IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 21, 2019, accepted March 31, 2019, date of publication April 9, 2019, date of current version April 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2909586

Mechanical Properties Prediction for Hot Rolled Alloy Steel Using Convolutional Neural Network

ZHI-WEI XU[1], XIAO-MING LIU[2], AND KAI ZHANG[2]
[1]School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China
[2]Hubei Province Key Laboratory of Intelligent Information Processing and Real-Time Industrial System, Wuhan 430065, China

Corresponding author: Kai Zhang (zhangkai@wust.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61472293, and in part by the Research Project of Hubei Provincial Department of Education under Grant 2016238.

ABSTRACT A convolutional neural network (CNN)-based method for predicting the mechanical properties of hot rolled steel using chemical composition and process parameters is proposed. The novel contribution of this research is to introduce the prediction method of CNN into the steel properties prediction field by converting the production data into two-dimensional data images. Compared with the traditional artificial neural network method, the CNN adopts the idea of local connection and weight sharing, which reduces the complexity of the network model, and uses convolution and pooling operations to extract local features for better prediction precision. The experiments in this paper show that the proposed CNN model with the optimal structure provides higher prediction accuracy and higher robustness when compared with other prediction model reported in the literature. Finally, the metallurgical phenomena in the steel rolling processes are verified by sensitivity analysis using the proposed CNN model. The results are consistent with the metallurgical properties of the steel materials used in the experiments. Therefore, the proposed CNN model has a guiding significance in predicting the mechanical properties of hot rolled steel products in practical applications.

INDEX TERMS Convolutional neural network, mechanical property prediction, deep learning, hot rolled alloy steel.

I. INTRODUCTION

Hot rolled alloy steel has been widely used in many industrial fields, including construction, bridges, highways, automotive industries, construction machinery, and so on. Since the reliability of these industrial products relies on the good mechanical properties, it is important to predict the mechanical properties of alloy steel accurately which are tensile strength (TS), yield strength (YS) and elongation (EL). Research on the mechanical properties prediction has great significance for reducing the sampling cost of alloy steel and improving product quality.

The mechanical properties of alloy steel are mainly related to chemical composition and the process parameters on the hot rolling processes [1]–[4]. However, the rolling process is a complex and dynamic nonlinearity system, it is extremely difficult to express the relationships mathematically [5], [6]. In previous studies, scholars have tried to use metallurgical mechanism models and statistical models to predict the mechanical properties [7]. Although the metallurgical mechanism model has a good theoretical basis and is suitable for most rolling processes, the model structure is very complicated and it needs to have a meticulous metallurgical understanding and cumbersome calculation or even experimental tests that are both time-consuming and expensive [8]. Due to the complexity and dynamic in steel manufacturing process, the statistical method of multivariate regression also cannot meet the requirement of accuracy [9].

In the past decades, Artificial Neural Network (ANN) and Support Vector Machine (SVM) are the two main methods to predict the mechanical properties due to the ability to solve complex nonlinear problems. Chou *et al.* [10] proposed a hybrid feedforward ANN with Taguchi particle swarm to optimize the chemical composition of steel

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano.

bars and improve the mechanical properties. Thankachan and Sooryaprakash [11] developed a multi-layer feedforward ANN model that uses the chemical composition of duplex stainless steel as input to predict the impact energy required for casting. Wang *et al.* [12] used chemical composition and hot rolling processes parameters as input parameters to predict the mechanical properties of alloy steel using SVM. The methods based on ANN and SVM can achieve high precision when the sample data volume is small, but their prediction effect needs further verification when dealing with massive data. And when ANN has many hidden layer nodes, it is easy to overfitting and difficult to converge. With the development of the steel industry, the processing technologies are more diversified, the relationships between input parameters are more complicated, the ANN and SVM prediction methods cannot deal with these difficulties. Since CNN can extract effective features from massive data and use a fewer number of parameters to construct the complex structure, it can effectively avoid overfitting and achieve much higher prediction accuracy than ANN and SVM when processing massive sample data [13].

In recent years, deep learning is used to solve prediction problems since it could learn and obtain deeper information from the raw data [14]. Among different deep neural network models, CNN can extract local features from complex data with less sensitive to noise in data, and it can achieve better effect and accuracy using fewer parameters using local connections and weight sharing [15]. Actually, CNN-based approaches have not only made tremendous progress in the image recognition [16] and the activity recognition fields [17], but also become the research hotspot in the fields of fault detection [18], financial forecasting [19], energy environmental protection [20], medical disease prediction [21], traffic management [22], and bioinformatics [23].

In this paper, a CNN-based method is proposed to predict the mechanical properties of alloy steel. Firstly, a new data preprocessing method is proposed to convert chemical composition and processing parameters into two-dimension images to extract features. Secondly, a CNN-based model for predicting mechanical properties is established, and the hyperparameters are studied to optimize network structure. Thirdly, our model is compared with the SVM [12] and Hybrid Artificial Neural Network with Particle Swarm Optimizer [10] approaches on 60,000 industrial data. The experiment results show that the proposed CNN is very promising in mechanical properties prediction for alloy steel. Finally, the relative importance of the input variable is evaluated by sensitivity analysis. This has great guiding significance for optimizing the chemical composition and the production processes of steel and developing new steel grades.

The remaining sections complete the presentation of this paper. Section II introduces the relationships between mechanical properties with chemical composition and the production processes. Section III presents the proposed CNN-based prediction method. In the Section IV, the hyperparameters of CNN structures are optimized through a series of experiments. The experiment results are given on more than 60,000 industrial data. Finally, a conclusion is drawn in Section V along with pertinent observations identified.

II. PROBLEM DESCRIPTION

A. CHEMICAL COMPOSITION OF HOT ROLLED ALLOY STEEL

Most of the chemical composition of alloy steel consist of iron (Fe), carbon (C), manganese (Mn), silicon (Si), phosphorus (P) and sulfur (S). After reheating, roughing rolling, finishing rolling, laminar cooling and down coiler, a steel slab becomes a coil of a thin sheet. The mechanical properties can be adjusted with different chemical composition or process parameters.

Complex interactions exist between the chemical composition and mechanical properties. For example, when the ratio of C in the steel is below 0.8Wt%, the YS and TS of the steel increase dramatically with the increases of C content, but the EL of steel decreases [24]. Furthermore, the content of S and P must be strictly controlled. S will reduce the hot workability and strength of steel and P will reduce the plasticity and toughness of the steel [25]. Mn has a strong ability of deoxidation and desulfurization, can greatly improve the hot workability and strength of steel. In steel production, molten iron is often mixed with oxygen (O_2) and nitrogen (N_2), which is detrimental to the mechanical properties of steel. For this reason, Mn, Si, aluminum (Al), vanadium (V), titanium (Ti) are often added as deoxidants to the steel [10]. The combination of Si, molybdenum (Mo), and chromium (Cr) not only makes the steel highly resistant to oxidation and corrosion but also enhances the strength and hardness of the steel. The addition of Al, niobium (Nb), and V to the steel can reduce the negative effects of N. The copper (Cu) in steel can improve the strength and toughness and resistance to atmospheric corrosion of steel. Boron(B) can improve the compactness and hot rolling properties of steel. Nickel (Ni) not only can significantly increase the strength and toughness of steel, Ni and Cr are also the main alloying elements of stainless steel [26].

