Computational Materials Science 148 (2018) 324–335

Contents lists available at ScienceDirect

Computational Materials Science

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

Advanced microstructure classification by data mining methods

Jessica Gola[a,*], Dominik Britz[a], Thorsten Staudt[b], Marc Winter[b], Andreas Simon Schneider[b], Marc Ludovici[c], Frank Mücklich[a]

[a] Functional Materials, Saarland University, Material Engineering Center Saarland, Campus D.3.3, 66123 Saarbruecken, Germany
[b] AG der Dillinger Hüttenwerke, Werkstrasse 1, 66763 Dillingen, Germany
[c] HTW Saar, Goebenstrasse 40, 66117 Saarbruecken, Germany

ARTICLE INFO

Keywords:
Microstructure classification
Data mining
Morphological parameter
Steel

ABSTRACT

The mechanical properties of modern multi-phase materials significantly depend on the distribution, the shape and the size of the microstructural constituents. Thus, quantification and classification of the microstructure are decisive in identifying the underlying structure-property relationship of a specific material. Due to the complexity of the microstructure in modern materials, a reliable classification of microstructural constituents remains one of the biggest challenges in metallography.

The present study demonstrates how data mining methods can be used to determine varying steel structures of two-phase steels by evaluating their morphological parameters.

A data mining process was developed by using a support vector machine as classifier to build a model that is able to distinguish between different microstructures of the two-phase steels. The impact of preprocessing and feature selection methods on the classification result was tested.

1. Introduction

The microstructure of advanced steels is usually controlled by sophisticated thermo-mechanical processing or heat treatments post hot rolling [1]. Depending on chemical composition and process control, the microstructure of such steels may consist of a range of different phases. If the microstructure consists of more than one phase, the properties of the material strongly depend on the type and distribution of the respective phases [2]. Therefore, it is crucial to determine the type and amount of the different phases in order to assess the underlying structure-property relationship. Traditionally, microstructures of steels are characterized by using standard metallographic procedures based on chemical etching and light optical microscopy (LOM) and they are classified by comparing the microscopy images with reference series.

Especially for steel and its complex microstructures the comparison with reference series is strongly dependent on the expert's subjective opinion. Nonetheless, steel is still one of the most widely used materials because of its excellent mechanical properties and the huge variety of applications [3]. Therefore, there is significant interest in the devolvement of objective quantification techniques for steels.

In order to characterize steel, the microstructures can be etched for example with a structure etching such as Nital [4] or color etching techniques like Beraha's etchant [5]. Due to different contrasts obtained by etching the ferritic matrix can be distinguished from a pearlitic, bainitic or martensitic second phase. However, these etchings are limited to empirical approaches and quickly reach their limits, especially for the discrimination of different phase constituents in steels that exhibit more than two phases. Furthermore, the microstructures of complex multi-phase steels are usually too fine to be resolved by light optical microscopy. A proper characterization requires modern metallographic techniques such as high resolution scanning electron microscopy (SEM) or electron back-scatter diffraction (EBSD) [6,7]. Therefore, any approach aiming at identifying the phase constituents of multi-phase steels has to rely on morphological or crystallographic parameters accessible by these techniques [8–13].

Recently, several studies have focused on EBSD for the microstructural characterization of steels, as this technique can provide direct information on the phase composition [6,7,8,14]. For example, in Ref. [14] a multitude of steel grades from different manufactures has been studied and an EBSD-based classification model was proposed. It was shown that the kernel average misorientation (KAM) deduced from EBSD measurements can be used to distinguish between ferrite, bainitic ferrite and martensite. Although those EBSD-based approaches have proven to work out for some steels, the phase separation by means of EBSD is very subjective as it strongly depends on a proper selection of

* Corresponding author.
E-mail addresses: jessica.gola@uni-saarland.de (J. Gola), d.britz@mx.uni-saarland.de (D. Britz), Thorsten.Staudt@dillinger.biz (T. Staudt), Marc.Winter@dillinger.biz (M. Winter), Andreas.Schneider@dillinger.biz (A.S. Schneider), marc.ludovici@uni-saarland.de (M. Ludovici), muecke@matsci.uni-sb.de (F. Mücklich).

https://doi.org/10.1016/j.commatsci.2018.03.004
Received 22 November 2017; Received in revised form 28 February 2018; Accepted 1 March 2018

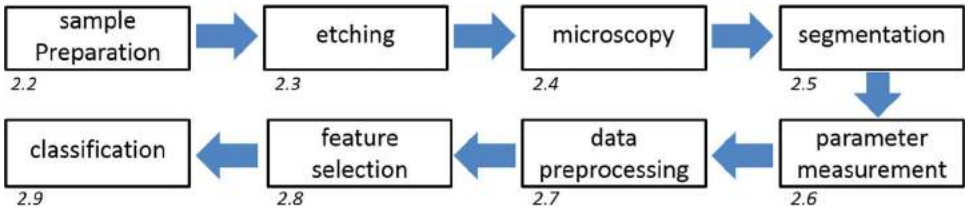

Fig. 1. Steps of the different methods used to get the data for the data mining classification.

the preparation, measurement and evaluation parameters [14].

A completely new approach for the classification of microstructures using data mining methods was presented by Velichko et al. [15]. Data mining is the process of knowledge discovery in datasets [16]. It summarizes all analysis procedures required in order to identify interesting trends and patterns within data and includes data preparation and data modeling. Various models can be constructed, depending on the research goal. In order to properly interpret the models, standard evaluation and statistical procedures are needed [16]. In their study, Velichko and coworkers have used data mining methods to classify the different graphite morphologies in cast iron by using a support vector machine (SVM) as a classifier. A SVM is a binary classification method that takes labeled data from different classes as input and outputs a model for classifying new unlabeled/labeled data into different classes. Basically, Velichko et al. have deduced morphological parameters from optical micrographs of several reference samples and analyzed the large quantity of data with a data mining tool. This enabled them to find trends, clusters or anomalies as well as relations between the morphological parameters in the data which were characteristic for the graphite morphologies and could be used to derive a classification model. This model was tested on independent samples and showed a classification accuracy of about 95% for most of the different graphite morphologies. Unlike classic metallographic procedures, the data mining approach has the advantage of not having a scope for a subjective interpretation of the microstructure. A similar data mining-based approach was used by Liu et al. for the classification of complex steel structures [17]. In their study, workflows were developed using a nearest neighbors (kNN) classifier and pixel-based parameters to classify steels using light microscopy images of single-phase pearlitic samples and samples with a mixed microstructure. For steel with two classes they found good agreement in terms of the phase fraction in comparison to manual classification results. For pearlite the classification workflow could not be used, so in a first approach artificially created structures were tested, which reached 93.8% accuracy. DeCost and Holm showed that a classification of 7 different materials (ductile cast iron, gray cast iron, malleable cast iron, annealing twins, brass hypereutectoid steel and superalloy) with visual features can be performed by using one SVM model for each image class. Their classifier system reached a cross-validation accuracy of 83% [18].

Besides data mining, deep learning methods can be applied for the classification of microstructures. For example, Chowdhury et al. showed in a case study that pre-trained neuronal networks could be used for feature extraction from images of alloys of varying Sn–Ag–Cu compositions showing dendritic structures. With these features they reached maximum classification accuracies of 97% by using linear SVM as a classifier in order to distinguish between microstructural images in terms of the presence of dendrites [19]. Although deep learning methods have recently received considerable interest and might be used for microstructure classification [20,21], with convolutional neural networks (CNNs) the features of the classification can no longer be accessed and thus the material-based background is no longer given.

The aim of this work is to prove that a data mining process in combination with traditional microstructural parameters can be used as a means to objectively classify the microstructure of two- phase steels with a ferritic matrix using the SVM as a classifier. The structure of the data mining process is described and different preprocessing and data split options are discussed. In contrast to the graphite morphologies in cast iron, typical microstructures of steels are more complex because of the substructure. Therefore, the classification is based on morphological parameters extracted from light optical microscopy and electron microscopy images. In order to reach a high degree of generalization a large amount of data is used to train the model. In a first step, a database is generated that contains a considerable variety of structures and objects representing the different classes of microstructures, which are then used to train the model.

2. Experimental

The principal approach of the data mining-based classification developed in this study is illustrated in Fig. 1. The different steps of sample preparation, etching and microscopy have to be performed to get an image of the microstructure. Next the image has to be converted to a binary image and the parameters to build the data mining model have to be measured. The data were preprocessed and the number of parameters was reduced by feature selection methods. With the final model microstructures can be classified by their parameters.

