International Journal of Minerals, Metallurgy and Materials
Volume 28, Number 8, August 2021, Page 1309
https://doi.org/10.1007/s12613-020-2168-z

Prediction of the Charpy V-notch impact energy of low carbon steel using a shallow neural network and deep learning

Si-wei Wu[1)], *Jian Yang*[1)], *and Guang-ming Cao*[2)]

1) State Key Laboratory of Advanced Special Steel, School of Materials Science and Engineering, Shanghai University, Shanghai 200444, China
2) State Key Laboratory of Rolling and Automation, Northeastern University, Shenyang 110819, China
(Received: 17 July 2020; revised: 6 August 2020; accepted: 10 August 2020)

Abstract: The impact energy prediction model of low carbon steel was investigated based on industrial data. A three-layer neural network, extreme learning machine, and deep neural network were compared with different activation functions, structure parameters, and training functions. Bayesian optimization was used to determine the optimal hyper-parameters of the deep neural network. The model with the best performance was applied to investigate the importance of process parameter variables on the impact energy of low carbon steel. The results show that the deep neural network obtains better prediction results than those of a shallow neural network because of the multiple hidden layers improving the learning ability of the model. Among the models, the Bayesian optimization deep neural network achieves the highest correlation coefficient of 0.9536, the lowest mean absolute relative error of 0.0843, and the lowest root mean square error of 17.34 J for predicting the impact energy of low carbon steel. Among the variables, the main factors affecting the impact energy of low carbon steel with a final thickness of 7.5 mm are the thickness of the original slab, the thickness of intermediate slab, and the rough rolling exit temperature from the specific hot rolling production line.

Keywords: prediction; shallow neural network; deep neural network; impact energy; low carbon steel

1. Introduction

A mechanical property prediction model is beneficial to guide the process design of steel at low cost in a short production cycle. Establishing a mechanical property prediction model based on a data-driven technique has attracted recent attentions [1]. As a typical data-driven modeling method, an artificial neural network (ANN) has advantages of excellent fitting ability, self-adaptive ability, and is independent of expertise knowledge. ANNs have been a common tool to relate the chemical compositions, process parameters, and mechanical property of metal materials. An ANN was successfully applied to predict the mechanical property of austenite stainless steel 304 [2–4]. Powar and Date [5] developed a method to predict the mechanical property of heat-treated components of 30CrMoNiV5-11 steel using ANN and an advanced thermal modeling tool, FLUENT. Lalam *et al.* [6] predicted the mechanical properties of galvanized steel using ANN. Thankachan and Sooryaprakash [7] exploited an ANN to predict the impact energy of cast duplex stainless steels, achieving good performance. Based on rough sets theory, Colas-Marquez *et al.* [8] proposed a new framework to establish a Charpy impact energy prediction model for alloy steels by combining a k-nearest neighbor and neural network model. Mahfouf and Yang [9] predicted the impact energy of steels by combining a genetic algorithm and neural network. ANN has also been successfully applied to predict fatigue damage of S420MC steel [10].

However, the above-mentioned research mainly focused on the shallow ANN model, which does not fully reflect the corresponding relationship among the chemical compositions, process parameters, and mechanical property of low carbon steel because of fewer hidden layers. Therefore, some researchers have developed a deep neural network (DNN) to address the complex correlation. Liu *et al.* [11] established a mechanical property prediction model of hot die steel using a four layer back-propagation neural network to investigate the effect of alloying elements and heat treatment parameters on the mechanical property. Xu *et al.* [12] applied a convolutional neural network to predict the mechanical property of hot rolled steel, achieving a high accuracy. A DNN has a similar structure as a shallow neural network. However, be-

Corresponding author: Jian Yang E-mail: yang_jian@t.shu.edu.cn
© University of Science and Technology Beijing 2021

cause of the increase in hidden layers of the DNN, the hyper-parameters that need to be determined increase rapidly, which makes determination of the optimal hyper-parameters of a multilayer neural network difficult. Therefore, it is difficult to establish the mechanical property prediction model of low carbon steel using a DNN.

In the present work, the data-driven impact energy prediction model of low carbon steel was investigated to realize a highly precise prediction. The shallow neural network and DNN were used to establish the prediction model. The influence of different activation functions and network structures on the model accuracy was analyzed. To address the problem of determination of optimal hyper-parameters in a DNN, the Bayesian optimization method was implemented to automatically determine the optimal hyper-parameters of the model. The correlation coefficient (R), mean absolute relative error (MARE), and root mean square error (RMSE) were used to evaluate the models. The influence of the process parameters on the impact energy of low carbon steel was analyzed based on the best performed model.

2. Materials and experiments

Data of 7211 cases were collected from hot rolling line including the chemical compositions, process parameters, and Charpy V-notch impact energy (Akv) of low carbon steel. The chemical composition included carbon (C), silicon (Si), manganese (Mn), sulfur (S), phosphorus (P), nitrogen (N), niobium (Nb), titanium (Ti), aluminum (Al), copper (Cu), chromium (Cr), and nickel (Ni). The process parameters included the furnace temperature (FT), rough rolling thickness for 1st pass to 5th pass (RRH 1–5), finish rolling thickness for 1st pass to 6th pass (FRH 1–6), rough rolling velocity for 1st pass to 5th pass (RRV 1–5), finish rolling velocity for 1st pass to 6th pass (FRV 1–6), rough rolling force for 1st pass to 5th pass (RRF 1–5), finish rolling force for 1st pass to 6th pass (FRF 1–6), rough rolling exit temperature (RRT), finish rolling exit temperature (FRT), coiling temperature (CT), width (w), and finish thickness specification (Spec). The distribution of chemical composition of the low carbon steel is listed in Fig. 1. The slabs were reheated in a furnace for 120 min, and then used for hot rolling. After laminar cooling, the strips were coiled and cooled in air to room temperature. The impact energy of low carbon steel was evaluated using a Charpy impact test. Defect-free specimens for impact testing were machined and prepared. Given the different target thicknesses of the hot rolled strips, there were three thickness specifications of the impact energy samples, 10 mm × 10 mm × 55 mm, 10 mm × 7.5 mm × 55 mm, and 10 mm × 5 mm × 55 mm, according to the final thickness of the strips. The test specimen geometry is shown in Fig. 2. The specimen has a V-shaped notch with a flank angle of 45° and a depth of 2 mm. The pendulum was equipped with a notch with a tip radius of 0.25 mm. For each strip, three samples were tested and the average value was calculated. The experiments were performed at 20°C in accordance with GB/T 229-2007. The impact energy of samples with different specifications cannot be directly compared. In order to consider the sample specification information with regard to the impact energy prediction model, three impact energy samples were marked as "−1" for a sample thickness of 5 mm, "0" for a sample thickness of 7.5 mm, and "1" for a sample thickness of 10 mm. Based on the extremely nonlinear fitting of a neural network, we attempted to build a prediction model that can predict the impact energy of three sample specifications at the same time. All of the data with chemical compositions and process parameters were combined with the impact energy according to the coil number of the steel. Abnormal data and outliners were eliminated to improve the data quality.

