Cross-evaluation of a parallel operating SVM – CNN classifier for reliable internal decision-making processes in composite inspection

Meister, Sebastian; Wermes, Mahdieu; Stüve, Jan; Groves, Roger M.

DOI
10.1016/j.jmsy.2021.07.022

Publication date
2021

Document Version
Final published version

Published in
Journal of Manufacturing Systems

Journal of Manufacturing Systems 60 (2021) 620–639

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

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

Technical Paper

Cross-evaluation of a parallel operating SVM – CNN classifier for reliable internal decision-making processes in composite inspection

Sebastian Meister [a,b,*], Mahdieu Wermes [a], Jan Stüve [a,b], Roger M. Groves [b]

[a] Center for Lightweight Production Technology (ZLP), German Aerospace Center (DLR), Ottenbecker Damm 12, Stade 21680, Germany
[b] Aerospace Non-Destructive Testing Laboratory, Delft University of Technology, Kluyverweg 1, Delft 2629, The Netherlands

ARTICLE INFO

Keywords:
Explainable Artificial Intelligence
Automated Fiber Placement
Inline inspection
Convolutional Neural Network
Laser Line Scan Sensor
Support Vector Machine

ABSTRACT

In the aerospace industry, automated fibre laying processes are often applied for economical composite part fabrication. Unfortunately, the current mandatory visual quality assurance process takes up to 50% of the entire manufacturing time. An automised classification of manufacturing deviations using Neural Networks potentially improves the inspection's effectiveness. Unfortunately, the automated decision-making procedures of machine learning approaches are challenging to trace. Therefore, we introduce an approach for evaluating the classifiers response for this use case.

For this purpose, we present a parallel classification approach of *Convolutional Neural Network* (CNN) and *Support Vector Machine* (SVM) with suitable intermediate checking stages between both classification processes. The particular novelty of this study is this intermediate comparison to trace the behaviour of the two classifiers along their image processing chains and to project the results back to the input image.

With respect to the SVM, we analyse their extracted input features via *t-Distributed Stochastic Neighbor Embedding* calculations and *parallel coordinates* plots. Moreover, the classification score of the SVM as well as the feature vector distances within the SVM are investigated. For the CNN, the outputs of its first joined convolutional layer are correlated with the raw input images of different classes using *Structural Similarity Index Measure* metrics. Additionally, also the CNN's classification rates are analysed. Accordingly, a suitable uncertainty confidence interval for the CNN is determined on the bases of its neural activations. Finally, the relevance of individual pixels for the CNN decision is determined through *Smooth Integrated Gradients* and linked to the manually extracted image features for the SVM Classifier.

The results of this paper are particularly valuable for developers and users of visual inspection systems in safety-critical domains.

1. Introduction

Structural components made of fibre composites are widely used in aerospace. This is especially the case since the mass production of the Airbus A350 XWB and Boeing 787 aircraft models [1,2]. For instance, 73 aircrafts were delivered of the Airbus A350 XWB in 2020 [3], and the company assumes a total delivery volume of this aircraft type in the period from 2019–2038 of 4116 aircraft [4]. Boeing predicts to sell about 7480 wide body aircrafts including the B787 between 2020–2039 [5]. Often aircraft parts are built from *Carbon Fiber Reinforced Plastic* (CFRP) and have complex geometries. This can substantially extend the production time and increase manufacturing costs. So, highly automated and hence cost efficient manufacturing techniques are applied. However, in safety-critical industries such as aerospace, a visual inspection of each fibre material layer is required. This additional inspection step can take up to 50% [6] of the production time. Automating this typically manual inspection would allow a reduction of costs while improving the quality of the inspection.

Automated inspection requires a trustworthy automated classification of fibre layup defects within a measurement image [7,8] and *Machine Learning* (ML) techniques have shown good results in related research for the prediction of such manufacturing defects [9,10]. However, the plausibility of a certain classification decision of such approaches is mostly difficult to comprehend, which is particularly challenging for deep learning classifiers [7,8].

Our first results on assessing the importance of individual image

* Corresponding author at: Center for Lightweight Production Technology (ZLP), German Aerospace Center (DLR), Ottenbecker Damm 12, Stade 21680, Germany.
E-mail address: sebastian.meister@dlr.de (S. Meister).

https://doi.org/10.1016/j.jmsy.2021.07.022
Received 17 May 2021; Received in revised form 17 July 2021; Accepted 17 July 2021
Available online 26 July 2021

regions for a *Convolutional Neural Network* (CNN) prediction as well as respective assessment metrics have already been presented in Meister et al. [11]. In this study, we have theoretically compared different *Explainable Artificial Intelligence* (xAI) methods and applied the three techniques *Deep Learning Important Features with Shapley Additive Explanations*, *Guided Gradient Class Activation Mapping* and *Smooth Integrated Gradients* for a detailed investigation of the classification of original example fibre layup defects for six defect classes. The classifiers behaviour for manipulated input data was investigated via the selected xAI calculations and the *Smooth Integrated Gradients* (Smooth IG) was found to be particularly valuable for this use case.

With the aim of reconstructing the origin of an *Artificial Neural Network* (ANN) machine decision as well as verifying its plausibility, we propose a parallel classification approach consisting of a CNN and a *Support Vector Machine* (SVM) with linear kernel in this paper. Therefore, several intermediate outcomes of the two very different classifiers are checked against each other. These intermediate tests enable the assessment of the quality of a machine decision and the evaluation of the respective origin. For this purpose, synthetic image data are applied for tests in this paper [9]. With respect to the SVM classification, a targeted feature extraction and selection with subsequent visualisation is carried out. Afterwards, the margin of a feature vector to the SVM's separating hyperplane is evaluated as an additional parameter. For the utilised classifying CNN, *Smooth IG*, as a feasible xAI procedure is applied for evaluating the importance of individual image areas [11]. Furthermore, the visual representation of selected feature maps are considered. The individual intermediate results of the SVM and CNN classifiers are then compared in detail.

In this study, synthetic topology data of *Automated Fiber Placement* (AFP) layup defects are analysed [9]. The AFP fabrication process is quite novel, but is becoming more and more common in mass production. Thus, the findings from this paper can be easily transferred to industrial manufacturing [12–14]. Moreover, a *Laser Line Scan Sensor* (LLSS) is frequently used to capture the topology information of an AFP layup surface. Respectively, we investigate synthetic greyscale depth maps of CFRP fibre layup defects of such LLSS [9,12,13,15].

Considering the challenges outlined above, we will address the following research questions in this study:

I. Which procedure is suitable to represent the decision making process of a machine classifier in order to assess its reliability?
II. How do neural network classifiers and model based classifiers need to be linked to compare their inherent classification processes and to evaluate the plausibility of classification results for different fibre layup defect images?

The methodology of this paper covers the visualisation of crucial image regions for the CNN prediction. These image regions are matched with the most relevant features of the input feature vector of the SVM classification. Furthermore, the location of the feature vector in the vector space of SVM is compared with the neural output activations of the CNN. Moreover, the feature maps of the CNN are visualised and likened to geometric attributes obtained from the input image. Besides, the normalised values of the features applied for the SVM prediction are analysed via *t-Distributed Stochastic Neighbor Embedding* (t-SNE) statistics and visualised in *parallel coordinates* and their overall context is discussed.

2. Related research

In this section, the necessary fundamentals as well as related research to this paper are discussed.

2.1. Manufacturing process and layup defect

Currently popular methods for fibre placement are *Automated Tape Laying* (ATL) [16] and AFP [16,17]. The AFP technology can be differentiated even further on the basis of the processed material [16,17]. Typically, with this technology, CFRP material is applied to a mould layer by layer, see Fig. 1 for a schematic representation of this procedure. Complex fibre composite structures are often manufactured with the AFP technology [18]. In the AFP process, up to 32 slim strips of material are deposited in parallel next to each other along a defined path [19] using an effector to guide the fibre material from the robot to the surface of the tool. The material is deposited while heat and pressure are applied [16]. This procedure can be used to flexibly fabricate different components. Accordingly, Rudberg [18] expects the use of AFP technology to increase in the future.

In the literature several defect types, which might occur during AFP fibre deposition, are discussed. Commonly used defect types for analysis purposes are *wrinkles*, *twists*, *foreign bodies*, *overlaps* and *gaps* [19–22]. These are exemplary illustrated in Fig. 2. Following Harik et al. [21] and Potter [23] all defects that appear in the fiber placement process result in geometric deformations, therefore in Table 1 the geometrical defect characteristics are given. *Twists* and *wrinkles* have different but distinct geometries and both defects protrude from the laminate surface. This causes considerable variations in height, which form distinct defect edges. In the longitudinal direction, *wrinkles* show one clearly visible edge. In contrast, *twists* have a slight increase in height along their length. *Gaps* and *overlaps* are rather similar to each other. These two defects are very flat and hardly show any topological changes. *Gaps* show two slightly prominent edges transversely to the fibre. *Overlaps* form three small edges perpendicular to the fibre orientation. These defects are usually a combination of a *gap* and an *overlap*. Along the fibre orientation, almost no edges are visible in *gaps* and *overlaps*. Due to their similarities, the correct automatic classification of these two defect types is often quite difficult. All five described defect categories are also frequently used as examples by other researchers. Oromiehie et al. [19] used them to analyse process tolerances and influences through layup paths, Harik et al. [21] provided a simulative and text-based structured description of these defects and Heinecke and Willberg [22] described the potential mechanical effects of these defects.

2.2. Sensor based inspection in composites

In the field of automated inspection of fibre composite components, different sensors are being investigated. The overview of Sun et al. [24] provides a brief summary of the different approaches. In principle, a fundamental differentiation can be made between 2D and 3D measuring techniques. On one hand, there are 2D image based approaches such as polarisation sensitive sensors examined at the *Fraunhofer Institute for Integrated Circuits* (IIS) [25,26] and thermographic imaging cameras for composite inspection were investigated by Denkena et al. [27] and Schmidt et al. [28] from the *Institute of Production Engineering and Machine Tools – University Hanover* (IFW).

On the other hand, InFactory Solutions [13], Profactor [29], Electroimpact [12] and Danobat Composites [30] have examined LLSS systems. Such systems acquire three dimensional topology data of a surface. As these companies have shown, this sensor technology can be used particularly well for inline inspection in the AFP process.

2.3. Advanced classifier approaches for inspection

Many different conventional and ANN based classifiers are currently available. Shrestha and Mahmood [31] presented a comprehensive review of ANN architectures, their corresponding application cases and their characteristics. Besides CNN classifiers, they examined *Residual Networks*, *Autoencoders*, *Restricted Boltzmann Machines* and *Long Short-Term Memory* (LSTM). Sen et al. [32] compared commonly used conventional supervised learning classifiers. For this purpose, they carried out performance evaluations of the individual conventional classifiers. Complementing SVM, their research looked at *Naive Bayes*

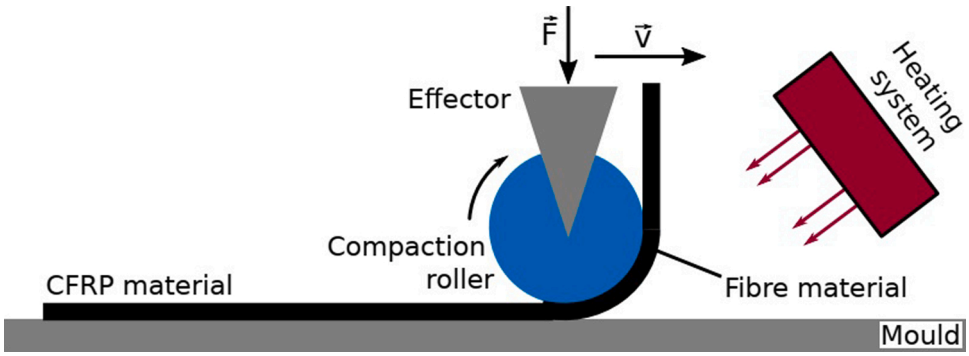

Fig. 1. AFP process including heating system and compacting roller. These components apply heat and pressure to the laid up fibre material. $\vec{v}$ is the effector velocity and $\vec{F}$ the compacting force. The figure is inspired by Meister et al. [9].

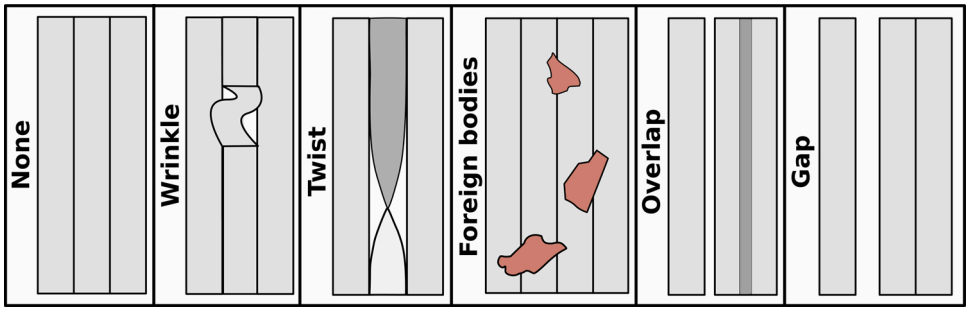

Fig. 2. Schematic representation of the fibre layup defects considered in this study, including a flawless reference sample. The figure is inspired by Meister et al. [9].

Table 1
An overview of the geometrical properties of the considered fibre layup defects schematically illustrated in Fig. 2 is presented. An estimated range of values for the length-to-width (l/w) proportion is given, due to geometrical variances within each class of defects. The considered *Cured Ply Thickness* (CPT) is around 0.125 mm. Referring to the thickness measures, + means a thickness increase and − indicates a thickness decrease. The table is inspired by Meister et al. [9].

	Wrinke	Twist	For. Bod.	Overlap	Gap
Typical ratio: (l/w)	0.5 to 2	5 to 10	unkn.	≤ course len.	≤ course len.
Thickn. dev. (±)	≥3× CPT (+)	≥2× CPT (+)	unkn.	≤1× CPT (+)	≤1× CPT (−)
Source	[21,19]	[21,22, 19]	[21, 22]	[20–22, 19]	[20–22, 19]

methods, *K-Nearest Neighbour* algorithms and *Decision Tree* (DT). From both references it can also be seen that CNN and SVM classifiers are particularly well suited for the classification of images. With respect to the investigations in our paper, several advanced or hybrid approaches between different classifiers are discussed below. These mostly aim to increase classification accuracy for specific use cases. Furthermore, approaches for visualisation and interpretation of more complex data features are described in the following literature. Here we should mention again that the objective of our work is on a classifier makes understanding how internal decisions are made rather than on the optimisation of the classification performance. This is also the fundamental difference to the hybrid classification approaches presented in the following research.