2.1. Material

In order to obtain a representative amount of data for the training and testing of the classification model a large number of samples with well-defined microstructures is required. For this purpose, samples from thermo-mechanically rolled steel plates with a carbon content of approximately 0.06 wt-% and two-phase microstructures consisting of a ferritic matrix and either a pearlitic, martensitic or bainitic second phase were used. In total, 2586 s phase grains – hereafter called objects – with a martensitic, 564 with a pearlitic and 1326 objects with a bainitic second phase were investigated on the basis of different samples.

2.2. Sample preparation

For the microstructural characterization, cross-sectional samples were taken from the as-received material with the plane of view parallel to the rolling direction. The specimen were ground with 400–2500 grit SiC abrasive paper and then polished successively using 9, 6 and 3 µm diamond suspension. Subsequently, a 120 s OP-S polish was used to obtain the final surface finish.

2.3. Etching

The samples were etched with a modified Beraha etching solution consisting of 3 g of potassium metabisulphite and 100 ml of water to contrast the second phase in the ferrite matrix. The etching was applied for an etching time of 35–50 s depending on the type of the second phase. The relatively short exposure time in combination with the Beraha etching had the advantage that no structural etching appeared on the matrix while the second phase objects were nicely visualized and appeared dark in the light microscope [22]. Furthermore, the substructure of the second phase was also accessible at higher magnifications in the SEM, which was crucial in order to distinguish the different types of second phase.

2.4. Microscopy

To extract the morphological parameters describing the second phase objects a DM6000M light microscope from Leica and a laserscanning microscope LEXT OSL4100 from Olympus were used. In order to achieve highly detailed images of the microstructure and gather a statistically significant number of objects, images were taken at predefined positions of the samples over an area of 300 µm × 350 µm at a magnification of 1000× and stitched together. These areas on the samples were marked with Vickers indents and additionally imaged by a Zeiss Merlin scanning electron microscope to resolve the substructure of the second phase. To cover the same area as in the light microscope, 24 images were taken with a magnification of 1000× and a store resolution of 4096 × 3072 pixel. As for the light microscope images, these images were stitched together using Microsoft Image Composite Editor (ICE). For the imaging, an Everhart-Thornley detector was used and the SEM was operated at an acceleration voltage of 5 kV, a probe current of 300 pA and a working distance of 5 mm.

2.5. Segmentation

Microscope images had to be converted to binary images using the image analysis program AxioVision from Zeiss in order for the computer to be able to process the data. The second phase objects were separated from the matrix by threshold value segmentation, which was carried out on the light microscopic images after standard filters such as shading correction and denoising had been applied. Several filters like dilatation were applied on the resulting binary images in order to delete artifacts and errors. Afterwards, the binarized light microscope images were overlaid with the corresponding SEM images using Adobe Photoshop CS6 and the area containing the second phase objects was cropped. The cropped SEM images were added to AxioVision and another threshold value segmentation on the combined images was applied to binarize and capture the substructure of the second phase [23].

2.6. Parameter measurement

In a next step, a broad variety of morphological parameters describing the structure and the substructure of the second phase objects were measured on the binary images [13] using AxioVision and A4i from Aquinto AG Version 5.10 a., where most of the parameters were already implemented. These traditional microstructural parameters are used in different fields like material science, but also biology or medicine for a quantitative analysis of microscopy images. In the case of material microstructure characterization, they are very helpful in quantitatively evaluating microstructures and their morphology [12]. Only two additional parameters had to be defined and calculated separately. The first one represented the area density of the substructure in a second phase object and the second one was the normalized gray value ratio of each second phase object. The substructure density was a measure of how much of the inner structure is contained in a single object and includes information on the substructure. The microscopic images itself also showed differences in the appearance of the second phase due to their appearance after etching. For the gray value ratio, the images were transformed into 8 bit grayscale images and the minimum, maximum and average grayscale of each object were determined and the new parameter was calculated. This parameter is thus associated with the texture of the phase in the light microscopic image. In total, 27 parameters were obtained for each object and are summarized in Table 1. These parameters, which were measured for the individual objects on each sample, were saved in an excel sheet and taken as input for the data mining model [13,15,24].

2.7. Data preprocessing

Data preprocessing influences the outcome of machine learning algorithms, especially if the feature values do not confirm to a common range or variance [27]. To address this issue, the values were normalized by a standardization method. This method transforms each of the feature vectors so as to get a zero mean (μ) and unit standard deviation (σ). Since for some of the parameters, feature visualization showed a standard distribution while for others it showed a skewed distribution, the differences in the data preprocessing before normalization were investigated by comparing the original data set of 27 parameters (dataset 1) with a data set transformed from the skewed distribution to the logarithm standard distribution (dataset 2).

2.8. Feature selection

Meaningful parameters for the model were selected by feature selection techniques. Feature selection is a technique to reduce dimensionality by finding a subset of the original features set that represents the dataset in the best way [28,29]. A dimensional reduction helps to reduce irrelevant information which may otherwise distort the model. Thus, the characteristic parameters that enable a distinction between the classes have to be identified in order to reduce the dimensionality and to guarantee a generalization of the results. One of the most common strategies for feature selection is a filter-based method which is used for preprocessing and ranking feature importance, regardless of the model selection.

As several morphological parameters deduced from the binarized microscope images correlated with each other such as the diameter and the perimeter of an object, which reduces the accuracy of the classification, a filter was applied such that only independent parameters remained. First, the number of parameters was reduced in order to find correlated parameters. For that reason, a parameter correlation matrix was built and all parameters with a correlation factor higher than 0.95 were removed. The remaining parameters were analyzed by different weight and feature ranking methods that were implemented in the data mining program Rapid Miner by Rapid-I GmbH [27,30]. The results of the different methods were combined to determine the importance of the parameters. After that a successive backward elimination starting with 15 parameters was performed to find the number of parameters with the highest performance. This data was used to compare the performance of the models with 27 parameters to a model of lower dimension.

2.9. Machine learning methods

In order to analyze the data extracted from the micrographs a data mining model was developed. The model was trained with the morphological parameters from the reference samples in order to identify characteristic microstructural patterns and objectively classify the microstructure. The microstructures of the second phase of the tested samples were categorized into three classes: pearlite, martensite and bainite. A SVM was implemented in the process of Rapid Miner by Rapid-I GmbH [27,30]. The SVM was chosen because this classifier can be used for a multiclass classification. The Radial Basis Function Kernel (RBF) was used because of the big amount of data and their nonlinear dependency [31]. The parameters C and γ define the model and they affect the quality of the classification by the SVM and describe the shape of the separating hyper plane. These parameters were analyzed to improve the generalization of the SVM model. *C* is a parameter that allows trading off training errors and model complexity [32].

Table 1
Traditional morphological parameters measured by two programs used in the data mining model [25,26].

Name	Description
Diameter	Diameter of a circle with an equal area.
Diameter filled	Same as diameter, but adjusts the area of an equivalent circle to account for filled holes in an object
Ellipse major	Length of the major axis of the ellipse with the same geometric moment of inertia as the region. The moment of inertia is related to the region centre of gravity
Elipse minor	Length of the minor axis of the ellipse with the same geometric moment of inertia as the region. The moment of inertia is related to the region centre of gravity
Fiber length	Length of a fiber-like region. In order to calculate the fiber length, a structure that is actually similar to a fiber is required
Feret maximum	Two straight lines are positioned on opposite sides of the object, like a sliding caliper, at 32 angle positions. The corresponding distance is measured for each angle position. The maximum value determined is the feret maximum
Feret minimum	The feret minimum parameter is measured in the same way as the feret maximum. The minimum value determined is the feret minimum
Feret ratio/stretching D	The ratio of feret minimum to feret maximum
Area	Area of a region excluding any holes it may contain.
Area filled	Area of a region including any holes it contains
Area convex	Area of convex hull of a region. The current region is surrounded by a convex polyline. The filled area of the resulting region is then measured
Stretching F	The ratio of area filled to the area of a circle with a diameter of feret maximum
Axial ratio	The ratio of ellipse minor to ellipse major.
Mittferet	Mean of the feret diameter
Euler number	Describes the topology of an object. It is calculated by the tangents of the object. An object without any holes is given the value 1
Shape factor	Describes the form of a region on the basis of its circularity. A perfect circle is given the value 1
Diameter max. Inscribed circle	The measurement of a circle, the diameter of which is expanded to its maximum, within the boundaries of the object
Radius	Radius of a circle with an equal area.
Perimeter	Perimeter of a region including the perimeter of holes it contains
Perimeter filled	Perimeter of a region excluding any holes it contains. The perimeter of the filled area is then measured.
Perimeter convex	Perimeter of convex hull of a region. The current region is surrounded by a convex polyline. The perimeter of the filled area of the resulting region is then measured
Gray value ratio	The ratio of the difference between the maximum gray value and the average gray value to the difference between the mean gray value and the minimum gray value multiplied by the ratio of the difference between maximum and minimum divided by 255
Area convex/area filled	The ratio of area convex to area filled
Perimeter convex/perimeter filled	The ratio of perimeter convex to perimeter filled
Area to total area	The ratio of area filled of an object to the sum of area filled of all objects in the image
Area density of substructure	The ratio between the total area of the substructure particles deduced from the SEM images and the area of the corresponding object as determined in the optical micrographs

