DTU Library

# Development of An Adversarial Transfer Learning Based Soft Sensor in Industrial Systems

**Li, Dong; Liu, Yiqi; Huang, Daoping; Lui, Chun Fai; Xie, Min**

*Published in:*
IEEE Transactions on Instrumentation and Measurement

*Link to article, DOI:*
10.1109/TIM.2023.3291771

*Publication date:*
2023

*Document Version*
Peer reviewed version



*Citation (APA):*
Li, D., Liu, Y., Huang, D., Lui, C. F., & Xie, M. (2023). Development of An Adversarial Transfer Learning Based Soft Sensor in Industrial Systems. *IEEE Transactions on Instrumentation and Measurement*, *72*, Article 3000610. https://doi.org/10.1109/TIM.2023.3291771

# Development of An Adversarial Transfer Learning Based Soft Sensor in Industrial Systems

Dong Li, Yiqi Liu, *Senior Member, IEEE,* Daoping Huang, Chun Fai Lui, Min Xie *Fellow, IEEE*

***Abstract*—Data-driven soft sensors are usually used to predict quality-related but hard-to-measure variables in industrial systems. However, the acceptable prediction performance mainly relies on the premise that training data are sufficient for model training. To acquire more training data, this paper proposes an adversarial transfer learning (ATL) methodology to enhance soft sensor learning. Firstly, a hierarchical transfer learning algorithm, which integrates a feature extraction method with model-based transfer learning, is proposed to refine the useful hidden information from both historical variables and samples. Then, a novel adversarial learning network is designed to prevent the deterioration of transferred results at each transfer learning stage. Thirdly, a Granger causality analysis (GCA)-based rationale analyzer is added to unfold the internal causality among input variables and between input and output variables simultaneously. Finally, the effectiveness of the proposed soft sensor and the rationale analyzer is validated in a simulated wastewater plant, Benchmark Simulation Model No.2 (BSM2), and a full-scale oxidation ditch (OD) wastewater plant. The experimental results demonstrate that the ATL-based soft sensor can achieve more accurate prediction in terms of RMSE and R, and the GCA-based rationale analyzer can provide a visual explanation for the corresponding model and prediction results.**

***Index Terms*—Soft sensor, Adversarial transfer learning, Granger causality analysis, Historical data, Industrial systems**

## I. Introduction

In recent years, accurate prediction for hard-to-measure variables has received widespread attention in industrial systems, such as biopharmaceutical, petrochemical, metal smelting and wastewater treatment, especially quality-related variables that can describe the operational and safe conditions [2, 3]. However, neither the offline analysis methods nor online measurement devices can meet the real-time prediction requirements totally [4-6]. Soft sensor technology provides an alternative way to address this issue [7, 8]. Soft sensor modeling methods are mainly categorized into data-driven modeling [9], mechanism modeling [10] and hybrid modeling [11]. Up to date, data-driven soft sensors are the most popular and powerful tools for prediction due to the ability that they can build proper soft-sensor models without having the necessary insight into hidden mechanism knowledge deeply. But prediction performance of data-driven soft sensors is mainly dependent on the quantity and quality of training data [12, 13]. In industrial systems, collections of sufficient and optimal training data are difficult and expensive, easily leading to the imbalance of data distribution [14]. Li et al. augmented training data by using the Co-training algorithm to select suitable unlabeled data as new labeled data, but the process of cross validation is complicated and has the risk of selecting the wrong unlabeled data [15]. Mohamed et al. provided an ensemble machine learning model, which can reduce the dependence on training data by implementing biological models to simulate reaction processes. However, the specific biological model is hard to generalize to other industrial systems [16]. Liu et al. proposed a bagging method to increase the sampling rate of training data, but resampling is not able to derive more valuable information for modeling [17].

Recently, the transfer learning algorithm gained popularity due to the fact there reserve a large amount of historical data in industrial systems [18,19]. Cai et al. proposed an instance-based transfer learning method working together with gradient boosting decision trees (GBDT) to establish a wind power quantile regression model. The main motive is to assign different weights to data from different sources in transfer processes. However, the premise of instance-based transfer learning is that the data among source and target domains are supposed to have a significant correlation [20]. Zhao et al. proposed a feature-based transfer learning method to detect unknown variants of network attacks. The main idea is to find the optimal feature variables between source and target domains, but transferred results are sensitive to the used feature extraction methods [21]. Roberto et al. proposed a model-based

This work was supported by National Natural Science Foundation of China (62273151, 62073145, 61873096, 71971181 and 72032005), Basic and Applied Basic Research Foundation of Guangdong Province (2020A1515011057, 2021B1515420003), Guangdong International Scientific Cooperation Research Foundation (2020A0505100024, 2021A0505060001), Horizon 2020 Framework Programme-Marie Skłodowska-Curie Individual Fellowships (891627), Hong Kong Innovation and Technology Commission (InnoHK Project CIMDA), Hong Kong Institute of Data Science (Project 9360163) and by the International Science and Technology Cooperation Program of Guangdong Province (Project #2022A0505050047) *(Corresponding author: Yiqi Liu).*

Dong Li is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China and The School of Automation, Central South University, Changsha, 410083, China (e-mail: lddscut@163.com).

