



Statistical evaluation of phase fractions in a MoSiBTiC alloy by machine-learning-assisted segmentation of scanning electron microscopy images

Chihana Kudo^a, Masahiko Demura^b, Akihiro Endo^b, Kenji Nagata^b, Kyosuke Yoshimi^{a,*}

^a Graduate School of Engineering, Tohoku University, 6-6-02 Aramaki Aza Aoba, Sendai, Miyagi, 980-8579, Japan

^b National Institute for Materials Science, 1-1 Namiki, Ibaraki, Tsukuba, 305-0044, Japan

ARTICLE INFO

Keywords:

MoSiBTiC alloy
Scanning electron microscopy
Image segmentation
Machine learning
Bootstrap resampling
Fracture toughness

ABSTRACT

MoSiBTiC alloys exhibit complex four-phase microstructures, and quantitative microstructural characterization requires accurate phase segmentation and statistically representative sampling. This study investigates how many SEM-BSE micrographs are required to obtain statistically reliable phase-fraction data from a heterogeneous multiphase MoSiBTiC alloy. A semi-automatic segmentation workflow based on Trainable Weka Segmentation was employed to enable consistent analysis of a large SEM-BSE image set. In a limited quantitative agreement evaluation using two manually labeled reference images, the Semi-automatic method achieved an average overall pixel accuracy of 0.984 and a weighted Intersection-over-Union (IoU) of 0.969. The total workflow time for preparing 100 analysis-ready segmentation images was reduced from approximately 300 days for full manual pixel-wise labeling to approximately one working day. Phase volume fractions obtained from the segmented images were analyzed by bootstrap resampling to evaluate the effect of image number on statistical reliability. The mean phase volume fractions were $42.7 \pm 0.46\%$ for the Mo solid solution, $36.1 \pm 0.63\%$ for Mo_5SiB_2 , $18.9 \pm 0.64\%$ for TiC, and $2.27 \pm 0.19\%$ for Mo_2C . The standard deviation of the mean phase volume fraction decreased with increasing image number following a power-law relationship. Analyses based on only a few images could misestimate phase volume fractions by more than 15%, particularly for TiC. Under the adopted empirical relationship between phase volume fractions and fracture toughness, approximately 60 images acquired at $2000\times$ were estimated to be necessary to reduce image-sampling-induced scatter in the fracture-toughness estimate to a level comparable to the experimental error.

1. Introduction

In materials engineering, composition, processing, structure, properties, and performance are five fundamental elements. Much of materials research can be viewed as an effort to elucidate one or more interrelationships among these five elements. In materials development, however, an integrated understanding that goes beyond isolated relationships—particularly along the composition-processing-structure-properties-performance chain—is required [1]. Structure is the key node linking composition and processing to properties and performance, and for structural materials, microstructure is of central importance. For a given composition, microstructure depends on the manufacturing process and strongly influences mechanical properties. For this reason, microstructural characterization has long been emphasized in structural

materials research, and descriptors such as grain size and phase volume fraction have been quantified from microscopy images to discuss mechanisms such as grain growth and strengthening. In recent years, data-driven approaches have also been introduced into materials science, and new ways of utilizing microstructural information are being explored, including the automated extraction of microstructural descriptors [2–7].

The MoSiBTiC alloy investigated in this study has a high melting point, a density comparable to that of Ni-based single crystal superalloys, excellent high-temperature creep strength, and good room-temperature fracture toughness, making it a promising next-generation ultra-high-temperature alloy [8–14]. Its microstructure consists of the Mo solid solution (Mo_{ss}) phase, the Mo_5SiB_2 (T_2) phase, the Mo_2C phase, and the TiC phase [8], and a fine lamellar structure

* Corresponding author.

E-mail address: yoshimiksk@tohoku.ac.jp (K. Yoshimi).

<https://doi.org/10.1016/j.jmrt.2026.07.282>

Received 26 May 2026; Received in revised form 15 July 2026; Accepted 29 July 2026

Available online 30 July 2026

2238-7854/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

composed of the Mo_{ss} and TiC phases is also frequently observed. This complex microstructure varies with processing and has been reported to affect mechanical properties [7]. To relate microstructure to processing and properties, it is essential to quantify geometric descriptors such as the volume fraction and connectivity of each constituent phase. For complex microstructures such as that of the MoSiBTiC alloy, two major challenges are efficient segmentation and statistically representative sampling.

Recent studies have increasingly explored machine-learning approaches to improve the efficiency and automation of microstructural segmentation, as summarized in Table 1 [15–34]. Table 1 summarizes representative examples of previous studies on microstructure segmentation to provide context for the present work. The purpose of this table is not to suggest that the present TWS-based workflow is technically superior to recent deep-learning-based methods, but to clarify the type of segmentation problem addressed here and the practical requirements for analyzing a large number of SEM-BSE images of a heterogeneous four-phase alloy. From the viewpoint of image characteristics, microstructural images can be broadly classified into those in which phases are distinguished by contrast differences [15–28] and those in which they are distinguished by texture differences [29–34]. The microstructure of the MoSiBTiC alloy is clearly observed in backscattered electron (BSE) images acquired by scanning electron microscopy (SEM), and the constituent phases are primarily distinguished by contrast. From a segmentation perspective, this microstructure has three notable characteristics. First, it contains four phases with different contrast levels, each essentially featureless and lacking texture arising from substructures; moreover, the contrast ratios among the phases vary from one field of view to another. Second, the constituent phases appear over a wide range of sizes and morphologies. For example, the TiC phase occurs as coarse primary crystals, as part of eutectic microstructures composed of two or three intermediate-sized phases, as fine needle-like precipitates, and as a constituent of the lamellar structure. Third, some precipitates differ in contrast but have similar morphologies. For example, the Mo_{ss} and TiC precipitates formed within the T₂ phase both exhibit prismatic

rod-like shapes. Previous studies on the segmentation of contrast-based microstructural images [15–28] generally addressed no more than three phases, with each phase exhibiting distinctive morphological characteristics. It is therefore not self-evident that previously reported segmentation methods can be directly applied to complex microstructures such as that of the MoSiBTiC alloy, in which four or more phases coexist with diverse morphologies.

The second challenge is identifying statistically representative sampling conditions. When microstructural descriptors are related to material properties, they must be measured over the material volume relevant to the manifestation of the target property. For example, fracture toughness is an important property for the practical application of the MoSiBTiC alloy, and consideration of specimen size indicates that statistical descriptors should ideally be extracted from a relatively large volume of approximately 1.2 mm³. In contrast, high magnification is required to observe the fine precipitates present in the microstructure, and a single SEM image covers only a very limited area. It is therefore necessary to observe multiple locations to obtain statistical values representative of the material as a whole. However, the extent of observation required to establish meaningful correlations with macroscopic material properties is not necessarily obvious. This is particularly true for the MoSiBTiC alloy, which exhibits a heterogeneous microstructure across multiple length scales, making it difficult to intuitively determine whether a given number of images is sufficient. A method is therefore needed to determine the minimum number of observations required for analyses linked to macroscopic material properties.

Based on the above, the primary objective of this study was to quantitatively evaluate the statistical representativeness of SEM-BSE micrographs for phase-fraction analysis in a heterogeneous multiphase MoSiBTiC alloy. For this purpose, a sufficiently large number of SEM-BSE images must be segmented with consistent criteria. The aim of using machine-learning-assisted segmentation in this study was therefore not to develop a new segmentation architecture, but to enable practical and reproducible processing of a large image set for statistical microstructural analysis. We employed Trainable Weka Segmentation (TWS) as an annotation-efficient tool that can be trained using limited annotations within selected images and implemented in the widely used Fiji/ImageJ environment. Using the segmented image set, we then evaluated how the number of analyzed micrographs affects the uncertainty in phase-fraction measurements and how this uncertainty propagates into fracture-toughness estimation through an empirical microstructure–property relationship.

2. Materials and methods

An alloy with a nominal composition of 65Mo–10Si–5B–10Ti–10C (at.%) was prepared by arc melting. The starting materials were 99.99 mass% Mo rods, 99.9999 mass% Si chips, 99.5 mass% B grains, and 99 mass% TiC powder, the latter serving as the Ti and C source. Melting was carried out under an Ar atmosphere. The alloy was then heat treated at 1800 °C for 24 h under Ar and furnace cooled to room temperature. Specimens for microstructural observation were sectioned from the heat-treated alloy by wire electrical discharge machining (EDM) and then mechanically polished. Diamond slurry buff polishing was selected as the final polishing procedure, and SEM images were acquired from the polished specimens. Fig. 1 shows SEM images obtained after final polishing by diamond-slurry buff polishing, Al₂O₃ buff polishing, and vibratory polishing using nanodiamond. In the specimens polished by Al₂O₃ buffing and vibratory polishing, some regions of the Mo_{ss} phase, which normally appeared white, were observed as gray. Because such gray-level variation could adversely affect contrast-based image segmentation, diamond slurry buff polishing, which did not exhibit this problem, was adopted as the final polishing method.

