ISSN 2616-5775 Vol. 8, Issue 6: 1-12, DOI: 10.25236/AJCIS.2025.080601

Prediction of Mechanical Properties of Rolled Steel Based on Multi-Scale Expert System

Wenkui Wu, Qiwen Zhang

School of Computer and Communication, Lanzhou University of Technology, Lanzhou, Gansu, 730050, China

Abstract: *The mechanical properties of carbon steel sheets directly depend on the chemical composition and process parameters of steel, how to establish the mapping relationship between key parameters and mechanical properties is the focus of predicting the mechanical properties of rolled steel. Traditional convolutional neural networks cannot effectively model the correlation between parameters, and there is a problem of feature loss during the feature transfer process. This paper proposed an expert system based on multi-scale for predicting carbon steel sheets' mechanical properties. A multi-scale module for the entire process was proposed to extract comprehensive features from the data, where a graph convolutional neural network captures the nonlinear causal relationships between production data, and the multi-scale convolution module ensures effective feature transfer. Secondly, the multi-activation module, composed of parallel channel and spatial attention, focuses on key features, effectively improving the network model's generalisation performance and computational efficiency.*

Keywords: *Carbon Steel Sheet; Multiscale; Graph Convolution; Mechanical Property Prediction; Attention Mechanism*

1. Introduction

Carbon steel sheets are widely used in the industrial field, and their production process is a nonlinear and complex system. The mechanical properties of carbon steel sheets largely determine the reliability of the final product. Yield strength, tensile strength, and elongation are three important mechanical property indicators. Mechanical properties not only directly affect the quality of the final product but also relate to its safety and durability in practical applications. Therefore, accurately predicting these properties is crucial [1].

Currently, model-based methods for predicting the mechanical properties of steel mainly include those based on metallurgical mechanism models and those based on data-driven approaches. Due to the complex dynamic nonlinearities in the production process of carbon steel sheets, it is very difficult to conduct a comprehensive investigation of the production parameters. In the past, methods for predicting mechanical properties primarily relied on trial and error to establish physical metallurgy mechanism models [2].During the modelling process, they were entirely dependent on human experience, which made it difficult to capture the nonlinear relationships in production data, increased production costs, and prolonged production cycles.

Convolutional neural networks (CNNs) [3] have achieved excellent results in the current field. For example, Li et al. proposed a model for predicting the tensile strength of hot-rolled steel using convolutional neural networks and experimentally validated the superiority of their model [4]. Mehdi et al. applied transfer learning using a pre-trained VGG19 model to predict the microhardness of threaded rolled steel [5]. Wang et al. established a convolutional neural network to predict the flatness of strip steel under different conditions [6]. Hu et al. used a multi-scale convolutional module to enhance the network's perceptual ability, effectively improving the prediction accuracy of the final cooling temperature of rolled steel [7]. Although the above methods provide some new ideas for establishing prediction models in the mechanical property prediction of rolled steel, they still have many shortcomings. As the depth of the network increases, some key information fails to be effectively transmitted to deeper layers. Shallow features play an important role in the early stages of the network by helping the model recognize basic patterns and structures. The loss of these features makes it difficult for the model to capture more complex features and relationships, thereby affecting overall performance. The attention mechanism can autonomously focus on key features, enhancing the model's sensitivity to important characteristics. Li et al. proposed a new attention module to capture global views and transmit features [8]. Zhang et al. used

ISSN 2616-5775 Vol. 8, Issue 6: 1-12, DOI: 10.25236/AJCIS.2025.080601

a channel-space concatenation-level attention mechanism for denoising mechanical fault signals and proposed a selective kernel convolutional neural network for mechanical fault diagnosis [9]. Cao et al. proposed integrating the SE-Res module into the convolutional network to enhance the model's ability to recognize important features [10]. In convolutional neural networks, most models adopt a hierarchical structure [3], where only the features from the last convolutional layer are used for prediction. As features pass through each convolutional layer, information is aggregated to form abstract features for prediction. This results in the loss of detailed features, affecting the model's performance and its ability to adapt to different data. In addition, in the production of carbon steel sheets, each process and parameter is compact and part of the entire process, with complex relationships between the parameters. One-dimensional data cannot fully capture these characteristics. To address the above issues, an expert system based on multi-scale(ESMS) was proposed for predicting the mechanical properties of carbon steel sheets. This system introduces a whole-process multi-scale convolution module (WP-MSC), consisting of a graph convolutional network (GCN) and a multi-scale convolution module (MSC). In this paper, one-dimensional data is converted into a causal relationship graph between variables using granger causality [11]. A graph convolutional network is then used to explore the process characteristics between variables and extract the causal association features of different variables. The multi-scale convolution module consists of convolution kernels of different scales, mining latent features at different levels while using lateral connections to fuse detailed features. This ensures that our model can adapt to different types of rolled steel, improving its generalization ability. This paper proposes a dual activation module (DAM), which consists of an adaptive convolutional channel attention mechanism [12] and a spatial activation module. The dual activation module simultaneously focuses on both the feature space and channels, which helps the model prioritize key features and improves the prediction accuracy. Extensive experiments have demonstrated the superiority of this method.

2. Related work

2.1 Convolutional neural network

Convolutional Neural Networks (CNNs) [3] are one of the most widely used models in deep learning. They reduce the number of model parameters by implementing parameter sharing through sliding convolutional kernels, which helps extract local features from images. Pooling layers are then used to reduce the dimensionality of the data while retaining important information. By repeating this process across the entire input data, a complete output feature map is generated. To extract rich features, CNNs typically employ multiple convolutional kernels, allowing the input data to be analyzed from different scales and perspectives, thereby enhancing the model's performance.

