

## Research Article

# A Novel Input Variable Selection and Structure Optimization Algorithm for Multilayer Perceptron-Based Soft Sensors

**Hongxun Wang** , **Lin Sui** , **Mengyan Zhang** , **Fangfang Zhang** , **Fengying Ma** ,  
and **Kai Sun** 

*School of Electrical Engineering and Automation, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China*

Correspondence should be addressed to Kai Sun; sunkai79@qlu.edu.cn

Received 6 January 2021; Revised 6 March 2021; Accepted 15 April 2021; Published 3 May 2021

Academic Editor: Adrian Neagu

Copyright © 2021 Hongxun Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A novel optimization algorithm for multilayer perceptron- (MLP-) based soft sensors is proposed in this paper. The proposed approach integrates input variable selection and hidden layer optimization on MLP into a constrained optimization problem. The nonnegative garrote (NNG) is implemented to perform the shrinkage of input variables and optimization of hidden layer simultaneously. The optimal garrote parameter of NNG is determined by combining cross-validation with Hannan-Quinn information criterion. The performance of the algorithm is demonstrated by an artificial dataset and the practical application of the desulfurization process in a thermal power plant. Comparative results demonstrated that the developed algorithm could build simpler and more accurate models than other state-of-the-art soft sensor algorithms.

## 1. Introduction

In complex industrial processes, important process parameters that influence product quality or energy consumption need to be monitored and controlled in real time and with high accuracy. However, some of them are difficult to be directly measured with hardware sensors due to the limitations of existing field conditions [1–3]. Soft sensors achieve the mathematical modeling of these hard-to-measure parameters through auxiliary variables that are easy to be measured [4, 5]. Basically, there are two categories of soft sensor techniques: mechanism analysis-based approaches and data-driven approaches. The mechanism analysis-based approaches require accurate understanding of the inherent mechanism of complex industrial processes, which is very difficult for the researchers. Data-driven algorithms provide advanced alternatives with statistical inference and machine learning techniques [6, 7]. In recent years, data-driven soft sensors including principal component regression (PCR), partial least squares (PLS) regression, support vector machine (SVM), extreme learning machine (ELM), and artificial neural networks (ANNs) have been widely studied [8–12].

Due to their powerful nonlinear modeling competence, ANNs have become the most popular nonlinear modeling techniques. There are a variety of ANNs such as convolutional neural networks (CNN) [13], generative adversarial networks (GAN) [14], radial basis networks, and recurrent neural network (RNN) [15], each of which has its own characteristics and advantages. Among them, multilayer perceptron (MLP) is the most widely used technique for nonlinear soft sensing owing to its outstanding nonlinear mapping capability and convenience of application. Heidari et al. [16] built an accurate predictive model of nanofluid viscosity with MLP. Shen et al. [17] presented an MLP-based recursive sliding mode dynamic surface control scheme for a fully actuated surface vessel with uncertain dynamics and external disturbances. In [18], MLP was applied to predict the water content of biodiesel and diesel blend in terms of temperature and composition.

With the rapid development of process automation, more and more variables are involved in the modern process industry. Redundant input variables increase the model complexity, delay the training time, and decrease the predictive accuracy of the model [19, 20]. Variable selection

technology provides a good solution to this problem and therefore is extensively studied [1, 21, 22]. Guo et al. [23] proposed an input variable selection method for a feed-forward neural network (FNN) by using partial auto-correlation function and successfully forecasted the wind speed. Fock [24] proposed a new algorithm for the selection of input variables, in which the global sensitivity analysis technique was used to select the optimal input variables. Adil et al. [25] presented a new variable selection algorithm that used the heuristic method and minimum redundancy maximum relevance, and the experimental results showed better accuracy than other algorithms. In [26], a neural network-based soft sensor was developed to predict effluent concentrations in a biological wastewater treatment plant, in which principal component analysis (PCA) was implemented to select optimal input variables.

Nonnegative garrote (NNG) is a linear coefficient shrinkage approach based on penalty likelihood function. In recent years, it has been widely used in the variable selection of ANNs [27]. Sun et al. [28] utilized the NNG to compress the input weights of the MLP to achieve nonlinear variable selection, and the superiority of the proposed algorithm was proved through two artificial dataset examples and a real industrial application. In [29], a local search strategy was incorporated into the NNG-MLP to improve its performance. However, these algorithms only consider the selection of input variables and ignore the optimization of the internal structure of the MLP network. Actually, the redundant nodes of hidden layers worsen the performance of MLP as the redundant input variables do and even lead to overfitting of the model. Pan et al. [30] proposed a novel approach of simplifying the structure of deep neural network through regularization of network architecture. Anbananthen et al. [31] presented a pruning procedure, by which redundant links were deleted from the trained network. Monika and Venkatesan [32] designed a divisive ANN clustering algorithm to prune the neurons of the hidden layer of MLP, which promoted model accuracy. Fan et al. proposed an algorithm that utilized the least absolute shrinkage and selection operator (LASSO) to perform the selection of input variables and the optimization of the hidden layer of MLP, named dLASSO [33]. However, the variable selection and hidden layer optimization of dLASSO are independent of each other, which may cause the omission of the optimal solution.

According to our investigation, few existing methods deal with the redundancy of input variables and hidden layers of ANN models synchronously. In this paper, a novel algorithm that performs global dimension reduction and structure simplification for MLP-based soft sensors is proposed by elaborately combining NNG and MLP. The MLP is implemented to cope with the nonlinear dynamics of the industrial processes, and NNG is devised to conduct the selection of the input variables and simplification of the hidden layers. To the best of our knowledge, this algorithm is a quite innovative design of a penalty function-based strategy for global optimizing the structure of ANNs. The effectiveness of the developed algorithm is validated by an

artificial dataset and application to a practical industrial process to provide informative analysis.

The remainder of this paper is organized as follows. The background theories of the approach are reviewed in Section 2. Section 3 describes the detailed principles and development of the proposed algorithm. The simulation results and analysis of artificial datasets and practical industrial process are presented in Section 4. Finally, some concluding remarks are given in Section 5.

