

## Article

# Prediction of Mechanical Properties of Cold-Rolled Steel Based on Improved Graph Attention Network

Xiaoyang Luo <sup>1</sup>, Rongping Guo <sup>2,\*</sup>, Qiwen Zhang <sup>2</sup> and Xingchang Tang <sup>3,4</sup>

<sup>1</sup> Gansu JISCO Group, Hongxing Iron & Steel Co., Ltd., Carbon Steel & Thin Slab Rolling Plant, Jiayuguan 735100, China; luoxiaoyang@jiugang.com

<sup>2</sup> School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China; qingwenzhang@lut.edu.cn

<sup>3</sup> State Key Laboratory of Advanced Processing and Recycling of Nonferrous Metals, Lanzhou 730050, China; tangxingchang@sohu.com

<sup>4</sup> College of Materials Science and Engineering, Lanzhou University of Technology, Lanzhou 730050, China

\* Correspondence: 212085400220@lut.edu.cn; Tel.: +86-17694834825

**Abstract:** The prediction of mechanical properties of cold-rolled steel is very important for the quality control, process optimization, and cost control of cold-rolled steel, but it is still a challenging task to predict accurately. For the existing graph structure of graph attention networks, it is difficult to effectively establish the complex coupling relationship and nonlinear causal relationship between variables. At the same time, it is considered that the process of cold-rolled steel has typical full-flow process characteristics and the graph attention network makes it difficult to extract the path information between the central node and its higher-order neighborhood. The neural Granger causality algorithm is used to extract the latent relationship between variables, and the basic graph structure of mechanical property prediction data is constructed. Secondly, the node embedding layer is added before the graph attention network, which leverages the symmetry nature of Node2vec method by incorporating both breadth-first and depth-first exploration strategies. This ensures a balanced exploration of diverse paths in the graph, capturing not only local structures but also higher-order relationships. The combined graph attention networks are then able to effectively capture the symmetry path information between nodes and dependencies between variables. The accuracy and superiority of this method are verified by experiments in real cold-rolled steel production cases.

**Keywords:** mechanical properties; graph attention networks; nonlinear causal relationship; neural Granger causality; Node2vec



**Citation:** Luo, X.; Guo, R.; Zhang, Q.; Tang, X. Prediction of Mechanical Properties of Cold-Rolled Steel Based on Improved Graph Attention Network. *Symmetry* **2024**, *16*, 188. <https://doi.org/10.3390/sym16020188>

Academic Editor: Luigi Nicolais

Received: 12 January 2024

Revised: 31 January 2024

Accepted: 1 February 2024

Published: 5 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Cold-rolled steel, known for its smooth surface and exceptional mechanical properties, finds extensive applications in various manufacturing sectors, including automobile, shipbuilding, and electrical appliances [1,2]. Mechanical properties serve as pivotal indicators reflecting the quality of cold-rolled steel during the rolling production process and act as crucial control parameters. The assessment of these properties typically involves destructive experiments, incurring significant time and resource costs [3]. Moreover, the intricate relationships between mechanical properties, process parameters, and chemical compositions make achieving accurate predictions a challenging endeavor. Hence, the development of a mechanical property prediction model based on chemical composition and process parameters holds paramount importance for reducing production costs, enhancing efficiency, and elevating product quality.

At present, the model-based methods for predicting mechanical properties of rolled steel mainly include the methods based on metallurgical mechanism model and the methods based on data drive. However, in the cold-rolling process, both austenitic phase and ferritic phase undergo complex microstructure evolution, so it is difficult to establish an

accurate mechanism model to explain the relationship between rolling process, chemical composition, and mechanical properties [4,5]. Compared with the metallurgical mechanism model, the data-driven method is based on the historical data of actual production, which does not rely on the physical model or theoretical assumptions of the actual construction, so it can be applied to different types of steel alloys and smelting processes, and is more flexible in dealing with the diversity and complexity of the problem, and has become an effective method for predicting the mechanical properties of rolled steel. For example, based on the multi-grain cascade forest architecture [6], the basic information of steel rolling and the local information collected by multi-grain are integrated to complete the performance prediction of steel rolling. Yan et al. [7] applied MSVR and particle swarm optimization to establish a multi-output prediction model for mechanical properties of cold-rolled steel and a multi-objective quality control method for cold-rolled products. In addition, with the rapid development of deep learning in extracting nonlinear representations, it has also been applied to the field of mechanical property prediction of rolled steel. Considering the time series correlation in the steel production process, the two-stage time series model [8] based on the two-stage attention mechanism extracts input features and time features by introducing input attention and time attention adaptive, so as to accurately predict the mechanical properties of rolled steel. The CNN-based method [9] realized the CNN-based steel rolling property prediction method by converting the actual production data into two-dimensional image data.

Although the above methods have been widely used in the prediction of steel rolling properties, they cannot accurately model the whole industrial process of cold rolling, which limits the practical application of the above methods in this field. Graph neural networks (GNN) provide an excellent solution for modeling industrial data dependencies due to their strong feature representation and permutation. Chen et al. [10] realized the long-term prediction of sintering temperature by introducing the adaptive adjacency matrix algorithm and the time-varying spatio-temporal correlation of process data accurately modeled by the spatio-temporal graph attention module. Sun et al. [11] realized accurate online prediction of key indicators of complex industrial processes through multi-modal clustering method of Gaussian mixture model and dynamic attribution graph attention network.