3. Brief descriptions of the modeling technique

3.1. A three-layer neural network

A shallow neural network is the ANN with less than two hidden layer numbers and is usually used to correlate nonlinear relationships among the dependent variable and independent variable. Generally, a shallow neural network is composed of an input layer, a hidden layer, and an output layer. Each layer contains several neurons. The neurons in the same layer connect with each other and the neurons in the different layer are not connected with each other [13]. In each neuron, there is an activation function for processing the data entering the neuron. These interconnected simple neurons form a complex network system for information processing. The prediction process of the neural network can be expressed as Eq. (1) [14],

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{1}$$

where x is the input data, y is the output data, w is the weight used to connect neurons in two adjacent layers, b is the threshold value for neurons, n is the number of neurons, and f is the activation function. The process of training the network is complicated and is aimed at minimizing the cost function by adjusting the weights and thresholds of the network [15]. The mean squared error (MSE) is used as a cost function to represent the error between the predicted value and measured value of the neural network, which is expressed as Eq. (2).

$$\mathrm{MSE} = \frac{1}{N'} \sum_{i=1}^{N'} (O_i - T_i)^2 \tag{2}$$

where N' is the number of data, O is the output value of neural network, and T is the target value. The activation function ("tansig," "sigmoid," "purelin"), training function ("trainbr," "trainlm," "trainscg"), and network structure are the main factors affecting the performance of the neural network. In our work, the different activation functions and training func-

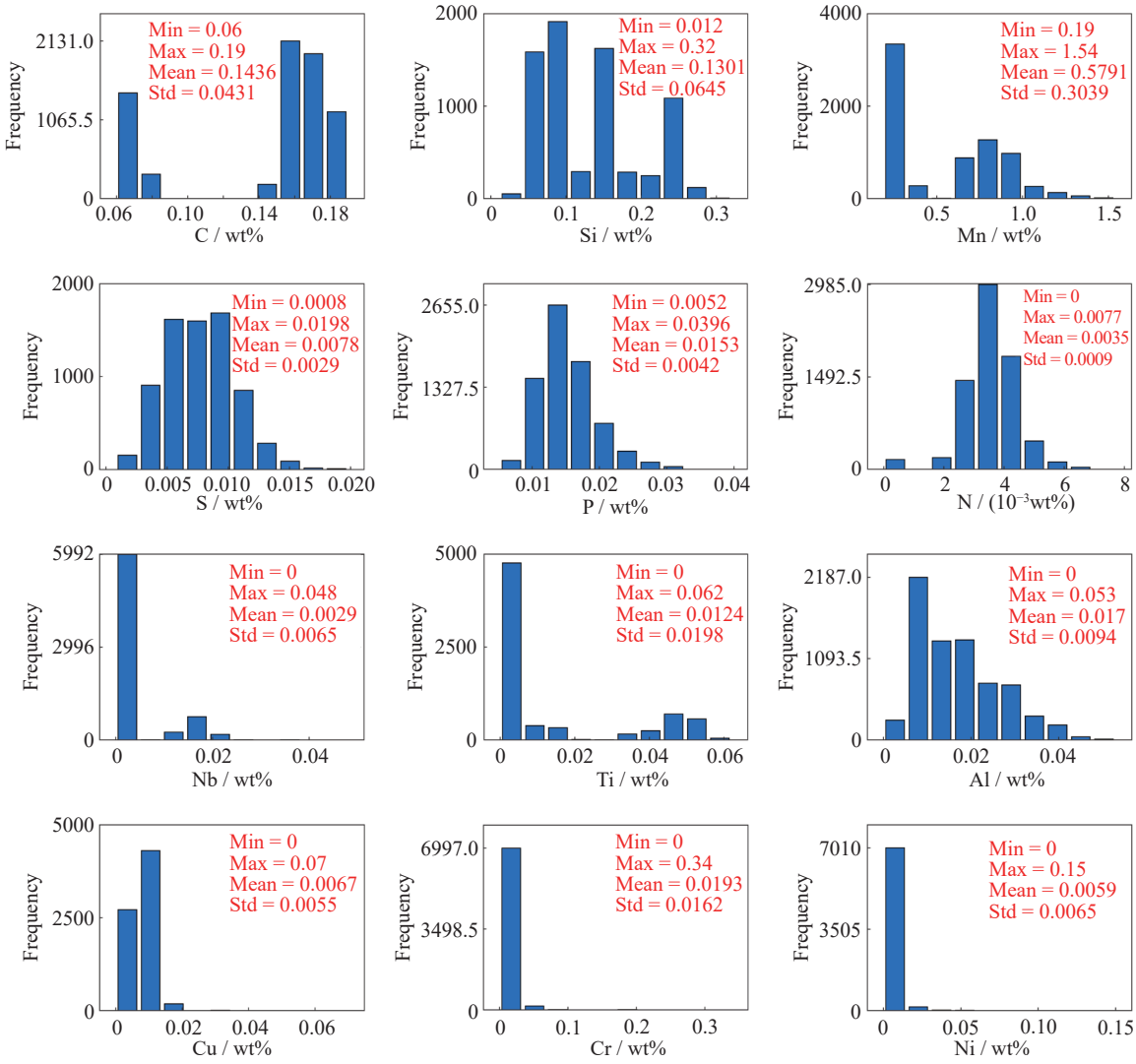

Fig. 1. Distribution of the chemical compositions of low carbon steel. Min, Max, Mean, and Std represent minimum, maximum, mean, and standard deviation of the data, respectively.

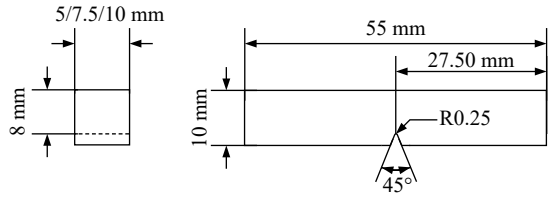

Fig. 2. Geometry of the Charpy V-notch impact specimen.

tions were used to establish the impact energy prediction models of low carbon steel with an optimal network structure.

3.2. Extreme learning machine

To improve the neural network training efficiency, an extreme learning machine (ELM) algorithm was proposed by Huang and the team [16–18]. An ELM, a revised shallow neural network, is a fast learning algorithm. Compared with a neural network, the input weights and hidden thresholds of the ELM algorithm are randomly generated and do not need to be fine-tuned [19]. The least square method is used to calculate the output weights instead of the iteration calculation in neural network training, which can help significantly reduce the modeling time [20]. Moreover, the ELM algorithm has better generalization capability and a faster learning rate than traditional gradient-based algorithms and it can avoid the network sinking in the local minimum [21].

3.3. Deep neural network

When the number of hidden layers of ANN is greater than two, the network is called a DNN. Compared with a shallow neural network, increasing the number of the hidden layers results in the phenomenon of the gradient disappearing in the process of neural network training. Additionally, increasing the number of the hidden layers introduces a large number of

hyper-parameters of the model, resulting in difficulty in selecting the optimal hyper-parameters [22]. The adam training function (derived from an adaptive moment estimation), which is a modification of the scaled conjugate gradient function, has excellent convergence for multilayer neural networks [23]. Therefore, the adam training function was used to minimize MSE. The structure of the network needs to be designed, including the number of hidden layers of network and the number of neurons in each hidden layer. The optimal hyper-parameters of DNN were determined by trial and error or a Bayesian optimization.

3.4. Bayesian optimization

Compared with trial and error methods, such as a grid search, that take significant time and cannot guarantee optimization of the network hyper-parameters, Bayesian optimization is a method to automatically determine the optimal hyper-parameters of the neural network [24]. The algorithm iterates the following steps: (1) establish the surrogate model for the objective function; (2) determine the hyper-parameters that perform best in the surrogate model; (3) evaluate the performance of the objective function with the optimal hyper-parameters in Step (2); (4) update the surrogate model; (5) repeat Steps (2)–(4) until the stop condition is reached. Bayesian optimization samples trial points sequentially, and each trial point is sampled utilizing all of the information in the history; specifically, what will be sampled next is actually determined by the previous samples [25].

