



Using deep neural network with small dataset to predict material defects

Shuo Feng^a, Huiyu Zhou^b, Hongbiao Dong^{a,*}



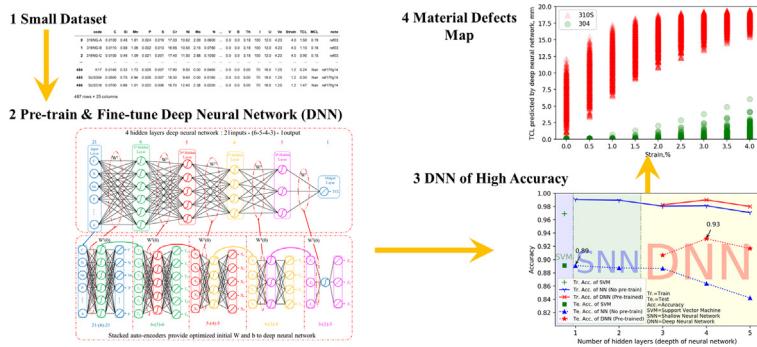
^a Department of Engineering, University of Leicester, Leicester LE1 7RH, UK

^b Department of Informatics, University of Leicester, Leicester LE1 7RH, UK

HIGHLIGHTS

- The deep neural network model for predicting solidification cracking susceptibility of stainless steels are developed.
- Stacked auto-encoder is used to pre-train deep neural network with a small dataset for optimization of initial weights.
- Deep neural network model shows better generalization performance than shallow neural network and support vector machine.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 17 October 2018

Received in revised form 16 November 2018

Accepted 28 November 2018

Available online 29 November 2018

Keywords:

Deep neural network
Small dataset
Pre-training
Defects

ABSTRACT

Deep neural network (DNN) exhibits state-of-the-art performance in many fields including microstructure recognition where big dataset is used in training. However, DNN trained by conventional methods with small datasets commonly shows worse performance than traditional machine learning methods, e.g. shallow neural network and support vector machine. This inherent limitation prevented the wide adoption of DNN in material study because collecting and assembling big dataset in material science is a challenge. In this study, we attempted to predict solidification defects by DNN regression with a small dataset that contains 487 data points. It is found that a pre-trained and fine-tuned DNN shows better generalization performance over shallow neural network, support vector machine, and DNN trained by conventional methods. The trained DNN transforms scattered experimental data points into a map of high accuracy in high-dimensional chemistry and processing parameters space. Though DNN with big datasets is the optimal solution, DNN with small datasets and pre-training can be a reasonable choice when big datasets are unavailable in material study.

© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Deep neural network

Neural network was firstly proposed in 1950s as a mathematical simplification and approximation of human neural cells in the form of single layer perception. The second-generation neural network, i.e. multi-layers perception or shallow neural network (SNN) commonly has 1 hidden layer (or ≤ 2 hidden layers) and is trained by the standard back propagation (BP) algorithm. It has extensive applications in

Abbreviations: BTR, brittle temperature range; DNN, deep neural network; LVT, longitudinal restraint test; PCA, principal component analysis; RBM, restricted Boltzmann machine; SAE, stacked auto-encoder; SCS, solidification cracking susceptibility; SNN, shallow neural network; SVM, support vector machine; TCL, total crack length.

* Corresponding author.

E-mail address: hd38@le.ac.uk (H. Dong).

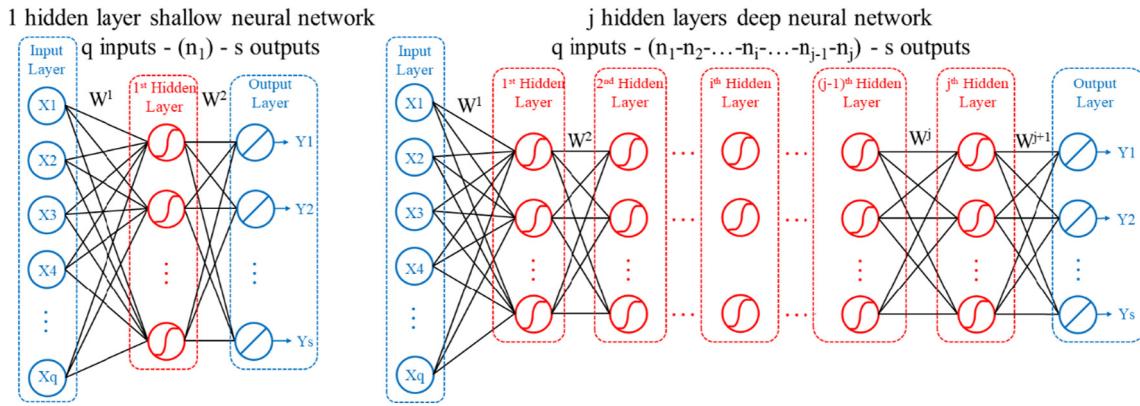


Fig. 1. Schematic diagrams of 1 hidden layer shallow neural network and j hidden layers fully connected deep neural network.

material science, e.g. predicting phase transformation behaviors [1–3] and material properties [4–15], optimizing processing parameters [16,17] etc. Research papers [1–18] and Reviews [19,20] about SNN applications in material science and engineering have been widely reported.

Since neural network was proposed, researchers in the artificial intelligence (AI) field have been exploring the possibility to train deep neural network (DNN) with many hidden layers like the human neural system [21,22]. But the success of the previous generation neural network is limited to SNN of 1 or 2 hidden layers, because training DNN is not easy and the resultant accuracy is usually worse than that of SNN [17]. The difficulties of training DNN lie in vanishing gradients with the increment of the hidden layer number, i.e. depth of network, and traps of poor local minimums [21–25]. Fig. 1 shows schematically the structure difference between SNN and DNN. The X represents input vector, Y represents output vector. The number n_i in the brackets

represents the neuron number of the i^{th} hidden layer. Circles represent neurons and lines represent the connections between them, W^j represents the weights matrix of the j^{th} layer. Deep neural network is similar to shallow neural network in structure but has more hidden layers and more obvious hierarchy structure. DNN can be perceived as an upgrade of SNN.

The proposal of restricted Boltzmann machine (RBM) and deep belief network (DBN) by Hinton in 2006 made it possible for the first time to train DNN which contains 3 or more hidden layers through a so-called greedy layer-wised pre-training [23,25]. After that, Bangio proposed stacked auto-encoder (SAE) as an alternative to RBM in pretraining [26]. DNN shows great advantage over SNN for its evident improvement of prediction accuracy on the unseen dataset or testing dataset. Since then, deep learning increasingly became prosperous and some particular types of deep neural network such as convolution neural network (CNN) and recurrence neural network (RNN) have

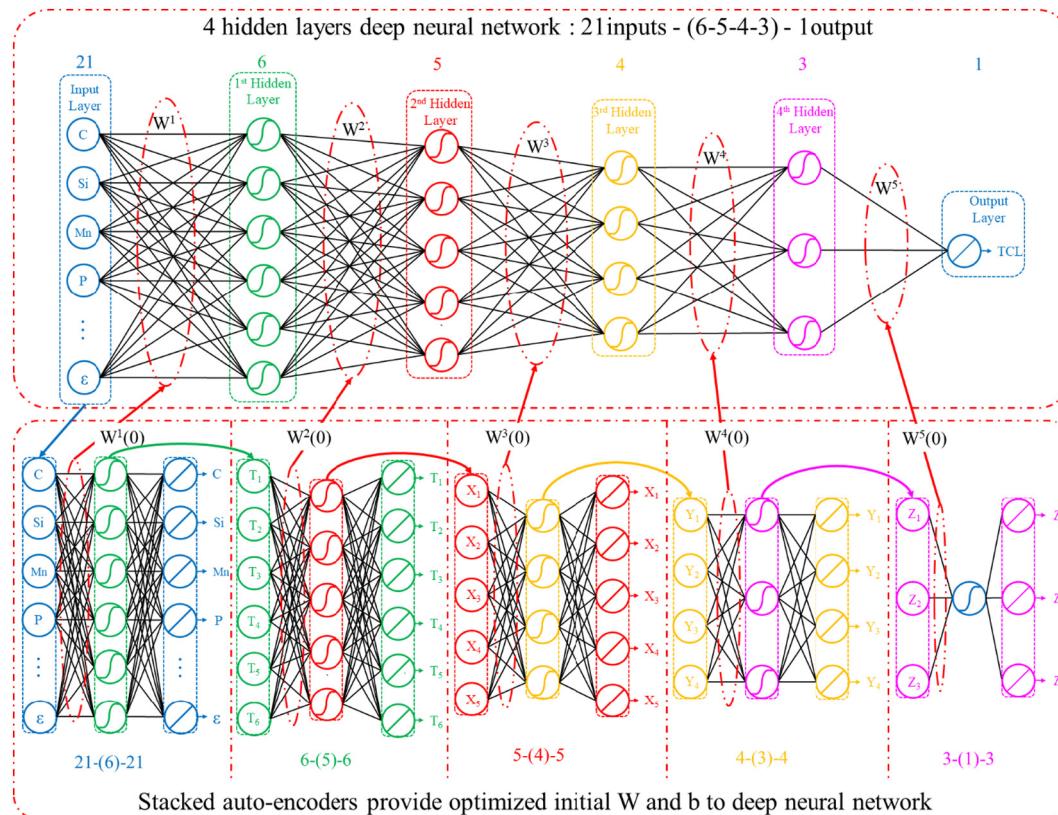


Fig. 2. Schematic diagram of the process of initializing DNN layer by layer and one autoencoder by one autoencoder, i.e. the pre-training process.

