MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **89** (2023) 199–222
ISSN: 0340–6253
doi: 10.46793/match.89-1.199H

Regression Prediction of Coupling Reaction Yield Based on Attention–Driven Convolutional Neural Network

Hexun Hou[a], Hengzhe Wang[a], Yanhui Guo[a], Puyu Zhang[b], Lichao Peng[c,*], Xiaohui Yang[a,*]

[a] *Henan Engineering Research Center for Artificial Intelligence Theory & Algorithms, School of Mathematics and Statistics, Henan University, Kaifeng, China, 475004*

[b] *College of Chemistry and Chemical Engineering, Henan University, Kaifeng, China, 475004*

[c] *National & Local Joint Engineering Research Center for Applied Technology of Hybrid Nanomaterials, Henan University, Kaifeng, China, 475004*

plc@henu.edu.cn, xhyanghenu@163.com

(Received May 17, 2022)

Abstract

The traditional method to improve the yield of Buchwald-Hartwig cross coupling reaction is to change the reactants or reaction conditions, but the reaction has many problems, such as harsh reaction conditions, complex synthetic route. In 2018, Doyle reported a yield prediction method based on random forest in ***Science***. However, the predicted value of the regression tree in the random forest is the average value of the target variable of the leaf node, which treats the feature as equally important. We focused on the important characteristic information in order to obtain a more accurate yield prediction value. Therefore, it is of interest to apply some advanced deep learning methods to the performance prediction of chemical reactions, during which less training data may be required.

*Corresponding author.

1 Introduction

Buchwald-Hartwig cross coupling reaction, also known as Buchwald-Hartwig amination reaction, is a reaction for the preparation of aromatic amines, mainly through the coupling of aryl halides or aryl sulfonates with amines. This reaction can directly form C-N bonds, and the resulting substances are mostly nitrogen-containing compounds. It is widely used in the processing of natural products and the preparation of medical articles.

In order to improve the yield of Buchwald-Hartwig cross coupling reaction, researchers in the field of chemistry have been committed to improving reaction conditions such as ligands and additives in the reaction [1–4]. However, the current Buchwald-Hartwig cross coupling reaction is obviously facing corresponding shortcomings. For example, The reaction conditions are harsh, the synthetic route is complex, the reaction time is long, and the solvent pollutes the environment [5]. Therefore, the design of green, simple and efficient chemical synthesis method is the focus of research on Buchwald-Hartwig cross coupling reaction.

Since 2015, the number of chemical publications related to artificial intelligence and machine learning has increased significantly. Despite the high number of literature studies in molecular chemistry, biochemistry and other fields, the intersection of organic synthesis and machine learning has been limited. It was found that previous research focused on synthetic planning, or the use of machine learning algorithms to predict the products of chemical reactions, while direct prediction of the performance of a given reaction (such as the specific value of yield) was very rare [6–9].

In 2018, Doyle et al. [10] first proposed the use of random forest to predict the reaction performance of C-N cross coupling, and the yield of Buchwald-Hartwig cross coupling reaction product was predicted with a goodness of fit (R^2) as high as 0.92. However, this method is point prediction based on feature descriptors and lacks feature learning. Based on this, our team previously proposed intelligent yield prediction based on quantile regression forest and deep forest respectively [11, 12].

With the rapid development of artificial intelligence, the application of some deep learning methods to organic synthesis is also worth pay-

ing attention to. Since 2006, deep learning has performed well in many fields, including image recognition, speech recognition and natural language processing. Convolutional neural network (CNN) [13] is a famous deep learning model, which is named after the convolutional operation is introduced into the network.

In 2021, Zhu et al. [14] used the data set reported by Doyle to predict the reaction performance of Buchwald-Hartwig cross coupling using an optimized deep convolutional neural network (DCNN). However, on the one hand, the processing method in this paper increases the system error and operation cost for the model. On the other hand, it also violates the learning characteristic of deep neural network from original data to a certain extent. Last but not lease, one-dimensional convolutional neural network has been well applied in data analysis.

Considering that the features extracted from the network may have different contributions to the current yield prediction task, it is necessary to assign weight to the feature information. Therefore, this paper first proposed a coupling reaction yield regression prediction method based on attention-driven convolutional neural network. During the regression prediction of yield, the weight of important features was increased, while the weight of unnecessary features was reduced, and the feature extraction ability of the network was improved. The aim is to obtain a more accurate yield prediction value, and screen the reaction based on the prediction results, which can be used for reference by chemists.

The model is improved on the database of Buchwald-Hartwig coupling reaction (Fig. 1) obtained by Doyle et al. Compared with Zhu et al., the network structure designed in this paper is very simple, using only three convolutional layers, three pooling layers, one attention layer and two full connection layers, with short training time and original data. Compared with the previous work of Doyle and our team, this paper focuses more on the importance of features. In general, this paper combines attention mechanism and convolutional neural network, and makes full use of important feature descriptor information to achieve regression prediction and analysis of chemical yield. The flow chart is shown in Figure 2.

