Article

Standard Deviation Effect of Average Structure Descriptor on Grain Boundary Energy Prediction

Ruoqi Dang and Wenshan Yu *

State Key Laboratory for Strength and Vibration of Mechanical Structures, Shaanxi Engineering Laboratory for Vibration Control of Aerospace Structures, School of Aerospace Engineering, Xi'an Jiaotong University, Xi'an 710049, China

* Correspondence: wenshan@mail.xjtu.edu.cn

Abstract: The structural complexities of grain boundaries (GBs) result in their complicated property contributions to polycrystalline metals and alloys. In this study, we propose a GB structure descriptor by linearly combining the average two-point correlation function (PCF) and standard deviation of PCF via a weight parameter, to reveal the standard deviation effect of PCF on energy predictions of Cu, Al and Ni asymmetric tilt GBs (i.e., Σ3, Σ5, Σ9, Σ11, Σ13 and Σ17), using two machine learning (ML) methods; i.e., principal component analysis (PCA)-based linear regression and recurrent neural networks (RNN). It is found that the proposed structure descriptor is capable of improving GB energy prediction for both ML methods. This suggests the discriminatory power of average PCF for different GBs is lifted since the proposed descriptor contains the data dispersion information. Meanwhile, we also show that GB atom selection methods by which PCF is evaluated also affect predictions.

Keywords: grain boundary; descriptor; pair distribution function; grain boundary energy; machine learning method

1. Introduction

Grain boundaries (GBs) are one of the most commonly seen planar defects in polycrystalline metals and alloys. Due to local atomic distortions and inconsistent atomic arrangement, GBs play important roles in determining the mechanical, thermal and electric, etc., properties of materials [1,2]. For example, GBs may act as the dislocation and point defect sources or sinkers, and they may block the dislocation motion and absorb them; thus the strength and ductility of materials can be greatly changed [3]. For an idealized GB, from the geometrical point of view, it can be completely governed by five parameters, usually represented by misorientation and a normal GB plane [4]. Unfortunately, the structures, as well as the properties of the GB, are hard to completely determine. This is simply because a GB may have numerous states due to the point defect absorptions and emissions. It means the structures of a given GB may no longer be unique for a given energy [5–11]. Thus, the connection between structure and property of a GB, such as energy, volume and mechanical behavior, etc., are usually built via atomistic simulations by using molecular dynamics (MD) and density functional theory (DFT) methods [12,13], which is also of great significance for the macroscopic modeling of material behavior [14,15]. Technically speaking, it is possible to do so using MD and DFT, but also needs a heavy workload if such connections for a large number of GBs are expected.

The ML method has been applied in many research fields [16–20], and provides an efficient technique by which to link the structure-property of a GB, particularly to extract correlations from high-dimensional datasets [21–25], and has been successfully applied in predicting GB energies [21,23–27], point defect segregation energies [28,29], GB structures [30] and damages and deformations in GB [31,32]. Usually, an appropriate ML method is employed according to the datasets and the expected correlations. Regardless of these, a problem is how to mathematically describe the GB structure, which should contain

the essential characteristics of GB structures. One of the representative examples is structure units (SUs) [33–36], which are usually used to describe GB structures, but are incapable of doing so for general GBs or GBs driven out of equilibrate states [11]. Furthermore, some studies have also focused on developing structural matrices [12,13,37]. We could indeed gain a unique insight from these studies. However, these descriptions cannot be readily applied in ML.

So far, some descriptors related to GB structures, which can be well used in ML so that the atomic structure-property relationships can be constructed, have been developed. Generally, these descriptors can be divided into two categories; i.e., local atom environment descriptor and descriptors for atom connectivity [21,22,38,39]. The former mostly considers the local atom environment, such as the atomic neighboring arrangement. One of the representative examples, developed by Banadaki and Patala [38,39], describes the local GB structure by using some polyhedral units in a way similar to SUs. In principle, this method can be used to represent any general GB. However, GB structures are usually not at a equilibrate state or subjected to deformation perturbations due to vacancies or self-interstitial diffusions, absorptions and emissions. It means this technique itself suffers from limitations. Banadaki et al. proposed the point-pattern matching algorithm to enhance the power of PU for describing GBs, particularly with local distortions [38]. Another example is the smooth overlap of atomic positions (SOAP) descriptor, which is essentially a combination of radial and spherical spectral bases, including spherical harmonics [22]. For the atom connectivity descriptors, they essentially correlate the positions of atoms in GB and in the vicinity of GB. Such descriptors usually vary smoothly upon the perturbations of atom positions [22]. An example is the pair correlation function (PCF) method due to Gomberg et al.'s study [21]. It was shown that the two-point correlation functions can reliably predict GB energies. Based on these descriptors, a variety of GB properties, such as energy, mobility and mechanical behavior, etc., can be successfully predicted by using ML methods.

Using ML methods to predict GB properties, a better quantitative prediction usually requires that a descriptor should carry the information of the GB structure as much as possible [38]. An extreme case is to consider all atoms composed of the GB structure by evaluating the neighboring atom distribution for each GB atom [22]. This not only drastically increases the dimension of data, but may also lead to data dimension inconsistency between GBs. To avoid such a dilemma, a straightforward method is to further compute the average quantity [21,38]. This way, the descriptor can be seen as an average structure representation (ASR) [22]. From the statistics, ASR does not contain the information of data scatter, such as the standard deviation. Moreover, it is still unclear how standard deviation affects the prediction.

In this study, we establish 464 asymmetric GB models for Cu, Al and Ni metals and relax all GB models using MD. We discuss the prediction of GB energies by comparing two ML methods, i.e., principal component analysis (PCA) -based linear regression [40] and recurrent neural networks (RNN) [41,42], in an effort to answer the following questions:

Based on two-point correlation functions, i.e., the PCF method proposed by Gomberg et al. [21], we introduce a new GB structure descriptor by linearly combining PCF and its standard distribution PCF_{std} via a parameter. How does such a descriptor affect the prediction?

GB atoms can be selected from GB using common neighbor analysis (CNA) [43] and centro-symmetric parameter (CSP) methods [44]. It can be imaged that a different number of GB atoms could be selected for two methods when setting different critical values. Will this affect the prediction?

Comparing two ML methods, what happens to the prediction when considering the GB atom selection method and the GB structure descriptor? Can we predict GB energy without clearly distinguishing the tilt axis of the GB?

The content of this paper is organized as follows. Firstly, we introduce the GB models and the establishment of the GB structure descriptor. Secondly, the PCA-based linear

regression of energies of Cu, Al and Ni GBs according to the tilt axis of GB considering full data and partition data is discussed. Thirdly, the effect of standard deviation of PCF on the RNN-based prediction of GB energies is discussed. Finally, the prediction comparisons between the two ML methods and conclusions of this study are made.