Zhao et al. [33] introduced an approach which applied *Structural Similarity Index Measure* (SSIM) metrics for comparison and feature extraction in the textile industry. Joshi et al. [34] investigated hybrid SVM – ANN approaches for the inspection of tiny parts by examining four different hybrid classifier designs. In their setup the SVM is directly integrated in the ANN classifier. Concluding, they have reached a classification accuracy up to 95% with a *Supervised Artificial Neural Networks – Unsupervised Support Vector Machine* (SANN-USVM) classifier architecture. Additionally, in their study they performed a feature assessment via *parallel coordinates*. Malaca et al. [35] carried out investigations of systems for textile inspection in the automotive industries. Here they evaluated different CNN and SVM configurations. For this, they also compared different features extraction methods more closely. Basly et al. [36] developed a hybrid CNN – SVM classifier approach for human gesture recognition. They first used a CNN to extract higher dimensional image features. These high level features were then forwarded to a SVM. In their study, they achieved an accuracy of 99.9% with their combined CNN – SVM assembly. This classification rate is even better than for the additionally examined hybrid classifiers of CNN + LSTM or CNN + *Multi-Layer Perceptron* (MLP). Xu et al. [37] introduced a hybrid CNN – SVM classifier approach for bearing fault diagnosis. Similar to Basly et al. [36] they utilised the CNN for high level feature extraction and a subsequently attached SVM for precise classification. As input for their combined classifier they applied a sequence of 1D vibration signals which they transformed into a 2D image. Furthermore, they used t-SNE plots for visualising the complex features from the CNN output. Such t-SNE visualisation of high dimensional data sets was introduced by van der Maaten and Hinton [38]. Based on the Euclidean distances between high dimensional data points, their method assigns a probability to a data point that describes its similarity to its neighbours. These probabilities serve to assign costs to each point, which are then plotted according to their importance in relation to their neighbouring data points. Lee et al. [39] followed a very similar approach as Basly et al. [36]. For different scenarios in manufacturing they recorded a 1-D dimensional input signal, which is utilised as input for different ANN classifiers. These classifiers extract complex signal features, which are then visualised via t-SNE plots. Finally, these complex features serve as input for a one-class SVM for different classification scenarios. Mohamed et al. [40] also applied a combined CNN – SVM design, where the CNN serves for feature extraction and the SVM for classification. They have tested

their approach for the classification of everyday objects with 98.5% accuracy. Sun et al. [41] used a combined SVM – CNN for the sensor data fusion and subsequent classification of multi sensor data for vehicle navigation. They first applied a SVM for data fusion and then a connected CNN for the decision-making. Long et al. [42] explained a hybrid deep learning structure consisting of an autoencoder and a SVM, where the autoencoder served for extracting more complex features and multiple SVM for fault classification. Zhang et al. [43] described a compact CNN for fault diagnosis, which first extracted more complex features, but then reduces the network dimension rapidly. Meng et al. [44] made use of a CNN based approach for the recognition of textile patterns. In particular, they applied SSIM and *Mean Squared Error* (MSE) estimates as a loss function for their CNN. In contrast, Lee et al. [45] compare different xAI methods to visualise characteristics of a categorised defect for domain experts in the field of display panel inspection. In their research, they applied the *Layer-wise Relevance Propagation* (LRP) method as most promising for their use case. Furthermore, they passed the class probabilities belonging to each class from their pre-trained classifier to a DT. From this, they derived a set of probably human understandable rules for a classification decision. Finally, they questioned domain experts about the benefit and comprehensibility of their derived rules and visualisations in order to enable these experts to understand the machine's decision. The overall focus of their study is on the preparation of classification information for a domain expert.

In the following section, procedures for synthetically generating a sufficiently large and balanced training and test data set are outlined.

2.4. Synthetic image data generation

Various methods are available to generate synthetic data for inspection. Yun et al. [46], for instance, used autoencoder-based methods for the application in metal surface inspection. In our related study from Meister et al. [9], we have carried out detailed investigations on the artificial generation of fibre layup defects. In this former research we utilised a conditional *Deep Convolutional Generative Adversarial Network* (DCGAN), which is an ANN based synthesis technique to synthesise LLSS depth maps of AFP layup defects. The respective DCGAN architecture is designed on the basis of concepts from the literature. Furthermore, parameters with an expected major impact on the outcome are examined. These are varied in a reasonable range of values for this scenario [47, 48]. Radford et al. [47] presented the default setup of the DCGAN. Starting from this, certain settings were varied. The mentioned parameters have been similarly used from Radford et al. [47] and Brownlee [48] for the design of a robust DCGAN. Accordingly, in our former paper from Meister et al. [9], three preliminary tests were considered to identify suitable test settings for the batch size, layer structure and DCGAN parameters. For training the DCGAN, 5000 images were utilised. This amount of training data is based on the literature as discussed in their study. Respectively the applied DCGAN parameters are set as follows: Batch size = 64, Layer structure (Generator/Discriminator) = (6/5), Learning rate = 0.0001, $\beta_1 = 0.5$, $\beta_2 = 0.999$, dropout = 0. The DCGAN parametrised in this way can be utilised as a data generator for height profile images of fibre placement defects, which are analysed in this study. Furthermore, we have described in this former research that, depending on the defect class, between 25 and 47 defect samples are sufficient to represent an original fibre layup defect within a CNN.

Subsequently, typical feature extraction and selection methods for such images are discussed.

2.5. Common feature extraction and selection techniques from the literature

This section presents an overview of different, commonly used feature extraction and selection methods from the literature.

2.5.1. Feature extraction methods

The literature provides various methods for the manual extraction of suitable features from images. Fundamentally, these can be divided into two groups. These are discussed below, where the methods' or developers' names as well as the corresponding reference are given. In addition, the individual methods are subdivided into categories according to their functional principle [49]. Furthermore, the resulting amount of features per method is given. Finally, the operating principle of the approaches are briefly described.

On the one hand, these are features that describe globally the texture of an entire input image. These features use few attributes to describe a full image. An overview of common methods of these category from the literature is given in Table 2. Texture features can be divided into seven categories. Statistical methods use statistical descriptors to express the distribution of a feature in an image. In structural methods, the texture of an image is defined as an assembly of repeating small patterns. In model based techniques, textures are characterised via mathematical

Table 2
Overview of texture sensitive image feature extraction techniques from the literature. Gravity: GV, Model: MO, Entropy: E, Machine Learning: ML, Statistics: ST, Spectrum: SP, Graph: GP.

Method	Ref.	Cat.	Dim.	Operating principle
STD per Cell	[15]	ST	1	Standard Deviation of pixel values per grid cell
de Mesquita &Backes	[50]	GV	92	Decay description through lacunarity and fractal geometry
de Mesquita	[51]	GV	>500	Decay description through lacunarity
Francos et al.	[52]	Mo	10-135	Superposition of three random fields
DistrEn2D	[53]	E	1	Shannon's entropy of two-dimensional distributions
Cimpoi et al.	[54]	ML	–	Convolutional layers of a CNN acting as a filter bank
de Mesquita	[55]	ML	30-180	Weights of ELM hidden layers
HGM	[56]	GV	16	Normalised histogram of gradient magnitudes
GLC	[57]	ST	3-10	Two-dimensional *long correlation* autocorrelation function
Wu &Wei	[58]	ST, SP	–	Autocorrelation functions in one dimension of subbands from a helically sampled input image
Backes et al.	[59]	MO, GP	5	Descriptors which characterise the topology of a *Complex Network*
SPG	[60]	GP	20	Standard deviation + mean along the shortest paths through grid cells
Backes et al.	[61]	GP	2-6	Entropy and statistical descriptors of deterministic graphs
Thewsuwan &Horio	[62]	GP	30-81	Rotational invariant evolution of *LBP* applied to *Complex Networks*
ELGS	[63]	GP	4	Histograms from graph-based binary descriptors
2D-MA	[64]	MO	16	Texture pattern approximation in the frequency domain
SampEn2D	[65]	E	1	Similarity analysis of a reference region to neighbour areas
GLCM	[66]	ST	28	Statistical descriptors of grey value matrix
LBP	[67]	ST	–	Local patterns are coded. Their distribution is given as histogram
Tamura et al.	[68]	ST	6	Global statistical descriptors
Zhang et al.	[69]	SP	2	Descriptors from spectral decomposition and gradient orientation
Maani et al.	[70]	SP	3	Two-dimensional spectral analysis of the frequency components
Riaz et al.	[71]	SP	48	Spectral analysis of sorted Gabor filters
Manjunath	[72]	SP	48	Position and statistical descriptors of Gabor filter amplitudes
MRIR	[73]	SP	–	Wavelet decomposition of an image. Descriptors from multiple spatial resolutions

models. This approach is similar to the structure based methods. Transformation based procedures map the textures contained in an image into an image area. These methods utilise two-dimensional discrete Fourier analyses, filters, wavelet transforms or decompositions. Descriptors are calculated from this image area, which serve as features. For the generation of features via ML algorithms, such ML techniques are initially applied for calculating an abstract feature vector. Approaches based on graph theory define a graph in which the nodes correspond to the image pixels. The features are the parameters which characterise this graph. In entropy based methods, the texture is specified through a combination of statistical parameters and entropy from information theory [49].

On the other hand, features which analyse local areas of an image and extract corresponding information can be considered. In this case, the actual information consists of the feature description at a certain image position and the connection of several analysis locations of the feature. The expression of an individual feature in a feature vector can be traced back to its respective analysis position. An overview of such usual, local features is presented in Table 3[74]. Local features can be divided into two categories. These refer to features that utilise brightness gradients in the image to describe the feature or directly the image intensity in a certain image area. Depending on the method, this description is carried out depending on the gradient rotation or linked to an image sequence [75–77].

Especially the invariance of a feature to brightness differences, position variations and rotations are desired properties of such a feature [83]. The priority of the individual invariances obviously strongly depends on the application case.

2.5.2. Feature selection approaches

In this section, several methods for evaluating the importance of individual features in a feature vector are presented. These approaches are designed to select the features that are most beneficial for a certain use case. In this respect, it should be noted that many rule- and model-based classifiers require significantly less training data or yield a better performance if they are trained with feature vectors that contain only few but meaningful features [84]. Obviously, this also indicates that the importance of individual features and their influence on the classification rate might depend on the considered classifier. For this reason, a suitable feature selection is necessary.

In the paper of Li et al. [85] they provided a comprehensive summery of various feature selection methods and their detailed operating principles. Jovic et al. [86] presented different feature selection methods and described their performance for different data sets and application cases. Sheikhpour et al. [87] give a hierarchical structured survey of semi-supervised feature selection techniques. In their study, they basically divided the feature selection procedures into filter, wrapper and embedded methods. This categorisation is also carried out by Zhang et al. [88]. In addition, they distinguish the methods according to the availability of label information on the data under consideration. Accordingly, for an in-depth overview we refer to these four papers. However, in Table 4, very common feature selection methods are listed and the corresponding references are given. The techniques are categorised with respect to their operating principle which is also briefly described. A very promising way to evaluate the influence of individual features on the response of a classifier is the *Analysis of variance* (ANOVA) method. This method analyses the mean value of the features of one category compared to the mean values of the other categories. Then this technique looks for the feature combinations which separate the different categories in the best way. To achieve this, a statistical hypothesis F-test is conducted, which compares the variance of the values within a certain class with the variances between different classes. This is represented as a F-value, whereby a larger F-value indicates a better distinguishability between the classes [103]. In the following, different classifiers are discussed, which potentially make use of such selected features.

2.6. Fundamentals of classifiers

2.6.1. Support Vector Machine

The SVM classifier can be applied to a wide range of applications with excellent classification results [104]. This approach belongs to the model-based supervised learning methods. Basically, a SVM operates in such a way that in a preliminary training process, pre-generated feature vectors for different classes are placed in a corresponding vector space. Subsequently, based on these training vectors, a hyperplane is aligned in this vector space, which separates the pre-trained feature vectors of the different training classes from each other [105]. This is schematically illustrated for a 2D soft margin SVM classifier in Fig. 3, where ξ_i

Table 3
Overview of locally operating image feature extraction techniques from the literature. Key point: KP. Some features create *n* dimensional features per KP, where the number of KP can be chosen as desired.

Method	Ref.	Cat.	Dim.	Operating principle
CD-HOG	[75]	Gradients	16/ KP	Background detection from image sequence, HOG only on foreground image regions
LMGO-HOG	[75]	Gradients	16/ KP	Directional weighting of the histogram values of an image area and thus noise tolerant
BIG-OH	[76]	Gradients	128/ KP	Binary encoded magnitude of oriented gradients, robust for image deviations
MROGH	[77]	Intensity	192	Descriptors are calculated from gradients at the key point using a rotation invariant coordinate system
MRRID	[77]	Intensity	256	Similarly to MROGH, descriptors directly characterise the intensity surrounding the KP
SIFT	[78]	Gradients	128/ KP	Scale invariant, rotation sensitive KPs
PCA-SIFT	[79]	Gradients	20/ KP	Reduce SIFT feature vector via PCA
SURF	[80]	Gradients	64/ KP	Fast, noise tolerant KPs from Haar wavelet descriptors
HOG	[81]	Gradients	9/KP	Histogram of gradient magnitudes and orientations
CS-LBP	[82]	Intensity	72-256	Position resolved combination of SIFT and LBP

Table 4
Summary of very common feature selection methods from the literature. Li et al. [85] give a comprehensive overview of feature selection methods. Information Theory: InfT; Statistics: ST; Brute Force: BF; Similarity; SI.

Method	Ref.	Cat.	Operating principle
Information Gain	[89]	InfT	Sorts features according to their transinformation
CHI^2	[90, 91]	ST	Chi-Square Test is used to evaluate the independence of individual features
Variance score	[92, 93]	ST	Euclidean distance between results from two different feature selection methods
Analysis of Variance (ANOVA)	[94, 95]	ST	Analyses the mean of two different feature vectors via F-test
Correlation Feature Selection	[96, 97]	ST	Searches subset of features which have low correlation among each other but high correlation to a class.
Drop Column feature importance	[98]	BF	Removes features from the feature vector sequentially and evaluates the performance of the classifier
Feature Importance Ranking Measure	[99]	ST	Analyses the variance of the conditional expected value of a classifier
Low variance	[100]	ST	Removes features with low variance to each other
Group Lasso	[101]	ST	Analysis different feature clusters and their relationship
Laplacian Score	[102]	ST	Evaluates geometric relationships between features

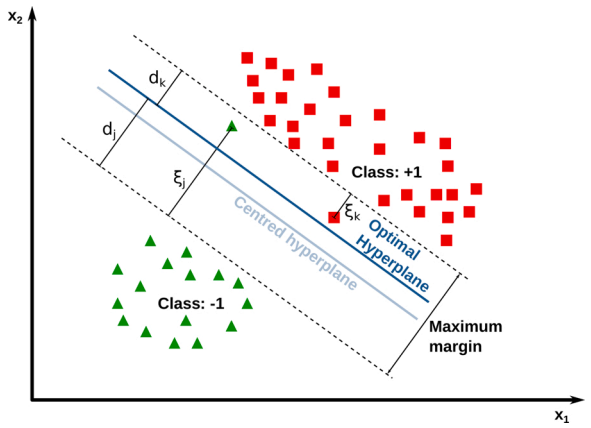

Fig. 3. As example of a non-linearly separable, two-dimensional, two-class classification problem is presented. ξ_j (class: -1) and ξ_k (class: $+1$) are the slack variables introduced for solvability of the problem. d_j (class: -1) and d_k (class: $+1$) are the distances between each class boundary and the linear hyperplane.

represents a soft margin slack parameter, which equals zero for hard margin classifiers. Moreover, the parameter d represents the Euclidean distance between the separating hyperplane and the respective class boundaries. This technique is well known. Hence, we refer to Abe [105] as well as Chang and Lin [106] for a detailed description of the mathematical principles. Beyond this, it should be mentioned, that the separating hyperplane can have any shape, if is not limited to the linear case shown above.

2.6.2. Convolutional Neural Network

Due to their inherent structure, CNN techniques are often used for the classification of image data. By using kernels, a CNN can substantially decrease the number of training parameters required. This network structure analyses individual regions of the input image incrementally. The number of trainable parameters correlates with the amount and size of the implemented kernels. Thus, less weights need to be trained than for an ANN with no kernels. This leads to increased efficiency of the classifier [107,108].

In general, a CNN is an ANN, which determines the individual image features by convolution calculations with different sized convolution matrices. This approach consists of several feed-forward connected convolution and subsampling layers. Characteristic data features are extracted from the input data as it passes through the individual CNN layers. The complexity of the determined features typically increases with the number of layers. A convolution layer takes advantage of multiple kernels to generate a suitable number of feature maps from the input data. Each kernel contains an individual set of weights, which is trained. Through the use of different kernels, several feature maps are generated. Finally, a fully connected layer is installed to generate the desired output. In case a classification task is to be solved in image processing by means of CNN, the result of the CNN is the corresponding class affiliation of the image content [107,108].

In order to guide the selection of an appropriate ANN architecture and parametrisation, Shrestha and Mahmood [31] gave a comprehensive review on several ANN setups, their corresponding application cases and their characteristics. In particular, for the inspection and classification of fibre layup defects in the AFP process, Chen et al. [109] outlined a suitable approach.

In this regard, below, a method for assessing the relevance of individual image areas for a particular ANN decision is discussed.

2.7. xAI methods theory

This section summarises briefly available xAI methods from the literature. Then the basics of the applied xAI procedure and the corresponding evaluation metrics are described. For this purpose, the operating principle of the *Smooth IG* algorithm is first discussed and then the metrics *Maximum Sensitivity* (SenseMAX) and *Infidelity* (INFD) are explained.

2.7.1. Overview xAI methods

In this section, a brief overview of common xAI procedures from the literature is given. Following Müller [110], popular algorithms are listed in Table 5. Since the aim of this paper is not a detailed analysis of the various xAI methods, only the most common methods are listed here. Due to the particularly advantageous properties for defect inspection in composites as well as the beneficial smoothing characteristics and the applied reference, the *Smooth IG* algorithm is explained in detail in the following section.