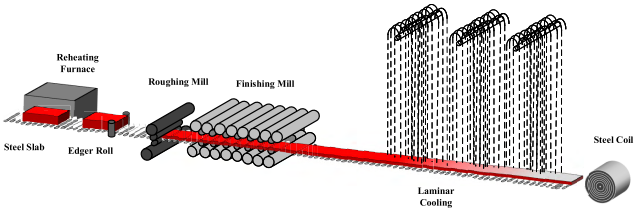

FIGURE 1. Hot rolled processing of alloy steel.

B. HOT ROLLED PROCESSING OF ALLOY STEEL

The hot rolled processing of steel slab can be divided into five steps, reheating, roughing rolling, finishing rolling, laminar cooling, and down coiling, as shown in Fig. 1. Firstly,

the steel slab is reheated in furnace to a high temperature about 1100° to 1250°. Secondly, the hot steel slab is transfer to edger mill and roughing mill to reduce the width and thickness. The steel slab would become longer and thinner. Thirdly, the steel slab will pass through the finishing mill to control the thickness with high precision. Fourthly, the steel strip will pass through the laminar cooling region to reduce the temperature quickly. Finally, the steel strip is coiled by the down coiler.

In fact, a series of complex microstructure changes occur in the manufacturing processes, which could determine the mechanical properties of the alloy steel. First, the reheating process provides a uniform temperature to the slab to provide a uniform initial austenite grain size Then, the roughing and finishing processes refine austenite by dynamic and static recrystallization. Furthermore, the steel sheet is continuously cooled by the laminar cooling system for refining transformed ferrite and pearlite grain [27]. The size and volume fraction of these grains determine the mechanical properties of the steel. These heat treatment temperatures play an important role in mechanical properties prediction. Therefore, the furnace temperature (FT), the roughing rolling temperature (RRT), the finishing rolling temperature (FRT) and the coiling temperature (CT) have a significant influence on the mechanical properties of hot rolled steel.

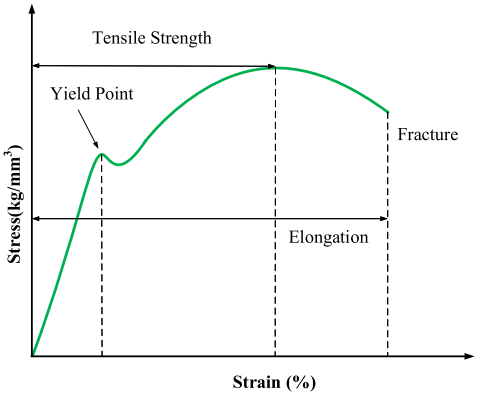

FIGURE 2. Relationship between tensile strength, yield strength and elongation of alloy steel.

C. MECHANICAL PROPERTIES OF HOT ROLLED ALLOY STEEL

The mechanical properties of alloy steel are TS, YS, and EL. TS is defined as the maximum tensile stress that the steel can withstand before breaking, and the YS is defined as the maximum stress of the steel can withstand before plastic deformation begins. EL is defined as the percentage of stretched length to the original length after the steel is broken. Fig. 2 shows the relationship between TS, YS, and EL of alloy steel.

III. PROPOSED APPROACH

This section describes the proposed prediction method based on CNN for mechanical properties of hot rolled alloy steel. First, the data-image converting method is introduced. Then, the CNN model for predicting the mechanical properties of the alloy steel is presented.

A. CONVERTING CHEMICAL COMPOSITION AND PROCESS PARAMETERS INTO IMAGES

The raw data of alloy steel consist of sixteen kinds of chemical compositions and four kinds of heat treatment process parameters. Our method converts the raw data into a two-dimensional matrix, and integrates the CNN model features learning.

Firstly, the raw data is normalized to eliminate the largely distinct scales in different fields. Let the raw data be $X = (x_1, x_2, \ldots, x_i, \ldots, x_{20})$, the normalization function is shown as Equation (1).

$$x_i' = \frac{x_i - Min(x_i)}{Max\,(x_i) - Min(x_i)} \tag{1}$$

where x_i' represents the normalized value of the input parameter x_i, and $Max\,(x_i)$ and $Min(x_i)$ are the maximum and minimum values of input parameter x_i, respectively.

Then, the normalized data X is treated as a column vector and multiplied by its own transposition X^T to obtain a production data matrix. In statistics, the data matrix is the information matrix of the raw data, which contains all the information of the raw data and can reflect the relative size of the data variance and covariance. Constructing a two-dimensional production data information matrix can maintain the data features and spatial correlations while meeting the input requirements of the CNN. Let S denote the production data matrix, which can be expressed as:

$$S = XX^T = \begin{bmatrix} x_1^2 & x_1x_2 & x_1x_3 & \ldots & x_1x_{20} \\ x_2x_1 & x_2^2 & x_2x_3 & \ldots & x_2x_{20} \\ x_3x_1 & x_3x_2 & x_3^2 & \ldots & x_3x_{20} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{19}x_1 & x_{19}x_2 & x_{19}x_3 & \cdots & x_{19}x_{20} \\ x_{20}x_1 & x_{20}x_2 & x_{20}x_3 & \cdots & x_{20}^2 \end{bmatrix} \tag{2}$$

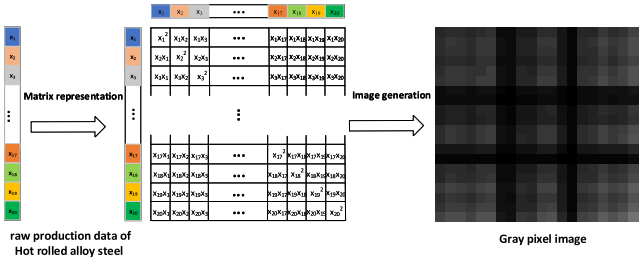

FIGURE 3. Data-image conversion method.

Finally, the raw data is converted to a two-dimension data matrix, as a gray pixel image. Fig. 3 illustrates the relations among the raw data of alloy steel, the production data matrix, and the gray pixel image.