The Bayesian optimization is used to find the model hyper-parameters that produce the best score on the measure of validation data. The objective function can be written as Eq. (3),

$$x^* = \mathrm{argmin}\phi(x),\ x \in X \tag{3}$$

where x is the hyper-parameter that can be real, integer, or categorical, x^* is the best hyper-parameter of the model, ϕ is the objective function, and X is the search space of x. Generative models of the Gaussian process were selected as the surrogate.

The expected improvement was selected as acquisition function because of its intuitive and good performance [26]. The expected improvement provides a combination that successfully balances between local and global search, which evaluates the expected improvement of the objective function and ignores values that may increase the objective function. The expected improvement acquisition function finds the next hyper-parameter solutions with high variance and high mean value. The expected improvement can be expressed as Eq. (4),

$$\mu(x) = E(\max(0, \phi(x) - \phi_{\mathrm{best}})) \tag{4}$$

where x is the hyper-parameter of the model, ϕ_{best} is the current best objective function value, E is the expected value, and $\mu(x)$ is the value of the acquisition function. The degree of improvement is the difference between the objective function value at the sampling point value and current optimum value. If the objective function value at the sampling point value is less than the current optimum value, the improvement function is zero [27].

4. Establishment of a prediction model for impact energy

4.1. Data preparation

As the sample thickness of the low carbon steel has three specifications, the number of data with different sample thickness specifications distributes unbalance. Establishing a model based on these data will make the model accurate in some areas but inaccurate in others. For the model to learn all of the impact energy information, the impact energy of low carbon steel was presented as a natural logarithm to improve its distribution. Fig. 3 compares the original and transformed impact energy data distribution. The three peaks correspond to the impact energy of three thicknesses. The max frequency of the impact energy data is different for the samples with different thicknesses, while after the data is transformed,

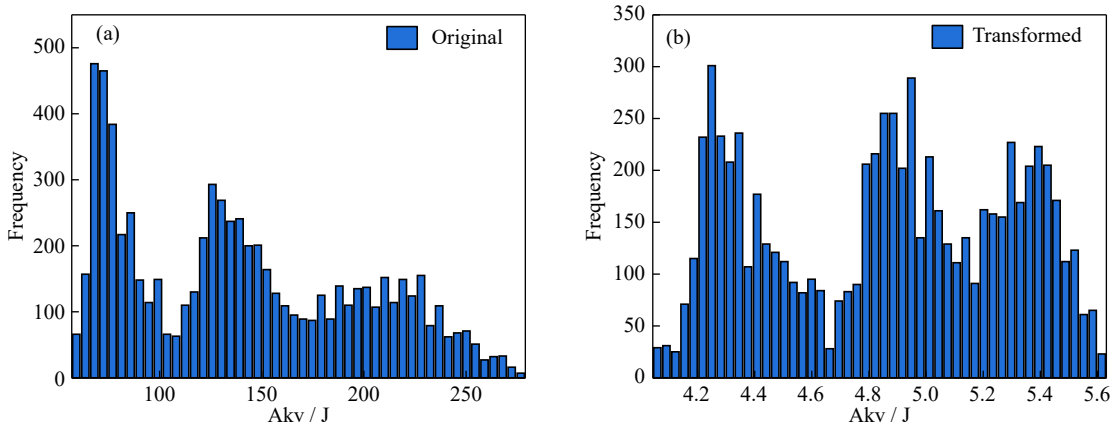

Fig. 3. Comparison of the (a) original and (b) transformed impact energy data distribution.

the distribution of the data improves.

Before modeling, all of the variables were normalized to [0, 1] by Eq. (5) to eliminate the different magnitudes of the variables and accelerate the training efficiency,

$$X_i^{\text{norm}} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{5}$$

where X_i^{norm} is the normalized data, X_i is the original data, and $X_{\min}$ and $X_{\max}$ are the minimum and maximum of the original data. The correlation coefficient of each variable is shown in Fig. 4. The dataset was divided into three parts: 70% of the data (5051) was selected as training data, 15% of the data (1080) was selected as validation data, and the remaining 15% of data (1080) was selected as testing data. Chemical compositions and process parameter variables were selected as input variables, and the impact energy was selected as an output variable.

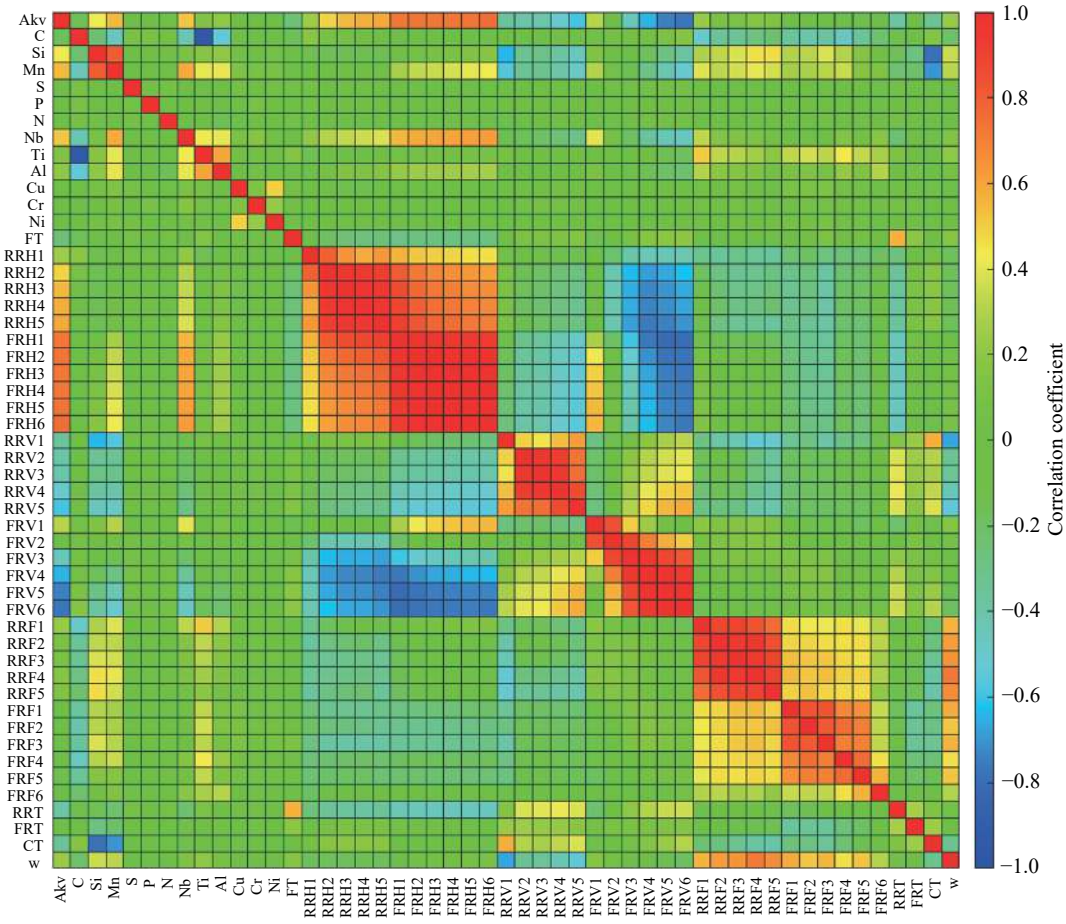

Fig. 4. Correlation coefficient of each variable.