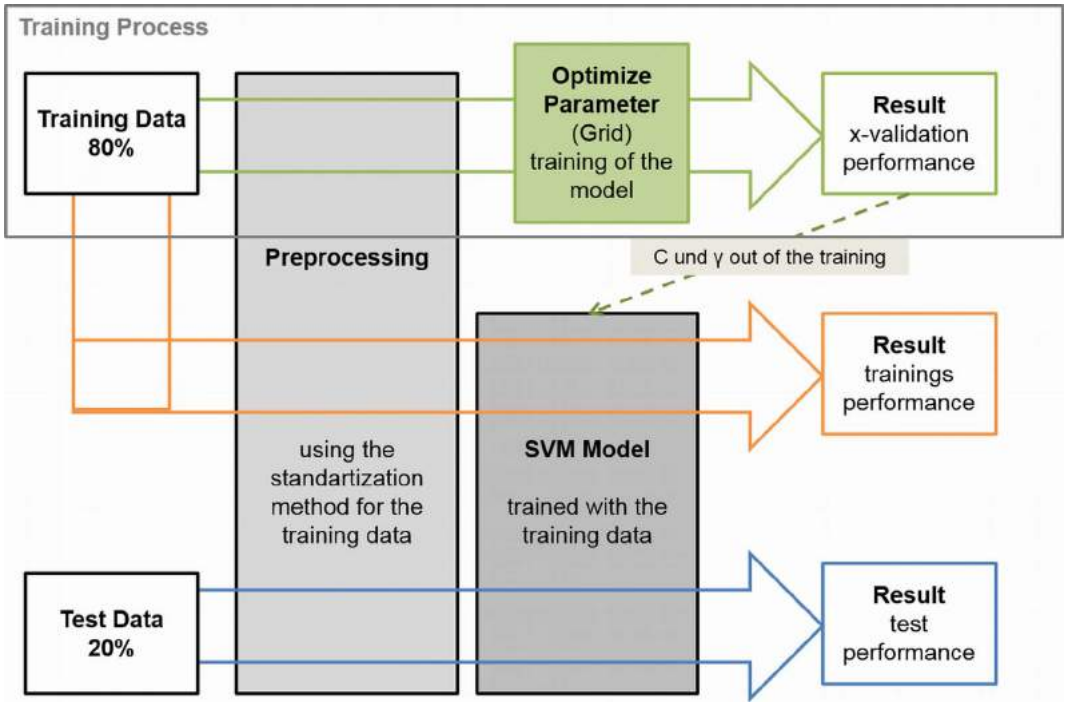

Fig. 2. Overall process chain in Rapid Miner with the three different performances for testing a model to classify microstructures. The training process used 80% of the whole data set to find the parameters C and γ in a grid search; with these parameters the final model is built. The final model is tested using the training data to get the training performance. The other part of the data (20%) was used to test the model with independent data and to get the test performance. The layers for training and testing are connected and fully automated.

2.9.1. Classification process

For the development of the classification process, the data was divided into a training and a test dataset by using two different data splitting methods. The data was shuffled and divided into training and test data in a ratio of 80:20 and compared using the sample-wise data split. Using the sample-wise data split, the data was divided into a test and a training set, i.e. some of the samples were only used for training and other samples only for testing. In that case, test and training data were completely independent because of the different samples. Using the shuffled data split, the whole variety of structure appearances in one class of different samples can be maintained and leads to a higher generalization of the model compared to a sample-wise data split. In case of the sample-wise data split there was less variety because only half of the samples were used to build the model. The test data was only used in the end to test the final model. By using the training data the best parameters for the SVM could be determined and the final model was built. The process implemented for the training of the model using the SVM as a classifier has two sub-level processes inside the *Optimize Parameter* operator [24]. Subsequently, the training data had to be imported from the Excel sheet into Rapid Miner and an operator was used to normalize the data using the standardization method [16]. In the next step, the best combination of the parameters C and γ from the SVM had to be found in a grid search using the *Optimize Parameter* operator. With this operator, numerous predefined combinations of C and γ were tested in a loop of its subprocesses. In these subprocesses, a cross-validation (x-validation) split the data into 5 parts. From these 5 parts, 4 were used for training and one for validation. On the basis of all 4 training data sets, the SVM successively created a model and predicted the validation performance for a given set of C and γ combinations. The result of the *Optimize Parameter* operator was the x-validation performance. This performance is the arithmetic mean of the accuracy of the 5 validation data sets with the best combination of C and γ out of the grid search.

On the basis of the best combination of C and γ, the SVM built the final classification model using all training data. In a next step, the model was tested with the whole data of the training set (80% of the whole data set) to get the training performance. The training performance shows how accurately the model is able to predict the data used to train the model. To get the test performance of the test data (20% of the whole data set), the test data set was normalized using the standardization model of the training data and subsequently the final model was tested.

The overall process with the three different types of results and the related performances is shown in Fig. 2. The different layers for grid search, training and testing are connected and fully automated.

The performance value of the trainings and test result shows how well the models were able to perform on the training data and the new test data. In a previous work, we investigated data mining models that were only trained by an x-validation and found that the x-validation performance is not meaningful enough to check the performance of the trained model. For this reason, it is necessary to test independent data on the final model to check the performance [31]. If the three performances in our data mining process (Fig. 2) are close together, our model has a high generalization and is able to predict new data as good as the training data.

2.9.2. Common performance metrics

The three different classification results of the developed process were summarized in confusion matrixes [33]. The major diagonal of the matrix shows the correctly classified data and the outside shows the incorrectly classified data. The overall accuracy in the confusion matrix was calculated by dividing the sum of the entries from the major diagonal by the total number of samples with the best combination of C and γ [34]. Fig. 3 shows a confusion matrix for a two-class problem.

As the overall accuracy does not indicate how the accuracy is distributed in the individual classes, the parameters, the true positive rate (TP), the false positive rate (FP), the true negative rate (TN) and the false negative rate (FN) defined in Fig. 3 were calculated from the confusion matrix. The true positive rate (TP) is the rate of objects correctly classified as positive. The false positive rate (FP) refers to negative objects incorrectly classified as positive. The true negative rate (TN) corresponds to negative objects that are correctly classified as negative and the false negative rate (FN) represents the negative objects that are incorrectly classified [35,36]. The common performance metrics calculated from the confusion matrix like accuracy, precision, recall and kappa are explained in [37].

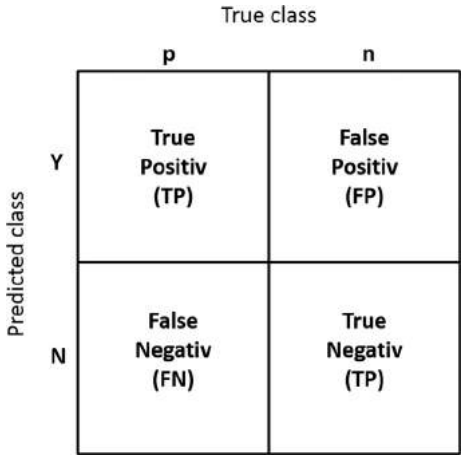

Fig. 3. Confusion matrix for a two-class problem.

3. Results

This section reports the results from the data mining process after different data preprocessing methods and two different data split techniques have been applied. Preprocessing, feature selection and data split are compared in Section 3.1. and the different parameters are compared in Section 3.2. The success of the different methods is compared based on the test performance and the kappa value of the test data for the developed data mining models.