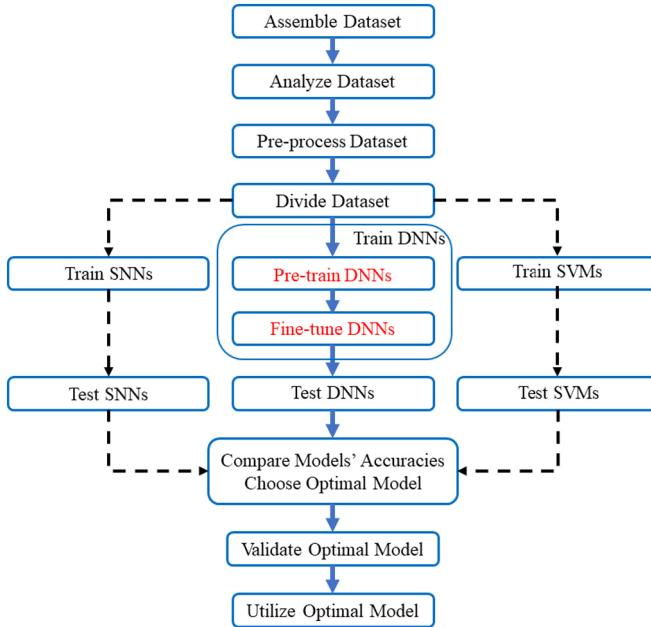


Fig. 3. The workflow of the case study on solidification cracking susceptibility prediction.

achieved astonished success in image and voice recognition and natural language processing [24]. Now deep learning has become the main stream in machine learning (ML).

1.2. Challenges of studying material defects and assembling big dataset

The formation of material defects, e.g. segregation, shrinkage pores, solidification crack, is a complex multiple variables nonlinear problem. In this paper, **solidification cracking susceptibility (SCS)** prediction was chosen as a case study of applying DNN in material study with small dataset. Solidification crack is one of the most serious defects which occurs widely in welding [27,28], casting [29–31] and additive manufacturing (AM) [32,33], which occurs at the last stage of solidification when liquid films exist between dendrites boundaries where local strains cannot be accommodated by liquid feeding and solid deformation. The interactions of material, mechanical restraint, thermal factors

Table 1

The accuracies of SVM, SNN and DNN based on dataset without PCA pre-process and datasets after PCA pre-process.

	Raw dataset 21 inputs train/test	After PCA 15 inputs train/test	After PCA 10 inputs train/test
SVM	0.969/0.891	0.925/0.862	0.914/0.859
SNN	0.990/0.891	0.951/0.883	0.959/0.874
DNN	0.989/0.931	0.977/0.901	0.964/0.900

bring about the final crack nucleation and propagation [34,35]. Metallic material processability, e.g. weldability and castability, is closely linked with SCS. Due to its complexity and importance, solidification crack has been investigated extensively with experimental methods [36,37], qualitative and quantitative theories [38,39], and computer numeric models [40–42] but still not been solved. To further study these complex problems, data-based and data-driven machine learning methods, e.g. neural network, increasingly become an alternative to physics based analytical and numerical methods [19,20]. Based on reliable experimental data, neural network is able to depict alloy SCS as a function of alloy composition and processing parameters in high-dimensional space, such as Ichikawa's work [18]. Because the experimental study of solidification cracking is time-consuming and expensive, only a few important alloys and parts can be studied, and the data collected are scattered and sparse, indicating that assembling big datasets on solidification cracking and training them with machine learning methods is a big challenge.

1.3. Potential for applying DNN with small dataset

Although convolution neural network for microstructure image recognition [43–45] and variational auto-encoder for text mining [46] have been introduced into material science, the applications of DNNs in material science are still limited. It is due to the difficulties of training DNN and assembling big dataset (e.g. $>10^4$ samples' data) in material science. The developments of machine learning in recent years made DNN trainable and provided some useful tools, like pre-training with RBM and SAE [21,24], to cope with small dataset.

On the other hand, the nature of problems in material science also indicates a feasibility of applying DNNs with small dataset: DNNs used in image recognition usually have $>10^4$ input variables (e.g. a small image of 100×100 pixels needs 10^4 input variables) and a huge number

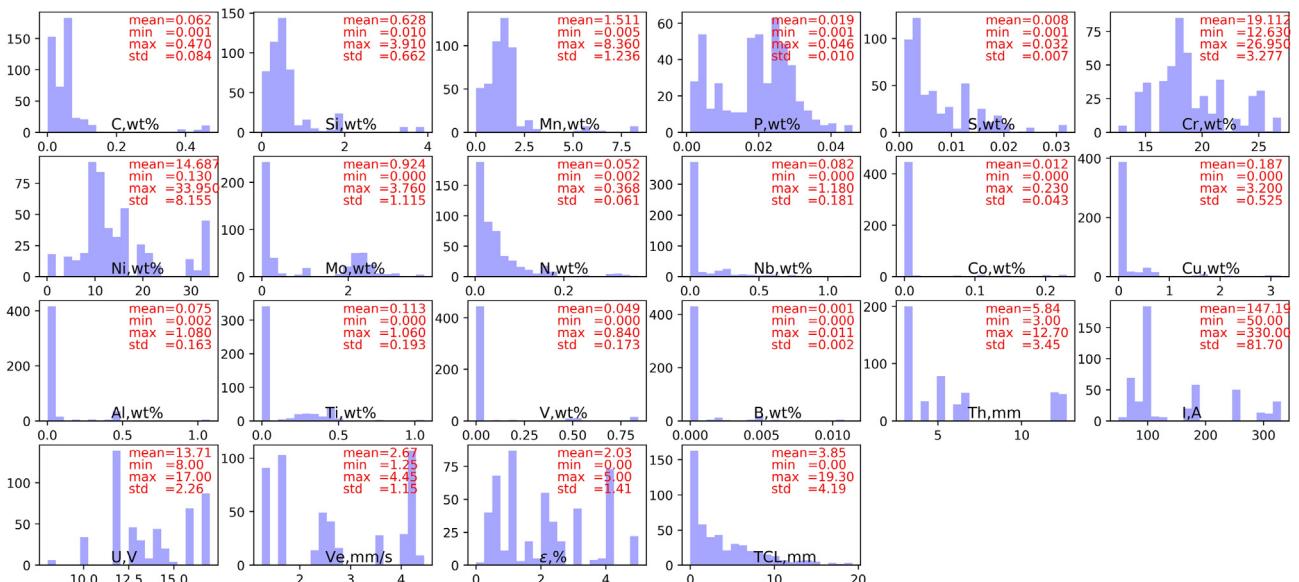


Fig. 4. Histograms of the 22 variables in the final dataset (samples count: 487), statistical information like mean, minimum, maximum, std. are also shown on the histograms.

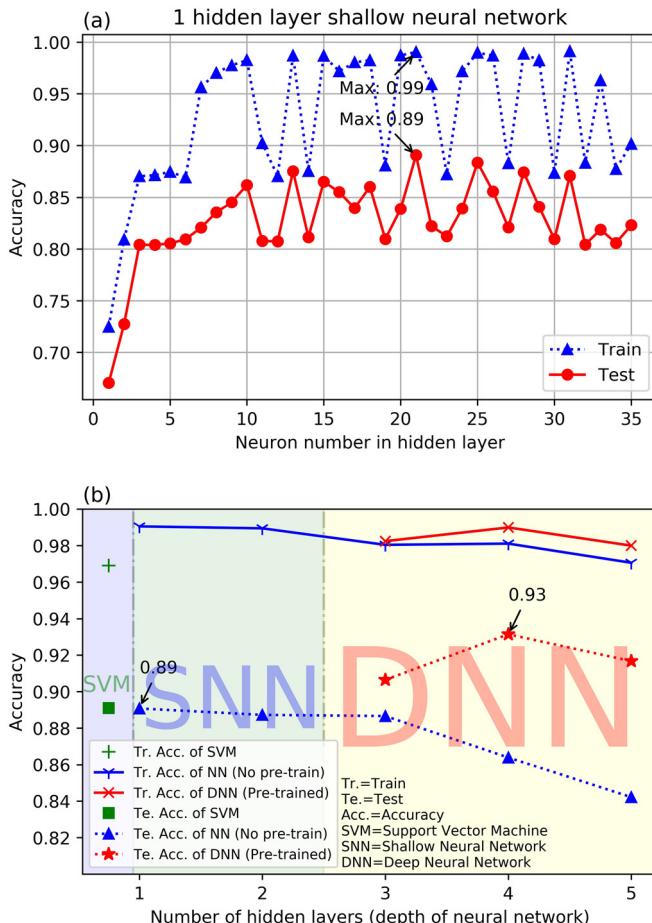


Fig. 5. (a) Shallow neural network's training and testing accuracies as a function of neuron number. (b) Neural networks' training and testing accuracies as a function of their depth. The support vector machine's accuracies are also shown on figure.

of parameters need to be determined (e.g. millions of parameters), thus big datasets are necessary; but DNNs for material problems commonly have no >100 input variables (including composition, processing, property variables) and fewer parameters need to be determined, thus small/narrow DNNs (a few hidden layers and small number of neurons in each layers) may be enough for most material problems.

