Exploration of outliers in strength-ductility relationship of dual-phase steels

Takayuki Shiraiwa, Shoya Kato, Fabien Briffo and Manabu Enoki

Department of Materials Engineering, School of Engineering, The University of Tokyo, Tokyo, Japan

ABSTRACT
To overcome the trade-off relationship between tensile strength and elongation of dual-phase steels, three exploratory techniques were utilized: Bayesian optimization (BO), BoundLess Objective-free eXploration (BLOX), and one-class support vector machine (OCSVM). The BO is an optimization method using Gaussian process regression, and the BLOX and OCSVM are designed for finding out-of-trend materials. Initially, a large number of synthetic phase distributions of ferrite and martensite were prepared by the Gaussian random field method. Feature importance analysis was then performed to extract microstructural descriptors used as explanatory variables in the explorations. The results revealed that the volume fraction of martensite, the higher-order principal component of the spatial correlation function, and the first principal component of the persistent homology are important in predicting the product of strength and elongation (TS × EL). In the three exploration methods, crystal plasticity finite element analyses with a ductile damage model were repeatedly performed to predict strength and elongation. The BO and OCSVM successfully found microstructures with TS × EL much higher than those of random search in a limited number of iterations. The microstructures discovered in the explorations indicated that two approaches to material design are effective in improving tensile properties of dual-phase steels: uniformly dispersing the fine hard phase and constructing a lamellar structure of the hard and soft phases.

1. Introduction
Steel materials are used in various structures in our modern society, such as ships, aircraft, and other transportation equipment, buildings, and bridges, and are deeply related to our daily lives. These structures are becoming larger, lighter, and more complex to meet the social demand, and the required properties are becoming diverse. In the automotive industry, weight reduction and collision safety are strongly required from the viewpoint of global environmental protection and occupant safety. In order to achieve both weight reduction and collision safety, it is necessary to develop steel materials with high tensile strength (TS) and elongation at failure (EL). However, increasing TS generally tends to decrease EL, which is well known as a ‘strength-ductility trade-off.’ To overcome this trade-off relationship, research and...
development of advanced high strength steels (AHSS) are actively conducted. Currently, a product of TS and EL (TS × EL) of 30 GPa-% or higher is recognized as an accepted benchmark for next-generation automotive steels [1]. To achieve this goal, transformation-induced-plasticity (TRIP) steels with retained austenite [2] and quenching and partitioning (Q&P) steels [3] are being developed as third-generation AHSSs. Dual-phase (DP) steels composed of soft ferrite and hard martensite phases are also one of the candidates for next-generation automotive steels. The DP steels have a history of roughly 50 years [4], but they still leave unsolved questions and have the potential to improve TS × EL at a low manufacturing cost.

The tensile properties of the DP steels are affected not only by the respective properties of the ferrite and martensite phases but also by the spatial arrangement of the two phases. Experimental studies have shown that the chain-like networked dual-phase structure improves the tensile strength compared to an isolated structure [5]. The strength is also enhanced by increasing the martensite volume fraction and grain refinement, but total elongation is less affected by grain refinement [6,7]. To achieve the goal of TS × EL ≥ 30 GPa-%, it is necessary to maximize total elongation as well as tensile strength. The EL is a sum of uniform elongation and local elongation. In general, uniform elongation plays a vital role in formability when processing sheet metal, while total elongation is a decisive factor in other processes such as bending and counter-sinking processes [8]. It has been reported that uniform elongation has a certain relationship with yield strength from a large amount of tensile experimental data on steel materials [9]. In contrast, local elongation does not correlate well with other tensile properties (yield stress, tensile strength, uniform elongation) and is one of the most difficult properties to predict. One of the reasons is that the local deformation is a process of void formation, growth, and coalescence in the later stage of the tensile test and is strongly affected by the microstructure. In the tensile tests of the DP steels, in-situ observation [10,11], digital image correlation analysis [12–14], and X-ray CT [15,16] revealed that the distribution of the two phases with different hardnesses causes local heterogeneous stress and strain, resulting in cracking of martensite grains, decohesion at ferrite-martensite interface, and fracture within the ferrite grains. These fracture mechanisms change depending on the martensite volume fraction and grain size [4]. Therefore, in order to predict the elongation of DP steels, simple equations are not sufficient, and numerical analysis that takes into account strain localization due to microstructure and multiple fracture mechanisms is necessary. Recently, by incorporating the two-phase microstructure and the ductile fracture models based on these experimental results into finite element method (FEM), prediction of local elongation is becoming possible [17], but the detailed relationship with the morphology of microstructure and the product of TS and EL has not been clarified.

Against this background, our group aims to optimize the microstructures of DP steels by forward and inverse analysis of the structure-property linkage of DP steel. The strategy is shown in Figure 1. The strategy consists of developing a forward analysis method using a numerical model based on experimental observations, and proposing an inverse analysis method to optimize the microstructure by repeatedly performing the forward analysis. Our previous work has developed the forward analysis method to predict the void formation in DP steels by reproducing the polycrystalline structure and incorporating the ductile damage model into the FEM [18]. Although the computational cost is higher than conventional FEM, it is possible to perform several tens to several hundreds of calculations in a realistic amount of time with today’s computational resources. Therefore, it is expected to find microstructures with higher TS × EL by

![Figure 1](image-url)
performing calculations multiple times with randomly changing the microstructure using an efficient exploration method. Bayesian optimization (BO) is currently one of the most commonly used tools in optimization problems and is suitable for finding the optimal point for such computationally expensive functions. The BO has been used in a wide range of fields, including optimization of hyperparameters in machine learning models [19], and has also been used in the materials field for optimization of welding conditions [20,21]. From another point of view, since TS and EL of DP steels follow the so-called banana curve (TS × EL is constant), the problem corresponds to the search for microstructures with properties that deviate from the trend. Terayama et al. proposed boundless objective-free exploration (BLOX) as a method to search for out-of-trend materials, and successfully found light-absorbing molecules with high molar absorption coefficients [22]. Therefore, BLOX may help find outliers that deviate from the banana curve of DP steels. One-class support vector machine (OCSVM) is another method of outlier detection, which uses a support vector machine (SVM) [23], and has been applied to anomaly detection of electrocardiography [24], and structural health monitoring [25]. In materials fields, support vector machine has also been shown to be suitable for optimizing the conditions of additive manufacturing (AM) [26]. The number of trials is more limited for AM experiments due to the high cost of each trial, but by using SVM, they have succeeded in extracting the appropriate manufacturing conditions in only 11 trials.

To explore efficiently with a limited number of trials, it is generally necessary to select appropriate explanatory variables. As mentioned above, it is known that the volume fraction of the martensite phase and grain morphologies affect TS and EL in DP steels. Alaie et al. reported on the effect of martensite grain width on elongation by combining in-situ observation and finite element analysis (FEM) [11]. In addition, Abid et al. performed finite element analysis by changing the morphology of the martensite phase and pointed out that the larger aspect ratio of the martensite grain provides the higher uniform elongation (UEL) [27]. Therefore, it is necessary to consider the geometric shape parameters of the martensite phase as candidate explanatory variables. Recently, spatial correlation function [28] and persistent homology [29], which are more statistical descriptors for phase distributions, have also been proposed as effective descriptors to determine material properties. It has been reported that the spatial correlation function is useful in predicting the yield strength of dual-phase metallic composite [30]. The persistent homology has been reported to be helpful in characterizing amorphous materials and phase-field analysis results [31]. However, few studies have compared the spatial correlation function and the persistent homology in exploration or optimization problems. To extract appropriate explanatory variables, it would be necessary to statistically examine the importance of various descriptors for predicting TS × EL.

In this work, two existing methods and one new method were applied to explore the DP steels with high TS × EL: BO, BLOX, and OCSVM-based methods. The strategy is to create a large number of microstructures with a wide range of morphology in advance, and then search within the microstructural database by crystal plasticity finite element method (CPFEM) with ductile damage model. Since the effective explanatory variables in the exploration were not clear, various geometric parameters, topological quantities, and statistic values related to the phase distribution were prepared as candidates for explanatory variables, and descriptors with a larger impact on TS × EL were extracted by conducting importance analysis. Using the extracted descriptors, microstructures with higher TS × EL were explored by three methods. The differences in the exploration methods were also examined.