3.1. Classification using morphological parameters

Fig. 4 shows representative laser microscopy images from tested samples with either a pearlitic, martensitic or bainitic second phase in a ferritic matrix as well as the corresponding binary images obtained by threshold value segmentation. SEM micrographs in Fig. 4c, f and i shows examples of typical second phase objects for each of the three different types of samples. In the overview images, it can be seen that size and distribution of the second phase objects vary strongly among the different types of test samples. The samples with the pearlitic second phase, as shown in Fig. 4a–c, exhibit objects with diameters up to 100 μm, which are partially connected and aligned in the rolling direction. On the contrary, the second phase objects in the samples with the bainitic second phase are much smaller and homogeneously distributed in the microstructure. The second phase objects in the martensitic dual-phase structure also exhibit diameters in the range of 100 μm and are strongly elongated in the rolling direction.

The morphological parameters (as defined in Table 1) of each object were measured and taken as input for the data mining process. In total, the data contains 4476 objects, from which 2586 objects were allocated to the martensite class, 564 to the pearlite class and 1326 to the bainitic class.

3.1.1. Classification results of the shuffled data split

The data was shuffled and divided into training and test data in a ratio of 80:20. For the shuffled data, the martensite class contained 517

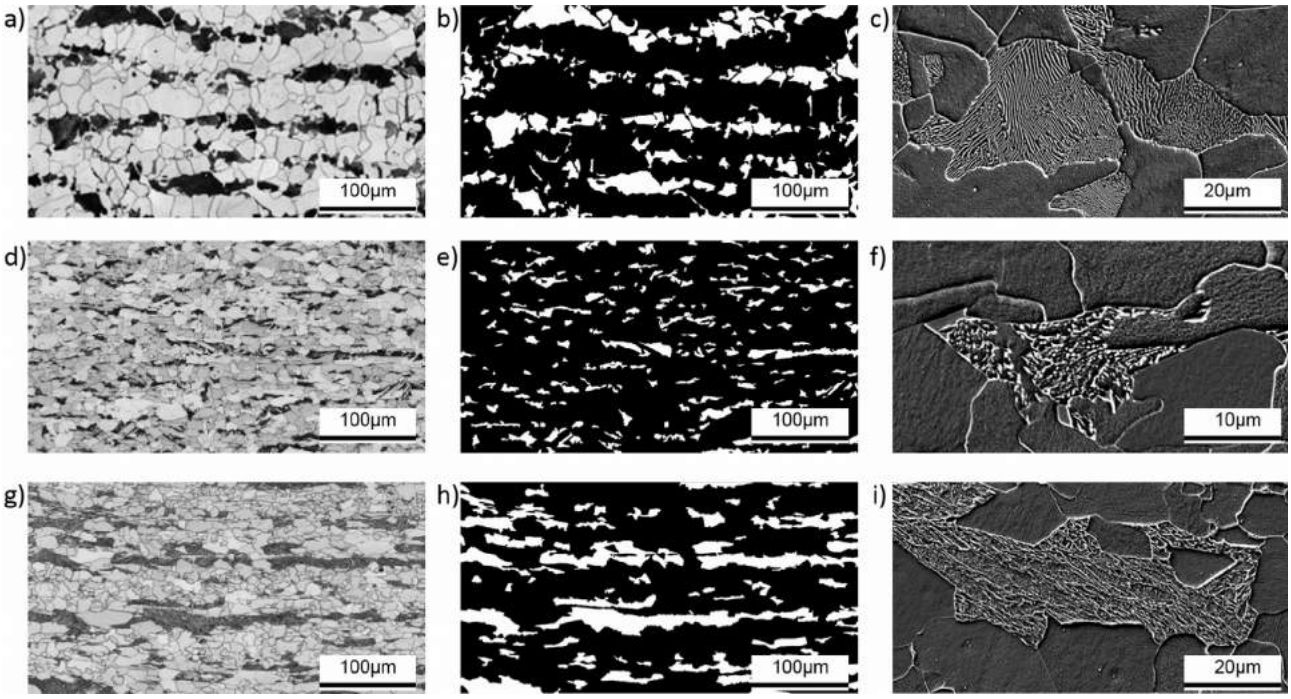

Fig. 4. Representative laser microscopy images of a pearlitic (a), a bainitic (d) and a martensitic (g) sample as well as corresponding binary images (b, e, and h); the scanning electron microscopy images show examples of the different substructures for (c) pearlite, (f) bainite, and (i) martensite.

Table 2
Summary of the different datasets that were used to train and test the data mining model.

Name	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Number of parameters	27	27	10	10
Preprocessing	Raw	Log transformed	Raw	Log transformed

objects, whereas 112 objects were considered for the pearlite class and 265 for the bainite class, respectively. From this data, different datasets were used for the experiments, as summarized in Table 2.

As shown in Table 3, dataset 1 yielded an x-validation performance of 85.03% ± 1.24% and a training performance of 89.78%. The total accuracy of the test data was 84.80%. The values of test and training performance were close to each other implying that the model can perform a reproducible classification on unknown test data as well as on training data.

For dataset 2, which was transformed to log standard distribution before standardization, the x-validation performance was 87.60% ± 0.87%. The training performance was 91.57%. The total test data accuracy was close to the training data accuracy with 86.82%.

The higher performance values of dataset 2 demonstrated that the log transformation of the input data improves the result of the classification and the quality of the model. In both cases the performance measurement kappa showed a good result with values larger than 0.7 (Table 3). The value for dataset 2, which was larger than for dataset 1, also indicated that the model was more efficient in handling data with a normal distribution than log distributed parameters. The test performance of the log transformed data of 86.8% represented a 2.38% increase compared to the raw data by using the same amount of data and number of parameters.

Table 4
Morphological parameters after removing correlated parameters and parameter ranking.

Parameter	Weight
Gray value ratio	1
Area density of substructure	
Diameter	
Ellipse major	
Euler number	
Diameter max. Inscribed circle area	
Area to total area	
Ellipse minor	
Fiber length	
Feret ratio	
Stretching F	
Shape factor	
Area convex/area filled	
Perimeter convex/perimeter filled	0

In the next steps, the parameters listed in Table 1 were reduced to further improve the classification model. After removing the correlated parameters, 15 independent parameters remained. These parameters

Table 3
Classification results for the three performance measurements of x-validation, test and training and the kappa values for dataset 1 and dataset 2 for the shuffled data.

Dataset	Number of parameters/preprocessing/data split	*Kappa* x-validation	Accuracy x-validation [%]	Accuracy training data [%]	Accuracy test data [%]	*Kappa* test data
Dataset 1	27/raw/shuffled	0.725 ± 0.026	85.03 ± 1.24	89.78	84.80	0.719
Dataset 2	27/log transformed/shuffled	0.774 ± 0.016	87.60 ± 0.87	91.57	86.82	0.758

Table 5
Classification results for the three performance measurements of x-validation, test and training and kappa values for dataset 3 and dataset 4 for the shuffled data.

Dataset	Number of parameters/preprocessing/ data split	*Kappa* x-validation	Accuracy x-validation [%]	Accuracy training data [%]	Accuracy test data [%]	*Kappa* test data
Dataset 3	10/raw/shuffled	0.750 ± 0.005	86.29 ± 0.29	90.81	84.58	0.716
Dataset 4	10/log transformed/shuffled	0.774 ± 0.016	87.60 ± 0.87	90.89	87.15	0.764

were weighted using different methods to determine the importance of the parameters. The results of the weighting process are shown in Table 4.

According to the weighting processes, the gray value ratio was the most relevant parameter with a value of 1, followed by the area density of the substructure and the diameter. The ratio of area convex to area filled and the ratio of perimeter convex to perimeter filled with values close to 0 were less relevant. These 15 parameters were further reduced by a successive backwards elimination according to their weight in Table 4. The best performance was reached using 10 parameters.

For dataset 3, the x-validation performance was 86.29% ± 0 .29%. The accuracy was 90.81% for training and 84.58% for testing. The classification results for dataset 4 showed higher performance values. As shown in Table 5 the x-validation performance was 87.60% ± 0.87%, the training performance was 90.89% and the test performance was 87.15%.

The results of dataset 4 showed that the performance of the x-validation, training data and test data was higher than that of dataset 3. The performance measurement kappa showed a higher value for dataset 4, which also indicates that the model is more efficient in handling the log transformed parameters than the raw data (Table 5). In both cases the kappa value for the test data was greater than 0.7 and in the same order as in dataset 1 and dataset 2 using 27 parameters. It can be seen that the test performance of dataset 4 was increased by 3% compared to dataset 3. Performance comparisons of classification models using 27 parameters and 10 parameters showed that the test performance was equal in the case of raw data (84.8% with dataset 1 and 84.58% with dataset 4) and increased from 86.82% (dataset 2) to 87.58% (dataset 4) for log transformed data. The best classification result in testing with 87.15% accuracy was reached with dataset 4, as shown in Table 6.