The potentials for applying DNN with small datasets in material study are clear: extensive regression/classification problems formerly treated by traditional machine learning methods (like SNN, random forest [47,48], support vector machine (SVM) [49], etc.) with small dataset

can be solved by DNN with higher accuracy and better generalization performance.

In this paper we use SCS prediction as an example to show that SAE pre-training is an effective method in coping with small datasets in DNN regression, and fully connected DNN shows higher accuracy and better generalization performance than SNN and SVM.

2. Method and workflow

2.1. Training algorithm and Bayesian regularization

The performance function mse and regularized performance function mse_{reg} of a neural network are defined as follows:

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

$$\text{mse}_{\text{reg}} = \gamma \text{mse} + (1-\gamma) \text{msw} \quad (2)$$

$$\text{msw} = \frac{1}{n} \sum_{j=1}^n W_j^2 \quad (3)$$

where mse is the mean square error, msw is the mean square weight, t_i is the target value and a_i is the prediction value of the network based on the input data, γ is the performance ratio (a hyper-parameter either selected manually or determined by optimization algorithm).

To improve generalization and reduce the possibility of overfitting, regularized performance function mse_{reg} have been used in training SNN after random initialization (SNN are initialized with random numbers) and in fine-tuning DNN after pre-training (i.e. DNN are initialized with optimized values). Training SNN and fine-tuning DNN involve the minimization of mse_{reg} through regulating weights and biases along the negative of the mse_{reg} gradient (by gradient descent algorithm) and the determination of optimal performance ratio γ (by Bayesian regularization proposed by MacKay). The training and regularization details are described in [50–52].

2.2. DNN pre-training

Pre-training initializes DNN with optimized weights and biases values that are close to the global optimal solution and it helps the follow-on fine-tuning step to bypass the traps of the local optimal solution. Fig. 2 schematically shows the pre-training process, i.e. initializing DNN by SAE. The upper half shows the DNN's structure and the lower half shows the 5 auto-encoders' structures. C, Si, Mn, ..., ε are 21 input variables and TCL is the output variable of the DNN. $W^j(0)$ is the initial weights matrix of DNN's j^{th} layer. Long arrows represent the directions

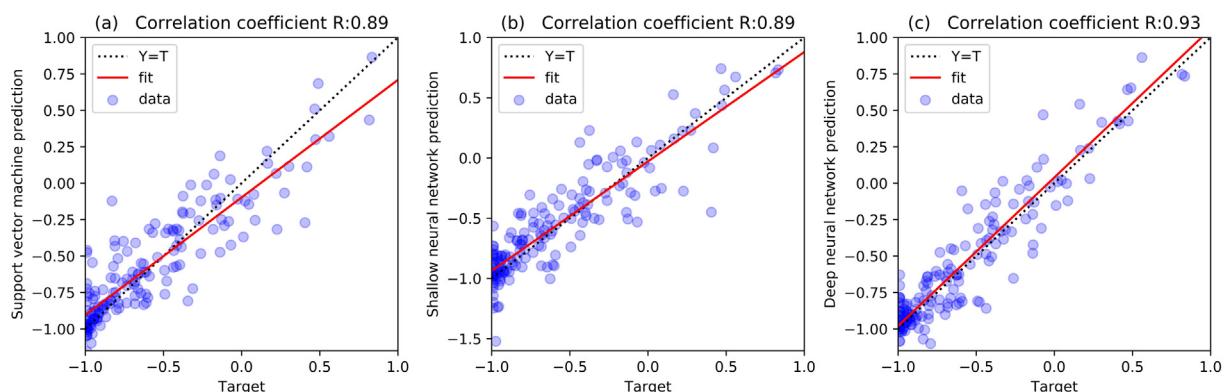


Fig. 6. The regression analysis on the unseen testing dataset between experimental data (horizontal axis, T values) and support vector machine's and neural networks' predictions (vertical axis, Y values). (a) SVM's predictions, (b) SNN's predictions and (c) DNN's predictions.

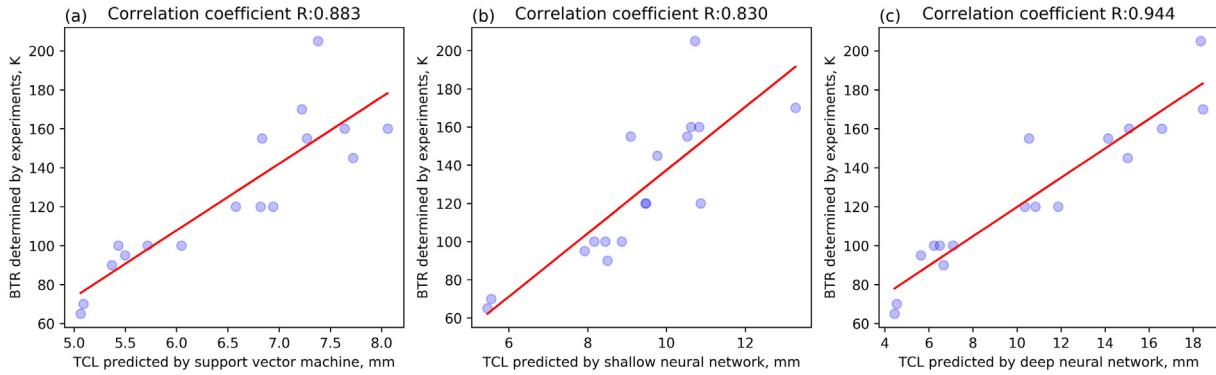


Fig. 7. Regression analysis between TCL predictions and BTR experimental results. (a) SVM's predictions, (b) SNN's predictions and (c) DNN's predictions.

of data transfer. Auto-encoder is a special one hidden layer shallow neural network which has the same input layer and output layer. Taking the 4 hidden layers 21-(6-5-4-3)-1 DNN as an example, 5 auto-encoders which have structures of 21-(6)-21, 6-(5)-6, 5-(4)-5, 4-(3)-4, 3-(1)-3 were built in sequence according to the DNN's structure. The neuron number of each auto-encoder is the same as the neuron number of the DNN's corresponding layer. The first auto-encoder takes the input of the DNN as its input and output, and the output of its hidden layer provides the second auto-encoder with the input and output, i.e. the output of the previous auto-encoder's hidden layer provides the next auto-encoder with the input and output. Each trained auto-encoder provides the DNN's corresponding layer with initial weights and biases values. After having been initialized by SAE, DNN was trained (fine-tuned) by the gradient descent algorithm as that used in SNN.

2.3. Workflow

The workflow of the following case study of SCS prediction is shown in Fig. 3, which includes: assembling and analyzing dataset, preprocessing and dividing training/testing dataset, training SVMs/SNNs/DNNs, testing the trained machine learning models, comparing models'

accuracy and choosing the optimal model, validating the optimal model using different SCS indicators and metallurgical experience, and applying the optimal model to make prediction. DNNs' training includes two steps: pre-training and fine-tuning.

3. Case study: Prediction of solidification cracking susceptibility

3.1. Dataset assembling

Longitudinal varestraint test (LVT) is a reliable quantitative SCS test of good repetition. In LVT, a definite bending strain is applied on the weld of sheet metal 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 indicators/index for SCS. A large amount of LVT data are available in the literature. These were why we chose LVT data in the published literature to compile our SCS dataset.