2. Methodology and GB Structure Descriptor

We consider a total number of 464 asymmetric tilt GBs (ATGBs) with misorietations $\Sigma3$, $\Sigma5$, $\Sigma9$, $\Sigma11$, $\Sigma13$ and $\Sigma17$ as the dataset for the subsequent GB energy prediction study. Each GB model is a bi-crystal composed of two grains with specified orientations. To construct it with periodic boundary conditions (PBCs) applicable, crystalline orientations of two grains are needed. Usually, a given GB misorientation Σ can be defined by the overlapped lattices of two crystals with one of them rotated around a specified axis ρ with a certain angle θ. Namely, Σ is equivalent to (ρ, θ), which can also be represented by a rotation matrix $\mathbf{R}_\Sigma$. Take $\Sigma = 3$ as an example, (ρ, θ) related to $\Sigma3$ equals ([110], 70.53°) [33], and the corresponding unit cell of $\Sigma3$ coincidence site lattice (CSL) is spanned by [1$\overline{1}$0], [111] and [11$\overline{2}$]. Then, the normal GB $\mathbf{m}$ of a series of $\Sigma3$ ATGBs expressed in one grain are linear combinations of [111] and [11$\overline{2}$], i.e., i [111] + j [11$\overline{2}$] for different integers i and j. GBs normally expressed in another grain can be obtained by $\mathbf{R}_{\Sigma3}.\mathbf{m}$. The angle between $\mathbf{m}$ and [111] defines the inclination angle, denoted as ϕ. Using $\mathbf{m}$ and $\mathbf{R}_{\Sigma3}.\mathbf{m}$, along with the tilt axis [1$\overline{1}$0], the crystal orientations of two grains of all $\Sigma3$ ATGBs can be defined. By following this approach, asymmetric GBs of all other misorientations can be readily created.

For $\Sigma3$, $\Sigma5$, $\Sigma9$, $\Sigma11$, $\Sigma13$ and $\Sigma17$, lattice symmetry requires the inclination angle ϕ varying from 0° to 90° for $\Sigma3$, $\Sigma9$ and $\Sigma11$, with ϕ varying from 0° to 45° for $\Sigma5$, $\Sigma13$ and $\Sigma17$, respectively. PBCs are imposed within the GB plane for all bicrystal models. Two grains (i.e., Grains A and B) terminate with free surfaces in the direction perpendicular to the GB plane by setting two 10Å thick vacuum spaces on the top and bottom ends of the bicrystal model, as schematically shown in Figure S1 in Supplementary Materials. This allows us to release the stress possibly produced in the z direction during the GB structure optimization. Embedded atom method (EAM) potentials [45,46] are used to model the atomic interactions in Al [47], Cu [48] and Ni [47]. We relax all 464 ATGBs via the conjugate gradient (CG) method using LAMMPS [49] and compute the average energies of all GBs. Tables S1 and S2 in Supplementary Materials list the variations in atom numbers and energies for $\Sigma3$, $\Sigma5$, $\Sigma9$, $\Sigma11$, $\Sigma13$ and $\Sigma17$ GB models of each metal. Atomic structures are visualized using Ovito [50].

With all GBs relaxed, we are in a position to introduce the descriptor by which the GB structure and structure differences between GBs can be described and distinguished. Herein, we employ the pair correlation function (PCF) method proposed Gomberg et al. [21] as a GB structure descriptor. In doing so, a primary concern is how many atoms in GB should be considered when evaluating the average PCF ($PCF_{mean}(r)$). In other words, an appropriate method for selecting GB atoms should include a certain number of atoms in the vicinity of the GB carrying structure information, but exclude other atoms. Herein, we consider common neighbor analysis (CNA) [43] and centro-symmetric parameter (CSP) methods [44]. For the CSP method, two CSP critical values are considered (i.e., CSP > 0.1 and 0.5). As exemplified in Figure S2 in Supplementary Material, the three methods identify a different number of GB atoms. Such effects on GB energy predictions will be discussed in the following section.

According to the approach of Gomberg et al. [21], the PCF of a given GB is computed by averaging the radial distribution function $g_a(r)$ of all N_{GB} GB atoms selected out of the GB using CNA, $CSP_{0.1}$ and $CSP_{0.5}$, which can be expressed as

$$\text{PCF}_{\text{mean}}(r) = \frac{1}{N_{GB}} \sum\nolimits_{\alpha \in GB} g_a(r) \tag{1}$$

where the radial distribution function $g_a(r)$ of a GB atom can be calculated by considering all of its N_{in} neighboring atoms within a specified cut-off radius for three metals.

$$g_\alpha(r) = \sum_{k=1}^{N_{in}} \frac{K^e \left(r - \|R_\alpha^k\|, h^e \right)}{4\pi r^2 n_0} \tag{2}$$

where a kernel function K^e with bandwidth h^e is used to smoothen the radial distribution function. n_0 is the atom density in the FCC lattice and $\|R_\alpha^k\|$ is the distance between atom a and its kth neighboring atom. The parameters needed in Equation (2) are listed in Table S3 in Supplementary Material.

Figure 1a compares PCF_{mean} of the Al single crystal with results taken from [21]. Good agreement validates our algorithm for computing PCF_{mean}. In fact, PCF_{mean} is an averaged radial distribution function (RDF) curve of each GB atom, by which the PCF data fluctuations of different GB atoms cannot be well considered, as evidenced by the variation in standard deviation of PCF ($\text{PCF}_{\text{std}}(r)$) for three GBs in Cu in Figure 1b. In order to incorporate the data fluctuation into the averaged PCF, we further propose a PCF_{comb} by combining $\text{PCF}_{\text{mean}}(r)$ and $\text{PCF}_{\text{std}}(r)$ as

$$\text{PCF}_{comb}(r) = (1 - \zeta)\frac{\text{PCF}_{meam}(r)}{\max(\text{PCF}_{meam}(r))} + \zeta \frac{\text{PCF}_{std}(r)}{\max(\text{PCF}_{std}(r))} \tag{3}$$

where parameter ζ is introduced to weigh the portions of $\text{PCF}_{\text{mean}}(r)$ and $\text{PCF}_{\text{std}}(r)$ in PCF_{comb}. PCF_{comb} is reduced to PCF by letting $\zeta = 0$. As an example, Figure 1c,d shows the $\text{PCF}_{\text{comb}}(r)$ of $\sum5(310)$ and $\sum9(114)$ Cu GBs for three values of ζ. Clearly, the variation trends of $\text{PCF}_{\text{comb}}(r)$ changes as ζ varies. In the following, how the variation of ζ influences the prediction will be discussed. The PCF_{comb} curve of each GB is further represented as 512 discrete points, serving as the input data for the ML methods.

In the following, we adopt two ML methods to predict GB energies, i.e., principal component analysis (PCA)-based linear regression [40] and recurrent neural networks (RNN) [41,42]. PCA is usually implemented in two steps, a dimensionality-reduction of data and regression based on the principle component, which are essentially the eigenvalues of the covariance matrix of raw data. Therefore, the regression of PCA is achieved only using a few principle components. The principle components for regression are selected by considering the explained variance percentage of each principle component. However, RNNs do not require dimensionality reduction. The training and prediction are performed by using raw data. To quantitatively compare the predictions, mean absolute error (MAE) and mean relative error (MRE) are assessed via