The microstructure was observed using SEM (JEOL JSM-IT800) in backscattered electron (BSE) mode. A magnification of 2000× was selected so that fine features such as precipitates could be clearly

Table 1

Examples of previous studies on the application of segmentation to microstructure images.

Materials	Segmentation objects	Method (Automatic = A/ Semi-automatic = S)
DP800 steel [27]	Martensite, ferrite, damage site (3 objects, 3 Gy scale)	• nnU-Net (A) • DB-GAN ensemble (A)
HP40–Nb stainless steel [28]	Intergranular/boundary carbides, intragranular/pepper carbides, matrix (3 objects, 2 Gy scale and 2 different shapes)	• Trainable Weka Segmentation with image preprocessing (S)
Continuous cooling steel [29]	Ferrite, pearlite, degenerate pearlite, Widmanstätten ferrite (4 objects, 3 Gy scale)	• Trainable Weka Segmentation (S)
Low carbon steel [30]	Grain boundary ferrite, Ferrite side plate, Pearlite, Martensite (4 objects, 3 Gy scale and 2 different morphology)	• Unsupervised method with a feature extraction using texture analysis filters and a clustering process using BGMM (A)
Ultrahigh carbon steel [31]	Network carbide, Ferritic denuded zone, Widmanstätten carbide, spheroidite matrix (4 objects, 4 different shapes)	• Pre-trained convolutional neural networks (A)
FeCrMo sintered steel [32]	Lower bainite, upper bainite, phosphorus-rich ferrite, fine pearlite, martensite, pores (6 objects, 3 Gy scale and 4 different morphology)	• Image Classifier (software package integrated with MicroGOP2000/S system) (S)
Steel (synthetic weld-heat affected zone) [33,34]	Ferrite, pearlite, bainite (3 objects, 3 different patterns)	• Expert-supervised Machine Learning by using Random Forest classification (S)

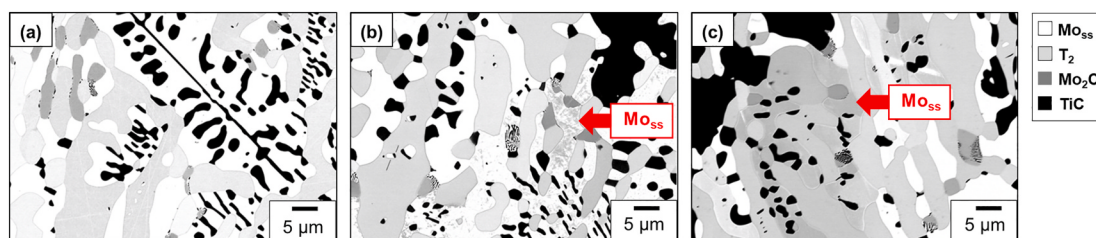


Fig. 1. SEM images obtained by different polishing conditions, (a) diamond slurry buffing, (b) Al_2O_3 buffing, (c) vibratory polishing.

resolved. The accelerating voltage and working distance were 10 kV and 10 mm, respectively. Each image covered an area of $64\ \mu\text{m} \times 48\ \mu\text{m}$ with a resolution of 1280×960 pixels. A total of 10×11 images were acquired at contiguous positions so that a large area of $640\ \mu\text{m} \times 528\ \mu\text{m}$ could be reconstructed. This reconstructed area was assumed to be sufficiently representative of the alloy as a whole for discussion of its macroscopic material properties, and the degradation in the precision of statistical measurements was then examined by progressively reducing the number of images used. Because several images contained voids or surface contamination, 100 BSE images free from these artifacts were used for the subsequent analysis. Images containing large gaps, stains, voids, or surface contamination were excluded because these artifact regions are not constituent phases of the MoSiBTiC alloy. The present TWS-based classifier was trained as a four-class model for the Mo_{ss} , T_2 , Mo_2C , and TiC phases, and no artifact or background class was included. Therefore, if artifact-containing images were directly input into the current segmentation model, the artifact regions would be forced to be classified as one of the four constituent phases according to their grayscale contrast and local image features. In particular, dark artifact regions such as voids, gaps, or stains would likely be misclassified as TiC or Mo_2C because these phases appear black or dark gray in SEM-BSE images. Bright or intermediate-contrast artifacts could also be misclassified as Mo_{ss} or T_2 , depending on their local contrast. Such misclassification would artificially change the measured phase fractions. Therefore, only artifact-free SEM-BSE images were used for the quantitative phase-fraction analysis in this study.

Segmentation of the constituent phases was performed for all 100 SEM images. Details of the segmentation procedure used in this study are given in Section 3. The segmented images were then used to determine the phase area fraction of each phase. The calculations were carried out in Python, and the fraction of pixels assigned to each phase in the segmented two-dimensional SEM-BSE images was calculated as the phase area fraction. In this study, these phase area fractions were used as stereological estimates of phase volume fractions. Hereafter, the term “phase volume fraction” refers to the volume fraction estimated from the measured phase area fractions in the SEM-BSE images. This stereological estimation is based on the Delesse principle, according to which the expected areal fraction of a phase on a random planar section is equal to its volume fraction [35]. The SEM images and the corresponding segmentation images were managed using RDE [36], a data structuring system provided by the National Institute for Materials Science (NIMS).

The dependence of the estimated phase volume fractions on the number of analyzed images was evaluated by bootstrap resampling. When all 100 images were used, the resulting dataset, consisting of 100 images with a resolution of 1280×960 pixels each, was defined as the full field of view. For each bootstrap trial, 100 images were randomly sampled with replacement from the 100-image dataset, and the phase volume fractions for the full field of view were calculated. This procedure was repeated 1000 times, and the variability of the resulting values was evaluated. Following the same procedure, for a given sample size N , N images were randomly resampled from the 100-image dataset, and the phase volume fractions for a field of view corresponding to N images (1280×960 pixels per image) were calculated 1000 times. N was varied from 1 to 99. In this statistical analysis, each SEM-BSE image

was treated as an equal-area spatial sampling unit. These image units should not be interpreted as completely independent material specimens because the images were acquired from contiguous positions. The purpose of the resampling analysis was to evaluate how the estimated mean phase fraction varies when a finite number of fields of view are selected from the observed 100-image dataset.

In addition to the image-wise bootstrap analysis, a 2×2 block bootstrap analysis was performed to examine the possible influence of local spatial correlation among neighboring SEM-BSE images. In this analysis, four spatially adjacent SEM-BSE images were treated as one sampling block and resampled with replacement. The effective number of images was defined as four times the number of sampled blocks. The standard deviations obtained by the 2×2 block bootstrap were compared with those obtained by the image-wise bootstrap. In addition, the correlation coefficients of image-based phase fractions between neighboring image regions were evaluated to estimate the strength of local spatial correlation.

3. Machine-learning-based segmentation workflow

Manual segmentation of SEM-BSE images of the four-phase MoSiBTiC alloy is extremely time-consuming. In this study, we used a machine-learning-assisted segmentation workflow based on the Trainable Weka Segmentation (TWS) plugin in Fiji/ImageJ. TWS is a supervised pixel-classification method implemented through Weka and is based on random-forest classification [37,38]. The method was selected not because it represents a new or state-of-the-art segmentation architecture, but because it can be trained using limited annotations within selected images and is suitable for preparing a large number of consistently segmented images for quantitative phase-fraction analysis. This practical advantage is important for the present study, where the central question is how many SEM-BSE micrographs are required to obtain statistically reliable phase-fraction data from a heterogeneous multi-phase microstructure.

Manual reference labels were prepared based on SEM-BSE contrast and the characteristic morphology of each phase. In the SEM-BSE images, the Mo_{ss} phase appears white, the T_2 phase appears light gray, the Mo_2C phase appears dark gray, and the TiC phase appears black. This phase assignment is consistent with the established phase identification in previous studies on MoSiBTiC alloys. Phase boundaries were determined based on the local contrast transition between neighboring phases. When the contrast transition was sharp, the boundary was placed along the steepest intensity change. When a boundary region exhibited a gradual grayscale transition, the boundary was assigned approximately to the middle of the transition region on a pixel basis. When the boundary was ambiguous, the local continuity and characteristic morphology of the neighboring phases were also considered. These manually labeled images were used as reference images for evaluating the semi-automatic segmentation. The segmentation workflow consisted of four steps: (1) level adjustment, (2) annotation and model training, (3) model application, and (4) model refinement. A schematic illustration of the workflow, including the operations performed in each step, is shown in Fig. 2.