Samples of various thickness and a range of welding process parameters with different applied strains are used in the LVT experiments. This leads to the difficulty in comparing data from different studies for a consolidated conclusion. But it's not a problem for DNN. On the

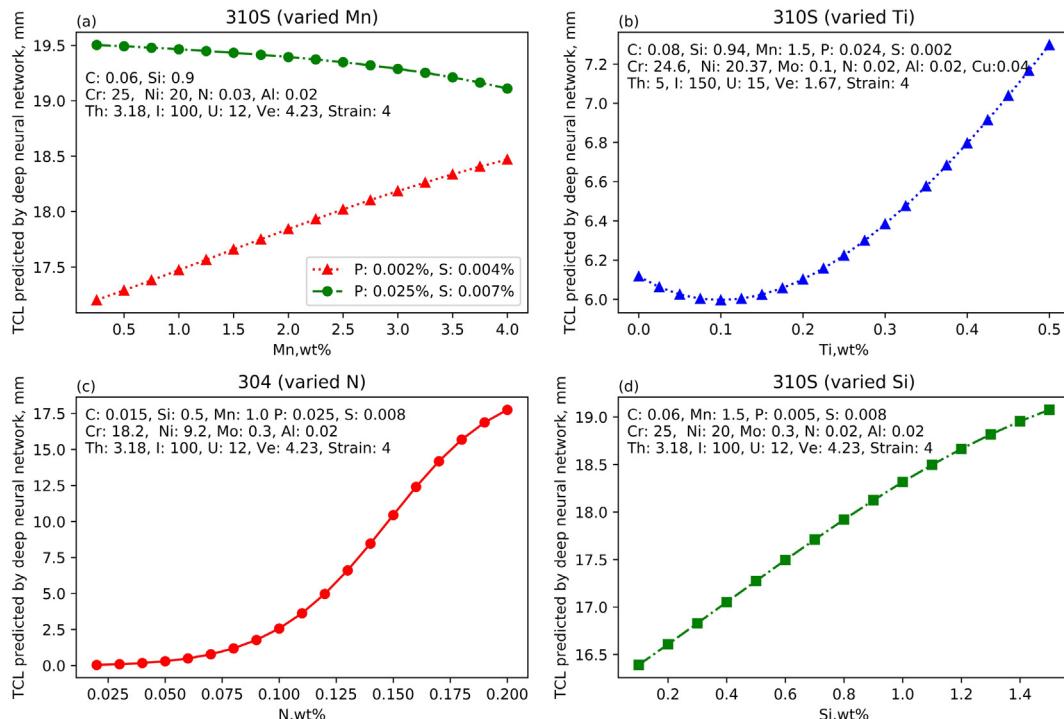


Fig. 8. Predicted TCL as functions of (a) Mn, (b) Ti, (c) N, (d) Si, the compositions used in prediction are shown on the figures.

$W^{(1)}$	$b^{(1)}$																																																						
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	0.89075	0.27322	0.25324	-0.64984	0.07199	0.40058																																		
-0.08141	0.11454	0.28791	-0.47611	-0.14114	1.15287	-0.12534	0.89128	-0.62873	0.03279	0.00436	0.54762	-0.76461	0.02923	0.57712	0.07656	-0.14809	-0.65051	1.11834	0.39843	0.12635	0.80163																																		
-0.96144	0.05044	-0.33331	-0.79264	0.36823	-0.01755	-0.09993	0.20156	-0.92305	0.68631	-0.15720	-0.28002	0.02075	-1.10695	-0.36385	0.55782	0.06162	-0.58075	-0.26965	-0.03545	1.72419	1.18896																																		
-0.28734	-0.39078	-0.53766	-0.37878	0.18985	0.24885	0.20358	-0.21241	0.19951	-0.01238	0.25977	0.06853	0.91162	0.21018	-0.42271	0.11499	-0.64558	0.23226	-1.04965	-0.49250	0.21684	0.67331																																		
-0.33686	0.19210	0.72565	0.67674	0.11436	-0.89480	-1.13927	-0.80494	0.02992	-0.02757	0.20245	-0.74528	0.10759	0.20419	-0.09976	-0.17644	0.49419	0.34569	-0.07508	0.49617	-0.20636	-0.56453																																		
0.59653	0.28000	0.24845	0.00150	-0.36076	-0.48237	0.16221	-0.02181	0.21926	-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																																		
$W^{(2)}$	$b^{(2)}$																																																						
-0.94332	-1.71903	1.21223	-1.83675	-0.62682	0.75407	-0.11556	$h_i^{(1)} = \tanh(\sum_j W_{ij}^{(1)} X_j + b_i^{(1)})$																																																
0.18031	0.50986	-0.43079	0.39878	0.82744	0.63286	-0.07297	$h_i^{(k)} = \tanh(\sum_j W_{ij}^{(k)} h_j^{(k-1)} + b_i^{(k)}), k = 2, 3, 4$																																																
0.30440	0.36401	-0.19122	0.59981	0.41342	0.88375	-0.23361	$Z = h_i^{(5)} = \sum_j W_{ij}^{(5)} h_j^{(4)} + b_i^{(5)}$																																																
0.15303	-0.51368	0.56099	-0.30175	-0.96129	-0.73518	0.06317	X is the normalized input vector,																																																
0.33916	0.24617	-0.10838	0.53502	0.25802	0.92477	-0.24121	Z is the normalized output value,																																																
$W^{(3)}$	$b^{(3)}$														$h_i^{(k)} (k=1, 2, \dots, 5)$ is the output of the kth layer,																																								
2.43093	-0.71937	-0.38082	0.79999	-0.21163	-0.41123	$W_{ij}^{(k)}$ and $b_i^{(k)}$ are the weights and bias matrix of the kth layer.																																																	
-0.98336	0.91622	0.75527	-0.99166	0.64225	0.11795																																																		
0.40206	-0.11321	-0.31942	0.08885	-0.35521	-0.04758																																																		
-0.85122	0.59576	0.88247	-0.59981	0.90241	-0.10502																																																		
$W^{(4)}$	$b^{(4)}$																																																						
-0.04615	0.06806	-0.11679	0.19370	0.00189																																																			
0.70474	-0.68230	-0.07769	-0.10960	-0.25479																																																			
2.40762	-1.66337	0.62795	-1.60591	-0.32387																																																			
$W^{(5)}$	$b^{(5)}$																																																						
-0.21672	-0.50473	1.25758	-0.08385																																																				

Fig. 9. The weights and biases matrix and the mathematical expressions of the optimal 21-(6-5-4-3)-1 DNN.

contrary, when we add these variables into the dataset, the more variations in the data space, the more opportunities for DNN to discover a complicated hidden relationship.

In this study, a dataset containing 575 LVT results of stainless steel was collected from the literature [53–70]. The dataset contains the information of ferritic, austenitic, duplex, precipitation hardening stainless steels, and comprises the following information: steel composition information of 16 elements (contents of C, Si, Mn, P, S, Cr, Ni, Mo, N, Nb, Co, Cu, Al, Ti, V, B, all compositions are expressed in wt% in this paper), samples thickness (Th, mm) and 3 welding parameters (welding current: I, A; voltage: U, V; velocity: Ve, mm/s), restraint information (applied strain: ε , %) and SCS indicator (TCL, mm).

The elements contents which were not specified in the literature were assumed as follows: Al = 0.02, N = 0.02, Mo = 0, Nb = 0, Cu = 0, V = 0, B = 0. Data points miss too much information was discarded, and the final size of the dataset is 487.

3.2. Dataset analysis

Fig. 4 shows the histograms of the 22 variables with minima, maxima and mean values in the final dataset. Most variables, except elements Co, Al, V, B, are well-distributed and suitable for the modeling.

3.3. Dataset pre-processing

Principal component analysis (PCA) is known to reduce data dimensions and noise in the input data [71]. For this study, 21 raw input datasets were pre-processed by PCA method that resulted in two datasets, one with 15 principle components (percentage of the total variance >99%) and the other with 10 principle components (percentage of the total variance >95%). These two pre-processed datasets as well as raw dataset were used as input in the training and testing of SVM, SNN and DNN.

Following the standard practice in neural network modeling, input variables and target variables were normalized before applying them to the network [50]. The input variables were normalized between -0.5 and +0.5, and the target variables were normalized between -0.99 and +0.99.

3.4. Dataset division

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

3.5. Training SVMs/SNNs/DNNs

Support vector machine and shallow neural network were also trained and tested in this study to validate the accuracy advantage of DNN.

In this study, SVM regression with linear/Gaussian/(order 2 and 3) polynomial kernels was used and the hyper-parameters of the SVMs were optimized with Bayesian optimization [72].

