SCIENTIFIC REPORTS

OPEN

Pattern recognition with machine learning on optical microscopy images of typical metallurgical microstructures

Dmitry S. Bulgarevich[1,2], Susumu Tsukamoto[1], Tadashi Kasuya[3], Masahiko Demura[1] & Makoto Watanabe[1,3]

For advanced materials characterization, a novel and extremely effective approach of pattern recognition in optical microscopic images of steels is demonstrated. It is based on fast Random Forest statistical algorithm of machine learning for reliable and automated segmentation of typical steel microstructures. Their percentage and location areas excellently agreed between machine learning and manual examination results. The accurate microstructure pattern recognition/segmentation technique in combination with other suitable mathematical methods of image processing and analysis can help to handle the large volumes of image data in a short time for quality control and for the quest of new steels with desirable properties.

Depending on cooling rate of steels, the ferrite (F), pearlite (P), bainite (B), and martensite (M) microstructures could be formed due to the displacive and reconstructive transformations of austenite (A) crystal structure, which are accompanied with cementite precipitation at different diffusion rates (except for M)[1]. This is shown in Fig. 1(a) with continuous cooling transformation (CCT) diagrams[2]. For example, the various combinations of such micro-structures could be formed and observed at different microscope magnifications in a weld heat-affected zone (HAZ). The appearance and additional nomenclature for some of them related to present work are shown in Fig. 1(b), which are not limited to welding process only. Here it must be stressed that Fig. 1(b) is for summary only since not all different microstructures could be observed at once in particular steel as can be seen from Fig. 1(a).

It is well known that the volume fraction, the dimension, and the morphology of these microstructure phases are greatly responsible for the mechanical properties of metallurgical materials[3–5]. For instance, the sensitivity to weld-cold cracking is in order of $M > B > F$ volume fractions. So far, however, such microstructural information was quantified manually for industrial or scientific purposes. Obviously, this was/is a very time-consuming process even for skilful personal. In addition, it must be noted that with the manual quantification, it is virtually impossible to obtain the statistical information on phases having the complex spatial distributions.

The outline of complexity of the problem is depicted in Fig. 2(a,b), where the optical microscopy images of steel microstructures with same chemical composition, but cooled at 1 and 3 °C/s are shown. In Fig. 2(a), the P, ferrite side plate (F_{sp}), and grain boundary allotriomorphic ferrite (F_{all}) could be detected compared to additional B phase in Fig. 2(b). These differences are manifested in changes of their macroscopic mechanical properties, i.e. the higher tensile and yield strengths for steel with B. In Fig. 2 and subsequent discussion, the fine colony P and M-A constituent are included into the F_{sp} since they could nucleate together in some cases between F_{sp}. In present samples, it is difficult to distinguish them by optical microscopy alone. In addition, the lamellar and degenerated P types from Fig. 1(b) are assigned to P. The lower B also cannot be seen with present optical magnification without scanning electron microscopy (SEM), so B corresponds to upper B in present work. If necessary to get more detailed information, the method of image analysis presented below can be equally applied to SEM images.

[1]Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki, 305-0047, Japan. [2]Research Center for Development of Far-Infrared Region, University of Fukui, Fukui, 3-9-1, Bunkyo, 910-8507, Japan. [3]School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan. Correspondence and requests for materials should be addressed to D.S.B. (email: bulgarevich@fir.u-fukui.ac.jp)
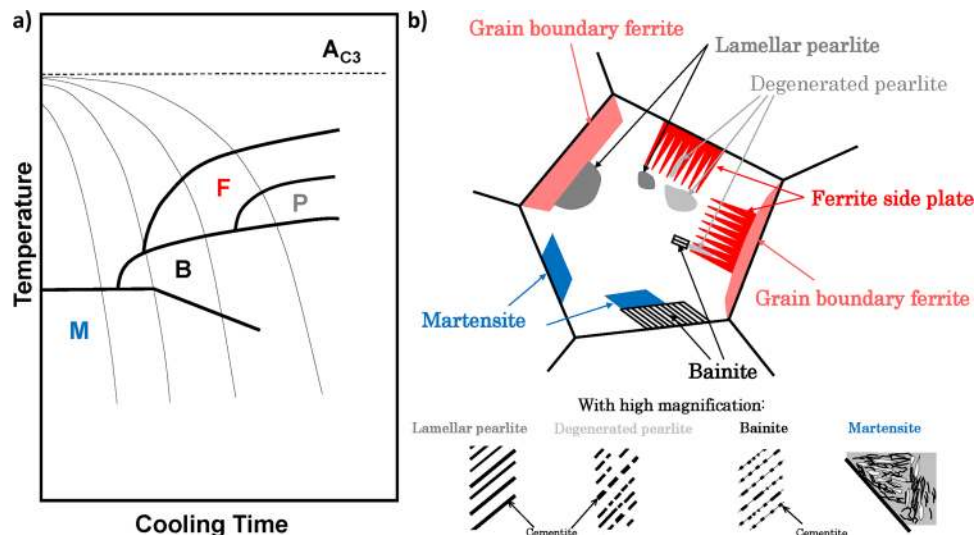
Figure 1. (**a**) The typical CCT diagrams with depicted regions of formed microstructures in low-carbon A-type structural steels. (**b**) The schematic drawing of different steel microstructures, which are formed in synthetic weld-heat affected zone during CCT process. See text for more details and abbreviations.
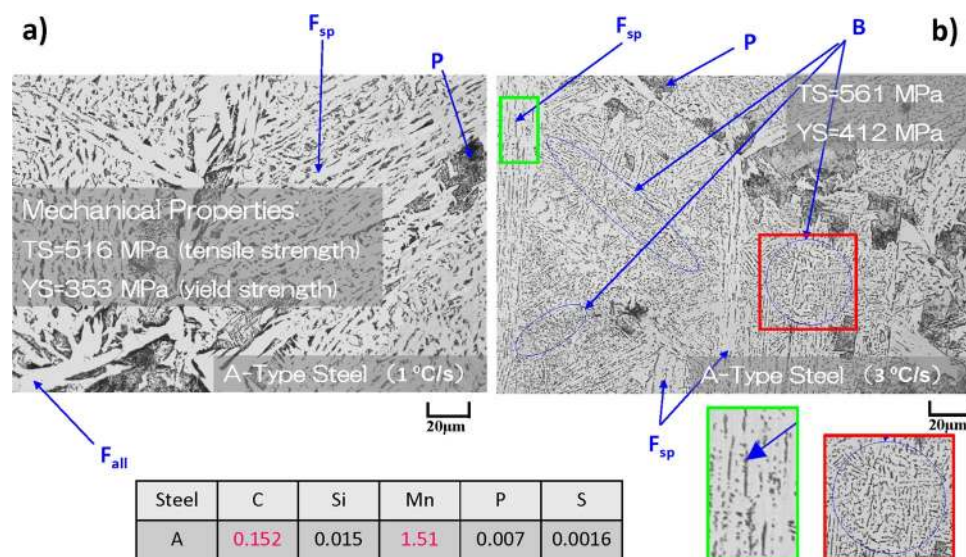

