

Review

Overview: Machine Learning for Segmentation and Classification of Complex Steel Microstructures

Martin Müller ^{1,2,*}, Marie Stiefel ^{1,2} , Björn-Ivo Bachmann ^{1,2} , Dominik Britz ^{1,2} and Frank Mücklich ^{1,2} 

¹ Department of Materials Science, Campus D3.3, Saarland University, 66123 Saarbruecken, Germany; marie.stiefel@uni-saarland.de (M.S.); bjoernivo.bachmann@uni-saarland.de (B.-I.B.); d.britz@mec-s.de (D.B.); frank.muecklich@uni-saarland.de (F.M.)

² Material Engineering Center Saarland, Campus D3.3, 66123 Saarbruecken, Germany

* Correspondence: martin.mueller1@uni-saarland.de; Tel.: +49-681-302-70548

Abstract: The foundation of materials science and engineering is the establishment of process–microstructure–property links, which in turn form the basis for materials and process development and optimization. At the heart of this is the characterization and quantification of the material’s microstructure. To date, microstructure quantification has traditionally involved a human deciding what to measure and included labor-intensive manual evaluation. Recent advancements in artificial intelligence (AI) and machine learning (ML) offer exciting new approaches to microstructural quantification, especially classification and semantic segmentation. This promises many benefits, most notably objective, reproducible, and automated analysis, but also quantification of complex microstructures that has not been possible with prior approaches. This review provides an overview of ML applications for microstructure analysis, using complex steel microstructures as examples. Special emphasis is placed on the quantity, quality, and variance of training data, as well as where the ground truth needed for ML comes from, which is usually not sufficiently discussed in the literature. In this context, correlative microscopy plays a key role, as it enables a comprehensive and scale-bridging characterization of complex microstructures, which is necessary to provide an objective and well-founded ground truth and ultimately to implement ML-based approaches.



Citation: Müller, M.; Stiefel, M.; Bachmann, B.-I.; Britz, D.; Mücklich, F. Overview: Machine Learning for Segmentation and Classification of Complex Steel Microstructures. *Metals* **2024**, *14*, 553. <https://doi.org/10.3390/met14050553>

Academic Editor: João Manuel R. S. Tavares

Received: 28 March 2024

Revised: 25 April 2024

Accepted: 3 May 2024

Published: 7 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The term microstructure refers to the inner structure of a material which, on one hand, stores its genesis and processing history, and, on the other hand, determines all its physical and chemical properties [1]. The microstructure is considered the central information carrier [2]. Therefore, the phases contained in the microstructure, including their distribution, shapes, and sizes, are decisive, and a correct analysis of the microstructure is crucial, in consideration of the overarching objective of understanding process–microstructure–property links. In fact, the materials science and engineering (MSE) community is in the midst of a paradigm change from empirical process–property correlation to microstructure-based development of new materials [1]. Consequently, characterization, analysis and quantification of the microstructure are all the more important. A deep understanding of the microstructure plays an increasingly important role for both quality assurance and research and development for materials and process optimization. Today’s challenge of an ever-growing economy, paired with the preservation of resources and the need for a circular economy in order to counteract climate change, makes the ability to optimize materials more valuable than ever.

Today, microstructure analysis is still frequently performed manually and often provides qualitative statements only, making them a bottleneck in microstructure-based materials development and process control. A microstructure analysis pipeline generally consists

of a metallographic preparation step, followed by contrasting, a segmentation of the microstructural components and their quantification and classification. Especially for these latter analysis steps, machine learning (ML) approaches promise substantial benefits [2]. In fact, ML has already replaced conventional solutions in computer vision and is employed, for example, for obstacle recognition in autonomous driving [3]. As many human vision tasks cannot be adequately solved using a simple deterministic, rule-based solution, the significance of ML lies in the fact that it makes problems accessible to automatic processing by computers for which full mathematical modeling is hopeless.

Recently, a major focus of ML in MSE has been on the ML-aided materials discovery and design [4–9] which involves discovering new compounds with promising combinations of properties. However, most of the work explores the “chemical space”, i.e., the possible chemical combinations of novel materials, with (so far) little emphasis on their microstructure [10,11]. Additionally, the capabilities of ML-aided analysis of big data seem to encourage researchers to again focus on process–property–correlations and shift away from process–microstructure–property-links. Yet the microstructure is an essential pillar of materials design [12] and its incorporation into materials design processes is significant, with process–microstructure–property relationships being more relevant than process–property only relationships [13]. Accordingly, microstructure analysis and quantification are relevant in many different respects: the optimization of new materials with regard to their microstructure, the quality control of existing materials and the continuous microstructure-based improvement of existing materials.

While materials become more advanced and their microstructures become increasingly smaller in size and more complex, consisting of a combination of different phases or constituents with different substructures [1], tolerances become narrower and quality requirements more strict, resulting in existing characterization methods reaching their limits. Microstructure analyses, notably microstructure classification and segmentation, are still often carried out manually. Not only is this time-consuming, but also subjective and poorly reproducible. In addition, simple assessments using reference images and comparison charts, or qualitative estimations of phase fractions do not extract all available information from the microstructure. While some simple computer-based approaches to automating these tasks exist, e.g., threshold segmentation, they quickly reach their limits with complex microstructures.

Accordingly, these approaches represent the limitations for establishing process–microstructure–property correlations and for microstructure-based material development. This issue can be tackled with ML-based approaches, allowing an automatization of microstructure analyses with the ability to treat large amounts of high-dimensional data, achieving increased objectivity and reproducibility, extracting all relevant information from the microstructure and enabling new analysis approaches of complex microstructures and thus bearing previously unused potential.

In fact, recently, ML has experienced a downright hype, reflected in the myriad of publications concerning artificial intelligence (AI). Despite the resulting acceleration of AI technology development, this includes examples where ML is applied only as an end in itself and viewed as a panacea, but with lack of consideration for the required amount and quality of data, the origin of the ground truth and the complex material-specific questions and the required domain expertise [14].

This review aims to showcase the possibilities of ML in microstructure analysis and at the same time to provide more clarity on the ground truth needed to train ML models as well as the data foundation, e.g., quantity and quality of training data and metallographic aspects that need to be considered during data acquisition. This is demonstrated using the examples of complex steel microstructures, where the classic tasks of grain size determination and phase analysis are investigated.

2. Domain Related Challenges in Materials Science

The aim of this section is to outline the domain-specific challenges of applying ML methods adopted from computer science to microscopic, microstructural images. In addition, a clear terminology of terms related to artificial intelligence should be defined first [15].

2.1. Machine Learning Methods for Microstructure Analysis

- Data science: roughly encompasses all steps in generating meaning and knowledge from data (data collection, data preparation, model building (including machine learning methods), application of the models). Data mining denotes the use of statistical or machine learning methods to detect interesting correlations or knowledge in large amounts of data.
- Computer vision (CV): broad term for a technology that enables machines to automatically recognize and describe images. Today, the main methods that are used for this purpose are artificial intelligence and machine learning.
- Artificial intelligence (AI): a branch of computer science which can be understood as the digitization of human knowledge skills and aims at performing tasks that normally require human intelligence. Subfields of AI are planning, reasoning and machine learning.
- Machine learning (ML): Subfield and fundamental method of AI where computers are enabled to independently learn patterns and regularities from available data without being specifically programmed for this task.
- ML can generally be divided into two sub-classes, namely, supervised and unsupervised learning. While supervised learning refers to learning based on annotated examples, unsupervised learning corresponds to finding patterns in the data autonomously, i.e., training without knowledge of a ground truth or human intervention.
- Deep learning (DL): Method of ML using so-called artificial neural networks (ANNs). ANNs can not only learn patterns and regularities from available data, but also detect important features in the data independently.
- Model training and model inference: In supervised ML, the algorithm receives a set of data and associated expected outcomes and learns a function that maps the input data to the output. The assignment of this so-called ground truth, i.e., the linking of the data with the associated expected outcomes by a human expert, is a crucial step in supervised ML. This learning phase results in a trained model, which can now be applied to new, unknown data of the same type (inference).

In microstructure analysis, ML tasks, derived from CV tasks, usually fall under the following categories [16], as visualized in Figure 1:

- Image classification: Identification of the content of an image, e.g., classification of single-phase microstructures or classification of defects.
- Semantic segmentation: Classification of each pixel of the image, e.g., for segmentation and subsequent quantification of multi-phase microstructures.
- Object detection: Location of individual objects within the image, e.g., for finding precipitates in microstructure. In everyday life, object detection is applied in autonomous driving for obstacle detection.
- Instance segmentation: Assignment of pixels to individual objects by identification and bounding of an object followed by segmentation of the pixels that belong to said object, often applied in combination with object detection.

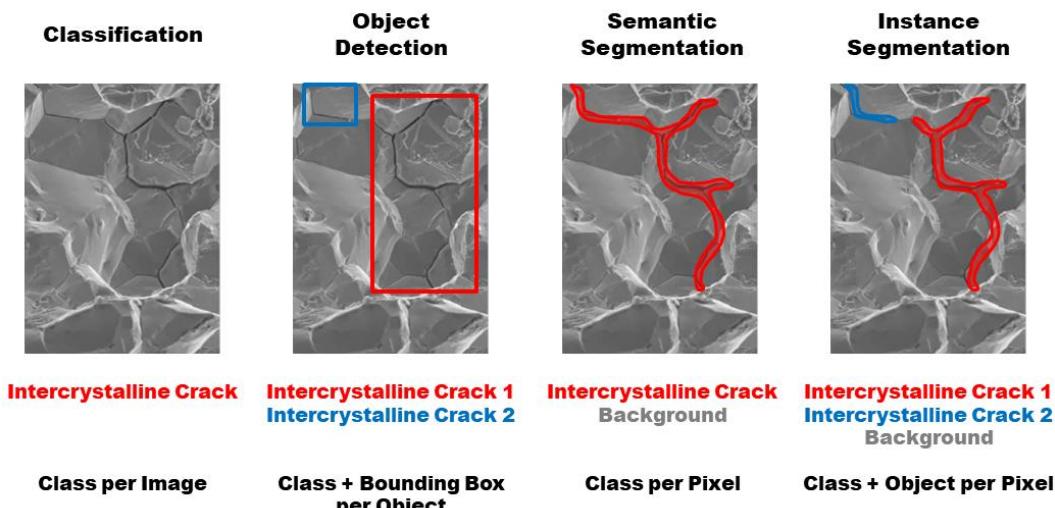


Figure 1. Visualization of the main categories of ML tasks in CV.

2.2. Domain Challenges and the Role of the Ground Truth

In supervised ML, a set of training data including the corresponding expected outcomes is required, with which the model will learn a function projecting the input on the output. This assignment of the expected outcomes to training data is called ground truth assignment and is the fundament for supervised machine learning. The terms label and annotations are also used synonymously with ground truth.

In MSE, as in other disciplines, the assignment of the ground truth represents the bottleneck for implementing an ML-based analysis. This not only concerns the time required, which is especially true for manual annotations for segmentation tasks, but also the challenge of providing an accurate and objective ground truth in view of the complex fine microstructures that may not always be fully determined. In MSE, this comes with the need for human domain expertise to decide a priori what to measure or analyze and create purpose-built datasets and ML models. However, due to its subjective nature, the required human input can impair the correctness of the algorithm. Simple and straightforward microstructures usually pose no problems, but this issue becomes relevant as microstructures are more complex with an inherent increased scope for interpretation. This data poses a higher risk for subjectiveness, which, in turn, can propagate through the entire ML workflow. Thus, objectiveness cannot be guaranteed for the final results of the model, e.g., class definition, image and class assignment or image region annotation. In the literature, the role of the ground truth is rarely discussed [14], and some publications lack a relevant materials science background, which can result in questionable methods and false interpretations [17,18]. Additionally, both methodological guidelines and widespread understanding within the community about required data are missing.

To better illustrate these challenges, and in order to discuss the dataset size needed for training a ML model, it is helpful to compare microscopic microstructural images to the natural scene, that is, everyday images of computer science datasets that are used to develop and benchmark CV related ML approaches.