The SNN and DNN structures used in this study are similar to that shown in Fig. 1. The SNNs and DNNs consisted of 21 input neurons (one neuron corresponds to one input variables), a few hidden neurons (the neurons number is varied), and one output neurons for TCL. The transfer function of the hidden layers is a hyperbolic tangent function $\tanh(x)$. The transfer function of the output layer is a linear function $y = x$. We trained the following SNNs and DNNs: 1 hidden layer SNNs of 21-(N₁)-1 structure with N₁ from 1 to 35; 2 hidden layers SNNs of 21-(N₁-N₂)-1 structure with N₁ from 4 to 20 and N₂ from 2 to 20; 3 hidden layers DNNs of 21-(4-3-2)-1 and 21-(5-4-3)-1 structure; 4 hidden layers DNNs of 21-(5-4-3-3)-1 and 21-(6-5-4-3)-1 structure; and 5 hidden layers DNNs of 21-(6-5-4-3-3)-1 and 21-(5-4-3-3-3)-1 structure. The numbers in the bracket represent the neuron numbers of the corresponding hidden layers. All SNNs and DNNs were trained using random initialization, and DNNs were also trained through SAE pre-training and fine-tuning. The maximum training epoch was 1000.

Each SVM/SNN/DNN configuration was trained >100 times using different random seeds, then the optimal SVM, SNN and DNN were chosen to make comparison. All calculations were carried out in Matlab

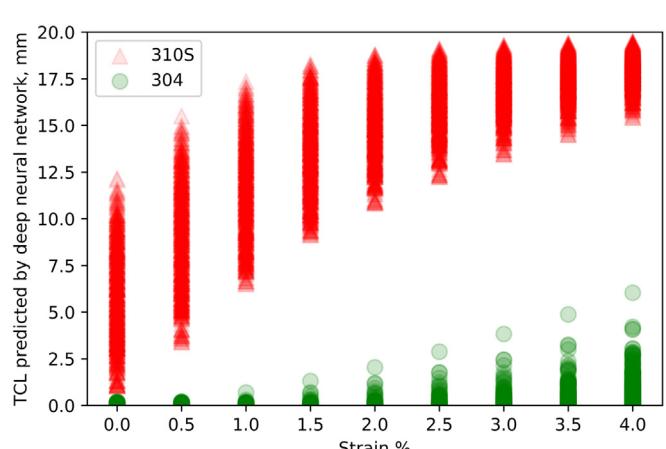


Fig. 10. TCLs of 304 and 310S stainless steel vary with strain, Cr, Ni, P, S contents vary in specification.

Table 2

The specifications (wt%) of 304 and 310S stainless steel and the testing parameters used for 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

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 DNN configuration with one random seed on a personal computer (@2.6 GHz processor) varied from several seconds to dozens of seconds, which is comparable with that of SNN and is obviously shorter than that of training DNN for image recognition where many GPUs and many hours or even many days are needed. Pearson correlation coefficients R_s of the target values and SVM/SNN/DNN prediction values were calculated as the index of training/testing accuracy.

3.6. Comparison of models' accuracy

After training SVMs, SNNs and DNNs, their training/testing accuracy is summarized and compared. Table 1 shows that PCA pre-process play

a deleterious effect on the training and testing accuracy, which may be caused by loss of nonlinear information in PCA pre-process [50]. The rest of this study is all based on dataset without PCA pre-process.

Fig. 5(a) shows the 1 hidden layer SNNs' training accuracy and testing accuracy as functions of the neuron number in the hidden layer. The SNN'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 accuracy is 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 DNN, for

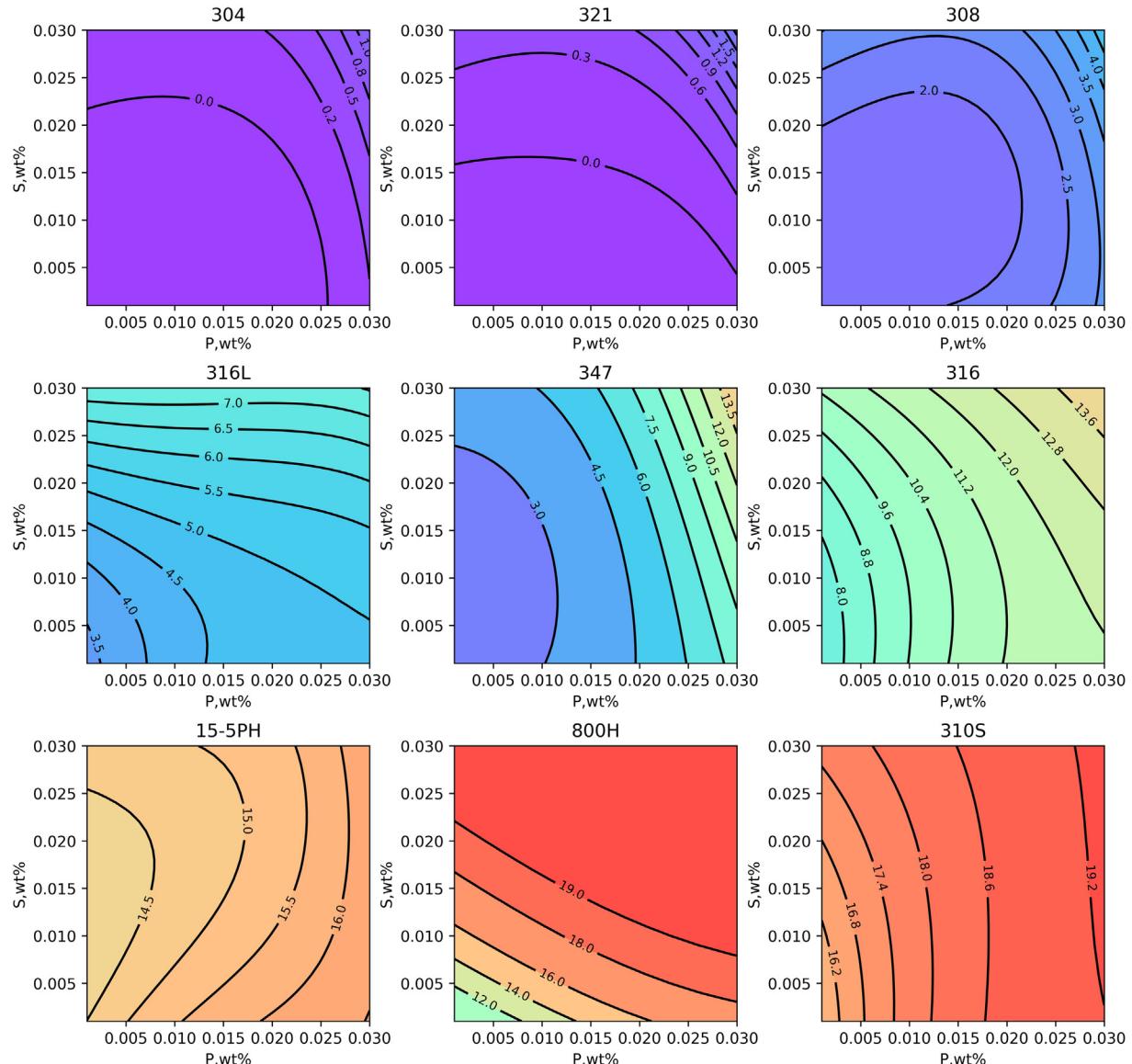


Fig. 11. The predicted TCL contours distribution of 304, 321, 308, 316L, 347, 316, 15-5PH, 800H, 310S stainless steel in P (0–0.03 wt%) and S (0–0.03 wt%) content space.

Table 3

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

Code	C	Si	Mn	Cr	Ni	Mo	N	Nb	Cu	Al	Ti	Rank
304	0.04	1	1	19	8.5	0	0.06	0	0	0.02	0	1
321	0.04	1	1	18	10	0	0.06	0	0	0.02	0.3	2
347	0.04	1	1	18	10	0	0.06	0.3	0	0.02	0	5
308	0.04	1	1	20	11	0	0.06	0	0	0.02	0	3
316 L	0.01	0.5	1	17	13	2	0.06	0	0	0.02	0	4
316	0.06	0.5	1	17	13	2	0.06	0	0	0.02	0	6
15-5PH	0.02	1	1	15	5	0	0.06	0.25	3	0.02	0	7
800H	0.02	1	1	23	30	0	0.06	0	0	0.4	0.6	8
310S	0.02	1	1	25	20	0	0.06	0	0	0.02	0	9

higher testing accuracy, which represents the real learning ability (the training accuracy can be perceived as the memory ability). The best training accuracy of SNNs is 0.99, and the best testing accuracy of SNNs is 0.89. So, the SNN of 21-(21)-1 structure was chosen as the final optimal SNN that was used for comparison in the next step.

Fig. 5(b) shows the optimal training and testing accuracy of SVM, SNNs and DNNs. When three DNNs of 3 to 5 hidden layers were trained like SNNs, they show lower testing accuracy than SNNs and SVM. But

three DNNs show higher testing accuracy than SNNs and SVM through SAE pre-training and fine-tuning. The training accuracy of three DNNs and SNN are close but higher than SVM. The 4 hidden layers DNN 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 SVM and SNN i.e. 0.89.

