters. Prime examples of this are max pooling and average pooling layers [28,29].

ResNet is a deep neural network architecture that was introduced in 2015 to address the problem of vanishing gradients in very deep CNNs [30]. ResNet allows for the training of much deeper networks with hundreds of layers by introducing shortcut connections between the layers, which skip over certain layers and allow the network to learn residual functions. The residual connections enable the network to better propagate gradients through the entire network, leading to improved performance and faster training time than simpler CNNs [5]. Another advantage of using a well-known architecture, such as ResNet, is that it is possible to start the training with the values of its weights learned from a different task or dataset. For example, ImageNet [31] is a large-scale image database containing over 14 million images, widely used for training and evaluating computer vision models. In this paper, ResNet weights trained on the ImageNet dataset were used as the initial starting values before the network was trained on the dataset. This approach is known as transfer learning [32]. Another subfield of ML is few-shot learning that deals with the problem of learning from a limited

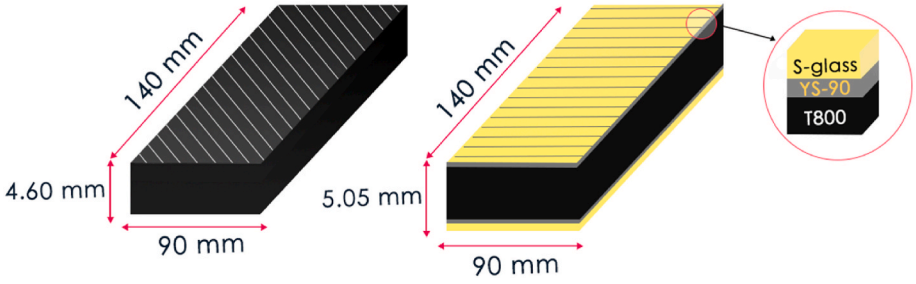

Fig. 2. Schematic of manufactured composite panels: reference samples (left) and sensor-integrated samples (right).

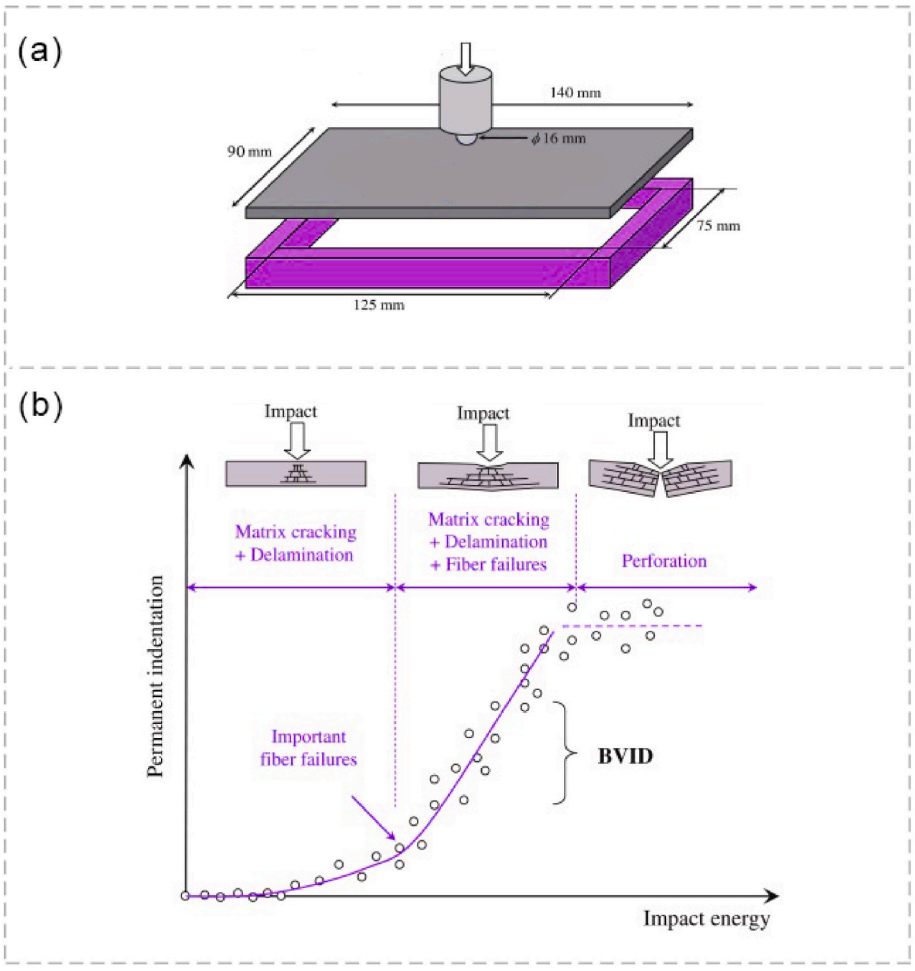

Fig. 3. (a) Schematic of impact test setup, (b) different impact damage stages with respect to the impact energy and permanent indentation size [21].

amount of labelled data [33]. In traditional supervised DL, a large amount of labelled data is required to train a model that can accurately classify new (unseen) data. However, in few-shot learning, the goal is to learn from a small number of labelled examples. Few-shot learning can be combined with metric learning to learn from small datasets. Metric learning is another subfield of ML that focuses on learning a vector representation of an image, known as embedding, in the feature space [34]. The few-shot learning and metric learning can be combined in Prototypical networks. Prototypical networks are a type of neural network architecture commonly used in few-shot learning tasks. They are designed to learn a feature space in which examples from the same class are closer to each other than to examples from other classes. The dataset splitting here is more complex than in standard supervised ML. Firstly, the dataset is split into meta-train and meta-test sets. Then, each of these sets is divided into shot-images and query-images. This approach is effective in few-shot learning tasks, where only a few labelled examples are available for each class [35].

A binary classification contains four entries: True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). In the case of damage detection, these terms can be defined as follows:

TP: The number of images that the network has correctly identified as containing damage.

TN: The number of correctly identified images without damage.

FP: The number of images without damage that are incorrectly identified as containing damage.

FN: The number of images containing damage that are incorrectly identified as images without damage.

In addition to the confusion matrix (see Table 1), other common metrics to evaluate the model's performance are precision, recall, F-1 score, and accuracy. Also, measuring a model's training time can be

Table 1
Confusion matrix.

Truth \ Predicted	Defect	No defect
Defect	TP	FN
No defect	FP	TN

important. The equations below show the definition of the evaluation metrics.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$F_1 = 2\frac{precision * recall}{precision + recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{4}$$

In this research, four different supervised networks are used as follows:

- Model one is a simple two-layer CNN
- Model two is a ResNet with 18 trainable layers pretrained on ImageNet dataset
- Model three is a Prototypical network, a more complex CNN that can use other simpler CNNs as a base. In the third model, the simple two-layer CNN from the first approach was used as a base for the Prototypical network.
- In the fourth model, the ResNet from the second approach was used as a base for the Prototypical network (pretrained on ImageNet dataset).

Also, to pre-process the data, instead of using raw pixels, the mean and standard deviation of each of the three RGB channels[1] of an input image were calculated and the raw image data were standardised before starting any of these pipelines. Note that the mean and the standard deviation for standardisation were only calculated on the training set, and this information was saved and fixed for test data. The reason to choose these four networks is that the first model is a simple CNN that is commonly used for image-based dataset. Then, ResNet-18 is used as the second model. This is a powerful deep neural network that has shown a promising performance in the field of computer vision and image recognition and can presumably capture complex data [1]. Given the small size of dataset in this research, a pretrained version of ResNet is used. Another type of DL models which work well with small dataset is a Prototypical network. This class of DL models uses few-shot learning, enabling an efficient generalisation of unseen data based on only a small number of samples. Therefore, it is used in the third and fourth models of this study. All models have been implemented in Python programming language (version 3.9) using a popular library, PyTorch, for implementation and experimentation with artificial neural networks.

2.2.1. Description of the first supervised network

The first network used a simple CNN with only two convolutional layers. Both convolutional layers output 64 feature maps, have 3x3 kernels, and are followed by: batch normalisation[2] layers, Rectified Linear Unit (ReLU) activation functions [36], and max pooling layers with a 3x3 kernel. The output of the CNN is flattened and used as an input to the MLP with FL layers. MLP also consists of two layers, where the first layer outputs a 512-dimensional representation, and the second layer takes that and outputs a one-dimensional representation of the image. This value is then passed on to the Sigmoid layer to get a value between 0 and 1, which classifies an image in one of the two classes. Finally, the outputted value is rounded up to either 0 or 1 to receive the final classification of the network (see Fig. 4). Furthermore, Adam is used as an optimiser [5,37], and BCE function is used as a loss function.