2. Numerical methods

2.1. Database of synthetic phase distributions

The material investigated in this study is DP steel, which consists of ferrite and martensite phases. To create finite element models, it was assumed that there was no third phase, there was no chemical heterogeneity in each phase, and each phase was composed of multiple crystal grains. Under these assumptions, a two-dimensional (2D) digital image of two-color patterns was called a ‘phase map’ and was used to define the spatial distribution of each phase. In order to enable the exploration of a wide range of phase morphologies, it is necessary to create a large number of phase maps. The open-source software DREAM.3D [32], which is often used for the reconstruction of polycrystalline models, can reproduce a variety of microstructures by applying multiple filters, but the number of models generated is limited in terms of computational cost because it is based on Voronoi tesselation. Therefore, the Gaussian random field (GRF) method [33,34] with the advantage of faster computation was selected for the generation of the phase maps, rather than the point-based method such as Voronoi tessellation or the physics-based method such as the phase-field method. In addition, it was considered necessary to include anisotropic shapes and extreme geometries such as uniform and lamellar distributions in the phase maps to further expand the search range. For this reason, anisotropy and periodicity were introduced into the GRF method. The details are described hereafter.
The martensite volume fraction was determined randomly between 0.25 and 0.9. This range was determined with reference to the DP steels used in the experimental calibration of crystal plasticity parameters [18]. Considering the computational cost and the microstructure of the DP steels, the size of the phase map was set to 128 × 128 pixels. In this study, 1 pixel was equivalent to 1 μm. To change the periodicity of the phase maps, the size of the GRF was defined as \(2^m \times 2^n\) pixels, where \(n\) and \(m\) were integer values obtained by randomly sampling from an exponential distribution with a mean of 1.5 and rounding the values toward zero. The upper limit of \(n\) and \(m\) was set to 7. These \(n\) and \(m\) correspond to the number of subimages for tiling, which will be explained later. Under these conditions, the size of the GRF was determined to be a rectangle with each side length of 1, 2, 4, 8, 16, 32, 64, or 128. Figure 2 shows an example when a size of the GRF is 64 × 128 pixels. After determining the GRF size, uncorrelated Gauss random noise, \(U(z)\), was obtained by randomly sampling values from a Gaussian distribution with a mean of 0 and a variance of 1 (standard normal distribution) and placing them in the 2D space (Figure 2(a)). A weight function, \(\omega(h)\) was then defined by a 2D probability distribution as shown in Figure 2(b). Although the isotropic 2D distribution was used as \(\omega(h)\) in the original papers [33,34], it was defined by a two-variable Gaussian distribution with horizontal and vertical variances and a rotation angle in order to generate an anisotropic phase distribution. The two variances were randomly selected from a lognormal distribution with a mean of 20 and a variance of 10². The rotation angle was determined from a uniform random distribution from 0° to 90°. This procedure is expected to provide an anisotropic phase distribution where each phase is elongated in the major axis of the weight function. The rotation angle corresponds to the elongated direction of the phase map. The Gaussian random field, \(G(z)\), was calculated by convolution of the uncorrelated Gauss random noise, \(U(z)\), and the weight function, \(\omega(h)\), assuming horizontal and vertical periodicity at the edges of the 2D space (Figure 2(c)). The \(G(z)\) was binarized by setting a threshold to satisfy a predefined volume fraction of martensite (Figure 2(d)). By tiling the binary image generated by the GRF method, a phase map with a size of 128 × 128 pixels was finally obtained (Figure 2(d)). The size of the phase map was fixed at 128 × 128 pixels regardless of the size of the GRF. By repeating the above procedure, a total of 100,000 phase maps were prepared and stored in a database.

### 2.2. Microstructural descriptors

Since the phase maps generated in the previous section are digital data of 128 × 128 dimensions, appropriate low-dimensional explanatory variables need to be extracted from the phase map for the exploration. These explanatory variables obtained from the phase

---

**Figure 2.** Procedure for constructing a phase map by anisotropic Gaussian random field method and tiling. (a) uncorrelated Gaussian random noise, (b) weight function, (c) random gaussian field assuming periodicity, (d) binarization, (e) tiling. The size of the unit 2D model generated by the GRF method is 64 × 128 pixels in this example.
maps are referred to as microstructural descriptors hereafter. The microstructural descriptors used in this study are summarized in Table 1.

### 2.2.1. Geometric descriptors

One of the most commonly used descriptors for quantifying the microstructure of DP steels is the volume fraction of martensite ($V_{FM}$). It has been widely reported that tensile strength strongly depends on $V_{FM}$ in DP steels. Also, the fracture mechanism of DP steels varies with $V_{FM}$ to martensite cracking and martensite-ferrite decohesion [4], which may affect the elongation to failure. In addition to $V_{FM}$, the stress–strain curve changes significantly depending on the morphology of the martensite grains. This has been confirmed both experimentally and by numerical calculations. One of the reasons is that the stress concentration induced by martensite grains on the adjacent ferrite phase is dependent on the shape of the martensite grains. To consider the effect of stress concentration, the local curvature of martensite grains was calculated as shown in Figure 3(a), and the minimum value in the phase map was recorded as a microstructural descriptor. The anisotropy of the martensite shape also affects the stress–strain curve [17]. To account for the anisotropy, all martensite grains in a phase map were fit into ellipses, and the mean value of the angle between the loading direction and the major axis was calculated (Figure 3(b)). Also, the square root of the area of the martensite grains was calculated to consider the fineness of the martensite grains. In addition, perimeter and circularity were extracted as typical parameters representing the geometry of martensite grains. These average values were recorded as descriptors.

### 2.2.2 Topological descriptors

Recent studies have reported that the connectivity of martensite grains has a considerable effect on tensile strength. A chain-like networked dual-phase structure improves tensile strength compared to an isolated structure with the same $V_{FM}$, without significant loss of elongation [5]. Topological parameters were considered useful for quantifying such connectivity. As typical topological parameters, 0th Betti number ($\beta_0$), 1st Betti number ($\beta_1$), and Euler characteristic ($\chi$) of the martensite phase were computed using Matlab programs. The $\beta_0$ represents the number of connected martensite grains, and the $\beta_1$ represents the number of holes in the martensite phase. The Euler characteristic is the alternating sum of the Betti numbers, i.e., $\chi = \beta_0 - \beta_1$ for the 2D models.

### 2.2.3. Covariogram

With the recent development of applied mathematics, more statistical analysis methods have been proposed to quantify the microstructural morphology. In this paper, three types of statistical analysis methods were examined: covariogram, spatial correlations, persistent homology. The covariogram was proposed by Kregel et al. [35] to quantify the banded structures of ferrite and martensite in DP steels. The principle of the covariogram is based on the covariance function proposed by Matheron [36] in the 1960s. The covariance function, $C(X, r)$, is the probability that both extremities of vector $r$ belong to phase $X$. Fixing the direction of $r$ to be perpendicular to the loading direction, the evolution of $C(X, r)$ as a function of the magnitude of $r$, is defined as covariogram. An example of a covariogram of martensite is shown in Figure 3(c). The covariogram oscillates with several periods when the phase distribution has a banded structure parallel to the loading direction. The mean distance between bands is estimated by the position of the first maximum ($HM$), and the band intensity is measured by the difference between the first maximum value and the first minimum value ($BI$). Random structures show lower $BI$ but greater than zero. These $BI$ and $HM$ were used as the microstructural descriptors.