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left| \gamma_i^{Pred} - \gamma_i^{MD} \right| \tag{4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left| \gamma_i^{Pred} - \gamma_i^{MD} \right| / \gamma_i^{MD} \tag{5}$$

where γ_i^{Pred} and γ_i^{MD} are GB energies predicted by ML methods and computed via MD.
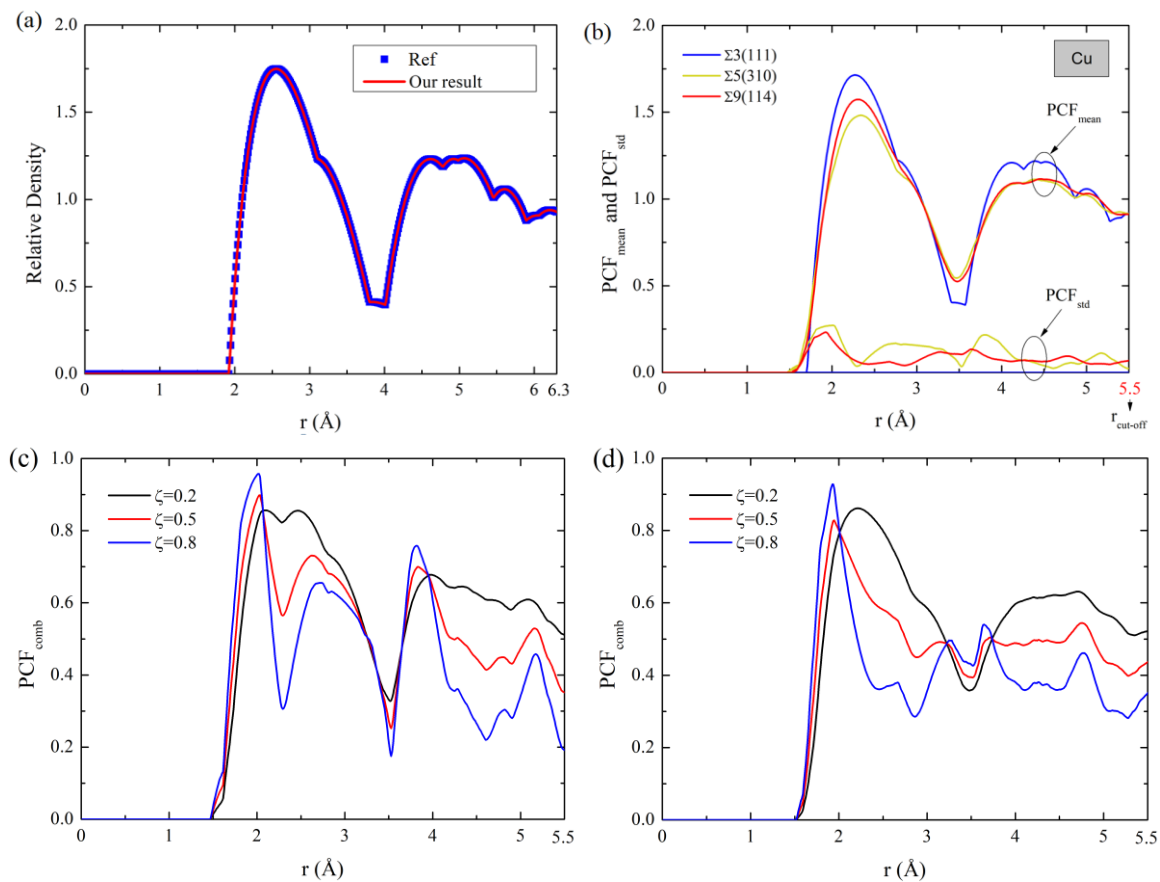
Figure 1. (**a**) PCF of Al single crystal compared with the results taken from [16], (**b**) PCF$_{mean}$ and PCF$_{std}$(r) of Cu GBs $\sum$5(310), $\sum$9(114) and $\sum$3(111). PCF$_{comb}$(r) of Cu GBs (**c**) $\sum$5(310) and (**d**) $\sum$9(114) for ζ = 0.2, 0.5 and 0.8. Results of figure (**b**) for Al and Ni are shown in Figure S3 in Supplementary Material.

3. PCA-Based Prediction

To implement PCA, we need to determine which principle components will be used in the regression. To do so, we analyze the explained variance percentage of the first ten principle components for Cu, as shown in Figure S4 in Supplementary Materials. It turns out that the explained variation of the first PC is up to 93%, while those of the other nine PCs are lower than 3%. This suggests that only the first few PCs accounting for higher explained variance percentages retain most of the original data, while the rest only keep a small amount of the data. It is therefore unnecessary to consider many PCs in the subsequent GB energy regression. Because of this, only the first three PCs, e.g., PC$_1$, PC$_2$ and PC$_3$, are used in the regression. With the multiple linear regression method, GB energy can be written as

$$\gamma_{GB} = a.PC_1 + b.PC_2 + c.PC_3 + d \tag{6}$$

where a, b, c and d are fitting parameters. PC$_1$, PC$_2$ and PC$_3$ are obtained from data training, which is dependent on the dataset. The dataset in this study consists of GBs with <100> and <110> tilts axes. Thus, there are two possible ways to obtain PCs by reducing the dimensionality of data when considering all data together and two data subsets corresponding to <100> and <110> tilt axes, denoted as *full data* and *partitioned data* methods, respectively. Thus, two sorts of PCs PC$_i$ through training different datasets can be obtained.

The PCs obtained from data training are actually a representation of data in a lower dimensional space. Although these PCs retain most of the original data, it is still challenging to impart each PC with possible physical interpretability. By following the approach by

Gomberg et al. [14], it is possible to correlate each PC to a geometrical parameter of GB by interpolating each PC as a function of the geometrical parameter. Herein, such a geometrical parameter is considered to be inclination angle ϕ related to asymmetric GBs for a specified Σ. In this study, PC_1, PC_2 and PC_3 are assumed to be cubic polynomial interpolation functions of ϕ.

$$PC_i(\varphi) = A_i\varphi^3 + B_i\varphi^2 + C_i\varphi + D_i \tag{7}$$

where A_i, B_i, C_i and D_i are fitting parameters. Such cubic polynomial interpolation can well characterize the variation of calculated PC_i vs ϕ for most GBs and PCs. From the above analysis, there are four ways of predicting GB energies in terms of different approaches of obtaining PCs, denoted as *full data*, *full data-fitting*, *partitioned data* and *partitioned data-fitting* methods, respectively. For the partitioned data, Equation (4) corresponds to three metals, shown in Equations (S2)–(S8) in Supplementary Material. The parameters in Equation (5) are listed in Tables S4–S6 in Supplementary Material.

For comparison, Figure 2 exemplifies energy predictions of $\Sigma3$ and $\Sigma5$ asymmetric copper GBs related to <110> and <100> tilt axes. In order to assess the prediction improvement due to the data partition, MAEs of all predictions are calculated, as shown in Figure 2. For $\Sigma3$ GBs, MAEs for PCi- and PCi(ϕ)-based linear fittings using full data are 64.57 mJ/m^2 and 113.77 mJ/m^2, but those for partitioned data are 45.49 mJ/m^2 and 87.42 mJ/m^2. For $\Sigma5$ GBs, MAEs for PCi- and PCi(ϕ)-based linear fittings using full data are 25.58 mJ/m^2 and 29.62 mJ/m^2, but those for partitioned data are 4.29 mJ/m^2 and 10.66 mJ/m^2. From further inspection of the variation of MAEs due to the data partition, MAEs for PCi- and PCi(ϕ)-based linear fittings are reduced by ~30% and ~23%, respectively, but, for $\Sigma3$ GBs, they are up to ~83% and ~64%. This suggests that a better prediction can be achieved by separately considering <110> and <100> GB datasets, which is particularly more prominent for <100> GBs. From the MAEs results of PCi and PCi(ϕ) linear fittings for $\Sigma3$ and $\Sigma5$ GBs, PCi(ϕ) linear fittings indeed lead to a larger MAE than PCi linear fittings, which is understandable since PCi(ϕ) is approximately obtained from cubic interpolation. Nevertheless, these results show that PCi is capable of being correlated with inclination angle ϕ.