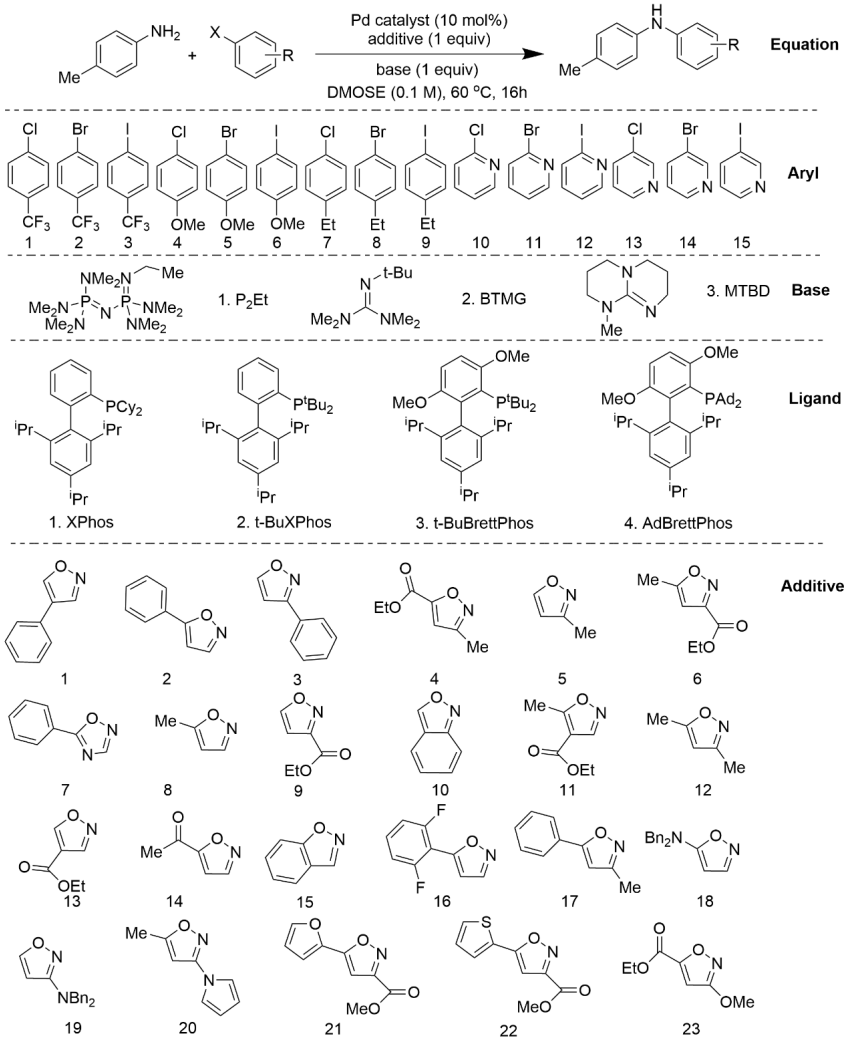

Figure 1. Reaction formula and structural formula for Buchwald-Hartwig coupling reaction.

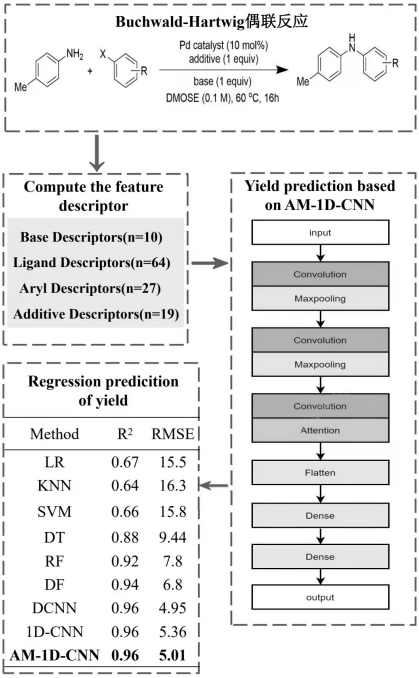

Figure 2. Regression prediction flow chart of coupling reaction yield.

2 Method

2.1 Convolutional neural network (CNN)

CNN is a feedforward neural network. According to different data types of network processing, it is divided into one-dimensional convolutional neural network (1D-CNN) for processing sequential data, two-dimensional convolutional neural network (2D-CNN) for processing image data or text data, and three-dimensional convolutional neural network (3D-CNN) for processing medical image or video data. As the feature descriptor data extracted in this paper is one-dimensional sequential data, 1D-CNN is selected for data processing.

In practical applications, there are many application scenarios of 2D-CNN, which we are familiar with. However, since 1D-CNN is slightly different from 2D-CNN, the basic principle of 1D-CNN is introduced.

2.1.1 Convolution layer

The convolution layer is the process of feature extraction through the convolution operation of the filter. The data of each convolution layer is obtained by the convolution operation of the input data and the convolution kernel. A convolution kernel can only generate one set of feature data, and it needs to set several more convolution kernels to extract multiple sets of feature data.

Since the input of 1D-CNN is a one-dimensional vector, the convolution kernel is also one-dimensional. We take the number of input channels of the convolution layer as 1, and the number of output channels as 1, to show the schematic diagram of the convolution operation, as shown in Figure 3 below:

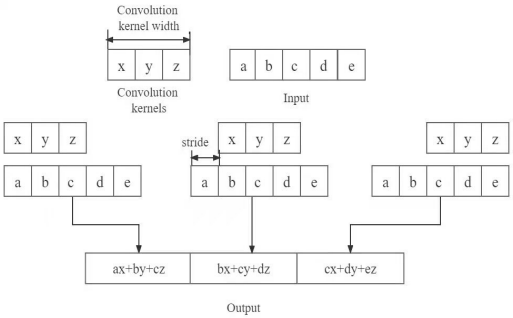

Figure 3. Operation diagram of convolution.