The accuracy rate for the martensite class was 94.20% with a precision of 85.74%, whereas values of 77.80% and 89.90% were found for the pearlite class and of 77.30% and 89.52% for the bainite class.

3.1.2. Classification results of the sample-wise data split

Moreover, a sample-wise data split was used for dataset 1 and dataset 2 to compare the result with the shuffled data split. The data was divided into a test and a training set, i.e. some of the samples were only used for training and other samples only for testing. In that case, the test and the training data sets were completely independent because of the different samples. In order to train the model, half of the samples of each class were used. In total, the training data contained 2511 objects, from which 1203 objects were allocated to the martensite class, 289 to the pearlite class and 1019 to the bainitic class. The remaining data was used for testing the model. The test set contained 1894 objects in total, 1290 objects of the martensite class, 252 of the pearlite class and 352 of the bainitic class.

Table 6
Confusion matrix for the classification of the morphological parameters (dataset 4) of the second phase objects for the three classes (martensite, pearlite and bainite) in Rapid Miner.

Accuracy: 87.15%				
↓Prediction /true→	Martensite	Pearlite	Bainite	Class precision
Martensite	487	24	57	85.74%
Pearlite	7	88	3	89.90%
Bainite	23	1	205	89.52%
Class recall	94.20%	77.80%	77.30%	

As shown in Table 7, dataset 1* yielded an x-validation performance of 82.39% ± 0.67%, a training performance of 85.19% and a total test data accuracy of 72.03%. The values for test and training performance were divergent and the model was not able to classify unknown test data as efficiently as training data. For the dataset 2*, the x-validation performance was 86.85% ± 1.64% and the training performance 93.76%. The total test data accuracy was 56.66% and hence much lower than in the training, which means that for independent data the log transformed data model had produced a less reproducible classification.

The results of dataset 1 in the first experiments using the shuffled data split showed a higher classification result for the test data than the results of dataset 2. But in this case the x-validation and training performance for dataset 2* was higher than for dataset 1*. Test and training performance showed a discrepancy of 37.10%, which means that the model is not able to classify independent data as efficiently as known data. The gap between test and training performance indicated that the model is overfitting. This could also be seen from the kappa value of 0.171 (Table 7).

The different sample splits demonstrated that the performance of the x-validation and training data was lower for the sample-wise data split than for the shuffled data split. But for both techniques the performance values for the training were greater than 80% and in the same order. The test performance of the sample-wise split was 72.03% and hence much lower. In addition, the kappa values for the sample-wise data split both showed a lower value.

To sum up, the best results for the test data with the developed data mining process and morphological parameters was reached by using the shuffled data split and 10 parameters which were log transformed (dataset 4) with a test accuracy of 87.15%. The results of the classification indicated that the morphological parameters of the objects form a satisfactory basis for a classification, but solely were not sufficient to perform a stable classification of the different classes. The most important morphological parameters for classification were the gray value ratio and the substructure density. Both parameters already contain information about the substructure. Therefore, in order to improve the classification performance, additional parameters of the substructure were tested in the data mining process.

3.2. Classification using substructure parameters of the second phase

In the next step, substructure parameters of the second phase objects were used for the classification process. The substructure of the second phase objects differed in appearance for the different classes for a certain etching. This offered further possibilities for separating the phases from each other. In order to see the very fine differences of the phases, high resolution images were necessary, as these structures could not be resolved in light microscopy images. Electron microscopy offers the possibility to dissolve these structures and therefore it is used today as one of the standard methods for the characterization of texture and crystallographic properties of microstructures [6–12]. In order to allow for the substructure to be used by data mining models, a methodology was developed to describe the substructure of an object using the same morphological parameters as for the second phase objects. The data for the substructure was created by overlaying the SEM image with the

Table 7
Classification results for the three performance measurements of x-validation, test and training and the kappa values for dataset 1 and dataset 2 for the sample split data.

Dataset	Number of parameters/preprocessing/ data split	*Kappa* x-validation	Accuracy x-validation [%]	Accuracy training data [%]	Accuracy test data [%]	*Kappa* test data
Dataset 1*	27/raw/sample-wise	0.694 ± 0.011	82.39 ± 0.67	85.19	72.03	0.510
Dataset 2*	27/log transformed/sample-wise	0.774 ± 0.029	86.85 ± 1.64	93.76	56.66	0.171

segmented light microscope image as a mask. This was necessary in order to be able to analyze individual objects separately [23]. In the combined image, the substructure of each second phase object could be binarized and then be measured. Fig. 5 shows the result of the binarization of a pearlitic object, in which individual, non-connected cementite particles were marked in different colors. It is important to note that for pearlite and bainite objects the binarized structure more or less represents the cementite particles, while for martensitic objects the binarized structure is related to lath boundaries and cementite particles. On the basis of the binarized substructure the morphological parameters defined in Table 1 were measured.

The average value and the standard variance of the logarithmic distribution for each of the morphological parameters of all sub-objects of a single object, as described in Table 1 were used as parameters in order to describe the substructure of a whole object. These parameters were used to build the data mining model. Only representative second phase objects were chosen to demonstrate that the second phase parameters can be used to distinguish between the different classes more effectively. Because of the high number of parameters in relation to the number of objects, only significant parameters were chosen to train the data mining model. After removing correlated attributes and weight of the parameters as described in Section 3.1, 5 parameters (area, fiber length, feret maximum, feret minimum as well as shape factor) were determined by a backward selection. The average value and the standard variance of these parameters were used for the classification. It is important to note that these parameters represent only geometrical properties of the imaged structures and contain neither information on physical properties nor on the production process parameters.

Fig. 6a and b shows examples of the binarized substructure of a pearlite and a martensite object. The pearlite object exhibits its typical, lamellar structure composed of alternating layers of ferrite and cementite in different colors. Martensite shows a fine irregular network structure related to the lath structure as well as small cementite particles. The average values of the logarithmic distribution for the maximum feret diameter and area parameters of all analyzed objects were plotted in Fig. 6c and d. It can be seen that for both parameters the pearlite class shows larger values than the martensite class. The differences in the data suggest that these characteristics enable the distinction is possible with these characteristics.

The 10 most important parameters were used as input for the data mining model and their influence on the classification result was analyzed. The data contained parameters of 123 objects that reveal the characteristic structure of the different phases. The model was trained for three classes. The classes were evenly distributed, hence each class contained 41 objects. The data were shuffled and 50% of the objects were used for training and 50% for testing. The x-validation performance was 97.15% ± 5.71% and a kappa value of 0.957 ± 0.086 was found. The training performance was 100%. The results of the classification model on the basis of the test data of 88.33% were shown in Table 8.

For the martensite class, the accuracy rate was 90% with a precision of 78.26%. In total, only two objects were classified incorrectly and both were assigned to the bainite class. Both accuracy rate and precision rate of the pearlite class were 100%, whereas for the bainite class the corresponding values were 75% and 88.24% respectively.

Overall, the classification performances considering the substructure parameters showed a higher accuracy rate than the one only based on the object parameters. The kappa value of the substructure parameters was 0.825 and hence higher than the kappa values for the

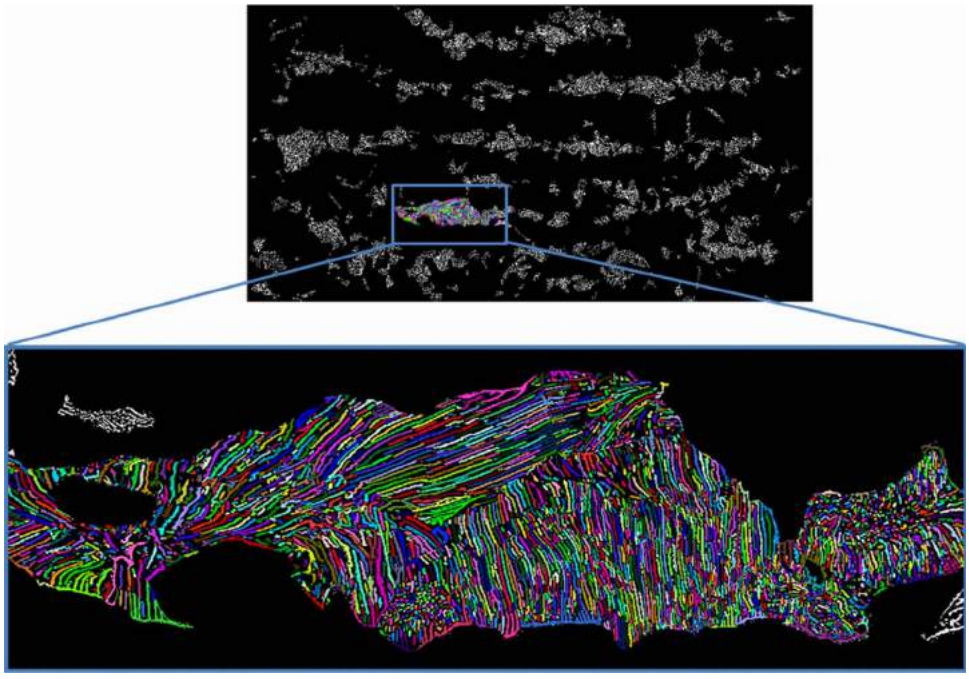

Fig. 5. Combined segmented LOM and SEM image; the individual cementite particles of the pearlite are shown in different colors.

