



UNIVERSITY OF
LEICESTER

**Prediction of material defects and phase formation using
deep learning and transfer learning**

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Shuo Feng

School of Engineering

University of Leicester

United Kingdom

September 2020

Abstract

Machine learning has been successfully employed in computer vision, speech processing, and natural language processing. However, when machine learning is applied to materials study, many challenges remain, and they include small datasets, tricky manual feature (descriptor) engineering, isolated models and poor interpretability. In this project, possible solutions to these challenges have been explored using deep learning and transfer learning.

For the challenge of small datasets, fully connected deep neural network and tree-based models were used to predict solidification cracking susceptibility of stainless steels with a dataset of 487 samples. It is found that deep neural network with pre-training and fine-tuning improves prediction accuracy, and tree-based models reveal the relative importance of input variables.

To overcome the challenge of tricky manual feature engineering in predicting phase formation in inorganic substances and compounds properties, I proposed a general and transferable deep learning framework as follows: (1) mapping raw data to pseudo images with periodic table structure, (2) automatically extracting features through convolutional neural networks, (3) transferring knowledge by sharing features extractors between models. The proposed deep learning models outperformed previous models in predicting glass-forming ability using a medium dataset of 16k samples and compounds properties using a big dataset of 228k samples. The developed transfer learning model for multi-principal element alloys can distinguish five phases (BCC, FCC, HCP, amorphous, mixture) with high scores (0.94) in a small dataset of 345 samples. The transfer learning model for phase prototypes can discriminate 170 phase prototypes with an accuracy of 0.9 in a dataset of 17k inorganic substances. Periodic table knowledge embedded in data representations and knowledge shared between models is beneficial for tasks with small datasets.

Acknowledgment

Firstly, I would like to thank Prof Hongbiao Dong for supervising and supporting me throughout the PhD. I would also like to thank Prof Huiyu Zhou for supervising and encouraging me on the way of machine learning. Special thanks go to Prof Simon Gill and Charlotte Ratcliffe for helping me in the past four years.

I would also like to thank everyone, especially my friends in IMPACT CDT, Prof Hongbiao Dong's group, and Prof Huiyu Zhou's group, who helped me in one way or another during the PhD.

Last but not least, I wish to acknowledge EPSRC CDT (Grant No: EP/L016206/1) in Innovative Metal Processing (IMPACT) for providing a PhD studentship for this study.

To Emma, Yun, and Jie

Publications

- [1] **Feng, S.**, Zhou, H.Y., & Dong, H.B. (2021). Application of deep transfer learning to predicting crystal structures of inorganic substances, **Computational Materials Science**, 195, 110476. <https://doi.org/10.1016/j.commatsci.2021.110476>.
- [2] **Feng, S.**, Fu, H.D., Zhou, H.Y., Wu, Y., Lu, Z.P., & Dong, H.B. (2021). A general and transferable deep learning framework for predicting phase formation in materials, **npj Computational Materials**, 7, 1–10. <https://doi.org/10.1038/s41524-020-00488-z>.
- [3] **Feng, S.**, & Dong, H.B. (2020). Predicting solidification cracking susceptibility of stainless steels using machine learning. **IOP Conference Series: Materials Science and Engineering**, 861, 012073. <https://doi.org/10.1088/1757-899X/861/1/012073>
- [4] **Feng, S.**, Zhou, H.Y., & Dong, H.B. (2019). Using deep neural network with small dataset to predict material defects. **Materials and Design**, 162, 300–310. <https://doi.org/10.1016/j.matdes.2018.11.060>
- [5] Dong, H.B. & **Feng, S.** (2019). Chapter 17 - Applications of artificial intelligence in material design and processing, in Yike Guo (editor) **Artificial intelligence and future social development** (in Chinese), Scientific and technical documentation press, Beijing, 2019, 315-336.
- [6] Dong, Z. H., Sergeev, D., Kobertz, D., D’Souza, N., **Feng, S.**, Müller, M., & Dong, H. B. (2019). Vaporization of Ni, Al and Cr in Ni-Base alloys and its influence on surface defect formation during manufacturing of single-crystal components. **Metallurgical and Materials Transactions A**. <https://doi.org/10.1007/s11661-019-05498-1>
- [7] Aucott, L., Dong, H.B., Mirihanage, W., Atwood, R., Kidess, A., Gao, S., Wen, S., Marsden, J., **Feng, S.**, Tong, M., Connolley, T., Drakopoulos, M., Kleijn, C. R., Richardson, I. M., Browne, D. J., Mathiesen, R. H., & Atkinson, H. V. (2018). Revealing internal flow behavior in arc welding and additive manufacturing of metals. **Nature Communications**, 9(1). <https://doi.org/10.1038/s41467-018-07900-9>

Abbreviations

AI: artificial intelligence

AMR: amorphous ribbon

BCC: body-centered cubic

BMG: bulk metallic glass

BTR: brittle temperature range

CR: crystalline

CSS: concentrated solid solution alloy

CNN: convolutional neural network

DBN: deep belief network

DFT: density functional theory

DL: deep learning

DNN: deep neural network

Ef: energy of formation

FC: fully connected

FCC: face-centered cubic

GFA: glass-forming ability

HCP: hexagonal closest packed

HEA: high-entropy alloy

LDA: linear discriminant analysis

LVT: longitudinal varestraint test

ML: machine learning

MPEA: multi-principal element alloy

MCL: maximum crack length

OQMD: open quantum materials database

PCA: principal component analysis

PTR: periodic table representation

RBM: restricted Boltzmann machine

RF: random forest

ROC: receiver operating characteristic

SAE: stacked auto-encoder

SCS: solidification cracking susceptibility

SNN: shallow neural network

SVM: support vector machine

TCL: total crack length

TVT: transverse varestraint test

V: specific volume

Table list

Table 2.1 Challenges in machine learning for materials research.....	15
Table 4.1 The best testing accuracies of five models used in prediction.	43
Table 4.2 The specifications (wt%) of 304 and 310S stainless steel and the testing parameters used in prediction.....	53
Table 4.3 Compositions (wt%) of stainless steels used in prediction (varied Ni and Cr).....	55
Table 4.4 Compositions (wt%) of stainless steels used in prediction (varied P and S).....	57
Table 4.5 Compositions (wt%) of stainless steels used in prediction (varied C and N).	58
Table 5.1 The statistics of labels in our dataset. CR, AMR, and BMG denote crystalline, amorphous ribbon, and bulk metallic glasses, respectively.....	65
Table 5.2 Matrix for transforming original labels of ternary classification to new labels of binary classification. CR and AM denote crystalline state, amorphous state, respectively.	66
Table 5.3 The list of manual features used in building machine learning models. Features used in SNN1 and SNN3 are denoted with *. Other features combinations were also tested at the beginning but did not show noticeable improvement in accuracy, so they were not used in final model.	71
Table 5.4 Hyperparameter details of CNN1 (atom table representation), CNN2 (randomized periodic table representation) and CNN3 (periodic table representation), CNN2 and CNN3 are assigned with the same hyperparameters.	74
Table 5.5 The summary of our VGG-like convolutional neural network used in the prediction of compounds' forming energy Ef and specific volume V.	75
Table 5.6 Comparison of average accuracy among different models under 10-fold cross-validation in predicting glass-forming ability	77
Table 5.7 Comparison of models' prediction accuracy on unseen alloy systems.	80
Table 5.8 Comparison of experimental results with the predictions by CNN1, CNN2, CNN3, SNN3, and SNN4 in Ir-Ni-Ta-(B), Mg-Cu-Yb alloys. The red indicates the wrong prediction, and the green indicates the right prediction.	82
Table 5.9 Comparison of experimentally results with the predictions by CNN1, CNN2, CNN3, SNN3, and SNN4 in sulfur-bearing alloys. The red indicates the wrong prediction, and the green indicates the	

right prediction.	83
Table 5.10 Comparison of experimental results with the predictions by CNN1, CNN2, CNN3, SNN3 and SNN4 in RE ₆ Fe ₇₂ B ₂₂ . The red indicates the wrong prediction, and the green indicates the right prediction. Superscript GT denotes this composition is in our dataset; superscript * denotes this composition is not in our dataset, but some other compositions of this alloy system are in our dataset.	85
.....	85
Table 5.11 Comparison of experimentally measured glass-forming ability with the predictions by CNN1, CNN2, CNN3, SNN3 and SNN4 in 18 binary outliers. The red indicates the wrong prediction, and the green indicates the right prediction. Superscript GT denotes this composition is in our dataset; superscript * denotes this composition is not in our dataset, but some other compositions of this alloy system are in our dataset.	85
.....	85
Table 6.1 Prototype names and sample numbers.....	99
Table 6.2 The scores of transfer learning on multi-principal element alloys dataset under 5-fold cross-validation with three data representations	103
Table 6.3 Confusion matrices of transfer learning models for predicting multi-principal element alloy phases on testing datasets under 5-fold cross-validation.....	104
Table 6.4 The performance metrics of transfer learning models for predicting multi-principal element alloy phases under 5-fold cross-validation	105

Figure list

Figure 1.1 The structure of this thesis.	2
Figure 2.1 Schematic diagram for a parametric machine learning system, Θ represents parameter vector.	4
Figure 2.2 Workflow for (a) conventional machine learning, (b) deep learning, (c) transfer learning.	4
Figure 2.3 Schematic diagrams for (a) best fit, (b) underfit, (c) overfit, and (d) training error/ testing error vs. model complexity ¹⁷	6
Figure 2.4 Typical algorithms in machine learning.....	8
Figure 2.5 Schematic diagram for a neuron, the basic unit of a neural network.....	9
Figure 2.6 Schematic diagrams of one hidden layer fully connected shallow neural network and j hidden layers fully connected deep neural network.....	10
Figure 2.7 Schematic diagram for the architecture differences between shallow network and deep network for the same task.	10
Figure 2.8 A schematic diagram for support vector machine.	13
Figure 2.9 (a) Single task architecture and (b-d) architectures for three transfer learning techniques based on conventional machine learning ⁶¹	14
Figure 2.10 Architecture for the transfer learning technique based on convolutional neural network. ...	14
Figure 3.1 Schematic diagram for the complex interaction between thermal, material, mechanical factors, which cause solidification cracking.	20
Figure 3.2 Stages of dendrites growth ⁹⁵	22
Figure 3.3 Ductility curve according to Senda's theory ⁸⁹	24
Figure 3.4 Schematic diagram for longitudinal varestraint test ¹⁰⁸	27
Figure 3.5 Comparison of physical-based methods and data-based methods for a materials science problem. The goal of material research is to determine the relationship between material property, composition, and processing. To achieve the goals, traditional material science completed it manually by structure analysis and computer modeling, modern data science can complete it automatically through mining the relationship between X and Y by machine learning algorithms.....	28
Figure 4.1 The workflow of predicting solidification cracking susceptibility using deep neural network (DNN).	35

Figure 4.2 Histograms of the 22 variables in the dataset (487 samples in total). Mean, minimum, maximum, std are also shown in the histograms.....	37
Figure 4.3 Scatter matrix of total crack length (TCL) and other 21 variables, the values of Pearson correlation coefficients Rs are also shown in the figures.....	40
Figure 4.4 Schematic diagram of deep neural network initiation by stacked auto-encoder.....	42
Figure 4.5 The regression analysis on the unseen testing dataset between experimental data (horizontal axis, T values) and neural network predictions (vertical axis, Y values). The black dotted line is Y=T. Pearson correlation coefficients Rs are also shown in the figures. (a) deep neural network's predictions, (b) shallow neural network's predictions, (c) support vector machine's predictions, (d) random forest's predictions, and (e) decision tree's predictions.....	44
Figure 4.6 (a) Shallow neural network's training and testing accuracies as a function of neuron number. (b) Neural networks' training and testing accuracies as a function of their depth. The support vector machine's accuracies are also shown in figure.	45
Figure 4.7 Regression analysis between total crack length predictions and brittle temperature range experimental results. (a) predictions of support vector machine, (b) predictions of shallow neural network and (c) predictions of deep neural network.	47
Figure 4.8 Predicted total crack length as functions of (a) Mn, (b) Ti, (c) N, (d) Si, the compositions used in prediction are shown in the figures.	48
Figure 4.9 (a) The optimal shallow neural network's structure; (b) the optimal deep neural network's structure.	50
Figure 4.10 The weights and biases tensors of the optimal 21-(21)-1 shallow neural network.	51
Figure 4.11 The weights and biases tensors of the optimal 21-(6-5-4-3)-1 fully connected deep neural network.	51
Figure 4.12 A decision tree for predicting solidification cracking susceptibility.	51
Figure 4.13 The feature importance of solidification cracking susceptibility derived from the random forest model.	52
Figure 4.14 Total crack lengths of 304 and 310S stainless steel vary with strain, Cr, Ni, P, S contents vary in the specification.....	53

Figure 4.15 The prediction of 8 stainless steels' total crack length (mm) dependence on Ni and Cr (their Ni and Cr contents are in AISI specification), compositions used in predictions are shown in Table 4.3, processing parameters are the same: Th=3.18 mm, I=100 A, U=12 V, Ve=4.23 mm/s, strain=3%.	55
Figure 4.16 The prediction of 8 stainless steels' total crack length (mm) dependence on P and S, compositions used in predictions are shown in Table 4.4, processing parameters are the same: Th=3.18 mm, I=100 A, U=12 V, Ve=4.23 mm/s, strain=3%.....	57
Figure 4.17 The prediction of stainless steels' total crack length (mm) dependence on C and N, compositions used in predictions are shown in Table 4.5. Processing parameters are the same: Th=3.18 mm, I=100 A, U=12 V, Ve=4.23 mm/s, strain=3%.....	58
Figure 4.18 The prediction of AISI 316 stainless steel's total crack length (mm) dependence on strain with C and N vary in specification, processing parameters are: Th=3.18 mm, I=100 A, U=12 V, Ve=4.23 mm/s.....	59
Figure 5.1 The workflow of predicting glass-forming ability using convolutional neural network plus periodic table representation, MF, PTR and CNN denote manual features, periodic table representation and convolutional neural network, respectively.	63
Figure 5.2 The statistics of elements distribution in our dataset. Many elements appear in crystalline alloys (CR), amorphous ribbon alloys (AMR), and bulk metallic glasses (BMG) simultaneously.....	65
Figure 5.3 The statistics of alloy systems' size in our dataset. Most alloy systems (918 out of 1134) have nine entries or less, the deficiency and sparsity of the data are apparent.....	65
Figure 5.4 Elements distribution in OQMD dataset. The occurrence numbers of elements in the dataset are given under element symbols. The white squares, <i>i.e.</i> square with element symbol and without occurrence number, indicate these elements are not in the dataset.	67
Figure 5.5 Mapping composition to 2-D representation.....	68
Figure 5.6 Schematic diagram for periodic table representation used in CNN3.....	69
Figure 5.7 Schematic diagram for randomized periodic table representation used in CNN2.....	69
Figure 5.8 Schematic diagram for atom table representation used in CNN1.....	70
Figure 5.9 Schematic diagram for our VGG-like convolutional neural network for glass-forming ability prediction.....	73

Figure 5.10 Schematic diagram for our VGG-like convolutional neural network for compound properties prediction.....	73
Figure 5.11 (a) Experimental data points of Al-Ni-Zr ternary system in our dataset and the predictions of (b) CNN3, (c) CNN2, (d) CNN1, (e) SNN4, (f) SNN3, (g) SNN2, and (h) SNN1.....	79
Figure 5.12 The prediction of Al-Ni-Zr ternary system by the re-trained models using a dataset that excluded Al-Ni, Al-Zr, Ni-Zr binary alloys and Al-Ni-Zr-containing multi-component alloys. (a) ground truth, (b) CNN3, (c) CNN2, (d) CNN1, (e) SNN3, and (f) SNN4.....	80
Figure 5.13 Experimental results and the predictions of S-bearing alloys $Ti_xZr_{75-x}Ni_{12}Cu_5S_8$, $Ti_xZr_{75-x}Ni_{17}S_8$, $Ti_xZr_{75-x}Cu_{17}S_8$ ($x=0-35$).....	84
Figure 5.14 The regression analysis between ground truth (horizontal axis, T values) and convolutional neural networks' predictions (vertical axis, Y values) of (a) compounds' forming energy Ef (b) compounds' specific volume. Testing accuracy R is also shown in the figure.....	87
Figure 5.15 The electron configuration, physical and chemical properties of elements show periodic variation in periodic table, e.g. atomic radius decreases and Pauling electronegativity increases from left to right within a period; atomic radius increases and electronegativity decreases from top to bottom within a group.....	87
Figure 5.16 Illustration of the 192-dimensional feature vectors of 108 elements extracted by CNN3. The values of features are represented by color (see the color bar to the right).	89
Figure 5.17 Illustration of the 192-dimensional feature vectors of 108 elements extracted by CNN2. The values of features are represented by color (see the color bar to the right).	89
Figure 5.18 Illustration of the 128-dimensional feature vectors of 108 elements extracted by CNN1. The values of features are represented by color (see the color bar to the right).	90
Figure 5.19 Features analysis of the glass-forming ability prediction model after principal component analysis (PCA). Projection of the feature vectors of 108 elements onto the plane spanned by the first and second principal axes. The percentage represents the ratio of the variance on the principal axis direction. Elements are colored according to their groups. (a) periodic table representation; (b) randomized periodic table representation. The superscript 1-18 on element symbol represents the element's group number; superscripts 19 and 20 represent lanthanide and actinides, respectively.	

Figure 5.20 Visualization of the first three principal features (after principal component analysis) of 80 pure elements. The percentage represents the ratio of the variance on the principal axis direction. Elements are colored according to their elemental groups. The superscript 1-17 on the element symbol represents the element's group number; superscript * represents lanthanide..... 93

Figure 6.1 The workflow of transfer learning (a) GFA to MPEA; (b) transfer from OQMD to phase prototype. GFA, AM, CR, MEPA, Conv, OQMD, CNN, SNN, RF denotes glass-forming ability, amorphous, crystalline, multi-principal element alloys, convolutional operation, fully connected layer, open quantum materials database, convolutional neural network, shallow neural network, random forest, supported vector machine, respectively. 95

Figure 6.2 Statistics of the 355 multi-principal element alloys in the dataset. (a) The numbers of binary to nonary multi-principal element alloys and the proportions of different phases. (b) The occurrence numbers of elements in the dataset are given under periodic table background. The blank squares, *e.g.* squares for noble gases, signify the elements not in the dataset. 97

Figure 6.3 Elements distribution in our phase prototypes dataset. The occurrence numbers of elements in the dataset are given under element symbols. The white squares, *i.e.* square with element symbol and without occurrence number, indicate these elements are not in the dataset..... 98

Figure 6.4 Statistics of each phase prototype's entry number in the dataset. In 170 phase prototypes, 93 prototypes have entries less than 56, and only 3 prototypes have entries more than 800. The label distribution of our dataset is unbalanced. 98

Figure 6.5 ROC curves of the transfer learning models for predicting multi-principal element alloy phases under 5-fold cross-validation. 103

Figure 6.6 Accuracy vs. testing ratio of three transfer learning models using feature vectors $\text{Vector}^{\text{Ef}\&\text{V}}$, $\text{Vector}^{\text{Ef}}$, Vector^{V} , and one non-transfer-learning model using composition vector $\text{Vector}^{\text{comp}}$ 106

Figure 6.7 The multi-principal element alloys' first four principal features generated by glass-forming ability model. Alloys are colored according to their phases. The percentage represents the ratio of the variance on the principal axis direction. 108

Figure 6.8 Visualization of the first three principal features (after principal component analysis) of

substances belonging to four common phase prototypes in our dataset. Inorganic substances are colored according to their phase prototypes. The percentage represents the ratio of the variance on the principal axis direction..... 110

Figure 6.9 Visualization of the first three principal features (after principal component analysis) of substances belonging to 15 uncommon phase prototypes in our phase prototypes dataset. Inorganic substances are colored according to their phase prototypes. The percentage represents the ratio of the variance on the principal axis direction..... 110

Contents

1	Introduction	1
2	Literature review on machine learning.....	3
2.1	Definition of machine learning	3
2.2	Machine learning workflow	4
2.3	Basic conceptions of machine learning	5
2.3.1	Data, classification & regression	5
2.3.2	Training & testing datasets	5
2.3.3	Overfit & underfit	6
2.3.4	Loss function, backpropagation algorithm & regularization	7
2.4	Typical machine learning models.....	8
2.4.1	Fully connected shallow & deep neural network	8
2.4.2	Convolutional neural network	11
2.4.3	Decision tree	12
2.4.4	Support vector machine	13
2.4.5	Ensemble and bagging	13
2.5	Transfer learning	14
2.6	Challenges in machine learning for materials research.....	15
2.6.1	Data-wise challenges	16
2.6.2	Descriptor-wise challenges	17
2.6.3	Task-wise challenges	17
3	Literature review on solidification cracking & phase formation.....	19
3.1	Solidification cracking	19
3.1.1	Hot cracking and solidification cracking	19

3.1.2	The complexity of solidification cracking.....	19
3.1.3	Material factors.....	20
3.1.4	Mechanical factors	23
3.1.5	Thermal factors	25
3.1.6	Solidification cracking of stainless steel.....	25
3.1.7	Solidification cracking susceptibility test	26
3.1.8	Machine learning for solidification cracking.....	27
3.2	Phase formation prediction	29
3.2.1	Amorphous phase and glass-forming ability	29
3.2.2	Multi-principal element alloys	31
3.2.3	Phase prototypes of inorganic substances	32
4	Application of deep neural network and pre-training to predicting solidification cracking susceptibility with small dataset	34
4.1	Dataset	36
4.2	Data pre-process & dataset division	40
4.2.1	Data pre-process	40
4.2.2	Training & testing datasets division	40
4.3	Machine learning models	40
4.3.1	Fully connected shallow & deep neural networks.....	40
4.3.2	Support vector machine	42
4.3.3	Tree-based models.....	42
4.4	Training details	43
4.5	Training and testing results	43
4.6	Validation of solidification cracking susceptibility prediction	47
4.7	Visualization of prediction and interpretability	49
4.7.1	Neural network equations	49
4.7.2	Feature importance	51
4.7.3	Visualization of prediction	52

4.8	Summary	60
5	Application of convolutional neural network & periodic table representation to predicting glass-forming ability & compound properties	62
5.1	Datasets	64
5.1.1	Glass-forming ability dataset	64
5.1.2	Open quantum materials database dataset	67
5.2	Data representation	67
5.2.1	2-D representation for data	68
5.2.2	Manual features engineering.....	70
5.2.3	Magpie descriptors	71
5.3	VGG-like convolutional neural network.....	73
5.4	Training details	75
5.5	Results of glass-forming ability prediction.....	76
5.5.1	Training and testing results	76
5.5.2	Leave-one-system-out cross-validation results	80
5.5.3	Validation using alloy systems outside our dataset	81
5.6	Results of compound properties prediction	86
5.7	The benefits of periodic table representation	87
5.8	Summary	93
6	Applications of transfer learning to predicting phase formation in multi-principal element alloys and phase prototypes with small datasets.....	95
6.1	Datasets	96
6.1.1	Multi-principal element alloys dataset	96
6.1.2	Phase prototype dataset.....	97
6.2	Details of transfer learning	100
6.3	Testing results	102
6.3.1	Testing accuracy on multi-principal element alloys dataset.....	102
6.3.2	Testing accuracy on phase prototypes dataset.....	106

6.4	Visualization of the automatically-extracted features.....	107
6.4.1	Features analysis for multi-principal element alloys	108
6.4.2	Features analysis for phase prototypes	109
6.5	Summary.....	111
7	Conclusions and future work	113
7.1	Conclusions	113
7.2	Future work.....	114
	Data and code availability.....	116
	Reference	116

1 Introduction

Machine learning is a powerful tool and an important complement to experiment, theory, and modeling. However, there are many challenges in machine learning for materials research: *e.g.*, small datasets, tricky manual feature engineering, poor interpretability, poor transferability. These challenges demand some tailored approaches. How to fully exploit limited data, automatic feature engineering, existing models and domain expertise is the key to effectively applying machine learning in materials research. This thesis shows our attempts to tackle these challenges through deep learning and transfer learning.

Figure 1.1 shows the structure of this thesis. Chapter 1 introduces the basics of machine learning and the challenges in machine learning for materials problems. Chapter 2 summarizes the theory of solidification cracking and phase formation, the applications of machine learning to predicting them in literature. Chapter 3 presents our attempts to improving model performance with a small dataset through deep neural network and pre-training. Case study about solidification cracking susceptibility prediction was used as a successful example. Chapter 4 presents our attempts to bypass the tricky manual feature engineering through periodic table representation and convolutional neural network. Case studies about predictions of glass-forming ability and compounds' properties (formation energy and specific volume) were used as evidences. Chapter 5 presents our attempts to tackle manual feature engineering and small datasets through deep transfer learning (with the help of the well-trained models in Chapter 4). Case studies about predictions of phase formation in multi-principal element alloys and compounds' phase prototypes were used as two successful examples.

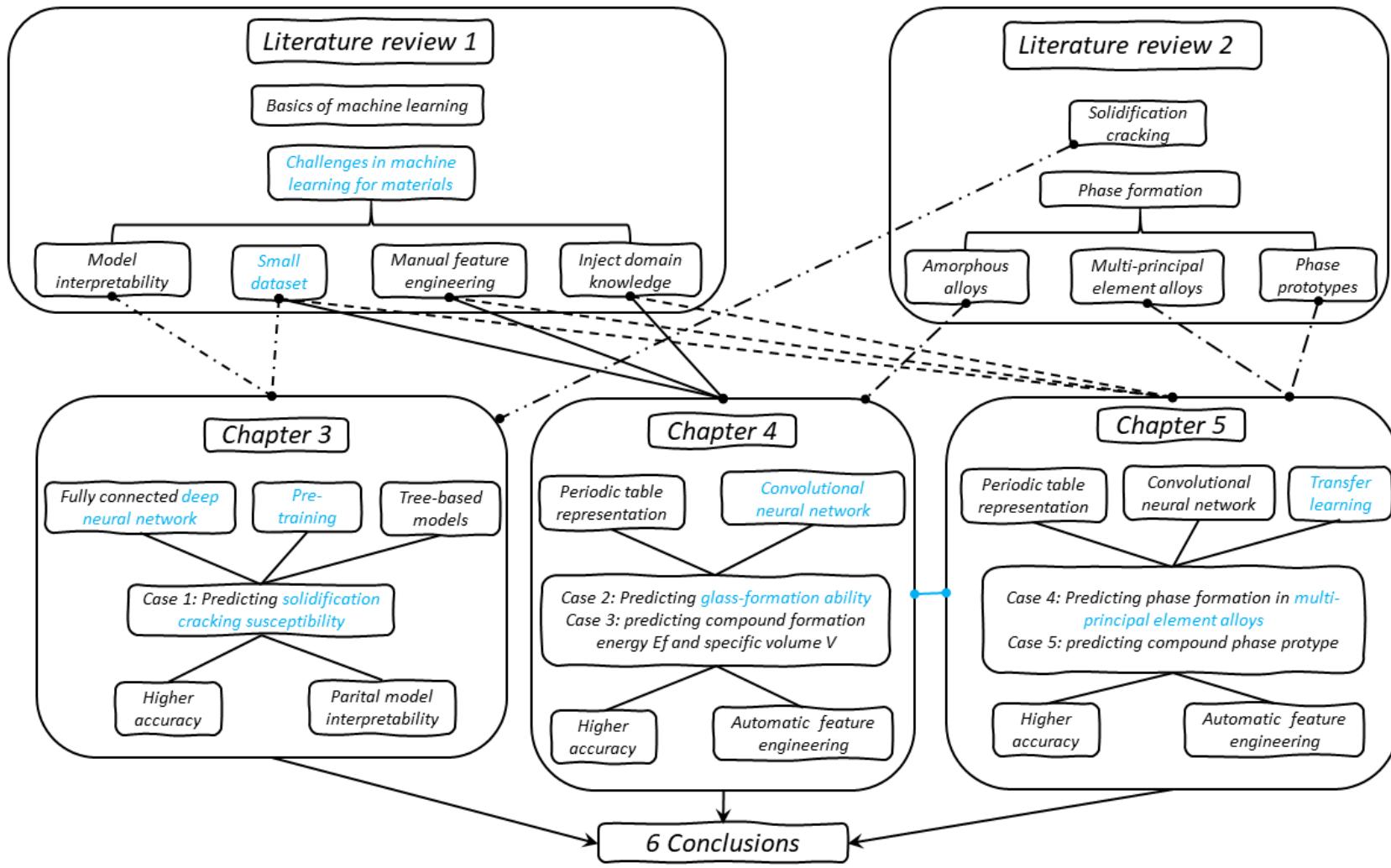


Figure 1.1 The structure of this thesis.

2 Literature review on machine learning

2.1 Definition of machine learning

Machine learning (ML), which is also known as statistical learning, data mining, and pattern recognition, is an important branch of artificial intelligence (AI)^{1,2}. It is a powerful technique that enables us to build adaptive mathematical models that automatically improve performance through example data or experience without explicit programming. Machine learning, especially its newest progress deep learning (DL), has achieved state-of-art performance in a variety of applications, *e.g.* search engine, spam email filtering, recommendation system, autonomous vehicles, computer vision (CV), speech recognition, natural language processing (NLP)^{3,4}. It is increasingly being exploited in science researches^{5–13}. The data-based and data-driven method has been recognized as the 4th paradigm¹⁴ of science and a essential complement to experiment, theory, and simulation.

Figure 2.1 shows the schematics for a parametric machine learning system (*e.g.*, logistic regression, naive Bayes, neural network): machine learning algorithms approximate the given dataset using a set of base functions, and the obtained surrogate models are used in new example prediction. A generic parametric training process must find the best parameter vector which minimizes the regression/classification error given a specific training dataset and it should also generate a model that can correctly generalize when unknown samples are provided.

Nonparametric machine learning (*e.g.*, k-nearest neighbors, Gaussian process, decision trees, support vector machines) is instance-based and does not make initial assumptions about the set of base functions. It makes predictions (without pre-computing parameter values) based on given labeled samples (instance set) and the concept of neighborhoods.

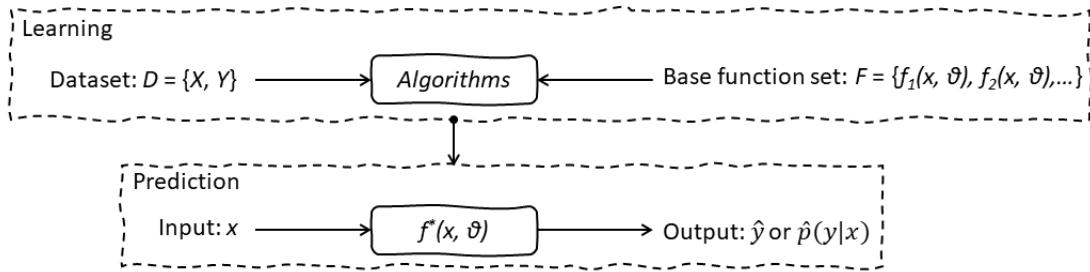


Figure 2.1 Schematic diagram for a parametric machine learning system, θ represents parameter vector.

2.2 Machine learning workflow

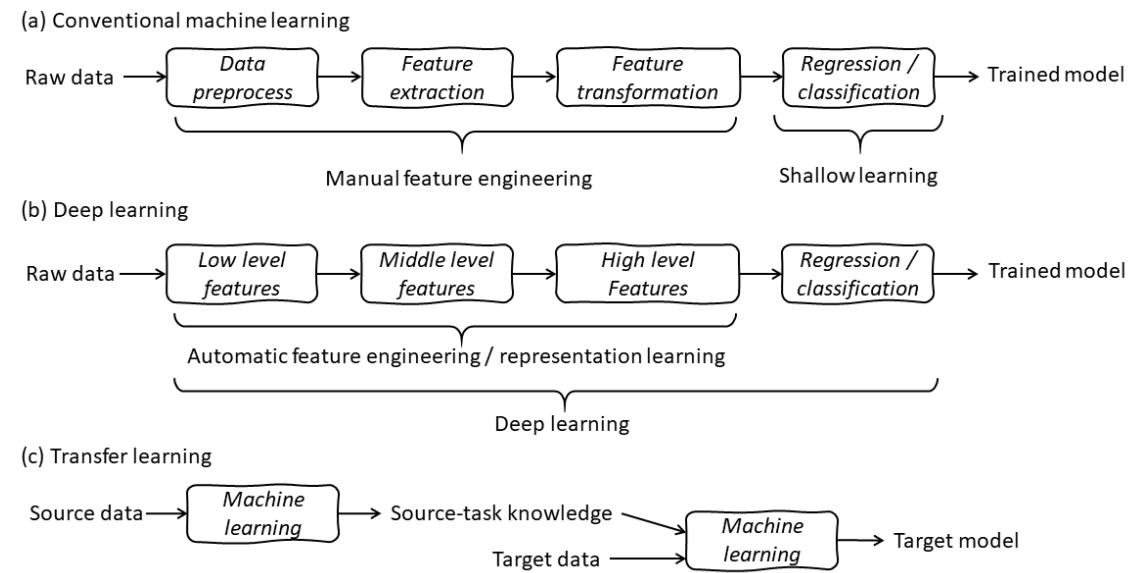


Figure 2.2 Workflow for (a) conventional machine learning, (b) deep learning, (c) transfer learning.

Figure 2.2 shows the workflows for conventional machine learning, deep learning, and transfer learning. Data preprocess involves denoise, token, filling missing data. Feature extraction includes some standard processing e.g., scale invariant feature transform (SIFT) and some special processing based on domain expertise and empirical knowledge. Feature transformation involves increasing/decreasing features dimensional e.g., principal component analysis (PCA) and linear discriminant analysis (LDA). Conventional machine learning is a combination of manual feature engineering and shallow leaning. Deep learning consists of automatic feature engineering and shallow learning. Thus, raw data, e.g., images, are used as direct input. Transfer learning was proposed to share

knowledge between different tasks^{15,16}.

2.3 Basic conceptions of machine learning

2.3.1 Data, classification & regression

Data is the most important component (lifeblood) for a machine learning system. It can be structured data *e.g.*, Excel datasheets, comma-separated values (CSV) data, and unstructured data *e.g.*, image, text, sound, time serial data.

A piece of data is also known as one entry, or an example, or an instance, depending on different scenarios. A dataset is a finite set of examples and their corresponding labels (targets). If labels exist for all entries in a dataset, it is a supervised learning task; if no label is available, it is an unsupervised learning task; if only partial labels exist, it is called a semi-supervised learning process. Labels can be numerical-continuous or categorical. In the first case, the process is called regression, while in the second, it is called classification. A dataset used in supervised learning can be represented as $D = \{X, Y\}$. In a dataset of structured data, input (features, which are also known as attributes and descriptors) can be represented as $X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$ where feature vector $\vec{x}_i \in \mathbb{R}^m$. Output (target) can be represented $Y = \{y_1, y_2, \dots, y_n\}$ where $y_i \in \mathbb{R}^1$ in single-task problems and $y_i \in \mathbb{R}^+$ in multi-task problems).

2.3.2 Training & testing datasets

In machine learning, a dataset is often split into training and testing datasets. The former is used for training a machine learning model and the latter is used for testing the performances of a trained model. If early-stopping technique is used, a small part of training dataset will be held out as validation dataset. In dataset division, training and testing datasets must reflect the original distribution of the dataset.

2.3.3 Overfit & underfit

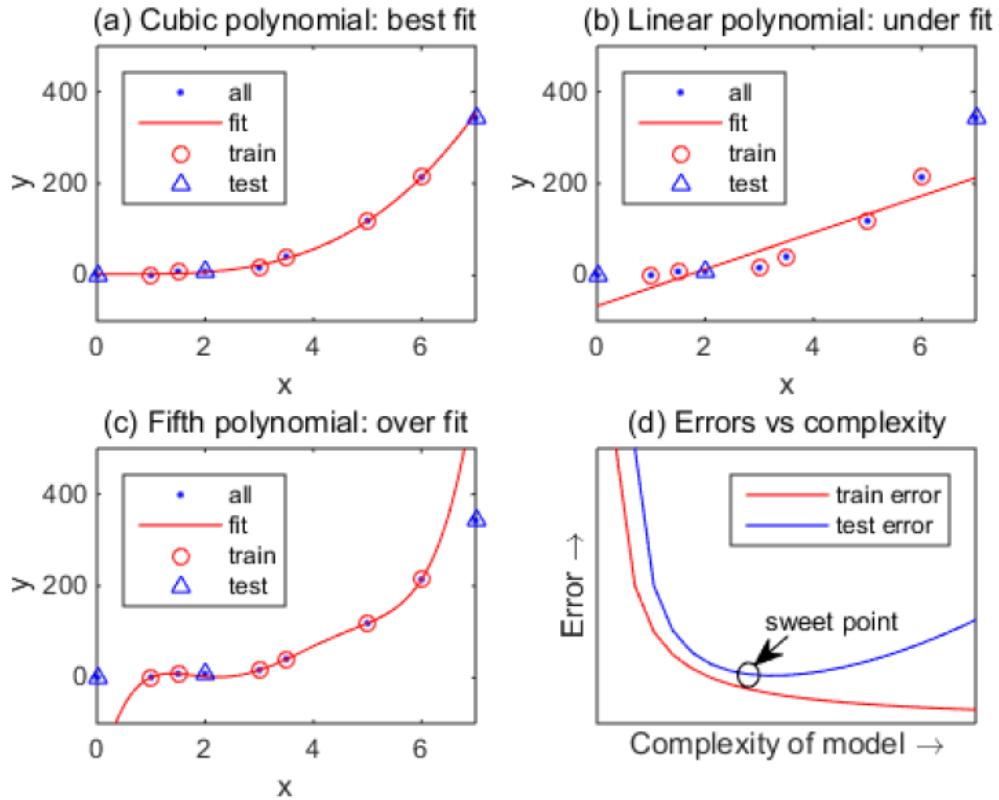


Figure 2.3 Schematic diagrams for (a) best fit, (b) underfit, (c) overfit, and (d) training error/testing error vs. model complexity¹⁷.

In machine learning, we need to select the best fit model which is neither underfit nor overfit (see Figure 2.3). The problem of underfitting will appear if we use an excessively simple model to approximate a complex dataset. For example, when the number of neurons is small in a neural network regression, underfitting occurs because the network lacks sufficient internal parameters (ability) to capture the complete characteristics of a dataset. Both training error and testing error are large in underfitting. It can be easily solved by increasing the complexity of model (*e.g.*, increase the number of neurons in a network).

In overfitting, the training error can be lowered to a very small value, but testing error turns out to be large. It indicates that the network only memorizes the training data, but it has not learned to generalize to new instances. A few techniques have been designed to avoid overfitting and improve the generalization, *e.g.*, early stopping, regularization,

drop out, batch normalization^{3,18}.

2.3.4 Loss function, backpropagation algorithm & regularization

The training process of many machine learning models *e.g.*, neural network, is a process of minimizing the loss function though updating the internal parameters (*e.g.* weights and biases) of models. A typical loss function often used in machine learning is mean square error (MSE):

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (t_i - a_i)^2$$

where t_i is the target value and a_i is the prediction value of the model based on the input data.

To improve generalization and reduce the risk of overfitting, a regularized loss function MSE_{reg} , which combines mean square error with mean square weight (MSW), is also usually employed:

$$\text{MSE}_{\text{reg}} = \gamma \text{MSE} + (1 - \gamma) \text{MSW}$$

$$\text{MSW} = \frac{1}{n} \sum_{j=1}^n W_j^2$$

where γ is the performance ratio (a hyper-parameter either given by researchers or automatically determined by optimization algorithm); W_j is the j-th weight. Minimization of mean square weight entails small values of internal parameters, which assures the model responses smoothly and reduces the risk of overfitting.

Gradient descent algorithm¹⁹ (backpropagation algorithm) updates the parameters of a model in the direction in which the loss function decreases most rapidly, *i.e.* the opposite direction of gradient. In an iteration, parameters are updated as follows:

$$x^{k+1} = x^k - \alpha^k g^k$$

where x^{k+1} is the updated parameters, x^k is the old parameters, g^k is the gradient, and α^k is the learning rate.

Automatic determination of the optimal regularization parameters is always desired.

Bayesian regularization proposed by MacKay^{20,21} gives us a good solution of this type. MacKay assumed that the internal parameters of a model are randomly distributed, and the regularization parameters can be estimated by Bayesian inference.

2.4 Typical machine learning models

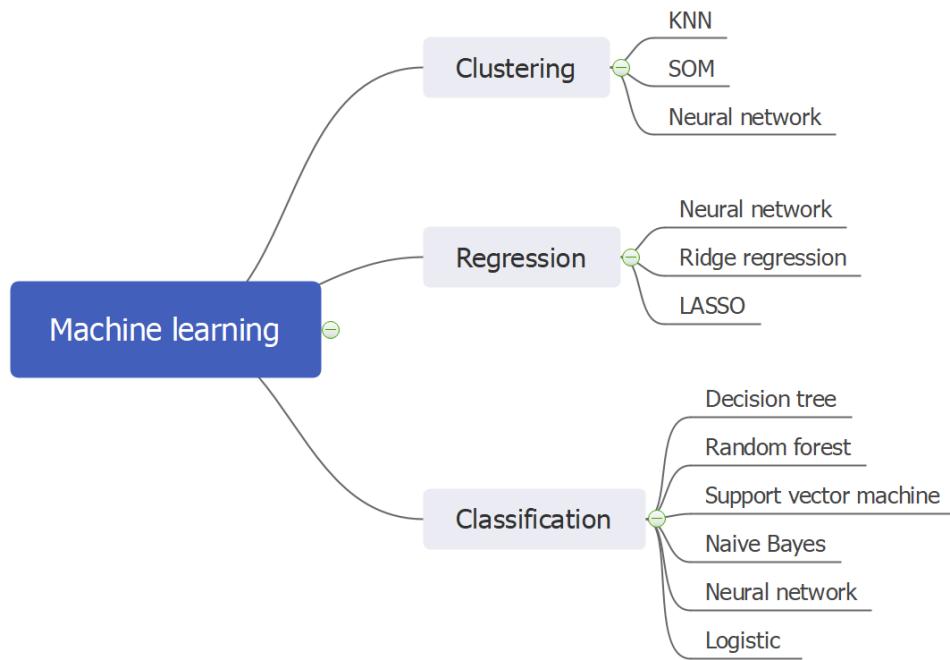


Figure 2.4 Typical algorithms in machine learning.

Figure 2.4 shows some typical algorithms in machine learning^{1,2,22}. Neural network, decision tree, nearest neighbor, Naïve Bayes, support vector machine are usually used in data classification. Neural network, Gaussian process, linear regression are usually used in data regression. Neural network, hidden Markov model, self-organizing map (SOM), K-nearest neighbors (KNN), K-means are usually used in data clustering.

2.4.1 Fully connected shallow & deep neural network

The first-generation neural network, *i.e.* single layer perception²³, was firstly proposed in 1950s as a mathematical approximation of human neural cells. The second-generation neural network, *i.e.* multi-layers perception or shallow neural network (SNN), commonly has one hidden layer (or no more than two hidden layers). The latest generation neural network, *i.e.* deep neural network (DNN), usually has three or more hidden layers. The deepest neural network now can have hundreds of layers.

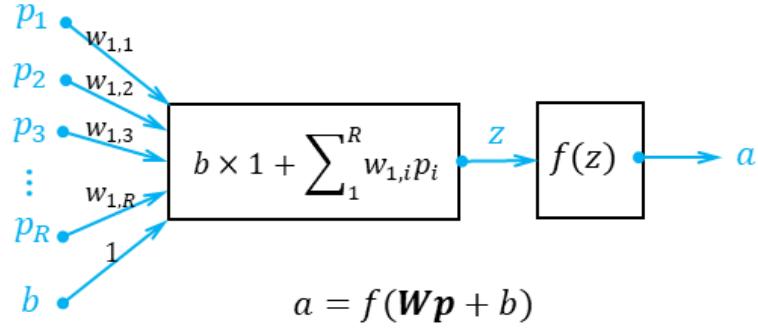


Figure 2.5 Schematic diagram for a neuron, the basic unit of a neural network.

Figure 2.5 shows the basic unit of a neural network (*i.e.* a neuron): the input vector [p_1, p_2, \dots, p_R] are multiplied by corresponding weights [$w_{1,1}, w_{1,2}, \dots, w_{1,R}$] and their summation (include a bias b , a special weight with constant input value 1) z is: $z = w_{1,1}p_1 + w_{1,2}p_2 + \dots + w_{1,R}p_R + b * 1$. The mathematical expression can be abbreviated by matrix notation: $n = W_{1 \times R}p_{R \times 1} + b$, where $W_{1 \times R}$ is the 1 row and R columns weight matrix, and $p_{R \times 1}$ is the R row and 1 column input vector. Then n is processed by the transfer function $f(z)$ and it yields the neuron's output a : $a = f(z) = f(Wp + b) = f(w_{1,1}p_1 + w_{1,2}p_2 + \dots + w_{1,R}p_R + b)$. Based on the unit, we can build a complex neural network (multiple neurons, multiple layers, and multiple inputs) through increasing the number of neurons and layers.

Figure 2.6 shows a shallow neural network of one hidden layer and a deep neural network of j hidden layers schematically. The X represents input vector, Y represents output vector. The number n_i in brackets signifies the number of neurons in the i -th hidden layer. Circles represent neurons and lines represent the links between them, W^i represents the weights matrix of the j -th layer. The biases are omitted for simplicity. The structures of deep neural network and shallow neural network are similar, except that deep neural network has more layers and a more obvious hierarchy structure.

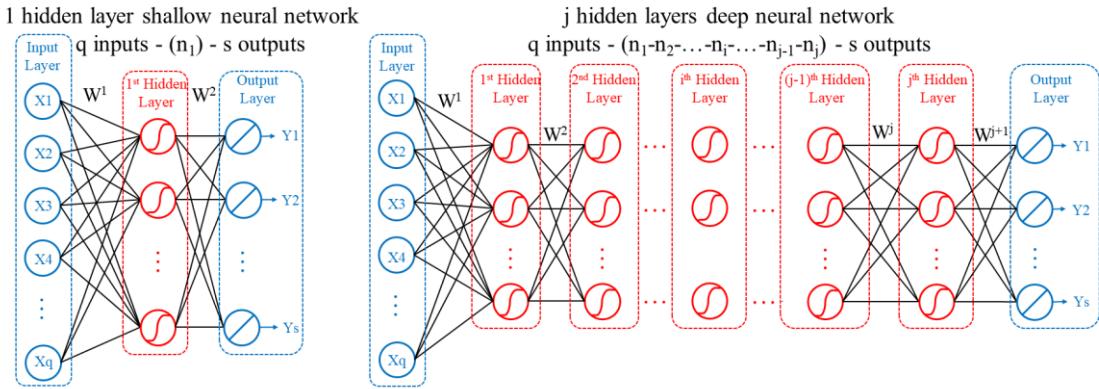


Figure 2.6 Schematic diagrams of one hidden layer fully connected shallow neural network and j hidden layers fully connected deep neural network.