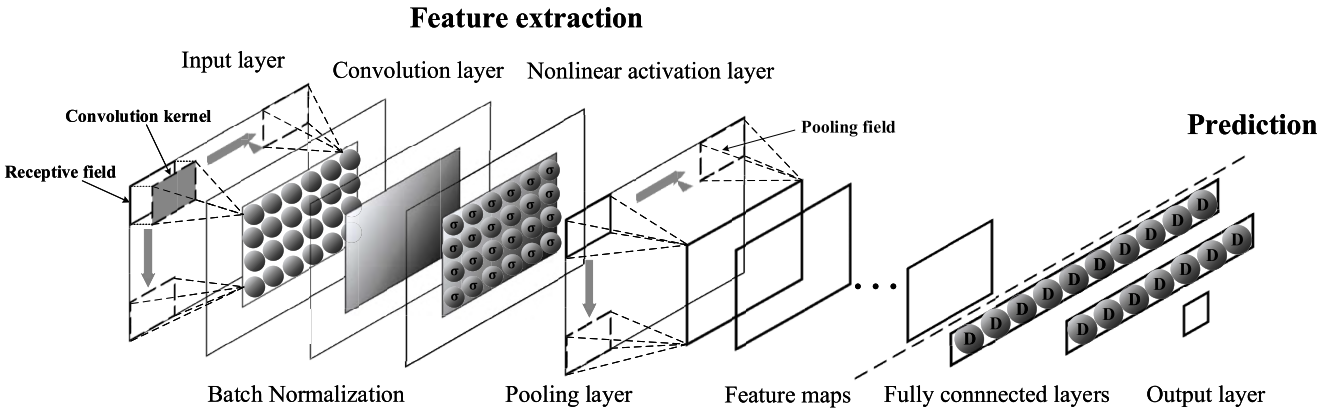

FIGURE 4. Structure of CNN-based prediction model for mechanical properties.

B. PROPOSED CNN METHOD FOR STEEL MECHANICAL PROPERTIES PREDICTION

The proposed CNN-based prediction model is composed of two parts, feature extraction part and prediction part. As shown in Fig. 4, the feature extraction part is composed of an input layer and several feature extractors. The feature extractor is stacked by convolutional layer, batch normalization (BN) operation, nonlinear activation layer, and pooling layer. Each feature extractor will extract its input features to get a feature map which will become the input data of the subsequent feature extractor. The prediction part contains two fully connected (FC) layers and an output layer. The feature maps output by the last feature extractor will be transferred to the FC layers and perform the prediction task. Finally, the predicted steel mechanical properties value will be output by the output layer.

The input layer receives the converted two-dimension data image, and the convolutional layer uses the convolution kernel which consists of a weight matrix with the same size as the receptive field on the input layer and a bias value to establish a local connection. As the receptive field slides from the top left of the input layer to the bottom right, the convolutional layer obtains the feature map of the input layer filtered by the convolution kernel, which can be expressed as Equation (3).

$$y_{jk} = (\sum_{p=1}^{F}\sum_{q=1}^{F} w_{pq} x_{(p+j*s)(q+k*s)} + b),$$
$$0 \le j \le \frac{H-F}{S}, \quad 0 \le k \le \frac{W-F}{S} \tag{3}$$

y_{jk} represents the value of the node positioned at (j, k) on the feature map, F represents the height and width of the receptive field, H and W represent the height and width of the input data, S indicates the stride of the receptive field. $x_{(p+j*s)(q+k*s)}$ represents the input data with the coordinate at $(p + j * s, q + k * s)$, w_{pq} and b denote the weight located at (p, q) on the weight matrix and the bias, respectively.

To obtain sufficient characteristics of the mechanical properties, a set of convolutional kernels are used to perform the convolution operation. By the convolution operation, each convolution kernel can get a feature map, and different nodes on the feature map correspond to different receptive fields when the convolution kernel sweeps across the input layer. The connection pattern that each node on the same feature map connects to its receptive field by the same convolution kernel is called parameter sharing. The local connection and the parameter sharing are two important characteristics of CNN, which can reduce the number of parameters, extract the features in raw data effectively, and enhance the generalization ability of the model [28]. For the lth convolutional layer with P_l convolutional kernels, the output can be denoted as Equation (4).

$$u_l^i = \sum_{k=1}^{P_{l-1}} \left(W_l^i x_l^k + b_l^i \right), \quad \mathrm{i} \in [1, P_l] \tag{4}$$

x_l^k and u_l^i represent the input matrix and output matrix of the lth layer respectively, where i and k denote the channel index in the lth convolutional layer and the l-1th convolutional layer, respectively. The weight matrix and bias contained in the kth convolution kernel of the lth convolutional layer are denoted by W_l^i and b_l^i respectively. The size of the feature map will shrink after convolution operation. The zero-padding method is applied to keep the size of the output feature map, which centers the output feature map and adds zero values at all the edges of the output feature map.

BN operation is added after each convolutional layer to improve training speed and achieve higher prediction accuracy, which allows us to be less careful about the initialization method and use much higher learning rates [29]. First, each dimension of the input is normalized into a stable distribution with the mean of zero and variance of one. Then, the normalized value is scaled and shifted by a pair of learnable variables γ and β to restore the data distribution that should be learned in the previous layer. The BN operation can be expressed as Equation (5).

$$y^{(k)} = \gamma^{(k)} \frac{x^{(k)} - E\left(x^{(k)}\right)}{\sqrt{\mathrm{Var}\left(x^{(k)}\right) + \epsilon}} + \beta^{(k)}, \quad k \in [1, N] \tag{5}$$

where N represents the total dimension of the input, $x^{(k)}$ and $y^{(k)}$ represent the kth dimension of the input and output of the layer respectively, $E(x^{(k)})$ and $Var\left(x^{(k)}\right)$ represent the mean value and variance of the kth dimension of the input respectively. A very small real number ϵ is added, to avoid the denominator is zero. The learnable variables for scaling and shifting the normalized value are represented by $\gamma^{(k)}$ and $\beta^{(k)}$ respectively.

Rectified linear unit (ReLU) is applied to the proposed CNN model as the nonlinear activation function, which can prevent the vanishing gradient and exploding gradient problems in the neural network and enhance the training speed [30]. Let max denote the function to select the larger value between x and zero, the ReLU activation function can be indicated as Equation (6).

$$\mathrm{ReLU}(x) = \max(x, 0) \tag{6}$$

In the pooling layer, downsampling is applied to scale down and concentrate feature map to obtain the most significant features in the input feature map. The max-pooling method is used as the pooling method by selecting the maximum value in the pooling field.

The information of the concentrated feature maps obtained from the final pooling layer is transmitted to the prediction part by flattening the condensed feature maps into a dense vector. Each node on the first FC layer is connected to the dense vector and the output is passed to the second FC layer and finally transmitted to the output layer. The dropout [31] which can effectively avoid the proposed model from overfitting by randomly selecting some nodes to skip weight updates in each iteration of training is used at all the FC layers to enhance the generalization ability of the model.