Even from a purely computer science perspective and in relation to CV of the natural scene, everyday images, ML is more than “just some ML code” [19]. From an MSE point of view, this is all the more true. Whereas data generation, i.e., image acquisition and the assignment of ground truth, is simple for datasets of natural scenes and everyday images (e.g., ImageNet, Stanford Dogs Dataset, Cityscapes Dataset [20–22]), the steps for obtaining the image alone are significantly more complex for microscopic microstructure images. Image acquisition includes sample selection, sample preparation (grinding and polishing), sample contrasting (typically chemical etching) and the image acquisition itself via a suitable microscope (e.g., light optical microscope (LOM) or scanning electron microscope (SEM)). On the one hand, the final appearance of the microstructure under the

microscope depends on the experimental settings (e.g., choice of etching reagent, etching time, microscopy settings) and, on the other hand, may not be fully determined due to the complexity of the structures and may be evaluated differently by various experts. This subjective perception can be reinforced by the lack of consistent terminology for describing microstructures, e.g., in the steel domain for bainitic microstructures [23].

When implementing an ML-based microstructure analysis, the following questions must therefore also be answered: How should we deal with variances in metallographic processes and ambiguities in expert judgements in order to still exploit the advantages ML promises? Thus, it is paramount to view the ML implementation in a holistic way, with a focus on metallographic processes and variances and including a well-defined ground truth, to actually reach the desired accuracy, objectivity and reproducibility of the ML model.

One of the reasons for using ML is that it can handle variances well, as long as they are represented in the training data. However, although ML has already demonstrated its potential for a variety of tasks in microstructural analysis, the majority of publications deal with well-curated datasets, generated under laboratory conditions and exhibiting little variation (e.g., [17,24–29])—and a general understanding of robustness and generalization, as well as dataset size, occurring variances and the maximum variance an ML model can handle, is therefore still lacking.

It is to be assumed that ML models can handle variances well for tasks with clearly separable foreground and background (e.g., segmentation of grain boundaries), whereas variances become more critical the more complex the task is (e.g., differentiation of complex and similar microstructures such as lower bainite and tempered martensite [30]). Nevertheless, it is important to understand metallographic processes in terms of the variances that may occur and to restrict variances if necessary. On the one hand, a parameter space can be defined for the safe application of the ML model, and on the other hand, by restricting variances, a faster and more systematic implementation of the ML evaluation is made possible even with few data. The origin and degree of the variances and the amount of training data will be discussed in the later application examples.

It is commonly believed that small datasets are generally insufficient for ML model training, but this is only true to some extent. In fact, excellent results have been achieved with as few as 10 training images [16]. In terms of the amount of training data, the comparison of natural scenes and everyday images with microscopic microstructural images actually shows an advantage for the micrographs. Micrographs can be viewed as “data-rich”; i.e., they are rich in relevant information. They contain more regions of interest than natural scene images (e.g., several hundred particles or grains), lack background that does not contribute to the analysis task, are statistically representative of the material and its microstructure and are usually captured from a fixed field of view [16]. In this context, it is also beneficial that microscopic microstructural images usually have higher resolution, and therefore several image tiles can be created out of one image (or even have to be, depending on the available GPU RAM), as long as an image tile still covers the representative microstructural scale [25]. Ultimately, controlling and understanding material science aspects provides the biggest leverage for a successful implementation of ML-based microstructural analysis.

2.3. Ground Truth Assignment in a Holistic Approach

For a holistic ML approach in microstructure analysis, the assignment of the ground truth is an essential step. Simultaneously, complex microstructures pose a problem of causing disagreements between domain experts, giving rise to the question of how an objective ground truth can be obtained. Some examples of ground truths are class definition and the assignment of images to said classes, as well as image region annotation for segmentation, all fairly difficult for a human domain expert relying solely on visual appearance. One possible strategy is the performance of round robin tests, consisting of having a group of several experts individually judge the same images, followed by a majority voting for the image class [31,32]. While this reduces ambiguities to an extent, round robin tests are still

limited by uncertainties caused by excessive disagreement between domain experts, such as regarding bainitic structures in steel [14,23].

Another, more objective, approach for ground truth assignment in micrographs is correlative microscopy, consisting of a combination of different, usually scale-bridging, microscopy methods that complement each other to cancel out disadvantages and limits. For complex microstructures, one single characterization method often is not sufficient to record all microstructural features relevant for ML, justifying the relevance of a correlative approach. Generally, the increased expenditure of correlative microscopy is envisaged for the creation of the ML training dataset only, for creating references, with the goal being the reduction of the serial evaluation to the simplest microscopy method.

A prominent example of correlative microscopy is the combination of light optical microscopy (LOM) and/or scanning electron microscopy (SEM) images capturing the visual appearance of the microstructure with electron backscatter diffraction (EBSD) maps, the latter providing complementary structural information like misorientation parameters, grain and phase boundaries, etc. [25,30,33,34]. As the initial measured crystallographic phases and orientations from EBSD are not based on visual appearance to a human expert eye, they can be regarded as objective measurement data and thereby as an ideal complementary source of information.

To ensure that the exact same sample location is imaged with the different methods, a region of interest (ROI) can be marked, for example with hardness indentations, or microscopes with shuttle systems can be used. The EBSD measurement is carried out first, then the sample is etched, and the ROI is imaged in LOM and SEM. Image registration must then be carried out so that the images taken with the different methods are properly aligned and can be superimposed. Feature extraction using scale-invariant feature transformation (SIFT [35]) and feature alignment using the algorithm bUnwarpJ [36] have proven to be reliable approaches for image registration. Further details on the experimental procedure, data preparation and registration can be found in [33,37].

On the one hand, the comprehensive, scale-bridging description of the microstructure based on the correlative data enables the most accurate and objective ground truth assignment possible and is the fundament for implementing ML-based analysis of complex microstructures. On the other hand, it can be systematically investigated which method in the correlative approach can achieve the highest accuracy for the given task or which method is sufficient or necessary for which level of detail of the analysis. In general, the goal is always to perform the additional experimental effort of correlative characterization only once when generating the training data, thereby creating understanding and references, and reducing the series application, i.e., the inference of the ML model, to the simplest method (here: LOM). In the following application examples, the way in which correlative microscopy was used to assign the ground truth is explained in each case.

3. Overview of ML Applications in Microstructure Analysis

The first works on ML in microstructure analysis date back about 15 years and deal, among other things, with the classification of different graphite morphologies of cast iron [38,39]. Other milestones in terms of different microstructure representations or the use of DL that have made a decisive contribution to progress and further development of ML in microstructure analysis include works from DeCost et al. [40–42], Chowdhury et al. [43], Gola et al. [44,45] or Azimi et al. [1]. In the meantime, ML methods have been applied to a wide range of microstructure analysis tasks, and it is difficult to summarize them concisely. Without claiming to be comprehensive, it can be stated that mainly metallic materials are dealt with, and there is still a clear focus on steel microstructures [1,25,30,46,47]. However, there are also case studies on non-ferrous metals such as copper [48], titanium [49–51] or magnetic compounds [52]. Microstructures of non-metallic materials are only dealt with occasionally [53]. In the broader context of microstructural analysis, the evaluation of fracture surfaces (both macroscopic and microscopic, [54,55]) or surface defects [56,57] can

also be mentioned as areas of application in which ML methods have proven their potential and will continue to establish themselves.

Due to the large number of publications and the dynamic field of research, it is difficult to maintain an overview and summarize the methods applied. Nevertheless, a certain consensus on the employed methods can be observed.

While there are also some compelling approaches to microstructure segmentation using unsupervised learning [58–62], supervised learning is used for the most part. This makes sense, because complex microstructures in particular require the algorithm to learn more complex concepts that can no longer be distinguished by means of unsupervised learning. Earlier work predominantly used conventional ML, i.e., manual extraction or engineering of features, in combination with classic ML algorithms (e.g., decision trees, random forest, support vector machine). The use of image texture parameters as features for ML is widespread, e.g., Haralick textural features or local binary pattern [45,63–67]. Other features that are used for classification are, for instance, morphological parameters [45], or so-called bags of visual words [40]. All these approaches to feature extractions are primarily used to classify entire images, image sections, or individual objects from an image. In the context of this conventional ML, so-called trainable segmentations are used for segmentation. Various filters (e.g., noise reduction, edge detection, texture filter and membrane detection) are used to extract feature vectors per pixel from the microscope image, which are then classified using an ML algorithm. A prominent example is the trainable Weka segmentation of the open-source image processing tool Fiji/ImageJ (release 2.15.1) [68], which opens up a wide range of applications [24,69]. It works well if the microstructure constituents can be distinguished on the basis of colors or grey values, edges or simple textures. It reaches its limits with more complex segmentation tasks, e.g., when microstructure constituents consist of several textures or differ purely in their shape.

More recent work predominantly uses deep learning. For classification, the typical convolutional neural network (CNN) architectures like VGG, Inception and Xception; densely connected neural networks (DenseNet); and deep residual networks (ResNet) are used. For semantic segmentation, the popular and widespread U-Net architecture is used, usually with one of the aforementioned CNN as the backbone of the encoder being responsible for feature extraction as well as condensing the visual information into a representative vector, the so-called bottleneck of the U-Net. Since large amounts of data are required to train a CNN and we are usually operating in a low-data regime in MSE, transfer learning is predominantly used, typically with the ImageNet data source. A model trained on a dataset as huge as ImageNet has learned a good representation of low-level features such as corners, edges, illumination or shapes, and these features can be used collectively to enable knowledge transfer from source to target domain (here: microstructure analysis) for which few labeled data exist [70,71]. This type of transfer learning using a non-domain database has proven successful in practice, but it is still being discussed whether domain-specific pre-training would be more appropriate. Some studies show that with a sufficient amount of data, the same results can be achieved with a randomly initialized network as with ImageNet pre-training [25], or that ImageNet pre-training works better than random initialization if there are too few training data [25]. Other studies show that a multi-stage pre-training to bridge the domain gap can bring a slight improvement [72]. Stuckner et al. [73] in turn carried out domain-specific pre-training on a dedicated microscope dataset. This showed that domain-specific pre-training is better than ImageNet pre-training in a very-low-data regime, while the results are comparable with sufficient training data. The amount of data that can be considered sufficient in this context cannot be generalized but depends much more on the complexity of the individual problem to be solved.

To summarize the application examples, classification and semantic segmentation are mainly used for microstructural analysis tasks. Object detection and instance segmentation are only used occasionally [16].

A general recommendation for the use of conventional ML or DL cannot be given. Conventional ML can also be used with smaller amounts of data, as tuning the hyperparameters is easier and fewer computing resources are required. DL, on the other hand, tends to yield better results than ML for larger data quantities, can learn feature extraction independently and offers more options for image processing tasks. However, there are also so-called hybrid approaches that have become popular over the past several years [41,60,74]. Here, a pre-trained CNN (usually on ImageNet) serves as a feature extractor, and this feature vector is then classified using conventional ML algorithms (e.g., support vector machine, random forest).

In principle, it can be stated that ML/DL approaches that have already proven and established themselves in computer science and in other disciplines are predominantly used for microstructural analysis. New, “cutting-edge” methods from computer science tend to be used only sporadically. Some examples include the application of semi-supervised ML [75,76], super-resolution approaches [77,78], the use of generative models [79–82], or the incorporation of Meta’s Segment Anything model [83] into microstructure segmentation workflows [84]. However, current trends in computer science, e.g., neuro-symbolic AI or transformer architectures, give us reason to believe that a lot more can be expected here in the future.

The following examples for ML use in microstructure analysis concern complex steel microstructures. A summary of these examples can be found in the Supplementary Materials (e.g., used machine learning approach, ground truth, dataset size). Due to its excellent property combinations and ability to specifically adjust tailor-made microstructures, steel is still one of the world’s most important engineering and construction materials. Although the many commercial steel grades differ partially in their chemical composition, the main difference is in their microstructure. In turn, the microstructure largely determines the mechanical properties. These fine-tunable microstructures are becoming increasingly fine and complex, therefore requiring a deep understanding which, in turn, calls for reliable analysis methods. Therefore, steel is an ideal material for ML case studies.

