Leveraging EBSD data by deep learning for bainite, ferrite and martensite segmentation

S. Breumier, T. Martinez Ostormujof, B. Frincu, Nathalie Gey, A. Couturier, N. Loukachenko, P.E. E Aba-Perea, L. Germain

Leveraging EBSD data by deep learning for bainite, ferrite and martensite segmentation

S. Breumier [a,b], T. Martinez Ostormujof [b,c], B. Frincu [d], N. Gey [b,c], A. Couturier [d], N. Loukachenko [d], P.E. Aba-perea [a], L. Germain [b,c,*]

[a] Institut de Recherche Technologique Matériaux, Métallurgie et Procédés, 4 rue Augustin Fresnel, F-57078, Metz, France
[b] Université de Lorraine, CNRS, Arts et Métiers Paris Tech, LEM3, F-57000 Metz, France
[c] Laboratory of Excellence on Design of Alloy Metals for Low-mAss Structures (DAMAS), Université de Lorraine, France
[d] INDUSTEEL (ArcelorMittal), Centre de Recherche des Matériaux du Creusot (CRMC), Le Creusot, France

Keywords:
Machine-learning
Convolutional neural network EBSD
Low-carbon steel
Phase segmentation

A B S T R A C T

A U-Net model was trained to perform the segmentation of bainite, ferrite and martensite on EBSD maps using the kernel average misorientation and the pattern quality index as input. The manual labeling work was eased by introducing an "unknown" class that is ignored by the model during training. The influence of providing maps with different acquisition steps, indexation quality and constituent content to the model during training was investigated to demonstrate the importance of training the model with a wide range of configurations. The model can differentiate the three constituents with an 92% mean accuracy. An additional channel containing the map acquisition step was provided to the model and helped it generalize to various EBSD acquisition steps.

1. Introduction

Estimation of phase transformation products in low-carbon steels (martensite, ferrite, bainite, pearlite) is crucial to understand and predict their mechanical behavior. Several authors have therefore focused on automating this task using machine learning approaches, taking Light Optical Microscopy (LOM) or scanning electron microscope (SEM) secondary electron pictures of etched samples as input [1–8].

In particular, Larmuseau et al. [4] and Decost et al. [2] trained Convolutional Neural Networks (CNN) to classify microstructures containing various constituents using LOM and SEM pictures. The authors then used feature reduction techniques to represent, in a two dimensional space, the correlation between the features extracted by those models. Such works revealed that CNNs can extract features from those microstructures, which can be mapped into separable clusters in a well-chosen 2D space. CNNs are therefore suitable to extract discriminating features of steel transformation products for further automatic classification and segmentation, generalized to a broad range of microstructures.

Azimi et al. [1] performed automated segmentation of microstructures containing bainite, ferrite, martensite and pearlite on SEM images of etched samples, using a fully CNN with a max-voting scheme. Their model achieved an overall pixel-wise accuracy of 93.9% despite a very low accuracy for the bainite class (37%). This resulted from ferrite and bainite grains presenting similar textures and from the low amount of bainite in the dataset. Also, chemical etching does not reveal all subgrain boundaries and the etchant efficiency often depends on the studied steel grade. This could mislead the distinction of certain constituents such as ferrite with upper bainite [9]. The ground truth used to asses such model accuracy may thus not be totally accurate.

Electron BackScatter Diffraction (EBSD) provides the local orientation of the crystal probed by an electron beam. It therefore grants access to the observed surface microstructure, as well as the misorientation between points of an analyzed map. Several methods have thus been developed to manually differentiate phase transformation products using EBSD-derived data such as, misorientation distributions [10,11], grain characteristics (grain average misorientation or morphology) [12,13], variant organization [14] obtained by parent grain reconstruction [15], or the diffraction pattern quality indicators (Band Slope (BS), Band Contrast (BC)) [9,10,16,17].

However such methods often require the manual choice of thresholds to distinguish each constituent, which have to be redefined for every new steel grade studied. The labeling of each constituent thus remains a challenging and tedious task. Few authors have therefore combined

machine learning approaches with EBSD analyses for automatic constituent characterization [18–20].

Tsutsui et al. [20] leveraged the organization of variants within prior austenite grains to classify microstructures obtained with different heat treatments. This approach used conventional machine learning on EBSD data for the classification of a whole map into different categories corresponding to a given heat treatment. However, their model could not provide a segmentation of the maps into the different constituents.

Kim et al. [19] used an unsupervised machine learning approach to estimate the constituents fraction of advanced high strength steel containing different fractions of ferrite, martensite, bainite and pearlite. Their model predicted similar constituent fractions as those estimated manually by LOM observations. However, their model only provided an average constituent fraction over several grains. It is therefore difficult to assess the local accuracy of the model and build confidence on its generalization abilities, especially when different EBSD acquisition conditions and grain detection parameters are used.

The present work aims at using EBSD data and machine learning methods to perform automatic segmentation of bainite, ferrite and martensite. The ability of the model to generalize over different acquisition conditions (acquisition step, diffraction pattern acquisition parameters) was also addressed. Indeed, it is necessary to ensure that the developed model will work consistently whatever the EBSD/SEM setup for optimized data acquisition [21–23].

In a previous work, Martinez Ostormujof et al. [24] developped a U-Net model that relied on EBSD data to separate ferrite from martensite in a dual phase steel. This work demonstrated the possibility to leverage EBSD data by machine learning to perform constituent segmentation. However this problem did not require the model to fully harness the wealthiness of information carried by orientation data, as the contrast between the studied constituents was strong, either on the pattern quality or on the misorientation maps.

The objective of the present work is to extend this approach to a more complex problem, involving three constituents: bainite, ferrite and martensite. The distinction between ferrite and bainite cannot be solely based on the pattern quality and requires additional information carried by the misorientation maps. This work therefore demonstrates that a U-Net architecture can extract relevant features from different EBSD data and combine them to differentiate bainite from ferrite and martensite in low-carbon steels. It also demonstrates that this distinction can be performed consistently for different acquisition steps and diffraction pattern qualities. In particular, the acquisition step used to obtain the EBSD maps was integrated to the model input. It revealed that providing the model with contextual data helps it generalize to acquisition step unseen during training.

The findings of this research may provide interesting perspectives for the use of EBSD data in the industry. It is another step toward the use of CNN architectures for automated interpretation of microstructures with higher performances than a human expert. In the future, the resulting segmentation may not only provide a local estimate of the phase fraction but also extract statistics relative to each constituent (grain size, morphology, misorientation content, variant organization).

2. Materials and methods

2.1. Material preparation

Dilatometry samples of SuperElso® 533 E and API5L X65QS PSL2 low carbon steel were manufactured. The two steel nominal composition is given in Table 1. The alloys differ in Ni content and have therefore different transformation behavior. The samples were first heated above the austenitization temperature and then continuously cooled at different rates, detailed in Table 2. The resulting microstructures are mixtures of ferrite/bainite/martensite. The amount of each transformation product depends on the cooling rate and can be estimated from the time-temperature-transformation (TTT) diagram. Quenching austenite below martensite start temperature (M_s) produces typically lath-like martensite formed by a displacive and diffusionless transformation. For the slower cooling rate, austenite transforms first to polygonal ferrite grains nucleating and growing from the prior austenite grains by a diffusive mechanism. With the decreasing temperature, diffusion is hindered and the transformation progresses in the bainite domain and finally carbon-enriched austenite transforms to martensite or martensite/austenite islands below M_s. Bainite observed in our samples was mostly upper bainite that nucleates and grows in the high temperature bainite range. It consists of coarse ferrite laths formed by a displacive transformation followed by short range of carbon diffusion to form enriched austenite in the lath surrounding and finally cementite precipitates between laths. [25]

The samples were cut in half and mechanically polished using P1000 *SiC* grit paper, then successively with a 9 μm and a 3 μm diamond suspension and finally with a colloidal silica solution (OP-S).

2.2. EBSD acquisition

EBSD maps were acquired using an Oxford Instrument Symmetry camera in a JEOL F100 SEM with a step of 0.1 μm and an accelerating voltage of 20 kV. The raw maps were indexed as ferrite and austenite. The indexing rate was above 95% for all maps.

On each sample, two 252×189 μm maps were obtained with a binning 8×8, 0.6 ms integration time, and no pattern averaging (condition Q_{ref}). Additionally, a single map was acquired on the API-1 sample with a binning 2×2, 3 ms integration time, 2 averaged patterns per point and the Aztec'Refined accuracy' option for indexing (condition Q_{alt}).

The different constituents were identified using the band contrast (BC), grain boundaries misorientations and Kernel Average Misorientation (KAM) maps. The BC is an indicator of the diffraction pattern quality. Its contrast carries morphological information of the analyzed microstructure that could help the model to differentiate constituents with different grain shapes.

Bainite and martensite grains usually present specific grain boundary misorientations resulting from their orientation relationship with their prior austenite grain. The analysis of such specific grain boundary misorientations provides the organization of the transformed microstuctures in packets, blocks and sub-blocks within prior austenite grains. Such organization is specific to each transformation products [26].

Finally, the KAM represents the intragranular averaged

Table 2
Sample's designation and corresponding heat treatments.

Sample designation	Material	Cooling rate
SA-1	SA533 E	0.28 °C s^{-1}
API-1	API5L x65QS	2.8 °C s^{-1}
API-2	API5L x65QS	2.8 °C s^{-1} down to 620 °C + Quench
API-3	API5L x65QS	8.3 °C s^{-1}
API-4	API5L x65QS	19.4 °C s^{-1}
API-5	API5L x65QS	Quench

Table 1
Steel sample nominal composition (weight %).

Designation	C	Mn	P	S	Si	Cr	Mo	Ni	V	Other
SuperElso® 533 E	0.10	1.42	0.007	0.001	0.275	0.60	0.425	0.8	0.03	V + Nb: 0.02
API5L X65QS	0.18	1.2	0.007	0.001	0.3	0.25	0.20	0.10	0.01	Nb: 0.015

misorientation over a kernel and could be useful to distinguish ferrite, bainite and martensite, as these constituents usually present different low angle boundary content [17]. The KAM maps were computed on the raw EBSD data using a 3×3 kernel ignoring misorientations greater than 3° and non-indexed pixels. In Fig. 1, the KAM scale has been limited to 1° to better visualize the contrast between the different constituents.