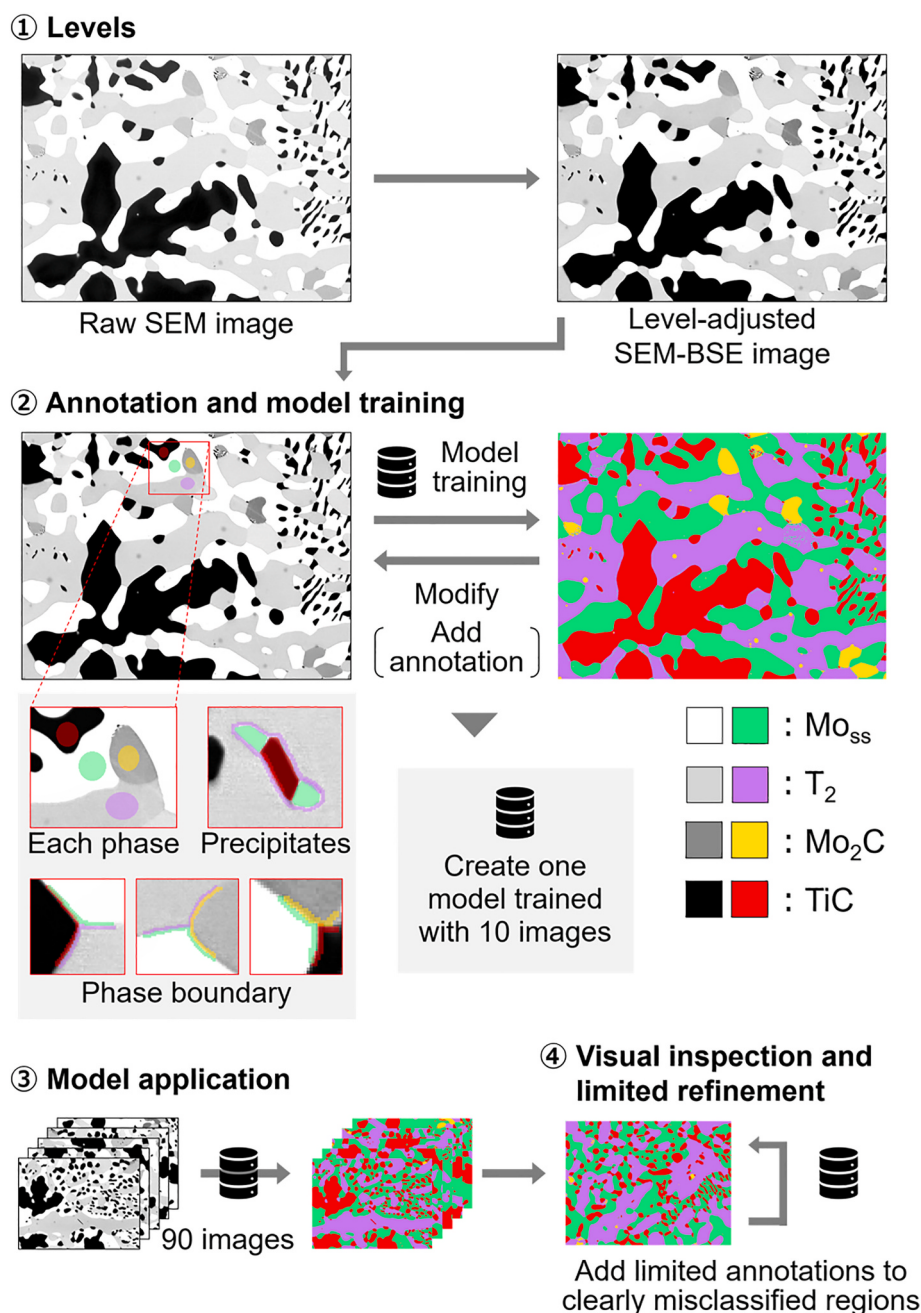


Fig. 2. Schematic diagram of the semi-automatic TWS-based segmentation workflow used in this study, including level adjustment, annotation and model training, model application, visual inspection, and limited refinement.

(1) Level adjustment

When 100 SEM-BSE images are segmented using the same classifier, the same phase may exhibit different gray levels due to differences in image acquisition conditions, local contrast variation, and observation position. This variation adversely affects phase recognition, particularly for the white Mo_{ss} , light-gray T_2 , and dark-gray Mo_2C phases. Therefore, all 100 SEM-BSE images were first subjected to level adjustment using PhotoDirector 365 to reduce large image-to-image tonal variation before TWS segmentation. In this preprocessing step, the highlight and shadow parameters were fixed at +100 and −100, respectively, for all images. The midtone parameter was set to −60 for most images. For images in which excessive brightness made the light-gray T_2 phase difficult to distinguish from the white Mo_{ss} phase, the midtone parameter was adjusted within the limited range from −60 to −80. No other image-

specific tonal adjustment was applied. The exact level-adjustment parameters used for the analyzed images are listed in [Supplementary Table S1](#). This preprocessing step was therefore limited to a defined parameter range, rather than arbitrary image-by-image adjustment. Nevertheless, because selecting the midtone value within this range still involves limited human judgment, the present workflow should be regarded as a semi-automatic workflow, not as a fully automatic segmentation procedure.

(2) Annotation and model training

As illustrated in [Fig. 2](#), one SEM image was first selected, and representative annotations were placed on regions corresponding to each phase and on selected interphase boundaries within the field of view. A classifier was then constructed from these initial training data

using the random forest algorithm. First, the input image was processed with multiple filters to generate a feature vector for each pixel. Next, a large number of decision trees were trained using these feature vectors [39]. Prediction was made by majority vote, and the class predicted by the largest number of decision trees was assigned to the corresponding pixel. Gaussian blur, Sobel filter, Hessian, Difference of Gaussians, and Membrane projections were used to construct the feature vectors. These features were used to extract image information relevant to contrast-based SEM-BSE segmentation of the present four-phase MoSiBTiC alloy. Gaussian blur was used to obtain smoothed grayscale-intensity information and to reduce high-frequency noise. The Sobel filter was used to emphasize local intensity gradients at phase boundaries. Hessian and Difference of Gaussians filters were used to capture local curvature- and scale-dependent contrast features associated with precipitates and interphase regions. Membrane projections were used to assist recognition of thin, lamellar, or boundary-like features.

Because the classifier trained on the initial annotations still produced misclassified regions, an iterative training and refinement procedure was introduced. The classifier was progressively improved by adding annotations from SEM-BSE images in which obvious misclassification was observed. During construction of the final classifier, annotations from 10 SEM-BSE images were used in total for the training/refinement stage. In addition, 5 SEM-BSE images were used as qualitative checking images to visually assess whether the classifier could be applied to images acquired from different observation positions. Because fully manually labeled reference images were not available for every intermediate training stage, the evolution of classification accuracy during iterative training could not be quantified at each step using pixel-wise metrics. Therefore, the training progress was assessed mainly by visual inspection, and the final segmentation accuracy was quantitatively evaluated using two manually labeled reference images, as described in Section 4.1. These 10 training/refinement images and 5 qualitative-checking images were part of the 100-image dataset used for the subsequent statistical analysis; no images beyond the 100-image dataset were required. Specifically, misclassified regions were newly annotated, added to the training data, and used for retraining. This process was repeated until the segmentation across the entire image was judged by visual inspection to be broadly satisfactory. The same procedure was then applied to other SEM images, thereby progressively improving the segmentation accuracy. In other words, classification errors gradually decreased as training proceeded from the second image to the third and subsequent images. It was confirmed that once this procedure was applied to 10 images, classification errors on subsequent images became almost negligible. The 10 images used in this training stage were segmented during the iterative training process itself.

(3) Model application

The final classifier was then applied to the remaining SEM-BSE images to obtain segmentation results. These results were subsequently visually inspected as part of the quality-checking process.

(4) Model refinement

During classifier development, segmentation results obtained from the training/refinement images were visually inspected. When obvious misclassification was observed, particularly in regions where the T_2 phase exhibited strong contrast variation due to channeling or local brightness differences, limited additional annotations were added within the training/refinement image set and the classifier was refined. After the final classifier was established, it was applied to the remaining SEM-BSE images, and the resulting segmentation images were visually inspected for quality checking. Therefore, the proposed method should be regarded as a semi-automatic procedure involving limited manual inspection and refinement, rather than a fully automatic segmentation

method.

By following the above procedure, 100 analysis-ready segmented images were obtained while manually annotating only limited regions of the image set. Hereafter, the conventional approach in which each SEM image was manually labeled pixel by pixel is referred to as Manual, whereas the method described in this section is referred to as Semi-automatic.

4. Results

4.1. Segmentation results obtained using the semi-automatic method

An example of the segmentation results obtained using the Semi-automatic method is shown in Fig. 3. In the segmented image, the Mo_{ss} , T_2 , Mo_2C , and TiC phases are colored green, purple, yellow, and red, respectively. The corresponding SEM-BSE image and the manually labeled reference image are also shown for comparison. The Semi-automatic method captured the main microstructural features observed in the manually labeled reference image, including fine precipitates and interphase boundaries. Because visual comparison alone is insufficient to evaluate segmentation quality, a quantitative agreement evaluation was performed using two manually labeled reference images. The Semi-automatic segmentation results were compared with the manually labeled reference images using overall pixel accuracy, precision, recall, F1-score, and Intersection-over-Union (IoU). The quantitative results are summarized in Table 2. The Semi-automatic method achieved an average overall pixel accuracy of 0.984, an average weighted precision of 0.985, an average weighted recall of 0.984, an average weighted F1-score of 0.984, and an average weighted IoU of 0.969. These values indicate high agreement between the Semi-automatic segmentation and the manually labeled reference images.