The predicted value is obtained by using feature extractions to gain the features from raw data and applying FC layers to process the feature information. The model output can be expressed as Equation (7).

$$\hat{y} = w_{fc2}\sigma(w_{fc1}fl(pool(\sigma(bn(\sum_{k=1}^{P_{L-1}}\left(W_L^i x_L^k + b_L^i\right)))) + b_{fc1}) + b_{fc2} \tag{7}$$

where the weight and the bias of the first FC layer are denoted by w_{fc1} and b_{fc1} respectively, w_{fc2} and b_{fc2} are the weight and the bias of the second FC layer, σ represents the nonlinear activation function ReLU, fl denotes the flattening operation which flattens features into a dense vector, $pool$ denotes the max-pooling method, bn denotes the BN operation, L represents the total number of convolutional layers, P_{L-1} denotes the number of convolutional kernels in the L-1th convolutional layer, the channel index of the Lth and L-1th convolutional layer are denoted by i and k respectively and x_L^k represents the input of the Lth layer.

The mean squared error (MSE) is applied as the loss function to measure the distance between predicted values and actual values. Minimizing MSE is taken as the training goal of the proposed model and the minimum MSE is achieved by continuously adjusting the weight and bias of each neuron. The MSE can be expressed as Equation (8).

$$\mathrm{MSE} = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{y}_i - y_i\right)^2 \tag{8}$$

where $\hat{y}_i$ represents the predicted value, y_i represents the actual value, and M represents the sample size in the data sets.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The structure of the CNN is crucial to the performance of the prediction model. Small size of filters cannot only achieve the same effect as the large size but also reduce the number of parameters, thereby improving the generalization performance of the model. The proposed prediction model uses 3∗3 convolutional kernels and 2∗2 pooling fields. The stride size of the convolutional layer and the pooling layer are set to 2 in this model, convolutional kernels of the proposed CNN-based model increase gradually as convolutional layers increases.

Deep neural network model with too much layers or too much nodes in hidden layer are not only difficult to converge but also lead to redundancy in the learned features and overfit the training set [32]. The hyperparameters setting of CNN is also important to model performance. Glorot and Bengio [33] is adopted as the weight initialization method, the batch size is set to 128, and parameters in the CNN model are updated by Adam optimizer. Referring to the paper of Adam optimization algorithm, 0.001 is used as the initial learning rate of the proposed CNN model. Since the BN operation and the dropout operation both have regularization effects, the BN operation is only used in the convolutional layers, the dropout operation is only used in the FC layers and the dropout rate is set to 0.4 in the proposed model, which is based on the solution to deal with the disharmony between dropout and batch normalization proposed by Li *et al.* [34].

A. DATA DESCRIPTION

In this study, 60,000 experimental steels are collected from two hot roll factories in Wuhan and are reprocessed into cylindrical samples of Φ 8mm × 15mm in the laboratory. The sample conforms to the GB/T700-2006 standard which is roughly equivalent to ASTM A36, but worse than it, as shown in Table 1:

Fig. 5(a)-(b) show the microstructure of the specimens made of the standard experimental slab with the same FT and different FRT under the optical microscope using standard metallographic techniques. The Gleeble 1500 CNC dynamic thermal-mechanical physical simulation machine is used to simulate the hot rolling experiment, and Axiplan 2 Imaging Zeiss optical microscope is used to observe the microstructure of the sample. Fig. 5(c)-(d) show the tensile fracture morphology of two samples of different YS after tensile testing. The CMT5105 microcomputer controlled electronic tester is used for tensile testing and the fracture morphology is observed

TABLE 1. Chinese national standard chemical composition and mechanical properties of experimental steel.

National standard	Chemical composition					Mechanical properties		
	C (Wt%)	Mn (Wt%)	Si (Wt%)	P (Wt%)	S (Wt%)	TS (MPa)	YS (MPa)	EL (%)
GB/T700-2006	⩽0.20	≤1.40	⩽0.35	⩽0.045	⩽0.045	375-500	≥235	≥26

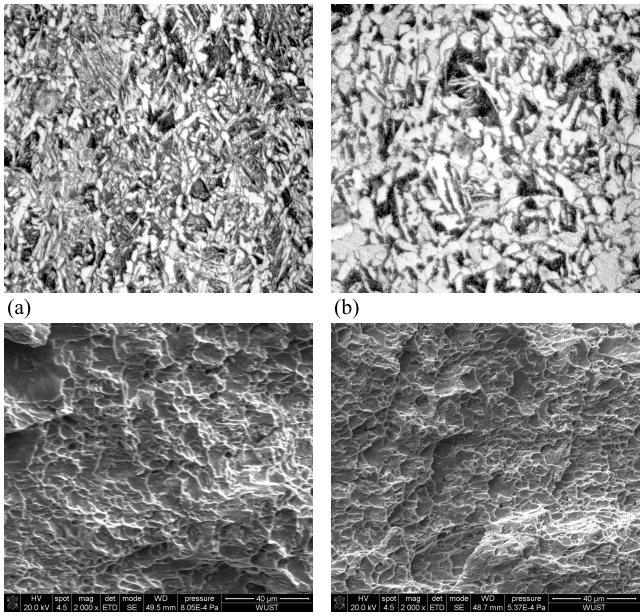

FIGURE 5. Metallographic microstructure of samples with the FT at 1150° and different FRT; (a) 830°, (b) 860°, tensile fracture morphology of samples with different YS, (c) 325 MPa, and (d) 345 MPa.

TABLE 2. Descriptive statistics of input parameters.

Parameter	Unit	Mean	Standard deviation	Minimum	Maximum
C	Wt%	0.1491	0.0085	0.0051	0.1936
Mn	Wt%	0.4682	0.0386	0.2362	1.3696
Si	Wt%	0.1941	0.0201	0.0063	0.2994
P	Wt%	0.0173	0.0038	0.0062	0.0364
S	Wt%	0.0130	0.0038	0.0012	0.0250
Cu	Wt%	0.0405	0.0093	0.0124	0.0826
Al	Wt%	0.0069	0.0150	0.0003	0.5654
Als	Wt%	0.0062	0.0148	0.0002	0.5652
Ni	Wt%	0.0168	0.0066	0.0012	0.0673
Cr	Wt%	0.0298	0.0108	0.0082	0.0942
Ti	Wt%	0.0015	0.0011	0.0001	0.0691
Mo	Wt%	0.0053	0.0023	0.0012	0.0184
V	Wt%	0.0015	0.0006	0.0012	0.0056
Nb	Wt%	0.0012	0.0004	0.0010	0.0194
N	Wt%	0.0030	0.0013	0.0008	0.0653
B	Wt%	0.0002	0.0001	0.0004	0.0031
FT	°C	1242.6082	19.9593	1188	1291
RRT	°C	1050.3915	14.0244	976	1142
FRT	°C	878.8846	27.3961	791	1056
CT	°C	647.8934	27.1412	200	680
TS	Mpa	439.9289	15.4781	321	568
YS	Mpa	295.2638	20.6944	215	477
EL	%	32.1944	3.2702	14	55

using a Nova 400 NanoSEM field emission scanning electron microscope.