Fig. 1 shows representative maps plotting those EBSD outputs, for the different samples. Three main types of constituents can be observed:

- Ferrite presenting equiaxed grains with only few intragranular low-angle boundaries, high BC values and randomly misoriented grain boundaries.
- Upper bainite corresponding to coarse lath structures separated by low-angle inter-lath boundaries and presenting lower BC values and a higher density of intragranular misorientations.
- Martensite grains consisting of packed laths separated by high angular grain boundaries and containing the lowest BC value and a high density of intragranular misorientations.

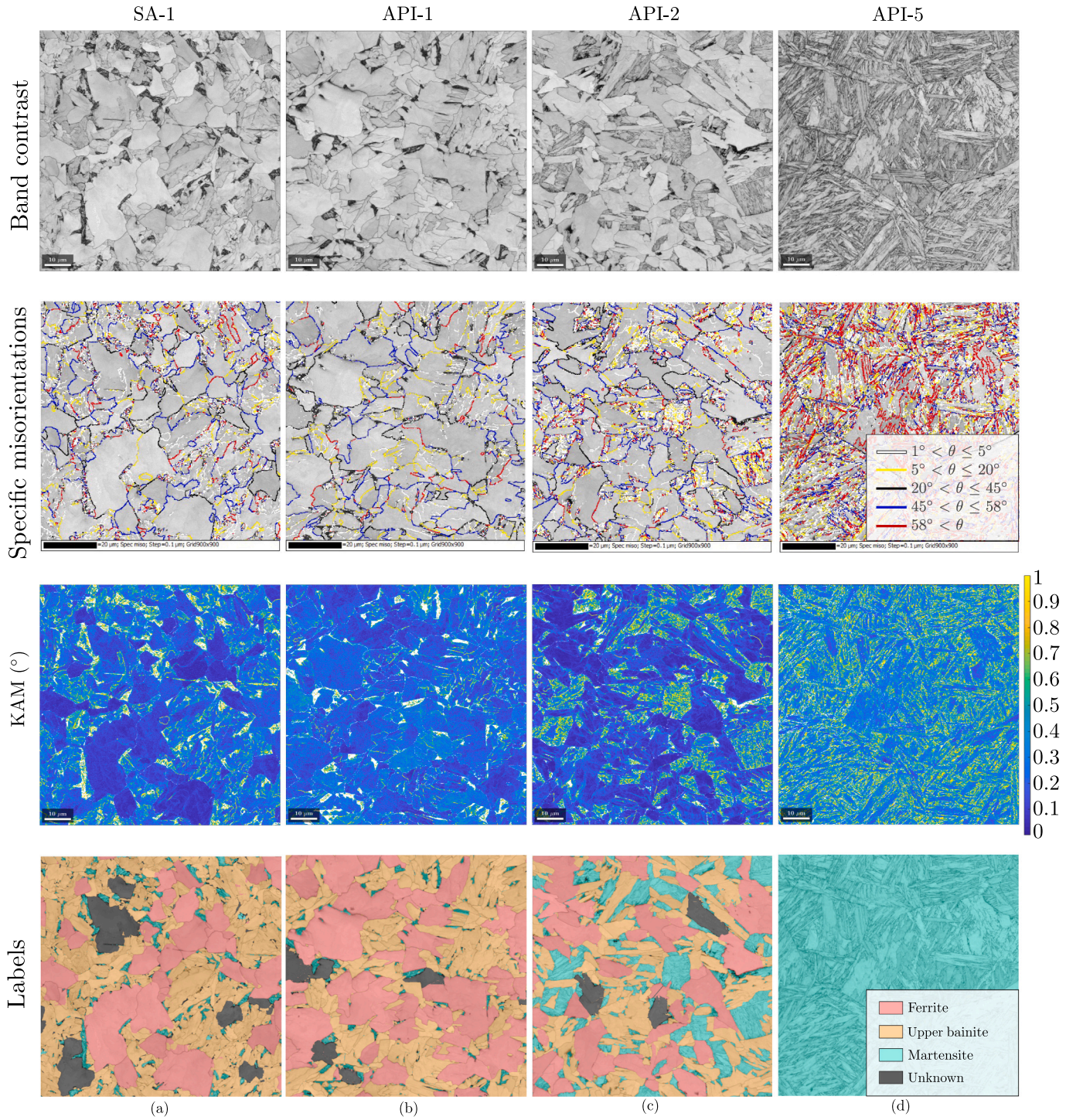

Fig. 1. Band contrast, specific misorientations, KAM and label maps of four labeled samples: (a) SA-1 (b) API-1 (c) API-2 (d) API-5. The KAM scale has been limited to 1° to better vizualize the contrast between the different constituents.

Martensite/Austenite (M/A) islands can also be observed in most samples except in the API-5, which contains only martensite.

2.3. Labeling strategy

The SA-1, API-1, API-2, API-5 maps were labeled manually using the mtex toolbox [27] and the Oxford Instrument Channel 5 software. Each pixel was labeled either as ferrite, bainite, martensite or 'unknown'.

The labeling was performed manually based on grains' morphologies, inter- and intragranular misorientations and BC levels. First, all the maps were labeled by thresholding the BC value to extract the martensite grains, and the KAM values to differentiate bainite and ferrite grains. These first labels were then manually improved by relabeling grains based on their morphology, their amount of intragranular low-angle boundaries and surrounding grains' properties.

Non-indexed pixels were located either at the grain boundaries or in M/A islands. They were therefore labeled by performing a one pixel dilatation followed by an erode operation on the manual labels. The remaining non indexed pixels were assigned to martensite.

Pixels for which it was difficult to assess the constituent were assigned to an "unknown" class. This was particularly the case for some ferrite grains that continuously transformed to bainite during cooling, with no detectable boundary. A representative example of such grain can be seen in Fig. 8. The pixels labeled as 'unknown' were ignored during training. This class allows to ease the labeling task and ensures consistency by not imposing a ground truth for ambiguous grains. The few pixels indexed as austenite were also considered as 'unknown', to be ignored by the model.

A first model was then trained with this first manual labeling, using the methodology described in Section 2.4 and 2.5. Comparing the predictions of this model with the labels revealed some differences resulting from labeling mistakes. Each of the differences between the labels and the model predictions were therefore successively analyzed to determine whether the label should be corrected or not. The final model was trained with the corrected dataset.

Fig. 1 shows representative labeled examples. Fig. 2 shows the distribution of the model inputs (BC and KAM) over all the labeled samples for each constituent. Overall, ferrite pixels have higher BC and lower KAM values than the other constituents, as they present only few misorientations and therefore have a higher pattern quality. Inversely, martensite pixels have darker BC and larger KAM misorientation values than those of the other constituents.

The labeling is therefore consistent with the characteristics observed qualitatively for these three constituents in Section 2.2. Also, the overlap observed between the constituents in Fig. 2 shows that it was not possible to separate those three constituents using a simple thresholding operation on the BC and the KAM. This further motivates the use of a machine learning model.

In total, eight 252×189 μm maps were labeled. The final dataset was composed of 28% of ferrite, 25% of martensite, 44% of bainite and 3% of 'unknown' pixels.

2.4. Deep learning model

The model used was a U-Net convolutional network, initially developed by Martinez Ostormujof et al. [24] to leverage orientation data for constituent discrimination in dual-phase microstructures. The model input is a picture with two channels containing the BC and KAM maps.

Note that alternative EBSD data indicators could be combined to vary the model input (Band Slope (BS), misorientations, orientation quaternions, …). The model trained with the BC and the KAM was used in this study as it provided the most consistent results.

The step size of EBSD maps is usually adapted to the microstructure and to the amount of time available on the microscope. Therefore, the model should have the ability to perform consistent predictions over a wide range of EBSD map resolutions and data acquisition conditions. To account for the EBSD map scale, an additional channel containing a constant value corresponding to the EBSD acquisition step was concatenated to the model input.

As in the work of Martinez Ostormujof et al. [24] the model starts with an encoder, composed of different convolution and dropout layers, followed by pooling and ReLU activations, progressively reducing the image resolution while increasing its depth. Each pooling operations requires padding of the image border by repeating the information of the nearest pixel (padding 'same'). The reduced image then passes through a decoder, composed of upsampling, convolution and dropout layers, linked to the encoder part by residual connections. This architecture allows to extract features of the input maps at different scales and combine them to perform the segmentation. The output image then passes through a softmax layer predicting the probability of each pixel to belong to a given constituent. The constituent with the highest softmax probability is assigned to each pixel.

The number of convolutional layers in the model was chosen to maximize the model testing accuracy, while keeping a reasonable training time. The best compromise was reached by training a model

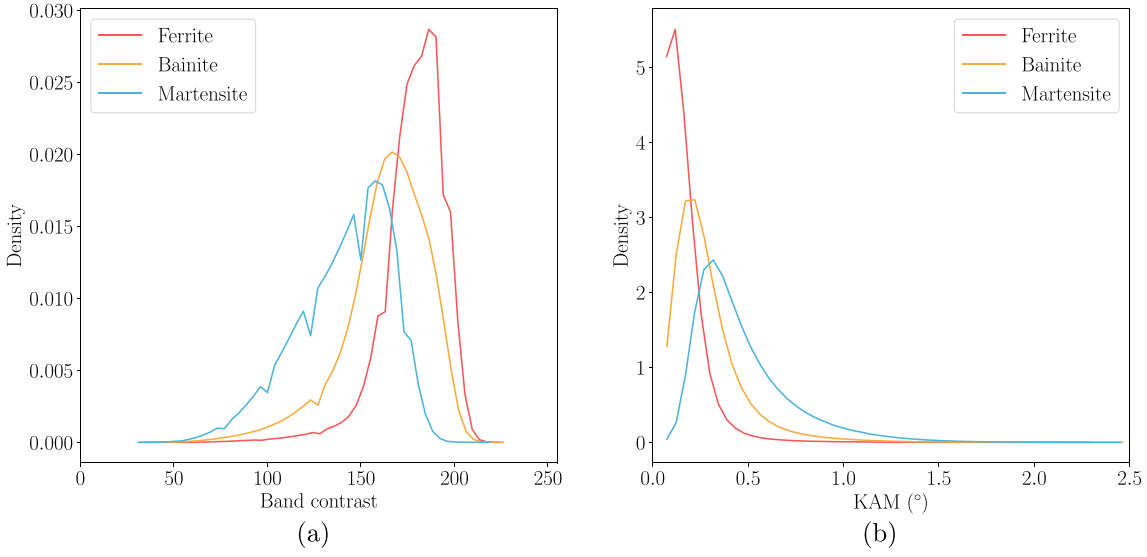

Fig. 2. Distribution of the (a) BC and (b) KAM values over all the labeled samples for each constituent.

composed of 11 convolutional blocks. The model was implemented using the Keras python library [28].