Shallow neural network is widely used in function approximation, nonlinear regression, pattern recognition and data clustering. It is a mature tool in materials research. The applications of it in materials science include: predicting physical properties^{24–29} (e.g., stacking fault energy, phase transformation temperatures); mechanical properties^{30–35} (e.g., strength, ductility, creep properties, fatigue properties); alloys optimization^{36–39}; facilitating quality control^{40–47} (e.g., discriminating weld defect type, forecasting weld pool shape in welding and mold level control in continuous caster). The ISIJ special issue of neural network⁴⁸ and Bhadeshia's works^{17,49} show us more examples of the applications of shallow neural networks in material science.

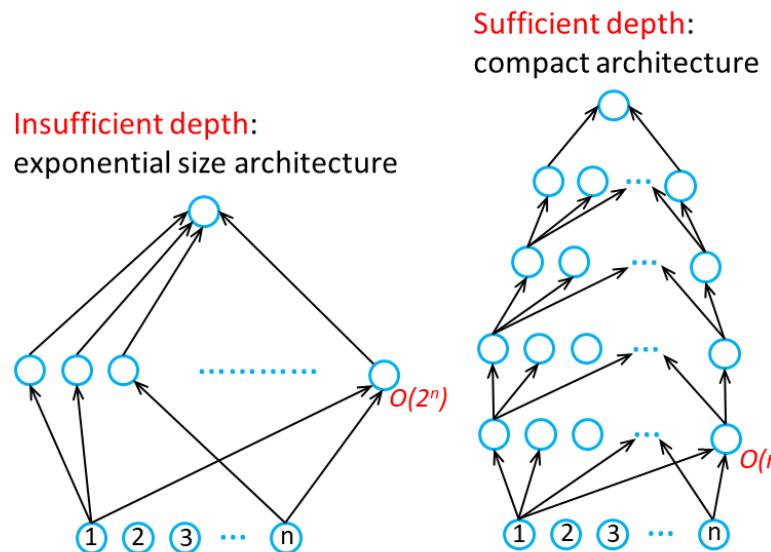


Figure 2.7 Schematic diagram for the architecture differences between shallow network and deep network for the same task.

Researchers believe that the power of a neural network increases with its depth. Bengio⁵⁰ proved that deep neural network uses fewer parameters (neurons) than shallow neural network to fit a task (dataset). Figure 2.7 schematically shows the difference. Researchers have been exploring the effective ways to train deep neural network which have thousands of layers like humans' long connected neural system since neural network was firstly proposed^{3,23}. Training a deep neural network like training a shallow neural network generally usually does not work: the performance of deep neural network always worse than that of shallow neural network⁵¹. Researchers believed that gradient vanishing with increasing depth of network and traps of poor local minimums were the main difficulties in training deep neural network^{4,52,53}. Hinton proposed restricted Boltzmann machine (RBM) and deep belief network (DBN) in 2006. Deep neural network then can be successfully trained through a so-called greedy layer-wise pre-training⁵³. Later, Bengio proposed stacked auto-encoder (SAE) as an alternative to restricted Boltzmann machine in pre-training⁵⁴. Since the success of deep belief network, deep neural network is thriving. Deep neural networks, especially the convolution neural network, accomplished unprecedented successes in image recognition. Deep learning (deep neural network) has become the mainstream in machine learning now.

2.4.2 Convolutional neural network

Convolutional neural network (CNN), a special type of neural network, is widely employed in processing images^{55,56}, e.g. images classifications, object detection, image semantic segmentation. A Convolutional neural network³ usually has a series of convolution layers with filters (kernels), pooling layers, ReLU layers, fully connected layers (FC). Softmax function layer is usually used as output layer in classification. Convolution kernels can realize complicate image operations such as image edge detection, sharpen and blur. Stride is the number of pixels we move the filters at a time over the input matrix. Sometimes filter does not perfectly fit the input image. We commonly pad the picture with zeros (zero-padding). ReLU represents rectified linear unit for a nonlinear operation. Its mathematic expression is $f(x) = \max(0, x)$. Pooling (subsampling or downsampling) layers (max pooling, average pooling, sum pooling) reduce the number of parameters but keep important information.

2.4.3 Decision tree

Tree-based machine learning models are a family of non-parametric supervised methods which are widely used in materials science^{57–60} due to their good interpretability. A decision tree¹ looks like an upside-down tree, with the first decision rule at the top and following decision rules spreading out below. In a decision tree, every decision rule (e.g., “If A > B...”) that produces the greatest decrease in impurity index occurs at a decision node, with the rule creating branches leading to new nodes. A branch without a decision rule at the end is called a leaf. Two frequently used impurity index are Gini impurity index $I_{Gini}(j) = \sum_i p(i|j)(1 - p(i|j))$ and cross-entropy $I_{Cross-entropy}(j) = -\sum_i p(i|j)\log(p(i|j))$. Here, $p(i|j)$ is the probability of the samples that belong to class i for a particular node j. Decision tree regression model attempts to find a decision rule that produces the greatest decrease in mean squared error (MSE) at a node. One benefit of tree-based models is their interpretability. Visualization of decision trees can give us an intuitive knowledge of the decision process. Decision trees can be used to evaluate the relative importance of each feature based on the impurity reduction determined by every single feature. The relative importance of feature x_i is determined as follows:

$$Importance\ of\ x_i = \sum_k \frac{N_k}{N} \Delta I_{x_i}$$

The sum is extended to all nodes where x_i is used, and N_k is the number of samples reaching the node k, ΔI_{x_i} is the decrease of impurity index. Therefore, the importance is a weighted sum of all impurity reductions computed considering only the nodes where the feature is used to split them.

When decision trees are built, many of the branches may reflect noise or outliers in the training data. Tree pruning attempts to identify and remove such branches to improve model's generalization.

2.4.4 Support vector machine

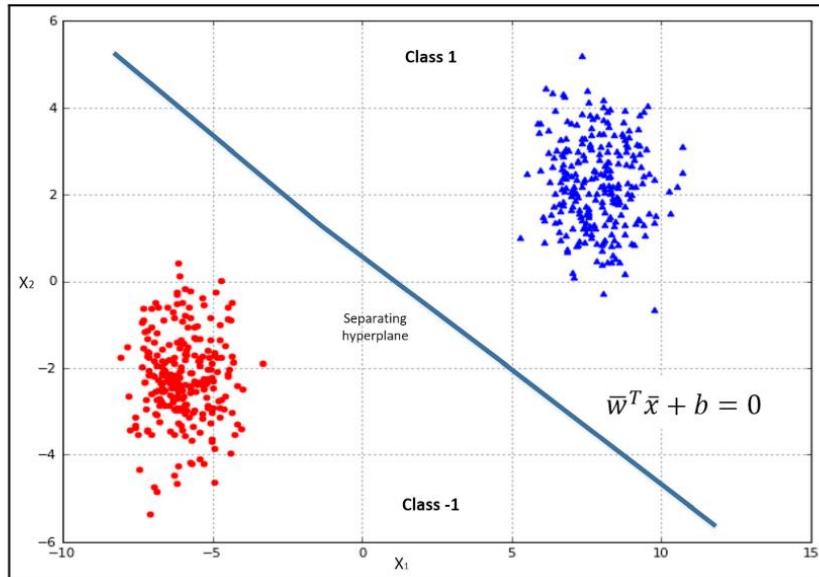


Figure 2.8 A schematic diagram for support vector machine.

Support vector machine² (SVM) accomplishes classification tasks by creating a hyperplane to segregate the data. This hyperplane is determined by maximizing the vector normal to the hyperplane (usually labelled W) and the closest data point to create the largest gap possible. A schematic diagram for support vector machine is shown in Figure 2.8. When working with nonlinear problems, it is useful to map original vectors into a higher dimensional space where they can be linearly separated. The so-called kernel trick² realizes this nonlinear transformation. The value of the kernel for two feature vectors is the product of the two projected vectors: $K(\bar{x}_i, \bar{x}_j) = \phi(\bar{x}_i)^T \phi(\bar{x}_j)$. Linear kernel, radial basis function (RBF) kernel, and polynomial kernel are frequently used in machine learning.

2.4.5 Ensemble and bagging

The models introduced previously are all so-called strong learners which solve a specific problem by looking for the best possible solution. Another approach is based on a set of weak learners that can be trained in parallel or sequentially (with slight modifications on the parameters) and used as an ensemble based on a majority vote or the averaging of results^{2,22}. These methods can be classified into two main categories: bagging (or bootstrap) and boost. In bagging, the ensemble is built completely. The training process

is based on a random selection of the splits and the predictions are based on a majority vote. Random forests are an example of bagged tree ensembles. In boost, the ensemble is built sequentially, focusing on the samples that have been previously misclassified. AdaBoost and gradient tree boosting are examples of boosted trees.

2.5 Transfer learning

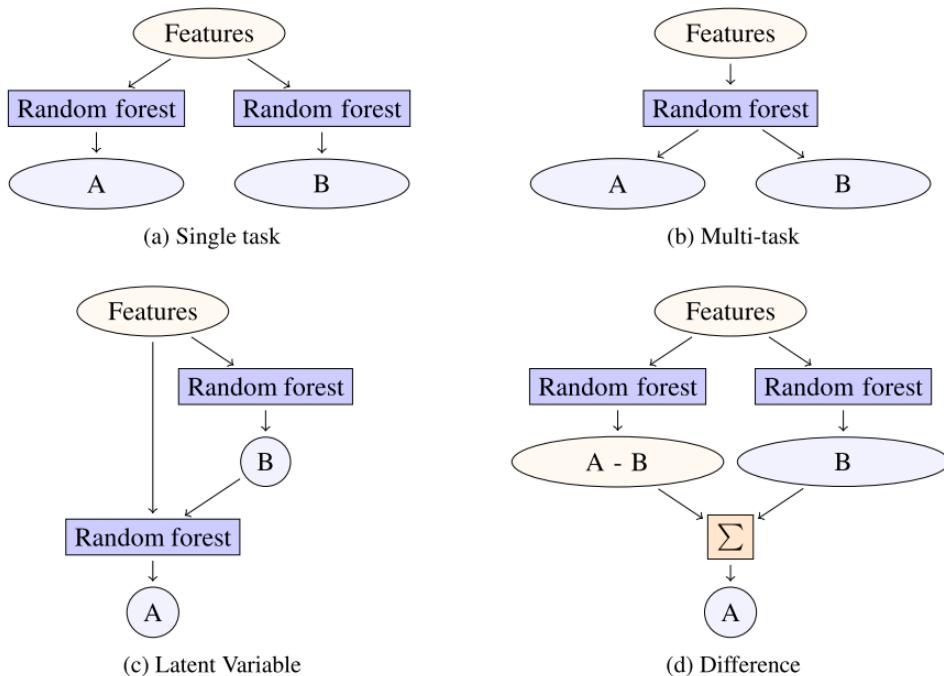


Figure 2.9 (a) Single task architecture and (b-d) architectures for three transfer learning techniques based on conventional machine learning⁶¹.

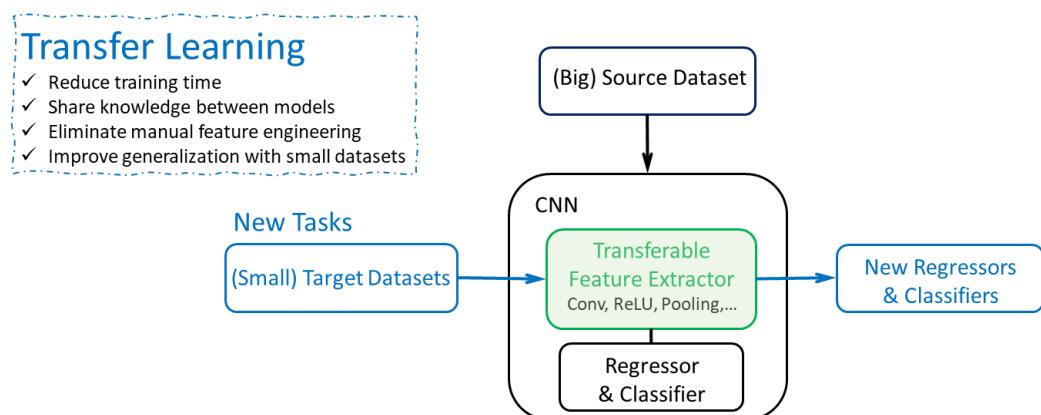


Figure 2.10 Architecture for the transfer learning technique based on convolutional neural network.

Transfer learning, where knowledge is shared between datasets and models, can be an

effective tool for sparse and small datasets. Figure 2.9 shows the architecture for single task and architectures for transfer learning based on conventional machine learning⁶¹. Figure 2.10 shows the architecture for the transfer learning technique based on convolutional neural network^{62,63}. The benefits of transfer learning include reducing training time, improving generalization with small datasets, sharing knowledge between related tasks, eliminating manual feature engineering.

2.6 Challenges in machine learning for materials research

Table 2.1 Challenges in machine learning for materials research.

Elements	Aspects	Typical ML applications (CV and NLP)	Material science applications
Data	Dataset size	Big dataset (usually $>10^4$ examples, up to 10^8 are available)	Small dataset (usually between 10^1 - 10^3 examples, $>10^4$ are rare)
	Data distribution X	Dense, balanced	Sparse, correlated, clustered, unbalanced X
	Label Y distribution	Balanced; sample bias often present	Unbalanced labels Y; biased labels (bad/failure results unreported)
	Error/noise in data X and label Y	Not important	Always exist, essential for model performance
	Data type	Often standardized	Rarely standardized, heterogeneous
	Data augment	Applicable	Usually not applicable
Features	Descriptors, data representation	Standard methods are available, e.g. SIFT; rely on data to learn patterns; can often be optimized by algorithms	Manual feature engineering requires deep domain knowledge
Tasks	Aims	Accurately pattern-match common cases, Interpolation	Need to predict unusual or “extreme” materials, extrapolation, optimization
	Prediction accuracy	Rough results are accepted	Need high prediction accuracy
	Prediction error estimation	Usually unimportant	Uncertainty and error estimation are desired
	Model interpretability	Usually unimportant	Interpretability is desired if possible

Machine learning was firstly designed and customized for its traditional application fields⁴, e.g. recommendation system, computer vision, natural language processing.

However, the tasks in material science show different characteristics from typical machine learning applications. Thus, existing machine learning tools need adjustments to make it work in materials science. Table 2.1 compared machine learning for its traditional applications and machine learning for materials from the points of data, descriptors, and tasks.

2.6.1 Data-wise challenges

Big data is a phrase commonly associated with artificial intelligence and machine learning. Companies, *e.g.*, Amazon, Google, Facebook, can quickly and cheaply gather big datasets of billions of samples. In materials science and engineering, the rate of generating data is very slow. In curation of material datasets, it is often completed by hand. The Materials algorithms project of Cambridge University, NIMS (Japan national institute for materials science) materials database provide dozens of materials datasets which are available only after decades of accumulation. Materials genome initiative (MGI)^{64–66}, high-throughput computing^{67–71} and material chips have increased the speed of generating data by dozens to hundreds of folds. However, it is still impossible to assemble big datasets like that in the internet and e-commerce. Thus, datasets in materials fields are often small (datasets usually have dozens of examples to several thousand of examples) for a specific task^{72,73}.

Machine learning applications, *e.g.*, product recommendation and consumer behavior analysis, acquire records from structured sources like website analytics tools. Images and videos used in computer vision can easily transform file formats. Materials data come from different sources (*e.g.*, test results, simulation data, handbook data, supplier datasheets) and in different formats (*e.g.*, microstructure images, datasheets of composition/processing/properties, processing history records, spectrum). Thus, material datasets are often highly heterogeneous. Curating a reliable dataset from distinct data sources requires deep understanding and careful processing of the data.

Sample bias in datasets is common in materials science. Failed experiments and materials of poor properties are commonly discarded and cannot be accessed publicly. The biased distribution of data is detrimental to the generalization of the built machine learning models without doubt.

The error and variation from processing and measurement are inevitable in material data. How to take into account of their influence on the achieved models is a big challenge for researchers in applying machine learning to material problems.

2.6.2 Descriptor-wise challenges

With a set of suitable descriptors, conventional machine learning can perform very well even with small dataset⁷⁴. However, the optimal set of descriptors for a specific job in material research is not out-of-shelf. It is selected by trials and errors and adding new pertinent descriptors is always be considered if models' performance is not met requirement⁷⁵. Building new applicable descriptors entails deep understanding of mechanisms, which is very challenging in developing new materials. For example, Ward et al. first used 145 general-purpose Magpie descriptors (descriptive statistics, *e.g.*, average, range, and variance of the constituent elements) in predicting ternary amorphous ribbon alloys⁵⁸. Later they used 210 descriptors (including 145 Magpie descriptors and new descriptors derived from physical models and empirical rules developed by amorphous alloys community, *e.g.* cluster packing efficiency and formation enthalpy) in optimizing Zr-based bulk metallic glass⁵⁷. Some descriptors derived from physical models and empirical rules are sensitive to alloying and temperature; obtaining precise values of them is difficult; using simplified models to calculate them (*e.g.*, utilize ideal solution model in estimating alloy mixing enthalpy instead of Miedema model or experimental results) might weaken the final machine learning models' performance. A lot of work in applying machine learning to materials research is spent on building a set of appropriate descriptors.

2.6.3 Task-wise challenges

Typical machine learning applications are tasks that are simple for humans but difficult to code explicitly, like speech recognition and image recognition. The tasks of material science and engineering are exploring the relationship between chemistry, processing, microstructure, properties, and performance; and to synthesis better materials by exploiting this knowledge. It is a task that is difficult for humans and hard to code explicitly. It demands prediction with high accuracy (even in extrapolation^{76,77} for

materials scientists are more interested in developing materials of better performance) and error estimation^{78–82}, models with interpretability^{56,83–86} (a "black box" model cannot meet our requirements in the material field) and transferability. If a machine learning model cannot meet these requirements, its applications will be constrained. Machine learning for materials science and engineering is more complicated than typical machine learning applications. It therefore requires a more sophisticated approach (special techniques and processes) at many points within the machine learning workflow.

In many typical machine learning applications, error estimation is not essential. For example, in the movies recommendation system, an error of 20% in prediction only means ± 1 point difference in a 0-5 scale. This usually has little impact on audiences. However, a 20% error in a prediction of material properties could mean the prediction is unreliable, and we need to carry out more costly experiments. Error estimation help researchers understand predictions and decide the next step actions.

3 Literature review on solidification cracking & phase formation

3.1 Solidification cracking

3.1.1 Hot cracking and solidification cracking

Hot cracking, which is also termed as hot tearing, hot shortness or hot brittleness, is one of the most severe defects formed at high temperature range in casting and welding, especially in Ni-based alloys and some Al-based alloys⁸⁷. Hot cracking can be defined as material separations that occur at high temperature range (above half the melting point of alloys) along the grain boundaries (dendrite boundaries) when the local strain and the local strain rate attain critical values⁸⁸. Hot cracking can be subdivided into three easily-confused types: solidification cracking, liquation cracking and ductility dip cracking^{89–91}. Ductility dip cracking is caused by reduced hot ductility of solid phase due to diffusion related compositional changes and precipitation of brittle secondary phases. Liquation cracking takes place in areas of liquation in heat-affected zone or in multi-pass welds. Solidification cracking occurs in mush zone at the last stage of solidification. The fractography of solidification cracks exhibits dendrites characteristics. The solidification cracks can develop to be surface cracks or remain to be subsurface / internal cracks⁸⁹.

3.1.2 The complexity of solidification cracking

Solidification crack, which widely occurs in casting, welding, and additive manufacturing, is a very serious defect for it cannot be healed through following processing e.g. rolling, forging and heat treatment. It appears at the last stage of solidification when liquid films exist between dendrites boundaries and the local strains cannot be accommodated through liquid phase feeding and solid phase deformation. It is a complex phenomenon for material, mechanical, and thermal factors work together to incur the nucleation and propagation of cracks and these factors interact with each other, see Figure 3.1⁸⁹. The weldability and castability of metallic materials are closely correlated with their solidification cracking susceptibility (SCS). The complexity and harmfulness of solidification cracking rationalize the existence of so many experimental techniques^{92,93}, physical models^{94–96}, and numerical models^{97–99}. Solidification cracking tests are time-

consuming and costly. Thus, only very few alloys and parts can be tested. The difficulty of direct in-situ observation of solidification cracking phenomena caused by high temperature and opaque metallic matrix determines that indirect measurements are the main approaches of studying solidification cracking. Theories about solidification cracking susceptibility based either on metallurgical features, or strain, or strain rate, or stress criteria are either over-simplified (no existing theory can take account of all factors mentioned before) or lack solid physical basis thus cannot work effectively in practice. Complex numerical models⁹⁷, which simultaneously modeled thermal, fluid, mechanical fields and intergranular network geometry, only able to predict the most susceptible positions of solidification cracking but not the solidification cracking susceptibility of alloys.

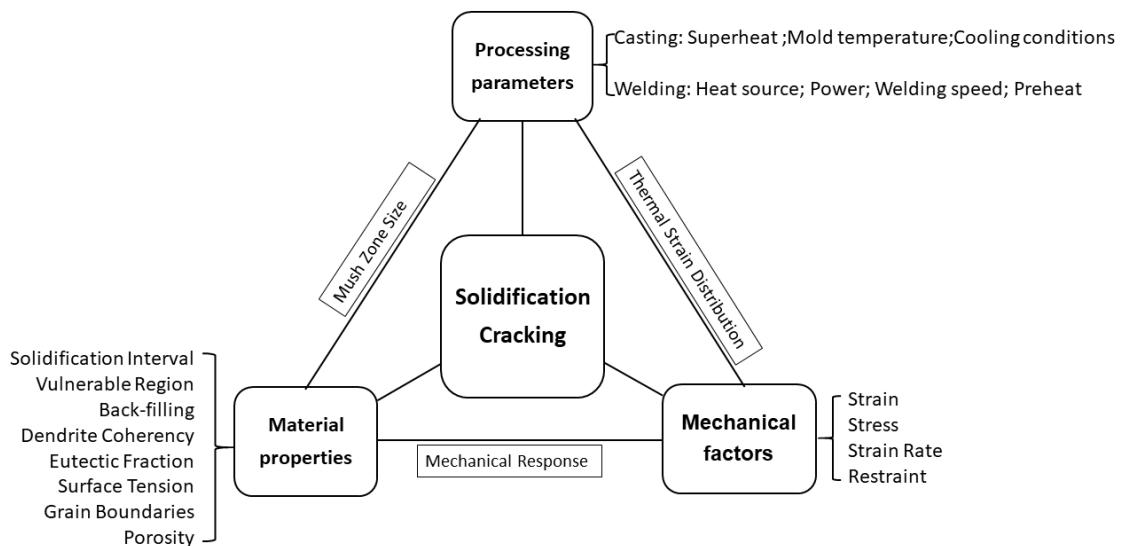


Figure 3.1 Schematic diagram for the complex interaction between thermal, material, mechanical factors, which cause solidification cracking.

3.1.3 Material factors

(1) Solidification interval

Solidification interval or solidification temperature range⁸⁹ ΔT of an alloy is defined as $\Delta T=T_L-T_S$ or $\Delta T= T_L-T_{Eut}$, where T_L is liquidus temperature, T_S is solidus temperature (use eutectic temperature T_{Eut} instead in nonequilibrium solidification). Solidification cracking occurs during liquid transforming to solid. Impurity elements such as sulfur and phosphorus remarkably enlarge the solidification interval of Fe-based and Ni-based

alloys by forming low melting point eutectics *e.g.* FeS-Fe ($T_{Eut} = 985^{\circ}\text{C}$), $\text{Fe}_3\text{P}-\text{Fe}$ ($T_{Eut} = 1050^{\circ}\text{C}$), NiS-Ni ($T_{Eut} = 644^{\circ}\text{C}$), and $\text{Ni}_3\text{P}-\text{Fe}$ ($T_{Eut} = 880^{\circ}\text{C}$). Thus, some researchers correlated solidification interval with solidification cracking susceptibility and proposed that solidification cracking susceptibility is in proportion to the solidification interval. However, many solidification cracking phenomena cannot be explained using this simple criterion, even use the solidification interval modified by Scheil model and back-diffusion model.

(2) Vulnerable Region

Clyne and Davies defined solidification cracking sensitivity (SCS)⁸⁷ based on thermal analysis: $SCS = t_v/t_r$, where t_v is the vulnerable time period during which local strain in mush zone is difficult to be accommodated and cracks can easily propagate, and t_r is the immune time period during which local stress and strain can be easily relaxed through liquid flow and solid deformation. The biggest problem of Clyne's solidification cracking sensitivity criterion is they simply choose (without much reason) t_v and t_r as: $t_v = t_{0.99}-t_{0.90}$, $t_r = t_{0.90}-t_{0.40}$, where the subscript numbers represent the volume fraction of solid. Different researchers may choose different t_v and t_r region, to make prediction results fit experiment results better.

(3) Solidification cracking temperature range (SCTR)

In varestraint test, the maximum crack distances (MCD) are quantified through imposing augmented strain on samples. If the cooling rate is tested by inserting a thermocouple into weld pool, a characteristic temperature range which is called solidification cracking temperature range (SCTR)¹⁰⁰ is defined as follow: $SCTR = [\text{Cooling rate}] \times [MCD/V]$, where V is the welding speed (m/s). SCTR is a useful parameter when comparing the relative weldability between different alloys.

(4) Back-filling

Solidification shrinkage causes shrinkage porosity⁸⁷. Porosity is a favorable site for solidification cracking initiation. Back-filling can reduce solidification porosity and heal solidification cracks. Risers are designed in casting to mitigate or eliminate solidification cracks, and the weld pool in welding plays the same role. The back-filling ability is closely connected with solidification cracking susceptibility. The back-filling ability is determined

by dendrite tortuosity, melt viscosity, interface energy between melt and solid, and solid fraction.

(5) Dendrite coherency

The difficulty of bridging or coalescence of dendrite arms during the last stage of solidification plays an important role in solidification cracking¹⁰¹.

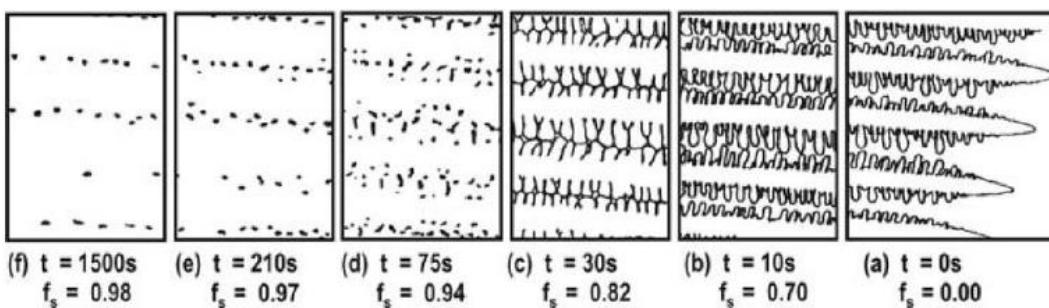


Figure 3.2 Stages of dendrites growth⁹⁵

Figure 3.2 shows different stages that dendrites experience in directional solidification growth⁹⁵: in stage (a - b) solid fraction increases with time but liquid back-filling is still available, dendrites begin to touch each other and solid bridges start to form at the end of this stage; in stage (c) only continuous liquid films exist between dendritic arms, it is the solidification cracking susceptible region; in stage (d-f) discontinuous residual liquid decreases and vanish at last, in this stage macro stain and stress gradually appear. Two characteristic temperatures¹⁰²⁻¹⁰⁴, *i.e.* coherency temperature and rigidity temperature, separate the three phases. Coherency temperature is the temperature that dendrites begin to touch each other. Rigidity temperature is the temperature that dendrites begin to be able to withstand tensile stress. During coherency temperature and rigidity temperature, tensile stress in mush zone will induce liquid film rupturing and cracking.

(6) Eutectic fraction

Scheil equation gives us the weight fraction f_{Eut} of inter-dendritic eutectic in a binary alloy with a composition of C_0 : $f_{Eut}=[C_0 / C_{Eut}]^{1/(1-k)}$, where C_{Eut} is the eutectic composition of the binary alloy, $k=C_s/C_L$ is the partition coefficient. Alloys with large fraction eutectic are generally less susceptible to solidification cracking^{89,105}, because eutectic structure is more open than dendrite structure and thus very good back-filling ability. Al-Si cast alloys with a large amount of eutectic own exceptional castability and

weldability. Exceptions always existed⁸⁹: Al-Mg alloys with large solidification interval and small eutectic fraction are expected to have high solidification cracking sensitivity, but in fact it is not. The cause might be relative to their high degree of coherency.

(7) Surface tension

Solidification cracking involves the separation of liquid film between grains, so surface tension^{87,89} plays a role in the process. Alloys of bad wettability (*i.e.* high interface energy $\gamma_{L/S}$ and bridge between grains is strong) and very good wettability (*i.e.* low interface energy $\gamma_{L/S}$ and back-filling ability is high) are not solidification cracking susceptible. Alloys of medium wettability are solidification cracking susceptible.

(8) Grain boundaries

Grain boundaries⁸⁷ are the areas where solidification cracking takes place. So the characteristics of grain boundaries have an obvious influence on solidification cracking susceptibility. These characteristics include grain shape (*i.e.* columnar vs equiaxed grains), grain structure (*i.e.* primary austenite stainless steel owns more straight grain boundaries and worse weldability than primary ferrite stainless steel) and grain size (fine grain means more grain interface need to be separated in the crack propagation compared with coarse grain. Thus fine grain castings and welds are less solidification cracking susceptible).

(9) Porosity

The formation of porosity^{87,89} is similar to the formation of solidification crack in that both involve the separation of liquid and the formation of new interface between liquid and vapor. So, it is very reasonable to correlate porosity with solidification crack, inferring the porosity can be the nucleus of solidification cracks. On the other hand, the formation of porosity reduces macro shrinkage, thus decrease solidification strain and solidification cracking susceptibility.

3.1.4 Mechanical factors

(1) Strain

Pellini proposed the ‘strain theory’ of solidification cracking: solidification cracking occurs when the local strain beyond a critical strain. Based on the strain theory, Senda

et al. proposed the conception of ‘ductility curve’ (shown in Figure 3.3) which can be determined by Gleeble experiment and other methods^{89,100,106}.

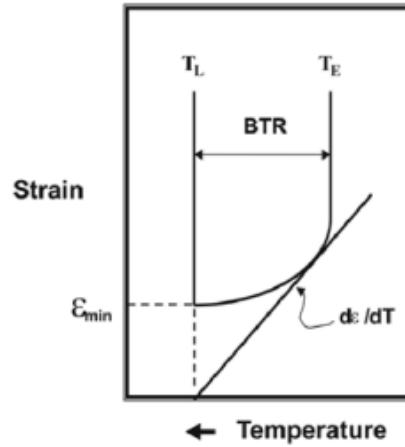


Figure 3.3 Ductility curve according to Senda’s theory⁸⁹

(2) Stress

Stress and strain are closely correlated through constitutive equation in solid mechanics. So alternative methods which based on mush zone strength for ‘ductility curve’ is proposed. Gleeble test can be used to determine the mush zone inherent strength¹⁰⁶.

(3) Strain rate

The tangent line on the deformation curve in Figure 3.3 can be transformed into the multiplication of strain rate ($d\varepsilon/dt$) and cooling rate (dT/dt) as follows: $d\varepsilon/dT = (d\varepsilon/dt)/(dT/dt)$. Rappaz proposed a critical deformation rate model which considers the influence of pressure drop caused by solid deformation and solidification shrinkage. The nucleation of a first void will take place when the local strain rate is larger than the critical deformation rate ($\dot{\varepsilon}_{p,max}$)⁹⁵.

(4) Restraints

Fixtures in weld and core in casting impose restraints on the mush zone in the process of solidification. Restraints can evolve into local strain/stress which affects solidification cracking susceptibility. Castings and welds of some structure, *e.g.* dog-bone shape workpieces, give themselves self-restraint¹⁰⁴.

3.1.5 Thermal factors

Process parameters *e.g.*, melt temperature (overheat), mold materials and preheat temperature, cooling conditions (*e.g.*, air cooling, furnace cooling) in casting, and the type of heat source (*e.g.*, laser, electrical arc, flame), input power (welding current and welding voltage), welding speed in welding are thermal factors that determine temperature field and temperature gradient field which influence the solidification cracking susceptibility with metallurgical and mechanical factors together.

3.1.6 Solidification cracking of stainless steel

Stainless steel is an important type of steels which is widely used in corrosion, marine, high temperature and cryogenic environment and always suffer solidification cracking in processing. Commercial stainless steel commonly contains at least a dozen elements. Its major elements include Fe, Ni, Cr, Si, Mn, Mo; and its minor elements include C, N, P, S, Ti, Nb, B. The phases in stainless steel include matrix *e.g.*, ferrite, austenite and martensite; and secondary phases *e.g.*, carbides, nitrides, sulfides and phosphides. Composition variation causes changes in solidification mode, solidification sequence, elements segregation and secondary phases precipitation.

Ferrite can accept more impurity elements (like S and P) than austenite. Irregular ferrite/austenite grain boundary is not in favor of the propagation of cracking cracks. Thus, a High ratio of Cr to Ni (or Cr equivalent to Ni equivalent) is good for solidification cracking resistance by forming a certain amount of ferrite like that in AISI 301 and 304. Fully austenitic stainless steels like AISI 316 is more susceptible to solidification cracking than stainless steels contain a certain amount of ferrite like AISI 304.

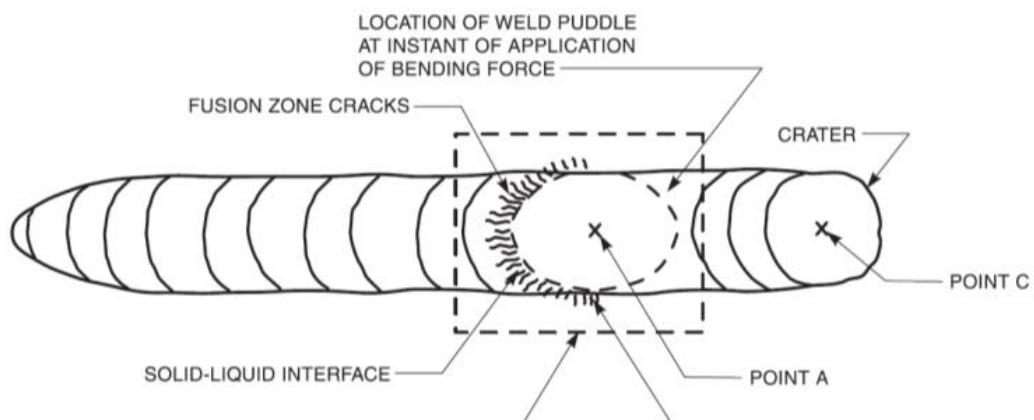
Impurity elements C, N, P and S, which tends to segregate to grains boundaries, increase cracking susceptibility. S have a strong effect on interfacial energies and is added to stainless steels to increase Marangoni effect. The extent of increasing susceptibility with impurity elements are different among different steels.

The increase of solidification cracking susceptibility with the applied strain is a nonlinear behavior.

3.1.7 Solidification cracking susceptibility test

Hundreds of test methods were proposed to qualitative and quantitatively evaluate the solidification cracking susceptibility^{92,107} of alloys. A controlled strain applied on a geometrically simple specimen is preferred for evaluation of cracking tendency, e.g. Gleebel hot ductility test^{108,109}, cast pin test^{110,111}, spot varestraint test (Sigmajig test)^{109,112}, transverse varestraint test (TVT)^{110,111,113–119}, longitudinal varestraint test (LVT)^{108,109,127–136,112,137–139,120–126}. Different researchers select one or some of them in research and modify test parameters according to their needs, and there is no standardized method for solidification cracking susceptibility test.

Longitudinal varestraint test (see Figure 3.4) is a reliable (non-standardized) quantitative solidification cracking susceptibility test of good repetition. In longitudinal varestraint test a definite bending strain is applied on the weld of sheet metals specimen, and total crack length (TCL) and maximum crack length (MCL), strain threshold, strain rate threshold and brittle temperature range (BTR) are measured as the index for solidification cracking susceptibility. A large amount of longitudinal varestraint test results are available in literature^{108,109,127–136,112,137–139,120–126}.



TOP SURFACE OF TEST WELD SHOWING LOCATION OF ARC, WELD PUDDLE, SOLID-LIQUID INTERFACE AT INSTANT OF APPLICATION OF BENDING FORCE AND WELD METAL AND HEAT-AFFECTED ZONE HOT CRACKS.

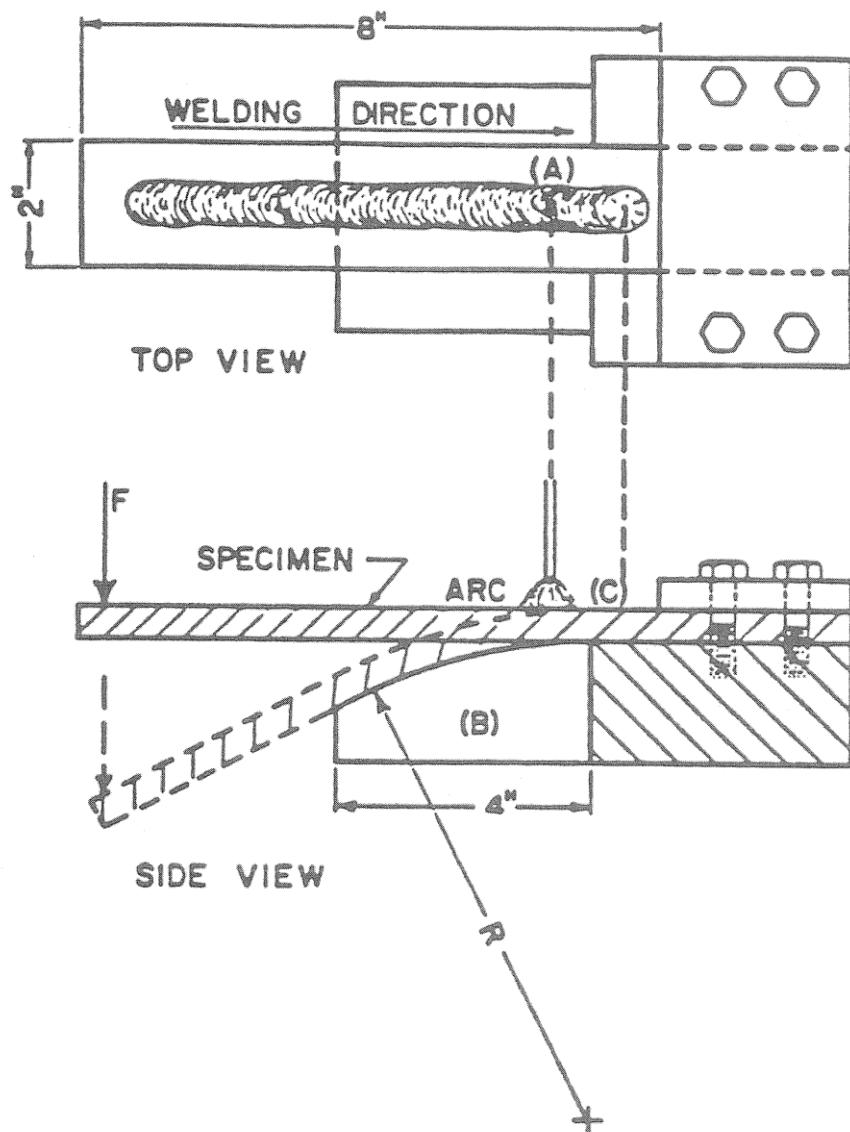


Figure 3.4 Schematic diagram for longitudinal varestraint test¹⁰⁸.

3.1.8 Machine learning for solidification cracking

In order to find the relationship between solidification cracking susceptibility, composition and processing parameters, researchers of material science commonly need the bridges of atomic/mesoscopic/micro /macro structure analysis. They explore the relationship between structure, composition, and processing parameters as well as the relationship between structure and solidification cracking susceptibility, and a lot of new terminologies and models are proposed like the ratio of chromium equivalent to nickel equivalent, primary ferritic/austenitic solidification mode, impurities content,

solidification region. Then an indirect relationship is built between them. But the main disadvantages of this method are the relationship found is always limited to specific compositions and processing parameters and most of them are qualitative descriptions. A universal relationship is rarely found, especially for complex nonlinear phenomena. Materials researchers use compositions, structures, and properties (which are called manual features in data science) to describe materials. The attractions of many data-based methods, especially the deep learning, are in that the hierarchical structure of them can complete features engineering automatically, we can obtain solidification cracking susceptibility as function of composition and processing parameters, *i.e.*, quantitative mappings from inputs to outputs, directly without the aids of structure analysis. Figure 3.5 shows schematically the difference between traditional material science route and data science route.

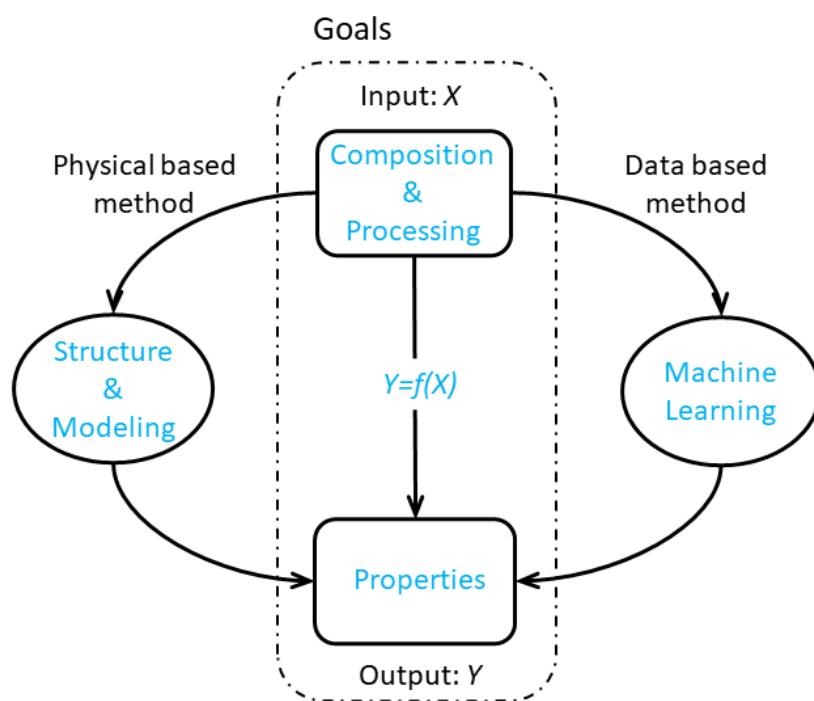


Figure 3.5 Comparison of physical-based methods and data-based methods for a materials science problem. The goal of material research is to determine the relationship between material property, composition, and processing. To achieve the goals, traditional material science completed it manually by structure analysis and computer modeling, modern data science can complete it automatically through mining the relationship between X and Y by machine learning algorithms.

Facing multiple variables nonlinear problem, data-based and data-driven statistical learning methods, *e.g.*, neural network (NN), increasingly become an alternative to physically based analytical and numerical methods. Based on reliable experimental data, neural network is able to depict alloy solidification cracking susceptibility as a function of composition and processing parameters in high-dimensional space, like that was done by Ichikawa¹⁴⁰. Ichikawa et al. employed a neural network to predict solidification cracking of welds in low-alloy steels. They assembled a small dataset of 154 samples. The input variables are chemistry (weight percent of C, Si, Mn, P, S, Ni, Cr, Mo) and welding parameters (the welding current in amperes; the welding voltage in voltages; the welding travel speed in cm/min; the weld preparation groove angle in degrees; the preheat temperature in centigrade). The output variable is solidification cracking sensitivity (0 represents not cracking and 1 represents cracking). The neural network model successfully reproduced the nonlinear relationship of solidification cracking sensitivity dependence on chemistry and welding parameters.

3.2 Phase formation prediction

Phases and their distribution in space determine the properties of a material. Predicting phase prototypes (crystal structures) of materials or phase formation in materials has been a fundamental task and a challenge for material science.

3.2.1 Amorphous phase and glass-forming ability

Amorphous ribbon alloys (AMRs) and bulk metallic glass (BMGs) extended metallic materials from typical crystalline (CR) materials to amorphous state materials. Owing to their unique mechanical^{141–143}, magnetic¹⁴⁴, catalytic¹⁴⁵, physical properties¹⁴⁶ and unique structure of lacking long-range order¹⁴⁷, amorphous alloys and metallic glasses have been widely studied and explored as a promising new material since they were first discovered in the 1960s¹⁴⁸. Two methods are commonly used to prepare amorphous alloys, *i.e.* single wheel melt-spun (cooling rate can be as high as $10^6\text{--}10^5\text{K/s}$) and copper mold casting (cooling rate in the range of $10^2\text{--}10^{-1}\text{K/s}$). Melt-spun is a rapid solidification process and its product is amorphous ribbon of about 20–100 μm in thickness and about 1mm to several mm in width. Copper mold casting is commonly used in preparing bulk metallic glasses with each dimension above 1mm.

