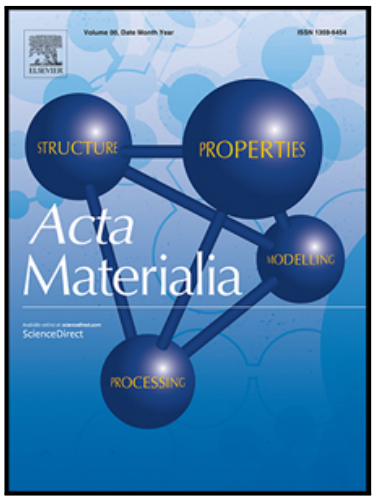


Dramatically Enhanced Combination of Ultimate Tensile Strength and Electric Conductivity of Alloys via Machine Learning Screening

Hongtao Zhang , Huadong Fu , Xingqun He , Changsheng Wang , Lei Jiang , Long-Qing Chen , Jianxin Xie

PII: S1359-6454(20)30762-X
DOI: https://doi.org/10.1016/j.actamat.2020.09.068
Reference: AM 16363

To appear in: *Acta Materialia*

Received date: 17 June 2020
Revised date: 24 August 2020
Accepted date: 19 September 2020

Please cite this article as: Hongtao Zhang , Huadong Fu , Xingqun He , Changsheng Wang , Lei Jiang , Long-Qing Chen , Jianxin Xie , Dramatically Enhanced Combination of Ultimate Tensile Strength and Electric Conductivity of Alloys via Machine Learning Screening, *Acta Materialia* (2020), doi: https://doi.org/10.1016/j.actamat.2020.09.068

Dramatically Enhanced Combination of Ultimate Tensile Strength and Electric Conductivity of Alloys via Machine Learning Screening

Hongtao Zhang[a,b,#], Huadong Fu[a,b,#], Xingqun He[a,b], Changsheng Wang[a,b], Lei Jiang[a,b], Long-Qing Chen[c,*], Jianxin Xie[a,b,*]

[a] Beijing Advanced Innovation Center for Materials Genome Engineering, Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing 100083, China
[b] Key Laboratory for Advanced Materials Processing of Ministry of Education, Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing 100083, China
[c] Department of Materials Science and Engineering, The Pennsylvania State University, University Park, PA 16802, USA

[*]Corresponding author and E-mail: Jianxin Xie (jxxie@mater.ustb.edu.cn), Long-Qing Chen (lqc3@psu.edu)

[#]These authors contributed equally: Hongtao Zhang, Huadong Fu

Abstract

Optimizing two conflicting properties such as mechanical strength and toughness or dielectric constant and breakdown strength of a material has always been a challenge. Here we propose a machine learning approach to dramatically enhancing the combined ultimate tensile strength (UTS) and electric conductivity (EC) of alloys by identifying a set of key features through correlation screening, recursive elimination and exhaustive screening of existing datasets. We demonstrate that the key features responsible for solid solution strengthened conductive Copper alloys are absolute electronegativity, core electron distance, and atomic radius, based on which, we discovered a series of new alloying elements that can significantly improve the combined UTS and EC. The

predictions are then validated by experimentally fabricating four new Cu-In alloys which could potentially replace the more expensive Cu-Ag alloys currently used in railway wiring. We show that the same set of key features can be generally applicable to designing a broad range of conductive alloys.

Keywords: machine learning, feature screening, alloy design, copper alloy, aluminum alloy

1. Introduction

It has always been a challenge to optimize a desirable combination of two conflicting properties of a material, e.g., the strength and toughness of an alloy or the dielectric constant and break-down strength of a dielectric.

Solid solution strengthened conductive copper alloys (SSCA) are a class of engineering alloys that have been widely used in electrical railway contact wires and other fields [1] due to their excellent mechanical and electrical properties [2,3]. However, to further improve the combination of their mechanical strength and electric conductivity, e.g., through alloying elements, is challenging. For example, SSCAs which possess higher ultimate tensile strength tend to have lower electric conductivity and vice versa. Here, instead of performing trial-and-error experimentations using different alloying elements, we explore machine learning approach to screen alloying elements which may significantly enhance the combination of their enhanced strength

and conductivity.

Machine learning has increasingly been employed to guide the design of materials [4-6]. For example, Lookman et al [7,8] successfully developed high-property shape memory alloys by combining a machine learning model and an adaptive learning strategy. Raccuglia et al [9] used a machine learning approach to learn the laws of material synthesis from failed experiments, and succeeded in designing new materials. We previously applied data-driven methods to establish a property-oriented alloy composition design system MLDS, which achieved inverse composition design of high-property copper alloys [10]. Similar machine learning approaches have been applied to the composition design of piezoelectric materials [11,12], inorganic superconducting materials [13], stainless steel [14], and high-entropy alloys [15], as well as to the structure and property predictions such as diffusion [16,17], density [18], lattice misfit [19], glass forming ability [20,21], Seebeck coefficient [22], atomic force field [23], elastic constant [24], strength [25], fatigue [26-28], etc.

Many of the existing machine learning efforts employed the amounts of particular elements as inputs or outputs to the establish composition-property regression models for the prediction and design of structures and properties. There have also been attempts [8,9,12,13,16,17,20-22] to use the physical and chemical features of the alloying elements as inputs to improve the model predictions. If we could identify a set of physical and chemical features of an element as the key factors that determine the microstructures and properties of alloys, it is possible to employ them to screen and

design alloys with desired properties or significantly improved properties.

In this work, we propose an approach for the compositional screening of elements and rational design of alloys using a set of key element features that we identified based on machine learning. We validated this approach by synthesizing alloys with the screened alloy compositions and experimental measurements of their properties.

2. Methods

2.1. Alloy Design Framework

Our proposed framework for the rational alloy design approach is shown in Figure 1. It consists of: data collection → feature extraction → feature screening → modeling → composition design and experimental verification.

(1) Data collection. The availability of high-quality data is a prerequisite to accurately screen the key element features that largely determine the alloy properties [8,11]. We collected 27 sets of data (Supporting Information Table S1 for details) on solid solution strengthened conductive copper alloys from the *ASM specialty handbook: copper and copper alloys* [29], including their composition, ultimate tensile strength (UTS), and electrical conductivity (EC). The final states of all these alloys were annealed with average grain sizes of about 25 μm (OS025 state). The composition ranges for the alloying element contents in the dataset are listed in Table 1.

Table 1. Alloying elements in dataset and their content ranges

Alloying element	Sn	Zn	Ni	Mg	P
Content (wt%)	0~11.4	0~36.4	0~15	0~0.4	0~0.2

(2) Feature extraction. We extracted a total of 69 types of physical and chemical features for the elements [30], including atomic number factor, electro-chemical factor, group number factor, cohesion-energy factor, Mendeleev number factor, and size factor, and we classified the features into 6 categories following reference [31]. The results are summarized in Supporting Information Table S2.

(3) Feature screening. We use f_{ij}, where i and j label a given feature and a certain element, respectively, to denote the element features, and α_j to represent the content of an element. We then compute the mean f_{mi} and variance f_{vi} for each alloy for the 69 features according to following equation (1) and equation (2),

$$f_{mi} = \Sigma (f_{ij} \alpha_j) / \Sigma \alpha_j \quad (1)$$

$$f_{vi} = \Sigma [(f_{ij} - f_{mi})^2 \alpha_j] / \Sigma \alpha_j \quad (2)$$

We will refer f_{mi} and f_{vi} as alloy factors below. We employ correlation screening, recursive elimination [32] and exhaustive screening [33] and utilize both f_{mi} and f_{vi} as inputs to screen the key alloy factors that have the largest impacts on the UTS and EC of the alloys.

(4) Establishing relations between key factors and properties. With the key alloy factors as inputs, we establish the analytical relationship or model between the key alloy factors and UTS and as well as the relationship or model between the key alloy factors and EC.

(5) Composition design and experimental verification. We down-select the alloying elements that are expected to improve both the mechanical and electric properties of the alloys and experimentally validate the models by synthesizing the down-selected alloys

and characterizing their properties.

