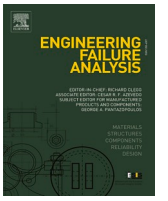
Engineering Failure Analysis 175 (2025) 109551

Contents lists available at ScienceDirect

Engineering Failure Analysis

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



Prediction of statistical force–displacement curves of Charpy-V impact tests based on unsupervised fracture surface machine learning

Johannes Rosenberger [a,b,*], Johannes Tlatlik [a], Nils Rump [a], Sebastian Münstermann [b]

[a] Fraunhofer Institute for Mechanics of Materials, Wöhlerstraße 11, D-79108 Freiburg, Germany
[b] Institute of Metal Forming, RWTH Aachen University, Intzestraße 10, D-52072 Aachen, Germany

ARTICLE INFO

Keywords:
Computer vision
K-means clustering
explainable AI
Quantitative fractography
Charpy Impact Test

ABSTRACT

While conventional pendulum impact tests only measure a material's integral energy absorption, the instrumented version of the test provides valuable additional insights by extracting force–displacement behaviour of the loaded specimen. The latter, however, requires auxiliary testing equipment, calibration procedures and evaluation methods. Therefore, this study aims to predict force–displacement behaviour of instrumented Charpy impact tests solely on the basis of analyzing images of specimen fracture surfaces postmortem. This is explored and achieved by using unsupervised machine learning techniques for computer vision. By using unsupervised computer vision on fracture images from 4 steels, we assess the feasibility of classifying fracture surfaces and deriving statistical force–displacement curves and provide crucial interpretability of the model's decision making. The results indicate the model's ability to learn the necessary representations without the need of supervision. The unsupervised model can extract significant insights from fracture images alone, supporting efficient, accurate, and interpretable materials testing, where confidence intervals of 97 % can already be met for the upper shelf. This allows detailed information about the mechanical behaviour of the material to be obtained from non-instrumented tests.

1. Introduction

Pendulum or Charpy impact testing is a widely used test method to determine the toughness of a material under impact loading by measuring the amount of absorbed energy by the material [1–3]. This is of particular importance for materials used in safety–critical components in nuclear power and safety industries such as reactor pressure vessels (RPV), piping and pipeline steels to avoid brittle failure. While conventional pendulum impact tests only measure a material's integral energy absorption, the instrumented version of the test provides valuable additional insights by extracting force–displacement behaviour of the loaded specimen. The instrumented Charpy V-notch pendulum impact test [4,5], however, requires auxiliary testing equipment, calibration procedures and evaluation methods. To complement the well-established and well-defined methods, this study aims to investigate whether the prediction of force–displacement curves from fracture surfaces alone is possible using unsupervised learning in order to avoid these additional

* Corresponding author at: Fraunhofer Institute for Mechanics of Materials, Wöhlerstraße 11, D-79108 Freiburg, Germany.
E-mail address: johannes.rosenberger@iwm.fraunhofer.de (J. Rosenberger).

https://doi.org/10.1016/j.engfailanal.2025.109551
Received 30 January 2025; Received in revised form 18 March 2025; Accepted 21 March 2025
Available online 22 March 2025

Nomenclature

A	Elongation at fracture
CV	Coefficient of variation
F	Force
F_a	Crack arrest force
F_{gy}	General yield force
F_{iu}	Unstable crack initiation force
F_m	Maximum force
l	Arc-length
$n_{specimens}$	Number of specimens
s	Displacement
s_a	Crack arrest displacement
s_{gy}	General yield displacement
s_{iu}	Displacement at unstable crack initiation
s_m	Displacement at maximum force
s_t	Total displacement
$s(t)$	Displacement of test piece at time t
S	Distance between anvils (span)
t	Time
T_0	Master Curve Reference Temperature ASTM E1921
T_{28J}	28 J transition temperature Charpy impact test
UTS	Ultimate tensile strength
v_0	Initial striker impact velocity
$v(t)$	Striker impact velocity at time t
W_a	Crack arrest energy
W_{iu}	Energy at unstable crack initiation
W_m	Energy at maximum force
W_t	Total impact energy
σ	Standard deviation
σ_y	Yield strength
Grad-CAM	Gradient-weighted Class Activation Mapping
HAZ	Heat-affected zone
PCA	Principal component analysis
RPV	Reactor pressure vessel
YOLO	You only look once

expenditures.

Although supervised machine learning approaches have proven to be very efficient, the requirement of an extensive and time-costly data labeling process constitutes a drawback and the reliance of the annotation process on human input can potentially introduce bias or errors [6]. In materials science the supervised computer vision is frequently used for classification of microstructures [7–9] and fracture mechanisms [10,11]. Semantic segmentation is used for crack size measurements [12] and fractography [13,14], whereas instance segmentation can be utilized for advanced microstructure analysis in cast iron [15] or martensitic steels [16,17]. In contrast, unsupervised machine learning describes a kind of machine learning where the model learns from data without human supervision. In this context the need for interpretability of the decision making is key for a trustworthy use [18,19]. In materials science, explainable AI is an eminently important to increase trust in such approaches [20–22]. Also, it intends to increase adaptation to challenges in the industrial context [9,23]. Already, unsupervised computer vision is used to predict fatigue crack growth metrics [24] and pore-scale structural analysis of fractured porous media [25]. These types of models also proved applicable to classify steel surface defects image data [26] or segment microstructures by mimicking metallurgists pattern recognition [27]. In fault detection, state of the art unsupervised models are used for wind turbines [28] as well as defect detection on an atomic scale [29]. Furthermore, Yu et al. [30] used unsupervised models for calibration of micromechanical fracture models of ultra-low cycle fatigue in steel structures. An overview of computer vision and machine learning for microstructural characterization and analysis was given by Holm et al. [31]. For impact testing, machine learning has been used to predict the total impact energy based on chemical composition, heat treatment, manufacturing parameters and mechanical properties [32,33]. Bianco et al. [34] trained deep convolutional neural networks to correlate the fractographic images with quantitative toughness values. Lee et al. [35] utilized k-means clustering to predict the ductile fracture area. However, no in-depth correlation between fractographic features and mechanical behaviour has been investigated. This is what our study tries to provide.

We therefore decided to utilize unsupervised computer vision on a database of Charpy V-notch fracture surface images tested in previous projects [36–40]. The database consists of a ferritic-bainitic reactor pressure vessel steel of type 22NiMoCr3-7 (see [41]),

including the heat-affected zone (HAZ), S3NiMo1 weld (see [42]) metal of this RPV and a quenched and tempered steel of type 20MnMoNi5-5 for comparison. Three main questions shall be addressed:

1) Is such a model able to cluster Charpy-V notch fracture surfaces without supervision (e.g., into lower or upper shelf experiments)?
2) Does the model allow us to go beyond and infer previously non-tangible information, like Charpy Impact Energy or force–displacement curve, from simply using fracture surface image data?
3) Can we determine which features of the fracture surface the model uses to identify the clusters and predict mechanical behaviour? Also, we would like to ascertain if this selection process can be comprehended by a human domain expert.

Our approach combines object detection-based data preprocessing, as well as feature extraction and reduction using state-of-the-art methods. To provide interpretability, we implemented a gradient-weighted class activation mapping (Grad-CAM) [19]. Selvaraju et al. [19] introduced this method to provide the possibility of good visual explanation. By providing an explainable reasoning of the model's decision, the trust in the proposed method shall be increased. Correlations and newfound causalities become reasonable and show which parts of the input data contribute most to a particular feature representation or cluster assignment. Thus, this study explores the inference of statistical force–displacement corridors [43,44] based on unsupervised fractography for Charpy-V notched pendulum impact testing. Our approach could be used to derive the materials energy absorption characteristics without the need for instrumentation, heavily simplifying the experimental setup and reducing costs. Additionally, we assess the potential use of these models in calibration of pendulum impact testing machines.

2. Materials and methods

2.1. Instrumented Charpy V-notch pendulum impact test

In comparison to the Charpy pendulum impact test described in ASTM E23 [1] or EN ISO 148–1 [2], the instrumented Charpy V-notch pendulum impact test [4,5] provides further information on the fracture behaviour of the material under impact testing conditions by measuring the force F at the time t after the beginning of the deformation at t_0. Instrumented pendulum impact tests are used to determine the toughness of a material under impact loading by measuring the amount of energy the material is able to absorb and characteristic values such as the maximum force F_m, unstable crack initiation force F_{iu} or crack arrest force F_a can be analyzed [4]. The test method is especially valuable in understanding how materials behave under dynamic loading conditions, such as sudden impacts, at various temperatures. The characteristic force, displacement and energy values are defined by their suffix, which are defined in Table 1. In general, a fit through the oscillations yields reliable characteristic values.

The general experimental workflow is shown in Fig. 1. In the instrumented Charpy V-notch impact test, a V-notched specimen is struck by a pendulum hammer, causing it to fracture. The force–displacement curve can be calculated as follows. The measured force–time relationship is proportional to the acceleration characteristic. By double numerical integration, the displacement of test piece at time $s(t)$ is calculated as

$$v(t) = v_0 - \frac{1}{m}\int_0^t F(t)dt, \tag{1}$$

$$s(t) = \int_0^t v(t)dt, \tag{2}$$

where m is the effective mass of the (assumed) rigid pendulum, v_0 is the initial impact velocity. The characteristic values of impact energy can then be calculated by integrating the F-s curve from

$s = 0$ mm to the characteristic displacement (e.g., the crack arrest energy W_a at s_a). By testing at different temperatures, defined characteristic transition temperatures from ductile-to-brittle behaviour (ductile-to-brittle transition temperature) can be found.

According to EN ISO 14556 [4], six characteristic types of F-s curve types can be distinguished (A to F). In turn ASTM E23 [1] divides the curves into three distinct regimes: lower shelf, upper shelf and ductile-to-brittle transition. In our database curves of type B, C, D, E and F are given. Fig. 2 shows exemplary F-s curves for each type in our database. In contrast to EN ISO 14556 [4] we do not define type B to show crack arrest. For the lower shelf experiments (type B and C), only unstable crack propagation occurs, the material behaviour is exclusively brittle and small displacements are observed. Both stable and unstable crack propagation, including crack arrest, can occur for types D and E (ductile-to-brittle transition). The upper shelf curves of type F show only stable crack propagation

Table 1
Suffix definitions for characteristic values.

Suffix	Definition
gy	General yield force / displacement
m	Maximum force, displacement / energy at maximum force
iu	Unstable crack initiation force, displacement / energy at unstable crack initiation
a	Crack arrest force / displacement / energy
t	Total displacement / impact energy

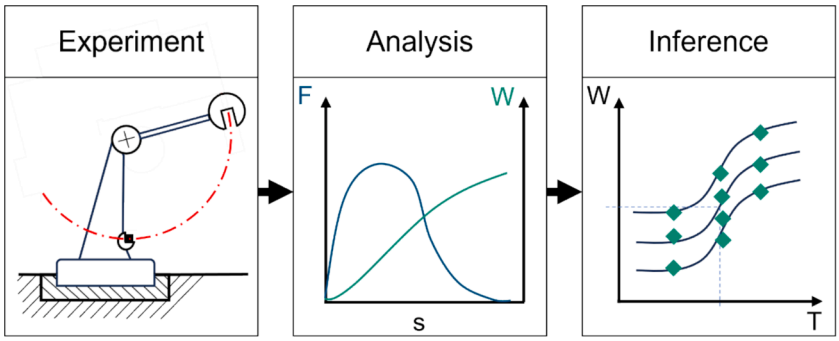


Fig. 1. General experimental workflow.

with exclusively ductile failure.