The glass-forming ability (GFA) of an alloy, *i.e.*, the critical cooling rate below which the alloy melt undergoes nucleation and growth and forms crystal (CR), is a fundamental problem in developing new amorphous alloys. In the past five decades, many empirical criteria and physical models about glass-forming ability have been postulated. Dimensionless parameters based on liquidus temperature T_L , glass transformation temperature T_g and crystallization temperature T_x , *e.g.*, Turnbull's T_g/T_L , Lu's $T_x/(T_g + T_L)$ ¹⁴⁹ were proposed to estimate the glass-forming ability of amorphous alloys. But these parameters can only be obtained after the successful preparation of amorphous alloys and cannot be used to predict new amorphous alloys. Inoue¹⁵⁰ proposed the famous three basic empirical rules of bulk metallic glass formation (high glass-forming ability): multicomponent alloys containing three or more elements, large negative mixing enthalpy, large atomic size difference. However, Louzguine-Luzgin's research¹⁵¹ shown that many alloys satisfy all three criteria but still cannot form an amorphous ribbon with rapid solidification process. Miracle^{152,153} linked the glass-forming ability with geometric packing and put forward the efficient cluster packing model. Greer formulated the confusion model based on the complexity of the solidification path^{154,155}. Other works related glass-forming ability to deep eutectic area¹⁵⁶, Pauling electronegativity¹⁵⁷. Due to the vague physical mechanisms of glass-forming ability, developing new amorphous alloys is still mainly through time-consuming and costly trials and errors.

Sun¹⁵⁸ et al. attempted to predict the glass-forming ability of binary alloys using support vector machine classification model. Manual features, *e.g.*, atomic weights, the mixing enthalpy, atomic radii, liquidus temperatures, the fictive temperatures, the difference in liquidus temperature were used as input descriptors (features, attributes). A dataset containing 31 binary alloy systems was assembled. Their results indicated that the difference between liquidus temperature and the fictive temperature is an important feature for predicting glass-forming ability of binary alloys.

Ward⁵⁸ et al. predicted the glass-forming ability of ternary alloys using random forest classification models with 145 'general purpose' attributes derived from the periodic table. A dataset of about 6000 entries was collected from handbooks¹⁵⁹. The model created could reach an accuracy of 90% under 10-fold cross-validation. Based on the work, Ward⁵⁷ et al. assembled a larger dataset of 8000 entries from literature. This new

dataset includes bulk metallic glass, amorphous ribbon, and crystalline data. The 145 ‘general purpose’ attributes plus cluster packing efficiency attributes, nearest special clusters, mean packing efficiency, and proximity to crystalline compound attributes. 201 attributes were used as input. The glass-forming ability (crystalline/amorphous ribbon/bulk metallic glass), the critical casting diameter, and the supercooled liquid range were the output. They employed these machine learning models to optimize commercial Zr-based bulk metallic glasses.

3.2.2 Multi-principal element alloys

Multi-principal element alloys (MPEAs), which are also known as high-entropy alloys (HEAs), and concentrated solid solution alloys (CSSs), extended metallic materials from corner and edge regions to the center regions of multi-component phase diagrams^{160–162}. Researchers attributed the formation of single-phase multi-principal element alloys to four core effects: high entropy, sluggish diffusion, high lattice distortion, and cocktail effects. There are challenges in understanding the evolution of phases in multi-principal element alloys. Early work on multi-principal element alloys was mainly focused on 3-d transition metals. The so-called Cantor alloy containing FeCrMnNiCo showed high strength and high ductility over a wide range of temperatures. With the interest to develop super high temperature alloys (next generation superalloys), a few researchers focused their attention on refractory multi-principal element alloys containing Nb, Hf, W, Cr, Zr¹⁶³. While many new equiatomic multi-principal element alloys have been investigated in the past decade, researchers now realized the need to develop non-equiatomic multi-principal element alloys (and with additions of minor elements) to improve various properties. New subclasses in multi-principal element (MPE) alloys such as eutectic MPE alloys, MPE superalloys, TWIP- and TRIP-type MPE steels, dual-phase MPEAs, light MPE alloys, and MPE oxides have been developed in recent years^{164–166}.

Researchers attempted to predict the phase formation in multi-principal element alloys through CALPHAD approaches¹⁶⁵. Thermo-Calc has built a new thermodynamic database TCHEA for multi-principal element alloys. Researchers also attempted to predict it through empirical criteria based on Hume-Rothery rules, thermodynamic parameters (*e.g.*, enthalpy of mixing ΔH_{mix} and entropy of mixing ΔS_{mix}), topological parameters (*e.g.*,

δ parameter and mismatch entropy ΔS_σ), Kinetic approach (viscosity), Pettifor map. Conventional machine learning was also be explored to predict the phase formation in multi-principal element alloys¹⁶⁷⁻¹⁷¹.

Li¹⁷¹ et al. proposed a support vector machine model that is capable of distinguishing stable body-centered cubic (BCC), face-centered cubic (FCC) phases, and the remaining phases out of the 322 as-cast samples with a cross validation accuracy over 90% after training and testing. The thermodynamic parameters (mixing enthalpy, mixing entropy, average valence electron concentration, average of melting points of constituent elements, atomic size difference, standard deviation of electronegativity) were chosen as input features.

Zhou¹⁶⁸ et al. used 13 features, *e.g.* mean atomic radius, atomic size difference, average of melting points of constituent elements, standard deviation of melting points, mixing enthalpy, standard deviation of mixing enthalpy, mixing entropy, average of electronegativity of constituent elements, standard deviation of electronegativity, average valence electron concentration, standard deviation of valence electron concentration, mean bulk modulus, standard deviation of bulk modulus, derived from physical models and empirical criteria to build artificial neural network which can distinguish solid solution, intermetallic, and amorphous phase.

Islam¹⁶⁷ et al. utilized five empirical materials descriptors, *i.e.* the valence electron concentration (VEC), difference in the Pauling negativities, atomic size difference, mixing enthalpy, and mixing entropy to build a neural network model to distinguish single-phase solid solution, amorphous structure and intermetallic compounds. On average, a predictive accuracy higher than 80% could be reached.

3.2.3 Phase prototypes of inorganic substances

Though density functional theory (DFT) and force field models have been successfully exploited to predict crystal structures of materials. Their applications are limited by computational time and cost. Recently, conventional machine learning models were also exploited to predict inorganic materials' structures. Graser¹⁷² et al. assembled phase prototype dataset from Pearson's Crystal database and used random forest to classify them. To take all data in model, labels of entries number below a threshold were

relabelled as “Other”. By selecting the threshold from 150 to 10, the classification performance metrics (accuracy, precision, and recall) vary as follows: accuracy ranged from 0.97 to 0.85; average precision ranged from 0.86 to 0.79, while average recall ranged from 0.73 to 0.54 for threshold from 150 to 10, respectively.

4 Application of deep neural network and pre-training to predicting solidification cracking susceptibility with small dataset

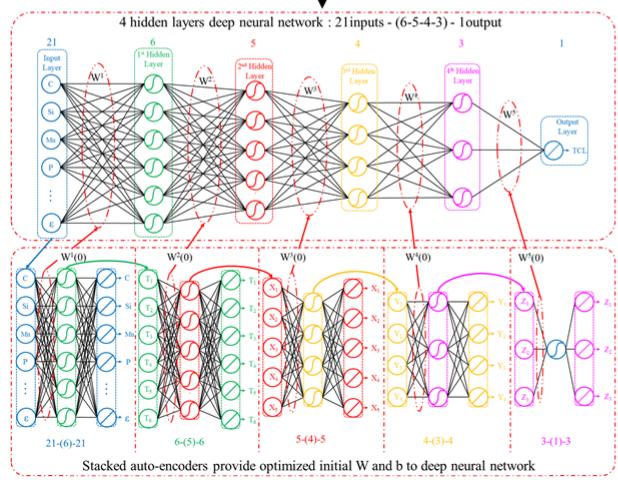
Accurately predicting solidification cracking susceptibility of alloys is challenging. Though the factors that affect solidification cracking are very clear (*i.e.*, descriptor is not a problem), the function is nonlinear, and experimental data are heterogeneous and insufficient. How to build machine learning models of good performance and interpretability based on our existing small dataset is the subject of this chapter. Figure 4.1 shows the workflow of predicting solidification cracking susceptibility using deep neural network (DNN).

1 Small Dataset

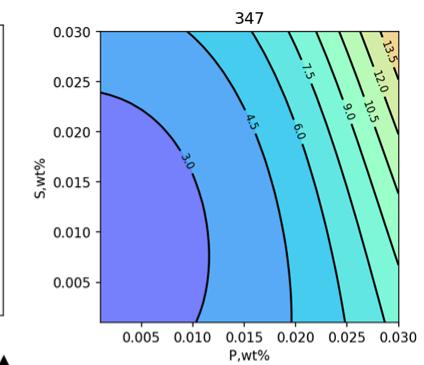
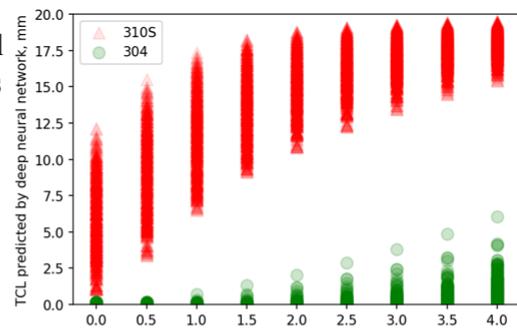
	code	C	Si	Mn	P	S	Cr	Ni	Mo	H	V	B	Th	I	U	Ve	Strain	TCL	MCL	note
0	316NG-A	0.0100	0.48	1.61	0.024	0.019	17.33	10.62	2.09	0.0600	...	0.0	0.0	3.18	100	12.0	4.23	4.0	1.50	0.19
1	316NG-B	0.0110	0.58	1.06	0.032	0.013	16.95	10.50	2.15	0.0780	...	0.0	0.0	3.18	100	12.0	4.23	4.0	1.10	0.18
2	316NG-C	0.0100	0.46	1.09	0.021	0.001	17.40	11.50	2.88	0.1050	...	0.0	0.0	3.18	100	12.0	4.23	4.0	0.90	0.15
484	K17	0.0140	0.33	1.73	0.026	0.007	17.90	9.50	0.00	0.0460	...	0.0	0.0	5.00	70	16.0	1.25	1.2	0.24	Nan ref17fig14
485	SUS316	0.0600	0.75	0.94	0.026	0.007	18.30	9.40	0.00	0.0160	...	0.0	0.0	5.00	70	16.0	1.25	1.2	0.00	Nan ref17fig14
486	SUS316	0.0700	0.66	1.01	0.020	0.006	16.70	12.40	2.38	0.0200	...	0.0	0.0	5.00	70	16.0	1.25	1.2	1.47	Nan ref17fig14

487 rows × 25 columns

2 Pre-train & Fine-tune DNN



4 Material Properties Maps



3 DNN of High Accuracy

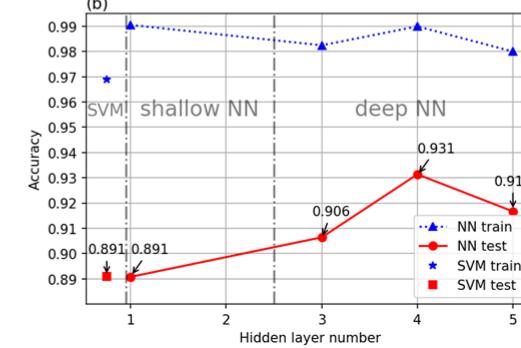


Figure 4.1 The workflow of predicting solidification cracking susceptibility using deep neural network (DNN).

4.1 Dataset

Longitudinal varestraint test (LVT, see Figure 3.4) is a reliable quantitative method for measuring solidification cracking susceptibility of alloys. Compared with other solidification cracking susceptibility tests, the amount of longitudinal varestraint test results in literature is significant. That is why we chose longitudinal varestraint test data in published literature to compile our solidification cracking susceptibility dataset. Different researchers used different sample thickness, welding process parameters and applied strains in their tests, so it is difficult to compare results from different researchers and get a unified conclusion. However, it is not a big problem for machine learning: when we add new variables into dataset, the more variation in data space, the more opportunities for machine learning models to discover the hidden nonlinear relationships.

In this study, a small dataset including about 600 longitudinal varestraint test results on stainless steels was collected from dozens of published literatures^{108,109,127–136,112,137–139,120–126}. The dataset contains longitudinal varestraint test data of ferritic, austenitic, duplex, precipitation hardening (wrought / casting) stainless steels. It encompasses four types data: composition information of 16 elements in stainless steels (weight percent of C, Si, Mn, P, S, Ni, Cr, Mo, N, Nb, Co, Cu, Al, Ti, V, B), sample thickness (Th, mm) and welding parameters (welding current: I, A; welding voltage: U, V; welding velocity: Ve, mm/s), restraints (applied strain: ϵ , %) and solidification cracking susceptibility index total crack length (TCL, mm). The elements contents which were not specified in literature were manually filled as follows: Al = 0.02, N = 0.02, Mo = 0, Nb = 0, Cu = 0, V = 0, B = 0. Entries that miss five or more variables' data were not used in training and testing models. The final dataset used for machine learning contains 487 entries. Figure 4.2 shows the histograms of the 22 variables in the dataset.

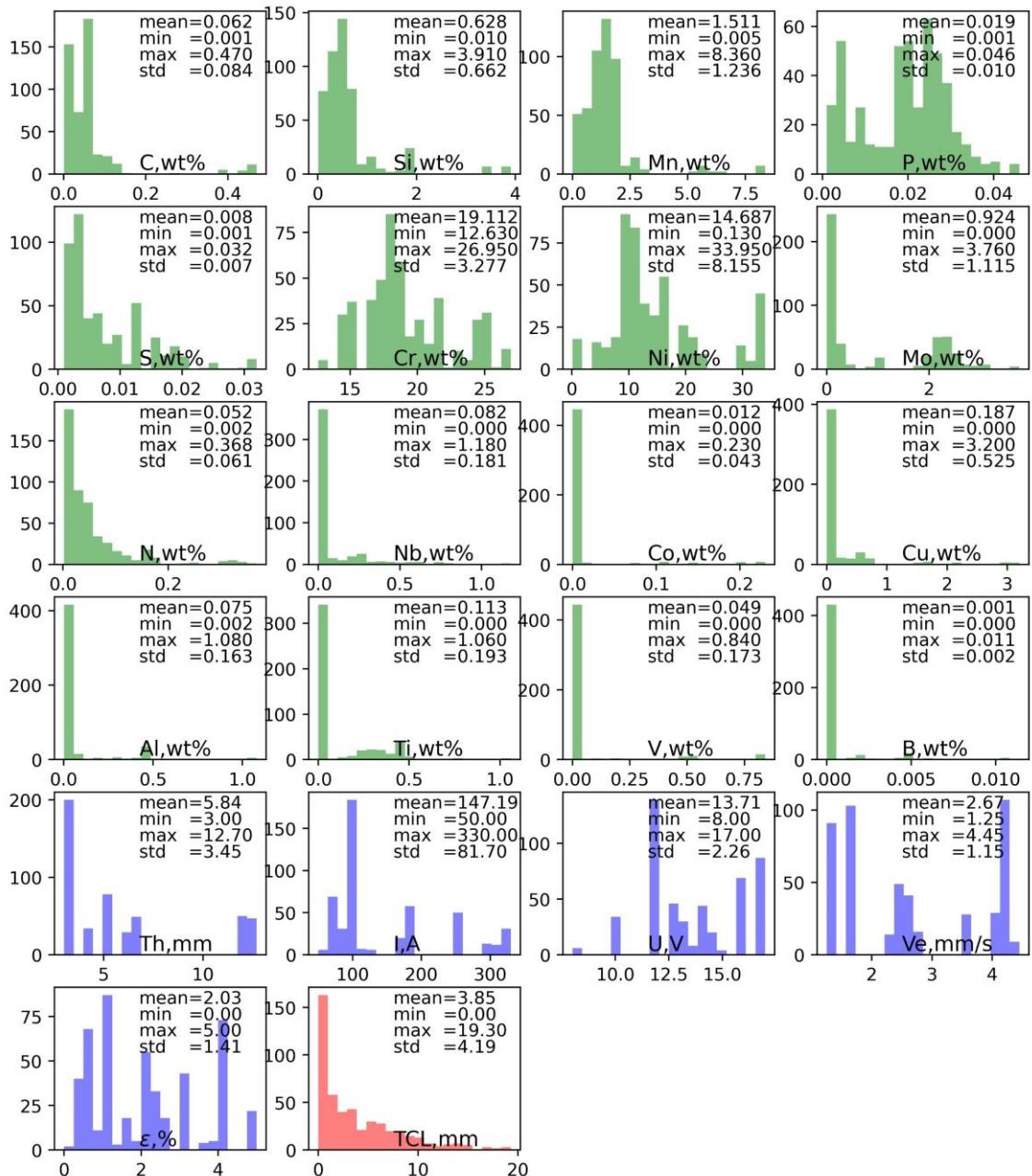


Figure 4.2 Histograms of the 22 variables in the dataset (487 samples in total). Mean, minimum, maximum, std are also shown in the histograms.

Pearson correlation coefficient R is usually used as a measure of linear dependence of two random variables. If variables X and Y are N -dimensional vectors, then the Pearson correlation coefficient $R(X, Y)$ of them is defined as follows:

$$R(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

$$\text{Cov}(X, Y) = \frac{1}{N-1} \sum_{i=1}^N (X_i - \mu_X)(Y_i - \mu_Y)$$

$$\mu_X = \frac{1}{N} \sum_{i=1}^N X_i$$

$$\mu_Y = \frac{1}{N} \sum_{i=1}^N Y_i$$

$$\sigma_X = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \mu_X)^2}$$

$$\sigma_Y = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - \mu_Y)^2}$$

where $\text{Cov}(X, Y)$ is the covariance of X and Y , μ_X and μ_Y are the mean values of X and Y respectively, σ_X and σ_Y are the standard deviation values of X and Y respectively.

Figure 4.3 shows the scatter matrix of total crack length (TCL) and the other 21 variables. The Pearson correlation coefficients R_s of total crack length and the other 21 variables are also shown in the figures. If there are obvious patterns or linear relationship between them, we can easily notice them from the scatter plot. They show no apparent patterns existed in the scatter matrix. The absolute values of correlation coefficients R_s are well below 0.5 which also indicates no linear relation existed between total crack length and the other 21 variables, *i.e.*, the relationship is nonlinear and cannot be expressed by a simple function.

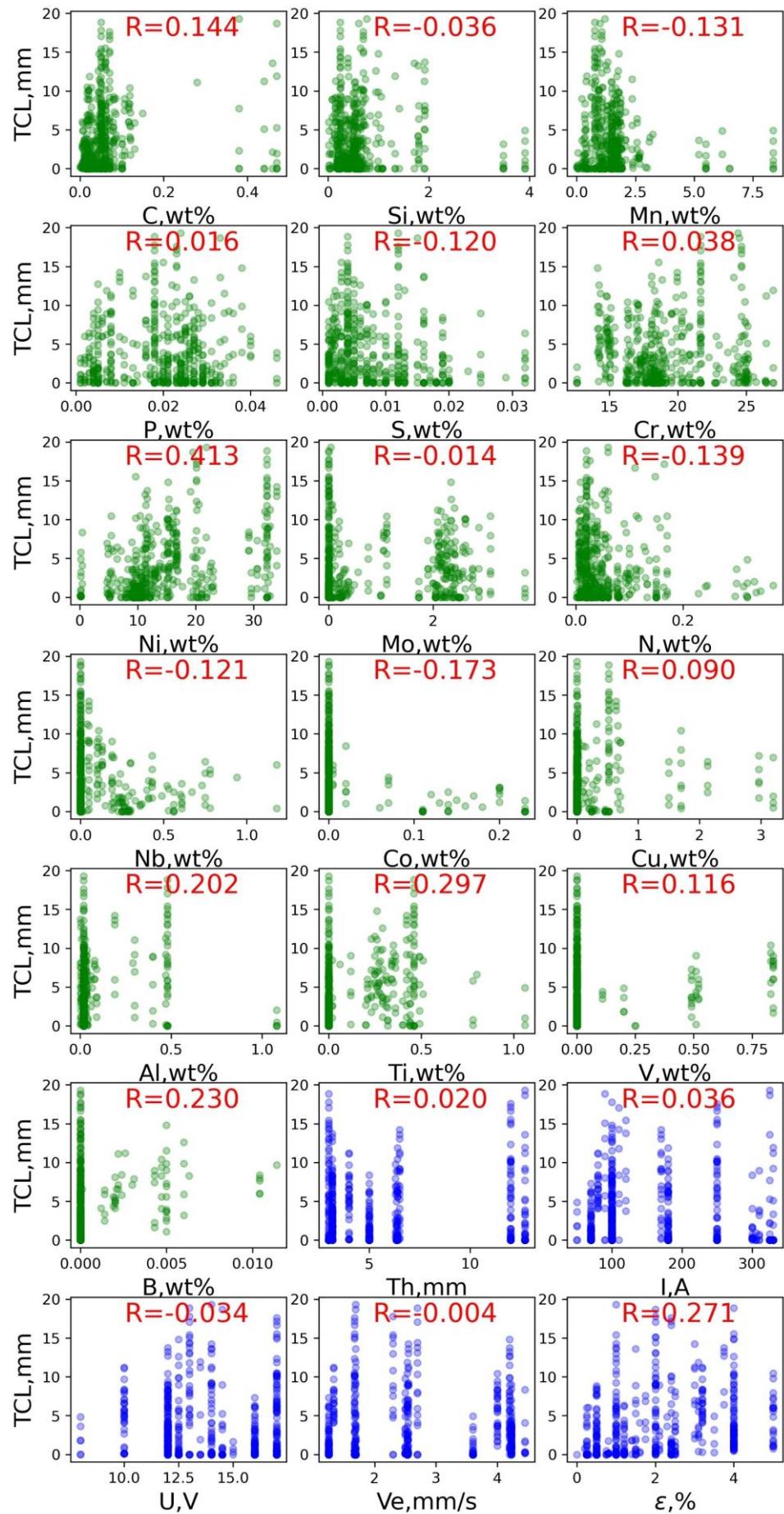


Figure 4.3 Scatter matrix of total crack length (TCL) and other 21 variables, the values of Pearson correlation coefficients Rs are also shown in the figures.

4.2 Data pre-process & dataset division

4.2.1 Data pre-process

The data used in shallow neural network and fully connected deep neural network were normalized before training. The input variables X were normalized between -0.5 and +0.5, as follows:

$$X_N = \frac{X - X_{min}}{X_{max} - X_{min}} - 0.5$$

Where X_N is the normalized value of X which has maximum and minimum values given by X_{max} and X_{min} respectively.

The target variables Y were normalized between -1 and +1, as follows:

$$Y_N = 2 * \frac{Y - Y_{min}}{Y_{max} - Y_{min}} - 1$$

Where Y_N is the normalized value of Y which has maximum and minimum values given by Y_{max} and Y_{min} respectively.

4.2.2 Training & testing datasets division

To achieve unbiased training/testing data division, all data were divided into groups first. In a group, only one or two variables vary. Then training data and testing data were randomly chosen in a 2 to 1 ratio. 2/3 data set (324 samples) was used for training. The remaining 1/3 data set (163 samples) was kept unseen by neural network in training to test the generalization performance (prediction accuracy on the unseen data set) of the trained neural network.

4.3 Machine learning models

4.3.1 Fully connected shallow & deep neural networks

The fully connected neural networks used in this study consisted of 21 input neurons (one neuron corresponds to one input variables), one to several hidden layers, a few

hidden neurons (the neurons number is variable) in each hidden layer, and one output neurons. The transfer function of hidden layer is hyperbolic tangent function $\tanh(x)$, and the mathematical expression of the $\tanh(x)$ is as follows:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

The transfer function of output layer is linear function $y = x$.

In this study, we tested shallow neural networks of neuron numbers from 1 to 35.

We also attempted three hidden layers fully connected deep neural network of 21-(4-3-2)-1 and 21-(5-4-3)-1 structure, four hidden layers deep neural networks of 21-(5-4-3-3)-1 and 21-(6-5-4-3)-1 structure, and five hidden layers deep neural networks of 21-(6-5-4-3-3)-1 and 21-(5-4-3-3-3)-1 structure. The numbers in bracket represent neuron numbers of hidden layers in sequence. Initiation is vital in training deep neural network (all you need is a good initiation¹⁷³). A selection of initiation values for stacked auto-encoder were used to generate different initiation conditions for deep neural networks.

Figure 4.4 shows the schematic diagram of the process of fully connected deep neural network initiation by stacked auto-encoder (SAE). We used the four hidden layers deep neural network of 21-(6-5-4-3)-1 structure (the upper half of the diagram) as an example. C, Si, Mn, ..., ε are 21 input variables of the deep neural network. Total crack length (TCL) is the output variable of the deep neural network. $W^j(0)$ is the initial weights matrix of deep neural network's j^{th} layer. Auto-encoder has the same input and output variables, i.e. it is a symmetry network. The neuron number of each autoencoder is equal to the neuron number of deep neural network's corresponding layer. Five auto-encoders (AE, the lower half of the diagram) of 21-(6)-21, 6-(5)-6, 5-(4)-5, 4-(3)-4, 3-(1)-3 structures were constructed in sequence. The first auto-encoder corresponds to the first hidden layer of deep neural network, and so on. Long arrows represent the directions of data transfer. The first auto-encoder took the input of deep neural network as its input and output. After training the weights and biases values of an auto-encoder's first hidden layer were transferred to the deep neural network's corresponding layer as its initial values, and the output of the previous auto-encoder's hidden layer were transferred to the next auto-encoder as its input and output. This process was carried on layer by layer and one autoencoder by one autoencoder. This unsupervised layer-wise pre-training

provide deep neural network with optimized initial weights and biases values.

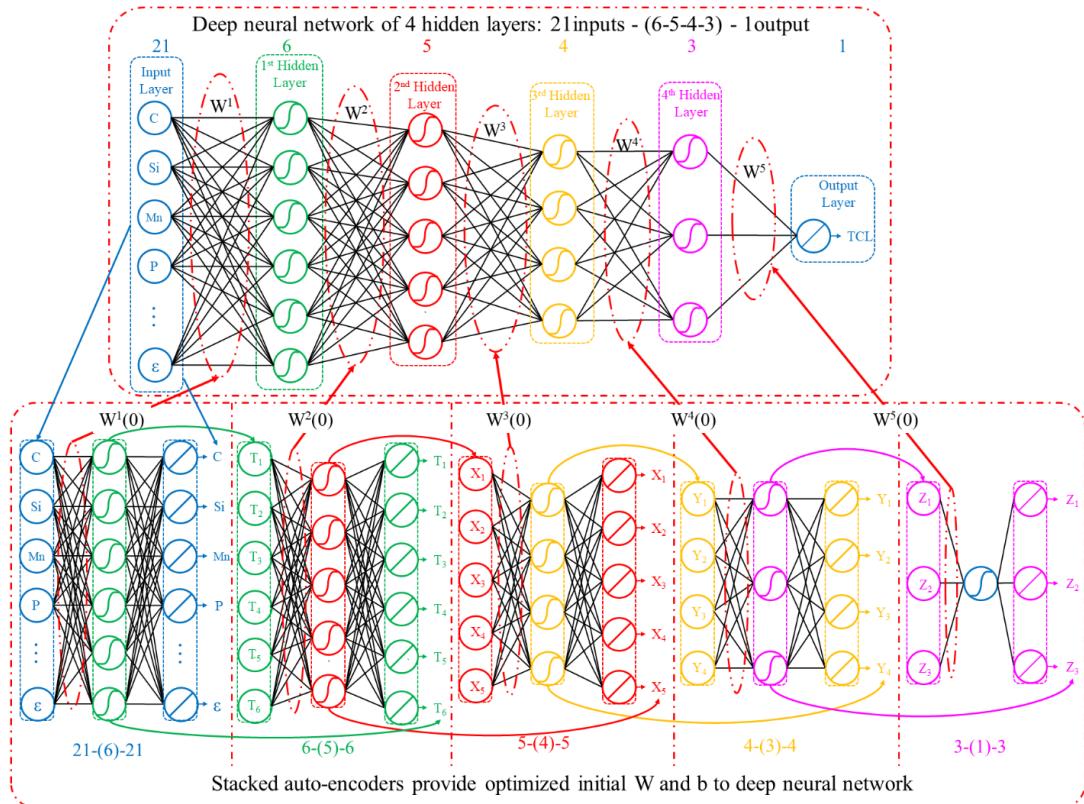


Figure 4.4 Schematic diagram of deep neural network initiation by stacked auto-encoder.

4.3.2 Support vector machine

Support vector machine was also trained and tested in this study to validate the accuracy advantage of deep neural network. In this study, support vector machine regression with linear/Gaussian/ (order 2 and 3) polynomial kernels was used and the hyper-parameters of the support vector machine were optimized with Bayesian optimization.

4.3.3 Tree-based models

Decision tree and random forest models were also trained and tested in this study to validate the accuracy advantage of deep neural network and to enhance the interpretability of our machine learning models (through analysing the relative importance of variables). The impurity criterion used in the tree-based models is ‘Gini’ index. We regulated the maximum depth of a tree (20 was used in our work) and the maximum decision tree number in the random forest (100 was used in our work).

4.4 Training details

Each support vector machine / shallow neural network / deep neural network configuration was trained more than 100 times using different random seeds. The optimal support vector machine, shallow neural network and deep neural network were chosen to make predictions and comparisons.

All machine learning works were carried out in Matlab 2018a with its statistics and machine learning toolbox and neural network toolbox. Due to small dataset and pre-training, the time required to train one deep neural network configuration with one random seed on a personal computer (@2.6GHz processor, 16GB RAM) varied from several seconds to dozens of seconds, which is comparable with that of shallow neural network and support vector machine and is obviously shorter than that of training deep neural network for image recognition where many GPUs and many hours or even many days are needed. The training time for tree-based models were quite short (only several seconds) and can be neglected.

4.5 Training and testing results

Table 4.1 The best testing accuracies of five models used in prediction.

Model	Decision tree	Random forest	Support vector machine	Shallow neural network	Deep neural network
Testing accuracy	0.75	0.88	0.89	0.89	0.93

After training, Pearson correlation coefficients Rs of the target values and decision tree / random forest / support vector machine / shallow neural network / deep neural network prediction values were calculated as the index of training/testing accuracy.

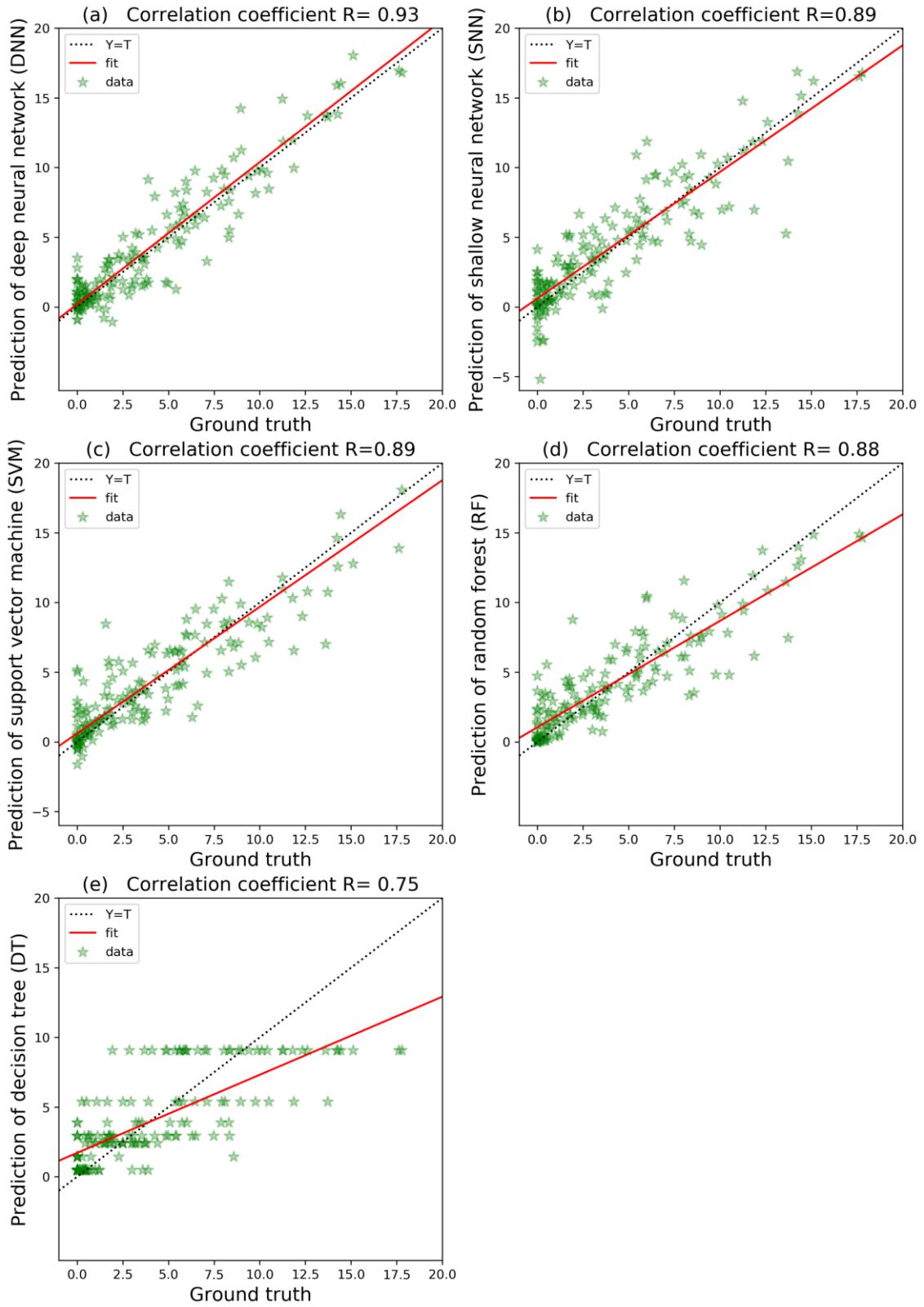


Figure 4.5 The regression analysis on the unseen testing dataset between experimental data (horizontal axis, T values) and neural network predictions (vertical axis, Y values). The black dotted line is $Y=T$. Pearson correlation coefficients R_s are also shown in the figures. (a) deep neural network's predictions, (b) shallow neural network's predictions, (c) support vector

machine's predictions, (d) random forest's predictions, and (e) decision tree's predictions.

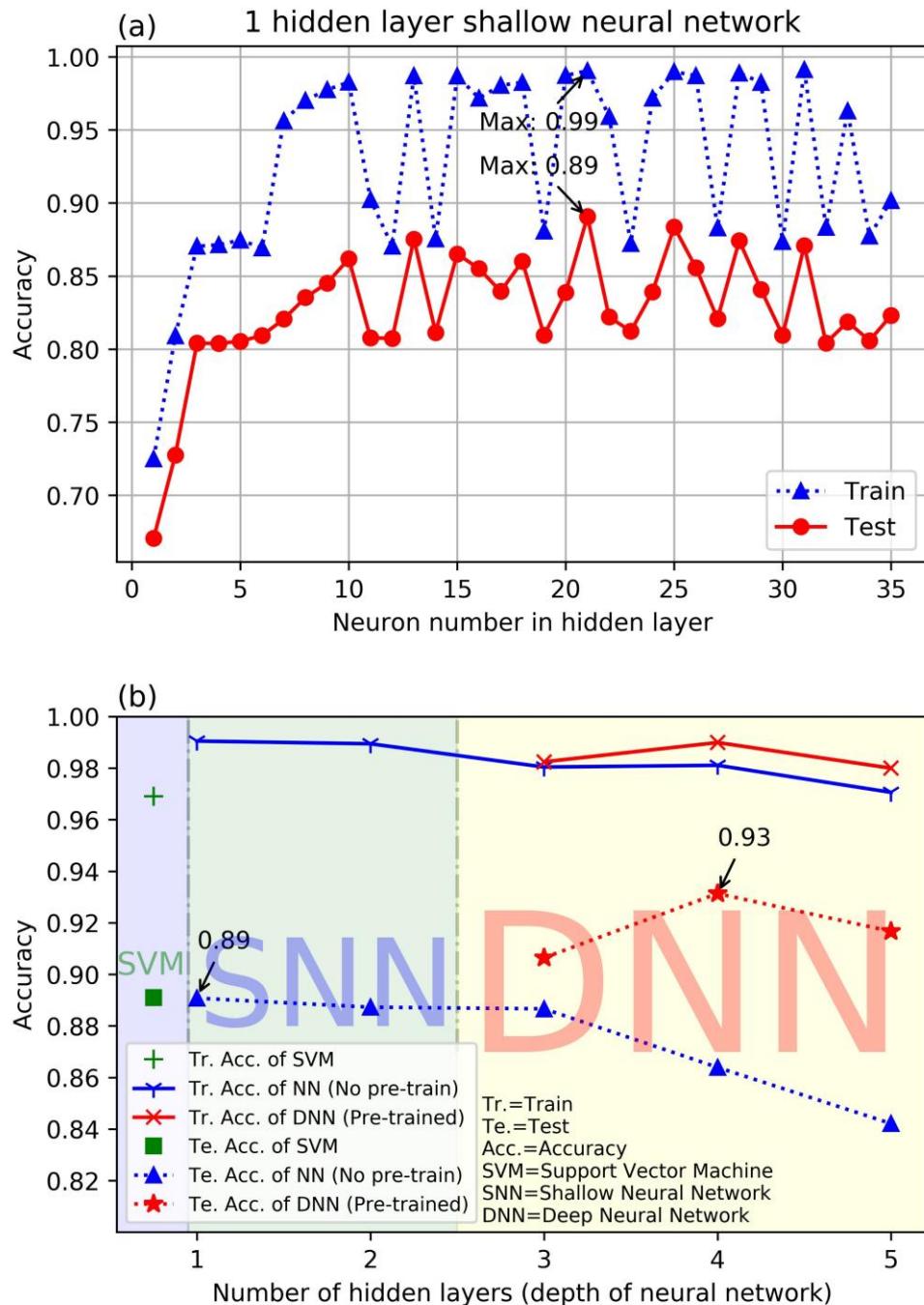


Figure 4.6 (a) Shallow neural network's training and testing accuracies as a function of neuron number. (b) Neural networks' training and testing accuracies as a function of their depth. The support vector machine's accuracies are also shown in figure.

Figure 4.5 shows the linear regression of target and prediction on testing data set, and the accuracy (Pearson correlation coefficient R) is also shown on figures. Table 4.1 shows the best accuracies of five machine learning models on testing dataset. Random forest's

accuracy (0.88) is better than that of a single decision tree (0.75) and is close to that of shallow neural network (0.89) and support vector machine (0.89). Deep neural network has the highest prediction accuracy (0.93) in the five models.

Figure 4.6(a) shows the one hidden layer neural network's training accuracy and testing accuracy as functions of the neuron number in the hidden layer. The shallow neural network's training accuracy increases from 0.73 to 0.98 and testing accuracy increases from 0.67 to 0.86 when the neuron number increases from 1 to 10. When the neuron number increases from 11 to 35, the training/testing accuracy fluctuates, and the best training and testing accuracies are achieved when the neuron number is 21 in this case. This is consistent with other researchers' results: when we increase the neuron number in the hidden layer, the training accuracy of a neural network easily reaches very a high score *e.g.*, 0.99, but the testing accuracy has a limit. That is one of the main reasons why researchers explore the deep neural network, for higher testing accuracy, which represents the real learning ability (the training accuracy can be perceived as the memory ability). The best training accuracy of shallow neural networks is 0.99, and the best testing accuracy is 0.89. So, the shallow neural network of 21-(21)-1 structure was chosen as the final optimal shallow neural network that was used for comparison in the next step.

Figure 4.6(b) shows the optimal training and testing accuracy of some models we tested in this study. When three deep neural networks of three to five hidden layers were trained with random initiation, they show lower testing accuracy than shallow neural networks and support vector machine. But three deep neural networks show higher testing accuracy than other models through pre-training and fine-tuning. Figure 4.6(b) also shows the testing accuracy of five hidden layers deep neural network is lower than that of four hidden layers deep neural network. This is not expected as deeper neural network should be more expressive and bring better performance according to the experience of deep neural network in image recognition, the possible cause of the behavior could be that the deep neural network structure is not optimal and the small dataset lacks enough training data to determine the optimal parameters of five hidden layers deep neural network.

The training accuracy of three deep neural networks and shallow neural networks are

close but higher than support vector machine. The four hidden layers deep neural networks of 21-(6-5-4-3)-1 structure shows the best testing accuracy 0.93, which is 0.04 higher in testing accuracy than that of the optimal support vector machine and shallow neural network *i.e.* 0.89.

4.6 Validation of solidification cracking susceptibility prediction

To further validate the generalization ability of the optimal deep neural network, their prediction results (based on input data that are not part of our training/testing dataset) are compared with other solidification cracking susceptibility indicators, *e.g.* brittle temperature range BTR (TCL has good linear correlation with BTR¹⁷⁴) and metallurgical experience.

Arata *et al.*¹¹⁹ measured the brittle temperature range of 17 stainless steels of different compositions using transverse varestraint test experiments. Figure 4.7 shows the comparison between Arata's results and total crack length values predicted by our optimal models. We can see a better linear relationship between brittle temperature range and total crack length predictions by our optimal deep neural network (correlation coefficient R=0.944) than that of shallow neural network (R=0.830) and support vector machine (R=0.883). This further confirms that deep neural network has a higher generalization performance than shallow neural network and support vector machine.

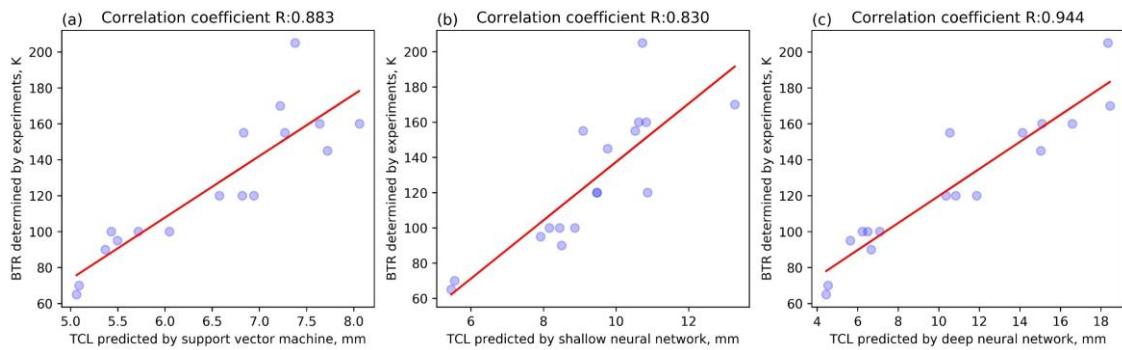


Figure 4.7 Regression analysis between total crack length predictions and brittle temperature range experimental results. (a) predictions of support vector machine, (b) predictions of shallow neural network and (c) predictions of deep neural network.

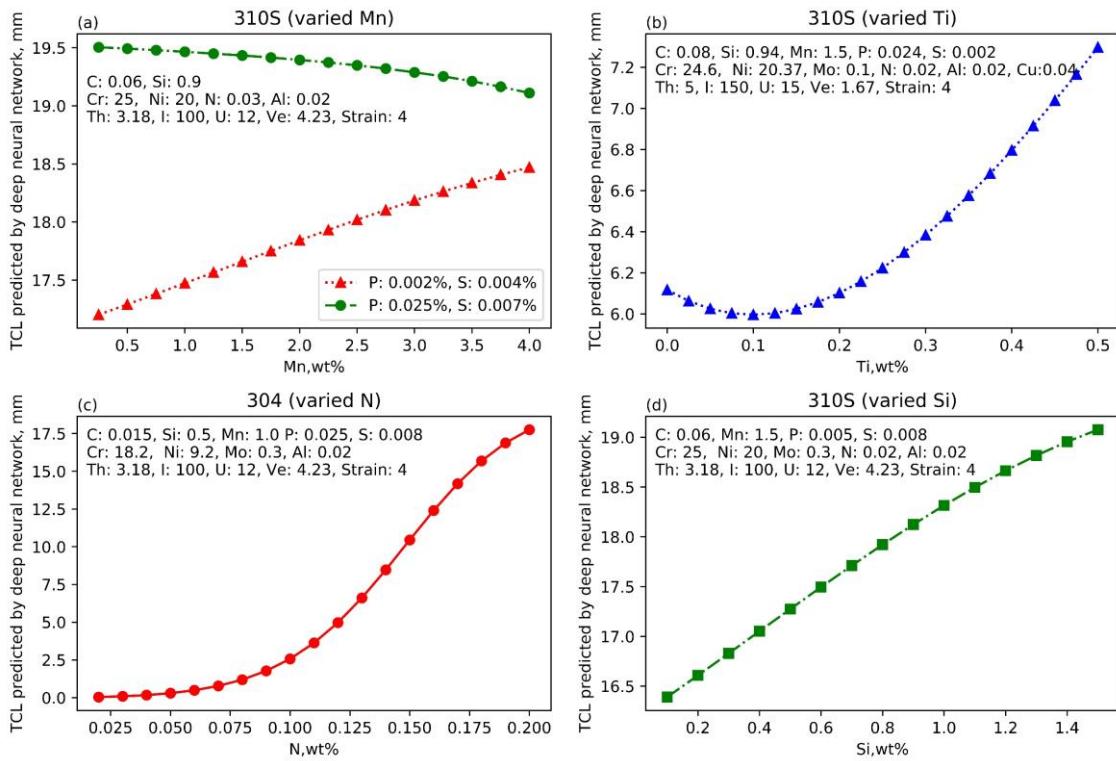


Figure 4.8 Predicted total crack length as functions of (a) Mn, (b) Ti, (c) N, (d) Si, the compositions used in prediction are shown in the figures.

Arata *et al.*¹¹⁴ reported that Mn addition to SUS310S with 0.002wt% P and 0.004wt% S exerts a harmful effect on cracking resistance under 4% applied-strain conditions using transverse varestraint test. On the other hand, under the same testing condition, Mn addition to SUS310S with 0.025wt% P and 0.007wt% S improves solidification cracking resistance. Prediction was carried out for the same SUS310S compositions published in Matsuda's paper. The two curves of Figure 4.8(a) represent the predictions of deep neural networks. The opposite tendency of the two curves verifies these behaviors: interactions between elements P, S, Mn causes final opposite solidification cracking susceptibility dependence on Mn.

Matsuda's research¹¹¹ shows that adding Ti to SUS310S of 0.024wt% P and 0.002wt% S improve solidification cracking resistance until reaching an optimal Ti content, then excessive Ti content becomes harmful to solidification cracking resistance. The prediction of deep neural network about the effect of Ti addition on solidification cracking susceptibility shown in Figure 4.8(b) is consistent with Matsuda's description.