4.2. Training model

The effects of the activation function, training algorithm, and network structure parameter on the network performance were investigated to train the impact energy prediction model of low carbon steel. For a shallow neural network, different activation functions were used to train the network, such as the "tanh," "sigmoid," and "purelin" for the three-layer network and "sine," "sigmoid," and "hardlim" for ELM. "trainbr," "trainlm," and "trainscg" functions were applied as the training function of the three-layer network. The numbers of hidden layers and neurons in each hidden layer were selected by the network performance on the validation data. For the three-layer neural network, the number of hidden layer neurons ranged from 2 to 20. For ELM, the number of hidden layer neurons ranged from 2 to the number of input variables. A trial and error method was applied to determine the best performance neural network [28]. In the three-layer neural network, to prevent the overfitting of the network, a L2 regularizer was applied. The training epoch was set to 2000.

For DNN, the activation functions commonly used to train the network are "tanh," "relu," and "leaky_relu." The training function parameters of adam remained as the default. To determine the structure parameters of the network, the optimal model selection was achieved by grid search for DNN by combining different numbers of hidden layers and neurons in each hidden layer. In our work, the number of hidden layers ranged from 1 to 4 and the number of neurons in each hidden

layer was set to 25, 50, 75, or 100. During DNN training, the learning rate was a hyper-parameter that determined the calculation efficiency, which was set to 0.001 [29]. The number of epochs for network training was set to 2000. Considering that the traditional grid search can only choose good hyper-parameters rather than the best hyper-parameters for DNN, our research applied Bayesian optimization to realize the hyper-parameter optimization. During the process, the real objective function was the MSE of the ANN on the validation data. The hyper-parameters to be determined included the activation functions, such as "tanh," "relu," and "leaky_relu," and the number of neurons in each hidden layer ranged from 2 to 100. Models with the number of hidden layers ranging from 1 to 4 were optimized. The max evolution epoch of the Bayesian optimization was set to 250. The other hyper-parameters remained as default.

For all of the shallow neural networks and DNN, each model was trained three times to avoid the effect of the initial weights on the network performance. All of the models investigated in this research are shown in Fig. 5. There are a total of 16 neural network models for predicting the impact energy of low carbon steel including a shallow neural network and DNN with different activation functions and training functions.

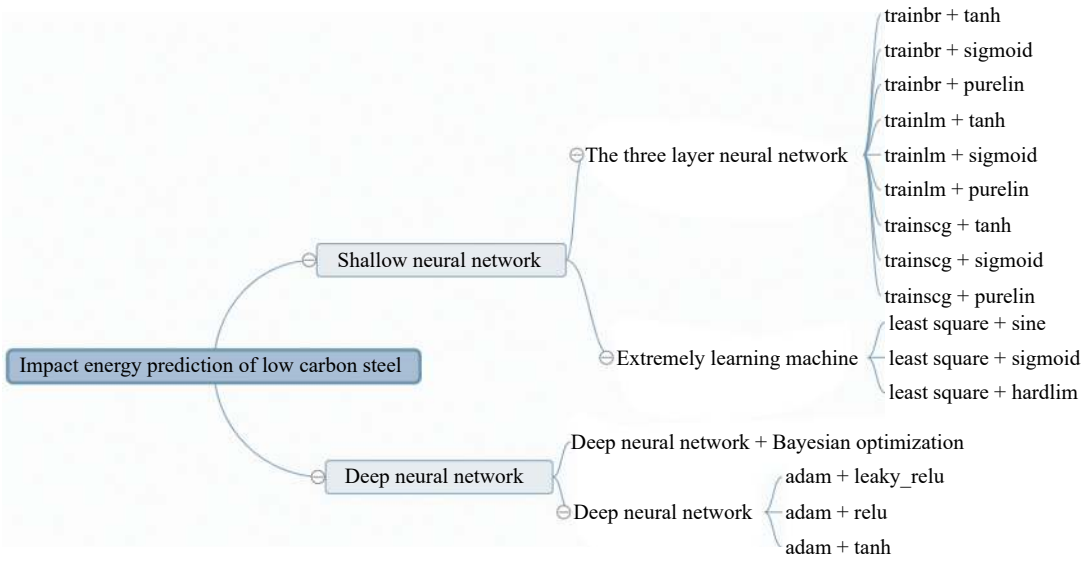

Fig. 5. Models investigated in this research.

The validation data were used to avoid overfitting of the neural network with training data. The prediction accuracy of the model with validation data was compared to select the best performance model. Here, the correlation coefficient and MSE were applied to evaluate the performance of all of the models, which could be written as Eqs. (6) and (2) [30],

$$R = \frac{\sum_{i=1}^{N}\left(M_i - \bar{M}\right)\left(P_i - \bar{P}\right)}{\sqrt{\sum_{i=1}^{N}\left(M_i - \bar{M}\right)^2 \sum_{i=1}^{N}\left(P_i - \bar{P}\right)^2}} \tag{6}$$

where P is the predicted Akv value of neural network, and M is the measured Akv value; $\bar{P}$ and $\bar{M}$ represent the mean values of the predicted and measured Akv values, respectively; N is the number of data.

Figs. 6 and 7 show the R and MSE values of all the shallow neural network models on the validation data. For the models with the same structure, the prediction results of the three-layer neural network models with training data and validation data are quite different, while the differences among the prediction results of the ELM model with training data and validation data are not obvious. For the three-layer neural network with training function of "trainbr," with an increase in the number of hidden layers, the prediction error of the model on the training data decreases gradually, and the prediction error in the validation data decreases first and then increases. The increasing prediction error in the validation data indicates the network is over fitted. When the number of hidden layer neurons is six, the network performs best with the highest R and lowest MSE. The optimal number of hidden layer neurons in other shallow neural networks can be determined in this way.

Fig. 8 shows the change in the minimum objective values with iterations. As the iterations increase, the mean objective value decreases and finally reaches 0.004936. The prediction results of the DNN models combined with different hyper-parameters were also evaluated. The optimal structure parameters of each network can be determined using these prediction results.

The prediction results of each best performance model are shown in Table 1. All of the models with optimal hyper-parameters and performance are presented. DNN produces lower MSE and higher R values than the shallow neural network in