As previously mentioned, γ_i^{Pred} should be dependent on ζ, therefore, both MAE and MRE are functions of ζ. As an example, Table 1 shows MAEs and MREs of all Σs for $\zeta = 0.5$. Comparing the MAE and MRE predictions for <100> and <110> GBs, both MAE and MRE are lower for <100> GBs, which further demonstrates that better predictions can be obtained for <100> GBs. In fact, this can be explained by considering the structure differences of <110> and <100> GBs. It is known that SUs for <100> GBs are composed of some [100] dislocations [2,51,52]. This brings simpler and mutually similar structures to <100> GBs. However, <110> GBs are composed of SUs much more complicated than those of <100> GBs [2,33,53–55]. Therefore, the structures of two <100> GBs may be quite different from each other. Thus, predictions for <100> GBs are better than those for their <110> counterparts, as also evidenced by the results of Al and Ni (see Figures S11, S12, S16 and S17 in Supplementary Material).

Table 1. MAE and MRE of Cu GB energy predictions. Note that this table lists the prediction for PCF$_{comb}$ computed for CNA-based GB atom selection and ζ is taken as 0.5.

Error	$\sum3$	$\sum9$	$\sum11$	$\sum5$	$\sum13$	$\sum17$
MAE (mJ/m^2)	99.72	50.86	70.48	31.70	34.22	64.19
MRE (%)	42.54	6.40	11.89	3.14	3.48	6.30
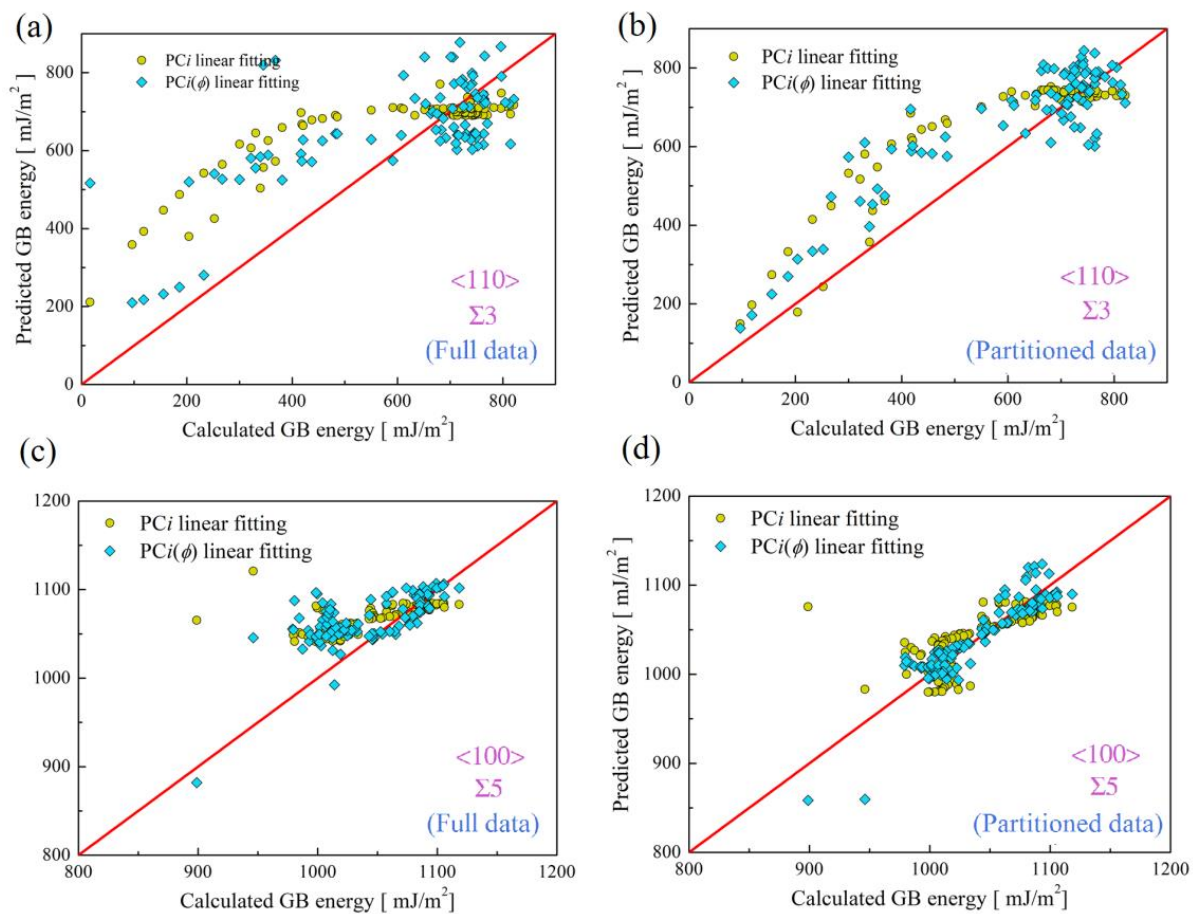
Figure 2. Energy predictions for (**a**,**b**) Σ3 and (**c**,**d**) Σ5 asymmetric copper GBs. Dimensionality reductions are performed based on full data (Figures (**a**,**c**)) and partitioned data (Figures (**b**,**d**)). Note that this figure exemplifies the predictions using PCF_{comb} computed for CNA-based GB atom selection and $\zeta = 0.5$. Results of Σ9, Σ11, Σ13 and Σ17 GBs based on full data and partitioned data are shown in Figures S5–S8 in Supplementary Material.

In order to compare the effects of the GB atom selection method (i.e., CNA, CSP = 0.1 and CSP = 0.5) on the prediction, Figure 3a,b exemplify the MRE of <110> and <100> Cu GBs vs. ζ. From Figure 3, with increasing ζ, the MRE of <110> GBs keep increasing; however, that of <110> GBs keep decreasing. Finally, MREs of both <110> and <100> GBs reach plateaus. Further inspection of Figure 3 reveals that the minimum values of MRE for <110> and <100> GBs for CAN and CSP0.1 methods correspond to $\zeta = 0.0$ and 1.0, respectively. For the CSP0.5 method, the minimum values of MRE for <110> and <100> GBs are $\zeta = 0.2$ and 0.1. Moreover, considering CAN, CSP0.1 and CSP0.5 alone, MREs also differ at $\zeta = 0.0$, but their general variation trends are similar. Therefore, it can be seen that a better prediction not only requires an appropriate GB atom selection method, but an appropriate value of ζ. In fact, the MREs of Al and Ni are also dependent on ζ, as seen from Figures S15 and S20 in Supplementary Material.
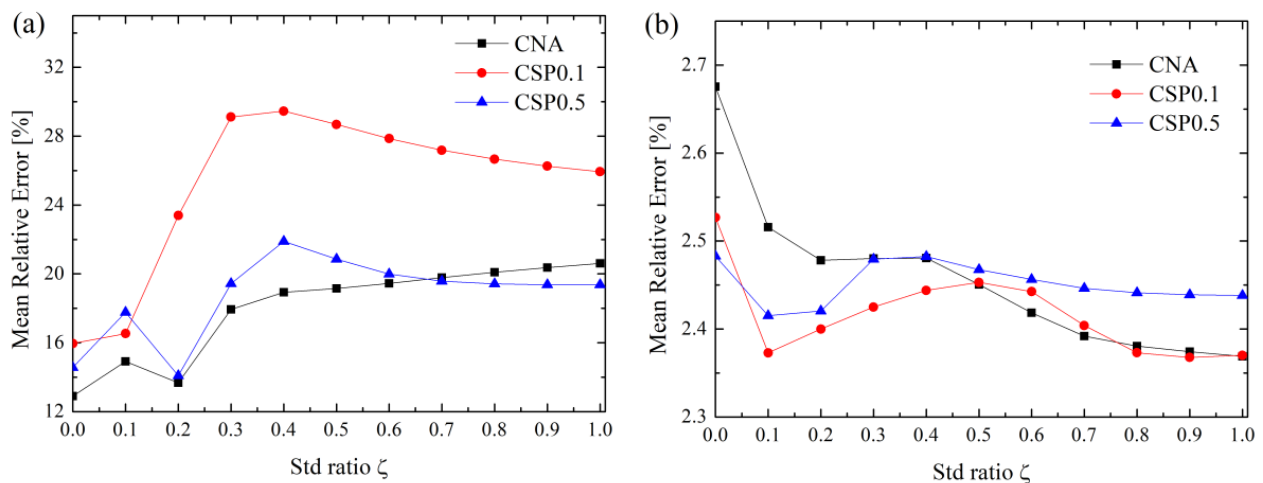
Figure 3. MRE of Cu GBs for three methods of selecting GB atoms (**a**) <110> and (**b**) <100> tilt axis. Results of Al and Ni are shown in Figures S15 and S20 in Supplementary Material.