Figure 2. The examples of complex microstructure patterns observed in optical microscopy images of two A-steels with same composition, but obtained with different cooling rates of 1 °C/s (**a**) and 3 °C/s (**b**) from 1400 °C and having different mechanical properties. The manual identification of different microstructures is shown with arrows.

Typically, the volume fractions of the phases are estimated with linear analysis. In this method, the length of all intercepts for a certain phase is measured along the straight line drawn on sample image. If the line length is long enough, it is assumed that ratio of the summed length for particular phase to the total line length is equal to the volume fraction of this phase. In metallurgy field, there is a great demand to automate the image processing on huge data volumes, which are already available and continue to grow. In principle, on modern PC workstations and supercomputers, the sophisticated machine learning techniques[6] for pattern recognitions in combination with other suitable mathematical methods of image processing and analysis[7–9] can help to handle such volumes of image data in a short time for materials quality control, for establishing of microstructure/mechanical property relations, and for the quest on new materials with desirable characteristics, i.e. for the goals of materials informatics. For example, there has been a dramatic progress in automated analysis of medical images with different segmentation techniques[10–12].

However, it must be stressed that for microscopy images of metallurgical samples with complex microstructures (as in Fig. 2) we are not aware of reported image segmentations with accurate automatic techniques. In reported attempts, the different segmentation methods, algorithms, or microstructure descriptors were used or
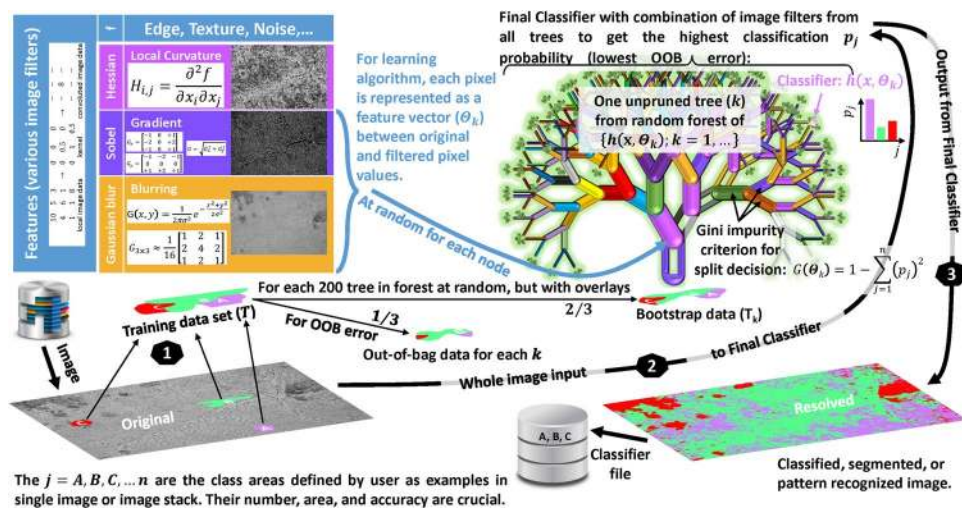
Figure 3. Scheme of image segmentation with machine learning Random Forest statistical algorithm (see text for more details).

derived, which were limited to particular problem, were not at the pixel level scale, or classification qualities were not examined in details[13–21]. Here it should be stressed that for industrial applications the automated phase segmentation quality must be in pair or better with manual analysis by experts. In this respect, the reported image segmentations by using the multilayer perceptron in backpropagation artificial neural network stands out, though the segmented microstructure types and their complexities were less demanding compared to present work[22–26]. The direct comparison of several machine learning classification techniques on image segmentation of graphite particles in metallurgical materials were also studied, but didn't include the methods used here[27].