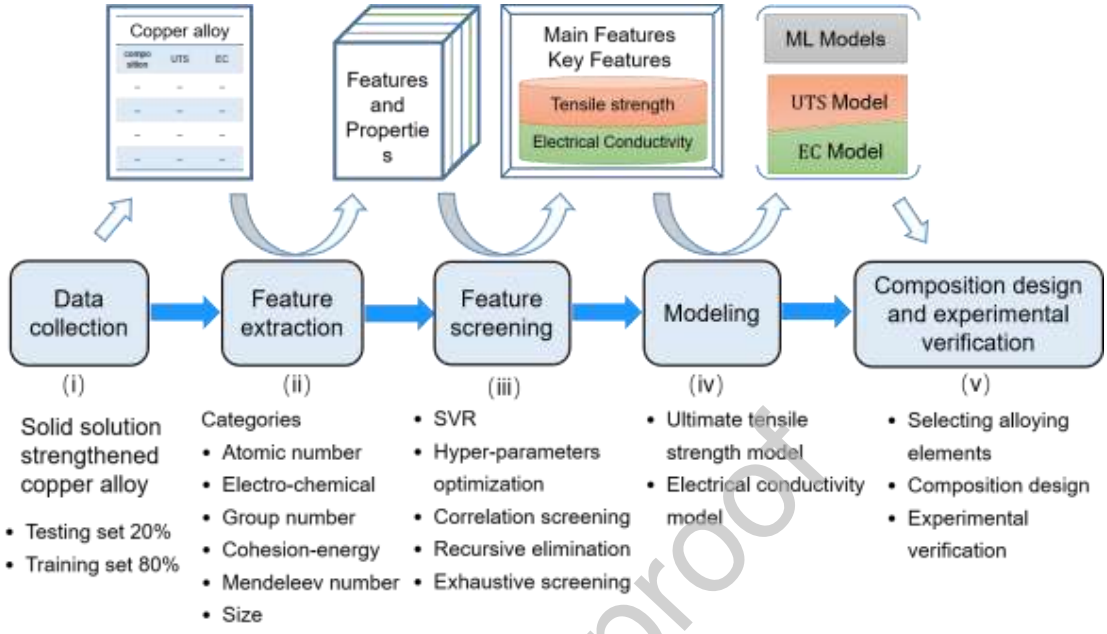


Figure 1. Alloy composition rational design strategy based on element features screening

2.2. Machine learning

We employed the Support Vector Regression (SVR) with the radial basis function as the kernel function for performing machine learning. We randomly selected 5 sets of data for testing from the 27 sets of sample data, and used the remaining 22 sets of data for training. We optimized the two hyper-parameters of SVR, cost of making mistakes and gamma controling the shape of the segmented hyperplane utilizing the 10-fold Cross Validation method [34]. The model error was determined by averaging the errors over 10 validation data sets using the following equation (3),

$$d = \frac{1}{n}\sum \frac{\left|P_{x-predict} - P_{x-experiment}\right|}{P_{x-experiment}} \times 100\% \quad (3)$$

where d is the error, and $P_{x-predict}$ and $P_{x-experiment}$ are the predicted properties and measured properties, respectively.

2.3. Feature screening

To screen the key alloy factors, we computed a total of 138 alloy factors using equations (1) and (2) and then apply a three-step feature screening method based on SVR. The first step is correlation screening to quickly remove redundant alloy factors. We calculated the Pearson correlation coefficient r [35] between any two alloy factors x and y based on the following equation (4),

$$r = \frac{\sum\left[\left(x_i - x_m\right)\times\left(y_i - y_m\right)\right]}{\sqrt{\sum\left(x_i - x_m\right)^2}\sqrt{\sum\left(y_i - y_m\right)^2}} \quad (4)$$

where x_i and y_i are the values of two alloy factors in the i-th alloy (i = 1, 2, …, 22), x_m and y_m are the average values of these two alloy factors in all 22 alloys. Values of $|r| > 0.95$ [12] imply a strong linear correlation between the two alloy factors x and y, indicating that they carry the same or similar information and hence have the same or similar influences on alloy properties. In this case, we use both alloy factors as input to establish models, but eliminate the alloy factor with the larger model error.

The second step is recursive elimination. We choose one of the n alloy factors in turn, and use the remaining n-1 alloy factors as inputs to build an SVR model. The alloy factor corresponding to the smallest model error was then removed, leaving n-1 alloy factors remained. We perform the elimination recursively until the minimum model error changes from a downward to an upward trend (See the red marks in Figs.2 (b) and (c)).

The third step is exhaustive screening. We establish a series of SVR models by using all exhaustive combinations of the remaining alloy factors from the recursive elimination step as inputs. We finally identify the key alloy factors that most significantly affect the alloy properties by comparing model errors.

2.4. Experimental procedures

Based on thee screened compositions, we produced four new In-containing copper alloys by performing vacuum intermediate frequency induction melting → 500 ℃ @ 72h homogenization → 90% deformation cold rolling with thickness 10 mm to 1 mm → 500 ℃ @ 15 min annealing treatment. The final state of the alloys has an average grain size of about 25 μm (OS025), consistent with the processing state of the 27 sample alloys employed in the machine learning.

3. Results

3.1. Feature screening and modeling

Figure 2 (a) and (b) show the correlation screening process and the correlation coefficients of the top 20 alloy factors with strong correlation. When the correlation coefficient of the two alloy factors $|r|>0.95$, two machine learning models were established with each one of the two alloy factors as input. The errors of the two models were compared, and the alloy factor with the larger model error was removed. After correlation screening, we obtained 29 alloy factors each associated with UTS and EC, respectively. The results are listed in Supporting Information Table S3.

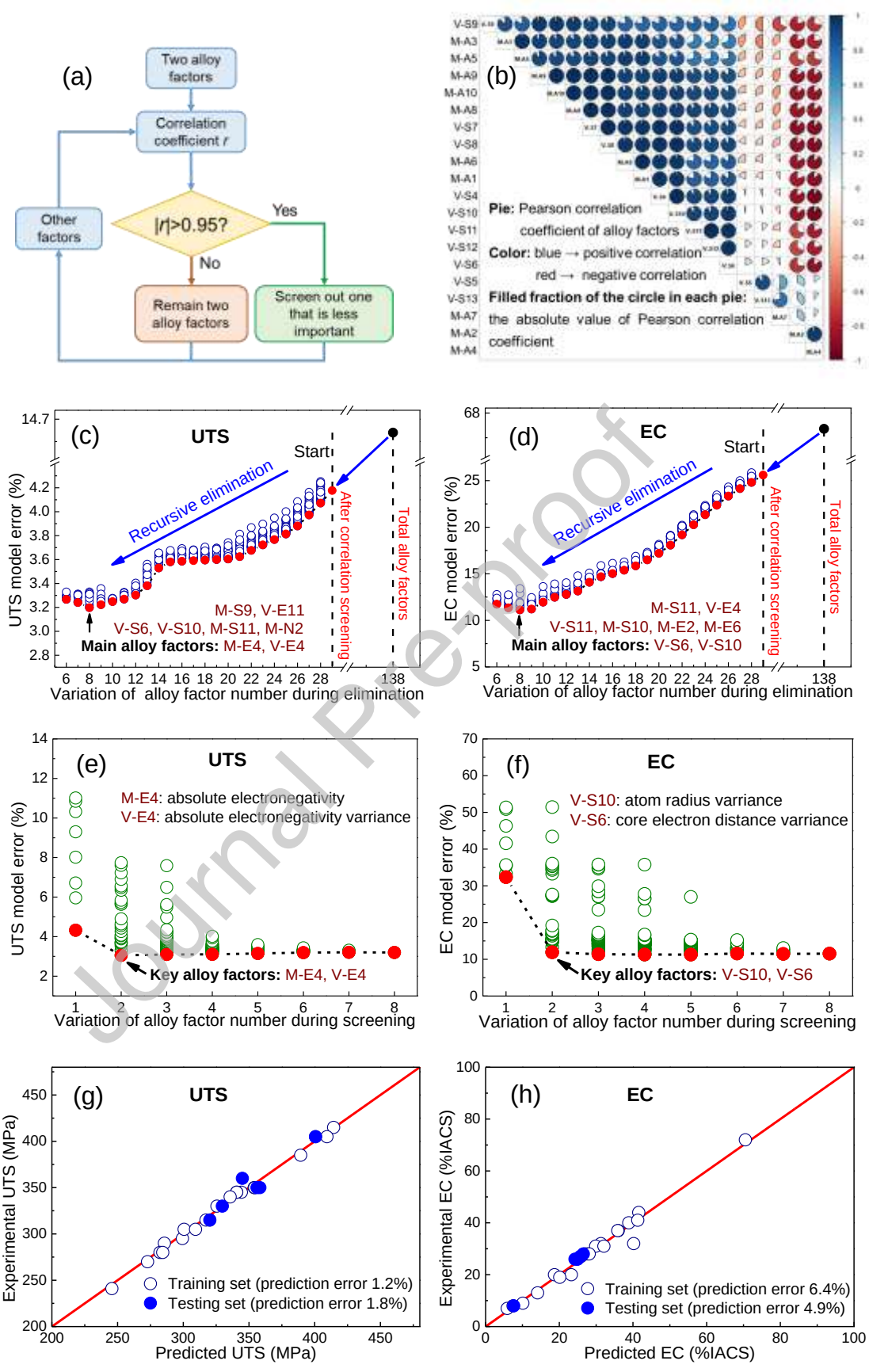


Figure 2. Results of alloy factor screening and modeling: (a) (b) Correlation screening. The flow chart (a) shows the correlation screening process, the pie chart (b) shows the Pearson correlation coefficients of alloy factors. Blue and red colors indicate positive and negative correlations, respectively, and the filled fraction of the circle in each of the pie charts corresponds to the absolute value of correlation coefficient. We obtained 29 alloy factors each associated with UTS and EC after correlation screening. **(c) (d) Recursive elimination.** Each dot indicates a combination of alloy factors and the solid dots are the combinations with minimum model error under the number of alloy factors after elimination. We have 8 main alloy factors affecting UTS and 8 main alloy factors affecting EC after recursive elimination. "M-" represents feature mean factor, and "V-" represents feature variance factor. **(e) (f) Exhaustive screening.** The dots indicate all combinations of 8 main alloy factors, and we have 2 key alloy factors affecting UTS and 2 key alloy factors affecting EC after exhaustive screening. **(g) (h) The performance of the UTS model and the EC model with the key alloy factors as inputs.**