4. RNN-Based Predictions

In this section, we discuss the predictions using the RNN method. Due to the dimension reduction in the PCA method, some data is lost. Moreover, a better prediction can be achieved provided that GB types are distinguished based on the tilt axis; i.e., <100> and <110>. In comparison to PCA, on the other hand, the RNN method is highly nonlinear. Considering all of these factors, we do not distinguish GB types for each metal; i.e., the prediction is performed using the full data for each metal. For each metal, 10-fold cross validation is performed, with each fit being performed on a training set consisting of 70% of the total training set selected at random, with the remaining 30% used as a holdout set for testing. This yields good convergence of MAE, as shown in Figure S21 in Supplementary Materials. Figure 4 shows the prediction results of the RNN method considering three GB atom selection methods. A preliminary comparison between PCA and RNN, as shown in Figures 2 and 4, reveals that the RNN method gives a better prediction. This is not surprising due to the higher nonlinearity of the RNN method.

We further evaluated the MAE of RNN predictions for three GB atom selection methods and three metals, as shown in Figure 5. Clearly, with increasing ζ, there is a sudden drop in the MAE. Meanwhile, such a drop in CNA, CSP0.1 and CSP0.5 for the same metal almost occurs at the same value of ζ. Moreover, MAEs for Cu, Al and Ni suddenly drop by ~75%, ~75% and ~70% at $\zeta_{crit} \approx 0.3$, 0.6 and 0.7. After the sudden drops, all curves approach plateaus with nearly the same MAE, regardless of the three GB atom selection methods. This evidences the significant dependence of ζ in RNN prediction, and also implies that considerable errors will be caused in RNN prediction when letting $\zeta = 0$. To avoid such errors, the standard deviation of PCFs (PCF_{std}) must be incorporated into the descriptor. Moreover, the ζ_{crit} of the three metals differ, which implies that the prediction accuracy can be enhanced only when more data scatter information of PCF is considered. ζ_{crit} of Cu is the smallest, while that of Ni is the largest. It suggests that data scatter of the average PCF for Cu is the lowest, but that of Ni is the highest.
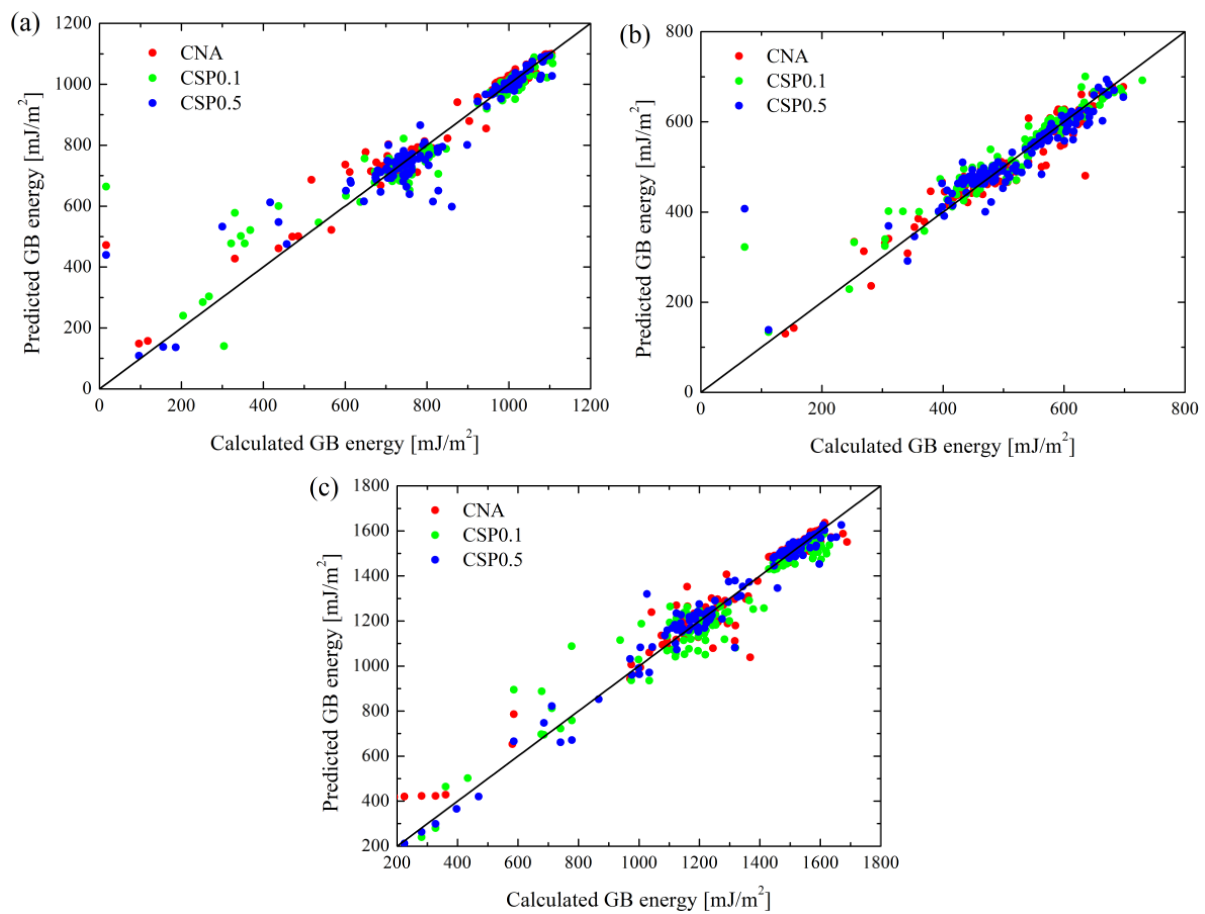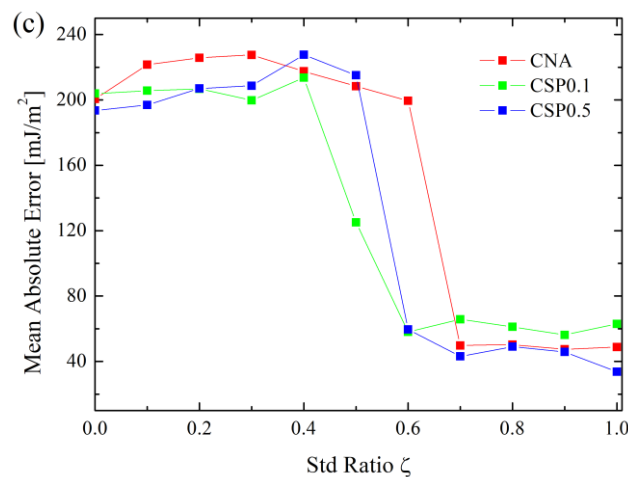
Figure 4. A comparison of RNN-based predictions and MD results for (**a**) Cu, (**b**) Al and (**c**) Ni. Note that predictions are made by considering the parameter ζ corresponding to the minimum MAE of three GB atom selection methods, as shown in Figure 5.
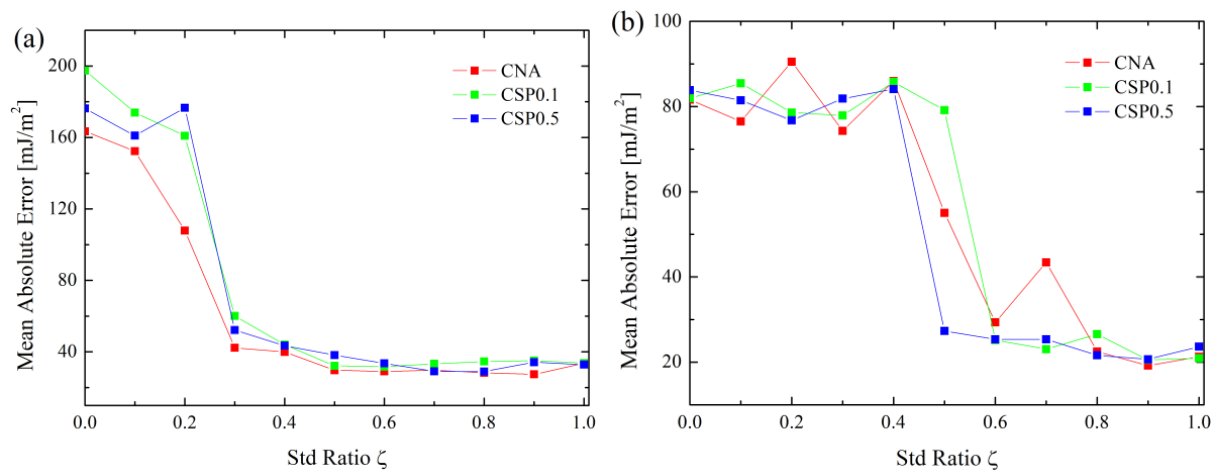