3.1. Phase Analysis of Two-Phase Steels

3.1.1. Initial Advances in DL Applications for Microstructural Classification in Steel

In 2018, among the earliest realizations of DL techniques in steel microstructure analysis, Azimi et al. [1] proposed segmentation of steel micrographs for quality appraisal, showing substantial gains over the previous state-of-the-art methods.

Two-phase steels are examined, their microstructure consisting of a ferritic matrix as a first phase, considered “background”, and a second “foreground” phase in the form of pearlitic, martensitic or bainitic objects. Microstructures were analyzed in a correlative approach, acquiring micrographs from the exact same sample positions in LOM and SEM, based on Britz et al. [37]. The ground truth was provided in the form of pixel-wise annotations. To avoid manually outlining the second-phase objects in the SEM image, the correlative LOM image could be segmented via thresholding into the foreground (all second-phase objects) vs. the background (ferritic matrix), based on an appropriate etching method. This binary mask provides location information regarding the objects and can also be used to further assign different classes. This ground truth for each second-phase object, performed by materials science experts, was assigned per sample respective micrograph: every second-phase object in one micrograph belonged to the same class. No bainitic subclasses were considered. Instead, samples that were neither pearlite nor martensite were assigned to bainite. The correlative data could be used to investigate which classification accuracies were possible with LOM and with SEM. In total, five classes were considered: the background class ferrite and the four foreground classes pearlite (333 objects), bainite (345 objects), martensite (1253 objects), and tempered martensite (274 objects).

DL was applied in the form of a semantic segmentation, i.e., a pixel-wise classification. The fully convolutional neural network (FCNN) employed was based on a work by Long et al. [85], who proposed an approach similar to that of the VGG16 architecture.

Data augmentation in form of rotations and a variation in the stride parameters were used to counteract the initially unbalanced dataset. To achieve a classification of the entire second-phase object rather than its pixels, a max-voting scheme was afterwards applied to each object, assigning it to the class of the majority of the pixels.

For SEM images, in terms of the number of correctly classified second-phase objects, an outstanding classification accuracy of 94% was achieved, which significantly outperformed previous state-of-the-art methods. In terms of pixel accuracy and mean intersection over union (mIoU), values of 93.9% and 67.9% were achieved, respectively. An excerpt of the results can be seen in Figure 2. For LOM images, the object classification accuracy amounted to only 70.1%, which is, considering the complexity of microstructures, still a good result, but also clearly shows the limitations of LOM in resolving fine, similar microstructures like bainite, martensite and tempered martensite. Overall, the proof of concept was highly successful, especially considering the novelty of DL technologies in MSE at the time.

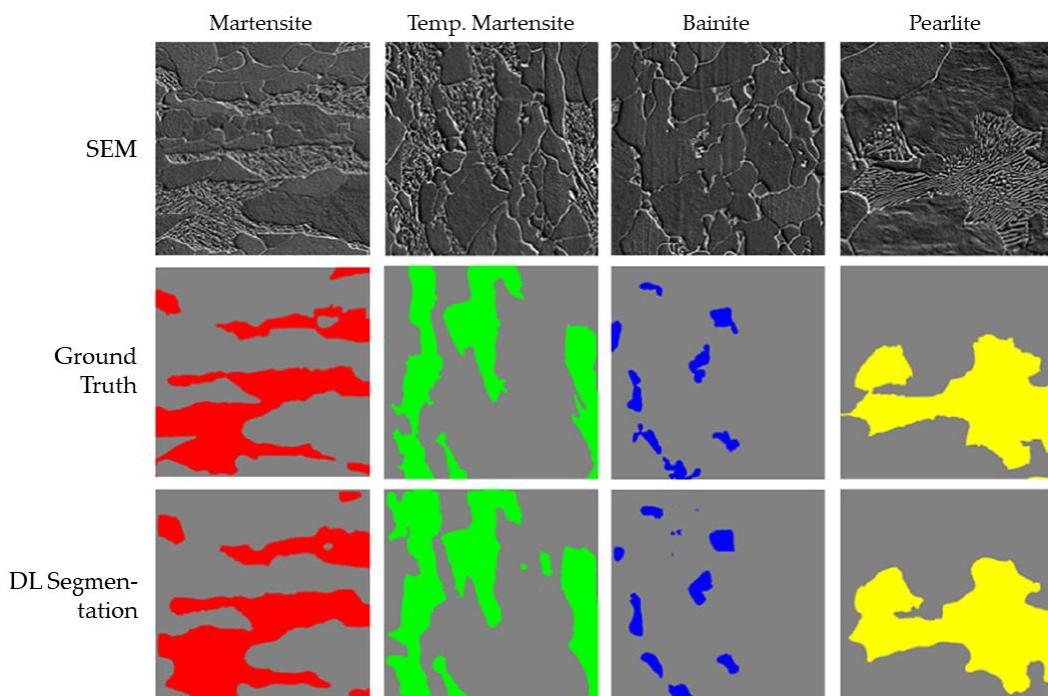


Figure 2. Examples of DL segmentation results for the four types of carbon-rich second phases (red: martensite, green: tempered martensite, blue: bainite, yellow: pearlite), modified according to [1]. Reprinted from [1].

3.1.2. ML-Based Classification of Bainitic Subclasses in SEM Micrographs

In continuation of the previous work on two-phase steels, Müller et al. [47] considered bainitic subclasses. The new classes required new annotations, which was the greatest difficulty in this ML implementation. In the field of steel, bainite has a special position due to its fine and complex structures, and since the large number of existing classification schemes also make a standardized description and assessment of bainitic structures difficult. In order to label the bainitic microstructures present in the samples, the classification scheme proposed by Zajac et al. [86] was chosen because it is the most convenient to use in common parlance and best fits the present bainitic structures. In total, seven classes were considered (see Figure 3): pearlite, martensite and degenerate pearlite, debris of cementite, incomplete transformation product, upper bainite and lower bainite as five bainitic subclasses, which can all be present simultaneously in one micrograph.

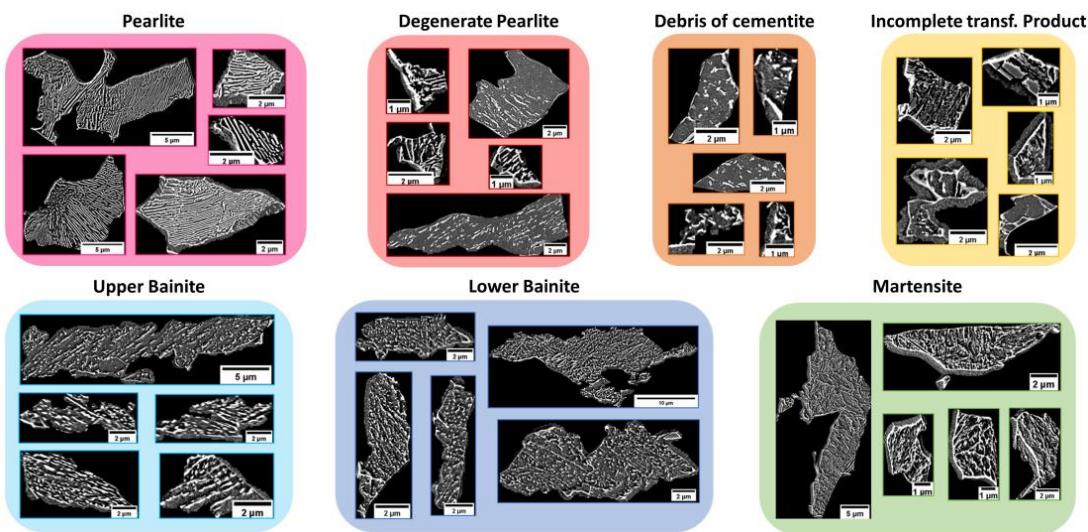


Figure 3. Seven microstructure classes considered for classification in [47], modified according to [47]. Reprinted from [47].

In contrast to the previous work of Azimi et al. [1], an object-wise ML classification based on initial works by Gola et al. [45] was used instead of a semantic segmentation. The pixel-by-pixel annotations required for semantic segmentation were in practice hardly realizable due to the required time but also due to uncertainty in manually marking and outlining regions and borders of these complex microstructures. By taking an object-by-object approach, however, objects that could not be clearly assessed could be filtered out before the ML model training, so that the model was only trained with unambiguous objects. In a later series application of the ML model, some uncertainty may remain in the classification of these unclear objects, but the assessment is always identical. With semantic segmentation, which uses the entire micrograph as input and not just individual objects, these uncertain objects would still be part of the training data. In combination with unbalanced classes, this represented too great an obstacle.

In the object-wise approach, first, the entire carbon-rich second phase was segmented. This segmentation was used as a binary mask to remove the ferritic matrix background from the micrograph and extract each second-phase object individually. For each individual second-phase object, image textural features, namely Haralick parameters and local binary patterns, as well as morphological characteristics of the substructure inside the objects, were extracted. These features could then be used for ML classification.

In the face of the complexity of bainitic structures and the disagreement between experts in labeling them, only relying on the visual appearance of the microstructures to the expert eye can easily introduce a subjective and non-reproducible component into the ML model. In this work, round robin tests were performed first to determine the best possible consensus within a group of experts. Furthermore, reference specimens were created, and correlative characterization was employed, combining SEM micrographs with EBSD maps showcasing misorientations and types of grain boundaries to increase the understanding of the different microstructural classes and assign the ground truth as objectively and reproducibly as possible. The final dataset consisted of almost 4000 annotated objects, although with imbalanced classes (Table 1).

Table 1. Summary of class distribution in final annotated dataset in [47]. Reprinted from [47].

Class	Pearlite (P)	Degenerate Pearlite (DP)	Debris of Cementite (DC)	Incomplete Transformation Product (ITP)	Upper Bainite (UB)	Lower Bainite (LB)	Martensite (M)	Total
No. of objects	772	321	964	856	212	228	550	3903

For the actual ML classification, the number of features was reduced by removing correlated features, and the data was standardized in order for all features to have the same data range, and split into an 80–20% train–test split. The classification was implemented in the MATLAB Classification Learner app (Version R2021a) by a support vector machine (SVM). Against the unbalanced classes, data augmentation at the image level could not be applied in a meaningful way, since the feature extraction process is invariant to typical augmentations such as flipping or rotating. However, approaches against imbalanced data at dataset level such as under-sampling, over-sampling or synthetic minority oversampling technique (SMOTE) did not show any significant effect.

Ultimately, the SVM performed well on the original, imbalanced dataset, and after feature selection (down to 40 features) and hyperparameter optimization, a classification accuracy of 82.9% was achieved (see confusion matrix, Table 2). Considering the complexity of the task and classes and the lack of consensus within the group of experts regarding bainitic microstructure characterization, this could be considered superhuman performance.

Table 2. Confusion matrix of the best classification model in [47]. Reprinted from [47].

Accuracy		82.9%				F1 Score		82.4%	
		Predicted							
		P	DP	DC	ITP	UB	LB	M	Precision [%]
True	P	148	3	1	2	1	0	0	96.1
	DP	7	33	19	5	0	0	0	50.8
	DC	0	11	156	16	3	3	3	81.3
	ITP	0	5	26	129	0	1	10	75.4
	UB	0	0	3	0	39	1	0	90.7
	LB	0	0	3	1	1	40	0	88.9
	M	0	1	0	7	0	0	102	92.7
Recall [%]		95.5	62.3	75.0	80.6	88.6	88.9	88.7	
F1 Score [%]		95.8	55.9	78.0	77.9	89.7	88.9	90.7	

Most misclassifications occurred in the classes degenerate pearlite, debris of cementite and incomplete transformation products, as those have weaker class boundaries compared to other classes. In fact, the uncertainty is somewhat inherent, as there are no explicitly defined class boundaries, and there is bias stemming from the choice of the classification scheme and the classes. With regard to a series application, it was assumed that in this special case with highly complex microstructures and classes and a sophisticated feature extraction process, the model only works with a certain data quality, and therefore certain procedures and parameters must be specified and adhered to during data generation (metallography, image acquisition).

