

Role of length-scale in machine learning based image analysis of ductile fracture surfaces

Xinzhu Zheng^a, Bekassyl Battalgazy^a, Abhilash Molkeri^a, Stylianos Tsopanidis^b, Shmuel Osovski^{b,*}, Ankit Srivastava^{a,*}

^a*Department of Materials Science and Engineering, Texas A&M University, College Station, TX 77845, USA*

^b*Faculty of Mechanical Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel*

Abstract

Recent advancements in machine learning (ML) techniques have opened up new opportunities for using image analysis to solve materials science problems. In this work, we have used an ML-based workflow to classify the fracture surfaces of dual-phase steels subjected to different stress states. This task is not straightforward, as the ductile fracture surfaces of many metallic materials exhibit similar features, such as dimples. The ML-based workflow uses a pre-trained convolution neural network in unsupervised mode to extract image features, which are then reduced in dimensionality using principal component analysis. Next, images are clustered and classified using K-Means and K-Nearest Neighbors algorithms, respectively. Our results show that the accuracy of the ML-based technique is sensitive to the length-scales of the fracture surface images, and the critical length-scale corresponding to the maximum accuracy depends on the typological categories being classified. A physical interpretation of the critical length-scales associated with the fracture surface images is provided through quantitative fracture surface roughness analysis. Our work demonstrates the potential of using unsupervised ML-based techniques for fractography of ductile materials, especially for typological classification. More importantly, it emphasizes the importance of length-scales in image analysis in materials science.

Keywords: Ductile Fracture; Fractography; Machine Learning; Unsupervised Learning; Image Analysis; Fracture surface roughness

1. Introduction

Qualitative analyses of fracture surfaces dates back to the sixteenth century when it was first used for quality assurance of metallurgical processes (ASM Handbook, 1987). The technological advancements in the twentieth century, such as high magnification imaging, enabled detailed typological analyses of the fracture surfaces and significantly advanced our understanding of material failure. For example, it enabled us to differentiate between brittle

*Corresponding authors: shmuliko@technion.ac.il (S. Osovski), ankit.sri@tamu.edu (A. Srivastava)

and ductile fracture, and overload and fatigue fracture by simply analyzing the fracture surfaces (Pineau et al., 2016a,b). At present, routine examination of fracture surfaces is utilized in both industrial and scientific research to correlate material, loading condition and fracture process. The fracture process, specifically, in metallic materials involves a series of micro-separation processes (e.g., nucleation, growth and coalescence of microscale voids). This in turn, results in highly irregular and topographically complex fracture surfaces. Thus, extracting meaningful information not only requires exhaustive analyses of the fracture surface images but also the outcome (in general) depends on the past experience of the practitioner.

More recently, the use of computer vision and machine learning (ML) techniques in materials science and engineering have enabled automated characterization and classification of materials' microstructures (DeCost et al., 2019; Holm et al., 2020; Choudhary et al., 2022). The use of ML techniques have also been extended to characterize and classify fracture surfaces (Naik and Kiran, 2019; Bastidas-Rodriguez et al., 2020; Kitahara and Holm, 2018; Tsopanidis et al., 2020; Tsopanidis and Osovski, 2021). In Naik and Kiran (2019) ductile and brittle fracture surfaces were classified using a supervised ML technique and hand crafted features. A similar typological classification of fracture surfaces using a supervised ML technique but without hand crafted features was carried out in Bastidas-Rodriguez et al. (2020), wherein they classified fatigue, brittle and ductile fracture surfaces. The classification task in both Naik and Kiran (2019) and Bastidas-Rodriguez et al. (2020) involved classification of the entire fracture surface image as belonging to one class or another. Tsopanidis et al. (2020), on the other hand, focused on differentiating between intergranular and transgranular features in the fracture surfaces of a brittle material. They followed a transfer learning approach and showed that once an ML technique is trained in a supervised manner to classify between intergranular and transgranular features in the fracture surfaces of a given material, it can be used for the same task for other similar materials.

Although these works have shown that ML techniques can be successfully used for typological classification of the fracture surfaces, these methods require a large labeled dataset of fracture surface images for training. To overcome this challenge, Kitahara and Holm (2018) explored the possibility of using an unsupervised ML technique to classify fracture surfaces of an additively manufactured In-718 alloy. Specifically, they used a Convolution Neural Network (CNN), VGG16 (Simonyan and Zisserman, 2014) trained on the ImageNet (Russakovsky et al., 2015) dataset as a features extractor and successfully classified fracture surfaces of Charpy impact specimens oriented parallel and normal to the build direction. Using the same method as in Kitahara and Holm (2018), Tsopanidis and Osovski (2021) classified fracture surfaces of five Tungsten heavy alloys (with Tungsten content varying from 90%-99%) showing that this method can be used to correlate material and fracture process. Moreover, they were also able to identify the specific topographical characteristics of the fracture surfaces that enabled classification which in this case was the signature of localized plastic deformation.

Building on the works of Kitahara and Holm (2018) and Tsopanidis and Osovski (2021), herein we attempt to not only classify the fracture surfaces of different materials but also differentiate between the loading conditions. To this end, fracture surfaces of flat dog-

bone (simple tension, ST) and single-edge notch tension (SENT) specimens of two (ferritic-martensitic) dual-phase steels, viz, DF140T and DP980 (Srivastava et al., 2016; Gerbig et al., 2018; Zheng et al., 2020), are analyzed and classified. The ST specimens subject the material to a uniaxial loading condition up until the onset of necking, whereas in the SENT specimens the loading is always heterogeneous. The microstructure of both the steels comprises martensite phase particles dispersed in ferrite phase matrix. However, their nominal chemical composition and amount of constituent phases in the microstructure are different. Nevertheless, both the ST and SENT specimens of the two steels predominantly undergo ductile fracture due to nucleation, growth and coalescence of microscale voids. Thus, it is challenging to classify the fracture surfaces of these fractured specimens into the four typological categories just by visual inspection, since they all have similar dimpled fracture surfaces.

Furthermore, in materials science, the application of ML techniques for image analysis, more often than not, only focuses on datasets characterized by the number of pixels in the images. This approach ignores the importance of length-scales that may not only improve the efficacy, but more importantly, provide physical insights into the classification process. Thus, understanding the role of length-scales in ML-based image analysis is a significant step towards going beyond the “black box” paradigm often associated with the application of ML in materials science and engineering (Holm, 2019). Following this, we thoroughly investigate the effect of length-scales associated with images on the accuracy of typological classification of fracture surfaces. Our results show that there are critical length-scales or simply the size of the images that maximize the classification accuracy. A physical interpretation of the critical length-scales associated with the fracture surface images is explored via quantitative fracture surface roughness analysis.

2. Materials and fracture surfaces

The materials considered in this work are two dual-phase (ferritic-martensitic) advanced high strength sheet steels, viz, DF140T and DP980 (Srivastava et al., 2016; Gerbig et al., 2018; Zheng et al., 2020). The nominal chemical composition of DF140T is 0.15wt%C – 1.45wt%Mn – 0.3wt%Si – rest Fe and that of DP980 is 0.09wt%C – 2.15wt%Mn – 0.6wt%Si – rest Fe. These steels were produced on a water – quenched continuous anneal line by ArcelorMittal. The ST and SENT specimens with axes parallel to the rolling direction of the steel sheets were machined using wire Electrical Discharge Machining (EDM). The length and width of the gauge section of the ST specimens were 8mm and 3mm, respectively. The SENT specimens had the same dimensions as the ST specimens, but they contained an edge notch of depth 1.5mm, machined using wire EDM in the center of the gauge section. The final tip radii of the EDM-machined notches were approximately $150\mu\text{m}$. Prior to mechanical testing, the surfaces of all the specimens were mechanically ground using 320-1200 grit SiC grinding papers and fine polished with $0.05\mu\text{m}$ alumina suspension. The specimens were then slightly chemically etched with 3% Nital for 15s to reveal their microstructure. The final thickness of the polished and etched specimens was approximately 1.35mm. All the mechanical tests were conducted using the same cross-head speed giving a nominal strain

rate of $0.5 \times 10^{-3} \text{s}^{-1}$ for the ST specimens.