Matsuda *et al.*¹¹³ also reported that as the N content gradually increased to about 0.2wt% in SUS 304 weld metal, the solidification cracking susceptibility increased and finally was

similar to that of SUS310S weld metal. The prediction of deep neural network shown in Figure 4.8(c) again reproduced this metallurgical experience.

Arata *et al.*¹¹⁷ reported that Si increases solidification cracking susceptibility in fully austenitic stainless steel SUS310S, for the element Si was likely to segregate to boundaries. Our deep neural network reproduced this behavior well, as shown in Figure 4.8(d).

Those validations using metallurgical experience confirm the correctness and generalization ability of our optimal deep neural network.

4.7 Visualization of prediction and interpretability

4.7.1 Neural network equations

The mathematical expressions of the optimal one hidden layer shallow neural network of 21inputs-(21)-1output structure are listed as follows:

$$h_i^{(1)} = \tanh \left(\sum_{j=1}^{j=21} W_{ij}^{(1)} X_j + b_i^{(1)} \right), i = 1, 2, \dots, 21$$

$$Z = h_1^{(2)} = \sum_{j=1}^{j=21} W_{1j}^{(2)} X_j + b_1^{(2)}$$

where X is the normalized input vector, Z is the normalized output value, $h_i^{(k)}$ ($k=1$ and 2) is the output of the k^{th} layer, $W_{ij}^{(k)}$ and $b_i^{(k)}$ are the weight and bias matrix of the k^{th} layer.

The mathematical expressions of the four hidden layers 21inputs-(6-5-4-3)-1output deep neural network are listed as follows:

$$h_i^{(1)} = \tanh \left(\sum_j W_{ij}^{(1)} X_j + b_i^{(1)} \right)$$

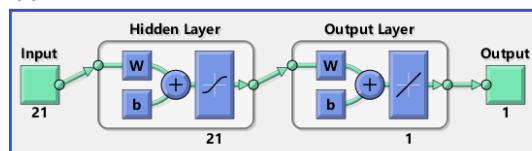
$$h_i^{(k)} = \tanh \left(\sum_j W_{ij}^{(k)} h_j^{(k-1)} + b_i^{(k)} \right), k = 2, 3, 4$$

$$Y = h_i^{(5)} = \sum_j W_{ij}^{(5)} h_j^{(4)} + b_i^{(5)}$$

where X is the normalized input vector, Y is the normalized output value, $h_i^{(k)}$ ($k=1, 2, \dots, 5$) is the output of the k^{th} layer, $W_{ij}^{(k)}$ and $b_i^{(k)}$ are the weights and bias matrix of the k^{th} layer.

The structure schematic diagram of the optimal shallow and deep neural network are shown in Figure 4.9 which demonstrates the optimal 21-(6-5-4-3)-1 deep neural network has a more compact structure (the hidden layer of deep neural network is narrower than that of shallow neural network). The weights and biases values of the optimal shallow and deep neural network are shown in Figure 4.10 and Figure 4.11 respectively. The optimal 21-(6-5-4-3)-1 deep neural network has fewer trainable parameters (210 trainable parameters) than the optimal 21-(21)-1 shallow neural network (484 trainable parameters). Neural network of fewer trainable parameters has less risk of overfitting. Combining neural network equations with the determined parameters together, the fitted functions can be easily used to estimate solidification cracking susceptibility dependence on compositions and strains (small strains can be utilized to predict solidification cracking susceptibility in casting and large strains can be chosen to predict solidification cracking susceptibility in welding and additive manufacturing) when the time consuming and expensive experimental results are not available, for example to select the optimal alloy composition, make comparison with experiment results, and so on. The following sections will provide several examples of applications. However, the applications of the machine learning models are not limited to those examples.

(a) shallow neural network



(b) deep neural network

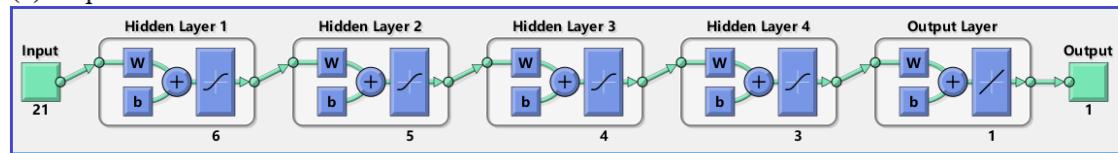


Figure 4.9 (a) The optimal shallow neural network's structure; (b) the optimal deep neural network's structure.

$W^{(1)}$	$b^{(1)}$
0.122305 0.041682 0.184067 0.024049 0.418877 -0.34953 0.235245 -0.67092 0.333255 0.580338 -0.36198 0.059561 0.23423 0.036927 0.319041 -0.27219 -0.51765 0.70446 -0.87996 -0.4319 -1.48351 0.044383	
-0.66465 0.112314 0.042773 0.64501 -1.47622 -0.36467 0.269783 -0.33558 0.021395 -0.26980 0.215942 -0.55381 -0.29275 0.023235 0.030957 -0.236 0.530267 1.158221 0.114724 0.78394 -1.66015 -0.60883	
0.041266 -0.54996 0.180791 0.629372 0.842539 -0.53402 -1.19575 -0.84247 0.438974 0.870567 0.526955 -0.6038 0.438601 0.179276 -0.43377 -0.68294 0.37247 -0.37711 0.580802 1.03805 -1.34166 -0.19311	
0.080763 0.560087 -0.26856 0.692645 -0.39051 -0.86796 1.90259 -0.89397 0.84932 -0.18147 -0.36584 -0.91564 -0.00696 -0.32926 0.570047 0.332912 -0.28026 0.205757 -0.0648 -0.45187 0.296556 0.548481	
-1.05444 -0.63805 0.2892 -0.95951 -1.22004 -0.3739 0.265728 0.168355 -0.34987 -0.92085 0.867467 0.449884 0.490472 0.633988 -0.4911 -0.7954 0.372266 0.719339 0.855423 0.076897 -1.25127 -0.98911	
-0.03402 -0.00999 0.004644 -0.01438 -0.0488 0.033796 0.021611 -0.00281 -0.04021 -0.05086 -0.07214 -0.0864 -0.08767 -0.03372 -0.06301 -0.03368 -0.06352 -0.08025 -0.0095 -0.01744 0.015923 0.138105	
1.094645 0.887032 -1.19865 0.594914 0.196927 -0.78111 0.65782 -0.89236 -0.04042 -0.13088 0.152894 0.102398 0.215653 0.377847 0.10433 0.004747 0.041248 0.633571 -0.79601 0.0503 1.000678 0.6121254	
-1.1469 -0.7809 -0.0996 -0.53638 1.336111 -0.47459 0.58343 0.139913 0.281488 0.325615 -0.00494 0.311361 0.01229 0.735363 0.432248 0.772624 0.815278 0.516235 -0.04014 0.910792 -0.103294 0.3979	
-0.03402 -0.00999 0.004644 -0.01438 -0.0488 0.033796 0.021611 -0.00281 -0.04021 -0.05086 -0.07214 -0.0864 -0.08767 -0.03372 -0.06301 -0.03368 -0.06352 -0.08025 -0.0095 -0.01744 0.015923 0.138105	
0.034023 0.009989 -0.00464 0.014383 0.048797 -0.0338 -0.02161 0.02811 0.040207 0.050858 0.072145 0.086403 0.087672 0.033723 0.063014 0.03368 0.063524 0.080251 0.0095 0.017445 -0.01592 -0.1381	
2.175348 0.675173 -0.55473 0.294712 0.504991 0.1316742 0.774976 0.192735 0.009625 0.462108 0.347845 0.419549 0.223045 0.53517 -0.17655 -0.2845 -0.12726 -0.45109 0.047497 -0.62911 -3.25217 -1.16322	
0.04307 -0.04685 -0.16961 0.133855 0.33306 0.674605 -0.11577 -0.49094 0.01928 -0.50831 0.085422 -0.04516 0.44464 0.598495 -0.22553 0.137545 0.284335 0.736631 -1.13491 -0.42261 -1.55419 0.592213	
-0.07613 0.319726 0.21357 0.621595 -0.73745 0.635644 0.591878 0.119229 0.409864 0.125602 0.419342 -0.24413 0.061087 -0.19626 0.351376 0.018343 0.238102 0.1365 1.338177 0.284735 -0.02397 -0.25718	
-0.03402 -0.00999 0.004644 -0.01438 -0.0488 0.033796 0.021611 -0.00281 -0.04021 -0.05086 -0.07214 -0.0864 -0.08767 -0.03372 -0.06301 -0.03368 -0.06352 -0.08025 -0.0095 -0.01744 0.015923 0.138105	
0.080446 -0.04658 0.032842 -0.62324 0.756491 0.077008 -0.10992 -0.15859 -0.37482 -0.41785 0.111666 -0.38903 -0.59242 -0.18709 0.27821 -0.10682 0.568941 0.794821 0.22528 0.197094 1.763416 -0.28739	
0.034023 0.009989 -0.00464 0.014383 0.048797 -0.0338 -0.02161 0.02811 0.040207 0.050858 0.072145 0.086403 0.087672 0.033723 0.063014 0.03368 0.063524 0.080251 0.0095 0.017445 -0.01592 -0.1381	
0.034023 0.009989 -0.00464 0.014383 0.048797 -0.0338 -0.02161 0.02811 0.040207 0.050858 0.072145 0.086403 0.087672 0.033723 0.063014 0.03368 0.063524 0.080251 0.0095 0.017445 -0.01592 -0.1381	
-0.38954 -1.29359 -0.22716 -1.02468 0.286273 0.82731 0.244069 -0.26472 1.53815 -0.79406 0.334348 -0.51296 0.857024 0.105553 -0.31286 0.777921 0.486651 -0.48237 0.069607 -0.05979 -1.37421 0.37902	
0.260505 -0.18824 0.344552 -0.91399 -0.77763 0.391862 -0.12299 -0.43166 1.193944 -0.44504 -0.247726 0.138279 0.10966 -0.60733 0.351743 0.149741 -0.39393 0.103202 1.024481 0.282358	
-0.2604 -0.0127 0.202823 -0.26465 0.893458 -0.00076 -0.20123 -1.00571 -0.15892 0.670825 0.028047 -0.04522 0.020482 0.392325 -0.65273 -0.22446 -0.90479 -0.19119 0.227911 -1.17926 -0.37842	
$W^{(2)}$	$b^{(2)}$
1.623502 2.163218 -2.10706 -2.48224 -1.94685 -0.28102 2.290641 -2.09624 -0.28102 0.281021 -3.35667 -1.82092 1.796284 -0.28102 -0.28102 1.837207 0.281021 0.281021 2.628007 -1.68422 1.735871 -0.63081	

Figure 4.10 The weights and biases tensors of the optimal 21-(21)-1 shallow neural network.

$W^{(2)}$	$b^{(2)}$
0.21302 -0.13728 -0.29130 0.30728 -0.22614 -0.18915 0.08463 0.83295 -0.09320 0.06928 0.04505 0.12717 0.27171 -0.08078 0.40400 0.07463 0.88075 0.27232 0.23234 -0.64984 0.07199 0.40058	
-0.08141 0.11454 0.28791 -0.47611 -0.14114 0.11387 -0.12534 0.89128 -0.62873 0.03279 0.0436 0.54762 -0.76461 0.02933 0.57712 0.07656 -0.14809 0.65051 1.11834 0.39843 0.12635 0.80163	
-0.96144 -0.05944 -0.33331 0.79264 0.36823 -0.07155 -0.09395 0.20156 -0.92035 0.68631 -0.13720 -0.28002 0.02075 -1.10695 0.36385 0.55782 0.06162 -0.58075 -0.26665 -0.03545 1.72419 1.18896	
-0.287374 -0.39078 -0.53766 -0.37878 0.18985 0.24883 0.20388 -0.21241 0.19951 -0.01238 0.25977 0.06853 0.91162 0.21018 -0.42271 0.11499 -0.64558 0.23226 -1.04965 -0.49230 0.21684 0.67331	
-0.33686 0.19210 0.72565 0.67674 0.11436 -0.89480 -0.19297 -0.80494 0.02992 -0.02757 0.20245 -0.74528 0.10759 0.20419 -0.69976 -0.17644 0.49419 0.34569 -0.07508 0.49617 -0.20636 -0.56453	
0.59653 0.28900 0.24845 0.00159 -0.36076 -0.48237 0.16221 -0.02181 0.21936 -0.25284 -0.80652 0.21373 0.72760 -0.56491 0.25185 -0.37342 -0.37807 -0.25855 0.41869 0.18635 -0.68136 -0.05009	
$W^{(3)}$	$b^{(3)}$
-0.94332 -1.71903 1.21223 -1.83675 -0.62682 0.75407 -0.11556	
0.18001 0.50986 -0.40379 0.39878 0.82744 0.63286 -0.07297	
0.30440 0.36401 -0.19122 0.59981 0.01432 0.88375 -0.23361	
-0.15903 -0.51368 0.56069 -0.30175 -0.96129 -0.73518 0.06317	
0.33916 0.24617 -0.10838 0.5302 0.25802 0.92477 -0.10502	
$W^{(4)}$	$b^{(4)}$
-0.43093 -0.71937 -0.38082 0.79999 -0.21163 -0.41123	
-0.98336 0.91622 0.75527 -0.99166 0.64225 0.11795	
0.40206 -0.11321 -0.31949 0.08848 -0.35521 -0.66348	
-0.85122 0.59576 0.88247 -0.59981 0.90241 -0.10502	
$W^{(5)}$	$b^{(5)}$
-0.04615 0.06806 -0.11679 0.19370 0.00189	
0.70474 -0.68250 -0.07769 -0.10960 -0.25479	
2.40762 -1.66337 0.62795 -1.66391 -0.32587	
-0.21672 -0.50473 1.23758 -0.08385	

Figure 4.11 The weights and biases tensors of the optimal 21-(6-5-4-3)-1 fully connected deep neural network.

4.7.2 Feature importance

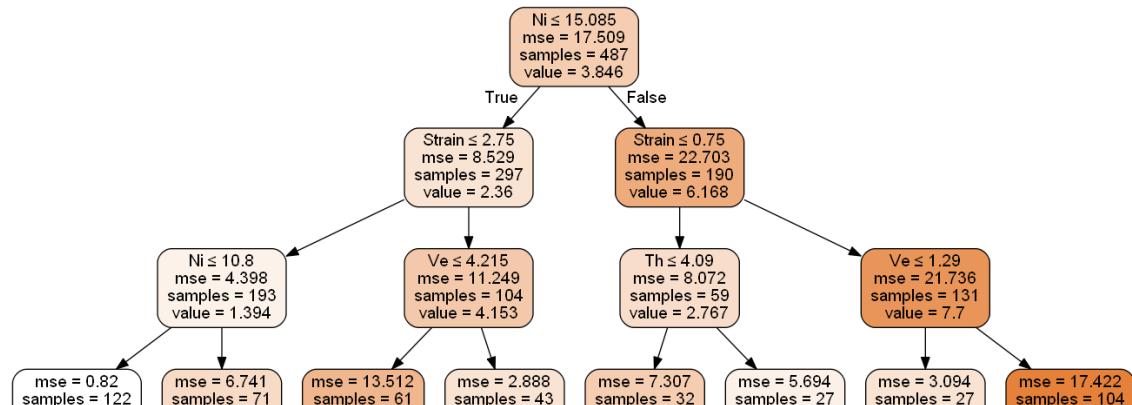


Figure 4.12 A decision tree for predicting solidification cracking susceptibility.

Decision tree can give us an intuitive knowledge of how solidification cracking susceptibility is calculated according to a serial of rules. Figure 4.12 visualizes a decision tree used for predicting solidification cracking susceptibility. The decision rule (e.g., $Ni \leq 15.085$, $Strain \leq 2.75$), mean square error (mse), sample number, the predicted TCL value

are shown in each node of the decision tree, The Ni content is used as the first decision rule, and the strain is used as the second decision rule. This shows the importance of the Ni content and strain in determining stainless steels' solidification cracking susceptibility.

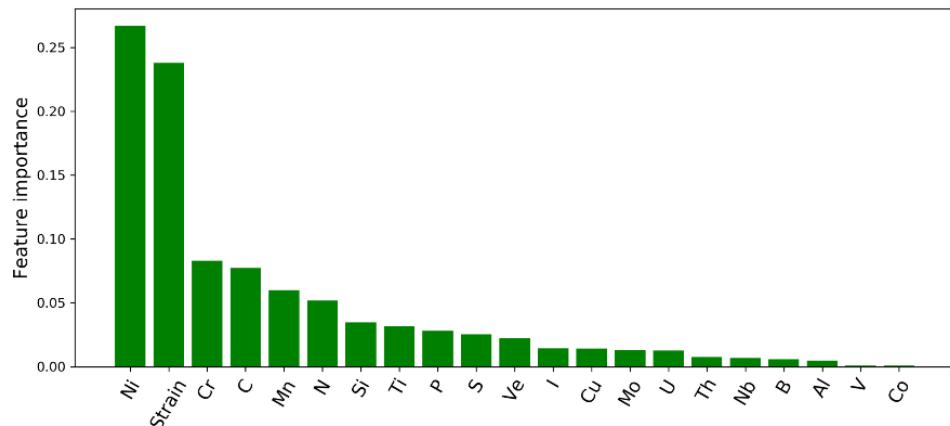


Figure 4.13 The feature importance of solidification cracking susceptibility derived from the random forest model.

Besides the intuitive impression of feature importance, tree-based machine learning models can give the value of feature importance. Figure 4.13 is the feature importance for solidification cracking susceptibility derived from the optimal random forest model. Major alloy elements Ni and Cr, minor alloy elements Mn and Si, impurity elements C, N, P, S, and the strain applied on the specimen play the most important roles in determining the solidification cracking susceptibility of stainless steels. The knowledge derived can be used to guide the next step predictions. A serial of predictions about solidification cracking susceptibility dependence on compositions and processing parameters in the following part will give us more details and a deeper understanding of this complex problem.

4.7.3 Visualization of prediction

(1) Solidification cracking susceptibilities of 304 and 310S stainless steels

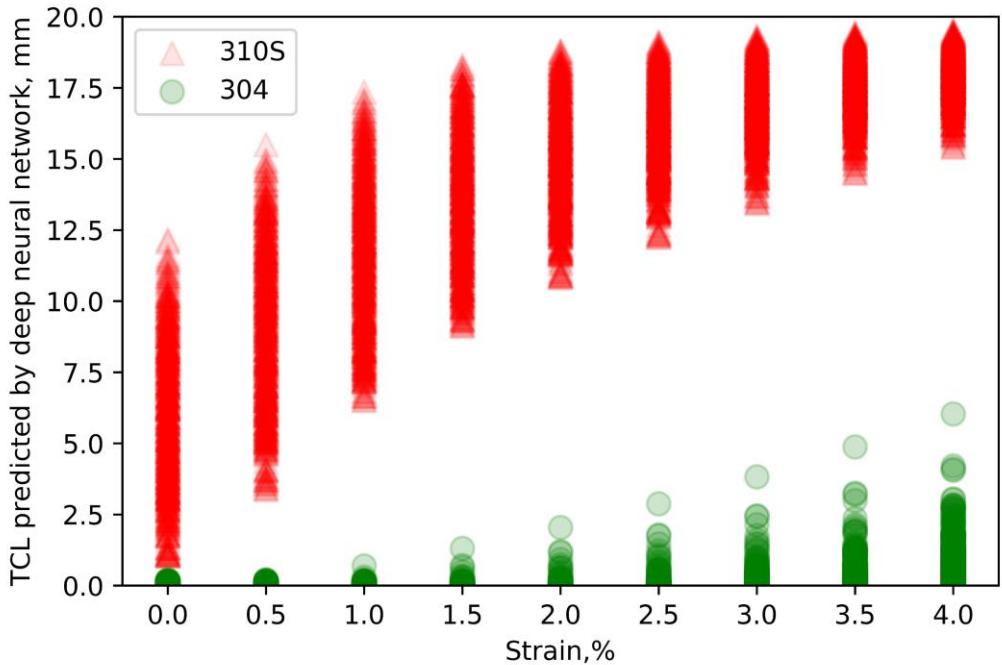


Figure 4.14 Total crack lengths of 304 and 310S stainless steel vary with strain, Cr, Ni, P, S contents vary in the specification.

Table 4.2 The specifications (wt%) of 304 and 310S stainless steel and the testing parameters used in prediction.

Code	C	Si	Mn	P	S	Cr	Ni	N	Al	Th	I	U	Ve
304	0.06	0.5	1.5	0.005-0.03	0.005-0.03	18-20	8-10.5	0.02	0.02	3.18	100	12	4.23
310S	0.01	0.5	1.5	0.005-0.03	0.005-0.03	24-26	19-22	0.02	0.02	3.18	100	12	4.23

Figure 4.14 shows the total crack length (an indicator of solidification cracking susceptibility) variation of 304 and 310S stainless steel when elements Ni, Cr, P, S vary in the specification (compositions are shown in Table 4.2). It shows that solidification cracking susceptibility is susceptible to composition variation. An alloy with a small difference in the composition can lead to a great difference in solidification cracking behaviors.

(2) Solidification cracking susceptibility dependence on Ni and Cr

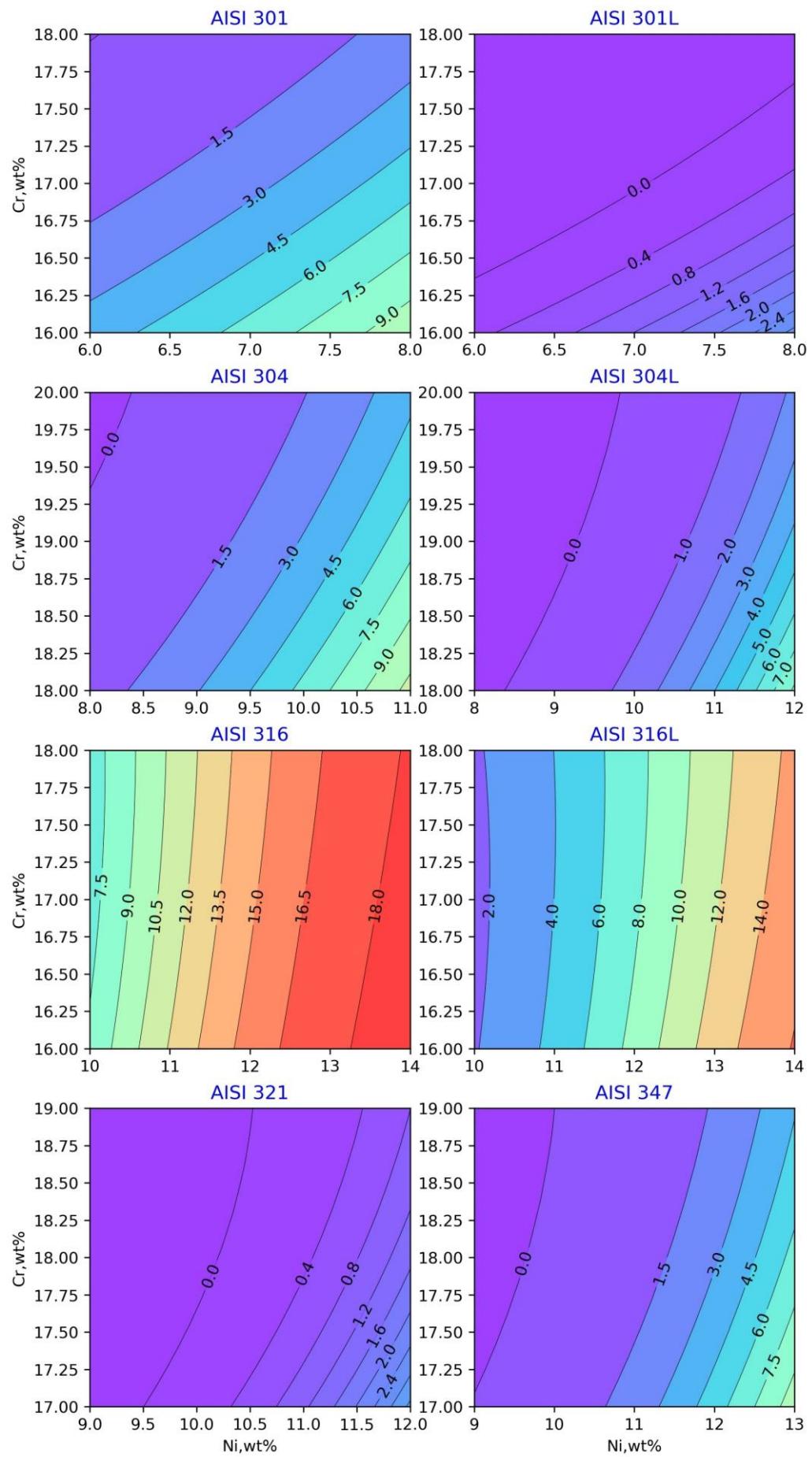


Figure 4.15 The prediction of 8 stainless steels' total crack length (mm) dependence on Ni and Cr (their Ni and Cr contents are in AISI specification), compositions used in predictions are shown in Table 4.3, processing parameters are the same: Th=3.18 mm, I=100 A, U=12 V, Ve=4.23 mm/s, strain=3%.

Table 4.3 Compositions (wt%) of stainless steels used in prediction (varied Ni and Cr).

Code	C	Si	Mn	P	S	Cr	Ni	Mo	N	Nb	Cu	Al	Ti
301	0.10	1.0	2.0	0.02	0.02	16.0-18.0	6.0-8.0	-	0.10	-	-	-	-
301L	0.03	1.0	2.0	0.02	0.02	16.0-18.0	6.0-8.0	-	0.10	-	-	-	-
304	0.08	1.0	2.0	0.02	0.02	18.0-20.0	8.0-11.0	-	0.10	-	-	-	-
304L	0.03	1.0	2.0	0.02	0.02	18.0-20.0	8.0-12.0	-	0.10	-	-	-	-
316	0.08	0.75	2.0	0.02	0.02	16.0-18.0	10.0-14.0	2.0	0.10	-	-	-	-
316L	0.03	0.75	2.0	0.02	0.02	16.0-18.0	10.0-14.0	2.0	0.10	-	-	-	-
321	0.08	0.75	2.0	0.02	0.02	17.0-19.0	9.0-12.0	-	-	-	-	-	0.5
347	0.08	0.75	2.0	0.02	0.02	17.0-19.0	9.0-13.0	-	-	0.5	-	-	-

The prediction of 8 stainless steels' total crack length (an index of solidification cracking susceptibility) dependence on the main alloy elements Ni and Cr is shown in Figure 4.15 (compositions are shown in Table 4.3). The predictions are in good accordance with metallurgical experience. Fully austenitic stainless steels like AISI 316 is more susceptible to solidification cracking than stainless steels contain a certain amount of ferrite like AISI 304. Element C, which tends to segregate to grains boundaries, increase cracking susceptibility (see the contour difference between AISI 316L and AISI 316). Machine learning regression has transformed scattered data points into an expressive high-dimensional map, and those contours are only some slices of it. The tendency of increasing solidification cracking susceptibility with the increase of Ni is illustrated in the figure. Ferrite can accept more impurity elements (like S and P) than austenite. Irregular ferrite/austenite grain boundary is not in favor of the propagation of cracking cracks. Thus, a High ratio of Cr to Ni (or Cr equivalent to Ni equivalent) is good for solidification cracking resistance by forming a certain amount of ferrite like that in AISI 301 and 304.

(3) Solidification cracking susceptibility dependence on P and S

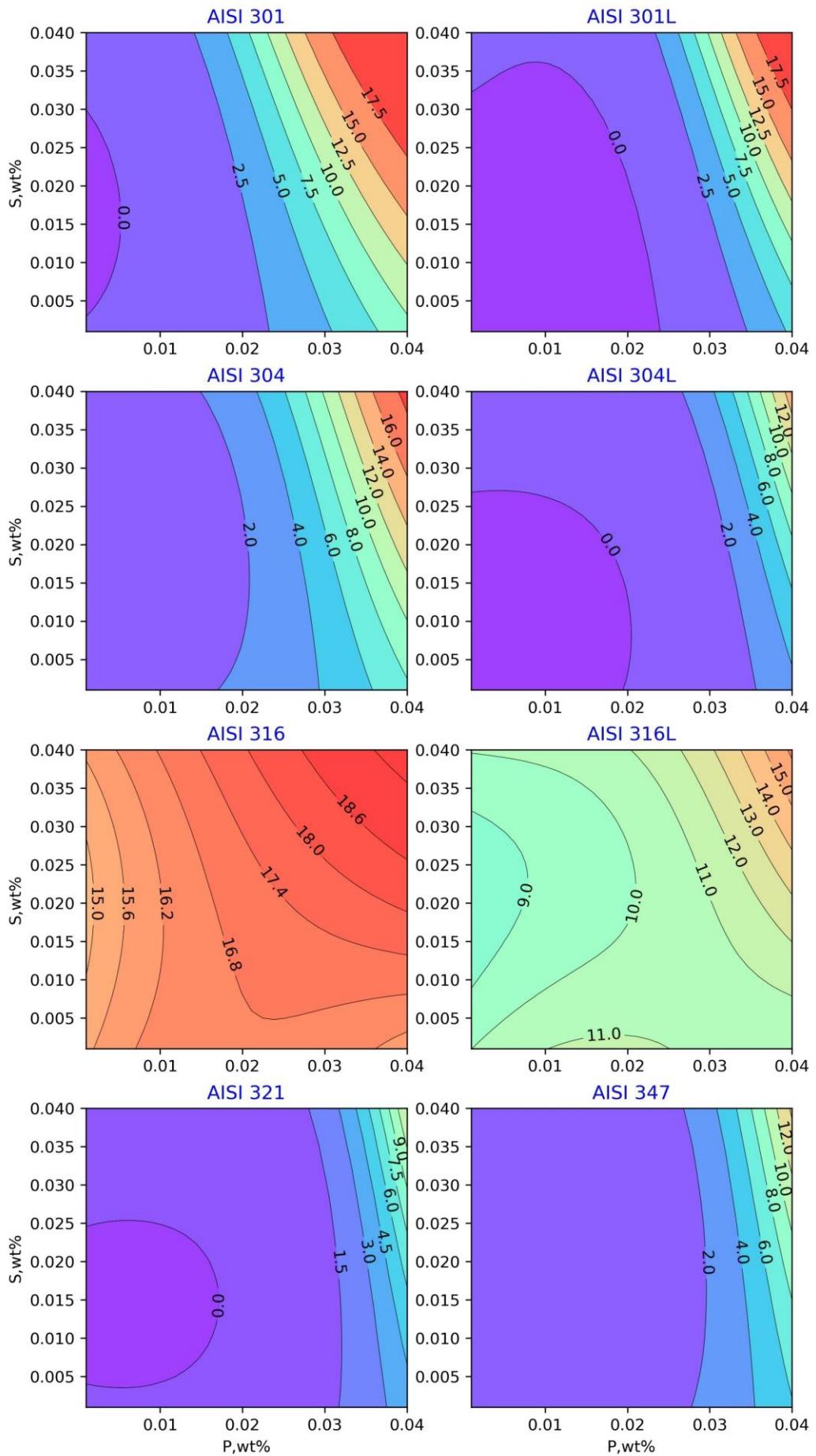


Figure 4.16 The prediction of 8 stainless steels' total crack length (mm) dependence on P and S, compositions used in predictions are shown in Table 4.4, processing parameters are the same: Th=3.18 mm, I=100 A, U=12 V, Ve=4.23 mm/s, strain=3%.

Table 4.4 Compositions (wt%) of stainless steels used in prediction (varied P and S).

Code	C	Si	Mn	P	S	Cr	Ni	Mo	N	Nb	Cu	Al	Ti
301	0.10	1.0	2.0	0.001-0.040	0.001-0.040	17.0	7.0	-	0.10	-	-	-	-
301L	0.03	1.0	2.0	0.001-0.040	0.001-0.040	17.0	7.0	-	0.10	-	-	-	-
304	0.08	1.0	2.0	0.001-0.040	0.001-0.040	19.0	9.5	-	0.10	-	-	-	-
304L	0.03	1.0	2.0	0.001-0.040	0.001-0.040	19.0	9.5	-	0.10	-	-	-	-
316	0.08	0.75	2.0	0.001-0.040	0.001-0.040	17.0	12.0	2.0	0.10	-	-	-	-
316L	0.03	0.75	2.0	0.001-0.040	0.001-0.040	17.0	12.0	2.0	0.10	-	-	-	-
321	0.08	0.75	2.0	0.001-0.040	0.001-0.040	18.0	9.5	-	-	-	-	-	0.5
347	0.08	0.75	2.0	0.001-0.040	0.001-0.040	18.0	9.5	-	-	0.5	-	-	-

The prediction of 8 stainless steels' total crack length (an indicator of solidification cracking susceptibility) dependence on impurity elements P and S is shown in Figure 4.16 (compositions are shown in Table 4.4). The tendency of increasing solidification cracking susceptibility with the increase of P and S is shown in the prediction. The extent of increasing susceptibility with P and S are different among different steels. The solidification cracking susceptibility of AISI 316 is very sensitive to the P and S contents. Small difference in C and N contents can show great difference in solidification cracking behavior. But that of AISI 316L with less C content is less sensitive to the P and S contents. The figure also shows that impurity elements P and S have a different impact on solidification cracking susceptibility (most contours are not symmetry along P=S diagonal), the commonly linear superposition P and S is improper.

(4) Solidification cracking susceptibility dependence on C and N

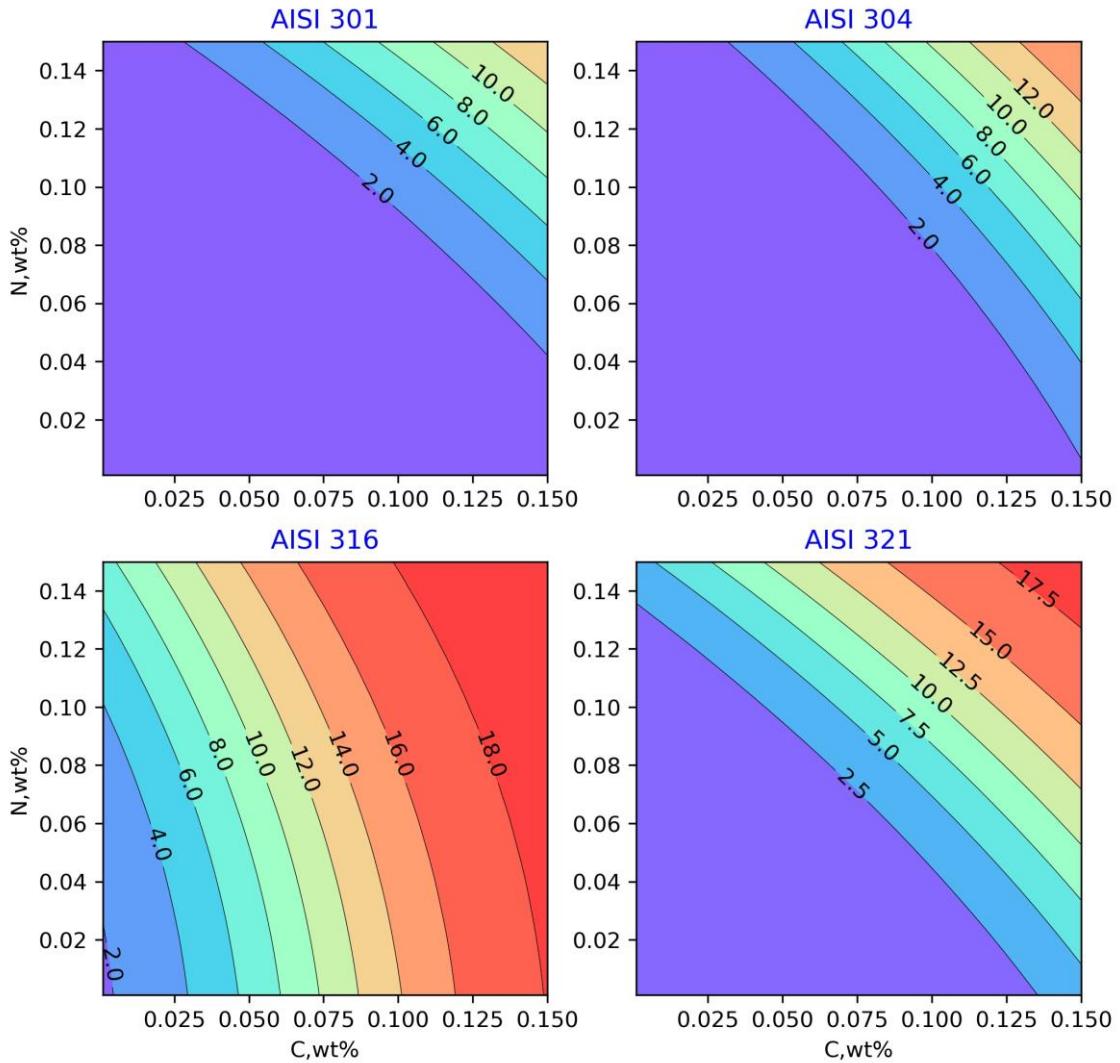


Figure 4.17 The prediction of stainless steels' total crack length (mm) dependence on C and N, compositions used in predictions are shown in Table 4.5. Processing parameters are the same: Th=3.18 mm, I=100 A, U=12 V, Ve=4.23 mm/s, strain=3%.

Table 4.5 Compositions (wt%) of stainless steels used in prediction (varied C and N).

Code	C	Si	Mn	P	S	Cr	Ni	Mo	N	Nb	Cu	Al	Ti
301	0.001-0.15	1.0	2.0	0.02	0.02	17.0	7.0	-	0.001-0.15	-	-	-	-
304	0.001-0.15	1.0	2.0	0.02	0.02	19.0	9.5	-	0.001-0.15	-	-	-	-
316	0.001-0.15	0.75	2.0	0.02	0.02	17.0	12.0	2.0	0.001-0.15	-	-	-	-
321	0.001-0.15	0.75	2.0	0.02	0.02	18.0	9.5	-	0.001-0.15	-	-	-	0.5

The prediction of 4 stainless steels' total crack length (an indicator of solidification cracking susceptibility) dependence on C and N is shown in Figure 4.17 (compositions are shown in Table 4.5). The tendency of increasing solidification cracking susceptibility with the increase of C and N is revealed in the prediction. The extent of increasing susceptibility with C and N varied among different steels due to the different alloy elements in corresponding steels. The solidification cracking susceptibility of AISI 316 is very sensitive to the C and N contents. Small difference in C and N contents can show great difference in solidification cracking behavior. But AISI301 and AISI304 are less susceptible to solidification cracking.

(5) Solidification cracking susceptibility dependence on strain

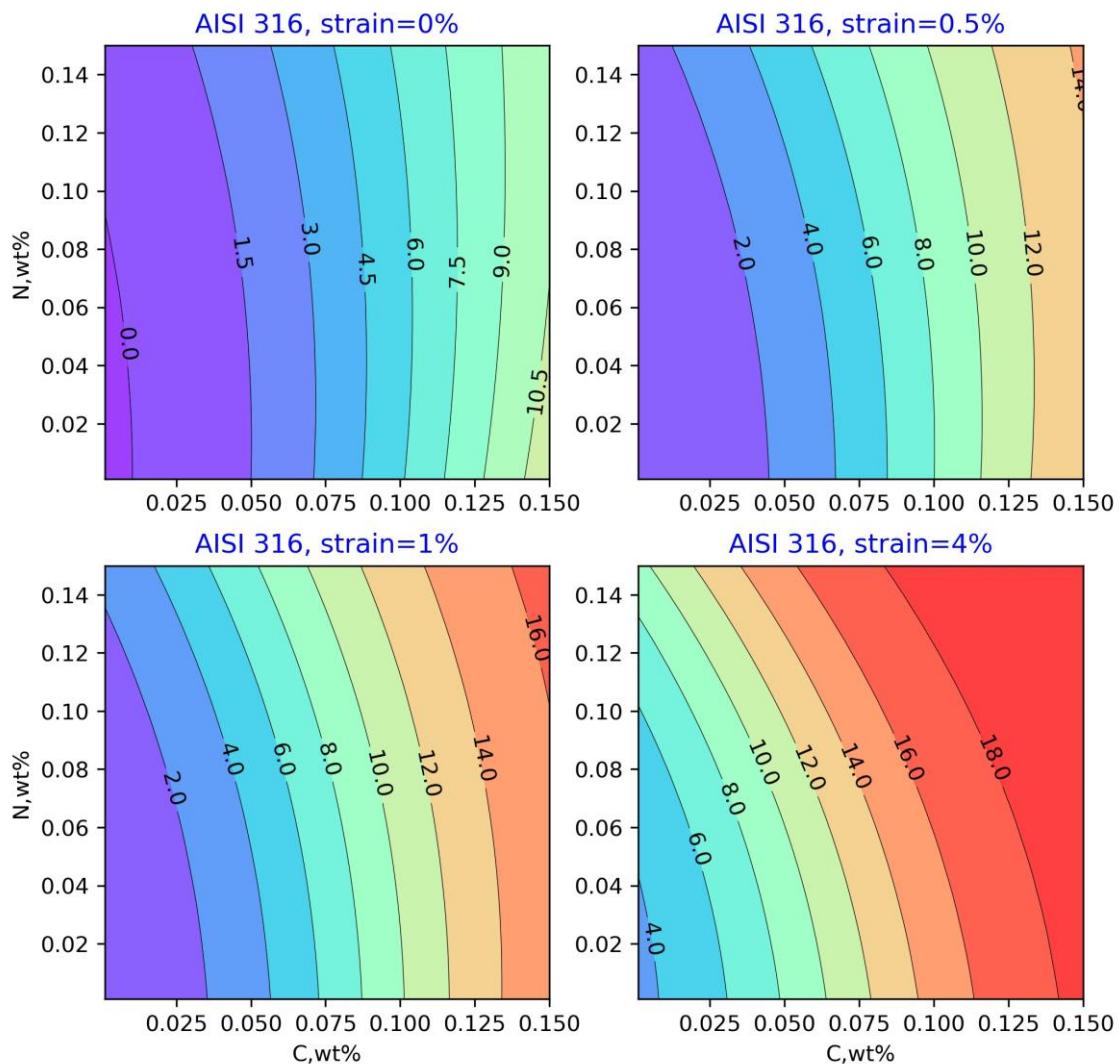


Figure 4.18 The prediction of AISI 316 stainless steel's total crack length (mm) dependence on strain with C and N vary in specification, processing parameters are: Th=3.18 mm, I=100 A, U=12

$V, Ve=4.23 \text{ mm/s}$.

The prediction of AISI 316 stainless steel's total crack length (an index of solidification cracking susceptibility) dependence on strain is shown in Figure 4.18 (compositions are shown in Table 4.5). The tendency of increasing solidification cracking susceptibility with the increase of strain is correctly revealed in the prediction. This tendency is also a nonlinear behavior, which is very difficult for physical models to depict but is relatively easy for machine learning models.

4.8 Summary

Five machine learning models, *i.e.*, shallow neural network, fully connected deep neural network, support vector machine, decision tree and random forest, were used to predict solidification cracking susceptibility of stainless steels as a function of composition and processing parameters. Fully connected deep neural network models (with pre-training and fine-tuning) outperformed shallow neural network, fully connected deep neural network models (without pre-training), support vector machine and tree-based models in prediction accuracy; tree-based models have accepted accuracy and better interpretability than support vector machine and neural network models; the optimal deep neural network has more compact structure and less parameters than the optimal shallow neural network (thus able to reduce the risk of overfitting and increase generalization).

Through machine learning regression, vast scattered experimental data hidden in literature can produce simple quantitative expression: specific material property as function of chemistry composition and processing parameters. The derived models can be used in material property prediction, new alloys development, and comparison with experimental results.

The combination of multiple machine learning models shows that solidification cracking susceptibility of stainless steels was mainly determined by Ni content / Cr content (or ferrite content / austenite content), impurity content and the value of applied strain.

Though deep neural networks plus big datasets are the best choice, deep neural networks with small datasets and pre-training can be a reasonable choice when big datasets are not available. In material science, small datasets are common, and the

problems to be solved have fewer input variables than that in image recognition. Thus, compact deep neural network (neuron number in each hidden layer is small) are suitable for material problems, such as solidification cracking susceptibility prediction in this study, and pre-training using stacked auto-encoder is effective and necessary in the fully connected deep neural network regression of small datasets. Our study demonstrates that compact / narrow fully connected deep neural network with a small dataset and special training methods has huge potential for extensive applications in material study, especially for those multivariate nonlinear problems.

5 Application of convolutional neural network & periodic table representation to predicting glass-forming ability & compound properties

The works of Sun¹⁵⁸ and Ward^{57,58} et al. show machine learning is a promising method for predicting glass-forming ability and developing new amorphous alloys. But the conventional machine learning models they used work well only when data is sufficient and features are well-designed. However, both data and descriptors are deficient in the materials field. Successful manual feature engineering, *i.e.*, transforming raw data into direct input for models, is very challenging and needs a deep understanding of the physical mechanisms of the problems. In this scenario, taking advantage of convolutional neural networks' automatic features extraction ability and adding domain expertise into data representations are better solutions^{3,56}.

Recently, Zheng et al. proposed a periodic table representation (PTR) for alloy compositions and combined it with convolutional neural networks (CNN) to predict compound formation energy¹⁷⁵. Zeng et al. proposed an atom table convolutional neural networks (ATCNN) for predicting superconducting critical temperature¹⁷⁶. Both methods mapped compositions to 2-D images and exploited convolutional neural networks to extract features and make regression automatically. Their results showed the advantage of high prediction accuracy and end-to-end learning. This chapter reported our attempt to predict glass-forming ability using convolutional neural networks with a modified periodic table representation. We used a medium dataset, however for most alloy systems in the dataset it is still a small dataset task. To show the advantage of this method, we compared it with convolutional neural networks using representations without the periodic table information and conventional shallow neural network (SNN) using manual feature engineering. The features (*i.e.* the intermediate results of convolutional neural networks) extracted were also visualized to prove the domain expertise was presented in the extracted features. Figure 5.1 shows the workflow of predicting glass-forming ability using convolutional neural network plus 2-D representation for compositions.