Figure 5. *Cont.*

Figure 5. MAE of RNN prediction vs. ζ of three GB atom selection methods for (**a**) Cu, (**b**) Al and (**c**) Ni.

5. Discussions

In this study, we predicted GB energies by using PCA-based linear regression and RNN. Figure 6 compares the GB energy prediction properties of two ML methods. From Figure 6a, RNN gives a better prediction than the PCA method. For PCA-based linear regression, linear fitting parameters of GB energy are obtained by dividing a full dataset into two separate datasets according to GBs of <100> and <110> tilt axis for three metals. In doing so, we attempted to weaken the effects of the mutual interference due to <100> and <110> GBs when obtaining linear fitting parameters. Indeed, such a way of treating a dataset for PCA-based predictions decreases the prediction errors (Figure 2). We also tried to further obtain the linear fitting parameters by expressing them as cubic polynomial interpolation functions of an inclination angle of GB ϕ, instead of obtaining them from data training. The purpose of doing so was to obtain an empirical GB energy prediction function and intensify the interpretability of ML prediction, which may be impossible for the RNN method. Such a method may work for PCA-based predictions, as evidenced in Figure 2.
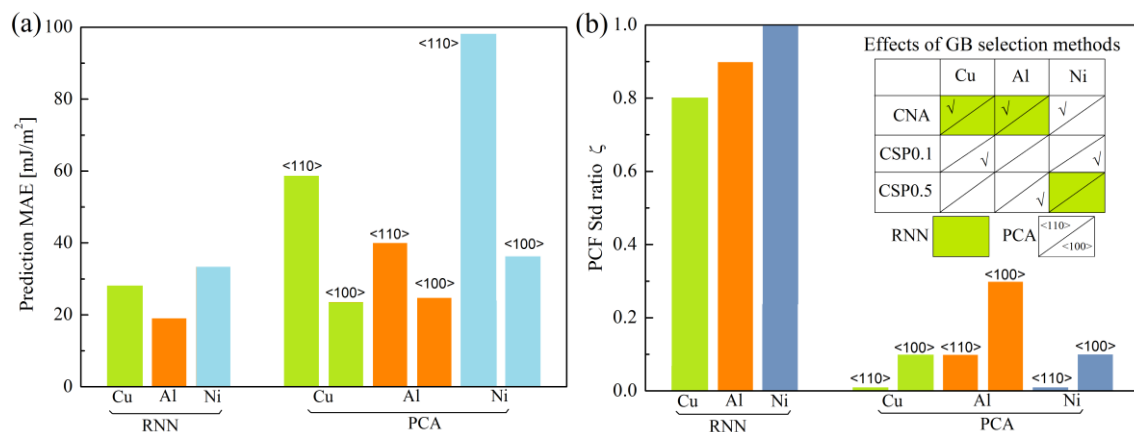
Figure 6. (**a**) Lowest MAE and (**b**) the variation of PCF standard deviation ratio ζ at the lowest MAE for PCA and RNN predictions. Inset in figure (**b**) shows which GB selection methods yield the lowest MAEs for PCA and RNN predictions.

An appropriate GB structure descriptor is vital in ML-based GB energy prediction, as whether the descriptor contains the essential information of GB structures or not determines the prediction accuracy. In fact, PCF as a GB structure descriptor [16], compared with those of polyhedral units [34,35], is much easier use. However, the definition of PCF shows that the GB structure is described by an average function. It is believed that the GB

structure may not be well described without considering the higher order moment of RDF data from the statistical point of view. Motivated by this, we further incorporated the standard deviation of PCF into the PCF function to further extend the descriptor of the GB structure (Equation (2)). Figure 6b shows that predictions using PCA and RNN are significantly dependent on the parameter and GB atom selection methods for three metals, also suggesting the necessity of considering PCF_{std} in GB structure descriptors.

6. Conclusions

In this paper, we studied the GB energies prediction properties of Cu, Al and Ni using two ML methods; i.e., PCA-based linear regression and RNN. We considered the asymmetric GBs $\Sigma3$, $\Sigma5$, $\Sigma9$, $\Sigma11$, $\Sigma13$ and $\Sigma17$ of <110> and <100> types. Atomistic models were constructed and relaxed using the MD method. By extending the PCF-based GB structure descriptor and using three methods of selecting GB atoms, we compared the prediction of two ML methods. The main conclusions of this study were drawn as:

For the three metals, the lowest MAE can be obtained when ζ is greater than 0.8 for RNN, while that should be smaller than 0.3 for PCA-based linear regression. This indicates the dependence of GB descriptors on the ML method. Meanwhile, PCF as an average function and GB structure descriptor needs to consider the PCF_{std}, by which ML prediction accuracy can be improved. The GB structure descriptor in the form of average structure representation (ASR) may need to further take into account the standard deviation of ASR.