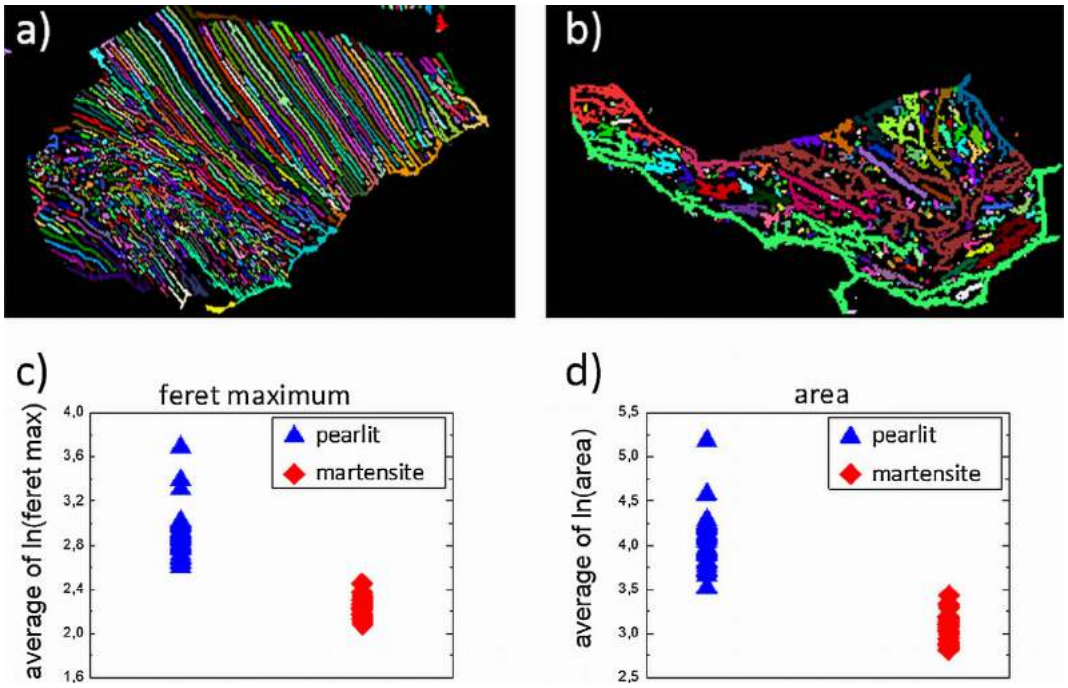

Fig. 6. Segmented and etched substructure of (a) a pearlite object and (b) a martensite object; comparison of the (c) Feret maximum and (d) area parameter for pearlite and martensite.

Table 8
Confusion matrix of the classification of the substructure parameters for the three classes martensite, pearlite and bainite in Rapid Miner.

Accuracy: 88.33%				
True pred.	Martensite	Pearlite	Bainite	Class precision
Martensite	18	0	5	78.26%
Pearlite	0	20	0	100.00%
Bainite	2	0	15	88.24%
Class recall	90.00%	100.00%	75.00%	

morphological parameters of the objects.

In Table 9: Summary of the classification results for different datasets whereby the number of parameters, preprocessing methods and data splits were changed. The SVM parameter C and γ, the total test accuracy and kappa of the classification are shown.Table 9 the different classification experiments described in Sections 3.1 and 3.2 are summarized together with the corresponding parameters C and γ as well as test accuracy and kappa.

As the table above shows, the results of the test data where the data are split sample-wise were much lower than with the shuffled data. In this case the kappa values support these results with values between 0.519 and 0.171 for the sample-wise split compared to values between 0.716 and 0.825 for shuffled data. The best results for the test data using the developed data mining process was reached for dataset 4 and the substructure parameters. The γ-values, which define the impact of a single training example, with low values meaning 'far' and high values meaning 'close', were low in all cases and showed values between 0.0272 and 0.00048. The lowest value was reached for the substructure. The C parameter, which is a parameter that allows one to trade off training errors and model complexity, showed lower values for datasets 3 and 4 than for datasets 1 and 2.

4. Discussion

The results of our data mining models demonstrate that the developed data mining process is able to build models that can classify different phases within steel samples. Based on the developed process chain it is possible to train and test the data gained from a quantitative analysis of the microstructure and its substructure in order to automatically build the final model.

The results of the classification on the basis of the given data with different datasets show that data preprocessing improves the results. The kappa values of the shuffled dataset indicate a good classification performance. The test accuracy of the classification can be improved by 2.4% for 27 parameters and by 3% for 10 parameters using the log transformed dataset instead of raw data. Furthermore, filter methods show a way to get a high performance while simultaneously reducing the complexity of the model. Decreasing the number of parameters from 27 to 10 leads to less complex models with smaller C values and provides for an equal performance both in training and testing with raw

Table 9
Summary of the classification results for different datasets whereby the number of parameters, preprocessing methods and data splits were changed. The SVM parameter C and γ, the total test accuracy and kappa of the classification are shown.

Morphological parameter	Number of parameters/preprocessing/data split	C	γ	Accuracy test data [%]	*Kappa* test data
Dataset 1	27/raw/shuffled	220,000	0.0036	84.80	0.719
Dataset 2	27/log transformed/shuffled	260,000	0.0019	86.82	0.758
Dataset 3	10/raw/shuffled	184,000	0.0272	84.58	0.716
Dataset 4	10/log transformed/shuffled	155,000	0.0059	87.15	0.764
Dataset 1*	27/raw/sample-wise	260,000	0.0010	72.03	0.510
Dataset 2*	27/log transformed/sample-wise	300,000	0.0050	56.66	0.171
Substructure	10/log transformed/shuffled	272,750	0.00048	88.33	0.825

data and a higher performance with log transformed data.

From these results it can be deduced that the most important 10 parameters contain meaningful information for the separation of the predefined phases. Table 4 shows that, according to the feature selection, the gray value ratio and the substructure density are the most important parameters followed by other morphological parameters. In contrast to the remaining morphological parameters both parameters describe the internal structure of the second phase objects. Furthermore, the size of the second phase objects, described for example by diameter and area, as well as their aspect ratio, which is described by the ellipse minor and major parameters, seems to play an important role in distinguishing between the different phases. The binary images of the bainitic two-phase microstructure used as samples show small second-phase sub-objects with a rather homogeneous aspect ratio compared to martensitic or pearlitic microstructures, as shown in Fig. 4. The second-phase objects in the martensitic and peralitic dual-phase structure exhibit diameters in the range of 100 μm and are strongly elongated in the rolling direction. Therefore the aspect ratio of these objects is higher than for bainitic objects.

However, the gray value ratio is the most important parameter. It describes the structure of the phase occurring in the light microscope on the basis of its gray values and thus contains information on the substructure of the objects. Of course, the appearance of the phases in light microscopy images depends on the type of etching and the environmental conditions, and it is partially susceptible to errors. However, under constant environmental conditions, the etching results can be compared and differences in texture and color, in case of color etching, can be identified between the classes on the images. In this work, only the gray value-based texture of the objects is evaluated as a parameter, not the color. Pearlitic structures, for example, show large gray value ranges between minimum and maximum gray value in the light microscope due to the contrast of the typical lamella structure. In bainitic and martensitic microstructures, however, the microscope images show a smaller gray value range with a homogeneous distribution of the gray values. The structures of bainitic and martensitic objects are often much finer than the pearlite lamellas and therefore cannot be completely resolved in the light microscope. Therefore, the blurring of the structures results in a low-contrast structure with medium gray values [38]. The second most important parameter is the density of the substructure. It is the only parameter that contains information from the SEM in addition to light-microscopic information. The extension of the microscopy methods by the electron microscopy has improved the classification of different phases also by material science experts. In the SEM, the very fine structures of the bainitic and martensitic structures can be resolved and the differences between the phases can be made visible [6,7]. In order to use this information for classification, the substructure density parameter is evaluated for each object. This adds a simple texture parameter with electron microscopic background to the morphology of the phases. The density of the substructure is shifted to smaller values for bainitic structures in our data, as these consist of individual cementite particles that are distributed inside the objects and do not have lamellar or network structures such as perlitic or martensitic structures with higher densities.