2.2 Attention mechanism

The attention mechanism [13] was first introduced in the field of natural language processing, particularly for machine translation tasks, to address issues in processing long sequence data. It dynamically assigns different emphasis weights to each element in the sequence, allowing the model to focus more effectively on relevant information, thereby improving performance and accuracy. With the development of deep neural networks, the attention mechanism has been widely applied and developed across various fields [14]. The channel attention mechanism [15] is used to adjust the importance between different channels. It compresses the feature information of each channel into a single scalar value through global average pooling and learns the weights of each channel using fully connected layers. The learned weights are then applied to the original feature map. Qin et al. [16] regarded global average pooling (GAP) as a special case of discrete cosine transform and proposed using a general 2D cosine transform to replace GAP, which significantly improves accuracy with almost no increase in computational cost. Li et al. [17] improved feature representation by selectively applying convolutional kernels of different scales, capturing multi-scale features at different levels. This approach helps enhance the network’s adaptability and generalization ability.

2.3 Graphical convolutional neural network

Graph-structured data contains rich information. In the industrial production domain, constructing graph data can describe the causal relationships between different data points. The input to a Graph Convolutional Network (GCN) [18] is a graph data structure containing nodes and edges. By stacking multiple convolutional layers, GCNs progressively aggregate the features of nodes and their neighbours,

ISSN 2616-5775 Vol. 8, Issue 6: 1-12, DOI: 10.25236/AJCIS.2025.080601

mining the latent associations between the data. Assume we are given an input feature matrix X and an adjacency matrix A. X contains the feature vectors of each node, and A describes the connectivity between the nodes. To enhance the model's expressiveness, the adjacency matrix A needs to be normalized, as shown in Equation (1), where I is the identity matrix and D is the degree matrix. The convolution process is shown in Equation (2).

$$\tilde{A} = D^{-\frac{1}{2}}(A + I)D^{\frac{1}{2}} \tag{1}$$

$$H^{l+1} = \sigma(\tilde{A}H^{l}W^{l}) \tag{2}$$

Where $H^{(0)}=X$ is the input node feature matrix, $H^{(l)}$ is the node feature matrix at the lth layer, $W^{(l)}$ is the learnable weight matrix at the lth layer, and σ is the nonlinear activation function. This convolution operation aggregates the features of each node and its neighbors through the adjacency matrix $\tilde{A}$, enabling the node features to capture both local and global information from the graph structure.

3. Methods

3.1 Whole process multi-scale convolution module

Steel rolling production is a whole-process, highly complex process [19], where every production parameter is crucial, and even small changes in parameter values can impact the mechanical properties of the steel. Changes in different parameters not only affect the mechanical properties of the steel but also influence the variations of other parameters. Chemical composition, as the most critical production factor, plays a decisive role in mechanical properties. A graph data structure can effectively reflect the correlations between parameters. To explore the whole-process association characteristics of carbon steel sheet production, we convert the one-dimensional production data into graph data with relational associations using neural granger causality [11]. We use graph convolutional networks to mine the latent relationships between nodes and edges and then apply node embedding methods to transform them into one-dimensional data, which will be used as part of the subsequent input. This process is illustrated in Equation (3)-(5).

$$\mathrm{g(v,e)=Gc(x)} \tag{3}$$

$$y = GCN(g(v,e)) \tag{4}$$

$$\hat{x} = node2vec(y) \tag{5}$$

Where $Gc()$ represents autoregressive granger causality analysis, which converts the one-dimensional data x into a directed graph g. Then, a stacked graph convolutional network is used to mine g and obtain y. Finally, *node2vec* is applied to transform the feature y into one-dimensional data $\hat{x}$.

A multi-scale convolution module (MSC) is proposed. This module uses convolution kernels with different receptive fields to mine the input data. Through lateral connections, it adds detailed features into convolution operations with larger receptive fields, effectively avoiding the feature loss caused by the hierarchical structure. The outputs of each convolution kernel are fused to obtain a multi-scale fused feature, which is then used as the input for the subsequent network layers. This module effectively preserves the detailed features of the original data while mining the latent relationships between different features. Suppose the input features are $x = [I_1, I_2, \ldots, I_n]$, as shown in Equation (6)-(10).

$$F = C^{1\times1}(\varphi(x)) \tag{6}$$

$$F_1 = C^{1\times1}(\varphi(F)) \tag{7}$$

$$F_2 = C^{3\times1}(\varphi(F \oplus F_1)) \tag{8}$$

$$F_3 = C^{5\times1}(\varphi(F \oplus F_2)) \tag{9}$$

$$\bar{x} = C^{1\times1}(\varphi(F_1 \oplus F_2 \oplus F_3)) \tag{10}$$

Where $C^{i\times1}$ represents the convolution operation with different receptive fields, φ denotes the nonlinear activation function, and $\oplus$ denotes the concatenation operation. The multi-scale convolution module generates a comprehensive feature $\bar{x}$ that contains both detailed and abstract features. Based on the above method, the whole-process multi-scale convolution module (WP-MSC) was proposed. As shown in Fig.1.