Fig. 6 shows the linear regression of the target values and the optimal SVM's, SNN's and DNN's prediction values on the unseen testing dataset, and the accuracy (Pearson correlation coefficient R) is also shown on figures. We can see the SVM and SNN show greater prediction errors than DNN.

3.7. Validation

To further validate the generalization ability of the optimal DNN, their prediction results (based on input data that are not part of our train/test dataset) are compared with other SCS indicators, e.g. BTR (TCL has good linear correlation with BTR [73]) and metallurgical experience.

Arata [74] measured the BTRs of 17 stainless steels of different compositions using transverse varestraint test experiments. Fig. 7 shows the

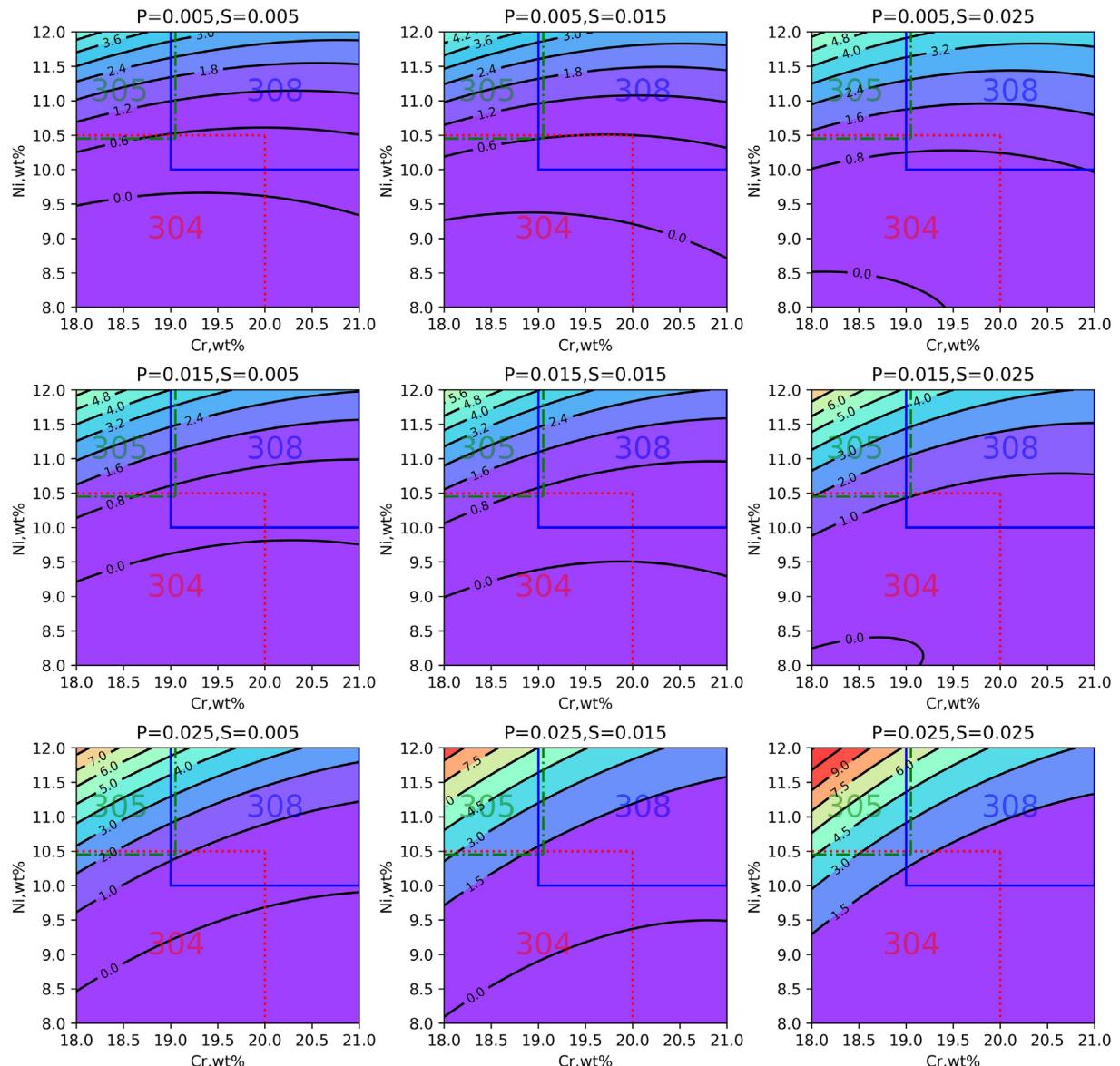


Fig. 12. The predicted effect of Ni (8–12 wt%) and Cr (18–21 wt%) on SCS with 9 different P and S content combinations.

comparison between Arata's BTR results and TCL values predicted by our optimal SNN and DNN. We can see a better linear relationship between BTR results and TCL predictions by our optimal DNN (correlation coefficient $R = 0.944$) than that of SNN ($R = 0.830$) and SVM ($R = 0.883$). This further confirms that DNN has a higher generalization performance than SNN and SVM.

Matsuda [75] reported that Mn addition to SUS310S with 0.002 wt% P and 0.004 wt% 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 0.025 wt% P and 0.007 wt% S improves solidification cracking resistance. Prediction was carried out for the same SUS310S compositions published in Matsuda's paper, the two curves of Fig. 8(a) represent the DNN's predictions. The opposite tendency of the two curves verifies these behaviors: interactions between elements P, S, Mn causes final opposite SCS dependence on Mn.

Matsuda's research [76] shows that adding Ti to SUS310S of 0.024 wt% P and 0.002 wt% S improve SC resistance until reaching an optimal Ti content, then excessive Ti content becomes harmful to SC resistance. The DNN's prediction of the effect of Ti addition on SCS shown in Fig. 8(b) is consistent with Matsuda's description.

Matsuda [77] also reported that as the N content was gradually increased to about 0.2 wt% in SUS 304 weld metal, the SCS increased and finally was similar to that of SUS310S weld metal. The DNN's prediction shown in Fig. 8(c) again reproduced this metallurgical experience.

Arata [78] reported that Si increases SCS in fully austenitic stainless steel SUS310S since the element Si was likely to segregate to boundaries, and our DNN reproduced this behavior well as shown in Fig. 8(d).

Those validations using metallurgical experience confirm the correctness and generalization ability of our optimal DNN.

3.8. Prediction of solidification cracking susceptibility (SCS) using the optimized DNN

3.8.1. Mathematical expressions of DNN
The weights and biases values and mathematical expressions of the optimal 21-(6-5-4-3)-1 DNN are shown in Fig. 9. With these we can estimate SCS dependence on compositions and strains (small strains can be utilized to predict SCS in casting and large strains can be chose to predict SCS in welding and additive manufacturing) which saves expensive and time consuming experiments, and enables us to select the best solidification cracking resistant alloy composition, make comparison with experiment results, and so on. The following subsections provide three examples of its applications. However, the applications of the DNN are not limited to these.

3.8.2. SCS vs. strain

Fig. 10 shows the TCL (the indicator of SCS) variation of 304 and 310S stainless steel when elements Ni, Cr, P, S vary in the specification (shown in Table 2). It shows that SCS is very sensitive to composition variances. An alloy with small difference in the composition can lead to great difference in solidification cracking behaviors.

3.8.3. SCS contours of 9 steels with varied P, S content

The SCS dependence of 9 different stainless steels on P and S is shown in Fig. 11. The compositions used in predictions are shown in Table 3, other parameters are the same: Th = 3.18 mm, I = 100 A, U = 12 V, Ve = 4.23 mm/s, and the rough rank of the 9 steels' SCS is also given in Table 3, the rank is identical with Arata's results [78]. We should bear in mind that the rank is based on the composition shown in the Table 3 and small change in the composition may result in a different rank.

Fig. 11 shows the complex patterns of the 9 steels' SCS contours which are caused by complex interactions of multiple elements in stainless steels. The predictions are in good accordance with metallurgical

experience: fully austenitic stainless steels like 310S is more susceptible to solidification cracking than stainless steels which contain a certain amount of ferrite like 304; the stabilizer elements Nb in 347 and Ti in 321 are detrimental to solidification cracking resistance, and Nb is more harmful than Ti; C which tends to segregate to grains boundaries increase SCS (see the SCS difference between 316 L and 316); low melting point element Cu which insoluble in matrix increases SCS (see 15-5PH); impurity elements P and S have different impact on SCS (most contours are not symmetry along P=S diagonal), the commonly linear superposition P and S is improper. DNN has transformed the scattered data points into an expressive high-dimensional map and those contours are only some slices of it.