### 2.2.4 Spatial correlation

A more general method to quantify the spatial correlation of multiple phases in two or three-dimensional space is the $n$-point correlation function [37]. This is also known as $n$-point statistics. Recently, Kalidindi and co-workers have developed a framework for quantifying microstructures based on the $n$-point correlation functions [38]. It has been demonstrated that the original image can be reconstructed from the 2-point correlation function if the probability of being in a certain phase at all points is 0 or 1. In other words, it is possible to uniquely determine higher-order correlation functions (3-point or 4-point correlation function, etc.) from the 2-point correlation function in such cases. Therefore, this study only dealt with the 2-point correlation function. When there are $N$ different phases, $N{H}_2$ types of 2-point correlation functions can be calculated, including autocorrelation and cross-correlation, where $H$ is the symbol of combinations with repetition. Due to the relationship of each 2-point correlation function, the number of independent functions is $N - 1$. Thus, only a 2-point correlation function between martensites, $f_r^MM$, was calculated. The definition of $f_r^MM$ is exactly the same as the covariance function, $C(X, r)$, described above, but it differs from the covariogram in that the $f_r^MM$ quantifies the phase distribution regardless of the loading direction. According to the literature [39], the $f_r^MM$ was calculated by the fast Fourier transform (FFT). Since the $f_r^MM$ has the same dimensional information as the original image, principal component analysis (PCA) was conducted to reduce the dimension of the $f_r^MM$ using the 100,000 phase maps created.
| Classification      | Symbol | Description                                                                 | Minimum | Maximum | Mean  | Variance |
|---------------------|--------|------------------------------------------------------------------------------|---------|---------|-------|----------|
| Geometric descriptors | VFₐ₈  | Volume fraction of martensite                                                  | 0.25    | 0.9     | 0.576 | 0.035    |
|                     | Cur    | Minimum value of local curvature of martensite grains (unit: µm)              | 0.01    | 2.835   | 1.391 | 0.089    |
|                     | Orient | Average value of the angle between the loading direction and the major axis of the ellipse fitted to the martensite grains (unit: °) | 0       | 90      | 43.6   | 496.0    |
|                     | M/√Area| Average of the square root of the area of martensite grains (unit: µm)        | 4.56    | 121.4   | 60.55 | 1751.9   |
|                     | Perimeter | Average of perimeter of martensite grains (unit: µm)                        | 6.83    | 2116    | 376.9 | 118576.1 |
|                     | Circularity | Average of circularity of martensite grains                                  | 0.01    | 1.00    | 0.463 | 0.094    |
| Topological descriptors | Betti0 | 0th Betti number of martensite phase                                          | 1       | 100     | 5.859 | 49.296   |
|                     | Betti1 | 1st Betti number of martensite phase                                          | 1       | 128     | 9.909 | 84.215   |
|                     | Euler  | Euler characteristic of martensite phase                                     | –127   | 96      | –4.051| 183.22   |
| Covariogram         | HM     | Position of the first maximum peak of the covariogram (unit: µm)            | 1       | 65      | 14.077| 101.22   |
|                     | BI     | Difference between the first maximum and minimum of the covariogram         | 0.01    | 0.297   | 0.03  | 0.002    |
| Spatial correlation | SC1    | 1st principal component of the 2-point correlation function between martensites | –37.49 | 33.24   | –11.21| 432.3    |
|                     | SC2    | 2nd principal component of the 2-point correlation function between martensites | –4.551 | 6.319   | 0.33  | 0.057    |
|                     | SC3    | 3rd principal component of the 2-point correlation function between martensites | –5.791 | 0.579   | –0.56 | 0.068    |
|                     | SC4    | 4th principal component of the 2-point correlation function between martensites | –2.501 | 7.992   | –0.066| 0.039    |
|                     | SC5    | 5th principal component of the 2-point correlation function between martensites | –7.094 | 7.692   | –0.035| 0.091    |
|                     | SC6    | 6th principal component of the 2-point correlation function between martensites | –0.673 | 4.922   | –0.019| 0.02     |
| Persistent homology | PH1    | 1st principal component of the 0th persistence diagram after applying PWG kernel | –31.8  | 155.8   | 11.2  | 381.3    |
|                     | PH2    | 2nd principal component of the 0th persistence diagram after applying PWG kernel | –34.7  | 131.6   | 11.6  | 132.3    |
|                     | PH3    | 3rd principal component of the 0th persistence diagram after applying PWG kernel | –72.1  | 93.1    | 2.805 | 87.17    |
|                     | PH4    | 4th principal component of the 0th persistence diagram after applying PWG kernel | –56.5  | 94.6    | 8.28  | 80.46    |
|                     | PH5    | 5th principal component of the 0th persistence diagram after applying PWG kernel | –30.4  | 80.1    | 8.372 | 47.96    |
|                     | PH6    | 6th principal component of the 0th persistence diagram after applying PWG kernel | –63.0  | 45.8    | –2.7  | 30.7     |
in the previous section, and the top six principal components (SC1, ..., SC6) were used as the microstructural descriptors.

2.2.5. Persistent homology

The last descriptor is persistent homology. This is a recently proposed concept for quantifying topological features. While dilating or eroding the phase of interest, the dilated values when the connected components and holes of the phase appear (birth, \( b \)) or disappear (death, \( d \)) are computed. The plot of \( (b, d) \) is referred to as the persistent diagram (PD). Here, martensite was chosen as the phase of interest. Similar to the Betti numbers, the plot of the birth and death of the connected components is referred to as 0th PD, and the plot of the holes is referred to as 1st PD. An example of the 0th PD of the martensite is shown in Figure 3(e). The PD was computed by HomCloud [40]. There are two main problems in using the PD for statistical analysis. One is that data points near the diagonal in the PD have short lifetimes and are sensitive to small perturbations of the digital images [41]. The other problem is that since the number of data points in the PD is limited due to the size of the image, the principal components obtained by PCA are excessively dependent on the positions of the data points. To overcome these two problems, the persistence weighted Gaussian (PWG) kernel [42] was applied to the 0th PD of the martensitic phase. The PWG kernel is a process that reduces the weights of data points close to the diagonal and applies

---

**Figure 3.** Schematic diagrams and examples of microstructural descriptors: (a) curvature of martensite, (b) orientation of martensite, (c) covariogram, (d) 2-point spatial correlation (e) 0th persistent homology, (f) 0th persistent homology after applying persistence weighted Gaussian (PWG) kernel. The same phase map was used in the calculations of (a)-(f) as examples.
2.3. Crystal plasticity finite element method (CPFEM)

For a phase map of interest, a regular mesh was created with 128 \times 128 8-node hexahedral elements (C3D8), and the phases were assigned to correspond perfectly to the phase map. The shape and orientation of the polycrystalline grains in each phase were determined by the previously developed method based on the anisotropic Voronoi tessellation and the random sequential addition algorithm [43]. Since the spatial distribution of the two phases is of interest in this work, the distributions of the grain size and the crystallographic orientation were fixed. The grain size distribution of the two phases followed a normal distribution with a mean value of 2 \mu m. The crystallographic orientations of the ferrite and martensite grains were sampled from orientation distribution functions obtained experimentally from the DP steels without a specific texture [18].

The tensile strength and elongation to failure were calculated by the crystal plasticity finite element method (CPFEM) with a damage model. A crystal plasticity model with 24 slip systems of the \{101\}\langle11\bar{1}\rangle and \{112\}\langle11\bar{1}\rangle was used to calculate the stress–strain curve. The shear rate on each slip system was phenomenologically modeled by a power law:

\[
\dot{\gamma}_s^\alpha = \dot{\gamma}_0 \left( \frac{\tau_s^\alpha}{\tau_c^\alpha} \right)^n \text{sgn}(\tau_s^\alpha) \quad (1)
\]

where \(\dot{\gamma}_0\) is the reference shear rate (\(\dot{\gamma}_0 = 10^{-2}\)s), \(n\) is the strain rate exponent (\(n = 20\)), \(\tau_s^\alpha\) is the resolved shear stress of slip system \(\alpha\), and \(\tau_c^\alpha\) is the critical resolved shear stress. To represent work hardening, the \(\tau_c^\alpha\) was defined as a function of slip–slip interaction:

\[
\tau_c^\alpha = \tau_0^\alpha + \int \dot{\tau}_s^\alpha \, dt \quad (2)
\]

with

\[
\dot{\tau}_s^\alpha = \frac{dr_s^\alpha}{dt} + \sum_{\beta=1}^{N_s} h_s^{\alpha \beta} \left| \dot{\gamma}_s^\beta \right| \quad (3)
\]

\[
\tau_s^\alpha = \tau_0^\alpha \left\{ 1 - \exp \left( \frac{\left| \dot{\gamma}_s^\alpha \right|}{\Gamma_s} \right) \right\} \quad (4)
\]

where \(\tau_0^\alpha\) is the initial CRSS, \(\tau_s^\alpha\) is the saturated CRSS, \(\Gamma_s\) is the initial hardening rate, and \(\Gamma_s\) is the accumulated shear strain on all the slip systems and \(h_s^{\alpha \beta}\) is the interaction coefficient of the slip system \(\beta\) over \(\alpha\) and assumed to be equal to 1. The parameters of the CP model for the two slip system families were assumed to be the same. The material constants used in the CPFEM are listed in Table 2. These parameters were taken from a previous work [18], where the crystal plasticity parameters were calibrated from the stress–strain curve of the DP steel with a tensile strength of approximately 1000 MPa. The material used in the experiment had a relatively random phase distribution. To express ductile fracture behavior, a single scalar damage variable, \(D\), ranging from 0 for the undamaged state to 1 for the fully damaged state was introduced into the CPFEM. Using the damage variable, the stress tensor is given by [44]:

\[
\sigma = (1 - D) \sigma_0 = (1 - D) C : \varepsilon_e \quad (5)
\]

where \(\sigma\) is effective stress tensor, \(C\) is stiffness tensor, and \(\varepsilon_e\) is elastic strain tensor. The damage variable was simply defined as a linear function of the equivalent plastic strain [45]:

\[
D = \min \left( \frac{\ddot{\varepsilon}_p^p - \ddot{\varepsilon}_p^0}{0.95}, 1 \right) \quad (6)
\]

*Table 2. Material constants used in crystal plasticity finite element analysis.*

|          | \(C_{11}/\text{GPa}\) | \(C_{12}/\text{GPa}\) | \(C_{44}/\text{GPa}\) | \(\tau_s/\text{MPa}\) | \(\tau_c/\text{MPa}\) | \(b_1/\text{MPa}\) | \(\ddot{\varepsilon}_0\) | \(\ddot{\varepsilon}_f\) |
|----------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| Ferrite  | 233.3                  | 135.5                  | 118.0                  | 100                    | 70                     | 1500                   | 0.11                   | 1.2                    |
| Martensite | 268.0                 | 111.0                  | 79.0                   | 600                    | 210                    | 10000                  | 0.05                   | 0.5                    |
where $\varepsilon_p$ is an equivalent plastic strain, $\varepsilon_p^d$ is the damage initiation strain, $\varepsilon_p^c$ is the strain at complete failure. The upper limit of 0.95 was defined for the convergence issues. The onset and critical damage variables for the two phases are shown in Table 2. The damage initiation strain ($\varepsilon_p^d$) was determined by fitting the equivalent plastic strain at damage initiation in previous experimental works [18]. Fracture strain ($\varepsilon_f^p$) in tensile testing of steels typically ranges from 0.4 to 1.3 in the literature [46,47]. Here, values of 1.2 for the ferrite phase and 0.5 for the martensite phase were used. Periodic boundary conditions were imposed on all pairs of faces of the model. The stress–strain curve was calculated by applying a horizontal displacement to the side surface of the model.