Nevertheless, it is difficult for existing graph attention networks to model complex nonlinear causality among variables. We introduce the neural Granger causality (NGC) algorithm to capture complex nonlinear causality among variables, establishing a graph structure for predicting the mechanical properties of cold-rolled steel. In order to further improve the prediction accuracy of mechanical properties of cold rolling, a node embedding layer is added in front of the graph attention network (GAT), which uses node embedding method to integrate node path information into GAT. Based on the above methods, a neural Granger causality and embedded graph attention network (NGC-EGAT) for predicting mechanical properties of cold-rolled steel is proposed.

## 2. Theoretical Foundation

### 2.1. Granger Causality

Given an  $N$ -dimensional time series  $X = (X_1, X_2, \dots, X_T) = (x_1, x_2, \dots, x_N)$ ,  $X \in \mathbb{R}^{N \times T}$ , Granger causality (GC) is defined as follows: when  $i \neq j$  and other sequences are given, the historical value of  $x_j$  is included in the current value of prediction  $x_i$ , which can improve the accuracy of prediction  $x_i$ , and then it can be concluded that there is GC between  $x_j$  and  $x_i$ . Model-based GC analysis usually uses a vector autoregressive (VAR) model where, at time  $t$ , the time series at  $X_t$  is a linear combination of  $K$  historical values of that series, then:

$$X_t = \sum_{k=1}^K A^{(k)} X(t-k) + \varepsilon_t S \quad (1)$$

where  $A^{(k)}$  is the coefficient matrix of  $N \times N$ ,  $k = 1, 2, \dots, K$ , and  $\varepsilon_t$  is 0 mean Gaussian noise.

In the VAR model, the sufficient condition for the different variables  $X_j$  of multivariate time series  $X$  to have no GC on  $X_i$  is  $A_{ij}^{(k)} = 0$ , that is, GC analysis can be determined in the VAR model by examining which values in all lag coefficient matrix  $A^{(k)}$  are 0.

## 2.2. Graph Attention Networks

GAT [12] is a novel network architecture that addresses the shortcomings of graph convolutional networks (GCN) by using a masked force layer to assign different weights to different nodes in the neighborhood. Assume that given a set of input node features,  $h = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_N\}$ ,  $\vec{h}_i \in \mathbb{R}^F$ , where  $N$  is the number of nodes and  $F$  is the number of features of nodes. The input features are passed through a self-attention layer with masking, employing the LeakyReLU activation function, and subsequently undergoing softmax normalization, resulting in attention coefficients:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\vec{a}^T \left[ W \vec{h}_i \parallel W \vec{h}_j \right] \right)\right)}{\sum_{k \in N_i} \exp\left(\text{LeakyReLU}\left(\vec{a}^T \left[ W \vec{h}_i \parallel W \vec{h}_k \right] \right)\right)} \quad (2)$$

where  $W$  is a parameterized weight matrix of linear change for each node,  $\parallel$  represents the joining operation of the vectors. After the attention coefficient  $\alpha_{ij}$  is obtained, the weighted summation of the feature vectors is performed as the final output feature of each node:

$$\vec{h}'_i = \sigma \left( \sum_{j \in N_i} \alpha_{ij} W \vec{h}_j \right) \quad (3)$$

In order to improve the stability of attention mechanism learning, multi-head attention is used to extend GAT, allowing the model to learn multiple different weight distributions, which helps to better capture multiple dependencies in complex graph data. Figure 1 shows the structure diagram of GAT with three attention heads.