2.5. Training strategy

The amount of data was augmented by performing a rotation of 90°, 180° and 270° and a vertical and a horizontal symmetry operations. Each map was also downsampled by a factor of three to provide the model with 0.1 μm and 0.3 μm step maps. The downsampling was performed by keeping only one pixel over N both horizontally and vertically, N being the reduction factor. The KAM maps were then recomputed on these reduced maps. A higher step size usually results in higher KAM values [29] and could artificially widen sub-grain boundaries. Feeding the network with maps acquired at different step sizes should therefore help it to account for this variability, as further discussed in Section 3.3.2. Before being fed to the network, the BC and KAM maps were normalized so that their distribution over all maps was centered at 0 with a standard deviation of 1.

Each 252×189 μm map was cut in half horizontally. The top half was used for the training and validation of the model and the bottom half was kept for testing the model on unseen data. Maps were then cropped into subsets of 224 × 224 pixels to reduce training time. The model was however tested on full maps after training, to reduce potential border effects resulting from the padding. The different figures presented in the results focus on specific region of interest taken from the fully predicted maps.

First, 20% of the training dataset was kept for validation during training to determine the number of iterations that should be performed to maximize the model accuracy without over-fitting the training data set. Then, the whole data set was then used for training, with the determined number of epochs.

The model was trained over 100 epochs, using the python keras 'adadelta' optimizer [28]. The training takes approximately 6 h using an Intel Core i5-8250U CPU.

3. Results

3.1. Model accuracy on labeled samples

Fig. 3 shows an example of the predictions performed by the model on a crop of the API-2 test map, with the associated input BC and KAM fields. The differences between the labeling and the prediction are highlighted in yellow in Fig. 3 (e). In general, the model classifies grains with few local misorientations as ferrite, grains with subgrain boundaries and tortuous grain boundaries or coarse laths as bainite and small laths with high KAM values and a dark BC as martensite. This is confirmed by Fig. 4 which shows the BC and KAM distributions for each predicted constituent over all the labeled samples. The predicted distributions match those of the labels shown in Fig. 2.

Table 3 shows the accuracy of the model. The mean pixel accuracy for each test map is greater than 86% for segmentation of the three constituents.

The model pixel accuracy for a given map depends on the sample complexity. As shown in Fig. 1, SA-1 presented the most pronounced KAM contrast between upper bainite and ferrite. It also had a high model testing accuracy (94.6%). The API-1 sample was more difficult to label, as distinction between upper bainite and ferrite was often more ambiguous. The model accuracy was accordingly lower than for the SA-1 sample (87.2%). The API-2 sample contained a larger proportion of martensite than that of the two first samples. It therefore represented a greater challenge both for labeling and prediction. However the average accuracy for this sample was similar to that of the API-2 sample (86.6%). The API-5 contained only martensite and was therefore easier to label and to classify.

Fig. 5 compares predicted and manually estimated constituent fractions for each constituent. A good agreement can overall be observed. Larger errors are found for the API-1 and API-2 samples, consistently with the lower pixel accuracy obtained for these samples.

Fig. 6 shows the model confusion matrix for the different constituents. The confusion between martensite and the other classes are negligible. The largest errors are found between ferrite and bainite. Indeed, 15% of the pixels labeled as ferrite were misclassified as bainite by the model. Inversely, 7% of the pixels labeled as bainite were misclassified as ferrite. A higher error is therefore found for ferrite rather than bainite. This could result from the fact that the dataset contains slightly more bainite examples than ferrite. It could also result from labeling mistakes, as, despite introducing the 'unknown' class, the difference between the two constituents was often ambiguous.

Fig. 7 shows representative examples of such confusions. Some small or non-equiaxed ferrite grains surrounded by bainite or martensite grains or presenting few intragranular misorientations were predicted as bainite. Large bainite grains surrounded by ferrite grains were inversely often misclassified as ferrite.

Fig. 8 (a) and (b) shows the KAM and the BC maps of a grain labeled as unknown on an API-2 EBSD map. This grain corresponds to a ferrite grain which continuously transformed into bainite during cooling. It could be divided in two parts: (i) a ferrite part containing a large equiaxed zone with very few intragranular misorientations and (ii) a bainite part corresponding to a lath like structure with a slightly higher intragranular misorientation density. However, due to the continuous nature of this transformation, no boundary can be identified between those two parts neither on the KAM nor on the BC map. Fig. 8 (d) shows the prediction of the model for this grain. The model has well identified the dual composition of this zone. However, parts of the bainitic lath-like

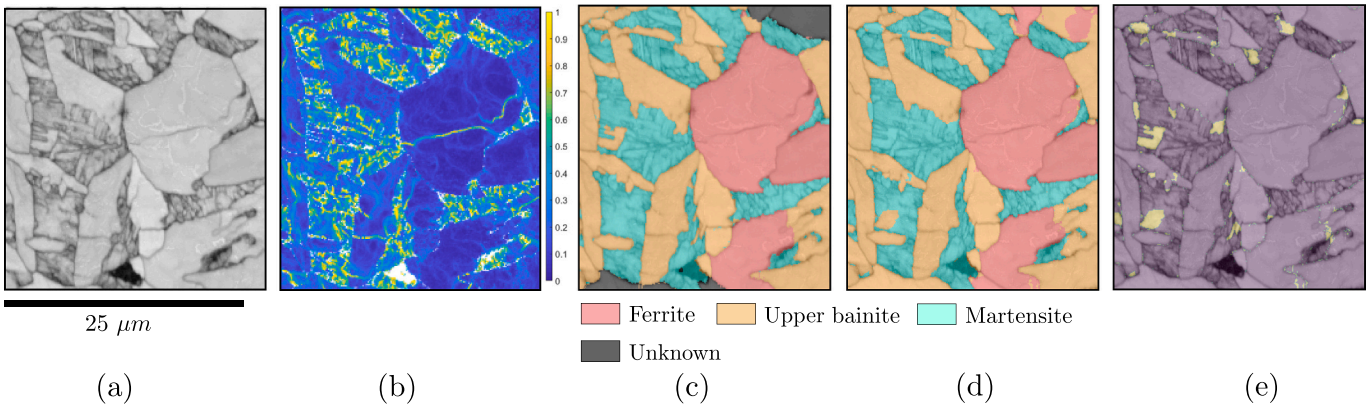

Fig. 3. Model prediction example on an API-2 test map. (a) BC map, (b) KAM map, (c) Manually labeled map, (d) Model predictions, (e) difference between labeling and prediction.

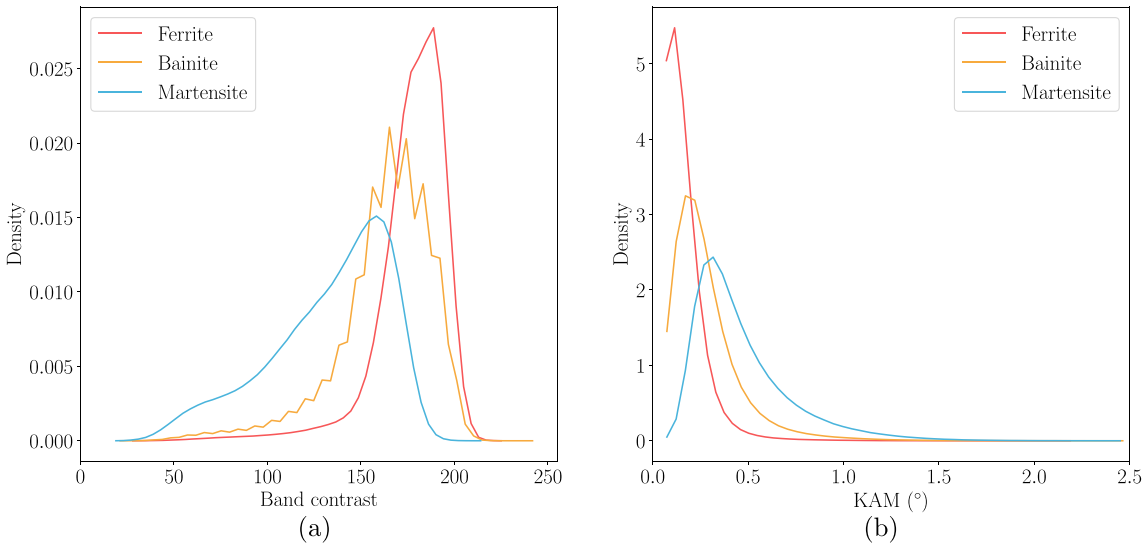

Fig. 4. Distribution of the (a) BC and (b) KAM values over all the labeled samples for each predicted constituent.

Table 3
Model train and test accuracy on the different samples.