2.2. Material and experimental data

The RPV material, including the heat-affected zone (HAZ), consists of the ferritic-bainitic steel of type 22NiMoCr3-7. The weld metal is a ferritic filler material of type S3NiMo1, which has been optimized for the submerged arc welding process. In addition, a quenched and tempered steel of type 20MnMoNi5-5 is used for comparison. The mechanical properties of the materials investigated in this study are given in Table 2 and were determined in previous research projects [36–40].

22NiMoCr3-7 and 20MnMoNi5-5 show similar yield and ultimate strength (σ_y and *UTS*), but the latter exhibits higher toughness in terms of 28 J transition temperature and Master Curve reference temperature T_0. The weld's yield strength and *UTS* are higher, whereas toughness is between that of 22NiMoCr3-7 and 20MnMoNi5-5. No tensile tests were performed for the HAZ, but toughness is very high as indicated by T_{28J}. In Fig. 3 the absorbed energy for a V-notch test piece K_V is given in dependency of the test temperature T for all considered materials. For reference, the transition temperature at 50 % of the absorbed energy of the upper shelf $T_{t50\%US}$ is given. The HAZ material shows the lowest transition temperature, while 22NiMoCr3-7 exhibits the highest $T_{t50\%US}$ values.

The specimen geometry is given in Fig. 4 with a specimen length of 55 mm and both the thickness and width are 10 mm. All specimens have been tested with the same experimental setup using a 300 J pendulum. The pendulum has a mass of 20.278 kg with an effective mass of the pendulum corresponding to its effective weight of 1.376 kg. The initial striker impact velocity v_0 is constant with 5.440 m/s. The distance between anvils (span) S is 42 mm. In Fig. 5 a series of tested Charpy-V notch weld material specimens is shown. The displayed fracture surfaces range from almost no plastic deformation to high deformation. In total, 150 specimens were captured. For 17 specimens no experimental data is available. Nevertheless, these specimens respectively their fracture surfaces are valuable input data for the machine learning model.

2.3. Preprocessing of the image data

As the ductile deformation increases with increased energy consumption (Fig. 5), the need for a semi-automated image handling becomes apparent. There are two challenges for the creation of the data set. First, for each specimen two fracture surfaces must be extracted from the image. To solve this, we decided to train a *you only look once* (YOLO) model [45], specifically YOLOv5, to detect the corners of each specimen half and crop them into separate images. Second, as the ductile deformation increases, the specimen does not fully separate in some cases (Fig. 5). For these fracture surfaces the crack plane and image plane are not parallel. To prevent any bias on the subsequent clustering due to this, affine transformations have to be applied to ensure that all parallel lines in the original image are still parallel in the output image. We implemented the affine transformations using OpenCV [46]. In Fig. 6 the initial workflow for the fracture surface detection and subsequent affine transformation is shown. After the affine transformation, all images are resized to a size of 512x512 pixels. This workflow is performed automatically for all 150 input images, resulting in a dataset of 300 fracture surface images. Before training, the images are converted to grayscale to cover for potential bias due to for example variances in the background color.

To artificially increase the size and diversity of our training dataset, we apply augmentations to our image data. The preprocessing and the augmentations (using OpenCV [46] and Torchvision [47]) are summarized in Table 3. The images are flipped using RandomHorizontalFlip and RandomVerticalFlip. To account for variability in the lightning conditions, ColorJitter adjusts image brightness, contrast, saturation, and hue with the specified ranges. All augmentations are applied with a probability of 50 %.

2.4. Unsupervised computer vision model

In contrast to supervised learning as in [7,10,11,13,48], unsupervised learning describes a kind of machine learning where the model learns from data without human supervision. Like this, the model is allowed to discover patterns without human bias, which could potentially introduce errors. Additionally, the requirement of extensive and time-costly data labeling can be neglected.

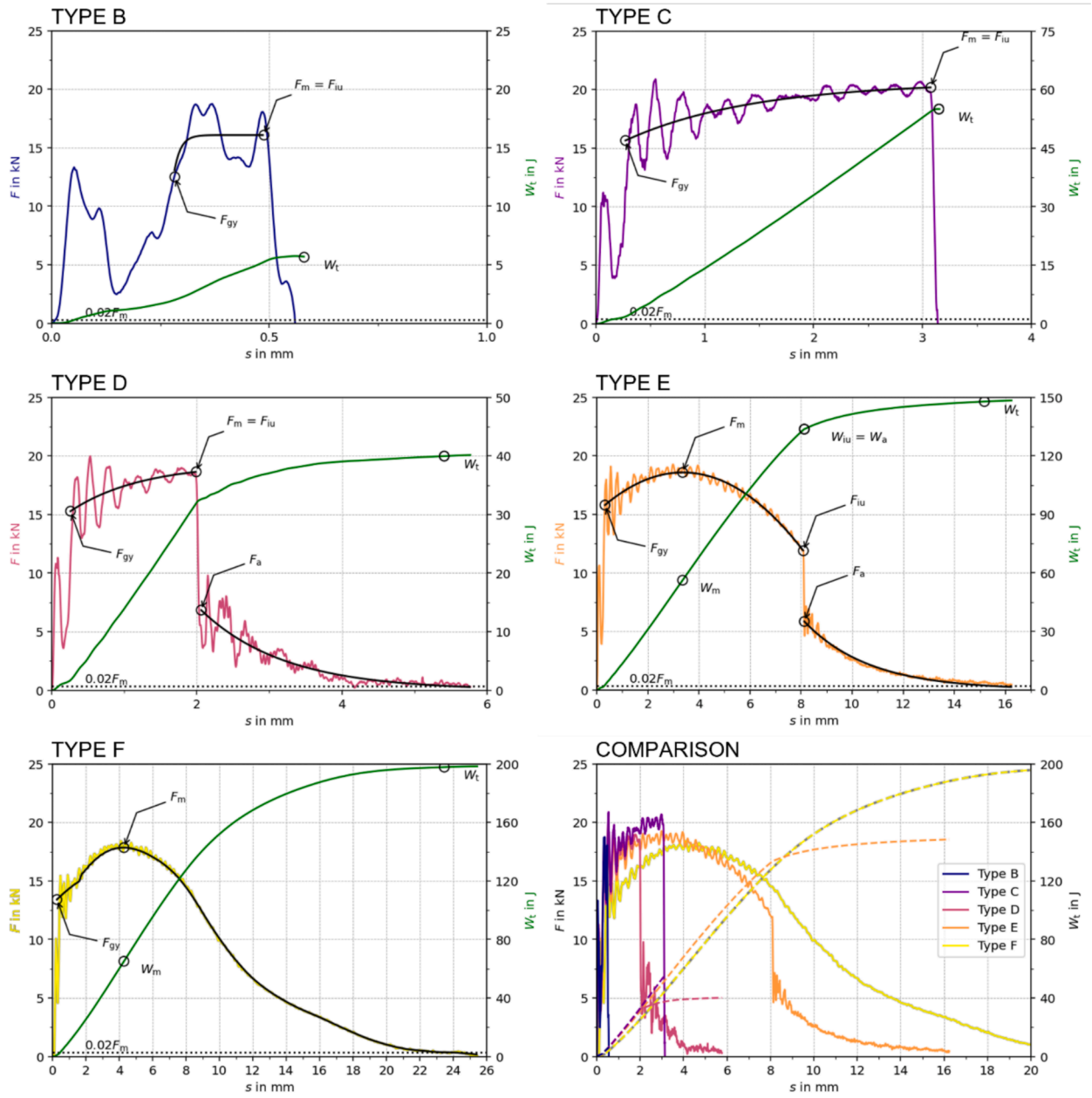


Fig. 2. Characteristic force–displacement curves of types B, C, D, E and F.

Table 2
Mechanical properties of the investigated materials.

Material	σ_y[MPa]	*UTS*[MPa]	*A*[%]	T_{28J}[°C]	T_0[°C]
22NiMoCr3-7	433	584	23	−44	−64
Weld material	557	634	29	−66	−100
HAZ	N.A.	N.A.	N.A.	−131	N.A.
20MnMoNi5-5	447	597	32	−98	−114

The general workflow of our approach is shown in Fig. 7. Our model utilizes a pre-trained ResNet50 architecture [48,49] to process the input images. The model was imported using PyTorch respectively Torchvision [47]. By removing the classification layer, the feature representations can be extracted from the input images. The features are extracted from the final layer of the network. As a

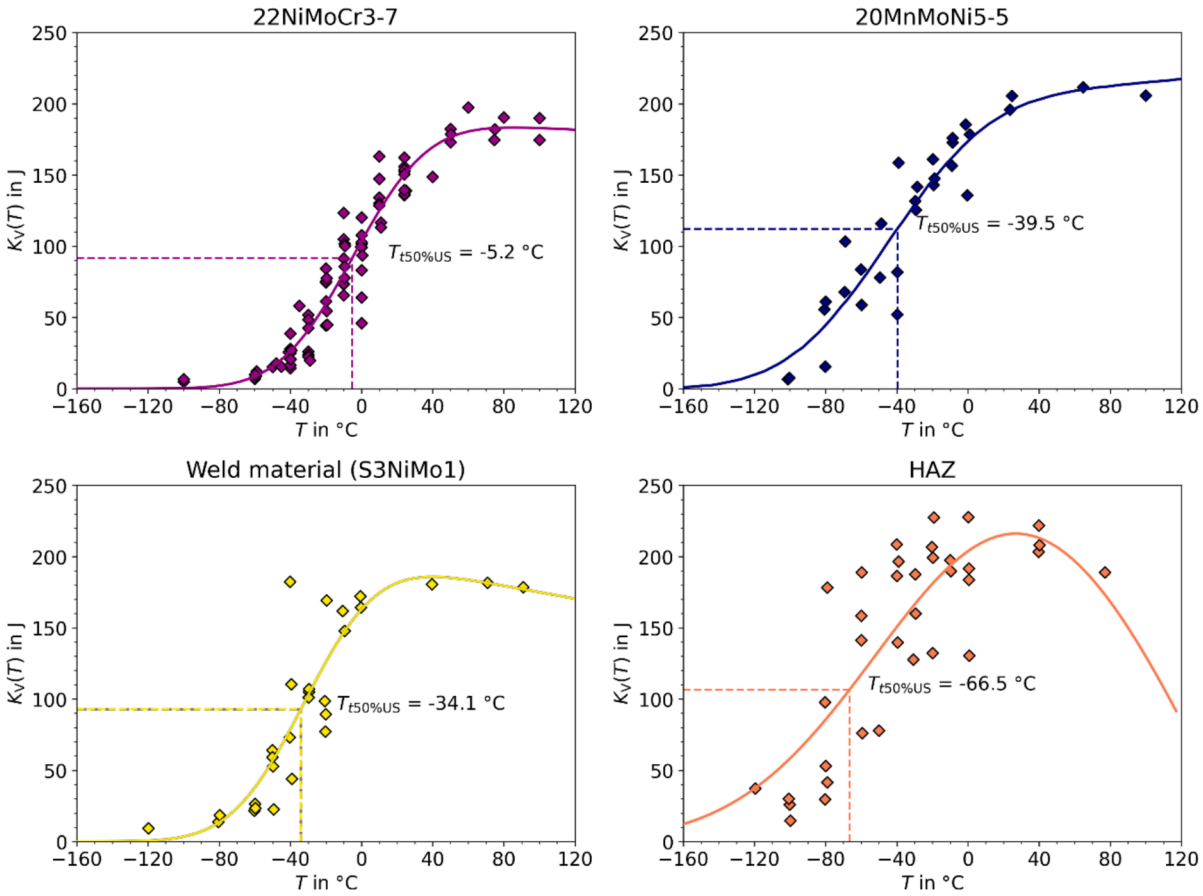


Fig. 3. K_V(T) curves for the considered materials (as specified in ISO 148–1).