2.7.2. Smooth Integrated Gradients

The *Smooth IG* approach is a combination of *Integrated Gradients* (IG) with an additional smoothing step. With respect to IG, for each input neuron the gradients of the neuronal activations along the paths within the neural network to the corresponding output neuron are calculated. These calculations serve as a decision criterion for the *Smooth IG* method, when compared with a reference activation.

For this we define: $F: R^n \rightarrow [0, 1]$ as the transfer function of an ANN. The input image is described by $x \in R^n$. $x' \in R^n$ gives a corresponding reference dataset. α denotes a setting parameter for the straight path from x' to x. Here c is the class of the input image x, where x_i specifies the pixel at position i of the image.

Thus, IG for x_i are calculated as follows:

$$\mathrm{IG}_i(x) = (x_i - x_i^{'}) \cdot \int_{\alpha=0}^{1} \frac{\partial F(x^{'} + \alpha \cdot (x - x^{'}))}{\partial x_i} d\alpha \tag{1}$$

Where $\frac{\partial F(x)}{\partial x_i}$ is the gradient of $F(x)$ along the i-th dimension [112]. Consequently, this equals the derivation of the ANN along the path from a given input neuron of a pixel to the corresponding output neuron of a class c. When smoothing is applied to the resulting pixel-based IG values, this is called *Smooth IG*. This smoothing is performed in accordance with Smilkov et al. [113] through adding normally distributed noise to the input dataset. Due to the integration and smoothing in this xAI procedure, the noise in the output is significantly reduced, which is beneficial. Moreover, this approach utilises identical reference across all classes of the classifier, which is a great advantages of this procedure in terms of comparison of the results [113].

2.7.3. Quality metrics for xAI results

Yeh et al. [123] in particular suggest two quantitative criteria for

Table 5
Overview of approaches for the visualisation of relevant pixels for an ANN prediction. This overview is based on Müller [110]. The name and the corresponding reference are given. These are common procedures from the reference mentioned before.

Algorithm	Reference
Deep Learning Important Features	[111]
Input * Gradient	[111]
Integrated Gradients	[112]
Smooth Integrated Gradients	[112,113]
Kernel Shapley Additive Explanations	[114]
Occlusion	[115]
Deconvnet	[115]
Guided Back Propagation	[116]
(Guided) Gradient Class Activation Mapping	[117]
Layer-wise Relevance Propagation	[118]
Saliency Maps	[119]
Local Interpretable Model-Agnostic Explanations	[120]
Meaningful Perturbations	[121]
PatternLRP	[122]

comparing the performance of xAI techniques for ANN. These are the *Explanation Sensitivity* and INFD.

For the following explanations I_m denotes the input image and R_m gives a suitable reference each for the m-th defect image. Whereas ϕ_f specifies the relevance of a certain pixel for the classification model f.

2.7.3.1. Maximum sensitivity metric. The metric *Explanation Sensitivity* quantitatively expresses the sensitivity of a method for infinitesimally small changes in the input data set. In this approach, Yeh et al. [123] first applied infinitesimally small changes to the input image under consideration. Then, this criterion was calculated from the normalised difference of the explainability results for this manipulated dataset and the explainability results for a reference dataset. Yeh et al. specified several variants for the calculation of this score. A commonly used variant is the SenseMAX calculation. This is inherently upper bounded and gives the maximum sensitivity of a method to perturbations. This SenseMAX criterion is defined as follows:

$$S_{\max}(\phi_f, I_m, R_m) = \begin{array}{l} \max \parallel \phi_f(R_m) - \phi_f(I_m) \parallel, \\ \text{with} \quad \parallel R_m - I_m \parallel \leq r \end{array} \tag{2}$$

The parameter r describes an adjustable range of values. The absolute $\|\cdots\|$ in this case is calculated according to the L_2 norm [123].

2.7.3.2. Infidelity metric. With reference to Yeh et al. [123], for large feature spaces, the INFD metric quantifies the correlation of an explanation with the model of an ANN. Thus, for an observed input image, the criterion expresses the importance of individual image pixels with respect to the behaviour of the ANN model. Consequently, this INFD metric can be formulated as the following expected value:

$$\mathrm{INFD}(\phi_f, f, I_m, R_m) = \mathrm{IE}_{R_m \sim \mu}\left[\left(R_m^T \phi_f(I_m) - (f(I_m) - f(I_m - R_m))\right)^2\right] \tag{3}$$

The reference in this scenario is given as:

$$R_m = I_m - X_0 \tag{4}$$

For this, X_0 is a random variable with the probability distribution μ. The expected value is estimated via a Monte-Carlo simulation [123]. In this context, it should be mentioned that also other references are conceivable for those calculations.

3. Methodology

The specific choice of methods and the experimental design for this paper is explained in this section.

3.1. Data acquisition and utilised image data sets

Suitable defect types were selected for the investigations in this study. In accordance with the defects presented in Section 2.1, *none*, *wrinkles*, *twists*, *foreign bodies*, *overlaps* and *gaps* were chosen for the subsequent experiments. Fig. 2 shows these defects schematically.

To investigate a sufficiently large and balanced training and test data set, a data set with 50 height profile defect images per class was synthetically generated. For six classes as well as one training and one test data set each, a total of 600 synthetic defect images were considered for the examinations in this paper. The generated images were randomly assigned to one of the two different test and training data sets. For this purpose, our DCGAN approach from our previous research [9] was applied. This procedure is described in detail in Section 2.4. The images generated in this way were used as input for the parallel classification architecture outlined below.

In our former research we also determined that, between 25 and 47 defect samples per defect class are sufficient to represent the mentioned defect types for the utilisation of the CNN [9]. This is the reason why 50 test samples per class were chosen for this work. Due to the simple SVM structure and a small feature set in this study, we assume that 50 input images per class are also sufficient for training and testing the SVM. The use of a larger amount of synthetic image data therefore hardly changes the core statement of the conducted investigations, but potentially increases the computational effort. Mixing synthetic and real data or synthetic data from different data generators will probably lead to inconclusive results when analysing the feature vectors as input for the SVM. Especially in order to be able to analyse these extended data sets sensibly with the SVM, significantly more complex features must be applied. In this paper, however, we focus on the basic investigation of the procedures described below. Various, significantly larger or manipulated data sets are therefore not considered in this paper.

3.2. Parallel classifier design

This section explains the structure of the parallel classifier. This setup is intended to perform the redundant classification of an input defect image with the two rather diverse classifiers CNN and SVM. This setup is illustrated schematically in Fig. 4.

It is aimed at comparing different intermediate results from both classifiers to determine suitable properties and parameters of the classifiers for a trustworthy and robust classification of fibre layup defects. For clarification, the aim of this paper is mainly the identification of suitable metrics to analyse the uncertainty of a classification, the origin of a machine decision and a potential behaviour for an unknown classification case.

For this reason, on the one hand, exemplary input images of the entire image processing chain were analysed for each defect type with respect to the output of the first joined convolutional layer. This procedure is described in detail in Section 3.5.4. Furthermore, the xAI calculation described in Section 3.5.3 was applied to the CNN in order to evaluate the relevance of individual image areas on the decision of the CNN and link them to the manually selected features. These results were also projected onto the exemplary input image. The associated methodology is described in Section 3.5.5, where the selected features and their influence on the informativeness of the entire feature vector were examined. The respective methodology is outlined in Section 3.5.1. Finally, the classification results of the CNN and SVM classifiers were compared and suitable indicators for assessing the classification robustness and confidence were investigated. This methodology is outlined in Section 3.5.2.

3.3. Feature vector composition

This section describes the calculation of suitable image features as well as the selection of a certain number of particularly advantageous image features. These serve as input for the training and usage of the SVM classifier in a subsequent stage.

3.3.1. Feature extraction

Since the origin of an image feature should later be traced back to its position in the input image, feature extraction was applied to the cells of an image spanning grid. For this, the input image was split into $4 \times 4 = 16$ square cells of equal size. The feature extraction algorithms were computed for each cell individually and then concatenated. The feature calculation was carried out in order, from left to right and from top to bottom. This means that an entire grid row was passed through before the calculations for the next row began.

Following the theoretical discussion in Section 2.5.1, several methods are suitable for the feature extraction. These primarily differ in describing the specific individual features in the image or in providing an indication value for the texture of a larger image area. For the analysis in this paper it is essential not to compute too many features per grid cell of the image. Furthermore, these features need to be linked to the image attributes of a cell in a human understandable way. This means in particular that features must be selected which can be traced

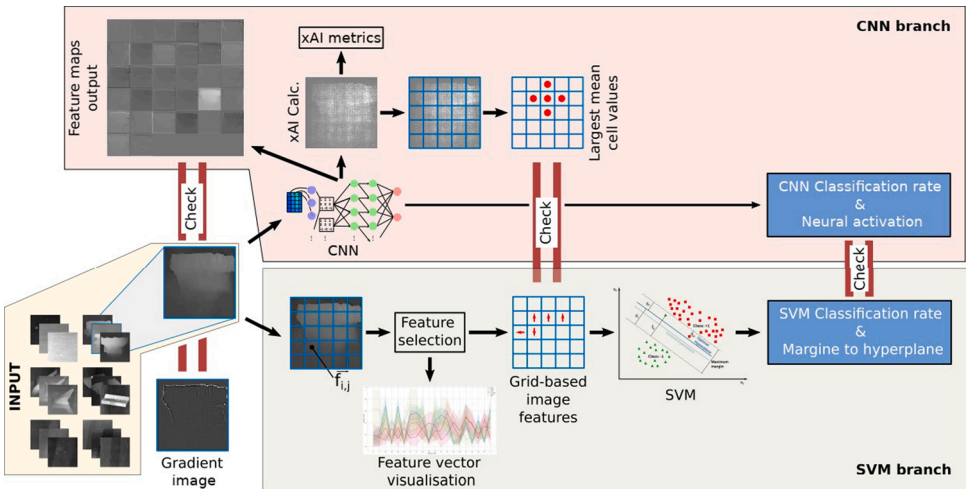

Fig. 4. The parallel classification approach is visualised which uses a SVM and CNN classifier. Certain intermediate stages for comparing the different classification processes are highlighted.

back to physical attributes in the height profile input image. These are features that can be recognised visually in an image without a major transfer operation. For this reason, relatively simple *Histogram of Oriented Gradients* (HOG) was used for the investigations in this paper to describe and visualise the edges of a defect. Moreover, the *Standard deviation* (STD) of each grid cell was used as an indicator for the homogeneity of the texture. This is particularly suitable for the application case outlined in this study, as this calculation only determines a single value per grid cell, in contrast to most other methods from Table 2. This might reduce the performance, but can very well be linked to the image area. In addition, the application of this feature reduces the length of the total feature vector compared to the application of alternative texture features. The methods DistrEn2D and SampEn2D from Table 2 also calculate just a single value to describe the texture. However, both methods are based on ML approaches and thus significantly increase the complexity of this feature extraction step. Furthermore, comprehending the origin of the associated feature values is quite challenging. Obviously, more complex features can also be handled with this introduced grid structure and traced back to the position in the image. However, their expression is often harder to relate to the geometric attributes of the defect. Thus, for the exemplary investigations in this study, only the described, very simple features were used. In the following, the procedure for the targeted selection of most informative features from the entire feature set is explained.

3.3.2. Feature selection

As described in the Section 2.5.2, various methods are available to carry out the feature selection. The difference between these approaches is in their functionality but also in their complexity. Since the goal of this paper is the traceability and uncertainty estimation of a machine decision, the added extra complexity in this step should be kept as low as possible. Accordingly, a simple but established and efficient feature selection procedure is recommended. For this reason, the Variance Score from Table 4 was applied as a selection criterion in this study. Specifically, the ANOVA method was used for this purpose. This is well suited for the partly well separating and partly very matching feature vector values due to its analysis of variance. The features selected in this way served as input for the SVM configuration described below.

3.4. Classifier configuration

This section outlines the configuration of each classifier applied to the conducted experiments. For this purpose, the SVM setup is described first and then the CNN setting is discussed.

3.4.1. SVM setting

This section describes the experimental setup for performing the defect classification experiments using SVM. These experiments served as input for the redundant defect classification. For this reason, a SVM with linear kernel was applied for the experiments, as already theoretically outlined in Section 2.6.1. This linear kernel has the advantage of outputting additional distance parameters which serve as robustness indicators for the classification decision besides the actual classification result. This advantage was also the reason for the selection of a linear kernel in our SVM setup. In the conducted experiments, these margin values of the individual test feature vectors to the trained hyperplane were calculated and the robustness of the respective SVM classification was evaluated. The required SVM parameter C was set to $C = 1$. The γ parameter was estimated automatically via internal cross validation over the training data set. This automatic determination of the parameter offers the possibility to determine the setting that yields the highest overall classification score on the basis of a training data set. The resulting parameter used in our experiments is $\gamma = 0.05$. As stated before, 50 different synthetic training and test samples were applied for each of the six defect categories considered. The training and test images were different from each other.

3.4.2. CNN setting

In this study, no pre-trained ANN were utilised for the experiments. As discussed in our former related study from Meister et al. [9] a pre-trained ANN offer no significant performance boost for the investigated scenario. Accordingly, a self-trained ANN was applied for the experiments in this paper.

The CNN applied in this study was based on the preliminary research of Chen et al. [109] and has already been used in our former research [9]. Both studies have already validated its functionality successfully for the visual defect recognition in the AFP manufacturing process. Table 6 shows the detailed architecture of this classifying CNN.

This consists of 18 layers in total, where the input is a 128×128 image matrix and the output is correspondingly the classification decision for one of the six defect categories. This result is based on the maximum neuronal activation of the CNN at the output of layer #16.

3.5. Experimental setup

The tests carried out in this paper were based on the individual comparison stages from the parallel classification approach introduced in Section 3.2. The required setup of the individual components has already been described above.

Table 6
Architecture of the classifying CNN for an 128 x 128 px input image, considering six different classes for categorisation. Total parameters: 1607686; Trainable parameters: 1606022; Non-trainable parameters: 1664.

#	Layer type	Output dimension
0	Input	None × 128 × 128 × 1
1	Convolution 2D	None × 64 × 64 × 32
2	Convolution 2D	None × 16 × 16 × 64
3	Max pooling	None × 16 × 16 × 64
4	ReLu	None × 16 × 16 × 64
5	Batch normalisation	None × 16 × 16 × 64
6	Convolution 2D	None × 8 ×8 × 128
7	Convolution 2D	None × 4 ×4 × 256
8	Max pooling	None × 2 ×2 × 256
9	ReLu	None × 2 ×2 × 256
10	Batch normalisation	None × 2 ×2 × 256
11	Flatten layer	None × 1024
12	Dense layer	None × 512
13	Dropout	None × 512
14	ReLu	None × 512
15	Batch normalisation	None × 512
16	Dense layer	None × 6
17	Softmax	None × 6

In all tests which served as input for the SVM or yielded their output, the two variants with 10 and 20 selected features were always considered. Both sets of features can still be visualised well, which is the reason why this number of features was chosen. Thus, the selection of even more features from the feature extraction does not seem reasonable, as these cannot be sensibly visualised with certain procedures and hence an interesting aspect of the analyses cannot be evaluated in this study. As already described, the analyses in this paper were not aimed at optimising the classification rate, but at the capability of interpreting the intermediate stages.

3.5.1. Feature vector analysis procedure

In this study the image features were first examined directly. They served as the input data for the SVM classifier. As already described in Section 2.5.1, robust image features are characterised by the fact that the individual feature values enable a reliable subdivision of the individual classes. Thus, the feature values per feature index should be strongly different for each of the considered defect categories. Therefore, first the statistical representation of the individual features through a t-SNE calculation was chosen in order to assess the suitable delimitation of the features in the multidimensional feature space more precisely [124,125]. Robustly distinguishable features appear as separate clusters in this t-SNE representation. This t-SNE plots first allow a qualitative evaluation of the calculated and applied image features for a robust defect classification [37,39].

Secondly, the normalised feature values of the selected feature vector were displayed over the selected feature indices. Therefore, the entire feature vector was normalised beforehand to the value range $[0, 1]$. The feature indices were presented from left to right, in the order of their estimated performance. These plots of *parallel coordinates* provide a direct way of analysing the individual feature values without any intermediate projection step [34]. Thus, the performance of the feature selection as well as the performance of individual features and their robustness can be assessed directly.