Sample designation	Material	Heat treatment	Training accuracy (%)	Test accuracy (%)
SA-1	SA533 E	0.28 °C s^{-1}	96.2	94.6
API-1	API5L x65QS	2.8 °C s^{-1}	93.0	87.2
API-2	API5L x65QS	2.8 °C s^{-1} down to 620 °C + Quench	91.3	86.6
API-5	API5L x65QS	Quench	100.0	99.9

structure was considered as ferrite by the model, in the zone where the phase transition occurs. This is probably because the model was fed with only few labeled examples of such transitions, due to the difficulty to label them.

3.2. Model behavior on unlabeled samples

The model was also tested on the unlabeled API-3 and API-4 samples, that were cooled after austenitization at 8.3 °C s^{-1} and 19.4 °C s^{-1}, respectively. Those samples presented a similar microstructure as that of SA-1 and API-1 but with a higher bainite fraction.

Fig. 9 shows the model predictions on those samples with their respective KAM maps. Large zones with few misorientations were classified as ferrite. Zones with a higher misorientation density and more tortuous grain boundaries, were classified as bainite. This is confirmed by Fig. 10 which shows the BC and KAM distribution for each predicted constituent over the two unlabeled samples. The M/A islands were also qualitatively well identified by the model as martensite.

Some coarse bainite laths containing very few intragranular misorientations were misclassified as ferrite, as can be seen for instance in the dotted frame in Fig. 9 (c). This results from the absence of such microstructures in the training dataset. This shows the importance of providing the model with a broad set of different microstructures to help it generalize.

To ensure the physical consistency of the model predictions over different cooling rates, the predicted constituent fractions were computed for the labeled API-1 and unlabeled API-3 and API-4 samples. The results are summarized in Table 4. The higher the cooling rate is, the lower the ferrite fraction and the higher the bainite and martensite fractions. This is consistent with the transformation mechanism of those constituents and further gives confidence in the abilities of the model.

These results reveal that the model generalizes well to microstructures of the same steel grade, with similar features as those seen during training but with different constituent fractions.

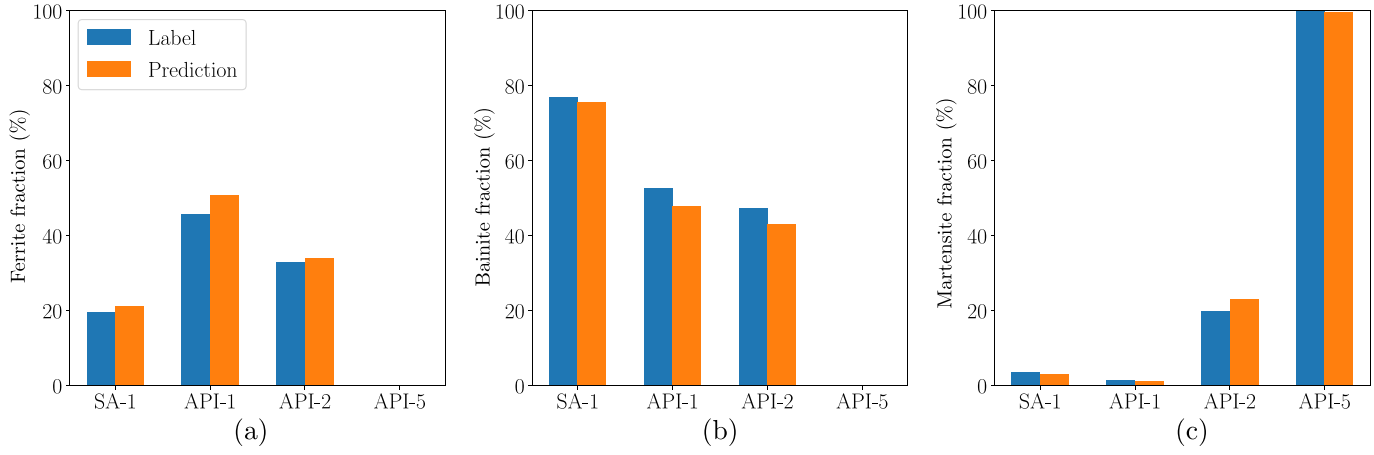

Fig. 5. Comparison of the predicted and manually labeled constituent fraction for (a) Ferrite, (b) Bainite and (c) Martensite on each test map.

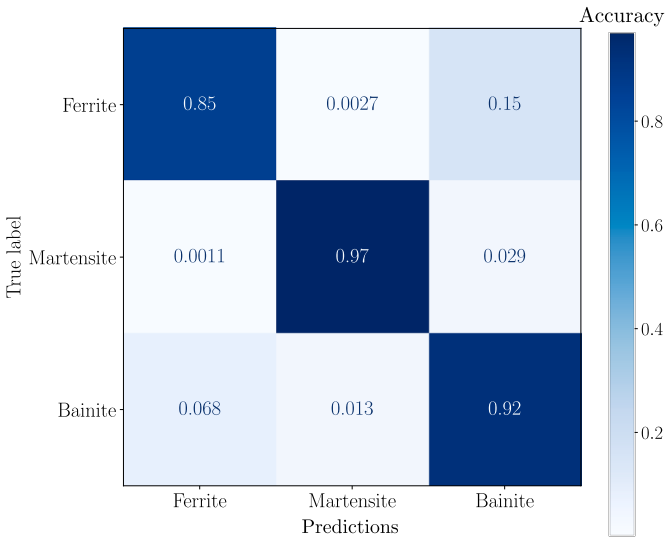

Fig. 6. Confusion matrix of the model predictions for the different classes.

3.3. *Influence of input data*

3.3.1. *Data quantity and diversity*

A CNN usually requires a large amount of data to reach the optimal accuracy. The amount of data used to train the model was therefore varied to assess that sufficient data was provided to the model.

Fig. 11 shows the evolution of the model testing accuracy with the number of 224 × 224 pixel crops used for training. Each point represents the average of 5 successive trainings using the same input data and hyper parameters but with different initial weights.

The testing accuracy saturates around 1000 crops. As the whole dataset was used for the final training, a sufficient amount of data was therefore given to the model to reach the optimal accuracy. Also, the model accuracy varies by 1% over five successive trainings. This results from the stochastic nature of the adadelta optimization algorithm as well as the random weight initialization. The network will therefore not necessarily converge systematically toward the same local equilibrium.

However, reaching a satisfying accuracy on the test set does not mean that the model is able to generalize to unseen microstructure types. To investigate this, another model was trained without the quenched sample map to demonstrate the importance of providing the model with a large diversity of microstructures. The training set in this case was uneven as it only contained 9% of martensite. Also, this model has not seen any microstructure with large martensite content as the

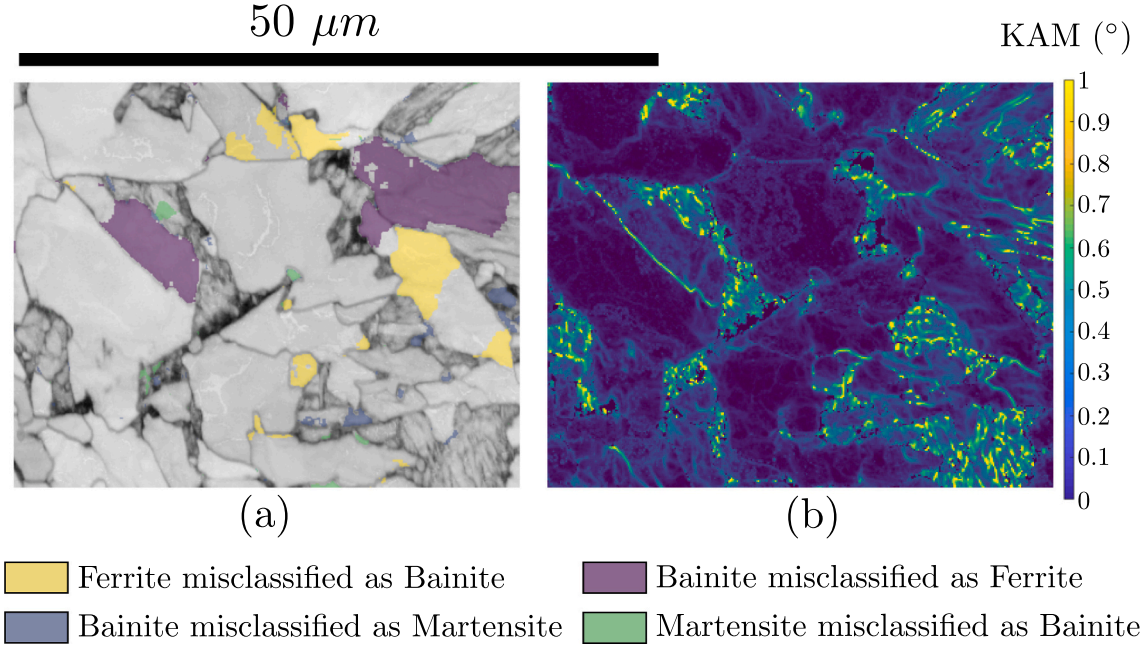

Fig. 7. (a) Example of local misclassification in a API-2 sample EBSD map overlap on BC and (b) the associated KAM map.

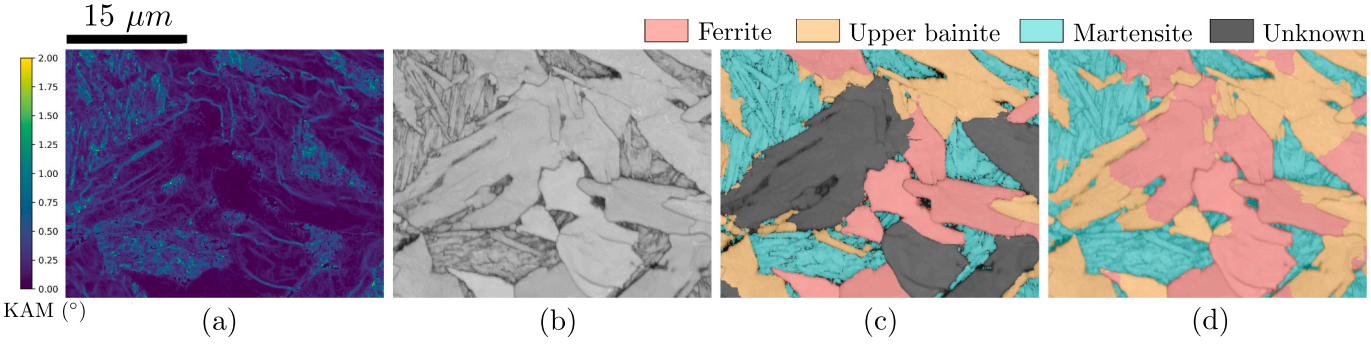

Fig. 8. Prediction of the model on grains labeled as 'unknown' on an API-2 EBSD map. (a) KAM map, (b) BC map, (c) Labels and (d) Predictions.