3.8.4. SCS contours with varied Ni, Cr contents

The SCS dependence on Ni and Cr is shown in Figs. 12, and 9 contours correspond to low/medium/high P and low/medium/high S content combinations. Other fixed parameters used in predictions are as follow: C = 0.04 wt%, Si = 0.8 wt%, Mn = 1 wt%, N = 0.02 wt%, Al = 0.02 wt%, Th = 3.18 mm, I = 100 A, U = 12 V, Ve = 4.23 mm/s, strain = 3%. The red dotted rectangle shows the Ni and Cr regions of 304 stainless steel, and the green dash-dotted rectangle shows that of 305 stainless steel and the blue solid rectangle shows that of 308 stainless steel. SCS increases with the increment of P and S contents, but different areas vary in different degrees of sensitivity. High ratio of Cr equivalent to Ni equivalent (see the bottom right of each contour) is good for solidification cracking resistance by forming a certain amount of ferrite like that in 304 and 308, because ferrite can accept more S and P than austenite and irregular ferrite/austenite grain boundary is not in favor of the propagation of cracks. The different contour line variation trends between a row (fixed P content, varied S content) and a column (fixed S content, varied P content) show that: stainless steels containing ferrite like 304 and 308 are more sensitive to the S, i.e. SCS of 304 and 308 varies more significantly in a row than that in a column; fully austenitic steels like 305 are sensitive to both P and S, but the influence of P is greater than S.

4. Conclusions

Fully connected DNN which consists of 3 or more hidden layers shows its advantage over shallow neural network and support vector machine in that it can achieve higher prediction accuracy and better generalization performance.

Through DNN regression, vast scattered experimental data in the literature can produce simple quantitative expression of specific material property as function of chemistry composition and processing parameters, etc. The derived mathematical expression can be used in the material defects prediction, new alloys development, and comparison with experimental results.

Though DNN with big datasets is the best choice, DNN with small datasets and pre-training can be a reasonable choice when big datasets are unattainable. In material science, small datasets are common, and the problems to be solved have fewer input variables than that in image recognition. Thus, deep and narrow (neuron number in each hidden layer is small) neural network are suitable for material problems, such as solidification cracking susceptibility prediction in this paper, and pretraining using stacked auto-encoder is effective and necessary in the DNN regression of numerical small datasets. Study in this paper demonstrates that small/narrow DNN with small dataset and special training methods has huge potential for extensive applications in material study, especially for those multivariable nonlinear problems.

CRediT authorship contribution statement

Shuo Feng: Data curation, Formal analysis, Writing – original draft. **Huiyu Zhou:** Supervision, Writing - review & editing. **Hongbiao Dong:** Supervision, Writing - review & editing.

Acknowledgements

Shuo Feng wishes to acknowledge EPSRC CDT (Grant No: EP/L016206/1) in Innovative Metal Processing for providing a PhD studentship for this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.matdes.2018.11.060>.