## 2. Theoretical Background

The architecture of a three-layer MLP discussed in the paper is demonstrated in Figure 1, which is composed of an output layer, a hidden layer, and an input layer. The number of neurons of input layer is dependent on the variables or columns of the input dataset, while that of the hidden layer is usually chosen by trial and error. The mathematical expression of the studied MLP is shown as

$$y = f(g(xw_I + b_I)w_O + b_O), \quad (1)$$

where  $g(\cdot)$  and  $f(\cdot)$  denote the activation functions of the hidden and output layer, respectively,  $x = [x_1, x_2, \dots, x_p]$  is the vector of input variables, and  $y$  is the output variable.

The weight  $w_I = \begin{bmatrix} w_{11} & \cdots & w_{1q} \\ \vdots & \ddots & \vdots \\ w_{p1} & \cdots & w_{pq} \end{bmatrix}$  is a  $p \times q$  matrix that links nodes of the input and hidden layer.  $b_I = [b_1, b_2, \dots, b_q]$  is the bias vector of the hidden nodes.  $w_O = [w_1^o, w_2^o, \dots, w_q^o]^T$  represents the matrix of output weights linking the hidden and the output layer. The output bias is denoted as  $b_O$ .

For the linear regression problem,

$$y = x\beta + \varepsilon, \quad (2)$$

where  $\beta = [\beta_1, \beta_2, \dots, \beta_p]^T$  is the vector of magnitude coefficients and  $\varepsilon$  is the random error. Breiman proposed a constraint consisting of the summation of shrinkage coefficients  $c = [c_1, c_2, \dots, c_p]$  and imposed it on the ordinary least squares (OLS) regression model [34]:

$$c^*(s) = \operatorname{argmin} \left\{ \sum (y - (c \cdot x)\hat{\beta})^2 \right\}, \quad \forall (x, y) \in \{X, Y\}$$

$$\text{subject to } c_i \geq 0, \quad \sum_{i=1}^p c_i \leq s, \quad (3)$$

in which  $\hat{\beta}$  represents the coefficient vector of OLS estimation and  $s$  is the garrote parameter.  $X$  is the input dataset, in which each column corresponds to a candidate input variable, and  $Y \in R^n$  is the dataset of output variable.

In [28], the NNG algorithm was devised to select the input variable of MLP by imposing  $c$  on the input layer:

$$\hat{y} = f(g(c \cdot xw_I + b_I)w_O + b_O), \quad (4)$$

and equation (3) is consequently reformulated as

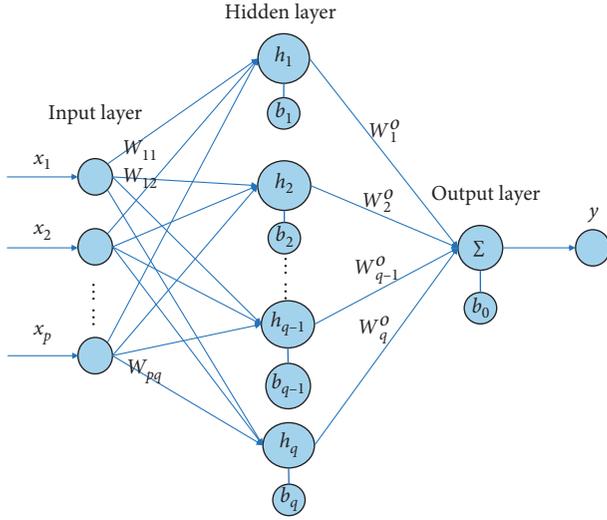


FIGURE 1: Architecture of a three-layer MLP.

$$c^*(s) = \operatorname{argmin} \left\{ \sum (y - f(g((c \cdot x)w_I + b_I)w_O + b_O))^2 \right\},$$

$\forall (x, y) \in \{X, Y\}$

subject to

$$c_i \geq 0, \quad \sum_{i=1}^p c_i \leq s. \quad (5)$$

$$[c_I^*; c_H^*](s) = \operatorname{argmin} \left\{ \sum (y - f(g((c_I \cdot x)w_I + b_I)c_H \cdot w_O + b_O))^2 \right\}$$

subject to  $c_i \geq 0, c_h \geq 0, \sum_{i=1}^p c_i + \sum_{j=1}^q c_h \leq s,$

(7)

where  $c_i^* = 0$  indicates that the input variable  $x_i$  is removed from the MLP and  $c_h^* = 0$  means that the hidden node  $h_i$  is excluded from the model. Equation (7) is a nonlinear quadratic optimization problem with constraints that can be solved with trust-region reflective optimization algorithm [35]. After that, the optimal predictive model of MLP is presented by

$$\hat{y} = f(g((c_I^* \cdot x)w_I + b_I)c_H^* \cdot w_O + b_O). \quad (8)$$

**3.2. Determination of Parameter  $s$ .** The choice of parameter  $s$  is very important for the developed algorithm because it can directly affect the extent of shrinkage on the MLP structure.  $s = 0$  implies that all input variables and hidden nodes will be eliminated. When  $s \geq p + q$ , all the input variables and the hidden nodes will be completely preserved. Therefore, the value of  $s$  directly determines the number of neurons and influences the performance of MLP. This paper adopts the enumeration approach to select the optimal  $s$  from the vector  $S = [s_1, s_2, \dots, s_u]$ . Herein,  $s_1$  is set to a constant close to zero, and  $s_u$  is set to

### 3. Development of GNNG-MLP Algorithm

**3.1. Design of Global Optimization for MLP.** In the study, a global optimization algorithm for MLP-based soft sensor, called GNNG-MLP, is proposed to reduce the redundancy of input and hidden layer simultaneously. The primary strategy of the proposed algorithm is to design a nonlinear quadratic optimization expression with NNG constraint that imposes the shrinkage coefficients on the input and hidden layers of MLP. The GNNG-MLP is implemented with the continuous adjustment of the garrote parameter. The schematic diagram of the proposed algorithm is illustrated in Figure 2, in which the nodes  $x_2$  and  $h_2$  have null impacts on the model and will be removed from the MLP. Meanwhile, the weight lines connected to them will also be invalid.