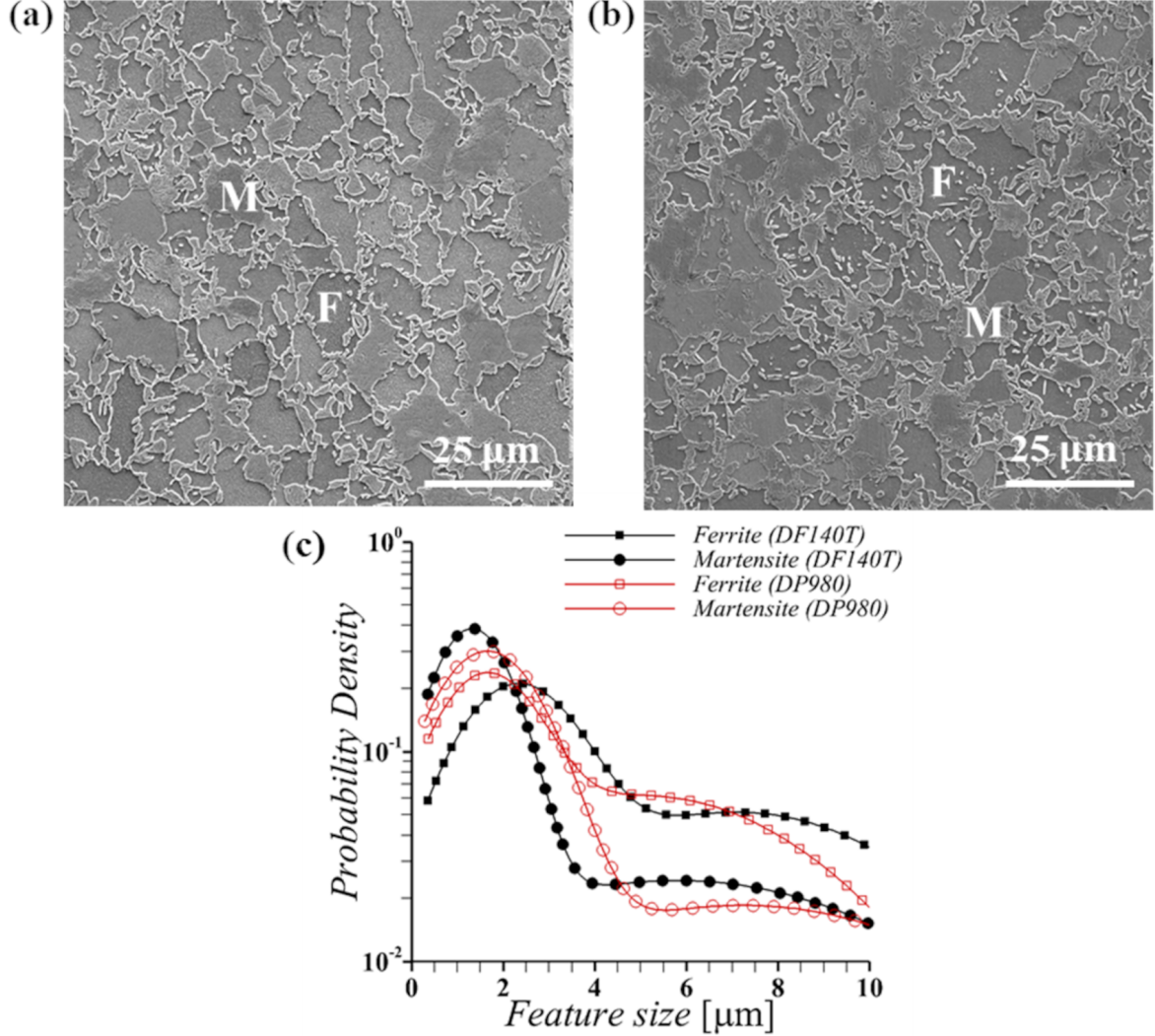


Figure 1: SE-SEM images showing the initial ferritic (F) - martensitic (M) microstructures of (a) DF140T and (b) DP980 dual-phase steels. (c) The probability density of size distribution of the features in the initial microstructures of the two steels.

The initial microstructures of the two steels were characterized using a Tescan FERA-3 HR Scanning Electron Microscope (SEM). All SEM images were captured using the Secondary Electron (SE) detector. Representative SE-SEM images of the initial microstructures of the DF140T and DP980 steels are shown in Figs. 1(a) and (b), respectively. The microstructure of both the steels comprises islands of martensite phase particles dispersed in the ferrite phase matrix. The microstructure of the DF140T contains 40% martensite and rest ferrite, while that of the DP980 contains 60% martensite and rest ferrite. The size of the individual features in the microstructures of the two steels were also measured following

Table 1: Parameters of the bimodal fit, Eq. (1), to the probability density plots of the size distribution of ferrite and martensite phases in DF140T and DP980 dual-phase steels shown in Fig. 1(c).

	p	μ_1	μ_2	σ_1	σ_2
Ferrite (DF140T)	0.6	2.3	7.1	1.2	3.3
Martensite (DF140T)	0.7	1.3	5.8	0.8	4.3
Ferrite (DP980)	0.5	1.6	5.1	1.0	3.1
Martensite (DP980)	0.8	1.7	7.2	1.1	4.12

the Heyn linear intercept procedure (ASTM E112-13, 2013). The probability density plot of the size distribution of the martensite and ferrite phases in the two steels are shown in Fig. 1(c). As shown in the figure, both the features in both the steels follow a bimodal size distribution. The bimodal size distribution of these features can be represented as a sum of two normal distributions, as follows:

$$f_{mix}(x) = p \times \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} + (1-p) \times \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2} \quad (1)$$

In Eq. (1), μ_i and σ_i are the mean and standard deviation, respectively, of the i^{th} normal distribution while p and $1-p$ are the proportion of the two distributions, respectively. The parameters of the bimodal fit to the probability density plots of the size distribution of both the features in the two steels are given in Table 1.

After mechanical testing, the fracture surfaces of the fractured specimens were imaged using a Tescan FERA-3 HR-SEM in the SE mode. The SE-SEM images were taken at 15kV with a beam intensity of 10. Also due to the highly irregular topography of the fracture surfaces (that leads to large height differences along the surface), the SEM was operated under the “depth mode”, allowing for an increased focal depth. To avoid image-specific characters produced by an SEM operator, special care was taken to ensure that the brightness/contrast setting used for all SEM images were the same. Additionally, due to the tendency of sharp edges to be over-exposed, all images were acquired such that the image histogram is fully contained within the range 5-250 (grayscale levels) thus avoiding regions with zero information.

Representative SE-SEM images of the entire fracture surfaces of fractured ST and SENT specimens of both the steels are shown in Figs. 2(a & d) and 3(a & d), respectively. High-resolution images (4096 pixels $\times$ 4096 pixels) of a 150 $\times$ 150 μm^2 area taken from the locations marked with dotted boxes on the full fracture surfaces are shown for all four cases in Figs. 2(b & e) and 3(b & e). Additionally, zoomed-in view of 37.5 $\times$ 37.5 μm^2 region from within the 150 $\times$ 150 μm^2 region are also shown in Figs. 2(c & f) and 3(c & f). As shown in Figs. 2 and 3, the fracture surfaces of both the specimens of both the steels exhibit features of ductile fracture, with large dimples surrounded by smaller dimples.

From each half of a fractured specimen, sixteen 150 $\times$ 150 μm^2 images were acquired from random locations, while avoiding the shear lips near the edges. All images were acquired as 4096 pixels $\times$ 4096 pixels, leading to a pixel size of 36.6nm. Thus, the initial database used in this work contains a total of (16 $\times$ 2=) 32 images for each typological category, viz, DF140T