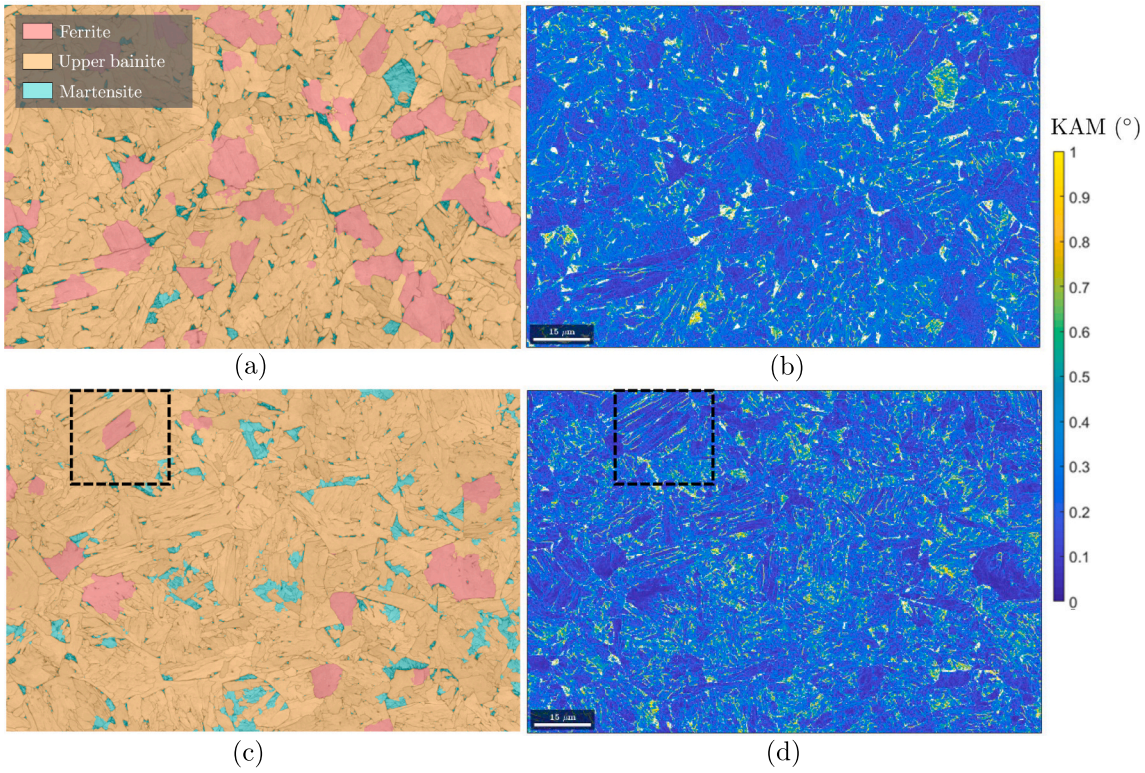

Fig. 9. (a) and (c) Model predictions on two unlabelled API5L x65QS EBSD maps cooled at 8.3 °C s^{-1} and 19.4 °C s^{-1} respectively. (b) and (d) Associated KAM maps. The dotted frame highlights coarse bainite laths misclassified as ferrite as they contained few KAM variations.

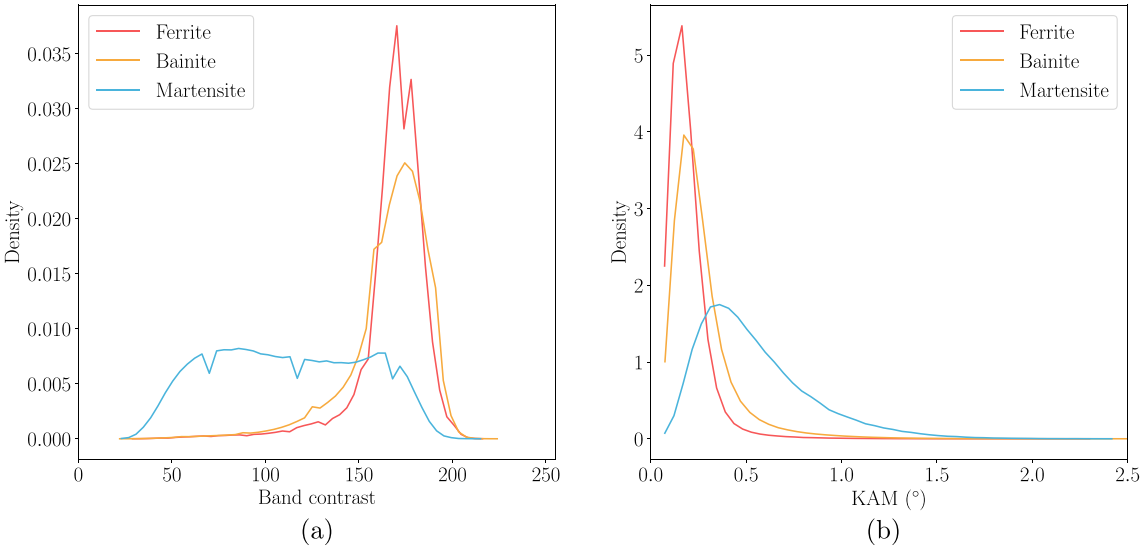

Fig. 10. Distribution of the (a) BC and (b) KAM values over all the unlabeled samples for each predicted constituent.

Table 4
Predicted constituent fraction evolution for samples transformed with different cooling rates.

Sample	Heat treatment	Ferrite (%)	Upper bainite (%)	Martensite (%)
API-1	2.8 °C s^{-1}	50.0	45.9	4.1
API-3	8.3 °C s^{-1}	26.2	70.1	3.7
API-4	19.4 °C s^{-1}	3.9	89.5	6.6

API-2 contained only 21% of martensite and as martensite was only present in M/A island on the API-1 and SA-1 samples.

Table 5 presents the testing accuracy of two models trained either with or without API-5 sample crops. The accuracy on the SA-1, API-1 and API-2 test set was not impacted.

However, the behavior of those models on maps with a high martensite content was different. Fig. 12 presents the prediction of the model trained without (a) and with (b) the API-5 map, on the API-5 test set. The model trained without any quenched sample maps predicted a higher bainite content in the quenched sample. The variety of the data provided to the model therefore maters as much as its quantity.

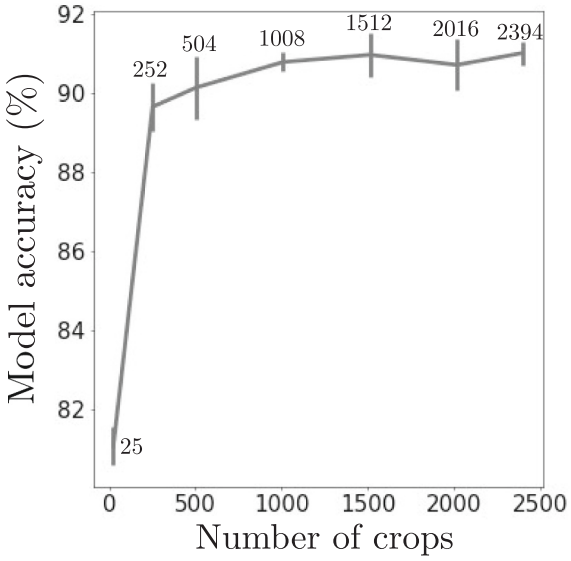

Fig. 11. Evolution of the testing accuracy with the number of 224 × 224 crops used for training. The number of crops used is displayed next to each evaluation point for better readability.

Table 5
Model test accuracy with and without the API-5 sample for training.

Sample	With API-5	Without API-5
SA-1	94.6	93.8
API-1	87.2	87.4
API-2	86.6	86.2
API-5	99.9	69.2

3.3.2. *Data acquisition conditions*

Maps were obtained on API-1 with two diffraction pattern qualities (Q_{ref} and Q_{alt}) to investigate the influence of acquisition conditions on the model performances. Fig. 13 shows the KAM distribution for each constituent on the maps obtained with the two conditions. A slightly higher misorientation contrast can be observed between the constituents on the Q_{alt} map than on the Q_{ref} map.

Table 6 compares the accuracies of two models trained with and without the map obtained with acquisition condition Q_{alt}. A 2% difference in accuracy can be observed between the API-1 (Q_{ref}) and (Q_{alt}) maps. However, the accuracy variability for a tested map, when training two different models is also around 2%, due to random weight initialization and to the stochastic nature of the adadelta algorithm. Therefore, the model reached similar classification accuracies on both maps, even when the higher quality map (Q_{alt}) was not used during training.

The influence of the acquisition step was also investigated using three training strategies. Fig. 14 (c) shows the prediction of the model on API-2 maps obtained using our reference strategy described in Sections 2.4 and 2.5. To better show the effect of the additional step channel, another model was trained on 0.1 μm and 0.3 μm step maps, with only the BC and KAM maps as input, without providing the acquisition step. The predicted segmentations of this second model for different acquisition steps are shown in Fig. 14 (b). Finally, Fig. 14 (a) shows the predictions of a model trained without the additional step channel and providing only 0.1 μm step maps to the model during training.