As can be seen from Figure 3, feature extraction using the convolution kernel is to extract small region information one by one, and the whole region information is obtained by the convolution kernel moving at a certain step and carrying out multiple convolution operations. The convolution calculation process is shown in equation (1) :

$$h_l^{out} = g\left(\sum \omega_l^i * x_l^i + b_l^i\right), \tag{1}$$

Where, ω_l^i represents the weight parameter of the i-th convolution kernel at layer l, x_l^i represents the input data of the i-th convolution region at layer l, b_l^i represents the bias parameter of i-th convolution region at layer

l, g$(\cdot)$ is the activation function (to be explained in detail in the next section), and h_l^{out} represents the output value after convolution operation at layer l.

2.1.2 Incentive layer

It can be seen from equation (1) of convolution operation that convolution operation is still a linear weighted sum process. In a real prediction model, we need the network to learn nonlinear features and map feature data into a nonlinear space. For this reason, researchers set the nonlinear function as the activation function. The activation functions used in this paper are ReLU function and Softplus function.

The analytical formula of ReLU function is $f(x) = max(0, x)$, where x is the eigenvalue, $f(x)$ is the value of the activation function, and the geometric figure is:

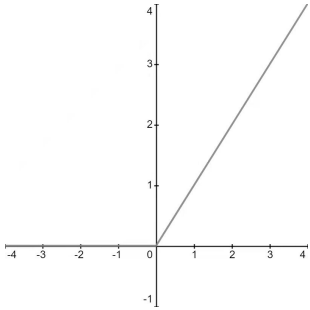

Figure 4. Geometry of ReLU function.

The ReLU function activates neurons on the positive interval and inhibits neurons on the negative interval, which alleviates the problems of gradient disappearance and gradient explosion in the positive interval to a certain extent. Moreover, the calculation speed and convergence speed are relatively fast.

The mathematical expression of Softplus function is $\zeta(x) = \log(1 + e^x)$ where x is the eigenvalue, $\zeta(x)$ is the activation function value, and the geometric figure is:

Softplus can be regarded as the smoothing of ReLU and conforms to

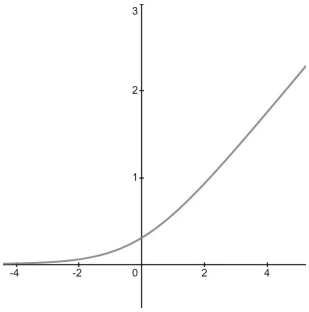

Figure 5. Geometry of Softplus function.

the activation model of brain neurons. In addition, Softplus function is a nonlinear activation function with a range of $(0, \infty)$, which is exactly consistent with the yield prediction interval in this paper. Therefore, Softplus activation function is used in the last output layer of this paper to perform regression prediction of coupling reaction yield.

2.1.3 Pooling layer

The pooling layer is usually behind the convolutional layer, which is used to reduce the number of network parameters to be trained, prevent over-fitting, reduce the number of features, and improve the fault tolerance of the model.

The common pooling methods include maximum pooling and average pooling. Maximum pooling is to extract the maximum value of several small regions segmented by a certain step as the representative value of the small region. Similarly, average pooling is to take the average value of a small region as the representative value.

In this paper, maximum pooling with step 3 is used, as shown in Figure 6 below:

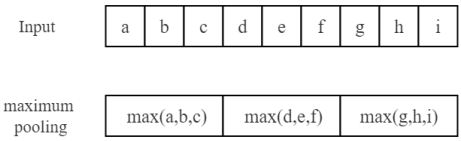

Figure 6. Operation diagram of maximum pooling.

As shown in Figure 6, the pooling operation divides the input data (that is the feature data extracted from the convolutional layer) into several small regions according to the step size, and there is no overlap between the small regions.

The pooling operation is performed on the data in each small region, and the pooling calculation operation is shown in equation (2) :

$$z_l^{out} = g(x_l^{i(p)}, x_l^{i(p+1)}, x_l^{i(p+2)}, \cdots), \tag{2}$$

Where, $x_l^{i(p)}$ represents the p-th value of the i-th region array at layer l, g$(\cdot)$ is the maximum pooling function or average pooling function, and z_l^{out} represents the output value after the pooling operation at layer l.

2.1.4 Full connection layer

The full-connection layer integrates feature extraction and regression prediction into one framework and optimizes the network globally through forward propagation and back propagation algorithms.

Since the shape of the convolutional layer and pooling layer is inconsistent with that of the fully connected layer, before the fully connected layer, we need to splicing and flattening the feature vector processed by multiple channels of the pooling layer into a one-dimensional vector, which meets the data format required by the input of the fully connected layer. Then it is connected with the full connection layer and the output layer to complete the regression prediction of coupling reaction yield, and the specific predicted value is given. The schematic diagram of the full connection layer is shown in Figure 7 below:

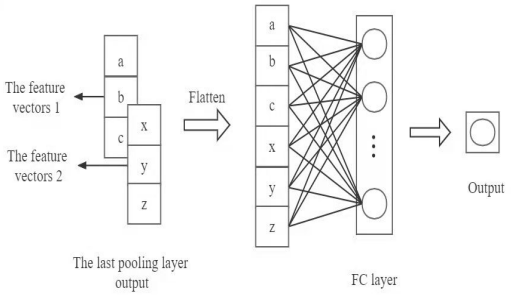

Figure 7. Schematic diagram of full connection layer.