The proposed algorithm is divided into two phases. In the first phase, a well-trained MLP network is presented with the conventional MLP training algorithm. At the second phase, a set of shrinkage coefficients are imposed on input and hidden layer of the obtained MLP. Consequently, the expression of MLP is reformulated as follows:

$$y = f(g((c_I \cdot x)w_I + b_I)c_H \cdot w_O + b_O), \quad (6)$$

where  $c_I = [c_1, c_2, \dots, c_p]$  and  $c_H = [c_1, c_2, \dots, c_q]$  denote the shrinkage coefficients of the nodes of input and hidden layer, respectively.  $c_I^*$  and  $c_H^*$  are obtained by solving the following formula:

$p + q$ . The other values of  $S$  are equally distributed between  $s_1$  and  $s_u$ .

In this paper, Hannan–Quinn information criterion (HQ) [36] that can balance the accuracy and complexity of a model is adopted as the model evaluation criterion that is formulated as

$$HQ = -2 \cdot \ln \sum (\hat{y} - y)^2 + k \cdot \ln(\ln n), \quad (9)$$

where  $n$  denotes the number of data samples,  $k$  represents the number of input variables, and  $y$  and  $\hat{y}$  are the actual and predictive value of the output variable, respectively. Considering the randomness of ANNs, the V-fold cross-validation (CV) method is taken to validate the model. The execution is described as follows. Firstly, the group of all datasets is evenly separated into V subdatasets. Secondly, a single subdataset is taken as the validation dataset, and the other V-1 subdatasets are used as the training dataset to acquire the trained MLP. The procedure is repeated V times, and these V results are averaged to present the ultimate estimate. In this work,  $s$  is chosen by V-fold CV with HQ, whose pseudocode is shown in Algorithm 1.

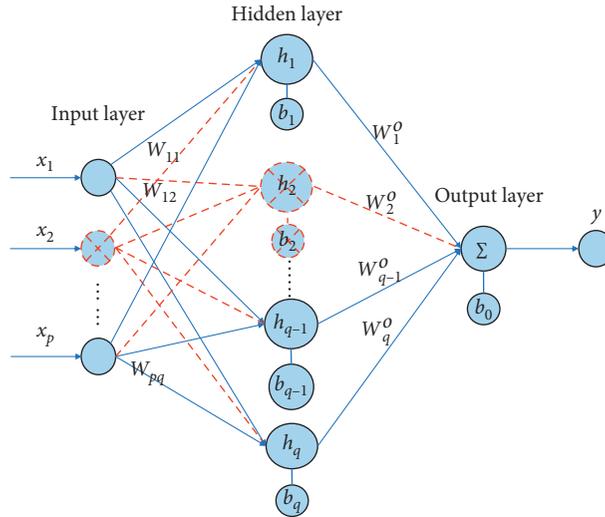


FIGURE 2: Schematic diagram of GNNG-MLP.

**Input:** dataset  $\{X, Y\}$   
**Output:** the optimal  $s^*$   
**Begin Algorithm**  
 Initialize  $S = [s_1, s_2, \dots, s_u]$ ;  
 Separate  $\{X, Y\}$  into  $V$  disjoint subdatasets  $\{X_1, Y_1\}, \{X_2, Y_2\}, \dots, \{X_V, Y_V\}$ ;  
 For  $i = 1: u$   
    $s = s_i$ ;  
   For  $v = 1: V$   
     Train a new MLP network with dataset  $\{X, Y\} = \{X, Y\} \setminus \{X_v, Y_v\}$ ;  
     Solve equation (7) with  $s$  to get  $c_I^*, c_H^*$ ;  
     Get the new MLP by equation (8);  
     Compute the HQ( $v$ ) with validation dataset  $\{X_v, Y_v\}$ ;  
   End for  
   CV\_HQ( $i$ ) = mean (HQ);  
 End for  
 Output the optimal  $s^*$  with the minimum CV\_HQ;  
**End Algorithm**

ALGORITHM 1: Pseudocode of choice of  $s$ .

**3.3. The Computational Procedure of Proposed Algorithm.** In this paper, a global optimization algorithm for MLP is developed. The advancement of the proposed algorithm is that it not only deals with the redundancy of input variables but also simplifies the internal structure of MLP. The overall computation flow of the algorithm is described as follows:

*Step 1.* Initialization: get a trained MLP with the training dataset  $\{X, Y\}$ .

*Step 2.* Impose the NNG coefficients on the input and hidden nodes of the MLP.

*Step 3.* Perform Algorithm 1 to obtain the optimal  $s$  as  $s^*$ .

*Step 4.* Acquire the shrinkage coefficient  $c_I^*$  and  $c_H^*$  by solving equation (7) with parameter  $s^*$ .

*Step 5.* Updated weights of input and hidden nodes by substituting  $c_I^*$  and  $c_H^*$  into equation (8).

*Step 6.* Remove the columns whose corresponding coefficient  $c_i^* = 0$  from  $\{X, Y\}$ , and delete the hidden nodes whose corresponding coefficient  $c_h^* = 0$ .

*Step 7.* Output the optimized MLP.

## 4. Simulation Results

**4.1. Experimental Setting.** In the paper, comprehensive simulations are implemented to verify the performance of the proposed algorithm, in which comparisons with other state-of-the-art variable selection algorithms such as SBS-MLP [37], NNGEO-MLP [29], and dLASSO-MLP [38] are performed. All algorithms are simulated under the same settings. The MLP structure in the case is a typical three-layer configuration, in which the activation function of hidden and output layer is hyperbolic tangent and linear, respectively. The initial number of hidden nodes is determined by

TABLE 1: The statistical results of artificial dataset.

	MLP	SBS-MLP	NNGEO-MLP	dLASSO-MLP	GNNG-MLP
MSE	0.1752	0.1653	0.1503	0.1223	0.1015
A_R <sup>2</sup>	0.8979	0.9033	0.9105	0.9335	0.9437
Neurons	62	29.15	24.55	20.10	12
FS+	40	7.1500	2.9000	3.1000	0.4500
FS-	0	0	0.3500	0	0.1000

some trial runs. Training and testing data take up 80% and 20% of the overall dataset, respectively. 5-fold CV is employed in the algorithm. The performance of the involved algorithms is assessed with the following five measures.