Accordingly, we are presenting the application of several image segmentation methods, which are suitable for discussed challenges and requirements. The segmentation with machine learning could be a core in many practical protocols for analysis of metallurgical samples due to its versatility and accuracy, but not the single one as it will be demonstrated below.

Principle

In this regard, we had implemented the particular machine learning with Random Forest statistical algorithm[28–30] to the typical image analysis demands in metallurgy field. In fact, there are several reports of successful use of this algorithm in other fields such as biological and medical sciences[31–33]. In short, this method creates an appropriate sequences of applied image filters which can be used to assign each pixel in sample image to the particular class (j = A, B, C, …, n) with highest probability.

The basic outline of the method is shown in Fig. 3. It belongs to an ensemble-type supervised learning. As such, the examples of image pixels (areas) for different j (microstructures) are given by user to form the training data set (T) from single image or image stack. The bootstrap data (T_k), which are used to construct each decision tree (k), are chosen at random from T, but with overlays. In each bootstrapping process, approximately one-third of the T are left unused, which forms the out-of-bag (OOB) dataset for later utilization in estimation of the prediction error of machine-learning segmentation.

The individual k is grown by applying the randomly selected feature from image filter/kernel set (N_f) to the T_k, i.e. by transforming the image input vectors ($\mathbf{x}$) to feature vectors (Θ_k). After such filtering, at every split of the node (branches), the threshold-based yes/no split function is used to assign each pixel to particular j. The probabilities (p_j) of correct classification are calculated for each j by dividing the numbers of assigned and actual pixels in T_k. From calculated p_j for parent and child nodes, the Gini impurity criteria [$G(\Theta_k) = 1 - \sum_{j=A}^{n}(p_j)^2$] are estimated. The k is allowed to grow only if child node has lower $G(\Theta_k)$ value for some j. In other words, only the selected image filters and optimized split function parameters at each node, that lead to information gain (better classification), are selected for the k to grow until the predefined threshold for $G(\Theta_k)$ or number of branches is reached.

After all k are grown, the leaf nodes [$L(l_1, …, l_n)$] contain the probability distributions $p(j|L)$ of pixels for each j. These statistics will be used during the decision-making process at classification stage. The k-Classifier [$h(\mathbf{x}, \Theta_k)$] stores the information on Θ_k for each node, the optimised binary test for each split, and $p(j|L_k)$. The collection of such $h(\mathbf{x}, \Theta_k)$ from all k constitutes the Random Forest Classifier [$\{h(\mathbf{x}, \Theta_k); k = 1, …\}$]. To check its quality, the OOB data are used and the OOB error is calculated by [$\sum_k I(h(\mathbf{x}, \Theta_k) = j; \quad (y, \mathbf{x}) \notin T_k)/\sum_k I((y, \mathbf{x}) \notin T_k)$] with indicator function [$I(\cdot)$], i.e. by estimation of the proportion of times averaged over all cases that estimated j is not equal to the true class (y) in the OOB dataset.