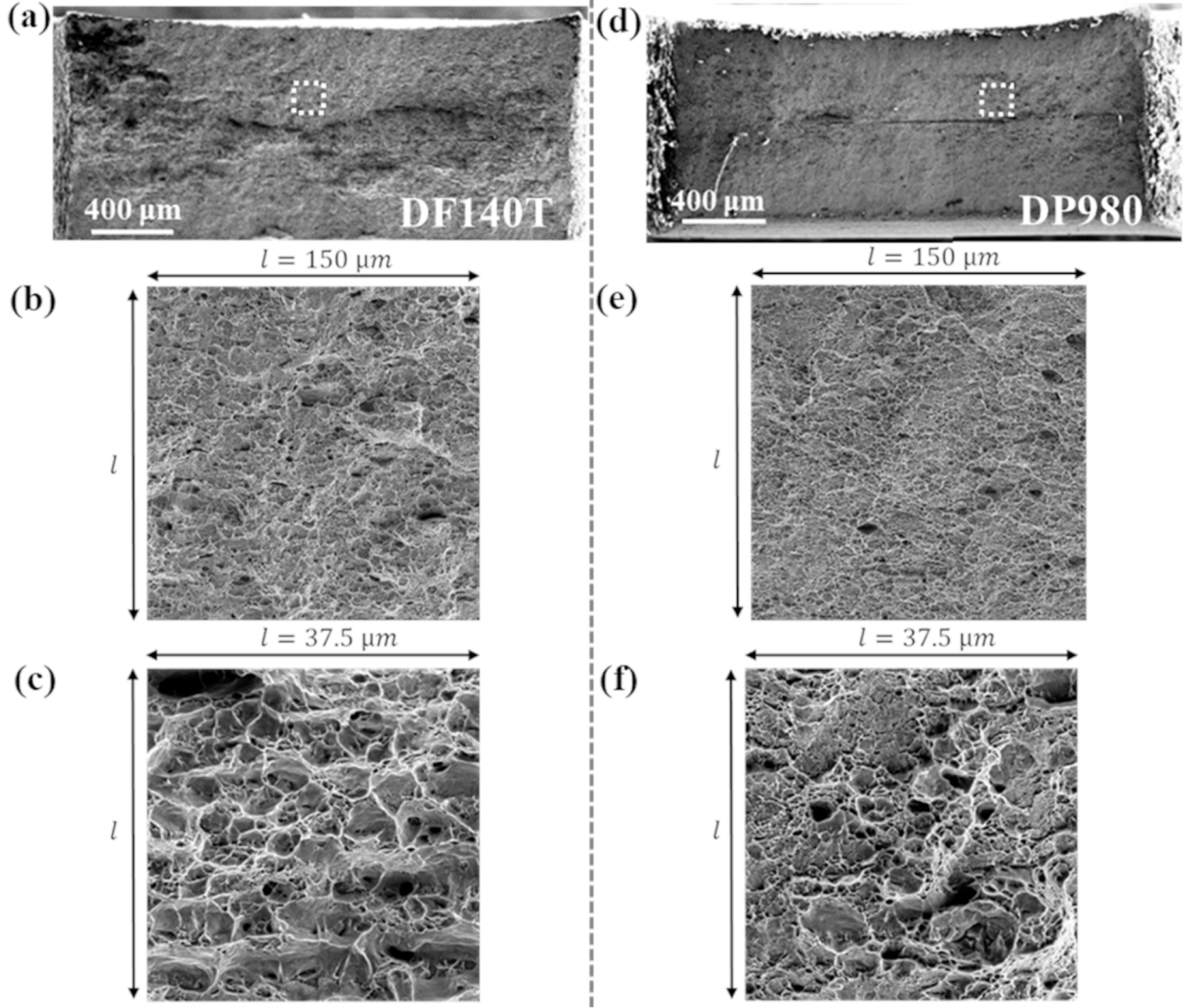


Figure 2: (a) SE-SEM image of the entire fracture surface and (b)-(c) high-resolution SE-SEM images of the region marked with a ‘white’ dotted box in (a) of a fractured ST specimen of DF140T. (d) SE-SEM image of the entire fracture surface and (e)-(f) high-resolution SE-SEM images of the region marked with a ‘white’ dotted box in (d) of a fractured ST specimen of DP980.

ST, DP980 ST, DF140T SENT and DP980 SENT. To unravel the impact of image size (i.e., length-scale) on the typological classification process, 11 datasets were then generated from the initial database. This was done by dividing each $150 \times 150 \mu\text{m}^2$ original image into sub-images of different sizes ranging from $10 \times 10 \mu\text{m}^2$ to $100 \times 100 \mu\text{m}^2$. Note that this procedure leads to different number of sub-images in different datasets. For example, the dataset of $100 \times 100 \mu\text{m}^2$ sub-images contains $(1 \times 1 \times 32 =)$ 32 images while the dataset of $10 \times 10 \mu\text{m}^2$ sub-images contains $(15 \times 15 \times 32 =)$ 7200 images.

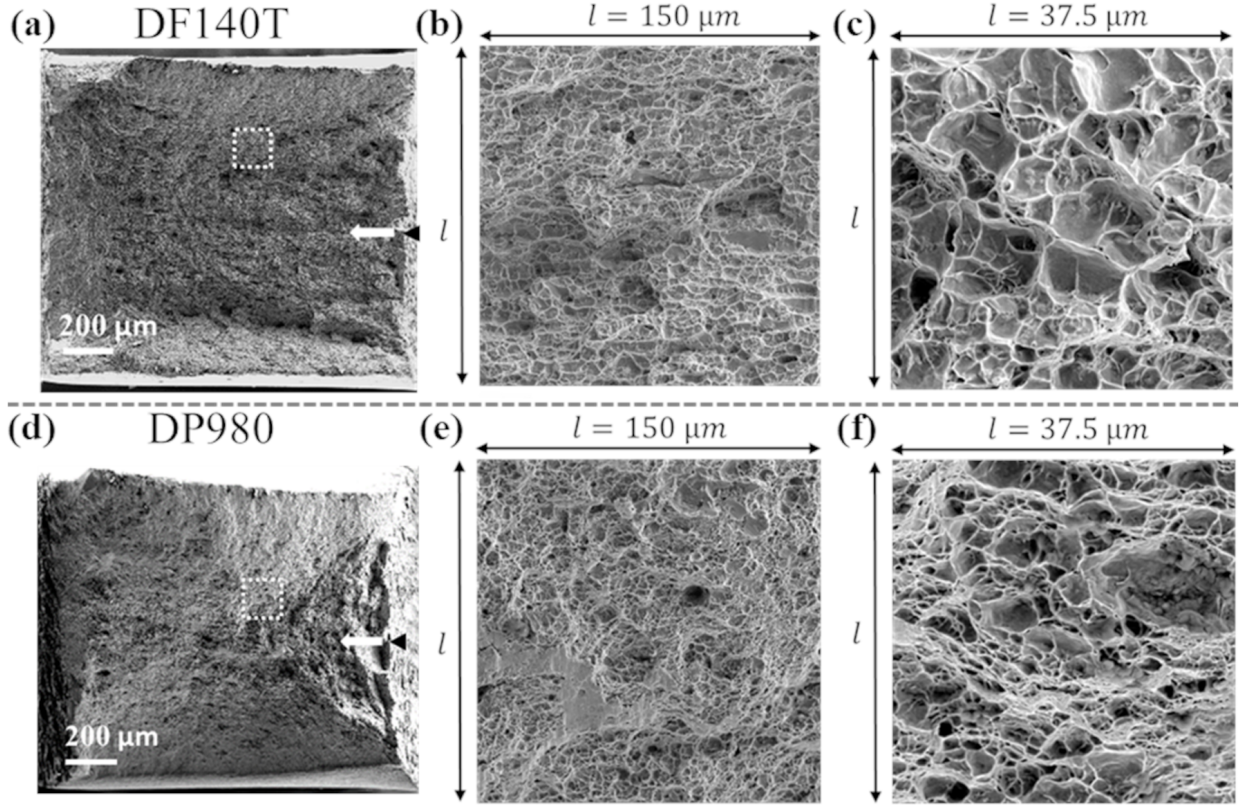


Figure 3: (a) SE-SEM image of the entire fracture surface and (b)-(c) high-resolution SE-SEM images of the region marked with a ‘white’ dotted box in (a) of a fractured SENT specimen of DF140T. (d) SE-SEM image of the entire fracture surface and (e)-(f) high-resolution SE-SEM images of the region marked with a ‘white’ dotted box in (d) of a fractured SENT specimen of DP980. The ‘black’ triangle in (a) and (d) marks the location of the initial notch tip while the ‘white’ arrow marks the direction of the crack growth.

3. Typological classification workflow

The workflow for the ML-based typological classification used in this work, follows from the work of Tsopanidis and Osovski (2021) and is schematically shown in Fig. 4. Prior to subjecting any image to the typological classification process, the resolution of the image was first reduced to $224 \text{ pixels} \times 224 \text{ pixels}$. This step, while not strictly mandatory, was performed to ensure that all images are of the same resolution as the images used in the training of the original CNN.

3.1. Features extraction

The first step of the typological classification workflow is features extraction. This was carried out using the VGG16 CNN (Simonyan and Zisserman, 2014) while utilizing the pre-trained weights obtained from training on the extensive ImageNet database (Russakovsky et al., 2015). The weights of the trained network, enable recognition and extraction of features which existed in the original ImageNet database. Building on the concept of transfer