The phase-wise IoU values were 0.974 for Mo_{ss} , 0.966 for T_2 , 0.970 for TiC, and 0.884 for Mo_2C . The lower IoU for the Mo_2C phase is attributed to its small area fraction and locally sparse distribution, which make its segmentation more sensitive to small boundary errors and minor misclassification. Nevertheless, the obtained agreement was considered sufficient for the purpose of quantitative phase-fraction analysis in this study. It should be noted that the present quantitative evaluation was performed using only two manually labeled reference images because preparing fully pixel-wise labeled reference images for this complex four-phase microstructure is highly time-consuming. Therefore, the accuracy values reported here should be interpreted as a limited quantitative agreement evaluation, rather than as a comprehensive evaluation of generalization performance over the entire 100-image dataset.

The Manual method required approximately 3 days per image for full pixel-wise labeling; thus, segmentation of 100 images would be expected to require approximately 300 days. In contrast, the total active workflow time required to prepare 100 analysis-ready segmentation images using the Semi-automatic method was approximately one working day when preprocessing, annotation, model training/refinement, model application, visual inspection, data export, and accuracy checking were included. Therefore, the proposed workflow should be regarded as a semi-automatic procedure involving limited human intervention, rather than as a fully automatic segmentation method. Even with this limitation, the workflow made it feasible to analyze 100 SEM-BSE images under consistent segmentation criteria, enabling the statistical evaluation of phase fractions described in the following sections.

Table 3 summarizes the segmentation-related challenges encountered in this study, the key elements of the workflow described in Section 3, and the corresponding improvements. Four points were found to be particularly important for successful segmentation. First, proper specimen surface preparation is essential for obtaining high quality SEM images. Specifically, diamond slurry buff polishing is desirable as the final polishing step for the MoSiBTiC alloy. Depending on the final

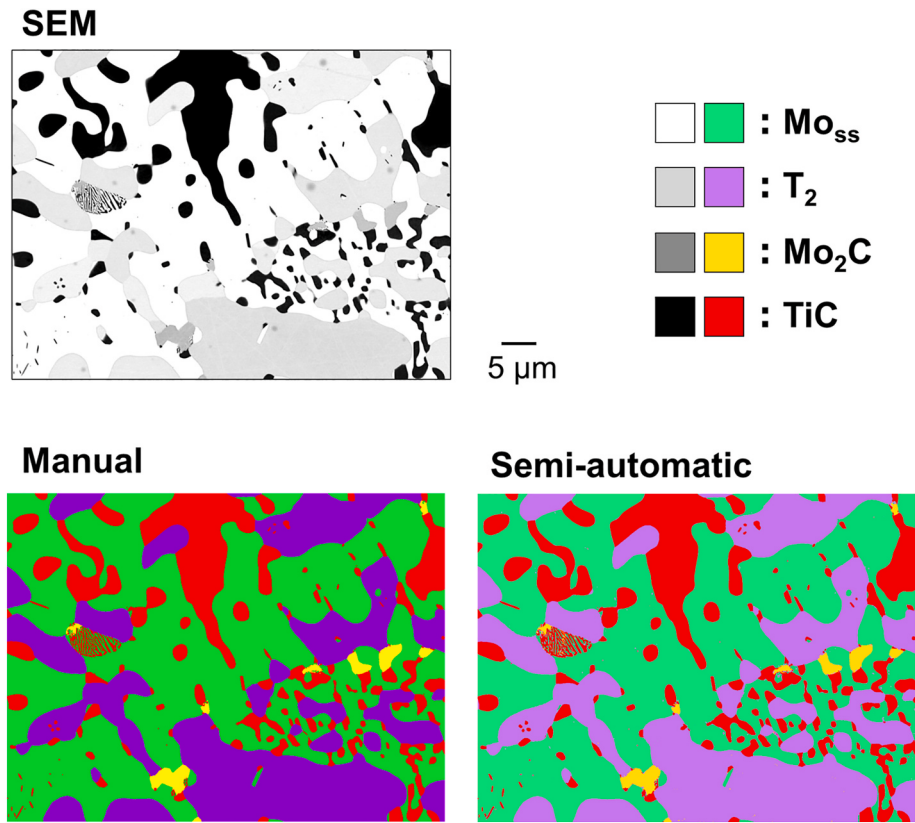


Fig. 3. Comparison of a representative SEM-BSE image, the manually labeled reference image, and the semi-automatic TWS-based segmentation result.

Table 2
Quantitative agreement between the Semi-automatic segmentation and manually labeled reference images.

Metric	Image A	Image B	Mean
Overall pixel accuracy	0.985	0.983	0.984
Weighted precision	0.986	0.984	0.985
Weighted recall	0.985	0.983	0.984
Weighted F1-score	0.985	0.983	0.984
Weighted IoU	0.971	0.967	0.969
Macro IoU	0.949	0.948	0.949

Note: Image A and Image B are the two manually labeled reference images. Weighted metrics account for the number of reference pixels belonging to each phase. Overall pixel accuracy is equivalent to the micro-averaged precision, recall, and F1-score for single-label segmentation.

polishing method, the contrast of the Mo_{ss} phase may appear darker, making it difficult to distinguish from the T₂ and Mo₂C phases. With diamond slurry buff polishing, by contrast, the Mo_{ss} phase exhibits bright contrast. Second, prior level adjustment of the SEM images is effective. Because the MoSiBTiC alloy contains the light gray T₂ phase and the dark gray Mo₂C phase, differences in imaging data and channeling contrast can cause large variations in the absolute contrast of these gray phases, making them difficult to distinguish. Prior adjustment reduces large image-to-image tonal variation and improves the model's accuracy when applied to images beyond the training set. Third, in the present dataset, annotations from 10 SEM-BSE images were used for the training/refinement stage, and 5 additional SEM-BSE images were used for qualitative checking of the classifier. This condition was adopted as a practical training/refinement procedure for the present MoSiBTiC alloy dataset. The optimal number of training/refinement images may depend on alloy composition, imaging condition, phase contrast, and microstructural complexity, and should therefore be evaluated separately when the workflow is applied to other alloy systems. Fourth, when high

accuracy is required, it is desirable to manually inspect the segmentation results after applying the trained model and refine the classifier accordingly.

4.2. Statistical evaluation of phase volume fractions

Bootstrap resampling was performed by randomly sampling 100 measurements with replacement from the 100 image-based phase-volume-fraction measurements. This procedure was used to evaluate the variability of phase volume fractions by repeatedly measuring over the full field of view, comprising 100 images (1280 × 960 pixels per image). Fig. 4 shows the histograms of the phase volume fractions obtained in this manner. The mean value for each phase is also shown in Fig. 4. The mean volume fractions were 42.7 ± 0.46% for the Mo_{ss} phase, 36.1 ± 0.63% for the T₂ phase, 18.9 ± 0.64% for the TiC phase, and 2.27 ± 0.19% for the Mo₂C phase. The resulting scatter can be interpreted as an estimate of sampling variability when 100 observation fields are used. The variability corresponded to approximately 1% of the mean value for the Mo_{ss} phase, which had the largest volume fraction, and approximately 8% of the Mo₂C phase, which had the smallest volume fraction. Whether 100 observation fields are sufficient to obtain statistically representative values for the alloy as a whole should therefore be considered in terms of the accuracy of the fracture toughness values predicted using these phase volume fractions.

Next, *N* measurements were randomly sampled from the 100 image-based phase-volume-fraction measurements. This enabled evaluation of the variability in phase volume fractions when measurements were repeated over a field of view corresponding to *N* images (1280 × 960 pixels per image). Fig. 5 shows the change in the scatter of the mean phase volume fractions as the number of images increases. For all constituent phases, the scatter in volume fraction decreased as the number of images increased. Fig. 6 shows the change in the standard deviation of the mean phase volume fractions as the number of images increases. In

Table 3
Summary of image segmentation methods, challenges, and improvements for MoSiBTiC alloy consisting of four phases.

Challenge		Improvement
White Mo _{ss} phase appears gray, leading to confusion with the T ₂ and Mo ₂ C phases	1. Polishing Diamond slurry buffing	Mo _{ss} phase is standardized to white only
Image-to-image tonal variation may cause confusion among the white Mo _{ss} , light-gray T ₂ , and dark-gray Mo ₂ C phases.	2. Pre-adjustment of images	Defined level adjustment was applied before TWS segmentation (highlight: +100; shadow: -100; midtone: -60 to -80; Supplementary Table S1)
Manual annotation is time-consuming, whereas too few training images may not capture image-to-image variations in phase contrast and microstructural morphology within the present dataset	3. Training and applying the model with 10 images	Annotations from 10 SEM-BSE images were used for classifier training and refinement, and 5 additional images were used for qualitative validation. The remaining images were segmented using the final classifier and visually inspected, with limited additional refinement where obvious misclassification was observed.
Color tones of the same phase vary by observation area, leading to low accuracy in parts of the model	4. Fine tuning of segmentation images	Segmentation accuracy can be ensured across all images

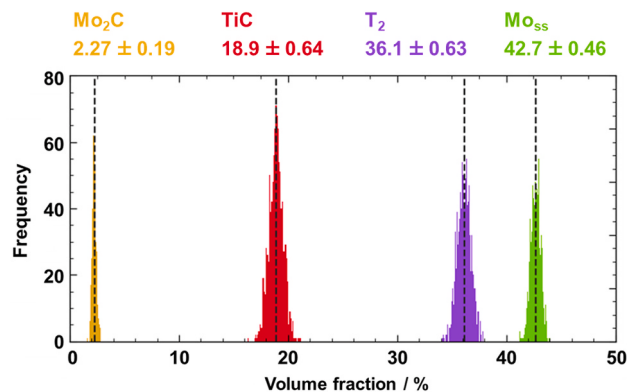


Fig. 4. Histogram of volume fractions and average volume fractions of each phase obtained by the bootstrap method.

the log-log plots, the standard deviation decreased linearly with increasing number of images. These relationships indicate that the number of images required to satisfy a prescribed precision criterion can be determined. For example, for the TiC phase, approximately 40 images are required to reduce the standard deviation to below 1%.

