3rd Conference on Production Systems and Logistics

Overcoming Data Scarcity In The Quality Control Of Safety-Critical Fibre-Reinforced Composites By Means Of Transfer And Curriculum Learning

Florian Brillowski[1], Vanessa Overhage[1], Thorsten Tegetmeyer-Kleine[2], Benjamin Hohnhäuser[3]

[1]Institut für Textiltechnik der RWTH Aachen University, Aachen, Germany
[2]MABRI.VISION GmbH, Aachen, Germany
[3]Gesellschaft zur Förderung angewandter Informatik e.V., Berlin, Germany

Abstract

Fibre-reinforced composites are one promising material class to provide a response to the increasing environmental awareness within society. Due to their excellent lightweight potential, fibre-reinforced composites are preferably employed in safety-critical applications, requiring extensive quality control (QC). However, commercially available QC systems are only able to measure fibre deviations, not directly detecting the error itself. In consequence, a worker is required to perform a manual inspection.

Artificial intelligence and especially convolutional neural networks (CNN) offer the opportunity to directly detect and classify defects. However, to train the corresponding algorithms large amounts of data are required, which are often inaccessible in production. Artificial augmentation of the available data is a popular approach to tackle this problem, yet, resulting most of the time in undesired overfitting of the CNN.

Therefore, in this contribution we examine the transfer of human learning behaviour elements to algorithms in form of transfer learning (TL) and curriculum learning (CL). The overall aim is to research, whether CL and TL are appropriate approaches to address data scarcity in e.g. production environments. Therefore, we perform our research on the error detection of three-dimensional shaped fibre-reinforced textiles.

Keywords

Machine Learning; Quality Control; Data Scarcity; Composites; Curriculum

1. Introduction

In light of increasing environmental awareness and a growing resource responsibility within society, lightweight construction solutions are becoming increasingly important. Besides in sports, these constructions are especially used in transportation or engineering to reduce moved masses and hence lower the pollution caused by emissions. Fibre-reinforced composites in particular are preferably employed in safety-critical lightweight applications due to their excellent mechanical properties in relation to their low weight [1]. As a result, many lives depend on the proper function and reliability of fibre-reinforced composites. To prevent fatal component failure, the quality of the manufactured parts is thoroughly tested during and after the production process. Various automated quality control (QC) systems already exist for this purpose. Yet, commercially available systems only measure deviations in fibre orientations [2]. Subsequently, a worker must determine the specific defect in a manual visual inspection. The inspection

DOI: https://doi.org/10.15488/12191

process requires a high level of concentration, is repetitive and exhausting for the worker, potentially resulting in errors the longer an inspection takes [3].

Convolutional neural networks (CNN) are a promising approach to directly detect varying defects without signs of fatigue. However, a large amount of training data is required for each material and defect. Especially for fibre-reinforced composites, many reinforcing textiles are available. Subsequently, this results in a vast amount of defect-material combinations. In industrial environments, data is often only available to a limited extent or not at all due to a lack of integrated sensors [4]. For this purpose, data augmentation techniques such as mirroring or rotating image data have already been developed but can quickly result in an overfitting of a CNN due to lacking data diversity [5]. As a consequence, there is a demand for a machine learning concept that allows the development of an adaptive QC system that uses limited amounts of data in an efficient way.

Therefore, in this contribution we investigate to what extent concepts comparable to the human learning behaviour (e.g. curricula with increasing complexity) can be transferred to algorithm-based learning. In this context, transfer learning (TL) as well as curriculum learning (CL) are examined for defect detection during three-dimensional shaping of reinforcing textiles. The overall aim is to research, whether CL and TL are appropriate approaches to tackle data scarcity in e.g. production environments. Different CNN architectures are evaluated during hyperparameter optimisation and thereafter the results of TL and CL are compared to a regular (vanilla) training approach. In conclusion, design recommendations and further research activities are derived.

2. Materials & Methods

The TL and CL approaches are strategies of machine learning (ML), which in turn is a subdomain of artificial intelligence. The aim of ML is to enable machines to recognize patterns and develop appropriate solutions [6]. In the field of image recognition, especially convolutional neural networks (CNN) are employed. With these, for instance, defects in textiles can be detected, yet a lot of image data is required. Currently, simple data augmentation operations (e.g., flipping or rotating) are commonly pursued to artificially increase the amount of data. Complimentary to data augmentation, the intention of TL and CL is to utilize the existing data more efficiently and thus enable the CNN to achieve faster learning success [7]. The dataset used and the two ML approaches are explained in more detail below.

