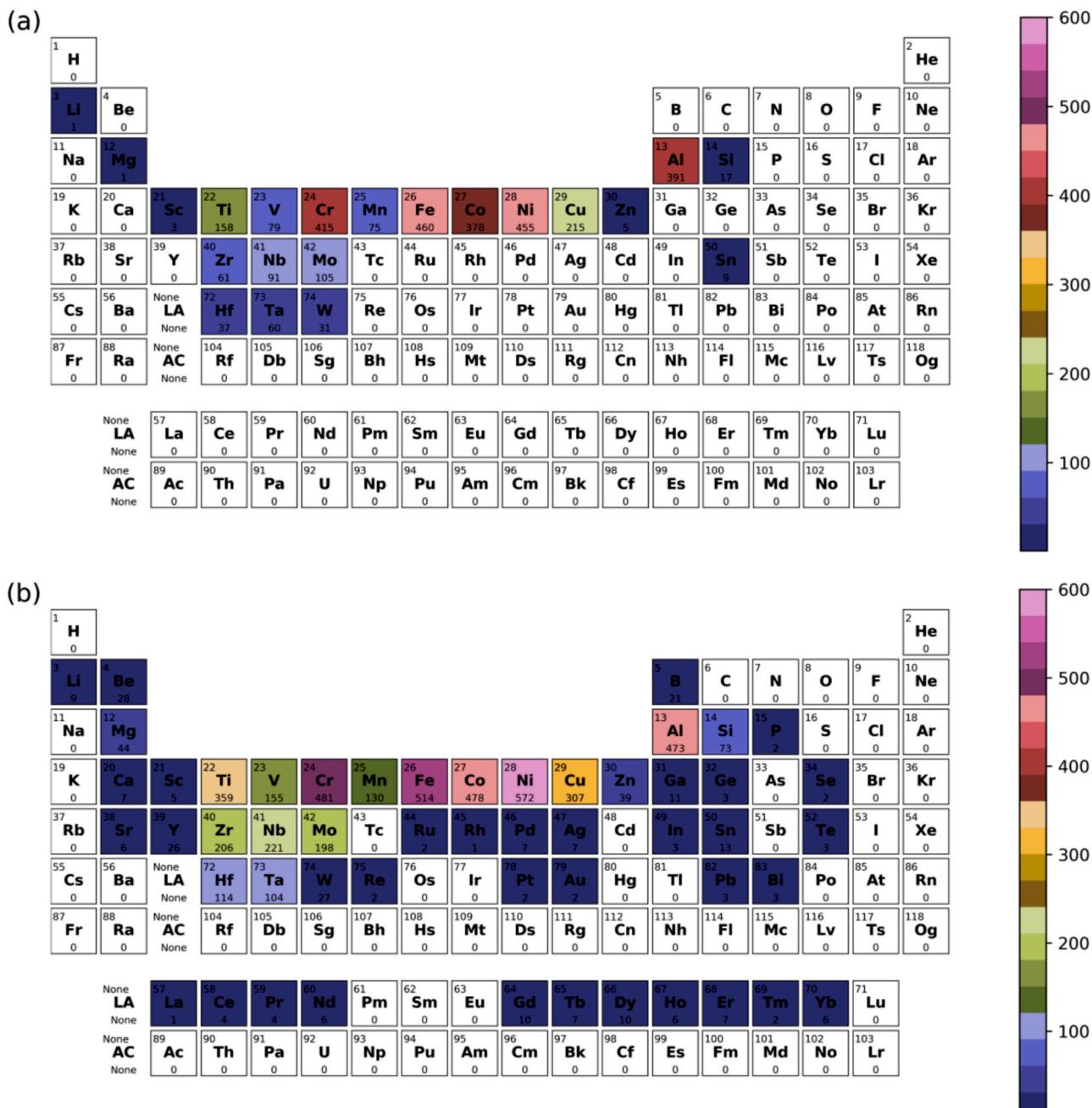




physical parameters as one-dimensional descriptors input of the selected ML model. These physical parameters are calculated based on the composition and properties of elements, such as the configuration entropy [45,46] and the mixing enthalpy [45,47]. In addition, the amount of data on HEAs subject to processing conditions other than as-cast is small, and most works ignore the processing conditions.