In comparison to RNN, it is indeed possible to intensify or realize the interpretability of ML prediction by using the PCF-based linear regression method, though how to generalize the fitting method of the linear regression function when considering a dataset of different GB misorientations still needs to be addressed.

For a specific ML method, the MAE of the prediction is determined by multiple factors, such as the GB atom selection method and a portion of PCF standard deviation. A better quantitative descriptor of GB structure is a trade-off between computation cost and complexity. It is expected that prediction accuracy can be enhanced by combining those comprehensive descriptors together. Moreover, it will be interesting to examine the performance of GB structure descriptors if we consider GBs of mixed types.

GBs for three methods of selecting GB atoms. Figure S16. PCA based prediction results of Ni <110> GBs using partitioned <110> data. Regression is performed using PCi based on dimensionality reduction and further interpolation as a function of φ, respectively. Figure S17. PCA based prediction results of Ni <100> GBs using portioned <100> data. Regression is performed using PCi based on dimensionality reduction and further interpolation as a function of φ, respectively. Figure S18. Comparison of Computed PCi with cubic interpolated PCi as a function of inclination angle for <100> type Ni GBs. Figure S19. Comparison of computed PCi with cubic interpolated PCi as a function of inclination angle for <100> type Ni GBs. Figure S20. MRE of Cu GBs for three methods of selecting GB atoms. Figure S21. MAE convergence curve for RNN prediction of Cu GBs. Table S1. Some additional information of all GB models. Table S2. Variation range of GB energies for all GBs of each metal (mJ/m^2). Table S3. Some material constants and parameter used for computing PCF. Table S4. Cubic interpolation of PCi as a function of inclination angle for all Cu GBs. Table S5. Cubic interpolation of PCi as a function of inclination angle for all Al GBs. Table S6. Cubic interpolation of PCi as a function of inclination angle for all Ni GBs.

Author Contributions: R.D.: Methodology, Investigation, Writing—original draft; W.Y.: Investigation, Supervision, Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data that supports the findings of this study are available in the supplementary material of this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rohrer, G.S. Grain boundary energy anisotropy: A review. *J. Mater. Sci.* **2011**, *46*, 5881. [CrossRef]
2. Cahn, J.W.; Mishin, Y.; Suzuki, A. Coupling grain boundary motion to shear deformation. *Acta Mater.* **2006**, *54*, 4953. [CrossRef]
3. Mishin, Y.; Asta, M.; Li, J. Atomistic modeling of interfaces and their impact on microstructure and properties. *Acta Mater.* **2010**, *58*, 1117. [CrossRef]
4. Cantwell, P.R.; Tang, M.; Dillon, S.J.; Luo, J.; Rohrer, G.S.; Harmer, M.P. Grain boundary complexions. *Acta Mater.* **2014**, *62*, 1. [CrossRef]
5. Suzuki, A.; Mishin, Y. Interaction of point defects with grain boundaries in fcc metals. *Interface Sci.* **2003**, *11*, 425. [CrossRef]
6. Bai, X.M.; Vernon, L.J.; Hoagland, R.G.; Voter, A.F.; Nastasi, M.; Uberuaga, B.P. Role of atomic structure on grain boundary-defect interactions in Cu. *Phys. Rev. B* **2012**, *85*, 204103. [CrossRef]
7. Tschopp, M.A.; Solanki, K.N.; Gao, F.; Sun, X.; Khaleel, M.A.; Horstemeyer, M.F. Probing grain boundary sink strength at the nanoscale: Energetics and length scales of vacancy and interstitial absorption by grain boundaries in a-Fe. *Phys. Rev. B* **2012**, *85*, 064108. [CrossRef]
8. Siegel, R.W.; Chang, S.M.; Balluffi, R.W. Vacancy Loss at Grain-Boundaries in Quenched Polycrystalline Gold. *Acta Met. Mater.* **1980**, *28*, 249. [CrossRef]
9. Dollar, M.; Gleiter, H. Point-Defect Annihilation at Grain-Boundaries in Gold. *Scr. Met.* **1985**, *19*, 481. [CrossRef]
10. Frolov, T.; Olmsted, D.L.; Asta, M.; Mishin, Y. Structural phase transformations in metallic grain boundaries. *Nat. Commun.* **2013**, *4*, 1899. [CrossRef]
11. Tucker, G.J.; McDowell, D.L. Non-equilibrium grain boundary structure and inelastic deformation using atomistic simulations. *Int. J. Plasticity* **2011**, *27*, 841. [CrossRef]
12. Olmsted, D.L.; Foiles, S.M.; Holm, E.A. Survey of computed grain boundary properties in face-centered cubic metals: I. Grain boundary energy. *Acta Mater.* **2009**, *57*, 3694. [CrossRef]
13. Olmsted, D.L.; Holm, E.A.; Foiles, S.M. Survey of computed grain boundary properties in face-centered cubic metals—II: Grain boundary mobility. *Acta Mater.* **2009**, *57*, 3704. [CrossRef]
14. Liu, M. Effect of uniform corrosion on mechanical behavior of E690 high-strength steel lattice corrugated panel in marine environment: A finite element analysis. *Mater. Res. Express* **2021**, *8*, 066510. [CrossRef]
15. Liu, M. Finite element analysis of pitting corrosion on mechanical behavior of E690 steel panel. *Anti-Corros. Methods Mater.* **2022**, *69*, 351. [CrossRef]
16. Elsheikh, A.H.; Sharshir, S.W.; Abd Elaziz, M.; Kabeel, A.E.; Guilan, W.; Haiou, Z. Modeling of solar energy systems using artificial neural network: A comprehensive review. *Sol. Energy* **2019**, *180*, 622. [CrossRef]
17. Elsheikh, A.H.; Elaziz, M.A.; Das, S.R.; Muthuramalingam, T.; Lu, S. A new optimized predictive model based on political optimizer for eco-friendly MQL-turning of AISI 4340 alloy with nano-lubricants. *J. Manuf. Process.* **2021**, *67*, 562. [CrossRef]

18. Elsheikh, A.H.; Katekar, V.P.; Muskens, O.L.; Deshmukh, S.S.; Elaziz, M.A.; Dabour, S.M. Utilization of LSTM neural network for water production forecasting of a stepped solar still with a corrugated absorber plate. *Process Saf. Environ. Prot.* **2021**, *148*, 273. [CrossRef]

19. Elsheikh, A.H.; Abd Elaziz, M.; Vendan, A. Modeling ultrasonic welding of polymers using an optimized artificial intelligence model using a gradient-based optimizer. *Weld. World* **2022**, *66*, 27. [CrossRef]

20. Elsheikh, A.H.; Shanmugan, S.; Sathyamurthy, R.; Kumar Thakur, A.; Issa, M.; Panchal, H.; Muthuramalingam, T.; Kumar, R.; Sharifpur, M. Low-cost bilayered structure for improving the performance of solar stills: Performance/cost analysis and water yield prediction using machine learning. *Sustain. Energy Technol. Assess.* **2022**, *49*, 101783. [CrossRef]

21. Gomberg, J.A.; Medford, A.J.; Kalidindi, S.R. Extracting knowledge from molecular mechanics simulations of grain boundaries using machine learning. *Acta Mater.* **2017**, *133*, 100. [CrossRef]