- (1) MSE: the mean square error between the predicted and the actual value with the testing dataset,  $MSE = 1/n \sum_{i=1}^n (y_i - \hat{y}_i)^2$ .
- (2) Adjusted R\_Square (A\_R<sup>2</sup>):  $A\_R^2 = 1 - ((n-1) * \sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-p-1) * \sum_{i=1}^n (y_i - \bar{y}_i)^2)$ , where  $\bar{y}_i$  is the mean value of output variable.
- (3) Neurons: the total number of the input and hidden nodes in the optimized MLP.
- (4) False-positive selection (FS+): the number of irrelevant variables included in the optimized MLP.
- (5) False-negative selection (FS-): the number of relevant variables excluded from the optimized MLP.

**4.2. Simulation Results of Artificial Dataset.** In this subsection, a nonlinear model that was proposed in [28] is applied to generate artificial datasets. The input dataset  $X$  was produced from a multivariate normal distribution with covariance matrix  $\Sigma$ , in which covariance between two different variables (columns)  $\sum_{i,j} = \rho^{|i-j|}$ ,  $\forall i \neq j$ . The mathematical expression of the model is

$$Y = \begin{cases} \sqrt{X_1\beta - 0.5} + \varepsilon, & X_1\beta \geq 0.5, \\ e^{X_1\beta+0.5} + \varepsilon, & X_1\beta \leq 0.5, \end{cases} \quad (10)$$

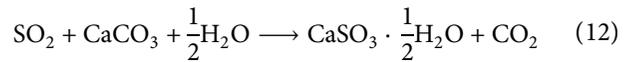
where  $X_1 \in \mathbb{R}^{1000 \times 10}$  are relevant variables,  $\varepsilon$  is white Gaussian noise, and  $\beta = [3.0, 1.5, 2.0, 4.0, 0.5, 1.3, -2.6, -3.5, -5.1, 2.0]^T$ . Besides the relevant variables, irrelevant dataset  $X_2 \in \mathbb{R}^{1000 \times 40}$  is produced to make this case a problem of selecting 10 relevant variables out of 50 variables.

Table 1 presents the statistical results of artificial dataset with different algorithms after 20 runs. In this case,  $\rho$  of the covariance matrix is set to 0.8, which generates a dataset with a high correlation between different variables. According to the numerical comparison of MSE and A\_R<sup>2</sup>, the GNNG-MLP has the highest prediction accuracy among all algorithms. Furthermore, FS+ is the smallest, which indicates that the GNNG-MLP selects fewer irrelevant variables than other approaches. By comprehensively comparisons of FS+ and FS-, it can be concluded that our algorithm could select relevant variables with more precision. Besides, statistical results of neurons show that GNNG-MLP can effectively remove the redundant nodes and then improve the performance of the model. It can be found from the

results that our algorithm solves the problems of input variables and model redundancy simultaneously.

In addition, the capability of different algorithms is further compared by changing the value of collinearity  $\rho$ . Figure 3 shows the comparison of the five indicators with different  $\rho$ s. It can be seen that the GNNG-MLP consistently yields the lowest MSE, meaning that our algorithm always has the best accuracy. The number of hidden layer nodes with GNNG-MLP is always the lowest, which proves the efficiency of reducing the redundancy with our approach. Moreover, our algorithm also performs the best on other indicators in most cases, which demonstrates that our algorithm has the best stability.

**4.3. Application to an Actual Desulfurization Process of Power Plant.** In this section, the developed algorithm was applied to forecast the SO<sub>2</sub> emissions from a desulfurization process of a thermal power plant in China. The structural diagram of the process is shown in Figure 4. This power plant adopts limestone-gypsum wet flue gas desulfurization technology, which includes SO<sub>2</sub> absorption system, flue gas system, and compressed air system. The technology mainly uses lime and limestone to absorb SO<sub>2</sub> by chemical reactions that are shown as follows:



The limestone slurry entering the primary absorption tower is dissolved in the absorption tower slurry pool. By adjusting the amount of limestone slurry entering the absorption tower or the concentration of the slurry discharged from the absorption tower, the pH value of the absorption tower slurry pool is maintained between 5.5 and 6.5 to ensure the limestone dissolution and SO<sub>2</sub> absorption. After the original flue gas first enters the primary absorption tower, it passes through the spray zone in countercurrent, is fully contacted with the slurry to absorb SO<sub>2</sub>, and then enters the secondary absorption tower. The remaining SO<sub>2</sub> and other harmful components in the flue gas are absorbed in the spray zone. Finally, the dust is removed by a wet dust collector and discharged to the chimney. The two absorption towers adopt almost the same structure that is demonstrated in Figure 5.

Table 2 presents the statistical results of 20 runs with different soft sensor algorithms. It can be found that GNNG-MLP has better prediction accuracy with a smaller number