3.2. Segmentation of Multi-Phase Steels

As opposed to the above-seen two-phase steels, multi-phase, or complex-phase, steels present more than two major components, usually containing polygonal ferrite, bainitic ferrite and finely dispersed carbon-rich second phases (e.g., cementite, martensite, austenite). Durmaz et al. [25] proposed a DL segmentation of lath-shaped bainite regions against a background of polygonal and irregular ferrite containing a carbon-rich second phase.

A particular emphasis was placed on the sample contrasting and the imaging with the goal of providing high-quality, low-variance data in order to train a DL model with few training data. Similarly to previous examples, correlative imaging was carried out using LOM, SEM and EBSD. EBSD provided complementary information on the grains, grain boundaries and misorientation parameters, which were decisive for a correct annotation of the lath-shaped bainite regions, overcoming the limitations of decision-making based solely on visual appearance. A relatively small dataset, consisting of 51 LOM and 36 SEM

images, from which several patches per image were extracted as training data (see Table 3), sufficed for the training of the ML model, taking into account high-quality micrographs with low variance among each other.

Table 3. Summary of the final annotated dataset used in [25]. Reprinted from [25].

Imaging Modality	No. of Micrographs	Resolution of Micrographs	No. of Extracted Patches	Resolution of Patches
LOM	51	1024 × 1024 px	754	256 × 256 px
SEM	36	2048 × 1433 px	413	512 × 512 px

Different DL models were used, one based on Vanilla U-Net architecture and trained from scratch, and the other consisting of a U-Net with a VGG16 backbone pre-trained on the ImageNet natural scene image dataset. All segmentation models yielded performances comparable to expert predictions, speaking in favor of the general robustness of the different architectures and training strategies. With the reproducibility of the dataset, lower amounts of data were required for training. The segmentation of the lath-shaped bainite regions was successful (Figure 4), and errors mainly occurred either at boundaries between regions or within regions where insecurities within the group of experts could not be ruled out despite the complementary EBSD information. However, when regarding the different phase fractions, the variance of the ML-based analysis (of 1–2%) was lower than the variance of the manual analysis carried out by experts.

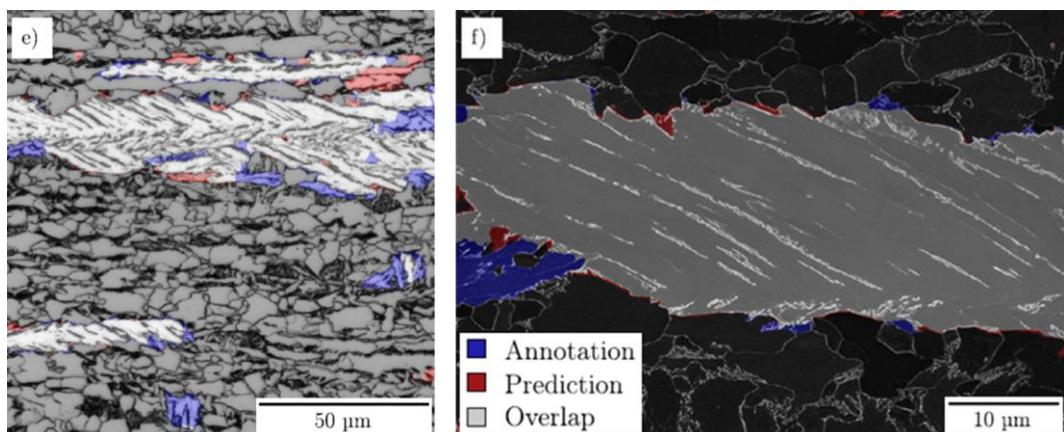


Figure 4. Light optical and scanning electron micrographs superimposed with lath–bainite predictions of the best segmentation models and annotated regions showing the comparison between model prediction (red) and manual expert annotation (blue), modified according to [25]. Reprinted from [25].

It should be noted that this dataset was tailored to a very specific application. Such applications are conceivable, for example, in daily quality control, where the same type of microstructure is always examined, with fixed workflows for sample preparation, sample contrasting and image acquisition. DL segmentation can then be implemented even with few training data.

At the same time, however, a weak generalization of this model is to be expected; i.e., in the case of deviations from the ideal state of the microstructure images, a lower model performance can be anticipated. In a follow-up work [72], an additional transferability study was performed in which a model trained with low-variance data was transferred to an alternate data domain, i.e., micrographs of the same material but with an alternative etching method. As a result, it was concluded that in order to carry out a successful domain transfer, higher variances are required within the source dataset, which can be achieved by employing image transformation methods more elaborate than basic brightness and contrast variation [25]. Additionally, the unsupervised domain adaptation was identified

as a potential useful tool to either increase the robustness of the present model without additional annotated data or to reduce the workload of manual annotations on the way to a higher-variance dataset.

3.3. Phase Analysis of Quenched and Tempered Steels

For quenched and tempered steels, the complexity of the analysis is further increased, partly because there are no clearly separable foreground and background in the micrographs. In the low- to medium-carbon steels investigated by Bachmann et al. [30], the microstructure constituents martensite, tempered martensite, lower bainite and upper bainite, which can all be present simultaneously in one micrograph, should be distinguished.

Despite the use of correlative microscopy (LOM and SEM combined with EBSD), pixel-by-pixel annotations for semantic segmentation were hardly feasible, as the microstructures contained many regions where a clear and doubt-free assignment to a class is not possible, and in particular, the identification of the boundaries between the different structural components was particularly difficult, as they frequently merged into one another due to their formation process. Instead, Bachmann et al. proposed a patch-wise classification that was implemented using a sliding window approach, reducing the complex problem of semantic segmentation to a simple classification task of the individual patches. The assignment of the ground truth for such individual patches within a micrograph was possible with certainty, thanks to the high resolution of the SEM and the correlative EBSD information (image quality, misorientations, grain boundary types). In this way, reference patches, representative of unambiguous microstructural constituents, were extracted, and two datasets, optimized for LOM and SEM respectively, were generated. The datasets contained over 6500 individual patches for SEM and over 2200 patches in the case of the LOM-optimized dataset.

For both imaging modalities, three separate CNN models were trained on these datasets, based on three different backbones, namely Xception, ResNet50 and DenseNet201. The basic idea behind these three models was to take into account the complexity of the present microstructures by later combining the three models in a majority voting scheme and thus increasing the confidence in the final prediction. For training, class weights were considered to counteract the class imbalances, and data augmentation and a categorical cross-entropy loss were used. All three models yielded similar performances, with accuracies up to 89% for the LOM and 94% for the high-resolution SEM dataset, reaching close to a superhuman performance (see Figure 5). It is particularly noteworthy that the ML models could also capture subtle differences, especially between lower bainite and tempered martensite, which even trained experts can often only anticipate.

With this patch-wise classification, entire micrographs can be analyzed using a sliding window approach. A window slides over the micrograph, and at every window position the extracted patch is classified with the max voting of the three trained models. By combining different step sizes of the sliding window, a finer classification resolution can be realized so that the final classification result almost approaches the pixel-level detail of a semantic segmentation. Low confidence predictions, e.g., ambiguous predictions of the different models, can also be assigned to a shared class by applying a certain confidence threshold, representing the uncertainty of specific regions in the micrograph. The pixelated appearance of the resulting images was then smoothed using a median filter. Figure 6 shows an example of the application to an entire, unseen micrograph for both imaging modalities.

For this example of quenched and tempered steels as well as for the previous examples of two-phase and multi-phase steels, the use of correlative microscopy with the additional EBSD information was crucial in order to obtain annotations for ML and thus to be able to implement ML in the first place. The EBSD analysis used was still comparatively simple, i.e., examining parameters of pattern quality, misorientation parameters, or different types of grain boundaries, and overlaying them with LOM and SEM images. However, there is still a lot of untapped potential in the analysis of EBSD data, especially in combination with ML. With regard to steel and especially martensite and bainite, it is possible to delve deeper

into the formation process and the crystallographic nature of the microstructures, analyzing for example misorientation angle distributions, variant pairings or Bain groups [87,88]. Variant pairings have already been used in some cases to determine bainite/martensite proportions globally for a specific measurement region [89,90]. The correlative approach described here could be taken further in order to automatically generate pixel-by-pixel annotations for segmenting the LOM/SEM image based on the correlative EBSD data, possibly also in combination with ML. In addition, the EBSD parameters could be used to verify existing steel classification schemes or even define alternative schemes using unsupervised learning.

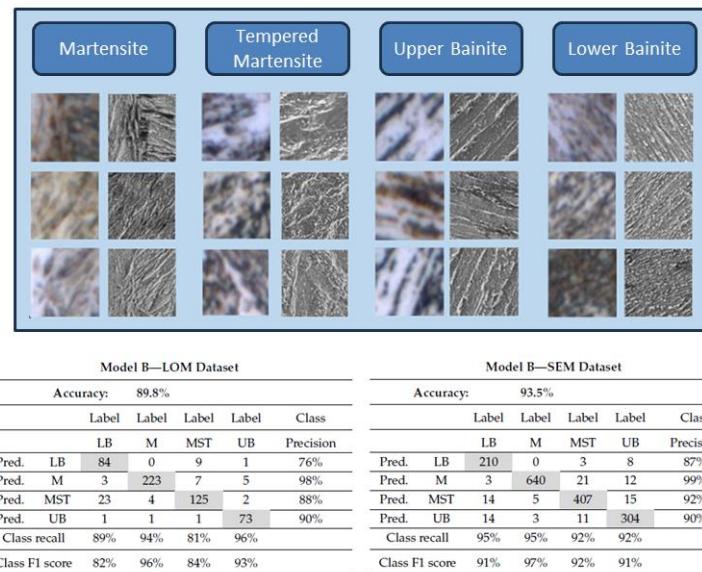


Figure 5. Illustration of the considered classes for microstructure classification (corresponding light optical microscope (LOM) and scanning electron microscope (SEM) images) as well as confusion matrixes for LOM model and SEM model, modified according to [30]. Reprinted from [30].

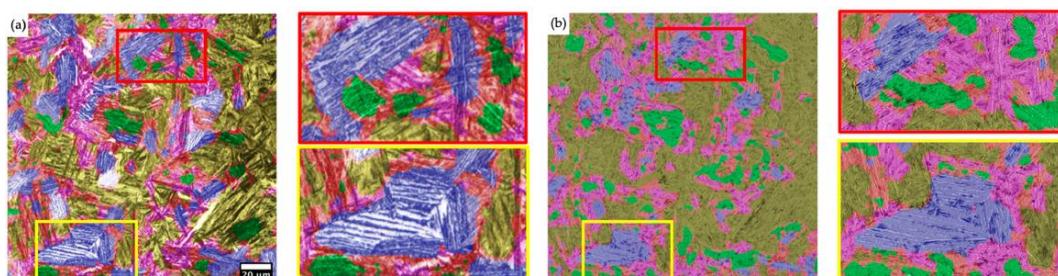


Figure 6. Quantification results from patch-wise classification using sliding window technique, with colored overlay based on LOM (a) and SEM (b), including magnified regions. Correlative microscopy makes it possible to compare the results from LOM and SEM models and thus to identify the discrepancies between the predictions based on the different imaging techniques. Colors correspond to the following classes: green-LB, yellow-M, purple-MST, blue-UB, and red-uncertain. Modified according to [30]. Reprinted from [30].

3.4. Measurement of Prior Austenite Grain Size

Hot-forming or any type of temperature treatment during steel production generally occurs in the high-temperature phase, that is, austenite. The development of both the austenite grain size during these processes and the resulting austenite grain size are highly significant for the final steel product because they influence the type and properties of the final microstructure, e.g., phase transformation behavior such as bainite or martensite formation or final grain size [46]. Additionally, knowledge about austenite

grain evolution is important for understanding and optimizing associated process like thermomechanical controlled processing or alloy designs [91]. Considering the microstructure of the final steel product, since austenite transforms into different room temperature phases upon cooling, the prior austenite grain (PAG) size cannot readily be measured but requires reconstruction [46,92].