2.2 Attention mechanism

The calculation formula of attention mechanism is shown.

$$\begin{aligned} e_t &= ReLU(W * X + b), \\ \alpha_t &= \frac{exp(e_t)}{\sum_{i=1}^{t} e_i} = softmax(e_t), \\ s_t &= \alpha_t X, \end{aligned} \tag{3}$$

Where, W is the weight, b is the bias, e_t is the attention distribution value at time t, α_t is the normalized weight value at time t, and s_t is the output value of the attention mechanism at time t.

As can be seen from equation (3), the attention model is essentially a fully connected layer. The fully connected layer of Softmax activation function can directly calculate the attention weight value, and then multiply it by the input of Dense to get the eigenvalue with the attention weight.

The specific processing process is shown in Figure 8 below:

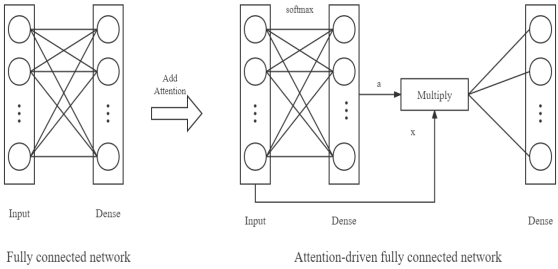

Figure 8. Schematic diagram of attention-driven full connection layer.

2.3 Regression prediction model of coupling reaction yield

2.3.1 One-dim convolutional neural network (1D-CNN)

Scholars in the field of deep learning have developed many classic CNN models and performed well in natural language processing, image recognition, speech recognition and other aspects. However, in specific studies, especially in one-dimensional data analysis, hidden layers can be simply stacked as required to build a network structure suitable for processing specific data and realize intelligent prediction of overall effectiveness.

Based on this idea, the hidden layer of 1D-CNN model constructed through experiments in this paper includes three convolution layers and three pooling layers, and two full-connection layers are added before the output layer for nonlinear mapping. As the whole network structure is simple, it is called lightweight 1D-CNN model. Through the use of convolution layer and pooling layer, 1D-CNN adaptively extracts features from feature descriptors and flattens the extracted features. Finally, two fully connected layers are connected to complete the construction of 1D-CNN model. The network topology is shown in Figure 9.

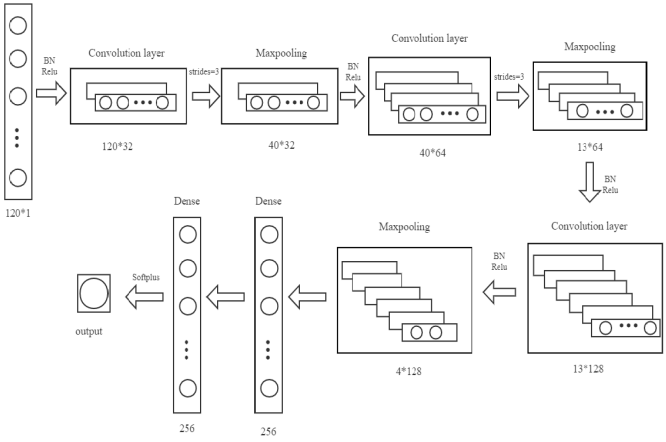

Figure 9. 1D-CNN structure diagram.

2.3.2 Attention–driven convolutional neural network (AM-1D-CNN)

The regression prediction model of yield of AM-1D-CNN coupling reaction constructed in this paper is to integrate the attention mechanism into the 1D-CNN to make it adaptively acquire the importance degree of features. According to this importance degree, the features that can represent the yield information can be enhanced and invalid features can be suppressed so that the network can obtain more prediction information. Its network structure is shown in Figure 10:

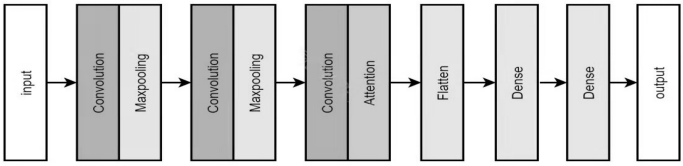

Figure 10. AM-1D-CNN structure diagram.

The attention mechanism layer of the AM-1D-CNN model is placed after the convolution layer of the third layer, and the third pooling layer is replaced by the attention mechanism layer. This is because the pooling

layer simply extracted the maximum feature value in feature screening, while the goal of this paper is to make the network pay attention to important features and reduce the loss of feature information as much as possible.

The entire network module in a series of convolution, pooling after operation so as to obtain the deepest descriptor information, and then use the attention mechanism technique to obtain weighted and adaptive to the characteristics of information filtering, and labeled with the characteristic of the filtered, then connect with good character sequence to link to all layers in the study again. Finally, the characteristic data extracted from the model are sent to Softplus regression function to achieve end-to-end coupling reaction yield intelligent prediction. The network parameters are shown in Table 1.