2.2.2. Description of the second supervised network

The second network used a pre-trained ResNet architecture with 18 trainable layers. Even though there are available pre-trained ResNet architectures with more layers, because of the small size of dataset in this paper, a bigger architecture was not used to avoid overfitting the data. ResNet 18 (available from Pytorch [38]) was pre-trained on more than a million images from the ImageNet database. However, its last FC layer was replaced with a new one that outputs only one number. After that, a Sigmoid activation function was added to turn it between 0 and 1, so that it could be used as a probability of the input image either being damaged or not. Furthermore, Adam was used as an optimiser, and the BCE as a loss function.

2.2.3. Description of the third and fourth supervised networks

The third and fourth models used a Prototypical network, which works very differently from the first two approaches. During the meta-training phase, a random batch is sampled from a subset of meta-training images. The meta-training subset contains 30% of the total number of images in the dataset. The base network used in the Prototypical networks outputs a 32-dimensional embedding. Any desirable network can be used as a base network to embed the image into this latent space with 32-dimensional representations. The 32-dimensional output is not a representation of a binary classification label, but it captures the semantical representation of an image in a high-dimensional latent space (which is used for binary classification later in the pipeline). To that note, 50% of the embeddings are used for training (shot-images) and 50% for testing (query-images). Since this is a supervised approach, the labels of the shot-images are known, and they are used to calculate the mean for each label, resulting in two 32-dimensional embeddings, one for each label (labels are 'damaged' and 'undamaged'). These are known as the prototypes of labels. Then, the distances from each query-image to each prototype are calculated. The query image takes the label of its closest prototype (in this latent space), which is how prototypical networks perform classification. The Euclidian distance is used to calculate the means, similar to Ref. [35]. Prototypical networks allow the use of any desirable CNN to encode the image into the 32-dimensional latent space. However, this network does not output the classification prediction of the image, but instead, its 32-dimensional embedding in the latent metric space, where the distances are calculated. This approach was used on the networks that had been used in the first and second models. They both have the same setup as the first two approaches, but this time instead of outputting a one-dimensional number, they output a 32-dimensional vector.

3. Results and discussion

3.1. LVI and NDE

The results of LVI tests are shown in Fig. 5. As seen in Fig. 5(a), except for 3J and 6J, a significant load drop can be seen in all energy levels associated with delamination initiation. Low energy levels (3J and 6J) do not cause any damage to the samples; therefore, there is no load drop for these energies. A considerable load drop occurred in 96J and 128J due to fibre failure. In these impact energies, there is a significant residual deflection. Energy absorption in composite structures under impact can cause various damage mechanisms. Therefore, damage mechanisms can be characterised by analysing energy absorption. As shown in Fig. 5(b), there is a significant change in the absorbed energy-

[1] Red-Green-Blue (RGB) channels. That is how images are stored in the computer.
[2] Batch Normalisation is written as (BN) in Fig. 4.

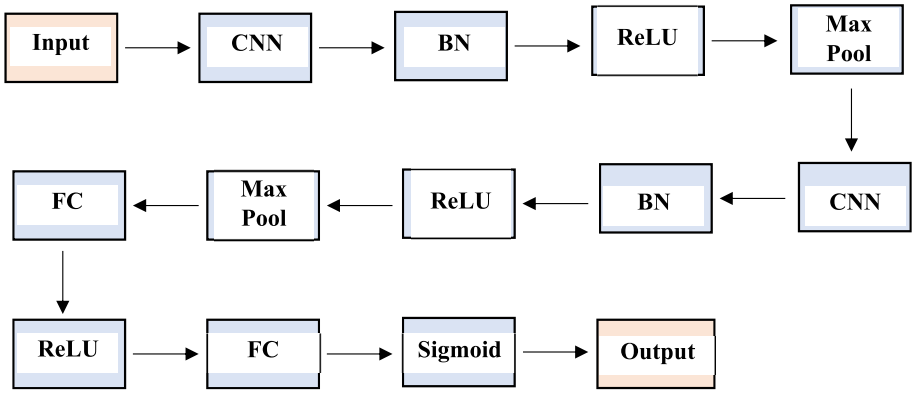

Fig. 4. A simple flowchart of the first network's architecture.

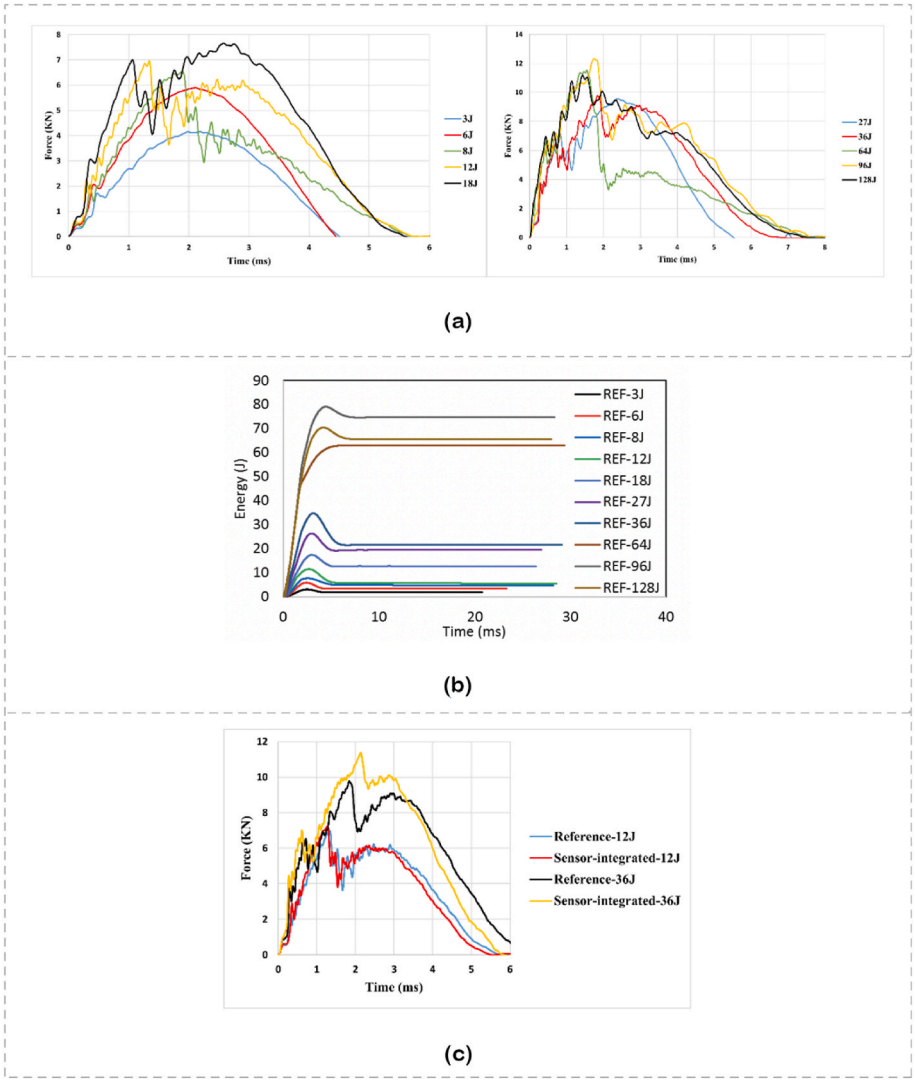

Fig. 5. Low velocity impact test results: a) Force-time response of reference samples (without a sensor) under different impact energies from 3J to 128J, b) Absorbed energy-time response of reference samples (without a sensor) under different impact energies from 3J to 128J, c) comparison of the impact response in reference and sensor-integrated samples at impact energies of 12J and 36J

time response of specimens, where energy absorption for impact energies lower than 64J is almost half of those higher than 64J. This is in line with the results of Fig. 5(a), suggesting that a specific range of energy from 8J to 64J is of great importance as the damage mechanism in this energy range is delamination, which causes BVID. Fig. 5(c) shows a comparison of impact response in reference and sensor-integrated samples for two impact energies within the BVID range (12J and 36J). The results suggest that adding the hybrid glass/carbon sensor does not significantly change the mechanical properties (impact response), as in both 12J and 36J, the same trend is followed for reference and sensor-

integrated samples.