There are various approaches for this [46]: (i) indirect measurement by laser-induced ultrasound, (ii) reconstruction from crystallographic orientation data, (iii) the McQuaid–Ehn method, i.e., highlighting PAG by preferred oxidation or precipitation, which can be visualized through etching, (iv) thermal etching and (v) chemical etching, each presenting its own process-specific disadvantages and a certain degree of uncertainty in the PAG determination.

Metallographic determination with picric acid-based etchants, primarily Bechet–Beaujard etching, is still the most widely used method. However, it is well known that these etchings can be difficult to reproduce, vary depending on the chemical composition of the steel being evaluated and are particularly challenging for low phosphorus and/or low carbon levels [46]. In addition, multiple etching steps combined with back polishing may be required to achieve sufficient contrast. Ultimately, final microstructural images frequently show an inhomogeneous contrasting of PAG, with an unwanted contrasting of the substructure inside the PAG, which is still sufficient for a trained expert to determine a comparative grain size by means of comparison charts, but not good enough to reliably segment the grain boundaries and determine a grain size distribution using conventional approaches.

3.4.1. Determination of Prior Austenite Grains after Picric Acid-Based Etching

For a reliable determination of prior austenite grain size (PAGS) and its distribution, Laub et al. [92] first proposed a modified picric acid-based etching, the micrographs of which were then segmented using semantic segmentation with deep learning. Steel types considered were low carbon (0.04 wt. %) that were micro-alloyed with different levels of niobium.

Due to the aforementioned difficulties of metallographic etching (incomplete contrasting of grain boundaries on the one hand, contrasting of the substructure inside the grains on the other hand), the annotations required for training the DL model posed the greatest challenge. Although an expert could anticipate the path of grain boundaries that are not fully contrasted, a significant subjective component would remain in the annotations. Therefore, a correlative microscopy approach in which light microscopy was combined with EBSD was applied for a subset of the dataset. With the EBSD data, a crystallographic reconstruction of the PAGS was performed, based on a workflow proposed by [93,94]. The image registration was performed using bUnwarpJ in the open-source image processing toolbox Fiji and assisted by manual feature selection. These EBSD reconstructions form the basis for annotations. Due to some artefacts of the EBSD reconstruction, some manual corrections are usually necessary. In the end, the highest quality of annotations could be achieved by a combination of EBSD reconstruction and manual outlining. Based on the experience and references gained with the correlative data, non-correlative data, i.e., LOM-only images, could also be manually annotated by the expert in the most accurate, objective and reproducible way. In addition, the EBSD reconstruction could later be used to validate the mean grain size and grain size distribution determined by DL segmentation. The dataset comprised micrographs of 30 samples, of which approximately 20% were in the form of correlative datasets.

For model training, the images were sliced into patches (8000 in total), and then divided into a 70–20–10% train–test–validation split and subjected to data augmentation. The ML model was implemented using the Keras library, and several backbones were tested—namely DenseNet, VGG16, VGG19, ResNet and Inception—of which DenseNet delivered the most promising results and was subsequently chosen for the final model. Jaccard loss and intersection over union (IoU) were chosen as metrics. Furthermore, the model's performance was validated on previously unseen entire micrographs. The model

yielded an accuracy (IoU) of 72–73%, and inference on previously unseen images was successful (Figure 7). Through post-processing, e.g., area-opening plus watershed, the segmentation result can be further improved.

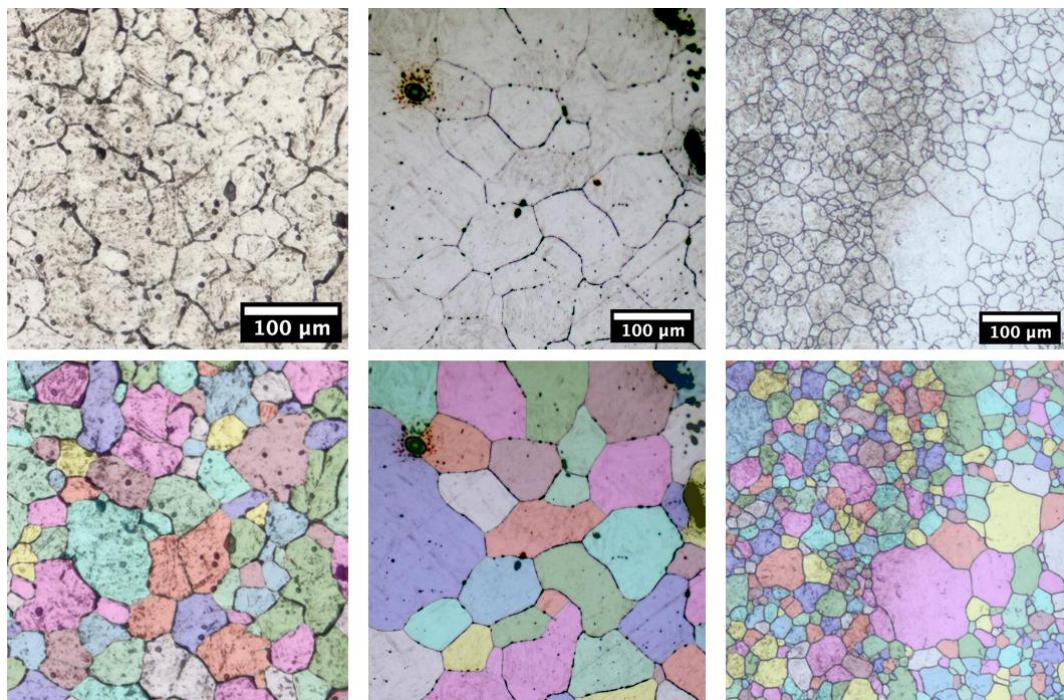


Figure 7. Picric-based etched micrographs containing different variations and etching artefacts as well as less pronounced PAG boundaries (**upper row**) and the corresponding results of the segmentation pipeline as overlays with the input image (**lower row**). Previously unpublished examples from the work of [92]. Adapted from Ref. [92].

Overall, the model bore decisive advantages over previous standard PAG reconstructions as it allowed a more detailed result and was efficient, reproducible and flexible to higher variances, improving its generalizability and robustness.

3.4.2. Determination of Prior Austenite Grains after Nital Etching

The combination of picric acid-based etching with DL segmentation allows for quantification of PAGS with unprecedented quality and robustness. Although established as a standard etchant for PAG analysis, picric acid-based etching still represents an additional effort compared to other steel etchants or is no longer used at all in some laboratories due to safety concerns [95]. Therefore, Bachmann et al. [46] investigated in a case study to what accuracy the PAGS can be determined with a simple Nital etching. Nital, a mixture of ethanol and nitric acid, is one of the standard etchants for analyzing the microstructures of non- and low-alloyed steels. It is very popular because of its simplicity and ease of use. Nital was not originally intended for PAG contrasting, but it still can reveal some information about it: PAG can either be seen directly due to topography differences or recognized indirectly based on orientations of martensitic and bainitic laths or sub-units. However, since all other hierarchical microstructural features are also contrasted, PAG segmentation in Nital-etched micrographs can also only be achieved by DL.

Compared to manual annotations of the PAG in LOM micrographs after picric acid etching, the annotations after Nital etching are subject to even greater uncertainty or are hardly possible at all. To perform annotations, and thus to enable a PAG determination in Nital-etched LOM micrographs in the first place, a correlative approach [33] was also used here. This time, LOM was combined with EBSD as well as SEM. As in the previous example, the EBSD reconstructions formed the basis of the annotations (Figure 8). Manual corrections

were also made here based on the grain boundaries visible in LOM and SEM. Due to the etching attack on all hierarchical structures by the Nital etching, SEM imaging was also required here, as its higher resolution enables a better topography contrast, by which some grain boundaries that are not visible in the LOM can still be recognized. Experimental procedure for data acquisition, EBSD reconstruction and registration were carried out as in the previous example. The final dataset was comprised of 13 samples with correlative data (LOM and SEM images for all 13 samples, EBSD reconstructions for 8 samples) and corresponding annotations, resulting in 1420 individual patches for ML model training after tiling.

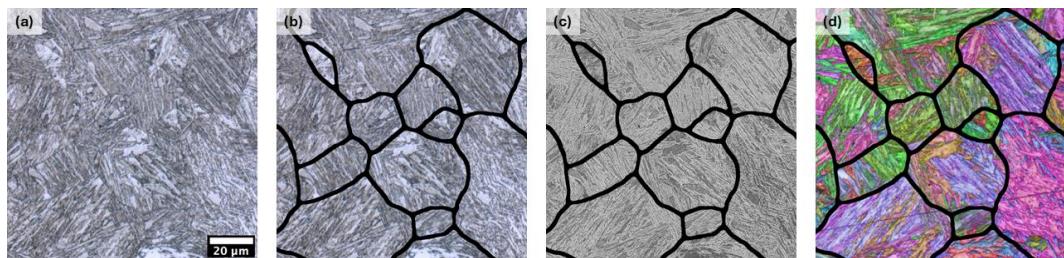


Figure 8. Correlative LOM (a,b), SEM (c) and EBSD images (image quality overlayed with inverse pole figure, (d)) of the identical sample region. Images b-d are overlayed with the final PAG annotations (black outlines), based on crystallographic reconstruction from EBSD and manual corrections. Previously unpublished examples from the work of [46]. Adapted from Ref. [46].

As pre-processing, the images are split into patches after downscaling them to achieve a higher density of features per patch. Data augmentation was employed to counteract data scarcity. The ML model consisted of a U-Net combined with the segmentation models package in Keras, using a DenseNet backbone (as the best from several backbones tested) pre-trained on ImageNet. Intersection over Union was used as accuracy metric, combined with a loss function of a weighted dice loss plus Jaccard loss to counteract the class imbalance (PAG boundaries vs. rest of the microstructure). The model yielded a 73% and 70% IoU for the training and the validation split, respectively, as well as corresponding respective F1 scores of 82% and 80% as a conservative estimation of the performance. Furthermore, the model was tested on unseen entire micrographs, with one example shown in Figure 9.



Figure 9. Original LOM image (left) with respective PAG determined by the DL workflow (middle) and the comparison (right) between segmentation result after postprocessing (blue) with the respective ground truth (green). Black grain boundaries represent the agreement between prediction and ground truth. Previously unpublished examples from the work of [46]. Adapted from Ref. [46].

Through post-processing, e.g., area-opening plus watershed, the segmentation map could be further improved. In addition to the established IoU metric, the model performance was assessed by comparing the mean grain size and grain size distribution from the post-processed DL segmentation map to the ground truth, i.e., mean grain size and

grain size distribution of the EBSD reconstruction. The grain size distributions showed a comparable pattern, with an average error (over three samples) of 9% for the number fraction mean grain size and 6.1% for the area fraction mean grain size, clearly showing the high quality of the DL model. It should be noted here that this is probably a systematic error towards a slight underestimation of the grain size, as some grain boundaries were erroneously reconstructed during post-processing by watershed. At this point, however, it should be noted that even the EBSD reconstruction does not represent the absolute truth and that every method for PAGS determination is associated with a certain degree of uncertainty, since austenite is no longer present in the final microstructure. The model also exhibited a notable robustness to etching artifacts, as illustrated by the example in Figure 10.