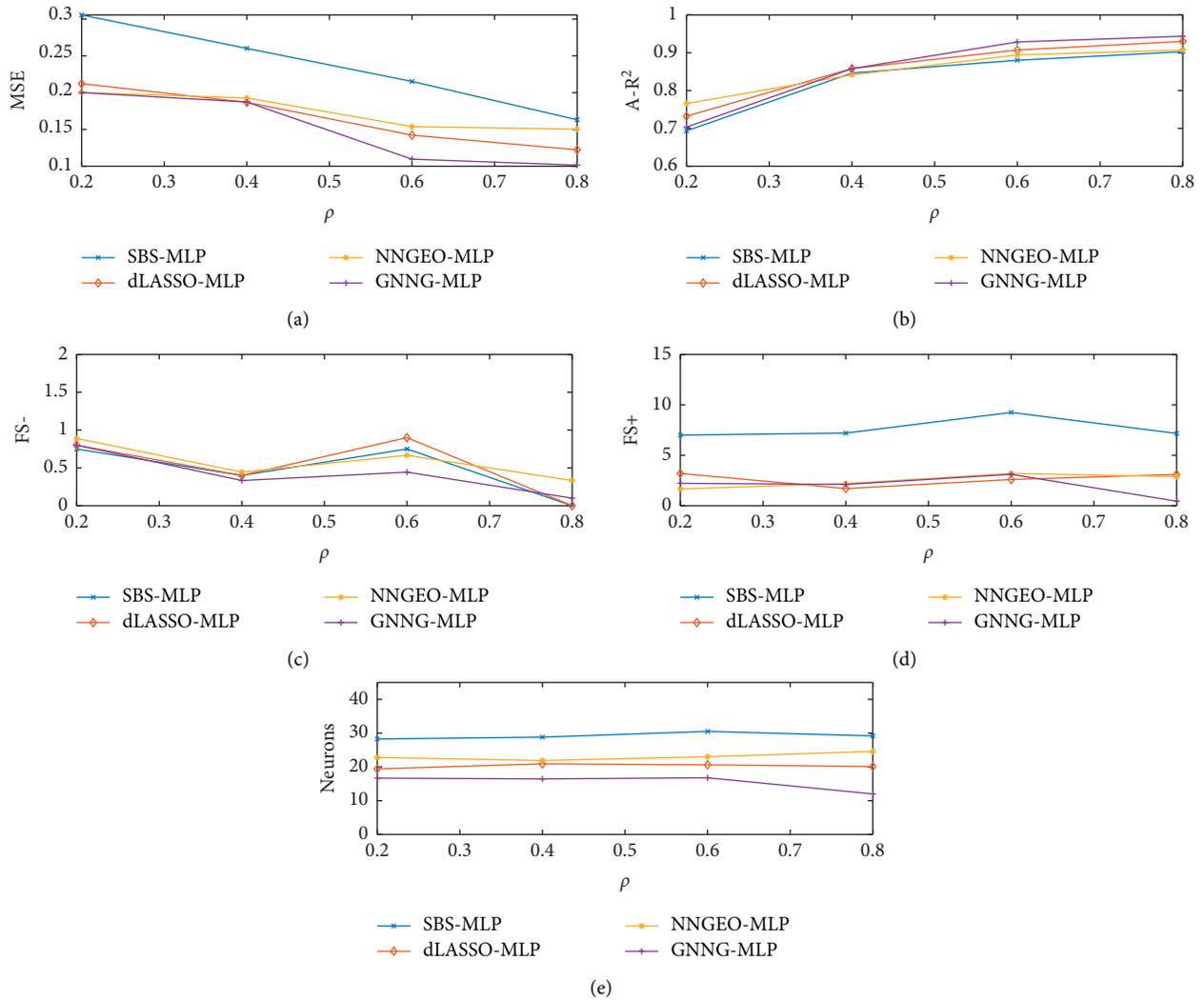


FIGURE 3: Comparison of performance with different  $\rho$ s. (a) Comparison of MSE. (b) Comparison of FS. (c) Comparison of neurons. (d) Comparison of A-R<sup>2</sup>. (e) Comparison of F5+.

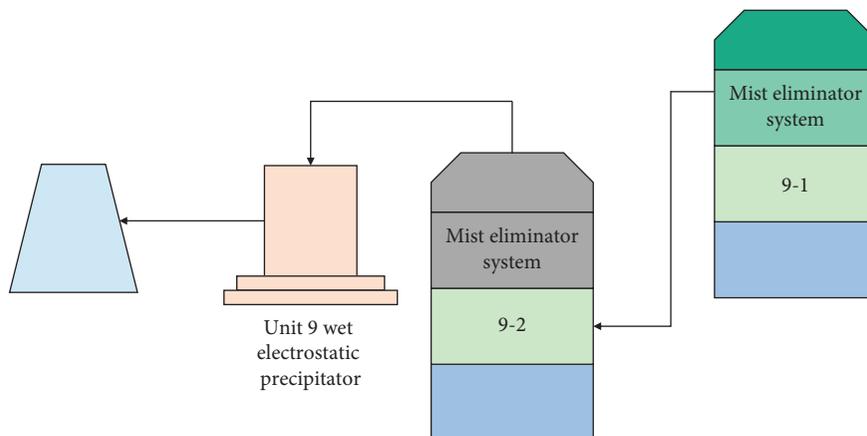


FIGURE 4: The framework of absorption system.

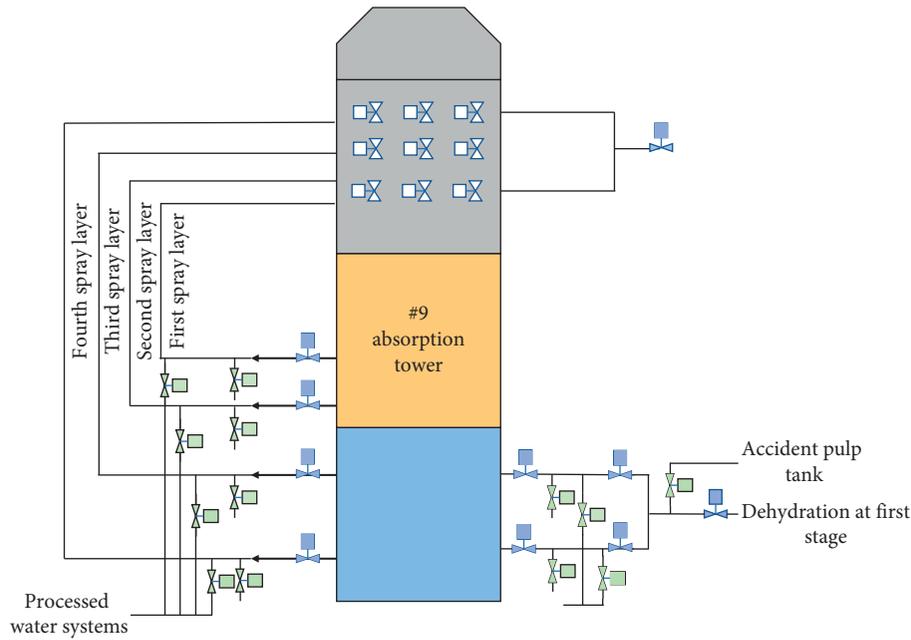


FIGURE 5: Schematic diagram of primary absorption tower.