ISSN 2616-5775 Vol. 8, Issue 6: 1-12, DOI: 10.25236/AJCIS.2025.080601

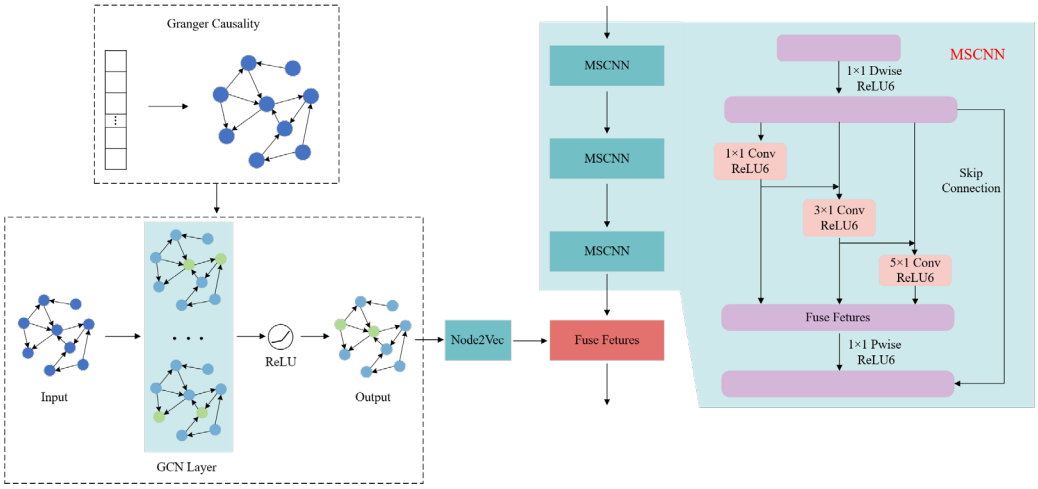


Fig.1 The whole process multi-scale convolution module

3.2 Dual activation module

In the production process of carbon steel sheets, numerous factors influence their mechanical properties. Each parameter contributes differently to the model. The variation of a certain parameter not only alters the mechanical properties of the steel but may also affect other parameters. However, relying solely on manual expertise makes it challenging to effectively distinguish the importance of these production parameters. The dual activation module (DAM) is proposed to simultaneously extract critical features along spatial and channel dimensions, enabling the model to focus on more important features.

In regression prediction tasks, the importance of different feature channels varies. Some features may have a greater impact on the prediction results, while others are relatively less important. Therefore, studying and understanding the global channel relationships between these features becomes particularly crucial. The channel attention mechanism [15] plays a key role in this regard. Through the channel attention mechanism, the weights of individual feature channels can be dynamically adjusted, thereby highlighting important features and suppressing irrelevant or minor features, which improves the overall prediction accuracy and stability. The adaptive convolutional channel attention mechanism [12] is used to extract the channel attention scores, as shown in Equation (11)-(13).

$$z = GMP(X_i) \tag{11}$$

$$k = |log_\gamma(c)|_{odd} \tag{12}$$

$$p^c = \varphi(C^{1\times1}(z,k)) \tag{13}$$

First, the global maximum pooling operation *GMP()* is used to explore the input feature X_i, obtaining the global information z for each channel. Equation (12) adapts the convolutional kernel k based on the number of feature channels. γ is a hyperparameter, whose optimal value is determined experimentally. The attention score p^c for each channel feature is obtained through 1D convolution, where φ represents the nonlinear activation function.

The spatial activation module plays a key role in the model by guiding it to integrate spatial position features from different channels, thereby improving the accuracy and effectiveness of feature representation. Specifically, these modules perform in-depth exploration of the input features, capturing subtle spatial information and utilizing this information to enhance the model's expressiveness. The following operations are performed on the input feature X_i.

$$q^c = C^{1\times1}(C^{3\times1}(C^{1\times1}(X_i))) \tag{14}$$

Specifically, a *1×1* convolution is used to compress the input feature channels to half of their original size, representing a lower-dimensional feature. Then, a *3×1* convolution is applied to perform activation mining on the low-dimensional abstract features. Finally, a *1×1* convolution is used to map the channel number back to the initial dimension to obtain the spatial activation feature q^c.

Based on the aforementioned channel attention scores and spatial activation features, dual activation is defined by Equation (15) – (17). Here, v is the dual attention score, and e is the normalized value. The weighted value $\hat{X}_i$ is obtained by performing element-wise multiplication between X_i and e and $\odot$

ISSN 2616-5775 Vol. 8, Issue 6: 1-12, DOI: 10.25236/AJCIS.2025.080601

represents the product of the elements. The DAM module is shown in Fig. 2.