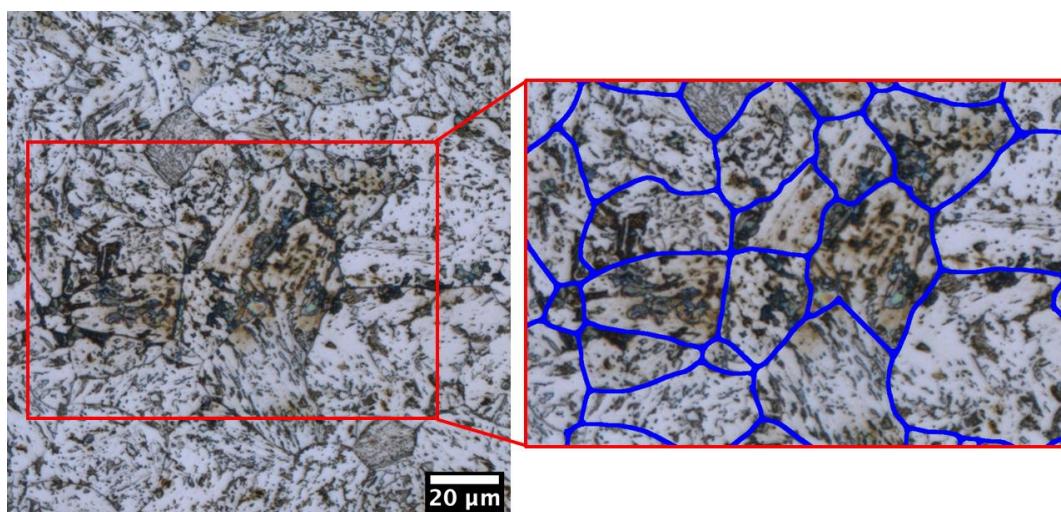


Figure 10. LOM image (left) with a blue overlay of the respective PAG determined by the DL workflow (right). Despite pronounced etching artifacts and stains, the model is able to detect the partially hidden PAG boundaries. Previously unpublished examples from the work of [46]. Adapted from Ref. [46].

The PAGS determination directly from Nital-etched LOM images is now the simplest and fastest method of metallographic PAGS measurement and is characterized by the simpler and more reproducible application of Nital etching compared to the commonly used picric acid-based etchings. Nevertheless, it must be noted that for the application of this model in a Nital-etched LOM image, a certain basic level of contrasting of PAG must be visible, and it still has to be investigated, depending on which chemical composition of the steel or manufacturing parameters this basic contrasting of PAG can be the case in Nital-etched images.

4. Application in Materials Development and Process Optimization

The combination of AI and ML technologies with domain expertise in materials science and engineering paves the way for new developments in both microstructure research and materials characterization. This allows the automatization of simple tasks on one hand, and, on the other, the analysis of complex microstructures where no satisfactory methods exist yet. It is important to note that, however, ML technologies are no panacea and are by no means supposed to replace the human expert, but more so to assist and relieve them. In fact, expert knowledge is necessary to implement any ML-based analysis in materials science. A holistic approach is required for a successful use of ML in the long run, comprising all steps needed to obtain micrographs, including approaches like correlative microscopy, and the assignment of the ground truth. This requires fundamental expertise in metallography and microstructure quantification. Furthermore, a deep understanding and control of materials science aspects have more leverage than ML parameter optimization.

Thus, ML-based segmentation and classification can form a solid base for an improved microstructure quantification that is automated, reliable, objective, and reproducible, paving the way towards process–microstructure–property correlations and microstructure-based materials development.

Using correlative microscopy allows to benchmark how exact a LOM-only quantification is, and ideally, provides a gain in knowledge that enables to reduce further analysis to a single—i.e., the simplest (here LOM)—characterization method. Even though LOM-based analysis might present some inaccuracies, ML allows us to further use LOM as a standard method in applications where, for example, a SEM examination is not possible due to time constraints or availability. Potential inaccuracies can be counterbalanced by the high speed at which large portions of data can be analyzed.

ML technologies also represent a step towards high-throughput microstructure analysis, which is automated, efficient, objective and reproducible—i.e., applicable to large amounts of data in a short time. This allows a rapid creation of a large database with which correlations can be established—essentially, showcasing the links between the microstructure and the processing history or further properties including, for instance, annealing processes [41], fracture energy [60], fatigue strength [96] or others. Based on ML-based microstructure analysis, ML methods can also be used to establish these process–microstructure–property links. Forward models for example predict the output from the current state (input data) in a system, essentially modelling forward dynamics, while inverse models aim to predict the input that leads to a desired output. When applied to process–microstructure–property links, a forward model can yield materials properties from descriptors, while an inverse model can be used to predict an optimal material or process from desired characteristics [97,98].

5. Challenges and Open Research Questions

ML has absolutely demonstrated its potential for a variety of tasks in microstructural analysis. With the help of this review, we hope not only to have provided an overview of possible applications, but also to have achieved a better understanding and sensitization with regard to ground truths, data quantities and data variances. These are points that are often not sufficiently discussed. Properly implemented, i.e., with a focus on objective ground truth and an understanding of the variances that occur, we can achieve robust, reliable ML models and use ML to open up completely new avenues in the quality, quantity and efficiency of microstructure analysis. However, just as with other applications in science, engineering and our everyday lives, the same applies to microstructure analysis: the quality of current and future AI systems can easily lure people into overestimating their capabilities, underestimating their weaknesses and limitations, and thus, using them in potentially problematic or even harmful ways. Therefore, we shall use ML models only within their intended limits and design spaces.

In the context of variances and robustness, it is interesting to investigate whether metadata, e.g., metallographic or microscopic metadata (e.g., etching, image acquisition conditions in the SEM) or manufacturing information, can improve ML classification and segmentation by explaining certain variances and eliminating the need for the ML model to learn them from additional training data. However, it must also be investigated whether the use of metadata can introduce a potential bias and how large the harmful effect of missing or incorrect metadata would be, e.g., with the help of adversarial examples. A first suggestion for the fusion of image data and microscopic metadata in a CNN has already been made by Stiefel et al. [99].

Since annotations continue to represent the bottleneck in the implementation of ML-based microstructure analysis, approaches that require less training data or simpler annotations or that can also use unlabeled data (e.g., weakly supervised learning, semi-supervised learning, unsupervised domain adaptation) are very interesting and relevant for future work [25,72,100,101]. In particular, a combination of such approaches with annotated data

from correlative microscopy, which involves a certain amount of additional experimental work, would be attractive.

With regard to the effort of manual annotations, synthetic data generation should also be mentioned. Its advantage is that the ground truth is usually also generated when the data are generated. In addition, data of rarely occurring classes or events can be created efficiently. However, there is still a lack of general understanding and established approaches as to which synthetic data of which complexity can be generated with which methods. Current work ([102,103] among others), which often uses data-driven methods such as generative models like GAN to create synthetic microstructures, is unsuitable because it also requires a minimum amount of data. Thus, they are not practically applicable to real-world use cases with little available data. Model- or rule-based approaches [104,105] or texture synthesis approaches, e.g., based on non-parametric example-based algorithms for image generation [106] seem more promising.

As the often-mentioned black box character of AI/ML systems is still sometimes cited as a reason not to use AI/ML [107], approaches from computer science to improve their trustworthiness, often-termed as “trustworthy AI” or “explainable AI” [108,109] are interesting. In the course of the continuous development and advancement of AI/ML approaches in informatics, trends such as neuro-symbolic AI are also to be observed and when or how they will find their way into microstructural analysis. A promising approach is the so-called Vision Transformer (ViT) [110], similar architectures for image processing inspired by the successful application of transformer models in natural language processing, which achieved better results in some benchmark datasets than the previously used CNN [110]. A competition in the further development of CNN and ViT architectures might be expected here [111].

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/met14050553/s1>. Supplementary materials include a table summarizing the presented application examples (used machine learning approach, ground truth, dataset size, data variance, design space for applying the machine learning model).

Author Contributions: Conceptualization, M.M., B.-I.B. and M.S.; methodology, M.M. and B.-I.B.; software, M.M. and B.-I.B.; validation, M.M. and B.-I.B.; formal analysis, M.M. and B.-I.B.; investigation, M.M. and B.-I.B.; resources, M.M. and B.-I.B.; data curation, M.M. and B.-I.B.; writing—original draft preparation, M.M., B.-I.B. and M.S.; writing—review and editing, M.M., B.-I.B. and M.S.; visualization, M.M. and B.-I.B.; supervision, D.B. and F.M.; project administration, D.B. and F.M.; funding acquisition, D.B. and F.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors thank steel manufacturer Aktien-Gesellschaft der Dillinger Hüttenwerke for the strategic collaboration in which the majority of the application examples presented were elaborated and implemented.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Azimi, S.M.; Britz, D.; Engstler, M.; Fritz, M.; Mücklich, F. Advanced steel microstructural classification by deep learning methods. *Sci. Rep.* **2018**, *8*, 2128. [[CrossRef](#)]
2. Britz, D.; Webel, J.; Gola, J.; Steimer, Y.; Muecklich, F. Practical Metallography of Low Carbon Steels—New approaches in Preparation, Imaging and Analysis of Microstructures. *Microsc. Microanal.* **2018**, *24*, 2226–2227. [[CrossRef](#)]
3. Khan, A.; Laghari, A.; Awan, S. Machine Learning in Computer Vision: A Review. *EAI Endorsed Trans. Scalable Inf. Syst.* **2018**, *8*, e4. [[CrossRef](#)]
4. Tehrani, A.M.; Oliynyk, A.O.; Parry, M.; Rizvi, Z.; Couper, S.; Lin, F.; Miyagi, L.; Sparks, T.D.; Brzoch, J. Machine Learning Directed Search for Ultraincompressible, Superhard Materials. *J. Am. Chem. Soc.* **2018**, *140*, 9844–9853. [[CrossRef](#)] [[PubMed](#)]