In order to strengthen the representation ability of PTR for HEAs hardness prediction, it is worth exploring that how to incorporate useful and hardness-related information, such as physical parameters, processing condition and phase information, into PTR images in appropriate way. Physical parameters commonly used in previous works are put forward based on the domain knowledge of HEAs hardness researches. Coupling them with PTR can enhance the material representation ability of PTR. The material properties are determined by the

composition and structure [48], thus the addition of processing condition and phase information would be beneficial to improve the prediction of material properties. Different processing conditions, such as as-cast and work hardening, would cause significant different in the internal structure of the material [49,50]. The phase represents the microstructure of the material. For alloys, face center cubic (FCC), L12, L10 phases generally possess relatively low hardness, body center cubic (BCC), B2, Heusler and intermetallic compound (IM), such as  $\sigma$ , Laves,  $\eta$  phases generally have relatively high hardness, and the mixed phase may produce uncertain properties [51]. The specific representation methods for these material information should also be explored. The four famous factors of the materials science tetrahedron include composition/structure, processing, property and performance, and they are interconnected according to potential physical laws [52]. The idea of



**Fig. 1.** The number of element occurrences for (a) HEAs hardness dataset and (b) HEAs phase dataset are shown against a periodic table background, and white squares indicate elements which are not included in the dataset.

coupling composition, structure, processing and physical parameters in periodic table representation to predict material property inspires this work.

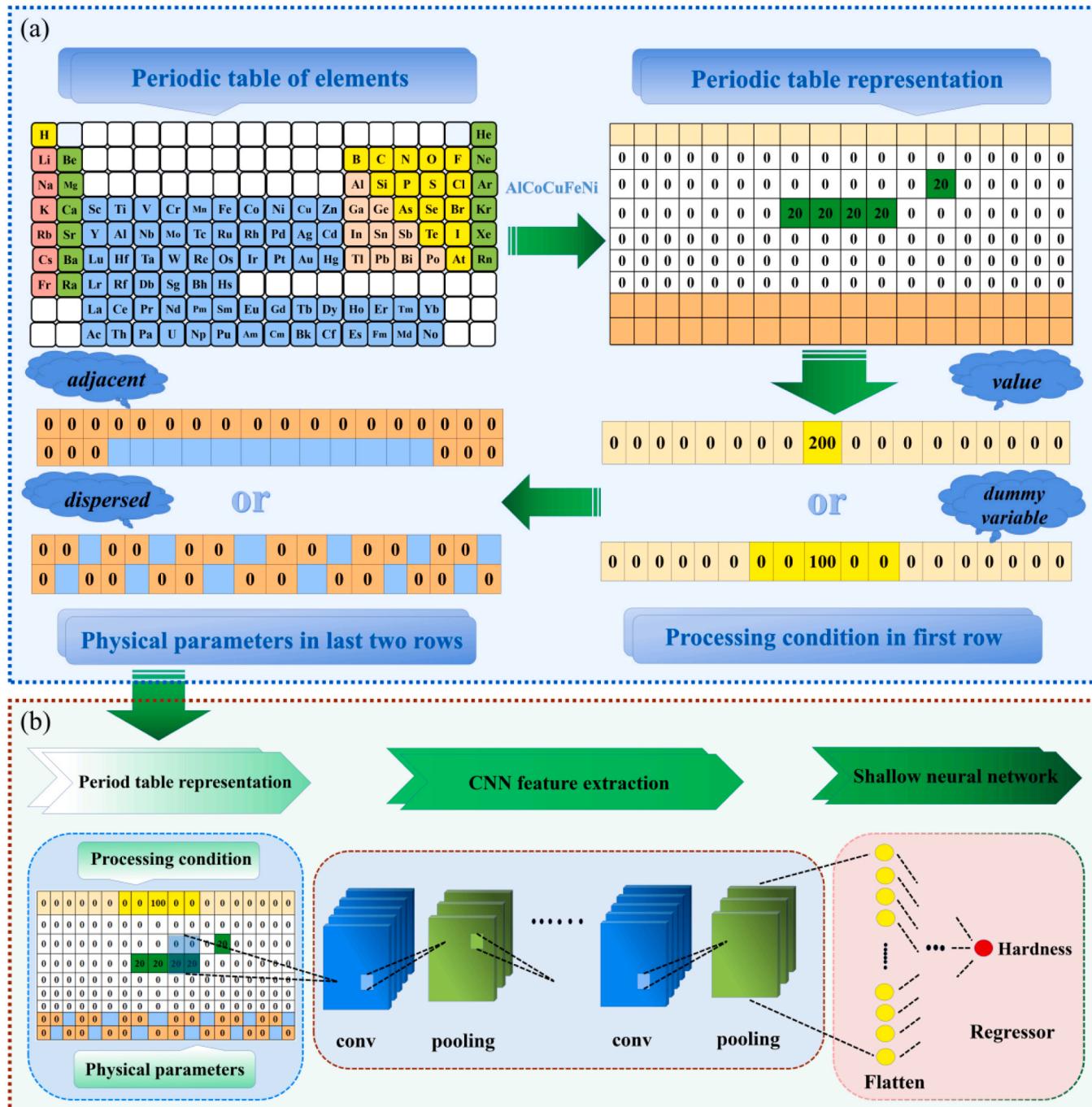
In this work, a ML framework based on the combination of CNN model and PTR was proposed for the prediction of HEAs hardness. Composition, processing condition, empirical physical parameters and phase information were incorporated into PTR, and different methods of representing these information were compared. Hardness regression, hardness classification and alloy system extrapolation of HEAs were further explored. Moreover, the stacking method, an ensemble learning strategy, was adopted to improve model prediction accuracy and