Figures 2 (c) and (d) represent the results from the recursive elimination of alloy factors associated with UTS and EC, respectively. As the recursive elimination process proceeds, the error of the SVR models gradually decreases, indicating that the eliminated alloy factors contribute little to the model, and reducing the variable dimensions can improve the prediction accuracy of the SVR model. After the number of remaining alloy factors decreases to 8, the SVR model error begins to increase with further recursive elimination. It indicates that the alloy factors eliminated afterwards contain important information, and the recursive elimination screening is considered finished. As a result, we have 8 main alloy factors affecting UTS and 8 main alloy factors affecting EC after recursive elimination. These main alloy factors are related to

lattice constant, atomic and electronic size, electronegativity, and electrochemical features, as shown in **Table 2**.

Table 2. Main alloy factors obtained by recursive elimination

Main alloy factors affecting UTS		Main alloy factors affecting EC	
Factors	Description	Factors	Description
M-S11	Lattice constant *a*	M-S11	Lattice constant *a*
V-E4	Absolute electronegativity variance	V-E4	Absolute electronegativity variance
V-S6	Core electron distance variance	V-S6	Core electron distance variance
V-S10	Atomic radius variance	V-S10	Atomic radius variance
M-E4	Absolute electronegativity	M-S10	Atomic radius
M-S9	Atomic environment number	M-E2	Pauling electronegativity
M-N2	Mendeleev Number	M-E6	Second energy ionization
V-E11	Nuclear charge effective variance	V-S11	Lattice constant *a* variance

Figures 2 (e) and (f) are the results from exhaustive screening using 8 main alloy factors each associated with UTS and EC, respectively. The figures indicate that many combinations with 2-8 main alloy factors lead to the same minimum error level. It shows that the exhaustive method can effectively reduce the variable dimension and improve the design efficiency. For 2 main alloy factors, the error associated with the UTS model is the smallest, 3.1%, for the combinations of M-E4 and V-E4, and the error of the EC model is the smallest, 11.9%, for the combination of V-S6 and V-S10. In this work we call M-E4, V-E4, V-S6, and V-S10 as key alloy factors, and the corresponding elemental features for the four combinations of factors are the absolute electronegativity E4, core electron distance S6, and atomic radius S10. The key element feature associated with UTS is absolute electronegativity E4 while the key element features associated with EC are core electron distance S6 (defined as the maximum distance

from nucleus to core electrons) and atomic radius S10 (the size of atoms containing nucleus, core electrons and valence electrons) [31].

Based on the above analysis, we established two efficient models for property predictions of alloys. One is the UTS prediction model incorporating the key alloy factors M-E4 and V-E4 as inputs, and the other is the EC prediction model using the key alloy factors V-S6 and V-S10 as inputs. The predicted properties and the corresponding experimentally measured properties of the two models are shown in Figures 2 (g) and (h). The results demonstrate that both the UTS and EC models yield satisfactory prediction accuracies, with less than 2% error for the UTS prediction and less than 7% error for the EC prediction.

3.2. Alloying element selections

In order to improve the comprehensive properties of the UTS and EC of the solid solution strengthened copper alloys, according to binary copper alloy phase diagrams [36], we selected alloying elements that can form solid solutions with Cu, and with their solubility limits in Cu greater than 0.5wt% at room temperature, which is to ensure that the application range of the alloy system is not so narrow. Then, a number of alloying elements were obtained, including In, Ga, Al, Ti, Sn, Rh, Co, Cd, Ni, Zn, Mg, Ge, Si, Sb, Pt, P and Au, a search space of Cu-0.5wt%Me (In, Ga, Al, Ti, Sn, Rh, Co, Cd, Ni, Zn, Mg, Ge, Si, Sb, Pt, P, Au) binary alloys was built and the mechanical and electrical properties of these alloys were predicted then.

Figure 3 (a) shows the predicted relationship between UTS and the key alloy factors M-E4 and V-E4, and Figures 3 (b) displays the predicted changes in EC with the key alloy factors V-S6 and V-S10. It can be seen that decrease in the values of M-E4 and V-E4, and V-S6 and V-S10 result in increases of both UTS and EC of the alloys. According to equations (1) and (2), the key alloy factors M-E4, V-E4, V-S6 and V-S10 of the Cu-Me, where Me represents an alloying element, binary alloy have a positive correlation with the absolute electronegativity $E4_{Me}$ of Me, the absolute electronegativity difference $|E4_{Me}-E4_{Cu}|$, the core electron distance difference $|S6_{Me}-S6_{Cu}|$, and the atomic radius difference $|S10_{Me}-S10_{Cu}|$ between Me and Cu, respectively. Therefore, decreasing the values of $E4_{Me}$ and $|E4_{Me}-E4_{Cu}|$ increases UTS of a Cu-Me alloy, while reducing the values of $|S6_{Me}-S6_{Cu}|$ and $|S10_{Me}-S10_{Cu}|$ enhances the EC of a Cu-Me alloy. The $E4_{Me}$, $|E4_{Me}-E4_{Cu}|$, $|S6_{Me}-S6_{Cu}|$ and $|S10_{Me}-S10_{Cu}|$ of various solid solution elements are shown in Figures 3 (c) and (d).

We calculated the key alloy factors M-E4, V-E4, V-S6 and V-S10 based on the $E4_{Me}$, $|E4_{Me}-E4_{Cu}|$, $|S6_{Me}-S6_{Cu}|$ and $|S10_{Me}-S10_{Cu}|$, then predicted UTS and EC of the binary Cu-0.5wt%Me shown in Figures 3 (e) and (f). We would like to point out that In, Ga, Al, Ti, Ge, Pt, Au etc. have not previously been reported as solid solution strengthening elements of conductive copper alloys, and hence we call them new solid solution strengthening elements. Figures 3 (e) and (f) demonstrate that Cu-0.5wt%In alloy has the highest predicted UTS of 322 MPa, and the highest EC which is 75.5% IACS (The International Annealed Copper Standard).