TABLE 2: Statistical results for SO<sub>2</sub> concentration prediction.

	MLP	SBS-MLP	NNGEO-MLP	dLASSO-MLP	GNNG-MLP
Neurons	36	24	15	15	10
MSE	32.5388	7.2795	4.5729	3.3518	2.4640
A_R <sup>2</sup>	0.1155	0.7105	0.8283	0.8895	0.9004

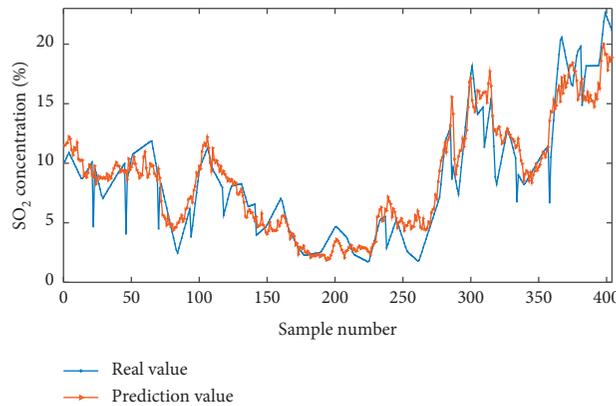


FIGURE 6: Measured and the predicted SO<sub>2</sub> concentration.

of neurons than other approaches. This result shows that GNNG-MLP can improve the accuracy of the model by simplifying the internal structure of the MLP.

Figure 6 shows the comparison of predictive and actual value of the target variable with our algorithm. Obviously, the proposed algorithm can effectively track the dynamic change of the target variable.

In order to further prove the accuracy of the proposed algorithm, error comparisons between the measured and the predicted SO<sub>2</sub> concentration with different algorithms are presented in Figure 7. The results show that the error of

GNNG-MLP is the lowest and within the range  $[-4.2, 4.2]$  in most instances, which can meet the requirements of the field operating. The performance of the developed soft sensor is fully compliant with the standards of industry demand.

Besides, comparative analyses based on the statistical results of variable selection and the actual industrial operating experience are given. Figure 8 presents the frequency of input variable selection over 100 runs. It can be found from Figure 8 that variable 13 is included in all solutions, and variables 17 and 30 are selected more than 80 times.

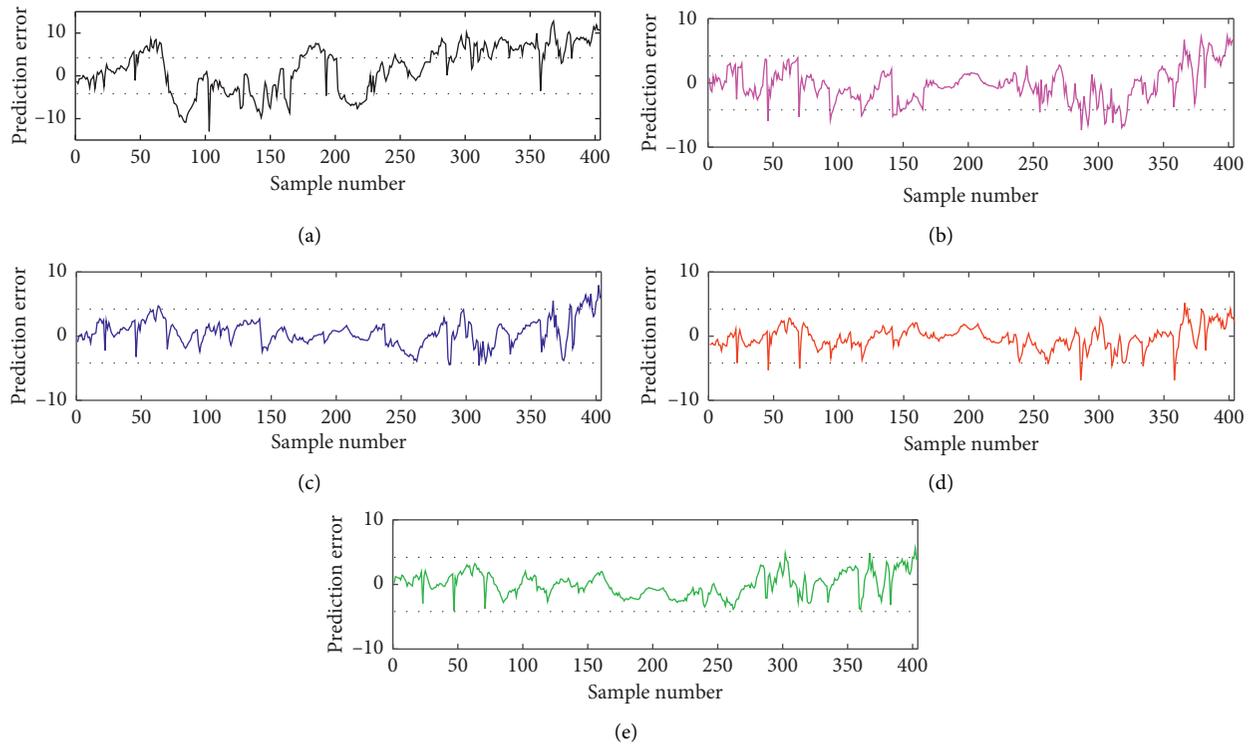


FIGURE 7: Comparison of prediction error with different algorithms. (a) MLP prediction error. (b) NNGEO-MLP prediction error. (c) GNNG-MLP prediction error. (d) SBS-MLP prediction error. (e) dLASSO-MLP prediction error.