At classification stage, the pixel that arrives to L_k after all filtering and binary tests stages has $p(j|L_k)$ for each k. To make a decision on j assignment for that pixel, each k casts a unit vote for the most popular class. These votes from all k cast a unit vote too. As a result, that pixel is finally classified to the particular j. If OOB error and the result of the image segmentation for the training data are satisfactory to the user, then such $\{h(\mathbf{x}, \Theta_k); k = 1, …\}$
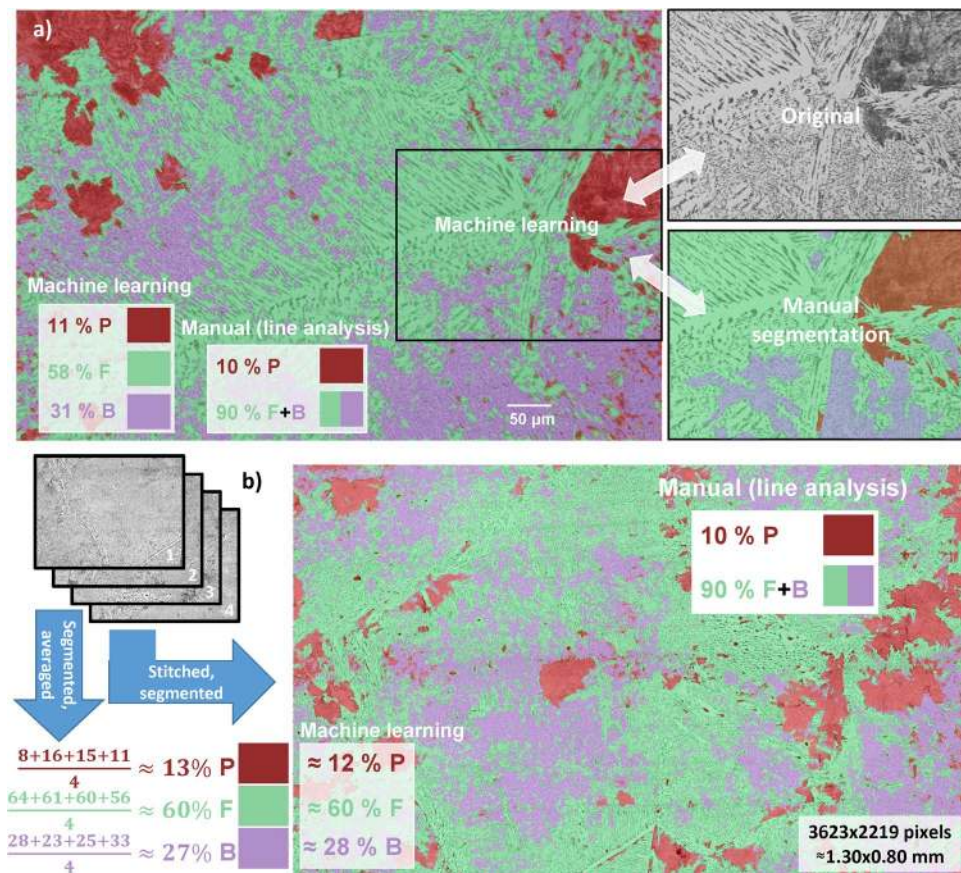
Figure 4. Examples of machine learning with Random Forest algorithm for automated pattern recognition in optical microscopy images of metallurgical samples. (**a**) The automated segmentation on single image is on the left. The small part of it and manual segmentation on this part are depicted in the right for comparison. (**b**) The application of the Random Forest Classifier on image stack or stitched image for accurate estimation of microstructure area/volume percentage from large imaging area.

can be applied to classify/segment the images of actual interest for statistical analysis. If not, the Classifier training process can be repeated by adding/revising of training data, image filters (edge, texture, noise, et. cet.), and tree/forest scales. Actually, each k is just a weak Classifier, but the forest of $\{h(\mathbf{x}, \Theta_k); k = 1, \ldots\}$ can be strong and unbiased thanks to an ensemble effect together with randomness in bootstrapping (de-correlating trees) and filter selections.

Though the expertise in steel microstructures is needed at the Classifier training step, but after that, even novice users can apply the well-trained Classifier for automatic segmentation of their image or image stack. This opens the possibilities to work with large image data volumes for materials informatics and materials integration goals across the fields. Such technique should also find the widespread use in industrial quality control and in research laboratories.

For appropriate image analysis of microstructures and from a viewpoint of practical industrial applicability, all samples in present work were prepared by a conventional polishing and subsequent etching[34] and observed by using an optical microscopy.

Results and Discussion

Segmentation with machine learning. Figure 4(a) shows the example of image segmentation with Random Forest algorithm on $F = F_{sp} + F_{all}$, B, and P microstructures for steel A from Fig. 2(b), but for different and larger area. The original image on the background is overlaid with segmented colour-coded areas for F, B, and P. The calculated phase percentages on inserts of Fig. 4(a) are also given for comparison with manual linear analysis of P and $F + B$ phases. The good agreement will be further validated below by using a much larger imaging area.

In addition, the part of the original image from Fig. 4(a) and its manual segmentation results are shown for comparison. Here, it should be stressed that manual area segmentation is not practical in real-world applications due to its significant labour and time demands. Nevertheless, it was confirmed that machine learning and expert-level manual segmentations are in excellent agreement with only minor inconsistencies in detecting of small microstructure areas.

Figure 4(b) shows the segmentation, which was performed on 3623×2219 pixel image (~1 mm²) by stitching[35] and slight cropping of four 1920×1440 pixel images with ~0.36 µm/pixel scale. The Classifier was created

CR K/s	Hv	V$_{manual}$ %					V$_{auto}$ %					Area mm^2
		P	F	B	F + B	M	P	F	B	F + B	M	
0.3		25	75		100		21(4)	79(4)		100		5.77
1		24	76		100		22(2)	78(2)		100		1.12
3		10	**	**	90		10(0)	58	32	90(0)		1.34
10	313			45*					43(2)		57	0.50
15	334			35*					31(4)		69	0.50

Table 1. Experimental conditions, mechanical properties and volume percentages of microstructure phases for several A-steels (see chemical composition in Fig. 2) measured by manual and automated image analysis. The CR, Hv, V$_{manual}$, V$_{auto}$, and Area are the cooling rate, Vickers hardness, volume percentage of phases by manual line analysis, volume percentage of phases by automated Random Forest segmentation, and analyzed image area for V$_{auto}$ estimation, respectively. The numbers in brackets are the errors between manual and automated analyses. *Calculated from empirical equation of B and M hardness[43]. **F and B were not separately counted.

on one of them and applied to others and to the stitched version. As demonstrated in Fig. 4(b), the accuracy of microstructure percentage estimation become comparable with manual linear analysis by averaging of segmentation results for image stack (see percentage deviations within the stack) or by using of stitched image with large imaging area. Both methods produce practically identical results. This demonstrates the good reproducibility and necessity to image the appropriate area size for accurate microstructure percentage estimation.