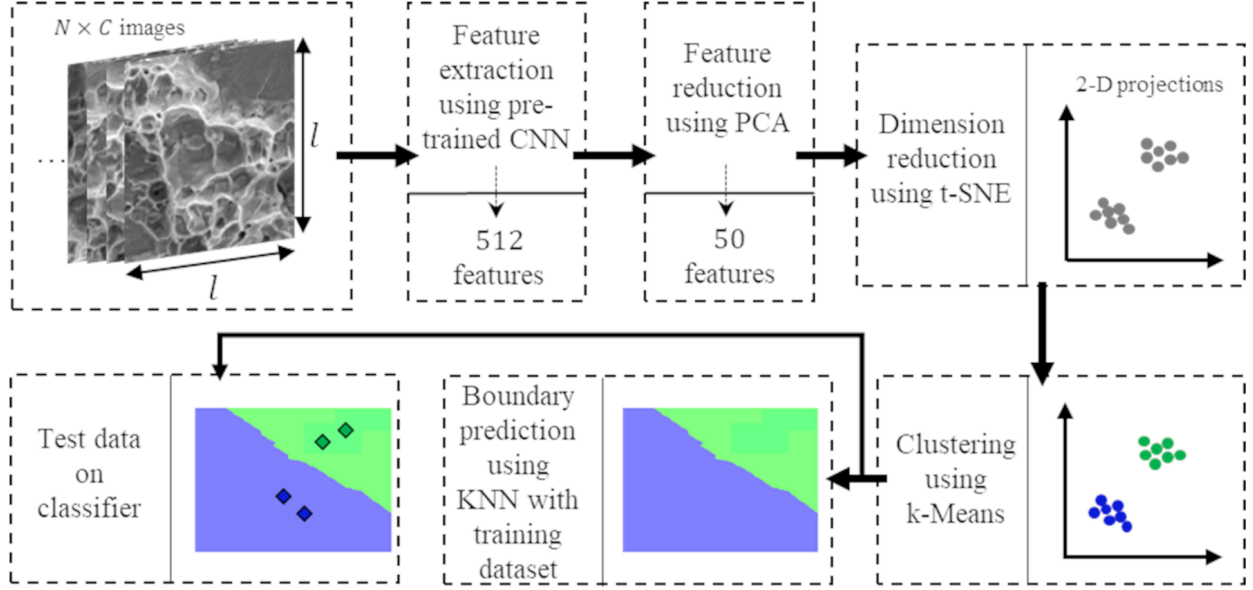


Figure 4: The workflow for ML-based typological classification.

learning it was previously shown that the extracted features suffice for the typological classification of fracture surfaces. For more details on the VGG16 architecture and its usage for fractography a reader is referred to Kitahara and Holm (2018); Tsopanidis et al. (2020); Tsopanidis and Osovski (2021).

As a result of passing through the VGG16 CNN, each image comprising the analyzed dataset is assigned a 512 features vector. Subsequently, the extracted features are subjected to a dimensionality reduction procedure using the Principal Component Analysis (PCA) (Pearson, 1901; Hotelling, 1933), as implemented in the scikit-learn package (Pedregosa et al., 2011). The PCA is set to reduce the 512 features vector of each image to 50 features, constructed by linear combination of the original features, such that the new features vector contains only the 50 most dominant descriptor used for explaining the variance between the different typological categories of images in the dataset.

Finally, the output of the PCA is projected onto a 2D space using t-distributed stochastic neighbor embedding (t-SNE) (Maaten and Hinton, 2008). The t-SNE algorithm is known for being sensitive to the chosen value of the perplexity. Hence, different perplexity values, ranging from 5 to 60 with an interval of 5 were tested. No significant difference was observed for the values of perplexity >30 and, thus, a value of 40 was chosen in this work.

3.2. Clustering

The t-SNE algorithm at the end of the features extraction step, embed each image in a 2D space based on its similarity (in terms of the features vector) to the rest of the images in the dataset. One may expect, that for a “well-behaved” dataset analyzed using a suitable feature extractor, the algorithm will map the images to a 2D space in which similar images (i.e., images belonging to a given typological category) will be grouped together while showing

some degree of separation from the images (i.e., images belonging to different typological categories) which exhibit greater dissimilarity. In other words, we expect that the images, once mapped to the 2D space will form clusters of images according to their typological category. If successful, this step will allow us to divide the 2D space into regions, with each region being occupied solely (or mostly) by images from one typological category. This process is commonly known as clustering.

Finally, according to the Euclidean distances in 2D space, the k-Means clustering algorithm (Lloyd, 1982; Forgey, 1965) assigns the data points exported by t-SNE into C clusters (each cluster correspond to a given typological category i.e., a combination of material and geometry). The evaluation of the clustering accuracy is based on the comparison of the label generated by the k-means algorithm against the ground-truth label assigned to each data point.

3.3. Classification

The end goal of the presented method is to predict the material and geometry of the specimen from which the input fracture surface image is obtained. To this end, the outputs of the t-SNE algorithm are shuffled and divided into two sub-sets used for training and testing while maintaining the ratio of images from each category to be 1:1, ensuring that the dataset is balanced. The training subset is used to train a predictor based on the k-nearest neighbors (KNN) algorithm (Fix and Hodges, 1989). The predictor, is expected to be able to classify previously unseen images (the testing sub-set) as belonging to one of the typological categories in the dataset on which it was trained. Here, a five-fold cross validation strategy was used for evaluating the accuracy of the classification process (Stone, 1974; Geisser, 1975).

4. Results: Clustering and classification

The objective of this work is twofold: (i) Extend the application of the unsupervised ML-based typological classification workflow shown in Fig. 4 (and described in Section 3) to not only classify the fracture surfaces of different materials but also differentiate between the loading conditions. (ii) Unravel the effect of length-scales associated with fracture surface images on the efficacy of typological classification process. To this end, we focus on the typological classification of the fracture surfaces of ST and SENT specimens of DF140T and DP980 dual-phase steels (described in Section 2).

We start our typological analysis of the fracture surface images by first considering the clustering results and accuracy of images originating from a fixed specimen geometry but different materials. The results of the clustering for the fracture surface images of the ST specimens of DF140T and DP980 steels are shown in Fig. 5(a) while that for the SENT specimens of the two steels are shown in Fig. 5(b). These results are obtained using a dataset of fracture surface images of size $18.75 \times 18.75 \mu\text{m}^2$. Figs. 5(a1) & (b1), show the 2D embedding of the entire image dataset obtained from the t-SNE algorithm, with colors representing the ground-truth labels for each point. In Fig. 5(a1), for the ST specimens there is a significant overlap between the two categories, while in Fig. 5(b1), for the SENT

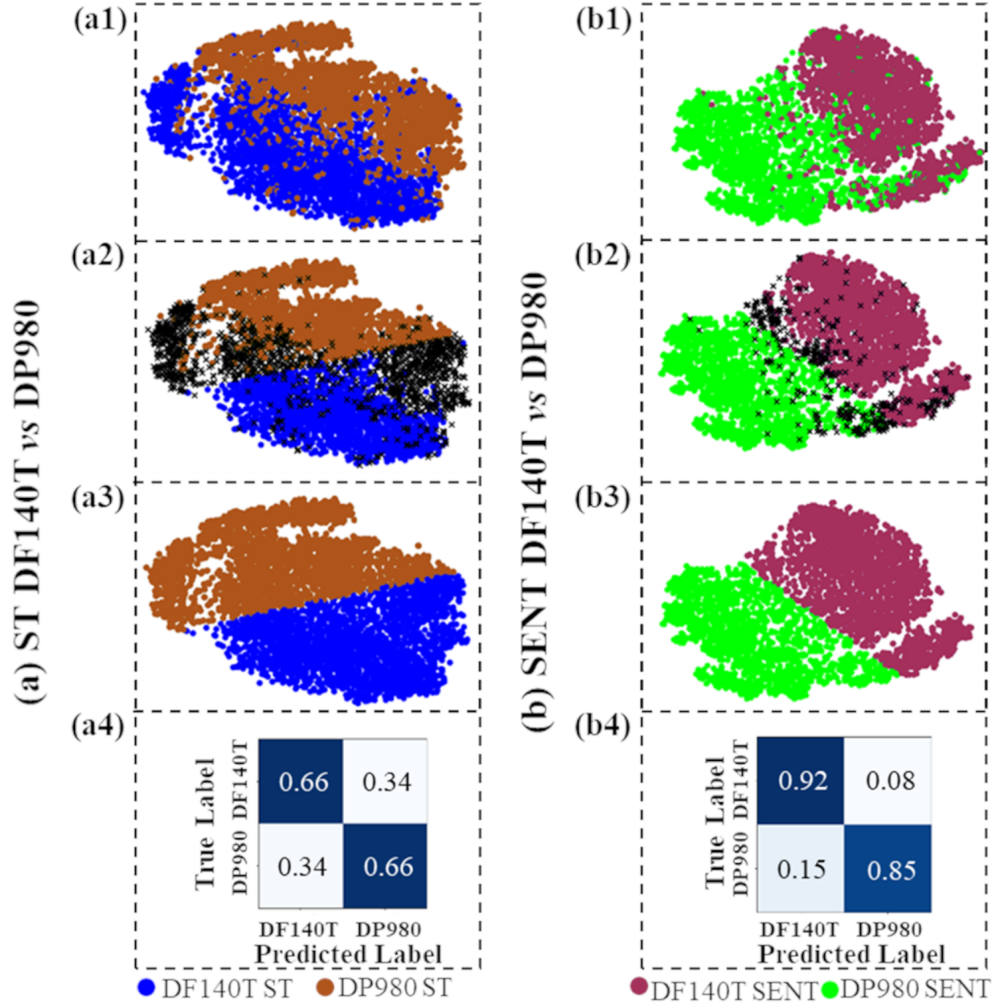


Figure 5: Clustering results for fracture surface images of size $18.75 \times 18.75 \mu\text{m}^2$ for typological analysis of (a) ST DF140T vs ST DP980 and (b) SENT DF140T vs SENT DP980 fracture surfaces. In (a) & (b), (a1) & (b1) show the results after t-SNE, (a2) & (b2) show the clustering results with confusion points (marked as 'x') of t-SNE and k-means, (a3) & (b3) show the k-means clusters with confusion points removed, and (a4) & (b4) show the confusion matrix.

specimens two distinct clusters can be seen with a small mixing zone between them. The errors resulting from the k-Means clustering are highlighted in Figs. 5(a2) and (b2), with the results shown without error markers in Figs. 5(a3) and (b3) for clarity. To evaluate the accuracy of the clustering, the ratio of mislabeled data points (marked as 'x' in Figs. 5(a2) & b2)) to the total number of samples in each category is used to construct the confusion matrices shown in Figs. 5(a4) and (b4).