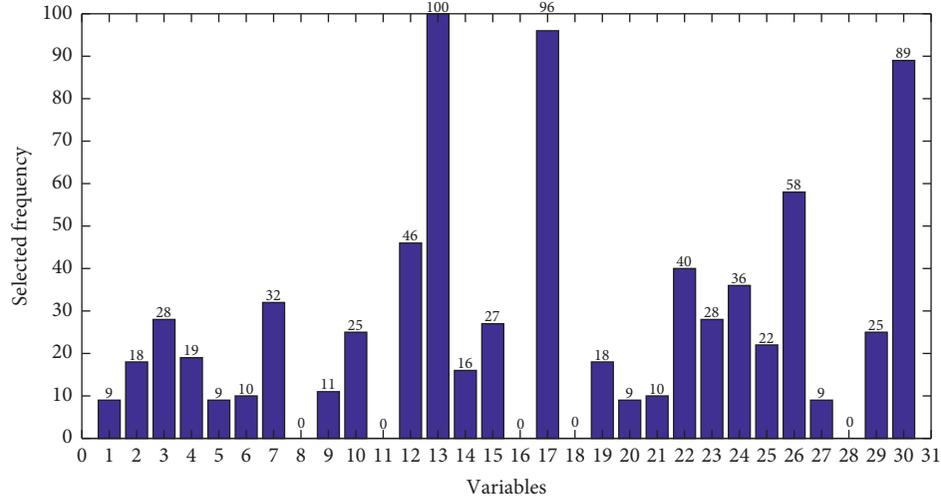


FIGURE 8: Variable selection frequency 100 runs by GNNG-MLP.

According to the statistics, the most relevant input variable to the output variable is variable 13. In terms of the manual book of the system, variable 13 is the  $\text{SO}_2$  concentration of #9-1AT outlet's flue gas and the  $\text{SO}_2$  concentration of #9-2AT inlet's flue gas. Obviously, this variable is highly related to the  $\text{SO}_2$  concentration of final emission. Variable 17, that is, the limestone slurry to #9 AT flow, has 90% of selection frequency. It can be seen from formulas (11) and (12) that the  $\text{CaO}$  and  $\text{CaCO}_3$  in limestone slurry can absorb the released  $\text{SO}_2$ . Therefore, variable 17 is included in the optimal solution. The

variable 30 is the pH value of the slurry in the tower 9-2. The slurry absorbs more  $\text{SO}_2$  when the  $\text{SO}_2$  concentration in the flue gas is relatively high. As a result of this, a large amount of hydrogen ions will be generated, and the pH value will decrease.

## 5. Conclusions

This paper proposed a new optimization algorithm for MLP-based soft sensors with NNG. The advantage of this algorithm is that it can simultaneously perform the selection

of the input layer and the optimization of the hidden layer for MLP and therefore has more tendency to get the global optimal model. The simulation results on the artificial datasets demonstrate that GNN-MLP has obvious advantages in both the number of neurons and the generalization performance of the model. In addition, the algorithm is applied to forecast the SO<sub>2</sub> emission in a desulfurization process to verify the reading of the online analyzer. Comprehensive results and comparisons prove that the developed soft sensor has remarkable model simplicity and accuracy. The proposed soft sensor can be further implemented for the optimization and control design of the desulfurization process.

### Data Availability

The data used to support the findings of this study are currently under embargo, while the research findings are commercialized. Requests for data 24 months after publication of this article will be considered by the corresponding author.

### Conflicts of Interest

The authors declare no conflicts of interest.

### Acknowledgments

The work was supported by the Key Research and Development Program of Shandong Province (Grant no. 2019GGX104037).

### References

- [1] C. A. C. Belchior, R. A. M. Araújo, F. A. A. Souza, and J. A. C. Landeck, "Sensor-fault tolerance in a wastewater treatment plant by means of ANFIS-based soft sensor and control reconfiguration," *Neural Computing and Applications*, vol. 30, no. 5, pp. 3265–3276, 2018.
- [2] X. Yuan, Z. Ge, Z. Song, Y. Wang, C. Yang, and H. Zhang, "Soft sensor modeling of nonlinear industrial processes based on weighted probabilistic projection regression," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 4, pp. 837–845, 2017.
- [3] X. Yi, R. Guo, and Y. Qi, "Stabilization of chaotic systems with both uncertainty and disturbance by the UDE-based control method," *IEEE Access*, vol. 8, no. 1, pp. 62471–62477, 2020.
- [4] Y. Calcada, "Development of an ANN-based soft-sensor to estimate the apparent viscosity of water-based drilling fluids," *Journal of Petroleum Science & Engineering*, vol. 150, pp. 69–73, 2017.
- [5] L. Yao and Z. Ge, "Deep learning of semisupervised process data with hierarchical extreme learning machine and soft sensor application," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1490–1498, 2017.
- [6] C. Cozin, F. E. C. Vicencio, F. A. de Almeida Barbuto et al., "Two-phase slug flow characterization using artificial neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 3, pp. 494–501, 2016.
- [7] Z. Ge, "Review on data-driven modeling and monitoring for plant-wide industrial processes," *Chemometrics & Intelligent Laboratory Systems*, vol. 171, pp. 16–25, 2016.
- [8] C. Shang, F. Yang, D. Huang, and W. Lyu, "Data-driven soft sensor development based on deep learning technique," *Journal of Process Control*, vol. 24, no. 3, pp. 223–233, 2014.
- [9] R. Ouyse, "Bayesian model averaging and principal component regression forecasts in a data rich environment," *International Journal of Forecasting*, vol. 32, no. 3, pp. 763–787, 2016.
- [10] P. Gottardo, M. Penasa, N. Lopez-Villalobos, and M. De Marchi, "Variable selection procedures before partial least squares regression enhance the accuracy of milk fatty acid composition predicted by mid-infrared spectroscopy," *Journal of Dairy Science*, vol. 99, no. 10, pp. 7782–7790, 2016.
- [11] H. Zhang, S. Wang, D. Li, Y. Zhang, and J. Hu, "Edible gelatin diagnosis using laser-induced breakdown spectroscopy and partial least square assisted support vector machine," *Sensors*, vol. 19, no. 19, pp. 4225, 2019.
- [12] B. Wang, S. Ch, S. Mathur, and J. Adamowski, "Estimation of in-situ bioremediation system cost using a hybrid Extreme Learning Machine (ELM)-particle swarm optimization approach," *Journal of Hydrology*, vol. 543, pp. 373–385, 2016.
- [13] M. C. Xie, X. Han, S. Luan, L. I. Fang, and C. X. Wang, "Brain tumor segmentation using convolutional neural networks feature extraction in MRI images," *Journal of Qufu Normal University*, vol. 43, no. 9, pp. 1–10, 2019.
- [14] M. Esmaeilpour, P. Cardinal, and A. Lameiras Koerich, "Unsupervised feature learning for environmental sound classification using Weighted Cycle-Consistent Generative Adversarial Network," *Applied Soft Computing*, vol. 86, pp. 105912, 2020.
- [15] M. Jiao, D. Wang, and J. Qiu, "A GRU-RNN based momentum optimized algorithm for SOC estimation," *Journal of Power Sources*, vol. 459, pp. 228051, 2020.
- [16] E. Heidari, M. A. Sobati, and S. Movahedirad, "Accurate prediction of nanofluid viscosity using a multilayer perceptron artificial neural network (MLP-ANN)," *Chemometrics and Intelligent Laboratory Systems*, vol. 155, pp. 73–85, 2016.
- [17] Z. Shen, Y. Bi, Y. Wang, and C. Guo, "MLP neural network-based recursive sliding mode dynamic surface control for trajectory tracking of fully actuated surface vessel subject to unknown dynamics and input saturation," *Neurocomputing*, vol. 377, no. 15, pp. 103–112, 2020.
- [18] Y. Wang and W. Gao, "Prediction of the water content of biodiesel using ANN-MLP: an environmental application," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 40, no. 8, pp. 987–993, 2018.
- [19] X. Yuan, B. Huang, Y. Wang, C. Yang, and W. Gui, "Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3235–3243, 2018.
- [20] L. Liu, B. Li, and R. Guo, "Consensus control for networked manipulators with switched parameters and topologies," *IEEE Access*, vol. 9, pp. 9209–9217, 2021.
- [21] K. Sun, S. H. Huang, S. H. Wong, and S. S. Jang, "Design and application of a variable selection method for multilayer perceptron neural network with LASSO," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 28, no. 6, pp. 1386–1396, 2016.
- [22] A. Rani, V. Singh, and J. R. P. Gupta, "Development of soft sensor for neural network based control of distillation column," *ISA Transactions*, vol. 52, no. 3, pp. 438–449, 2013.
- [23] Z. Guo, W. Zhao, H. Lu, and J. Wang, "Multi-step forecasting for wind speed using a modified EMD-based artificial neural