Table 1. Main parameters setting of network training process.

Main operation	Parameter
Dropout	0.3
Optimization Algorithm	Adam
Batch size	32
Epoch	350

3 Experiments and analysis

3.1 Data set

Taking Buchwald-Hartwig amination reaction as the model reaction for data processing, based on the prediction of yield of Buchwald-Hartwig amination reaction by random forest model published by Doyle. In this paper, 23 isoxazole additives, 15 aryl and heteroaryl halides, 4 Buchwald ligands and 3 bases have been obtained by high throughput experiment (HTE).

The schematic diagram of Buchwald-Hartwig amination reaction between reactants and reaction products is shown in Figure 11:

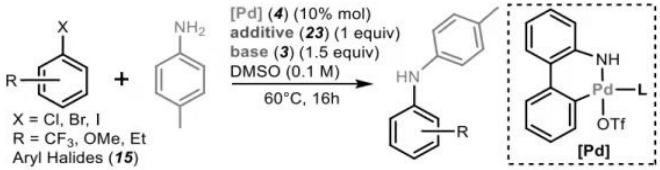

Figure 11. Diagram of Buchwald-Hartwig amination reaction.

First of all, according to the thought of permutation and combination, a total of 4140 kinds of reaction type, remove invalid response, remaining 3960 valid responses, and record the reaction yield.

Secondly, the chemical software Spartan was used to extract the characteristic descriptors of the reaction components, so that the compounds involved in each effective reaction could be described by the characteristic descriptors.

According to the complex operation process, a total of 120 feature descriptors were extracted, including molecular, atomic and vibration descriptors. Finally, we used 120 characteristic descriptors corresponding to reaction components and 3960 effective reaction corresponding samples.

3.2 Evaluation indicators

After the experiment, in order to measure the prediction performance of the model and evaluate the generalization performance of the learner, a measure needs to be set in advance. In this paper, goodness of fit and root mean square error are used to measure the regression prediction results of the model and compare the advantages and disadvantages of each learner.

Goodness of fit (R^2) is the degree of fitting between regression prediction results and experimental values, and its formula is as follows:

$$R^2 = 1 - \frac{\sum(y_i - \widehat{y}_i)^2}{\sum(y_i - \overline{y})^2}, \tag{4}$$

where y_i is the observed value of the sample, $\widehat{y}_i$ is the regression predicted value and $\overline{y}$ is the average value of the sample.

The value range of R^2 is [0,1]. The closer R^2 is to 1, the closer the

regression prediction point is to each observation point, and the better fitting degree is between the regression predicted value and the actual observed value.

Root Mean Square Error (RMSE) is the square root of the Mean Square Error(MSE) and represents the sample standard deviation of the difference between the predicted value and the true value. Its formula is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}, \tag{5}$$

where y_i is the observed value of the sample, $\widehat{y}_i$ is the regression predicted value, and n is the number of samples.

In addition, RMSE is of the same magnitude as data, making it easier to sense data. The smaller the RMSE value is, the smaller the regression prediction deviation is, indicating that the prediction effect of the model is better.

3.3 Comparison between the proposed method and traditional machine learning algorithm

In order to verify the adaptability of the proposed method in the Buchwald-Hartwig cross-coupling reaction and the improvement of the method on the 1D-CNN model, the training set and test set were divided into 7:3. The advantages and disadvantages of AM-1D-CNN, 1D-CNN, random forest and other traditional machine learning methods for regression prediction of cross-coupling reaction yield are explored, and the results are shown in Table 2.

As can be seen from Table 2, the AM-1D-CNN proposed in this paper has some improvement in the regression prediction of cross-coupling reaction yield compared with the 1D-CNN model alone, indicating that attention technology can be used to predict the performance of chemical reactions. Although the R^2 value given in the table is 0.96, this value is the result of rounding. Compared with the original value, the R^2 value of AM-1D-CNN is higher than that of 1D-CNN. Compared with the RMSE,

Table 2. Yield regression prediction results of different models.

Method	R^2	RMSE
Linear Regression	0.67	15.5
K-Nearest Neighbor	0.64	16.3
Support Vector Machine	0.66	15.8
Decision Tree	0.88	9.44
Random Forests	0.92	7.8
Deep Forests	0.94	6.8
DCNN	0.96	4.95
1D-CNN	0.96	5.36
AM-1D-CNN	**0.96**	**5.01**

the RMSE of the proposed method can reach 5.01, which is significantly lower than that of 1D-CNN and other machine learning methods, indicating that the proposed method has better performance in predicting the yield regression of Buchwald-Hartwig cross coupling reaction.

In addition, compared with DCNN proposed by Zhu et al. in 2021, the two methods have equivalent effects in the regression prediction of yield of Buchwald-Hartwig cross coupling reaction. The advantages of this paper lie in the simple structure of the designed network, the short training time, and the use of original data and the classic pooling layer of convolutional neural network.

3.4 Analysis of influencing factors of coupling reaction yield

When 120 feature descriptors were directly imported into the neural network for regression intelligent prediction, it was not clear which feature descriptors had a great impact on the yield of Buchwald-Hartwig cross coupling reaction, which was very important for chemical problems. Therefore, in this paper, principal component analysis was selected to extract the principal components of the data set, transform the original data into linearly independent, reduce the high dimension of the data, and most importantly, analyze the factor contribution rate of feature descriptors to find out the characteristic factors that affect the yield of Buchwald-Hartwig cross coupling reaction.