Fig. 6 shows the C-scan, front-face, and back-face images of the reference samples. The C-scan images show no internal damage associated with 3J and 6J impact energies, and delamination starts from 8J. It is also indicated that damage size varies in line with impact energy. Images of the front-face and back-face of samples clarify that damage is visible only at the energy levels of 96J and 128J, barely visible at 12J–64J, and invisible at 8J. These images suggest that the back-face may allow for better damage visibility than the front-face, particularly for higher energy levels such as 64J, 96J, and 128J. Fig. 7 demonstrates C-scan, front-face, and back-face images of sensor-integrated samples. Similar to the reference samples, there is no damage in 3J and 6J impact energies. However, a significant improvement in visual inspection of damage is achieved when adding sensors. The hybrid glass/carbon sensor allows the visually detecting of the damage caused by low energies such as 8J. A comparison of Figs. 6 and 7 suggests that this sensing technology can work well and improve visual inspection, particularly for BVID, ranging from 8J to 64J. More detail about these smart hybrid composite sensors can be found in Ref. [16].

BVID can be defined based on different metrics or standards. For example, based on the Airbus damage definition, "BVID is the minimum impact damage surely detectable by scheduled inspection. Dent depth is the damage metric for transverse impact. For an edge impact, where internal cracks and delamination become visible, the damage metric is the dent depth and/or the crack length" [39]. Boeing defines BVID as "small damages which may not be found during heavy maintenance general visual inspections using typical lighting conditions from a distance of five feet (1.524 m). The damage metric is typically a dent depth of 0.01–0.02 inches (0.254 mm–0.508 mm). Dent depth relaxation must be accounted for" [40]. According to general guidelines, permanent indentations between 0.3 mm and 0.5 mm can indicate BVID, whereas permanent indentations of 2 mm or perforations of 20 mm indicate minor Visible Impact Damage (VID) [21]. In this research, after conducting quasi-static indentation tests, impact tests were conducted at different energies, ranging from 3J to 128J, to generate a complete dataset of different impact-induced damage modes such as matrix cracks, delamination, fibre failure and perforation. The C-scan results revealed that the internal damage happens from 8J. However, visual inspection results suggested that damage (permanent indentation) is (barely) visible from 12J. Moreover, it was observed that impact energies higher than 64J can cause fibre failure, which was beyond the BVID (the damage was not an indentation but a significant fibre fracture). This was also confirmed by load-displacement curves obtained from the impact tests. Therefore, this range (12J–64J) was considered as the BVID range and images of samples impacted at 12J, 18J, 27J, 36J, and 64J were used used as an input for training the models.

Table 2 represents the classification of the dataset. After experimental investigations, it was concluded that the best way to train and test the CNNs is to classify the dataset into four main groups (or tasks), as shown in Table 2. In task 1, the model is trained and tested to recognise BVID using images of the back-face of reference samples. In task 2, the model uses images of the front-face of reference samples. Similarly, tasks 3 and 4 deal with images of the back-face and front-face of sensor-integrated samples, respectively. Therefore, a model can output one of the two possible classes (damaged or undamaged) in each task.

3.2. First DL model (ConvNet)

The results of the first model are presented in Tables 3–5 and Fig. 8. Note that evaluation parameters presented in these tables were introduced in Section 2.2. All evaluation parameters show that it is easier for the model to detect the damage on back-face than front-face, regardless of whether using a sensor or not. For example, the accuracy of damage prediction on back-face is 37% higher than on the front-face in reference samples. This could be due to different visible or barely visible damage patterns developed during the impact event on the two sides. This difference in sensor-integrated samples is only 19%, proving the effectiveness of adding a sensor in generating recognisable damage patterns on both sides. The beneficial effect of adding a sensor can also be seen through comparing the accuracy in back-faces of reference and sensor-integrated samples (tasks 1 and 3), and front-faces, similarly (tasks 2 and 4), where there has been an increase in damage prediction accuracy of 8% and 26% on back-face and front-face, respectively. Here, the higher improvement percentage achieved on the front-face is because when there is no sensor (reference samples), and the damage is within the BVID range, the damage pattern on the front-face is very difficult to detect. That explains the lowest accuracy of 51.25% (task 2) among all four different tasks. However, adding a sensor generates a significantly easier-to-detect damage pattern. The highest accuracy is achieved in task 3 (96.72%), where the damage is studied on the back-face of sensor-integrated samples. Note that here, only images of samples damaged within the BVID energy range are used, and different results may be achieved if images of higher impact energies are used as input to the model. Table 5 suggests that the network's training time for the two sides of the same samples is almost the same, while it is significantly faster in sensor-integrated ones. Therefore, adding a sensor can improve the damage detection process in terms of accuracy and time, when using this network.

3.3. Second DL model (ResNet)

Tables 6–8 and Fig. 9 show the results of the second model. Similar to the first model, all metrics confirm that the model can learn and predict damage on the back-face easier than front-face. However, the difference between the two sides is less significant than that of the first model. For example, the difference between the accuracy of the front-face and back-

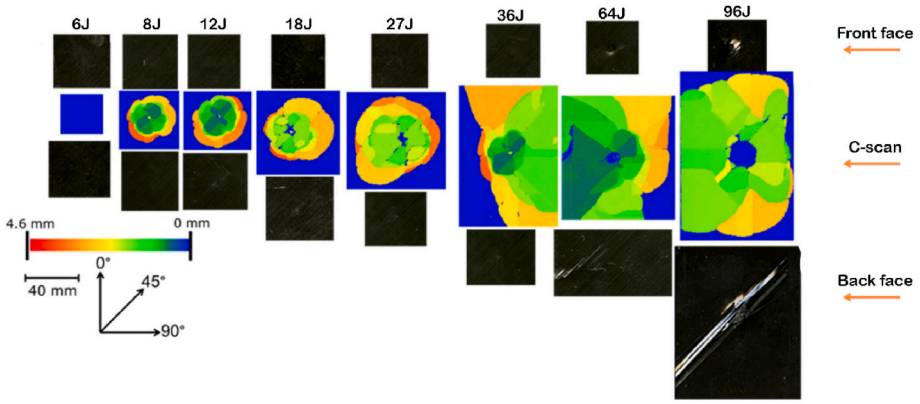

Fig. 6. Front-face, back-face, and C-scan images of reference samples at different impact energies [17].

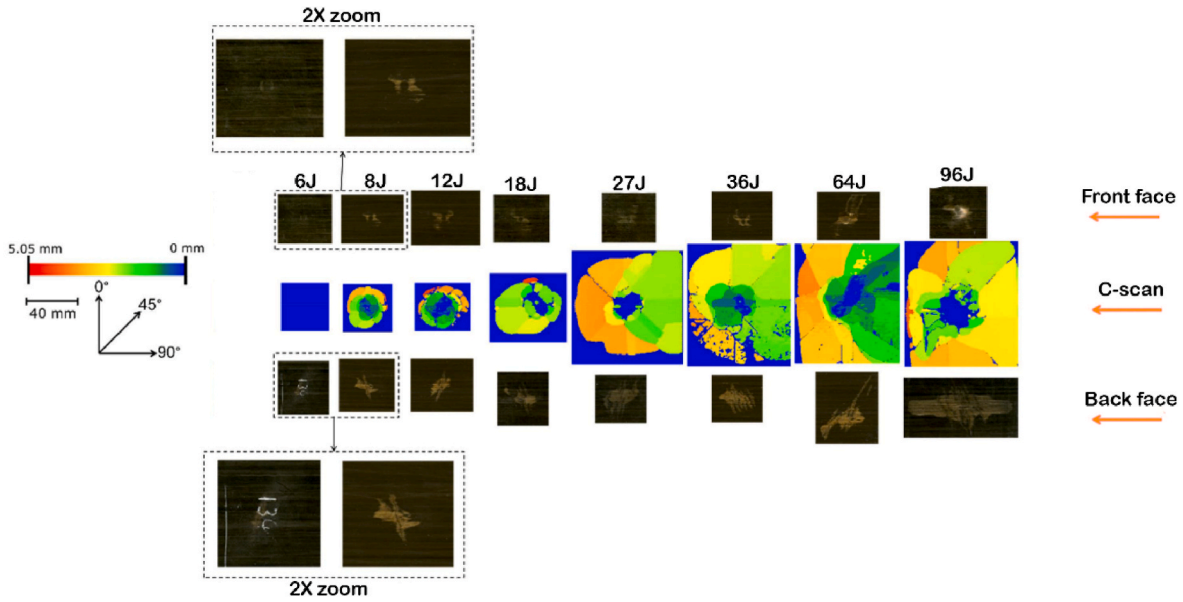

Fig. 7. Front-face, back-face, and C-scan images of sensor-integrated samples at different impact energies [17].