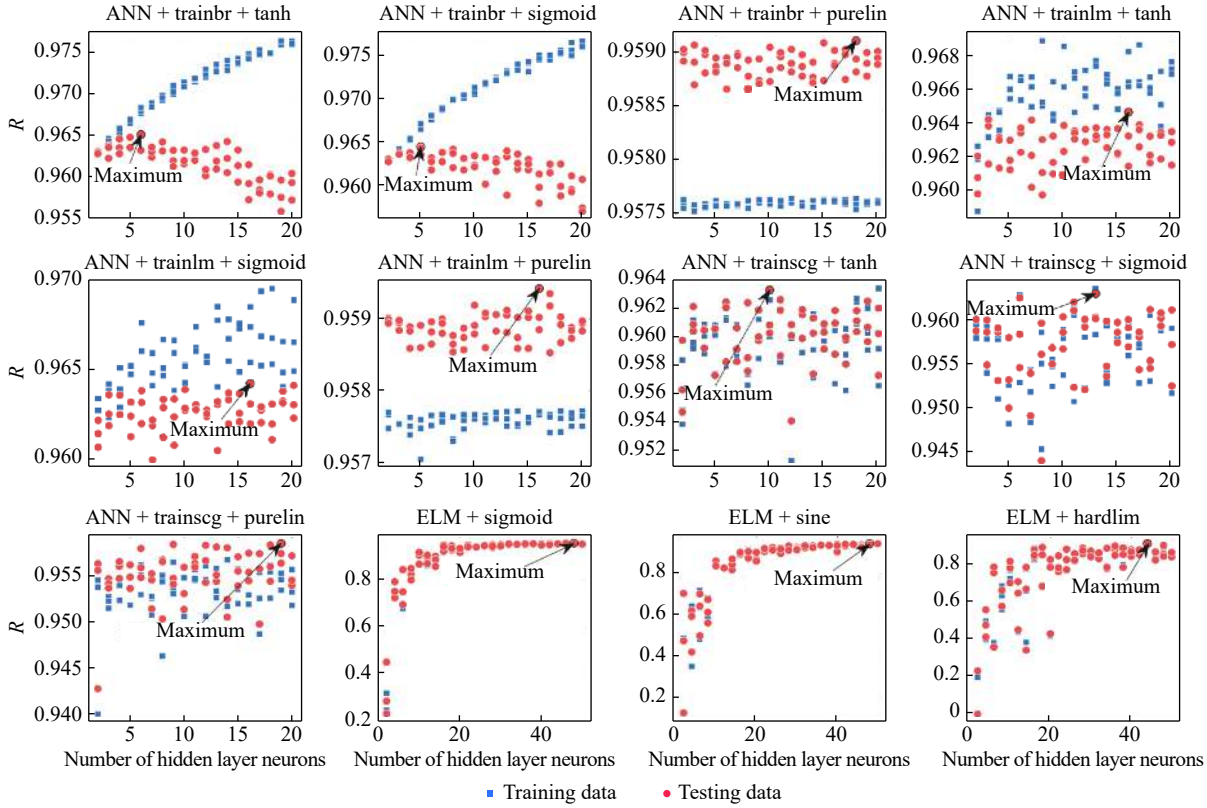

Fig. 6. ***R*** **values of all of the shallow neural network models on the validation data.**

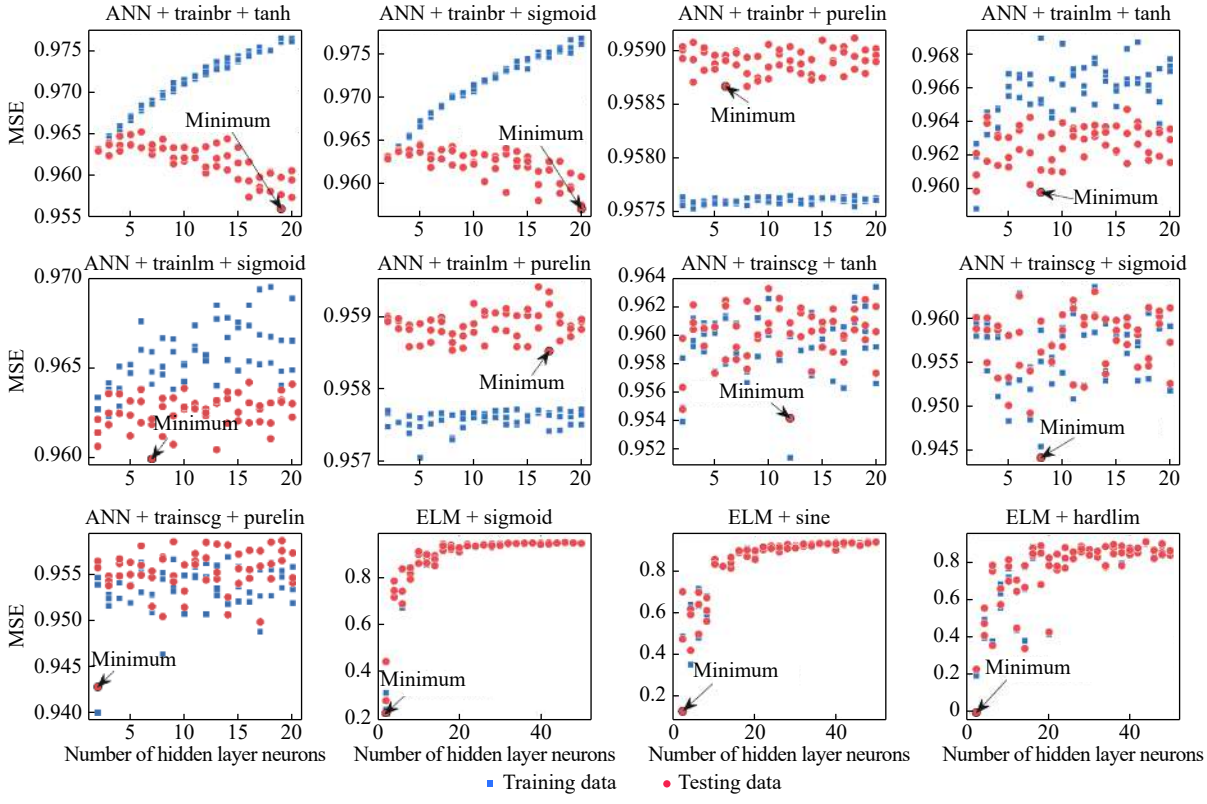

Fig. 7. MSE values of all of the shallow neural network models with validation data.

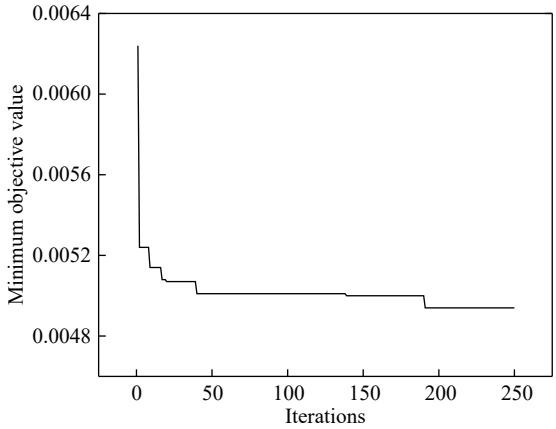

Fig. 8. Change in the minimum objective values with iterations.

both training data and validation data. To determine the optimal hyper-parameter combination, a Bayesian optimization algorithm has more advantages than a grid search to realize hyper-parameter fine tuning. The three hidden layer network with activation function of "leaky_relu" achieves the best performance with R of 0.9655 and MSE of 0.004936 using the Bayesian optimization algorithm.

5. Results and analyses

For a comparative analysis of all of the models, the testing data was used to test the performance of the model. Fig. 9 compares the predicted Akv values with the measured Akv values of all of the models. The red diagonal line in the figure indicates that the predicted value is equal to the measured value, and the points on this line mean that a high accuracy is achieved. All of the models have good accuracy.

To further quantify the model prediction error, R was used to measure the fit performance of the model. MARE and RMSE were applied to evaluate the model prediction accuracy, which can be calculated by Eqs. (7) and (8) [31–32].

$$\mathrm{MARE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{P_i - M_i}{M_i}\right| \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(P_i - M_i)^2}{N}} \tag{8}$$

Fig. 10 shows the R, MARE, and RMSE values of all of the models with testing data. The R of the three-layer neural network is better than that of ELM. DNN performs better than the shallow neural network. The Bayesian optimization DNN achieves the highest R value of 0.9536, lowest MARE of 0.0843, and lowest RMSE of 17.34 J to predict the impact energy of low carbon steel.

Fig. 11 shows the box plot of the prediction errors of all of the models with testing data. As shown by the median line, the prediction errors of Model 13 and Model 14 show a large offset from zero, which may be derived from the rectified linear units "relu" and "leaky_relu." Among all of the models, Model 1, Model 5, Model 7, Model 8, and Model 16 show small offsets. When considering the interval of 10%–90%, Model 2 and Model 16 show narrow ranges with values of 39 J and 39.52 J, respectively. Therefore, Model 16 performs the best.