The confusion matrices in Figs. 5(a4) and (b4) reveal that the overlap observed visually for the ST specimens (indicating poor separation in the clustering) indeed leads to lower accuracy of the clustering results. In contrast, the clustering of the SENT specimens shows

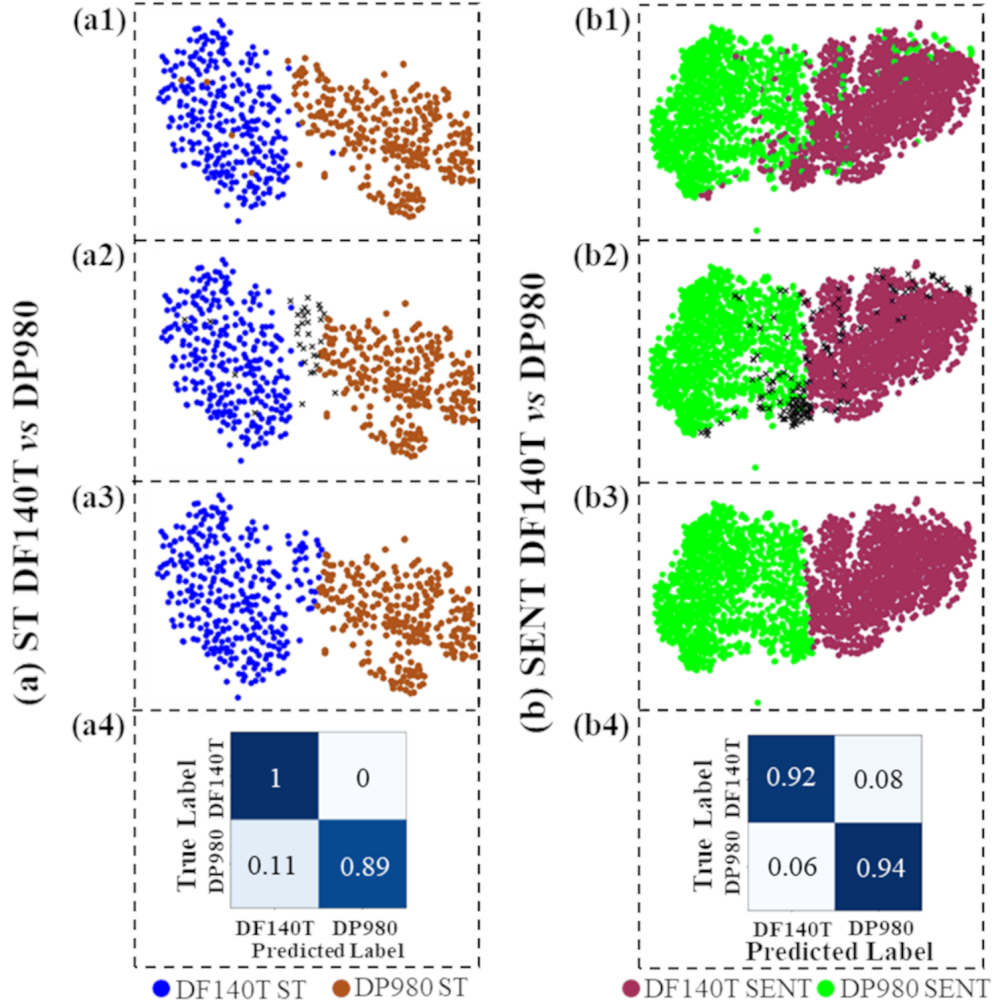


Figure 6: Clustering results for fracture surface images of size $50 \times 50 \mu\text{m}^2$ for typological analysis of (a) ST DF140T vs ST DP980 and that of size $25 \times 25 \mu\text{m}^2$ for typological analysis of (b) SENT DF140T vs SENT DP980 fracture surfaces. In (a) & (b), (a1) & (b1) show the results after t-SNE, (a2) & (b2) show the clustering results with confusion points (marked as 'x') of t-SNE and k-means, (a3) & (b3) show the k-means clusters with confusion points removed, and (a4) & (b4) show the confusion matrix.

significantly better accuracy. It is worth noting that in both Figs. 5(a) and (b), we are clustering the same two materials, with the only difference being the specimen geometry. This suggests that the influence of different loading conditions on the fracture process leads to differences in the embedded features in the fracture surfaces. This is despite the fact that both the specimens of the two steels undergo ductile fracture due to nucleation and growth of micro-scale voids.

To explore whether the image size has an effect on the clustering accuracy, we repeated the procedure used to generate Fig. 5 on a dataset of fracture surface images of size $50 \times 50 \mu\text{m}^2$ for the ST specimens and that of size $25 \times 25 \mu\text{m}^2$ for the SENT specimens. The

results of this exercise are summarized in Fig. 6. From the figure, it is clearly evident that the algorithm successfully creates two separate clusters for the images in the datasets with greater length-scales (Figs. 6(a1) and (b1)), which leads to a reduction in errors (Figs. 6(a2) and (b2)). The confusion matrices in Figs. 6(a4) and (b4) show that the clustering accuracy for the ST specimens has improved drastically from [0.66, 0.66] to [1, 0.89] while there is also a small improvement (from [0.92, 0.85] to [0.92, 0.94]) in the clustering accuracy for the SENT specimens.

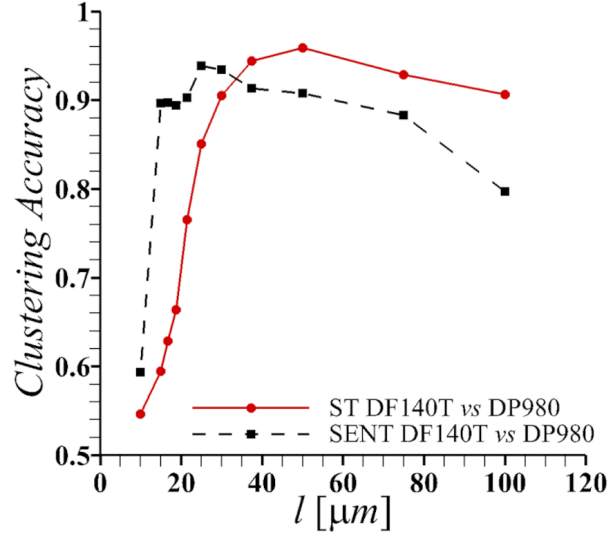


Figure 7: Variation of the average clustering accuracy with the length-scale (l) of the fracture surface images for typological analysis of ST DF140T vs ST DP980 and SENT DF140T vs SENT DP980 fracture surfaces.

The comparison of the results presented in Figs. 5 and 6 clearly shows that the clustering accuracy of the typological analysis of fracture surface images depends on the size of the images. This is further elucidated in Fig. 7 that presents the variation of the clustering accuracy with the length-scale (i.e., the image size) associated with the fracture surface images for typological analysis of ST DF140T vs ST DP980 and SENT DF140T vs SENT DP980 fracture surfaces. For simplicity, Fig. 7 only shows the average clustering accuracy for each case.

As shown in the figure, at smaller length-scales, the clustering accuracy is relatively poor. This is likely due to the fact that the salient features of the fracture surfaces of a given typological category are not fully contained within the smaller images. So the features extracted and reduced to capture the variance between different typological categories of fracture surfaces cannot resolve the differences between them. In other words, in scenarios in which the spatial arrangement of the features in the fracture surface images are of importance and not just the features themselves (e.g. location of small dimples with respect to larger ones), a large enough length-scale then must be set to resolve the differences between two typological categories. Using the same line of thought, we postulate that the larger images properly capture the salient features in the different categories of fracture surfaces allowing

the clustering algorithm to resolve the differences between them. Furthermore, as can be seen in Fig. 7, the critical length-scale corresponding to the maximum clustering accuracy depends on the combination of the typological categories being analyzed. The critical length-scale for the typological analysis of ST DF140T vs ST DP980 is $50\mu\text{m}$, while that for SENT DF140T vs SENT DP980 is $25\mu\text{m}$.