When 27 parameters are used and the data is split sample-wise, the test performance of the classification and the kappa values show lower accuracy for the given data. In case of a sufficiently large database the advantage of sample-wise data splitting is that it usually leads to models that are less susceptible to fluctuations within a class and show better generalization properties. However, a sufficiently large database must be available so that the training data can represent a maximum of diversity of a class. This is because different samples of the same class always show small variations. A large group of samples is therefore best suited to represent the variety of variants in a class. For a data set in combination with a sample-wise data split this means that part of the information of the data set on the variation of classes is lost for the data mining model. In this work, the given data set seems not sufficient to allow for the advantages of a sample-wise data split. The number of objects for training using a sample-wise data split is 2511, 30% less than the training dataset used for the shuffled data, which consists of 3583 objects. This leads to less variation in microstructures in the training data with a low generalization of the trained model with low performance values. In addition, the large amount of parameters compared to the amount of objects leads to an overfitting model. This is indicated by the significant difference between test and training accuracy. Overfitting means that a model describes noise instead of the underlying relationship. This occurs when a model begins to adapt to random variations in the training data rather than generalizing from tendencies. Usually the overfitting of the model results in a poor prediction performance, due to a high sensitivity to minor fluctuations in the training data [39]. The high values for C might be an additional indicator that there is some overfitting in our data mining model. In order to avoid overfitting of the model and improve the generalization, more data and a greater variation of samples are needed. Overall it can be seen that on the given data a shuffled data split lead to models that at the same time show higher generalization, less complexity because of fewer parameters and a higher classification performance.

Moreover, the results in Table 6 demonstrate that the classification of martensite objects works well with a rate of wrongly classified objects of 5.80%. The classification of pearlite and bainite is considered to be satisfying. For these classes, a higher fraction of objects was assigned to the martensite class and therefore classified incorrectly. One reason for this could be that although normally all three phases have different morphological properties; sometimes they look similar on binary images. This is the case especially when the object areas are below $5\,\mu m^2$, as demonstrated in Fig. 7 where non-typically objects for the three classes are shown whose area is smaller than the critical area. In this case, the form parameters, the density and the gray value ratio show similar values for all three classes.

In contrast, larger objects show typical appearances of the phase and are therefore easier to identify. For this reason, some small pearlitic and bainitic objects can be falsely classified as martensitic, as shown in Table 6. Therefore, a distinction between different phases based on the deduced morphological parameters cannot always be evaluated. The substructure density and gray value ratio parameters indicate that parameters describing the substructure of the objects are of great importance for the classification. In order to get a better distinction among the different types of second phase, it is important to find parameters which represent the substructure of the second phase objects more precisely like the lamellar structure of typical pearlite [17]. Furthermore, the fact that the data mining model has classified more martensitic objects than objects of the other two phases is related to the different amount of data available for the different classes. The high classification accuracy of the martensitic phase demonstrates that more data and more variation results in higher classification accuracy and in a better generalization of the data mining model. This tendency can also be seen from the classification results of sample-wise split data trained with less data. For all these reasons, more data is needed to get better statistics in general and to balance the amount of objects for the different classes.

To sum up, the results of the classification indicate that morphological parameters of the objects form a good basis for a classification, but solely are not enough to distinguish between different classes, especially if the data is split sample-wise. The morphological parameters that were most important for classification, the gray value ratio and the substructure density, both already contain information on the substructure. Therefore, additional parameters of the substructure are used to improve classification performance, especially for the pearlite and bainite classes.

The results of these substructure parameters demonstrate that pearlite and martensite can be classified successfully based on the binarized substructure. However, the classification performance for the bainite is still much lower than for the two other classes, as shown in

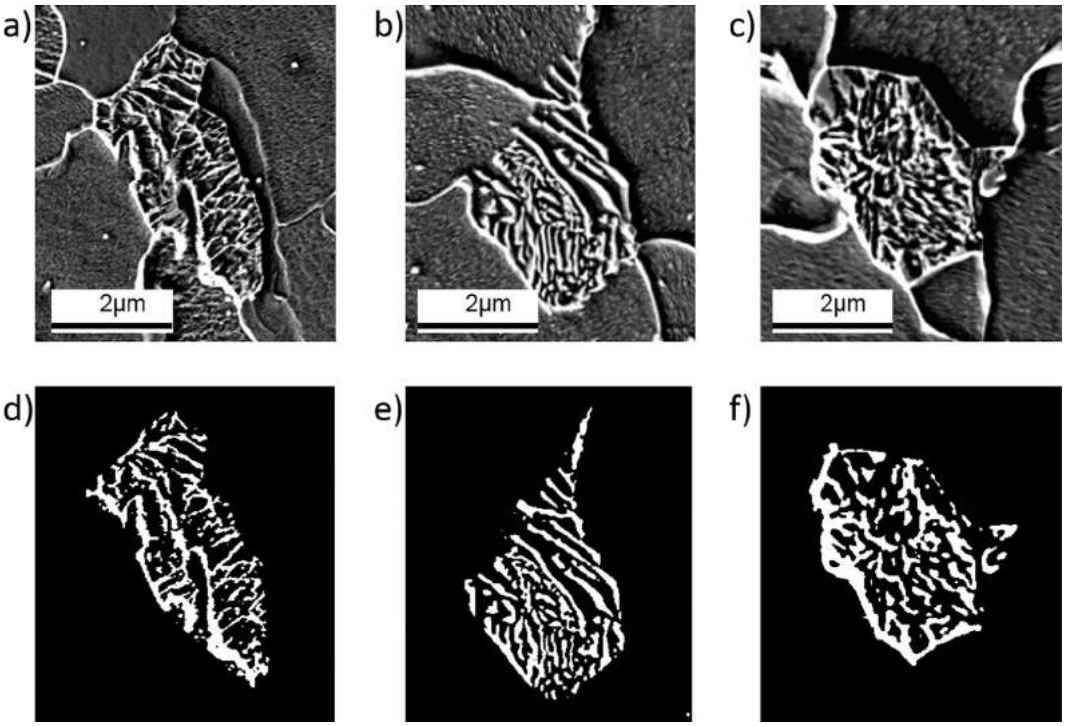

Fig. 7. Representative scanning electron microscopy images of a non-typical martensitic (a), a non-typical pearlitic (b) and a non-typical bainitic (c) object as well as corresponding binary images used to compute the area density of the substructure in an object (d–f).

Table 8. Most likely it results from the variance of structures representing this class [7]. Nevertheless, the performances of the classification based on substructure parameters show a high accuracy rate with 10 parameters.

It can be deduced from these results that the most important parameters contain meaningful information for a further differentiation of the predefined phases. In the case of the substructure, above all the mean length and areas and their standard deviation of the objects, are a suitable measure to distinguish the microstructure classes, as shown in Fig. 6. For example, the individual sub-objects of pearlite objects with the typical cementite lamellas have high feret maximum and fiber length values in the binary images, whereas the shape factor reaches small values due to its elongated morphology. In comparison, bainitic structures with finely distributed cementite precipitates have rather round particles in binary images with large shape factors and small areas with a more homogeneous ratio of feret maximum and minimum [7]. The network-like etching structure of martensite shows large values in area and fiber length for the binary structure. Thus, these parameters provide the basis for further differentiation parameters, which can be used in addition to the morphological parameters of the objects. The results indicate that the parameters of the substructure can help to further improve the accuracy. Nevertheless, it is important to note that comparable results for the substructure parameter out of the segmentation are only possible with reproducible etching and comparable segmentation methods. However, in this initial study, only representative substructure objects with areas greater than 5 μm^2 were considered for the different classes in order to produce the data in the first step. This leads to higher accuracy of the classification compared to using all objects because the model can be built with characteristic data. In the next step, the substructure parameters for all second-phase objects should be measured and tested in the SVM model so that the various appearances of the substructure in one class is included in the data. Furthermore, the different parameters for second-phase objects and substructures have to be combined in one model and the significant parameters of both have to be found in order to improve the classification accuracy. However, such information could also be used to further improve classification results. For the initial tests using morphological parameters for a microstructure classification in steel, the parameters show a good performance for second-phase objects and for the substructure.