achieve stable results. Finally, Diebold-Mariano (DM) test [53], a well-known statistical test method, was introduced in model comparisons to demonstrate the statistical reliability of model improvement caused by information addition. This framework could be extended to other material property prediction tasks.

## 2. Methods

### 2.1. Dataset briefing

The 585 HEAs hardness data collected from the literature [54–62]



**Fig. 2.** The construction process of the CNN+PTR model using AlCoCuFeNi as an example. (a) Adding processing condition and physical parameters to the periodic table representation of the HEAs hardness dataset. There are two ways to represent processing condition and physical parameters, respectively. These ways could be combined into different types of PTR representation methods. (b) Constructing the convolutional neural network hardness regression model based on one of the PTR representation methods.

are adopted as HEAs hardness dataset. This dataset involves composition, processing condition and Vickers hardness value. The processing conditions include five categories: as-cast, homogenized, work hardening, powder metallurgy and additive manufacturing. Among them, the number of as-cast with 483 is the highest. In addition, 12 empirical physical parameters are calculated and they are all related to the HEAs hardness [55]. The details of the physical parameters are given in the supplementary. The 895 HEAs phase data collected from the literature [63] are adopted as the phase dataset. The phase dataset involves composition and phase information with 68 specific phase types. These phases could be categorized into different classes, such as FCC, BCC and IM. The processing conditions of HEAs in phase dataset are all as-cast. For these two datasets, the occurrence numbers of the elements are shown in the Fig. 1. It could be seen that common alloy elements, such as Al, Co, Cr are relatively more, while some elements such as inert gases and halogens are not included in the HEAs datasets.