5. Liu, Y.; Niu, C.; Wang, Z.; Gan, Y.; Zhu, Y.; Sun, S.; Shen, T. Machine learning in materials genome initiative: A review. *J. Mater. Sci. Technol.* **2020**, *57*, 113–122. [CrossRef]
6. Szymanski, N.J.; Zeng, Y.; Huo, H.; Bartel, C.J.; Kim, H.; Ceder, G. Toward autonomous design and synthesis of novel inorganic materials. *Mater. Horiz.* **2021**, *8*, 2169–2198. [CrossRef]
7. Liu, Y.; Zhao, T.; Ju, W.; Shi, S. Materials discovery and design using machine learning. *J. Mater.* **2017**, *3*, 159–177. [CrossRef]
8. Raccuglia, P.; Elbert, K.C.; Adler, P.D.F.; Falk, C.; Wenny, M.B.; Mollo, A.; Zeller, M.; Friedler, S.A.; Schrier, J.; Norquist, A.J. Machine-learning-assisted materials discovery using failed experiments. *Nature* **2016**, *533*, 73–76. [CrossRef] [PubMed]
9. Cai, J.; Chu, X.; Xu, K.; Li, H.; Wei, J. Machine learning-driven new material discovery. *Nanoscale Adv.* **2020**, *2*, 3115–3130. [CrossRef]
10. Butler, K.T.; Davies, D.W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555. [CrossRef]
11. Choudhary, K.; DeCost, B.; Chen, C.; Jain, A.; Tavazza, F.; Cohn, R.; Park, C.W.; Choudhary, A.; Agrawal, A.; Billinge, S.J.L.; et al. Recent advances and applications of deep learning methods in materials science. *npj Comput. Mater.* **2022**, *8*, 59. [CrossRef]
12. Callister, W.D.; Rethwisch, D.G. *Fundamentals of Materials Science and Engineering: An Integrated Approach*; Wiley: Hoboken, NJ, USA, 2015; p. 964.
13. Molkeri, A.; Khatamsaz, D.; Couperthwaite, R.; James, J.; Arróyave, R.; Allaire, D.; Srivastava, A. On the importance of microstructure information in materials design: PSP vs. PP. *Acta Mater.* **2022**, *223*, 117471. [CrossRef]
14. Müller, M.; Britz, D.; Mücklich, F. Machine Learning for Microstructure Classification: How to Assign the Ground Truth in the Most Objective Way. *Am&P Tech. Artic.* **2021**, *179*, 16–21. [CrossRef]
15. Russel, S.; Norvig, P. *Artificial Intelligence: A Modern Approach*, 3rd ed.; Pearson Education, Inc.: Upper Saddle River, NJ, USA, 2010.
16. Holm, E.A.; Cohn, R.; Gao, N.; Kitahara, A.R.; Matson, T.P.; Lei, B.; Yarasi, S.R. Overview: Computer Vision and Machine Learning for Microstructural Characterization and Analysis. *Met. Mater. Trans. A* **2020**, *51*, 5985–5999. [CrossRef]
17. Park, H.; Öztürk, A. Machine Learning Approach on Steel Microstructure Classification. In Proceedings of the EKC 2019 Conference Proceedings: Science, Technology, and Humanity: Advancement and Sustainability, Vienna, Austria, 15–18 July 2019. [CrossRef]
18. Larmuseau, M.; Sluydts, M.; Theuwissen, K.; Duprez, L.; Dhaene, T.; Cottenier, S. Compact representations of microstructure images using triplet networks. *npj Comput. Mater.* **2020**, *6*, 156. [CrossRef]
19. Sculley, D.; Holt, G.; Golovin, D.; Davydov, E.; Phillips, T.; Ebner, D.; Chaudhary, V.; Young, M.; Crespo, J.F.; Dennison, D. Hidden technical debt in machine learning systems. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 2503–2511.
20. ImageNet. Available online: <https://www.image-net.org/update-mar-11-2021.php> (accessed on 26 March 2024).
21. Stanford Dogs Dataset. Available online: <https://www.kaggle.com/datasets/jessicali9530/stanford-dogs-dataset/> (accessed on 26 March 2024).
22. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 3213–3223. [CrossRef]
23. MAarnts, P.; Rijkenberg, R.A.; Twisk, F.A. *Microstructural Quantification of Multi-Phase Steels (Micro-Quant)*; European Comission: Luxembourg, 2011. [CrossRef]
24. Bulgarevich, D.S.; Tsukamoto, S.; Kasuya, T.; Demura, M.; Watanabe, M. Pattern recognition with machine learning on optical microscopy images of typical metallurgical microstructures. *Sci. Rep.* **2018**, *8*, 2078. [CrossRef]
25. Durmaz, A.R.; Müller, M.; Lei, B.; Thomas, A.; Britz, D.; Holm, E.A.; Eberl, C.; Mücklich, F.; Gumbsch, P. A deep learning approach for complex microstructure inference. *Nat. Commun.* **2021**, *12*, 6272. [CrossRef] [PubMed]
26. MetalDAM—DaSCI. Available online: <https://dasci.es/transferencia/open-data/metal-dam/> (accessed on 25 April 2024).
27. Han, B.; Wan, W.-H.; Sun, D.-D.; Dong, C.-C.; Zhao, L.; Wang, H.-Z. A deep learning-based method for segmentation and quantitative characterization of microstructures in weathering steel from sequential scanning electron microscope images. *J. Iron Steel Res. Int.* **2022**, *29*, 836–845. [CrossRef]
28. Muñoz-Rodenas, J.; García-Sevilla, F.; Miguel-Eguía, V.; Coello-Sobrino, J.; Martínez-Martínez, A. A Deep Learning Approach to Semantic Segmentation of Steel Microstructures. *Appl. Sci.* **2024**, *14*, 2297. [CrossRef]
29. Preußner, J.; Bellmer, M.J. *Mit dem Rasterelektronenmikroskop (REM) Erzeugte Aufnahmen von Bruchflächen von Metallen für eine Spätere Statistische Auswertung*; Fraunhofer-Gesellschaft FhG: Munich, Germany, 2018. [CrossRef]
30. Bachmann, B.-I.; Müller, M.; Britz, D.; Staudt, T.; Mücklich, F. Reproducible Quantification of the Microstructure of Complex Quenched and Tempered Steels Using Modern Methods of Machine Learning. *Metals* **2023**, *13*, 1395. [CrossRef]
31. Friess, J. *Entwicklung einer Reproduzierbaren Gefügeanalyse von Graphitmorphologien in Gusseisen mit Kugelgraphit*, Ergebnisse; Universitätsbibliothek der RWTH Aachen: Aachen, Germany, 2021. [CrossRef]
32. Iren, D.; Ackermann, M.; Gorfer, J.; Pujar, G.; Wesselmecking, S.; Krupp, U.; Bromuri, S. Aachen-Heerlen annotated steel microstructure dataset. *Sci. Data* **2021**, *8*, 140. [CrossRef] [PubMed]
33. Müller, M.; Britz, D.; Mücklich, F. Scale-bridging Microstructural Analysis—A Correlative Approach to Microstructure Quantification Combining Microscopic Images and EBSD Data. *Pract. Met.* **2021**, *58*, 408–426. [CrossRef]

34. Shen, C.; Wang, C.; Huang, M.; Xu, N.; van der Zwaag, S.; Xu, W. A generic high-throughput microstructure classification and quantification method for regular SEM images of complex steel microstructures combining EBSD labeling and deep learning. *J. Mater. Sci. Technol.* **2021**, *93*, 191–204. [[CrossRef](#)]
35. Feature Extraction—ImageJ. Available online: https://imagej.net/Feature_Extraction (accessed on 22 April 2024).
36. BUUnwarpJ—ImageJ. Available online: <https://imagej.net/BUUnwarpJ> (accessed on 22 April 2024).
37. Britz, D.; Webel, J.; Gola, J.; Mücklich, F. A Correlative Approach to Capture and Quantify Substructures by Means of Image Registration. *Pract. Met.* **2017**, *54*, 685–696. [[CrossRef](#)]
38. De Albuquerque, V.H.C.; Cortez, P.C.; de Alexandria, A.R.; Tavares, J.M.R. A new solution for automatic microstructures analysis from images based on a backpropagation artificial neural network. *Nondestruct. Test. Eval.* **2008**, *23*, 273–283. [[CrossRef](#)]
39. Velichko, A. *Quantitative 3D Characterization of Graphite Morphologies in Cast Iron Using FIB Microstructure Tomography*; Universität des Saarlandes: Saarbrücken, Germany, 2008.
40. DeCost, B.L.; Holm, E.A. A computer vision approach for automated analysis and classification of microstructural image data. *Comput. Mater. Sci.* **2015**, *110*, 126–133. [[CrossRef](#)]
41. DeCost, B.L.; Francis, T.; Holm, E.A. Exploring the microstructure manifold: Image texture representations applied to ultrahigh carbon steel microstructures. *Acta Mater.* **2017**, *133*, 30–40. [[CrossRef](#)]
42. DeCost, B.L.; Lei, B.; Francis, T.; Holm, E.A. High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel. *Microsc. Microanal.* **2019**, *25*, 21–29. [[CrossRef](#)]
43. Chowdhury, A.; Kautz, E.; Yener, B.; Lewis, D. Image driven machine learning methods for microstructure recognition. *Comput. Mater. Sci.* **2016**, *123*, 176–187. [[CrossRef](#)]
44. Gola, J.; Britz, D.; Staudt, T.; Winter, M.; Schneider, A.S.; Ludovici, M.; Mücklich, F. Advanced microstructure classification by data mining methods. *Comput. Mater. Sci.* **2018**, *148*, 324–335. [[CrossRef](#)]
45. Gola, J.; Webel, J.; Britz, D.; Guitar, A.; Staudt, T.; Winter, M.; Mücklich, F. Objective microstructure classification by support vector machine (SVM) using a combination of morphological parameters and textural features for low carbon steels. *Comput. Mater. Sci.* **2019**, *160*, 186–196. [[CrossRef](#)]
46. Bachmann, B.-I.; Müller, M.; Britz, D.; Durmaz, A.R.; Ackermann, M.; Shchyglo, O.; Staudt, T.; Mücklich, F. Efficient reconstruction of prior austenite grains in steel from etched light optical micrographs using deep learning and annotations from correlative microscopy. *Front. Mater.* **2022**, *9*, 1033505. [[CrossRef](#)]
47. Müller, M.; Britz, D.; Staudt, T.; Mücklich, F. Microstructural Classification of Bainitic Subclasses in Low-Carbon Multi-Phase Steels Using Machine Learning Techniques. *Metals* **2021**, *11*, 1836. [[CrossRef](#)]
48. Abouelatta, O.B. Classification of Copper Alloys Microstructure using Image Processing and Neural Network. *J. Am. Sci.* **2013**, *9*, 213–223. Available online: <http://www.jofamericanscience.orghttp://www.jofamericanscience.org.25> (accessed on 28 March 2024).
49. Rao, H.; Liu, D.; Nan, J.; Wang, J. Machine learning assisted microtextured regions segmentation in a near- α titanium alloy. *Mater. Lett.* **2024**, *363*, 136292. [[CrossRef](#)]
50. Baskaran, A.; Kane, G.; Biggs, K.; Hull, R.; Lewis, D. Adaptive characterization of microstructure dataset using a two stage machine learning approach. *Comput. Mater. Sci.* **2020**, *177*, 109593. [[CrossRef](#)]
51. Zhao, P.; Wang, Y.; Jiang, B.; Wei, M.; Zhang, H.; Cheng, X. A new method for classifying and segmenting material microstructure based on machine learning. *Mater. Des.* **2023**, *227*, 111775. [[CrossRef](#)]
52. Choudhary, A.K.; Jansche, A.; Grubesa, T.; Trier, F.; Goll, D.; Bernthal, T.; Schneider, G. Grain size analysis in permanent magnets from Kerr microscopy images using machine learning techniques. *Mater. Charact.* **2022**, *186*, 111790. [[CrossRef](#)]
53. Kondo, R.; Yamakawa, S.; Masuoka, Y.; Tajima, S.; Asahi, R. Microstructure recognition using convolutional neural networks for prediction of ionic conductivity in ceramics. *Acta Mater.* **2017**, *141*, 29–38. [[CrossRef](#)]
54. Bastidas-Rodriguez, M.X.; Polania, L.; Gruson, A.; Prieto-Ortiz, F. Deep Learning for fractographic classification in metallic materials. *Eng. Fail. Anal.* **2020**, *113*, 104532. [[CrossRef](#)]
55. Tsopanidis, S.; Moreno, R.H.; Osovski, S. Toward quantitative fractography using convolutional neural networks. *Eng. Fract. Mech.* **2020**, *231*, 106992. [[CrossRef](#)]
56. Zhou, S.; Chen, Y.; Zhang, D.; Xie, J.; Zhou, Y. Classification of surface defects on steel sheet using convolutional neural networks. *Mater. Teh.* **2017**, *51*, 123–131. [[CrossRef](#)]
57. Yun, J.P.; Shin, W.C.; Koo, G.; Kim, M.S.; Lee, C.; Lee, S.J. Automated defect inspection system for metal surfaces based on deep learning and data augmentation. *J. Manuf. Syst.* **2020**, *55*, 317–324. [[CrossRef](#)]
58. Tsopanidis, S.; Osovski, S. Unsupervised machine learning in fractography: Evaluation and interpretation. *Mater. Charact.* **2021**, *182*, 111551. [[CrossRef](#)]
59. Kunselman, C.; Sheikh, S.; Mikkelsen, M.; Attari, V.; Arróyave, R. Microstructure classification in the unsupervised context. *Acta Mater.* **2022**, *223*, 117434. [[CrossRef](#)]
60. Kitahara, A.R.; Holm, E.A. Microstructure Cluster Analysis with Transfer Learning and Unsupervised Learning. *Integr. Mater. Manuf. Innov.* **2018**, *7*, 148–156. [[CrossRef](#)]
61. Kim, H.; Inoue, J.; Kasuya, T. Unsupervised microstructure segmentation by mimicking metallurgists' approach to pattern recognition. *Sci. Rep.* **2020**, *10*, 17835. [[CrossRef](#)] [[PubMed](#)]