5. Discussion

5.1. Maximum deviation in phase volume fractions as a function of the number of images

The effect of the number of images used for analysis on the estimated volume fraction of each constituent phase was examined. First, one image was randomly selected from the 100-image dataset, and the phase volume fractions in that image were recorded. This procedure was repeated 1000 times. Because only 100 images were available, the 1000 trials inevitably included duplicate selections. Among these 1000 trials, the set of phase volume fractions corresponding to the trial in which the Mo_{ss} phase exhibited the maximum volume fraction was extracted and defined as Case 1 (Mo_{ss} max). Similarly, the set of phase volume fractions for the trial in which the Mo_{ss} phase had the minimum volume

fraction was defined as Case 2 (Mo_{ss} min). The same procedure was applied to the other three phases, yielding the phase volume fractions corresponding to the maximum and minimum values for each phase. In this way, eight cases were defined: Case 1, Mo_{ss} max; Case 2, Mo_{ss} min; Case 3, T₂ max; Case 4, T₂ min; Case 5, TiC max; Case 6, TiC min; Case 7, Mo₂C max; and Case 8, Mo₂C min.

Fig. 7(a)–(d) show, for each constituent phase, the volume fractions obtained in Cases 1–8 when only one image was used, i.e., when the analysis was based on a single field of view of 64 μm × 48 μm. For the Mo_{ss} phase, the measured volume fraction ranged from a maximum of 56.4% to a minimum of 32.1%, corresponding to a spread of 24.3 percentage points. Similarly large spreads were observed for the T₂ and TiC phases, amounting to 28.0 and 36.0 percentage points, respectively. Even for the Mo₂C phase, which had the smallest estimated phase volume fraction, a spread of 8.9 percentage points was observed. In some fields of view, no clearly recognizable Mo₂C region was observed, or the measured Mo₂C fraction was extremely small. Such near-zero measured values should be interpreted as fields of view in which no clearly resolvable Mo₂C region was observed at the present magnification and segmentation resolution. Because Mo₂C is a minor and locally sparse phase, its measured fraction is sensitive to both true spatial heterogeneity and small pixel-level segmentation errors.

The Mo₂C phase requires particular attention in this analysis because it has the smallest phase fraction and exhibits a heterogeneous spatial distribution. In the final 100-image dataset, the mean Mo₂C fraction was 2.27%, whereas the image-to-image standard deviation was 1.828%. The median and interquartile range were 1.917% and 0.737–3.491%, respectively, and the minimum and maximum values among individual images were 0.0127% and 8.851%. These values confirm that Mo₂C is a minor phase with large relative spatial heterogeneity. Accordingly, the bootstrap result for Mo₂C should be interpreted as an empirical estimate of sampling variability within the observed 100-image dataset, including possible contributions from both true spatial heterogeneity and minor segmentation uncertainty.

The maximum deviations from the mean volume fractions obtained from the 100 image analysis were 13.7 percentage points for the Mo_{ss} phase, 16.2 percentage points for the T₂ phase, 21.9 percentage points for the TiC phase, and 6.6 percentage points for the Mo₂C phase. These results demonstrate that when the analysis is based on a single image, the variation due to the observation location is large, and the measured

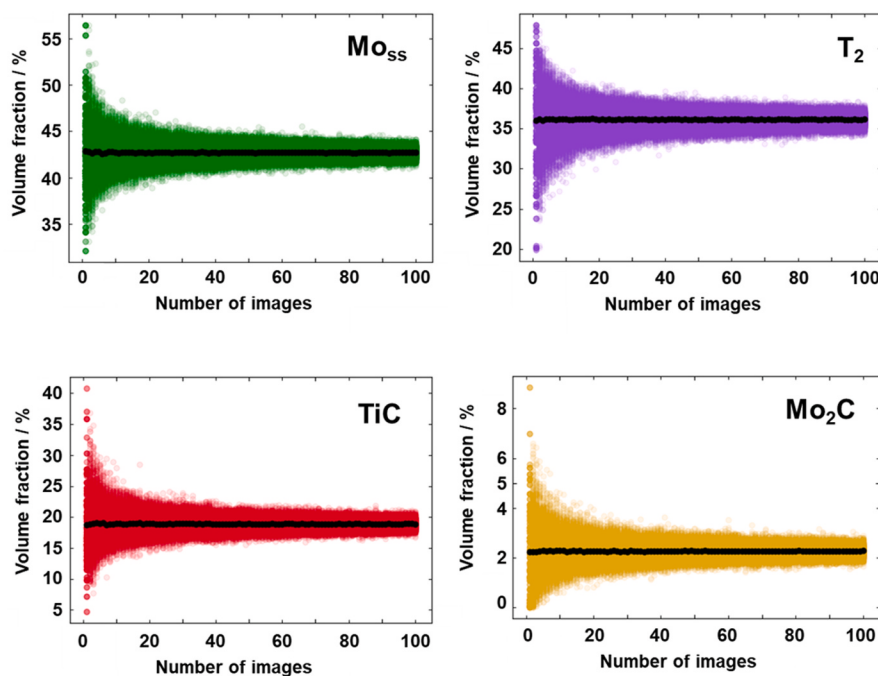


Fig. 5. Dispersion transition of average volume fraction with increasing number of images in each component phase.

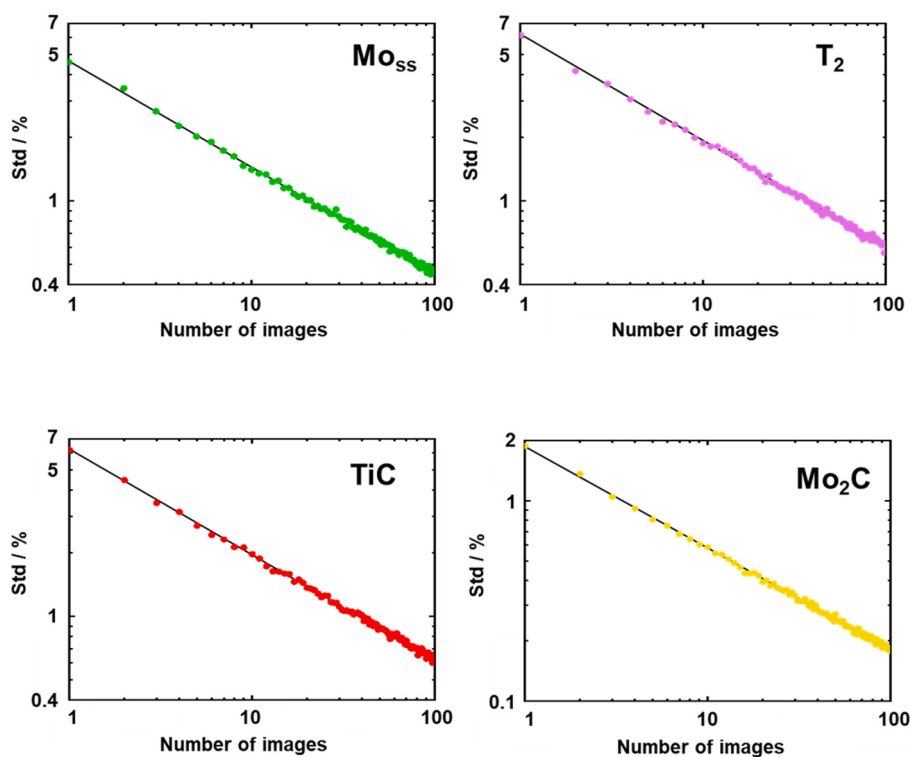


Fig. 6. Standard deviation transition of average volume fraction with increasing number of images in each component phase.

phase volume fractions can deviate substantially from the mean values obtained from the 100-image dataset.