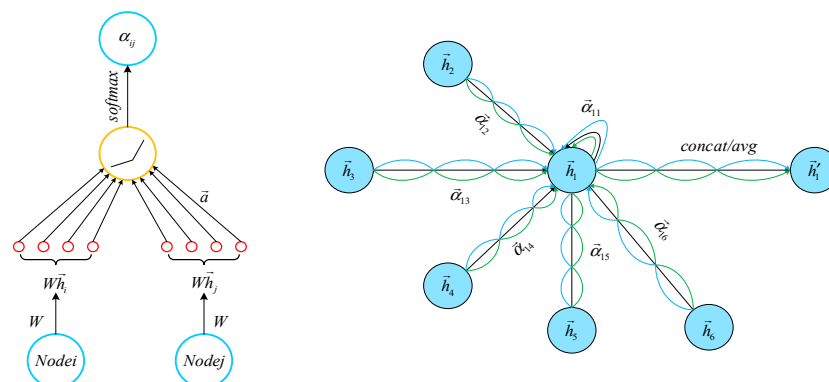


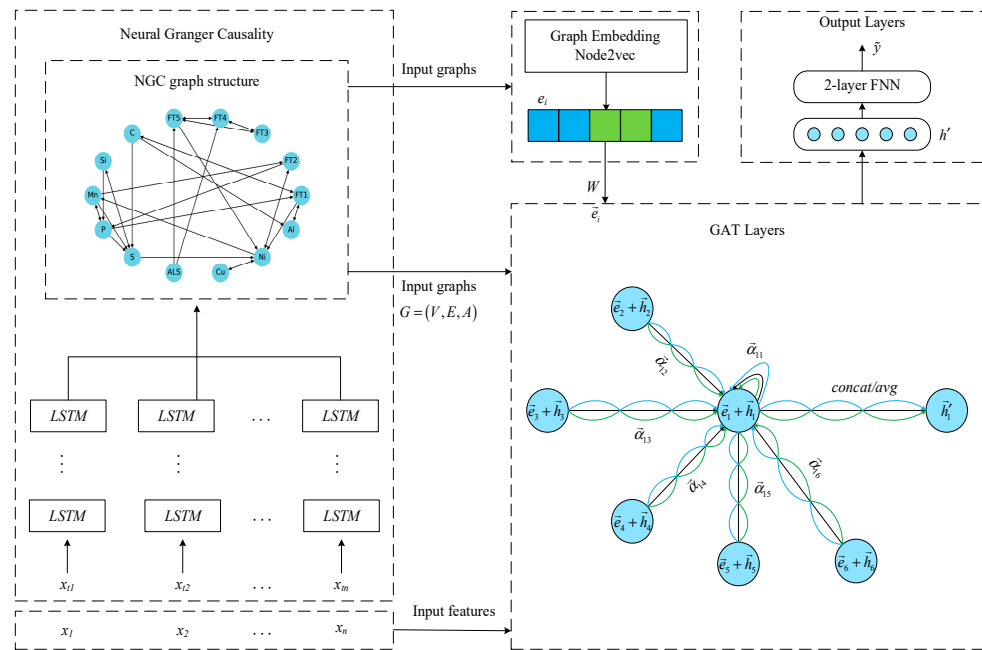
Figure 1. Schematic diagram of GAT structure.

## 3. NGC-EGAT Method

### 3.1. NGC-EGAT Overall Framework

As illustrated in Figure 2, NGC-EGAT primarily comprises two modules: the neural Granger causal graph construction module and graph attention network integrated with node embedding layer. Firstly, recognizing the limitations of existing methods in accurately establishing complex relationships between variables, a novel graph construction method based on the NGC algorithm is introduced. Secondly, a new embedding graph attention network (EGAT) is proposed. EGAT captures node path information by incorporating a node embedding layer before the GAT network. The extracted path information and node

characteristics are then fed into the GAT network to update node representations. Finally, a fully connected layer of two layers is used to obtain the final prediction result.



**Figure 2.** Structure diagram of NGC-EGAT.

### 3.2. Neural Granger Causal Graph Structure

The mechanical properties of cold-rolled steel are usually affected by a variety of process parameters and chemical compositions, and there are intricate relationships among them. Therefore, inspired by [13], this paper proposes NGC learning algorithm by combining long short-term memory (LSTM) with GC analysis, and uses LSTM for GC analysis to extract complex nonlinear causality between variables. The group lasso is used to reduce the introduction of redundant edges, and then the graph structure with efficient information transfer and correct sparsity is constructed.

Specifically, each input component is modeled using a separate LSTM. For the variable  $x_{ti}$  at time  $t$ , it has the following functional relationship:

$$x_{ti} = g_i(x_{>t1}, \dots, x_{>tn}) + \varepsilon_{ti} \quad (4)$$

where  $g_i$  is a nonlinear function that maps the past  $t - 1$  historical values to  $x_{ti}$ ,  $x_{>ti}$  is the past  $t - 1$  historical values of  $x_i$ , and  $\varepsilon_t$  is 0 mean gaussian noise.

LSTM networks are often employed to model intricate temporal dependencies. They incorporate gate mechanisms to regulate information flow, with the cell state  $c_t$  facilitating the transmission of long-term dependencies and the hidden state  $h_t$  aiding in the transmission of short-term dependencies. The computational formula is expressed as follows:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (5)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (6)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (7)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1}) \quad (8)$$

$$h_t = o_t \odot \tanh(c_t) \quad (9)$$

$f_t, i_t, o_t$  stand for forgetting gate, input gate, and output gate, respectively, which are used to control the information update of  $c_t$  and  $h_t$ . Where the forgetting gate  $f_t$  is used to

determine what information should be forgotten from the cell state  $c_t$ , the input gate  $i_t$  determines what information should be added to the cell state  $c_t$ , and the output gate  $o_t$  determines what information of the current input and cell state should be input to the hidden state of the next time step. LSTM can effectively capture the nonlinear dependence relationship between data. Finally, the sequence  $i$  at time  $t$  goes through the nonlinear evolution of LSTM, and its output  $x_{ti}$  can be updated from Equation (4) as follows:

$$x_{ti} = g_i(x_{>t}) + e_{ti} = W_i h_t + e_{ti} \quad (10)$$

where  $W_i$  is the output weight matrix. For  $W_i$ , if we assume that all lag values  $k$  in column  $j$  equal to 0 (i.e.,  $W_i^j = 0$ ), then sequence  $j$  is not the Granger cause of sequence  $i$ , indicating that  $x_{(t-k)j}$  does not affect  $h_t$ , and, thus, does not affect the output  $x_{ti}$ . As cold-rolled steel data are high-dimensional data, in order to select the most relevant causal relationship in  $W_i$  to ensure the sparsity of the graph, group lasso is used in each column of  $W_i$  to restrict the sparsity of the whole group and select the sequence with the most GC with sequence  $i$ . Group lasso is a method to deal with the selection of multiple groups of correlated variables. Variables are divided into several groups. In feature selection, the sparsity constraint is introduced to the whole variable group, and the whole feature group is selected or excluded at the same time, instead of just a single feature, so that variable selection can be more accurate in the case of intra-group correlation. This method can better balance variable selection and model interpretation when dealing with high dimensional data. By using group lasso on Equation (10), it is converted to solving the following optimization problem:

$$L = \min_w \sum_{t=1}^T (x_{ti} - W_i h_t)^2 + \lambda \sum_{j=1}^k \|W_i^j\| \quad (11)$$

where  $W_i^j$  is the  $j$ -th column of  $W_i$ ,  $\lambda$  is the penalty function used to control the sparsity between groups, and the variable  $x_{tj}$  with GC for  $x_{ti}$  is selected through the inter-group constraint term. GC analysis in the LSTM model can be transformed into a variable selection problem, that is, whether the elements in the matrix  $W_i^j$  are equal to 0. At the same time,  $W_i$  is penalized by introducing regularization terms for group lasso, thus, ensuring sparsity while ensuring correct correlation.