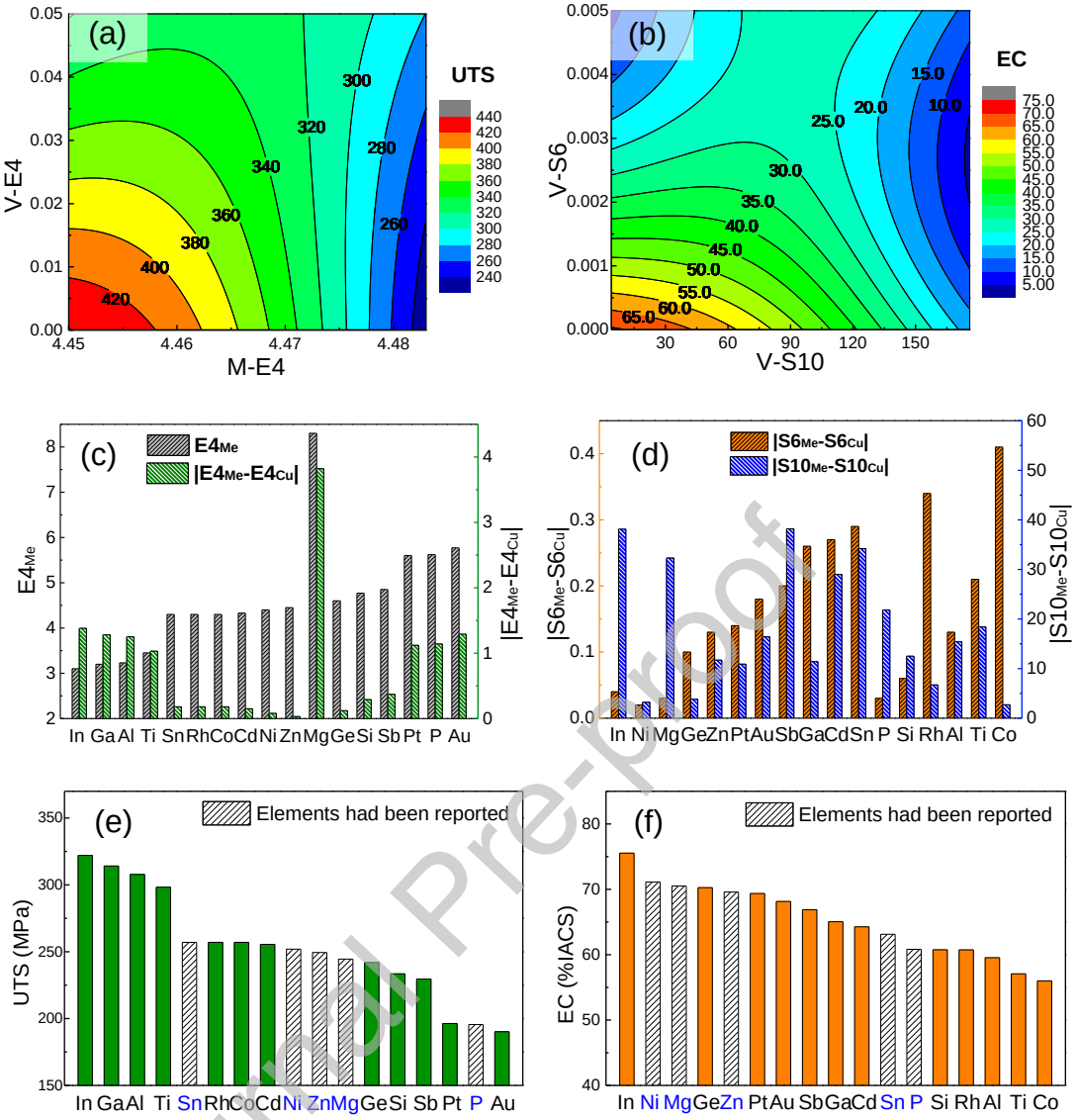


Figure 3. (a) and (b) show the relationships between alloy properties and key alloy factors M-E4, V-E4, V-S6 and V-S10 based on the UTS and EC prediction models; (c) and (d) depict the key element features of solid solution elements; (e) and (f) plot the predicted properties of Cu-0.5wt%Me binary alloys (Me=In, Ga, Al, Ti, …).

3.3. New alloys design and experiments

Based on Figures 3 (e) and (f), we designed four new solid solution strengthened

Cu-In alloys with low alloying element content. Three of them are binary solid solution copper alloys Cu-0.26In, Cu-0.42In, and Cu-0.68In, and the other is a ternary alloy Cu-0.30Sn-0.27In.

Table 3 lists the predicted and corresponding experimentally measured properties of the four new alloys at the OS025 state with UTSs ranging from 262 MPa to 298 MPa, and the ECs from 68.5% IACS to 90.4% IACS. The predicted UTSs are higher than the measured values with an average prediction error of +10.5%, the predicted ECs are lower than the measured values with an average prediction error of -8.2%, demonstrating that the machine learning models using key element features yield satisfactory predictions of the new alloy properties.

Table 3. Properties of new SSCAs containing element In (OS025 state)

Alloys (wt%)	UTS (MPa)		EC (%IACSA)		Prediction error of UTS	Prediction error of EC
	Predicted	Experimental	Predicted	Experimental		
Cu-0.26In	289	262±5	80.1	90.4±0.3	+10.3%	-11.4%
Cu-0.42In	313	277±2	77.8	86.7±0.5	+13.0%	-10.3%
Cu-0.68In	328	290±5	72.0	78.8±0.2	+13.1%	-8.6%
Cu-0.30Sn-0.27In	315	298±4	66.9	68.5±0.3	+5.7%	-2.3%

* A positive error indicates that the predicted value is higher than the measured value, and a negative error indicates that the predicted value is lower than the measured value.

We compared the measured properties of the four new In-containing Cu alloys with those of the sample alloys from the published handbook [29] and present the comparison in Figure 4. The red five-pointed stars are the measured UTSs and ECs of the four new alloys, and the blue dots represent those of the sample alloys. Figure 4

clearly demonstrate that the combined tensile strengths and electric conductivities of new alloys are significantly higher than the existing SSCAs.

We can further improve mechanical strengths by plastic deformation. For example, the UTS and EC of the Cu-0.26In alloy after cold rolling at 80% deformation are 438 ± 6 MPa and 89.6% ± 0.9% IACS, respectively, which are very close to those of Cu-0.28Ag alloy (453 MPa and 91.5% IACS) currently in use in electric railway contact wires with the same cold deformation state [37]. However, we should point out that the cost for producing Cu-0.26In alloys is expected to be significantly lower Cu-0.28Ag alloys since the price of In is only one third of that of Ag.

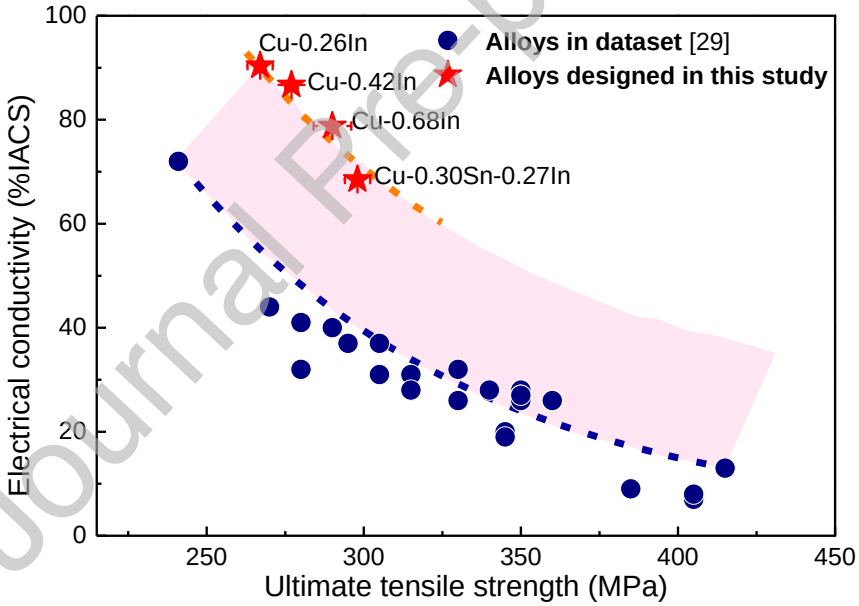


Fig. 4 Property comparison of SSCAs. All alloys are at OS025 state.

4. Discussion

4.1. Relationship between key element features and properties

We proposed a compositional alloy design approach by identifying the key element

features associated with the properties of interest. Application of this approach to SSCA demonstrate that the key elemental feature for UTS of an alloy is the absolute electronegativity, and the key element features for EC are the atomic radius and core electron distance.

The absolute electronegativity can be estimated from the first ionization energy and electron affinity [38].

$$x = k\,(E_i + E_{ea}) \qquad (5)$$

where E_i is first ionization energy which the amount of energy required to remove an electron from an atom to vacuum, E_{ea} is electron affinity energy which indicates the amount of energy released when an atom acquires an electron, and k is a constant which is often taken as 0.5. Therefore, the absolute electronegativity of an element can be used to characterize the ability of atoms to acquire or donate electrons. For example, a high value indicates a high ability of an atom to acquire an electron, and a low value implies the high ability of atom to donate an electron.

The mechanical strength of a metal is largely determined by the bond energy among atoms [39]. It is well known that metallic bonds result from the interaction between atomic cores and their free electrons, and thus the magnitude of free electron density of a metallic alloy is one of the important factors that control the strength of a metal [40]. The contribution of each atom to the free electron density is mainly determined by an atom's ability to donate electrons. The greater the ability of an atom to donate electrons, the larger the contribution of the atom to the free electron density, and

thus the stronger the metallic bonding. Therefore, one can use the absolute electronegativity of an element to estimate the strength of metal bonds; The smaller the absolute electronegativity of the alloying element, and hence the smaller the absolute electronegativity factor M-E4 of the corresponding alloy, the greater the contribution of the alloying element to the free electron density, the stronger the metal or alloy.

However, the absolute electronegativity of each element in binary or multicomponent alloys varies. A large variance in absolute electronegativity factor V-E4 indicates that the absolute electronegativities of each element in the alloy has a large variation. It is reasonably to speculate that an element with a larger absolute electronegativity contributes to the weakening of the alloy or lowering of the alloy strength. Therefore, it is not difficult to understand that the absolute electronegativity variance factor V-E4 of the alloy is also one of the most important factors affecting UTS.

The electric carriers in metal crystals are free electrons, and the scattering of electrons by the atomic lattice is the key factor that determine the EC of a metal [41,42]. The degree of electron scattering is characterized by the average scattering time τ, i.e., the moving time of an electron between two consecutive scatterings. The longer the average scattering time, the higher the conductivity σ of a metal, according to the following equation (6),

$$\sigma = ne^2\tau / m_e \qquad (6)$$

where n is the free electron density, m_e is the effective electron mass, and e is

elementary charge.