The study shows that this data mining process can distinguish between different classes of steel based on light optical and SEM images. This is one of the major advantages of this approach in comparison with EBSD-based classification, which is extremely time-consuming and cannot be used in daily quality control. This method allows closing the gap between the subjective metallographers classification and the precise EBSD classification. Additionally, the combination of a classification by using data mining methods and morphological parameters can help to understand the important parameters for different classes and could be used for quantitative microstructure analysis. Another important point to mention in this study is that the assignment of the images to each microstructure class (the so-called ground truth) was made by materials experts before the classification step. This means that in the first step the knowledge of the model is based on expert opinions. The application of the model remains the same so that the classification results have the same fundamental error. Due to that the classification results are independent of the user's knowledge and personal opinion. In order to find a completely objective way without relying on a preclassification by experts, data mining methods could be used to identify different classes in case of unsupervised learning and using other data mining methods like CNNs. Furthermore, new parameters for the substructure could be developed and tested. Otherwise the models could be extended by using EBSD data which could be integrated in the data mining process under the condition that sufficient data is available for the database. In addition, the bainitic class allocated all objects that are neither martensite nor pearlite. So the bainitic class has to be divided into subclasses according to Zajac et al. [6] and Gerdemann [10], enabling the model to distinguish more precisely between different types of bainite. In the future three dimensional structures could help to collect large amounts of data and offer the opportunity to cut the objects in different levels so that the connection to the polished surfaces of the samples is possible. In addition,

parameters from other techniques like EBSD could be added to the database in order to further improve the precision of the classification.

5. Conclusion

The results of the present work demonstrate the feasibility of an objective classification of different structures in steel on the basis of morphological parameters by data mining methods. The process in Rapid Miner with the SVM as classifier showed good classification results for the three classes martensite, pearlite and bainite. We were able to show how to develop a process for the classification of microstructures. Furthermore, data preprocessing and feature selection could at the same time improve the classification results in order to make the model less complex and increase the generalization. Additionally, we found significant differences in the results between shuffled data split and sample-wise data split. The integration of substructure parameters in the classification process has shown high accuracy using fewer parameters. Comparable results for the morphological microstructural parameter were only possible with reproducible etching and segmentation methods. In order to further improve the accuracy, characteristic parameters of the second-phase objects and the substructure could be combined and more data should be produced in order to get statistically sufficient results and improve generalization.

References

[1] C.C. Tasan, M. Diehl, D. Yan, M. Bechtold, F. Roters, L. Schemmann, C. Zheng, N. Peranio, D. Ponge, M. Koyama, K. Tsuzaki, D. Raabe, An overview of dual-phase steels: advances in microstructure-oriented processing and micromechanically guided design, Annual Rev. Mater. Res. 45 (1) (2014) 391–431.

[2] N. Sato, Y. Adachi, H. Kawata, K. Kaneko, Topological approach to ferrite/martensite dual-phase microstructures, ISIJ Int. 52 (7) (2012) 1362–1365.

[3] P. Khedkar, R. Motagi, P. Mahajan, G. Makwana, A review on advance high strength steels, international journal of current engeneering and technology, Special Is 6 (2016) 240–243.

[4] G.F. Vander Voort, Metallography: Principles and Practice, McGraw-Hill, New York, 1984, pp. 216–217.

[5] E. Beraha, B. Shpigler, Color Metallography, American Society for Metals, 1977.

[6] S. Zajac, J. Komenda, P. Morris, Quantitative structure-property relationships for complex bainitic steels, ECSC- 247, Draft final report, 2004.

[7] S. Zajac, V. Schwin, K.-H. Tacke, Characterisation and quantification of complex bainitic microstructures in high and ultra-high strength linepipe steel, Mater. Sci. Forum (2005).

[8] S.L. Shrestha, A.J. Breen, P. Trimby, G. Proust, S.P. Ringer, J.M. Cairney, An automated method of quantifying ferrite microstructures using electron backscatter diffraction (EBSD) data, Ultramicroscopy 137 (2014) 40–47.

[9] H.K.D.H. Bhadeshia, R.W.K. Honeycomb, Steels – Microstructure and Properties, Elsevier Ltd., Cambridge, 2006.

[10] F.L.H. Gerdemann, Bainite in Medium Carbon Steels, Shaker Verlag, Aachen, 2010.

[11] D.G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, Int. J. Comput. Vis. (2004).

[12] A. International, Practical Guide to Image Analysis, ASM International, (2000).

[13] J. Ohser, F. Mücklich, Statistical Analysis of Microstructures in Materials Science, Karlsruhe, Saarbrücken, 2001.

[14] M.P. Aarnts, R.A. Rijkenberg, F. Twisk, Microstructural Quantification of Multiphase Steels (Micro-quant), Brussel, 2011.

[15] A. Velichko, Quantitative 3D Characterization of Graphite Morphologies in Cast Iron using FIB Microstructure Tomography, Shaker Verlag, Saarbrücken, 2009.

[16] R.J. Roiger, Data Mining: A Tutorial-Based Primer, Taylor&Francis Group, 2017.

[17] X. Liu, Microstructural Classification of Pearlitic and Complex Phase Steels Using Image Analysis Methods, Birmingham, 2014.

[18] B.L. DeCost, E.A. Holm, A computer vision approach for automated analysis and classification of microstructural image data, Comput. Mater. Sci. 110 (2015) 126–133.

[19] A.C. Chowdhury, E. Kautz, B. Yener, D. Lewis, Image driven machine learning methods for microstructure recognition, Comput. Mater. Sci. 123 (2016) 176–187.

[20] N. Lubbers, T. Lookman, K. Barros, Inferring low-dimensional microstructure representation using convolution neural networks, Phys. Rev. E 96 (2017) 052111.

[21] S.M. Azimi, D. Britz, M. Engstler, M. Fritz, F. Mücklich, Advanced steel microstructure classification by deep learning methods, Sci. Rep. 8 (1) (2018).

[22] D. Britz, A. Hegetschweiler, M. Roberts, F. Mücklich, Reproducible surface contrasting and orientation correlation of low carbon steels by time resolved Beraha color etching, Mater. Perform. Charact. 5 (2016) 553–563.

[23] D. Britz, J. Webel, J. Gola, F. Mücklich, A Correlative approach to capture and quantify substructures by means of image registration, Practische Metallographie 54 (10) (2017) 685–696.

[24] N. Harste, Objektive Gefügeklassifizierung von Dual-Phasen-Stählen mittels DataMining, Saarbrücken, 2014.

[25] Usermanual AxioVision, Zeiss.

[26] Usermanual A4i applications for imaging, Aquinto.

[27] M. Hofmann and R. Klinkenberg, RapidMiner: Data Mining Use, Cases and Business Analytics Applications, 2014.

[28] I. Guyon, A. Elisseeff, Special issue on variable and feature selection, J. Mach. Learn. Res. 3 (2003).

[29] B. Schowe, Feature Selection for high-dimensional data with RapidMiner, Dortmund, 2011.

[30] A. Ben-Hur, J. Weston, A user's guide to support vector machines, Data Min. Techn. Life Sci. (2010) 223–239.

[31] C.W. Hsu, C.C. Chang, C.J. Lin, A Practical Guide to Support Vector Classification, National Taiwan University, Taipei, 2010.

[32] J. Weston, C. Watkins, Multi-class support vector machines, Royal Holloway technical report CSD-TR-98-04, 1998.

[33] J. Davis, M. Goadrich, The Relationship between Precision-Recall and ROC Curves, in: Proceedings of the 23rd International Conference on Machine Learning, ICML '06, 2006.

[34] M.J. Ortiz, A.R. Formaggio, J.C.N. Epiphanio, Classification of croplands through integration of remote sensing, GIS, and historical database, Int. J. Remote Sens. 18 (1) (1997) 95–105.

[35] G.H. Rosenfield, K. Fitzpatrick-Lins, A coefficient of agreement as a measure of thematic classification accuracy, Photogram. Eng. Remote Sens. 52 (2) (1986) 223–227.

[36] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Inf. Process. Manage. 45 (2009) 427–437.

[37] J. Richard Landis, Gary G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1) (1977) 159–174.

[38] H.K.D.H. Bhadeshia, Bainite in Steels, second ed., London, 2001.

[39] D.M. Hawkins, The problem of overfitting, J. Chem. Inf. Comput. Sci. 44 (1) (2004) 1–12.