Given  $N$  variables, the GC between variables is calculated by the above LSTM model to determine the connectivity between variables, so as to generate the causal graph structure. The unweighted digraph  $G = (V, E, A)$  is used to describe the causal relationship between the multivariate variables of cold-rolled steel data, where  $V \in \mathbb{R}^N$  is a set of multivariate node sets,  $E$  is a set of edge sets, used to represent the causal relationship between variables, adjacency matrix  $A \in \mathbb{R}^{N \times N}$  is used to describe the relationship between nodes, and the elements of  $A$  are composed of 0 and 1, indicating whether there is a causal relationship between nodes. Firstly, given sequence  $(x_1, x_2, \dots, x_N) \in \mathbb{R}^{N \times T}$ , for sequence  $x_i$ ,  $i = 1, \dots, N$ , the group lasso causal matrix with other associated sequences  $x_j$ ,  $j = 1, 2, \dots, N$  can be obtained according to Equation (11). If the sub-vector  $W_i^j \neq 0$ , then GC exists for  $x_i$  and  $x_j$ , and the corresponding element in the adjacency matrix  $A$  is set to 1, otherwise it is set to 0, indicating that there is connectivity between the two nodes. Finally, the above steps are repeated to obtain the adjacency matrix  $A$ , and the final graph structure of the cold-rolled steel performance prediction data is obtained.

### 3.3. Embedding Graph Attention Network

The attention mechanism of GAT allows the model to dynamically adjust its attention to node relationships in the face of different scenarios, which enables the model to better adapt to the relationship changes in the data, thereby improving the robustness of the model in the face of different input conditions. However, as a typical whole process, the cold-rolling process requires that the directed path information from the central node to other nodes of the graph can reflect the causal relationship of various influencing factors in

the whole production process. However, GAT tends to only focus on first-order adjacent nodes within the same layer, which is used to aggregate the topological information of the graph and update the characteristics of the central node [14]. Therefore, in order to comprehensively consider the information of nodes on the whole path, so as to learn the continuous feature representation of nodes in the network, a node embedding layer is added before GAT. The node embedding layer utilizes the Node2vec [15] algorithm, which leverages the symmetry in its exploration of various neighborhood paths through a random walk strategy. This process aims to learn the low-dimensional feature space mapping of nodes. Thus, the high-order neighborhood information of nodes is preserved to the greatest extent. At the same time, the path information learned by Node2vec is used in the calculation of GAT's attention score, so that GAT can consider the path information and data characteristics of graph nodes in the process of attention score calculation and information aggregation, so as to more effectively complete the task of cold-rolling performance prediction.

**Node embedding layer:** For a given graph structure  $G = (V, E, A)$ , Node2vec is used for node embedding to learn the continuous feature representation of nodes and map nodes to the low-dimensional feature space, so as to learn the path information of nodes in the network topology. First, given the current node  $v \in V$ , the probability of visiting the next node  $x$  is:

$$P(c_i = x | c_{i-1} = v) = \begin{cases} \frac{\pi_{vx}}{Z} & \text{if } (v, x) \in E \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

where  $\pi_{vx}$  is the non-normalized transition probability between nodes  $x$  and  $v$  in the causal graph of the mechanical properties of cold-rolled steel, and  $Z$  is the normalized constant. Subsequently, let the transition probability  $\pi_{vx} = \alpha_{pq}(t, x) \cdot w_{vx}$ ,  $w_{vx} = 1$  be the weight of the edges of the unweighted directed graph, and  $\alpha_{pq}(t, x)$  be:

$$\alpha_{pq}(t, x) = \begin{cases} \frac{1}{p}, & \text{if } d_{tx} = 0 \\ 1, & \text{if } d_{tx} = 1 \\ \frac{1}{q}, & \text{if } d_{tx} = 2 \end{cases} \quad (13)$$

where  $d_{tx}$  represents the shortest path distance of nodes  $t$  and  $x$ , Node2vec guides the random walk strategy to generate node sequence  $N$  by introducing two parameters  $p$  and  $q$ ,  $p$  and  $q$  are utilized to balance the symmetry between depth-first search and breadth-first search, so as to balance the capture of local structure and global structure of the graph in the learning process and learn the embedding of nodes more comprehensively.

Finally, the generated sequence of nodes is trained using a skip-gram model, which aims to maximize the conditional probability of a context node given a central node. By adjusting the node embedding vector, nodes appearing in the same context are closer in the embedding space. After training, each node will have a corresponding embedding vector, which represents the position of the node in the learning process. These embedding vectors capture the structural relationships and similarities between nodes. For node  $v_i$ , there exists the following embedding vector:

$$e_i = \left( e^1_{v_i}, e^2_{v_i}, \dots, e^d_{v_i} \right) \quad (14)$$

where  $e$  represents the embedding vector of nodes, which is used to map nodes into the low-dimensional feature space, and  $d$  is the embedding dimension.