$$v = p^c \odot q^c \tag{15}$$

$$e = sigmoid(v) \tag{16}$$

$$\hat{X}_i = X_i \odot e \tag{17}$$

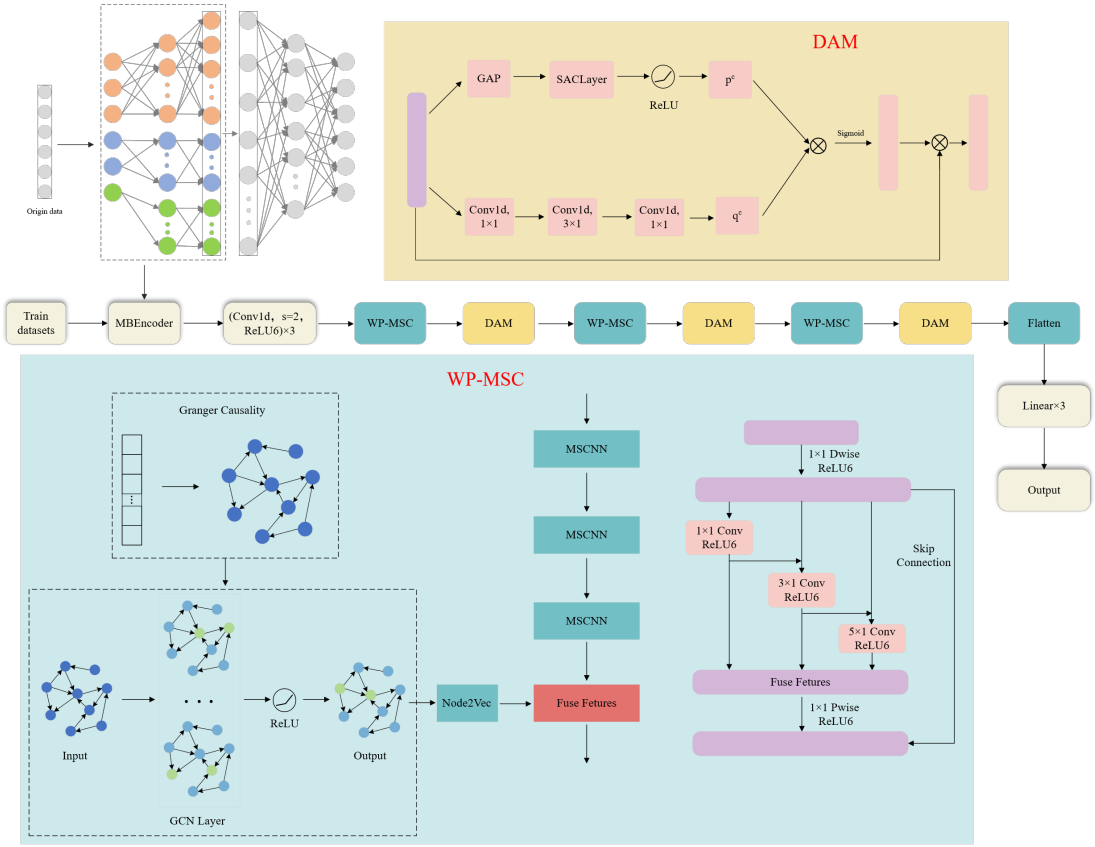


Fig.2 Expert system based on multi-scale

3.3 Expert system based on multi-scale

This paper proposes an expert system based on multi-scale (ESMS) for predicting the mechanical properties of carbon steel sheets, as shown in Fig. 2. The system consists of two parts: the feature extraction module and the prediction module. The normalized data first undergoes two-point convolution to expand the channels and enrich semantic information, which serves as the input to the feature extraction module. The feature extraction module is composed of multiple stacked WP-MSC and DAM modules, with the optimal number of network layers determined through experiments. WP-MSC uses graph convolutional networks to mine the causal relationships between production data, while multi-scale convolution is introduced to ensure the transmission of fine-grained features. Additionally, shortcut connections are incorporated to prevent model overfitting. The DAM focuses on both the spatial and channel dimensions of the features, enabling the model to prioritize important features, enhancing its generalization ability, and improving computational efficiency and usability. The output of the feature extraction module serves as the input to the prediction module, which maps it into the mechanical property prediction space using a support vector machine. Extensive experiments demonstrate that ESMS performs excellently and is highly robust in predicting the mechanical properties of carbon steel sheets.

4. Experiments

4.1 Description of data and indicators

The data used were based on the data collected from the production line of JISCO's carbon thin plate plant, which was in line with international standards and has practical significance. Before conducting

ISSN 2616-5775 Vol. 8, Issue 6: 1-12, DOI: 10.25236/AJCIS.2025.080601

the experiments, both datasets were screened, cleaned, and standardized, resulting in 16699 cold-rolled steel data samples and 7895 hot-rolled steel data samples. In the cold-rolled steel dataset, each sample contains 14 features and 3 main mechanical performance indicators, while each sample in the hot-rolled steel dataset contains 10 features and 3 main mechanical performance indicators. For the experiments, each dataset was randomly divided into a training set and a test set, with 80% used for training and 20% used for testing. Table 1 shows the data distribution statistics of the minimum, maximum, mean, and standard deviation for all input variables and labels in the cold-rolled steel dataset used in the experiment. Among them, C, Si, Mn, P, S, ALS, Cu, Ni and Al represent nine different chemical compositions, CRT, FE, BT, FRT, and CT represent five process parameters of cold rolled thickness, flatness elongation, billet thickness, final rolling temperature and coiling temperature, respectively, and YS, TS, EL represent the yield strength, tensile strength and elongation of the three mechanical property indicators. In addition, the experimental data were normalised using the z-score method to ensure the stability of the model.

Table 1 Cold rolled dataset

Parameter	Minimum	Maximum	Mean	Standard deviation
C	0.0018	0.0634	0.0440	0.0081
Si	0.0140	0.0740	0.0335	0.0087
Mn	0.0830	0.2790	0.1598	0.0296
P	0.0033	0.0258	0.0110	0.0029
S	0.0013	0.0095	0.0044	0.0013
ALS	0.0166	0.0524	0.0334	0.0056
Cu	0.0135	0.0631	0.0233	0.0050
Ni	0.0056	0.0568	0.0140	0.0063
Al	0.0003	0.0722	0.0032	0.0076
CRT	0.23	3.00	1.7339	0.8816
FE	0.48	1.39	1.0623	0.1809
BT	1.80	6.00	4.5938	1.2056
FTT	655.38	958.34	900.52	11.933
CT	590.00	730.00	623.49	13.690
YS	151.00	254.00	186.94	11.460
TS	286.00	373.00	318.45	11.661
EL	31.000	71.500	45.554	3.5066

To evaluate the effectiveness of the final prediction results, four performance metrics were used. The mean absolute error (MAE) measures the average absolute difference between the model's predicted values and the actual values. A smaller MAE indicates that the model's predictions are closer to the actual values. The mean squared error (MSE) and root mean squared error (RMSE) measure the difference between the predicted and actual values, with smaller values being better. The coefficient of determination (R^2) reflects the extent to which the model explains the variance of the target variable. Its value ranges from 0 to 1, with values closer to 1 indicating that the model explains the target variable more effectively and has a better fit. These are shown in Equation (18)-(21).

$$MAE = \frac{1}{n}\Sigma_{i=1}^{n}|y_i - \hat{y}_i| \tag{18}$$

$$MSE = \frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{19}$$

$$MSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{20}$$

$$R^2 = 1 - \frac{\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2}{\Sigma_{i=1}^{n}(y_i - \bar{y})^2} \tag{21}$$

4.2 Parameterization of the network

The hyperparameters involved in this study were determined using grid search, which explores every parameter combination to find the optimal values. Specifically, the adaptive factor γ of the adaptive convolutional channel attention mechanism is set to 4, the learning rate r is set to 0.001, and the batch size (epoch) is set to 300. Additionally, a multi-branch autoencoder [12] was used for data augmentation to expand the input dimensions of the original data. For the cold-rolled steel dataset and hot-rolled steel dataset, the number of autoencoder branches n was set to 3 and 2, respectively.