In order to further compare the prediction error distribution of each model, the hit rates of all of the models are listed

Table 1. Prediction results of the best performance models

Model	Train		Validation		h	m
	R	MSE	R	MSE		
ANN + trainbr + tanh	0.9683	0.004519	0.9651	0.004992	1	6
ANN + trainbr + sigmoid	0.9666	0.004758	0.9644	0.005089	1	5
ANN + trainbr + purelin	0.9576	0.006008	0.9591	0.005813	1	18
ANN + trainlm + tanh	0.9682	0.004526	0.9646	0.005055	1	16
ANN + trainlm + sigmoid	0.9650	0.004985	0.9642	0.005106	1	20
ANN + trainlm + purelin	0.9577	0.005998	0.9594	0.005770	1	16
ANN + trainscg + tanh	0.9626	0.005311	0.9633	0.005229	1	10
ANN + trainscg + sigmoid	0.9637	0.005166	0.9631	0.005260	1	13
ANN + trainscg + purelin	0.9554	0.006316	0.9586	0.005887	1	19
ELM + sigmoid	0.9557	0.006264	0.9581	0.005948	1	48
ELM + sine	0.9484	0.007277	0.9490	0.007219	1	48
ELM + hardlim	0.9179	0.011401	0.9206	0.011066	1	51
DNN + leaky_relu	0.9744	0.003805	0.9660	0.005042	4	100, 25, 25, 100
DNN + relu	0.9738	0.003784	0.9648	0.005105	2	50, 75
DNN + tanh	0.9731	0.003867	0.9654	0.004969	4	75, 75, 25, 25
DNN + Bayesian optimization	0.9696	0.004341	0.9655	0.004936	3	33, 9, 33

Note: h—Number of hidden layers; m—Number of hidden layer neurons.

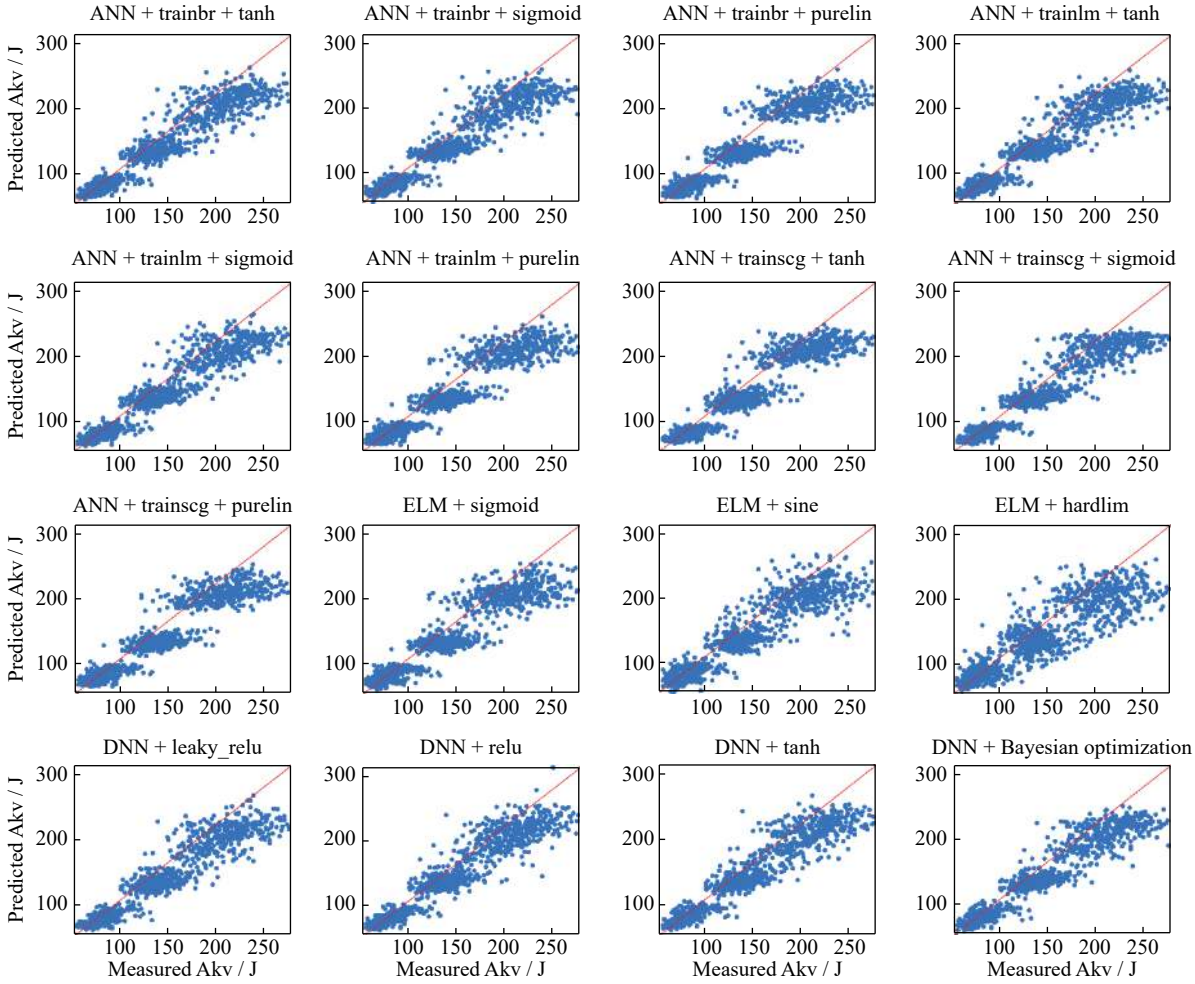

Fig. 9. Comparison of the predicted Akv with the measured Akv of all of the models.

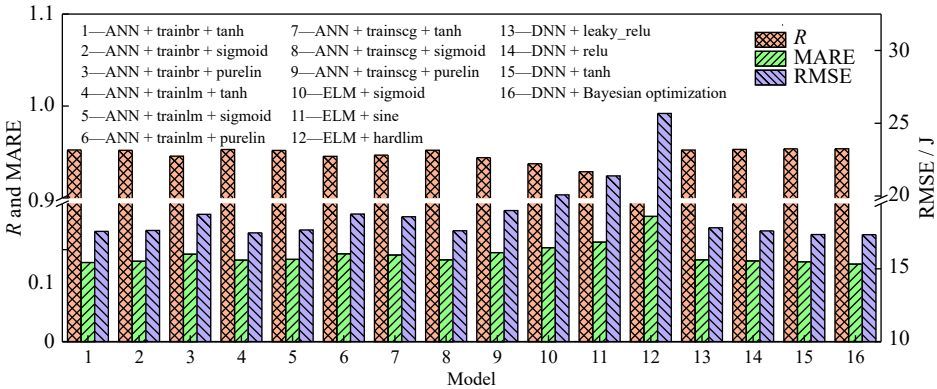

Fig. 10. *R*, MARE, and RMSE values for all of the models.

in Table 2. The hit rates of ±10, ±20, ±30, ±40, and ±50 J were calculated. The DNN models generally achieve a higher hit rate than ANN or ELM models. By using the Bayesian optimization method to determine the optimal hyper-parameter combination, a good performing DNN model was established with a hit rate of 91.20% and an absolute error of ±30 J.