**GAT layer with path information:** Considering that GAT only considers the first-order neighborhood, it is difficult to characterize the whole process characteristics of cold-rolled steel. Therefore, a GAT model with path information is proposed. Firstly, the path information features learned from the embedded model Node2vec are used to calculate the attention coefficient. The model can consider both the original features of nodes and the path information of nodes in the network topology in the attention mechanism, and assign





$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (21)$$

where  $y_i$  is the target value,  $\hat{y}_i$  is the predicted value, and  $\bar{y}$  is the average of the target value.

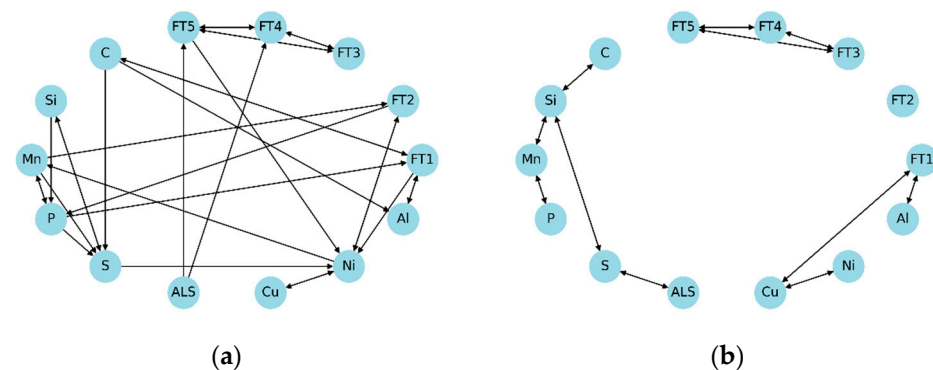
The hyperparameters of NGC-EGAT mainly include the hidden size (hid\_dim) of GAT, the number of heads (num\_heads); Table 1 lists the node embedding dimensions (emb\_dim), random walk step (walk\_len), context size (cont\_size), and number of random walks (walks\_pn) of each node in the Node2vector section.

**Table 1.** NGC-EGAT hyperparameters.

Parameters	Values	Parameters	Values
hid_dim	512	num_heads	14
emb_dim	128	walk_len	10
cont_size	10	walks_pn	10

#### 4.3. Structural Analysis of Neural Causality Graph

In order to verify the feasibility of the proposed NGC graph structure, this paper compares the structure of cold-rolled steel performance prediction graph constructed by Pearson correlation analysis. Figure 3 shows the graph structure based on NGC and Pearson correlation analysis. In Figure 3, the connections between nodes are determined solely by measuring the strength of the linear relationship between the two variables through a graph structure built based on Pearson correlation analysis. As a result, the connections present in the graph structure are limited, covering only a few nodes. However, the graph structure based on NGC analysis not only contains all the node connections in the graph structure based on Pearson correlation analysis, but also contains many potential node relationship connections. Through the analysis, it is found that these potential node relationships are more consistent with the whole process of cold-rolled steel production and the conclusions of previous studies. For example, ALS is an important parameter in the pickling stage, FT5 is an important parameter in the finishing machine stage, and the pickling stage and the finishing machine stage are usually adjacent processes. In addition, the literature [16] pointed out that the content of P, S, C, and Mn show a positive correlation trend with mechanical properties. The literature [17] proposes that simultaneous adjustment of composition and heat treatment conditions is an effective way to optimize mechanical properties, so there is a potential relationship between FT1, FT2, and most chemical components. In addition, in order to quantitatively analyze the validity of the NGC graph structure, this paper conducted experiments on the NGC-EGAT model using two graph structures. Table 2 shows the experimental results of the two graph structures on the NGC-EGAT model. It can be seen from Table 2 that the graph structure based on neural Granger causality has the best prediction accuracy in different evaluation indicators.



**Figure 3.** (a) Causal graph structure; (b) Pearson correlation graph structure.



**Table 2.** Comparative experimental results of different graph structures.

Methods	Target Variables	RMSE	R <sup>2</sup>	MAE
NGC	YS	2.926	0.930	1.859
	TS	2.501	0.947	1.518
	EL	0.596	0.970	0.356
Pearson	YS	2.947	0.926	1.842
	TS	2.538	0.945	1.573
	EL	0.692	0.960	0.481

#### 4.4. Model Performance Comparison

In order to compare the performance of the NGC–EGAT model, this paper selects five advanced methods for comparison. The detailed information of the comparison model is as follows:

1. GATv2 [18] is an improvement of GAT model. The key hyperparameters of GATv2 are set as the hidden size is 128, the number of heads is 7, and the learning rate is 0.01;
2. GCN [19] is a semi-supervised deep learning model designed for graph-structured data. The key hyperparameters of GCN are set to the hidden size of 256 and the learning rate of 0.01;
3. DeepGCNs [20], a variant of traditional GCN, defines a differentiable generalized aggregation function to unify different message aggregation operations, adopts a deeper structure, and solves the problem of information disappearance in graph-structured data. The key hyperparameters of DeepGCNs are set as the size of hidden is 128, the number of model layers is 4, and the learning rate is 0.01;
4. GraphUNet [21] is a U-Net model based on graph-structured data. It realizes the feature learning of graph nodes and hierarchical representation of graph data through the graph convolution operation of hierarchical structure. The key hyperparameters for GraphUNet are set to a hidden size of 256, a U-Net depth of 4, and a learning rate of 0.01;
5. CNN–LSTM [22] is a deep neural network that integrates CNN and LSTM models. The hidden layer size of CNN–LSTM is 64, the time step is 2, the model has 2 layers, and the learning rate is set to 0.01.