3.4.1 Principal component analysis

Principal Component Analysis (PCA) is to transform the variables that may be correlated in the original data into a group of linearly unrelated variables through orthogonal transformation, and select a few index variables from the transformed variables to reflect the information of the original data according to the research needs.

Table 3 shows the algorithm steps of principal component analysis [31].

Table 3. Principal component analysis algorithms.

Input:	sample set $D = x_1, x_2, \cdots, x_n$, low dimensional space dimension d'.
Process:	Centralization of all samples: $x_i \leftarrow x_i - \frac{1}{n}\sum_{i=1}^{n} x_i$; Calculate the covariance matrix of the sample XX^T; Do the eigenvalue decomposition of the covariance matrix XX^T; Take the eigenvector $\omega_1, \omega_2, \cdots, \omega_{d'}$ corresponding to the largest d' eigenvalues.
Output:	the projection matrix $W = (\omega_1, \omega_2, \cdots, \omega_{d'})$.

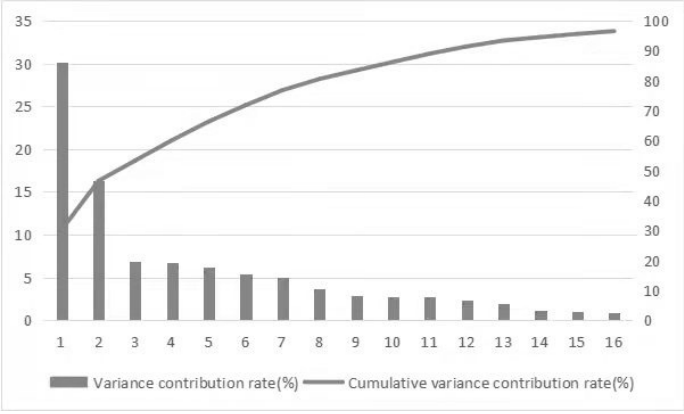

Figure 12. The combination diagram of variance contribution rate and cumulative variance contribution rate of 16 principal components.

When principal component analysis was used to extract common fac-

tors, 16 principal components were extracted according to the eigenvalue greater than 1. At this time, the cumulative variance contribution rate reached 96.49%, indicating that most information of the original data was retained by the extracted common factors.

3.4.2 Regression prediction of yield of cross-coupling reaction based on principal component

The extracted 16 principal component data were saved and a new data table was created to correspond the principal component data with the yield. The AM-1D-CNN model constructed above was used to perform regression prediction of Buchwald-Hartwig cross coupling reaction yield for the 16 principal component data.

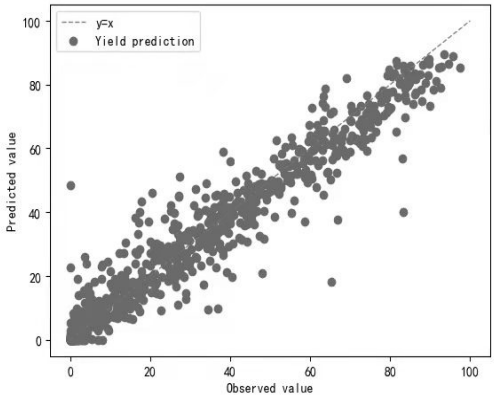

Figure 13. A fitting diagram of predicted yield and observed yield.

R^2 of AM-1D-CNN model on 16 principal component data is 0.93, and the RMSE is 7.13. The results show that the lightweight AM-1D-CNN model proposed in this paper is not only suitable for the case of large amount of data. When the amount of data is greatly reduced, the model can still extract data information well as long as the existing data can well represent the characteristic information. It also performed well in regression prediction of yield of Buchwald-Hartwig cross coupling reaction.

3.4.3 Analysis of influencing factors

In order to further investigate how different feature descriptors affect the yield of Buchwald-Hartwig cross-coupling reaction, the component matrices of each principal component are calculated to track the model analysis and trace back to the complete original feature descriptors, to illustrate the correlation between the extracted principal component and the original variable.

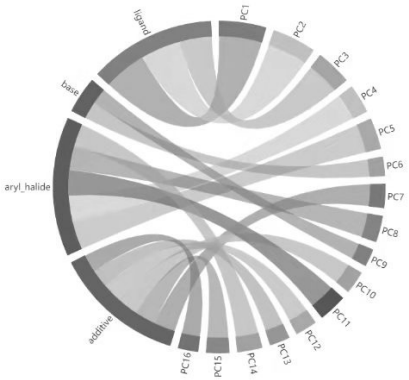

Figure 14. Diagram of correlation between principal components and primary variables.

As can be seen from Figure 14, the first principal component, second principal component and third principal component are mainly related to ligand descriptors in catalyst, indicating that the change of catalyst ligand has the greatest impact on yield. Ligand plays an important role in the whole cross-coupling reaction, and its complex with the central metal determines the catalytic activity of the reaction system. At the beginning of the reaction, the ligand binds with the Pd catalyst precursor to release the active catalyst LPd(0). Moreover, the ligand can increase the electron density of Pd atom and promote the oxidation addition reaction with halogenated aromatics.