ISSN 2616-5775 Vol. 8, Issue 6: 1-12, DOI: 10.25236/AJCIS.2025.080601

4.3 Ablation experiments

Table 2 Ablation experiments

Dataset	Target Variables	Metrics	Module	DAM	MSCNN-DAM	GCN-DAM	Proposed Model
			GCN	×	×	√	√
			MSCNN	×	√	×	√
			DAM	√	√	√	√
Cold-rolled carbon steel	YS	MAE		0.3523	0.2832	0.2713	0.1165
		MSE		0.2256	0.1462	0.1322	0.0362
		RMSE		0.4724	0.3824	0.3833	0.1826
		R^2		0.7363	0.8296	0.8315	0.9482
	TS	MAE		0.2945	0.2769	0.2719	0.1093
		MSE		0.1932	0.1368	0.1304	0.0367
		RMSE		0.4526	0.3756	0.3775	0.1722
		R^2		0.7627	0.8336	0.8362	0.9572
	EL	MAE		0.2638	0.2724	0.2649	0.1069
		MSE		0.1934	0.1262	0.1344	0.0219
		RMSE		0.4365	0.3723	0.3720	0.1690
		R^2		0.7749	0.8484	0.8494	0.9827
Hot-rolled carbon steel	YS	MAE		0.3643	0.2802	0.2785	0.1171
		MSE		0.2367	0.1533	0.1442	0.0473
		RMSE		0.4825	0.3902	0.3839	0.1917
		R^2		0.7261	0.8232	0.8303	0.9493
	TS	MAE		0.3036	0.2735	0.2741	0.1128
		MSE		0.1935	0.1362	0.1347	0.0324
		RMSE		0.4638	0.3742	0.3832	0.1884
		R^2		0.7471	0.8275	0.8331	0.9583
	EL	MAE		0.2723	0.2623	0.2667	0.1026
		MSE		0.1932	0.1473	0.1451	0.0327
		RMSE		0.4203	0.3621	0.3642	0.1894
		R^2		0.7689	0.8332	0.8472	0.9794

To verify the effectiveness of each module in the proposed model, ablation studies were conducted on the test set in this section. By removing individual modules and observing their impact on model performance, the contribution of each module is demonstrated. First, the importance of the WP-MSC module is discussed. The graph convolutional network was removed, and traditional convolutions were used instead of multi-scale convolutions to test the effectiveness of the WP-MSC module, as shown in Table 2.

4.4 Comparative tests

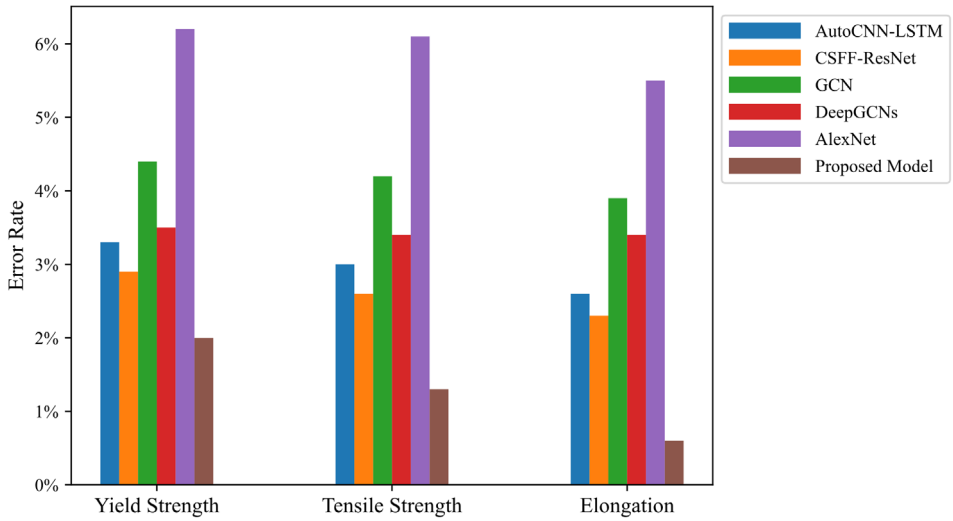
In this section, comparison experiments were conducted using the cold-rolled steel and hot-rolled steel production datasets. The ESMS model was compared with five other existing prediction models. The input data for both datasets were standardized, with the cold-rolled steel dataset containing 14 parameters, and the hot-rolled steel dataset containing 10 parameters. The output consisted of three mechanical performance indicators. The comparison models include the classical models AlexNet [20]and GCN [18], as well as newer models proposed in recent years. DeepGCNs [21] defines a differentiable generalized aggregation function that unifies different message aggregation operations, adopts a deeper structure, and addresses the problem of information vanishing in graph-structured data. CSFF-ResNet [22] constructed a channel-space parallel residual network. Additionally, the AutoCNN-LSTM hybrid network model proposed in reference [23] was used. The proposed network model demonstrated excellent prediction performance and robustness through multiple evaluation metrics, as shown in Table 3. To ensure fairness, all models used multi-branch autoencoders [12] for data augmentation.