3.5.2. Investigating the classifier performances

Previously the ability of the individual features for distinguishing the considered defect classes was examined. In this section, the actual classification behaviour of the introduced SVM and CNN classifiers were investigated. For this purpose, the classification result of both approaches for the entire test data set were first evaluated. Hence, the corresponding classification rates from the SVM and CNN classifiers were assigned to each ground truth class in a confusion matrix. Thus, for the applied test data set, the actual classification scores with their corresponding misclassifications were presented.

In order to additionally assess the SVMs robustness of a trained class to possible changes in the input features, the distances of the test feature vectors to the linear separating hyperplane of the SVM were calculated. The associated mean values and STDs were again presented in a confusion matrix to visualise the correlations between individual classes. This also demonstrates the great advantage of the linear kernel of the SVM. This kernel enables this distance determination without much effort, due to the inherent parameters of the classifier or the associated model, as theoretically explained in Section 2.6.1.

With the aim of investigating the robustness of the CNN classifier, the respective measured mean values and STDs of the neuronal activations at the last hidden layer #16 of the classifier were presented for all images of the test data set and the mean values and STDs were again displayed in a confusion matrix. Please note that theoretically each respective neuron can be activated in the range from 0% to 100%, independently of the other activation scores. The mean magnitude of the activation of a defect class in relation to the neuronal activations of the remaining classes gave a measure for the robustness of the classification outcomes for a certain defect type. The STD, on the other hand, indicated a measure of the varying CNN model response and thus, varying activations of the output neurons, due to a different appearance of the individual input defect images. Additionally, from these values, a bounded 95.5% confidence interval was derived and analysed. Based on the standard deviation across all neuronal activations of the test data set, this interval describes the range of values per output neuron in which 95.5% of all percentage activation values occur. Bounded means in this context that the interval does not exceed the minimum and maximum range of activations.

3.5.3. Sensitivity and fidelity evaluation

This section explains the sensitivity and fidelity analysis of the CNN model. Therefore, the SenseMAX and INFD measures indicate the sensitivity and fidelity of the CNN model with respect to the CNN response for the utilised test data set.

The SenseMAX values were determined in accordance with Equation (2). Following Yeh et al. [123], 50 Monte-Carlo calulations with Gaussian noise having a STD of $\sigma = 0.2$ were performed accordingly. In addition, a noisy baseline was used for the SenseMAX calculations.

Equation (3) specifies the calculation of the INFD values. For this purpose, after Yeh et al. [123], 1000 Monte-Carlo calculations were performed.

3.5.4. Input feature map analysis and visualisation

In order to review the input of the CNN, the calculated feature maps of the first joined convolutional layer #2 were compared with the geometric edge attributes of the input image. This comparison is feasible since the first joined convolutional layer in the CNN maps edges in the input image through a mathematical convolution. Respectively, a 3×3 kernel Sobel filter was applied to extract the gradients of the input image. The same Sobel filter was then applied to the individual feature maps of the first convolutional layer. This is necessary since the feature maps contain the image features themselves as well as their neural activations. Both gradient images are then compared with each other using the SSIM algorithm. This calculated a similarity index for the pairs of each feature map and the input gradient image. The similarity indexes were presented along with the actual neuronal activation of the individual feature maps. The matrix data order corresponds to the matrix representation of the individual feature maps of the first joined convolutional layer. The SSIM method is particularly well suited for the comparison of the individual images, as it does not require more complex auxiliary models and provides a simple measure of the match between two images. This is also indicated by the research of Zhao et al. [33] and Meng et al. [44].

3.5.5. Examining the matching between selected features and xAI relevance of individual image regions

In this section, the comparison of the *Smooth IG* xAI results with the selected image features is described. First the original input images and the *Smooth IG* output images were separated into 4×4 equally sized grid cells. As described above, the HOG and STD values from Table 3 were calculated for each grid cell of the original input image. In the subsequent feature selection, the most promising features were selected from this full set of features. For the *Smooth IG* output images, the mean values for each grid cell were computed. This denotes the averaged importance of a grid cell based on the xAI estimation. Subsequently, the grid positions of the manually selected image features and the most relevant grid cells from the xAI calculation were compared and displayed. This comparison is carried out each for one example defect per defect category.

4. Results

In this section, the results of the previously introduced experiments are presented and discussed.

4.1. Feature vector analysis

This section presents the results of the image feature analysis. The calculated t-SNE outcomes in Fig. 5 are considered first. The x_1 and x_2 axes denote the respective projection axes. The coloured dots represent the feature vectors of the test dataset for each class, projected onto the 2D plane. The respective initial vector space is 10 or 20 dimensions, corresponding to the number of selected features.

For the t-SNE plot for 10 selected image features in Fig. 5a, a quite clear separation of *wrinkles* and *foreign bodies* from the remaining classes is evident. Nevertheless, we realise a few intersections of the clusters from both classes. The respective clusters of the classes *overlaps* and *gaps* almost completely intersect. The clusters of the *none* and *twist* classes are very close to the accumulated *overlap – gap* cluster. For these two classes, some projection points are particularly striking here, since they are located within other clusters.

For the t-SNE plot in Fig. 5b covering a total of 20 selected features, a slightly more compact aggregation of the *none* and *twist* class is noticeable. All remaining classes have a similar projection behaviour as for 10 selected features. Please note that the 10 features of the t-SNE plot in Fig. 5a are fully incorporated in the 20 features of the t-SNE plot in Fig. 5b.

After analysing the t-SNE plots, using 20 instead of 10 features should mainly reduce the misclassification between *foreign bodies* and *wrinkles* as well as between *twists* and the accumulated *overlap – gap* cluster.

For a more detailed investigation of the influences of the individual features, the respective mean feature values and the associated STDs of the selected feature indices are examined in the format of a *parallel coordinates* plot in Fig. 6. The feature values are normalised between $[0, 1]$ and are given on the ordinate. The respective feature indexes are on the horizontal axis. The solid lines represent the normalise feature mean value for each defect type. The coloured bands give the respective STDs. The features associated with the feature indices are specified in Table 7. The features are arranged chronologically on the horizontal axis from left to right, according to their importance for the classification assigned via the ANOVA feature selection algorithm. The respective ANOVA F-values and p-values are given in the two right columns, where the critical F-value for the significance level $\alpha = 5\%$ is $F_{crit}(\alpha = 5\ \%) = 1.26$.

From the ANOVA F-values from Table 7 we recognise, that the first three values indexes have F-values >100. Furthermore, it can be seen that the F-value decreases from the top downwards, according to the assigned importance of the feature. Also, the strongest decrease of the F-value of 16.64 can be seen between index 11 and index 124, which describes the transition from the diagonal (-3) to the vertical (-4) orientation of the HOG feature. The F-values listed in Table 7 are all significantly greater than the critical F-value $F_{crit}(\alpha = 5\ \%) = 1.26$, indicating a large variance between classes when using these features. Moreover, the respective p-values are very small, ranging from $1.6 \cdot 10^{-38}$ to $1.43 \cdot 10^{-72}$, which means that the calculated F-values can be used validly. In the *parallel coordinates* plot in Fig. 6a we notice that the *wrinkles* can be particularly well characterised with diagonal features of HOG with the ending -3. Furthermore, with respect to Table 7 we can see that the cell-wise STDs of the features 135, 132, 136 are very well suited to separate the *foreign bodies* from the other defect classes. Far right located vertical HOG features labelled with ending -4 described *twists* quite well. The curves for *gaps* and *overlaps* are usually very close to each other in the *parallel coordinates* plot. These two classes are, expectedly, best characterised through vertically aligned HOG features having ending -0. This can be seen, for example, from the features 104 and 112. However, it is important to note that the STDs of the individual feature values intersect considerably between each other. Obviously, this can lead to uncertainties in the preformed image abstraction. Therefore, this STD feature is less suitable for these less prominent defect types. Accordingly, additional features that are more suitable for these classes should be incorporated into the feature set for this purpose.

In the following section, the classification scores for the CNN and the SVM are examined, based on the feature selection described above.

4.2. Classification accuracy and robustness

This section aims to analyse the classification rates for the SVM and the CNN. For training and prediction with the SVM, the previously described feature combination of HOG features with the cell-wise STD is applied. For these investigations, the selected feature vectors with dimension 10 and dimension 20 are examined respectively. Furthermore, the classification rate of the CNN for the specified ANN model is

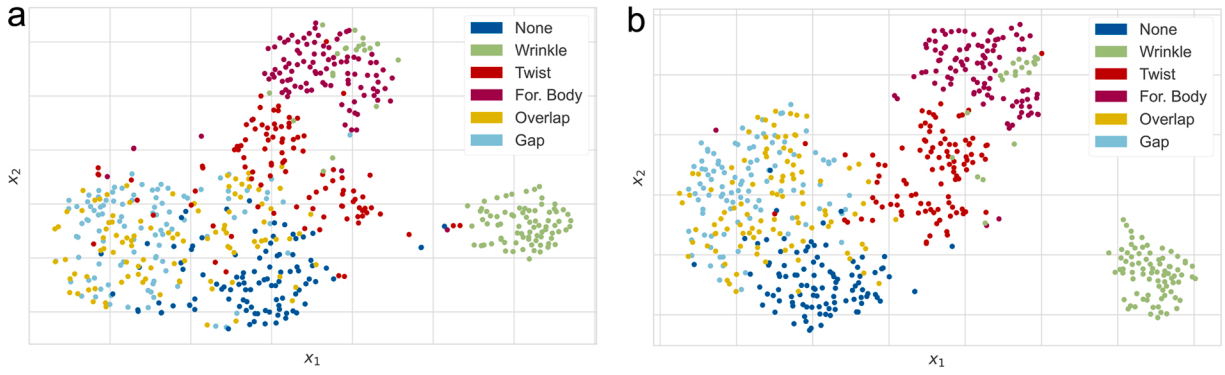

Fig. 5. The plots show the projection of the corresponding higher dimensional feature vectors from the manual feature extraction into the 2D space via the t-SNE calculations. The projection points are separated according to their defect category. x_1 and x_2 are the associated t-SNE projection axes.

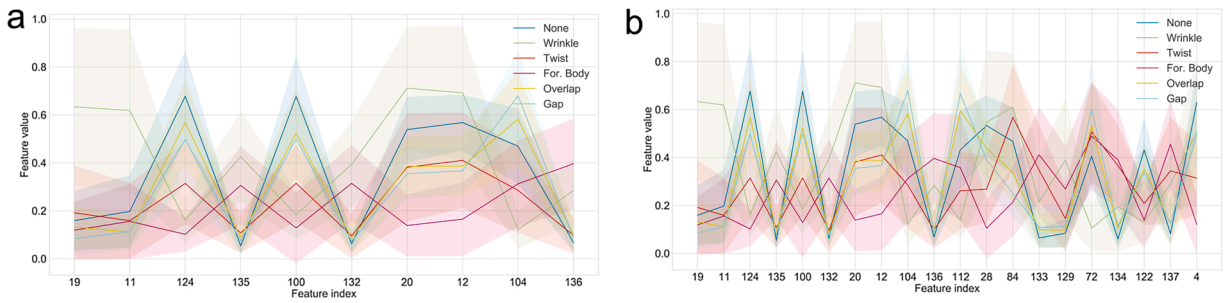

Fig. 6. The parallel coordinate values of the different selected feature indices normalised to the value range $[0, 1]$ are given. The presented features are arranged according to their assigned importance level from left to right. The attributes associated with the feature indexes are described in Table 7.

Table 7
The actual attributes belonging to the selected feature indexes from Fig. 6 are listed. x and y define the analysis cell on the input image. Index x: From left to right $[0, 3]$; Index y: From top to bottom $[0, 3]$; HOG orientation: Starting from vertical (-0) in steps of 45° in a clockwise rotation. Moreover, the respective F-values and p-values from the ANOVA feature selection are given. Critical F-value: $F_{crit}(\alpha = 5\ \%) = 1.26$.

Index	Feature name	F-val	p-val
19	hog-cell-x-0-y-2-orientation-3	130.78	$1.43 \cdot 10^{-72}$
11	hog-cell-x-0-y-1-orientation-3	120.30	$5.86 \cdot 10^{-69}$
124	hog-cell-x-3-y-3-orientation-4	103.66	$9.18 \cdot 10^{-63}$
135	std-cell-x-1-y-3	96.74	$5.32 \cdot 10^{-60}$
100	hog-cell-x-3-y-0-orientation-4	95.37	$1.93 \cdot 10^{-59}$
132	std-cell-x-1-y-0	92.80	$2.24 \cdot 10^{-58}$
20	hog-cell-x-0-y-2-orientation-4	89.32	$6.69 \cdot 10^{-57}$
12	hog-cell-x-0-y-1-orientation-4	84.61	$7.47 \cdot 10^{-55}$
104	hog-cell-x-3-y-1-orientation-0	74.62	$2.83 \cdot 10^{-50}$
136	std-cell-x-2-y-0	73.17	$1.38 \cdot 10^{-49}$
112	hog-cell-x-3-y-2-orientation-0	63.08	$1.49 \cdot 10^{-44}$
28	hog-cell-x-0-y-3-orientation-4	61.23	$1.4 \cdot 10^{-43}$
84	hog-cell-x-2-y-2-orientation-4	58.34	$4.8 \cdot 10^{-42}$
133	std-cell-x-1-y-1	58.09	$6.54 \cdot 10^{-42}$
129	std-cell-x-0-y-1	56.95	$2.72 \cdot 10^{-41}$
72	hog-cell-x-2-y-1-orientation-0	56.46	$5.03 \cdot 10^{-41}$
134	std-cell-x-1-y-2	55.39	$1.97 \cdot 10^{-40}$
122	hog-cell-x-3-y-3-orientation-2	55.15	$2.69 \cdot 10^{-40}$
137	std-cell-x-2-y-1	54.55	$5.78 \cdot 10^{-40}$
4	hog-cell-x-0-y-0-orientation-4	51.99	$1.6 \cdot 10^{-38}$

analysed. As mentioned before, the training and testing of both classifiers is performed based on synthetic image data. Obviously, separate training and testing dataset are used for each.

The classification rates of the different classifiers and configurations are given for the SVM in Fig. 7 and for the CNN in Fig. 8. The results are presented as confusion matrices where the vertical axis represents the ground truth class and the horizontal axis the predicted class.

In order to examine the robustness and confidence of a machine

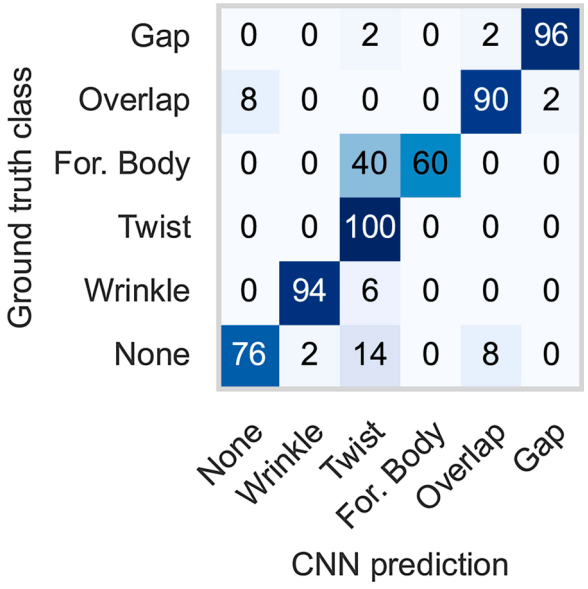

Fig. 8. The classification scores of the CNN are given as a confusion matrix. The vertical axis indicates the ground truth class, the horizontal axis the predicted class. The average classification rate is 86.0%.