### 2.4. Selection of microstructural descriptors

In general, the number of evaluations required to find the global optimum increases exponentially with the dimension of explanatory variables. A total of 23 microstructural descriptors were derived in Section 2.2, but it is difficult to explore using all of them with current limited computer resources. In Bayesian optimization (BO), which is currently one of the most promising exploration methods, the number of explanatory variables is practically limited to 10–20 even for relatively simple calculations [48]. In order to extract descriptors important for predicting tensile strength and elongation, CPFEM analysis was performed on 100 randomly selected phase maps, and the results were subjected to random forest (RF) regression to calculate the feature importance of each descriptor. The RF is an ensemble machine learning method that builds many regression trees as a weak learner. The regression trees were constructed by a method based on chi-square analysis [49]. The feature importances of descriptors were quantified by the permutation-based mean squared error reduction method [50]. It corresponds to the average value of the change in the mean squared error of the predictions when permuting out-of-bag data (data not used in training) of each decision tree. The descriptors with higher importance were used in the following exploration.

### 2.5. Exploring microstructure

#### 2.5.1. Bayesian optimization

Bayesian optimization (BO) is a method of efficiently searching for the explanatory variables that minimize (or maximize) the output of the objective function. The microstructural descriptors are the explanatory variables, the CPFEM is the objective function, and the output is TS × EL in this work. The BO was performed by the following procedure:

(i) Compute TS and EL of 10 randomly selected phase maps with CPFEM.

(ii) Perform Gaussian process regression (GPR) using the results of the CPFEM calculations to derive the mean and standard deviation of the TS × EL as a function of explanatory variables.

(iii) Calculate the acquisition function using the mean and standard deviation of the GPR model. Determine the point with the largest acquisition function among the unexplored points in the phase map database to be the next search point.

(iv) Perform CPFEM calculation using the microstructure corresponding to the search point obtained in step (iii), and save TS, EL, and explanatory variables. Return to (ii).

The steps (ii)–(iv) were repeated until reaching a certain number of iterations. An example of determining the next search point by BO is shown in Figure 4. The exploration by BO is performed in the microstructural descriptors space. The white circles in the figure represent the results of the CPFEM simulation, the gray points represent the unexplored points, and the colored distributions represent the acquisition function calculated from the GPR model. The efficiency of the optimization generally depends on the choice of the kernel function of the GPR and acquisition function. The ARD Matérn 5/2 kernel [51], which is widely used in the literature of Bayesian optimization, was used as the kernel function. For the acquisition function, the expected-improvement plus (EI+) acquisition function implemented in Matlab Statistics and Machine Learning Toolbox was used to avoid falling into the local maximum of the objective function. The EI+ acquisition function modifies the kernel function when the same point is explored repeatedly [52]. The exploration–exploitation ratio was set to the default value of 0.5. The number of iterations was set to 100. The first ten microstructures selected in a random search (RS) in Section 2.4.4 were used in the initialization step (i). In order to compare the three exploration methods avoiding the effect of randomness in the initialization step, identical microstructures were used in the initialization steps for the remaining two explorations described below.

#### 2.5.2. Boundless objective-free exploration (BLOX)

The second exploration method is a recently proposed method called Boundless objective-free exploration (BLOX) [22]. This is a method of searching for out-of-trend materials without setting an objective function. The exploration using BLOX was performed by the following procedure:
Figure 4. Example of determining the search point by Bayesian optimization. The red-colored point represents the next search point.

(i) Compute TS and EL of 10 randomly selected phase maps with CPFEM.
(ii) Apply random forest regression to the CPFEM results to generate machine learning-based models that predict TS and EL from the explanatory variables.
(iii) Derive TS and EL of all unexplored phase maps by substituting the explanatory variables into the machine learning-based models with random forest.
(iv) Calculate Stein discrepancy [53] of each unexplored point in the TS-EL space, and determine the next search point based on Stein novelty [22].
(v) Calculate TS and EL at the next search point by CPFEM. Return to (ii).

The steps (ii)–(v) were repeated up to a certain number of iterations. Since machine learning with random forest (RF) is much faster than CPFEM, most of the computation time was spent on step (v). The CPFEM provides a small number of relatively accurate predictions of TS and EL, whereas the RF provides its trends from a large number of TS and EL data. An example of determining the search point by BLOX is shown in Figure 5. The Stein novelty quantifies how much the point deviates from the existing trend in the TS-EL space. The BO searches for the next search point in the space of explanatory variables, but the BLOX searches for the next point in the space of objective variables (TS and EL). The first ten phase maps selected in initialization (step (i)) are the same as the BO.

2.5.3. One-class support vector machine (OCSVM) based exploration

The BO has the advantage of searching toward higher objective functions within the explanatory variable space. On the other hand, the BLOX has the advantage of searching in the space of objective variables and has the disadvantage that it may search for a lower TS × EL because there is no objective function. In this paper, an OCSVM-based exploration method is proposed to exploit the advantages of both methods simultaneously. In the OCSVM classification, it is assumed that all data belong to the +1 class and only the origin belongs to the −1 class [23]. Then, SVM finds the boundary that separates the two classes, and the data belonging to the −1 class side is regarded as an outlier. The signed
distance between the boundary and the data point is called the ‘classification score (CS)’ [54]. An example of calculating the classification score is shown in Figure 6(a). A negative value represents an anomaly. To find a higher TS × EL value, a product of normalized TS × EL and classification score was defined as ‘curiosity index (CI)’ for the exploration:

$$CI = \frac{TS \times EL}{\text{mean}(TS \times EL)} \cdot (-CS)$$ 

(7)

where the TS and EL of the unsearched points were obtained by SVM regression, mean(TS × EL) is the mean value of TS × EL of all phase maps, and CS is the classification score obtained by applying OCSVM to the (TS, EL) space. The exploration procedure is summarized as follows:

(i) Compute TS and EL of 10 randomly selected phase maps with CPFEM.
(ii) Apply support vector machine (SVM) regression to the CPFEM results to generate machine learning-based models that predict TS and EL from the explanatory variables.
(iii) Derive TS and EL of all unexplored phase maps by substituting the explanatory variables into the machine learning-based models with SVM.
(iv) Calculate classification score (CS) and normalized TS × EL in the TS-EL space, and determine the next search point based on curiosity index (CI).
(v) Calculate TS and EL at the next search point by CPFEM. Return to (ii).

An example of the distribution of the CI calculated in step (iv) is shown in Figure 6(b). It can be seen that CI is more prominent in the upper right of the figure. The red point in the figure has the highest CI among the unexplored points and is the next search point to be explored. In BLOX, the point with the lower TS × EL can be chosen, but in the OCSVM-based exploration, the point with the higher TS × EL is determined thanks to the normalized TS × EL in Equation (6). The exploration may change depending on the phase map selected in the initialization of step (i). To avoid such an effect of the initialization, the ten phase maps used in the initialization were the same as the other two methods.