ISSN 2616-5775 Vol. 8, Issue 6: 1-12, DOI: 10.25236/AJCIS.2025.080601

Table 3 Comparative experiments

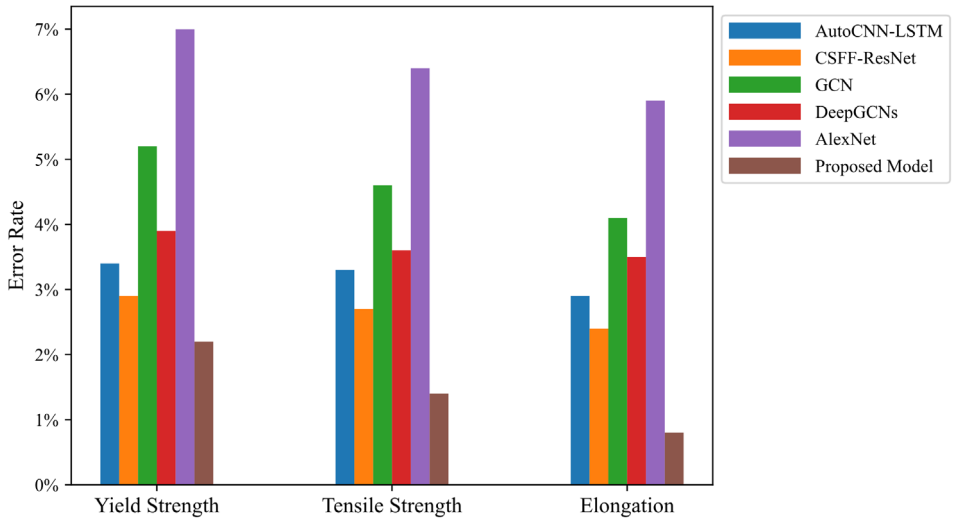
Dataset	Target Variables	Metrics	Model					
			AlexNet	GCN	DeepGCNs	CSFF-ResNet	AutoCNN-LSTM	ESMS
Cold-rolled carbon steel	YS	MAE	0.365	0.2823	0.2048	0.1717	0.1894	0.1165
		MSE	0.1738	0.1482	0.1068	0.0747	0.0820	0.0362
		RMSE	0.4172	0.3973	0.3263	0.2728	0.2829	0.1826
		R^2	0.7684	0.8109	0.8811	0.9248	0.9126	0.9482
	TS	MAE	0.2973	0.2896	0.1911	0.1479	0.1590	0.1093
		MSE	0.1872	0.1497	0.0902	0.0616	0.0729	0.0367
		RMSE	0.4383	0.3830	0.2999	0.2476	0.2529	0.1722
		R^2	0.7643	0.8146	0.9005	0.9297	0.9189	0.9572
	EL	MAE	0.3027	0.2790	0.1702	0.1435	0.1521	0.1069
		MSE	0.1747	0.1478	0.0791	0.0621	0.0792	0.0219
		RMSE	0.4172	0.3876	0.2795	0.2423	0.2527	0.1690
		R^2	0.7720	0.8272	0.9143	0.9304	0.9262	0.9827
Hot rolled carbon steel	YS	MAE	0.3104	0.2894	0.2133	0.1773	0.1849	0.1171
		MSE	0.1792	0.1593	0.1187	0.0794	0.0887	0.0473
		RMSE	0.4138	0.3980	0.3376	0.2810	0.2928	0.1917
		R^2	0.7559	0.8272	0.8741	0.9201	0.9192	0.9493
	TS	MAE	0.2927	0.2820	0.2052	0.1583	0.1650	0.1128
		MSE	0.1733	0.1481	0.0982	0.0695	0.0830	0.0324
		RMSE	0.4204	0.3930	0.3173	0.2593	0.2638	0.1884
		R^2	0.7628	0.8261	0.8925	0.9267	0.9198	0.9583
	EL	MAE	0.3128	0.2783	0.1774	0.1483	0.1849	0.1026
		MSE	0.1627	0.1528	0.0836	0.0703	0.0947	0.0327
		RMSE	0.4142	0.3783	0.2839	0.2572	0.2637	0.1894
		R^2	0.7719	0.8229	0.9060	0.9296	0.9262	0.9794

Based on the working environment conditions and conventional experience in the field, the deviation ranges for YS, TS, and EL were set to ±10 MPa, ±15 MPa, and ±3, respectively. Based on these error ranges, Fig. 3 shows the error rates of different models on two different datasets. The ESMS model achieves the lowest error rates for all three mechanical performance indicators. The dual activation module and graph convolution network demonstrate excellent data interpretability, adapting well to different datasets and exhibiting strong robustness. Compared to the CSFF-ResNet model, the error rates for YS, TS, and EL in the cold-rolled steel dataset decreased by 31.0%, 50.0%, and 73.9%, respectively, while in the hot-rolled steel dataset, the error rates for YS, TS, and EL decreased by 24.1%, 48.2%, and 66.7%, respectively. This further demonstrates that the ESMS model is better suited for practical industrial requirements.



(a) Cold Rolled Steel.

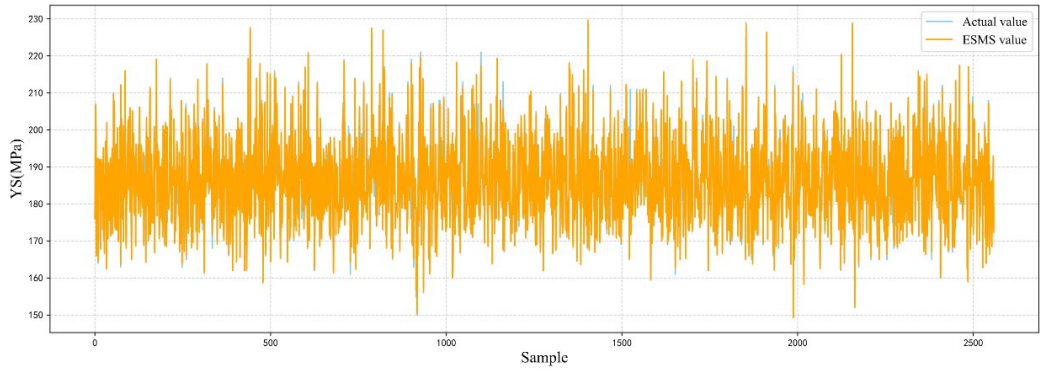
ISSN 2616-5775 Vol. 8, Issue 6: 1-12, DOI: 10.25236/AJCIS.2025.080601



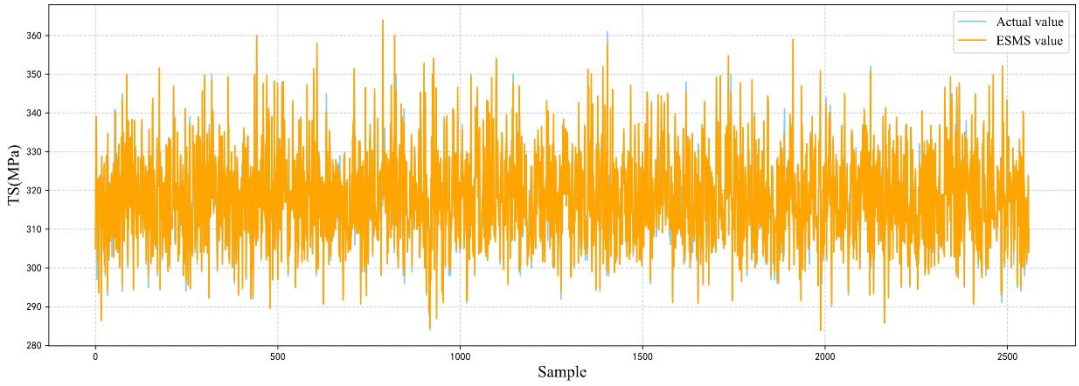
(b) Hot Rolled Steel.