Free electrons originate from the valence electrons of atoms, and those inner electrons closer to the atomic cores are core electrons. The core electron distance S6 largely determines the distances between atomic cores in a crystal, and the variance factor V-S6 represents the variation in spacings among atomic cores. Therefore, a smaller variation among the atomic spacings, characterized by V-S6, implies a longer relaxation time for the free electrons and thus higher EC of the alloy.

A large V-S10 implies a large variation in atomic radius of elements in an alloy, and thus a high distortion in a solid solution, and consequently a short electron means free path, and a low EC. Therefore, both the atomic radius variance V-S10 and the core electron variance V-S6 of the alloy are important factors for the EC of an alloy.

According to the values of $E4_{Me}$, $|E4_{Me}-E4_{Cu}|$, $|S6_{Me}-S6_{Cu}|$, $|S10_{Me}-S10_{Cu}|$ of the different alloying elements in Figures 3 (c) and (d), and the effects of alloying elements on the alloy properties in Figures 3 (e) and (f), we make the following two observations. First, M-E4 has a greater impact on the UTS than V-E4. For example, although the $|E4_{Me}-E4_{Cu}|$ of element In is large resulting in a large V-E4, its E4 is the smallest among all solid solution alloying elements, leading to the smallest M-E4 and the largest strengthening effect for the corresponding alloy. On the contrary, $|E4_{Me}-E4_{Cu}|$ of element Co is relatively small, but its E4 is larger, and thus the strengthening effect of Co is significantly smaller than that of In. Second, V-S6 has a greater impact on the EC than V-S10. For example, the larger $|S10_{Me}-S10_{Cu}|$ of element In leads to a larger V-S10, but

its $|S6_{Me}-S6_{Cu}|$ is very small. As a result, the corresponding V-S6 is very small, and the EC of Cu-In alloy is highest. On the contrary, although the $|S10_{Me}-S10_{Cu}|$ of element Co is the smallest, the EC of Cu-Co alloy is the lowest because its $|S6_{Me}-S6_{Cu}|$ is the largest.

4.2. Reliability and general applicability of the proposed approach

The feature screening and machine learning in this study are based on a rather small dataset of 27 SSCAs [29]. Our results indicate that a small amount of representative sample data can be employed to effectively screen the key features associated with a given set of properties if the sample data are of high quality.

To verify the reliability of the proposed approach and demonstrate its general applicability to other solid solution strengthened conductive alloys, we collected 22 sets of data on solid solution strengthened conductive aluminum alloys at an annealed state from *ASM specialty handbook: aluminum and aluminum alloys* [43], and the results are listed in Table S4 of the Supporting Information. The key alloy factors M-E4, V-E4, V-S10, and V-S6 of each aluminum alloy were calculated based on the amount and elemental features and of each element. We then establish the UTS prediction model with M-E4 and V-E4 as inputs and the EC prediction model using V-S10 and V-S6 as inputs by employing SVR. The performances of the models were evaluated by examining the direct modeling prediction error and 10-fold cross validation error. As shown in Figure 5, the direct modeling prediction error and the 10-fold cross validation

error for UTS are 6.9%, and 9.0%, respectively while the direct modeling prediction error and the 10-fold cross validation error for EC are 4.0% and 8.9%, respectively. The results confirm that the absolute electronegativity E4 is also the key element feature with the UTS, and the core electron distance S6 and atomic radius S10 are also the key element features associated with EC of the solid solution strengthened conductive aluminum alloys.

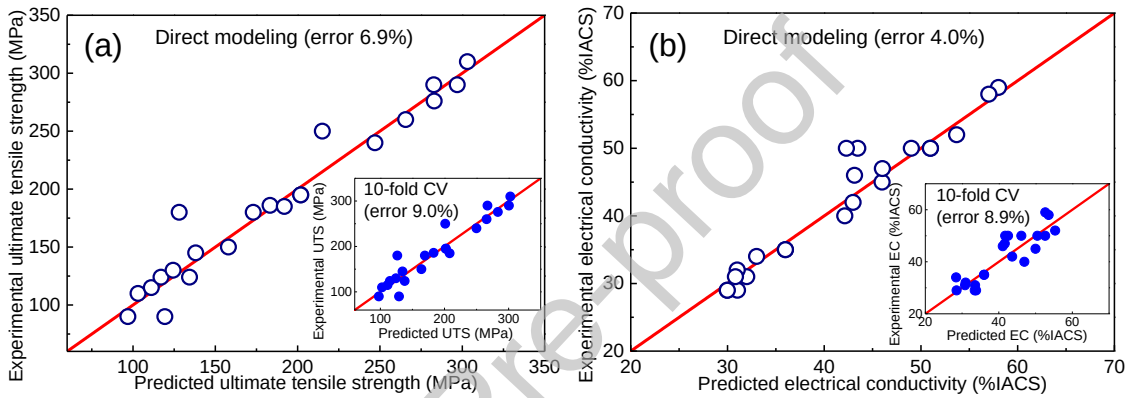


Figure 5. Direct modeling errors and 10-fold cross validation (10-fold CV) errors of aluminum alloy property prediction model. (a) UTS model, (b) EC model

5. Conclusion

A machine learning approach for screening alloying elements to improve the properties of solid solution strengthened conductive alloys is established. It identifies the absolute electronegativity of an element as the key feature associated with the ultimate tensile strength of an alloy and the core electron distance and atomic radius of an element as the two key features associated with the electric conductivity of an alloy. Based on the key features and by searching the periodic table for possible solid solution elements, a new alloying element In was predicted to significantly enhance both the

ultimate tensile strength and electric conductivity of solid solution strengthened conductive alloys. This prediction was experimentally validated by preparing four new copper alloys with In content of less than 0.7 wt%. The ultimate tensile strength and the electric conductivity of these alloys at an annealed state range from 262 MPa to 298 MPa and from 68.5% IACS to 90.4% IACS, comparable to those of much more expensive, medium-strength and high-conductivity Cu-0.28Ag alloys currently being used in electric railway contact wires. The identified three key element features associated with the ultimate tensile strength and electric conductivity of Cu alloys are expected to be generally applicable to identifying alloying elements to improve the properties of other solid solution strengthened conductive alloys.

Acknowledgement

This work was supported by the National Key Research and Development Program of China (No. 2016YFB0301300), and Beijing Municipal Science and Technology Commission (No. Z191100007219002), and Beijing Nova Programs (No. Z191100001119125), and the Funds for Creative Research Groups of China (No. 51921001).

Appendix A. Supplementary information

Supplementary information to this article can be found online or from the author.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Z. Li, Z. Xiao, Y. Jiang, Q. Lei, J. Xie, Composition design, phase transition and fabrication of copper alloys with high strength and electrical conductivity, Chin. J. of Nonferrous Met. 29 (2019) 2009-2049.

[2] S. Gorsse, B. Ouvrard, M. Gouné, A. Poulon-Quintin, Microstructural design of new high conductivity–high strength Cu-based alloy, J. Alloys Compd. 633 (2015) 42-47.

[3] K. Maki, Y. Ito, H. Matsunaga, H. Mori, Solid-solution copper alloys with high strength and high electrical conductivity, Scr. Mater. 68 (2013) 777-780.

[4] B. Sanchez-Lengeling, A. Aspuru-Guzik, Inverse molecular design using machine learning: Generative models for matter engineering, Science 361 (2018) 360-365.

[5] L. Himanen, A. Geurts, A. S. Foster, P. Rinke, Data‐Driven Materials Science: Status, Challenges, and Perspectives, Adv. Sci. 6 (2019) 1900808.

[6] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, Nature 559 (2018) 547.

[7] D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue, T. Lookman, Accelerated search for materials with targeted properties by adaptive design, Nat. Commun. 7 (2016) 11241.

[8] D. Xue, D. Xue, R. Yuan, Y. Zhou, P. V. Balachandran, X. Ding, J. Sun, T. Lookman, An informatics approach to transformation temperatures of NiTi-based shape memory alloys, Acta Mater. 125 (2017) 532-541.

[9] P. Raccuglia, K. C. Elbert, P. D. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, A. J. Norquist, Machine-learning-assisted materials discovery using failed experiments, Nature 533 (2016) 73.

[10] C. Wang, H. Fu, L. Jiang, D. Xue, J. Xie, A property-oriented design strategy for high performance copper alloys via machine learning, npj Comput. Mater. 5 (2019) 87.

[11] J. Gao, Y. Liu, Y. Wang, X. Hu, W. Yan, X. Ke, L. Zhong, Y. He, X. Ren, Designing high dielectric permittivity material in barium titanate, J. Phys. Chem. C 121 (2017) 13106-13113.