It is worth noting that, in Fig. 7, the clustering accuracy improves with an increase in length-scale, despite the corresponding decrease in dataset size. This suggests that the length-scale effect is the likely cause of any observed improvements in accuracy with increasing length-scale. As also shown in Fig. 7, beyond the critical length-scale, the clustering accuracy does not improve but rather tends to decrease. This decrease in the clustering accuracy is likely due to a combination of two factors: the loss of information resulting from rescaling of the images to $224 \text{ pixels} \times 224 \text{ pixels}$, and the reduction in the overall number of images in each category as the image size increases, which amplifies the impact of any clustering errors on the overall accuracy. For example, the rescaling of a $50 \times 50 \mu\text{m}^2$ image leads to the averaging of a 6×6 pixel window into a single pixel while the rescaling of $18.75 \times 18.75 \mu\text{m}^2$ images only averages a 2×2 pixel window. Similarly, the dataset of $50 \times 50 \mu\text{m}^2$ images contains 288 images per category while the dataset of $18.75 \times 18.75 \mu\text{m}^2$ images contains 2048 images per category. Nevertheless, it is worth noting that both of these factors are not random artifacts. Technologically, there is often a trade-off between resolution and field of view during imaging, and there is an obvious limit on the number of images of a given size that can be captured from a finite-size fracture surface.

We now extend our analysis to include all four typological categories of fracture surfaces. The results of this exercise are summarized in Fig 8. As evident from Fig. 8(a), for the subset of images with dimensions of $50 \times 50 \mu\text{m}^2$, the two geometries (ST, SENT) of the DF140T steel are embedded such that they produce two distinct clusters, separated in 2D space. On the other hand, the data points corresponding to the two geometries (ST, SENT) of the DP980 steels are not entirely localized as some of these points also populate the space occupied by each other as well as the space occupied by DF140T (mostly SENT) clusters. This lack of distinct spatial locality can be seen visually in Fig.8(b), where the mislabeled data points are marked with an ‘x’. To quantify the overlap of clusters and the resulting clustering accuracy, the confusion matrix is given in Fig.8(c). As can be expected from the visual inspection, the clustering accuracy for the two DF140T clusters is extremely high (0.86 and 0.94 for ST and SENT geometries, respectively), while the two DP980 categories exhibit decent but relatively lower accuracy (0.8 and 0.61 for ST and SENT geometries, respectively). Additionally, the confusion matrix shows that DP980 ST (if mislabeled) is mostly mislabeled as DP980 SENT, while DP980 SENT (if mislabeled) is mislabeled as either DP980 ST or DF140T SENT. Recall that the confusion between DF140T and DP980 SENT specimen was much smaller for the same image size when only the two categories were considered. This is because as the number of categories increases, the features extraction and related steps become more complex and challenging.

Next, to investigate the effect of image size or length-scale on the classification accuracy of all four categories, the average accuracy versus length-scale is plotted in Fig. 9(a) along with the minimum and maximum values obtained from five-fold cross-validation. As the

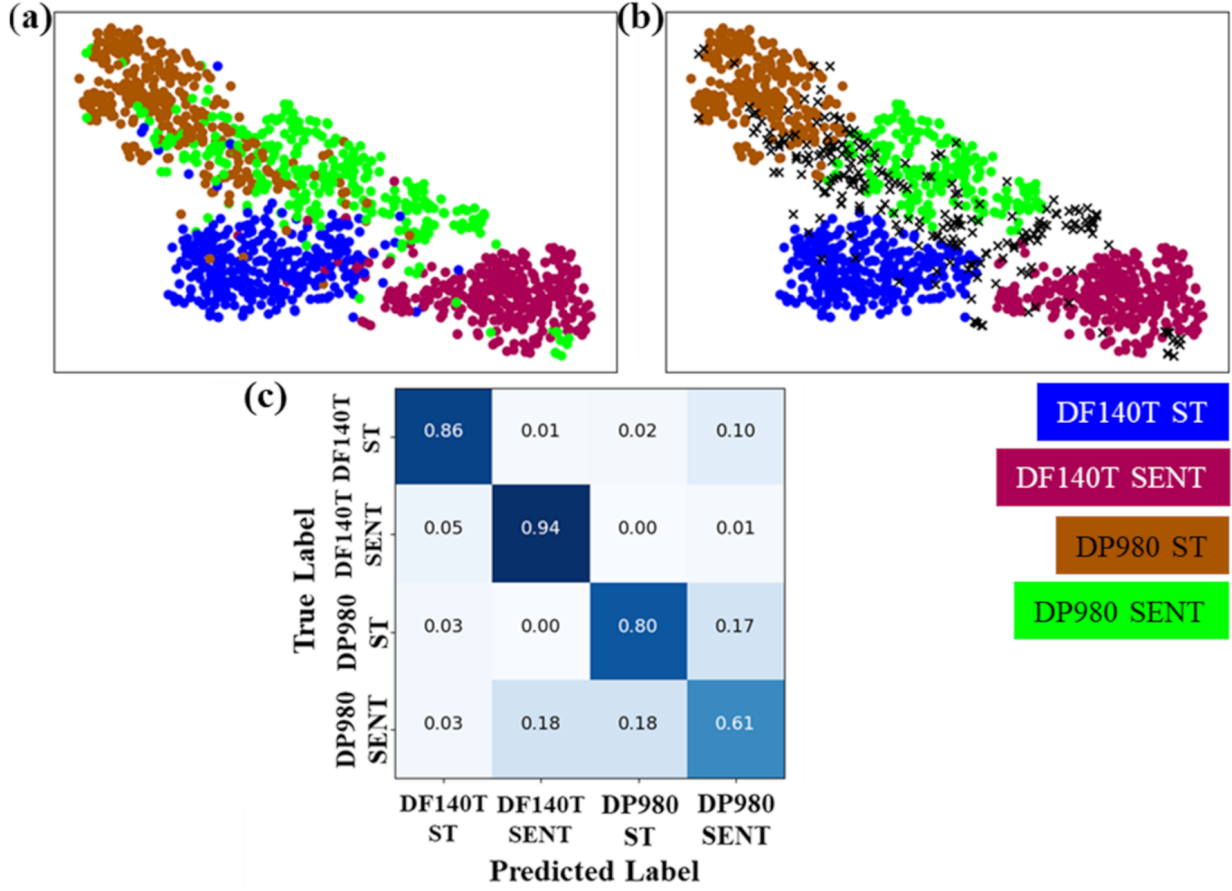


Figure 8: Clustering results for fracture surface images of size $50 \times 50 \mu\text{m}^2$ for typological analysis of all four categories, DF140T ST, DP980 ST, DF140T SENT and DP980 SENT, of fracture surfaces. (a) The results after t-SNE. (b) The clustering results with confusion points (marked as 'x') of t-SNE and k-means. (c) The confusion matrix.

image size increases, we see a continuous improvement in the average classification accuracy, reaching a maximum average accuracy of 0.82 for images of size $50 \times 50 \mu\text{m}^2$. For length-scales greater than $50 \mu\text{m}$, the classification accuracy, however, starts to decrease. The results of the classification of unseen data points (test dataset) predicted by the trained KNN algorithm for the $50 \times 50 \mu\text{m}^2$ images are also shown in Fig. 9(b). As can be seen in the figure, consistent with the clustering results, most of the errors are due to mislabeling of the DP980 ST fracture surfaces.