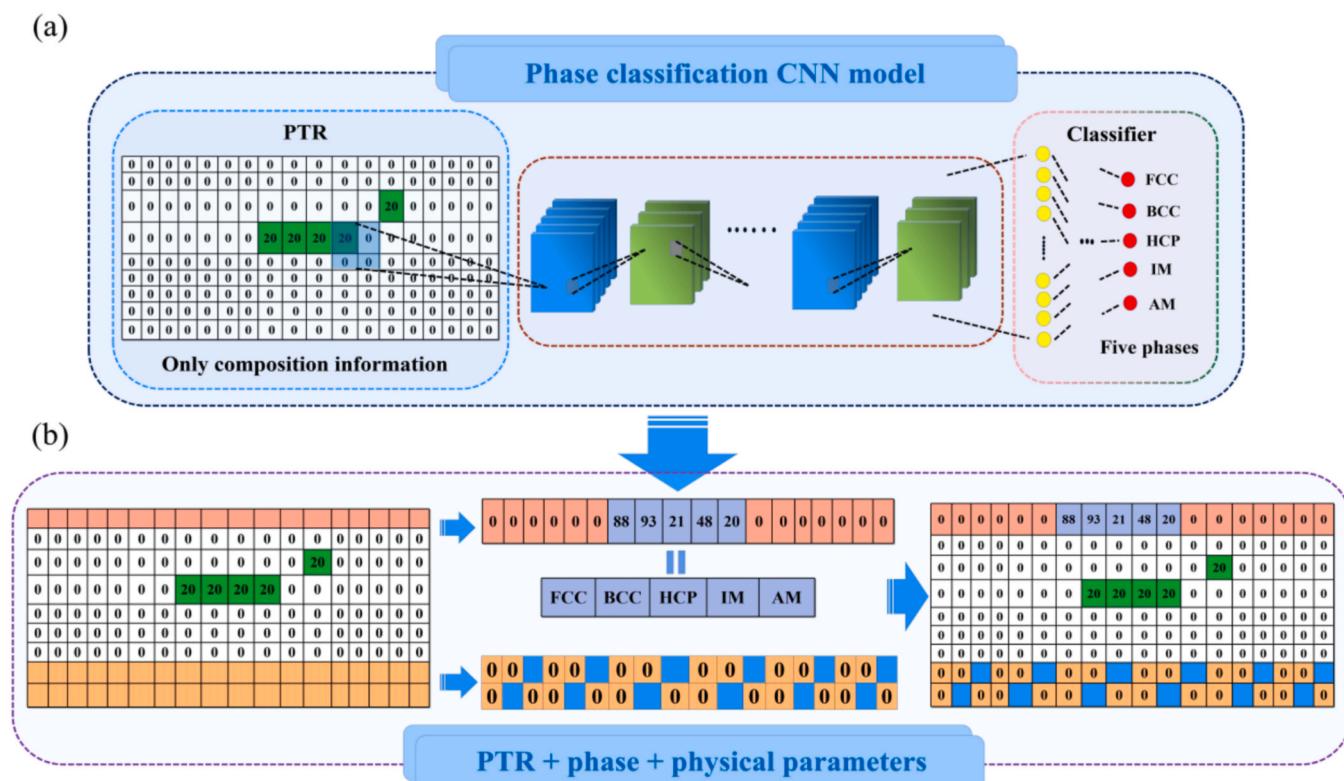
## 2.2. Periodic table representation of HEAs

This work mainly consists of two parts: (a) Implementing the periodic table representation with composition, processing condition and physical parameters to construct CNN hardness prediction model. (b) Constructing PTR+CNN classification model for HEAs phase classification, and then incorporating the phase information, predicted by the phase classification model, along with composition and physical parameters into PTR to construct CNN hardness prediction model.

In part (a), the mapping process of HEAs is shown in Fig. 2(a), which takes AlCoCuFeNi as an example. The PTR image consists of a matrix of  $9 \times 18$ , and its rows and columns correspond to periods and groups of the periodic table, respectively. Thus, the composition information of HEAs could be represented by filling the atomic percentage of each constituent element into its corresponding position in the matrix and other positions are filled with zeros. Different representation methods would affect the mining and learning of useful information. Five

processing conditions could be represented as different values of 0/100/200/300/400 in the middle position of the first row, or represented as dummy variables in the middle five positions of the first row, such as [0, 0, 100, 0, 0]. As shown in Fig. 1 and Fig. 2(a), the lanthanides (LA) and actinides (AC) are in the third subgroup of the sixth and seventh periods of the periodic table, respectively. For ease of presentation, they are shown in the eighth and ninth rows of the periodic table. As a result, their properties are not as closely related to their positions as the other elements. In addition, the lanthanides and actinides are not included in the hardness dataset. For these reasons, as shown in Fig. 2(a), the 12 physical parameters are presented in the eighth and ninth rows of the PTR. In order to scale the physical parameters to the same order of magnitude to avoid unequal use of physical parameters with different physical meanings and scales, each physical parameter is standardized by the z-score method,  $x' = \frac{x-\mu}{\sigma}$ , where  $x$  and  $x'$  are the physical parameters before and after standardization, respectively;  $\mu$  and  $\sigma$  are the mean and standard deviation of the physical parameter, respectively. There are two arrangement methods for physical parameters. One is to place them in adjacent positions in the middle of the last row. Another is to place them in dispersed positions which are separated by zero value pixels in the last two rows, each two closest physical parameters in the same row are separated by two pixels with zero value, and each two closest physical parameters in different rows are horizontally staggered.