In the experiment, 42000 sets of data are randomly selected as the training set for training the proposed model, 9000 sets of data are randomly selected as the validation set to determine the network structure and the hyperparameters of the proposed CNN, the remaining 9,000 sets of data are as the test set to test the performance of the optimal model. And to emphasize, the ten-fold cross validation method is used on the training and validation sets to ensure the robustness of the results when compared with other methods. Table 2 displays the data distribution statistics for the minimum, maximum, average, and standard deviation of all input parameters in the overall datasets.

The input parameters comprise sixteen kinds of chemical composition and four kinds of heat treatment parameters. The output parameters comprise three important mechanical properties which are TS, YS, and EL.

B. RESULTS OF PROPOSED CNN MODEL

In order to obtain optimal hyperparameters, different number of depths, width and convolutional kernels are attempted on network structure. Four indicators are adopted as the evaluation metrics to assess the prediction capability comprehensively, such as mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and coefficient of determination (R^2) [27].

The structures of the model with different depths are illustrated in Table 3, where LN denotes the Nth convolutional layer, the number indicates the number of convolution filters. TS, YS and EL represent the performance of the model in predicting TS, YS and EL respectively. The only one FC layer is set to highlight the influence of the convolutional layer structure on the proposed model, and the number of nodes is 256. The experiments are performed five times and the results are averaged. As can be seen from the results that the Depth-2 achieves the best result. The MSE, MAE, and MAPE of Depth-2 are all lower than those of other depths and the R^2 of Depth-2 is the highest. This may be because the Depth-1 model is too small to fully represent the relationship between the input parameters and the mechanical properties of steel but the Depth-3 model is too large, which reduces the generalization ability of the model.

TABLE 3. Prediction performances of the CNN with different depths.

Layer	Predicting Item	Evaluation Metric	Depth-1	Depth-2	Depth-3
L1			64	64	64
L2			128	64	64
L3			128	128	64
L4			256	128	128
L5			-	256	128
L6			-	256	128
L7			-	-	256
L8			-	-	256
	TS	MSE	0.0039	**0.0020**	0.0033
		MAE	0.0465	**0.0246**	0.0451
		MAPE	0.0924	**0.0612**	0.0825
		R^2	0.7423	**0.8746**	0.8254
	YS	MSE	0.0021	**0.0011**	0.0018
		MAE	0.0332	**0.0156**	0.0314
		MAPE	0.0984	**0.0694**	0.0912
		R^2	0.9124	**0.9434**	0.9267
	EL	MSE	0.0029	**0.0018**	0.0033
		MAE	0.0374	**0.0252**	0.0352
		MAPE	0.1042	**0.0612**	0.1174
		R^2	0.8741	**0.9142**	0.8578

Table 4 shows the results of models based on Depth-2 but with different numbers of convolutional kernels and the values are the average of the five experiments. The results show that Width-4, Width-3, and Width-2 achieve the best results for TS, YS, and EL respectively. The model scales of Width-4 and Width-3 are similar but the scale of Width-2 is smaller, which indicates that the impact of the input parameters on EL is smaller than TS and YS, and using a deeper and larger convolutional model to predict EL is more likely to produce overfitting.

Different network structures perform differently when predicting different mechanical properties. Three different structures are used to train the network model separately to improve accuracy when predicting different mechanical properties. As shown in Table 5, TS, YS, and EL denote the different optimal structures of the proposed CNN method for predicting the TS, YS, and EL, respectively. Conv(3∗3∗64) denotes a convolutional layer that filter size is 3∗3 and has 64 channels. Maxpool(2∗2) denotes a max-pooling layer with 2∗2 pooling field. FC1 represents the first FC layer, FC2 represents the second FC layer, and the followed numbers indicate the number of neuron nodes in this FC layer.

TABLE 5. Optimal structures of CNN.

Layer	TS	YS	EL
L1	Conv(3*3*64)	Conv(3*3*64)	Conv(3*3*32)
	Maxpool(2*2)	Maxpool(2*2)	Maxpool(2*2)
L2	Conv(3*3*64)	Conv(3*3*64)	Conv(3*3*32)
	Maxpool(2*2)	Maxpool(2*2)	Maxpool(2*2)
L3	Conv(3*3*128)	Conv(3*3*64)	Conv(3*3*64)
	Maxpool(2*2)	Maxpool(2*2)	Maxpool(2*2)
L4	Conv(3*3*128)	Conv(3*3*128)	Conv(3*3*64)
	Maxpool(2*2)	Maxpool(2*2)	Maxpool(2*2)
L5	Conv(3*3*256)	Conv(3*3*128)	Conv(3*3*128)
	Maxpool(2*2)	Maxpool(2*2)	Maxpool(2*2)
L6	Conv(3*3*256)	Conv(3*3*128)	Conv(3*3*128)
	Maxpool(2*2)	Maxpool(2*2)	Maxpool(2*2)
FC1	**1280**	**768**	**512**
FC2	**768**	**256**	**256**

Fig. 6(a)-(c) shows that the predicted values of the proposed CNN are in good agreement with the actual values of TS, YS, and EL on the test set. Fig. 6(d)-(f) shows the statistical distributions of the percentage errors between the predicted values of the proposed CNN model and the actual values on the test datasets. The percentage errors not only proximately accord with the normal distribution but also are distributed in the range of -10% to 10%, which proves that the proposed CNN model has a good prediction performance

The experiment results of the proposed CNN-based model are compared with SVM [12] and ANN [10] models by

TABLE 4. Prediction performances of the CNN with different widths.