[12] R. Yuan, Z. Liu, P. V. Balachandran, D. Xue, Y. Zhou, X. Ding, J. Sun, D. Xue, T. Lookman, Accelerated discovery of large electrostrains in BaTiO3 - based piezoelectrics using active learning, Adv. Mater. 30 (2018) 1702884.

[13] V. Stanev, C. Oses, A. G. Kusne, E. Rodriguez, J. Paglione, S. Curtarolo, I. Takeuchi, Machine learning modeling of superconducting critical temperature, npj Comput. Mater. 4 (2018) 29.

[14] W. Xu, P.E.J. Rivera Díaz del Castillo, S. van der Zwaag, A combined optimization

of alloy composition and aging temperature in designing new UHS precipitation hardenable stainless steels, Comput. Mater. Sci. 45 (2009) 467-473.

[15] C. Wen, Y. Zhang, C. Wang, D. Xue, Y. Bai, S. Antonov, L. Dai, T. Lookman, Y. Su, Machine learning assisted design of high entropy alloys with desired property, Acta Mater. 170 (2019) 109-117.

[16] Y. Zeng, Q. Li, K. Bai, Prediction of interstitial diffusion activation energies of nitrogen, oxygen, boron and carbon in bcc, fcc, and hcp metals using machine learning, Comput. Mater. Sci. 144 (2018) 232-247.

[17] H. Wu, A. Lorenson, B. Anderson, L. Witteman, H. Wu, B. Meredig, D. Morgan, Robust FCC solute diffusion predictions from ab-initio machine learning methods, Comput. Mater. Sci. 134 (2017) 160-165.

[18] Z. Deng, H. Yin, X. Jiang, C. Zhang, K. Zhang, T. Zhang, B. Xu, Q. Zheng, X. Qu, Machine leaning aided study of sintered density in Cu-Al alloy, Comput. Mater. Sci. 155 (2018) 48-54.

[19] X. Jiang, H. Q. Yin, C. Zhang, R. J. Zhang, K. Q. Zhang, Z. H. Deng, G. Liu, X. Qu, An materials informatics approach to Ni-based single crystal superalloys lattice misfit prediction, Comput. Mater. Sci. 143 (2018) 295-300.

[20] Y. T. Sun, H. Y. Bai, M. Z. Li, W. H. Wang, Machine learning approach for prediction and understanding of glass-forming ability, J. Phys. Chem. Lett. 8 (2017) 3434-3439.

[21] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine

learning framework for predicting properties of inorganic materials, npj Comput. Mater. 2 (2016) 16028.

[22] A. O. Furmanchuk, J. E. Saal, J. W. Doak, G. B. Olson, A. Choudhary, A. Agrawal, Prediction of seebeck coefficient for compounds without restriction to fixed stoichiometry: A machine learning approach, J. Comput. Chem. 39 (2018) 191-202.

[23] T. D. Huan, R. Batra, J. Chapman, S. Krishnan, L. Chen, R. Ramprasad, A universal strategy for the creation of machine learning-based atomistic force fields, npj Comput. Mater. 3 (2017) 37.

[24] J. Wang, X. Yang, Z. Zeng, X. Zhang, X. Zhao, Z. Wang, New methods for prediction of elastic constants based on density functional theory combined with machine learning, Comput. Mater. Sci. 138 (2017) 135-148.

[25] M. S. Ozerdem, S. Kolukisa, Artificial neural network approach to predict the mechanical properties of Cu–Sn–Pb–Zn–Ni cast alloys, Mater. Des. 30 (2009) 764-769.

[26] B. Wang, W. Zhao, Y. Du, G. Zhang, Y. Yang, Prediction of fatigue stress concentration factor using extreme learning machine, Comput. Mater. Sci. 125 (2016) 136-145.

[27] A. Rovinelli, M. D. Sangid, H. Proudhon, W. Ludwig, Using machine learning and a data-driven approach to identify the small fatigue crack driving force in polycrystalline materials, npj Comput. Mater. 4 (2018) 35.

[28] A. Agrawal, A. Choudhary, An online tool for predicting fatigue strength of steel alloys based on ensemble data mining, Int. J. Fatigue 113 (2018) 389-400.

[29] J. R. Davis, ASM Specialty Handbook: Copper and Copper Alloys, ASM Int., Ohio, 2001.

[30] P. Villars, Pearson's Handbook Desk Edition Crystallographic Data for Intermetallic Phases, ASM Int., Ohio, 1997.

[31] P. Villars, K. Brandenburg, M. Berndt, S. LeClair, A. Jackson, Y. H. Pao, B. Igelnik, M. Oxley, B. Bakshi, P. Chen, S. Iwata, Binary, ternary and quaternary compound former/nonformer prediction via Mendeleev number, J. Alloys Compd. 317 (2001) 26-38.

[32] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn. 46 (2002) 389-422.

[33] L. Wang, Y. Wang, Q. Chang, Feature selection methods for big data bioinformatics: A survey from the search perspective, Methods 111 (2016) 21-31.

[34] S. An, W. Liu, S. Venkatesh, Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression, Pattern Recognit. 40 (2007) 2154-2162.

[35] K. Pearson, Note on regression and inheritance in the case of two parents, Proc. R. Soc. London 58 (1895) 240-242.

[36] H. Landolt, R. Börnstein, Phase Equilibria, Crystallographic and Thermodynamic Data of Binary Alloys, Springer-Verlag, 1993.

[37] J. Ma, D. Wang, Y. Zheng, F. Liu, J. Shen, L. Peng, Work hardening and recrystallization temperature of Cu-Ag alloy, Spec. Cast. Nonferrous Alloys 35 (2015)

1006-1008.

[38] R. S. Mulliken, A new electroaffinity scale; together with data on valence states and on valence ionization potentials and electron affinities, J. Chem. Phys. 2 (1934) 782-793.

[39] R. G. Pearson, Absolute electronegativity and absolute hardness of Lewis acids and bases, J. Am. Chem. Soc. 107 (1985) 6801-6806.

[40] M. S. Daw, S. M. Foiles, M. I. Baskes, The embedded-atom method: a review of theory and applications, Mater. Sci. Rep. 9 (1993) 251-310.

[41] D. Drechsel, M. M. Giannini, Electron scattering off nuclei, Rep. Prog. Phys. 52 (1989) 1083.

[42] B. Frois, I. Sick, Modern Topics in Electron Scattering, World Scientific, 1991.

[43] J. R. Davis, ASM Specialty Handbook: Aluminum and Aluminum Alloys, ASM Int., Ohio, 1993.

Table 1. Alloying elements in dataset and their content ranges

Alloying element	Sn	Zn	Ni	Mg	P
Content (wt%)	0~11.4	0~36.4	0~15	0~0.4	0~0.2

Table 2. Main alloy factors obtained by recursive elimination

Main alloy factors affecting UTS		Main alloy factors affecting EC	
Factors	Description	Factors	Description
M-S11	Lattice constant *a*	M-S11	Lattice constant *a*
V-E4	Absolute electronegativity variance	V-E4	Absolute electronegativity variance
V-S6	Core electron distance variance	V-S6	Core electron distance variance
V-S10	Atomic radius variance	V-S10	Atomic radius variance
M-E4	Absolute electronegativity	M-S10	Atomic radius
M-S9	Atomic environment number	M-E2	Pauling electronegativity
M-N2	Mendeleev Number	M-E6	Second energy ionization
V-E11	Nuclear charge effective variance	V-S11	Lattice constant *a* variance

Table 3. Properties of new SSCAs containing element In (OS025 state)

Alloys (wt%)	UTS (MPa)		EC (%IACSA)		Prediction error of UTS	Prediction error of EC
	Predicted	Experimental	Predicted	Experimental		
Cu-0.26In	289	262±5	80.1	90.4±0.3	+10.3%	-11.4%
Cu-0.42In	313	277±2	77.8	86.7±0.5	+13.0%	-10.3%
Cu-0.68In	328	290±5	72.0	78.8±0.2	+13.1%	-8.6%
Cu-0.30Sn-0.27In	315	298±4	66.9	68.5±0.3	+5.7%	-2.3%

* A positive error indicates that the predicted value is higher than the measured value, and a negative error indicates that the predicted value is lower than the measured value.