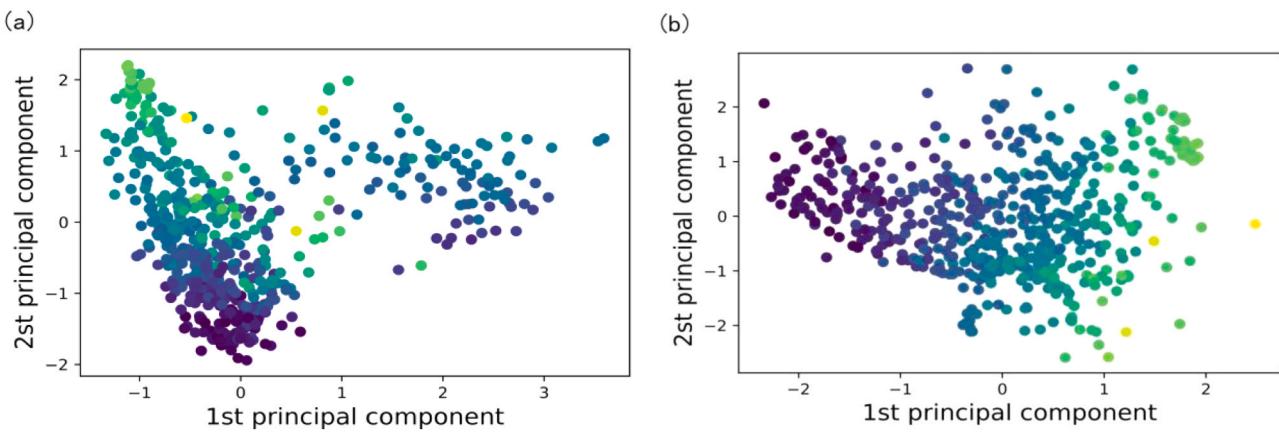
In part (b), all HEAs data used in phase classification and hardness regression are fixed as the as-cast processing condition. As shown in Fig. 3, in the phase classification CNN model which is trained in phase dataset, the input is the PTR which only contains composition information. As for the output, most ML phase classification works in previous researches classified HEAs into two to six mutually exclusive classes, i.e. each HEA is assigned to only one specific class [64–66], such as BCC or BCC+FCC. Therefore, these models could only predict the phases that appeared in the training set. An alternative approach is the multi-label classification scheme, which supports the prediction of



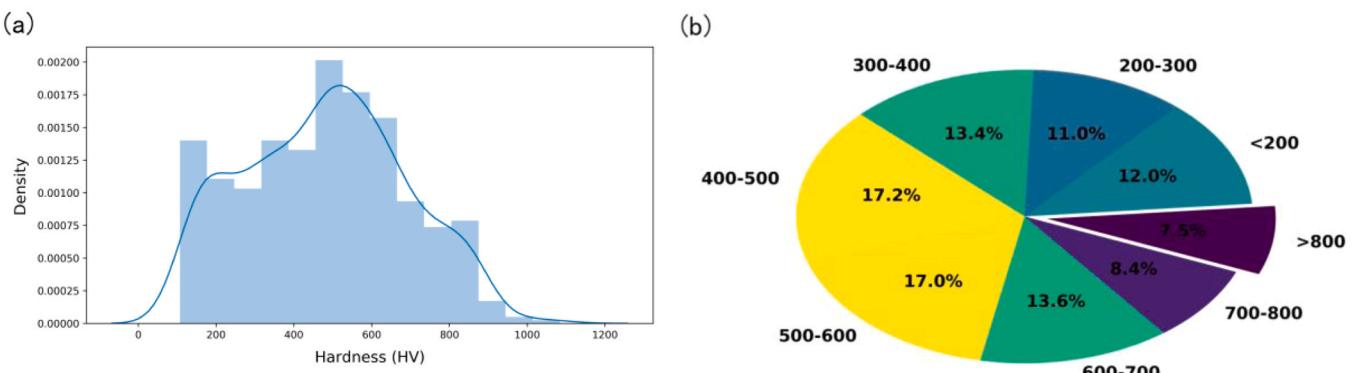
**Fig. 3.** Taking AlCoCuFeNi as an example, (a) constructing the CNN classification model for HEAs phase classification, and (b) adding predicted phase information and physical parameters into the PTR for HEAs hardness prediction.