2.1 Dataset

The used dataset includes 3,653 image captures of biaxial ±45° glass non-crimp fabrics (fringe) with 320 g/m² grammage. The captured images have a resolution of 2,048 x 1,536 pixels and were acquired during three-dimensional forming of the textiles with an Apodius HP-C-V3D vision sensor, mounted on a Hexagon ROMER Absolute Arm. The forming was performed on the shape shown in Figure 1 as it favours multiple defects due to its complex corners and varying curvatures.

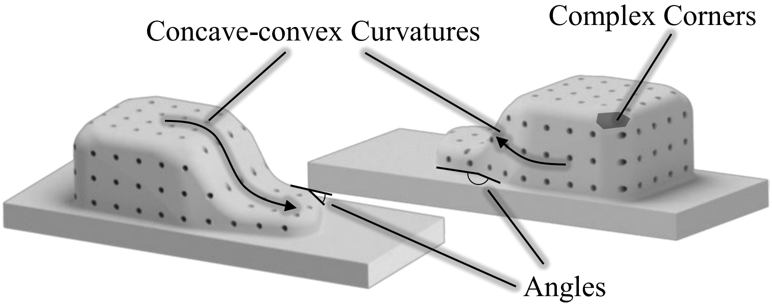


Figure 1: Geometry used for dataset generation

The raw images were segmented into 12 overlapping images with a resolution of 512 x 512 pixels each (Figure 2) and were then assigned to exactly one of the defect classes fold, flawless, gap, undulation, sling and distortion. Exemplary embodiments of each class are depicted in Figure 3. As some classes occur more frequently than others, data augmentation is used to create a balanced dataset. An unbalanced dataset could result in learning biased correlations and thus leading to erroneous classifications. Data augmentation is performed using the operations rescale, flip, brightness adjustment, rotation and zooming to create three distinct datasets with 500, 2,000 and 3,000 images for each class. All three datasets are used during a hyperparameter optimisation to determine a fitting network architecture as well as the best performing dataset. Hyperparameters are parameters that are not influenced by the dataset during training and therefore have to be specified before the network's training. The network's architecture is primarily determined by these hyperparameters, which are iteratively identified in a so-called hyperparameter optimisation. During our hyperparameter optimisation (HPO) we alter epochs, dropouts, learning rate, number of layers, number of neurons and batch size. After HPO, a validation dataset unknown to the network is applied to evaluate the generalisation capability (validation accuracy) of the network. The best performing augmented dataset is used for transfer and curriculum learning.