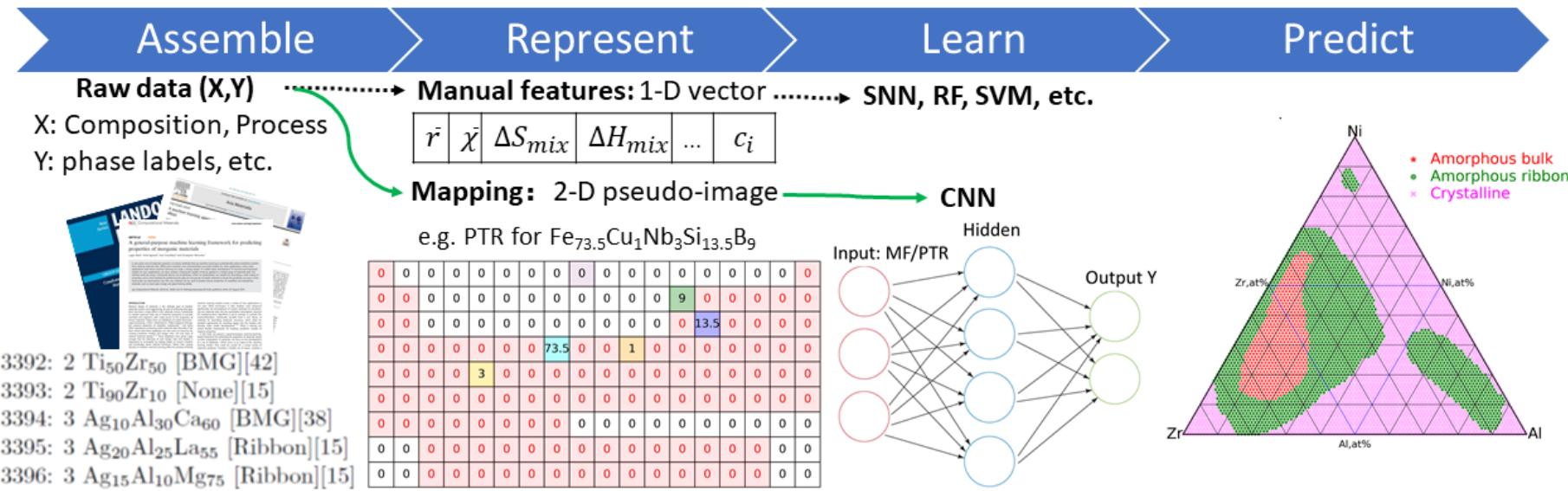


Figure 5.1 The workflow of predicting glass-forming ability using convolutional neural network plus periodic table representation, MF, PTR and CNN denote manual features, periodic table representation and convolutional neural network, respectively.

5.1 Datasets

5.1.1 Glass-forming ability dataset

Enough reliable data is vital for the performance of a machine learning model. In this work, we try to assemble a glass-forming ability dataset as large as possible. Our dataset includes Sun's binary alloys dataset¹⁵⁸ (about 3000 entries) and Ward's ternary amorphous ribbon alloys dataset^{58,159} (about 6000 entries) and bulk metallic glasses dataset⁵⁷ (about 800 entries) and Miracle's dataset¹⁷⁷ (about 300 entries). In those datasets, crystalline alloys data are in the minority, because amorphous ribbon alloys and bulk metallic glasses are the focus of research, crystalline compositions are commonly discarded and unpublished as the failed experimental results. In reality, the number of amorphous alloys is less than that of their crystalline counterparts. To compensate for this weakness and increase the variety of crystalline data in our dataset, we add 800 pieces of conventional crystalline metallic materials data (including steels, superalloys and Co, Al, Mg, Cu, Zn alloys) from <https://www.makeitfrom.com/> into our dataset. After sorting and removing redundancy, our dataset has 10440 entries. Figure 5.2 shows the statistics of our raw dataset. Our dataset contains 97 different elements in the periodic table, and many of these elements simultaneously present in entries of crystalline alloys (CR), amorphous ribbon alloys (AMR), and bulk metallic glasses (BMG). The statistics of labels are shown in Table 5.1. The distribution of elements and three labels is good for data division and model training.

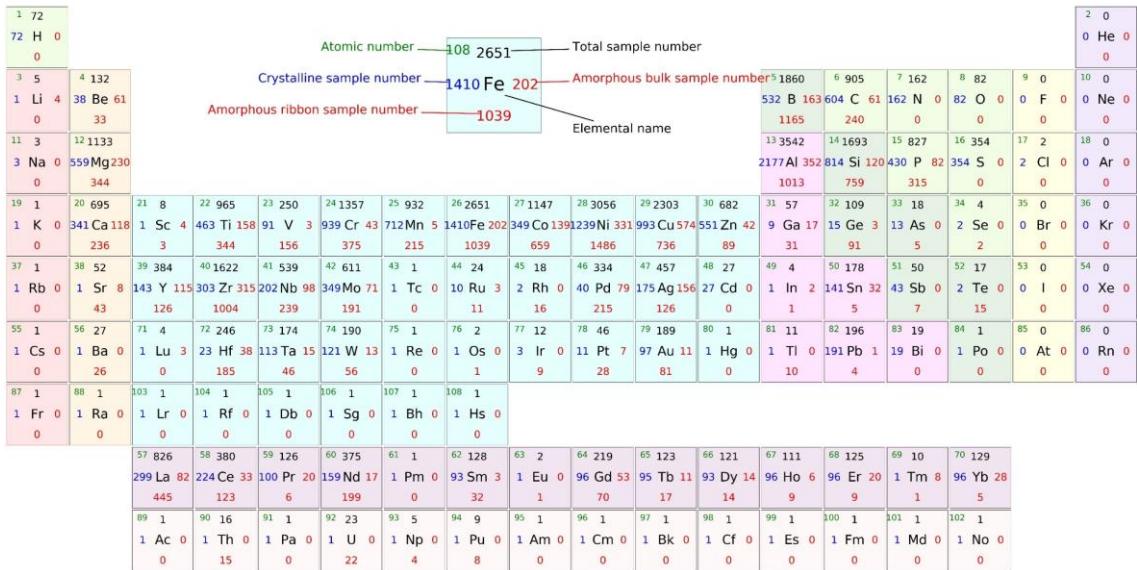


Figure 5.2 The statistics of elements distribution in our dataset. Many elements appear in crystalline alloys (CR), amorphous ribbon alloys (AMR), and bulk metallic glasses (BMG) simultaneously.

Table 5.1 The statistics of labels in our dataset. CR, AMR, and BMG denote crystalline, amorphous ribbon, and bulk metallic glasses, respectively.

	BMG	AMR	CR	total
Pure	0	0	84	84
Binary	17	1108	2607	3732
Ternary	279	3520	1369	5168
quaternary & other	650	2	804	1456
sum	946	4630	4864	10440

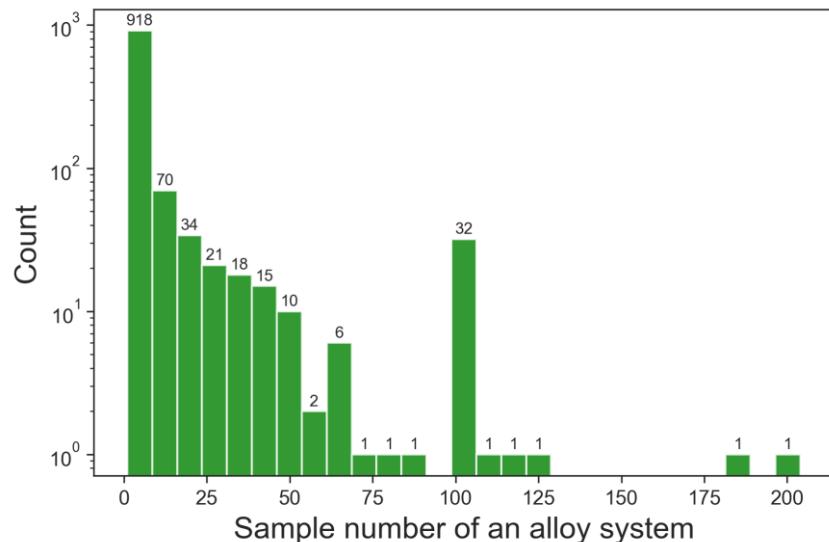


Figure 5.3 The statistics of alloy systems' size in our dataset. Most alloy systems (918 out of 1134)

have nine entries or less, the deficiency and sparsity of the data are apparent.

Figure 5.3 shows the statistics of alloy systems' size in our dataset. It shows most alloy systems (918 out of 1134 alloy systems) have less than nine entries, and only two systems, *i.e.*, Al-Ni-Zr (182 entries) and Al-La-Ni (204 entries) have over 150 entries. Though the dataset is not small, taking into consideration of the myriad of elemental combinations and compositional variation (*e.g.*, if elements A, B, C, D vary from 1 to 99 at% with 1at% compositional interval, there will be 4851 possible combinations in one ternary system A-B-C alone; there will be 156849 possible combinations in one quaternary system A-B-C-D alone), the deficiency and sparsity of the data is apparent. Alloys in our dataset only occupy a tiny part of the possible combinations, *i.e.*, most compositional space is unexplored. This can be seen as another form of small dataset and is common when applying machine learning models to material problems.

The ideal parameters to describe glass-forming ability of an alloy are critical cooling rate and critical thickness. Due to the incompleteness of them in our dataset, we simplified the regression problem as a classification problem. Considering that we should not arbitrarily declare that all amorphous ribbon alloys in our dataset cannot form bulk metallic glasses, we did not simply treat it as a problem of (CR / AMR / BMG) ternary classification. We added a process parameter (0 represents rapid solidification melt-spun, and 100 represent copper mold casting of normal cooling rate) into this problem to convert the ternary classification problem into a problem of (AM / CR) binary classification. AM represents forming an amorphous state, and CR represents forming a crystalline state, see the matrix for transforming the original labels to the new labels in Table 5.2. The size of our original dataset is 10440, and the size of our final dataset (after transformation) is 16250 (946*2 + 4630 + 4864*2).

Table 5.2 Matrix for transforming original labels of ternary classification to new labels of binary classification. CR and AM denote crystalline state, amorphous state, respectively.

	Melt-spun	Copper mold casting
CR	CR	CR
BMG	AM	AM
AMR	AM	-

5.1.2 Open quantum materials database dataset

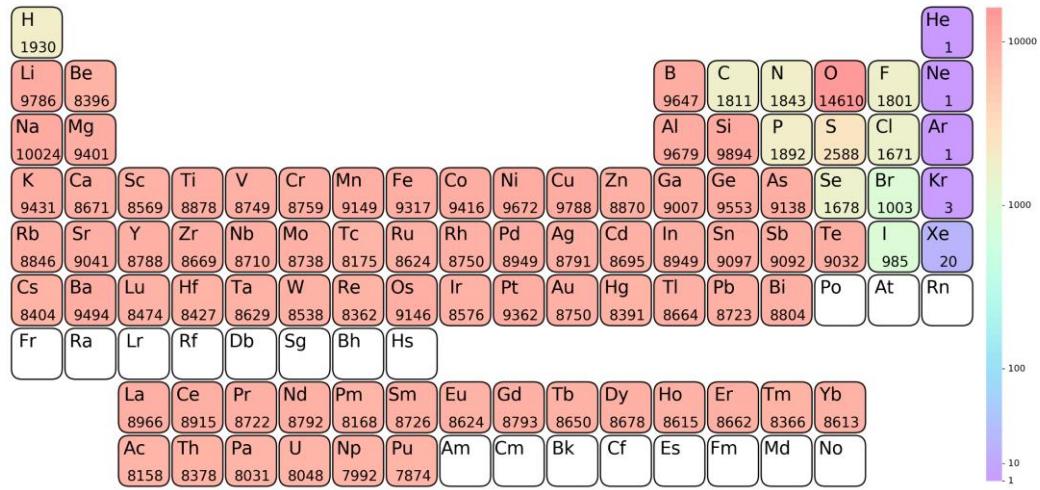


Figure 5.4 Elements distribution in OQMD dataset. The occurrence numbers of elements in the dataset are given under element symbols. The white squares, i.e. square with element symbol and without occurrence number, indicate these elements are not in the dataset.

Materials science community has made a great effort in accumulating big datasets of materials properties: e.g. open quantum materials database (OQMD)^{68,178}, automatic flow for materials discovery (AFLOWLIB)⁷⁰, the materials project⁷¹, crystallography open database (COD)¹⁷⁹, Springer Materials (<https://materials.springer.com/>). These big datasets can be exploited as source datasets for transfer learning (just like the role of ImageNet dataset in computer vision field) to build the general and transferable feature extractors. In this work, we used a dataset of 228,676 compounds from OQMD which was also used by Ward⁵⁸ before. They are lowest-energy compounds at their unique compositions. The dataset involves these compounds' chemistry and their physical properties calculated by density functional theory (DFT). Convolutional neural networks are trained on the dataset to predict formation energy (Ef) and specific volume (V) of these compounds. Figure 5.4 shows the elements distribution in the OQMD dataset. Most elements in periodic table (89 out of 108) occur in the dataset, and 73 elements occur more than 8000 times. The good distribution of elements and the size of the source dataset is a guarantee of the effectiveness of general and transferable feature extractors.

5.2 Data representation

5.2.1 2-D representation for data

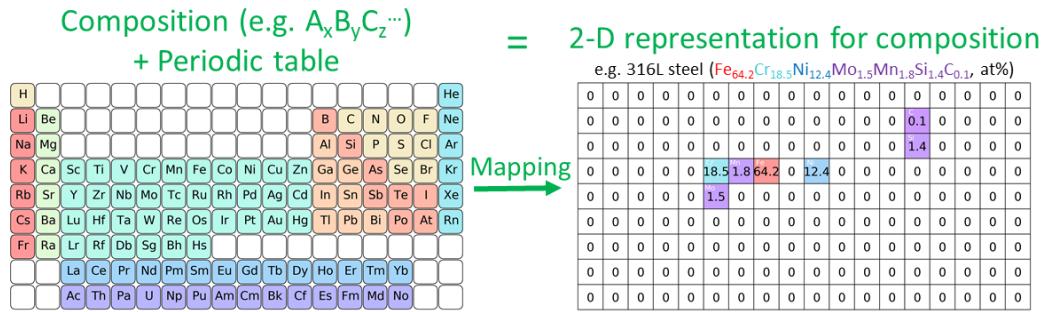


Figure 5.5 Mapping composition to 2-D representation.

To exploit the convenience of convolutional neural network's automatic features extraction, we mapped (see Figure 5.5) the raw data to 2-D images (increasing data dimensional) using a modified periodical table representation (PTR). We used the whole periodic table (see Figure 5.6) and mapped processing parameters to the unused area in the periodic table; on the contrary, Zheng's periodical table representation¹⁷⁵ did not contain rare earth elements and processing parameters. In order to validate the benefit of merging periodic table knowledge into data representation, two other mapping methods without periodic table structure were also used: *i.e.*, randomized periodic table representation (see Figure 5.7) and atom table representation¹⁷⁶ (see Figure 5.8).

Periodic table representations (used in model CNN3) mimic digital images. Alloy composition and preparation processes are mapped to a 2-dimensional image of 9 pixels by 18 pixels (162 pixels in total). Each square represents a pixel. The 108 blue squares correspond to 108 elements in the periodic table, *e.g.*, the 1st pixel/square in the 1st row is used to store the atomic percentage of element Hydrogen in an alloy. The 54 grey squares are the unused area in the periodic table. The alloy composition (in atomic percentage) is mapped to the corresponding blue squares, and the preparation process (0 represents melt-spun, and 100 represents copper mold casting) is mapped to a grey square (we arbitrarily chose the 9th pixel/square in the first row in this work). The rest pixels/squares are set to 0.

The randomized periodical table representation used in model CNN2 is almost the same with periodical table representation except 108 elements were randomly placed in the periodical table area (see Figure 5.7). The atom table representation used in model CNN1

are square images of 11*11 pixels, elements are placed in an atom table from left to right and from top to bottom according to the atomic number of elements (see Figure 5.8). The preparation process is mapped to the last pixel in the atom table and the rest unused pixels are set to 0.

Periodic table representation										
0: melt-spun 100: copper mold casting										
¹ H	0	0	0	0	0	0	0	0	0	² He
³ Li	⁴ Be	0	0	0	0	0	0	0	0	⁶ C
¹¹ Na	¹² Mg	0	0	0	0	0	0	0	⁷ N	⁸ O
¹⁹ K	²⁰ Ca	²¹ Sc	²² Ti	²³ V	²⁴ Cr	²⁵ Mn	²⁶ Fe	²⁷ Co	²⁸ Ni	³⁰ Zn
³⁷ Rb	³⁸ Sr	³⁹ Y	⁴⁰ Zr	⁴¹ Nb	⁴² Mo	⁴³ Tc	⁴⁴ Ru	⁴⁵ Rh	⁴⁶ Pd	⁴⁷ Ag
⁵⁵ Cs	⁵⁶ Ba	⁷¹ Lu	⁷² Hf	⁷³ Ta	⁷⁴ W	⁷⁵ Re	⁷⁶ Os	⁷⁷ Ir	⁷⁸ Pt	⁷⁹ Au
⁸⁷ Fr	⁸⁸ Ra	¹⁰³ Lr	¹⁰⁴ Rf	¹⁰⁵ Db	¹⁰⁶ Sg	¹⁰⁷ Bh	¹⁰⁸ Hs	0	0	0
0	0	⁵⁷ La	⁵⁸ Ce	⁵⁹ Pr	⁶⁰ Nd	⁶¹ Pm	⁶² Sm	⁶³ Eu	⁶⁴ Gd	⁶⁵ Tb
0	0	⁸⁹ Ac	⁹⁰ Th	⁹¹ Pa	⁹² U	⁹³ Np	⁹⁴ Pu	⁹⁵ Am	⁹⁶ Cm	⁹⁷ Bk

Figure 5.6 Schematic diagram for periodic table representation used in CNN3.

Randomized periodic table representation										
0: melt-spun 100: copper mold casting										
⁷⁸ Pt	0	0	0	0	0	0	0	0	0	⁸¹ Tl
⁵⁴ Xe	⁹ F	0	0	0	0	0	0	0	0	¹⁰⁸ Hs
⁷ N	⁵² Te	0	0	0	0	0	0	0	¹² Mg	¹⁰⁷ Bh
¹⁰ Ne	⁴⁶ Pd	³³ As	⁸⁷ Fr	⁵⁰ Sn	¹⁰⁰ Fm	²¹ Sc	³⁴ Se	¹⁸ Ar	¹⁶ S	⁶⁶ Dy
⁴⁷ Ag	⁶ C	⁹⁰ Zn	¹⁴ Si	²³ V	³¹ Ga	⁴⁹ In	⁴ Be	⁹⁹ Es	¹⁰⁵ Db	⁵⁸ Ce
³² Ge	⁴⁸ Cd	⁷⁵ Re	⁵⁵ Cs	⁶⁵ Tb	³⁵ Br	⁸⁰ Hg	²⁴ Cr	⁷⁷ Ir	²² Ti	⁴³ Tc
⁵³ I	³⁷ Rb	²⁶ Fe	³⁶ Kr	⁵⁶ Ba	²⁸ Ni	⁹³ Np	²⁵ Mn	0	0	¹³ Al
0	0	¹¹ Na	¹⁵ P	¹⁰⁴ Rf	¹⁰⁶ Sg	⁶⁸ Er	⁵ B	²⁹ Cu	⁴¹ Nb	² He
0	0	⁶⁰ Nd	⁶¹ Pm	⁹⁰ Th	⁸⁶ Rn	⁷⁴ W	⁴⁴ Ru	⁸⁸ Ra	⁶³ Eu	³⁹ Y

Figure 5.7 Schematic diagram for randomized periodic table representation used in CNN2.



Figure 5.8 Schematic diagram for atom table representation used in CNN1.

5.2.2 Manual features engineering

The 1-D vector of features (attributes/descriptors) used as input for shallow neural network include (a) statistics information of components' properties based on element properties, e.g., the maximum/minimum/average of atomic radius, Pauling electronegativity, elemental bulk modulus, elemental work function, melting point; (b) one-hot coded composition vector; (c) parameters derived from physical models and empirical criteria¹⁴¹, e.g., mixing entropy ΔS_{mix} , mixing enthalpy ΔH_{mix} , the atomic size difference ΔR , the electronegativity difference $\Delta \chi$, valence electron concentration VEC.

$$\Delta S_{mix} = -R \sum_{i=1}^n c_i \ln c_i$$

$$\Delta H_{mix} = \sum_{i=1, i \neq j}^n 4\Delta_{mix}^{AB} c_i c_j$$

$$\Delta R = \sqrt{\sum_{i=1}^n c_i (1 - \frac{r_i}{\bar{r}})^2}, \quad \bar{r} = \sum_{i=1}^n c_i r_i$$

$$\Delta\chi = \sqrt{\sum_{i=1}^n c_i(\chi_i - \bar{\chi})^2}, \quad \bar{\chi} = \sum_{i=1}^n c_i\chi_i$$

$$VEC = \sum_{i=1}^n c_i(VEC)_i$$

where c_i is the atomic fraction of the i -th component; Δ_{mix}^{AB} is the mixing enthalpy of alloy A-B; r_i is the Miracle's atomic radius of the i -th component; χ_i is the electronegativity of the i -th component; \bar{r} is the average atoms radius of the components in the alloy; $\bar{\chi}$ is the average electronegativity of the components in the alloy; $(VEC)_i$ is the valence electron concentration of the i -th component; VEC is the average valence electron concentration of the components in the alloy. The r_i were taken from Ref¹⁸⁰; Δ_{mix}^{AB} were taken from Ref¹⁵⁰; Pauling electronegativity, elemental bulk modulus, Elemental work function were taken from ref¹⁸¹.

Table 5.3 The list of manual features used in building machine learning models. Features used in SNN1 and SNN3 are denoted with *. Other features combinations were also tested at the beginning but did not show noticeable improvement in accuracy, so they were not used in final model.

Mixing entropy ΔS_{mix}^*	(Max / Min / Average) Miracle atomic Radius*	(Max / Min / Average) Pauling electronegativity*	(Max / Min / Average) elemental bulk modulus*
(Max / Min / Average) elemental work function*	(Max / Min / Average) melting temperature	(Max / Min / Average) boiling temperature	(Max / Min / Average) atomic number
(Max / Min / Average) atomic weight	(Max / Min / Average) covalent radius	(Max / Min / Average) ionic radius	(Max / Min / Average) elemental concentration
Mixing enthalpy ΔH_{mix}	Atomic size difference ΔR	Electronegativity difference $\Delta\chi$	Valence electron concentration VEC

5.2.3 Magpie descriptors

Ward et al. proposed a set of general-purpose descriptors, called Magpie descriptors⁵⁸, which fall into four categories:

(1) stoichiometric descriptors (6 in total)

They are based on L^p norms:

$$\|x\|_p = (\sum_{i=0}^n |x_i|^p)^{1/p}, p = 0, 2, 3, 5, 7 \text{ and } 10$$

x_i is the atomic fraction of the element i.

(2) Elemental-Property-Based Attributes (115 in total)

They are based on statistics of the elemental properties (atomic number, Mendeleev number, column, row, atomic weight, melting temperature, covalent Radius, electronegativity, number of s valence electrons, number of p valence electrons, number of d valence electrons, number of f valence electrons, total number of valance Electrons, number of unfilled s states, number of unfilled p states, number of unfilled d states, number of unfilled f states, total number of unfilled states, specific volume of 0 K ground state, band gap energy of 0 K ground state, magnetic moment per atom of 0 K ground state, space group number of 0 K ground state). For each property, the minimum, maximum, range, fraction-weighted mean, average deviation, and mode (*i.e.* the property of the most prevalent element) of the values of the properties of each element present in the material is computed. The mean \bar{f} and average deviation \hat{f} are calculated as follows:

$$\bar{f} = \sum x_i f_i$$

$$\hat{f} = \sum x_i |f_i - \bar{f}|$$

where f_i is the property of element i, x_i is the atomic fraction.

(3) Valance orbital occupation descriptors (4 in total)

The ratio of the average number of valance electrons in each orbital to the total number of valance electrons.

(4) Ionic compound descriptors (3 in total)

They are designed to determine whether a material is ionically bonded. The first measure is a Boolean denoting whether it possible to form an ionic compound. The other two attributes are based on the “ionic character” of a binary compound. The first

attribute we used is the maximum ionic character between any two constituent elements in the material:

$$I(\chi_i, \chi_j) = 1 - \exp(-0.25(\chi_i - \chi_j)^2)$$

where $I(\chi_i, \chi_j)$ is the fraction of ionic character, χ_i is the electronegativity of element i. The second attribute is the mean ionic character:

$$\bar{I} = \sum x_i x_j I(\chi_i, \chi_j)$$

5.3 VGG-like convolutional neural network

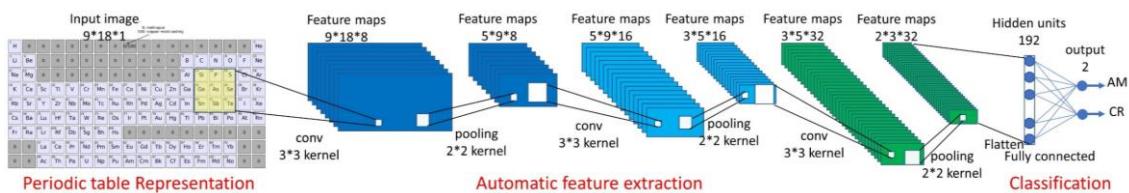


Figure 5.9 Schematic diagram for our VGG-like convolutional neural network for glass-forming ability prediction.

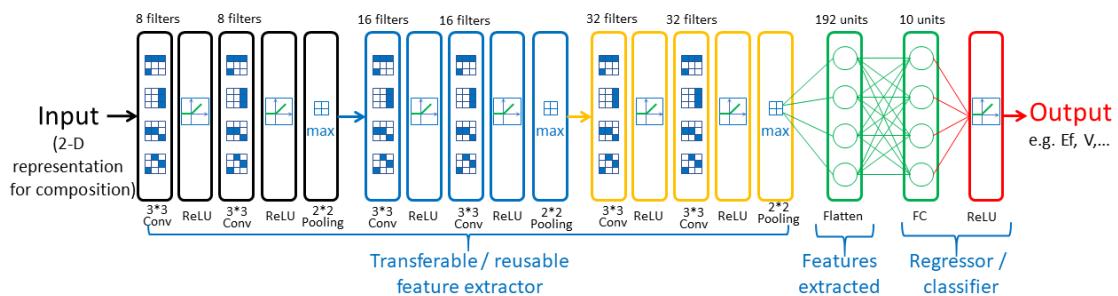


Figure 5.10 Schematic diagram for our VGG-like convolutional neural network for compound properties prediction.

A VGG-like convolutional neural network (in which only 3*3 convolutional kernel was used, and the channels number doubles from previous convolutional layer to the next convolutional layer)¹⁸² were used in automatically extracting features and making classification and regression. Due to the limit of dataset size (16250 for glass-forming ability dataset and 228,676 for OQMD dataset) and small input images, our convolutional neural networks have much fewer layers, channels, and trainable parameters (about 6k for glass-forming ability prediction and about 20k for OQMD) than the well-known VGG-16 (about 133 million)¹⁸². Figure 5.9 and Figure 5.10 show the VGG-

like convolutional neural network we used in this work. The networks consist of two parts. One part is the feature extractor involving the first three pairs of convolutional layers, pooling layers, and ReLU (Rectified Linear Unit) layers which have a nonlinear activation function $f(x) = \max(0, x)$. The other part is the classifier with one fully connected layer and one softmax classification layer in predicting glass-formation ability and the regressor with two full connection layer and one ReLU layer in predicting compounds' forming energy E_f and specific volume V .

The structure details of the VGG-like convolutional neural network are as follows: the size of convolutional filters was $3*3$ for all the three convolutional layers, and the stride was 1. The channel number in convolutional layer doubles from 8 to 16 to 32. Zero-padding was used for the input matrix of convolutional layers. A common type of convolution with a ReLU filter was used, and the value of each convolution kernel was learned through training. Other details of the convolutional neural network are shown in Table 5.4 and Table 5.5.

Table 5.4 Hyperparameter details of CNN1 (atom table representation), CNN2 (randomized periodic table representation) and CNN3 (periodic table representation), CNN2 and CNN3 are assigned with the same hyperparameters.

Order	Layer (type)	Kernel	Stride	CNN1 Param # (atom table representation of 11*11 pixels)	CNN2 & CNN3 Param # (periodic table representation of 9*18 pixels)
1	Conv2d	$8*3*3$	$1*1$	80	80
2	MaxPooling2D	$2*2$	$2*2$	0	0
3	Conv2d	$16*3*3$	$1*1$	1168	1168
4	MaxPooling2D	$2*2$	$2*2$	0	0
5	Conv2d	$32*3*3$	$1*1$	4640	4640
6	MaxPooling2D	$2*2$	$1*1$	0	0
7	Flatten	-	-	0	0
8	Dense	-	-	258	386
Total params				6146	6274
Trainable params				6146	6274
Non-trainable params				0	0

Table 5.5 The summary of our VGG-like convolutional neural network used in the prediction of compounds' forming energy Ef and specific volume V.

Layer (Type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 8, 9, 18)	80
conv2d_2 (Conv2D)	(None, 8, 9, 18)	584
max_pooling2d_1 (MaxPooling2D)	(None, 8, 5, 9)	0
conv2d_3 (Conv2D)	(None, 16, 5, 9)	1168
conv2d_4 (Conv2D)	(None, 16, 5, 9)	2320
max_pooling2d_2 (MaxPooling2D)	(None, 16, 3, 5)	0
conv2d_5 (Conv2D)	(None, 32, 3, 5)	4640
conv2d_6 (Conv2D)	(None, 32, 3, 5)	9248
max_pooling2d_3 (MaxPooling2D)	(None, 32, 2, 3)	0
flatten_1 (Flatten)	(None, 192)	0
dense_1 (Dense)	(None, 10)	1930
dense_2 (Dense)	(None, 1)	11
Total params: 19,981		
Trainable params: 19,981		
Non-trainable params: 0		

5.4 Training details

In predicting glass-forming ability, all models were created and tested using the Keras with Tensorflow as its backend. In the training phase, the output of the shallow neural network and convolutional neural network fitted the ground truth, and the categorical cross-entropy was used as the loss function to evaluate the fitness. The maximum epoch was set to 2000 (loss values almost remain unchanged), and 10-fold cross-validation (so ten models were created after training) was used to evaluate the training/testing accuracy. In prediction, the results of a committee consisted of 10 models were utilized. A series of shallow neural networks using different input vectors of features and hyperparameters, *e.g.*, neuron number in a hidden layer, were trained and optimized. We found shallow neural networks of 30 neurons in the hidden layer did not show better accuracy than shallow neural networks of 20 neurons. Four shallow neural networks (with 20 neurons in the hidden layer) using different feature vectors achieved the best accuracies were compared with convolutional neural networks. All models were trained on a laptop with 2.6 GHz i7 processor and 16 GB memory. Training a convolutional neural network using random Initialization and AdaDelta algorithm on glass-forming ability dataset which had over 16k entries took about 10h under 10-fold cross-validation (batch size 64, 2000 epochs in total).

In prediction of compounds' forming energy E_f and specific volume V , all convolutional neural networks were created and tested using Keras with Tensorflow as its backend. 80% dataset was used for training convolutional neural network and the rest 20% was kept out for testing to evaluate the performance of convolutional neural network. In the training phase, the output of the convolutional neural network fitted the ground truth, and the mean absolute error (MAE) was used as the loss function to evaluate the fitness. The maximum epoch was set to 2000 (loss values almost remain unchanged). All models were trained on a laptop with 2.6 GHz i7 processor and 16G memory. Training a convolutional neural network using random Initialization and AdaDelta algorithm on OQMD dataset which had over 228k entries took 12h (batch size 64, 2000 epochs in total).

5.5 Results of glass-forming ability prediction

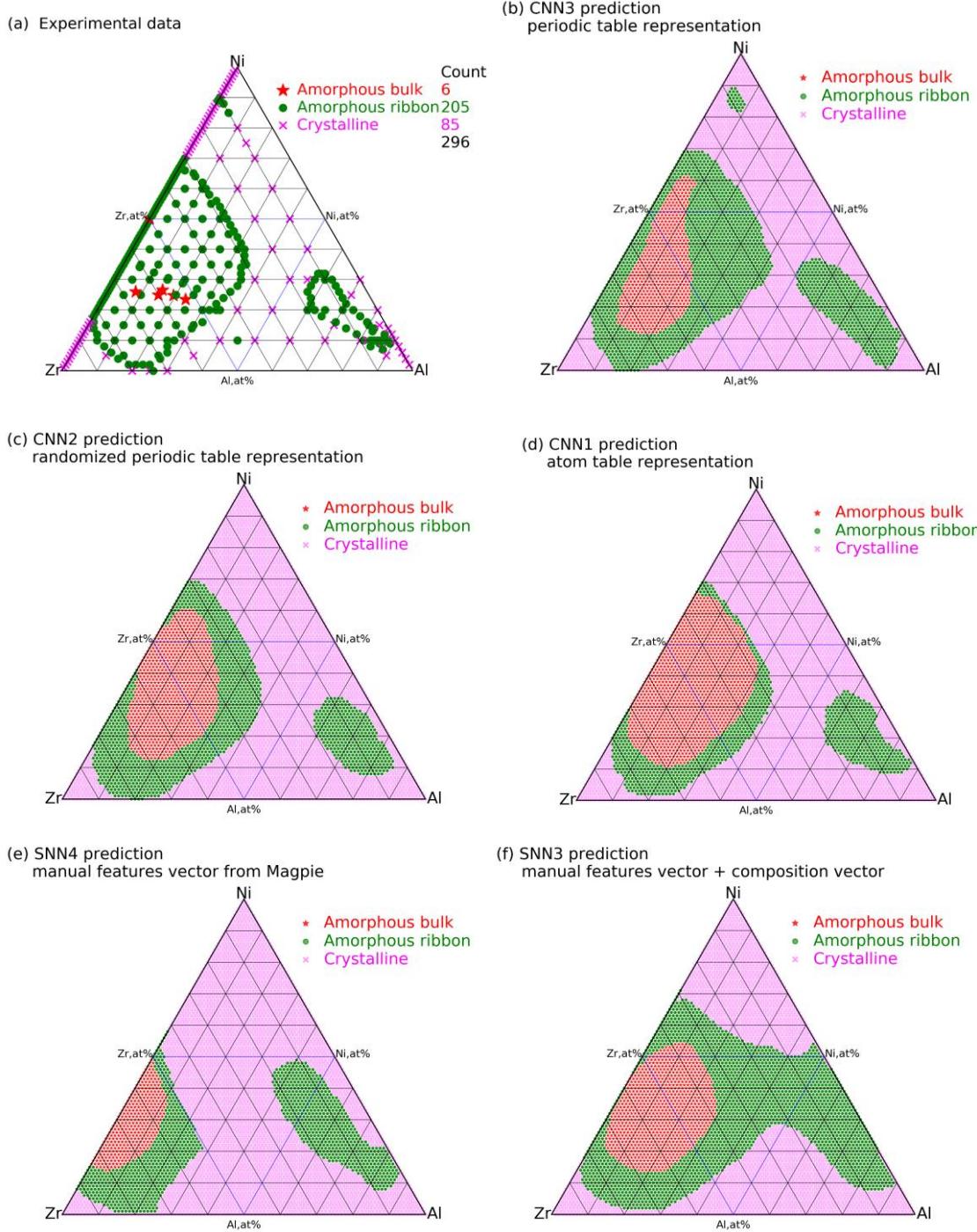
5.5.1 Training and testing results

Table 5.6 shows the average training and testing accuracies of four shallow neural networks (SNNs) and three convolutional neural networks (CNNs) under 10-fold cross-validation in predicting glass-forming ability. SNN1 (using only 14 features derived from empirical rules of metallic glass community) and SNN4 (using 145 general-purpose Magpie descriptors) show the lowest testing accuracy of about 90%. They show a marginal difference in accuracy with Ward's random forest model (89.9% vs. 90%). SNN4 and Ward's random forest model used 145 general-purpose Magpie features, and SNN1 only used 14 features (including one processing parameter, mixing entropy, the statistical information of atomic radius, Pauling electronegativity, bulk modulus, and work function). We found increasing features or even using the full list of features (see Table 5.3) did not bring noticeable improvement in accuracy. SNN2 only used one-hot composition vector as input, but it showed higher accuracy than SNN1 and SNN4. SNN3 used manual features vector plus composition vector as input and it improved the accuracy further. Due to our limited understanding of glass-forming ability and lacking precise property data as input (*e.g.*, ideal solution model and Miedema model were used to calculate alloy mixing entropy and mixing enthalpy, respectively), improving the model's accuracy by adding more pertinent features is impracticable. All four shallow

neural networks show lower accuracies than three convolutional neural networks. Besides convolutional neural networks' accuracy advantage over shallow neural networks, it is also quite convenient to use convolutional neural networks. They only need compositions and processing parameters as input, and they automatically extract features through convolutional layers. CNN3 which refers to a convolutional neural network with periodic table representation shows the highest testing accuracy of 96.3%. The only difference among the three convolutional neural networks is in that the data representations of CNN1 and CNN2 did not have periodic table structure. The advantage of CNN3 over CNN2 and CNN1 is not evident (only 1.3% higher). However, we will demonstrate that CNN3 has more obvious advantages over other models in predicting unseen alloys, *i.e.* better generalization.

Table 5.6 Comparison of average accuracy among different models under 10-fold cross-validation in predicting glass-forming ability .

Model	Data representation	Input size	Algorithms	Average accuracy	
				Training	Testing
Ward's work ⁵⁸	Magpie general-purpose descriptors	145	Random forest	-	90%
SNN1	Manual features vector + processing parameter	13+1	SNN	89.8%	89.9%
SNN2	Composition vector + processing parameter	73+1	SNN	93.2%	92.8%
SNN3	Manual features vector + composition vector + processing parameter	86+1	SNN	93.9%	93.5%
SNN4	Magpie general-purpose descriptors + processing parameter	145+1	SNN	90.1%	90.0%
CNN1	Atom table representation	11*11	CNN	96.4%	95.0%
CNN2	Randomized periodic table representation	9*18	CNN	96.7%	94.9%
CNN3	Periodic table representation	9*18	CNN	96.4%	96.3%



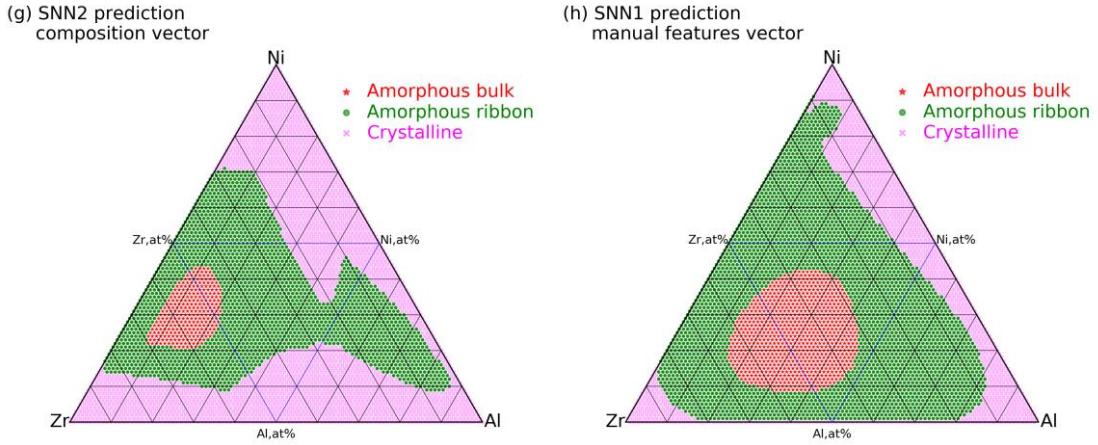


Figure 5.11 (a) Experimental data points of Al-Ni-Zr ternary system in our dataset and the predictions of (b) CNN3, (c) CNN2, (d) CNN1, (e) SNN4, (f) SNN3, (g) SNN2, and (h) SNN1.

Al-Ni-Zr ternary system has 296 entries (include 186 entries from Al-Ni-Zr ternary system, and 110 entries from Al-Ni, Al-Zr, Ni-Zr binary systems) in our dataset. The distribution of these data points is relatively uniform in composition space, see Figure 5.11(a) the ground truth of the Al-Ni-Zr system. So, the Al-Ni-Zr system is quite suitable to validate and compare models. Figure 5.11 (b-h) show the prediction of CNN3, CNN2, CNN1, SNN4, SNN3, SNN2, and SNN1. CNN3 successfully predicted three amorphous composition areas, and the shapes and boundaries of these areas are satisfied when compared with the ground truth. Other models did not predict all three areas. SNN3 did not predict the crystalline area between two amorphous composition areas, *i.e.*, the glass-forming ability of the area was overestimated. CNN2 successfully predicted two amorphous composition areas but missed the small amorphous composition areas near Ni corner. All models (except SNN4) correctly predicted the five bulk metallic glasses in the ground truth and the predicted bulk metallic glasses cover certain area (not some discrete points) around the ground truth points. It is reasonable, researchers commonly reported the optimal bulk metallic glasses only, and bulk metallic glasses candidates (especially before the discovery of bulk metallic glasses) are archived as amorphous ribbon alloys. This sparse and ununiform distribution of bulk metallic glasses data points usually induces bulk metallic glasses data points buried by surrounding densely distributed amorphous ribbon alloys points and omitted as a noise, see Figure 5.11(a). That is why we adjusted the ternary classification into binary classification, *i.e.*, ternary classification easily underestimates alloys' glass-forming ability.

5.5.2 Leave-one-system-out cross-validation results

Table 5.7 Comparison of models' prediction accuracy on unseen alloy systems.

Alloy system	SNN3	SNN4	CNN1	CNN2	CNN3
Leave one system out test on 160 ternary systems	-	76.5%	-	77.3%	83.8%
Leave one system out test on Al-Ni-Zr system	62.3%	63.5%	68.4%	66.4%	80.3%

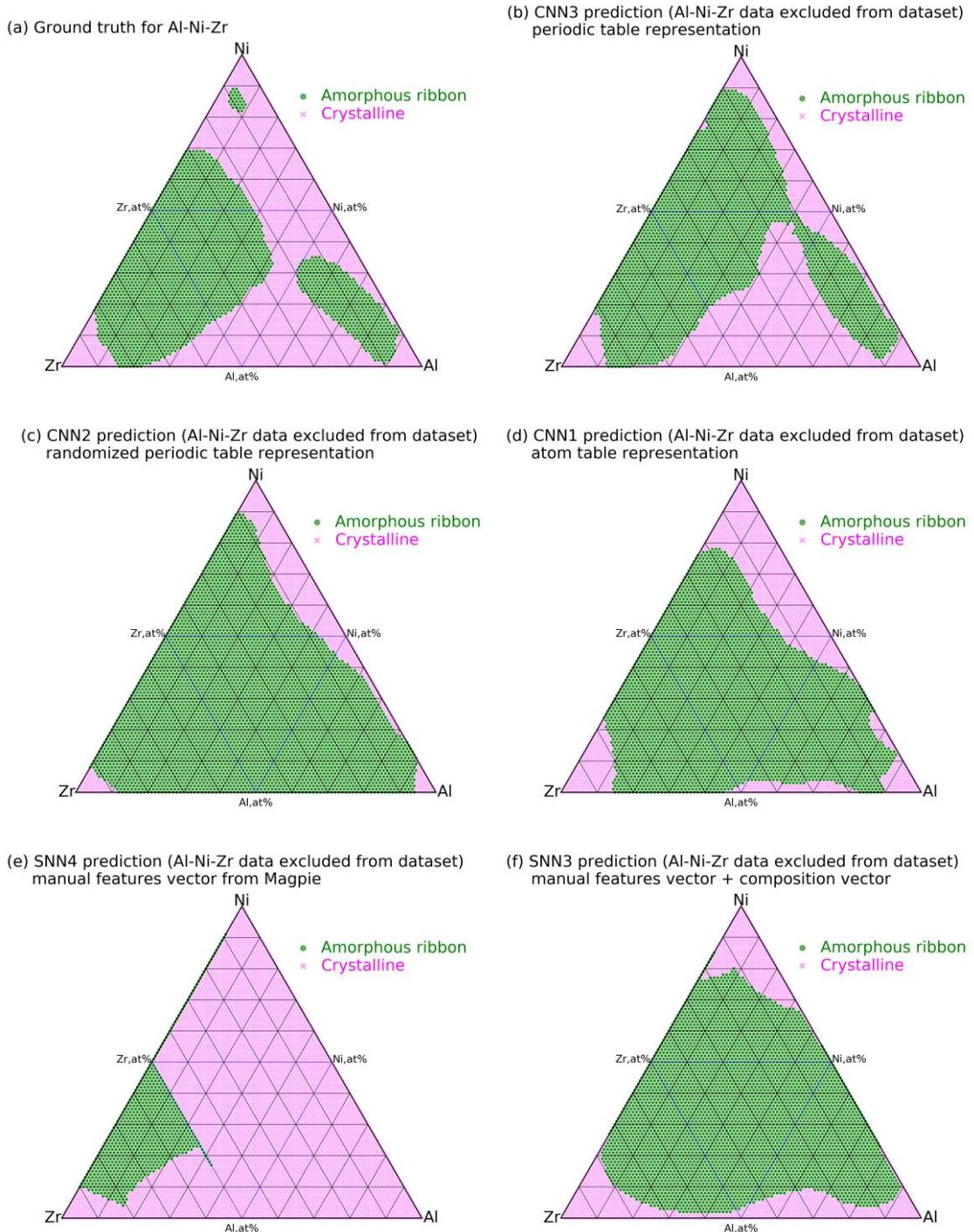


Figure 5.12 The prediction of Al-Ni-Zr ternary system by the re-trained models using a dataset that excluded Al-Ni, Al-Zr, Ni-Zr binary alloys and Al-Ni-Zr-containing multi-component alloys. (a)

ground truth, (b) CNN3, (c) CNN2, (d) CNN1, (e) SNN3, and (f) SNN4.