Yiqi Liu is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China and Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 227, 2800 Kgs. Lyngby, Denmark (e-mail: aulyq@scut.edu.cn).

Daoping Huang is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: audhuang@scut.edu.cn).

Chun Fai Lui is with is with the Department of Advanced Design and Systems Engineering, City University of Hong Kong, Hong Kong (e-mail: chunflui5-c@my.cityu.edu.hk).

Min Xie is with the Department of Advanced Design and Systems Engineering and the School of Data Science, City University of Hong Kong, Hong Kong (e-mail: minxie@cityu.edu.hk)

transfer learning method with an application for classification, but transferred results will deteriorate over time [22]. In a word, all the instance-based, feature-based and model-based transfer learning algorithms belong to single-level transfer learning algorithms, which are limited to their own disadvantages and are susceptible to other unobserved factors [23].

In addition, to improve the adaptive ability of soft sensors, Du et al. proposed a moving window partial least squares (MW-PLS) model to update informative regions, but it is susceptible to outliers [24]. Then, Liu et al. proposed an enhanced just-in-time learning (JITL) algorithm to cope with normal changes and abrupt noises in wastewater treatment processes, but the computational burden is unacceptable [25]. Li et al. used the long-short term memory recurrent neural network (LSTM) to keep the soft sensors having high prediction accuracy [26]. However, it requires too much training data to build the initial model due to the complex network structure. Adversarial learning network is a type of machine learning model usually composing of two neural networks, both of which compete with each other aiming to improve their own performance. In adversarial learning networks, one network is called the generator to create new data that resembles some known data distribution, whereas the other network is called the discriminator, which is used to distinguish between real and fake data generated by the generator. Two networks are trained in a zero-sum game framework, where one network's loss is another network's gain [27].

Finally, with more and more soft sensors being opaque due to the advent and development of neural networks, it is more important than ever to study rationally and visually explainable methods aiming to provide a reasonable explanation for the corresponding model and prediction results as well [28]. Traditional explainable methods only rely on mechanism knowledge of process industries to explore why the soft sensor can achieve better prediction performance. However, they lack visualization ability to show operators or managers why better performance can be achieved [29]. Then, some research scholars make the complexity of model as the explainable definition, so as to achieve visual explanation, but it is easy to make a risk of misleading predictions or scarify some predictive power of soft sensors [30, 31]. Therefore, it is urgently needed to design a reasonable and visually explainable method that can balance explanation and prediction abilities to quantify explanation.

To resolve these problems, we propose an adversarial transfer learning (ATL)-based soft sensor and a Granger causality analysis (GCA)-based rationale analyzer in this paper. The detailed contributions are provided in Table I and as follows:

TABLE I
The advantages of the ATL-based soft sensor and the GCA-based rationale analyzer.

| Contributions | Advantages |
| --- | --- |
| A hierarchical transfer learning algorithm | ● Utilize VIP to extract the optimal feature variables from data samples<br>● Build a SCG-NN-based generator with the optimal feature variables for transferring historical data |
| A novel adversarial learning network | ● The SCG-NN-based generator is used to transfer historical data<br>● The cross-entropy error-based discriminator is used to evaluate transferred results |
| A new rationale analyzer | ● Utilize the improved GCA to unfold the internal causality among input variables and between input and out variables simultaneously with low computational intensity |

(1) A hierarchical transfer learning algorithm is proposed to overcome the disadvantages of single-level transfer learning by integrating feature extraction with model-based transfer learning. In this algorithm, variable importance in projection (VIP) is first used to extract the optimal feature variables from data samples. Then, a scaled conjugate gradient (SCG) learning -based shallow neural network (SCG-NN) is built with the optimal feature variables and acts as a generator to transfer historical data and to generate the new training data.

(2) A novel adversarial learning network consisting of a SCG-NN-based generator and a cross-entropy error-based discriminator is proposed. The transferred historical data resembling the original training data distribution are first selected by calculating their cross-entropy error values, and are added into training data set. Then, a new SCG-NN-based generator is rebuilt with the augmented training data set and is used to transfer historical data again. As constantly adversarial learning occurs between the generator and the discriminator, the transferred results will be continuously optimized.

(3) An improved Granger causality analysis (GCA)-based rationale analyzer is added to provide a visual explanation for the corresponding model and prediction results. In this rationale analyzer, we utilize multiple linear regression (MLR) to replace the internal regression method, so the improved GCA-based rationale analyzer can unfold the internal causality among input variables and between input and output variables simultaneously with low computational intensity.