Table 2
Classification of the dataset.

Task number	Type of composite panel	Damage location	Damage range
1	Reference	Back-face	Within the BVID range
2	Reference	Front-face	Within the BVID range
3	Sensor-integrated	Back-face	Within the BVID range
4	Sensor-integrated	Front-face	Within the BVID range

Table 3
Evaluation metrics for four different tasks (1st model).

NAME	Values per dataset			
Task	1	2	3	4
Accuracy	88.46%	51.25%	96.72%	77.94%
Precision	90.90%	78.26%	100%	86.84%
Recall	88.89%	55.38%	93.93%	76.74%
F-1	89.88%	64.86%	96.87%	81.48%

face in reference and sensor-integrated samples is only 10% (tasks 1 and 2) and 1% (tasks 3 and 4), respectively. This model has the highest and lowest accuracy of 98.36% (task 3) and 86.25% (task 2). This aligns with the visual observation results, as damage on the back-face of sensor-integrated samples is the most visible, and on the front-face of reference samples is the least visible. The results also suggest that adding a sensor can improve this model's accuracy, which is more considerable on the front-face (comparing tasks 2 and 4). The results presented in Table 8 indicate that the difference between the training time for the back-face and front-face of reference samples is considerably higher than that of sensor-integrated ones. Also, overall, the training time for both sides of the sensor-integrated samples is lower than for reference samples, proving the effectiveness of adding a sensor on the second model's performance.

3.4. Third DL model (ProtoNet + convnet)

The results of the third model are presented in Tables 9–11 and Fig. 10. It is seen that when comparing the evaluation metrics values of

Table 4
Confusion matrix for four different tasks (1st model).

Task: 1		
Predicted \Truth	Defect	No defect
Defect	39	5
No defect	4	30
Task: 2		
Predicted \Truth	Defect	No defect
Defect	36	29
No defect	10	5
Task: 3		
Predicted \Truth	Defect	No defect
Defect	31	2
No defect	0	28
Task: 4		
Predicted \Truth	Defect	No defect
Defect	33	10
No defect	5	20

Table 5
Average training time of each task (1st model).

Task	1	2	3	4
Average Train Time	54 m 12s	40 m 55s	16 m 56s	13 m 18s

the front-face and back-face, the same trend as the two previous models is followed here, meaning that the model can predict the damage on back-face better than the front-face. Overall, the network has a better performance compared to the first model, which suggests that using the first model as a base for a Prototypical network, for this analysis, is a good choice. This is more notable when comparing the training time of the first and third models, where the third network offers a higher accuracy (or other evaluation metrics), while requiring the same training time. This model has the best and worst prediction performance in task 1 and task 4, respectively. An interesting result here is that the model's performance deteriorates when adding a sensor, as opposed to the two previous models. This might be explained by the different working principles of a Prototypical network compared to a simpler CNN. Adding

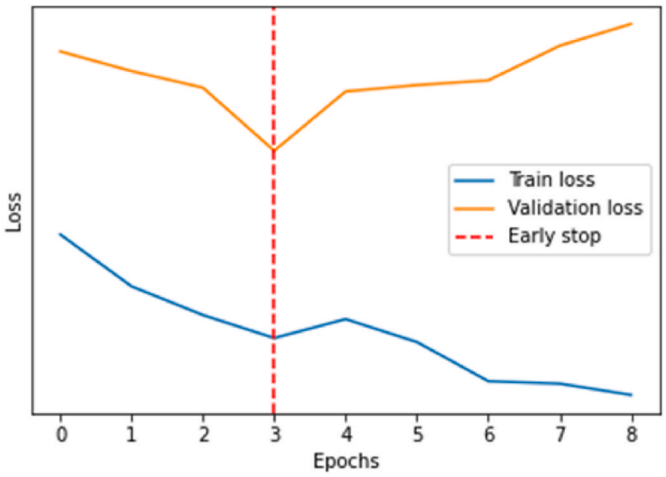

Fig. 8. Example of train and validation losses of the 1st model (task 1).

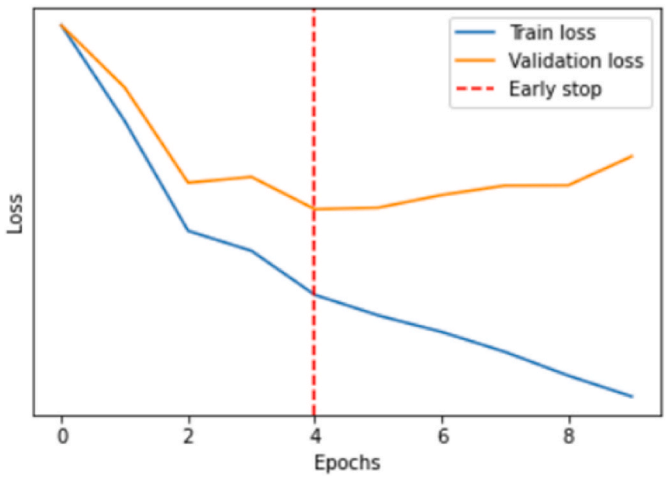

Fig. 9. Example of train and validation losses of the 2nd model (task 1).

Table 6
Evaluation metrics for four different tasks (2nd model).

NAME	Values per dataset			
Tasks	1	2	3	4
Accuracy	96.2%	86.25%	98.36%	97.05%
Precision	100%	91.3%	100%	97.37%
Recall	93.62%	85.71%	96.88%	97.37%
F-1	96.7%	88.42%	98.42%	97.37%

Table 7
Confusion matrix for four different tasks (2nd model).

Task: 1		
Predicted \ Truth	Defect	No defect
Defect	44	3
No defect	0	31
Task: 2		
Predicted \ Truth	Defect	No defect
Defect	42	7
No defect	4	27
Task: 3		
Predicted \ Truth	Defect	No defect
Defect	31	1
No defect	0	29
Task: 4		
Predicted \ Truth	Defect	No defect
Defect	37	1
No defect	1	29

Table 8
Average training time of each task (2nd model).

Task	1	2	3	4
Train Time	1h 2 m 2s	42 m 7s	17 m 10s	13 m 5s

Table 9
Evaluation metrics for four different tasks (3rd model).

NAME	Values per dataset			
Task	1	2	3	4
Accuracy	94.87%	84%	88.52%	69.12%
Precision	100%	81.25%	93.55%	78.95%
Recall	91.67%	72.22%	85.29%	69.77%
F-1	95.65%	76.47%	89.22%	74.11%

Table 10
Confusion matrix for four different tasks (3rd model).

Task: 1		
Predicted \ Truth	Defect	No defect
Defect	44	4
No defect	0	30
Task: 2		
Predicted \ Truth	Defect	No defect
Defect	13	5
No defect	3	29
Task: 3		
Predicted \ Truth	Defect	No defect
Defect	29	5
No defect	2	25
Task: 4		
Predicted \ Truth	Defect	No defect
Defect	30	13
No defect	8	17

Table 11
Average training time of each task (3rd model).

Task	1	2	3	4
Train Time	53 m 32s	36 m 20s	20 m 54s	16 m 3s

a sensor can improve the visibility of damage by causing a clear colour change and a nice damage pattern. However, when adding a sensor, the difference between the damage patterns of a sample impacted at 12J and 64J is much higher than those of 12J and 64J in the case of a reference sample (without any sensor). Given that the dataset includes images of all samples damaged within the BVID range, in sensor-integrated samples, a more different damage pattern between various images can adversely influence the network's performance, thus, achieving a slightly lower accuracy, precision, or recall. Even though adding the sensor makes the damage on the images easier to predict, this shows how

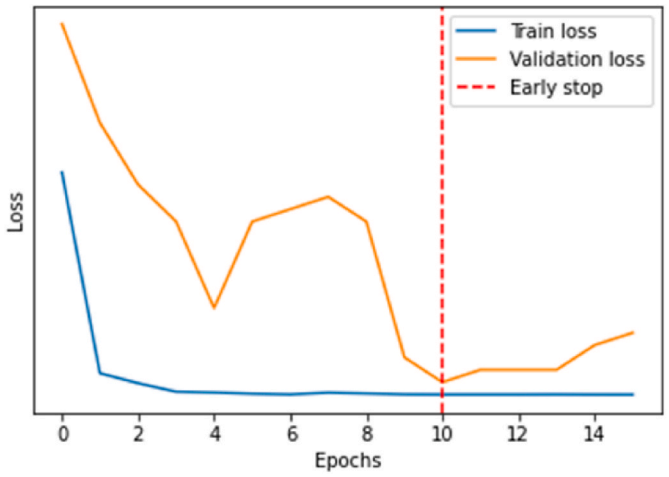

Fig. 10. Example of train and validation losses of the 3rd model (task 1).