For comparison, Fig. 7(e)–(h) plot the results obtained when 100 images were randomly sampled with replacement from the 100-image dataset and this procedure was repeated 1000 times. When 100 images were used, i.e., when the analysis was based on an observation area

of $640 \mu\text{m} \times 480 \mu\text{m}$, the variability across all phases was greatly reduced. The spreads were limited to 2.8 percentage points for the Mo_{SS} phase, 4.2 percentage points for the T_2 phase, 4.0 percentage points for the TiC phase, and 1.1 percentage points for the Mo_2C phase. Bias was also reduced, and the deviations from the mean values obtained from the 100-image dataset were approximately symmetric. Thus, when 100

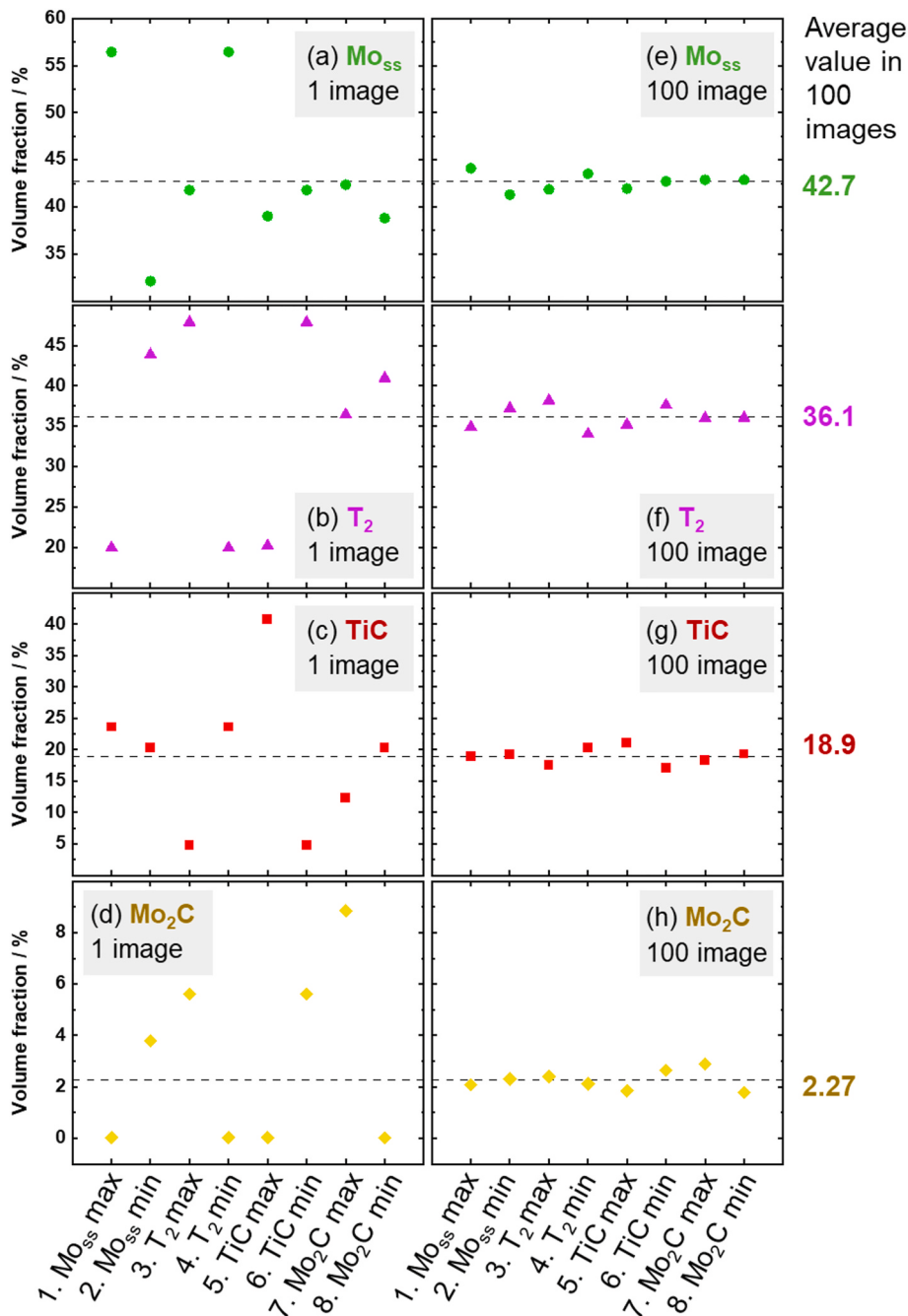


Fig. 7. Volume fractions for the cases where each phase showed the minimum and maximum volume fraction in the analysis. (a–d) Correspond to the single-image analysis, and (e–h) correspond to the analysis using 100 images. (a, e) Show the volume fraction of Mo_{ss} phase, (b, f) T₂, (c, g) TiC, and (d, h) Mo₂C.

fields are observed, the errors in the measured phase volume fractions remain sufficiently small for all phases even if the observation locations are selected at random, indicating that adequate precision can be ensured.

The same analysis was also performed for cases in which 3, 10, or 20 images were sampled. Table 4 lists, for each image number, the phase

Table 4					
Maximum deviation from the average for each number of images.					
Number of images	1	3	10	20	100
Phases with the largest deviation	TiC	T ₂	TiC	TiC	TiC
Maximum deviation	+22%	−15%	+6.6%	+5.2%	+2.1%

that shows the largest deviation from the mean and its magnitude. Table 4 suggests that analyses based on only a few images may misestimate the phase volume fractions by more than 15%. It also shows that the TiC phase is the phase most strongly affected by the number of images used. This is likely because the TiC phase occurs in various morphologies, including coarse primary crystals, intermediate-sized eutectic microstructures, and fine precipitates. Because of this morphological diversity, the measured volume fraction of the TiC phase is particularly sensitive to the observation location.

The possible influence of local spatial correlation among neighboring SEM-BSE images was examined by 2×2 block bootstrap analysis. In this analysis, four spatially adjacent images were treated as one block and resampled together. The standard deviations obtained by the 2×2 block bootstrap were comparable to those obtained by the image-wise

bootstrap. At an effective image number of 100, the ratios of the standard deviation obtained by the 2×2 block bootstrap to that obtained by the image-wise bootstrap were 1.066 for Mo_{ss} , 0.857 for T_2 , 0.993 for TiC , and 0.868 for Mo_2C . At an effective image number of 60, the corresponding ratios were 1.099, 0.934, 1.108, and 0.891, respectively. In addition, the correlation coefficients between neighboring image regions were weak for all phases, with the largest coefficient being 0.202. These results indicate that, although neighboring images are not strictly independent, local spatial correlation did not substantially change the convergence behavior of the mean phase-fraction estimates in the present dataset. For some phases, the standard deviation obtained by the 2×2 block bootstrap was slightly larger than that obtained by the image-wise bootstrap (ratios of 1.066–1.108 for Mo_{ss} and TiC), which is consistent with the weak positive spatial correlation among neighboring images; these differences were nonetheless within about 10%.

The artificial truncation of microstructural features at image

boundaries would be important if descriptors such as phase connectivity, particle size, aspect ratio, or interfacial length were evaluated. In the present study, however, the target descriptor was the phase area fraction estimated by pixel counting. Therefore, partitioning the observed area into equal-area image units does not change the total phase fraction measured from the same total area, although it affects the statistical unit used for resampling.

5.2. Effect of the number of microstructural images on fracture toughness estimation

In MoSiBTiC alloys, the volume fractions of the constituent phases are known to affect the mechanical properties [8]. In particular, for fracture toughness, a regression equation of the form shown in Eq. (1) can be obtained from previously reported experimental data:

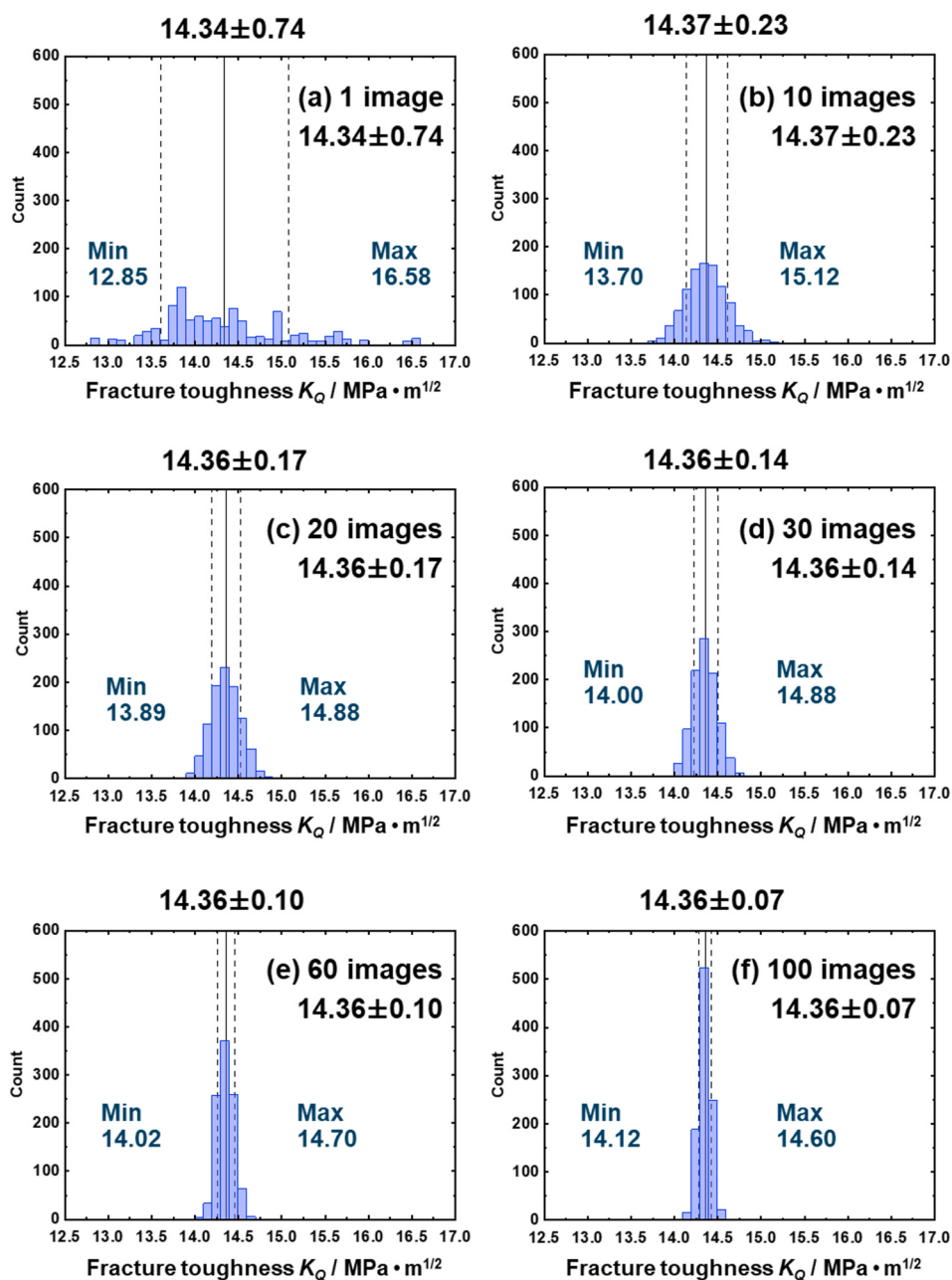


Fig. 8. Histogram of fracture toughness calculated from the volume fraction of constituent phases for each number of (a) 1 image, (b) 10, (c) 20, (d) 30, (e) 60, and (f) 100 images.

$$K_Q = 20.0V_f^{\text{Mo}_{\text{ss}}} + 7.3V_f^{\text{Ti}_2} + 16.5V_f^{\text{TiC}} + 4.8V_f^{\text{Mo}_2\text{C}} \quad (1)$$

where K_Q denotes the fracture toughness and V_f^X denotes the volume fraction of phase X .

To investigate how the number of images used in the analysis affects the fracture toughness estimated from Eq. (1), bootstrap sampling was performed. Specifically, N images were randomly sampled with replacement from the 100-image dataset, and this procedure was repeated 1000 times for each of $N = 1, 10, 20, 30, 60$, and 100. The phase-volume-fraction data obtained in each trial were substituted into Eq. (1), yielding 1000 estimated fracture toughness values for each N . The maximum, minimum, and mean values of these 1000 fracture toughness values were then examined. In this manner, the extent to which the estimated fracture toughness could vary when N images were repeatedly sampled from the 100-image dataset was evaluated.

Fig. 8 shows histograms of the fracture toughness values estimated from N images. The scatter in the estimated fracture toughness decreased as the number of images increased. In particular, when 100 images were used, the standard deviation of the estimated fracture toughness decreased to $0.07 \text{ MPa(m)}^{1/2}$, corresponding to a coefficient of variation of about 0.5% of the mean (Fig. 8(f)).

By comparing the image-sampling-induced scatter with the experimental uncertainty, a conditional estimate of the required image number was obtained. It has been reported that the experimental error in fracture-toughness measurement for a single alloy is approximately $0.3 \text{ MPa(m)}^{1/2}$ [8]. As shown in Fig. 8, under the adopted empirical relationship of Eq. (1), the scatter in fracture toughness caused by image sampling approached approximately $0.3 \text{ MPa(m)}^{1/2}$ when about 60 images were used. Therefore, under the assumption that Eq. (1) is valid for the present comparison, approximately 60 images acquired at $2000\times$ were estimated to be necessary to reduce the image-sampling-induced scatter in the fracture-toughness estimate to a level comparable to the experimental error. In this comparison, the image-sampling-induced scatter is defined as three times the standard deviation (3σ) of the 1000 bootstrap estimates. At 60 images, the standard deviation of the estimated fracture toughness was $0.10 \text{ MPa(m)}^{1/2}$ (Fig. 8(e)), corresponding to a 3σ scatter of approximately $0.30 \text{ MPa(m)}^{1/2}$, which is comparable to the reported experimental error of approximately $0.3 \text{ MPa(m)}^{1/2}$ [8].

Fig. 9 shows Eq. (1) used in this study together with the nine experimental data points from which it was derived. It should be emphasized that the estimated number of approximately 60 images is a conditional estimate and does not include the uncertainty associated with the regression model itself. Equation (1) contains four phase-fraction terms and was derived from only nine experimental data points; therefore, the regression model may be affected by overfitting. In this study, Eq. (1) was used only as an empirical relationship to examine how uncertainty in phase-fraction measurements propagates into the estimated fracture toughness. In addition, the imaging conditions used to determine the phase fractions for deriving Eq. (1), such as magnification, resolution, and the number of SEM images, were not reported in the original literature. A more robust microstructure–property model will require a larger dataset of MoSiBTiC alloys with fracture-toughness measurements and phase-fraction data obtained under standardized imaging and segmentation conditions.

6. Conclusions

In this study, a semi-automatic TWS-based workflow was applied to the segmentation of SEM-BSE images of a four-phase MoSiBTiC alloy. In a limited quantitative agreement evaluation using two manually labeled reference images, the workflow showed high agreement with the reference labels, with an average overall pixel accuracy of 0.984 and an average weighted IoU of 0.969. The total workflow time required to prepare 100 analysis-ready segmentation images was reduced from

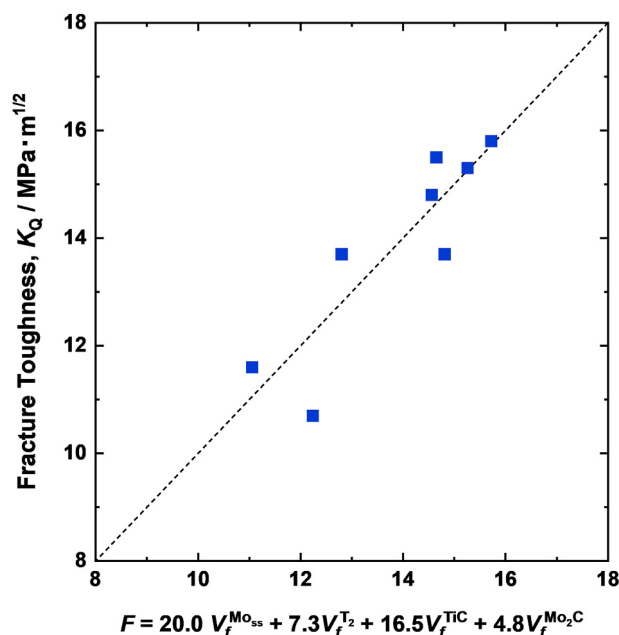


Fig. 9. Empirical relationship between the estimated phase volume fractions and fracture toughness used in this study, together with the nine experimental data points from which it was derived. This relationship was used only to evaluate the propagation of image-sampling-induced uncertainty into the fracture-toughness estimate and should not be interpreted as a universal fracture-toughness prediction model.

approximately 300 days for full manual pixel-wise labeling to approximately one working day. Phase volume fractions were then determined from the resulting segmentation images. Using bootstrap resampling, the number of microstructural images required to obtain statistically reliable phase volume fractions representative of the whole alloy was investigated. The bootstrap-estimated mean volume fractions for the entire alloy were $42.7 \pm 0.46\%$ for the Mo_{ss} phase, $36.1 \pm 0.63\%$ for the Ti_2 phase, $18.9 \pm 0.64\%$ for the TiC phase, and $2.27 \pm 0.19\%$ for the Mo_2C phase. As the number of images used in the analysis increased, the standard deviation of the mean volume fraction for each phase decreased in a power-law relationship. This trend was linear on a log-log plot, indicating that the number of images required to satisfy a prescribed precision criterion can be determined from this relationship. When the number of images used for analysis was reduced, the estimated volume fraction of the TiC phase in particular could deviate markedly from its mean value, suggesting that analyses based on only a few images may misestimate phase volume fractions by more than 15%. Fracture toughness was then estimated from the microstructural analysis using an experimentally derived relationship between phase volume fractions and fracture toughness. The calculated fracture toughness exhibited scatter arising from the limited observation area. Under the assumption that the adopted empirical relationship is valid, approximately 60 images acquired at $2000\times$ were estimated to be necessary to reduce the image-sampling-induced scatter in the fracture-toughness estimate to a level comparable to the experimental error. Because the empirical relationship was derived from a limited dataset, this number should be regarded as a conditional estimate for the present alloy, imaging condition, segmentation workflow, and empirical relationship, rather than as a universal requirement for fracture-toughness prediction.