To validate the predicting ability of models on unseen alloy systems, we carried out a leave-one-system-out (LOSO, like the leave-one-cluster-out cross-validation used by Meredig et al)⁷⁶ cross-validation on 160 ternary systems which has over 40 entries in our dataset. In LOSO cross-validation for a ternary system A-B-C, entries of A-B, A-C, B-C binary alloys and A-B-C ternary alloys were held out as testing dataset. Models were trained with the remaining dataset. The average testing accuracies of SNN4, CNN2 and CNN3 under LOSO cross-validation are shown in Table 5.7. CNN3 outperforms CNN2 and SNN4 in predicting unseen alloy systems by about 7%.

Table 5.7 also shows the LOSO cross-validation results for the Al-Ni-Zr system. Here, we used Al-Ni-Zr amorphous ribbon alloys results (5151 composition points in total, not the 296 Al-Ni-Zr entries in our dataset) in Figure 5.11(b) as ground truth to calculate prediction accuracy. The predictions of CNN3, CNN2, CNN1, SNN4 and SNN2 are shown in Figure 5.12. We can see CNN3 shows accuracy advantage over other CNNs and SNNs by at least 12% when no Al-Ni-Zr data is in training dataset. Overall, this exclusion test strongly verified the advantage of convolutional neural network plus periodic table representation over other models using representation without domain knowledge and using manual featuring engineering. CNN3 can be used to predict the glass-forming ability in alloy systems that are entirely unassessed.

5.5.3 Validation using alloy systems outside our dataset

(1) Validation using Ir-Ni-Ta-(B) and Mg-Cu-Yb bulk metallic glasses

To further validate the prediction ability of the models, we collected some data from newly reported alloy systems, *e.g.* high-temperature Ir-Ni-Ta-(B)¹⁸³ and Mg-Cu-Yb¹⁸⁴ bulk metallic glasses, which are outside our dataset. Our dataset only has one entry about In-Ni-Ta system, *i.e.*, amorphous ribbon composition $In_1Ni_{59}Ta_{40}$; our dataset does not have any data about Mg-Cu-Yb ternary amorphous ribbon alloys or bulk metallic glasses. These two alloys systems were developed using trials and errors guided by empirical criteria. Table 5.8 shows the experimental results and the predictions of our models. The comparison agrees with that in the Al-Ni-Zr system: CNN3 using periodic table representation plus automatic feature engineering shows the highest prediction ability

in unseen systems. SNN4 based on Magpie general purpose descriptors failed in predicting Ir-Ni-Ta-(B) and Mg-Cu-Yb bulk metallic glasses. On the other hand, SNN3 based on manual features derived from empirical criteria successfully predicted all Ir-Ni-Ta-(B) and Mg-Cu-Yb bulk metallic glasses. It shows the influence of feature engineering on performance of a machine learning model. CNN2 using automatic feature engineering but lack domain knowledge in data representation show lower accuracy than CNN3 with periodic table knowledge in representation.

Table 5.8 Comparison of experimental results with the predictions by CNN1, CNN2, CNN3, SNN3, and SNN4 in Ir-Ni-Ta-(B), Mg-Cu-Yb alloys. The red indicates the wrong prediction, and the green indicates the right prediction.

		Experiment	SNN3	SNN4	CNN1	CNN2	CNN3
Ir-Ni-Ta-(B)	Ir ₃₅ Ni ₂₀ Ta ₄₀ B ₅	BMG	BMG	CR	BMG	BMG	BMG
	Ir ₃₅ Ni ₂₅ Ta ₄₀	BMG	BMG	AMR	BMG	CR	BMG
	Ir ₃₃ Ni ₂₈ Ta ₃₉	BMG	BMG	AMR	BMG	AMR	BMG
	Ir ₃₀ Ni ₃₀ Ta ₄₀	BMG	BMG	AMR	BMG	AMR	BMG
	Ir ₂₅ Ni ₃₅ Ta ₄₀	BMG	BMG	AMR	BMG	BMG	BMG
	Ir ₂₅ Ni ₄₀ Ta ₃₅	BMG	BMG	AMR	BMG	BMG	BMG
	Ir ₂₀ Ni ₄₀ Ta ₄₀	BMG	BMG	AMR	BMG	BMG	BMG
Mg-Cu-Yb	Mg ₅₅ Cu ₃₆ Yb ₉	BMG	BMG	BMG	AMR	BMG	BMG
	Mg _{54.5} Cu _{27.3} Yb _{18.2}	BMG	BMG	AMR	AMR	BMG	BMG
	Mg ₅₀ Cu _{36.4} Yb _{13.6}	BMG	BMG	BMG	BMG	BMG	BMG
	Mg ₅₀ Cu ₃₀ Yb ₂₀	BMG	BMG	AMR	AMR	BMG	BMG
	Mg _{45.5} Cu _{36.4} Yb _{18.2}	BMG	BMG	AMR	AMR	BMG	BMG
	Mg ₄₄ Cu _{40.5} Yb _{15.5}	BMG	BMG	AMR	AMR	BMG	BMG
	Mg ₄₃ Cu ₃₀ Yb ₂₇	BMG	BMG	AMR	AMR	BMG	BMG
	Mg _{40.8} Cu _{36.4} Yb _{22.9}	BMG	BMG	AMR	AMR	BMG	BMG
	Mg _{36.4} Cu _{27.3} Yb _{36.4}	BMG	BMG	AMR	AMR	BMG	BMG
	Mg _{33.4} Cu _{33.3} Yb _{33.3}	BMG	BMG	AMR	AMR	BMG	BMG
Score			17/17	2/17	8/17	14/17	17/17

(2) Validation using sulfur-bearing alloys

Since sulfur is a direct neighbor of phosphorous in the periodic table and due to their similar atomic diameter and similar electronegativity on Pauling's scale, sulfur may be assumed to act like phosphorous in glass-forming compositions. Recently, Kuball et.al reported on the formation of sulfur-bearing bulk glass-forming alloys^{185,186}. Figure 5.2 shows that our dataset did not contain any S-containing amorphous ribbon alloys or bulk

metallic glasses. Table 5.9 and Figure 5.13 show the convolutional neural network plus periodic table representation can infer sulfur's beneficial effect on glass-forming ability from its position in periodic table. Researchers have already known the similarity of Ni and Cu in affecting glass forming. Figure 5.13 shows the glass-forming ability curves of $Ti_xZr_{75-x}Ni_{12}Cu_5S_8$, $Ti_xZr_{75-x}Ni_{17}S_8$, and $Ti_xZr_{75-x}Cu_{17}S_8$ almost coincide. SNN3 that based on manual feature engineering did not predict the glass-forming ability of S-containing alloys. It shows the advantage of convolutional neural network plus periodic table representation over models based on manual featuring engineering in predicting glass-forming ability when relevant information is rare or not provided.

Table 5.9 Comparison of experimentally results with the predictions by CNN1, CNN2, CNN3, SNN3, and SNN4 in sulfur-bearing alloys. The red indicates the wrong prediction, and the green indicates the right prediction.

		Experiment	SNN3	SNN4	CNN1	CNN2	CNN3
Sulfur-bearing alloys	$Ti_{75}Ni_{17}S_8$	AMR	CR	AMR	AMR	AMR	AMR
	$Ti_{70}Zr_5Ni_{12}Cu_5S_8$	BMG	CR	AMR	AMR	AMR	AMR
	$Ti_{60}Zr_{15}Ni_{12}Cu_5S_8$	BMG	CR	AMR	BMG	AMR	BMG
	$Ti_{50}Zr_{25}Ni_{12}Cu_5S_8$	BMG	CR	BMG	BMG	BMG	BMG
	$Ti_{40}Zr_{35}Ni_{12}Cu_5S_8$	BMG	CR	BMG	BMG	BMG	BMG
	$Pd_{31}Ni_{42}S_{27}$	BMG	CR	BMG	BMG	BMG	BMG
	$Pd_{37}Ni_{37}S_{26}$	BMG	CR	BMG	BMG	BMG	BMG
	$Zr_{56.5}Ti_{13.3}Ni_{13.6}Cu_{9.6}S_7$	BMG	CR	BMG	BMG	BMG	BMG
	$(Cu_{47}Ti_{34}Zr_{11}Ni_8)_{98.5}S_{1.5}$	BMG	CR	AMR	BMG	BMG	BMG
	$(Zr_{52.5}Cu_{17.9}Ni_{14.6}Al_{10}Ti_5)_{98}S_2$	BMG	CR	BMG	BMG	BMG	BMG
	$(Ni_{62}Nb_{38})_{97}S_3$	BMG	CR	AMR	BMG	BMG	BMG
Score		0/11	7/11	10/11	9/11	10/11	

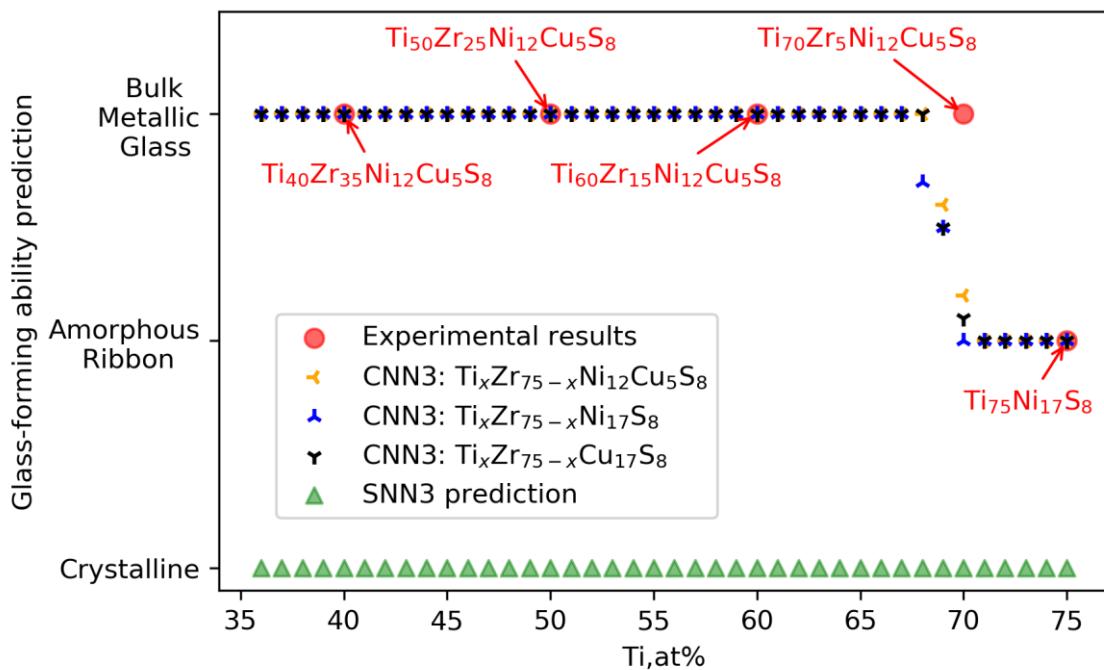


Figure 5.13 Experimental results and the predictions of S-bearing alloys $Ti_xZr_{75-x}Ni_{12}Cu_5S_8$, $Ti_xZr_{75-x}Ni_{17}S_8$, $Ti_xZr_{75-x}Cu_{17}S_8$ ($x=0-35$).

(3) Validation using $RE_6Fe_{72}B_{22}$ and specially designed binary alloys

In this part, we used two groups of easily confused alloys to validate the models' generalization ability. The first group includes 13 $RE_6Fe_{72}B_{22}$ alloys (RE : Sc, Y, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er)¹⁸⁷. Rare earth elements are easily confused elements for their close physical and chemical properties. Experimental results show the simple substitution of similar elements causes the glass-forming ability variations of $RE_6Fe_{72}B_{22}$ alloys. The second group reported by Louzguine-Luzgin contains 18 binary alloys outliers which should be good glass-formers according to empirical criteria but in reality they cannot form an amorphous state even in rapid solidification¹⁵¹. SNN3 based on manual feature engineering shows the worst in predicting glass-forming ability of these two groups alloys (SNN3 overestimates the GFA of 18 alloys in 31 alloys). CNN1 and CNN2 based on automatic feature engineering show better results than SNN3. CNN3 based on automated feature engineering and periodic table knowledge shows the best results (shown in Table 5.10 and Table 5.11).

Table 5.10 Comparison of experimental results with the predictions by CNN1, CNN2, CNN3, SNN3 and SNN4 in $\text{RE}_6\text{Fe}_{72}\text{B}_{22}$. The red indicates the wrong prediction, and the green indicates the right prediction. Superscript GT denotes this composition is in our dataset; superscript * denotes this composition is not in our dataset, but some other compositions of this alloy system are in our dataset.

		Experimental	SNN3	SNN4	CNN1	CNN2	CNN3
$\text{RE}_6\text{Fe}_{72}\text{B}_{22}$	$\text{Sc}_6\text{Fe}_{72}\text{B}_{22}$	BMG	BMG	AMR	BMG	BMG	BMG
	$\text{Y}_6\text{Fe}_{72}\text{B}_{22}^{\text{GT}}$	BMG	BMG	AMR	BMG	BMG	BMG
	$\text{La}_6\text{Fe}_{72}\text{B}_{22}^*$	AMR	BMG	AMR	BMG	BMG	BMG
	$\text{Ce}_6\text{Fe}_{72}\text{B}_{22}$	AMR	BMG	AMR	AMR	BMG	BMG
	$\text{Pr}_6\text{Fe}_{72}\text{B}_{22}$	AMR	BMG	AMR	AMR	AMR	BMG
	$\text{Nd}_6\text{Fe}_{72}\text{B}_{22}^*$	AMR	BMG	AMR	BMG	AMR	AMR
	$\text{Sm}_6\text{Fe}_{72}\text{B}_{22}$	AMR	BMG	AMR	BMG	BMG	AMR
	$\text{Eu}_6\text{Fe}_{72}\text{B}_{22}$	AMR	BMG	AMR	AMR	BMG	AMR
	$\text{Gd}_6\text{Fe}_{72}\text{B}_{22}$	AMR	BMG	AMR	AMR	AMR	AMR
	$\text{Tb}_6\text{Fe}_{72}\text{B}_{22}$	AMR	BMG	AMR	AMR	AMR	AMR
	$\text{Dy}_6\text{Fe}_{72}\text{B}_{22}$	BMG	BMG	AMR	AMR	BMG	BMG
	$\text{Ho}_6\text{Fe}_{72}\text{B}_{22}^*$	BMG	BMG	AMR	AMR	BMG	BMG
	$\text{Er}_6\text{Fe}_{72}\text{B}_{22}$	BMG	BMG	AMR	AMR	AMR	BMG
Score		5/13	8/13	7/13	8/13	10/13	

Table 5.11 Comparison of experimentally measured glass-forming ability with the predictions by CNN1, CNN2, CNN3, SNN3 and SNN4 in 18 binary outliers. The red indicates the wrong prediction, and the green indicates the right prediction. Superscript GT denotes this composition is in our dataset; superscript * denotes this composition is not in our dataset, but some other compositions of this alloy system are in our dataset.

		Experimental	SNN3	SNN4	CNN1	CNN2	CNN3
Outliers according to empirical criteria	$\text{Ag}_{88.5}\text{Y}_{11.5}$	CR	CR	CR	CR	CR	CR
	$\text{Ag}_{71}\text{Y}_{29}$	CR	CR	CR	CR	CR	CR
	$\text{Ag}_{65}\text{Y}_{35}$	CR	CR	CR	CR	CR	CR
	$\text{Ag}_{50}\text{Y}_{50}$	CR	CR	CR	AMR	CR	CR
	$\text{Ag}_{27.5}\text{Y}_{72.5}$	CR	CR	CR	AMR	CR	CR
	$\text{Cu}_{56}\text{Y}_{44}^*$	CR	AMR	AMR	AMR	AMR	AMR
	$\text{Si}_{57}\text{Mg}_{43}$	CR	AMR	CR	CR	CR	CR
	$\text{Si}_{88}\text{Sm}_{12}$	CR	CR	CR	CR	CR	CR
	$\text{Si}_{52}\text{Pd}_{48}^*$	CR	AMR	AMR	CR	CR	AMR
	$\text{Si}_{83}\text{Nd}_{17}$	CR	CR	CR	CR	AMR	CR
	$\text{Si}_{82}\text{Y}_{18}$	CR	CR	CR	CR	AMR	CR
	$\text{Ge}_{84}\text{La}_{16}^*$	CR	BMG	CR	CR	CR	CR
	$\text{Ge}_{64}\text{Mg}_{36}$	CR	BMG	CR	BMG	CR	CR
	$\text{Ge}_{84}\text{Nd}_{16}$	CR	BMG	CR	CR	CR	CR

	$\text{Ge}_{83}\text{Ce}_{17}$	CR	BMG	CR	BMG	CR	CR
	$\text{Ge}_{88}\text{Y}_{12}$	CR	BMG	CR	CR	AMR	CR
	$\text{Ge}_{85}\text{Pr}_{15}$	CR	BMG	CR	CR	AMR	CR
	$\text{Ge}_{85}\text{Sm}_{15}$	CR	BMG	CR	CR	CR	CR
Score			8/18	16/18	13/18	13/18	16/18

In summary, when dataset is large enough (*e.g.* the Al-Ni-Zr system), though the benefit of adding domain expertise to representation is not obvious, the advantage of automatic feature engineering over manual feature engineering is evident; when data is rare or no data is available, domain expertise is vital. Periodic table structure plus CNN, like CNN3, not only brings the convenience of automatic feature engineering but also improves the generalization by introducing background knowledge.

5.6 Results of compound properties prediction

Figure 5.14 shows our convolutional neural network regression accuracy (defined as Pearson correlation coefficient between prediction values and target values) on OQMD testing dataset (20% the dataset, about 45k entries). Based on compositions only, the convolutional neural network can achieve high accuracy of 0.997 in predicting compound forming energy (Ef) and 0.993 in compound specific volume (V). The networks' mean absolute error (MAE) for predicting Ef is 0.083eV/atom and for predicting V is 0.457 Å³/atom, which is better than Ward's result⁵⁸ (MAE is 0.088 eV/atom and 0.563 Å³/atom for Ef and V, respectively) using the same dataset with 145 Magpie manual features plus random forest model. The results show convolutional neural network regression not only eliminates the need for manual features engineering but also improve the testing accuracy (generalization). We believe that feature extractors of these convolutional neural networks can extract a rich set of features (which is the basis for next step transfer learning) as these convolutional neural networks have been well trained (testing accuracy can be as high as 0.99) on a big dataset involving 228k substances and 89 elements.

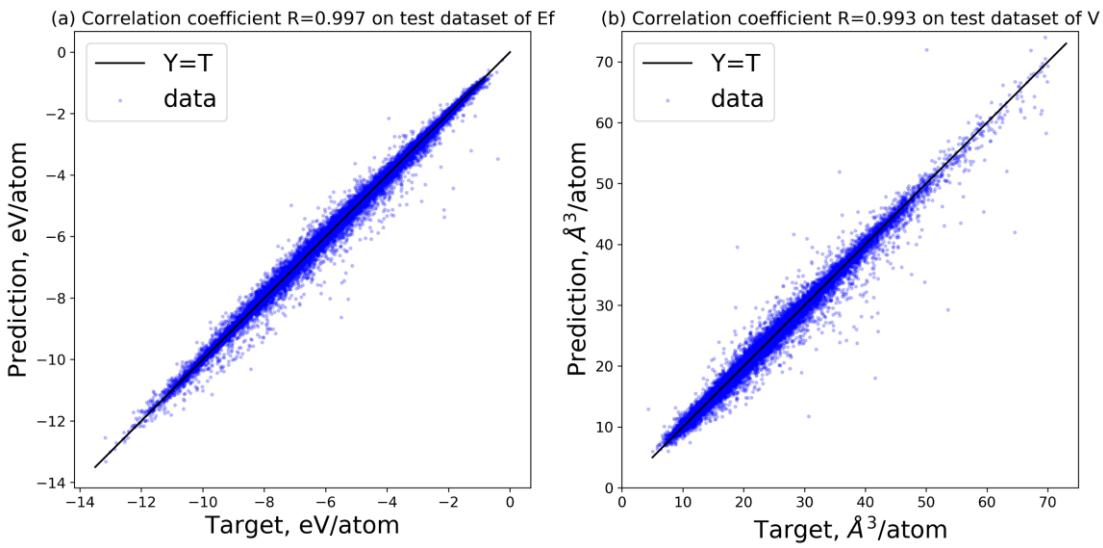


Figure 5.14 The regression analysis between ground truth (horizontal axis, T values) and convolutional neural networks' predictions (vertical axis, Y values) of (a) compounds' forming energy Ef (b) compounds' specific volume. Testing accuracy R is also shown in the figure.

5.7 The benefits of periodic table representation

1 H	2 He	3 Li	4 Be	5 B	6 C	7 N	8 O	9 F	10 Ne									
.. 2.2	1.519 0.98	1.128 1.57	1.976 1	0.773 2.55	0.750 3.04	0.730 3.44	.. 3.98									
.. 14	1.1 453	1.30 1551	1.1 453	1.4 2573	.. 63	.. 54	.. 53	.. 24									
[He]2s ¹	2s ²																	
11 Na	12 Mg	13 Al	14 Si	15 P	16 S	17 Cl	18 Ar	19 Ne	20 Kr									
1.857 0.93	1.601 1.31	1.432 1.61	1.153 1.9	1.060 2.19	1.020 2.58	.. 3.16									
6.3 370	45 922	76 933	100 1683	11 317	7.7 386	1.1 172	.. 83									
[Ne]3s ¹	3s ²																	
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr	
2.310 0.82	1.976 1	1.641 1.36	1.462 1.54	1.316 1.63	1.249 1.66	1.350 1.55	1.241 1.83	1.251 1.88	1.246 1.91	1.278 1.9	1.395 1.65	1.392 1.81	1.240 2.01	1.15 2.18	1.400 2.55	.. 2.96	.. 3
3.1 336	17 1112	57 1814	110 1933	160 2160	160 2130	120 1517	170 1808	180 1768	180 1726	140 1356	70 692	56 302	75.8 1210	22 1090	8.3 490	1.9 265	.. 116
2.5 312	11.89 1042	41 1795	97 2125	170 2741	230 2890	380 2393	180 1825	100 1235	42 429	58 505	42 903	65 722	7.7 386	.. 161
[Ar]4s ¹	4s ²	3d ² 4s ²	3d ² 4s ²	3d ² 4s ²	3d ² 4s ¹	3d ² 4s ²	3d ¹⁰ 4s ²	3d ¹⁰ 4s ²	3d ¹⁰ 4s ²	3d ¹⁰ 4s ²	3d ¹⁰ 4s ²	3d ¹⁰ 4s ²	3d ¹⁰ 4s ²	3d ¹⁰ 4s ²	3d ¹⁰ 4s ²			
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe	
2.440 0.82	2.152 0.95	1.802 1.22	1.603 1.33	1.429 1.6	1.363 2.16	1.360 1.9	1.338 2.2	1.345 2.28	1.375 2.2	1.445 1.93	1.568 1.69	1.659 1.78	1.620 1.96	1.550 2.05	1.452 2.1	.. 2.66	.. 2.6
2.5 312	11.89 1042	41 1795	97 2125	170 2741	230 2890	380 2393	180 1825	100 1235	42 429	58 505	42 903	65 722	7.7 386	.. 161
[Kr]5s ¹	5s ²	4d ⁵ s ²	4d ² 5s ²	4d ⁴ s ¹	4d ⁵ s ¹	4d ⁵ s ²	4d ⁵ s ¹	4d ⁸ s ¹	4d ¹⁰	4d ¹⁰ s ²	4d ¹⁰ s ¹	4d ¹⁰ s ⁵	4d ¹⁰ s ⁵					
55 Cs	56 Ba	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn	
2.650 0.79	2.176 0.89	1.735 1.27	1.578 1.3	1.430 1.5	1.367 2.36	1.375 1.9	1.352 2.2	1.357 2.2	1.387 2.28	1.442 2.54	1.52 2	1.716 1.62	1.750 1.87	1.62 2.02	1.530 2	.. 2.2	.. 2.2	
1.6 301	9.6 1002	48 1936	110 2506	200 3290	310 3695	370 3459	462 3306	2739 230	2041 220	1337 25	234 43	577 46	600 31	544 ..	527 ..	575 ..	202	
[Xe]6s ¹	6s ²	4f ¹⁴ 5d ¹ 6s ²	4f ¹⁴ 5d ² 6s ²	4f ¹⁴ 5d ³ 6s ²	4f ¹⁴ 5d ⁴ 6s ²	4f ¹⁴ 5d ⁵ 6s ²	4f ¹⁴ 5d ⁶ 6s ²	4f ¹⁴ 5d ⁷ 6s ²	4f ¹⁴ 5d ⁸ 6s ²	4f ¹⁴ 5d ⁹ 6s ¹	4f ¹⁴ 5d ¹⁰ 6s ²	[Hg]6p ¹	[Hg]6p ²	[Hg]6p ³	[Hg]6p ⁴	[Hg]6p ⁵		
87 Fr	88 Ra	103 Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs											
.. 0.7	.. 0.9	
.. 300	.. 973	
[Rn]7s ¹	7s ²	5f ¹⁴ 7s ²	5f ¹⁴ 6d ⁷ 7s ²	5f ¹⁴ 6d ⁷ 7s ²	5f ¹⁴ 6d ⁷ 7s ²	5f ¹⁴ 6d ⁷ 7s ²	5f ¹⁴ 6d ⁷ 7s ²	5f ¹⁴ 6d ⁷ 7s ²	5f ¹⁴ 6d ⁷ 7s ²	5f ¹⁴ 6d ⁷ 7s ²	5f ¹⁴ 6d ⁷ 7s ²	5f ¹⁴ 6d ⁷ 7s ²	5f ¹⁴ 6d ⁷ 7s ²	5f ¹⁴ 6d ⁷ 7s ²	5f ¹⁴ 6d ⁷ 7s ²	5f ¹⁴ 6d ⁷ 7s ²		
57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb					
1.879 11	1.825 11.2	1.650 1.13	1.640 1.14	1.630 1.13	1.810 1.17	1.984 1.2	1.801 1.2	1.774 1.22	1.766 1.23	1.756 1.24	1.560 1.25	1.700 1.11						
28 1193	22 1071	29 1204	32 1294	33 1373	38 1345	8.3 1095	38 1586	38.7 1629	41 1685	40 1747	44 1770	45 1818	31 1092					
[Xe]5d ¹ 6s ²	4f ⁵ d ¹ 6s ²	4f ⁶ 6s ²	4f ⁶ 6s ²	4f ⁶ 6s ²	4f ⁶ 6s ²	4f ⁷ 5d ¹ 6s ²	4f ⁶ 6s ²	4f ¹¹ 6s ²	4f ¹² 6s ²	4f ¹³ 6s ²	4f ¹⁴ 6s ²							
89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No					
.. 1.800	.. 1.610	.. 1.420	
..	
[Rn]6d ¹ 7s ²	6d ² 7s ²	5f ² 6d ¹ 7s ²	5f ² 6d ¹ 7s ²	5f ⁴ 6d ¹ 7s ²	5f ⁷ 7s ²	5f ⁷ 7s ²	5f ⁸ 7s ²	5f ¹¹ 7s ²	5f ¹² 7s ²	5f ¹³ 7s ²	5f ¹⁴ 7s ²							

Figure 5.15 The electron configuration, physical and chemical properties of elements show periodic variation in periodic table, e.g. atomic radius decreases and Pauling electronegativity increases from left to right within a period; atomic radius increases and electronegativity

decreases from top to bottom within a group.

Materials properties originated from electrons' behaviors, and periodic table embodies the electron distribution of outer shells. Periodic table has abundant physical and chemical knowledge (see Figure 5.15). Atomic radius, Pauli electronegativity, valence electrons density, and other physical chemistry properties display periodic variations in periodic table. When developing new amorphous alloys, periodic table is often used as a map; similar atom substitution and column substitution are common strategies for improving glass-forming ability. The spatial information or elements' relative position information is difficult to be fully described by manual features engineering (a 1-D vector). The solution is keeping the periodic table structure in representation. So, adding periodic table structure into data representations affords models the ability to infer useful information from the periodic table when direct data is insufficient.

To improve model interpretability and prove the benefit of using representations with periodic table structure, we illustrated the information convolutional neural networks extract from different representations. Visualizing the high dimensional features extracted by convolutional layers, *i.e.*, the intermediate results of convolutional neural networks, is a good way to explore the extracted features. Figure 5.16 - Figure 5.18 show the feature vectors of 108 elements in vector space for (a) CNN3, (b) CNN2, and (c) CNN1. The values of features are represented by color (see the color bar to the right). However, finding the visual and intuitive relationship between elements from these high dimensional features is still very challenging: dimensionality reduction is necessary. Those high dimensional features were compressed by principal component analysis (PCA) and the first two principal components were visualized.

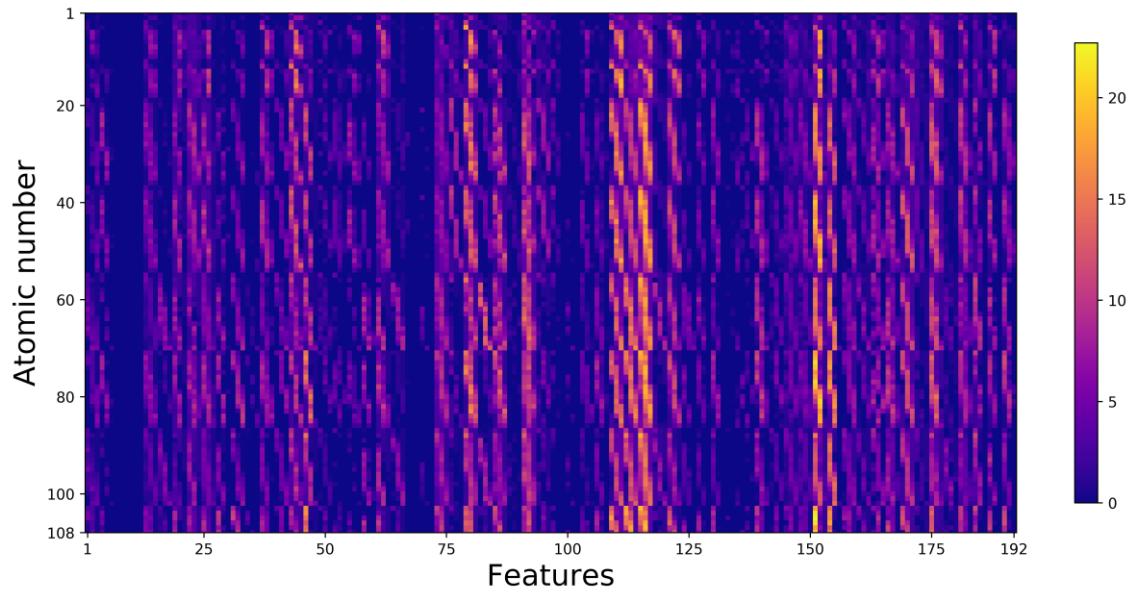


Figure 5.16 Illustration of the 192-dimensional feature vectors of 108 elements extracted by CNN3. The values of features are represented by color (see the color bar to the right).

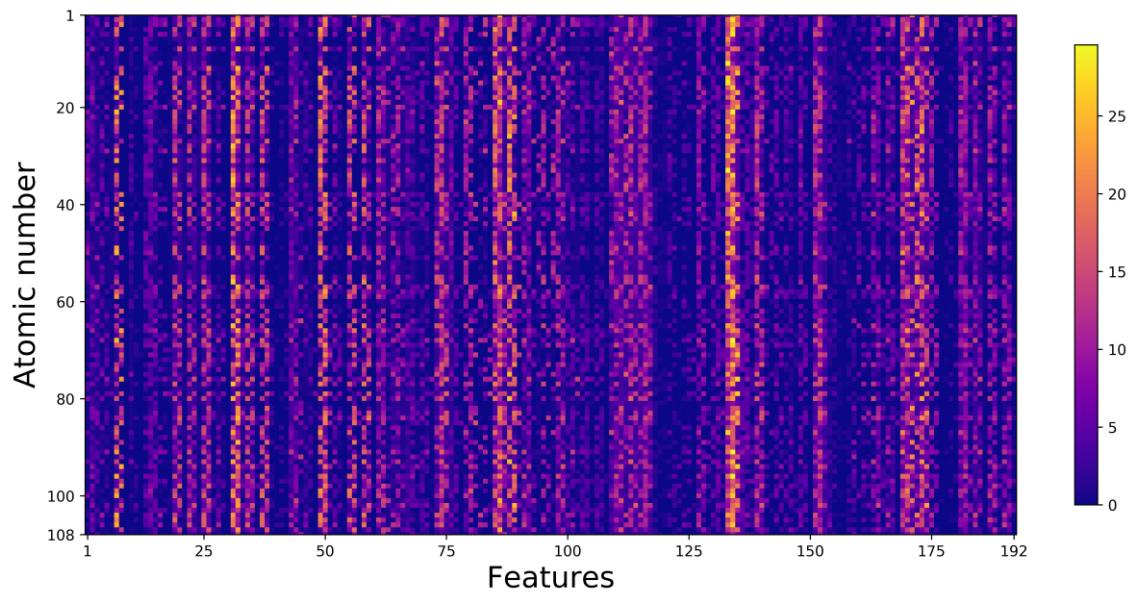


Figure 5.17 Illustration of the 192-dimensional feature vectors of 108 elements extracted by CNN2. The values of features are represented by color (see the color bar to the right).

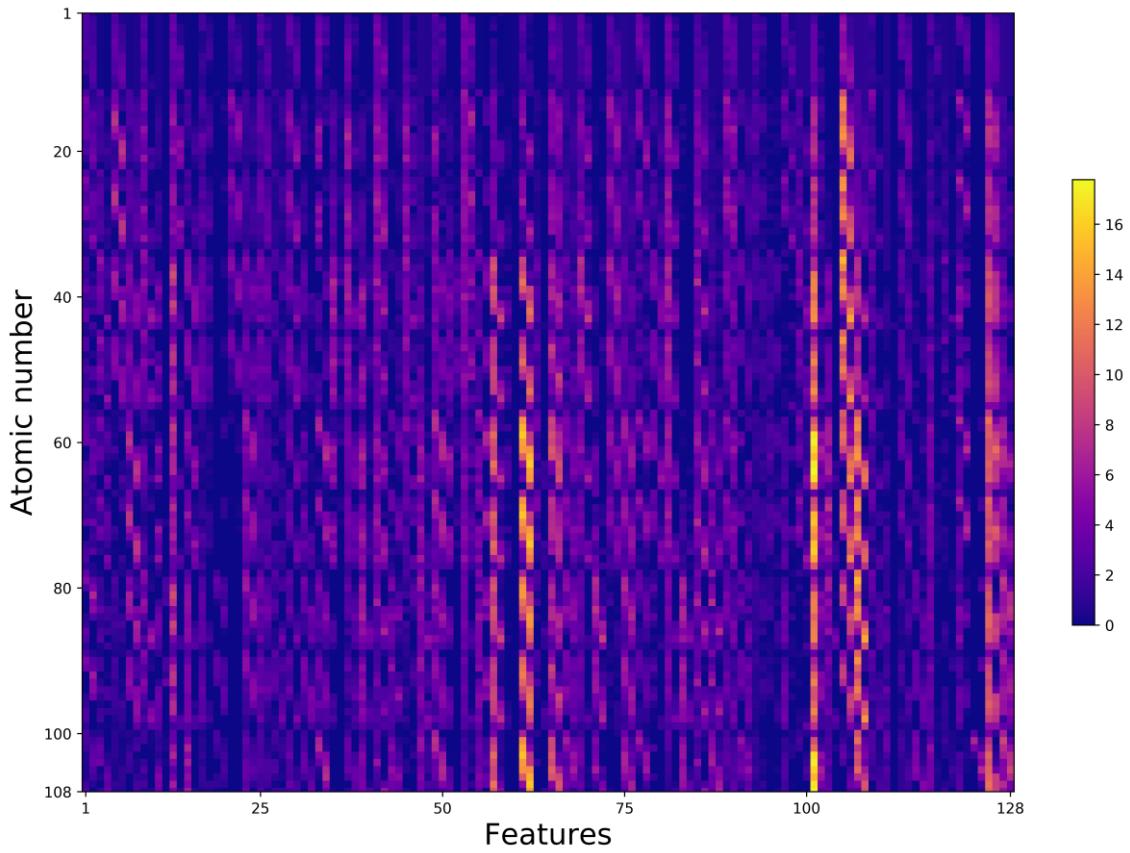
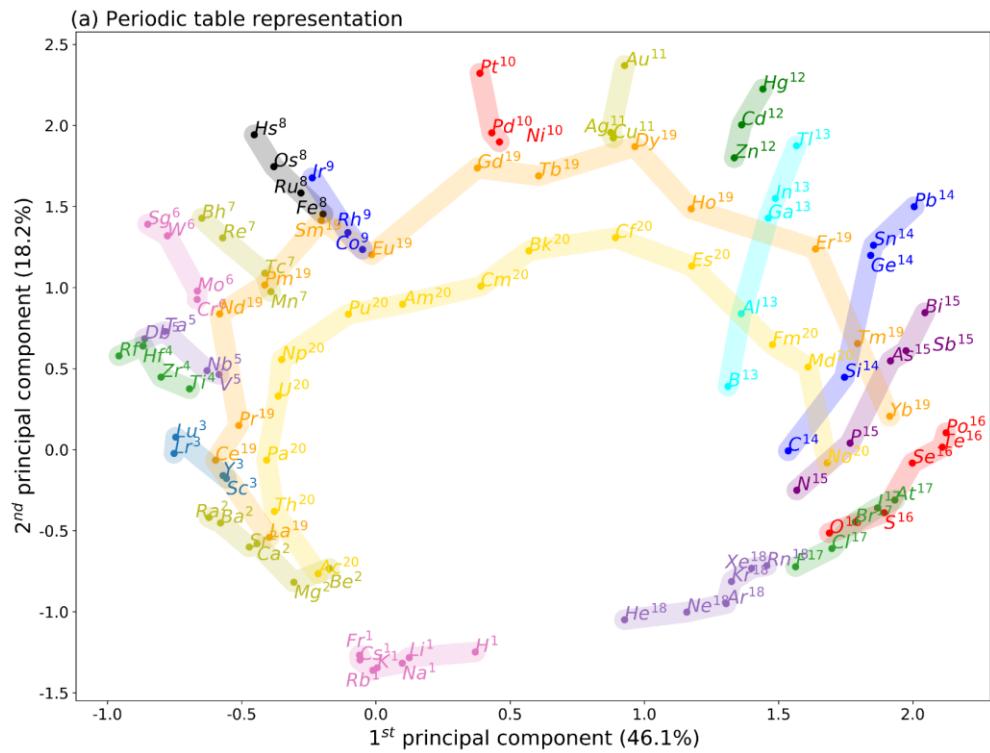


Figure 5.18 Illustration of the 128-dimensional feature vectors of 108 elements extracted by CNN1. The values of features are represented by color (see the color bar to the right).



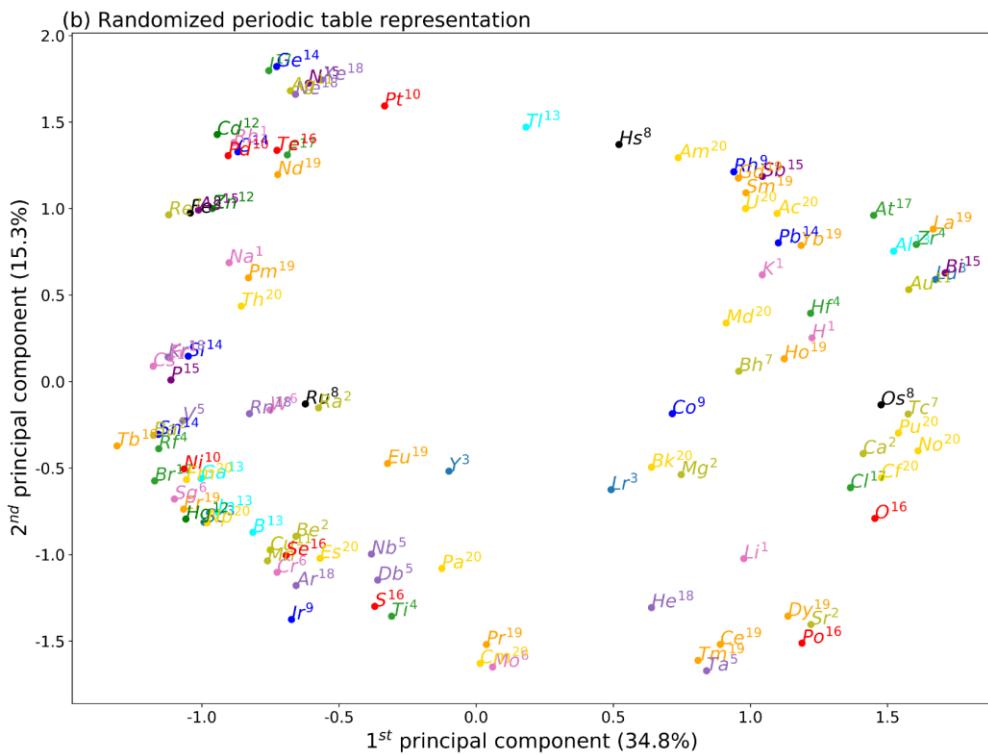


Figure 5.19 Features analysis of the glass-forming ability prediction model after principal component analysis (PCA). Projection of the feature vectors of 108 elements onto the plane spanned by the first and second principal axes. The percentage represents the ratio of the variance on the principal axis direction. Elements are colored according to their groups. (a) periodic table representation; (b) randomized periodic table representation. The superscript 1-18 on element symbol represents the element's group number; superscripts 19 and 20 represent lanthanide and actinides, respectively.

Figure 5.19(a) illustrates the glass-forming ability knowledge of 108 elements in periodic table extracted by convolutional neural network with periodic table representation and it shows apparent periodic trends: elements from 18 groups, lanthanide (group 19) and actinides (group 20) are clustered in different regions (marked with different colors); 18 groups distribute along a semicircle from group 1 to group 18 in sequence; elements from lanthanide and actinides distribute in 2 semicircles with atomic number sequence; elements in one group distributes from semicircle's inside to outside according to ascending atomic number. Our dataset has limited data of more than half the elements in the periodic table and does not have data about alkali metals (group 1), halogens (group 17), noble gases (group 18), but the trends of them are consistent and reasonable. It indicates the periodic table representation transfers the knowledge of periodic table

to the glass-forming ability knowledge, *i.e.*, background knowledge was absorbed by the machine learning models. Figure 5.19(b) illustrates the knowledge extracted by convolutional neural network from representation without periodic table background: no apparent periodic trends can be observed. The differences between Figure 5.19(a) and Figure 5.19(b) indicate the knowledge extracted from randomized periodic table representation is incomplete and fragmented due to our limited training data. The difference explained why CNN3 shows better performance in predicting new data than CNN1 and CNN2: CNN3 using periodic table representation can exploit periodic table' knowledge and make a reasonable inference.

Figure 5.19 (a) can be used for amorphous alloys design. Element substitution is one of the most effective ways to improve the glass-forming ability of amorphous alloys. However, selecting suitable substituting elements is very challenging. Researchers used “column substitution” (substitution with elements of the same group) and “row substitution” (substitution with neighboring elements in the same period) to simplify the process. Neighboring elements in Figure 5.19 (a) have close features, so they are good candidates for element substitution.

Figure 5.20 illustrates the knowledge of 80 elements in periodic table (actinides, noble gases, elements of 7th period are excluded for clarity) generated by feature extractors trained on OQMD dataset. Elements of a group (the group number of each element is shown to the top right of the element symbol) cluster in a small region (different regions / groups marked with different colors in the figure). There are apparent periodic trends in the distribution of these groups: group 1 (alkali metal) to group 17 (halogens) distribute along a circle (not shown in figure) one by one; elements in a group distribute along radial direction according to ascending atomic number; lanthanides distribute in the center of the circle and close to group 3 (elements Sc and Y which are similar to lanthanides in many aspects); all transition metallic elements situate at the up left part; all nonmetallic elements locate at the bottom part. The spatial distribution of elements in Figure 5.20 resemble a 3-D periodic table reflecting some features of elements. These patterns indicate feature extractors have successfully learned knowledge of elements.

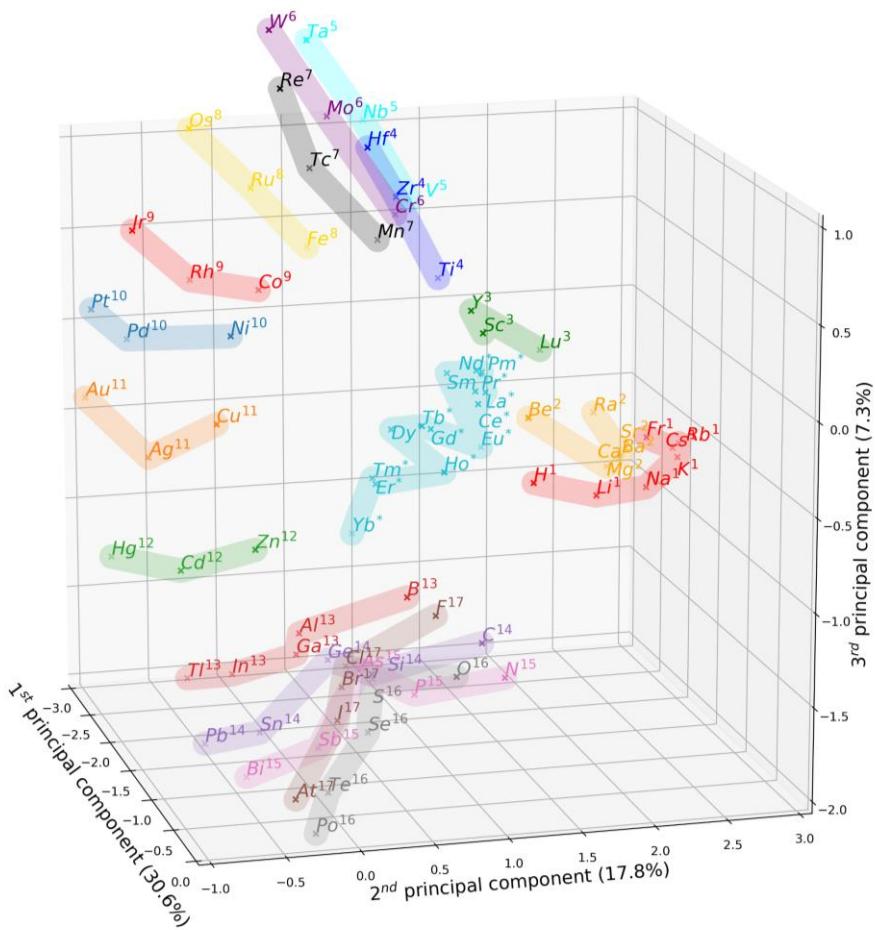


Figure 5.20 Visualization of the first three principal features (after principal component analysis) of 80 pure elements. The percentage represents the ratio of the variance on the principal axis direction. Elements are colored according to their elemental groups. The superscript 1-17 on the element symbol represents the element's group number; superscript * represents lanthanide.