AI works differently from the human brain, and sometimes a task that is easier for humans is harder for an AI. It should be noted that the sensor can still be beneficial when looking at the training time of different tasks (Table 11). Therefore, using the third model for the sensor-integrated dataset may still be advantageous, particularly if images of only specific impact energy are used as input.

3.5. Fourth DL model (ProtoNet + ResNet)

This section presents the results of the fourth model in Tables 12–14 and Fig. 11. Like all previous models, here, detection and prediction of damage on the back-face is a more straightforward task than the front-face. This model shows the best and worst performance in task 1 and task 4, respectively. As this model employs a Prototypical network in its architecture, its performance is more similar to the third model, than the first two. This is particularly true when comparing the results of reference and sensor-integrated samples (tasks 1&3 and 2&4), where a slight accuracy reduction can be seen in the latter. The reason for this poorer performance was discussed in the previous section. However, here, the performance deterioration due to adding a sensor is less considerable compared to the third model, suggesting that the ResNet can act better as a base for the Prototypical network compared to the first simpler CNN. With an exception for task 4, the results of the training time (Table 14) also indicate the efficiency of the fourth model for different damage prediction tasks. Given the complex architecture of this model, it might be more beneficial when dealing with more challenging damage detection tasks, for example, composite structures without a sensor with lower impact damage energies.

3.6. Comparison and discussion

Herein, two important evaluation metrics, accuracy (Fig. 12) and training time (Table 15), are chosen to compare and analyse the results of the four models. The reason for selecting accuracy is that it considers both TF and TN (all correct predictions) when considering the whole dataset (not just a part). Therefore, in the case of damage detection, this

Table 12
Evaluation metrics for four different tasks (4th model).

NAME	Values per dataset			
Task	1	2	3	4
Accuracy	96.15%	88.24%	91.8%	85.29%
Precision	100%	94.12%	96.77%	92.1%
Recall	93.6%	84.2%	88.24%	83.3%
F-1	96.69%	88.88%	92.3%	87.48%

Table 13
Confusion matrix for four different tasks (4th model).

Task: 1		
Predicted \ Truth	Defect	No defect
Defect	44	3
No defect	0	31
Task: 2		
Predicted \ Truth	Defect	No defect
Defect	32	6
No defect	2	28
Task: 3		
Predicted \ Truth	Defect	No defect
Defect	30	4
No defect	1	26
Task: 4		
Predicted \ Truth	Defect	No defect
Defect	35	7
No defect	3	23

Table 14
Average training time of each task (4th model).

Task	1	2	3	4
Train Time	55 m 33s	37 m 59s	23 m 2s	39 m 45s

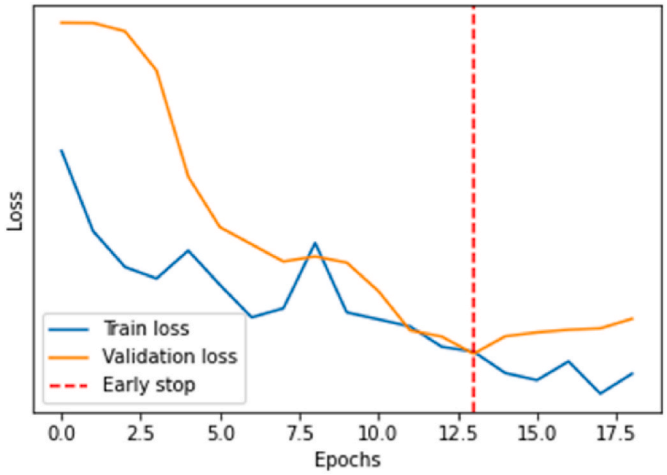

Fig. 11. Example of train and validation losses of the 4th model (task 1).

criterion may allow for a better understanding of the system's performance. Moreover, training time can be an important parameter, as in some real-life or controlled laboratory monitoring scenarios, the structure's health should be assessed rapidly. For example, short-term monitoring might take only a few hours [41]. Furthermore, it may be necessary to apply the NDE part by part in very large composite structures through different steps, highlighting the necessity of developing a fast damage recognition tool. Accordingly, in real-world applications, longer training times can make it harder to deploy the model since it can impact its scalability and practicality. Therefore, finding the optimal training time that balances accuracy and efficiency is crucial in developing effective DL models.

A comparison of all results suggests that the first convolutional

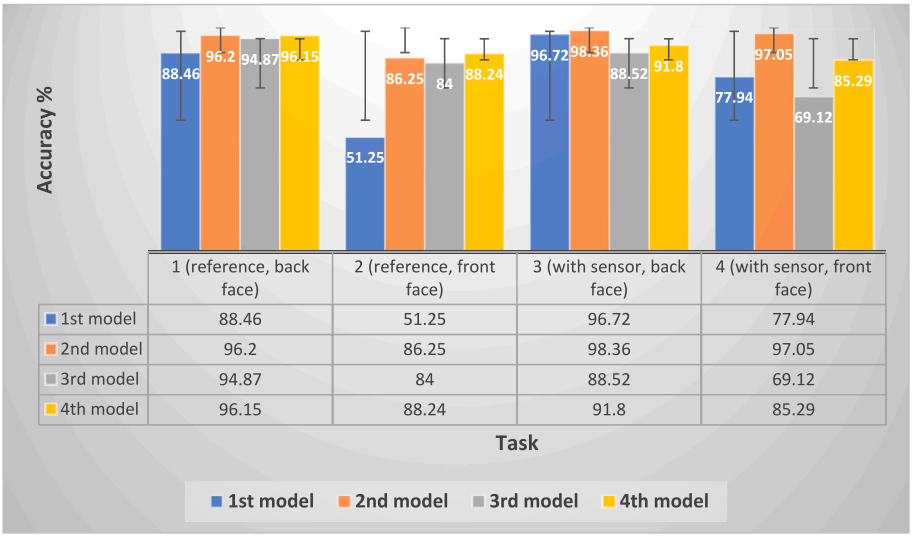

Fig. 12. Comparison of the accuracy for different tasks in all four models.

Table 15
Comparison of the training time for different tasks in all four models.

Model \ Task Number	1 (reference, back face)	2 (reference, front face)	3 (with sensor, back face)	4 (with sensor, front face)
1st model	54 m 12s	40 m 55s	16 m 56s	13 m 18s
2nd model	62 m 2s	42 m 7s	17 m 10s	13 m 5s
3rd model	53 m 32s	36 m 20s	20 m 54s	16 m 3s
4th model	55 m 33s	37 m 59s	23 m 2s	39 m 35s

network achieves the lowest score while the ResNet or combining a Prototypical network with a ResNet achieves the highest scores. This is as expected because the first convolutional network is the simplest one, whereas the fourth one was adapted for the case of a 'few-shot' classification. Prototypical networks are a good choice for a small dataset and a simple task (such as a binary classification on this dataset). Comparison of the outcomes of the first and third models, as well as those of the second and fourth models, shows a great improvement in damage recognition accuracy both on front-face and back-face due to applying a Prototypical network. It is seen that both in the second model (when ResNet is used as a classification model) and in the fourth model (when ResNet is used as a base model for the Prototypical network), the ResNet performs better than a simpler convolutional network. This is because of the better architecture of the ResNet (by using the residual connections) and the fact that it has been pre-trained. Using pre-trained models that have been trained on other datasets can improve their performance because the network does not start learning from scratch. Instead, it starts learning with the set of parameters where the last learning converged. Also, a deeper network such as ResNet offers a higher capacity to understand and analyse the image.