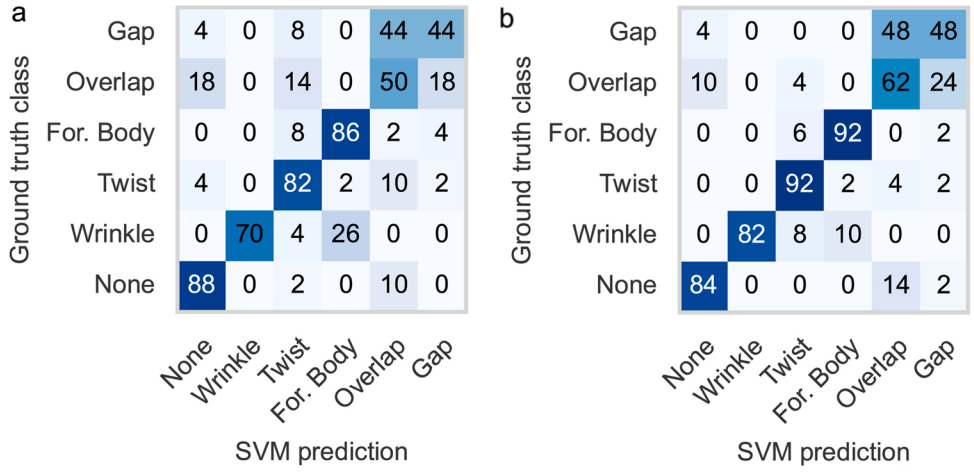

Fig. 7. The classification scores of the SVM for feature vectors of different lengths are displayed as a confusion matrix. The vertical axis indicates the ground truth class, the horizontal axis the predicted class.

classification more closely, the mean distances and the STDs of the feature vectors from the test data set to the pre-trained separating hyperplane of the SVM are displayed in the confusion matrices in Fig. 9.

Accordingly, we again differentiate between the feature vectors with different dimensions. For the CNN classification output, the mean percentage of neural activations for each class neuron, together with the STDs, is displayed for the images of the test data set. Keep in mind that each output neuron can be activated almost independently between 0% and 100%. According to the max pooling operation applied, the class with the highest activation is predicted. In addition, Fig. 10 presents the 95.5% confidence interval for each cell of the confusion matrix. Due to the minimum and maximum possible activations, the confidence intervals are restricted respectively.

When looking at the SVM classification results, we first notice that the classification rate for a 20 dimensional feature vector is 76.7%, only 6.7% higher than the classification rates for the application of a 10 dimensional feature vector. This basically implies that the first 10 features of the feature vector already contain a very large amount of information. As assumed before in Section 4.1, the last 10 entries of the feature vector primarily lead to a better differentiability between *wrinkles* and *foreign bodies*. The classification rate for *wrinkles* increases by 12% and that for *foreign bodies* by 6%. Furthermore, *twists* are better separated from *overlaps* and *gaps*. The classification rate for *twists* increases by 10%. Less obvious from the above analysis of the t-SNE outcomes and the *parallel coordinates* plots is the improvement in the classification rate of *overlaps* by 12%, when using 20 features instead of only 10 features. The changes in classification rates for *none* and *gaps* are not significant. Especially the less distinct defect types *none*, *gap* and *overlap* can only be classified to a limited extent with the current manual feature extraction and SVM. However, the focus in this study is primarily on the traceability of individual machine decision stages and not on the optimisation of the classification rate.

When looking at the classification rates of the CNN, classification rates of $\geq$90% are evident for all classes except *foreign bodies* and *none*. For *none* and *foreign bodies*, the classification rates are significantly lower at 76% and 60%. However, these two classes are particularly well classified with a 20 dimensional feature vector and the SVM classifier, having 84% and 92% accuracy.

As described above, the confusion matrices in Fig. 9 display the mean distances of the test feature vectors to the pre-trained linear separating hyperplane. Thus, a larger distance is an indicator for a more robust classification decision in the case of degraded input data. For both vector spaces $\mathbb{R}^n$ with $n = 10$ and $n = 20$, a quite similar behaviour is apparent with respect to the distance values. The differentiation of very prominent defect types from rather inconspicuous defect types is basically more robust than within these two prominence groups. Particularly striking are the distances for the differentiation between *foreign body* and *none* in both classifier vector spaces. *Wrinkles*, on the other hand, have a large distance to the class *gaps*, but the distance to all other classes is very similar to the distance to their own class. They also have relatively high STDs. For *twists*, the distances to all other classes, except their own, are quite similar. For *overlaps* and *gaps*, increased distances to *wrinkles* and *foreign bodies* are again evident. As described above, the distance values for $n = 10$ and $n = 20$ are very similar, but a slight increase in distances for $n = 20$ is noticeable across basically all cells of the confusion matrix. Thus, this implies that the application of the 20 features probably increases the classification robustness. This reflects the classification behaviour described above and the representation in the t-SNE plot from Fig. 5. This distance analysis accordingly also gives a feasible measure of whether a longer feature vector potentially leads to overfitting or rather increases the robustness of the classification. In the case of overfitting in the training phase, the distances would tend to decrease as the amount of information increases. Thus, in the use case considered here, significantly more features can be utilised for training the SVM. However, this makes the analysis of the classification process more complex, which is why we restrict ourselves to $n = 10$ and $n = 20$.

Similar to the distance values of the SVM, the neuronal activations at the output of layer #16 of the CNN can be used as a robustness indicator for the classification decision. The corresponding results are given in the confusion matrices in Fig. 10. Accordingly, Fig. 10a first presents the mean percentage neuronal activations for different classes across the full test data set. As expected, the largest mean activations are present on the matrix diagonal where the ground truth class equals the predicted class. In comparison with the classification results of the CNN from Fig. 8, we clearly see that the defect types *wrinkle* and *gap*, which have been reliably categorised with a very high classification rate, also have a great mean neuronal activation combined with a low STD. However, what is striking is that for *twists* with the highest classification rate of 100%, a lower mean neuronal activation of only 92.53% with a significantly increased STD of $\sigma = 13.89\%$ is evident. Additionally, the incorrect classes are significantly activated with up to a mean value of 2.62% and a maximum STD of $\sigma = 7.86\%$. Conversely, for *foreign bodies*, the high misclassification as *twists* is indicated through the relatively large mean neuronal activation of the *twist* class of 8.83% with a very high STD of $\sigma = 12.9\%$. From this we conclude two things. First, all neuronal activations must be considered jointly in order to be able to make a statement about the trustworthiness of a classification decision. In addition to the actual neuronal activations with the associated STDs, the activations of the neighbouring classes are also relevant. Deviations from this activation pattern indicate difficulties of the ANN to assign the input data accordingly. However, the reduction of the maximum activation is tolerable and by itself is only a limited indicator for unsuitable input

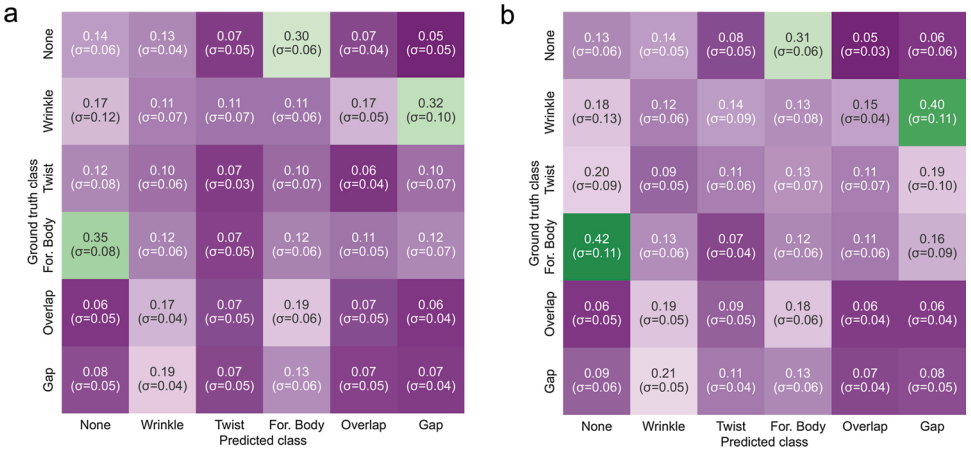

Fig. 9. The mean distances and the associated STD of a feature vector of the test data set to the previously trained separating hyperplane of the SVM classifier are presented. Different confusion matrices are displayed with respect to the different dimensions of the corresponding vector space.

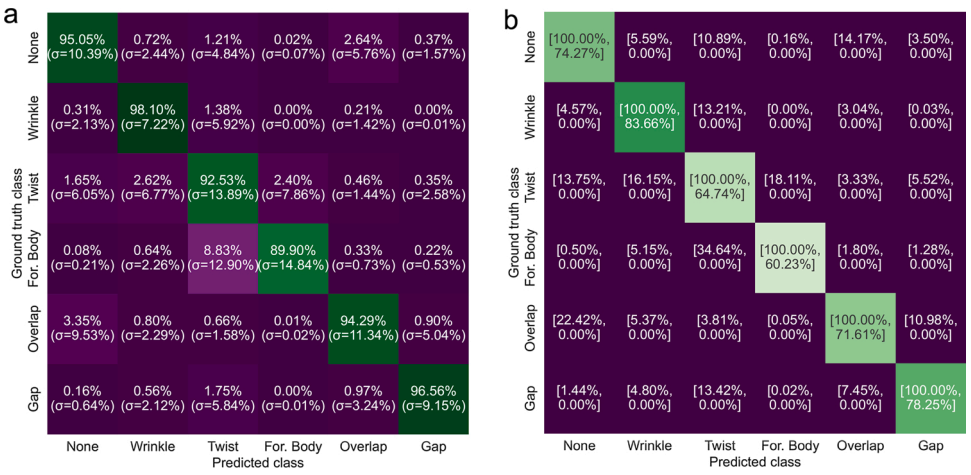

Fig. 10. The mean neuronal activations of layer #16 of the applied CNN model are given for the utilised test data set together with the corresponding STD (a) as well as the derived and restricted 95.5% confidence interval (b).

data. In addition, some defect types also seem to cause larger activations of the surrounding classes. Such a behaviour can of course also be used to introduce an "unknown" class for all defect types that have an excessive activation of multiple classes. This has the advantage that even completely unfamiliar input data can be handled within the ANN.

Based on the neuronal activations from Fig. 10a, the 95.5% confidence interval for each individual cell is given in the confusion matrix in Fig. 10b. As described above, the permissible neuronal activation is obviously limited to the value range $[0, 100]$%. However, because the confidence interval only indicates which proportion of the occurring defects is within this interval, the confidence estimation for a single defect type is rather difficult. The assessment of the confidence that a particular defect type is not present in the input image is considerably more straightforward and achievable through analysing the given statistical values. For this, we have to take into account that the 95.5% confidence interval is determined for each matrix cell. But as mentioned

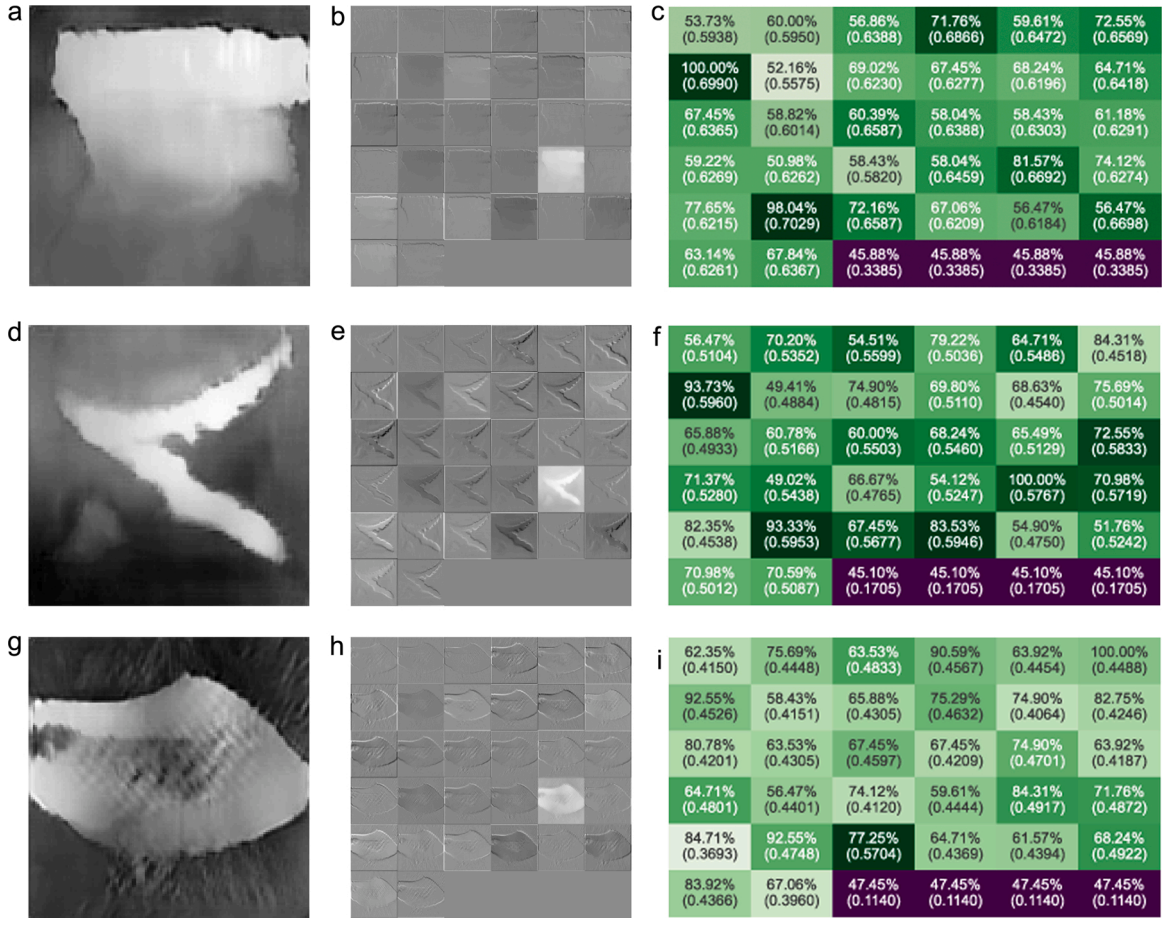

Fig. 11. The figure shows the input image for each distinct defect type on the left, all corresponding feature maps of the first joined convolutional layer #2 of the CNN in the middle and the respective comparison matrix on the right. Each comparison matrix gives the percentage of neural activation and the SSIM value for each feature map, in relation to the input image of each class. The values are given as: Neural activation% (SSIM score).

before, all individual activations have to be considered jointly. Hence, for the six classes, the confidence level decreases to $95 \,.\, 5^{6}\% = 75.8\ \%$. Consequently, we are able to conclude that for the prediction of a certain defect type there is a 75.8% confidence that this defect type is misclassified if one of the six neuronal activations exceeds the given confidence intervals. The reverse conclusion is not necessarily valid. However, this analysis provides a feasible indicator for an exception handling in the automated classification procedure of manufacturing defects.

4.3. Feature map output and similarity to input image

In this section, the CNN output data of the first joined convolutional layer #2 are examined for their validity in comparison with the input image. For this purpose, first the input image as well as the feature maps belonging to the first joined convolutional layer are visualised. Subsequently, the matching between the respective gradient images are determined via the SSIM algorithm and displayed in the confusion matrix together with the percentage neuronal activation of each feature map. These representations are displayed separately for distinct defect types in Fig. 11 and less distinct defect types in Fig. 12.

Initially, we would like to qualitatively evaluate the visual appearance of the individual feature maps in comparison to the input image. First, we examine the distinct defects from Fig. 11. We can see clearly that the edges of the distinct defects are represented in many feature maps quite well. However, we observe a difference in brightness across the feature maps. Furthermore, edges with varying orientations are represented differently in the individual feature maps. This can be attributed to the respective convolution matrices applied within this first convolution layer.

For the feature maps of the less distinct defects *overlap* and *gap* in Fig. 12 we recognise that the lateral edges of these defect types are presented very well in several feature maps. In these cases a convolution kernel which is very sensitive in the horizontal image direction is applied. In some other feature maps, hardly any edges of both defect types are visible. However, the edges of non-defect-specific artefacts are sometimes emphasised in the measurement image. This obviously enables a proper differentiation of such pre-processing artefacts to the actual defect type. Hence, this explains the significantly improved classification rates of the CNN in Fig. 8 for *overlaps* and *gaps* compared to the SVM classification scores from Fig. 7 for the same defect types. For flawless measurement images, only a slight change in image texture and brightness of the feature maps is visible.

Noticeable across all classes are the nearly uniform bright, homogeneous four feature maps in the bottom right corner of each of the aggregate overviews of the individual feature maps. The convolution matrices applied in these cases seem to have only little sensitivity to the actual image features. Nevertheless, the actual activation within these four feature maps varies with the distinctness and attributes of the input defect image.