References

- [1] L. Guo, H. Roelofs, M.I. Lembke, H.K.D.H. Bhadeshia, Modelling of transition from upper to lower bainite in multi-component system, *Mater. Sci. Technol.* 33 (4) (2017) 430–437.
- [2] M. Vasudevan, A.K. Bhaduri, B. Raj, K.P. Rao, Artificial neural network modelling of solidification mode in austenitic stainless steel welds, *Mater. Sci. Technol.* 23 (4) (2007) 451–459.
- [3] D.R. Cassar, A.C.P.L.F. de Carvalho, E.D. Zanotto, Predicting glass transition temperatures using neural networks, *Acta Mater.* 159 (2018) 249–256.
- [4] R.C. Dimitriu, H.K.D.H. Bhadeshia, C. Fillon, C. Poloni, Strength of ferritic steels: neural networks and genetic programming, *Mater. Manuf. Process.* 24 (1) (2008) 10–15.
- [5] S. Malinov, W. Sha, Application of artificial neural networks for modelling correlations in titanium alloys, *Mater. Sci. Eng. A* 365 (1) (2004) 202–211.
- [6] W. Solano-Alvarez, M.J. Peet, E.J. Pickering, J. Jaiswal, A. Bevan, H.K.D.H. Bhadeshia, Synchrotron and neural network analysis of the influence of composition and heat treatment on the rolling contact fatigue of hypereutectoid pearlitic steels, *Mater. Sci. Eng. A* 707 (2017) 259–269.
- [7] H.K.D.H. Bhadeshia, D.J.C. MacKay, L.E. Svensson, Impact toughness of C-Mn steel arc welds – Bayesian neural network analysis, *Mater. Sci. Technol.* 11 (10) (1995) 1046–1051.
- [8] T. Sourmail, H.K.D.H. Bhadeshia, D.J.C. MacKay, Neural network model of creep strength of austenitic stainless steels, *Mater. Sci. Technol.* 18 (6) (2002) 655–663.
- [9] G. Allegri, Modelling fatigue delamination growth in fibre-reinforced composites: power-law equations or artificial neural networks? *Mater. Des.* 155 (2018) 59–70.
- [10] Z. Anjum, F. Qayyum, S. Khushnoor, S. Ahmed, M. Shah, Prediction of non-propagating fretting fatigue cracks in Ti6Al4V sheet tested under pin-in-dovetail configuration: experimentation and numerical simulation, *Mater. Des.* 87 (2015) 750–758.
- [11] B.D. Conduit, N.G. Jones, H.J. Stone, G.J. Conduit, Design of a nickel-base superalloy using a neural network, *Mater. Des.* 131 (2017) 358–365.
- [12] M.W. Dewan, D.J. Huggett, T. Warren Liao, M.A. Wahab, A.M. Okeil, Prediction of tensile strength of friction stir weld joints with adaptive neuro-fuzzy inference system (ANFIS) and neural network, *Mater. Des.* 92 (2016) 288–299.
- [13] G. Liu, L. Jia, B. Kong, K. Guan, H. Zhang, Artificial neural network application to study quantitative relationship between silicide and fracture toughness of Nb-Si alloys, *Mater. Des.* 129 (2017) 210–218.
- [14] O. Palavar, D. Özürek, A. Kalyon, Artificial neural network prediction of aging effects on the wear behavior of IN706 superalloy, *Mater. Des.* 82 (2015) 164–172.
- [15] X. Xia, J.F. Nie, C.H.J. Davies, W.N. Tang, W. Xu, N. Birbilis, An artificial neural network for predicting corrosion rate and hardness of magnesium alloys, *Mater. Des.* 90 (2016) 1034–1043.
- [16] L. Ajun, L. Hejun, L. Kezhi, G. Zhengbing, Applications of neural networks and genetic algorithms to CVI processes in carbon/carbon composites, *Acta Mater.* 52 (2) (2004) 299–305.
- [17] S. Malinov, W. Sha, J.J. McKeown, Modelling the correlation between processing parameters and properties in titanium alloys using artificial neural network, *Comput. Mater. Sci.* 21 (3) (2001) 375–394.
- [18] K. Ichikawa, H.K.D.H. Bhadeshia, D.J.C. MacKay, Model for solidification cracking in low alloy steel weld metals, *Sci. Technol. Weld. Join.* 1 (1) (1996) 43–50.
- [19] H.K.D.H. Bhadeshia, Neural networks in materials science, *ISIJ Int.* 39 (10) (1999) 966–979.
- [20] H.K.D.H. Bhadeshia, R.C. Dimitriu, S. Forsik, J.H. Pak, J.H. Ryu, Performance of neural networks in materials science, *Mater. Sci. Technol.* 25 (4) (2009) 504–510.
- [21] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT press, Cambridge, 2016.
- [22] J. Schmidhuber, Deep learning in neural networks: an overview, *Neural Netw.* 61 (2015) 85–117.
- [23] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science* 313 (5786) (2006) 504–507.
- [24] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436.
- [25] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, *Neural Comput.* 18 (7) (2006) 1527–1554.
- [26] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, Greedy layer-wise training of deep networks, *Adv. Neural Inf. Proces. Syst.* (2007) 153–160.
- [27] D. Kotecki, J. Lippold, Welding Metallurgy and Weldability of Stainless Steels, John Wiley & Sons, 2005.
- [28] S. Kou, Welding Metallurgy, John Wiley & Sons, 2003.
- [29] J. Campbell, Chapter 8 - Cracks and tears, Complete Casting Handbook, Butterworth-Heinemann, Oxford, 2011 465–497.
- [30] J.A. Dantzig, M. Rappaz, Solidification, EPFL press, 2009.
- [31] S. Li, D. Apelian, Hot tearing of aluminum alloys, *Int. J. Met.* 5 (1) (2011) 23–40.
- [32] W.J. Sames, F.A. List, S. Pannala, R.R. Dehoff, S.S. Babu, The metallurgy and processing science of metal additive manufacturing, *Int. Mater. Rev.* 61 (5) (2016) 315–360.
- [33] D.D. Gu, W. Meiners, K. Wissenbach, R. Poprawe, Laser additive manufacturing of metallic components: materials, processes and mechanisms, *Int. Mater. Rev.* 57 (3) (2012) 133–164.
- [34] L. Aucott, D. Huang, H.B. Dong, S.W. Wen, J. Marsden, A. Rack, A.C.F. Cocks, A three-stage mechanistic model for solidification cracking during welding of steel, *Mettal. Mater. Trans. A* 49 (5) (2018) 1674–1682.
- [35] L. Aucott, D. Huang, H. Dong, S. Wen, J. Marsden, A. Rack, A. Cocks, Initiation and growth kinetics of solidification cracking during welding of steel, *Sci. Rep.* 7 (2017), 40255..
- [36] D.G. Eskin, L. Katgerman Suyitno, Mechanical properties in the semi-solid state and hot tearing of aluminium alloys, *Mater. Sci. Eng. R* 49 (5) (2004) 629–711.
- [37] T. Soysal, S. Kou, A simple test for assessing solidification cracking susceptibility and checking validity of susceptibility prediction, *Acta Mater.* 143 (2018) 181–197.
- [38] S. Kou, A criterion for cracking during solidification, *Acta Mater.* 88 (2015) 366–374.
- [39] M. Rappaz, J.M. Drezet, M. Gremaud, A new hot-tearing criterion, *Mettal. Mater. Trans. A* 30 (2) (1999) 449–455.
- [40] C. Bordreuil, A. Niel, Modelling of hot cracking in welding with a cellular automaton combined with an intergranular fluid flow model, *Comput. Mater. Sci.* 82 (2014) 442–450.
- [41] A. Niel, C. Bordreuil, F. Deschaux-Beaume, G. Fras, Modelling hot cracking in 6061 aluminium alloy weld metal with microstructure based criterion, *Sci. Technol. Weld. Join.* 18 (2) (2013) 154–160.
- [42] B.C. Thomas, Modeling of hot tearing and other defects in casting processes, *ASM Handbook* 22 (2009) 362–374.
- [43] A. Cecen, H. Dai, Y.C. Yabansu, S.R. Kalidindi, L. Song, Material structure-property linkages using three-dimensional convolutional neural networks, *Acta Mater.* 146 (2018) 76–84.
- [44] B.L. Decost, T. Francis, E.A. Holm, Exploring the microstructure manifold: image texture representations applied to ultrahigh carbon steel microstructures, *Acta Mater.* 133 (2017) 30–40.
- [45] R. Kondo, S. Yamakawa, Y. Masuoka, S. Tajima, R. Asahi, Microstructure recognition using convolutional neural networks for prediction of ionic conductivity in ceramics, *Acta Mater.* 141 (2017) 29–38.
- [46] E. Kim, K. Huang, S. Jegelka, E. Olivetti, Virtual screening of inorganic materials synthesis parameters with deep learning, *npj Comp. Mater.* 3 (1) (2017) 53.
- [47] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, *npj Comp. Mater.* 2 (2016), 16028..
- [48] L. Ward, S.C. O'Keefe, J. Stevick, G.R. Jelbert, M. Aykol, C. Wolverton, A machine learning approach for engineering bulk metallic glass alloys, *Acta Mater.* 159 (2018) 102–111.
- [49] Y.T. Sun, H. Bai, M.-Z. Li, W. Wang, Machine learning approach for prediction and understanding of glass-forming ability, *J. Phys. Chem. Lett.* 8 (14) (2017) 3434–3439.
- [50] H.B. Demuth, M.H. Beale, O. De Jess, M.T. Hagan, Neural Network design, Martin Hagan, 2014.
- [51] D.J. MacKay, A practical Bayesian framework for backpropagation networks, *Neural Comput.* 4 (3) (1992) 448–472.
- [52] D.J. MacKay, Bayesian interpolation, *Neural Comput.* 4 (3) (1992) 415–447.
- [53] Y. Arata, F. Matsuda, S. Saruwatari, Varestraint test for solidification crack susceptibility in weld metal of austenitic stainless steels, *Trans. JWRI* 3 (1) (1974) 79–88.
- [54] J.A. Brooks, W.M. Garrison, Weld microstructure development and properties of precipitation-strengthened martensitic stainless steels, *Weld. J.* 78 (8) (1999) 2805–2915.
- [55] M.J. Cieslak, W.F. Savage, Weldability and solidification phenomena of cast stainless-steel, *Weld. Sci.* 59 (5) (1980) S136–S146.
- [56] D.H. Kah, D.W. Dickinson, Weldability of ferritic stainless-steels, *Weld. J.* 60 (8) (1981) S135–S142.
- [57] L. Li, R.W. Messler, Effects of phosphorus and sulfur on susceptibility to weld hot cracking in austenitic stainless steels, *WRC Bul.* 480 (2003).
- [58] J.C. Lippold, An investigation of weld cracking in alloy 800, *Weld. J.* 63 (3) (1984) S91–S103.
- [59] C. Lundin, C. Qiao, Y. Kikuchi, C. Shi, T. Gill, Investigation of Joining Techniques for Advanced Austenitic Alloys, Oak Ridge National Lab, TN (United States); Tennessee Univ., Knoxville, TN (United States), 1991.
- [60] C.D. Lundin, C.H. Lee, C.Y.P. Qiao, Evaluation of backfilled solidification cracks in austenitic stainless welds in relationship to evaluation of hot cracking, *Weld. J.* 72 (7) (1993) S321–S328.
- [61] C.D. Lundin, C. Qiao, C. Lee, G. Batten, Weldability and hot ductility behavior of nuclear grade austenitic stainless steels, *WRC Bul.* 509 (2006).
- [62] C.D. Lundin, C.Y.P. Qiao, T.P.S. Gill, G.M. Goodwin, Hot ductility and hot cracking behavior of modified 316 stainless-steels designed for high-temperature service, *Weld. J.* 72 (5) (1993) S189–S200.
- [63] H. Madarambe, T. Sukegawa, H. Inoue, Cracking susceptibility of stainless-steel subjected to plasma disruption, *Fusion Eng. Des.* 27 (1995) 499–506.
- [64] T. Ogawa, T. Koseki, Weldability of newly developed austenitic alloys for cryogenic service. 2. High-nitrogen stainless-steel weld metal, *Weld. J.* 67 (1) (1988) S8–S17.
- [65] T. Ogawa, E. Tsunetomi, Hot cracking susceptibility of austenitic stainless-steels, *Weld. J.* 61 (3) (1982) S82–S93.
- [66] A.M. Ritter, W.F. Savage, Solidification and solidification cracking in nitrogen-strengthened austenitic stainless-steels, *Mettal. Mater. Trans. A* 17 (4) (1986) 727–737.
- [67] C.V. Robino, J.R. Michael, M.C. Maguire, The solidification and welding metallurgy of gallium-resistant stainless steels, *Weld. J.* 77 (11) (1998) 446S–457S.

- [68] S. Shono, O. Matsumoto, S. Kawaguchi, T. Urabe, Weldability of 18-percent chromium ferritic stainless-steel with titanium addition, *Trans. Iron Steel Inst. Jpn.* 24 (7) (1984) 557–565.
- [69] I. Varol, W.A. Baeslack, J.C. Lippold, Characterization of weld solidification cracking in a duplex stainless-steel, *Metallography* 23 (1) (1989) 1–19.
- [70] M. Zaghloul, A. Sadek, A. El-Batahgy, M. Hanafy, Effect of welding parameters on hot cracking susceptibility of alloy 800, *Yosetsu Gakkai Ronbunshu* 12 (3) (1994) 335–341.
- [71] R. May, G. Dandy, H. Maier, Review of input variable selection methods for artificial neural networks, *Artificial Neural Networks-methodological Advances and Biomedical Applications*, InTech, 2011.
- [72] Statistics and Machine Learning Toolbox™ User's Guide., The MathWorks, Inc., 2018
- [73] A.K. Bhaduri, S.K. Albert, B. Raj, 2 - robust welding technologies for ferrous alloys, in: R. Singh (Ed.), *Weld Cracking in Ferrous Alloys*, Woodhead Publishing 2009, pp. 34–95.
- [74] Y. Arata, F. Matsuda, H. Nakagawa, S. Katayama, 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 (2) (1978) 169–172.
- [75] F. Matsuda, H. Nakagawa, S. Katayama, Y. Arata, 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 (2) (1982) 79–85.
- [76] F. Matsuda, S. Katayama, Y. Arata, Solidification crack susceptibility in weld metals of fully austenitic stainless steels (report IX): effect of titanium on solidification crack resistance, *Trans. JWRI* 12 (2) (1983) 247–252.
- [77] F. Matsuda, H. Nakagawa, S. Katayama, Y. Arata, Solidification crack susceptibility in weld metals of fully austenitic stainless steels (report VIII): effect of nitrogen on cracking in SUS 304 weld metal, *Trans. JWRI* 12 (6) (1983) 89–95.
- [78] Y. Arata, F. Matsuda, S. Katayama, 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 (1) (1977) 105–116.