Table 1 summarises the comparison between manual (linear) and automated (Random Forest segmentation and empirical calculation) image analyses for several A-steels obtained at different cooling rates. Note that additional accurate segmentation on F and B areas was also achieved with machine learning compared to the manual line analysis (see Fig. 4 and Table 1). For the variety of the steel microstructures such as P + F, P + F + B, and B + M, the present automated segmentations show a very good agreement with manual analysis, which proofs the practical aspect of used segmentation method for various steels (see listed errors, i.e. the cross-validation by manual linear analysis).

In Fig. 4 the OOB error was ~1% with more than 70 features and class homogenization (balanced distribution of j pixels) used[30]. Due to the nature of the Random Forest, the larger N_f produces the lower OOB error[28]. For example, without class homogenization, the Classifier created by using original image in Fig. 4(a) with only Gaussian Blur, Hessian, Sobel, Difference of Gaussians, and Membrane Projection filters gave the OOB error of ~6%. By using all N_f, the OOB error decreased to ~2%. In above comparison, the same T (P = 16346, F = 86957, and B = 3850 pixels), k = 200, two random features per node, two decimal places for calculated precision, and filter parameter (standard deviation, etc.) ranges were used. However, the larger N_f increased the computation time for Classifier creation by ~5 times. Among used N_f, the Anisotropic Diffusion, Gabor, Kuwahara, and Bilateral filters were mostly accounted for the time difference. The runtime and OOB error also scale with k[33,36], but later one typically saturates for k > 100. At classification stage, with present ~250 MB Classifier file size, it takes ~4 min and ~10 GB additional RAM to segment the 2352 × 1568-pixel image on two-CPU Optiron 6128 workstation.

For industrial applications, when the computation time is important, the Classifier could be fine-tuned with particular T, N_f, filter parameters, k, and OOB error, which are suitable for the task of acceptable segmentation quality. In principle, the cross-validation is performed internally in Random Forest algorithm during the run by using the OOB dataset and lower OOB error is a good indicator of better Classifier quality. However, the main challenge is at Classifier application stage on datasets, which may differ strongly in image quality. In this regard, the development/use of appropriate cross-validation protocols by automated or human inspection with sampling analysis may still be needed. The major limitation for practical industrial applications is in supervised-learning nature of the Random Forest algorithm. Therefore, the best way is, first, to analyse the image data in terms of j types with some automated unsupervised-learning algorithm or statistical image analysis, and then to choose the appropriate Random Forest Classifier for this j combination from corresponding library. These are the interesting and important topics for future research and development.

Apart from general image quality and colour scale, the Classifier is sensitive to the image spatial scale, contrast, and j present. For example, it will be wrong to apply the Classifier for sample in Fig. 4(a) on Fig. 2(a) due to different μm/pixel scale and absence of B microstructure. From practical view point, the high numerical aperture (NA) microscope objective (NA = 0.8–0.9) could be used to collect the image dataset over ~1 mm^2 sample area. Typically, this will allow to create one Classifier for all phases present in particular sample zone of interest for statistical analysis of metallurgical material. As for the contrast variations, the image histograms should not differ much during picture collections and images should be run through the standardized brightness/contrast adjustments afterwards[37]. Nowadays, the automatic illumination/exposure control for consistently good image quality is a common feature with industrial optical microscopes. Nevertheless, more robust split functions could be also used or developed in this regard[38].