In the following, the SSIM values are examined. Those indicate the correlation between the two gradient images of the feature map and the input image. These values are analysed with respect to the actual percentage of the CNNs neural activations. The SSIM values are given in the confusion matrix in brackets below the percentage of neural activation.

First, the analysis is performed again for the distinct defect types

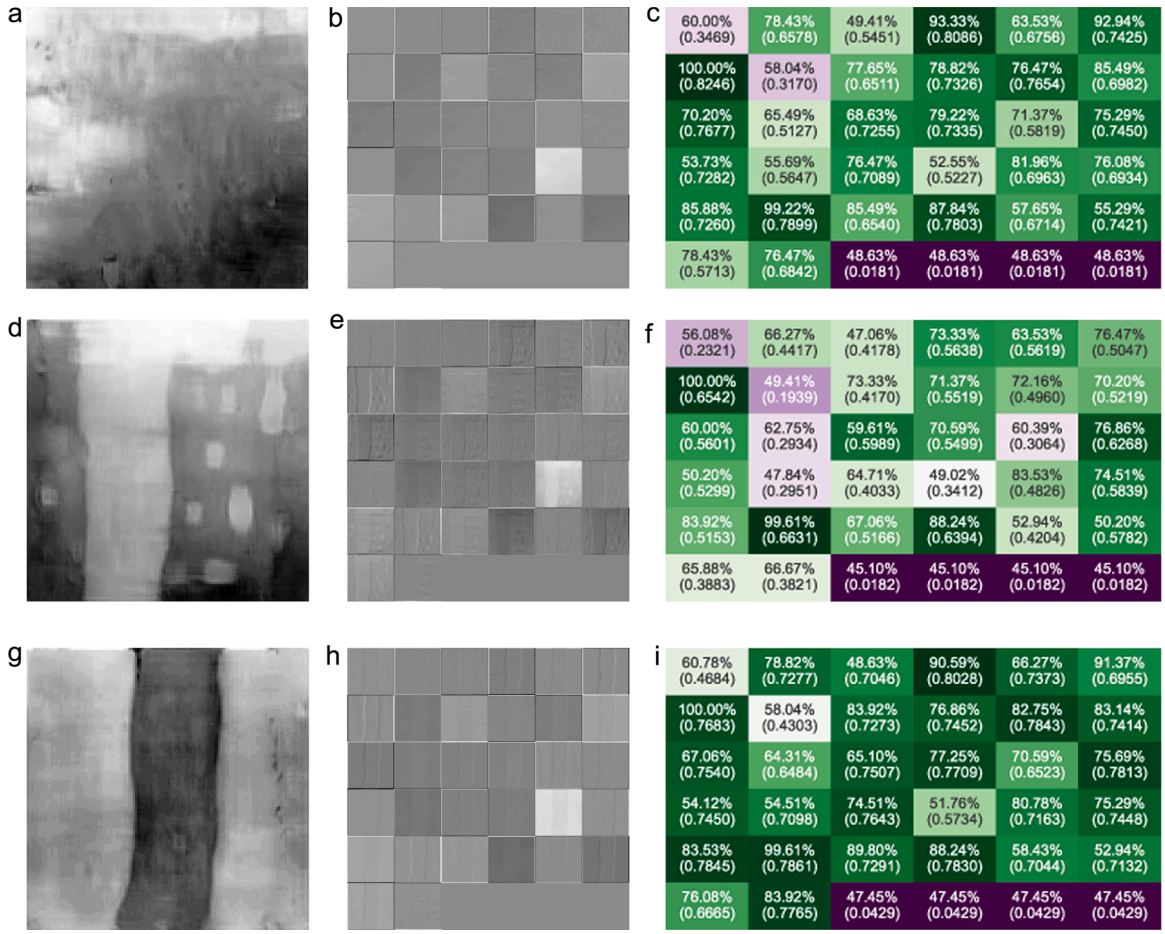

Fig. 12. The figure shows the input image for each less distinct defect type on the left, all corresponding feature maps of the first joined convolutional layer #2 of the CNN in the middle and the respective comparison matrix on the right. Each comparison matrix gives the percentage of neural activation and the SSIM value for each feature map, in relation to the input image of each class. The values are given as: Neural activation% (SSIM score).

from Fig. 11. For *wrinkles* in Fig. 11c, there is close correlation between the SSIM values and the neural activations for very strong and very weak neuronal activations. The four strongest and four weakest activated feature maps correspond to the highest and lowest SSIM values. For moderate neural activations, the SSIM values have only little significance. For *twists* in Fig. 11f and *foreign bodies* in Fig. 11i, on the other hand, only a very low correlation of the SSIM values with the stronger neuronal activations is evident. The very weak neuronal activations, however, are again well represented through the SSIM. This behaviour is probably due to the fact that the CNN performs significantly more diversified convolution operations for edge detection than the horizontal and vertical Sobel filter operation applied for this comparison. Alternative and, eventually, differently oriented convolution matrices or advanced edge detection methods could contribute to an improvement of the correlation. For *wrinkles*, the agreement between the examined neuronal activations and SSIM values are quite high, since this defect type has very pronounced vertical and horizontal edges, which are extracted as features from the CNN.

For the *none* class in Fig. 12c, the SSIM values correlate closely with the neuronal activations of the CNN. This also applies roughly to the moderate neuronal activations and not only to very strong and very weak neuronal activations as for *wrinkles*. This is plausible since the light textures in the gradient image are arranged relatively randomly and thus have a similar amount with similar magnitudes of vertical and horizontal gradient components. Thus, they are detected equally efficiently from both the CNN and the horizontally and vertically sensitive Sobel filter. This also strengthens the assumption that an extended filter kernel or an advanced edge detection algorithm might improve the linke between the SSIM scores and the neuronal activations of the CNN. For *overlaps* in Fig. 12f, the neuronal activations correlate well with the corresponding SSIM scores for many cells in the confusion matrix. This correlation even applies fairly well to moderately strong neuronal activations. In case of this example image, however, this behaviour might be strongly influenced from the rectangular pre-processing artefacts in the sample image. Nevertheless, this again confirms the assumption described several times before that the applied gradient filter has a great influence on this evaluation. For *gaps* in Fig. 12i, no meaningful correlation between the SSIM values and the neuronal activations of the CNN is apparent. The reason for this is probably again the operating principle of the Sobel filter used for the manual gradient calculation.

4.4. Matching of xAI importance of pixel regions with chosen features from feature selection

In this section, the matching of important pixels from the *Smooth IG* computation with relevant pixels from the feature selection are examined. This analysis is carried out for the example defects presented in the previous section. Please be aware that the feature selection is only performed once over the entire training dataset. Thus, the index of the selected features remains unchanged for all test images. The amount of features equals the quantity of selected features. For the sake of clarity in visualisation, we only consider the 10 most relevant features from the feature selection. The exact composition of the feature vector is given in Table 7. In contrast, the *Smooth IG* results are calculated individually for each CNN input image. Fig. 13 presents the outcomes of the conducted investigation. The top row from Fig. 13a to 13 f shows the output images of the *Smooth IG* computation. In the bottom row, from Fig. 13g to 13 l, the matches between xAI outcomes and the 10 selected features are marked cell by cell. The important image cells from the *Smooth IG* method are labelled with *x*. *o* marks direction-dependent features from the HOG and *s* represents the STD features. The *n* brightest cells of the *Smooth IG* calculation were labelled. *n* corresponds to the number of different cells addressed via the manual feature selection. The xAI markers are directly overlaid with the original input image of the classifier to demonstrate the respective geometrical correlations with the image.

Initially, we notice that the pixel relevance of the *Smooth IG* calculations strongly depends on the brightness of the input image at the respective location. This is due to the operating principle of *Smooth IG*, which is described in Section 2.7.2. In contrast, the calculated image features are more sensitive to edges or inhomogeneities in the image. Since the input images usually contain the entire manufacturing defect, the edges and conspicuous areas tend to be located at the edges of the input images. As can be seen in Fig. 13g–l, this also corresponds to the image areas implicitly selected via the feature selection. All selected image features are located in the border cells of the 4×4 cell structure. The right and left borders are described with HOG features, whereas the top and bottom border cells are represented via STD features. Keeping in mind that the same feature set is used for all defect types equally, this concrete feature selection might be related to the fact that in particular the lateral edges of the *gap* and *overlap* defects can be described very well with HOG features. At the top and bottom of the image, there are different brightness variations in the images for the different defect types, which might be meaningfully characterised through the STD. However, this agrees only to a limited extent with the *Smooth IG* assessments. For classes with rather flat shapes or a central intensity maximum such as *none*, *wrinkles*, *twists* and *foreign bodies*, the important image areas are therefore located in the centre of the image. For *gaps* and *overlaps*, these important image cells tend to be located laterally, along the defect, correspondingly where the defect type indicates a respective increase in brightness. Both considerations are obviously combinable, since edges in an image are consistent with a change in pixel brightness in a certain image direction. Accordingly, the output of the *Smooth IG* method can represent exactly such image edges when combining the observation of very bright xAI results with a neighbouring cell with very dark outcomes.

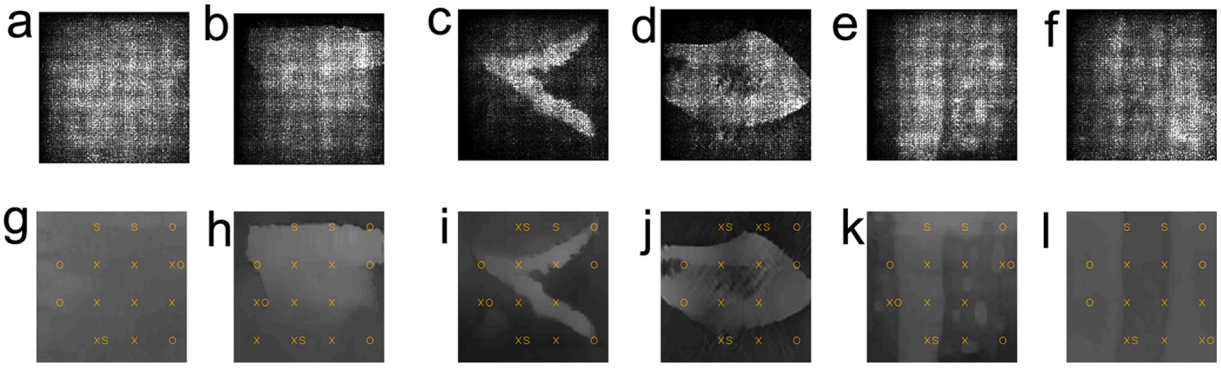

Fig. 13. The result images of the *Smooth IG* calculations for each example defect are displayed in the first row. The second row presents the original input images with a marking of the most important cells resulting from the *Smooth IG* procedure (*x*), as well as the relevant features determined from the feature selection (HOG: *o*; STD: *s*). A 10 dimensional feature vector is considered.

In order to assess the validity of the *Smooth IG* approach in relation to the CNN model, the SenseMAX and INFD values can be used. SenseMAX describes the sensitivity of the xAI procedure to manipulated input data. INFD denotes the actual correlation of the behaviour of the xAI procedure with respect to the CNN model. The corresponding results for the full test data set are calculated according to the methodology in Section 3.5.3 and listed in Table 8.

We notice that *Smooth IG* responds more strongly to changes in the input image for less distinct defect types such as *none*, *overlaps* and *gaps* than for more distinct defects like *wrinkles*, *twists* and *foreign bodies*. However, we must take into account that the *Smooth IG* corresponds more closely to the behaviour of the CNN for *foreign bodies*, *overlaps* and *gaps* than for *none* and *twists*. *Wrinkles* are in the middle range in this respect. In principle, the *Smooth IG* xAI method applied above can be utilised for all defect types and the respective findings are meaningfully.

Below, the overall results of this study are discussed jointly.

5. Discussion

In this section, the results from this study are discussed and linked to related research. Furthermore, the research questions are answered.

The hybrid classifiers of Joshi et al. [34], Basly et al. [36], Xu et al. [37], Lee et al. [39] and Mohamed et al. [40] outlined in Section 2.3 indicate the principal value of combining different classifiers in terms of improving the classification performance. Sun et al. [41] focuses on the advantages of a hybrid classification approach for the fusion of input data from different sensors. We were able to make use of this potential for the investigation of the reliability and comprehensibility of machine decisions in this study. So far, such an approach has only been considered peripherally in the related research and thus represents a great novelty in the field of manufacturing inspection systems.

Based on our previous research [9], we were able to show that the provided synthetic data from the DCGAN is suitable for training the utilised CNN and SVM. The applied CNN architecture from Chen et al. [109] yields with 86% very convincing classification results for the utilised synthetic defect samples. However, these classification rates are strongly reduced due to the extensive misclassification of *foreign bodies*. Without the consideration of this class, the classification rate is 91.2%. The classification rates for the used SVM is adequate with 76.7% despite its simplistic design for better comprehensibility of the classification. The classification rates of the SVM are most strongly reduced through the misclassification of *gaps* and *overlaps*. Assuming these two classes are combined, the classification rate of the linear SVM for the synthetic data is 89%. This is close to the classification accuracy of the CNN. The combined use of CNN and SVM can lead to improvements in classifying less distinct defects like *none*, *gap*, *overlap* more accurately. This is probably due to the varying sensitivity of both methods for smaller image gradients and slighter edges. Moreover, the confidence intervals of the CNNs neural activations per class are usable for estimating the confidence that a certain prediction does not correspond to the considered class. Additionally, the visual impression of the outcomes of the first joined convolutional layer of the CNN indicate the validity of the input data and the sensitivity of the CNN for processing these images. But the quantitative comparison between feature maps and input images via SSIM strongly depends on the applied gradient calculation method. However, for certain types of defects or a gradient calculation adapted to the specific application case, such values provide a meaningful quantitative indicator for the evaluation of the validity of the CNNs input data.

In order to assess the direct input of the SVM, the determined raw image features can be evaluated via the t-SNE statistics in combination with *parallel coordinates* plots. This agrees very well with the findings from the studies of Xu et al. [37] and Lee et al. [39]. This approach is usable to quantitatively evaluate the performance of a feature vector for utilisation within a SVM. Moreover, we have shown that the selected image features as well as the results of the *Smooth IG* calculations can be mapped onto the input image very well and the validity of the conclusions can be illustrated. However, in order to achieve a direct match of the *Smooth IG* results with the manually extracted image features, further adjustments are necessary. Respectively, the SenseMAX and INFD metrics additionally provide a sound way to describe the informativeness and robustness of a xAI method for a given use case. These metrics also give the ability to assess the trustworthiness of the xAI outputs. The research of Lee et al. [45] is a valuable complement to the findings presented in this study. In a modified manner, their approach might contribute to consolidate the diverse outcomes from our research and prepare them for specific groups of people from the composite manufacturing sector. In this context, our conducted research findings provide additional valuable indicators for the traceability of a machine decision, which expand the scope of the study of Lee et al. [45]. However, in its current form, our presented approach is strongly focused on input images that can be visually interpreted through humans. This is also a limitation of this approach. For instance, for non-imaging input data or the direct line-by-line interpretation of the height profile scans of the LLSS, the presented concept is not directly applicable and needs to be adapted accordingly. Furthermore, although the application of a very simple SVM setup with few simple input features offers the possibility of a sound understanding of the decision making, its performance is limited. This must be taken into account in principle, but does not contradict the overall aim of this paper.

Finally, we would like to explicitly address the two research questions. In order to answer the first research question, the t-SNE calculations in combination with a *parallel coordinates* plot for given image features are very well suited to assess the performance of a SVM. The defect type specific evaluation of all neuronal output activations of the CNN and the examination of the resulting 95.5% confidence interval are a reasonable way for estimating the uncertainty of the CNN classification for individual defect types and unknown input data. Accordingly, exception handlings can be derived from this observation when the decision uncertainty gets too large.

To answer the second research question, the comparison between the gradient images of the feature maps of the first joined convolutional layer of the CNN with the gradient images of the original input images via the SSIM metric is suitable to check the validity of the input data of the CNN classifier. However, it is essential to perform a suitable gradient calculation for the particular use case. Furthermore, a linking of the *Smooth IG* outputs for the used CNN model with the input features of the SVM and subsequent projection onto the original input image is well suited to assess the sensitivity of both classification streams for certain image attributes.