Overall, all models perform quite well on the third and fourth

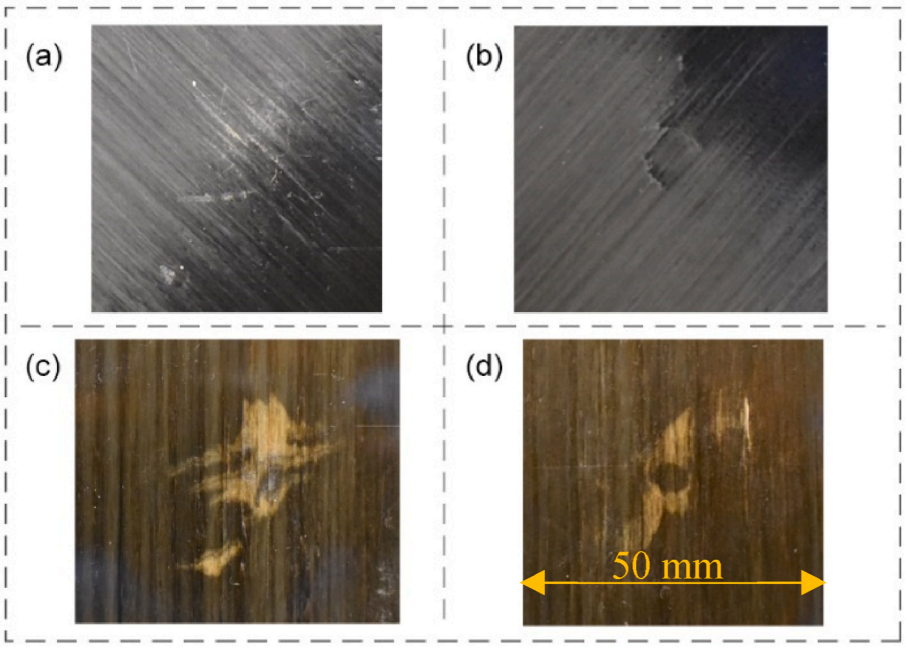

Fig. 13. Damage patterns on different datasets (tasks): a) back-face of reference samples (task 1), b) front-face of reference samples (task 2), c) back-face of sensor-integrated samples (task 3), d) front-face of sensor-integrated samples (task 4).

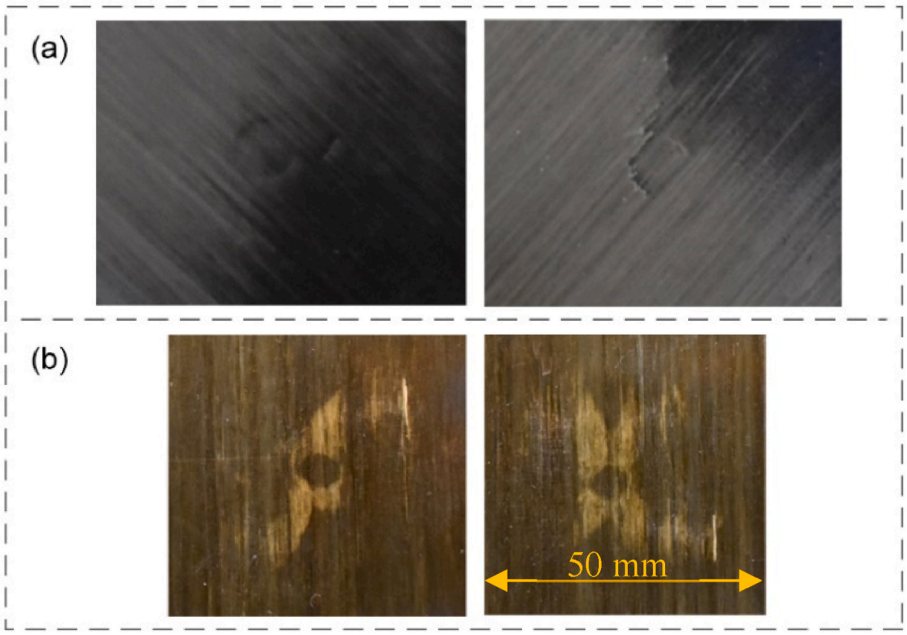

Fig. 14. Graphical explanation of a lower accuracy in the third and fourth models (Prototypical networks) when adding a sensor. Impact damage on (a) front-face of reference samples impacted at 36J (left) and 64J (right), and (b) front-face of sensor-integrated samples impacted at 18J (left) and 27J (right). The visual difference between the damage pattern of reference samples impacted at two different energies is significantly less identical than that of sensor-integrated samples. This is true, even though a bigger energy gap exists in (a) than in (b). Given the strong pattern recognition of Prototypical networks, more different damage patterns in sensor-integrated samples impacted at various energies (within the BVID range) may adversely influence the network's performance.

datasets. This is due to adding a hybrid glass/carbon sensor on the back-face and front-face, achieving images with higher contrast between the damaged and undamaged areas (the damaged area is brighter, and the background is darker) (see Fig. 13). The damage recognition accuracy of the first model on the back-face and front-face increases by 8.26% and 26.74% when using a hybrid glass/carbon sensor. Similarly, adding a sensor improves the accuracy of the second model by 10.8% and 2.16% on the front-face and back-face, respectively. In both cases, the front-face benefits more from adding a sensor. A more complex nature of Prototypical networks in the third and fourth models may capture very detailed features of damage patterns on each side of the sensor-integrated samples, which explains a slightly lower accuracy of these two models on the third and fourth datasets (see Fig. 14). Regarding the training time, it is almost the same for all models. If the training time is significantly higher for a model while not improving the performance, it can also be an important factor when selecting the model. Given that ResNet has more layers, it is expected to have a longer training time in models with this network. This has proved to be the case here, but the

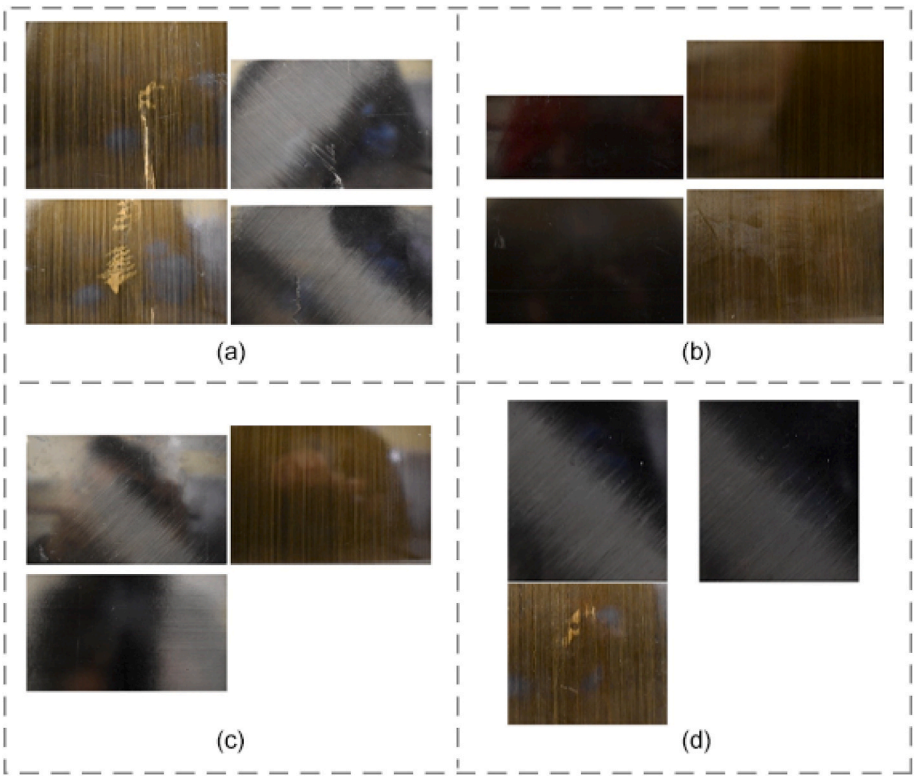

Fig. 15. Some examples of the predictions that are most prevalent across networks:
a) true positive, b) true negative, c) false positive, d) false negative.