Segmentation with Euclidean distance conversion. To extract more information from Fig. 4(a), additional image processing and analysis techniques developed for different scientific fields could be also applied here. For example, the scheme of protocol with Euclidean distance (ED) conversion is shown in Fig. 5 for highlighting/segmenting of F$_{all}$. In this method, each foreground pixel (grey one) in original image is replaced with a colour value equal to that pixel's distance from the nearest background pixel (black one)[39,40].
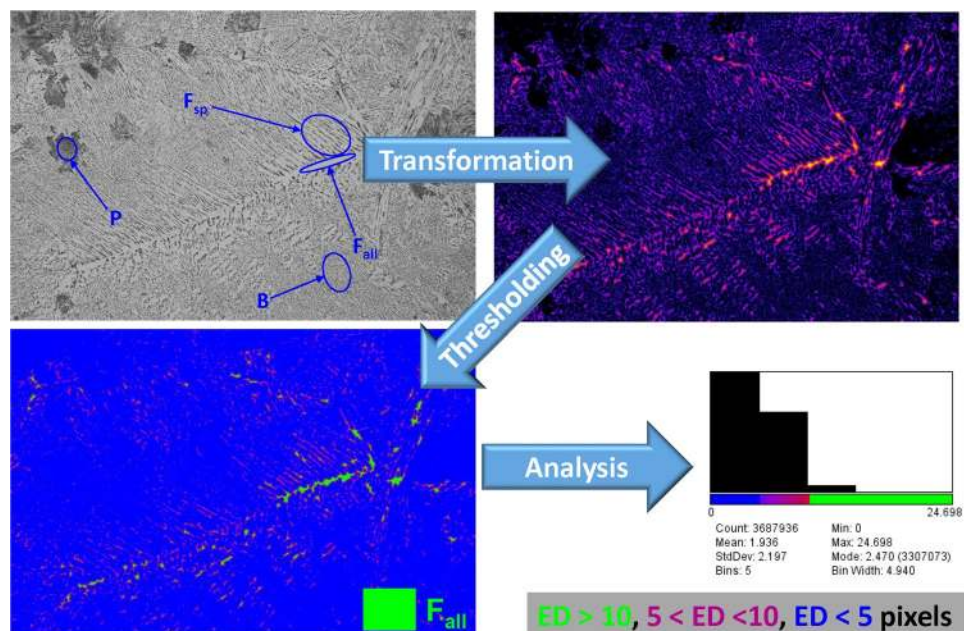
Figure 5. The image analysis protocol with Euclidean distance conversion technique for ferrite sub-phase segmentation in steels.
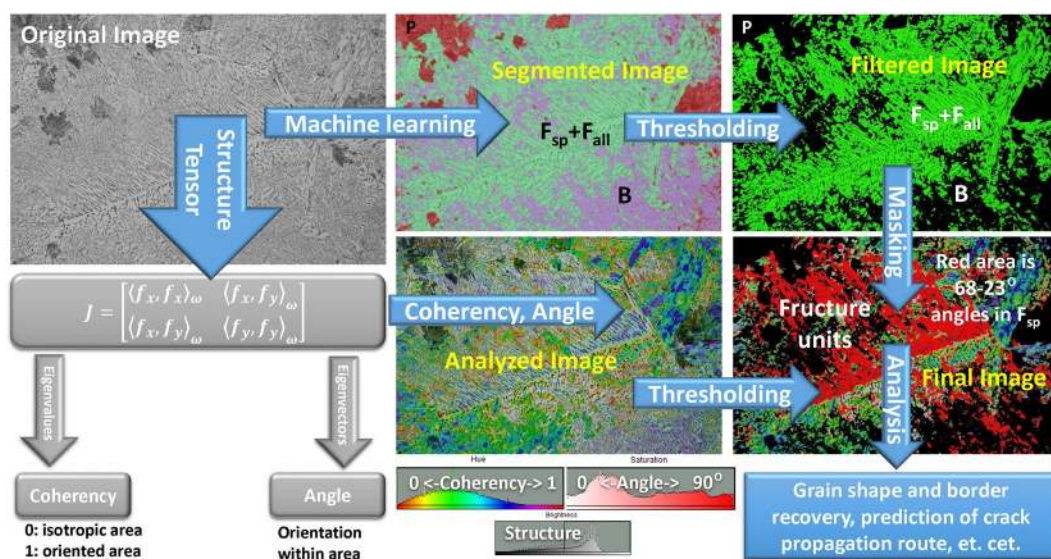

Figure 6. The image analysis protocol with structure tensor estimation for spatial anisotropy segmentation of fracture units in steel grains.

Because of higher nucleation temperature of F_{all} compared to F_{sp} and slow cooling rate[1], the diffusion distance of carbon from F_{all} to A becomes larger than the width of F_{sp} and it precipitates as cementite (black pixels) between F_{sp}. As a result, the F_{all} areas appears brighter in ED map (see Fig. 5), i.e. the F_{all} corresponds to wider areas between black pixels. Then, depending on user's criteria/thresholding on estimated ED map, the coloured areas could be assigned to the different types of F and studied quantitatively, for example, by additional segmentation of F_{all} from F_{sp}.

This analysis is important since plastic deformation of the soft F_{all} can relax the local stress concentration and could affect the crack initiation/propagation in steels[41]. As it is seen from Fig. 4(a), there are no clear boundaries between F_{all}, F_{sp}, and B. Therefore, the manual segmentation could be a challenge even for an expert. In this regard, the further correlations between mechanical properties and ED analysis results could help to find the mathematical criteria for correct segmentation of some phases.

Problems	Tools	Comments/Examples
Segmentation of metal microstructures	Expert-supervised Machine Learning by using Random Forest classification with a suitable set of image filters	Most versatile method to segment F, P, B, M, M-A, et. cet. of steel microstructures from image data (see Fig. 4). The segmentation quality is reasonably practical to obtain the statistics on the volume fraction of each phase.
Highlighting of particular microstructure	Euclidean distance conversion	Powerful image processing tool to segment, visualize, and analyze the different F sub-phases (see Fig. 5), which is a challenge even for an expert-level machine learning.
Spatial orientation/anisotropy of microstructures	Structure tensor extraction	Useful image processing tool to visualize, segment, and analyze the steel microstructures with anisotropic orientations of cementite or dislocations in F_{sp}, B, P, and M microstructures as well as of rolled steel grains (see Fig. 6).
Quantitative characterization of microstructures from large image data volumes	Combination of above tools with thresholding and various filtering techniques	It is important to create the classifier file library and protocols with above tools for the fast, reliable, and routine segmentations by common user (ongoing work, see Acknowledgments).

Table 2. Summary of the image segmentation methods for metal materials paradigm.