In further research, it is certainly useful to investigate additional SVM features. In this context, ways to visualise such more complex feature sets as input to the SVM should also be examined. Building on this, larger, manipulated or mixed data sets from different data sources should also be considered. Furthermore, the uncertainty analysis of the CNN should be continued in order to determine the very precise uncertainty of a classification decision for each defect type. For this

Table 8
The mean SenseMAX and INFD values are given together with their associated STDs σ for the *Smooth IG* algorithm under consideration of the given CNN model and the full test data set.

Metric	None	Wrinkle	Twist	For. Body	Overlap	Gap
SenseMAX	0.1026 ($\sigma = 0.0042$)	0.0915 ($\sigma = 0.0078$)	0.0869 ($\sigma = 0.0086$)	0.0875 ($\sigma = 0.0084$)	0.1002 ($\sigma = 0.0051$)	0.1003 ($\sigma = 0.0059$)
INFD	1.0832 ($\sigma = 0.0132$)	0.59 ($\sigma = 0.0077$)	1.057 ($\sigma = 0.0131$)	0.082 ($\sigma = 0.0122$)	0.0736 ($\sigma = 0.0094$)	0.0757 ($\sigma = 0.0127$)

purpose, the integration of advanced statistical models into the classification model can probably be beneficial in order to perform an inherent uncertainty estimations already during the prediction. The application of a Bayesian ANN can be a starting point for this. In addition, it is reasonable to perform a benchmarking between different ANN classifiers, conventional ML methods and xAI techniques. For instance, recurrent classifiers can be useful for line-by-line image analysis. In addition, DT based methods can potentially provide an advantage in creating a set of human understandable rules for conventional classifiers. Furthermore, investigating other, even more complex, xAI methods in the context of this parallel classification architecture is exciting. In particular, ML methods such as *Local Interpretable Model-Agnostic Explanations* (LIME) are very interesting for future analysis. These add further complexity to the system, but potentially offer better performance. However, this requires more detailed investigations. In the following section, the major findings of this study are summarised again and the added value for the community is highlighted.

6. Conclusion

This section summarises the key findings of this study and highlights their added value for the field.

The analysed *t-Distributed Stochastic Neighbor Embedding* and *parallel coordinates* plots are well suited for direct feature assessment of a manually crafted feature set of the *Support Vector Machine*. Moreover, the evaluation of all *Convolutional Neural Network* neural output activations for a certain defect class can be used as an uncertainty indicator for the respective *Convolutional Neural Network* decision. Especially the visualisation the feature maps of the first joined convolutional layer of the *Convolutional Neural Network* are feasible to check the validity of the input data of this classifier. Furthermore, mapping the *Smooth Integrated Gradients* outputs for a certain *Convolutional Neural Network* model together with the *Support Vector Machine* feature onto the raw input images provides suitable indication of the sensitivity of both classifiers to certain image attributes.

The methodology of this paper can also be applied to different image-based inspection processes in composite manufacturing and beyond for gaining understanding and describing the behaviour, robustness and validity of machine decision processes. Accordingly, the findings from this research are valuable for developers and applicants of such image-based inspection systems.

Funding

This research is part of the project DHiiP-AIR and was financially supported by the Federal Ministry for Economic Affairs and Energy. This project has received funding from the Federal Ministry for Economic Affairs and Energy under the funding code No. 20W1911F.

Supported by:

Federal Ministry for Economic Affairs and Energy

on the basis of a decision by the German Bundestag

Author contributions

Sebastian Meister: Conceptualization; Methodology; Software; Formal analysis; Investigation; Writing – Original Draft.

Mahdieu Wermes: Methodology; Software; Formal analysis; Data Curation; Visualization.

Jan Stüve: Resources; Writing – Review & Editing; Supervision; Funding acquisition.

Roger Groves: Conceptualization; Writing – Review & Editing; Supervision.

All authors have approved the final manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

Declaration of Competing Interest

The authors report no declarations of interest.

References

[1] Marsh G. Airbus A350 XWB update. Reinf Plast 2010;54(6):20–4. https://doi.org/10.1016/s0034-3617(10)70212-5.
[2] McIlhagger A, Archer E, McIlhagger R. Manufacturing processes for composite materials and components for aerospace applications. In: Irving P, Soutis C, editors. Polymer composites in the aerospace industry. Elsevier; 2020. p. 59–81. https://doi.org/10.1016/b978-0-08-102679-3.00003-4.
[3] Airbus SE. Airbus 2020 deliveries demonstrate resilience. Airbus SE; 2021 (accessed 02.06.21 (January 2021)), https://www.airbus.com/newsroom/press-releases/en/2021/01/airbus-2020-deliveries-demonstrate-resilience.html.
[4] Airbus SAS. Global market forecast 2019–2038. AIRBUS S.A.S; 2019 (accessed 02.06.21 (2019)), https://www.airbus.com/aircraft/market/global-market-forecast.html.
[5] Boeing. Commercial market outlook 2020–2039, Boeing. 2020 (accessed 02.06.21 (2020)), http://www.boeing.com/commercial/market/commercial-market-outlook/index.page.
[6] Eitzinger C. Inline inspection helps accelerate production by up to 50 %. Lightweight Design worldwide; 2019.
[7] European Union Aviation Safety Agency. Human intelligence roadmap – a human-centric approach to ai in aviation, techreport Vers. 1.0, European union aviation safety agency, vers. 1.0 (February 2020). 2020. https://www.easa.europa.eu/ai.
[8] EASA. AI task force, daedalean AG, concepts of design assurance for neural networks (codann), techreport Vers. 1.0, European union aviation safety agency and Daedalean AG, vers. 1.0 (March 2020). 2020. https://www.easa.europa.eu/document-library/general-publications/concepts-design-assurance-neural-networks-codann.
[9] Meister S, Möller N, Stüve J, Groves RM. Synthetic image data augmentation for fibre layup inspection processes: techniques to enhance the data set. J Intell Manuf 2021. https://doi.org/10.1007/s10845-021-01738-7.
[10] Schmidt C, Hocke T, Denkena B. Artificial intelligence for non-destructive testing of CFRP prepreg materials. Prod Eng 2019. https://doi.org/10.1007/s11740-019-00913-3.
[11] Meister S, Wermes MAM, Stüve J, Groves RM. Explainability of deep learning classifier decisions for optical detection of manufacturing defects in the automated fiber placement process. In: Beyerer J, Heizmann M, editors. SPIE optical metrology – OM106 – automated visual inspection and machine vision IV. SPIE; 2021. https://doi.org/10.1117/12.2592584. https://spie.org/EOM/conferencedetails/automated-visual-inspection.
[12] Cemenska J, Rudberg T, Henscheid M. Automated in-process inspection system for AFP machines. SAE Int J Aerosp 2015;8(2):303–9. https://doi.org/10.4271/2015-01-2608.
[13] Weimer C, Friedberger A, Helwig A, Heckner S, Buchmann C, Engel F. Increasing the productivity of CFRP production processes by robustness and reliability enhancement. In: CAMX 2016 – the composites and advanced materials expo and conference, airbus group innovations, 81663 Munich Germany; 2016. https://www.researchgate.net/profile/Christian_Weimer/publication/308778487_increasing_the_productivity_of_cfrp_production_processes_by_robustness_and_reliability_enhancement/links/57efa78208ae886b8975147a.pdf.
[14] Black S. Improving composites processing with automated inspection, compositesworld. 2018 (accessed 19.06.19) (January 2018), https://www.compositesworld.com/articles/improving-composites-processing-with-automated-inspection.
[15] Meister S, Wermes MAM, Stueve J, Groves RM. Algorithm assessment for layup defect segmentation from laser line scan sensor based image data. In: Zonta D, Huang H, editors. Sensors and smart structures technologies for civil, mechanical, and aerospace systems 2020. SPIE; 2020. https://doi.org/10.1117/12.2558434.
[16] Lengsfeld H, Fabris FW, Krämer J, Lacalle J, Altstädt V. Faserverbundwerkstoffe. Hanser Fachbuchverlag; 2014. https://www.ebook.de/de/product/22746074/hauke_lengsfeld_felipe_wolff_fabris_johannes_kraemer_javier_lacalle_volker_altstaedt_faserverbundwerkstoffe.html.
[17] Maass D. Automated dry fiber placement for aerospace composites. Composites manufacturing 2012 2012.
[18] Rudberg T, Nielson J, Henscheid M, Cemenska J. Improving AFP cell performance. SAE Int J Aerosp 2014;7(2):317–21. https://doi.org/10.4271/2014-01-2272.
[19] Oromiehie E, Prusty BG, Compston P, Rajan G. Automated fibre placement based composite structures: review on the defects, impacts and inspections techniques. Compos Struct 2019;224. https://doi.org/10.1016/j.compstruct.2019.110987.

[20] Nardi D, Abouhamzeh M, Leonard R, Sinke J. Detection and evaluation of prepreg gaps and overlaps in glare laminates. Appl Compos Mater 2018;25(6):1491–507. https://doi.org/10.1007/s10443-018-9679-z.
[21] Harik R, Saidy C, Williams SJ, Gurdal Z, Grimsley B. Automated fiber placement defect identity cards: cause, anticipation, existence, significance, and progression. SAMPE 18 2018. https://www.researchgate.net/publication/326464139_Automated_fiber_placement_defect_identity_cards_cause_anticipation_existence_significance_and_progression.
[22] Heinecke F, Willberg C. Manufacturing-induced imperfections in composite parts manufactured via automated fiber placement. J Compos Sci 2019;3(2):56. https://doi.org/10.3390/jcs3020056.
[23] Potter K. Understanding the origins of defects and variability in composites manufacture. ICCM international conferences on composite materials 2009. http://iccm-central.org/Proceedings/ICCM17proceedings/Themes/Plenaries/P1.5%20Potter.pdf.
[24] Sun S, Han Z, Fu H, Jin H, Dhupia JS, Wang Y. Defect characteristics and online detection techniques during manufacturing of FRPs using automated fiber placement: a review. Polymers 2020;12(6):1337. https://doi.org/10.3390/polym12061337.
[25] Atkinson GA, Thornton TJ, Peynado DI, Ernst JD. High-precision polarization measurements and analysis for machine vision applications. In: 2018 7th European workshop on visual information processing (EUVIP); 2018. https://doi.org/10.1109/euvip.2018.8611762.
[26] Schöberl M, Kasnakli K, Nowak A. Measuring strand orientation in carbon fiber reinforced plastics (CFRP) with polarization. 19th World conference on non-destructive testing 2016 2016.
[27] Denkena B, Schmidt C, Völtzer K, Hocke T. Thermographic online monitoring system for automated fiber placement processes. Compos Part B Eng 2016;97:239–43. https://doi.org/10.1016/j.compositesb.2016.04.076.
[28] Schmidt C, Hocke T, Denkena B. Deep learning-based classification of production defects in automated-fiber-placement processes. Prod Eng 2019;13(3–4):501–9. https://doi.org/10.1007/s11740-019-00893-4.
[29] Gardiner G. Zero-defect manufacturing of composite parts. CompositesWorld; 2018 (accessed 18.06.19 (November 2018)), https://www.compositesworld.com/blog/post/zero-defect-manufacturing-of-composite-parts.
[30] Black S. Improving composites processing with automated inspection, Part II, Compositesworld. 2018 (accessed 19.06.19 (June. 2018)), https://www.compositesworld.com/articles/improving-composites-processing-with-automated-inspection-part-ii.
[31] Shrestha A, Mahmood A. Review of deep learning algorithms and architectures. IEEE Access 2019;7:53040–65. https://doi.org/10.1109/access.2019.2912200.
[32] Sen PC, Hajra M, Ghosh M. Supervised classification algorithms in machine learning: a survey and review. Advances in intelligent systems and computing. Springer Singapore; 2019. p. 99–111. https://doi.org/10.1007/978-981-13-7403-6_11.
[33] Zhao X, Shi X, Liu K, Deng Y. An intelligent detection and assessment method based on textile fabric image feature. Int J Cloth Sci Technol 2019;31(3):390–402. https://doi.org/10.1108/ijcst-01-2018-0005.
[34] Joshi KD, Chauhan V, Surgenor B. A flexible machine vision system for small part inspection based on a hybrid SVM/ANN approach. J Intell Manuf 2018;31(1):103–25. https://doi.org/10.1007/s10845-018-1438-3.
[35] Malaca P, Rocha LF, Gomes D, Silva J, Veiga G. Online inspection system based on machine learning techniques: real case study of fabric textures classification for the automotive industry. J Intell Manuf 2016;30(1):351–61. https://doi.org/10.1007/s10845-016-1254-6.
[36] Basly H, Ouarda W, Sayadi FE, Ouni B, Alimi AM. CNN-SVM learning approach based human activity recognition. Lecture notes in computer science. Springer International Publishing; 2020. p. 271–81. https://doi.org/10.1007/978-3-030-51935-3_29.
[37] Xu J, Ma L, Zhang W, Yang Q, Li X, Liu S. An improved hybrid CNN-SVM based method for bearing fault diagnosis under noisy environment. In: 2019 Chinese control and decision conference (CCDC); 2019. https://doi.org/10.1109/ccdc.2019.8832683.
[38] van der Maaten L, Hinton G. Visualizing data using t-sne. J Mach Learn Res 2008;9(86):2579–605. http://jmlr.org/papers/v9/vandermaaten08a.html.
[39] Lee J, Lee YC, Kim JT. Fault detection based on one-class deep learning for manufacturing applications limited to an imbalanced database. J Manuf Syst 2020;57:357–66. https://doi.org/10.1016/j.jmsy.2020.10.013.
[40] Mohamed O, Khalid EA, Mohammed O, Brahim A. Content-based image retrieval using convolutional neural networks. Advances in intelligent systems and computing. Springer International Publishing; 2018. p. 463–76. https://doi.org/10.1007/978-3-319-91337-7_41.
[41] Sun J, Wu Z, Yin Z, Yang Z. SVM-CNN-based fusion algorithm for vehicle navigation considering atypical observations. IEEE Signal Process Lett 2019;26(2):212–6. https://doi.org/10.1109/lsp.2018.2885511.
[42] Long J, Mou J, Zhang L, Zhang S, Li C. Attitude data-based deep hybrid learning architecture for intelligent fault diagnosis of multi-joint industrial robots. J Manuf Syst 2020. https://doi.org/10.1016/j.jmsy.2020.08.010.
[43] Zhang K, Chen J, Zhang T, Zhou Z. A compact convolutional neural network augmented with multiscale feature extraction of acquired monitoring data for mechanical intelligent fault diagnosis. J Manuf Syst 2020;55:273–84. https://doi.org/10.1016/j.jmsy.2020.04.016.
[44] Meng S, Pan R, Gao W, Zhou J, Wang J, He W. A multi-task and multi-scale convolutional neural network for automatic recognition of woven fabric pattern. J Intell Manuf 2020;32(4):1147–61. https://doi.org/10.1007/s10845-020-01607-9.
[45] Lee M, Jeon J, Lee H. Explainable AI for domain experts: a post hoc analysis of deep learning for defect classification of TFT-LCD panels. J Intell Manuf 2021. https://doi.org/10.1007/s10845-021-01758-3.
[46] Yun JP, Shin WC, Koo G, Kim MS, Lee C, Lee SJ. Automated defect inspection system for metal surfaces based on deep learning and data augmentation. J Manuf Syst 2020;55:317–24. https://doi.org/10.1016/j.jmsy.2020.03.009.
[47] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks; 2016. ICLR; 2016. arxiv:1511.06434.
[48] Brownlee J. Generative adversarial networks with python: deep learning generative models for image synthesis and image translation. Machine Learning Mastery; 2019. https://books.google.de/books?id=YBimDwAAQBAJ.
[49] Humeau-Heurtier A. Texture feature extraction methods: a survey. IEEE Access 2019;7:8975–9000. https://doi.org/10.1109/access.2018.2890743.
[50] de Mesquita Sá Junior JJ, Backes AR. A gravitational model for grayscale texture classification applied to the pap-smear database. Image Analysis and Processing – ICIAP 2015 2015:332–9. https://doi.org/10.1007/978-3-319-23234-8_31.
[51] de Mesquita Sá Junior JJ, Backes AR. A simplified gravitational model for texture analysis. Comput Anal Images Patterns 2011:26–33. https://doi.org/10.1007/978-3-642-23672-3_4.
[52] Francos J, Meiri A, Porat B. A unified texture model based on a 2-d wold-like decomposition. IEEE Trans Signal Process 1993;41(8):2665–78. https://doi.org/10.1109/78.229897.
[53] Azami H, Escudero J, Humeau-Heurtier A. Bidimensional distribution entropy to analyze the irregularity of small-sized textures. IEEE Signal Process Lett 2017;24:1338–42. https://doi.org/10.1109/lsp.2017.2723505.
[54] Cimpoi M, Maji S, Kokkinos I, Vedaldi A. Deep filter banks for texture recognition, description, and segmentation. Int J Comput Vis 2016;118(1):65–94. https://doi.org/10.1007/s11263-015-0872-3.
[55] de Mesquita Sá Junior JJ, Backes AR. ELM based signature for texture classification. Pattern Recognit 2016;51:395–401. https://doi.org/10.1016/j.patcog.2015.09.014.
[56] Sharma M, Ghosh H. Histogram of gradient magnitudes: a rotation invariant texture-descriptor. 2015. https://doi.org/10.1109/icip.2015.7351681.
[57] Bennett J, Khotanzad A. Modeling textured images using generalized long correlation models. IEEE Trans Pattern Anal Mach Intell 1998;20(12):1365–70. https://doi.org/10.1109/34.735810.
[58] Wu W-R, Wei S-C. Rotation and gray-scale transform-invariant texture classification using spiral resampling, subband decomposition, and hidden markov model. IEEE Trans Image Process 1996;5(10):1423–34. https://doi.org/10.1109/83.536891.
[59] Backes AR, Casanova D, Bruno OM. Texture analysis and classification: a complex network-based approach. Inf Sci 2013;219:168–80. https://doi.org/10.1016/j.ins.2012.07.003.
[60] de Mesquita Sá JJ, Backes AR, Cortez PC. Texture analysis and classification using shortest paths in graphs. Pattern Recognit Lett 2013;34(11):1314–9. https://doi.org/10.1016/j.patrec.2013.04.013.
[61] Backes AR, Martinez AS, Bruno OM. Texture analysis using graphs generated by deterministic partially self-avoiding walks. Pattern Recognit 2011;44(8):1684–9. https://doi.org/10.1016/j.patcog.2011.01.018.
[62] Thewsuwan S, Horio K. Texture classification by local spatial pattern maping based on complex network model. Int J Innov Comput Inf Control 2018. https://doi.org/10.24507/ijicic.14.03.1113.
[63] Bashier HK, Hoe LS, Hui LT, Azli MF, Han PY, Kwee WK, et al. Texture classification via extended local graph structure. Optik 2016;127(2):638–43. https://doi.org/10.1016/j.ijleo.2015.10.096.
[64] Chanyagorn P, Eom K. Texture segmentation using moving average modeling approach. 2000. https://doi.org/10.1109/icip.2000.899241.
[65] Silva LEV, Filho ACSS, Fazan VPS, Felipe JC, Junior LOM. Two-dimensional sample entropy: assessing image texture through irregularity. Biomed Phys Eng Express 2016;2(4). https://doi.org/10.1088/2057-1976/2/4/045002.
[66] Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern SMC-3 1973;(6):610–21. https://doi.org/10.1109/TSMC.1973.4309314.
[67] Ojala T, Pietikainen M, Harwood D. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. 1994. https://doi.org/10.1109/icpr.1994.576366.
[68] Tamura H, Mori S, Yamawaki T. Textural features corresponding to visual perception. IEEE Trans Syst Man Cybern 1978;8(6):460–73. https://doi.org/10.1109/tsmc.1978.4309999.
[69] Zhang J, Liang J, Zhang C, Zhao H. Scale invariant texture representation based on frequency decomposition and gradient orientation. Pattern Recognit Lett 2015;51:57–62. https://doi.org/10.1016/j.patrec.2014.08.002.
[70] Maani R, Kalra S, Yang Y-H. Noise robust rotation invariant features for texture classification. Pattern Recognit 2013;46(8):2103–16. https://doi.org/10.1016/j.patcog.2013.01.014.
[71] Riaz F, Hassan A, Rehman S, Qamar U. Texture classification using rotation- and scale-invariant gabor texture features. IEEE Signal Process Lett 2013;20(6):607–10. https://doi.org/10.1109/lsp.2013.2259622.
[72] Manjunath B, Ma W. Texture features for browsing and retrieval of image data. IEEE Trans Pattern Anal Mach Intell 1996;18(8):837–42. https://doi.org/10.1109/34.531803.
[73] Dong Y, Feng J, Liang L, Zheng L, Wu Q. Multiscale sampling based texture image classification. IEEE Signal Process Lett 2017;24(5):614–8. https://doi.org/10.1109/lsp.2017.2670026.