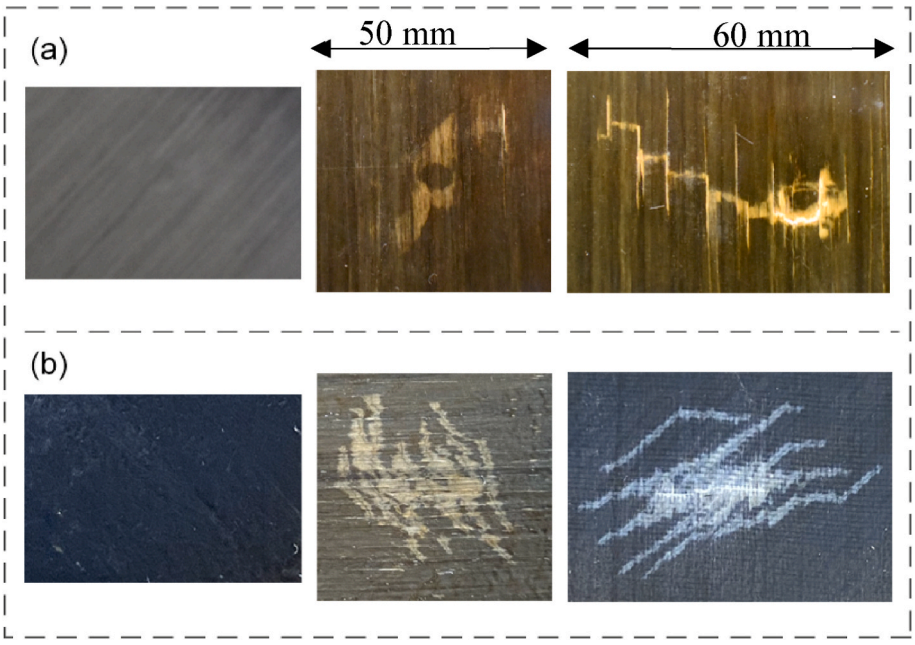

Fig. 16. Potential strategies to generate various damage patterns with higher detection possibility: a) using an HS40 carbon layer as the sensing layer (right). The damage pattern is larger and different from a sample without a sensor (left), or with a sensor composed of a YS-90 carbon layer (middle) (these images were taken after the impact test), b) using a woven glass layer as the outer layer (right). The white color and bi-directional fibres of this glass layer may generate a more recognisable and larger pattern than a sample without a sensor (left), or with a sensor composed of a unidirectional S-glass layer (middle) (these images were taken after the indentation test). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

difference is not big. The results show a faster training time in the third and fourth tasks, confirming the beneficial effect of the hybrid glass/carbon sensing technology in the autonomous detection of BVID using a CNN model.

Finally, some points to note: first, when it comes to an understanding of false positives (FPs) and false negatives (FNs), there are a few relevant factors. It is noticeable on the FP images that there are prevalent light reflection anomalies. Images with defects also tend to be brighter on average. This can sometimes give the network the impression that there is a correlation between image brightness and the presence of damage. Similarly, most FN images are darker because images without damage tend to be darker. Some examples of this can be seen in Fig. 15. For future work, a more in-depth pre-processing can be performed on images to ensure that the influence of brightness is minimised across the entire image dataset. Second, this study focused only on BVID; all results are based on the performance of the different models on the BVID image dataset. Note that due to the small size of dataset (limited number of images), images of all samples damaged at or below BVID energies (12J, 18J, 27J, 36J, 64J) were included. In this case, the results represent the average model performance on a range of impact energies within the BVID range. This can be seen as a good assessment of the CNN model's performance because the impact damage in real-world examples does not always happen at a well-known energy. However, depending on the specific application and damage monitoring scenario, it might be interesting for future studies to focus on only a particular impact energy, for example, the threshold of BVID and VID, to achieve a more precise prediction result. It would allow not only damage detection and localisation, but classification and severity estimation as well. Another potential route for future work is exploring different fibre types and architectures in the sensor structure to generate a different damage pattern on composite surfaces. For example, our preliminary investigations show that if using HS40 carbon prepregs as the sensing layer (instead of YS-90 carbon) or woven glass prepregs as the outer layer (instead of unidirectional glass), a better contrast between the damage pattern of a reference and sensor-integrated sample can be achieved, which accordingly improves the DL or ML model's performance (see Fig. 16). The proposed SHM method in this research can be further automated by adding a Remotely Operated Vehicle (ROE) that takes surface images of large composite structures, including inaccessible parts, in real-life applications.

4. Conclusions and future work

This research successfully applied four DL models to detect BVID on both impacted and non-impacted surfaces of composite panels with and without a hybrid glass/carbon sensor. The input in all DL models was simple surface images; thus, this study proposed an autonomous BVID recognition system in composite materials using DL models, then improved the DL model's performance by introducing hybrid glass/carbon sensors. The following conclusions can be drawn from the results:

- Depending on the dataset, different CNNs can be trained to detect BVID using simple images of the composite structure's surface. The results of all four CNNs in both reference and sensor-integrated composites of this study confirmed that the AI-based BVID detection accuracy is higher on the back-face than the front-face. This is due to a more recognisable damage pattern on the back-face of these thin-walled composite structures. Therefore, where the non-impacted side is accessible, using back-face images as input to an AI network is preferred.
- All four DL models could learn and identify BVID quite well. ResNet outperformed other models in most of the tasks. This is because of the flexible and robust architecture and using residual connections. Also, a Prototypical network can offer higher accuracy than a simple CNN, especially for small datasets, but it may require a higher computational time.
- The hybrid glass/carbon sensing technology provides a higher contrast between the damaged and undamaged areas, thus, improving the AI model's performance to a great extent, especially on the front-face. The results of the first and second models showed an improvement of 26.74% and 10.8% in damage detection accuracy when adding a sensor on the front-face. It should be noted that a very complex network, such as a Prototypical network, may capture very detailed features of damage patterns. In this case, as the sensor may add to the variance of the damage pattern, it can slightly decrease the model's accuracy. Nevertheless, the training time of models confirmed that the glass/carbon sensing system could significantly reduce the training time, enhancing the computational efficiency of AI-based damage detection systems.

In future work, the following research directions could be considered:

- A more in-depth pre-processing could be performed on input data to refine the design and further optimise the AI model's performance. For example, the BVID range can be narrowed, closer to the threshold of VID and BVID. Also, more efforts can be made to ensure that the influence of the brightness is minimised in the image dataset.
- Exploration of the potential of AI-based damage detection methods in a real-world setting, including scalability and reliability testing. For example, the proposed SHM method can be further automated by adding an ROE that takes surface images of large composite structures, including inaccessible parts.
- Design of sensors to create more recognisable damage patterns, achieving a higher contrast between the damaged and undamaged areas. Further analysis of different DL models with input data, including images of composites surface with various sensors, can help find a better relationship between the network performance and damage pattern caused by the sensor.

CRediT authorship contribution statement

Ali Tabatabaeian: Conceptualization, Investigation, Methodology, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing.

Bruno Jerkovic: Software, Methodology, Writing – review & editing.

Philip Harrison: Supervision, Writing – review & editing.

Elena Marchiori: Supervision, Writing – review & editing, Funding acquisition.

Mohammad Fotouhi: Project administration, Supervision, Writing – review & editing, Funding acquisition.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

The research presented in this paper was financially supported by the UK Engineering and Physical Sciences Research Council (EPSRC) Grants EP/V009451/1, which focused on the development of high-performance impact-resistant composites with visible damage. Authors are also grateful to the Radboud-Glasgow partnership scheme for supporting this work. The data that support the findings of this study are available on request from the corresponding author.

List of Abbreviations

(AI) Artificial Intelligence
(DL) Deep Learning
Carbon Fibre Reinforced Polymer (CFRP)
(ML) Machine Learning
(LVI) Low-Velocity Impact
(SHM) Structural Health Monitoring
(CNN) Convolutional Neural Network
(BVID) Barely Visible Impact Damage
(VID) Visible Impact Damage
(NDE) Non-Destructive Evaluation
(BCE) Binary Cross Entropy
(CCE) Categorical Cross Entropy
(ANN) Artificial Neural Network
(MLP) Multi-Layer Perceptron
(FC) Fully Connected
(ReLU) Rectified Linear Unit
(CNN) Convolutional Neural Network
(ResNet) Residual Network
(TP) True Positives
(TN) True Negatives
(FP) False Positives
(FN) False Negatives
(ROE) Remotely Operated Vehicle