The proposed ATL-based soft sensor and the GCA-based rationale analyzer are described in Section II. Section III provides a simulation wastewater plant (BSM2) and a full-scale OD wastewater plant to verify their effectiveness. Section IV discusses the advantages and disadvantages. Finally, the conclusion and future research are presented in Section V.

## II. ATL-based Soft Sensor and GCA-based Rationale Analyzer

To refine the useful hidden information from historical data for modeling, and provide a visual explanation for the corresponding model and prediction results, an ATL-based soft sensor and a GCA-based rationale analyzer are proposed in this paper. The flowchart is shown in Fig. 1 and the detailed procedure is as follows:

**(1) VIP-based feature extraction:** In industrial systems, there are space-time differences between the historical data set and training data set due to the influences of internal reactions and external environment. We first utilize VIP to extract the optimal feature variables between source domain and target domain (historical data set and training data set).

**(2) Model-based transfer learning:** After VIP-based feature extraction, two new data sets with equal dimension are derived, training data set $D_t$ and historical data set $D_h$. Then, we build a SCG-NN using $D_t$ and denote it as a generator to transfer $D_h$ and to produce new training data.

**(3) Adversarial learning:** To ensure the quality of transferred results, a cross-entropy error-based discriminator is designed to evaluate the transferred results. The proper transferred historical data resembling the original training data distribution can be selected by calculating their cross-entropy error values, and be added into $D_t$. Then, a new SCG-NN-based generator is rebuilt with the new $D_t$ and is used to transfer $D_h$ again.

**(4) Soft sensor establishing:** As constantly adversarial learning occurs between the generator and the discriminator, transferred results are optimized and $D_t$ is augmented continuously. Until satisfying the termination condition of adversarial learning, we use the final $D_t$ to build a SCG-NN-based model for prediction.

**(5) Causality analysis:** To provide a visual explanation for the corresponding model and prediction results, we utilize the GCA-based rationale analyzer unfold the internal causality among input variables and between input and out variables simultaneously, and show the results in the form of figures.

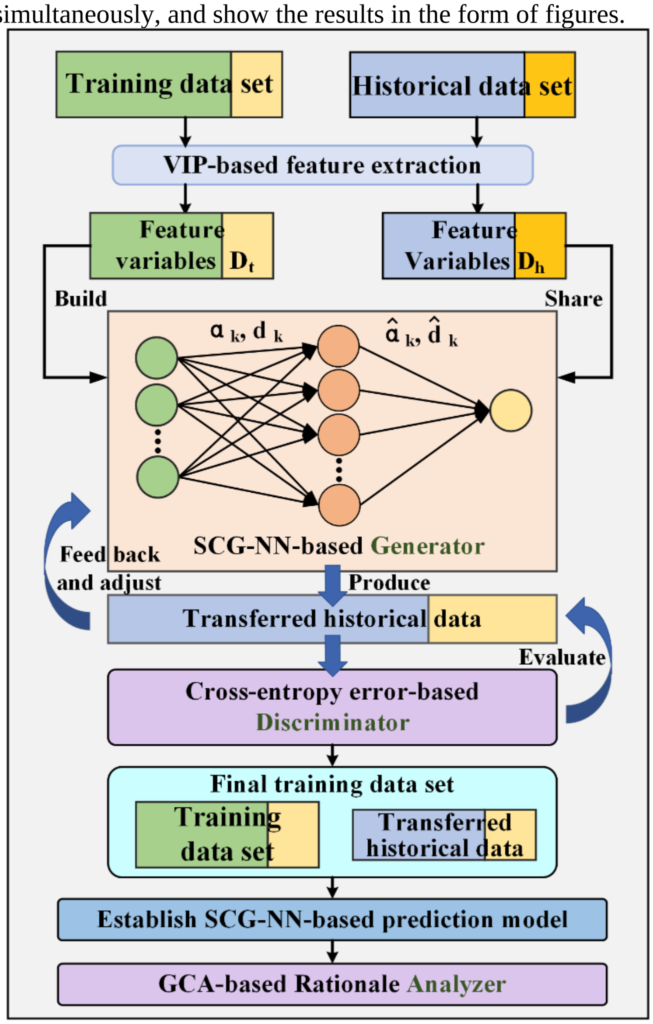


Fig. 1 The flowchart of ATL-based soft sensor and the GCA-based rationale analyzer.