Fig. 14 (a) reveals that, the larger the step, the lower this latter model accuracy is over the whole API-2 map. In particular, with increasing step size, ferrite is completely misclassified as bainite. This results from the model being fed with a different resolution than that used for training. The model in Fig. 14 (a) was only provided with ferrite grains presenting low KAM values and few intragranular low-angle grain boundary during training. As the step size increases, the KAM values globally increases. Also sub-grain boundaries are artificially widen, as there are less pixels on the maps. This model could therefore confuse ferrite grains with bainite as the density of sub-grain boundaries appears greater. This results in an overestimation of bainite at the expense of ferrite for the 0.2 μm step. Note that the martensite prediction remains relatively constant for the different steps as the model may rely on the BC value to discriminate this constituent from the others, which does not vary much with the step size.

As shown in Fig. 14 (b), the model trained without the additional channel but with data obtained with 0.1 μm and 0.3 μm steps generalizes better to different scales. Nevertheless, it does not seem to interpolate well between the steps provided during training as the accuracy reached for 0.2 μm was only 66.4% over the whole API-2 test map. This could again result from the features learnt by the model to discriminate ferrite from bainite changing as the step size increases. The model was not provided with examples acquired at 0.2 μm. More importantly, the model does not have any information on the context with which the data were acquired, *e.g.* the step size. It therefore cannot adapt the features learnt during training to those observed for a new step size.

Finally, Fig. 14 (c) shows that the reference model, trained with the additional step channel, adapts better its criteria to the scale and provides consistent results over different scales. The step channel model reaches higher accuracy than the single scale model for the 0.2 μm and 0.4 μm scale, which were not shown to the models at training time. This reveals that the step channel model better interpolates and extrapolates to acquisition steps not seen during training. This results from the model being provided with more context related to the acquisition. However lower accuracies are still reached by the model at those scales when compared to that provided during training. This could result from the

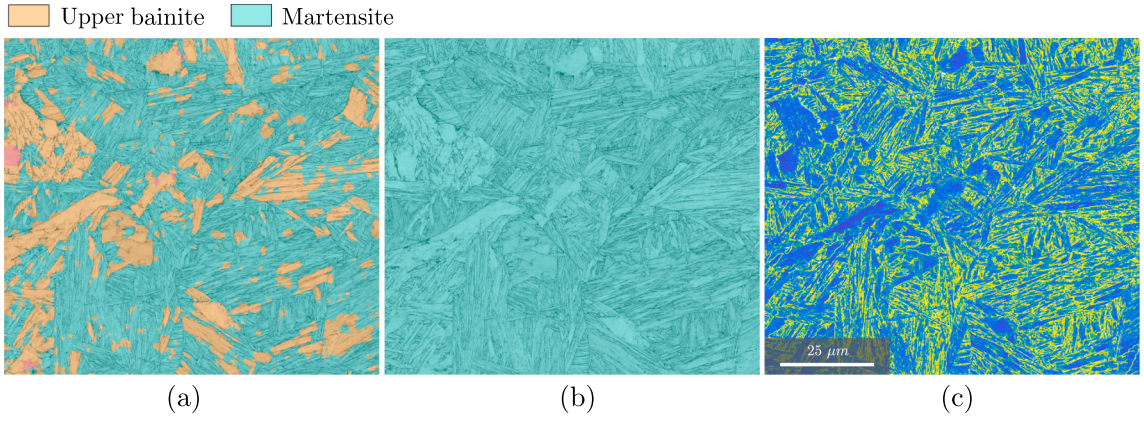

Fig. 12. Model prediction on an API-5 EBSD map with a model trained (a) without and (b) with API-5 EBSD maps examples and (c) the associated KAM map.

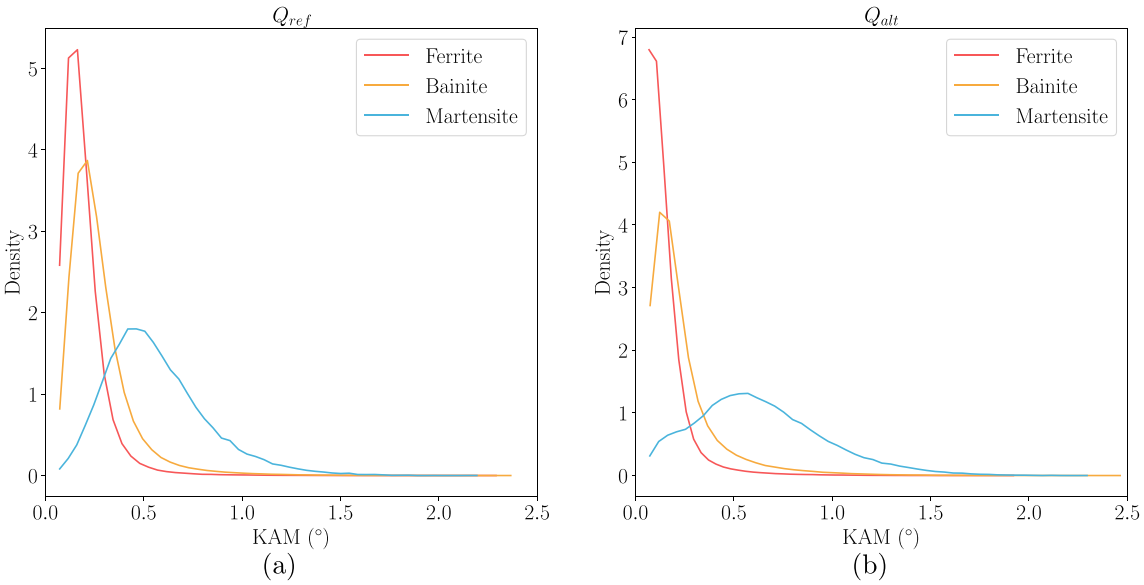

Fig. 13. KAM distribution for each of the labeled constituents on the EBSD map acquired with conditions (a) Q_{ref} and (b) Q_{alt}.

Table 6
Testing accuracy of the models trained with and without the map acquired with Q_{alt}.

Sample	Training with Q_{alt}	Training without Q_{alt}
SA-1	94.6	95.8
API-1 (Q_{alt})	87.6	87.6
API-1 (Q_{ref})	86.8	85.6
API-2	86.6	85.4

fact that the model was only trained with two different scales (0.1 μm and 0.3 μm).

3.3.3. Classification accuracy according to the pixel location in the map

Additionally the influence of the grain location in the map (in the center or at the edges) on the model accuracy, was also investigated. Indeed grains at the border of the crop may be cut out. The size, intragranular misorientation content or the neighbouring of those grains are therefore inaccessible which makes their constituent difficult to assess, even during manual labeling.

Fig. 15 shows the prediction of the model on API-2 crops of different sizes and the same central position. The figure reveals that the network predictions are rather consistent for the different crop sizes, especially for grains at the middle of the crop. Some grains, such as the one highlighted in frames 1 and 2 are predicted differently for the different crop sizes. In Fig. 15 (d), only a third of the grain in frame 1 can be seen. It is therefore misclassified as ferrite. The grain is nevertheless rightly classified in larger crops where it is fully represented.

The grain in frame 2 is partially misclassified as ferrite, even in Fig. 15 (c) where it is fully present in the crop. Fig. 15 (b) however shows that this grain is rightly classified as bainite, once it is situated in the middle of the crop. This results from the various pooling and padding operations performed in the different layers of the network. Due to such operations, the missing information is replaced by the nearest known pixel (padding 'same'), which amount to change the morphology of the cropped features as well as the gradient of information. The pixel position in the crop has therefore some importance for constituent fraction estimation using a UNET model.

This effect is however minimized here for two reasons. First, the model was trained on a large number of small crops extracted from larger maps. The grains at the border of those crops were therefore rightly labeled. The model could have learned to adapt its decision parameters depending on the position of the grain on the crop.

Also, the influence of padding is the strongest on the deepest layers of the network, for which only low level features are detected. This border effect should therefore not impact too much the detection of high level features (such as the presence of low-angle boundaries) and their use in the constituent prediction.

4. Discussion

The objective of this work was to extend the EBSD data-based model developed by Martinez Ostormujof et al. [24] to the segmentation of upper bainite, ferrite and martensite in a low-carbon steel. Fig. 3 and Table 3 show that the model accurately differentiates those three constituents. This model is the first in the literature to perform pixel-wise segmentation of those three constituents with such accuracy, on EBSD data. Using a CNN architecture also enables to obtain a prediction in a short amount of time. One prediction on a 2520×945 pixels map takes approximately 35 s on an Intel Core i5-8250U CPU, 1.60GHz, if the KAM map is already computed.

Many works distinguished bainite from martensite, pearlite and ferrite on either SEM or LOM pictures [1,3,5–8] or EBSD data [20,30]. Few authors made the distinction between upper bainite and ferrite [3,7]. However, most of these models were only assigning a single class to a whole EBSD map and were not performing a segmentation of the analyzed microstructure.

Azimi et al. [1] performed pixel-wise segmentation of bainite using LOM pictures. However, they only obtained a 37.32% pixel-wise accuracy for the bainite class which is below the 92% bainite class accuracy evidenced in Fig. 6. Also, as the presented U-Net model uses EBSD data as input, it does not rely on an etching procedure, which does not always reveal all subgrain boundaries and could induce errors in the ground truth.

The most challenging task for developing this model was to manually label the EBSD maps, as the difference between ferrite and upper bainite is often ambiguous. Introducing the 'unknown' class, ignored by the model during training, eased this process by allowing the expert to exclude ambiguous cases from the training.