Although ligands play an important role in the success of coupling reactions, the electronic properties of aryl and heteroaryl halides and additives still seem to be important. In addition, the additive was correlated with

the six principal components, indicating that the change of additive had a great influence on the yield.

3.5 Out–of–sample prediction

Out-of-sample prediction refers to the use of models to predict sample values beyond the training data. If effective out-of-sample prediction can be achieved, the results of the model presented in this paper will be able to predict the effect of a novel isoxazole additive on the results of the Buchwald-Hartwig cross coupling reaction. In addition, the reaction performance of a coupled reaction can be predicted prior to preparation because calculated characteristic descriptor information is used, avoiding the need to obtain any spectral information. To this end, we evaluated whether the results of twenty additives could be used to predict the results of three other different additives.

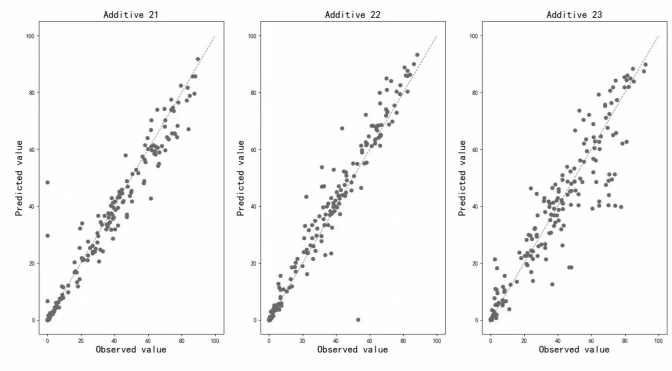

Figure 15. A scatter plot of real and predicted values.

As can be seen from Figure 15, each point in the out-of-sample predicted scatter plot of the three additives is almost distributed near the fitting line, indicating that the predicted value of the model is very close to the real measured value in the test set, and the sum of squares of residuals is small. The regression prediction result of the coupling reaction yield of isoxazole additives under the AM-1D-CNN model is good.

In order to evaluate the effect of out-of-sample prediction more accurately, we calculated the R^2 and RMSE of out-of-sample regression pre-

Table 4. Prediction results of additive out-of-sample regression.

Type of additives	R^2	RMSE
Additive 21	0.94	6.50
Additive 22	0.93	6.81
Additive 23	0.90	8.52
Mean	0.923	7.277

diction for the three additives, and the results are shown in Table 4. The mean R^2 and RMSE of the additive regression predicted by the three out-of-sample tests were 0.923 and 7.277 respectively. We concluded that the effects of the three additives in the test set on the yield of Buchwald-Hartwig cross coupling reaction had no significant systematic deviation from the predicted results of the AM-1D-CNN model.

4 Conclusion

In this paper, 120 feature descriptors extracted from additives, aryls, bases and ligands were used as inputs. Based on the characteristics of data heterogeneity, 1D-CNN method was first proposed to make regression prediction of the yield of Buchwald-Hartwig cross coupling reaction. Then, in order to make the model pay more attention to important feature information, to suppress unimportant features, this paper proposes to integrate the attention mechanism into 1D-CNN, and the constructed AM-1D-CNN model can adaptively obtain the importance degree of features and enhance the prediction performance of the network. The main purpose of this paper is to predict and analyze the yield of coupling reaction intelligently and get more accurate yield value.

In order to explore the influencing factors of yield prediction, AM-1D-CNN method based on PCA was proposed in this paper. By calculating the component matrix of each principal component to track the model, it was found that the change of ligand had the greatest impact on the reaction yield, while the change of additives had a greater impact on the reaction yield. Finally, this paper tests the performance of the model on the out-of-sample data set, and proves that the AM-1D-CNN model can predict

the performance of unknown additives, thus improving the efficiency of chemical research and reducing the pressure of chemical workers.

References

[1] P. K. Michael, W. C. Ramali, B. G. Nancy, C. Bae, Synthesis of anion conducting polymer electrolyte membranes by Pd-catalyzed Buchwald-Hartwig amination coupling reaction, *Tetrahedron* **75** (2019) 4150–4155.

[2] X. Li, C. Zhang, C. Wang, W. Ye, Q. Zhang, Z. Zhang, J. Su, Y. Chen, H. Tian, Modular synthesis of (C-10 to C-13)-substituted-9,14-diaryl-9,14-dihydrodibenzo[a,c]phenazines via a subsequent Buchwald-Hartwig amination and C-H amination strategy, *Chem. Commun.* **56** (2020) 2260–2263.

[3] T. Taeufer, J. Pospech, Palladium-catalyzed synthesis of N, N-dimethylanilines via Buchwald-Hartwig amination of (hetero)aryl triflates, *J. Org. Chem.* **85** (2020) 7097–7111.

[4] M. Kucharek, A. Danel, Palladium-catalyzed amino group arylation of 1,3-disubstituted 1H-pyrazol-5-amine based on Buchwald-Hartwig reaction, *Chem. Heterocyc. Comp.* **57** (2021) 633–639.

[5] M. M. Heravi, Z. Kheilkordi, V. Zadsirjan, Masumeh Heydari, M. Malmir, Buchwald-Hartwig reaction: An overview, *J. Org. Chem.* **861** (2018) 17–104.