*A. VIP-based Feature Extraction*

VIP is a simple but powerful feature extraction method, which can effectively identify the connection between input and output variables [32]. The detailed procedure is shown in Fig. 2. $X \in R^{n\times m}$ is initial input matrix, $Y \in R^{n\times 1}$ is output matrix. The structure of the *PLS* calibration model between the input and output matrix can be defined as follows:

$$X = TP' + E \tag{1}$$

$$Y = TQ' + F \tag{2}$$

where $T$ represents the relation matrix between $X$ and $Y$, $P \in R^{m\times n}$ and $Q \in R^{1\times n}$ are loading matrices, $E \in R^{n\times m}$ and $F \in R^{n\times 1}$ matrices contain the projection residuals. Then, a common latent variables pace $W^*$ is derived as follows:

$$T = XW^* \tag{3}$$

$$W^* = (P')^{-1} = W(P'W)^{-1} \tag{4}$$

where $W$ represents the $X$-weights matrix. The final prediction output matrix $\hat{Y} \in R^{n\times 1}$ can be described as follows:

$$\hat{Y} = XW^*Q' \tag{5}$$

$$B_{PLS} = W^*Q' \tag{6}$$

where $B_{PLS}$ is named as the *PLS* regression coefficient. It is one of the most frequently used metrics for a multilinear regression issue.

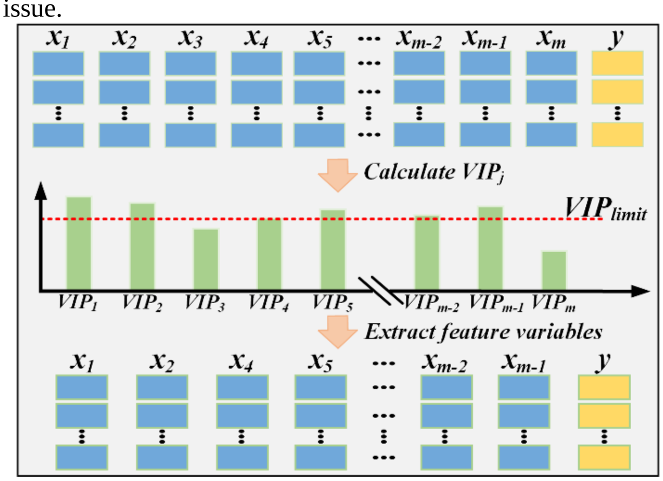


Fig. 2 The procedure of VIP-based feature extraction.

Then, depending on Eq. (1) to Eq. (6), the VIP score of the j-th input variable for $Y$ is calculated:

$$VIP_j = \sqrt{m\sum_{i=1}^{a}[(q_i^2 t_i' t_i)(w_{ij}/\|w_j\|^2)] \,/\, \sum_{i=1}^{a}(q_i^2 t_i' t_i)} \tag{7}$$

where $m$ is the number of variables of the input matrix $X$, $a$ is the size of the matrix, $q_i$, $t_i$ and $w_j$ represent the $i$-th and $j$-th column vectors of $Q$, $T$ and $W$, respectively.

By comparing the *VIP* score with the specific baseline $VIP_{limit}$, the feature variables with high connection to the output variable can be selected. The baseline $VIP_{limit}$ plays an important role in the feature extraction process. If $VIP_j \geq VIP_{limit}$, the $j$-th variable will be selected as an input feature variable. Otherwise, the $j$-th variable will be removed.

Different from the traditional statistical feature extraction methods such as PCA, the VIP metric weights the contribution of each variable according to the variance explained by each *PLS* principal component. In addition, compared with neural network-based feature extraction methods such as

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3291771



Autoencoders (AE), VIP has more highly computational efficiency. Moreover, the extracted features from VIP have the true physical meaning and can reflect the physical correlation among the variables [33].

*B. SCG-based Transfer Learning*

SCG-NN is one of feedforward shallow neural networks. It has the same framework as other feedforward shallow neural networks. The optimization process follows the well-known gradient descent method similarly, and the purpose is to find the proper step size $\alpha_k$ and the descent direction $p_k$ for optimization [34].

$$x_{k+1} = x_k + \alpha_k p_k \quad (8)$$

$$p_k = -g_k + \beta_{k-1} p_{k-1} \quad (9)$$

where $x_k$ represents the $k$-th step iterative value, $p_k$ is the $k$-th step direction of descent, $-g_k$ is the negative gradient of $p_{k-1}$, $\beta_{k-1}$ is determined by different methods in different feedforward neural networks.

Unlike other gradient descent algorithms, SCG-NN utilizes a fully-automated variable gradient method to derive the optimal step size $\alpha_k$ and the descent direction $p_k$ [35], where the searching of step size $\alpha_k$ and the descent direction $p_k$ is determined by the unconstrained optimization problem:

$$\min f(x_k + \alpha_k p_k) \quad (10)$$

where step size $\alpha_k$ is calculated:

$$\alpha_k = -\frac{g_k' p_k}{p_k' H_k P_k} \quad (11)$$

where $H_k$ represents the Hessian matrix, $H_k = \frac{\partial^2 E_k}{\partial x_k^2}$. $E_k$ is the total error of the $k$-th iteration output. $H_k > 0$ is a necessary premise for the continuation of iteration.