5.8 Summary

Convolutional neural network can be easily used in predicting materials properties: mapping raw data, *e.g.* composition and processing parameters, to pseudo images with some special 2-D structure, *e.g.* periodic table; automatically extracting features and making classification/regression through convolutional neural network.

Case studies on glass-forming ability (GFA), compounds' formation energy (Ef) and specific volume (V) prediction show that: this method not only eliminated the need of the challenging manual features engineering, but also outperformed the known models based on manual feature engineering.

Convolutional neural networks get domain knowledge (*e.g.* periodic table knowledge)

embedded in 2-D representation through learning. Applying convolutional neural networks with enough data, the benefit of adding domain expertise (periodic table structure) to data representation is not obvious. However, domain expertise is vital if data is insufficient.

6 Applications of transfer learning to predicting phase formation in multi-principal element alloys and phase prototypes with small datasets

Overfitting and low accuracy can be anticipated when applying convolutional neural network (that we used in glass-forming ability and OQMD prediction) to small datasets of multi-principal element alloys. In this scenario, we can resort to transfer learning to tackle the challenge of small datasets. In transfer learning, two correlated datasets, *i.e.*, source dataset (usually a big/medium dataset) and target dataset (usually a small dataset) are assembled. The source dataset is used for the transferable feature extractor (the convolutional layers of a convolutional neural network). The feature extractor generates feature vectors for the target dataset. The feature vectors are then used for training shallow classifier/regressor, *e.g.*, random forest, shallow neural network. Figure 6.1 shows workflow for transfer learning from predicting glass-forming ability to predicting phase formation of multi-principal element alloys and transfer learning from predicting inorganic compound's forming energy Ef and specific volume V to predicting inorganic compound's phase prototype.

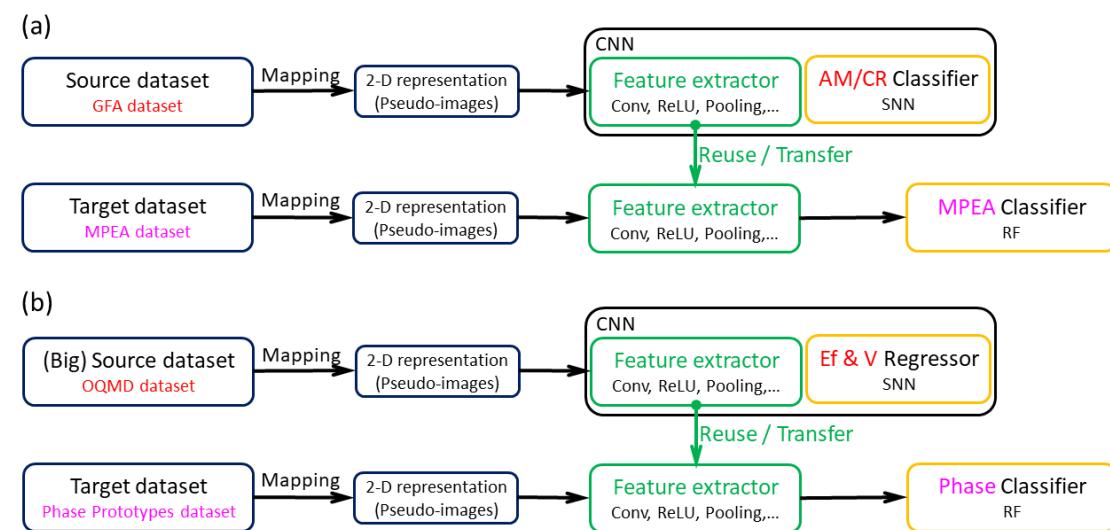


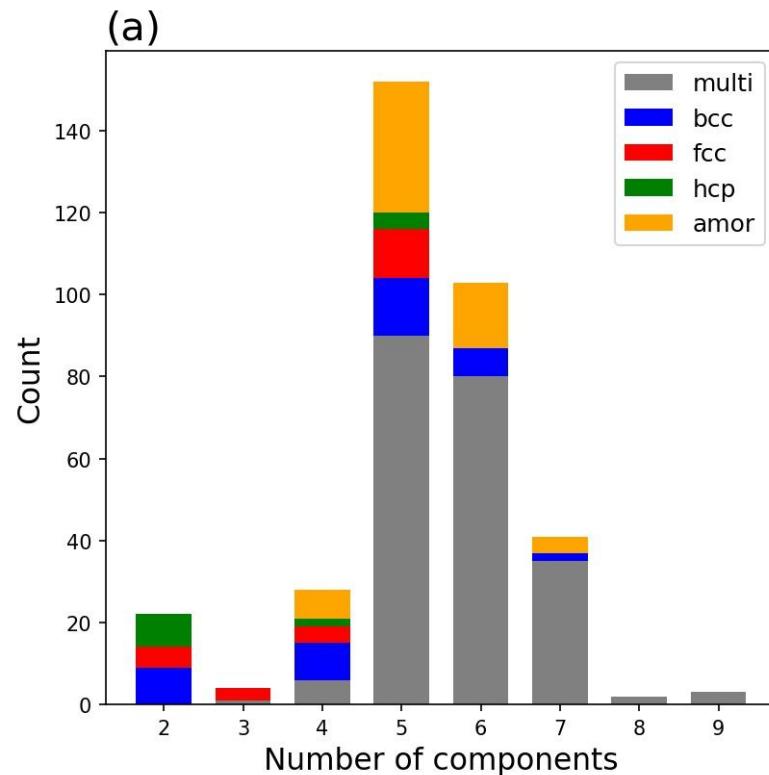
Figure 6.1 The workflow of transfer learning (a) GFA to MPEA; (b) transfer from OQMD to phase prototype. GFA, AM, CR, MEPA, Conv, OQMD, CNN, SNN, RF denotes glass-forming ability, amorphous, crystalline, multi-principal element alloys, convolutional operation, fully connected layer, open quantum materials database, convolutional neural network, shallow neural network,

random forest, supported vector machine, respectively.

6.1 Datasets

6.1.1 Multi-principal element alloys dataset

We used Gao's dataset¹⁸⁸ of multi-principal element alloys (MPEAs) in this study. It has 355 experimental data of multi-principal element alloys. Therein, 41 samples have single body-centered cubic (BCC) phase, 24 samples have single face-centered cubic (FCC) phase, 14 samples have single hexagonal closest packed (HCP) phase, and 59 samples have single amorphous phase. The remaining 217 samples are mixture of multiple phases. Numbers of binary to nonary MPEAs and the proportions of BCC, FCC, HCP, amorphous and multiple phases are shown in Figure 6.2(a). Most of the samples consist of 5 or 6 elements and the single-phase multi-principal element alloys only account for a small fraction. There are 50 elements in the dataset and their occurrence frequencies are shown in Figure 6.2(b). Elements Al, Co, Cr, Fe and Ni appear in more than 190 samples, while elements Ga, Ge Sc, Tc, and Tm only occur once. To build a machine learning model with such a small dataset, so many elements and unbalanced data distribution, is quite difficult.



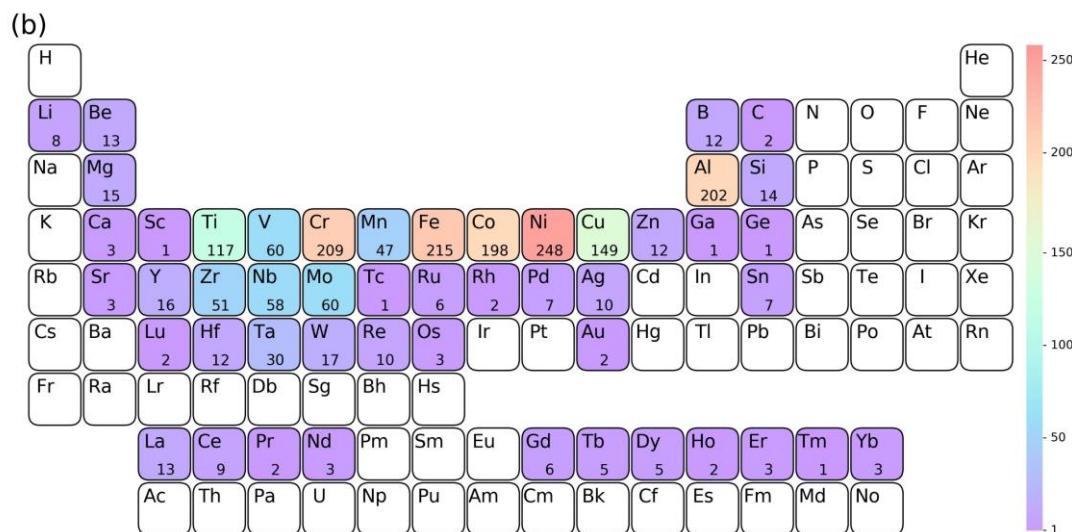


Figure 6.2 Statistics of the 355 multi-principal element alloys in the dataset. (a) The numbers of binary to nonary multi-principal element alloys and the proportions of different phases. (b) The occurrence numbers of elements in the dataset are given under periodic table background. The blank squares, e.g. squares for noble gases, signify the elements not in the dataset.

6.1.2 Phase prototype dataset

We selected phase prototypes data from ‘Handbook of inorganic substance 2016’¹⁸⁹. The handbook which has phase prototypes of more than 160k inorganic substance, belongs to the well-known Pauling files project¹⁹⁰. In the Pauling files project, a phase prototype (crystal structures) was specified in three parts: type-defining compound (e.g., ‘Fe’, ‘ZnS’), Pearson symbol (e.g., ‘cF4’, ‘tP8’) and space group number (1 to 230). For example, the phase prototype of single-phase solid solution alloy ‘Fe_{0.4}Co_{0.2}Ni_{0.4}’ is ‘Cu, cF4, 225’. We selected a subset from the handbook: only prototypes that have more than 20 entries were selected; only room temperature stable phases were kept (high temperature phases, high pressure phases, and low temperature phases were discarded); substances of one element were not included. The final dataset has 17,762 inorganic substances and involves 170 phase prototypes. The elements distribution is shown in Figure 6.3. Most elements (92 out of 108) occur in our dataset, and the frequency of element occurrence is not so uniform like that in OQMD dataset. Figure 6.4 shows the statistics of each phase prototype’s entry number in the dataset. In 170 prototypes, 93 prototypes have entries

less 56, only 3 prototypes *i.e.* ‘MgCu₂, cF24, 227’ (896 entries), ‘NaCl, cF8, 225’ (892 entries) and ‘CeAl₂Ga₂, tI10, 139’ (833 entries) have over 800 entries. The 170 prototypes are detailed in the Table 6.1.

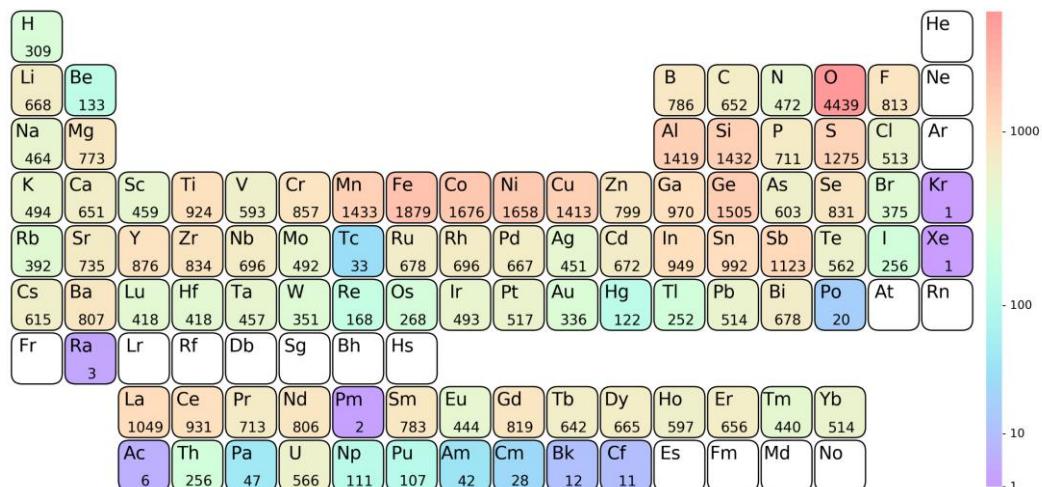


Figure 6.3 Elements distribution in our phase prototypes dataset. The occurrence numbers of elements in the dataset are given under element symbols. The white squares, i.e. square with element symbol and without occurrence number, indicate these elements are not in the dataset.

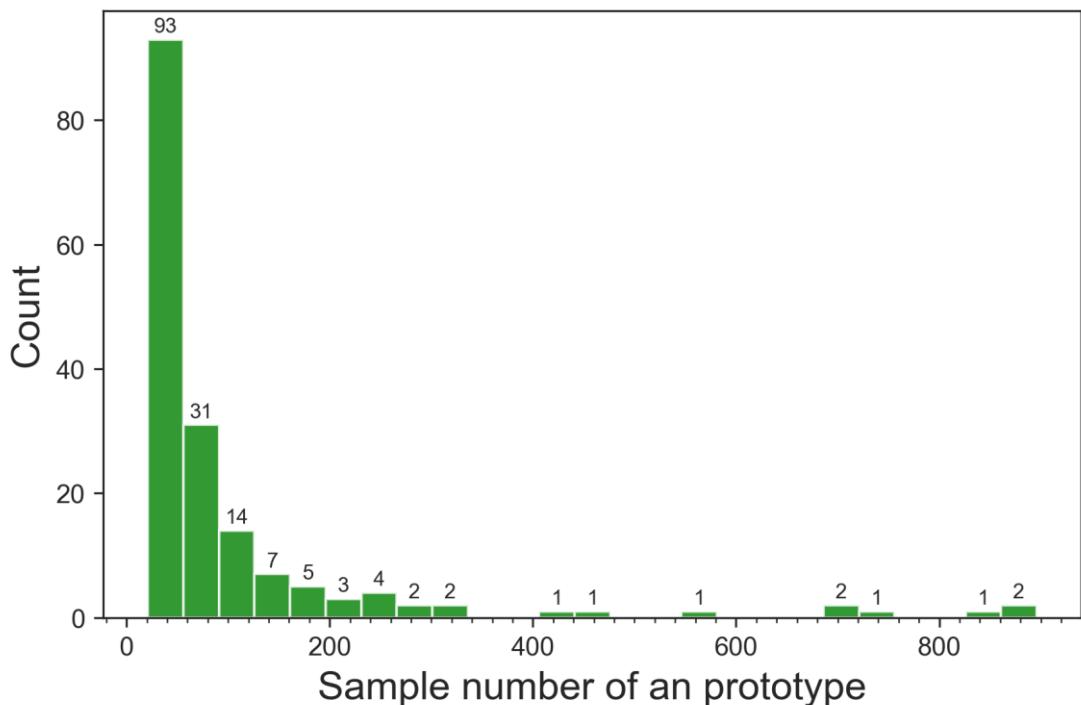


Figure 6.4 Statistics of each phase prototype's entry number in the dataset. In 170 phase

prototypes, 93 prototypes have entries less than 56, and only 3 prototypes have entries more than 800. The label distribution of our dataset is unbalanced.

Table 6.1 Prototype names and sample numbers

Prototype	Size	Prototype	Size
MgCu ₂ ,cF24,227	896	NaCl,cF8,225	892
MgAl ₂ O ₄ ,cF56,227	744	CeAl ₂ Ga ₂ ,tI10,139	833
Ba ₂ CaWO ₆ ,cF40,225	570	CaF ₂ ,cF12,225	701
Cu,cF4,225	443	Ca ₂ Nb ₂ O ₇ ,cF88,227	687
K ₂ NiF ₄ ,tI14,139	319	Cu ₂ MnAl,cF16,225	417
NaFeO ₂ ,hR12,166	275	PbClF,tP6,129	316
Ce ₂ SO ₂ ,hP5,164	255	Mo ₂ FeB ₂ ,tP10,127	290
MgAgAs,cF12,216	213	CeNiSi ₂ ,oS16,63	250
Cr ₃ Si,cP8,223	188	ZnS,cF8,216	245
TiO ₂ ,tP6,136	180	CuHf ₅ Sn ₃ ,hP18,193	243
BeZrSi,hP6,194	165	Gd ₄ RhIn,cF96,216	208
Nd ₂ Fe ₁₄ B,tP68,136	162	CuZrSiAs,tP8,129	208
K ₂ PtCl ₆ ,cF36,225	159	La ₃ Mn _{0.5} Si ₇ ,hP24,173	164
CeCr ₂ Al ₂ O,cF184,227	156	MgCu ₄ Sn,cF24,216	157
CuHfSi ₂ ,tP8,129	149	Al ₂ [OH] _{0.5} F _{0.5} ...,cF72,227	153
LaFe ₄ P ₁₂ ,cI34,204	142	Pr ₂ Mn ₁₇ C _{1.77} ,hR66,166	118
Y ₂ HfS ₅ ,oP32,62	127	K ₂ UF ₆ ,hP9,189	117
KCuZrS ₃ ,oS24,63	121	CuAl ₂ ,tI12,140	117
Zr ₂ Fe ₁₂ P ₇ ,hP21,174	120	YCrB ₄ ,oP24,55	111
CaCu ₃ Mn ₄ O ₁₂ ,cI40,204	112	BaNiO ₃ ,hP10,194	100
CuFeS ₂ ,tI16,122	110	BaNiSn ₃ ,tI10,107	97
MgCuAl ₂ ,oS16,63	106	Ce ₃ Al _{1.67} S ₇ ,hP24,173	95
MgFe ₆ Ge ₆ ,hP13,191	96	Y ₂ Co ₃ Ga ₉ ,oS56,63	94
Ba ₃ (W _{0.5} Fe _{0.5}) ₂ ...,hP30,194	90	Mg ₆ Cu ₁₆ Si ₇ ,cF116,225	88
Yb ₃ Rh ₄ Sn ₁₃ ,cP40,223	89	(Ca _{0.5} La _{0.5}) ₂ Ga...,tP24,113	82
Ho ₆ Mo ₄ Al ₄₃ ,hP106,193	88	FeS ₂ ,cP12,205	81
W ₃ Fe ₃ C,cF112,227	84	Na ₄ Si ₂₃ ,cP54,223	77
Cr ₂ AlC,hP8,194	78	DyFe ₂ SiC,oS20,63	75
CuLa ₃ SiS ₇ ,hP24,173	75	K ₃ NaFeCl ₆ -b,hR66,167	73
CuCe ₂ Ge ₆ ,oS18,38	73	Au ₃ Y ₃ Sb ₄ ,cI40,220	70
Bi ₁₂ SiO ₂₀ ,cI66,197	69	Mg ₂ Ir ₅ SiB ₂ ,tP20,127	68
(Cd _{0.2} Tm _{0.8}) ₅ Mo...,cP96,222	65	Li ₂ AgSb,cF16,216	68
CeCo ₄ B,hP12,191	63	HoCoGa ₅ ,tP7,123	64
Ho ₆ Co ₂ Ga,ol36,71	61	U ₄ Re ₇ Si ₆ ,cI34,229	59
CeCo ₃ B ₂ ,hP6,191	60	CaMgSi ₂ O ₆ ,mS40,15	58
Y ₆ UO ₁₂ ,hR57,148	59	Mo ₆ PbS ₈ ,hR45,148	56
Sc ₇ NCl ₁₂ ,hR60,148	57	CaB ₆ ,cP7,221	56
LiSbF ₆ ,hR24,148	56	Sc ₂ CoSi ₂ ,mS20,12	56
Ho ₂ (Ho _{0.67} Fe _{0.33}) ₃ S ₇ ,mS24,12	56	K ₂ [SO ₄],oP28,62	52
UAs ₂ ,tP6,129	53	LuNi ₂ B ₂ C,tI12,139	52
CuFeO ₂ ,hR12,166	50	Nb ₅ Sn ₂ Si,tI32,140	47
PbSb ₂ O ₆ ,hP9,162	48	Sc ₅ Co ₄ Si ₁₀ ,tP38,127	47

Y ₂ ReB ₆ ,oP36,55	47	Pu ₃ Pd ₅ ,oS32,63	47
U ₂ Mn ₃ Si ₅ ,tP40,128	47	Bi ₂ Te ₃ ,hR15,166	46
Ca ₃ Pl ₃ ,cl56,214	46	Zn ₁₄ Gd ₂ Co ₃ ,hR57,166	45
Ho ₂ CoGa ₈ ,tP11,123	45	LaMnSi ₂ ,oS16,63	45
ZrFe ₄ Si ₂ ,tP14,136	44	NaBe ₄ SbO ₇ ,hP26,186	44
Cs ₃ CoCl ₅ ,tI36,140	43	U ₂ PbSe ₅ ,mP32,14	41
NdB ₂ ClO ₄ ,tP8,123	43	W ₆ Fe ₇ ,hR39,166	41
DyMn ₂ O ₅ ,oP32,55	41	TlPd ₅ As,tP7,123	40
U ₂ Pt ₂ Sn,tP20,136	40	Rb _{0.81} W ₃ O ₉ ,hP26,193	39
Ho ₁₁ Ge ₁₀ ,tI84,139	40	Ca ₃ (Ga _{0.67} Ge _{0.3} ...,hP23,150	38
CeNiC ₂ ,oS8,38	39	Al ₂ O ₃ ,hR30,167	37
La ₆ Co ₉ (Co _{0.5} Ga _{0.5}) ₄ Ga,tI80,140	39	YbFe ₂ Al ₁₀ ,oS52,63	37
HgMo ₆ Cl ₁₄ ,cP84,201	38	Zr(Zr _{0.67} Bi _{0.33}) ₃ F ₁₅ ,hR114,167	36
Sb ₂ S ₃ ,oP20,62	38	Tl _{0.39} V ₃ S ₄ ,hP16,176	36
Sc ₃ Ce ₂ Si ₄ ,oP36,62	37	Y ₄ PdGa ₁₂ ,cl34,229	34
Tl _{0.83} V ₅ S ₈ ,mS28,12	36	Sm ₁₂ Ni ₆ In,cl38,204	34
KTb ₃ F ₁₂ ,tI32,87	35	Y ₆ RuI ₁₀ ,aP17,2	33
Ni ₃ P,tI32,82	35	YCo ₅ P ₃ ,oP36,62	33
Sc ₅ Co ₁₉ P ₁₂ ,hP37,189	34	Cs ₃ Cr ₂ Cl ₉ ,hP28,194	33
K ₃ Nb ₈ O ₂₁ ,hP64,193	33	Pu ₃ Pd ₄ ,hR42,148	33
Lu ₅ Ni ₂ In ₄ ,oP22,55	32	CeNi ₈ (Ni _{0.5} Si _{0.5})Si ₄ ,tI56,140	32
FeClO,oP6,59	32	Zr ₆ Ni ₂₀ P ₁₃ ,hP39,174	32
La ₃ Ni ₂ Ga ₂ ,oP28,57	31	Mg ₂ [SiO ₄],oP28,62	32
Pu ₂ C ₃ ,cl40,220	30	LuMnO ₃ ,hP30,185	32
Hf ₉ Mo ₄ B,hP28,194	30	CeAsS,mP12,14	31
La ₁₂ Re ₅ C ₁₅ ,hP32,189	30	KGdNb ₆ Cl ₁₈ ,hR81,148	31
CeFe ₂ Al ₈ ,oP44,55	29	Er ₃ CrB ₇ ,oS44,63	30
YNi ₂ Al ₃ ,hP18,191	29	Ba ₂ Cu ₄ YO ₈ ,oS30,65	30
Na ₃ (Na _{0.67} Sr _{0.3} ...,hP22,186	29	ScU ₃ S ₆ ,oP40,58	30
U ₃ Ni ₄ Si ₄ ,oI22,71	29	Cu ₂ Sb,tP6,129	30
Tl ₄ PbTe ₃ ,tI32,140	28	K _{0.5} Mn ₄ O ₈ ,tI26,87	30
NaPt ₃ O ₄ ,cP16,223	28	BaHg ₁₁ ,cP36,221	29
BaTiSi ₃ O ₉ ,hP28,188	28	CsMn ₂ P ₆ O ₁₈ ,mS56,12	29
Pr ₃ WCl ₃ O ₆ ,hP26,176	27	Sc _{0.6} Fe ₂ Si _{4.9} ,hP20,194	29
FeAs ₂ ,oP6,58	26	Lu ₂ Co ₃ Si ₅ ,mS40,15	28
La ₂ Sb,tI12,139	26	K ₂ Mg ₂ [SO ₄] ₃ ,cP76,198	27
CaC ₂ ,tI6,139	24	CaNb ₂ Bi ₂ O ₉ ,oS56,36	25
LiCaAlF ₆ ,hP18,163	23	SmZrF ₇ ,mP18,4	25
MoS ₂ ,hP6,194	22	Cs ₃ Tl ₂ Cl ₉ ,hR84,167	24
La _{4.67} [SiO ₄] ₃ O,hP42,176	22	Cu ₃ Ti,oP8,59	24
KAsF ₆ ,hR24,148	22	Na _{0.5} CoO ₂ ,hP10,194	23
UCl ₃ ,hP8,176	21	Ba[SO ₄],oP24,62	23
Bil ₃ ,hR24,148	21	GdSi _{1.4} ,oI12,74	21

6.2 Details of transfer learning

Two types of datasets were assembled in transfer learning: the big source dataset (e.g. OQMD dataset) and the smaller target dataset (phase prototypes dataset); source

dataset was used to train convolutional neural networks (in order to obtain the transferable and reusable feature extractors), feature vectors of the target dataset were generated by the reusable feature extractor; classification/regression was made based on the generated features by some shallow learning machines (random forest in this work) *e.g.* random forest (RF), support vector machine (SVM), shallow neural network (SNN). The features extractors can be directly used in many small new target datasets. Training new convolutional neural networks from scratch based on the new datasets is not necessary. Only new classifiers and regressors need to be built and trained.

The two machine learning tasks, *i.e.* predicting glass-forming ability (GFA) and predicting phases of multi-principal element alloys (MPEAs), have different output domain (amorphous/crystalline binary classification in glass-forming ability prediction and 5 phases labels in multi-principal element alloys prediction) and highly correlated (or overlapped) input domain from the point of transfer learning: Figure 5.2 and Figure 6.2(b) show common elements in those alloys are similar; some amorphous alloys are also multi-principal element alloys; the descriptors developed in conventional machine learning for glass-forming ability and multi-principal element alloys can be shared (*e.g.* atomic size difference, mixing enthalpy, mixing entropy, difference in Pauling electronegativities, and valence electron concentration). So, we believe that the automatic feature extractors of the well-trained convolutional neural networks, which have outperformed known manual features in glass-forming ability prediction, will work in multi-principal element alloys prediction too. In transfer learning from glass-forming ability to multi-principal element alloys, compositions of multi-principal element alloys were fed into the well-trained CNN1, CNN2, CNN3, and the intermediate results (high dimensional features yielded from convolutional layers) of these convolutional neural networks were extracted. Then these features were used in new classifier (here we used random forest for its good interpretability, and it needs very little hyperparameters optimization) as input. Stratified data division strategy (to ensure training and testing dataset have similar data distribution) and Scikit-learn package were used in training.

The well-trained feature extractors (the left part of the convolutional neural networks for predicting Ef and V) on OQMD dataset were reused in distinguishing 170 phase prototypes of inorganic materials. In transfer learning, chemistry composition of a

substance was mapped to a 2-D pseudo-image and fed into the two feature extractors (one from convolutional neural network for Ef and the other from convolutional neural network for V), and two feature vectors of 192-dimensional were yielded. Then these features vectors were used as input in training phase prototypes classifier (random forest in our work). We trained three random forest models, the first model took feature vector generated by Ef feature extractor (denoted as $\text{Vector}^{\text{Ef}}$) as input, the second model took feature vector generated by V feature extractor (denoted as Vector^{V}) as input, the third one with a input of the combination of $\text{Vector}^{\text{Ef}}$ and Vector^{V} (denoted as $\text{Vector}^{\text{Ef}\&\text{V}}$). For comparison, we also trained a random forest model using composition vector (one-hot coding), a 108-dimensional vector representing 108 elements' atomic percentage in a substance, as input (denoted as $\text{Vector}^{\text{comp}}$). Random forest in Scikit-learn package was used in transfer learning for prediction of phase prototypes because it has good interpretability, and it need very little hyper-parameters optimization. We only regulated the maximum decision tree number in the random forest (100 was used in our work) and the maximum depth of a tree (20 was used in our work). Stratified data division strategy was used in dataset division to ensure training and testing dataset have similar data distribution.

All models were built with Scikit-learn and trained on a laptop with 2.6 GHz i7 processor and 16G memory. Training a convolutional neural network using random initialization and AdaDelta algorithm on OQMD dataset which had over 228k entries took 12h (batch size 64, 2000 epochs in total). Training a random forest on our multi-principal element alloys dataset which had 355 entries and 5 categories only took less than 1s. Training a random forest on our phase prototypes dataset which had over 17k entries and 170 categories only took about 10s. The transfer learning has great speed advantage over training machine learning models from scratch.

6.3 Testing results

6.3.1 Testing accuracy on multi-principal element alloys dataset

Table 6.2 shows the scores of our transfer learning models on multi-principal element alloys dataset under 5-fold cross-validation. Figure 6.5, Table 6.3, and Table 6.4 show the ROC (receiver operating characteristic) curves, confusion matrix, and performance

metrics, respectively. Our model without resorting to any manual features is capable of distinguishing BCC, FCC, HCP, amorphous, and multiple-phase mixture with 5-fold cross validation scores (accuracy, recall, precision and F1) over 94% after training and test. We should bear in mind that when labels' distribution is unbalanced like that of our multi-principal element alloys data, achieving high recall, high precision and high accuracy at the same time is very difficult. We can see model transferred from CNN3 has the highest scores which indicates that periodic table representation is also beneficial for transfer learning. The proposed transfer learning model is an upgrade for conventional machine learning relying on manual feature engineering. It could serve as an effective guide for designing new multi-principal element alloys.

Table 6.2 The scores of transfer learning on multi-principal element alloys dataset under 5-fold cross-validation with three data representations

Data representation	Transfer from	Accuracy	Precision	Recall	F1 Score
Periodic table representation	CNN3	0.935	0.940	0.935	0.934
Randomized periodic table representation	CNN2	0.854	0.837	0.854	0.831
Atom table representation	CNN1	0.884	0.888	0.884	0.875

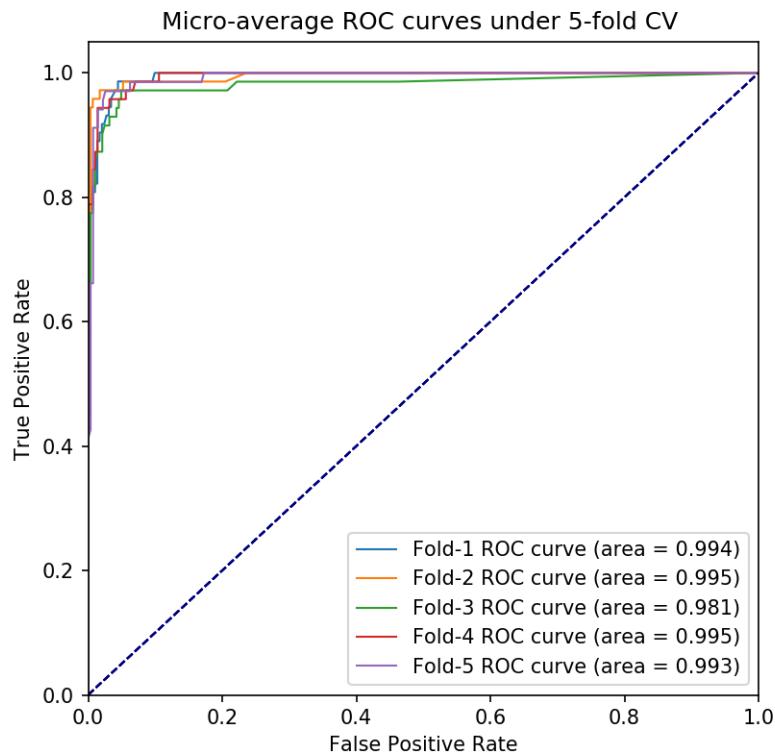


Figure 6.5 ROC curves of the transfer learning models for predicting multi-principal element alloy

phases under 5-fold cross-validation.

Table 6.3 Confusion matrices of transfer learning models for predicting multi-principal element alloy phases on testing datasets under 5-fold cross-validation.

Fold 1		Predicted				
		multi	fcc	amorphous	bcc	hcp
Actual	multi	42	1	0	1	0
	fcc	1	4	0	0	0
	amor	2	0	10	0	0
	bcc	1	0	0	8	0
	hcp	0	0	0	0	3
Fold 2		Predicted				
		multi	fcc	amorphous	bcc	hcp
Actual	multi	44	0	0	0	0
	fcc	0	4	0	0	1
	amor	0	0	12	0	0
	bcc	1	0	0	7	0
	hcp	0	0	0	0	3
Fold 3		Predicted				
		multi	fcc	amorphous	bcc	hcp
Actual	multi	41	0	1	1	0
	fcc	2	3	0	0	1
	amor	1	0	11	0	0
	bcc	1	0	0	7	0
	hcp	1	0	0	0	2
Fold 4		Predicted				
		multi	fcc	amorphous	bcc	hcp
Actual	multi	43	0	0	0	0
	fcc	0	5	0	0	0
	amor	2	0	10	0	0
	bcc	2	0	0	6	0
	hcp	0	0	0	0	3
Fold 5		Predicted				
		multi	fcc	amorphous	bcc	hcp
Actual	multi	41	0	0	2	0
	fcc	1	3	0	0	0
	amor	1	0	10	0	0
	bcc	0	0	0	8	0
	hcp	0	0	0	0	2

Table 6.4 The performance metrics of transfer learning models for predicting multi-principal element alloy phases under 5-fold cross-validation

Fold 1	precision	recall	f1-score	support
multi	0.91	0.95	0.93	44
fcc	0.80	0.80	0.80	5
amorphous	1.00	0.83	0.91	12
bcc	0.89	0.89	0.89	9
hcp	1.00	1.00	1.00	3
macro avg	0.92	0.90	0.91	73
weighted avg	0.92	0.92	0.92	73
accuracy	0.92			73
Fold 2	precision	recall	f1-score	support
multi	0.98	1.00	0.99	44
fcc	1.00	0.80	0.89	5
amorphous	1.00	1.00	1.00	12
bcc	1.00	0.88	0.93	9
hcp	0.75	1.00	0.86	3
macro avg	0.95	0.93	0.93	72
weighted avg	0.98	0.97	0.97	72
accuracy	0.97			72
Fold 3	precision	recall	f1-score	support
multi	0.89	0.95	0.92	43
fcc	1.00	0.60	0.75	5
amorphous	0.92	0.92	0.92	12
bcc	0.88	0.88	0.88	8
hcp	1.00	0.67	0.80	3
macro avg	0.94	0.80	0.85	71
weighted avg	0.91	0.90	0.90	71
accuracy	0.90			71
Fold 4	precision	recall	f1-score	support
multi	0.91	1.00	0.96	43
fcc	1.00	1.00	1.00	5
amorphous	1.00	0.83	0.91	12
bcc	1.00	0.75	0.86	8
hcp	1.00	1.00	1.00	3
macro avg	0.98	0.92	0.94	71
weighted avg	0.95	0.94	0.94	71
accuracy	0.94			71
Fold 5	precision	recall	f1-score	support
multi	0.95	0.95	0.95	43
fcc	1.00	0.75	0.86	4
amorphous	1.00	0.91	0.95	11
bcc	0.80	1.00	0.89	8
hcp	1.00	1.00	1.00	2
macro avg	0.95	0.92	0.93	68
weighted avg	0.95	0.94	0.94	68
accuracy	0.94			68

6.3.2 Testing accuracy on phase prototypes dataset

Figure 6.6 shows the four models' classification accuracy versus test ratio (the ratio of test dataset to the entire dataset) on phase prototypes dataset. All four models are built on chemistry compositions only. However, the accuracy advantage of models based on transfer learning over the model based on simple composition vector is obvious. When testing ratio is 0.1 (*i.e.* 90% of the dataset are for training and 10% of the dataset are hold out for testing), transfer learning model based on $\text{Vector}^{\text{Ef}\&\text{V}}$ can reach accuracy of 0.9, while non-transfer-learning model based on $\text{Vector}^{\text{comp}}$ only got accuracy of 0.54. The accuracy of transfer learning model is relatively insensitive to the test ratio. When the test ratio increases from 0.1 to 0.5 (*i.e.* the ratio of training dataset to entire dataset decreases from 0.9 to 0.5), the accuracy only decreases from 0.9 to 0.85. The insensitivity of transfer learning models to the size of training dataset indicates it has good generalization and suitable for small datasets. The accuracy differences between 3 transfer learning models are small (about 0.01 - 0.03). The model based on $\text{Vector}^{\text{Ef}\&\text{V}}$ shows the highest accuracy because it can access features from both feature extractors.

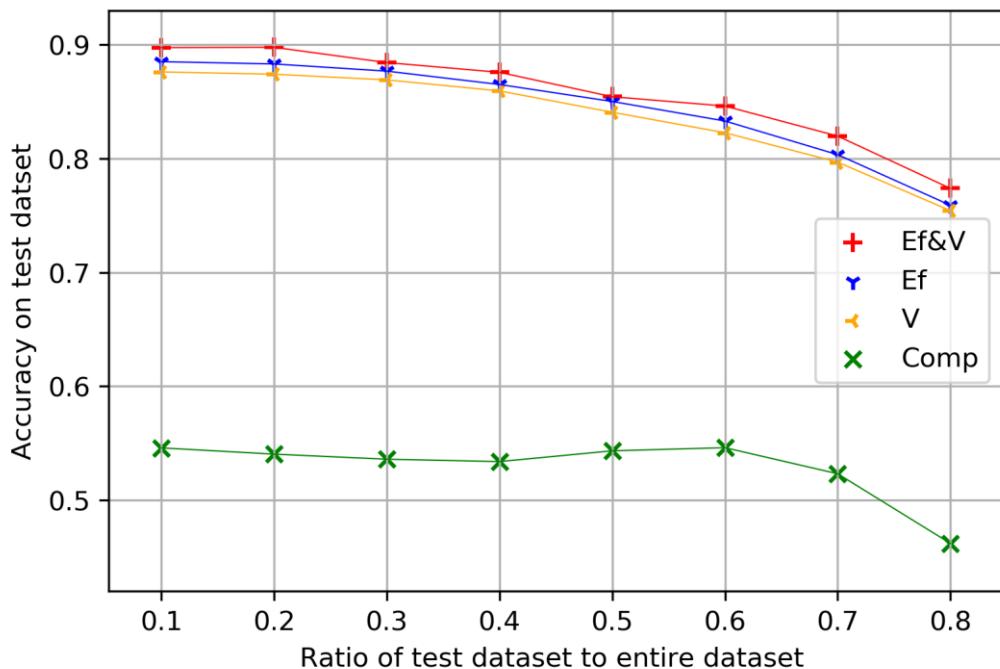


Figure 6.6 Accuracy vs. testing ratio of three transfer learning models using feature vectors $\text{Vector}^{\text{Ef}\&\text{V}}$, $\text{Vector}^{\text{Ef}}$, Vector^{V} , and one non-transfer-learning model using composition vector $\text{Vector}^{\text{comp}}$.

6.4 Visualization of the automatically-extracted features

Challenges in predicting phase prototypes based on composition include small sample volume (small dataset) for half prototypes in our dataset, imbalanced data distribution (see Figure 6.4 and Table 6.1), and large number of categorical labels (170 labels in our work, as far as we know in the published works¹⁷² researchers discriminated no more than ten structures in a machine learning task). Manual feature engineering plus conventional machine learning were considered at the beginning. However, our dataset involves a variety of inorganic substances, *e.g.* solid solution alloys, metal intermetallics, oxides, sulfide, halide. It is quite challenging to build features manually (which require a deep understanding of these substances). Furthermore, some physical data, *e.g.* atomic radius, bulk modulus used in manual features engineering is unavailable for some elements in our dataset which is difficult to fill reliably by hand. For these reasons, we resorted to transfer learning.

To gain insight into the extracted features and explaining why transfer learning is effective, we attempted to visualize features vectors generated by the transferable feature extractors and to find some patterns from these visualizations. However, uncovering the relationship visually and intuitively from these high dimensional features is quite challenging. There will be a 192*17,762 matrix (*i.e.* one substance has a 192-dimensional feature vector) for our phase prototypes dataset, and it necessities a dimensionality reduction of these features. We compressed these high dimensional features data by principal component analysis (PCA). We visualized the first 4 principal components to (partially) represent the multi-principal element alloy knowledge yielded by the well-trained feature extractors on glass-forming ability dataset. We visualized the first 3 principal components to (partially) represent the phase prototypes knowledge yielded by the well-trained feature extractors on OQMD dataset.

6.4.1 Features analysis for multi-principal element alloys

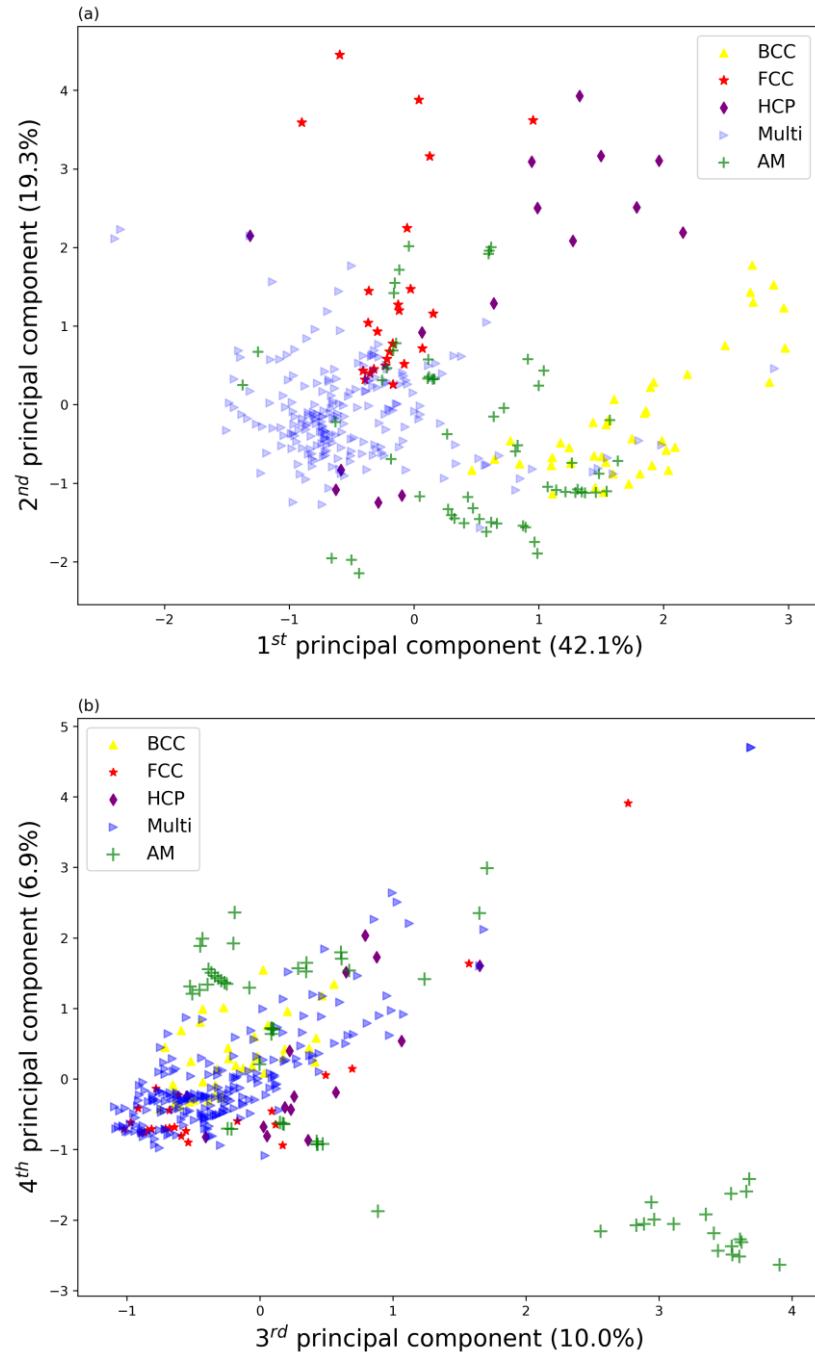


Figure 6.7 The multi-principal element alloys' first four principal features generated by glass-forming ability model. Alloys are colored according to their phases. The percentage represents the ratio of the variance on the principal axis direction.

The features of 355 multi-principal element alloys generated by glass-forming ability model were shown in Figure 6.7. We can see that alloys of the same phases tend to cluster in the diagrams. Based on the first and second principal features, we can

intuitively distinguish stable BCC, FCC, HCP and multi-phase alloys. Most alloys of metastable amorphous phases can be discriminated from other alloys of stable phases by the third and fourth principal features visually. It indicates transfer learning from glass-forming ability to multi-principal element alloys is successful and justifies the high scores of our model for multi-principal element alloys.

6.4.2 Features analysis for phase prototypes

Because it is impossible to illustrate the knowledge of all 170 prototypes and over 17k substances in a figure (even impossible in many figures due to the nature of a large number of high-dimensional vectors), we choose a few phase prototypes (some common prototypes and some uncommon prototypes) to illustrate their distribution in 3-D space using the first three principal components of their feature vectors. The four common phase prototypes (shown in Figure 6.8) which have hundreds of entries in our dataset (they are shown in big symbols in the word cloud, see [Error! Reference source not found.](#)) are ‘MgCu₂, cF24, 227’, ‘Ca₂Nb₂O₇, cF88, 227’, ‘Cu, cF4, 225’, and ‘ZnS, cF8, 216’. Many thousands of inorganic substances in total contained in four different prototype clusters to four different regions (marked with different colors). They can easily be discriminated only by the first three principal components, which shows that feature extractors are highly effective.