Segmentation with structure tensor extraction.   Figure 6 shows another example of protocol scheme for detecting the spatial anisotropy in images of steels. It utilizes the evaluation of the structure tensor (J) in a local neighbourhood, which is commonly used in the field of image processing and is further developed for biological applications[42]. In this method, the continuous spatial derivatives of cubic B-spline interpolation along X and Y axes for the input image are used in evaluations of coherency and orientation maps from J eigenvalues and eigenvectors, respectively.

For metallurgical samples, this analysis could highlight/segment the areas with isotropic (B and P) and anisotropic (F_{sp}) microstructure patters (depending on image scale and J Gaussian window). In combination with machine learning results of image segmentation in Fig. 4(a), the B and P areas are first masked out and then the dimensions of microfracture unit, which is responsible for cleavage fracture (F_{all} area oriented in same direction on Final Image in Fig. 6), could be highlighted. This was done by filtering of the microstructure orientation angles at 60° from grain boundaries. This angle corresponds to the cementite and dislocation growths during F_{sp} formation.

Since steel cracking is affected by grain geometry and rolled steels display spatial anisotropy of grain alignments, such fracture unit and grain visualizations could be used to correlate the macroscopic mechanical and microscopic structural properties in more details in order to predict the crack propagation routes.

Concluding Remarks

To our best knowledge, we are reporting for the first time the highly accurate, practical, and fast image segmentation with machine learning for microscopy images of metallurgical samples. This technique in combination with appropriate image processing/segmentation protocols, including image filtering, thresholding, and other mathematical/statistical methods, could revolutionize the image analysis in metallurgy field by creating the classifier/protocol libraries/databases for typical microstructure patterns. Table 2 summarizes the problems, methods, and applicability. In summary, we can conclude that automated expert-level analysis is getting to be quite feasible by using an appropriate method.

References

1. Bhadeshia, H. K. D. H. & Honeycombe, R. *Steels. Microstructure and Properties* (Elsevier, 2006).
2. Krauss, G. *Steels: Processing, Structure, and Performance* (ASM International, 2005).
3. Bramfitt, B. L. *Structure/Property Relationships in Irons and Steels, Metals Handbook Desk Edition* (ed. Davis, J. R.) 153–173 (ASM International, Second Edition, 1998).
4. Kumar, A., Singh, S. B. & Ray, K. K. Influence of bainite/martensite-content on the tensile properties of low carbon dual-phase steels. *Mater. Sci. & Eng. A* **474**, 270–282 (2008).
5. Zare, A. & Ekrami, A. Influence of martensite volume fraction on tensile properties of triple phase ferrite–bainite–martensite steels. *Mater. Sci. & Eng. A* **530**, 440–445 (2011).
6. Witten, I. H. & Frank, E. *Data mining: practical machine learning tools and techniques* (Morgan Kaufmann series in data management systems, Second Edition, 2005).
7. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).
8. Collins, T. J. ImageJ for microscopy. *BioTechniques* **43**, S25–S30 (2007).
9. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676–682 (2012).
10. Ma, Z., Tavares, J. M. R. S., Jorge, R. N. & Mascarenhas, T. A review of algorithms for medical image segmentation and their applications to the female pelvic cavity. *Comput. Methods Biomech. Biomed. Engin.* **13**, 235–246 (2010).
11. Oliveira, R. B. *et al.* Computational methods for the image segmentation of pigmented skin lesions: a review. *Comput. Methods Programs Biomed.* **131**, 127–141 (2016).
12. Jodas, D. S., Pereira, A. S. & Tavares, J. M. R. S. A review of computational methods applied for identification and quantification of atherosclerotic plaques in images. *Expert Syst. Appl.* **46**, 1–14 (2016).
13. Burikova, K. & Rosenberg, G. Quantification of Microstructural Parameter Ferritic-Martensite Dual Phase Steel by Image Analysis. *Metal 2009, Hradec nad Moravici, May* **19–21**, 185–189 (2009).
14. Dengiz, O., Smith, A. E. & Nettleship, I. Grain boundary detection in microstructure images using computational intelligence. *Comput. Ind.* **56**, 854–866 (2005).
15. Chmiela, J., Słota, D. & Szala, J. Multiscale description of the inhomogeneity of multiphase materials. *Mater. Charact.* **60**, 1145–1150 (2009).
16. Kim, D., Liu, J. J. & Han, C. Determination of steel quality based on discriminating textural feature selection. *Chem. Eng. Sci.* **66**, 6264–6271 (2011).
17. Dutta, S. *et al.* Characterization of micrographs and fractographs of Cu-strengthened HSLA steel using image texture analysis. *Measurement* **47**, 130–144 (2014).
18. Komenda, J. Automatic recognition of complex microstructures using the Image Classifier. *Mater. Charact.* **46**, 87–92 (2001).