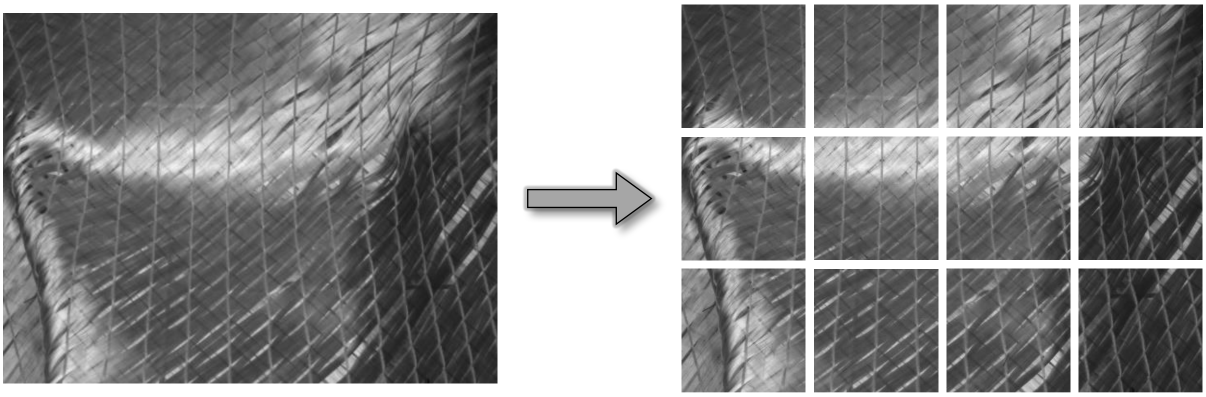


Figure 2: Segmentation of raw images into 12 parts

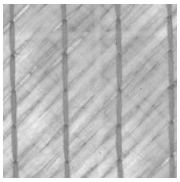
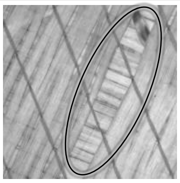
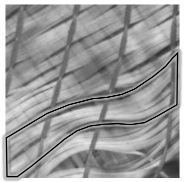
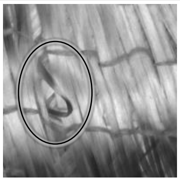
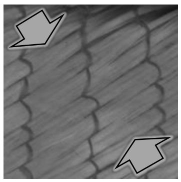
Fold	Flawless	Gap	Undulation	Sling	Distortion
					

Figure 3: Examples for each defect class

2.2 Transfer Learning

In transfer learning, the applied neural network is not designed from scratch, but a pre-trained network is used as a starting point. In the pre-trained network, the weights between the individual neurons are already preset. Starting from this, the weights can be adjusted more quickly and thus high validation accuracies can be achieved in fewer epochs [8].

In this contribution, the Xception [9] and VGG16 [10] networks are used for transfer learning because they exhibit high top 5 accuracy while having widely varying depth, data size, and parameter specification. A version of each network pretrained with the ImageNet dataset is used, a hyperparameter optimization is performed and the existing network architecture is specifically adapted. Based on this, design guidelines for transfer learning are derived and the validation results are compared with those of the regular (vanilla) network.

2.3 Curriculum Learning

Curriculum learning is a learning concept for neural networks in which content is processed in increasing complexity analogous to human teaching. In the context of neural networks, this means that during training not all data is fed to the network at once, but that the training dataset is gradually extended in increasing complexity until all data is included. Bengio et al. were able to demonstrate a performance improvement of a network by applying curriculum learning in different use cases [11]. However, this has not yet been done in the context of defect detection within the fibre-reinforced composite domain.

For CL, the regular trained vanilla network is used. Based on this, three different curricula are investigated in three variants each and then an HPO is performed for each curriculum. In Curriculum 1, the number of classes to be learned is gradually increased from 3 to 6 in four stages during training (Table 1). In Curriculum 2, six classes are introduced and differentiated from the beginning. Initially, this curriculum uses only images with horizontal fibres. In the course of the curriculum the diversity of the fibre trajectories increases. In addition, the defect characteristics of folds and loops are subjectively divided into four difficulty levels. A third variation of Curriculum 2 has an increased amount of steps (eight) in which the data is fed gradually during the network's training. The third curriculum is partitioned based on the illumination intensities of the images. In the first stage, bright images are fed, in the second stage dark images are added and in the third stage images with bright and dark areas are additionally provided, which have so-called shadow edges. In the last stage, the remaining images are fed. The three variants of each curriculum change the order in which classes, illumination intensities, or fibre orientations are fed to the training. These are randomized as well as descending and ascending according to their brightness (Curriculum 3), fibre orientation (Curriculum 2) or precision of each class of the vanilla mesh (Curriculum 1). Precision of a class is defined as the number of correctly labelled images of a class divided by the sum of correctly labelled images and the falsely allocated images to this class.

Table 1: Overview of examined curriculum strategies

	First stage of training	**Changes for following stages**
Curriculum 1	Starting with three classes	Adding one class every stage
Curriculum 2	Six classes, starting with horizonzal fibres only, subjective difficulty for loops and folds	Increasing deviation of fibre trajectories, increasing difficulty for loops and folds
Curriculum 3	Only bright images are used	Decreasing illumination within pictures

3. Results

Following, the results of the vanilla network's as well as the transfer and curriculum learning networks' HPO and validation are described.

3.1 Vanilla Approach

In the course of the vanilla approach, a CNN architecture is determined and all data is fed to the vanilla network at once during training. The HPO provides the highest validation accuracy and lowest loss function with the dataset consisting of 3,000 images. This dataset is subsequently used for curriculum and transfer learning. The HPO of the vanilla network results in a validation accuracy of 90.3 % and a loss value of 0.314.

The network consists of five convolution and one flattening layer, each followed by a dropout layer to avoid overfitting. The hyperparameters are 64-128-128-256-512 neurons per layer, 25 training epochs, a dropout value of 0.2, a learning rate of 0.0002 and a batch size of 20. The vanilla network's validation accuracy of 90.3 % is used as reference value.