CRediT authorship contribution statement

Chihana Kudo: Writing – original draft, Visualization, Methodology, Investigation, Data curation. **Masahiko Demura:** Writing – review & editing, Methodology, Conceptualization. **Akihiro Endo:**

Methodology, Software, Formal analysis. **Kenji Nagata:** Methodology. **Kyosuke Yoshimi:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Data availability

Data will be made available on request.

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.

Acknowledgements

This work was supported by JSPS Grant-in-Aid for Scientific Research (A) (Grant Number JP21H04606) and the MEXT Program: Data Creation and Utilization-Type Material Research and Development Project (Grant Number JPMXP1122684766). Chihana KUDO acknowledges support from the International Joint Graduate Program in Materials Science (GP-MS) at Tohoku University. This work was conducted in part using the RDE [36] provided by the Materials Data Platform (MDPF) of the National Institute for Materials Science (NIMS).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jmrt.2026.07.282>.

References

- Demura M. Challenges in materials integration. *ISIJ Int* 2024;64:503–12.
- Song L, Wang D, Liu X, Yin A, Long Z. Prediction of mechanical properties of composite materials using multimodal fusion learning. *Sens Actuators A Phys* 2023;358:114433.
- Mototake Y, Ito K, Demura M. Quantitative prediction of fracture toughness (KIC) of polymer by fractography using deep neural networks. *Sci Technol Adv Mater Methods* 2022;2:310–21.
- Sawai K, Chen T-T, Sun F, Ogawa T, Adachi Y. Image regression analysis for linking the microstructure and property of steel. *Results Mater* 2024;21:100526.
- Wang ZL, Ogawa T, Adachi Y. Property predictions for dual-phase steels using persistent homology and machine learning. *Adv Theory Simul* 2020;3:1900227.
- Li Y, Ma Y, Tan KH, Qian H, Liu T. Microstructure-informed deep learning model for accurate prediction of multiple concrete properties. *J Build Eng* 2024;98:111339.
- Gallagher B, Rever M, Loveland D, Mundhenk TN, Beauchamp B, Robertson E, et al. Predicting compressive strength of consolidated molecular solids using computer vision and deep learning. *Mater Des* 2020;190:108541.
- Moriyama T, Yoshimi K, Zhao M, Masnou T, Yokoyama T, Nakamura J, et al. Room-temperature fracture toughness of MoSIBTiC alloys. *Intermetallics* 2017;84:92–102.
- Miyamoto S, Yoshimi K, Ha SH, Kaneko T, Nakamura J, Sato T, et al. Phase equilibria, microstructure, and high-temperature strength of TiC-added Mo-Si-B alloys. *Metall Mater Trans A* 2014;45:1112–23.
- Yoshimi K, Nakamura J, Kaneko D, Yamamoto S, Maruyama K, Katsui H, et al. High-temperature compressive properties of TiC-added Mo-Si-B alloys. *JOM* 2014;66:1930–8.
- Yamamoto S, Yoshimi K, Kim J, Yokoyama K. Effects of microstructure on high-temperature strength of TiC-added Mo-Si-B alloys. *J Jpn Inst Met Mater* 2016;80:51–9.
- Kamata SY, Kaneko D, Lu Y, Sekido N, Maruyama K, Eggeler G, et al. Ultrahigh-temperature tensile creep of TiC-reinforced Mo-Si-B-based alloy. *Sci Rep* 2018;8:10487.
- Yoshimi K. MoSIBTiC alloy continuing to evolve through microstructural control: current status of development and material properties. *J Gas Turb Soc Jpn* 2022;50:318–26.
- Hatakeyama T, Sekido N, Yoshimi K. Effect of TiSi3 phase on mechanical properties and oxidation resistance of MoSIBTiC-based alloys. *J Gas Turb Soc Jpn* 2018;46:486–94.
- Dengiz O, Smith AE, Nettleship I. Grain boundary detection in microstructure images using computational intelligence. *Comput Ind* 2005;56:854–66.
- Kim D, Liu JJ, Han C. Determination of steel quality based on discriminating textural feature selection. *Chem Eng Sci* 2011;66:6264–71.
- Dutta S, Barat K, Das A, Das SK, Shukla AK, Roy H. Characterization of micrographs and fractographs of Cu-strengthened HSLA steel using image texture analysis. *Measurement* 2014;47:130–44.
- Ajioka F, Wang ZL, Ogawa T, Adachi Y. Development of high accuracy segmentation model for microstructure of steel by deep learning. *ISIJ Int* 2020;60:954–9.
- Shen C, Wang C, Huang M, Xu N, van der Zwaag S, Xu W. A generic high-throughput microstructure classification and quantification method for regular SEM images of complex steel microstructures combining EBSD labeling and deep learning. *J Mater Sci Technol* 2021;93:191–204.
- Shen C, Wei X, Wang C, Xu W. A deep learning method for extensible microstructural quantification of DP steel enhanced by physical metallurgy-guided data augmentation. *Mater Charact* 2021;180:111392.
- Ostomujof TM, Purohit RPR, Breumier S, Gey N, Salib M, Germain L. Deep learning for automated phase segmentation in EBSD maps: a case study in dual-phase steel microstructures. *Mater Charact* 2022;184:111638.
- Dergaoui A, Tu S, Vo P. Framework for generative artificial intelligence enhanced microscopy image analysis automation of metallic materials: a case study. *Microsc Microanal* 2024;30(Suppl 1):170.
- Hirabayashi Y, Iga H, Ogawa H, Tokuta S, Shimada Y, Yamamoto A. Deep learning for three-dimensional segmentation of electron microscopy images of complex ceramic materials. *npj Comput Mater* 2024;10:46.
- Strohmann T, Bugelnig K, Breitbarth E, Wilde F, Steffens T, Germann H, et al. Semantic segmentation of synchrotron tomography of multiphase Al-Si alloys using a convolutional neural network with a pixel-wise weighted loss function. *Sci Rep* 2019;9:19611.
- Soboleva N, Mushnikov A. Improving the accuracy of semantic segmentation of carbides in the microstructure of composite coatings by the neural network. *Mater Today Commun* 2024;38:108276.
- Zhou P, Zhang X, Shen X, Shi H, He J, Zhu Y, et al. Multi-phase material microscopic image segmentation for microstructure analysis of superalloys via modified U-Net and rectify strategies. *Comput Mater Sci* 2024;242:113063.
- Medghalchi S, Kortmann J, Lee SH, Karimi E, Kerzel U, Korte-Kerzel S. Automated segmentation of large image datasets using artificial intelligence for microstructure characterisation and damage analysis. *Mater Des* 2024;243:113031.
- Ghauri H, Tafreshi R, Mansoor B. Toward automated microstructure characterization of stainless steels through machine learning-based analysis of replication micrographs. *J Mater Sci Mater Eng* 2024;19:4.
- Taguchi M, Hirokawa S, Yasuda I, Tokuda K, Adachi Y. Microstructure detection by advanced image processing. *Tetsu-To-Hagane* 2017;103:142–8.
- Kim H, Arisato Y, Inoue J. Unsupervised segmentation of microstructural images of steel using data mining methods. *Comput Mater Sci* 2022;201:110855.
- Komenda J. Automatic recognition of complex microstructures using the image classifier. *Mater Charact* 2001;46:87–92.
- Holm EA, Cohn R, Gao N, Kitahara AR, Matson TP, Lei B, et al. Overview: computer vision and machine learning for microstructural characterization and analysis. *Metall Mater Trans A* 2020;51:5985–99.
- Bulgarevich DS, Tsukamoto S, Kasuya T, Demura M, Watanabe M. Pattern recognition with machine learning on optical microscopy images of typical metallurgical microstructures. *Sci Rep* 2018;8:2078.
- Bulgarevich DS, Tsukamoto S, Kasuya T, Demura M, Watanabe M. Automatic steel labeling on certain microstructural constituents with image processing and machine learning tools. *Sci Technol Adv Mater* 2019;20:532–42.
- Underwood EE. Quantitative stereology. Reading, MA: Addison-Wesley Publishing Company; 1970.
- Fujima J, Yoshikawa H, Nagao H, Tosaka H, Kadohira T, Demura M. Research data express: a data accumulation and sharing system for digital transformation in materials research. *Sci Technol Adv Mater Methods* 2025;5:2597702.
- Arganda-Carreras I, Kaynig V, Rueden C, Eliceiri KW, Schindelin J, Cardona A, et al. Trainable weka segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* 2017;33:2424–6.
- Kanuri N, Abdelkarim AZ, Rathore SA. Trainable WEKA (Waikato environment for knowledge analysis) segmentation tool: machine-learning-enabled segmentation on features of panoramic radiographs. *Cureus* 2022;14.
- Polan DF, Brady SL, Kaufman RA. Tissue segmentation of computed tomography images using a random forest algorithm: a feasibility study. *Phys Med Biol* 2016;61:6553–69.