62. Na, J.; Lee, J.; Kang, S.-H.; Kim, S.-J.; Lee, S. Label-free grain segmentation for optical microscopy images via unsupervised image-to-image translation. *Mater. Charact.* **2023**, *206*, 113410. [[CrossRef](#)]
63. Gupta, S.; Sarkar, J.; Kundu, M.; Bandyopadhyay, N.; Ganguly, S. Automatic recognition of SEM microstructure and phases of steel using LBP and random decision forest operator. *Measurement* **2019**, *151*, 107224. [[CrossRef](#)]
64. Tsutsui, K.; Terasaki, H.; Uto, K.; Maemura, T.; Hiramatsu, S.; Hayashi, K.; Moriguchi, K.; Morito, S. A methodology of steel microstructure recognition using SEM images by machine learning based on textural analysis. *Mater. Today Commun.* **2020**, *25*, 101514. [[CrossRef](#)]
65. Webel, J.; Gola, J.; Britz, D.; Mücklich, F. A new analysis approach based on Haralick texture features for the characterization of microstructure on the example of low-alloy steels. *Mater. Charact.* **2018**, *144*, 584–596. [[CrossRef](#)]
66. Liu, X. Microstructural Characterisation of Pearlitic and Complex Phase Steels Using Image Analysis Methods. Ph.D. Thesis, University of Birmingham, Birmingham, UK, 2014.
67. Müller, M.; Britz, D.; Ulrich, L.; Staudt, T.; Mücklich, F. Classification of Bainitic Structures Using Textural Parameters and Machine Learning Techniques. *Metals* **2020**, *10*, 630. [[CrossRef](#)]
68. Arganda-Carreras, I.; Kaynig, V.; Rueden, C.; Eliceiri, K.W.; Schindelin, J.; Cardona, A.; Seung, H.S. Trainable Weka Segmentation: A machine learning tool for microscopy pixel classification. *Bioinformatics* **2017**, *33*, 2424–2426. [[CrossRef](#)] [[PubMed](#)]
69. Müller, M.; Britz, D.; Mücklich, F. Application of Trainable Segmentation to Microstructural Images Using Low-alloy Steels as an Example. *Pract. Met.* **2020**, *57*, 337–358. [[CrossRef](#)]
70. Aggarwal, C.C. *Neural Networks and Deep Learning*; Springer International Publishing AG: Berlin/Heidelberg, Germany, 2018. [[CrossRef](#)]
71. Tammina, S. Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. *Int. J. Sci. Res. Publ.* **2019**, *9*, 143–150. [[CrossRef](#)]
72. Goetz, A.; Durmaz, A.R.; Müller, M.; Thomas, A.; Britz, D.; Kerfriden, P.; Eberl, C. Addressing materials' microstructure diversity using transfer learning. *npj Comput. Mater.* **2022**, *8*, 27. [[CrossRef](#)]
73. Stuckner, J.; Harder, B.; Smith, T.M. Microstructure segmentation with deep learning encoders pre-trained on a large microscopy dataset. *npj Comput. Mater.* **2022**, *8*, 200. [[CrossRef](#)]
74. Ling, J.; Hutchinson, M.; Antono, E.; DeCost, B.; Holm, E.A.; Meredig, B. Building data-driven models with microstructural images: Generalization and interpretability. *Mater. Discov.* **2017**, *10*, 19–28. [[CrossRef](#)]
75. Kunselman, C.; Attari, V.; McCleyny, L.; Braga-Neto, U.; Arroyave, R. Semi-supervised learning approaches to class assignment in ambiguous microstructures. *Acta Mater.* **2020**, *188*, 49–62. [[CrossRef](#)]
76. Chen, D.; Sun, D.; Fu, J.; Liu, S. Semi-Supervised Learning Framework for Aluminum Alloy Metallographic Image Segmentation. *IEEE Access* **2021**, *9*, 30858–30867. [[CrossRef](#)]
77. Jung, J.; Na, J.; Park, H.K.; Park, J.M.; Kim, G.; Lee, S.; Kim, H.S. Super-resolving material microstructure image via deep learning for microstructure characterization and mechanical behavior analysis. *npj Comput. Mater.* **2021**, *7*, 96. [[CrossRef](#)]
78. Wang, W.; Wang, H.; Yang, S.; Zhang, X.; Wang, X.; Wang, J.; Lei, J.; Zhang, Z.; Dong, Z. Resolution enhancement in microscopic imaging based on generative adversarial network with unpaired data. *Opt. Commun.* **2022**, *503*, 127454. [[CrossRef](#)]
79. Pütz, F.; Henrich, M.; Fehlemann, N.; Roth, A.; Münstermann, S. generating input data for microstructure modelling: A deep learning approach using generative adversarial networks. *Materials* **2020**, *13*, 4236. [[CrossRef](#)]
80. Molitor, D.A.; Kubik, C.; Becker, M.; Hetfleisch, R.H.; Lyu, F.; Groche, P. Towards high-performance deep learning models in tool wear classification with generative adversarial networks. *J. Mater. Process. Technol.* **2022**, *302*, 117484. [[CrossRef](#)]
81. Fokina, D.; Muravleva, E.; Ovchinnikov, G.; Oseledets, I. Microstructure synthesis using style-based generative adversarial networks. *Phys. Rev. E* **2020**, *101*, 043308. [[CrossRef](#)]
82. Nguyen, P.C.H.; Vlassis, N.N.; Bahmani, B.; Sun, W.; Udaykumar, H.S.; Baek, S.S. Synthesizing controlled microstructures of porous media using generative adversarial networks and reinforcement learning. *Sci. Rep.* **2022**, *12*, 9034. [[CrossRef](#)]
83. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment Anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 3992–4003. [[CrossRef](#)]
84. Li, C.; Han, X.; Yao, C.; Ban, X. MatSAM: Efficient Extraction of Microstructures of Materials via Visual Large Model. *arXiv* **2024**, arXiv:2401.05638.
85. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [[CrossRef](#)]
86. Zajac, S.; Schwinn, V.; Tacke, K. Characterisation and Quantification of Complex Bainitic Microstructures in High and Ultra-High Strength Linepipe Steels. *Mater. Sci. Forum* **2005**, *500–501*, 387–394. [[CrossRef](#)]
87. Morito, S.; Huang, X.; Furuhara, T.; Maki, T.; Hansen, N. The morphology and crystallography of lath martensite in alloy steels. *Acta Mater.* **2006**, *54*, 5323–5331. [[CrossRef](#)]
88. Takayama, N.; Miyamoto, G.; Furuhara, T. Effects of transformation temperature on variant pairing of bainitic ferrite in low carbon steel. *Acta Mater.* **2012**, *60*, 2387–2396. [[CrossRef](#)]
89. Terasaki, H.; Miyahara, Y.; Hayashi, K.; Moriguchi, K.; Morito, S. Digital identification scheme for steel microstructures in low-carbon steel. *Mater. Charact.* **2017**, *129*, 305–312. [[CrossRef](#)]

90. Tsutsui, K.; Terasaki, H.; Maemura, T.; Hayashi, K.; Moriguchi, K.; Morito, S. Microstructural diagram for steel based on crystallography with machine learning. *Comput. Mater. Sci.* **2019**, *159*, 403–411. [[CrossRef](#)]
91. Li, X.; Lu, G.; Wang, Q.; Zhao, J.; Xie, Z.; Misra, R.D.K.; Shang, C. The Effects of Prior Austenite Grain Refinement on Strength and Toughness of High-Strength Low-Alloy Steel. *Metals* **2022**, *12*, 28. [[CrossRef](#)]
92. Laub, M.; Bachmann, B.-I.; Detemple, E.; Scherff, F.; Staudt, T.; Müller, M.; Britz, D.; Mücklich, F.; Motz, C. Determination of grain size distribution of prior austenite grains through a combination of a modified contrasting method and machine learning. *Pract. Met.* **2022**, *60*, 4–36. [[CrossRef](#)]
93. Bachmann, F.; Hielscher, R.; Schaeben, H. Texture analysis with MTEX—Free and open source software toolbox. *Solid State Phenom.* **2010**, *160*, 63–68. [[CrossRef](#)]
94. Niessen, F.; Nyysönen, T.; Gazder, A.A.; Hielscher, R. Parent grain reconstruction from partially or fully transformed microstructures in MTEX. *J. Appl. Crystallogr.* **2022**, *55*, 180–194. [[CrossRef](#)]
95. Britz, D.; Steimer, Y.; Mücklich, F. A New Approach for Color Metallography: Through Controlled Conditions to Objective Microstructure Analysis of Low-Carbon Steels by LePera-Etching. In *Symposium Commemorating 100 Years of E04 Development of Metallography Standards*; ASTM International: West Conshohocken, PA, USA, 2019; pp. 130–151. [[CrossRef](#)]
96. Noraas, R.; Somanath, N.; Giering, M.; Oshin, O. Structural Material Property Tailoring Using Deep Neural Networks. *arXiv* **2019**. [[CrossRef](#)]
97. Pei, Z.; Rozman, K.A.; Doğan, N.; Wen, Y.; Gao, N.; Holm, E.A.; Hawk, J.A.; Alman, D.E.; Gao, M.C. Machine-Learning Microstructure for Inverse Material Design. *Adv. Sci.* **2021**, *8*, 2101207. [[CrossRef](#)]
98. Tran, A.; Mitchell, J.A.; Swiler, L.P.; Wildey, T. An active learning high-throughput microstructure calibration framework for solving inverse structure–process problems in materials informatics. *Acta Mater.* **2020**, *194*, 80–92. [[CrossRef](#)]
99. Stiefel, M.; Müller, M.; Bachmann, B.-I.; Guitar, M.A.; Nayak, U.P.; Mücklich, F. Enhancing machine learning classification of microstructures: A workflow study on joining image data and metadata in CNN. *MRS Commun.* **2024**, *1*–9. [[CrossRef](#)]
100. Kolesnikov, A.; Lampert, C.H. Seed, expand and constrain: Three principles for weakly-supervised image segmentation. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016. [[CrossRef](#)]
101. Zhang, C.; Tavanapong, W.; Kijkul, G.; Wong, J.; De Groen, P.C.; Oh, J. Similarity-Based Active Learning for Image Classification under Class Imbalance. In Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 17–20 November 2018; pp. 1422–1427. [[CrossRef](#)]
102. Chun, S.; Roy, S.; Nguyen, Y.T.; Choi, J.B.; Udaykumar, H.S.; Baek, S.S. Deep learning for synthetic microstructure generation in a materials-by-design framework for heterogeneous energetic materials. *Sci. Rep.* **2020**, *10*, 13307. [[CrossRef](#)]
103. Lee, J.; Goo, N.H.; Park, W.B.; Pyo, M.; Sohn, K. Virtual microstructure design for steels using generative adversarial networks. *Eng. Rep.* **2021**, *3*, e12274. [[CrossRef](#)]
104. Dahmen, T.; Trampert, P.; Bougħorbel, F.; Sprenger, J.; Klusch, M.; Fischer, K.; Kübel, C.; Slusallek, P. Digital reality: A model-based approach to supervised learning from synthetic data. *AI Perspect.* **2019**, *1*, 2. [[CrossRef](#)]
105. Trampert, P.; Rubinstein, D.; Bougħorbel, F.; Schlinkmann, C.; Luschkova, M.; Slusallek, P.; Dahmen, T.; Sandfeld, S. Deep neural networks for analysis of microscopy images—Synthetic data generation and adaptive sampling. *Crystals* **2021**, *11*, 258. [[CrossRef](#)]
106. Gutierrez, P.; Luschkova, M.; Cordier, A.; Shukor, M.; Schappert, M.; Dahmen, T. Synthetic training data generation for deep learning based quality inspection. *arXiv* **2022**. [[CrossRef](#)]
107. Holm, E.A. In defense of the black box. *Science* **2019**, *364*, 26–27. [[CrossRef](#)]
108. Dwivedi, R.; Dave, D.; Naik, H.; Singhal, S.; Omer, R.; Patel, P.; Qian, B.; Wen, Z.; Shah, T.; Morgan, G.; et al. Explainable AI (XAI): Core Ideas, Techniques, and Solutions. *ACM Comput. Surv.* **2023**, *55*, 1–33. [[CrossRef](#)]
109. Holzinger, A.; Saranti, A.; Molnar, C.; Biecek, P.; Samek, W. Explainable AI Methods—A Brief Overview. In *International Workshop on Extending Explainable AI beyond Deep Models and Classifiers*; Springer: Cham, Switzerland, 2022; pp. 13–38. [[CrossRef](#)]
110. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
111. Tan, M.; Le, Q.V. EfficientNetV2: Smaller Models and Faster Training. *arXiv* **2021**, arXiv:2104.00298.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.