The graph structure constructed by NGC proposed in this paper is used for prediction of GNN-based networks. Table 3 shows the experimental results of NGC–EGAT and five comparison models on the test set. As can be seen in Table 3, the proposed NGC–EGAT method is largely due to the existing advanced methods. In addition, GNN model showed better performance than the CNN–LSTM model, which also verified that the graph structure was more able to express the complex dependencies between variables in the cold-rolled steel data set. Compared with other advanced GNN models, NGC–EGAT has achieved the best results under different evaluation indexes. This is mainly because NGC–EGAT not only uses GAT to self-adaptively weight the neighbors of nodes, but also integrates the structural information of the network topology through Node2Vec. The model fully considers the path information from the central node to other nodes, so that the model learns the continuous feature representation of nodes.

In order to visually demonstrate the effectiveness of the NGC–EGAT model, Figure 4a–c shows the fitting results between the predicted values and the real values of the three mechanical properties indexes of NGC–EGAT. It can be seen from the figure that the predicted value of NGC–EGAT fits most of the true value, which also shows the effectiveness of NGC–EGAT.

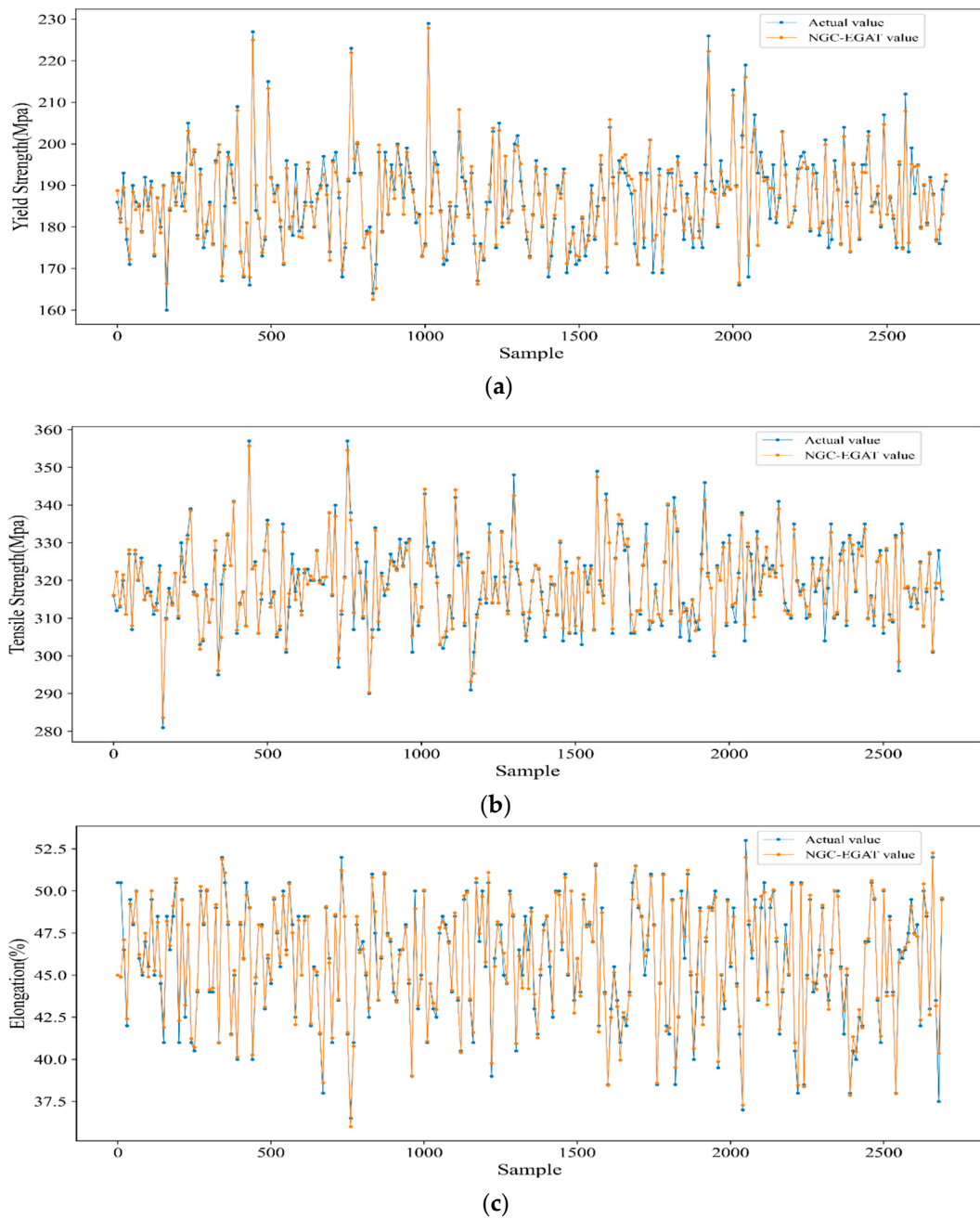


Figure 4. Prediction results of NGC-EGAT: (a) YS, (b) TS, (c) EL.