5. Discussion: Critical length-scale

Herein, we have demonstrated the applicability of an unsupervised ML-based typological classification workflow to carry out classification of fracture surfaces of two specimen geometries of two materials. To this end, fracture surfaces of ST and SENT specimens of

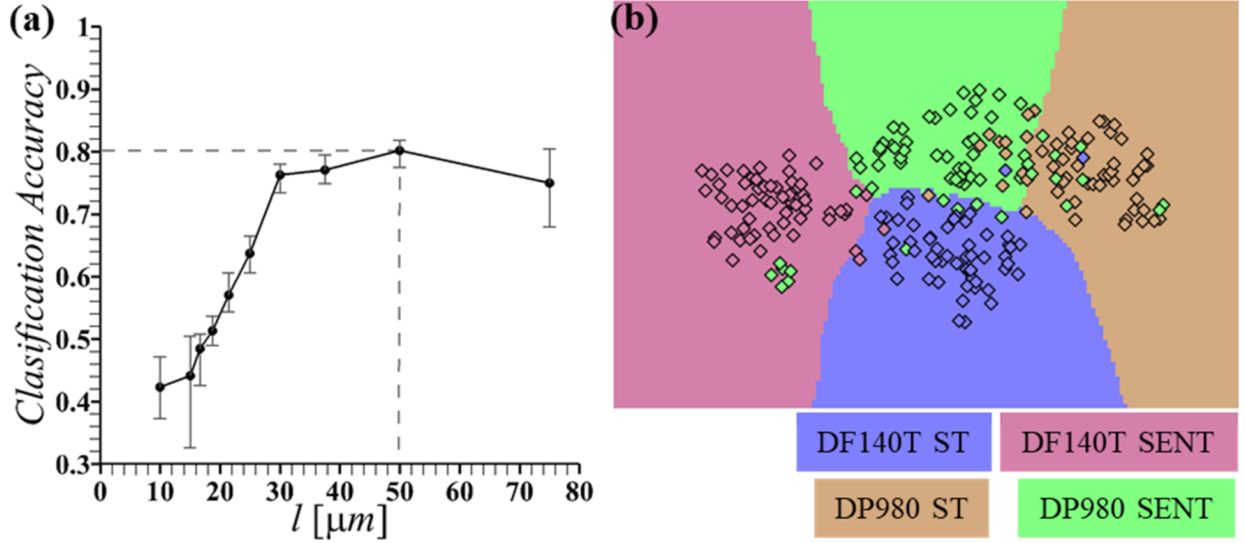


Figure 9: (a) Variation of the classification accuracy with the length-scale (l) of the fracture surface images for typological classification of all four categories, DF140T ST, DP980 ST, DF140T SENT and DP980 SENT, of fracture surfaces. The error bars in (a) show the minimum and the maximum values obtained from the five-fold cross validation. (b) The classification map for fracture surface images of size $50 \times 50 \mu\text{m}^2$ for typological classification of all four categories, DF140T ST, DP980 ST, DF140T SENT and DP980 SENT, of fracture surfaces.

DF140T and DP980 dual-phase steels are analyzed and classified. The ST and SENT specimens of both the steels undergo ductile fracture due to nucleation, growth and coalescence of micro-scale voids. Note that it is not easy to classify their fracture surfaces into the four typological categories by simple visual inspection as they all exhibit similar dimpled fracture surfaces. More importantly, we discovered that the accuracy of the clustering and classification algorithms is sensitive to the length-scale (i.e., the image size) of the fracture surface images. There is a critical length-scale for which the accuracy of the typological clustering and classification process is maximum. We also found that the critical length-scale corresponding to the maximum accuracy is not unique and depends on the typological categories being classified, meaning that the critical length-scale differs when classifying ST DF140T vs ST DP980, SENT DF140T vs SENT DP980, or all four categories (ST DF140T, ST DP980, SENT DF140T, and SENT DP980) of fracture surfaces.

In general, the critical length-scales corresponding to maximum clustering and classification accuracy are greater than the microstructural length-scales (see Table 1) of the two materials. This is not surprising, as the ductile fracture process involves the interaction of the deformation fields at both the structural (determined by the specimen geometry and loading conditions) and microstructural (determined by the material microstructure and microscale property distributions) length-scales (Zheng et al., 2020). However, the resulting fracture surfaces are expected to contain information about the interaction of these deformation fields. In fact, quantitative correlations between length-scales extracted from

the fracture surface roughness (that are greater than the microstructural length-scales) and fracture toughness have been obtained both computationally (Srivastava et al., 2014; Osowski et al., 2015) and experimentally (Barak et al., 2019) for ductile materials. Following this, we analyzed the roughness of the fracture surfaces of all four specimens. To this end, 3D digital elevation maps of $550 \times 550 \mu\text{m}^2$ region (with a resolution of $3300 \text{ pixels} \times 3300 \text{ pixels}$) from the center of the fracture surfaces of all four specimens were captured using a high-resolution Olympus DSX 510 digital microscope. An example of 3D digital elevation map is shown in Fig. 10(a).

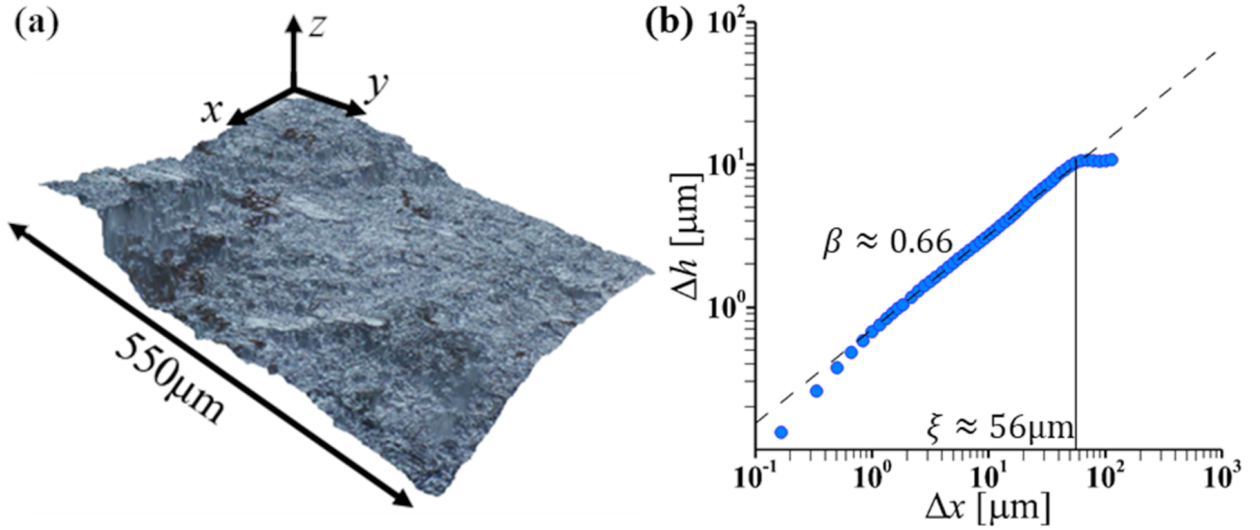


Figure 10: (a) A representative 3D digital elevation map of a $550 \times 550 \mu\text{m}^2$ region extracted from the center of the fracture surface of a fractured DP980 ST specimen. The x -axis in (a) is along the width while y -axis is along the thickness of the specimen. (b) A representative height-height correlation function, Eq. (2), of the fracture surface roughness in (a) for a line profile along the x -axis. The roughness exponent (β) and the correlation length (ξ) are also marked in (b).

Next, from each 3D digital elevation map of the fracture surfaces, we extracted line profiles of the variation in the out-of-plane height of the fracture surface along the width of the specimens (x -axis, which is also the crack growth direction in the SENT specimens) and calculated the height-height correlation function:

$$\Delta h(\Delta x) = \sqrt{\langle [h(x + \Delta x) - h(x)]^2 \rangle_x} \quad (2)$$

where h is the height at a location x and $\langle \rangle_x$ denotes average over x . To avoid any artifacts near the edges, the height profiles were extracted at intervals of $5 \mu\text{m}$ from within a $275 \times 275 \mu\text{m}^2$ region in the center of the digital elevation maps.

The quantity $\Delta h(\Delta x)$ in Eq. (2) is simply the difference in height between two points separated by a distance Δx . A representative log-log plot of the correlation function is shown in Fig. 10(b). The correlation functions of the fracture surface roughness exhibit a power law behavior:

$$\Delta h(\Delta x) \propto \Delta x^\beta \quad (3)$$

where β is the roughness or the Hurst exponent and lies between 0 and 1. The value $\beta = 0$ corresponds to a straight line with zero slope, while $\beta = 1$ corresponds to a straight line with a nonzero slope. The value $\beta = 0.5$ corresponds to a random walk suggesting no correlation between Δh and Δx . The value of $\beta > 0.5$ indicate persistence i.e. an increase (decrease) in the value of Δh with increasing (decreasing) value of Δx , while $\beta < 0.5$ indicate anti-persistence i.e. a decrease (increase) in the value of Δh with increasing (decreasing) value of Δx . The value of β (estimated as the moving average slope of a correlation function on the log-log scale) for the correlation functions extracted from all fracture surfaces was found to vary with Δx (especially at smaller and greater values of Δx). Thus, a correlation length, ξ , is defined as the value of Δx at which the value of β first decreases below 0.45. In other words, the correlation function is persistent for $\Delta x < \xi$ and is anti-persistent for $\Delta x > \xi$.

The average values ($\pm$ standard error) of ξ for DF140T ST, DP980 ST, DF140T SENT and DP980 SENT are estimated to be $22.6(\pm 2.5)\mu\text{m}$, $53.2(\pm 4.7)\mu\text{m}$, $17.0(\pm 3.0)\mu\text{m}$ and $31.7(\pm 2.8)\mu\text{m}$, respectively. Recall, that the critical length-scale corresponding to the maximum clustering accuracy for typological analysis of ST DF140T vs ST DP980 fracture surfaces is found to be $50\mu\text{m}$, while that for SENT DF140T vs SENT DP980 fracture surfaces is found to be $25\mu\text{m}$. This shows that the critical length-scale for maximum clustering accuracy is close to the greatest value of ξ among the typological categories being analyzed. Similarly, the maximum classification accuracy for typological analysis of all four categories of fracture surfaces is found to be $50\mu\text{m}$. This is close to the greatest value of ξ among the four typological categories being analyzed.