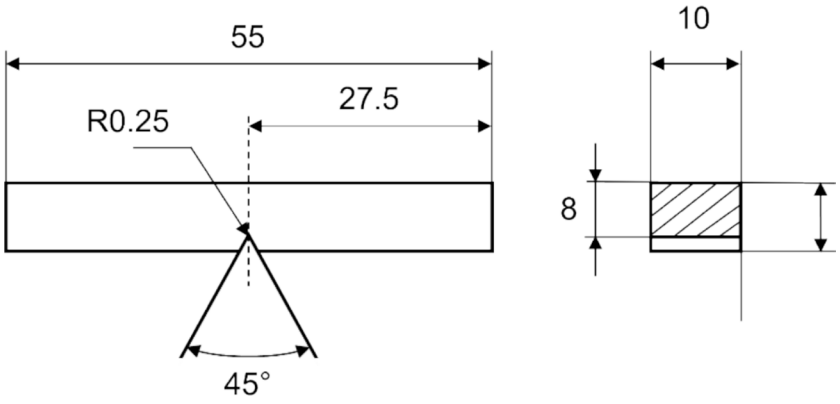


Fig. 4. Charpy V-notch Impact specimen dimensions (all values in mm).

result, a 2048-dimensional feature vector can be extracted for every image. To reduce the dimensionality of the feature vector, a principal component analysis (PCA) is performed, retaining 95 % of the overall feature variance using Scikit-learn [50]. Like this, the dimensionality is reduced to 56 features per image. Our working hypothesis is that the model is able to distinguish between the 5 different *F-s* curve types based on the fracture surfaces. Therefore, a k-means clustering is performed, to group the images into 5 clusters [50]. To visualize the results, the feature vectors are further reduced into two dimensions (PCA components) by performing another PCA.

Fig. 5. Tested Charpy-V notch weld material specimens with an increase in measured total impact energy from left to right.

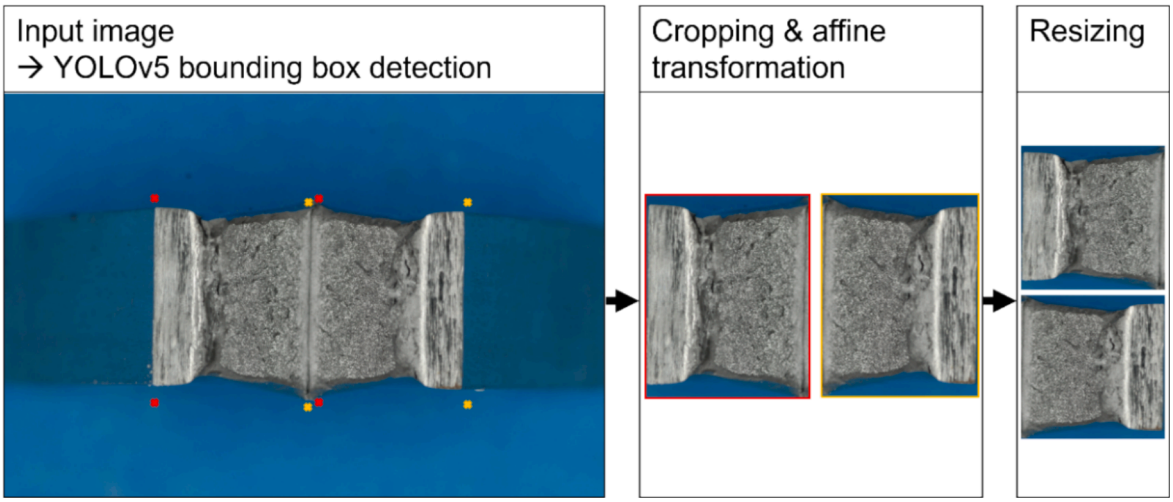


Fig. 6. Initial image data preprocessing.

Table 3
Summary of the preprocessing steps and applied augmentations.

Preprocessing steps	Augmentations
cv2.resize(image, (512, 512)) transforms.RandomGrayscale(p = 1) transforms.Normalize(mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225])	transforms.RandomHorizontalFlip() transforms.RandomVerticalFlip() transforms.ColorJitter(brightness = 0.2, contrast = 0.2, saturation = 0.2, hue = 0.1)

3. Results and discussion

3.1. Clustering

After 21 iterations the k-means clustering yields the results shown on the left in Fig. 8. Each fracture surface half is represented by a scatter point. The clusters are decoded by their color and their marker shape. Throughout this study, the marker shapes remain constant for each cluster (0: ○, 1: ▷, 2: ◇, 3: ◁, 4: □). The 5 clusters are clearly separable from each other. Along the cluster borders only little scattering, for example between cluster 3 and 4, can be observed. As mentioned, no experimental data is available for 17 specimens respectively 34 fracture surfaces. Nevertheless, these specimens respectively their fracture surfaces are valuable input data, as the model has more data available to identify the principal components for clustering. For the following correlation analysis, the corresponding 34 fracture surfaces are removed. The results are displayed on the right in Fig. 8.

The key properties of the remaining 133 experiments are summarized in Fig. 9. Regarding the characteristic force–displacement curve type, the dataset contains 37 experiments associated with lower shelf experiments (type B and C). Most of the experiments (77) belong to the ductile-to-brittle transition region (type D and E) and 19 are associated with the upper shelf region (type F). The material types are equally distributed with the most experiments being available for 22NiMoCr3-7 (52). For the other materials at least 26

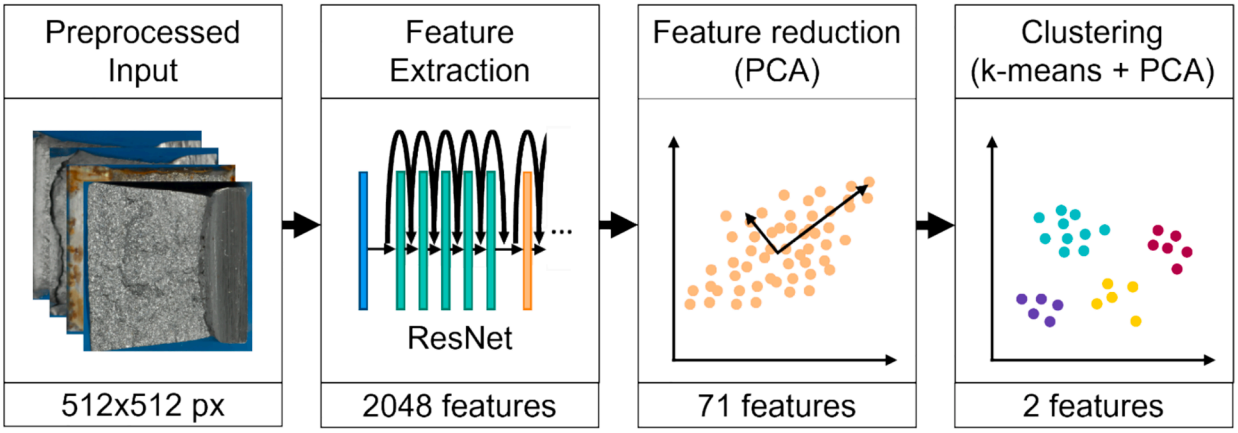


Fig. 7. Workflow for unsupervised clustering.

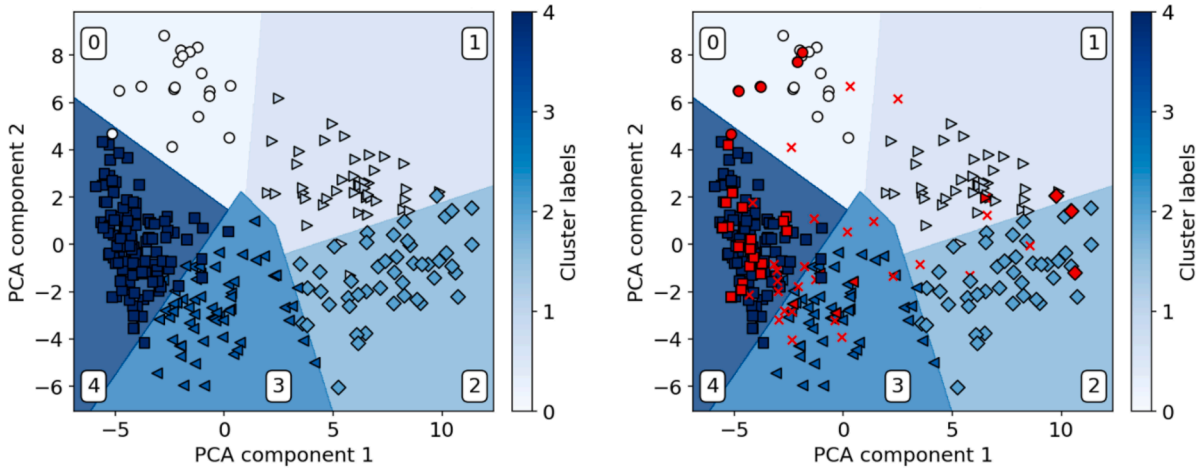


Fig. 8. Left: Results of the k-means clustering (5 labels) for the 300 fracture surface images dataset. Right: Dropouts that cannot be considered in the evaluation due to no experimental data (F-s curves) available () or unequal predictions for the different specimen halves (×).

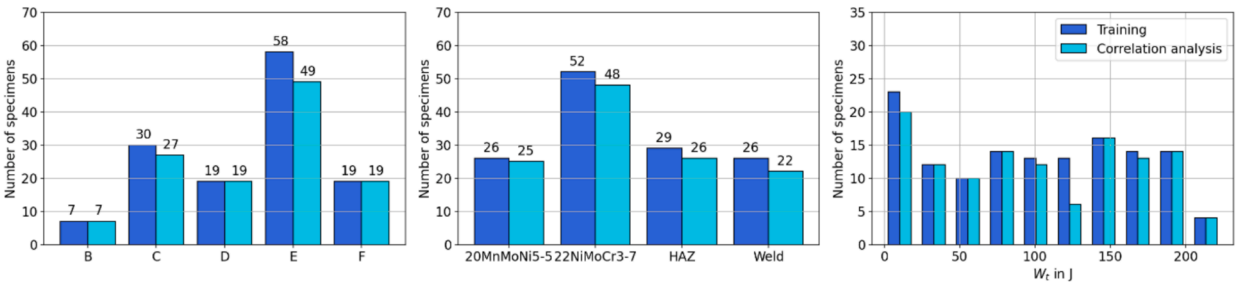


Fig. 9. Dataset composition for training and correlation analysis regarding the experiment type, the material type and the total impact energy.

specimens can be considered. The total impact energy is equally distributed amongst the data. The removed specimens mainly belong to cluster 4.