The sensitivity analysis of variables [33] was performed to investigate the influence of process parameters on the impact energy prediction of low carbon steel. As the Bayesian optimization DNN model was considered to be the most successful model in predicting the impact energy of low carbon steel, the sensitivity analysis of the process parameters was investigated based on the optimal DNN model. The mean impact value (MIV) was applied to calculate the variable sensitivity. MIV is an indicator used to determine the influence of

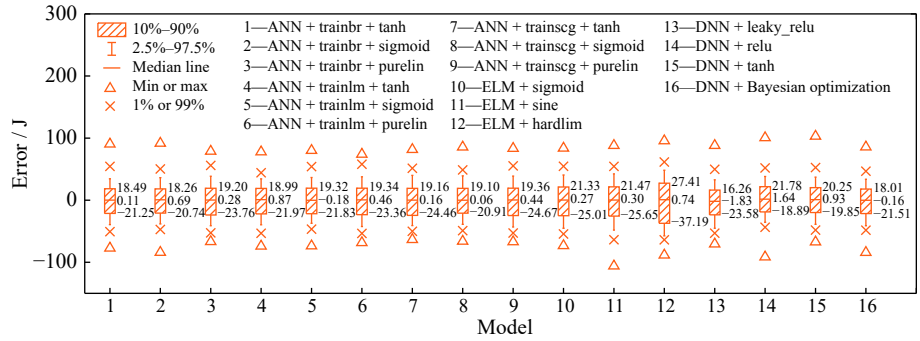

Fig. 11. Box plot of the prediction errors of all of the models with testing data.

Table 2. Hit rates of the models

Model	Hit rate / %				
	±10 J	±20 J	±30 J	±40 J	±50 J
ANN + trainbr + tanh	57.41	81.02	90.46	95.83	97.78
ANN + trainbr + sigmoid	54.07	81.20	90.19	95.74	98.33
ANN + trainbr + purelin	53.98	77.04	88.70	95.46	97.31
ANN + trainlm + tanh	56.20	79.91	91.02	95.83	98.15
ANN + trainlm + sigmoid	54.81	78.80	90.00	96.20	98.06
ANN + trainlm + purelin	52.59	76.76	89.35	94.91	97.22
ANN + trainscg + tanh	53.15	77.69	89.17	94.44	97.87
ANN + trainscg + sigmoid	55.65	80.37	90.56	95.28	98.24
ANN + trainscg + purelin	52.50	76.39	88.61	94.63	97.50
ELM + sigmoid	50.46	75.28	86.85	93.43	96.76
ELM + sine	48.89	73.33	86.30	92.78	95.83
ELM + hardlim	38.98	64.54	78.52	87.31	92.31
DNN + leaky_relu	55.28	79.35	90.28	95.09	97.87
DNN + relu	56.30	79.54	91.02	96.30	98.15
DNN + tanh	55.93	79.81	91.57	95.93	97.96
DNN + Bayesian optimization	57.78	79.07	91.20	96.02	98.52

input variables on output variables. The absolute value represents the relative importance of the influence. After network training, each independent variable in the training dataset D increased or decreased 10% on the basis of its original value to form two new training datasets D_1 and D_2. The prediction results, R_1 and R_2, were obtained using the training datasets D_1 and D_2, respectively. The mean difference between R_1 and R_2 is the importance of the independent variable. In this way, MIV was calculated using Eq. (9) [34],

$$\mathrm{MIV} = |R_1 - R_2|/N \tag{9}$$

The analysis results of the process parameter variable importance for a sample thickness of 7.5 mm are shown in Fig. 12. Under the current process, the main factors affecting the impact energy are the thickness of the original slab (RRH2,

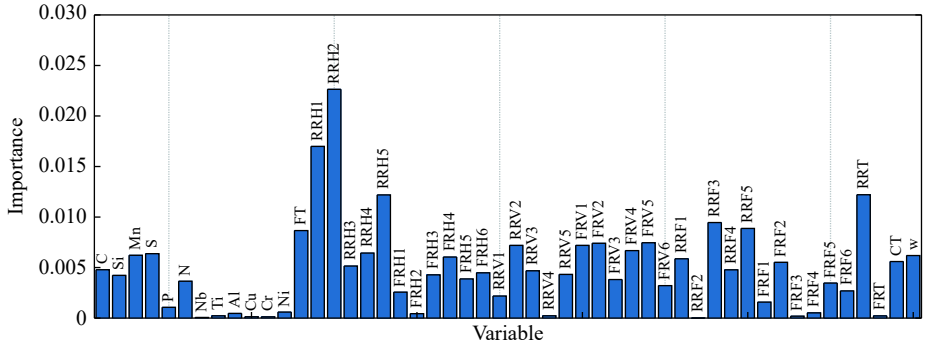

Fig. 12. Analysis results of the process parameter variable importance for a sample thickness of 7.5 mm.

RRH1), the thickness of the intermediate slab (RRH5), and the RRT. RRH2 and RRH1 reflect the influence of the original thickness of the strip on the final impact toughness. For the same product thickness, the thicker the original thickness of the strip, the lower the impact toughness. RRH5 reflects the thickness of the intermediate slab. For the same product thickness, the larger the thickness of the intermediate slab, the more serious band structure and inclusions that appear in the center of the strip, which will reduce the impact energy. The basic reason for the formation of the ferrite pearlite banded structure is the dendrite segregation of manganese, sulfur, and other elements during the solidification of steel. Secondly, the cooling rate and size of austenite grains are also important factors for the formation of the banded structure. The RRT directly affects the finish rolling entry temperature. If the finish rolling entry temperature is greater than the recrystallization termination temperature of the slab, part recrystallization will occur during finish rolling, resulting in a mixed crystal microstructure and finally decreasing the impact energy of the steel.

6. Conclusions

(1) The impact energy prediction model of low carbon steel was investigated based on the industrial data of chemical compositions, process parameters, and properties. Different activation functions, structure parameters, and training functions of the three-layer neural network, ELM, and DNN were compared. The results show that all of the models produce high accuracy predictions, excluding ELM with an activation function of "hardlim." The DNN produces better prediction results than the shallow neural network, potentially because the multiple hidden layers improve the learning ability of the model.

(2) Bayesian optimization can be applied to optimize the hyper-parameters in a DNN. It is helpful to determine the optimal hyper-parameters faster than using a trial and error method. The Bayesian optimization DNN achieves the highest R of 0.9536, lowest MARE of 0.0843, and lowest RMSE of 17.34 J for predicting the impact energy of low carbon steel. The hit rate of the model is 91.20% with an absolute error of ±30 J.

(3) The model with the best performance was applied to investigate the importance of process parameter variables on the impact energy of low carbon steel with a final thickness of 7.5 mm. Among all of the variables, the main factors affecting the impact energy are the thickness of the original slab, the thickness of the intermediate slab, and the RRT of the specific hot rolling production line.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (No. U1960202), the China Post-doctoral Science Foundation funded Projects (No. 2019M651467), and the Natural Science Foundation Joint Fund Project of Liaoning Province, China (No. 2019-KF-25-06).

References

[1] S. Guo, J.X. Yu, X.J. Liu, C.P. Wang, and Q.S. Jiang, A predicting model for properties of steel using the industrial big data based on machine learning, *Comput. Mater. Sci.*, 160(2019), p. 95.

[2] R.K. Desu, H.N. Krishnamurthy, A. Balu, A.K. Gupta, and S.K. Singh, Mechanical properties of austenitic stainless steel 304L and 316L at elevated temperatures, *J. Mater. Res. Technol.*, 5(2016), No. 1, p. 13.

[3] A.A. Lakshmi, C.S. Rao, M. Srikanth, K. Faisal, K. Fayaz, Puspalatha, and S.K. Singh, Prediction of mechanical properties of ASS 304 in superplastic region using artificial neural networks, *Mater. Today: Proc.*, 5(2018), No. 2, p. 3704.