22. Rosenbrock, C.W.; Homer, E.R.; Csányi, G.; Hart, G.L.W. Discovering the building blocks of atomic systems using machine learning: Application to grain boundaries. *NPJ Comput. Mater.* **2017**, *3*, 29. [CrossRef]

23. Ye, W.; Zheng, H.; Chen, C.; Ong, S.P. A Universal Machine Learning Model for Elemental Grain Boundary Energies. *Scr. Mater.* **2022**, *218*, 114803. [CrossRef]

24. Homer, E.R.; Hart, G.L.W.; Braxton Owens, C.; Hensley, D.M.; Spendlove, J.; Serafin, L.H. Examination of computed aluminum grain boundary structures and energies that span the 5D space of crystallographic character. *Acta Mater.* **2022**, *234*, 118006. [CrossRef]

25. Song, X.; Deng, C. Atomic energy in grain boundaries studied by machine learning. *Phys. Rev. Mater.* **2022**, *6*, 043601. [CrossRef]

26. Snow, B.D.; Doty, D.D.; Johnson, O.K. A Simple Approach to Atomic Structure Characterization for Machine Learning of Grain Boundary Structure-Property Models. *Front. Mater.* **2019**, *6*, 120. [CrossRef]

27. Guziewski, M.; Montes de Oca Zapiain, D.; Dingreville, R.; Coleman, S.P. Microscopic and Macroscopic Characterization of Grain Boundary Energy and Strength in Silicon Carbide via Machine-Learning Techniques. *ACS Appl. Mater. Interfaces* **2021**, *13*, 3311. [CrossRef]

28. Huber, L.; Hadian, R.; Grabowski, B.; Neugebauer, J. A machine learning approach to model solute grain boundary segregation. *NPJ Comput. Mater.* **2018**, *4*, 64. [CrossRef]

29. Wagih, M.; Larsen, P.M.; Schuh, C.A. Learning grain boundary segregation energy spectra in polycrystals. *Nat. Commun.* **2020**, *11*, 6376. [CrossRef]

30. Homer, E.R.; Hensley, D.M.; Rosenbrock, C.W.; Nguyen, A.H.; Hart, G.L.W. Machine-Learning Informed Representations for Grain Boundary Structures. *Front. Mater.* **2019**, *6*, 168. [CrossRef]

31. Vieira, R.B.; Lambros, J. Machine Learning Neural-Network Predictions for Grain-Boundary Strain Accumulation in a Polycrystalline Metal. *Exp. Mech.* **2021**, *61*, 627–639. [CrossRef]

32. Zhang, S.; Wang, L.; Zhu, G.; Diehl, M.; Maldar, A.; Shang, X.; Zeng, X. Predicting grain boundary damage by machine learning. *Int. J. Plasticity* **2022**, *150*, 103186. [CrossRef]

33. Tschopp, M.A.; McDowell, D.L. Structural unit and faceting description of Sigma 3 asymmetric tilt grain boundaries. *J. Mater. Sci.* **2007**, *42*, 7806. [CrossRef]

34. Sutton, A.P.; Vitek, V. On the Structure of Tilt Grain Boundaries in Cubic Metals I. Symmetrical Tilt Boundaries. *Philos. Trans. R. Soc. London. Ser. A Math. Phys. Sci.* **1983**, *309*, 1.

35. Sutton, A.P.; Vitek, V. On the Structure of Tilt Grain Boundaries in Cubic Metals II. Asymmetrical Tilt Boundaries. *Philos. Trans. R. Soc. London. Ser. A Math. Phys. Sci.* **1983**, *309*, 37.

36. Sutton, A.P.; Vitek, V. On the Structure of Tilt Grain Boundaries in Cubic Metals. III. Generalizations of the Structural Study and Implications for the Properties of Grain Boundaries. *Philos. Trans. R. Soc. London. Ser. A Math. Phys. Sci.* **1983**, *309*, 55.

37. Francis, T.; Chesser, I.; Singh, S.; Holm, E.A.; De Graef, M. A geodesic octonion metric for grain boundaries. *Acta Mater.* **2019**, *166*, 135. [CrossRef]

38. Patala, S. Understanding grain boundaries—The role of crystallography, structural descriptors and machine learning. *Comp. Mater. Sci.* **2019**, *162*, 281. [CrossRef]

39. Banadaki, A.D.; Patala, S. A three-dimensional polyhedral unit model for grain boundary structure in fcc metals. *NPJ Comput. Mater.* **2017**, *3*, 13. [CrossRef]

40. Jolliffe, I.T. *Principal Component Analysis*, 2nd ed.; Springer: New York, NY, USA, 2002.

41. Aggarwal, C.C. *Neural Networks and Deep Learning: A Textbook*; Springer: New York, NY, USA, 2018.

42. Nielsen, M. *Neural Networks and Deep Learning*; Determination Press: Los Angeles, CA, USA, 2015.

43. Faken, D.; Jónsson, H. Systematic analysis of local atomic structure combined with 3D computer graphics. *Comp. Mater. Sci.* **1994**, *2*, 279. [CrossRef]

44. Kelchner, C.L.; Plimpton, S.J.; Hamilton, J.C. Dislocation nucleation and defect structure during surface indentation. *Phys. Rev. B* **1998**, *58*, 11085. [CrossRef]

45. Daw, M.S.; Baskes, M.I. Embedded-Atom Method—Derivation and Application to Impurities, Surfaces, and Other Defects in Metals. *Phys. Rev. B* **1984**, *29*, 6443. [CrossRef]

46. Daw, M.S.; Foiles, S.M.; Baskes, M.I. The embedded-atom method: A review of theory and applications. *Mater. Sci. Rep.* **1993**, *9*, 251. [CrossRef]

47. Voter, A.F.; Chen, S.P. Accurate Interatomic Potentials for Ni, Al and Ni3Al. In *MRS Proceedings*; Materials Research Society: Warrendale, PA, USA, 1986.

48. Mishin, Y.; Farkas, D.; Mehl, M.J.; Papaconstantopoulos, D.A. Interatomic potentials for monoatomic metals from experimental data and ab initio calculations. *Phys. Rev. B* **1999**, *59*, 3393. [CrossRef]

49. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comp. Phys.* **1995**, *117*, 1. [CrossRef]

50. Alexander, S. Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool. *Model Simul. Mater. Sci. Eng.* **2010**, *18*, 015012.

51. Sansoz, F.; Molinari, J.F. Incidence of atom shuffling on the shear and decohesion behavior of a symmetric tilt grain boundary in copper. *Scr. Mater.* **2004**, *50*, 1283. [CrossRef]

52. Vitek, V.; Sutton, A.P.; Gui Jin, W.; Schwartz, D. On the multiplicity of structures and grain boundaries. *Scr. Met.* **1983**, *17*, 183. [CrossRef]

53. Spearot, D.E.; Tschopp, M.A.; Jacob, K.I.; McDowell, D.L. Tensile strength of < 100 > and < 110 > tilt bicrystal copper interfaces. *Acta Mater.* **2007**, *55*, 705.

54. Tschopp, M.A.; Tucker, G.J.; McDowell, D.L. Structure and free volume of < 110 > symmetric tilt grain boundaries with the E structural unit. *Acta Mater.* **2007**, *55*, 3959.

55. Spearot, D.E.; Jacob, K.I.; McDowell, D.L. Nucleation of dislocations from [0 0 1] bicrystal interfaces in aluminum. *Acta Mater.* **2005**, *53*, 3579. [CrossRef]