3.2 Transfer Learning Approach

An HPO is performed with the Xception ImageNet. Analogous to the vanilla training, epochs, batch size, dropout, learning rate, number of layers and number of neurons per layer are varied. None of the 142 trained networks shows comparable convergence or accuracy as the vanilla network. Either the training and validation accuracies are lower than 75 % or a difference between validation and training functions is evident (overfitting). Since none of the trained networks has a sufficiently high accuracy, the approach with the Xception network is not investigated further.

Analogously, an HPO with the same parameters is also performed with the VGG16-ImageNet [10]. The highest validation accuracy of 93.06% with a loss of 0.5147 provides 15 epochs, a batch size of 5, a 0.3 dropout and a learning rate of 0.0002 (Figure 4). The number of layers and neurons per layer from [10] can be confirmed during the HPO, resulting in only adding one dropout layer at the end of the model and adjusting the output layer accordingly.

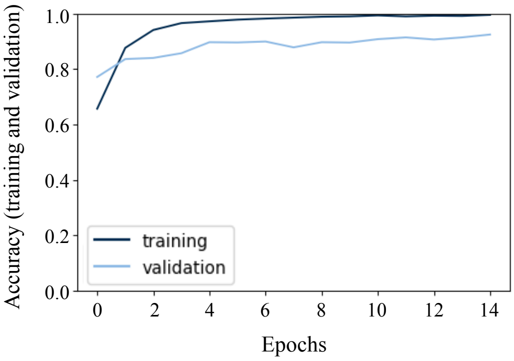


Figure 4: Training and Validation for VGG 16's Transfer Learning

Since the VGG16 is a very deep network with 13 convolutional layers, we omitted and added individual layers to investigate how this affected training. In total, we removed 2, 4, 6, and 9 layers and trained the network. For all networks with a reduced number of layers, a significant difference between training and validation data is evident, suggesting overfitting. The more layers removed, the larger the difference. In contrast, adding 2, 4, 6, and 9 layers shows an increase in the generalisation ability of the network. The difference between validation and training data decreases as the number of layers increases. From 6 additional layers onwards, the effect reverses and a reduction in validation accuracies and an increase in loss occurs.

3.3 Curriculum Learning Approach

The HPO results in the same hyperparameters being applied for Curriculum 1 as in the vanilla network. The three variants achieve a validation accuracy of 87.15 %, 91.07 % and 95.25 %. The highest validation accuracy (95.25 %) with a loss of 0.2202 is achieved by the curriculum in which the classes are gradually added to the training with increasing precision. Compared to the vanilla approach, the curriculum increases the precision of the class flawless from 74.17 % to 96.14 % during training and 90.39 % during validation. All other classes also show values above 90 %. In total, fewer images are misclassified. (Figure 5).

Predicted Class
(Vanilla)

Actual Class

	FO	FL	GA	UN	SL	DI
FO	2,113	7	2	0	0	25
FL	1	2,036	0	0	0	8
GA	21	359	1,680	1	3	44
UN	1	234	3	1,709	4	188
SL	27	55	0	0	1,878	134
DI	3	54	0	0	0	2,082

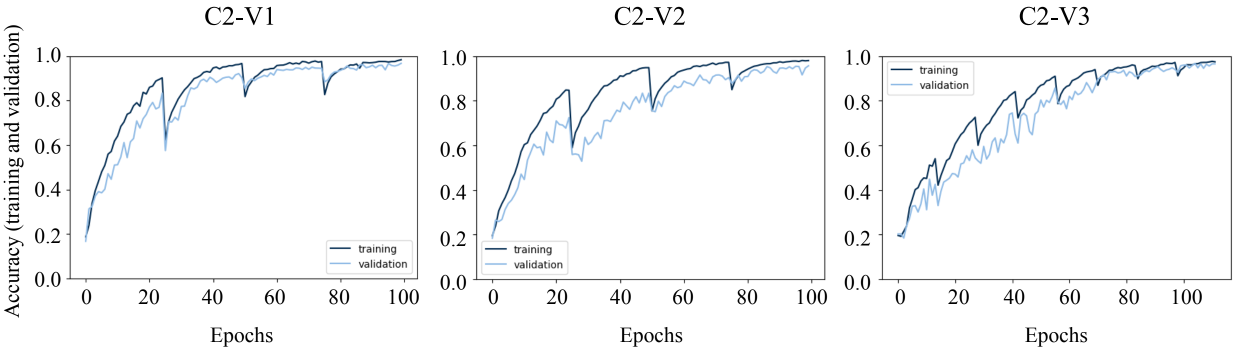
Predicted Class
(C1-V3)