Depending on the definition: $s_k = H_k P_k$, $\delta_k = p_k' s_k$ and $u_k = -g_k' p_k$, so $\alpha_k = \frac{u_k}{\delta_k}$. $s_k$ can be calculated by using information from the second-order approximation:

$$s_k = \frac{E'(x_k + \sigma_k P_k) - E'(x_k)}{\sigma_k} + \lambda_k P_k \quad (12)$$

where $E'$ represents the first-order approximation of the total error $E_k$. So $\frac{E'(x_k+\sigma_k P_k)-E'(x_k)}{\sigma_k}$ can describe the second-order approximation of $E_k$ when $\sigma_k \to 0$. According to the newly introduced scale factor $\lambda_k$ and $\alpha_k$ are redefined as:

$$\alpha_k = \frac{u_k}{\delta_k} = \frac{u_k}{p_k' s_k + \lambda_k |P_k|^2} \quad (13)$$

In this paper, SCG-NN is used as the transferred learning model. Its detailed training process is shown as follows:

**Step 1:** The number of neurons in each layer of neural network is determined relying on the distinguish of data, and then we initialize the network parameters randomly.

**Step 2:** We calculate the total error $E_k$ of the $k$-th iteration output, and depending on Eq. (11) to Eq. (13), derive the proper step size $\alpha_k$ and the descent direction $p_k$.

**Step3:** The network parameters are updated through coping with the unconstrained optimization problem Eq. (10).

**Step4:** We recalculate the total error $E_{k+1}$ with the updated network parameters, and derive the more proper step size $\alpha_{k+1}$ and the descent direction $p_{k+1}$ again.

**Step 5:** Repeat the above steps until satisfying the termination condition of iteration.

In SCG-NN, the fully-automated variable gradient method is an optimization methodology based on conjugate directions. In other words, the directions are orthogonal to each other with respect to the quadratic function $H_k$ [36]. SCG-NN has a clearer optimization goal in training processes, so leading to that it can denote better computation efficiency.

*C. A Novel Adversarial Learning Network*

Inspired by the adversarial learning network [37], we design a novel adversarial learning network to prevent the deterioration of transferred results in this paper. Its generator is based on the proposed hierarchical transfer learning algorithm, aiming to transfer the historical data and to produce new training data, whereas the cross-entropy error is involved as the evaluation criterion setting up the discriminator, which can be used to evaluate the quality of transferred data. Then, we adjust the parameters of SCG-NN depending on the feedback evaluation results. The detailed cross-entropy error function is as follows:

$$L_{cee}(y_i, \hat{y}_i) = -[y_i \ln(\hat{y}_i) + (1 - y_i)\ln(1 - \hat{y}_i) \quad (14)$$

where $y_i$ and $\hat{y}_i$ are real and predictive values, respectively. $L_{cee}(y_i, \hat{y}_i)$ represents the cross-entropy error value between $y_i$ and $\hat{y}_i$. Since the cross-entropy error function can describe the difference from probability distribution of variables, it is suitable to assess the quality of transferred data in this paper [38].

*D. GCA-based Rationale Analyzer*

GCA is a multivariable linear analysis method based on time series. The primary purpose is to identify the causality by exploring whether one time series is useful for predicting another [39]. Different from other analysis methods, GCA can identify the causality between input and output variables rather than only analyzing why the soft sensor can achieve predictions for target variables.

In this paper, the problem of space-time differences between two data sets has been resolved through the proposed ATL-based soft sensor, so the multiple regression issue is simplified. MLR is involved as the new regression method to replace the internal regression method of GCA. MLR can not only decrease the computational intensity but also analyze multiple causalities among input variables and between input and output variables simultaneously. The detailed process is as follows:

Firstly, all the input and output variables are used to build an MLR-based regression model:

$$\begin{gathered} x_1 = a_{1\,1}x_1 + a_{1\,2}x_2 + \cdots + a_{1\,n}x_n + a_{1\,n+1}y + e_1 \\ x_2 = a_{2\,1}x_1 + a_{2\,2}x_2 + \cdots + a_{2\,n}x_n + a_{2\,n+1}y + e_2 \\ \dots\dots \\ x_n = a_{n\,1}x_1 + a_{n\,2}x_2 + \cdots + a_{n\,n}x_n + a_{n\,n+1}y + e_n \\ y = a_{n+1\,1}x_1 + a_{n+1\,2}x_2 + \cdots + a_{n+1\,n}x_n + a_{n+1\,n+1}y + e_{n+1} \end{gathered} \quad (15)$$

where $(x_1, x_2, \dots, x_n)$ is the complete set of input variables, $y$ is the output variable, $a_{p\,q}$ is the regression coefficient, $x_q$ is the $q$-th variable, $e_p$ represents the prediction error. The regression model Eq. (15) is called the unrestricted model or the full model [40].