### 3. Results and discussion

#### 3.1. Random search

To identify the microstructural descriptors that have significant effects on the tensile properties, 100 phase maps were randomly selected from the database of phase maps, and the stress–strain curves were calculated by CPFEM. The scatter plot of the tensile strength (TS) and total elongation to failure (EL) is illustrated with the curves of TS × EL = 10, 20, 30, 40 GPa-% in Figure 7(a). Most of the search points showed values close to TS × EL = 10 GPa-%, which is the same as the material used in the calibration of crystal plasticity parameters. In this random search, no microstructure exceeded 20 GPa-% in 100 iterations. The relationships of TS, EL, and TS × EL to the martensite volume fraction (VFₘ) are plotted in Figure 7(b–d). With the increase of VFₘ, TS increased, and EL decreased. This trend is consistent with the trend observed in experiments [55]. Figure 7(b) shows that TS and VFₘ are correlated, while Figure 7(c) shows that EL and VFₘ are inversely correlated. Linear regression of the relationships between VFₘ and TS [MPa] and between inverse VFₘ and EL [%] yielded the following equations:

$$TS = 2268.8VFₘ - 20.71$$

(8)

$$EL = \frac{5.887}{VFₘ} - 3.691.$$  

(9)

The coefficients of determination R² were 0.9185 and 0.7582, respectively. The values of R² did not change much when regressing with a quadratic

---

**Figure 6.** Example of determining the search point by OCSVM (one-class support vector machine): (a) classification score and (b) curiosity index. The red-colored point in (b) indicates the next search point.
equation or higher order equation. The product of TS and EL was relatively insensitive to $V_{FM}$ and showed large scatter, suggesting the difficulty of maximizing $TS \times EL$ by $V_{FM}$ alone.

The points indicated by A-C in Figure 7(a) represent the search points with maximum elongation (A), maximum tensile strength (C), and the intermediate between them (B). The microstructural descriptors and tensile properties of microstructures A-C are listed in Tables 3 and 4, respectively. Figure 8 shows crystal plasticity finite element models, distributions of damage variables, and stress–strain curves of these points. The elongation was maximum when the martensite volume fraction was low ($V_{FM} = 0.332$, Figure 8(a)). In this microstructure A, the fracture occurred mainly in the ferrite phase. As shown by the arrow in the figure, severe localization of damage was observed primarily in the ferrite grains sandwiched between two martensite grains along the loading direction. Such fracture behavior has been reported in in-situ observations during tensile testing of DP steels with similar $V_{FM}$ [11,56]. The ductile damage in the ferrite phase initiated from the early stage, and the slope of the stress–strain curve was lower than the elastic modulus. This microstructure also showed a maximum value for $TS \times EL$ ($TS \times EL = 18.4$ GPa·%). In microstructure B with intermediate martensite volume fraction ($V_{FM} = 0.573$, Figure 8(b)), the damage was observed in both ferrite and martensite phases. The martensite grains connected in the loading direction were subjected to higher stress than ferrite due to their harder properties, resulting in higher tensile strength than the microstructure A. The fracture occurred at the necked area of martensite, as indicated by the arrow. This is consistent with in-situ observations by phase-contrast

### Table 3. Microstructural descriptors of microstructures A-I.

| Microstructure | Exploration method | Cur/¹/µm | Orientation/° | Ms/Area/µm | Perimeter/µm | Circularity | Bettii0 | Bettii1 | Euler | HM/µm | B | SC6 | PH1 |
|----------------|--------------------|----------|----------------|-------------|--------------|-------------|---------|---------|-------|-------|---|----|----|
| A              | RS                 | 0.332    | 1.22           | 49.61       | 15.2         | 65.6        | 1.15    | 16      | 15    | 15    | 24 | 0.033 | 0.044 | 53.14 |
| B              | 0.573              | 1.65     | 80.41          | 97.0        | 1039.3       | 0.11        | 1       | 9       | 15    | 0.063 | 0.039 | 8.69  |
| C              | 0.847              | 1.22     | 11.25          | 38.1        | 271.0        | 0.27        | 8       | 0.24    | 0.099 | 4.77  |
| D              | BO                 | 0.201    | 0.97           | 67.66       | 6.1          | 20.5        | 1.33    | 78      | 1.77  | 2     | 0.000 | 0.356 | 176.02 |
| E              | 0.359              | 2.37     | 0.01           | 27.1        | 278.5        | 0.12        | 8       | 8       | 0.239 | 4.458 | 20.80 |
| F              | BLOX               | 0.271    | 0.97           | 6.08        | 5.9          | 42.4        | 1.63    | 35      | 3.20  | 0.063 | 0.003 | 28.12 |
| G              | 0.725              | 1.32     | 0.06           | 76.9        | 495.6        | 0.31        | 2       | 64      | 13    | 0.033 | 0.224 | 29.60 |
| H              | OCSVM              | 0.320    | 1.11           | 36.83       | 9.0          | 41.4        | 0.65    | 64      | 8     | 0.076 | 1.565 | 92.48 |
| I              | 0.539              | 2.29     | 11.26          | 32.0        | 269.7        | 0.18        | 8       | 8       | 0     | 0.150 | 3.503 | 28.55 |
X-ray microtomography, which confirmed the nucleation and growth of microvoids located in necked martensitic grains [57]. The elongation was lower than microstructure A. The value of TS × EL of microstructure B (TS × EL = 17.9 GPa-%) was almost the same as that of microstructure A. The tensile strength was maximized in microstructure C, which has a high volume fraction of martensite (VF_M = 0.847).

### Table 4. Tensile properties calculated from volume fraction of martensite VF_M (Equationa (8)–(9)) and calculated from CPFEM for microstructures A-I.

| Microstructure | Exploration method | Calculated from VF_M (Equations (8)–(9)) | Calculated from CPFEM |
|---------------|--------------------|------------------------------------------|-----------------------|
|               |                    | TS/GPa GPa | EL/ % | TS × EL/GPa. % | TS/GPa | EL/ % | TS × EL/GPa. % |
| A             | RS                 | 0.733       | 14.04 | 10.29          | 0.863   | 21.35 | 18.42          |
| B             |                    | 1.279       | 6.58  | 8.42           | 1.412   | 12.66 | 17.88          |
| C             |                    | 1.901       | 3.26  | 6.20           | 2.294   | 5.09  | 11.67          |
| D             | BO                 | 0.435       | 25.60 | 11.14          | 0.762   | 34.72 | 26.46          |
| E             |                    | 0.794       | 12.71 | 10.09          | 1.105   | 26.90 | 29.72          |
| F             | BLOX               | 0.594       | 18.03 | 10.71          | 0.917   | 20.54 | 18.83          |
| G             |                    | 1.624       | 4.43  | 7.79           | 1.908   | 9.38  | 17.89          |
| H             | OCSVM              | 0.705       | 14.71 | 10.37          | 1.000   | 30.58 | 30.59          |
| I             |                    | 1.202       | 7.23  | 8.69           | 1.470   | 19.42 | 28.55          |

Figure 8. Crystal plasticity finite element models, distributions of damage variable and stress–strain curves at search points A–C (a–c) in Figure 7. The volume fractions of martensite are 0.332, 0.573, and 0.847, respectively.
3.2. Microstructural descriptors important for predicting tensile properties

Using the results of 100 CPFEM calculations in random search (RS), the importance of each descriptor to TS, EL, and ES × EL was calculated by RF regression, respectively (Figure 9). The coefficient of determination ($R^2$) is also described in the figures. For tensile strength, the importance of the three descriptors $VF_M$, Betti1, and SC1 were significantly high. For elongation, PH1 was relatively more important than the others, but no significantly important descriptors were found. The $R^2$ of EL was lower than that of TS, indicating that the prediction of EL was more difficult than TS. For TS × EL, the feature importance of five descriptors, $VF_M$, Betti0, SC1, SC6, and PH1, were higher than the others, exceeding 0.4. The $R^2$ value was between those of TS and EL. From these results, it is unclear how many descriptors should be used for the explorations. To consider the number of descriptors required for the prediction of each tensile property, the values of $R^2$ obtained from RF regression using the top $N$ most important descriptors were investigated for different values of $N$ (Figure 9(d)). Interestingly, TS could be predicted with good accuracy using only one descriptor, but for EL, $R^2$ did not saturate until approximately ten descriptors. For TS × EL, approximately five descriptors were sufficient to reach a saturated value of $R^2$. Therefore, the five most important descriptors for TS × EL ($VF_M$, Betti0, SC1, SC6, and PH1) were focused on. In order to improve the efficiency of the exploration in the next section, it is necessary to remove the highly correlated explanatory variables. Figure 10 shows the correlation between the five descriptors. As already reported [58], when there is a large variation in the volume fraction in the database, SC1 (the first principal component of the spatial correlation function) has a one-to-one relationship with the volume fraction. Since $VF_M$ has a clear physical meaning, $VF_M$ was left

![Figure 9](image-url)
as an explanatory variable, and SC1 was not used in the exploration. There is also a strong correlation between Betti0 (the 0th Betti number) and PH1 (the first principal component of the 0th persistence diagram). This is probably because both descriptors count the holes in the martensite phase. Since Betti0 is a discrete variable and PH1 is a continuous variable, PH1 containing more information was selected as the explanatory variable. In summary, three descriptors (VF_M, SC6, PH1) were finally chosen as explanatory variables in the explorations in the next section.