Table 3. Comparative experimental results of different models.

Target Variables	Metrics	Model					
		GATv2	GCN	DeepGCNs	GraphUNet	CNN-LSTM	NGC-EGAT
YS	RMSE	3.803	3.903	3.035	3.486	4.764	<b>2.926</b>
	$R^2$	0.878	0.871	0.922	0.897	0.831	<b>0.930</b>
	MAE	2.704	2.674	1.921	2.407	2.920	<b>1.859</b>
TS	RMSE	3.135	3.086	2.658	3.122	4.673	<b>2.501</b>
	$R^2$	0.916	0.919	0.940	0.917	0.839	<b>0.947</b>
	MAE	1.969	2.066	1.637	2.117	3.554	<b>1.518</b>
EL	RMSE	0.712	0.699	0.642	0.712	1.221	<b>0.596</b>
	$R^2$	0.957	0.959	0.965	0.957	0.884	<b>0.970</b>
	MAE	0.482	0.480	0.432	0.497	0.930	<b>0.356</b>

In addition, Figure 5 presents the error rates of different models within the range of actual industrial production error requirements. The actual industrial production error ranges of YS, TS, and EL are  $\pm 10$  Mpa,  $\pm 10$  Mpa, and  $\pm 3$ , respectively. By looking at Figure 5, it is clear that the NGC-EGAT model has the lowest error rate in these areas. It is further confirmed that NGC-EGAT is more suitable for actual industrial production needs.

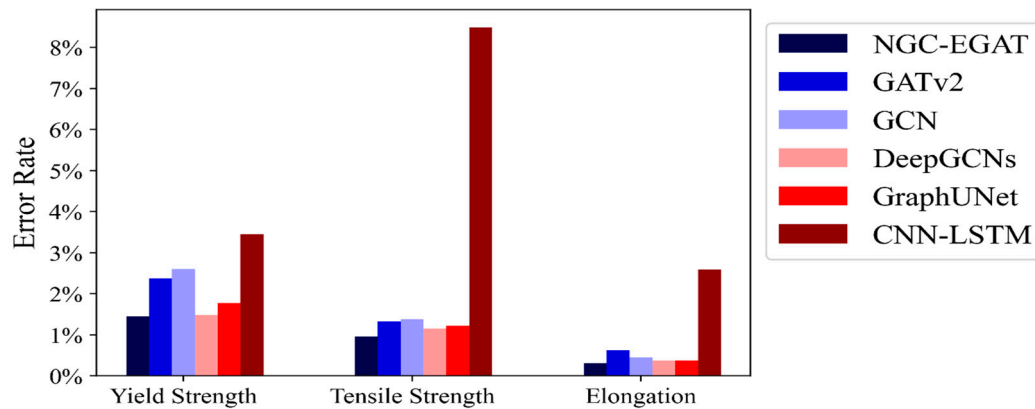


Figure 5. Prediction error rates of different models.

#### 4.5. Ablation Experiment

To verify the effectiveness of each module in the proposed method for model performance, ablation studies were performed on the test set. In the NGC-GCN method, GCN is used to replace GAT, and the effectiveness of GAT is verified. In NGC-GAT method, Node2vec node embedding is removed, and the NGC graph structure training test GAT is used directly to verify the effectiveness of using Node2vec node embedding to integrate structural information. The experimental results are shown in Table 4. It can be clearly observed from Table 4 that compared with the GCN method, the GAT method shows significant improvement in different mechanical properties. After removing node embeddings, the performance of the NGC-GAT model decreases significantly compared with NGC-EGAT model, which also verifies the importance of integrating structural information, and also proves that integrating structural information through Node2vec node embeddings is more in line with the requirements of cold-rolled steel whole-process production process.

Table 4. Ablation experiment results.

Methods	Target Variables	RMSE	$R^2$	MAE
NGC-GCN	YS	3.903	0.871	2.674
	TS	3.086	0.919	2.066
	EL	0.699	0.959	0.480
NGC-GAT	YS	3.827	0.876	2.466
	TS	3.029	0.922	1.911
	EL	0.685	0.960	0.430
NGC-EGAT	YS	<b>2.926</b>	<b>0.930</b>	<b>1.859</b>
	TS	<b>2.501</b>	<b>0.947</b>	<b>1.518</b>
	EL	<b>0.596</b>	<b>0.970</b>	<b>0.356</b>

## 5. Conclusions

In this paper, an NGC-EGAT model for predicting mechanical properties of cold-rolled steel is presented. Firstly, the model employs the NGC algorithm to systematically construct the graph structure for predicting mechanical properties based on data related to cold-rolled steel. Subsequently, the Node2vec node embedding method utilizes a random walk strategy to explore diverse neighborhood paths of nodes, extract the path information

features of graph nodes, and fuse the path information features with the original data features. After that, the GAT is used to learn the dependencies between multiple variables and the path information of the graph nodes simultaneously. Finally, the validity and robustness of the NGC–EGAT method are verified by experiments on real data sets, and the feasibility and superiority of NGC–EGAT in predicting the properties of cold-rolled steel are also proved.