In addition, the model categorizes corresponding pairs of fracture surfaces of 12 specimens into different clusters. Since there is no ground truth in unsupervised training, these specimens cannot be further evaluated and must be removed for the correlation analysis. The removed fracture surfaces are marked in Fig. 8. The most inter cluster mix-up for corresponding fracture surfaces can be observed for cluster 3 and 4 (8 specimens). For each other cluster combination only one fracture surface pair must be removed. A gradient-weighted class activation mapping (Grad-CAM) analysis shows which parts of the input data contribute most to a particular feature

representation or cluster assignment. Selvaraju et al. [19] invented this method to provide the possibility of good visual explanation by being class-discriminative (localization) and providing a high resolution. In Fig. 10 three Grad-CAM examples for the inter cluster mix-ups are given. All examples have in common that one fracture surface is highly activated (unstable or ductile fracture region), while the corresponding half of the specimen is significantly less activated. Accordingly, the model assigns them different clusters. The remaining 121 experiments are summarized in Fig. 9. In comparison to the unfiltered dataset, no significant change in dataset composition can be observed, as mainly 22NiMoCr3-7 and weld material specimens were removed but the total impact energy distribution remains uniform. Therefore, no correlation can be found between inter cluster mix-ups and the material properties. Also, the cluster size is not equally distributed, as cluster 0 is the smallest with only 6 specimens respectively 12 fracture surface images. Cluster 1, 2 and 3 are equally large with 19, 21 and 24 specimens. Cluster 4, however, is by far the largest, with 51 specimens.

3.2. Correlation analysis

The following correlation analysis corresponds to the 121 specimens where the corresponding pairs of fracture surfaces are categorized in the same clusters. In Fig. 11 the results of the correlation analysis plots regarding the material types (left) and *F-s* curve types (right) are given. Although cluster 0 solely corresponds to weld material, all other clusters show a material type mix. In cluster 1 and 3 the majority of the points correlates to 22NiMoCr3-7 with some weld material specimens present. Cluster 2 mainly responds to HAZ material and 20MnMoNi5-5, while cluster 4 hosts all four material types present in the dataset. Due to the material mix in clusters 2 and 4, the possibility of the model "simply" categorizing the materials, e.g. because of their color or fracture surface features, is ruled out. However, for clusters 0 (weld material) as well as 1 and 3(22NiMoCr3-7) a differentiation of material types is recognizable.

For the *F-s* curve types, a correlation to the clusters can be seen. Cluster 4 represents the upper shelf with curve types F and E. Cluster 3 shows a wide range of ductile-to-brittle transition experiments with 4 only fracture surfaces corresponding to type F. For clusters 0, 1 and 2 all *F-s* types except F are represented. The further the points are located from the center, the more *F-s* curves of type B and C occur. Therefore, these clusters can be correlated to a behaviour in between the ductile-to-brittle transition and the lower shelf.

To further examine this observation, the correlation with the total impact energy W_t was performed in Fig. 12. In doing so, a range of numerical values can be correlated to the clusters qualitatively. For the total impact energy W_t (plot on the left), a distinct gradient can be seen from high W_t values (bright) on the left to low W_t values (dark) on the right. Here, cluster 4 corresponds to upper shelf specimens. Cluster 3 and the left parts of clusters 0, 1 and 2 show mid-range total impact energies. These experiments can therefore be associated with the ductile-to-brittle transition. The scatter points on the right of clusters 0, 1 and 2 show low W_t values and therefore correspond to the lower shelf regime.

To correlate the characteristic values of the *F-s* curves and the clusters, the difference of unstable crack initiation force F_{iu} and crack arrest force F_a denoted as ΔF was calculated. The results are shown in the right plot of Fig. 12. The results are congruent and emphasize the results for the total impact energy, but the gradient is reversed. Upper shelf experiments show only little to no force drop (no unstable crack initiation), as the specimen fail mostly ductile. These specimens are once again located in cluster 4. In the ductile-to-brittle transition region, the unstable crack is initiated, but the crack can be arrested (e.g., *F-s* curve type D). This is resembled by

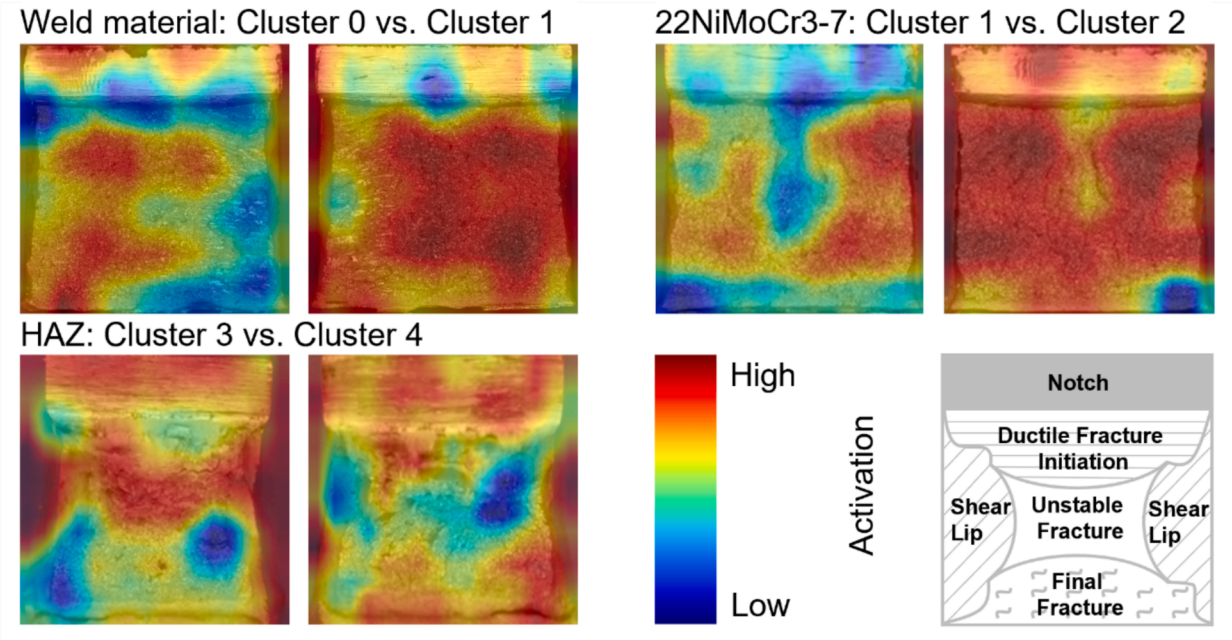


Fig. 10. Three Grad-CAM examples for the inter cluster mix-ups (regions defined according to [1]).

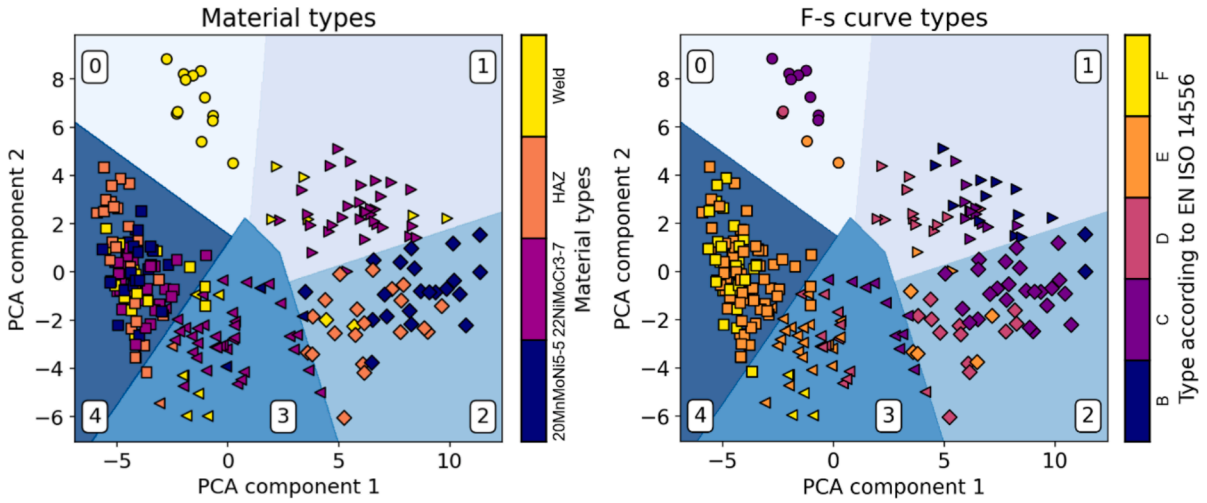


Fig. 11. Correlation analysis for the material types (left) and the F-s curve types (right).

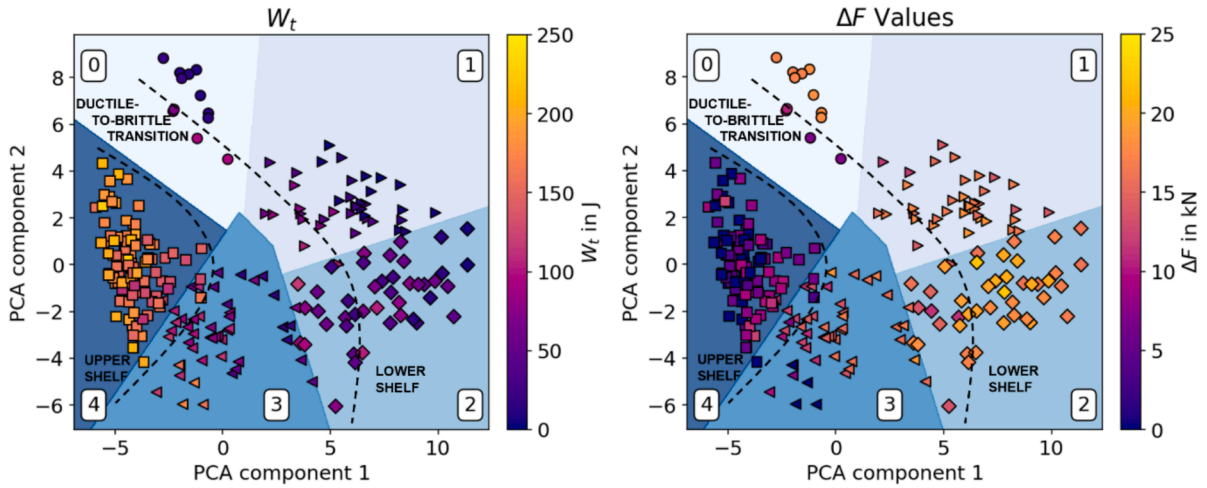


Fig. 12. Correlation analysis for the total impact energy W_t (left) and the difference of unstable crack initiation force F_{iu} and crack arrest force F_a denoted as ΔF (right).

medium range ΔF values. If the unstable crack cannot be arrested, a high drop in force will be measured as there is no crack arrest force measurable. To plot the according scatter points, F_a was set to 0 kN. This applies to the *F-s* curve types B and C. The scatter points are located on the right side of clusters 0, 1 and 2.