[4] L. Kanumuri, D.V. Pushpalatha, A.S.K. Naidu, and S.K. Singh, A hybrid neural network - Genetic algorithm for prediction of mechanical properties of ASS-304 at elevated temperatures, *Mater. Today: Proc.*, 4(2017), No. 2, p. 746.

[5] A. Powar and P. Date, Modeling of microstructure and mechanical properties of heat treated components by using artificial neural network, *Mater. Sci. Eng. A*, 628(2015), p. 89.

[6] S. Lalam, P.K. Tiwari, S. Sahoo, and A.K. Dalal, Online prediction and monitoring of mechanical properties of industrial galvanised steel coils using neural networks, *Ironmaking Steelmaking*, 46(2019), No. 1, p. 89.

[7] T. Thankachan and K. Sooryaprakash, Artificial neural network-based modeling for impact energy of cast duplex stainless steel, *Arabian J. Sci. Eng.*, 43(2018), No. 3, p. 1335.

[8] R. Colas-Marquez and M. Mahfouf, Data mining and modelling of Charpy impact energy for alloy steels using fuzzy rough sets, *IFAC-Papersonline*, 50(2017), No. 1, p. 14970.

[9] M. Mahfouf and Y.Y. Yang, A GA-optimised ensemble neural network model for Charpy impact energy predictions, *IFAC Proc. Vol.*, 43(2010), No. 9, p. 62.

[10] M. Jimenez-Martinez and M. Alfaro-Ponce, Fatigue damage effect approach by artificial neural network, *Int. J. Fatigue*, 124(2019), p. 42.

[11] Y. Liu, J.C. Zhu, and Y. Cao, Modeling effects of alloying elements and heat treatment parameters on mechanical properties of hot die steel with back-propagation artificial neural network, *J. Iron Steel Res. Int.*, 24(2017), No. 12, p. 1254.

[12] Z.W. Xu, X.M. Liu, and K. Zhang, Mechanical properties prediction for hot rolled alloy steel using convolutional neural network, *IEEE Access*, 7(2019), p. 47068.

[13] J.F. Deng, J. Sun, W. Peng, Y.H. Hu, and D.H. Zhang, Application of neural networks for predicting hot-rolled strip crown, *Appl. Soft Comput.*, 78(2019), p. 119.

[14] S.W. Wu, J. Yang, R.H. Zhang, and H. Ono, Prediction of endpoint sulfur content in KR desulfurization based on the hybrid algorithm combining artificial neural network with SAPSO, *IEEE Access*, 8(2020), p. 33778.

[15] S.W. Wu, J.K. Ren, X.G. Zhou, G.M. Cao, Z.Y. Liu, and J. Yang, Comparisons of different data-driven modeling techniques for predicting tensile strength of X70 pipeline steels, *Trans. Indian Inst. Met.*, 72(2019), No. 5, p. 1277.

[16] G.B. Huang, Q.Y. Zhu, and C.K. Siew, Extreme learning machine: Theory and applications, *Neurocomputing*, 70(2006), No. 1-3, p. 489.
[17] G.B. Huang, Q.Y. Zhu, and C.K. Siew, Extreme learning machine: A new learning scheme of feedforward neural networks, [in] *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, Vol. 2, Budapest, 2004, p. 985.
[18] M.B. Li, G.B. Huang, P. Saratchandran, and N. Sundararajan, Fully complex extreme learning machine, *Neurocomputing*, 68(2005), p. 306.
[19] X.Y. Sui and Z.M. Lv, Prediction of the mechanical properties of hot rolling products by using attribute reduction ELM, *Int. J. Adv. Manuf. Technol.*, 85(2016), No. 5-8, p. 1395.
[20] X.L. Su, S. Zhang, Y.X. Yin, Y.N. Liu, and W.D. Xiao, Data-driven prediction model for adjusting burden distribution matrix of blast furnace based on improved multilayer extreme learning machine, *Soft Comput.*, 22(2018), No. 11, p. 3575.
[21] X.L. Su, S. Zhang, Y.X. Yin, and W.D. Xiao, Prediction model of hot metal temperature for blast furnace based on improved multi-layer extreme learning machine, *Int. J. Mach. Learn. Cybern.*, 10(2019), No. 10, p. 2739.
[22] S. Feng, H.Y. Zhou, and H.B. Dong, Using deep neural network with small dataset to predict material defects, *Mater. Des.*, 162(2019), p. 300.
[23] D.P. Kingma and J.L. Ba, Adam: A method for stochastic optimization, [in] *3rd International Conference for Learning Representations*, San Diego, 2015.
[24] Y. Yoo, Hyperparameter optimization of deep neural network using univariate dynamic encoding algorithm for searches, *Knowledge-Based Syst.*, 178(2019), p. 74.
[25] X.Q. Zeng and G. Luo, Progressive sampling-based Bayesian optimization for efficient and automatic machine learning model selection, *Health Inf. Sci. Syst.*, 5(2017), No. 1, art. No. 2.
[26] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, Algorithms for hyper-parameter optimization, [in] J. Shawe-Taylor, R.S. Zemel, P.L. Bartlett, F. Pereira, and K.Q. Weinberger, eds., *Proceedings of the 24th International Conference on Neural Information Processing Systems*, Granada, 2011, p. 2546.
[27] J. Wu, X.Y. Chen, H. Zhang, L.D. Xiong, H. Lei, and S.H. Deng, Hyperparameter optimization for machine learning models based on Bayesian optimization, *J. Electron. Sci. Technol.*, 17(2019), No. 1, p. 26.
[28] T.F. Awolusi, O.L. Oke, O.O. Akinkurolere, A.O. Sojobi, and O.G. Aluko, Performance comparison of neural network training algorithms in the modeling properties of steel fiber reinforced concrete, *Heliyon*, 5(2019), No. 1, art. No. e01115.
[29] J.S. Kim, D.Y. Kim, and Y.T. Kim, Experiment on radial inflow turbines and performance prediction using deep neural network for the organic Rankine cycle, *Appl. Therm. Eng.*, 149(2019), p. 633.
[30] J. Pazhoohan, H. Beiki, and M. Esfandyari, Experimental investigation and adaptive neural fuzzy inference system prediction of copper recovery from flotation tailings by acid leaching in a batch agitated tank, *Int. J. Miner. Metall. Mater.*, 26(2019), No. 5, p. 538.
[31] A.H. Elsheikh, S.W. Sharshir, M.A. Elaziz, A.E. Kabeel, G.L. Wang, and H.O. Zhang, Modeling of solar energy systems using artificial neural network: A comprehensive review, *Sol. Energy*, 180(2019), p. 622.
[32] K.Q. Zhang, H.Q. Yin, X. Jiang, X.Q. Liu, F. He, Z.H. Deng, D.F. Khan, Q.J. Zheng, and X.H. Qu, A novel approach to predict green density by high-velocity compaction based on the materials informatics method, *Int. J. Miner. Metall. Mater.*, 26(2019), No. 2, p. 194.
[33] C.C. Qi, A. Fourie, G.W. Ma, X.L. Tang, and X.H. Du, Comparative study of hybrid artificial intelligence approaches for predicting hangingwall stability, *J. Comput. Civ. Eng.*, 32(2018), No. 2, art. No. 04017086.
[34] S.W. Wu, Z.Y. Liu, X.G. Zhou, and N.A. Shi, Prediction of mechanical properties and process parameters selection based on big data, *J. Iron Steel Res.*, 28(2016), No. 12, p. 1.