[74] Chorás RS. Image feature extraction techniques and their applications for cbir and biometrics systems. Int J Biol Biomed Eng 2007:349–56.
[75] Su X, Lin W, Zheng X, Han X, Chu H, Zhang X. A new local-main-gradient-orientation HOG and contour differences based algorithm for object classification. 2013. https://doi.org/10.1109/iscas.2013.6572483.
[76] Baber J, Dailey MN, Satoh S, Afzulpurkar N, Bakhtyar M. BIG-OH: BInarization of gradient orientation histograms. Image Vis Comput 2014;32(11):940–53. https://doi.org/10.1016/j.imavis.2014.08.006.
[77] Fan B, Wu F, Hu Z. Rotationally invariant descriptors using intensity order pooling. IEEE Trans Pattern Anal Mach Intell 2012;34(10):2031–45. https://doi.org/10.1109/tpami.2011.277.
[78] Lowe DG. Distinctive image features from scale-invariant keypoints. Int J Comput Vis 2004;60:91–110. https://doi.org/10.1023/b:visi.0000029664.99615.94.
[79] Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors. 2004. https://doi.org/10.1109/cvpr.2004.1315206.
[80] Bay H, Ess A, Tuytelaars T, Gool LV. Speeded-up robust features (SURF). Comput Vis Image Underst 2008;110(3):346–59. https://doi.org/10.1016/j.cviu.2007.09.014.
[81] Dalal N, Triggs B. Histograms of oriented gradients for human detection. 2005. https://doi.org/10.1109/cvpr.2005.177.
[82] Heikkilä M, Pietikä inen M, Schmid C. Description of interest regions with local binary patterns. Pattern Recognit 2009;42(3):425–36. https://doi.org/10.1016/j.patcog.2008.08.014.
[83] Nixon MS, Aguado AS. Feature extraction and image processing for computer vision. Elsevier; 2002. https://doi.org/10.1016/c2017-0-02153-5.
[84] Kadhim AI. Survey on supervised machine learning techniques for automatic text classification. Artif Intell Rev 2019;52(1):273–92. https://doi.org/10.1007/s10462-018-09677-1.
[85] Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, et al. Feature selection: a data perspective; 2017. ACM Computing Surveys; 2017. https://doi.org/10.1145/3136625. arXiv:1601.07996v5.
[86] Jovic A, Brkic K, Bogunovic N. A review of feature selection methods with applications. In: 2015 38th international convention on information and communication technology electronics and microelectronics (MIPRO); 2015. https://doi.org/10.1109/mipro.2015.7160458.
[87] Sheikhpour R, Sarram MA, Gharaghani S, Chahooki MAZ. A survey on semi-supervised feature selection methods. Pattern Recognit 2017;64:141–58. https://doi.org/10.1016/j.patcog.2016.11.003.
[88] Zhang R, Nie F, Li X, Wei X. Feature selection with multi-view data: a survey. Inf Fusion 2019;50:158–67. https://doi.org/10.1016/j.inffus.2018.11.019.
[89] Kamalov F, Thabtah F. A feature selection method based on ranked vector scores of features for classification. Ann Data Sci 2017;4(4):483–502. https://doi.org/10.1007/s40745-017-0116-1.
[90] Thaseen IS, Kumar CA. Intrusion detection model using fusion of chi-square feature selection and multi class SVM. J King Saud Univ – Comput Inf Sci 2017;29(4):462–72. https://doi.org/10.1016/j.jksuci.2015.12.004.
[91] Mesleh AMA. Chi square feature extraction based svms arabic language text categorization system. J Comput Sci 2007:430–5.
[92] Liu L, Kang J, Yu J, Wang Z. A comparative study on unsupervised feature selection methods for text clustering. In: International conference on natural language processing and knowledge engineering; 2005. https://doi.org/10.1109/nlpke.2005.1598807.
[93] Wang H, Hong M. Distance variance score: an efficient feature selection method in text classification. Math Probl Eng 2015:1–10. https://doi.org/10.1155/2015/695720. https://www.hindawi.com/journals/mpe/2015/695720/.
[94] Kumar M, Rath NK, Swain A, Rath SK. Feature selection and classification of microarray data using MapReduce based ANOVA and k-nearest neighbor. Proc Comput Sci 2015;54:301–10. https://doi.org/10.1016/j.procs.2015.06.035.
[95] Madhavi Bharatbhai Desai BP, Patel SV. Anova and fisher criterion based feature selection for lower dimensional universal image steganalysis. Int J Image Process 2016;10:145–60.
[96] Senliol B, Gulgezen G, Yu L, Cataltepe Z. Fast correlation based filter (fcbf) with a different search strategy. 23rd international symposium on computer and information sciences 2008:1–4. https://doi.org/10.1109/ISCIS.2008.4717949. https://ieeexplore.ieee.org/document/4717949.
[97] Nguyen H, Franke K, Petrovic S. Optimizing a class of feature selection measures. NIPS 2009 workshop on discrete optimization in machine learning 2009. https://www.researchgate.net/publication/231175763_Optimizing_a_Class_of_Feature_Selection_Measures/citation/download.
[98] Huang N, Lu G, Xu D. A permutation importance-based feature selection method for short-term electricity load forecasting using random forest. Energies 2016;9(10):767. https://doi.org/10.3390/en9100767.
[99] Zien A, Kraemer N, Sonnenburg S, Raetsch G. The feature importance ranking measure. ECML PKDD 2009: machine learning and knowledge discovery in databases 2009. arXiv:0906.4258v1.
[100] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–30. https://www.jmlr.org/papers/v12/pedregosa11a.html.
[101] Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. J R Stat Soc Ser B (Stat Methodol) 2006;68(1):49–67. https://doi.org/10.1111/j. 1467-9868.2005.00532.x.
[102] He X, Cai D, Niyogi P. Laplacian score for feature selection. NIPS'05: proceedings of the 18th international conference on neural information processing systems 2005.
[103] Gajawada SK. Anova for feature selection in machine learning, towards data science. 2021 (accessed 03.06.21 (October 2019)), https://towardsdatascience.com/anova-for-feature-selection-in-machine-learning-d9305e228476.
[104] Chandra MA, Bedi SS. Survey on SVM and their application in image classification. Int J Inf Technol 2018. https://doi.org/10.1007/s41870-017-0080-1.
[105] Abe S. Support vector machines for pattern classification (advances in pattern recognition). Secaucus, NJ, USA: Springer-Verlag New York, Inc; 2010.
[106] Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2013;2:1–27.
[107] Khan A, Sohail A, Zahoora U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev 2020. https://doi.org/10.1007/s10462-020-09825-6.
[108] Vasilev I, Slater D, Spacagna G. Python deep learning. Second Ed. Packt Publishing; 2019. https://www.ebook.de/de/product/35342345/ivan_vasilev_daniel_slater_gianmario_spacagna_python_deep_learning_second_edition.html.
[109] Chen M, Jiang M, Liu X, Wu B. Intelligent inspection system based on infrared vision for automated fiber placement. In: 2018 IEEE international conference on mechatronics and automation (ICMA); 2018. https://doi.org/10.1109/icma.2018.8484646.
[110] Müller K-R. Understanding ML models. Technical University of Berlin; 2019. http://helper.ipam.ucla.edu/publications/mlpws3/mlpws3_15932.pdf.
[111] Shrikumar A, Greenside P, Kundaje A. Learning important features through propagating activation differences. In: Precup D, Teh YW, editors. Proceedings of the 34th international conference on machine learning, vol. 70 of proceedings of machine learning research; 2017. Sydney, Australia: PMLR, International Convention Centre; 2017. p. 3145–53. http://proceedings.mlr.press/v70/shrikumar17a.html, arXiv:1704.02685v2.
[112] Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. Proceedings of the 34th international conference on machine learning 2017.
[113] Smilkov D, Thorat N, Kim B, Martin Wattenberg FV. Smoothgrad: removing noise by adding noise. 2017. arXiv:1706.03825.
[114] Lundberg S, Lee S-I. A unified approach to interpreting model predictions. Advances in neural information processing systems 30 (NIPS 2017). 2017.
[115] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. European conference on computer vision 2014.
[116] Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M. Striving for simplicity: the all confolutional net. ICLR 2015 2015.
[117] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localizatio. Proceedings of the IEEE international conference on computer vision 2017.
[118] Montavon G, Binder A, Lapuschkin S, Samek W, Müller K-R. Layer-wise relevance propagation: an overview. Explainable AI: interpreting, explaining and visualizing deep learning 2019:193–209. https://doi.org/10.1007/978-3-030-28954-6_10.
[119] Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualising image classification models and saliency maps. 2013 (CoRR abs/1312.6034 (2013)), http://dblp.uni-trier.de/db/journals/corr/corr1312.html#SimonyanVZ13.
[120] Ribeiro MT, Singh S, Guestrin C. Model-agnostic interpretability of machine learning. ICML workshop on human interpretability in machine learning 2016.
[121] Fong RC, Vedaldi A. Interpretable explanations of black boxes by meaningful perturbation. Proceedings of the IEEE international conference on computer vision (ICCV) 2017. https://openaccess.thecvf.com/content_iccv_2017/html/Fong_Interpretable_Explanations_of_ICCV_2017_paper.html.
[122] Kindermans P-J, Schütt KT, Alber M, Müller K-R, Erhan D, Kim B, et al. Learning how to explain neural networks: patternNet and patternAttribution. 2017. arXiv:1705.05598.
[123] Yeh C-K, Hsieh C-Y, Suggala AS. On the (in)fidelity and sensitivity of explanations. NeurIPS; 2019.
[124] Luca Massaron AB. Python data science essentials. Packt Publishing; 2018. https://www.ebook.de/de/product/34617776/luca_massaron_alberto_boschetti_python_data_science_essentials.html.
[125] Quandoo G. Mastering machine learning algorithms: expert techniques to implement popular machine learning algorithms and fine-tune your models. Birmingham: Packt Publishing; 2018.

Sebastian Meister is a PhD Candidate in optical inspection at the Delft University of Technology, The Netherlands, at the Aerospace Non-Destructive Testing Laboratory. He is also working as a Researcher at the German Aerospace Center, Center for Lightweight-Production-Technology in Stade, Germany. In 2017 he received his Master's degree in Mechanical Engineering from Friedrich-Alexander-University of Erlangen-Nürnberg. His field of research is the computer vision and machine learning for the optical inspection in automated composite manufacturing.

Mahdieu Wermes is a Master student at the Ilmenau University of Technology and a working student at the German Aerospace Center, Center for Lightweight-Production-Technology in Stade. In 2021 he received his Bachelors's degree in Electrical Engineering from the TU Ilmenau. As a working student he focuses on machine learning techniques for automated industrial inspection.

Dr. Jan Stüve is Associate Professor in composite manufacturing at the Delft University of Technology, The Netherlands. His PhD was on weave technologies from the RWTH Aachen University, Germany. After he was working as a researcher at RWTH Aachen, he became the managing director of Bergal Erfurter Flechttechnik GmbH. Since 2016, he has been

Head of Department, Composite Process Technology at the German Aerospace Center, Center for Lightweight-Production-Technology in Stade.

Dr. Roger M. Groves is Associate Professor in Aerospace NDT/SHM and Heritage Diagnostics at Delft University of Technology, The Netherlands. His PhD is in Optical Instrumentation from Cranfield University (2002) and he was a Senior Scientist at Institute for Applied Optics, University of Stuttgart, before joining TU Delft in 2008 as an Assistant Professor. Dr Groves heads a team of approximately 20 researchers in the Aerospace NDT Laboratory at TU Delft. His research interests are Optical Metrology, Fibre Optic Sensing and Ultrasonic Wave Propagation in Composite Materials. He has approximately 200 journal and conference publications in these topics. In 2020 he was awarded Fellow of SPIE.