Then, to identify the causality between input and output variables, we derive an incomplete regression model lacking a certain variable. For example, $q \neq 1$:

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3291771



$$\begin{aligned} x_1 &= b_{1\,2}x_2 + b_{1\,3}x_3 + \cdots + b_{1\,n}x_n + b_{1\,n+1}y + e_{1(1)} \\ x_2 &= b_{2\,2}x_2 + b_{2\,3}x_3 + \cdots + b_{2n}x_n + b_{2\,n+1}y + e_{2(1)} \\ &\cdots\cdots \\ x_n &= b_{n\,2}x_2 + b_{n\,3}x_3 + \cdots + b_{n\,n}x_n + b_{n\,n+1}y + e_{n(1)} \\ y &= b_{n+1\,2}x_2 + b_{n+1\,3}x_3 + \cdots + b_{n+1\,n}x_n + b_{n+1\,n+1}y + e_{n+1(1)} \end{aligned} \quad (16)$$

where $b_{p\,q}$ is the regression coefficient, $e_{p(1)}$ is the prediction error of the $p$-th variable by excluding the first variable.

Finally, the GCA-index from $x_q$ to $x_p$ is defined to be the log-likelihood ratio:

$$F_{q\to p} = \ln \frac{var\,(e_{p(q)})}{var\,(e_p)} \quad (17)$$

where $var$ is the variance, $F_{q\to p}$ represents the influence of variable $x_q$ on $x_p$. Eq. (16) quantifies the causality between variables. When $F_{q\to p}$ value is larger, it means that $x_q$ has a significant causality to $x_p$. Otherwise, $x_q$ invalid to $x_p$.

The contribution rate (CR) is defined as follows:

$$CR_{q\to p} = \frac{F_{q\to p}}{sum(F_{j\to p})}\ j = 1,2,\dots,n+1 \quad (18)$$

By comparing $CR_{q\to p}$ values, we can provide a more intuitive result to identify the causality among input and between input and output variables.

## III. Case Studies

To validate the prediction performance of the ATL-based soft sensor and the explanation ability of GCA-based rationale analyzer, two sets of data are collected from a simulated wastewater plant and a full-scale OD wastewater plant, respectively. The root means square error (RMSE) and correlation coefficient (R) are selected as evaluation indexes of prediction results. The detailed introduction of two wastewater plants and evaluation indexes can be referred to the Appendix Information.

### *A. Benchmark Simulation Model No.2 (BSM2)*

*(1) Feature Extraction and Parameter Setting:* In this case study, chemical oxygen demand (COD) is the output variable, which is a typically quality-related but hard-to-measure variable, other 36 process variables are initial input variables. To extract the optimal feature variables, we calculate the $VIP$ score of each initial input variables using Eq. (7). At the same time, the threshold baseline $VIP_{limit} = 1.2$ is provided relying on the distribution of data. After VIP-based feature extraction, the final input and output variables are derived and shown in Appendix Information Table S1. In addition, with the iterative adjustment, the final structure of SCG-NN is set up as 21-15-1. The suitable network structure can not only contribute to reducing overfitting occurs but also improve the computational efficiency.

TABLE II
The feature extraction results and time consumption of PCA, AE and VIP.

| Method | Initial inputs | Feature extraction | Time(s) |
|---|---|---|---|
| PCA | 36 | 25 | 0.38 |
| AE | 36 | 30 | 12.36 |
| VIP | 36 | 21 | 0.41 |

*(2) Performance of different feature extraction methods:* As profiled in Table II, PCA, AE and VIP extracted 25, 30 and 21 feature variables from the 36 initial input variables, respectively. Obviously, VIP derived the lowest-dimensional feature variables. This is because VIP can extract the optimal feature variables and remove the irrelevant variables through calculating the VIP score of each initial input variable. Then, by comparing their time consumption, we found that the time consumption of VIP is lower than AE significantly. This is mainly because VIP only needs to consider the statistical correlation between input and output variables rather than using the complex neural networks like AE to exploring the unexplainable relationship. To further demonstrate the superiority of VIP, Fig. 3 and Fig. 4 show the fitting profile between the prediction and true values as well as prediction errors with different feature extraction methods. It can be observed from Fig. 3 that VIP-SCG has better stability and achieves more accurate prediction, especially under the peak and valley locations. This illustrates that VIP can extract the most proper feature variables for prediction. Also, as can be seen from Fig. 4, occasional and unacceptable prediction errors occur frequently when using PCA-SCG and AE-SCG, but the prediction errors of VIP-SCG are always close to zero. Therefore, VIP-SCG can predict COD better than other soft sensors.