Layer	Predicting Item	Evaluation Metric	Width-1	Width-2	Width-3	Width-4	Width-5	Width-6
L1			32	32	64	64	128	128
L2			32	32	64	64	128	128
L3			32	64	64	128	128	256
L4			64	64	128	128	256	256
L5			64	128	128	256	256	512
L6			64	128	128	256	256	512
	TS	MSE	0.0028	0.0025	0.0022	**0.0020**	0.0028	0.0035
		MAE	0.0294	0.0263	0.0255	**0.0246**	0.0297	0.0449
		MAPE	0.0736	0.0703	0.0644	**0.0612**	0.0759	0.0905
		R^2	0.8175	0.8349	0.8448	**0.8746**	0.8043	0.7841
	YS	MSE	0.0018	0.0017	**0.0011**	**0.0011**	0.0016	0.0018
		MAE	0.0329	0.0284	**0.0152**	0.0156	0.0203	0.0316
		MAPE	0.0972	0.0851	**0.0578**	0.0694	0.0822	0.0927
		R^2	0.9156	0.9207	**0.9562**	0.9434	0.9217	0.9186
	EL	MSE	**0.0008**	**0.0008**	0.0014	0.0018	0.0025	0.0029
		MAE	0.0160	**0.0156**	0.0215	0.0252	0.0317	0.0338
		MAPE	0.0481	**0.0448**	0.0583	0.0612	0.0824	0.0941
		R^2	0.9619	**0.9683**	0.9385	0.9142	0.9074	0.8914

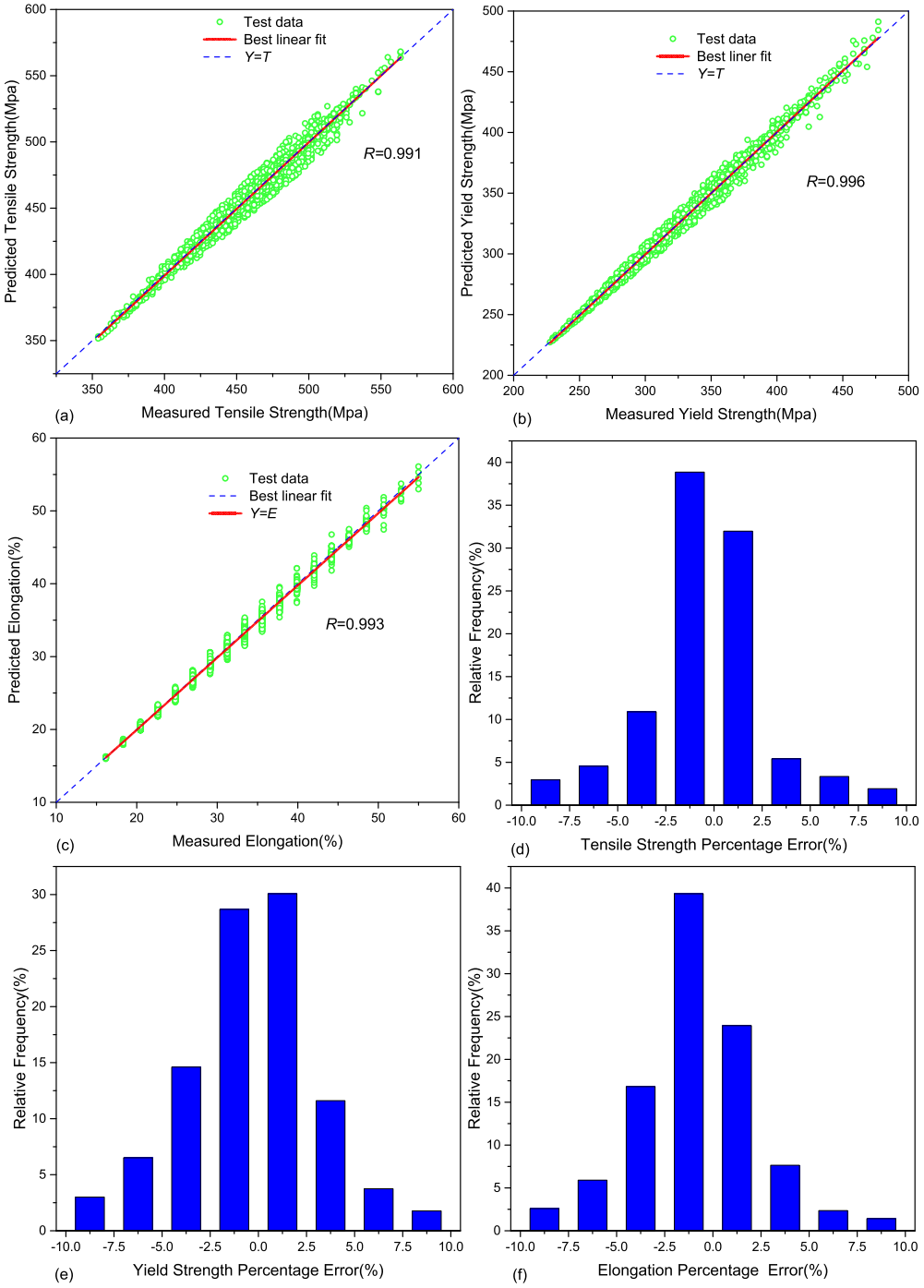

FIGURE 6. **Comparisons of predicted and actual (a) tensile strength, (b) yield strength, (c) elongation and distributions on percentage errors of (d) tensile strength, (e) yield strength, and (f) elongation.**

ten-fold cross validation method on the training and validation sets and the average are recorded. The comparison results are shown in Table 6. The SVM uses a radial basis function as the kernel function, and the ANN uses a single hidden layer structure and uses particle swarm optimization to optimize it.

It can be seen from the results that the proposed CNN method obtains competitive results against with SVM and ANN. For the prediction of TS, although both ANN and CNN have achieved an MSE of 0.004, the CNN method is better for other indicators, with MAE, MAPE, and R2 of the proposed CNN are 0.0113, 0.0218, 0.9835 respectively. In the case of the prediction of YS, the MSE of CNN and ANN are both 0.0003, but the MAE and MAPE of the proposed CNN which are 0.0106 and 0.0229 respectively are