Figure captions

Fig.1 Alloy composition rational design strategy based on element features screening

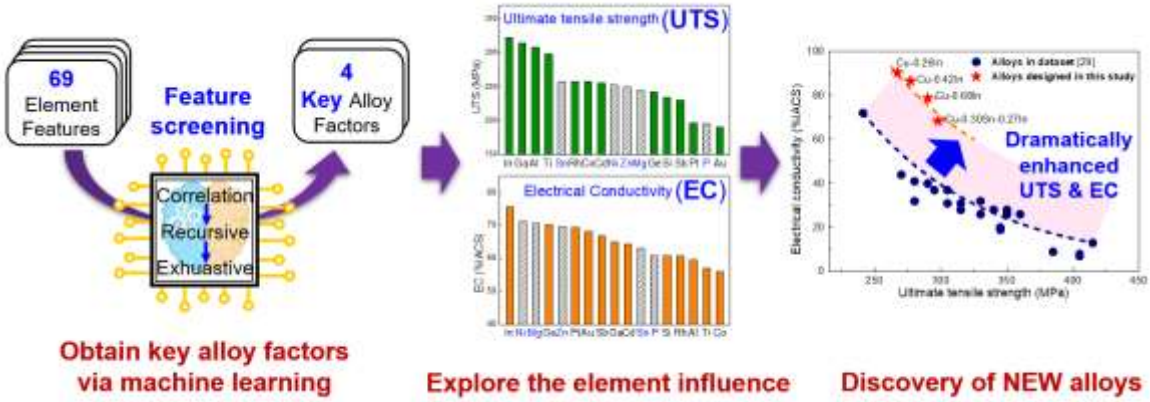
Fig. 2 Results of alloy factor screening and modeling: (a) (b) Correlation screening. The flow chart (a) shows the correlation screening process, the pie chart (b) shows the Pearson correlation coefficients of alloy factors. Blue and red colors indicate positive and negative correlations, respectively, and the filled fraction of the circle in each of the pie charts corresponds to the absolute value of correlation coefficient. We obtained 29 alloy factors each associated with UTS and EC after correlation screening. **(c) (d) Recursive elimination.** Each dot indicates a combination of alloy factors and the solid dots are the combinations with minimum model error under the number of alloy factors after elimination. We have 8 main alloy factors affecting UTS and 8 main alloy factors affecting EC after recursive elimination. "M-" represents feature mean factor, and "V-" represents feature variance factor. **(e) (f) Exhaustive screening.** The dots indicate all combinations of 8 main alloy factors, and we have 2 key alloy factors affecting UTS and 2 key alloy factors affecting EC after exhaustive screening. **(g) (h) The performance of the UTS model and the EC model with the key alloy factors as inputs.**

Fig. 3 (a) and **(b)** show the relationships between alloy properties and key alloy factors M-E4, V-E4, V-S6 and V-S10 based on the UTS and EC prediction models; **(c)** and **(d)** depict the key element features of solid solution elements; **(e)** and **(f)** plot the predicted properties of Cu-0.5wt%Me binary alloys (Me=In, Ga, Al, Ti, …).

Fig. 4 Property comparison of SSCAs. All alloys are at OS025 state.

Fig. 5 Direct modeling errors and 10-fold cross validation (10-fold CV) errors of aluminum alloy property prediction model. (a) UTS model, (b) EC model

Graphical abstract



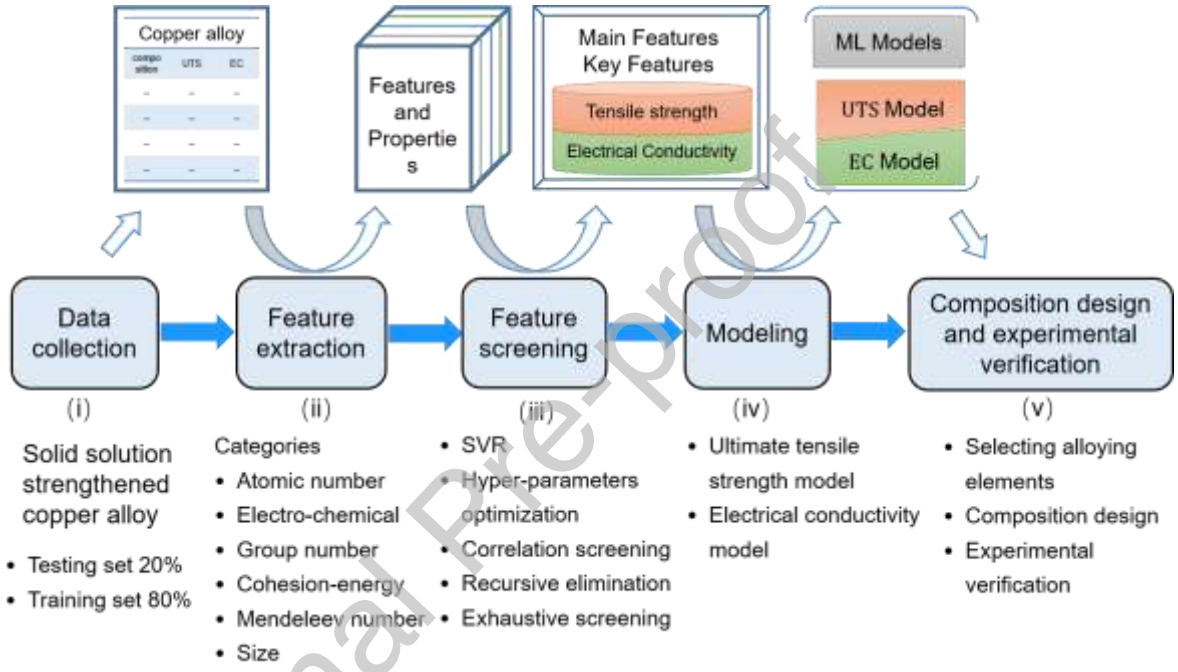


Fig.1 Alloy composition rational design strategy based on element features screening

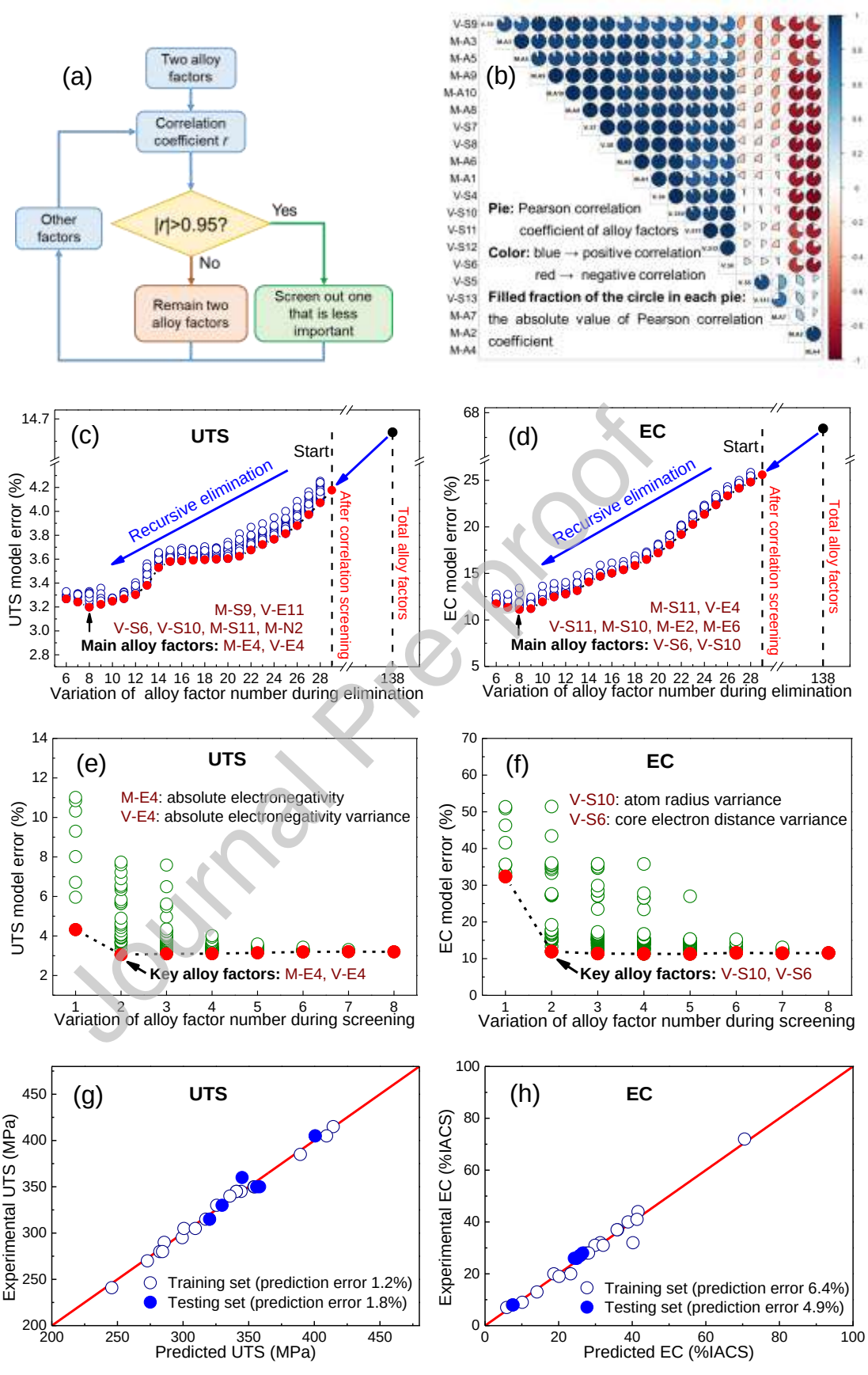


Fig. 2 Results of alloy factor screening and modeling: (a) (b) Correlation screening. The flow chart (a) shows the correlation screening process, the pie chart (b) shows the Pearson correlation coefficients of alloy factors. Blue and red colors indicate positive and negative correlations, respectively, and the filled fraction of the circle in each of the pie charts corresponds to the absolute value of correlation coefficient. We obtained 29 alloy factors each associated with UTS and EC after correlation screening. **(c) (d) Recursive elimination.** Each dot indicates a combination of alloy factors and the solid dots are the combinations with minimum model error under the number of alloy factors after elimination. We have 8 main alloy factors affecting UTS and 8 main alloy factors affecting EC after recursive elimination. "M-" represents feature mean factor, and "V-" represents feature variance factor. **(e) (f) Exhaustive screening.** The dots indicate all combinations of 8 main alloy factors, and we have 2 key alloy factors affecting UTS and 2 key alloy factors affecting EC after exhaustive screening. **(g) (h) The performance of the UTS model and the EC model with the key alloy factors as inputs.**