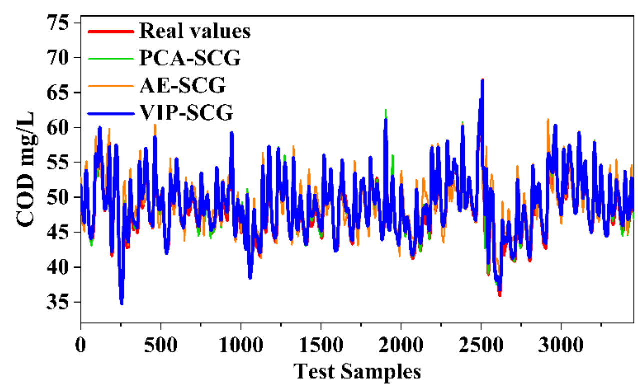


Fig. 3 The fitting profile between the prediction and true values of PCA-SCG, AE-SCG and VIP-SCG.

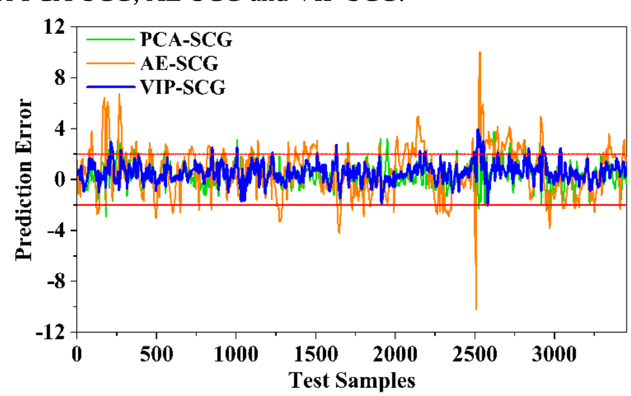


Fig. 4 The prediction errors of PCA-SCG, AE-SCG and VIP-SCG.

*(3) Performance of soft sensors with or without ATL:* To verify the advantages of ATL-based soft sensor, Table III tabulates the prediction results in terms of RMSE, R and time consumption of different soft sensors with or without ATL. By performing ATL, the prediction results of all soft sensors have

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3291771



been improved significantly than without ATL in terms of RMSE values are reduced by 0.21, 0.81 and 0.47, respectively. This is mainly because ATL transferred the proper historical data to augment the training data set, leading to that we can establish better soft sensor model to predict COD. Then, by comparing the time consumption, it can be found that although the time consumption of soft sensors with ATL is larger than without ATL, the computational intensity is still acceptable.

TABLE III
The prediction results in terms of RMSE, R and time consumption of soft sensors with and without ATL.

| Method | RMSE | R | Time(s) | Method | RMSE | R | Time(s) |
|---|---|---|---|---|---|---|---|
| PCA-SCG | 0.99 | 97.69 % | 1.53 | PCA-SCG-ATL | 0.78 | 98.30 % | 5.57 |
| AE-SCG | 3.89 | 90.04 % | 1.56 | AE-SCG-ATL | 3.28 | 92.32 % | 5.87 |
| VIP-SCG | 0.82 | 98.61 % | 1.51 | VIP-SCG-ATL | 0.35 | 99.18 % | 5.31 |

*(4) Performance of different soft sensors:* Fig. 5 shows the prediction results of VIP-SCG-ATL and some other state-of-the-art regression methods-based soft sensors such as VIP-LSTM and VIP-SAE, and Table IV depicts their RMSE, R and time consumption. As profiled in Fig. 5, the prediction values of VIP-SCG-ATL are still the closest to the true values. This is mainly because it can refine the useful hidden information from historical data through ATL and use them to build better models for prediction. In addition, it is worthy to note that other two soft sensors, VIP-LSTM and VIP-SAE, have the poor prediction performance during the initial phase. This indicates that their optimization efficiency is inferior to VIP-SCG-ATL. By comparing the time consumption in Table IV, we can more intuitively discover VIP-SCG-ATL has the cheapest time consumption. The major reason is that SCG-NN belongs to one of feedforward shallow neural networks and uses the fully-automated variable gradient method for training.

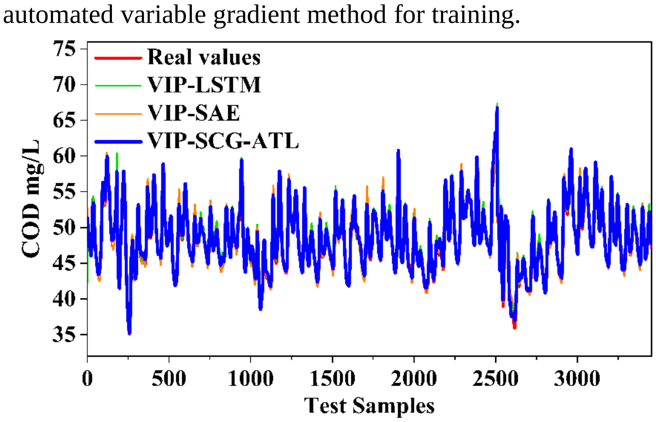


Fig. 5 The fitting profile between the prediction and true values of VIP-LSTM, VIP-SAE and VIP-SCG-ATL.