Providing explainable reasoning or some kind of interpretability of the model's decision making plays a key role in building trustworthiness for the use of (unsupervised) machine learning models. Based on this, correlations and newfound causalities become reasonable. Therefore, we once more applied Grad-CAM to our model. In Fig. 13 the results of the Grad-CAM analysis for the five different *F-s* curve types in our dataset are given. The fracture surfaces correspond to the *F-s* curves given in Fig. 2. In addition, the assigned cluster is given for each specimen. Dark red corresponds to high activation. Low activation mapping is decoded in blue. For the type B specimen, no ductile fracture is visible on the fracture surface. The Grad-CAM shows a high activation for the center of the fracture surface (brittle fracture region) itself, as well as the corners of the notch. For the fracture surfaces corresponding to the type C *F-s* curve, solely the brittle fracture is activated. For the specimen associated to type D, ductile tearing is clearly distinguishable from the unstable fracture in the center of the fracture surface. The Grad-CAM analysis shows that the model still focuses on the unstable fracture region for its decision making. Additionally, to this, also the area left and right of the shear lips is slightly activated. As the total impact energy rises further (type E and F) the regions with the highest activation are changing. For type E the fracture initiation region (ductile) is more activated than the unstable fracture region (brittle). Parts of the brittle fracture are even marked as low activated regions, which is in contrast to the behaviour observed earlier. Remarkably, the activation of the area next to the shear lips is becoming stronger as the ductility grows further (type F).

This means the model tends to classify the lower shelf by its visible unstable or brittle fracture, where the activation of the brittle

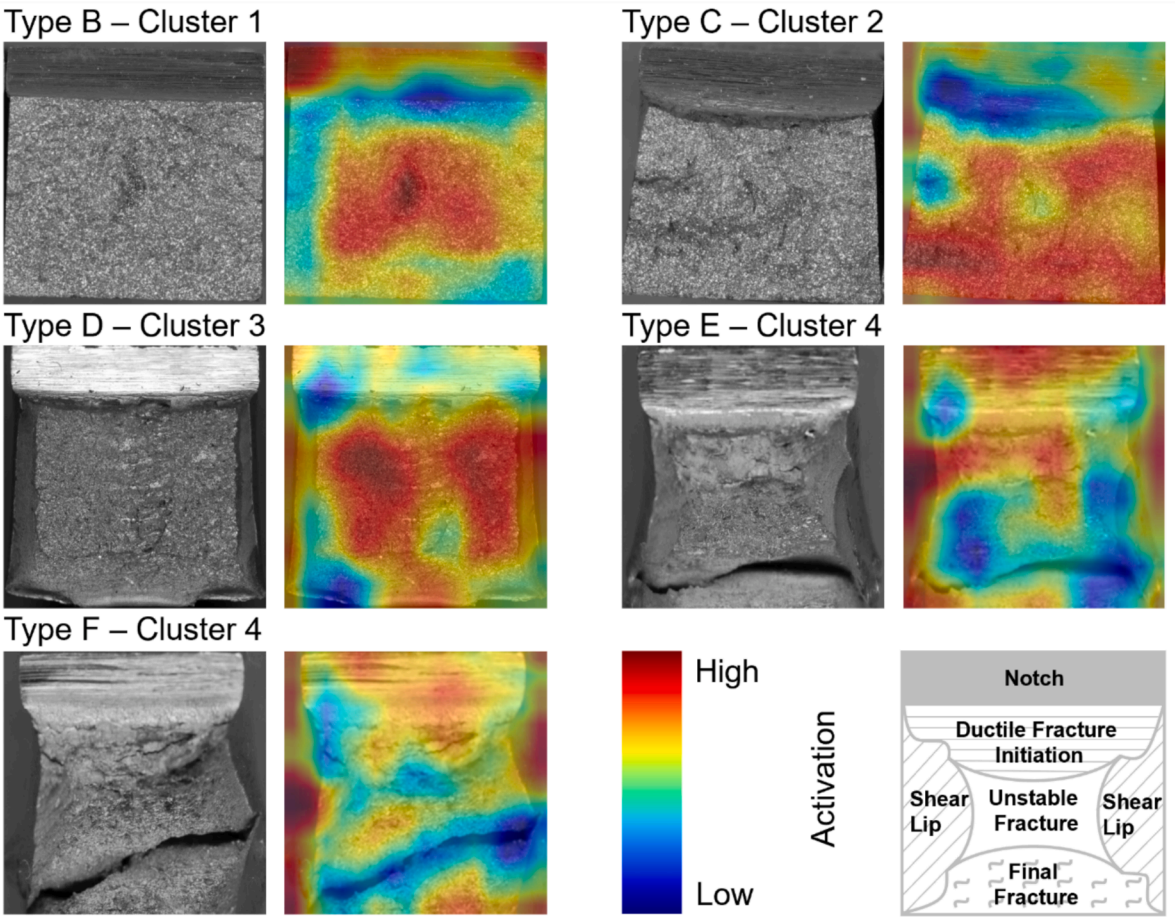


Fig. 13. Grad-CAM analysis for different F-s curve types (regions defined according to [1]).

fracture regions for the clustering is reasonable from a domain expert point of view. The decision-making changes for the higher energy specimen, where the focus on the area next to the shear lips is a clear indicator on the decision making of the model being rather simple. Rather than solely learning the complex patterns of ductile and brittle fracture, the model instead choses to distinguish the specimen shape after testing. This could be reminiscent of the behaviour of non-domain experts (e.g., children) when sorting fractured specimens according to their shape first rather than focusing on other, more complex features. However, this also corresponds to domain knowledge, since toughness is also commonly determined by lateral expansion measurements [1,51,52], where a higher expansion corresponds to higher toughness. Specimens that showed more ductility (types E and F) are mostly grouped into cluster 4 because of their distorted shape from the original dimensions. As this feature seems predominant, no further distinction between type E and F is

Table 4
Material specific statistical analysis of the clusters.

Cluster	Material	$n_{specimens}$	$\Delta\overline{F}$[kN]	$\overline{W}_t$[J]
0	Weld	6	14.7 (±4.3)	37.2 (±33.0)
1	22NiMoCr3-7	16	16.6 (±1.4)	35.5 (±31.9)
	Weld	3	13.7 (±1.6)	35.2 (±22.9)
	Weld	1	11.5	54.6
2	HAZ	10	18.3 (±2.4)	43.9 (±24.9)
	20MnMoNi5-5	10	17.8 (±1.5)	55.4 (±27.7)
3	22NiMoCr3-7	19	14.1 (±2.6)	75.4 (±33.3)
	Weld	3	3.8 (± 5.3)	127.6 (±56.8)
	HAZ	1	9.9	119.2
	20MnMoNi5-5	1	15.8	72.1
4	22NiMoCr3-7	13	5.8 (±3.1)	153.8 (±17.7)
	Weld	9	4.3 (± 1.9)	154.2 (±24.2)
	HAZ	15	3.4 (±4.4)	188.7 (±21.5)
	20MnMoNi5-5	14	5.7 (±5.1)	174.5 (±26.0)

achieved. This also explains the sheer size of the 4th cluster (51 specimens). The specimens with *F-s* curves of type B, C and D are more distinguishable for the model but still there is a mix of types in clusters 1, 2 and 3. The model learns to recognize the different representations of unstable (brittle) fracture rather than basing its decision on the specimen shape. The Grad-CAM analysis shows that the model is basing its decision on domain knowledge without being supervised.

In order to combine the image data with quantitative conclusions and to draw added value from the unsupervised clustering, we performed a cluster-sensitive and material-specific statistical analysis of the data. In Table 4, the mean ΔF ($\Delta\overline{F}$) and W_t ($\overline{W}_t$) values are given for each material in each cluster. The standard deviation σ is given in brackets. These two characteristic parameters describe the overall material behaviour during the pendulum impact test in a meaningful way. The model's material sensitivity regarding the weld material (cluster 0) and the RPV steel (clusters 1 and 3) becomes apparent, as it was observed in Fig. 11 (left). The mean ΔF values noticeably differ for the five clusters. While for cluster 4 $\Delta\overline{F}$ ranges from 3.4 to 5.8 kN for all considered materials, $\Delta\overline{F}$ lies above 14 kN

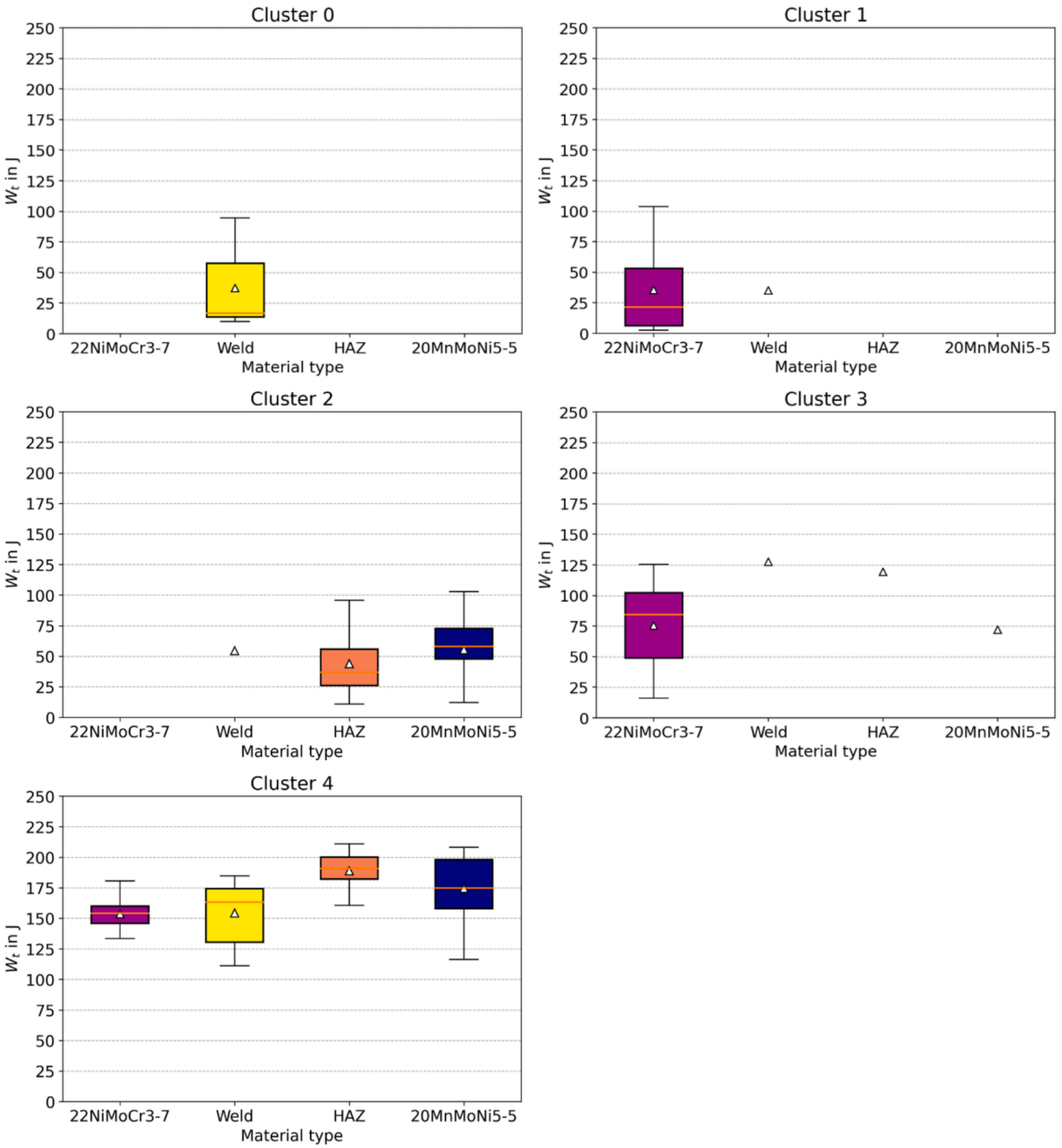


Fig. 14. Cluster and material sensitive boxplots regarding the total impact energy W_t (boxplot requires min. 5 specimens): triangles indicate the mean, orange lines the median value.

for the other clusters if at least 6 or more specimens are considered per material. This is in agreement to the observations made for Fig. 12, where cluster 4 shows the smallest ΔF values. For cluster 2 the $\Delta\overline{F}$ values are the highest, followed by cluster 1 and clusters 0 and 3 slightly below. Considering the total impact energy W_t similar observations can be made. For cluster 4 $\overline{W}_t$ ranges from 153.8 to 188.7 J, while for the other clusters considerably lower values are observed. Also, a clear distinction between the 22NiMoCr3-7 specimens in cluster 1 and in cluster 3 can be made, as cluster 3 lies above cluster 1. For $\Delta\overline{F}$ no such clear difference can be found. In general, HAZ and 20MnMoNi5-5 show higher total impact energies than 22NiMoCr3-7 and the weld material for cluster 4. Comparing the clusters 0, 1 and 2, the highest $\overline{W}_t$ values can be found for 20MnMoNi5-5, followed by the HAZ material. Weld and 22NiMoCr3-7 show similar but the overall lowest $\overline{W}_t$ values.