The distribution of the three descriptors of the phase map database is shown in Figure 11(a). To understand the physical meaning of SC6, PH1, a scatter plot of SC6 and PH1 in the range of VF_M 0.4–0.6 is plotted with several typical examples of the microstructures in Figure 11(b). The random microstructure is located near the center point indicated by the cross mark. It can be seen that a larger SC6 means an elongated microstructure in the loading direction, and a smaller SC6 means an elongated microstructure perpendicular to the loading direction. Therefore, SC6 represents the intensity of the lamellar structure and its orientation. The PH1 expresses the fineness of the two-phase pattern because microstructures with a fine distribution of martensite phase are located at a higher PH1.

3.3. Exploration by BO, BLOX, and OCSVM

The results at 20 iterations for each exploration method are shown in Figure 12. In BO and OCSVM, the search points exceeded the curve of TS × EL = 20 GPa% by the first 20 iterations indicating that the search efficiency is much higher than RS. The BLOX did not show a clear difference from RS at this early stage. Comparing the ten initialization points and the subsequent 11–20 iteration points, it can be seen that all three methods tried to avoid the initialization points. This is because the acquisition function of BO takes a large value at unsearched points, and BLOX and OCSVM aim to find points farther from the searched points in the TS-EL space. One of the differences between BLOX and OCSVM is that the Stein novelty in BLOX does not include TS × EL,
while the curiosity index in OCSVM does. For this reason, BLOX searched in all directions from the points explored by RS. Another difference is that BLOX uses RF and OCSVM uses SVM as the machine learning to derive the TS-EL trend of the searched points. Since RF consists of a large number of decision trees, it excels at complex interpolation prediction but has difficulty in extrapolation prediction. On the other hand, the SVM used in OCSVM is a binary classifier, which makes predictions with relatively simple formulas and prevents overfitting. It is considered that OCSVM worked more effectively than BLOX due to such a difference in the machine learning methods.

The results of 100 iterations for each search method are shown in Figure 13. The search points of BO were most widely distributed in the TS-EL space among the three methods. In BO, the highest EL value (EL = 34.7%) was observed through all exploration methods. This microstructure is referred to as microstructure D. The maximum TS × EL value in BO was 29.72 GPa·% at search point E. This is much higher than the maximum value in RS (18.4 GPa·%). The maximum value of TS × EL in BLOX was not much different from that of RS. In RS, the points were distributed evenly along the curve of TS × EL = 10 GPa·% (Figure 7(a)). While in BLOX, the search points were

![Figure 11](image1_url)  
**Figure 11.** Scatter plot of phase maps in selected microstructural descriptors space. (a) all 100000 synthetic phase maps, (b) phase maps with volume fraction of martensite of 0.4–0.6. The cross mark in (b) represents the center position of the distribution.

![Figure 12](image2_url)  
**Figure 12.** Searched points in iterations 1–20. (a) Bayesian optimization, (b) Boundless objective-free exploration (BLOX), (c) One-class support vector machine (OCSVM). The light gray points represent the initialization steps.
clustered in several regions. The microstructures with large EL and TS × EL are shown in F and G in the figure, respectively. The search points of OCSVM were biased to the high EL region, and few high TS microstructures were found. This is because BO suppresses repeated searches for similar microstructures by modifying the kernel function of the acquisition function, but OCSVM-based exploration has no algorithm to suppress such repeated searches. The maximum values of EL and TS × EL were both obtained in microstructure H. The EL value was 30.6%, which was lower than in BO, and the TS × EL value was 30.6 GPa·%. Among the three methods, only OCSVM found the microstructure with TS × EL > 30 GPa. The microstructure with the second-highest TS × EL in the OCSVM had a relatively good balance of TS and EL, which is referred to as microstructure I.

To compare the efficiency of the four explorations, including random search, the histories of the calculated TS × EL and its maximum value are plotted in Figure 14. The history of maximum TS × EL suggests that BO and OCSVM explorations have higher search performance than RS, and BLOX has similar performance to RS. The regression methods used in each exploration were GPR for BO, RF for BLOX, and SVM for OCSVM. These differences in regression methods may have caused differences in search performance. Especially, OCSVM exploration found the microstructure with the highest TS × EL in the early iterations. The TS × EL trace (Figure 14(a)) shows that OCSVM was always searching for high TS × EL. Therefore, OCSVM is the most suitable for discovering microstructures with higher tensile properties in a small number of iterations. On the other hand, if a large number of numerical calculations are allowed, BO is possibly more suitable than OCSVM. The BO continued to increase the maximum value of TS × EL by exploring a wide range of microstructures, while the OCSVM mainly explored the high EL region and the maximum value of TS × EL was not updated after 20 iterations. In the OCSVM search, the distribution of CI did not change significantly after 20 iterations, which may be related to the saturation of the search. To solve this problem in OCSVM, it would be adequate to add a penalty term to the CI that reduces the CI value when repeatedly searching for similar microstructures, similar to the exploration–exploitation ratio in BO. However, it would introduce a new problem of tuning the hyperparameter in the penalty term. Another approach to achieve such a rapid search would be to directly use the TS × EL values predicted by the SVM, rather than CI, to determine the next search.

3.4. Microstructures with high TS×EL

To consider the strengthening mechanism of the microstructures with high TS × EL, the calculated results of microstructures D-I found in the three explorations are shown in Figures 15–17, respectively. The microstructural descriptors and tensile properties for each microstructure are shown in Tables 3 and 4, respectively. To examine the effect of volume fraction

![Figure 13](image-url)
Figure 14. Traces of (a) TS×EL and (b) maximum TS×EL in four exploration methods: random search (RS), Bayesian optimization (BO), boundless objective-free exploration (BLOX), one-class support vector machine (OCSVM).

Figure 15. Crystal plasticity finite element models, distributions of damage variable and macroscopic stress–strain curves at search points D and E in Bayesian optimization: (a) point D (maximum elongation), (b) point E (maximum TS×EL).
of martensite (VFₘ), the TS and EL values calculated from the regression equation based on VFₘ (Equations (8)–(9)) are also listed as TS_VF and EL_VF in Table 4. Among them, microstructures D (Figure 15(a)) and H (Figure 17(a)) showed particularly high EL (EL > 30%). These EL values were much higher than EL_VF, indicating that the total elongation was improved not only by the effect of the volume fraction of the martensite phase but also by its spatial distribution. The microstructures D and H consisted of a ferrite matrix with martensite grains dispersed at regular spacing. In these microstructures, a large number of deformation bands oblique to the loading direction occurred almost simultaneously and were distributed throughout the microstructures. This is in contrast to the microstructure found in RS (Figure 8), where one or two main deformation bands develop obliquely to the loading direction, leading to eventual failure. To examine the mechanism of the high elongation in more detail, the evolution of the equivalent plastic strain in the microstructure H was plotted (Figure 18). At applied strain of 10%, a relatively large plastic strain was generated in the ferrite phase sandwiched by martensite grains similar to the other microstructures, but due to the regular distribution of the two phases, small deformation bands appeared synchronously in the microstructure, and the plastic strains showed a zigzag pattern. The plastic strain developed keeping this geometrical pattern, and strain localization was suppressed until rupture. Thus, it was shown that the microstructure with regularly aligned hard grains improves elongation through a synchronized void growth mechanism. To experimentally prove the ductility improvement due to this mechanism, it is necessary to find the thermo-mechanical treatment conditions that uniformly disperse the martensite grains.

Microstructures E (Figure 15(b)) and I (Figure 17(b)) showed a well-balanced improvement in TS and EL. These microstructures had a lamellar structure elongated in the loading direction. In contrast to the lamellar structure found in RS (microstructure C, Figure 8(c)), the lamellar spacing was relatively uniform and fine. For all of the lamellar microstructures C, E and I, the EL values calculated by CPFEM were higher than the EL_VF. The ratio of EL to EL_VF was approximately 1.6 for microstructure C, but exceeded 2 for microstructures E and I. These results suggest that uniform, fine lamellae are more effective in improving the total elongation. In the microstructures E and I, damage occurred in both the ferrite and martensite phases and the fracture eventually occurred in the grain with the most favorable crystallographic orientation for plastic deformation. Since the ferrite-martensite interface is parallel to the loading direction, the lamellar structure has a smaller stress concentration than the other structures, which may have resulted in the unique tensile properties. The equivalent plastic strain distribution of the microstructure I is plotted in Figure 19. The microstructure with large lamella thickness (Figure 8(c)) showed brittle fracture

![Figure 16. Crystal plasticity finite element models, distributions of damage variable and macroscopic stress–strain curves at search points F and G in boundless objective-free exploration (BLOX): (a) point F (maximum elongation), (b) point G (maximum TS×EL).]
Figure 17. Crystal plasticity finite element models, distributions of damage variable and macroscopic stress–strain curves at search points H and I in one-class support vector machine (OCSVM) based exploration: (a) point H (maximum $TS \times EL$), (b) point I (second maximum $TS \times EL$).