- network model,” *Renewable Energy*, vol. 37, no. 1, pp. 241–249, 2012.
- [24] E. Fock, “Global sensitivity analysis approach for input selection and system identification purposes-A new framework for feedforward neural networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 8, pp. 1484–1495, 2014.
- [25] B.-H. Adil, G. Youssef, and E. Q. Abderrahim, “Hybrid method HVS-MRMR for variable selection in multilayer artificial neural network classifier,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 5, pp. 2773–2781, 2017.
- [26] J. F. D. Canete, P. D. Saz-Orozco, R. Baratti, M. Mulas, and A. Garcia-Cerezo, “Soft-sensing estimation of plant effluent concentrations in a biological wastewater treatment plant using an optimal neural network,” *Expert Systems with Applications*, vol. 63, pp. 8–19, 2016.
- [27] H. Sun, X. Deng, K. Wang, and R. Jin, “Logistic regression for crystal growth process modeling through hierarchical non-negative garrote-based variable selection,” *Iie Transactions*, vol. 48, no. 8, pp. 787–796, 2016.
- [28] K. Sun, J. Liu, J.-L. Kang, S.-S. Jang, D. S.-H. Wong, and D.-S. Chen, “Development of a variable selection method for soft sensor using artificial neural network and nonnegative garrote,” *Journal of Process Control*, vol. 24, no. 7, pp. 1068–1075, 2014.
- [29] K. Sun, X. Wu, J. Xue, and F. Ma, “Development of a new multi-layer perceptron based soft sensor for SO<sub>2</sub> emissions in power plant,” *Journal of Process Control*, vol. 84, pp. 182–191, 2019.
- [30] W. Pan, H. Dong, and Y. Guo, *DropNeuron: Simplifying the Structure of Deep Neural Networks*, vol. 12, pp. 160–172, 2016.
- [31] S. K. Anbananthen, G. Sainarayanan, A. Chekima, and J. Teo, “Data mining using pruned artificial neural network tree (ANNT),” *Journal of Process Control*, vol. 1, pp. 1350–1356, 2006.
- [32] P. Monika and D. Venkatesan, “DI-ANN clustering algorithm for pruning in MLP neural network,” *Indian Journal of Ence & Technology*, vol. 8, no. 16, pp. 1, 2015.
- [33] Y. Fan, B. Tao, Y. Zheng, and S.-S. Jang, “A data-driven soft sensor based on multilayer perceptron neural network with a double LASSO approach,” *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 7, pp. 3972–3979, 2020.
- [34] L. Breiman, “Better subset regression using the nonnegative garrote,” *Technometrics*, vol. 37, no. 4, pp. 373–384, 1995.
- [35] T. F. Coleman and Y. Li, “An interior trust region approach for nonlinear minimization subject to bounds,” *Siam Journal on Optimization*, vol. 6, no. 2, pp. 418–445, 1993.
- [36] Y. Miche and A. Lendasse, “A faster model selection criterion for OP-ELM and OP-KNN: hannan-quinn criterion,” *European Symposium on Esann*, vol. 9, pp. 177–182, 2009.
- [37] E. Romero and J. M. Sopena, “Performing feature selection with multilayer perceptrons,” *IEEE Transactions on Neural Networks*, vol. 19, no. 3, pp. 431–441, 2008.
- [38] Y. Fan, B. Tao, Y. Zheng, and S.-S. Jang, “A data-driven soft sensor based on multilayer perceptron neural network with a double LASSO approach,” *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 7, pp. 3972–3979, 2019.