[6] M. A. Kayala, C. A. Azencott, J. H. Chen, P. Baldi, Learning to predict chemical reactions, *J. Chem. Inf. Model.* **51** (2011) 2209-–2222.

[7] J. N. Wei, D. David, A. Alán, Neural networks for the prediction of organic chemistry reactions, *ACS Cent Sci.* **2** (2016) 725–732.

[8] C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, K. F. Jensen, Prediction of organic reaction outcomes using machine learning, *ACS Cent Sci.* **3** (2017) 434–443.

[9] B. Liu, R. Bharath, K. Prasad, S. Jade, G. Joseph, L. N. Quang, H. Stephen, S. Jack, W. Paul, P. Vijay, Retrosynthetic reaction prediction using neural sequence-to-sequence models, *ACS Cent Sci.* **3** (2017) 1103–1113.

[10] D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, A. G. Doyle, Predicting reaction performance in C-N cross-coupling using machine learning, *Science* **360** (2018) 186–190.

[11] X. Mu, J. Dong, L. Peng, X. Yang, Deep forest-based intelligent yield predicting of Buchwald-Hartwig coupling reaction, *MATCH Commun. Math. Comput. Chem.* **88** (2022) 5–27.

[12] L. Peng, J. Dong, X. Mu, Z. Zhang, Y. Zhang, X. Yang, P. Zhang, Intelligent predicting reaction performance in multi-dimensional chemical space using quantile regression forest, *MATCH Commun. Math. Comput. Chem.* **87** (2022) 299–318.

[13] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition. *Proceed. IEEE* **86** (1998) 2278–2324.

[14] Y. Zhao, X. Liu, H. Lu, X. Zhu, T. Wang, G. Luo, R. Zheng, Y. Luo, An optimized deep convolutional neural network for yield prediction of Buchwald-Hartwig amination. *Chem. Phys.* **550** (2021) #111296.

[15] V. Mnih, N. Heess, A. Graves, K. Koray, Recurrent models of visual attention, *Proceed. 27th Int. Conf. Neur. Inf. Process. Sys.* **2** (2014) 2204–2212.

[16] K. Fukushima, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, *Biol. Cyber.* **36** (1980) 193–202.

[17] D. H. Hubel, T. N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *J Physiol.* **160** (1962) 106–154.

[18] C. Coley, W. Jin, L. Rogers, T. Jamison, T. Jaakkola, W. Green, R. Barzilay, K. Jensen, A graph-convolutional neural network model for the prediction of chemical reactivity, *Chem. Sci.* **10** (2019) 370–377.

[19] S. Loffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, *Proceed. 32nd Int. Conf. Machine Learn.* **37** (2015) 448–456.

[20] G. E. Hinton, S. Nitish, K. Alex, S. Ilya, S. Ruslan, Improving neural networks by preventing co-adaptation of feature detectors, *Comput. Sci.* **3** (2012) 212–223.

[21] J. W. Kim, J. S. Jang, M. S. Yang, J. H. Kang, K. W. Kim, Y. J. Cho, J. W. Lee, A study on fault classification of machining center using acceleration data based on 1D CNN algorithm, *J. Korean Soc. Manufact. Process Engin.* **18** (2019) 29–35.

[22] G. Hu, H. Wan, X. Li, A high-precision magnetic-assisted heading angle calculation method based on a 1D convolutional neural network (CNN) in a complicated magnetic environment, *Micromachines* **11** (2020) #642.

[23] B. Shin, S. Park, K. Kang, J. C. Ho, Self-attention based molecule representation for predicting drug-target interaction, *Proceed. 4th Machine Learn. Healthcare Conf.* **106** (2019) 230–248.

[24] M. Manica, A. Oskooei, J. Born, V. Subramanian, S. Julio, M. R. Martínez, Towards explainable anticancer compound sensitivity prediction via multimodal attention-based convolutional encoders, *Mol. Pharm.* **16** (2019) 4797–4806.

[25] J. N. Wei, D. David, A. Alán, Neural networks for the prediction of organic chemistry reactions, *ACS Cent. Sci.* **2** (2016) 725–732.

[26] L. Peng, X. Yang, H. Hou, J. Dong, Z. Wang, Y. Zhao, Intelligent prediction method of coupling reaction yield based on attentional convolutional neural network, *National Invention Patent*, Issue number: CN113380346A.

[27] K. Ladan, H. Farzad, A. Falah, E. Ali, Combining design of experiments, machine learning, and principal component analysis for predicting energy consumption and product quality of a natural gas processing plant, *Int. J. Energy Res.* **45** (2021) 5974–5987.

[28] P. Kang, Z. Liu, Reaction prediction via atomistic simulation: from quantum mechanics to machine learning, *iScience* **24** (2020) #102013.

[29] Q. Zhang, L.Yang, F. Zhou, Attention enhanced long short-term memory network with multi-source heterogeneous information fusion:An application to BGI Genomics, *Inf. Sci.* **553** (2021) 305–330.

[30] J. Ma, R. Sheridan, A. Liaw, G. E. Dahl, V. Svetnik, Deep neural nets as a method for quantitative structure-activity relationships, *Comput. Sci.* **55** (2015) 263–274.

[31] Z. Zhou, *Machine Learning*, Tsinghua Univ. Press, 2016, pp. 229–232.