Fig.3 Error rate comparison of different models

To visually demonstrate the effectiveness of the ESMS network model, Fig. 4(a-c) and Fig. 5(a-c) show the fitting results between the predicted and actual values for the three mechanical performance indicators on the cold-rolled steel and hot-rolled steel datasets, respectively. From the figures, it can be seen that the predicted results closely match the actual values, which further validates the effectiveness of the model.

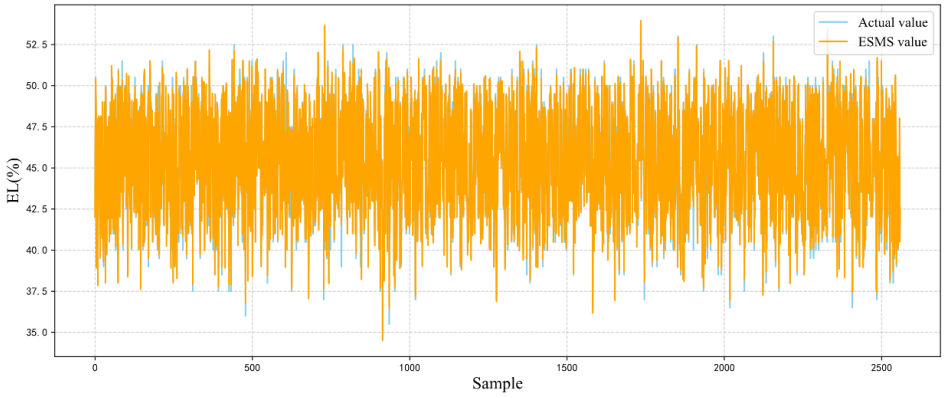


(a) YS.



(b) TS.

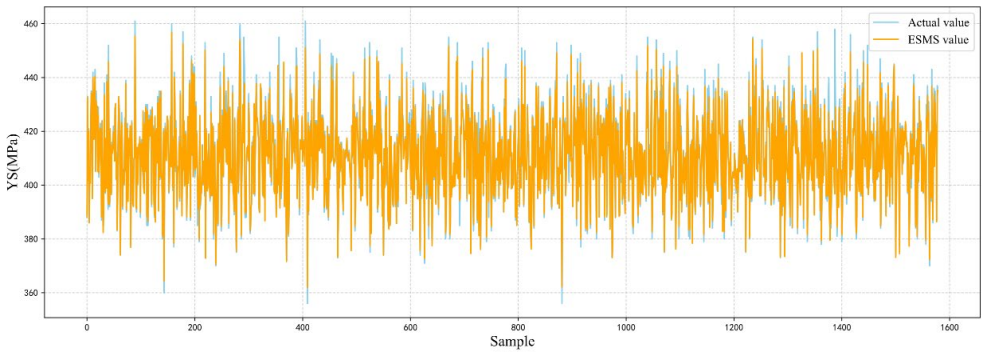
ISSN 2616-5775 Vol. 8, Issue 6: 1-12, DOI: 10.25236/AJCIS.2025.080601



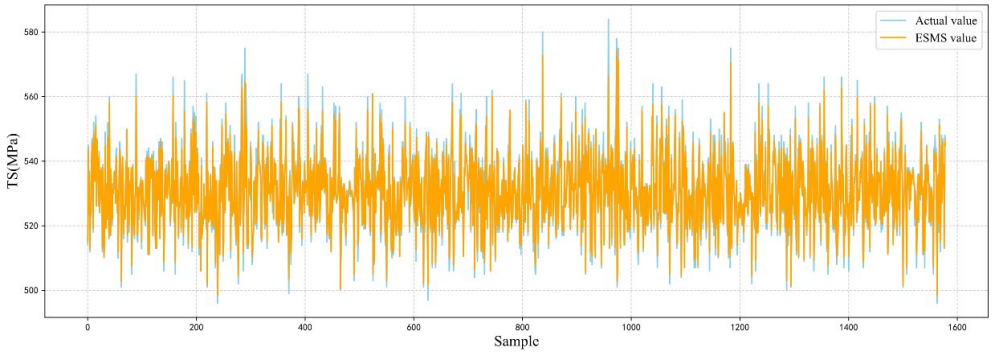
(c) EL.

(a)YS (b)TS (c)EL

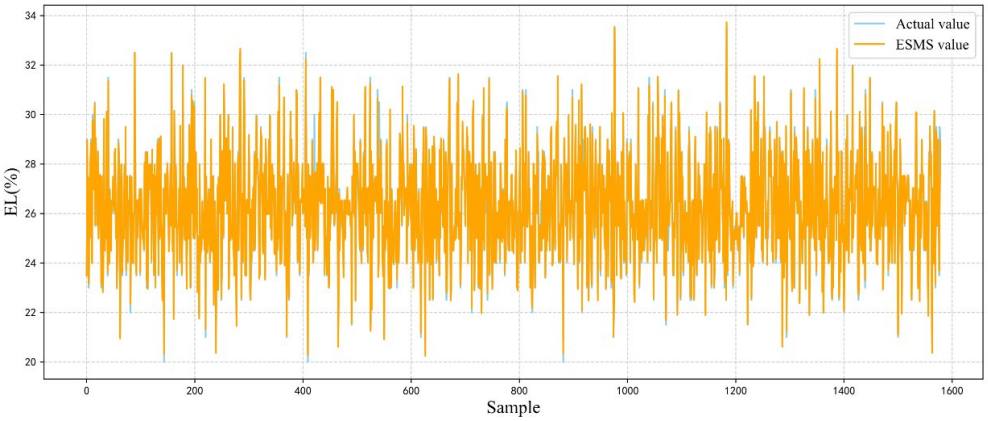
Fig.4 Fitting Curves for the Cold-Rolled Steel Dataset



(a) YS.