References

[1] Azimi M, Eslamlou AD, Pekcan G. Data-driven structural health monitoring and damage detection through deep learning: state-of-the-art review. Sensors 2020;20:2778. https://doi.org/10.3390/S20102778. 2020;20:2778.
[2] Nelon C, Myers O, Hall A. The intersection of damage evaluation of fiber-reinforced composite materials with machine learning: a review. J Compos Mater 2022;56:1417–52. https://doi.org/10.1177/00219983211037048/ASSET/IMAGES/LARGE/10.1177_00219983211037048-FIG19.JPEG.
[3] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, et al. Recent advances in convolutional neural networks. Pattern Recogn 2018;77:354–77. https://doi.org/10.1016/J.PATCOG.2017.10.013.
[4] Sony S, Dunphy K, Sadhu A, Capretz M. A systematic review of convolutional neural network-based structural condition assessment techniques. Eng Struct 2021;226:111347. https://doi.org/10.1016/J.ENGSTRUCT.2020.111347.
[5] Deng K, Liu H, Yang L, Addepalli S, Zhao Y. Classification of barely visible impact damage in composite laminates using deep learning and pulsed thermographic inspection. Neural Comput Appl 2023:1–15. https://doi.org/10.1007/s00521-023-08293-7.
[6] Hasebe S, Higuchi R, Yokozeki T, Takeda S. Multi-task learning application for predicting impact damage-related information using surface profiles of CFRP laminates. Compos Sci Technol 2022;231:109820. https://doi.org/10.1016/j.compscitech.2022.109820.
[7] Alhammad M, Avdelidis NP, Ibarra-Castanedo C, Torbali ME, Genest M, Zhang H, et al. Automated impact damage detection technique for composites based on thermographic image processing and machine learning classification. Sensors 2022;22. https://doi.org/10.3390/s22239031.
[8] Hasebe S, Higuchi R, Yokozeki T, Takeda S ichi. Internal low-velocity impact damage prediction in CFRP laminates using surface profiles and machine learning. Compos B Eng 2022;237:109844. https://doi.org/10.1016/j.compositesb.2022.109844.
[9] Zargar SA, Yuan FG. Impact diagnosis in stiffened structural panels using a deep learning approach. Struct Health Monit 2021;20:681–91. https://doi.org/10.1177/1475921720925044.
[10] Wei Z, Fernandes H, Herrmann HG, Tarpani JR, Osman A. A deep learning method for the impact damage segmentation of curve-shaped CFRP specimens inspected by infrared thermography. Sensors 2021;21:395. https://doi.org/10.3390/S21020395. 2021;21:395.
[11] Jung KC, Chang SH. Advanced deep learning model-based impact characterization method for composite laminates. Compos Sci Technol 2021;207:108713. https://doi.org/10.1016/j.compscitech.2021.108713.
[12] Ahmed O, Wang X, Tran MV, Ismadi MZ. Advancements in fiber-reinforced polymer composite materials damage detection methods: towards achieving energy-efficient SHM systems. Compos B Eng 2021;223:109136. https://doi.org/10.1016/J.COMPOSITESB.2021.109136.
[13] Fotouhi S, Pashmforoush F, Bodaghi M, Fotouhi M. Autonomous damage recognition in visual inspection of laminated composite structures using deep learning. Compos Struct 2021;268:113960. https://doi.org/10.1016/J.COMPSTRUCT.2021.113960.
[14] Delaney MP, Fung SYK, Kim H. Dent depth visibility versus delamination damage for impact of composite panels by tips of varying radius. J Compos Mater 2018;52:2691–705. https://doi.org/10.1177/0021998317752502.
[15] FLIR. How does emissivity affect thermal imaging?. 2021. https://www.flir.co.uk/discover/professional-tools/how-does-emissivity-affect-thermal-imaging/. [Accessed 4 April 2023].
[16] Tabatabaeian A, Liu S, Harrison P, Schlangen E, Fotouhi M. A review on self-reporting mechanochromic composites: an emerging technology for structural health monitoring. Composites Part A Appl Sci Manuf 2022;163:107236. https://doi.org/10.1016/J.COMPOSITESA.2022.107236.
[17] Fotouhi S, Jalalvand M, Wisnom MR, Fotouhi M. Smart hybrid composite sensor technology to enhance the detection of low energy impact damage in composite structures. Composites Part A Appl Sci Manuf 2023;172:107595. https://doi.org/10.1016/J.COMPOSITESA.2023.107595.
[18] Kaczmarek H, Maison S. Comparative ultrasonic analysis of damage in CFRP under static indentation and low-velocity impact. Compos Sci Technol 1994;51:11–26. https://doi.org/10.1016/0266-3538(94)90152-X.
[19] Leroy A, Scida D, Roux E, Toussaint F, Ayad R. Are there similarities between quasi-static indentation and low velocity impact tests for flax-fibre composites. 2021. https://doi.org/10.1016/j.indcrop.2021.113840.

[20] Sutherland LS, Guedes Soares C. The use of quasi-static testing to obtain the low-velocity impact damage resistance of marine GRP laminates. Compos B Eng 2012; 43:1459–67. https://doi.org/10.1016/J.COMPOSITESB.2012.01.002.
[21] Bouvet C, Rivallant S. Damage tolerance of composite structures under low-velocity impact. In: Silberschmidt VV, editor. Dyn. Deform. Damage fract. Compos. Mater. Struct.; 2016. p. 7–33. https://doi.org/10.1016/b978-0-08-100080-9.00002-6.
[22] Choudhary K, DeCost B, Chen C, Jain A, Tavazza F, Cohn R, et al. Recent advances and applications of deep learning methods in materials science. npj Comput Mater 2022;8:1–26. https://doi.org/10.1038/s41524-022-00734-6.
[23] Semenov A, Boginski V, Pasiliao EL. Neural networks with multidimensional cross-entropy loss functions. Comput Data Soc Networks CSoNet 2019 Lect Notes Comput Sci 2019:57–62. https://doi.org/10.1007/978-3-030-34980-6_5/TABLES/1. 11917 LNCS.
[24] Chen CH, Lin PH, Hsieh JG, Cheng SL, Jeng JH. Robust multi-class classification using linearly scored categorical cross-entropy. Proc 3rd IEEE Int Conf Knowl Innov Invent 2020. https://doi.org/10.1109/ICKII50300.2020.9318835. ICKII 2020 2020:200–3.
[25] Singh J, Banerjee R. A study on single and multi-layer perceptron neural network. Proc 3rd Int Conf Comput Methodol Commun ICCMC 2019;2019:35–40. https://doi.org/10.1109/ICCMC.2019.8819775.
[26] Rautela M, Senthilnath J, Monaco E, Gopalakrishnan S. Delamination prediction in composite panels using unsupervised-feature learning methods with wavelet-enhanced guided wave representations. Compos Struct 2022;291:115579. https://doi.org/10.1016/J.COMPSTRUCT.2022.115579.
[27] Elyasi N, Hosseini moghadam M. Classification of skin lesions by topological data analysis alongside with neural network. 2020.
[28] Giusti A, Cireşan DC, Masci J, Gambardella LM, Schmidhuber J. Fast image scanning with deep max-pooling convolutional neural networks. 2013 IEEE Int Conf Image Process ICIP 2013 - Proc 2013. https://doi.org/10.1109/ICIP.2013.6738831. 4034–8.
[29] Yu D, Wang H, Chen P, Wei Z. Mixed pooling for convolutional neural networks. Lect Notes Comput Sci 2014;8818:364–75. https://doi.org/10.1007/978-3-319-11740-9_34/COVER.
[30] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 2016:770–8.
[31] Deng J, Dong W, Socher R, Li L-J, Li Kai, Fei-Fei Li. ImageNet: a large-scale hierarchical image database. IEEE Conf. Comput. Vis. Pattern Recognit., Institute of Electrical and Electronics Engineers (IEEE) 2009:248–55. https://doi.org/10.1109/cvpr.2009.5206848.
[32] Weiss K, Khoshgoftaar TM, Wang DD. A survey of transfer learning. J Big Data 2016;3:1–40. https://doi.org/10.1186/S40537-016-0043-6/TABLES/6.
[33] Wang Y, Yao Q, Kwok JT, Ni LM. Generalizing from a few examples: a survey on few-shot learning. ACM Comput Surv 2020;53. https://doi.org/10.1145/3386252.
[34] Kulis B. Metric learning: a survey. Found Trends® Mach Learn 2013;5:287–364. https://doi.org/10.1561/2200000019.
[35] Snell J, Swersky K, Zemel TR. Prototypical networks for few-shot learning. Adv Neural Inf Process Syst 2017;30.
[36] Li H, Li J, Guan X, Liang B, Lai Y, Luo X. Research on overfitting of deep learning. Proc - 2019 15th Int Conf Comput Intell Secur CIS 2019;2019:78–81. https://doi.org/10.1109/CIS.2019.00025.
[37] Kingma D.P., Ba J.L. Adam: a method for stochastic optimization 2014, CoRR. abs/1412.6980.
[38] Paszke A, Gross S, Massa F, Lerer A, Bradbury Google J, Chanan G, et al. PyTorch: an imperative style, high-performance deep learning library. Adv Neural Inf Process Syst 2019;32.
[39] Fualdes C. Composites @ Airbus - damage tolerance methodology presentation at the composite damage tolerance and maintenance workshop. 2006. Chicago.
[40] Fawcett AJ, Oakes GD. Boeing composite airframe damage tolerance and service experience. Chicago: Presentation at the Composite Damage Tolerance and Maintenance Workshop; 2006.
[41] Entezami A, Shariatmadar H, De Michele C. Non-parametric empirical machine learning for short-term and long-term structural health monitoring. Struct Health Monit 2022;21:2700–18. https://doi.org/10.1177/14759217211069842.