	FO	FL	GA	UN	SL	DI
FO	2,147	0	0	0	0	0
FL	0	2,045	0	0	0	0
GA	1	62	2,042	1	1	1
UN	0	18	0	2,129	0	1
SL	0	2	0	0	2,092	0
DI	0	0	0	0	0	2,139

FO – Fold, FL – Flawless, GA – Gap, UN – Undulation, SL – Sling, DI - Distortion

Figure 5: Confusion Matrices during Training of Vanilla Approach (Vanilla, left) and Curriculum 1 Variant 3 (C1-V3, right)

For Curriculum 2, which is structured according to fibre orientations and defect characteristics, 25 training epochs, a learning rate of 0.0002, a dropout of 0.2 and a batch size of 20 are derived from HPO. With these parameters, validation accuracies of 95.00 % (random order), 93.06 % (decreasing fibre orientation and precision) and 90.79 % (increasing fibre orientation and precision, 8 stages) are achieved (Figure 6). The class flawless together with the class distortion shows the most classification errors and thus the lowest precision. However, the curriculum can increase the precisions to 89.46 % (distortion) and 93.17 % (flawless). Especially in the first stages of the curriculum, many images are misclassified as flawless or distortion during training, while other classes show less or hardly any classification errors. Accordingly, the precision values of flawless and distortion are below 60 % during the early stages.


C2-V1

C2-V2

C2-V3

Figure 6: Accuracies of Training and Validation of Curriculum 2's (C2) three variants (V1-V3, left to right)

Analogous to the other two curricula, an HPO is also performed for the brightness-based Curriculum 3. The resulting hyperparameters are 50 training epochs, a learning rate of 0.0002, a dropout of 0.2 and a batch size of 20. All three variants achieve comparable validation accuracies, differences are in the per mil range (V1: 94.59 %, V2: 94.71 %, V3: 95.34 %).

Compared to the other two curricula, Curriculum 3 achieves the highest validation accuracy overall, but the precision of the class flawless with 84.38 % is lower than for Curriculum 1 (90.39 %) and Curriculum 2 (93.17 %). Consequently, the classification errors in Curriculum 1 and Curriculum 2 are distributed more evenly across all classes, while in Curriculum 3 the class flawless is the most frequently specified class for classification errors. Overall, higher validation accuracies are achieved with all three curricula ($Acc_{Vanilla}$ = 90.3 %; $Acc_{Curricula}$ > 94.59 %). The precision values for the frequently incorrectly specified class flawless are increased as well due to the curricula ($Prec_{Vanilla}$ = 74.17 %; $Prec_{Curricula}$ > 84.38 %).

4. Discussion & Conclusion

The results of transfer and curriculum learning indicate that both strategies can improve the generalisation ability of a convolutional neural network. Overall, fewer classification errors occur when one of the two approaches is employed. However, especially the defect class flawless seems to be difficult to distinguish from the other classes and contains the most incorrect classifications. In our opinion, the difficulty of differentiation is due to the great similarity to images of the class gaps. Images of the class flawless often have small spaces that can be misinterpreted as gaps. We assume that this is why gaps are most often incorrectly categorised as flawless. Interestingly, flawless images are not or hardly ever assigned to the class gaps. A similar observation can be made for the class distortion, where we attribute the incorrect classifications to superpositions of several defects that may additionally arise due to the distortion, thus posing further challenges to the network. Nevertheless, due to transfer and curriculum learning, the networks exhibit accuracies, precision and recalls of over 90 %.

In the context of the TL, pretrained networks (with ImageNet) can in some cases be adopted one-to-one for the reinforcement textile defects dataset. Whether this is possible depends strongly on the use case and the respective network structure. The more convolutional layers a pre-trained network has, the more data is needed to adapt all weights to the new use case in the course of training. Therefore, if the goal is to achieve maximum training success of the network with as little data as possible, lean networks with few convolutional layers should be selected. However, the results of the TL suggest that additional layers enable better feature recognition and thus contribute to higher classification accuracies. At this point, further research is needed to investigate causal relationships between network depth, further hyperparameters and validation accuracies. Based on these yet to be gained insights, elaborated design recommendations for TL in production related context can be derived in the future.