(b) TS.



(c) EL.

(a)YS (b)TS (c)EL

Fig.5 Fitting Curves for the hot-Rolled Steel Dataset

ISSN 2616-5775 Vol. 8, Issue 6: 1-12, DOI: 10.25236/AJCIS.2025.080601

5. Conclusion

This paper proposes an expert system based on multi-scale for predicting the mechanical properties of different types of steel. First, a whole-process multi-scale convolution module is introduced to mine the graph-structured production process data, capturing the correlation features between various production parameters, while multi-scale convolution ensures the transmission of fine-grained features. Secondly, the dual activation module focuses on the spatial and channel dimensions of features, making the model more attentive to key features, enhancing its generalization ability, and improving computational efficiency and usability. Through extensive experiments, key factors such as hyperparameters and model depth were determined, and the model's prediction performance was evaluated using multiple metrics and different datasets. On various datasets, ESMS achieved the highest prediction accuracy. Specifically, for the cold-rolled steel dataset, the accuracy for YS, TS, and EL were 98%, 98.7%, and 99.4%, respectively, while for the hot-rolled steel dataset, the accuracy for YS, TS, and EL were 97.8%, 98.6%, and 99.2%, respectively. Experimental results show that the ESMS model demonstrates exceptional prediction performance and strong robustness, providing valuable insights and references for research and applications in related fields.

References

[1] Xia, J. S., et al. "Using feed-forward perceptron Artificial Neural Network (ANN) model to determine the rolling force, power and slip of the tandem cold rolling." ISA transactions 132 (2023): 353-363.
[2] Cui, Chunyuan, et al. A strategy combining machine learning and physical metallurgical principles to predict mechanical properties for hot rolled Ti micro-alloyed steels, Journal of Materials Processing Technology.311(2023):117810.https://doi.org/10.1016/j.jmatprotec.2022.117810
[3] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324. https://doi.org/10.1109/5.726791
[4] Li, Wei-gang, et al. "Prediction model for mechanical properties of hot-rolled strips by deep learning." Journal of Iron and Steel Research International 27 (2020): 1045-1053.
[5] Soleymani, Mehdi, Mohammad Khoshnevisan, and Behnam Davoodi. "Prediction of micro-hardness in thread rolling of St37 by convolutional neural networks and transfer learning." The International Journal of Advanced Manufacturing Technology 123.9 (2022): 3261-3274.
[6] Wang, Yu, et al. "Application of convolutional neural networks for prediction of strip flatness in tandem cold rolling process." Journal of Manufacturing Processes 68 (2021): 512-522.
[7] Hu, Xiao, et al. "Controlled cooling temperature prediction of hot-rolled steel plate based on multi-scale convolutional neural network." Metals 12.9 (2022): 1455. https://doi.org/10.3390/met12091455
[8] Li, Anni, JinQing Qi, and Huchuan Lu. "Multi-attention guided feature fusion network for salient object detection." Neurocomputing 411 (2020): 416-427.
[9] Zhang, Shuo, et al. "Selective kernel convolution deep residual network based on channel-spatial attention mechanism and feature fusion for mechanical fault diagnosis." ISA transactions 133 (2023): 369-383.
[10] Cao, Xincheng, Binqiang Chen, and Nianyin Zeng. "A deep domain adaption model with multi-task networks for planetary gearbox fault diagnosis." Neurocomputing 409 (2020): 173-190.
[11] Luo, Xiaoyang, et al. "Prediction of Mechanical Properties of Cold-Rolled Steel Based on Improved Graph Attention Network." Symmetry 16.2 (2024): 188.
[12] Zhang, Qiwen, Wenkui Wu, and Xingchang Tang. "Data enhanced lightweight network-based prediction of cold-rolled steel mechanical properties." Materials Today Communications (2024): 109640.
[13] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473 (2014).
[14] Niu, Zhaoyang, Guoqiang Zhong, and Hui Yu. "A review on the attention mechanism of deep learning." Neurocomputing 452 (2021): 48-62.
[15] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
[16] Qin, Zequn, et al. "Fcanet: Frequency channel attention networks." Proceedings of the IEEE/CVF international conference on computer vision. 2021.
[17] Li, Xiang, et al. "Selective kernel networks." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. https://doi.org/10.1109/CVPR.2019.00060
[18] Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. arXiv 2016, arXiv:1609.02907.
[19] Li X, Lu N, Jiang B, et al. A frequent pattern mining based shape defect diagnosis method for cold

ISSN 2616-5775 Vol. 8, Issue 6: 1-12, DOI: 10.25236/AJCIS.2025.080601

rolled strip products[C]//2017 6th International Symposium on Advanced Control of Industrial Processes (AdCONIP). IEEE, 2017: 90-94.
[20] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks." Communications of the ACM 60.6 (2017): 84-90.
[21] Li, Guohao, et al. "Deepergcn: All you need to train deeper gcns." arXiv preprint arXiv:2006.07739 (2020).
[22] Zhang, Shuo, et al. "Selective kernel convolution deep residual network based on channel-spatial attention mechanism and feature fusion for mechanical fault diagnosis." ISA transactions 133 (2023): 369-383.
[23] Ren, Lei, et al. "A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life." IEEE Transactions on Industrial Informatics 17.5 (2020): 3478-3487. https://doi.org/10.1109/TII.2020.3008223.