The model learned itself which criteria to use to differentiate the three constituents. The benefit of using such model is that these criteria will remain constant, which is not the case for human labeling. One could wonder about their physical relevance, as the model was trained on maps labeled mainly by a single expert. Future works should

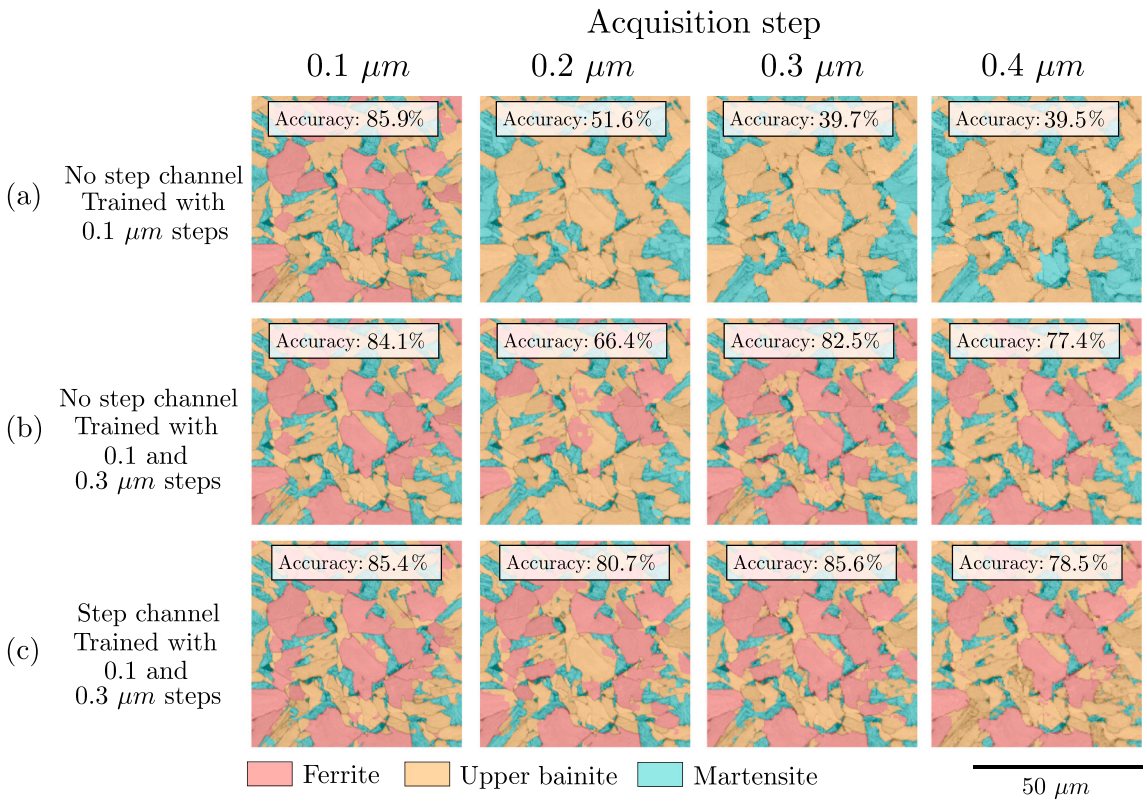

Fig. 14. Comparison of predictions on a API-2 maps crop with three different models, for an acquisition step of 0.1 μm, 0.2 μm, 0.3 μm and 0.4 μm. (a,b) Models using only the KAM and BC as input, trained with (a) 0.1 μm steps or (b) 0.1 μm and 0.3 μm steps. (c) Model using an additional input containing the acquisition step and trained with 0.1 μm and 0.3 μm steps.

therefore focus on aggregating different experts labeling to train the model with data representative of the material science community.

As shown in Section 3.2, the model generalizes well to microstructure presenting the same constituents than those seen during training but with different fractions. Neural network learns which features of an image is discriminating enough to assess a pixel's class with the highest accuracy. As shown recently by Durmaz et al. [31] using secondary electron picture, such features could be the presence of subgrain boundaries, carbides or the grain shape. In the present case, the model could alternatively rely on the presence of sub-grain low-angular boundaries, the grain shape, the BC value. The model prediction should therefore remain valid, as long as the studied constituents present such similar features, which could not be the case on a different steel grade.

As part of the main industrial study, the model was tested on a different steel grade presenting similar constituents and maps acquired with another camera. A lower accuracy was obtained with the as-trained model. However, adding only a single map of the new steel grade to the training set was sufficient to reach reasonable accuracies on these new samples. Providing the model with various examples of the three constituents obtained on different steel grades would therefore help its generalization, as suggested by Section 3.3.1.

As revealed by Table 5, the model does not seem to be sensitive to the acquisition conditions, provided that they are sufficiently good to evidence a contrast between the different constituents on the model input. Machine learning model robustness over different acquisition conditions was recently addressed by few authors for microstructures segmentation [21–23,32]. Figs. 14 (a) and (c) however show that CNNs do not always generalize well to different acquisition step if no contextual data were provided to the model, especially if maps acquired with only a single step are used for training.

Larmuseau et al. [22] provided the microscope magnification as an input of a triplet-network trained to classify microstructure crops into different constituent classes. In their case, the magnification was concatenated to a fully connected layer at the end of the network. Such approach was however not compatible with the U-Net segmentation model used in this study, as it is mostly composed of convolutional layers.

Fig. 14 (a) shows that when provided with the acquisition step as a third input channel, the model better learns to adapt its classification criteria to different resolutions. The developed model can therefore be applied to a wide variety of acquisition conditions. Such approach is similar to that performed by Facil et al. [33] to generalize a CNN prediction to different cameras. To the author knowledge, explicitely providing acquisition parameters as an input to a UNET model had however never been applied to the segmentation of microstructures.

5. Conclusion

The main objective of this work was to develop a machine learning model able to leverage EBSD data for the segmentation of different transformation products in low-carbon steels. The main contributions are as follows:

- A U-Net model was developed for the semantic segmentation of ferrite, bainite and martensite using EBSD data
- The model has an overall test accuracy of 92% and an accuracy per sample that varies between 86.6% to 95% depending on its microstructure complexity
- Providing the acquisition step as a model input helped it to generalize well to different acquisition steps.
- The model also generalizes well to microstructures with constituents presenting the same features than those seen during training, for the studied steel grades, and presenting various fractions of ferrite, bainite and martensite.

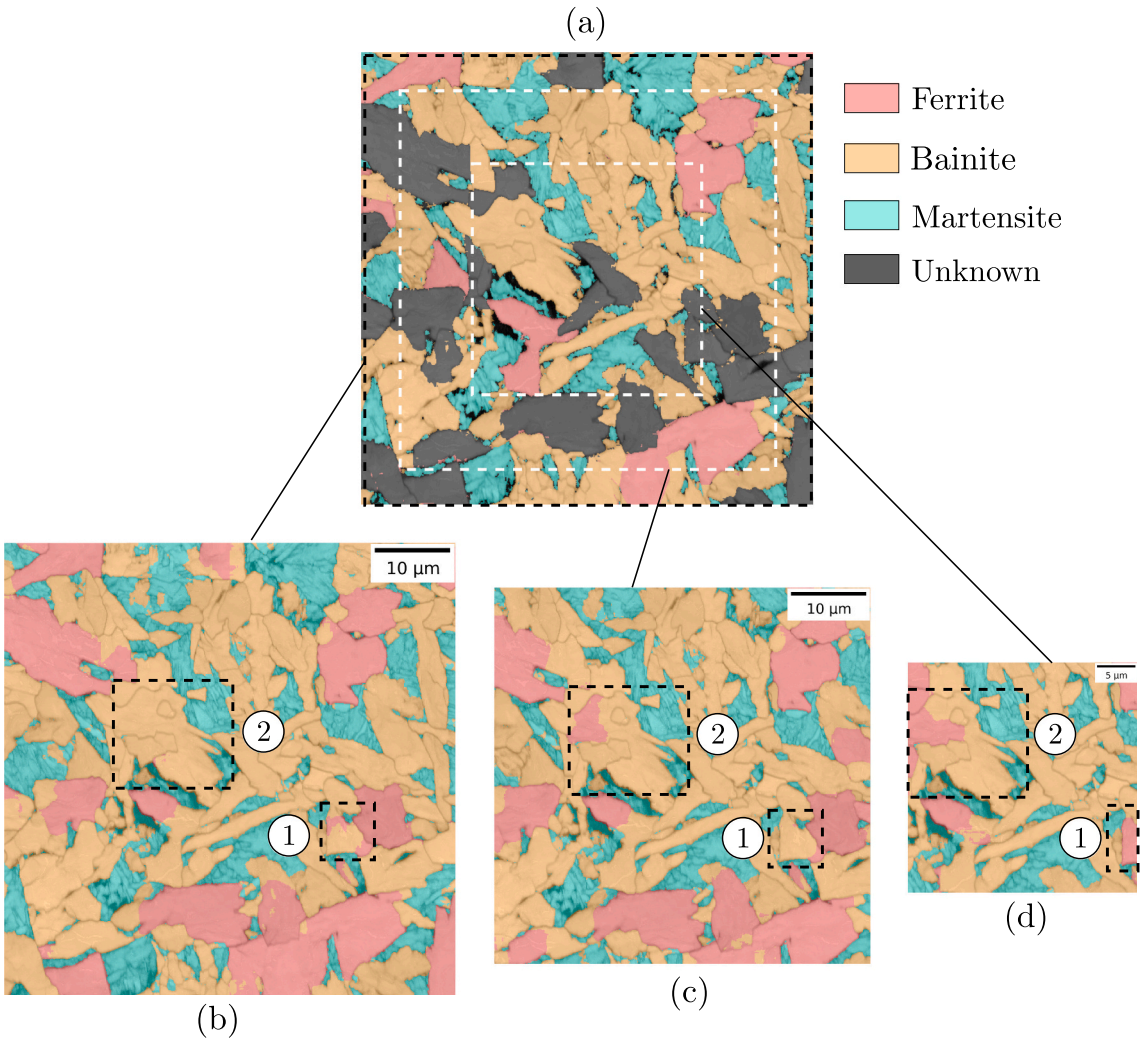

Fig. 15. Model predictions on an API-2 crop with different definitions (a) Labels (b) 300×300 pixels, (c) 400×400 pixels (d) 600×600 pixels.

The labeling step represents the main challenge to overcome, for the development of a more general model, that could differentiate different types of bainite (upper, lower, granular). This could be performed using unsupervised learning approaches that would learn to extract the statistical tendencies of a dataset.

Alternatively, the labels performed by many different experts could be aggregated. This would reduce the subjectivity of the manual process and reduce the impact of potential mistakes. Future works will focus on these approaches.