Figure 18. Distribution of plastic equivalent strain of microstructure H (TS 1.00 GPa, EL 30.6%) with uniformly dispersed martensite particles at applied strain of (a) 10%, (b) 20%, (c) 30%. Gray lines represent the ferritic-martensitic boundaries.

Figure 19. Distribution of plastic equivalent strain of microstructure I (TS 1.47 GPa, EL 19.4%) with lamellar structure at applied strain of (a) 10%, (b) 15%, (c) 19%. Gray lines represent the ferritic-martensitic boundaries.
due to strain localization in a specific location, whereas the microstructure I with multiple layers of relatively thin lamellae showed high ductility due to the distribution of plastic strain throughout the microstructure. This approach of using a lamellar structure of soft and hard phases has already been realized in multilayered steels [59]. It has been experimentally confirmed that the fracture behavior of the hard layer clearly transitions from brittle cleavage to ductile shear when the layer thickness is reduced [60], which is consistent with the calculation results of this work. The multilayer steel sheets composed of a hard martensitic stainless steel and a soft austenitic stainless steel exhibited a good strength–ductility balance, with a tensile strength of 1.4 GPa and uniform elongation of 20% [61]. Microstructures F and G found in BLOX had a unique network-type distribution of martensite and ferrite phases (Figure 16), but the values of TS and EL were not much different from those of RS (Figure 8(a,b)). The reason for the lack of improvement in tensile properties is because intense stress concentration occurred at the wavy ferrite-martensite interfaces, leading to fracture.

In this work, two-dimensional microstructure models with a simple damage law were used. In a future study, three-dimensional (3D) models with the stress triaxiality dependence of the fracture strain will be used for optimizing the 3D microstructure. Focusing on the other tensile properties, yield stress and uniform elongation (or work hardening rate) play an essential role in sheet metal formability. Since there is also a trade-off relationship between yield stress and uniform elongation [62], the approaches to search for outliers using BLOX and OCSVM proposed in this study can be applied to such problems.

4. Conclusions

In this study, CPFEM with ductile damage model was introduced into the framework of various exploration methods: random search (RS), Bayesian optimization (BO), BoundLess Objective-free eXploration (BLOX). Inspired by the advantages of BO and BLOX, a new exploration method was proposed by defining a curiosity index from the classification scores of OCSVM. The following conclusions can be drawn:

1. The results of RS showed that tensile strength (TS) correlates with the volume fraction of martensitic phase (VF_M) and total elongation (EL) correlates with the inverse of VF_M. The product of strength and total elongation (TS × EL) was relatively insensitive to VF_M and showed large scatter. The maximum TS × EL value of the microstructures analysed in RS was 18.4 GPa.%.
2. The feature importance analysis revealed that the VF_M, the 6th principal component of the spatial correlation function (SC6), and the first principal component of the persistent homology (PH1) are important in predicting the product of strength and total elongation (TS × EL) of dual-phase steels. The SC6 represents the intensity of the lamellar structure, and the PH1 represents the fineness of the two-phase pattern.

3. Using the above three important microstructural descriptors, explorations of microstructures were conducted by BO, BLOX, and OCSVM. The BO and OCSVM found microstructures exhibiting TS × EL much higher than those found in RS. The maximum TS × EL values in BO and OCSVM were approximately 30 GPa.% The search performance of BLOX was comparable to that of RS.

4. The CPFEM results of the microstructures with high TS × EL values found in BO and OCSVM were investigated in detail. These microstructures were homogeneously dispersed fine martensitic particles or a lamellar structure of martensitic and ferritic phases. It was indicated that strain partitioning and reduction of stress concentration led to improved tensile properties in these microstructures. These findings are consistent with known findings in materials engineering. It was shown that the proposed methodology of inverse problem analysis can be used to derive the optimal microstructures for improving the tensile properties of DP steels.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Council for Science, Technology and Innovation (CSTI), Cross ministerial Strategic Innovation Promotion Program (SIP), “Materials Integration” for Revolutionary Design System of Structural Materials (Funding Agency: Japan Science and Technology agency).

ORCID

Takayuki Shiraiwa  http://orcid.org/0000-0002-0999-0798

References

[1] Adachi T, Ito A, Adachi H, et al. Effect of prior structure to intercritical annealing on rapid formation of ultrafine ferrite + austenite structure and mechanical properties in 0.1%C-2%Si-5%mm Steels. ISIJ Int. 2020;60(4):764–773. DOI:10.2355/isijinternational.ISIJINT-2019-401.
[2] Zackay VF, Parker ER, Fahr D, et al. The enhancement of ductility in high-strength steels. ASM Trans Q. 1967;60(2):252–259.

[3] Speer JG, De Moor E, Findley KO, et al. Analysis of microstructure evolution in quenching and partitioning automotive sheet steel. Metall Mater Trans a. 2011;42(12):3591–3601. DOI:10.1007/s11661-011-0869-7.

[4] Tasan CC, Diehl M, Yan D, et al. An overview of dual-phase steels: advances in microstructure-oriented processing and micromechanically guided design. Annu Rev Mater Res. 2015;45(5):391–431.

[5] Park K, Nishiya M, Nakada N, et al. Effect of the martensite distribution on the strain hardening and ductile fracture behaviors in dual-phase steel. Mater Sci Eng A. 2014;604:135–141. DOI:10.1016/j.msea.2014.02.058.

[6] Calcagnotto M, Adachi Y, Ponge D, et al. Deformation and fracture mechanisms in fine- and ultrafine-grained ferrite-martensite dual-phase steels and the effect of aging. Acta Materialia. 2011;59(2):658–670. DOI:10.1016/j.actamat.2010.10.002.

[7] Son YI, Lee YK, Park K-T, et al. Ultrafine grained ferrite–martensite dual phase steels fabricated via equal channel angular pressing: microstructure and tensile properties. Acta Materialia. 2005;53(11):3125–3134.

[8] Soliman M, Pakowski H. Tensile properties and bake hardening response of dual phase steels with varied martensite volume fraction. Mater Sci Eng A. 2020;777:139044. DOI:10.1016/j.msea.2020.139044.

[9] Oka M. Title in Japanese. J High Pressure Inst Jpn. 1978;16(2):102–110.

[10] Kadkhodapour J, Butz A, Ziaei R, et al. Mechanisms of void formation during tensile testing in a commercial, dual-phase steel. Acta Materialia. 2011;59(7):2575–2588.

[11] Alaee A, Rad SZ, Kadkhodapour J, et al. Effect of microstructure pattern on the strain localization in DP600 steels analyzed using combined in-situ experimental test and numerical simulation. Mater Sci Eng A. 2015;638:251–261. DOI:10.1016/j.msea.2015.04.071.

[12] Yan D, Tasan CC, Raabe D. High resolution in situ mapping of microstrain and microstructure evolution reveals damage resistance criteria in dual phase steels. Acta Materialia. 2015;96:399–409. DOI:10.1016/j.actamat.2015.05.038.

[13] Uehata N, Koyama M, Takagi S, et al. Optical microscopy-based damage quantification: an example of cryogenic deformation of a dual-phase steel. ISIJ Int. 2018;58(1):179–185. DOI:10.2355/isijinternational.ISIJINT-2017-468.

[14] Matsuno T, Maeda D, Shutoh H, et al. Effect of martensite volume fraction on void formation leading to ductile fracture in dual phase steels. ISIJ Int. 2014;54(4):938–944. DOI:10.2355/isijinternational.54.938.

[15] Sun X, Choi KS, Liu WN, et al. Predicting failure modes and ductility of dual phase steels using plastic strain localization. Int J Plast. 2009;25(10):1888–1909.

[16] Landron C, Maire E, Bouaziz O, et al. Validation of void growth models using X-ray micromotography characterization of damage in dual phase steels. Acta Materialia. 2011;59(20):7564–7573.

[17] Matsuno T, Yoshioka T, Watanabe I, et al. Three-Dimensional finite element analysis of representative volume elements for characterizing the effects of martensite elongation and banding on tensile strength of ferrite-martensite dual-phase steels. Int J Mech Sci. 2019;163:105133.

[18] Brifod F, Shiraïwa T, Enoki M. Micromechanical investigation of the effect of the crystal orientation on the local deformation path and ductile void nucleation in dual-phase steels. Mater Sci Eng A. 2021;826:141933. DOI:10.1016/j.msea.2021.141933.

[19] Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: theory and practice. Neurocomputing. 2020;415:295–316. DOI:10.1016/j.neucom.2020.07.061.

[20] Shiraïwa T, Enoki M, Goto S, et al. Data assimilation in the welding process for analysis of weld toe geometry and heat source model. ISIJ Int. 2020;60(6):1301–1311. DOI:10.2355/isijinternational.ISIJINT-2019-720.