In Fig. 14 material sensitive boxplots are drawn for W_t for each cluster if a minimum number of specimens $n_{specimens}$ of 5 are available. Otherwise, solely the mean value is given indicated by a triangle. Overall, W_t rises with the cluster label. For cluster 0 and 1 only one material each can be analyzed (weld material and 22NiMoCr3-7). However, in combination with cluster 2, all four material types are covered, resembling the lower shelf. In contrast, cluster 4, shows much higher total impact energy values (upper shelf). The boxplots for clusters 0, 1 show similar values for weld and 22NiMoCr3-7. In comparison, the HAZ and 20MnMoNi5-5 (cluster 2) boxes are slightly elevated and narrower, indicating less scatter. The standard deviation values in Table 4 support this. For 22NiMoCr3-7 three ranges can be distinguished for W_t, corresponding to the lower shelf (cluster 1), the ductile-to-brittle transition (cluster 3) and the upper shelf (cluster 4).

In summary, the correlation analysis connecting clusters with material properties and the interpretability provided by the Grad-CAM analysis allow us to use the model for quantitative inference, where from a fracture surface image a prediction regarding the characteristic *F-s* curve parameters like F_{iu} or F_a and the total impact energy W_t can be made.

3.3. Inference of Force-Displacement behaviour

To fully utilize the potential of the methodology, we inferred material type sensitive statistical *F-s* curves for each cluster, using the characteristic force–displacement values and derived a statistical response corridor. It is calculated using the ARCGen python library by Hartlen et al. [43]. The method consists of three stages: arc-length reparameterization, signal registration and statistical analysis.

The arc-length reparameterization (first stage) scales a set of n input signals (*F-s* curves) and computes the arc-length l for each point in the signal. The components $s_{i,j}$ and $F_{i,j}$ are scaled by the mean extrema ($\overline{s}_{max}$, $\overline{s}_{min}$, $\overline{F}_{max}$, $\overline{F}_{min}$) of all signals

$$\widehat{s}_{i,j} = \frac{s_{i,j}}{\overline{s}_{max} - \overline{s}_{min}}, \tag{3}$$

$$\widehat{F}_{i,j} = \frac{F_{i,j}}{\overline{F}_{max} - \overline{F}_{min}}, \tag{4}$$

where j denotes an arbitrary point of the i^{th} signal ($i = 1,\cdots,n$). This maintains the relative shape and size of each *F-s* curve relative to one another. Starting at the first point of each individual signal and assuming linear behaviour between discrete points, the arc-length of each j^{th} data point $l_{i,j}$ can be computed as

$$l_{i,j} = \sum_{k=2}^{j} \sqrt{\left(\widehat{s}_{i,k} - \widehat{s}_{i,k-1}\right)^2 + \left(\widehat{F}_{i,k} - \widehat{F}_{i,k-1}\right)^2}. \tag{5}$$

Following, the arc-length is normalized (denoted as $\widehat{l}_{i,j}$) to the total arc-length of each individual signal. Now both components (axes) of each signal can be defined with respect to the normalized arc-length, yielding $s_i(\widehat{l}_i)$ and $F_i(\widehat{l}_i)$. By resampling at regular intervals with respect to $\widehat{l}$, all signals have the same number of points. These points are also located at the same normalized arc-lengths, which is necessary for the statistical analysis in stage three.

In the second stage (signal registration), critical features such as F_{gy}, F_m or F_a are aligned using warping functions to prevent distortion, when determining the characteristic averages. This assumes that for all signals the critical features occur at approximately the same normalized arc-length. The normalized arc-length was remapped using strictly monotonic and signal-specific warping functions $h_i(\widehat{l})$ in the form of monotonic cubic Hermite interpolation splines. Two exterior control points ([0,0] and [1,1]) were defined. In this study, the number of interior control points used in the warping functions was set to 2. The penalty factor, that influences the amount of warping, was set to 10^{-2}. Larger values will minimize the degree of warping. After re-parameterization and signal registration, the normalized arc-length is used to infer a point-wise statistical analysis (third stage) under the assumption, that all points at a given normalized arc-length are uncorrelated and normally distributed. This includes computing the mean of all signals at each normalized arc-length and calculating the standard deviation for s and F at each point. To compute the uncertainty, a two-dimensional uncorrelated normal distribution was used

$$f(s,F) = \frac{1}{2\pi\sigma_s\sigma_F}\exp\left(-\frac{1}{2}\left[\left(\frac{s-\overline{s}}{\sigma_s}\right)^2 + \left(\frac{F-\overline{F}}{\sigma_F}\right)^2\right]\right), \tag{6}$$

where $\overline{s}$ and $\overline{F}$ are the mean of all points at a given arc-length and their respective standard deviations are σ_s and σ_F. This results in a

confidence region (ellipse) centered at the mean values $\bar{s}$ and $\bar{F}$ with axes of $1\sigma_s$ and $1\sigma_F$. The corridors have a confidence of 68.3 % for one axes (solely considering *s* or *F*) and 39.4 % for the two-dimensional normal distribution [53]. Based on the statistical *F-s* curve, a characteristic average total impact energy could be inferred.

In Fig. 15 selected examples of inferred confidence interval of the *F-s* curves and corresponding total impact energy are given for all clusters and material types. For cluster 0, a statistical *F-s* curve corridor is shown for the weld material. Lower shelf experiments of type B and C with $\bar{s}_t = 1.03$ (±0.18) mm and $\overline{W}_t = 14.5$ (±3.1) J are considered. For cluster 1, the predicted 22NiMoCr3-7 *F-s* curve corridor is wider and considers more experiments. The total displacement s_t ranges roughly between 0.5 and 2 mm with a mean total displacement $\bar{s}_t$ of 1.22 (±0.88) mm. The expected total impact energy to be absorbed by a specimen classified in cluster 1 showing no crack arrest is low with 16.9 (±13.9) J. Cluster 2 is represented by HAZ data. The ductile-to-brittle transition (*F-s* curve types D and E) yields a mean total displacement of 5.74 (±2.43) mm and mean total impact energy of 63.5 (±20.2) J, which is slightly below the values for 22NiMoCr3-7 but with equal standard deviations. Here, $\bar{s}_t$ is inferred with 9.61 (±2.54) mm and $\overline{W}_t$ can be expected at 88.6 (±22.8) j. For cluster 4, the statistical *F-s* curve corridors for 20MnMoNi5-5 and the HAZ are given. Both show *F-s* curves of type F. Both materials show high total impact energies with $\overline{W}_t = 202.0$ (±3.9) J for 20MnMoNi5-5 and $\overline{W}_t = 196.5$ (±16.5) J for the HAZ, where the standard deviation is increased for the HAZ. The mean total displacement is slightly higher for 20MnMoNi5-5 compared to the HAZ (25.41 (±1.07) mm vs. 21.62 (±2.37) mm)

The results of the material sensitive inferred mean total impact energy values per cluster and *F-s* curve type are summarized in Table 5. 20MnMoNi5-5 has the toughest properties, ahead of those of the HAZ and 22NiMoCr3-7 and the weld. To display the variability in relation to the mean of the.

W_t population, we calculate the coefficient of variation *CV*. It is defined as

$$CV = \frac{\sigma}{\overline{W}_t}, \tag{7}$$

where σ is the standard deviation and $\overline{W}_t$ is the mean. The coefficient of variation shows that the statistical response corridor for the total impact energy is much narrower for the upper shelf (e.g., $CV = 1.9$ % for 20MnMoNi5-5 in cluster 4) and the ductile-to-brittle transition curves (e.g., $CV = 9.7$ % for 22NiMoCr3-7 in cluster 4) than for the lower shelf experiments (cluster 1 22NiMoCr3-7, cluster 2 HAZ and 20MnMoNi5-5), where *CV* reaches values up to 82.2 %. Only for the weld material, the *CV* is equal for both lower and upper shelf. However, the upper shelf weld material *F-s* curves exhibit much larger scatter ($CV = 16.6$ %) than the other materials due to the non-homogeneity.

The small statistical response corridors (low standard deviations) for the upper shelf experiments in cluster 4 for all materials except the weld material, can be explained by the homogeneity of the material [51,52]. The response of the material is predominantly plastic, resulting in a more uniform and stable fracture mechanism (microvoid nucleation and coalescence [54]) and therefore less variation in impact energy measurements compared to cleavage fracture which is governed by weakest-linkFor the lower shelf (e.g., clusters 1 and 2), the material undergoes brittle fracture, where little to no plastic deformation occurs, and cracks propagate unstable with minimal energy absorption. Brittle fracture is a statistical event [54] and highly sensitive to microstructural features [55,56], leading to more scatter in the measured impact energies. To improve the model's sensitivity and allow for a narrower statistical response corridor with high confidences (e.g., ± 3σ respectively 99.7 %) for all materials and clusters, further training on additional ductile-to-brittle transition and lower shelf region data would be required.

Nonetheless, our results indicate that this approach provides a feasible methodology for statistical inference of force–displacement curves and total impact energies from fractographic information alone. Our approach has enormous potential, as the fracture mechanical behaviour and the energy absorption characteristics of the material can be assessed without the need of instrumentation. The non-instrumented Charpy-V notch impact test in combination with a post-mortem fracture surface analysis simplifies the experimental setup and reduces costs, which enables broader application. For our data, this could already be applied for the quenched and tempered 20MnMoNi5-5 experiments in cluster 4 that fail solely ductile (upper shelf). For this specific material, the statistical corridor for W_t with a confidence of ± 3σ would be roughly ± 12 J, meaning that a test falling into this regime has a 99.7 % confidence of being representative for the materials upper shelf behaviour. Naturally, the scatter band for the transition region experiments is larger. Thinking of instrumented tests, our results also yield the possibility of extending the calibration procedure of the testing machine. Additionally to the specific statistical calibration described in [4] our approach also considers the fracture surface. A user could evaluate several Charpy-V notch specimens of a specific material type on an uncalibrated pendulum impact testing machine. The model could classify the specimens fracture surface following the experiment and suggest a statistical *F-s* curve corridor (inferred on a calibrated system) based on the cluster and material. If the measured force–displacement curve falls within this statistical range, it can be inferred that the testing machine operates with sufficient accuracy for use in this application.