7

19. DeCost, B. L. & Holm, E. A. A computer vision approach for automated analysis and classification of microstructural image data. *Comput. Mater. Sci.* **110**, 126–133 (2015).
20. Adachi, Y., Taguchi, M. & Hirokawa, S. Microstructure Recognition by Deep Learning. *Tetsu-to-Hagané* **102**, 62–69 (2016).
21. Taguchi, M., Hirokawa, S., Yasuda, I., Tokuda, K. & Adachi, Y. Microstructure Detection by Advanced Image Processing. *Tetsu-to-Hagané* **103**, 24–30 (2017).
22. De Albuquerque, V. H. C., Cortez, P. C., Alexandria, A. R. & Tavares, J. M. R. S. A new solution for automatic microstructures analysis from images based on a backpropagation artificial neural network. *NDT & E Int.* **23**, 273–283 (2008).
23. De Albuquerque, V. H. C., Alexandria, A. R., Cortez, P. C. & Tavares, J. M. R. S. Evaluation of multilayer perceptron and self-organizing map neural network topologies applied on microstructure segmentation from metallographic images. *NDT & E Int.* **42**, 644–651 (2009).
24. De Albuquerque, V., Tavares, J. M. R. S. & Cortez, P. Quantification of the microstructures of hypoeutectic white cast iron using mathematical morphology and an artificial neural network. *IJMMP* **5**, 52–64 (2010).
25. De Albuquerque, V. H. C., Reboucas Filho, P. P., Cavalcante, T. S. & Tavares, J. M. R. S. New computational solution to quantify synthetic material porosity from optical microscopic images. *J. Microsc.* **240**, 50–59 (2010).
26. Albuquerque, V. H. C. D., Silva, C. C., Menezes, T. I. D. S., Farias, J. P. & Tavares, J. M. R. S. Automatic evaluation of nickel alloy secondary phases from SEM images. *Microsc. Res. Tech.* **74**, 36–46 (2011).
27. Papa, J. P., Nakamura, R. Y. M., de Albuquerque, V. H. C., Falcao, A. X. & Tavares, J. M. R. S. Computer techniques towards the automatic characterization of graphite particles in metallographic images of industrial materials. *Expert Syst. Appl.* **40**, 590–597 (2013).
28. Breiman, L. Random forests. *Machine Learning* **45**, 5–32 (2001).
29. Breiman, L. Technical Note: Some Properties of Splitting Criteria. *Machine Learning* **24**, 41–47 (1996).
30. Arganda-Carreras, I., *et al.* Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* (Oxford, England) (2017).
31. Vyas, N., Sammons, R. L., Addison, O., Dehghani, H. & Walmsley, A. D. A quantitative method to measure biofilm removal efficiency from complex biomaterial surfaces using SEM and image analysis. *Scientific Reports* **6**, 32694 (2016).
32. Staniewicz, L. & Midgley, P. A. Machine learning as a tool for classifying electron tomographic reconstructions. *Adv. Struct. Chem. Imaging.* **1**, 1–15 (2015).
33. Ko, B. C., Kim, S. H. & Nam, J.-Y. X-ray image classification using Random Forests with local wavelet-based CS-local binary patterns. *J. Digit. Imaging* **24**, 1141–1151 (2011).
34. Ikawa, H., Oshige, H. & Tanoue, T. Study on the martensite-austenite constituent in weld-heat affected zone of high strength steel. *J. Jap. Weld. Soc.* **49**, 467–472 (1980).
35. Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **25**, 1463–5 (Oxford, England) (2009).
36. Wright, M. N. & Ziegler, A. A fast implementation of Random Forests for high dimensional data in C++ and R. *J. Stat. Softw.* **77**, 1–17 (2017).
37. Ferreira, T. & Rasband, W. ImageJ user guide IJ 1.46r. http://imagej.nih.gov/ij/docs/guide (2012).
38. Gall, J., Razavi, N. & Van Gool, L. *An Introduction to Random Forests for Multi-class Object Detection* (eds Dellaert, F., Frahm, J. M., Pollefeys, M., Leal-Taixé, L. & Rosenhahn, B.) *Outdoor and Large-Scale Real-World Scene Analysis. Lecture Notes in Computer Science* **7474**, 243–263 (Springer, Berlin, Heidelberg 2012).
39. Dougherty, B., Schindelin, J., Doube, M., Domander, R. & Hiner, M. Fiji: Local Thickness. https://imagej.net/Local_Thickness (2017).
40. Saito, T. & Toriwaki, J. New algorithms for Euclidean distance transformation on an n-dimensional digitized picture with applications. *Pattern Recognition* **27**, 1551–1565 (1994).
41. Xua, P., Bai, B., Yin, F., Fang, H. & Nagai, K. Microstructure control and wear resistance of grain boundary allotriomorphic ferrite/granular bainite duplex steel. *Mater. Sci. & Eng. A* **385**, 65–73 (2004).
42. Rezakhaniha, R. *et al.* Experimental investigation of collagen waviness and orientation in the arterial adventitia using confocal laser scanning microscopy. *Biomech. Model Mechanobiol.* **3-4**, 461–473 (2012).
43. Yurioka, N., Okumura, M., Kasuya, T. & Cotton, H. J. U. Prediction of HAZ hardness of transformable steels. *Metal Construction* **19**, 217R–223R (1987).

Acknowledgements

Author Contributions

D.S.B. developed the machine learning procedures, performed the image analyses, and wrote the manuscript. S.T. and T.K. observed the microstructures, performed the manual image analysis, and provided the training data for machine learning. M.D. and M.W. supervised this research and edited the manuscript. The manuscript was written up through contributions of all authors. All authors have given approval to the final version of the manuscript.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.