In Zheng et al. (2020) through in-situ quantitative SEM mechanical tests, it was shown that at the microstructural-scale the deformation in both ST and SENT specimens of the two steels is always heterogeneous with the extent of heterogeneity being more pronounced in DF140T compared to DP980 steel. Moreover, the interlacing of the heterogenous deformation at the structural and the microstructural length-scales in the SENT specimens of DF140T steel results in an even higher degree of localized deformation (largely due to the presence of softer non strain-hardening ferrite phase). Although this physical understanding of the deformation process leading to the final fracture of these steels still does not clearly explain the exact way in which a combination of material and loading conditions affects the morphology of the fracture surface. However, it can be hypothesized that the more localized the deformation at the microstructural-scale, the lower the value of ξ , and therefore, the lower the critical length-scale for maximum clustering and classification accuracy.

Our work clearly highlights the significance of length-scales in the application of ML-based techniques to materials science problems involving image analysis. The consideration of proper length-scales not only improves the efficacy but also provides physical insights into the image analysis process. The consideration of proper image length-scales in ML-based techniques can also enhance transfer learning by allowing the same ML algorithm to be trained on one dataset and then applied to another dataset that has different image sizes or

features. For example, as shown in Tsopanidis et al. (2020), changing the viewfield of the SEM while capturing the fracture surface images can allow an ML-based technique trained to analyze intergranular and transgranular features in the fracture surfaces of Mg-Spinel specimens to be applied to Alumina specimens. Thus, by considering proper length-scales, transfer learning can be improved in two ways. Firstly, the knowledge of critical length-scales can aid in selecting suitable image sizes for training and testing the ML algorithms. Secondly, the understanding of the effect of length-scales can be used to transform the images to a common scale or size to facilitate the transfer of learned features or models between different datasets. We hope that our work will inspire and instigate future works aimed at enhancing ML-based techniques by incorporating the notion of physical length-scales.

6. Concluding remarks

Using an ML-based typological classification workflow, we have carried out the typological classification of the fracture surfaces of simple tension and single-edge notch tension specimens of two dual-phase steels, DF140T and DP980. Both the specimens of the two materials undergo ductile fracture due to nucleation, growth and coalescence of microscale voids, and exhibit similar dimpled fracture surfaces. The ML-based workflow uses a pre-trained convolutional neural network in an unsupervised learning mode to extract image feature descriptors. The extracted features are then subjected to dimensionality reduction using principal component analysis. The workflow then clusters and classifies the images using K-Means and K-Nearest Neighbors algorithms, respectively.

Our results show that the accuracy of the ML-based workflow is sensitive to the length-scales of the fracture surface images, and the critical length-scales corresponding to the maximum accuracy depends on the typological categories being classified. To interpret the physical origin of critical length-scales, we performed quantitative analysis of the fracture surface roughness. To this end, the height-height correlation functions of the roughness of all the fracture surfaces were quantified, and a correlation length as the point at which the correlation function transitions from being persistent to anti-persistent was defined. Comparison of the critical length-scales and the correlation lengths revealed that the critical length-scales are always close to the greatest value of the correlation lengths among the typological categories being analyzed.

Our work demonstrates the potential of using unsupervised ML-based techniques for fractography of ductile materials, particularly for typological classification. More importantly, our work not only highlights the significance of length-scales in image analysis in materials science, but also provides motivation to enhance ML-based techniques by incorporating physical length-scales.

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.

Acknowledgments

This work was partially supported by the U.S. National Science Foundation grant CMMI - 1944496. SO and AS also acknowledge the European Union’s Horizon 2020 Programme (Excellent Science, Marie-Sklodowska - Curie Actions, H2020-MSCA-RISE-2017) under REA Grant agreement 777896 (Project QUANTIFY).

References

- ASM Handbook, 1987. Volume 12: Fractography. Handbook, ASM International.
- ASTM E112-13, 2013. Standard test methods for determining average grain size. Standard, ASTM International.
- Barak, Y., Srivastava, A., Osovski, S., 2019. Correlating fracture toughness and fracture surface roughness via correlation length scale. *International Journal of Fracture* 219 (1), 19–30.
- Bastidas-Rodriguez, M., Polania, L., Gruson, A., Prieto-Ortiz, F., 2020. Deep learning for fractographic classification in metallic materials. *Engineering Failure Analysis* 113, 104532.
- Choudhary, K., DeCost, B., Chen, C., Jain, A., Tavazza, F., Cohn, R., Park, C. W., Choudhary, A., Agrawal, A., Billinge, S. J., et al., 2022. Recent advances and applications of deep learning methods in materials science. *npj Computational Materials* 8 (1), 1–26.
- DeCost, B., Lei, B., Francis, T., Holm, E., 2019. High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel. *Microscopy and Microanalysis* 25 (1), 21–29.
- Fix, E., Hodges, J., 1989. Discriminatory analysis. nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique* 57 (3), 238–247.
- Forgey, E., 1965. Cluster analysis of multivariate data: Efficiency vs. interpretability of classification. *Biometrics* 21 (3), 768–769.
- Geisser, S., 1975. The predictive sample reuse method with applications. *Journal of the American statistical Association* 70 (350), 320–328.
- Gerbig, D., Srivastava, A., Osovski, S., Hector, L. G., Bower, A., 2018. Analysis and design of dual-phase steel microstructure for enhanced ductile fracture resistance. *International Journal of Fracture* 209 (1), 3–26.
- Holm, E., Cohn, R., Gao, N., Kitahara, A., Matson, T., Lei, B., Yarasi, S., 2020. Overview: Computer vision and machine learning for microstructural characterization and analysis. *Metallurgical and Materials Transactions A*, 1–15.
- Holm, E. A., 2019. In defense of the black box. *Science (New York, N.Y.)* 364 (6435), 26–27.
- Hotelling, H., 1933. Analysis of a complex of statistical variables with principal components. *J. Educ. Psy* 24, 498–520.
- Kitahara, A., Holm, E., 2018. Microstructure cluster analysis with transfer learning and unsupervised learning. *Integrating Materials and Manufacturing Innovation* 7 (3), 148–156.
- Lloyd, S., 1982. Least squares quantization in pcm. *IEEE transactions on information theory* 28 (2), 129–137.
- Maaten, L., Hinton, G., 2008. Visualizing data using t-sne. *Journal of machine learning research* 9 (11).
- Naik, D., Kiran, R., 2019. Identification and characterization of fracture in metals using machine learning based texture recognition algorithms. *Engineering Fracture Mechanics* 219, 106618.
- Osovski, S., Srivastava, A., Ponson, L., Bouchaud, E., Tvergaard, V., Ravi-Chandar, K., Needleman, A., 2015. The effect of loading rate on ductile fracture toughness and fracture surface roughness. *Journal of the Mechanics and Physics of Solids* 76, 20–46.
- Pearson, K., 1901. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2 (11), 559–572.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duch-

- esnay, E., 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
- Pineau, A., Benzerga, A. A., Pardoen, T., 2016a. Failure of metals i: Brittle and ductile fracture. *Acta Materialia* 107, 424–483.
- Pineau, A., McDowell, D. L., Busso, E. P., Antolovich, S. D., 2016b. Failure of metals ii: Fatigue. *Acta Materialia* 107, 484–507.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision* 115 (3), 211–252.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Srivastava, A., Bower, A., Hector, L., Carsley, J., Zhang, L., Abu-Farha, F., 2016. A multiscale approach to modeling formability of dual-phase steels. *Modelling and Simulation in Materials Science and Engineering* 24 (2), 025011.
- Srivastava, A., Ponson, L., Osovski, S., Bouchaud, E., Tvergaard, V., Needleman, A., 2014. Effect of inclusion density on ductile fracture toughness and roughness. *Journal of the Mechanics and Physics of Solids* 63, 62–79.
- Stone, M., 1974. Cross-validators choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)* 36 (2), 111–133.
- Tsopanidis, S., Moreno, R., Osovski, S., 2020. Toward quantitative fractography using convolutional neural networks. *Engineering Fracture Mechanics* 231, 106992.
- Tsopanidis, S., Osovski, S., 2021. Unsupervised machine learning in fractography: Evaluation and interpretation. *Materials Characterization* 182, 111551.
- Zheng, X., Ghassemi-Armaki, H., Srivastava, A., 2020. Structural and microstructural influence on deformation and fracture of dual-phase steels. *Materials Science and Engineering: A* 774, 138924.