4. Conclusions

Our study shows, how an unsupervised computer vision models can be used for quantitative fractography of Charpy-V notch pendulum impact tested specimens, where material type sensitive statistical *F-s* curves are inferred with a statistical response corridor of ± 1σ. The prior correlation analysis links the extracted clusters to material properties and model interpretability is provided by a Grad-CAM analysis.

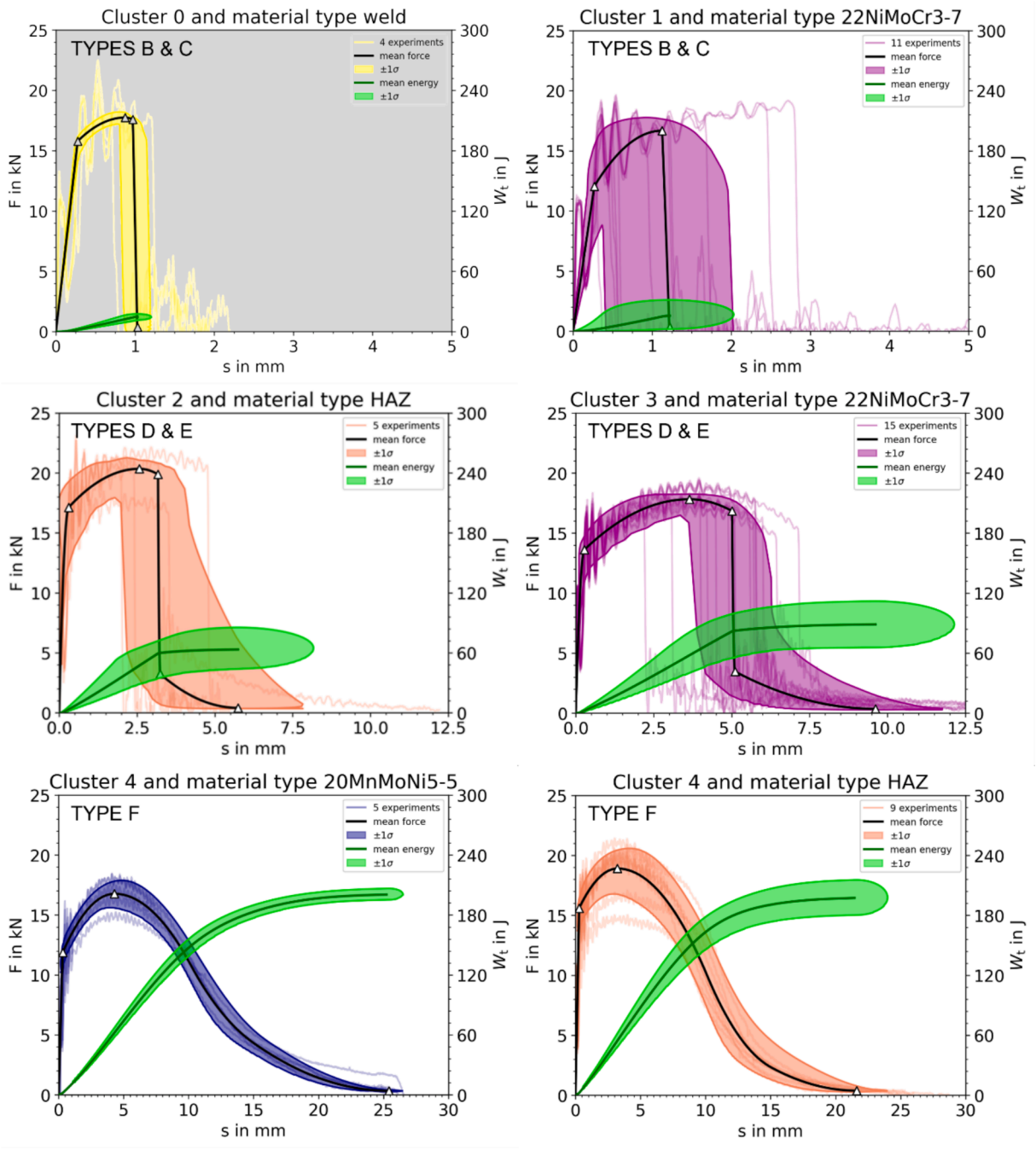


Fig. 15. Selected cluster sensitive inferred F-s curves and corresponding total impact energy W_t with a confidence interval of $\pm$ 1σ for all material types and F-s curve types.

- **Unsupervised learning**: A pre-trained ResNet50 architecture [48,49] is used to process the input images. By removing the classification layer, the feature representations can be extracted from the final layer of the network. A 2048-dimensional feature vector is extracted for every image. By performing a principal component analysis (PCA) is performed, the dimensionality is reduced to 56 features per image. The subsequent k-means clustering allows the distinction of 5 clusters in accordance with the number of *F-s* curve types in our data.
- **Correlation**: The clusters show high correlation to the *F-s* curve types and therefore the lower shelf, transition and upper shelf regime. By analyzing the correspondence to the unstable crack initiation force F_{iu} and the crack arrest force F_a expressed by their difference ΔF it could be seen that the clusters are highly responsive to characteristic *F-s* curve parameters. Using boxplots for

Table 5
Material sensitive summary of the inferred mean total impact energy per cluster and *F-s* curve type.

Material	Cluster	*F-s* curve type	$n_{specimens}$	$\overline{W}_t$[J]	σ[J]	*CV*[%]
22NiMoCr3-7	1	B & C	11	16.9	13.9	82.2
		D & E	5	76.5	19.8	25.9
	3	B & C	4	25.8	13.4	51.9
		D & E	15	88.6	22.8	25.7
	4	E	11	148.8	14.4	9.7
Weld	0	B & C	4	14.5	3.1	21.4
	4	E	8	152.8	25.3	16.6
HAZ	2	B & C	5	24.3	7.7	31.7
		D & E	5	63.5	20.2	31.8
	4	E	6	177.2	23.3	13.1
		F	9	196.5	16.5	8.4
	2	B & C	8	46.7	23.3	49.9
20MnMoNi5-5	4	E	8	164.6	13.5	8.2
		F	5	202.0	3.9	1.9

statistical visualization, our results show that the model is capable of classifying lower shelf, ductile-to-brittle transition and the upper shelf regions according to the total impact energy W_t values for all materials in the dataset.

- **Interpretability**: The Grad-CAM analysis reveals that the model is using two different approaches for classification. Lower shelf experiments are distinguished by the activation of the unstable (brittle) fracture region in the center of the fracture surface. As the total impact energy increases (e.g. for upper shelf experiments), the fracture surfaces are distinguished by the distortion of their shape. Here, the area next to the shear lips is highly activated, thus recalling lateral expansion measurements. Therefore, both methods can be linked to domain knowledge, indicating the model's ability to learn the necessary representations without the need of supervision.
- **Inference**: For each cluster, we inferred material type sensitive statistical *F-s* curves using the characteristic force–displacement values and derived a statistical response corridor with $\pm 1\sigma$. Based on the statistical *F-s* curve, a characteristic average total impact energy could be inferred. For the upper shelf, statistical response corridors with low standard deviations could be found for all materials except the weld material. This can be attributed to the inhomogeneity of the material [51,52]. For the ductile-to-brittle transition and the lower shelf experiments the standard deviations are increased, due to the microstructure sensitivity [54–56]. This leads to an increase of the predicted corridor size.
- **Potential use**: Our model is feasible for inference of statistical force–displacement curves and resulting total impact energies based on fracture surface images. Utilizing our approach non-instrumented Charpy-V notch impact tests could be supplemented with information on the fracture mechanical behaviour and energy absorption characteristics of the material. The fracture surface analysis simplifies the experimental setup and reduces costs. The need for instrumentation could even be eliminated for a sufficient statistical database. Based on the results, the approach has potential utility in validating pendulum impact testing machines based on the fracture surface in combination with the force displacement curve. To improve the model's sensitivity and allow for narrow statistical response corridors with high confidences (e.g., $\pm 3\sigma$), also for lower shelf regimes, further training on additional data, covering the ductile-to-brittle transition and lower shelf region would be required.

CRediT authorship contribution statement

Johannes Rosenberger: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Johannes Tlatlik:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nils Rump:** Visualization, Formal analysis, Data curation. **Sebastian Münstermann:** Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

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

Data availability

Data will be made available on request.

References

[1] ASTM International, American Society for Testing and Materials. ASTM E23-16b: Test Methods for Notched Bar Impact Testing of Metallic Materials. West Conshohocken, PA: ASTM International.