Data availability

The data required to reproduce these findings were made available [34].

Declaration of Competing Interest

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

Acknowledgement

The present work is part of a study led by LEM3 and IRT-M2P through the SERAPHIA project for the development of automated constituent segmentation in complex steel microstructures using deep learning models. It was supported by the French government's Plan d'Investissement d'Avenir (PIA) and INDUSTEEL ARCELORMITTAL.

References

[1] Seyed Majid Azimi, Dominik Britz, Michael Engstler, Mario Fritz, and Frank Mücklich. Advanced steel microstructural classification by deep learning methods. 8(1):2128. doi:https://doi.org/10.1038/s41598-018-20037-5.

[2] Brian L. DeCost, Toby Francis, and Elizabeth A. Holm. Exploring the microstructure manifold: Image texture representations applied to ultrahigh carbon steel microstructures. 133:30–40. doi:https://doi.org/10.1016/j.actamat.2017.05.014.

[3] Jessica Gola, Johannes Webel, Dominik Britz, Agustina Guitar, Thorsten Staudt, Marc Winter, and Frank Mücklich. Objective microstructure classification by support vector machine (SVM) using a combination of morphological parameters and textural features for low carbon steels. 160:186–196. doi:https://doi.org/10.1016/j.commatsci.2019.01.006.

[4] Michiel Larmuseau, Michael Sluydts, Koenraad Theuwissen, Lode Duprez, Tom Dhaene, and Stefaan Cottenier. Compact representations of microstructure images using triplet networks. 6(1):156–167. doi:https://doi.org/10.1038/s41524-020-00423-2.

[5] Tatsuya Maemura, Hidenori Terasaki, Kazumasa Tsutsui, Kyohei Uto, Shogo Hiramatsu, Kotaro Hayashi, Koji Moriguchi, and Shigekazu Morito. Interpretability of deep learning classification for low-carbon steel microstructures. 61(8):1584–1592. doi:https://doi.org/10.2320/matertrans.MT-M2020131.

[6] Martin Müller, Dominik Britz, Laura Ulrich, Thorsten Staudt, and Frank Mücklich. Classification of bainitic structures using textural parameters and machine learning techniques. 10(5):630. doi:https://doi.org/10.3390/met10050630.

[7] Angshuman Paul, Abhinandan Gangopadhyay, Appa Rao Chintha, Dipti Prasad Mukherjee, Prasun Das, and Saurabh Kundu. Calculation of phase fraction in steel microstructure images using random forest classifier. 12(8):1370–1377. doi:https://doi.org/10.1049/iet-ipr.2017.1154.

[8] Kazumasa Tsutsui, Hidenori Terasaki, Kyohei Uto, Tatsuya Maemura, Shogo Hiramatsu, Kotaro Hayashi, Koji Moriguchi, and Shigekazu Morito. A methodology of steel microstructure recognition using SEM images by machine learning based on textural analysis. 25:101514–101527. doi:https://doi.org/10.1016/j.mtcomm.2020.101514.

[9] Kangying Zhu, David Barbier, and Thierry Iung. Characterization and quantification methods of complex BCC matrix microstructures in advanced high strength steels. 48(1):413–423. doi:https://doi.org/10.1007/s10853-012-6756-9.

[10] Raphael Fialho Tomaz, Dagoberto Brandão Santos, Kenji Camey, Ronaldo Barbosa, Margareth Spangler Andrade, and Diana Pérez Escobar. Complex phase quantification methodology using electron backscatter diffraction (EBSD) on low manganese high temperature processed steel (HTP) microalloyed steel. 8(2):2423–2431. doi:https://doi.org/10.1016/j.jmrt.2019.01.021.

[11] S Zajac and V Schwinn. Characterisation and quantification of complex bainitic microstructures in high and ultra-high strength linepipe steels. 500–501:387–394. doi:https://doi.org/10.4028/www.scientific.net/MSF.500-501.387.

[12] Sachin L. Shrestha, Andrew J. Breen, Patrick Trimby, Gwénaëlle Proust, Simon P. Ringer, and Julie M. Cairney. An automated method of quantifying ferrite microstructures using electron backscatter diffraction (EBSD) data. 137:40–47. doi:https://doi.org/10.1016/j.ultramic.2013.11.003.

[13] H. Zhao, B. P. Wynne, and E. J. Palmiere. A phase quantification method based on EBSD data for a continuously cooled microalloyed steel. 123:339–348. doi:https://doi.org/10.1016/j.matchar.2016.11.024.

[14] Naoki Takayama, Goro Miyamoto, and Tadashi Furuhara. Effects of transformation temperature on variant grouping of bainitic ferrite in low carbon steel. 172–174:155–160. doi:https://doi.org/10.4028/www.scientific.net/SSP.172-174.155.

[15] L. Germain, N. Gey, R. Mercier, P. Blaineau, and M. Humbert. An advanced approach to reconstructing parent orientation maps in the case of approximate orientation relations: application to steels. 60(11):4551–4562. doi:https://doi.org/10.1016/j.actamat.2012.04.034.

[16] Jun-Yun Kang, Seong-Jun Park, and Man-Been Moon. Phase analysis on dual-phase steel using band slope of electron backscatter diffraction pattern. 19:13–16. doi:https://doi.org/10.1017/S1431927613012233.

[17] S. Zaefferer, P. Romano, and F. Friedel. EBSD as a tool to identify and quantify bainite and ferrite in low-alloyed al-TRIP steels. 230(3):499–508. doi:https://doi.org/10.1111/j.1365-2818.2008.02010.x.

[18] Yi Han, R. Joey Griffiths, Hang Z. Yu, and Yunhui Zhu. Quantitative microstructure analysis for solid-state metal additive manufacturing *via* deep learning. 35(15):1936–1948. doi:https://doi.org/10.1557/jmr.2020.120.

[19] Sung Wook Kim, Seong-Hoon Kang, Se-Jong Kim, and Seungchul Lee. Estimating the phase volume fraction of multi-phase steel via unsupervised deep learning. 11(1):5902. Number: 1 Publisher: Nature Publishing Group. doi:https://doi.org/10.1038/s41598-021-85407-y.

[20] Kazumasa Tsutsui, Hidenori Terasaki, Tatsuya Maemura, Kotaro Hayashi, Koji Moriguchi, and Shigekazu Morito. Microstructural diagram for steel based on crystallography with machine learning. 159:403–411. doi:https://doi.org/10.1016/j.commatsci.2018.12.003.

[21] Kevin Kaufmann and Kenneth S. Vecchio. An acquisition parameter study for machine-learning-enabled electron backscatter diffraction. 27(4):776–793. doi:https://doi.org/10.1017/S1431927621000556.

[22] Michiel Larmuseau, Michael Sluydts, Koenraad Theuwissen, Lode Duprez, Tom Dhaene, and Stefaan Cottenier. Race against the machine: can deep learning recognize microstructures as well as the trained human eye? 193:33–37. doi:https://doi.org/10.1016/j.scriptamat.2020.10.026.

[23] Xiaoting Zhong, Brian Gallagher, Keenan Eves, Emily Robertson, T. Nathan Mundhenk, and T. Yong-Jin Han. A study of real-world micrograph data quality and machine learning model robustness. 7(1):1–11. doi:https://doi.org/10.1038/s41524-021-00616-3.

[24] Tomas Martinez Ostormujof, RRP. Purushottam Raj Purohit, Simon Breumier, Nathalie Gey, Mathieu Salib, and Lionel Germain. Deep learning for automated identification of steels transformation products in EBSD maps. a case of study in dual phase microstructures. 184:111638–111649. doi:https://doi.org/10.1016/j.matchar.2021.111638.

[25] H. K. D. H. Bhadeshia. Bainite in Steels: Theory and Practice. Maney Publishing, third edition.

[26] N. Takayama, G. Miyamoto, and T. Furuhara. Effects of transformation temperature on variant pairing of bainitic ferrite in low carbon steel. 60(5):2387–2396. doi:https://doi.org/10.1016/j.actamat.2011.12.018.

[27] F. Bachmann, Ralf Hielscher, and Helmut Schaeben. Texture analysis with MTEX – free and open source software toolbox. 160:63–68. doi:https://doi.org/10.4028/www.scientific.net/SSP.160.63.

[28] Francois Chollet et al. Keras. https://github.com/fchollet/keras.

[29] Masayuki Kamaya. Assessment of local deformation using EBSD: Quantification of accuracy of measurement and definition of local gradient. 111(8):1189–1199. doi:https://doi.org/10.1016/j.ultramic.2011.02.004.

[30] Wonjik Kim, Asako Kanezaki, and Masayuki Tanaka. Unsupervised learning of image segmentation based on differentiable feature clustering. 29:8055–8068. doi:https://doi.org/10.1109/TIP.2020.3011269.

[31] Ali Riza Durmaz, Martin Müller, Bo Lei, Akhil Thomas, Dominik Britz, Elizabeth A. Holm, Chris Eberl, Frank Mücklich, and Peter Gumbsch. A deep learning approach for complex microstructure inference. 12(1):6272. doi:https://doi.org/10.1038/s41467-021-26565-5.

[32] Chunguang Shen, Chenchong Wang, Minghao Huang, Ning Xu, Sybrand van der Zwaag, and Wei Xu. A generic high-throughput microstructure classification and quantification method for regular SEM images of complex steel microstructures combining EBSD labeling and deep learning. 93:191–204. doi:https://doi.org/10.1016/j.jmst.2021.04.009.

[33] Jose M. Facil, Benjamin Ummenhofer, Huizhong Zhou, Luis Montesano, Thomas Brox, and Javier Civera. CAM-Convs: Camera-Aware Multi-Scale Convolutions for Single-View Depth. p. 11818–11827. doi:arXiv:1904.02028.

[34] S. Breumier, T. Martinez Ostormujof, B. Frincu, A. Couturier, N. Loukachenko, P.E. Aba-perea, and L. Germain. EBSD Maps Acquired on Different Steels with Transformed Microstructures. doi:10.12763/RHWHD7.