In curriculum learning, the results show that the derivation of a semantically designed learning plan has a beneficial effect on the network's performance indicators (accuracies, precision, recall). However, it is not only the semantic, stage-by-stage division into a curriculum that is important, but also the order in which the stages are presented to the network. The analysis of the three curricula shows that, for example, the arrangement according to precision or the arrangement according to the subjective perception of difficulty apparently has no immediate correlation on an improvement. These approaches are therefore not suitable as general design recommendations for a curriculum. In this context, the ethically and morally motivated question arises whether a machine learns in the same way as a human being and is able to perform a nuanced differentiation? Based on this question, we propose to develop metrics for the perception of difficulty or relatedness of data units. We see initial starting points for this in the use of confidence scores as well as density-based clustering approaches to measure difficulty and identify semantically related data points.

In conclusion, we observe that both learning strategies achieve higher accuracies in training and validation as well as higher discriminatory power in classification. The observed learning effect is comparable to an increase in the amount of data in the vanilla approach. As a result, both CL and TL contribute to making big data approaches accessible for applications with few data available. Thus, the further investigation and successful enhancement of both approaches represents an essential milestone in making artificial intelligence accessible in data scarce environments.

Acknowledgements

Funded by the Federal Ministry of Education and Research (BMBF) in context of the project "KIQS - AI based quality control system to classify draping defects" (01IS20089C).

References

[1] Towsyfyan, H., Biguri, A., Boardman, R., Blumensath, T., 2020. Successes and challenges in non-destructive testing of aircraft composite structures. *Chinese Journal of Aeronautics*, *33*(3), 771-791. doi: 10.1016/j.cja.2019.09.017

[2] Fuhr, J. P., 2017. *Schichtbasierte Modellierung von Fertigungseffekten in der Struktursimulation von Faserverbundwerkstoffen*. Verlag Dr. Hut. Dissertation

[3] Luczak, H., Mueller, T., 1994. Worker attention and fatigue. *Design of work and development of personnel in advanced manufacturing*, 463-492.

[4] Tekin E., Kapan Ö., 2016. Composite manufacturing data management in aerospace industry. *Procedia CIRP 41*, pp. 1039-1042, doi: 10.1016/j.procir.2015.12.058

[5] Perez, L., Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621.*

[6] Erickson, B. J., Korfiatis, P., Akkus, Z., Kline, T. L., 2017. Machine learning for medical imaging. *Radiographics*, *37*(2), 505-515. doi: 10.1148/rg.2017160130

[7] Shorten, C., Khoshgoftaar, T. M., 2019. A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 1-48. doi: 10.1186/s40537-019-0197-0

[8] Neyshabur, B., Sedghi, H.,, Zhang, C., 2020. What is being transferred in transfer learning?. *arXiv preprint arXiv:2008.11687.*

[9] Chollet, F., 2016. Xception: Deep Learning with Depthwise Separable Convolutions. *arXiv preprint arXiv:1610.02357.*

[10] Simonyan, K.; Zisserman, A., 2014. Very deep convolutional networsk for large-scale image recognition. *arXiv preprint arXiv:1409.1556.*

[11] Bengio, Y., Louradour, J., Collobert, R., Weston, J., 2009. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning* (pp. 41-48). doi: 10.1145/1553374.1553380

Biography

Florian Brillowski is an engineer specialized in plastics processing and operations research. He is working at the Institute for Textile Technology of RWTH Aachen University as research associate since 2018. His research interests are AI, digitisation, automation, composites and prototyping. He is currently working on his dissertation about the acceptance of AI-based decision support systems.

Vanessa Overhage is a textile engineer specialised in composite materials. Since 2021 she is a research associate at the Institute of Textile Technology at RWTH Aachen University. Her research interests are composites, sustainability and AI. She is currently esearching the use of recycled carbon fibres in the construction industry.

Thorsten Tegetmeyer-Kleine graduated in electrical engineering from RWTH Aachen University with major in systems and automation. During his master's degree in 2019, he started specializing in Deep Learning. Currently, he is employed as Software Developer for AI at MaBri Vision GmbH

Benjamin Hohnhäuser is the head of the robotics research group, department Image Processing / Industrial Applications at the Society for advancement of applied computer science (GFaI e.V.). His major fields of research are 3D-data processing, automated coating, machine learning and robotics.