The 15 uncommon phase prototypes (shown in Figure 6.9) which have limited entries (20-40 entries) in our dataset (they are shown in very small symbols in the word cloud see [Error! Reference source not found.](#)) are ‘CaB₆, cP7, 221’, ‘YbFe₂Al₁₀, oS52, 63’, ‘Pr₃WCl₃O₆, hP26, 176’, ‘Y₄PdGa₁₂, cI34, 229’, ‘YCo₅P₃, oP36, 62’, ‘K₃Nb₈O₂₁, hP64, 193’, ‘Y₆RuI₁₀, aP17, 2’, ‘KGdNb₆Cl₁₈, hR81, 148’, ‘KAsF₆, hR24, 148’, ‘Cs₃Tl₂Cl₉, hR84, 167’, ‘Ba₂Cu₄YO₈, oS30, 65’, ‘Hf₉Mo₄B, hP28, 194’, ‘Er₃CrB₇, oS44, 63’, ‘CsMn₂P₆O₁₈, mS56, 12’, ‘U₃Ni₄Si₄, oI22, 71’. Substances of different prototype cluster to different regions (marked with different colors). The first three principal components can be easily used in distinguishing the 15 uncommon phase prototypes. It further validates the effectiveness of the features yielded by the transferable feature extractors.

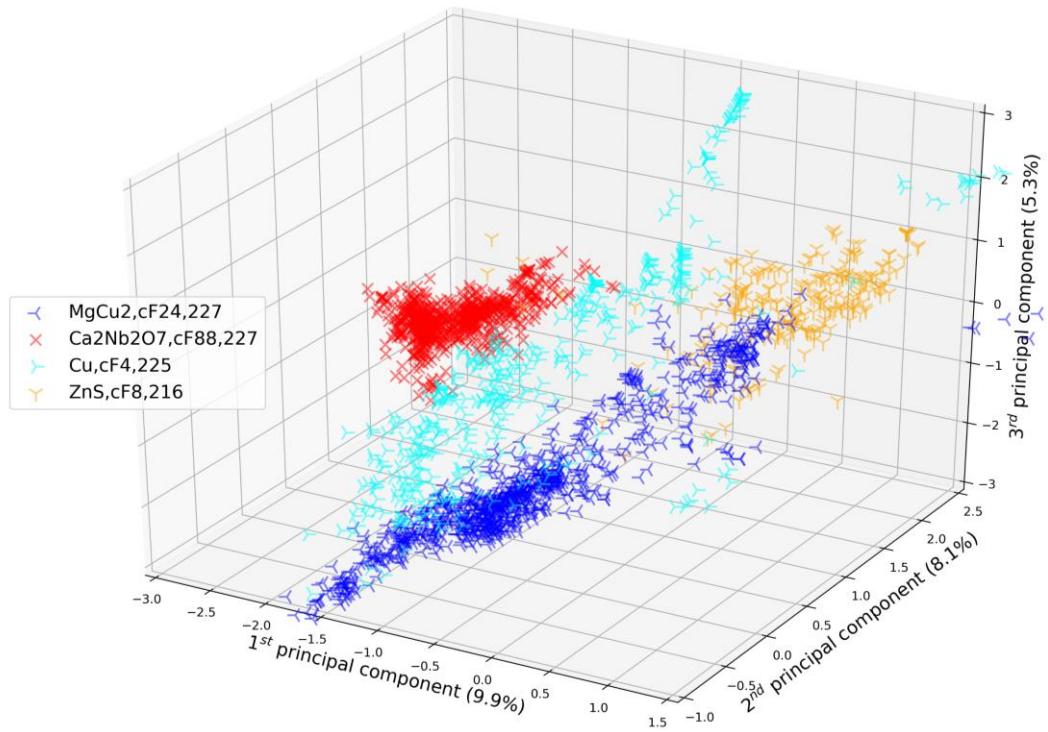


Figure 6.8 Visualization of the first three principal features (after principal component analysis) of substances belonging to four common phase prototypes in our dataset. Inorganic substances are colored according to their phase prototypes. The percentage represents the ratio of the variance on the principal axis direction.

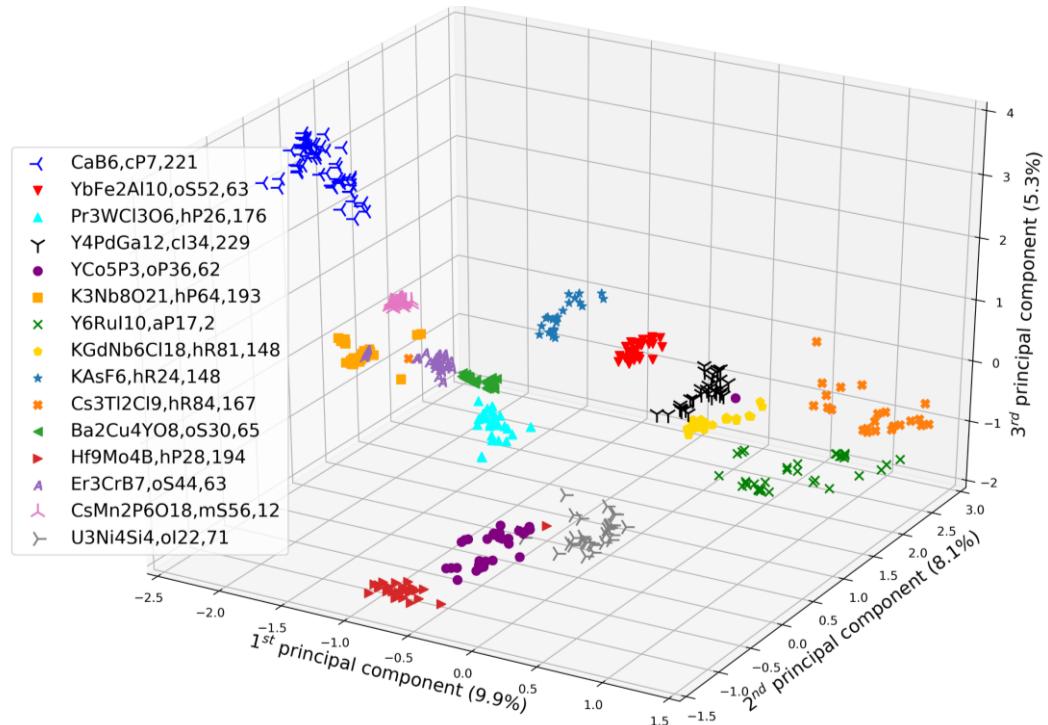


Figure 6.9 Visualization of the first three principal features (after principal component analysis) of substances belonging to 15 uncommon phase prototypes in our phase prototypes dataset.

Inorganic substances are colored according to their phase prototypes. The percentage represents the ratio of the variance on the principal axis direction.

Figure 6.8 and Figure 6.9 show that only using three principal components, *i.e.*, three automatic generated principal features, we can discriminate thousands of substances belonging to dozens of prototypes. It indicates transfer learning from predicting Ef and V using OQMD dataset to predicting phase prototypes is successful and justifies the high scores of our model for prediction of phase prototypes. It can be concluded that basis of good application of transfer learning in this case is the sharing of mutual features between two domains, one with a large dataset and other with a small dataset. The root of mutual features can be attributed to the fundamental laws of physics and chemistry that all inorganic substances follow (*i.e.*, different materials have some features in common). So, some general and mutual features exist that can be used for describing them, which is the basis of transfer learning. We should bear in mind that we have 384 features in total. Excluding some irrelevant features (*i.e.*, features only relevant to predicting forming energy Ef and specific volume V), the rest are common features and generally useful in new applications.

6.5 Summary

Convolutional neural network cannot work well on small datasets for its internal parameters of a large number need to be determined through learning. Transfer learning can partially solve this challenge.

The feature extractors of convolutional neural networks that are well-trained on big datasets (*e.g.*, OQMD, glass-forming ability) can be reused directly in new tasks to generate a rich set of general and effective features without the challenging manual feature engineering. In our study, transfer learning was effectively exploited to build a phase prototypes classifier which could accurately discriminate 170 phase prototypes and a phase classifier which could accurately distinguish BCC, FCC, HCP, amorphous, and multiple-phase mixture in multi-principal element alloys.

Transfer learning not only greatly decreases training time of new models, but also improves accuracy and generalization of new models with small datasets. In addition, periodic table knowledge embedded in data representations and knowledge shared

between models is beneficial for transfer learning tasks with small datasets. It can be easily used to create machine learning models quickly with small datasets and limited domain expertise for new materials designing, screening and development.

7 Conclusions and future work

7.1 Conclusions

In this thesis, three deep learning techniques (*i.e.*, compact fully connected deep neural network plus pre-training, 2D periodic table data representation plus convolutional neural network, convolutional neural network plus transfer learning) were proposed to tackle the challenges in machine learning for materials research: small datasets and challenging manual feature engineering. Five case studies were carried out to validate their effectiveness: predicting solidification cracking susceptibility with a small dataset using compact fully connected deep neural network; predicting glass-forming ability and compound's forming energy and specific volume using periodic table representation plus convolutional neural network; predicting compound's phase prototype and phases of multi-principal element alloys through transfer learning with small datasets. The conclusions are shown as follows.

Through machine learning, vast scattered experimental data hidden in literature can produce simple quantitative expression: material property as a function of chemistry composition and processing parameters.

Though big dataset plus deep learning is an ideal solution, small dataset plus compact deep neural network and pre-training is a reasonable choice for small datasets are common in material research.

Combination of deep neural network and tree-based models can bring us predictions of high accuracy and a certain insight of mechanisms.

Mapping chemistry composition and processing parameters of materials to 2-D pseudo-images enables convolutional neural network process conventional structured data. This bypass the challenging manual feature engineering in machine learning. The feature extractors of convolutional neural networks that are well-trained on big datasets can be reused directly in new tasks to generate a rich set of general features. Transfer learning not only dramatically decreases training time of new models, but also improves accuracy and generalization of new models with small datasets.

Convolutional neural networks get domain knowledge (*e.g.*, periodic table knowledge)

embedded in 2-D representation through learning. When dataset is large enough, the benefit of adding domain expertise to 2-D data representation is not apparent. However, domain expertise is vital for tasks of small datasets (limited data). Adding domain expertise to data representation improves models' generalization. Periodic table knowledge and is beneficial for the performance of convolutional neural networks and transfer learning models with limited data.

The proposed three deep learning techniques is an upgrade for conventional machine learning and could serve as an effective guide for the development of new materials.

7.2 Future work

This thesis mainly deals with the challenges of small datasets and manual feature engineering when applying machine learning to materials problems. Future works are listed as follows.

There are many alternatives to stacked auto - encoder in pre-training deep neural network, *e.g.* restricted Boltzmann machine and some empirical equations. Those should be tried and compared in future and some guidelines in training compact deep neural networks with small datasets can be summarized.

Periodic table representation is an effective method to inject knowledge into a machine learning model. To further improve the generalization of a deep learning model and to reduce the dependence on the dataset volume, we can add more domain expertise to data representation. How to effectively inject knowledge to 2-D data representation is an open question.

In this thesis, the uncertainty of predictions was not given due to our datasets came from multiple sources and had different fidelities. In this scenario, how to estimate the prediction error/uncertainty proposes a big challenge for machine learning.

We partially solved the problems of unbalanced/biased labels distribution in the glass-forming ability dataset through transforming ternary classification to binary classification. However, biased sampling is common in materials science. How to assign each sample with different weight (importance) is another challenge in machine learning for materials science.

How to gain more insights into machine learning models (improve models' interpretability) is a big challenge for machine learning. Materials researchers need more tools to interpret their models and to understand the physical mechanisms of materials behaviors (transform black-box models into grey-box models).

Data and code availability

The datasets and codes used to generate the results in this thesis are available at <https://github.com/sf254/SC>, <https://github.com/sf254/glass-forming-ability-prediction>, and <https://github.com/sf254/phase-prototypes-prediction>.

Reference

1. Alpaydin, E. *Introduction to Machine Learning*. (MIT Press, 2014).
2. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. (Springer New York, 2013).
3. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning*. (MIT Press, 2016).
4. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
5. Carrasquilla, J. & Melko, R. G. Machine learning phases of matter. *Nat. Phys.* **13**, 431–434 (2017).
6. Ma, J. *et al.* Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods* **15**, 290–298 (2018).
7. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nat. Biotechnol.* **36**, 460–468 (2018).
8. Radovic, A. *et al.* Machine learning at the energy and intensity frontiers of particle physics. *Nature* **560**, 41–48 (2018).
9. Van Nieuwenburg, E. P. L., Liu, Y. H. & Huber, S. D. Learning phase transitions by confusion. *Nat. Phys.* **13**, 435–439 (2017).
10. Zhu, B., Liu, J. Z., Cauley, S. F., Rosen, B. R. & Rosen, M. S. Image reconstruction by domain-transform manifold learning. *Nature* **555**, 487–492 (2018).
11. Isayev, O. Text mining facilitates materials discovery. *Nature* **571**, 42–43 (2019).
12. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).

13. Prashun, G., Vladan, S. & Eric, S. T. Computationally guided discovery of thermoelectric materials. *Nat. Rev. Mater.* **2**, 1–16 (2017).
14. Hey, T., Tansley, S. & Tolle, K. *The Fourth Paradigm: Data-Intensive Scientific Discovery. The Fourth Paradigm: Data-Intensive Scientific Discovery* (Microsoft Research, 2009).
15. Torrey, L. & Shavlik, J. Transfer learning. in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques* 242–264 (IGI Global, 2010).
16. Pan, S. J. & Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**, 1345–1359 (2010).
17. Bhadeshia, H. K. D. H. Neural Networks in Materials Science. *ISIJ Int.* **39**, 966–979 (1999).
18. Demuth, H. B., Beale, M. H., De Jess, O. & Hagan, M. T. *Neural network design*. (Martin Hagan, 2014).
19. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
20. MacKay, D. J. C. A Practical Bayesian Framework for Backpropagation Networks. *Neural Comput.* **4**, 448–472 (1992).
21. MacKay, D. J. C. Bayesian interpolation. *Neural Comput.* **4**, 415–447 (1992).
22. James, G., Witten, D., Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning: with Applications in R*. (Springer New York, 2013).
23. Lucci, S. & Kopec, D. *Artificial Intelligence in the 21st Century*. (Mercury Learning & Information, 2015).
24. Cassar, D. R., de Carvalho, A. C. P. L. F. & Zanotto, E. D. Predicting glass transition temperatures using neural networks. *Acta Mater.* **159**, 249–256 (2018).
25. Wang, J., Van Der Wolk, P. J. & Van Der Zwaag, S. Effects of carbon concentration and cooling rate on continuous cooling transformations predicted by artificial neural network. *ISIJ Int.* **39**, 1038–1046 (1999).

26. Guo, L., Roelofs, H., Lembke, M. I. & Bhadeshia, H. K. D. H. Modelling of transition from upper to lower bainite in multi-component system. *Mater. Sci. Technol. (United Kingdom)* **33**, 430–437 (2017).
27. Das, A. Revisiting Stacking Fault Energy of Steels. *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.* **47**, 748–768 (2016).
28. Dobrzański, L. A. & Trzaska, J. Application of neural networks for prediction of critical values of temperatures and time of the supercooled austenite transformations. *J. Mater. Process. Technol.* **155–156**, 1950–1955 (2004).
29. Trzaska, J. & Dobrzański, L. A. Modelling of CCT diagrams for engineering and constructional steels. *J. Mater. Process. Technol.* **192–193**, 504–510 (2007).
30. Tancret, F., Bhadeshia, H. K. D. H. & Mackay, D. J. C. Comparison of artificial neural networks with Gaussian processes to model the yield strength of nickel-base superalloys. *ISIJ Int.* **39**, 1020–1026 (1999).
31. Kong, L. X. & Hodgson, P. D. Application of constitutive and artificial neural network models to predict the hot strength of steels. *ISIJ Int.* **39**, 991–998 (1999).
32. Warde, J. & Knowles, D. M. Application of neural networks to mechanical property determination of Ni-base superalloys. *ISIJ Int.* **39**, 1006–1014 (1999).
33. Narayan, V., Abad, R., Lopez, B., Bhadeshia, H. K. D. H. & Mackay, D. J. C. Estimation of hot torsion stress strain curves in iron alloys using a neural network analysis. *ISIJ Int.* **39**, 999–1005 (1999).
34. Dumortier, C. & Lehert, P. Statistical modelling of mechanical tensile properties of steels by using Neural Networks and Multivariate Data Analysis. *ISIJ Int.* **39**, 980–985 (1999).
35. Datta, S., Sil, J. & Banerjee, M. K. Petri neural network model for the effect of controlled thermomechanical process parameters on the mechanical properties of HSLA steels. *ISIJ Int.* **39**, 986–990 (1999).
36. Warde, J. & Knowles, D. M. Use of neural networks for alloy design. *ISIJ Int.* **39**, 1015–1019 (1999).

37. Conduit, B. D., Jones, N. G., Stone, H. J. & Conduit, G. J. Design of a nickel-base superalloy using a neural network. *Mater. Des.* **131**, 358–365 (2017).
38. Trzaska, J. & Dobrzański, L. A. Application of neural networks for designing the chemical composition of steel with the assumed hardness after cooling from the austenitising temperature. *J. Mater. Process. Technol.* **164–165**, 1637–1643 (2005).
39. Aggour, K. S. *et al.* Artificial intelligence/machine learning in manufacturing and inspection: A GE perspective. *MRS Bulletin* **44**, 545–558 (2019).
40. Thomson, R. C., Perez-Perez, F. J., Warters, A. D. & Thewlis, G. Neural network approach to the prediction of submerged arc weld metal chemistry. *ISIJ Int.* **39**, 1096–1105 (1999).
41. Watanabe, T., Omura, K., Konishi, M., Watanabe, S. & Furukawa, K. Mold level control in continuous caster by neural network model. *ISIJ Int.* **39**, 1053–1060 (1999).
42. Dilthey, U. & Dickersbach, J. Application of neural networks for quality evaluation of resistance spot welds. *ISIJ Int.* **39**, 1061–1066 (1999).
43. Vitek, J. M. Neural networks applied to welding: Two examples. *ISIJ Int.* **39**, 1088–1095 (1999).
44. Aoki, K. & Suga, Y. Application Artificial Neural Network to Discrimination of Defect Type in Automatic Radiographic Testing of Welds. *ISIJ Int.* **39**, 1081–1087 (1999).
45. Suga, Y., Shimamura, T., Usui, S. & Aoki, K. Measurement of molten pool shape and penetration control applying Neural Network in TIG welding of thin steel plates. *ISIJ Int.* **39**, 1075–1080 (1999).
46. Otsuka, Y., Konishi, M., Hanaoka, K. & Maki, T. Forecasting heat levels in blast furnaces using a neural network model. *ISIJ Int.* **39**, 1047–1052 (1999).
47. Dilthey, U. & Heidrich, J. Using AI-methods for parameter scheduling, quality control and weld geometry determination in GMA-welding. *ISIJ Int.* **39**, 1067–1074 (1999).

48. Fujii, H. & Bhadeshia, H. K. D. H. Preface to the Special Issue on 'Application of Neural Network Analysis in Materials Science'. *ISIJ Int.* **39**, 965 (1999).
49. Bhadeshia, H. K. D. H. D. H., Dimitriu, R. C., Forsik, S., Pak, J. H. & Ryu, J. H. Performance of neural networks in materials science. *Mater. Sci. Technol.* **25**, 504–510 (2009).
50. Bengio, Y. *Learning deep architectures for AI*. (Now Publishers Inc, 2009).
51. Malinov, S., Sha, W. & McKeown, J. J. Modelling the correlation between processing parameters and properties in titanium alloys using artificial neural network. *Comput. Mater. Sci.* **21**, 375–394 (2001).
52. Schmidhuber, J. Deep Learning in neural networks: An overview. *Neural Networks* **61**, 85–117 (2015).
53. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science (80-).* **313**, 504–507 (2006).
54. Bengio, Y., Lamblin, P., Popovici, D. & Larochelle, H. Greedy layer-wise training of deep networks. in *Advances in neural information processing systems* 153–160 (2007).
55. Cecen, A., Dai, H., Yabansu, Y. C., Kalidindi, S. R. & Song, L. Material structure-property linkages using three-dimensional convolutional neural networks. *Acta Mater.* **146**, 76–84 (2018).
56. Kondo, R., Yamakawa, S., Masuoka, Y., Tajima, S. & Asahi, R. Microstructure recognition using convolutional neural networks for prediction of ionic conductivity in ceramics. *Acta Mater.* **141**, 29–38 (2017).
57. Ward, L. *et al.* A machine learning approach for engineering bulk metallic glass alloys. *Acta Mater.* **159**, 102–111 (2018).
58. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2**, 16028 (2016).
59. Stanev, V. *et al.* Machine learning modeling of superconducting critical temperature. *npj Comput. Mater.* **4**, 29 (2018).

60. Feng, S. & Dong, H. Predicting solidification cracking susceptibility of stainless steels using machine learning. *IOP Conf. Ser. Mater. Sci. Eng.* **861**, 012073 (2020).
61. Hutchinson, M. L. *et al.* Overcoming data scarcity with transfer learning. *arXiv preprint arXiv:1711.05099* (2017).
62. Ferguson, M., Ak, R., Lee, Y.-T. T. & Law, K. H. Detection and Segmentation of Manufacturing Defects with Convolutional Neural Networks and Transfer Learning. *Smart Sustain. Manuf. Syst.* **2**, 20180033 (2018).
63. Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. *How transferable are features in deep neural networks?* (2014).
64. De Pablo, J. J., Jones, B., Kovacs, C. L., Ozolins, V. & Ramirez, A. P. The Materials Genome Initiative, the interplay of experiment, theory and computation. *Current Opinion in Solid State and Materials Science* **18**, 99–117 (2014).
65. de Pablo, J. J. *et al.* New frontiers for the materials genome initiative. *npj Computational Materials* **5**, 1–23 (2019).
66. White, A. The materials genome initiative: One year on. *MRS Bulletin* **37**, 715–716 (2012).
67. Pizzi, G., Cepellotti, A., Sabatini, R., Marzari, N. & Kozinsky, B. AiiDA: automated interactive infrastructure and database for computational science. *Comput. Mater. Sci.* **111**, 218–230 (2016).
68. Kirklin, S. *et al.* The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies. *npj Comput. Mater.* **1**, 15010 (2015).
69. Lederer, Y., Toher, C., Vecchio, K. S. & Curtarolo, S. The search for high entropy alloys: A high-throughput ab-initio approach. *Acta Mater.* **159**, 364–383 (2018).
70. Curtarolo, S. *et al.* AFLOW: An automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226 (2012).
71. Jain, A. *et al.* Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).

72. Zhang, Y. & Ling, C. A strategy to apply machine learning to small datasets in materials science. *npj Comput. Mater.* **4**, 25 (2018).
73. Feng, S., Zhou, H. & Dong, H. Using deep neural network with small dataset to predict material defects. *Mater. Des.* **162**, 300–310 (2019).
74. Murdock, R., Kauwe, S., Wang, A. & Sparks, T. Is domain knowledge necessary for machine learning materials properties? *ChemRxiv* (2020). doi:10.26434/chemrxiv.11879193.v1
75. Saal, J. E., Oliynyk, A. O. & Meredig, B. Machine learning in materials discovery: Confirmed predictions and their underlying approaches. *Annu. Rev. Mater. Res.* **50**, 49–69 (2020).
76. Meredig, B. *et al.* Can machine learning identify the next high-temperature superconductor? Examining extrapolation performance for materials discovery. *Mol. Syst. Des. Eng.* **3**, 819–825 (2018).
77. Kauwe, S. K., Graser, J., Murdock, R. & Sparks, T. D. Can machine learning find extraordinary materials? *Comput. Mater. Sci.* **174**, 109498 (2020).
78. Lookman, T., Balachandran, P. V., Xue, D. & Yuan, R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Computational Materials* **5**, 1–17 (2019).
79. Balachandran, P. V., Xue, D., Theiler, J., Hogden, J. & Lookman, T. Adaptive Strategies for Materials Design using Uncertainties. *Sci. Rep.* **6**, 1–9 (2016).
80. Ling, J., Hutchinson, M., Antono, E., Paradiso, S. & Meredig, B. High-Dimensional Materials and Process Optimization Using Data-Driven Experimental Design with Well-Calibrated Uncertainty Estimates. *Integr. Mater. Manuf. Innov.* **6**, 207–217 (2017).
81. Janet, J. P., Duan, C., Yang, T., Nandy, A. & Kulik, H. J. A quantitative uncertainty metric controls error in neural network-driven chemical discovery. *Chem. Sci.* **10**, 7913–7922 (2019).
82. Ryan, K. M. *et al.* A Gaussian Process Modeling Approach for Fast Robust Design With Uncertain Inputs. in (ASME International, 2018). doi:10.1115/gt2018-

83. Le, T. D. *et al.* Critical Temperature Prediction for a Superconductor: A Variational Bayesian Neural Network Approach. *IEEE Trans. Appl. Supercond.* **30**, (2020).
84. Childs, C. M. & Washburn, N. R. Embedding domain knowledge for machine learning of complex material systems. *MRS Communications* **9**, 806–820 (2019).
85. Wan, A. *et al.* NBDT: Neural-Backed Decision Trees. (2020).
86. Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **5**, 83 (2019).
87. Campbell, J. *Complete Casting Handbook: Metal Casting Processes, Metallurgy, Techniques and Design*. (Elsevier Science, 2015).
88. Farrar, J. C. M. Hot cracking tests—The route to International Standardization. in *Hot cracking phenomena in welds* 291–304 (Springer, 2005).
89. Cross, C. E. On the origin of weld solidification cracking. in *Hot cracking phenomena in welds* 3–18 (Springer, 2005).
90. Bernasovský, P. Contribution to HAZ Liquation Cracking of Austenitic Stainless Steels. in *Hot cracking phenomena in welds* 84–103 (Springer, 2005).
91. Ramirez, A. J. & Lippold, J. C. New insight into the mechanism of ductility-dip cracking in Ni-base weld metals. in *Hot Cracking Phenomena in Welds* 19–41 (Springer, 2005).
92. Eskin, D. G. G., Suyitno & Katgerman, L. *Mechanical properties in the semi-solid state and hot tearing of aluminium alloys. Progress in Materials Science* **49**, 629–711 (Pergamon, 2004).
93. Soysal, T. & Kou, S. A simple test for assessing solidification cracking susceptibility and checking validity of susceptibility prediction. *Acta Mater.* **143**, 181–197 (2018).
94. Kou, S. A criterion for cracking during solidification. *Acta Mater.* **88**, 366–374

(2015).

95. Rappaz, M., Drezet, J.-M. M. & Gremaud, M. A new hot-tearing criterion. *Metall. Mater. Trans. A* **30**, 449–455 (1999).
96. Li, S. & Apelian, D. Hot tearing of aluminum alloys a critical literature review. *Int. J. Met.* **5**, 23–40 (2011).
97. Bordreuil, C. & Niel, A. Modelling of hot cracking in welding with a cellular automaton combined with an intergranular fluid flow model. *Comput. Mater. Sci.* **82**, 442–450 (2014).
98. Niel, A., Bordreuil, C., Deschaux-Beaume, F. & Fras, G. Modelling hot cracking in 6061 aluminium alloy weld metal with microstructure based criterion. *Sci. Technol. Weld. Join.* **18**, 154–160 (2013).
99. Thomas, B. G. Modeling of hot tearing and other defects in casting processes. *ASM Handb.* **22**, 362–374 (2009).
100. Lippold, J. C. Recent developments in weldability testing. in *Hot cracking phenomena in welds* 271–290 (Springer, 2005).
101. Dantzig, J. A. & Rappaz, M. *Solidification: 2nd Edition - Revised & Expanded*. (EPFL Press, 2016).
102. Rappaz, M., Jacot, A. & Boettinger, W. J. Last-stage solidification of alloys: theoretical model of dendrite-arm and grain coalescence. *Metall. Mater. Trans. A* **34**, 467–479 (2003).
103. Drezet, J.-M., Mireux, B., Szaraz, Z. & Pirling, T. In situ neutron diffraction during casting: determination of rigidity point in grain refined Al-Cu alloys. *Materials (Basel)*. **7**, 1165–1172 (2014).
104. Drezet, J.-M., Mireux, B., Szaraz, Z. & Pirling, T. Determination of coherency and rigidity temperatures in Al-Cu alloys using in situ neutron diffraction during casting. *JOM* **66**, 1425–1430 (2014).
105. Heydari, D., Fard, A. S., Bakhshi, A. & Drezet, J. M. Hot tearing in polycrystalline Ni-based IN738LC superalloy: Influence of Zr content. *J. Mater. Process. Technol.* **214**, 681–687 (2014).

106. Mandziej, S. T. Testing for susceptibility to hot cracking on Gleeble™ physical simulator. in *Hot cracking phenomena in welds* 347–376 (Springer, 2005).
107. Kannengiesser, T. & Boellinghaus, T. Hot cracking tests—an overview of present technologies and applications. *Weld. World* **58**, 397–421 (2014).
108. Lundin, C. D., Qiao, C. Y. P., Lee, C. H. & Batten, G. Weldability and hot ductility behavior of nuclear grade austenitic stainless steels. *Weld. Res. Counc. Bull.* **1** (2006).
109. Lundin, C. D., Qiao, C. Y. P., Gill, T. P. S. & Goodwin, G. M. Hot ductility and hot cracking behavior of modified 316 stainless steels designed for high-temperature service. *Weld. JOURNAL-NEW YORK-* **72**, 189-s (1993).
110. Matsuda, F., Nakagawa, H., Katayama, S. & Arata, Y. Solidification Crack Susceptibility in Weld Metals of Fully Austenitic Stainless Steels (Report VI): Effect of La or REM Addition on Solidification Crack Resistance (MATERIALS METALLURGY AND WELDABILITY). *Trans. JWRI* **11**, 79–94 (1982).
111. Matsuda, F., Katayama, S. & Arata, Y. Solidification Crack Susceptibility in Weld Metals of Fully Austenitic Stainless Steels (Report IX): Effect of Titanium on Solidification Crack Resistance (Materials, Metallurgy & Weldability). *Trans. JWRI* **12**, 247–252 (1983).
112. Ogawa, T. & Tsunetomi, E. Hot Cracking Susceptibility of Austenitic Stainless Steels. *Weld. J.* **61**, 82 (1982).
113. Matsuda, F., Nakagawa, H., Katayama, S. & Arata, Y. Solidification Crack Susceptibility in Weld Metals of Fully Austenitic Stainless Steels (Report VIII): Effect of Nitrogen on Cracking in SUS 304 Weld Metal (Materials, Metallurgy & Weldability). *Trans. JWRI* **12**, 89–95 (1983).
114. Arata, Y., Matsuda, F., Nakagawa, H. & Katayama, S. Solidification crack susceptibility in weld metals of fully austenitic stainless steels (report VII) : effect of Mn and N on solidification crack resistance. *Trans. JWRI* **11**, 79–85 (1982).
115. Matsuda, F., Katayama, S. & Arata, Y. Solidification crack susceptibility in weld metals of fully austenitic stainless steels (report V): solidification crack

- susceptibility and amount of phosphide and sulphide in sus 310s weld metals. *Trans. JWRI* **10**, 201–212 (1981).
116. Arata, Y., Matsuda, F., Nakagawa, H., Katayama, S. & Ogata, S. Solidification crack susceptibility in weld metals of fully austenitic stainless steels (Report III): effect of strain rate on cracking threshold in weld metal during solidification. *Trans. JWRI* **6**, 197–206 (1977).
117. Arata, Y., Matsuda, F. & Katayama, S. Solidification crack susceptibility in weld metals of fully austenitic stainless steels (Report II): Effect of ferrite, P, S, C, Si and Mn on ductility properties of solidification brittleness. *Trans. JWRI* **6**, 105–116 (1977).
118. Arata, Y., Matsuda, F. & Katayama, S. Solidification crack susceptibility in weld metals of fully austenitic stainless steels (Report I): fundamental investigation on solidification behavior of fully austenitic and duplex microstructures and effect of ferrite on microsegregation. *Trans. JWRI* **5**, 135–151 (1976).
119. Arata, Y., Matsuda, F., Nakagawa, H. & Katayama, S. Solidification Crack Susceptibility in Weld Metals of Fully Austenitic Stainless Steels (Report IV): Effect of Decreasing P and S on Solidification Crack Susceptibility of SUS 310S Austenitic Stainless Steel Weld Metals. *Trans. JWRI* **7**, 169–172 (1978).
120. Arata, Y., Matsuda, F. & Saruwatari, S. Varestraint test for solidification crack susceptibility in weld metal of austenitic stainless steels. *Trans. JWRI* **3**, 79–88 (1974).
121. Cieslak, M. J. & Savage, W. F. Weldability and Solidification Phenomena of Cast Stainless Steels. *Weld. J.* **59**, S136–S146 (1980).
122. Kah, D. H., DH, K. A. H. & DW, D. Weldability of ferritic stainless steels. *Weld. J.* **60**, S135–S142 (1981).
123. Lippold, J. C. An investigation of weld cracking in Alloy 800. *Weld. J.* **63**, 91 (1983).
124. Ritter, A. M. & Savage, W. F. Solidification and solidification cracking in nitrogen-strengthened austenitic stainless steels. *Metall. Trans. A* **17**, 727–737 (1986).

125. Ogawa, T. & Koseki, T. Weldability of newly developed austenitic alloys for cryogenic service: Part II-High-nitrogen stainless steel weld metal. *Weld. J.* **67**, 8s-17s (1988).
126. Varol, I., Baeslack III, W. A. & Lippold, J. C. Characterization of weld solidification cracking in a duplex stainless steel. *Metallography* **23**, 1–19 (1989).
127. Lundin, C. D., Qiao, C. Y. P., Kikuchi, Y., Shi, C. & Gill, T. P. S. *Investigation of joining techniques for advanced austenitic alloys*. (Oak Ridge National Lab., TN (United States); Tennessee Univ., Knoxville, TN ..., 1991).
128. Lundin, C. D., Lee, C. H. & Qiao, C. Y. P. Evaluation of Backfilled Solidification Cracks in Austenitic Stainless Welds in Relationship to Evaluating Hot Cracking. *Weld. JOURNAL-NEW YORK-* **72**, 321-s (1993).
129. MB, Z. *et al.* Effect of welding parameters on hot cracking susceptibility of alloy 800. *Q. J. JAPAN Weld. Soc.* **12**, 335–341 (1994).
130. Madarame, H., Sukegawa, T. & Inoue, H. Cracking susceptibility of stainless steel subjected to plasma disruption. *Fusion Eng. Des.* **27**, 499–506 (1995).
131. Robino, C. V, Michael, J. R. & Maguire, M. C. The solidification and welding metallurgy of galling-resistant stainless steels. *Weld. JOURNAL-NEW YORK-* **77**, 446-s (1998).
132. Shankar, V., Gill, T. P. S., Mannan, S. L. & Sundaresan, S. Evaluation of hot cracking in nitrogen-bearing and fully austenitic stainless steel weldments. *Weld. Journal-Including Weld. Res. Suppl.* **77**, 193S (1998).
133. Brooks, J. A. & Garrison, W. M. Weld Microstructure Development and Properties of Precipitation-Strengthened Martensitic Stainless Steels. *Weld. JOURNAL-NEW YORK-* **78**, 280-s (1999).
134. Li, L. & Messler Jr, R. W. The effects of phosphorus and sulfur on susceptibility to weld hot cracking in austenitic stainless steels. *Weld. JOURNAL-NEW YORK-* **78**, 387-s (1999).
135. Shankar, V., Gill, T. P. S., Mannan, S. L. & Sundaresan, S. Criteria for hot cracking evaluation in austenitic stainless steel welds using longitudinal varestraint and

- transvarestraint tests. *Sci. Technol. Weld. Join.* **5**, 91–97 (2000).
136. Brooks, J. A., Robino, C. V, Headley, T. J. & Michael, J. R. Weld solidification and cracking behavior of free-machining stainless steel. *Weld. JOURNAL-NEW YORK*-**82**, 51-S (2003).
 137. Li, L. & Messler, R. W. Effects of Phosphorus and Sulfur on Susceptibility to Weld Hot Cracking in Austenitic Stainless Steels. *Weld. Res. Counc. Bull.* (2003).
 138. Shankar, V., Gill, T. P. S., Mannan, S. L. & Sundaresan, S. Effect of nitrogen addition on microstructure and fusion zone cracking in type 316L stainless steel weld metals. *Mater. Sci. Eng. A* **343**, 170–181 (2003).
 139. Srinivasan, G., Bhaduri, A. K., Albert, S. K. & Klenk, A. Effect Of Phosphorous and Silicon On Hot Cracking Susceptibility Of 14Cr-15Ni-2.3 Mo Ti-Modified Fully-Austenitic Stainless Steel. *Weld. World* **56**, 2–13 (2012).
 140. Ichikawa, K., Bhadeshia, H. & MacKay, D. J. C. Model for solidification cracking in low alloy steel weld metals. *Sci. Technol. Weld. Join.* **1**, 43–50 (1996).
 141. Inoue, A. & Takeuchi, A. Recent development and application products of bulk glassy alloys. *Acta Mater.* **59**, 2243–2267 (2011).
 142. Jia, H. *et al.* Fatigue and fracture behavior of bulk metallic glasses and their composites. *Prog. Mater. Sci.* **98**, 168–248 (2018).
 143. Sun, B. A. & Wang, W. H. The fracture of bulk metallic glasses. *Prog. Mater. Sci.* **74**, 211–307 (2015).
 144. Herzer, G. Modern soft magnets: Amorphous and nanocrystalline materials. *Acta Mater.* **61**, 718–734 (2013).
 145. Zhang, L.-C., Jia, Z., Lyu, F., Liang, S.-X. & Lu, J. A review of catalytic performance of metallic glasses in wastewater treatment: Recent progress and prospects. *Prog. Mater. Sci.* **105**, 100576 (2019).
 146. Wang, W. H. Bulk metallic glasses with functional physical properties. *Adv. Mater.* **21**, 4524–4544 (2009).
 147. Qiao, J. C. *et al.* Structural heterogeneities and mechanical behavior of

- amorphous alloys. *Prog. Mater. Sci.* **104**, 250–329 (2019).
148. Klement, W., Willens, R. H. & Duwez, P. Non-crystalline Structure in Solidified Gold–Silicon Alloys. *Nature* **187**, 869–870 (1960).
149. Lu, Z. P. & Liu, C. T. A new glass-forming ability criterion for bulk metallic glasses. *Acta Mater.* **50**, 3501–3512 (2002).
150. Takeuchi, A. & Inoue, A. Classification of bulk metallic glasses by atomic size difference, heat of mixing and period of constituent elements and its application to characterization of the main alloying element. *Mater. Trans.* **46**, 2817–2829 (2005).
151. Louzguine-Luzgin, D. V. *et al.* Role of different factors in the glass-forming ability of binary alloys. *J. Mater. Sci.* **50**, 1783–1793 (2015).
152. Miracle, D. B., Sanders, W. S. & Senkov, O. N. The influence of efficient atomic packing on the constitution of metallic glasses. *Philos. Mag.* **83**, 2409–2428 (2003).
153. Laws, K. J., Miracle, D. B. & Ferry, M. A predictive structural model for bulk metallic glasses. *Nat. Commun.* **6**, 8123 (2015).
154. Greer, A. L. Confusion by design. *Nature* **366**, 303–304 (1993).
155. Perim, E. *et al.* Spectral descriptors for bulk metallic glasses based on the thermodynamics of competing crystalline phases. *Nat. Commun.* **7**, 12315 (2016).
156. Dasgupta, A. *et al.* Probabilistic Assessment of Glass Forming Ability Rules for Metallic Glasses Aided by Automated Analysis of Phase Diagrams. *Sci. Rep.* **9**, 357 (2019).
157. Louzguine, D. V. & Inoue, A. Electronegativity of the constituent rare-earth metals as a factor stabilizing the supercooled liquid region in Al-based metallic glasses. *Appl. Phys. Lett.* **79**, 3410–3412 (2001).
158. Sun, Y. T., Bai, H. Y., Li, M. Z. & Wang, W. H. Machine learning approach for prediction and understanding of glass-forming ability. *J. Phys. Chem. Lett.* **8**, 3434–3439 (2017).

159. Kawazoe, Y., Yu, J.-Z., Tsai, A.-P. & Masumoto, T. *Nonequilibrium Phase Diagrams of Ternary Amorphous Alloys. Nonequilibrium Phase Diagrams of Ternary Amorphous Alloys* (Springer-Verlag, 1997). doi:10.1007/b58222
160. Murty, B. S., Yeh, J. W., Ranganathan, S. & Bhattacharjee, P. P. *High-Entropy Alloys*. (Elsevier Science, 2019).
161. Gao, M. C., Liaw, P. K., Yeh, J. W. & Zhang, Y. *High-entropy alloys: Fundamentals and applications. High-Entropy Alloys: Fundamentals and Applications* (Springer International Publishing, 2016). doi:10.1007/978-3-319-27013-5
162. George, E. P., Raabe, D. & Ritchie, R. O. High-entropy alloys. *Nat. Rev. Mater.* **4**, 515–534 (2019).
163. Senkov, O. N., Senkova, S. V., Woodward, C. & Miracle, D. B. Low-density, refractory multi-principal element alloys of the Cr-Nb-Ti-V-Zr system: Microstructure and phase analysis. *Acta Mater.* **61**, 1545–1557 (2013).
164. Miracle, D. B. & Senkov, O. N. A critical review of high entropy alloys and related concepts. *Acta Mater.* **122**, 448–511 (2017).
165. Senkov, O. N., Miller, J. D., Miracle, D. B. & Woodward, C. Accelerated exploration of multi-principal element alloys for structural applications. *Calphad* **50**, 32–48 (2015).
166. Senkov, O. N., Miller, J. D., Miracle, D. B. & Woodward, C. Accelerated exploration of multi-principal element alloys with solid solution phases. *Nat. Commun.* **6**, 6529 (2015).
167. Islam, N., Huang, W. & Zhuang, H. L. Machine learning for phase selection in multi-principal element alloys. *Comput. Mater. Sci.* **150**, 230–235 (2018).
168. Zhou, Z. *et al.* Machine learning guided appraisal and exploration of phase design for high entropy alloys. *npj Comput. Mater.* **5**, 128 (2019).
169. Zhang, Y. *et al.* Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. *Acta Mater.* (2019). doi:10.1016/j.actamat.2019.11.067
170. Wen, C. *et al.* Machine learning assisted design of high entropy alloys with

- desired property. *Acta Mater.* **170**, 109–117 (2019).
171. Li, Y. & Guo, W. Machine-learning model for predicting phase formations of high-entropy alloys. *Phys. Rev. Mater.* **3**, 095005 (2019).
 172. Graser, J., Kauwe, S. K. & Sparks, T. D. Machine Learning and Energy Minimization Approaches for Crystal Structure Predictions: A Review and New Horizons. *Chemistry of Materials* **30**, 3601–3612 (2018).
 173. Mishkin, D. & Matas, J. All you need is a good init. *4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.* (2015).
 174. Bhaduri, A. K., Albert, S. K. & Raj, B. Robust welding technologies for ferrous alloys. in *Weld Cracking in Ferrous Alloys* 34–95 (Elsevier Ltd., 2008).
doi:10.1533/9781845695453.1.34
 175. Zheng, X., Zheng, P. & Zhang, R. Machine learning material properties from the periodic table using convolutional neural networks. *Chem. Sci.* **9**, 8426–8432 (2018).
 176. Zeng, S. *et al.* Atom table convolutional neural networks for an accurate prediction of compounds properties. *npj Comput. Mater.* **5**, 84 (2019).
 177. Miracle, D. B., Louzguine-Luzgin, D. V, Louzguina-Luzgina, L. V & Inoue, A. An assessment of binary metallic glasses: correlations between structure, glass forming ability and stability. *Int. Mater. Rev.* **55**, 218–256 (2010).
 178. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD). *JOM* **65**, 1501–1509 (2013).
 179. Graulis, S. *et al.* Crystallography Open Database - An open-access collection of crystal structures. *J. Appl. Crystallogr.* **42**, 726–729 (2009).
 180. Miracle, D. B. Efficient local packing in metallic glasses. *J. Non. Cryst. Solids* **342**, 89–96 (2004).
 181. Guo, S. & Liu, C. T. Phase stability in high entropy alloys: formation of solid-solution phase or amorphous phase. *Prog. Nat. Sci. Mater. Int.* **21**, 433–446 (2011).

182. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. in *3rd International Conference on Learning Representations, {ICLR} 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (eds. Bengio, Y. & LeCun, Y.) (2015).
183. Li, M.-X. *et al.* High-temperature bulk metallic glasses developed by combinatorial methods. *Nature* **569**, 99–103 (2019).
184. Shamlaye, K. F., Laws, K. J. & Löffler, J. F. Exceptionally broad bulk metallic glass formation in the Mg–Cu–Yb system. *Acta Mater.* **128**, 188–196 (2017).
185. Kuball, A., Gross, O., Bochtler, B. & Busch, R. Sulfur-bearing metallic glasses: A new family of bulk glass-forming alloys. *Scr. Mater.* **146**, 73–76 (2018).
186. Kuball, A. *et al.* On the bulk glass formation in the ternary Pd-Ni-S system. *Acta Mater.* **158**, 13–22 (2018).
187. Lin, C.-Y., Tien, H.-Y. & Chin, T.-S. Soft magnetic ternary iron-boron-based bulk metallic glasses. *Appl. Phys. Lett.* **86**, 162501 (2005).
188. Gao, M. C. *et al.* Thermodynamics of concentrated solid solution alloys. *Curr. Opin. Solid State Mater. Sci.* **21**, 238–251 (2017).
189. Villars, P. & Cenzual, K. *Handbook of Inorganic Substances 2016*. (Walter De Gruyter Incorporated, 2016).
190. Blokhin, E. & Villars, P. The PAULING FILE Project and Materials Platform for Data Science: From Big Data Toward Materials Genome. in *Handbook of Materials Modeling 1837–1861* (Springer International Publishing, 2020).
doi:10.1007/978-3-319-44677-6_62