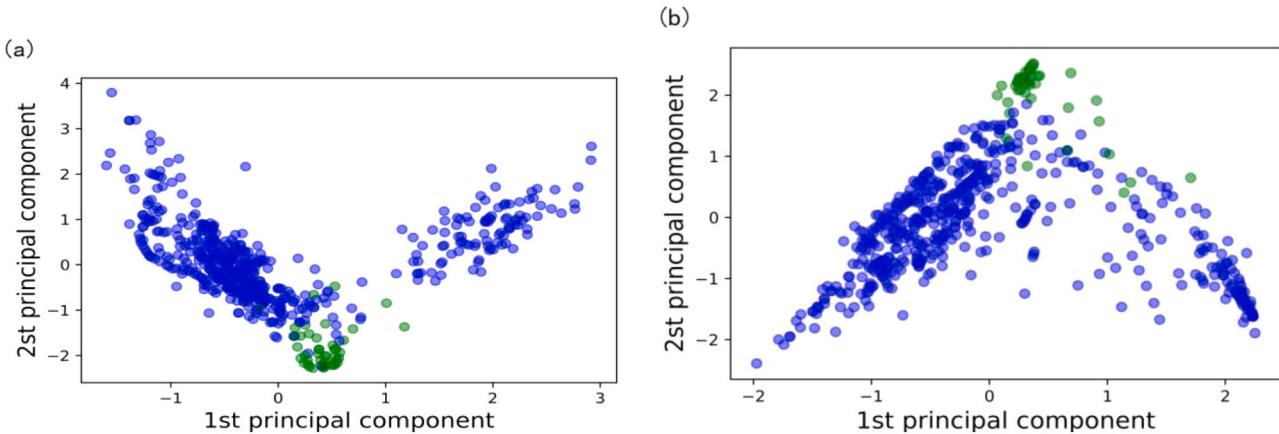




**Fig. 5.** The two-dimensional plots of PCA-compressed feature data for (a) Model-0 and (b) Model-1.



**Fig. 6.** The distribution of HEAs hardness. (a) The distribution densities of hardness. (b) The pie chart of hardness zones.



**Fig. 7.** The two dimensions features data diagrams based on (a) model-0 input and (b) model-1 input. Blue points represent the HEAs with hardness values below 800 HV, and green points represent the opposite.

tightening the concept of alloy system extrapolation, the HEAs data with the specific element are adopted as the extrapolation data. The prediction performance of model-0 and model-1 in extrapolation data is estimated by the RMSE and  $R^2$  mean value of twenty times training and testing. The specific extrapolation cases and the extrapolation results are shown in the Table 3.

It could be seen that the model-1 is obviously better than model-0 for three different alloy system extrapolation schemes. As for model-0, it extracts information only from material composition and element positions in the periodic table. It is easy to understand that the information

extracted from component is insufficient for alloy system extrapolation. Although the element property related information which is extracted from the element positions may be helpful, the extrapolation ability is still weak because this information is implicitly contained in the PTR.

As for model-1, it could extract information from empirical physical parameters. The empirical physical parameters are calculated based on the material composition and physical properties of the constituent elements, and they generally represent the same physical information for different HEAs composition systems. For example, it is generally believed that for any alloy system, the greater the electronegativity