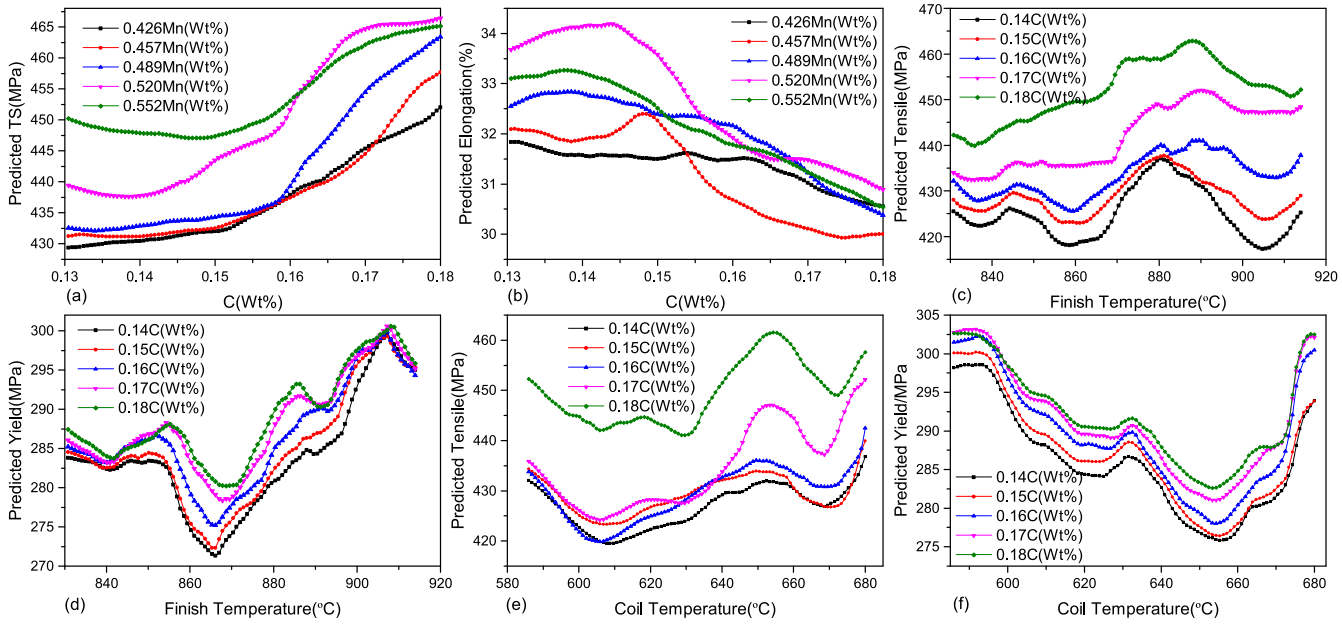

FIGURE 7. Combined effects of (a) C and Mn content on tensile strength, (b) C and Mn content on Elongation, (c) C content and finishing temperature on tensile strength, (d) C content and finishing temperature on yield strength, (e) C content and coiling temperature on tensile strength, and (f) C content and coiling temperature on yield strength.

TABLE 6. Prediction performances of the CNN and other algorithms.

Predicting Item	Evaluation Metric	CNN	SVM	ANN
TS	MSE	**0.0004**	0.0005	**0.0004**
	MAE	**0.0113**	0.0141	0.0127
	MAPE	**0.0218**	0.0252	0.0223
	R^2	**0.9835**	0.9678	0.9814
YS	MSE	**0.0003**	0.0004	**0.0003**
	MAE	**0.0106**	0.0132	0.0114
	MAPE	**0.0229**	0.0325	0.0267
	R^2	**0.9904**	0.9786	0.9883
EL	MSE	**0.0004**	0.0005	**0.0004**
	MAE	**0.0116**	0.0147	0.0131
	MAPE	**0.0243**	0.0332	0.0287
	R^2	**0.9865**	0.9762	0.9819

both lower than those of SVM and ANN, and the proposed CNN achieves the best R2 with 0.9904. For the prediction of EL, although the CNN and ANN obtain the same MSE with 0.0004, the MAE and MAPE of the proposed CNN are both lower than the other two methods, which are 0.0116 and 0.0243 respectively, and the highest R2 of 0.9865 is obtained by the CNN method. Therefore, the proposed CNN method has an excellent ability to predict the mechanical properties of steel and outperforms the other methods.

C. SENSITIVITY ANALYSIS

The synergistic effects of the input parameters are studied by changing two input parameters simultaneously and holding the remaining parameters at their average values. C has a significant influence on the mechanical properties of steel, which can enhance the strength of hot rolled alloy steel by means of interstitial solid solution strengthening. Mn enhances the strength of steel by stabilizing austenite and solid solution strengthening, and its strengthening effect is better as the carbon content increases, but the EL of steel also decreases as the strength increases. Fig. 7(a)-(b) shows the combined effects of C and Mn content on the TS and EL of steel.

FRT and CT are two important processing technology parameters. Fig. 7(c) shows the combined effect of C content and FRT on TS, that the TS of the steel increases first and then decreases with the FRT increases and the effect of the change of C content on the TS is obvious. The combined effect of C content and FRT on YS is shown in Fig. 7(d), which illustrates the YS of the steel decreases first and then increases with the FRT increases, the FRT has a greater influence on the YS than the C content, and the YS reaches a maximum of 300.592 MPa when the FRT is 908.120°C and the carbon content is 0.18Wt%. Fig. 7(e) demonstrates the effect of C content and CT on TS, as can be seen from the graph, the TS of the steel first decreases and then increases with the CT increases, and when the carbon content is 0.18Wt%, the TS of the steel increases greatly. Fig. 7(f) shows the effect of C content and CT on YS, that the YS of the steel first decreases and then increases with the CT increases, and the CT has greater influence on the YS than the C content, and when the CT is 591.650°C and the carbon content is 0.17Wt%, the maximum YS is achieved with 303.193 MPa. In summary, only considering the parameters of C content, FRT, and CT, the TS is affected more by C content, and the YS is more affected by the FRT and the CT.

V. CONCLUSION

In this paper, a CNN based model for predicting the mechanical properties of alloy steel by using sixteen kinds of chemical composition and four kinds of hot rolling production processes of hot rolled steel is proposed. The raw data is processed by normalization to keep the data in the same range, then is converted into an image that the CNN can process using the proposed data-image conversion method. Small convolutional kernels and small pooling fields are used in the proposed CNN model to reduce the number of parameters and improve the generalization ability of the model. The BN operation is used to increase the convergence speed of the model, and the dropout operation is used to enhance the generalization capability of the model and avoid overfitting on the training set. The best CNN structures of predicting TS, YS, and EL are determined by comparison experiments, and the model prediction results are compared with the SVM model and the ANN model. The results show that the prediction accuracy of the proposed CNN model for predicting hot rolled steel mechanical properties is greatly improved. Through correlation analysis, it is verified that the increase of C and Mn elements can increase the TS and YS of steel to a certain extent. For the parameters of carbon content and processing temperatures, it is verified that the TS is more affected by the change of carbon content and the YS is more affected by the FRT and the CT.

The future research works can focus on the following aspects. First, the standard data set of steel performance prediction can be established to accelerate research progress in steel field. Second, the metallographic microstructure of steel can be further studied to predict steel type and mechanical properties. Finally, the model can be improved for online steel mechanical properties prediction.

REFERENCES

[1] T. Thankachan, K. S. Prakash, and M. Kamarthin, "Optimizing the tribological behavior of hybrid copper surface composites using statistical and machine learning techniques," *J. Tribol.*, vol. 140, no. 3, Jan. 2018, Art. no. 031610.

[2] M. Beghini, L. Bertini, B. D. Monelli, C. Santus, and M. Bandini, "Experimental parameter sensitivity analysis of residual stresses induced by deep rolling on 7075-T6 aluminium alloy," *Surf. Coat. Technol.*, vol. 254, pp. 175–186, Sep. 2014.