[21] Kusano M, Kitano H, Watanabe M. Novel calibration strategy for validation of finite element thermal analysis of selective laser melting process using bayesian optimization. Materials. 2021 Sep;14(17):4948.

[22] Terayama K, Sumita M, Tamura R, et al. Pushing property limits in materials discovery via boundless objective-free exploration. Chem Sci. 2020;11(23):5959–5968.

[23] Scholkopf B, Platt JC, Shawe-Taylor J, et al. Estimating the support of a high-dimensional distribution. Neural Comput. 2001;13(7):1443–1471. DOI:10.1162/089976601750264965.

[24] Fujiwara K, Abe E, Suzuki Y, et al., editors. Epileptic seizure monitoring by one-class support vector machine. Signal and Information Processing Association Annual Summit and Conference (APSIPA). 2014 Asia-Pacific, Siem Reap, Cambodia; 2014 Dec 9–12.

[25] Kats V. Structural health monitoring system of construction facilities: enhanced training approach. J Phys. 2020;1694(1):012014.

[26] Aoyagi K, Wang H, Sudo H, et al. Simple method to construct process maps for additive manufacturing using a support vector machine. Additive Manuf. 2019;27:353–362. DOI:10.1016/j.addma.2019.03.013.

[27] Abid NH, Abu Al-Rub RK, Paluzotto AN. Micromechanical finite element analysis of the effects of martensite morphology on the overall mechanical behavior of dual phase steel. Int J Solids Struct. 2017;104:8–24. DOI:10.1016/j.ijsolstr.2016.11.005.

[28] Kalidindi SR, Niezgoda SR, Salem AA. Microstructure informatics using higher-order statistics and efficient data-mining protocols. JOM. 2011;63(4):34–41.

[29] Hiraoka Y, Nakamura T, Hirata A, et al. Hierarchical structures of amorphous solids characterized by persistent homology. PNAS. 2016;113(26):7035–7040. DOI:10.1073/pnas.1520877113.

[30] Kim H, Yamamoto T, Sato Y, et al. Establishment of structure-property linkages using a Bayesian model selection method: application to a dual-phase metallic composite system. Acta Materialia. 2019;176:264–277. DOI:10.1016/j.actamat.2019.07.006.

[31] Avalos E, Xie S, Akagi K, et al. Bridging a mesoscopic inhomogeneity to macroscopic performance of amorphous materials in the framework of the phase field modeling. Phys D. 2020;409:132470. DOI:10.1016/j.physd.2020.132470.

[32] Grober MA, Jackson MA. DREAM.3D: a digital representation environment for the analysis of microstructure in 3D. Integr Mater Manuf Innov. 2014;3(1):56–72.
[33] Khristenko U, Constantinescu A, Le Tallec P, et al. A statistical framework for generating microstructures of two-phase random materials: application to Fatigue analysis. Multiscale Model Simul. 2020;18 (1):21–43. DOI:10.1137/19M1259286.

[34] Moussaoui H, Laurencin J, Gavet Y, et al. Stochastic geometrical modeling of solid oxide cells electrodes validated on 3D reconstructions. Comput Mater Sci. 2018;143:262–276. DOI:10.1016/j.commatsci.2017.11.015.

[35] Krebs B, Hazotte A, Germain L, et al. Quantitative analysis of banded structures in dual-phase steels. Image Anal Stereol. 2011;29(2):85. DOI:10.5566/ias.v29.p85-90.

[36] Matheron G. Éléments Pour une théorie des milieux poreux. Paris: Masson et Cie; 1967.

[37] Torquato S. Random heterogeneous materials. New York: Springer-Verlag; 2002.

[38] Fast T, Kalindiri SR. Formulation and calibration of higher-order elastic localization relationships using the MKS approach. Acta Materialia. 2011 Jun;59 (11):4595–4605.

[39] Fullwood DT, Kalindiri SR, Niezgoda SR, et al. Gradient-Based microstructure reconstructions from distributions using fast fourier transforms. Mater Sci Eng A. 2008;494(1–2):68–72.

[40] Obayashi I, Hiraoka Y, Kimura M. Persistence diagrams with linear machine learning models. J Appl Comput Topol. 2018;1(3):421–449.

[41] Fasy BT, Lecci F, Rinaldo A, et al. Confidence sets for persistence diagrams. Ann. Stat. 2014;42 (6):2301–2339. DOI:10.1214/14-AOS1252.

[42] Kusano G, Hiraoka Y, Fukumizu K, editors. Persistence weighted Gaussian kernel for topological data analysis. International Conference on Machine Learning. New York: PMLR; 2016.

[43] Briffod F, Shiraiva T, Enoki M. Microstructure modeling and crystal plasticity simulations for the evaluation of fatigue crack initiation in α-iron specimen including an elliptic defect. Mater Sci Eng A. 2017;695:165–177. DOI:10.1016/j.msea.2017.04.030.

[44] Addessi D, Marfia S, Sacco E. A plastic nonlocal damage model. Comput Meth Appl Mech Eng. 2002;191(13–14):1291–1310.

[45] Boeff M, Gutknecht F, Engels PS, et al. Formulation of nonlocal damage models based on spectral methods for application to complex microstructures. Eng Fract Mech. 2015;147:373–387.

[46] Bao YB, Wierzbicki T. On fracture locus in the equivalent strain and stress triaxiality space. Int J Mech Sci. 2004 Jan;46(1):81–98.

[47] Coppola T, Cortese L, Folgarait P. The effect of stress invariants on ductile fracture limit in steels. Eng Fract Mech. 2009 Jun;76(9):1288–1302.

[48] Moriconi R, Deisenroth MP, Kumar KSS. High-Dimensional Bayesian optimization using low-dimensional feature spaces. Mach Learn. 2020 Sep;109(9–10):1925–1943.

[49] Loh WY. Regression trees with unbiased variable selection and interaction detection. Stat Sinica. 2002 Apr;12(2):361–386.

[50] Genuer R, Poggi J-M, Tuleau-Malot C. Variable selection using random forests. Pattern Recognit Lett. 2010;31(14):2225–2236.

[51] Snoek J, Larochelle H, Adams R, editors. Practical bayesian optimization of machine learning algorithms. Advances in neural information processing systems. Red Hook, New York: Curran & Associates Inc.; 2012.

[52] Bull AD. Convergence rates of efficient global optimization algorithms. J Mach Learn Res. 2011 Oct;12 (10):2879–2904.

[53] Stein C. Miscellaneous Frontmatter. Approximate computation of expectations. Hayward, California: Institute of Mathematical Statistics; 1986; p. i–iii.

[54] Yan C, Bai X, Ren P, et al. Band weighting via maximizing interclass distance for hyperspectral image classification. IEEE Geosci Remote Sens Lett. 2016;13 (7):922–925. DOI:10.1109/LGRS.2016.2553699.

[55] Lai QQ, Brassart L, Bouaziz O, et al. Influence of martensite volume fraction and hardness on the plastic behavior of dual-phase steels: experiments and micro-mechanical modeling. Int J Plast. 2016;80:187–203.

[56] Alaie A, Kadkhodapour J, Rad SZ, et al. Formation and coalescence of strain localized regions in ferrite phase of DP600 steels under uniaxial tensile deformation. Mater Sci Eng A. 2015;623:133–144. DOI:10.1016/j.msea.2014.11.042.

[57] Toda H, Takijiri A, Azuma M, et al. Damage micromechanisms in dual-phase steel investigated with combined phase- and absorption-contrast tomography. Acta Materialia. 2017;126:401–412. DOI:10.1016/j.actamat.2017.01.010.

[58] Kunselman C, Attari V, McClenny L, et al. Semi-Supervised learning approaches to class assignment in ambiguous microstructures. Acta Materialia. 2020;188:49–62. DOI:10.1016/j.actamat.2020.01.046.

[59] Koseki T, Inoue J, Nambu S. Development of multilayer steels for improved combinations of high strength and high ductility. Mater Trans. 2014;55(2):227–237.

[60] Inoue J, Nambu S, Ishimoto Y, et al. Fracture elongation of brittle/ductile multilayered steel composites with a strong interface. Scr Mater. 2008;59(10):1055–1058. DOI:10.1016/j.scriptamat.2008.07.020.

[61] Seok M-Y, Lee J-A, Lee D-H, et al. Decoupling the contributions of constituent layers to the strength and ductility of a multi-layered steel. Acta Materialia. 2016;121:164–172. DOI:10.1016/j.actamat.2016.09.007.

[62] Wu X, Yuan F, Yang M, et al. Nanodomained Nickel unite nanocrystal strength with coarse-grain ductility. Sci Rep. 2015;5(1):11728.