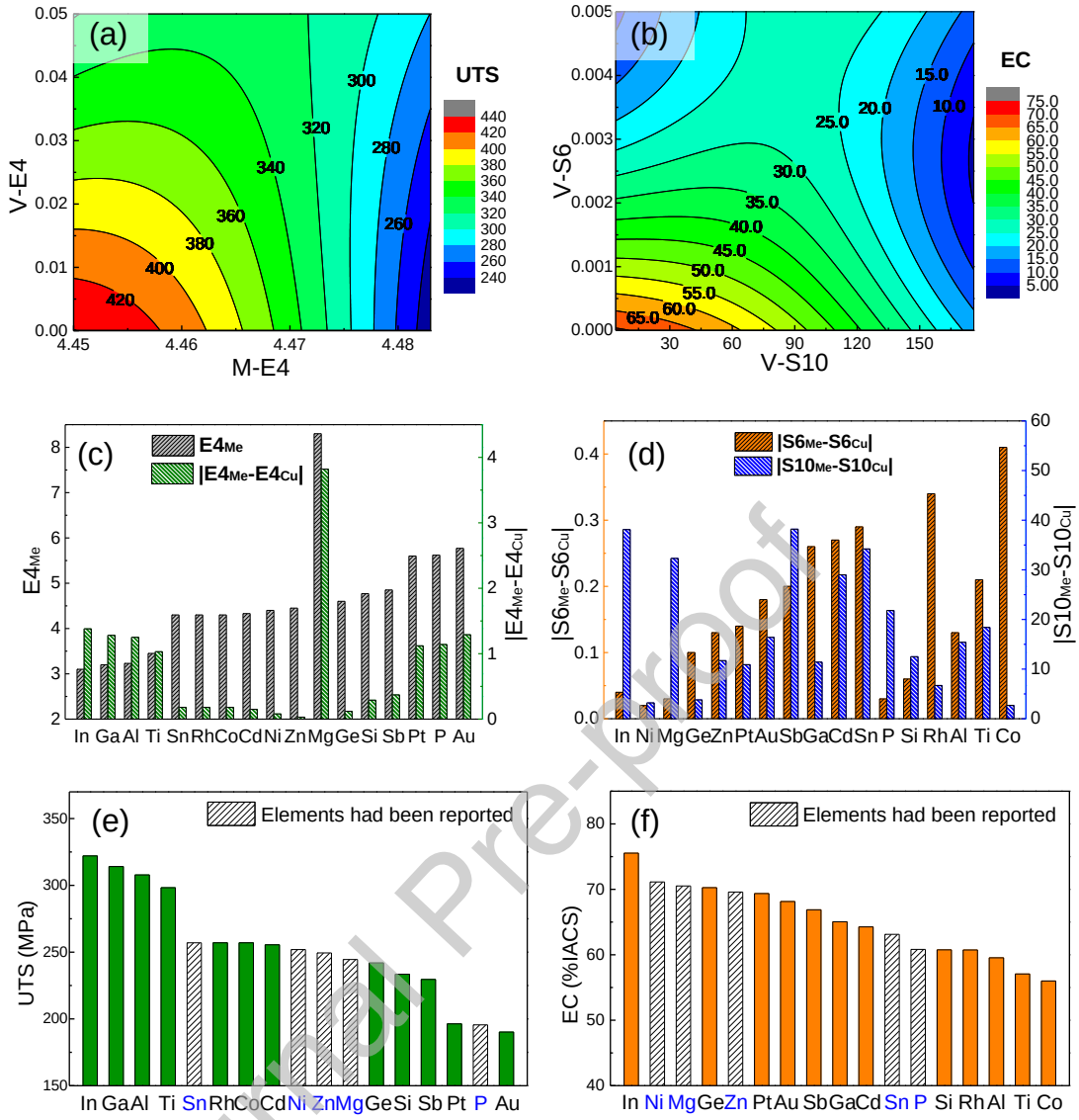


Fig. 3 (a) and **(b)** show the relationships between alloy properties and key alloy factors M-E4, V-E4, V-S6 and V-S10 based on the UTS and EC prediction models; **(c)** and **(d)** depict the key element features of solid solution elements; **(e)** and **(f)** plot the predicted properties of Cu-0.5wt%Me binary alloys (Me=In, Ga, Al, Ti, …).

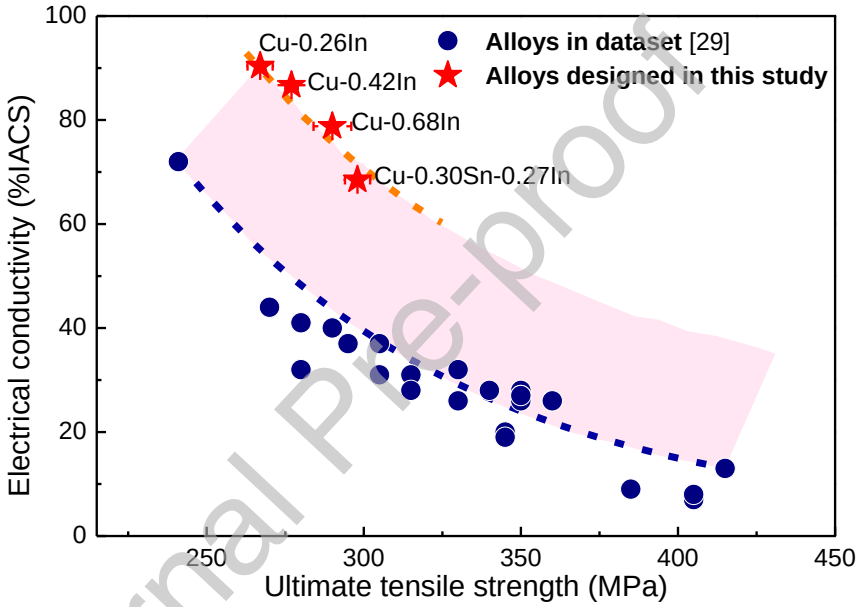


Fig. 4 Property comparison of SSCAs. All alloys are at OS025 state.

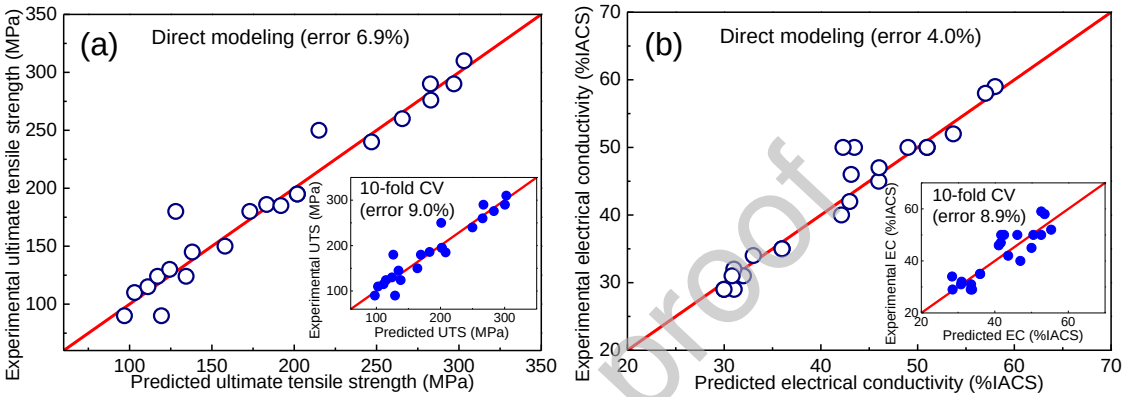


Fig. 5 Direct modeling errors and 10-fold cross validation (10-fold CV) errors of aluminum alloy property prediction model. (a) UTS model, (b) EC model