[3] C.-K. Cheng, J.-T. Tsai, T.-T. Lee, J.-H. Chou, and K.-S. Hwang, "Modeling and optimizing tensile strength and yield point on steel bar by artificial neural network with evolutionary algorithm," in *Proc. IEEE Int. Conf. Automat. Sci. Eng. (CASE)*, Gothenburg, Sweden, Aug. 2015, pp. 1562–1563.

[4] Q. Zou, L. Chen, N. Xiong, S. Zou, and C. Wang, "Prediction and key computer programming of mechanical properties of hot rolled plate based on BP neural network," in *Proc. Int. Conf. Comput. Sci. Eng.*, Vancouver, BC, Canada, Aug. 2009, pp. 967–971.

[5] H.-T. He and H.-M. Liu, "The research on integrated neural networks in rolling load prediction system for temper mill," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Guangzhou, China, vol. 7, Aug. 2005, pp. 4089–4093.

[6] Z.-H. Wang, D.-Y. Gong, X. Li, G.-T. Li, and D.-H. Zhang, "Prediction of bending force in the hot strip rolling process using artificial neural network and genetic algorithm (ANN-GA)," *Int. J. Adv. Manuf. Technol.*, vol. 93, nos. 9–12, pp. 3325–3338, Dec. 2017.

[7] Z.-H. Guo, Q.-L. Zhang, Y.-C. Su, and Y. Xia, "Thoughts on mechanical property prediction of hot rolled strip," *Metall. Ind. Automat.*, vol. 33, no. 2, p. 1, 2009.

[8] R.-C. Hwang, Y.-J. Chen, and H.-C. Huang, "Artificial intelligent analyzer for mechanical properties of rolled steel bar by using neural networks," *Expert Syst. Appl.*, vol. 37, no. 4, pp. 3136–3139, Apr. 2010.

[9] J. Ghaisari, H. Jannesari, and M. Vatani, "Artificial neural network predictors for mechanical properties of cold rolling products," *Adv. Eng. Softw.*, vol. 45, no. 1, pp. 91–99, Mar. 2012.

[10] P.-Y. Chou, J.-T. Tsai, and J.-H. Chou, "Modeling and optimizing tensile strength and yield point on a steel bar using an artificial neural network with taguchi particle swarm optimizer," *IEEE Access*, vol. 4, pp. 585–593, 2016.

[11] T. Thankachan and K. Sooryaprakash, "Artificial neural network-based modeling for impact energy of cast duplex stainless steel," *Arabian J. Sci. Eng.*, vol. 43, no. 3, pp. 1335–1343, Mar. 2018.

[12] L. Wang, Z. Mu, and H. Guo, "Application of support vector machine in the prediction of mechanical property of steel materials," *J. Univ. Sci. Technol. Beijing, Mineral, Metall., Mater.*, vol. 13, no. 6, pp. 512–515, Dec. 2006.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[14] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[17] Z. Tu *et al.*, "Multi-stream CNN: Learning representations based on human-related regions for action recognition," *Pattern Recognit.*, vol. 79, pp. 32–43, Jul. 2018.

[18] K. B. Lee, S. Cheon, and C. O. Kim, "A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 2, pp. 135–142, May 2017.

[19] Z. M. Li, S. W. Xu, L. G. Cui, G. Q. Li, X. X. Dong, and J. Z. Wu, "The short-term forecast model of pork price based on CNN-GA," *Adv. Mater. Res.*, vol. 628, pp. 350–358, Dec. 2012.

[20] S. Li and V. Ajjarapu, "Real-time monitoring of long-term voltage stability via convolutional neural network," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Chicago, IL, USA, Jul. 2017, pp. 1–5.

[21] Q. Suo *et al.*, "Personalized disease prediction using a CNN-based similarity learning method," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Kansas City, MO, USA, Nov. 2017, pp. 811–816.

[22] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.

[23] S. Wang, J. Peng, J. Ma, and J. Xu, "Protein secondary structure prediction using deep convolutional neural fields," *Sci. Rep.*, vol. 6, no. 1, May 2016, Art. no. 18962.

[24] G. Xia, C. Bernhard, S. Ilie, and C. Fuerst, "A study about the influence of carbon content in the steel on the casting behavior," *Steel Res. Int.*, vol. 82, no. 3, pp. 230–236, Mar. 2011.

[25] X.-W. Yang *et al.*, "Prediction of mechanical properties of A357 alloy using artificial neural network," *Trans. Nonferrous Met. Soc. China*, vol. 23, no. 3, pp. 788–795, Mar. 2013.

[26] M. B. Esfahani, M. R. Toroghinejad, and S. Abbasi, "Artificial neural network modeling the tensile strength of hot strip mill products," *ISIJ Int.*, vol. 49, no. 10, pp. 1583–1587, 2009.

[27] I. Mohanty, S. Sarkar, B. Jha, S. Das, and R. Kumar, "Online mechanical property prediction system for hot rolled IF steel," *Ironmaking Steelmaking*, vol. 41, no. 8, pp. 618–627, Sep. 2014.

[28] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision—ECCV*, vol. 8689, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 818–833.

[29] S. Ioffe and C. Szegedy. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift." [Online]. Available: https://arxiv.org/abs/1502.03167

[30] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.

[31] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[32] M. Liu, J. Shi, Z. Li, C. Li, J. Zhu, and S. Liu, "Towards better analysis of deep convolutional neural networks," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 1, pp. 91–100, Jan. 2017.
[33] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
[34] X. Li, S. Chen, X. Hu, and J. Yang. (2018). "Understanding the disharmony between dropout and batch normalization by variance shift." [Online]. Available: https://arxiv.org/abs/1801.05134

ZHI-WEI XU received the B.S. degree in computer science from the Wuhan University of Science and Technology, Wuhan, China, in 2017, where he is currently pursuing the M.S. degree with the School of Computer Science and Technology.

His research interests include multi-objective evolutionary computation, neural networks, DNA computing, data analysis, and quality engineering.

XIAO-MING LIU received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2007.

From 2008 to 2009, he was a Postdoctoral Research Associate with the Image Processing and Bioimaging Research Laboratory, Alcorn State University, Alcorn, MS, USA. He is currently a Postdoctoral Researcher with the Huazhong University of Science and Technology, Wuhan, China, and as a Professor with the School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China. His research interests include medical image processing, pattern recognition, and machine learning.

KAI ZHANG received the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2008.

From 2008 to 2010, he was a Postdoctoral Researcher with Peking University, Beijing, China. He is currently a Professor with the School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan. His research interests include evolutionary computation, DNA computing, intelligent control and systems, neural networks, and quality engineering.