**Author Contributions:** Conceptualization, X.L., R.G. and Q.Z.; methodology, X.L., R.G., Q.Z. and X.T.; software, R.G.; validation, R.G.; formal analysis, X.L., R.G., Q.Z. and X.T.; investigation, X.L., R.G. and Q.Z.; resources, R.G. and X.T.; data curation, X.L., Q.Z. and X.T.; writing—original draft preparation, R.G.; writing—review and editing, X.L., R.G. and Q.Z.; visualization, R.G.; supervision, X.L., Q.Z. and X.T.; project administration, X.L. and Q.Z.; funding acquisition, X.L., Q.Z. and X.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the National Natural Science Foundation of China 62063021 (Research on HMS Scheduling Optimization and Control and Intelligence System in Manufacturing IoT Environment), and the National Natural Science Foundation of China 62162040 (Research on Terrain Representation and Dissemination Adaptive Scheduling Strategy for Large-scale Social Network Influence Adaptability).

**Data Availability Statement:** Data available on request from the authors.

**Conflicts of Interest:** Author Xiaoyang Luo was employed by the company Gansu JISCO Group, Hongxing Iron & Steel Co., Ltd., Carbon Steel & Thin Slab Rolling Plant. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Diao, Y.; Yan, L.; Gao, K. A strategy assisted machine learning to process multi-objective optimization for improving mechanical properties of carbon steels. *J. Mater. Sci. Technol.* **2022**, *109*, 86–93. [[CrossRef](#)]
- Jacobs, L.J.M.; Atzema, E.H.; Moerman, J.; de Rooij, M.B. Quantification of the influence of Anisotropic Plastic Yielding on Cold Rolling Force. *J. Mater. Process. Technol.* **2023**, *319*, 118055. [[CrossRef](#)]
- Sheng, H.; Wang, P.; Tang, C. Predicting mechanical properties of cold-rolled steel strips using micro-magnetic ndt technologies. *Materials* **2022**, *15*, 2151. [[CrossRef](#)] [[PubMed](#)]
- Kano, M.; Nakagawa, Y. Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry. *Comput. Chem. Eng.* **2008**, *32*, 12–24. [[CrossRef](#)]
- Li, W.; Gu, J.; Deng, Y.; Mu, W.; Li, J. New comprehension on the microstructure, texture and deformation behaviors of UNS S32101 duplex stainless steel fabricated by direct cold rolling process. *Mater. Sci. Eng. A* **2022**, *845*, 143150. [[CrossRef](#)]
- Li, F.; He, A.; Song, Y.; Wang, Z.; Xu, X.; Zhang, S. Deep learning for predictive mechanical properties of hot-rolled strip in complex manufacturing systems. *Int. J. Miner. Metall. Mater.* **2022**, *30*, 1093–1103. [[CrossRef](#)]
- Yan, Y.F.; Lü, Z.M. Multi-objective quality control method for cold-rolled products oriented to customized requirements. *Int. J. Miner. Metall. Mater.* **2021**, *28*, 1332–1342. [[CrossRef](#)]
- Liu, X.; Cong, Z.; Peng, K.; Dong, J.; Li, L. DA-CBGRU-Seq2Seq based soft sensor for mechanical properties of hot rolling process. *IEEE Sens. J.* **2023**, *23*, 14234–14244. [[CrossRef](#)]
- Xu, Z.W.; Liu, X.M.; Zhang, K. Mechanical properties prediction for hot rolled alloy steel using convolutional neural network. *IEEE Access* **2019**, *7*, 47068–47078. [[CrossRef](#)]
- Chen, H.; Jiang, Y.; Zhang, X.; Zhou, Y.; Wang, L.; Wei, J. Spatio-Temporal Graph Attention Network for Sintering Temperature Long-Range Forecasting in Rotary Kilns. *IEEE Trans. Ind. Inform.* **2022**, *19*, 1923–1932. [[CrossRef](#)]
- Sun, B.; Lv, M.; Zhou, C.; Li, Y. A multimode structured prediction model based on dynamic attribution graph attention network for complex industrial processes. *Inf. Sci.* **2023**, *640*, 119001. [[CrossRef](#)]
- Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv* **2021**, arXiv:1710.10903.
- Tank, A.; Covert, I.; Foti, N.; Shojaie, A.; Fox, E.B. Neural granger causality. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 4267–4279. [[CrossRef](#)] [[PubMed](#)]
- Yang, Y.; Wang, X.; Song, M.; Yuan, J.; Tao, D. Spagan: Shortest path graph attention network. *arXiv* **2021**, arXiv:2101.03464.
- Grover, A.; Leskovec, J. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 855–864.
- Choudhury, A. Prediction and analysis of mechanical properties of low carbon steels using machine learning. *J. Inst. Eng. (India) Ser. D* **2022**, *103*, 303–310. [[CrossRef](#)]
- Li, X.; Zheng, M.; Yang, X.; Chen, P.; Ding, W. A property-oriented design strategy of high-strength ductile RAFM steels based on machine learning. *Mater. Sci. Eng. A* **2022**, *840*, 142891. [[CrossRef](#)]

18. Brody, S.; Alon, U.; Yahav, E. How attentive are graph attention networks? *arXiv* **2021**, arXiv:2105.14491.
19. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.
20. Li, G.; Xiong, C.; Thabet, A.; Ghanem, B. Deepergcn: All you need to train deeper gcns. *arXiv* **2020**, arXiv:2006.07739.
21. Gao, H.; Ji, S. Graph u-nets. In *International Conference on Machine Learning*; PMLR: Long Beach, CA, USA, 9–15 June 2019; pp. 2083–2092.
22. Kim, T.Y.; Cho, S.B. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy* **2019**, *182*, 72–81. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.