[2] European Committee For Standardization. EN ISO 148-1:2017-05: Metallic Materials - Charpy pendulum impact test - Part 1: Test method; 2017 2017.
[3] K. Wallin, P. Nevasmaa, T. Planman, M. Valo, Evolution of the Charpy-V test from a quality control test to a materials evaluation tool for structural integrity assessment, From Charpy to Present Impact Testing (european Structural Integrity Society) 30 (2002) 57–68, https://doi.org/10.1016/S1566-1369(02)80006-1.
[4] European committee for standardization, EN ISO 14556:2023 E: Metallic Materials - Charpy pendulum impact test - Instrumented test method, CEN-CENELEC Management Centre, Brussels, 2023, p. 2023.
[5] ASTM International, American Society for Testing and Materials. ASTM E2298-15: Test Method for Instrumented Impact Testing of Metallic Materials. West Conshohocken, PA: ASTM International.
[6] S. Tsopanidis, S. Osovski, Unsupervised machine learning in fractography: Evaluation and interpretation, Mater Charact 182 (2021) 111551, https://doi.org/10.1016/j.matchar.2021.111551.
[7] S.M. Azimi, D. Britz, M. Engstler, M. Fritz, F. Mücklich, Advanced Steel Microstructural Classification by Deep Learning Methods, Sci Rep 8 (1) (2018) 2128, https://doi.org/10.1038/s41598-018-20037-5.
[8] B.L. DeCost, E.A. Holm, A computer vision approach for automated analysis and classification of microstructural image data, Comput. Mater. Sci 110 (2015) 126–133, https://doi.org/10.1016/j.commatsci.2015.08.011.
[9] A.R. Durmaz, S.T. Potu, D. Romich, J.J. Möller, R. Nützel, Microstructure quality control of steels using deep learning, Front. Mater. 10 (2023), https://doi.org/10.3389/fmats.2023.1222456.
[10] M.X. Bastidas-Rodriguez, F.A. Prieto-Ortiz, E. Espejo, Fractographic classification in metallic materials by using computer vision, Eng. Fail. Anal. 59 (2016) 237–252, https://doi.org/10.1016/j.engfailanal.2015.10.008.
[11] M.X. Bastidas-Rodriguez, L. Polania, A. Gruson, F. Prieto-Ortiz, Deep Learning for fractographic classification in metallic materials, Eng. Fail. Anal. 113 (2020) 104532, https://doi.org/10.1016/j.engfailanal.2020.104532.
[12] J. Rosenberger, J. Tlatlik, S. Münstermann, Deep learning based initial crack size measurements utilizing macroscale fracture surface segmentation, Eng. Fract. Mech. 293 (2023) 109686, https://doi.org/10.1016/j.engfracmech.2023.109686.
[13] L. Schmies, et al., Classification of fracture characteristics and fracture mechanisms using deep learning and topography data, Practical Metallography 60 (2) (2023) 76–92, https://doi.org/10.1515/pm-2022-1008.
[14] L. Schmies, M. Hemmleb, D. Bettge, Relevant input data for crack feature segmentation with deep learning on SEM imagery and topography data, Eng. Fail. Anal. 156 (2024) 107814, https://doi.org/10.1016/j.engfailanal.2023.107814.
[15] J. Rosenberger, J. Tlatlik, C. Beckmann, B. Rohrmüller, S. Münstermann, Extracting ductile cast iron microstructure parameters from fracture surfaces: A deep learning based instance segmentation approach, Eng. Fract. Mech. 311 (2024) 110586, https://doi.org/10.1016/j.engfracmech.2024.110586.
[16] M. Ackermann, D. Iren, S. Wesselmecking, D. Shetty, U. Krupp, Automated segmentation of martensite-austenite islands in bainitic steel, Mater Charact 191 (2022) 112091, https://doi.org/10.1016/j.matchar.2022.112091.
[17] N. Fehlemann, A.L. Suarez Aguilera, S. Sandfeld, F. Bexter, M. Neite, D. Lenz, M. Könemann, S. Münstermann, Identification of martensite bands in dual-phase steels: a deep learning object detection approach using faster region-based-convolutional neural network, Steel Res. Int. 94 (7) (2023) 2200836, https://doi.org/10.1002/srin.202200836.
[18] C. Zednik, H. Boelsen, Scientific Exploration and Explainable Artificial Intelligence, Mind. Mach. 32 (1) (2022) 219–239, https://doi.org/10.1007/s11023-021-09583-6.
[19] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization, Int J Comput vis 128 (2) (2020) 336–359, https://doi.org/10.1007/s11263-019-01228-7.
[20] T. Liu, A.S. Barnard, The emergent role of explainable artificial intelligence in the materials sciences, Cell Rep. Phys. Sci. 4 (10) (2023) 101630, https://doi.org/10.1016/j.xcrp.2023.101630.
[21] X. Zhong, B. Gallagher, S. Liu, B. Kailkhura, A. Hiszpanski, T.Y.-J. Han, Explainable machine learning in materials science, Npj Comput Mater 8 (1) (2022) 204, https://doi.org/10.1038/s41524-022-00884-7.
[22] D. Melching, T. Strohmann, G. Requena, E. Breitbarth, Explainable machine learning for precise fatigue crack tip detection, Sci Rep 12 (1) (2022) 9513, https://doi.org/10.1038/s41598-022-13275-1.
[23] A.-Y.-T. Wang, M.S. Mahmoud, M. Czasny, A. Gurlo, CrabNet for Explainable Deep Learning in Materials Science: Bridging the Gap Between Academia and Industry, Integr Mater Manuf Innov 11 (1) (2022) 41–56, https://doi.org/10.1007/s40192-021-00247-y.
[24] K. Jones, et al., Predicting fatigue crack growth metrics from fractographs: Towards fractography by computer vision, Int. J. Fatigue 177 (2023) 107915, https://doi.org/10.1016/j.ijfatigue.2023.107915.
[25] A. Singh, A. Rabbani, K. Regenauer-Lieb, R.T. Armstrong, P. Mostaghimi, Computer vision and unsupervised machine learning for pore-scale structural analysis of fractured porous media, Adv. Water Resour. 147 (2021) 103801, https://doi.org/10.1016/j.advwatres.2020.103801.
[26] R. Cohn, E. Holm, Unsupervised Machine Learning Via Transfer Learning and k-Means Clustering to Classify Materials Image Data, Integr Mater Manuf Innov 10 (2) (2021) 231–244, https://doi.org/10.1007/s40192-021-00205-8.
[27] H. Kim, J. Inoue, T. Kasuya, Unsupervised microstructure segmentation by mimicking metallurgists' approach to pattern recognition, Sci Rep 10 (1) (2020) 17835, https://doi.org/10.1038/s41598-020-74935-8.
[28] J. Maldonado-Correa, J. Torres-Cabrera, S. Martín-Martínez, E. Artigao, E. Gómez-Lázaro, Wind turbine fault detection based on the transformer model using SCADA data, Eng. Fail. Anal. 162 (2024) 108354, https://doi.org/10.1016/j.engfailanal.2024.108354.
[29] Y. Guo, S.V. Kalinin, H. Cai, K. Xiao, S. Krylyuk, A.V. Davydov, Q. Guo, A.R. Lupini, Defect detection in atomic-resolution images via unsupervised learning with translational invariance, Npj Comput Mater 7 (1) (2021) 180, https://doi.org/10.1038/s41524-021-00642-1.
[30] M. Yu, X. Xie, Z. Fang, J.B. Lim, A novel machine-learning based framework for calibrating micromechanical fracture model of ultra-low cycle fatigue in steel structures, Eng. Fract. Mech. 306 (2024) 110200, https://doi.org/10.1016/j.engfracmech.2024.110200.
[31] E.A. Holm, et al., Overview: Computer Vision and Machine Learning for Microstructural Characterization and Analysis, Metall Mater Trans A 51 (12) (2020) 5985–5999, https://doi.org/10.1007/s11661-020-06008-4.
[32] Y. Chen, et al., Identifying facile material descriptors for Charpy impact toughness in low-alloy steel via machine learning, J. Mater. Sci. Technol. 132 (2023) 213–222, https://doi.org/10.1016/j.jmst.2022.05.051.
[33] S. Wu, J. Yang, G. Cao, Prediction of the Charpy V-notch impact energy of low carbon steel using a shallow neural network and deep learning, Int J Miner Metall Mater 28 (8) (2021) 1309–1320, https://doi.org/10.1007/s12613-020-2168-z.
[34] N. Bianco, et al., Toughness from Imagery: Extracting More from Failure Analysis using Deep Convolutional Neural Networks, J Fail. Anal. and Preven. (2024), https://doi.org/10.1007/s11668-024-02045-y.
[35] S.-J. Lee, G.-M. Kim, S.-D. Choi, A Machine Learning Program for Impact Fracture Analysis, KSMPE 20 (1) (2021) 95–102, https://doi.org/10.14775/ksmpe.2021.20.01.095.
[36] Böhme W, *et al.* Überprüfung und Weiterentwicklung von Bewertungsmethoden für dynamische Rissinitiierung und Rissarrest : Abschlussbericht/final report. IWM-Bericht 2012; 665. doi: 10.2314/GBV:745745466.
[37] T. Reichert, J. Tlatlik, W. Böhme, S.D. Verbundprojekt, Analyse und Validierung von bruchmechanischen Bewertungsmethoden für dynamische Beanspruchungen: Teilprojekt: Untersuchungen mit stoßbelasteten Biegeproben und Identifikation von statistisch abgesicherten dynamischen Bruchzähigkeits-Grenzkurven, Abschlussbericht: Reaktorsicherheitsforschung - Vorhaben Nr.: (2017), https://doi.org/10.2314/GBV:1027454739.
[38] Mayer U, Allmendinger K. Verbundprojekt: Analyse und Validierung von bruchmechanischen Bewertungsmethoden für dynamische Beanspruchungen - Teilprojekt: Versuche mit Kompaktzugproben und Identifikation eines Parameters für die Rissbeanspruchungsrate : Abschlussbericht : Reaktorsicherheitsforschung - Vorhaben-Nr. 1501472B 2017. doi: 10.2314/GBV:1026853893.
[39] J. Tlatlik, D. Discher, Verbundprojekt Bruchdynamik Phase 3: Untersuchung des Masterkurven-Konzepts bei dynamische Rissbeanspruchung; Teilvorhaben: Untersuchung anhand von dynamischen SE(B)-Versuchen: Abschlussbericht, Reaktorsicherheitsforschung - Vorhaben Nr.: (2023), https://doi.org/10.2314/KXP:1877629707.

[40] D. Discher, J. Tlatlik, F. Huberth, Analyse und Bewertung lokaler Arrestereignisse bei Komponenten unter dynamischer Belastung Abschlussbericht, Reaktorsicherheitsforschung - Vorhaben Nr.: (2021), https://doi.org/10.2314/KXP:181468929X.
[41] J. Tlatlik, U. Mayer, Impact of elevated loading rates on the shape of the Master Curve (ASTM E1921) for a German RPV steel, Eng. Fract. Mech. 311 (2024) 110588, https://doi.org/10.1016/j.engfracmech.2024.110588.
[42] Tlatlik J. Verification of the Master Curve Conecpt (ASTM 1921) and Inhomogeneity Analysis of a German RPV Weld for Various Loading Rates. Engineering Fracture Mechanics 2023 p. 108998. doi: 10.1016/j.engfracmech.2022.108998.
[43] D.C. Hartlen, D.S. Cronin, Arc-Length Re-Parametrization and Signal Registration to Determine a Characteristic Average and Statistical Response Corridors of Biomechanical Data. Front Bioeng, Biotechnol 10 (2022) 843148, https://doi.org/10.3389/fbioe.2022.843148.
[44] N. Christ, et al., Characterization and Simulation of the Interface between a Continuous and Discontinuous Carbon Fiber Reinforced Thermoplastic by Using the Climbing Drum Peel Test Considering Humidity, Polymers (basel) 16 (7) (2024), https://doi.org/10.3390/polym16070976.
[45] Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection; 2015 Jun 8. doi: 10.48550/arXiv.1506.02640.
[46] G. Bradski, The OpenCV Library, Dr. Dobb's Journal of Software Tools (2000).
[47] Paszke A, *et al.* PyTorch: An Imperative Style, High-Performance Deep Learning Library. doi: 10.48550/arXiv.1912.01703.
[48] Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A Large-Scale Hierarchical Image Database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA, 2009 p. 248–255. doi: 10.1109/CVPR.2009.5206848.
[49] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition (2015), https://doi.org/10.48550/arXiv.1512.03385.
[50] F. Pedregosa, et al., Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[51] W. Böhme, Results of a DVM round robin on instrumented Charpy testing, From Charpy to Present Impact Testing (european Structural Integrity Society) 30 (2002) 189–196, https://doi.org/10.1016/S1566-1369(02)80020-6.
[52] Böhme W. Experience with Charpy tests obtained by a DVM round-robin and further developmens. Evaluating Material Properties by Dynamic Testing : papers presented at the ESIS TC5 Workshop on the Physical Understanding and Use of Material Properties Determined by Dynamic Testing and Standards (ESIS 20), Belgian nuclear research centre SCK/CEN, April 28, 1994 / ed. by E. van Walle 1996 p. 1–23.
[53] B. Wang, W. Shi, Z. Miao, Confidence analysis of standard deviational ellipse and its extension into higher dimensional euclidean space, PLoS One 10 (3) (2015) e0118537, https://doi.org/10.1371/journal.pone.0118537.
[54] K. Wallin, Macroscopic nature of brittle fracture, Le Journal De Physique IV 3 (7) (1993) 575–584, https://doi.org/10.1051/jp4:1993795.
[55] K. Wallin, Master curve analysis of ductile to brittle transition region fracture toughness round robin data: The "EURO" fracture toughness curve, Technical Research Centre of Finland, Espoo, 1998.
[56] B. Strnadel, P. Haušild, Statistical scatter in the fracture toughness and Charpy impact energy of pearlitic steel, Mater. Sci. Eng. A 486 (1–2) (2008) 208–214, https://doi.org/10.1016/j.msea.2007.08.079.