TABLE IV
The prediction results in terms of RMSE, R and time consumption of VIP-LSTM, VIP-SAE and VIP-SCG-ATL.

| Method | RMSE | R | Time(s) |
|---|---|---|---|
| VIP-LSTM | 0.97 | 98.26% | 33.64 |
| VIP-SAE | 1.01 | 97.27% | 23.42 |
| VIP-SCG-ATL | 0.35 | 99.18% | 5.31 |

*(5) Result of GCA-based rationale analyzer:* Fig. 6 shows the contribution rate of each input variable to COD using the GCA-based rationale analyzer. The first-highest and second-highest causalities to COD are Q-in and Q-p, respectively. Obviously, the result conforms to the actual wastewater treatment process. When increasing or decreasing the water flow at the influence and primary locations, the efficiency of the whole wastewater treatment will be indeed affected, COD-e is no exception. Besides the influence of water flow, SI and SO also play significant roles in nitrification and denitrification processes. As displayed in Fig. 6, they have large CR values on COD-e such as SI-r2 and SO-r5.

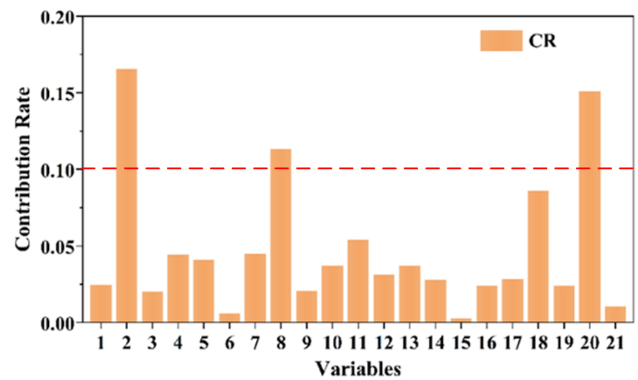


Fig. 6 The contribution rate (CR) of each input variable to COD.

*B. A Full-scale OD Wastewater Treatment Plant*

*(1) Feature Extraction and Parameter Setting:* In this case study, biological oxygen demand (BOD) that can describe the reaction efficiency of biological activity is the output variable. And there are 36 process variables in the full-scale OD wastewater treatment process, which are all selected as initial input variables. To eliminate the negative influences of unrelated process variables and reduce the complexity of model, like the first case study, we determine the threshold baseline $VIP_{limit} = 1$ relying on the distribution of data, and calculate $VIP$ score of each initial input variables using Eq. (7). After VIP-based feature extraction, the final input and output variables are derived and shown in Appendix Information Table S2. Similarly, the structure of SCG-NN is determined through continual trial-and-error. The final structure is set up as 9-12-1.

TABLE V
The feature extraction results and time consumption of PCA, AE and VIP.

| Method | Initial inputs | Feature extraction | Time(s) |
|---|---|---|---|
| PCA | 36 | 15 | 0.063 |
| AE | 36 | 20 | 8.721 |
| VIP | 36 | 9 | 0.096 |

*(2) Performance of different feature extraction methods:* We compare the results and time consumption of different feature extraction methods. As profiled in Table V, PCA, AE and VIP reduce the dimension of the input variables significantly, especially VIP reduces the dimension from 36 to 9. Also, even though the time consumption of VIP is not cheapest, it is still acceptable. This is because VIP can refine the correlation between input and output variables only through simple computation. To further illustrate the superiority of VIP,

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3291771



Fig. 7 and Fig. 8 show the fitting profile between prediction and the true values as well as prediction errors of different feature extraction methods. As displayed in Fig. 7, the soft sensor after VIP-based feature extraction can derive the best prediction results, especially under the peak and valley locations. This further illustrates that VIP captures the feature variables with high correlation to BOD. In addition, it can be seen from Fig. 8, the prediction errors of all soft sensors are highly dynamic in the whole wastewater treatment process, but the VIP-SCG still derived the best prediction results with the most stable prediction errors between -0.1 and 0.1.

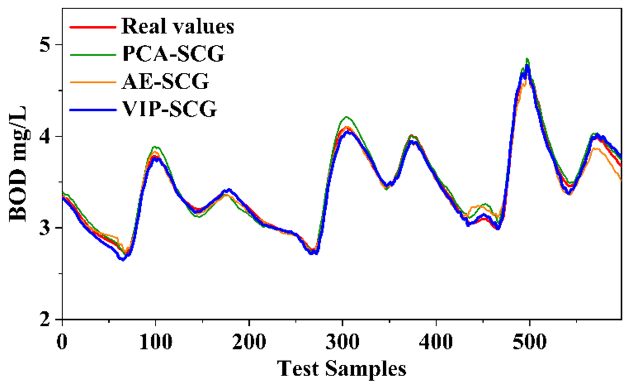


Fig. 7 The fitting profile between the prediction and true values of PCA-SCG, AE-SCG and VIP-SCG.

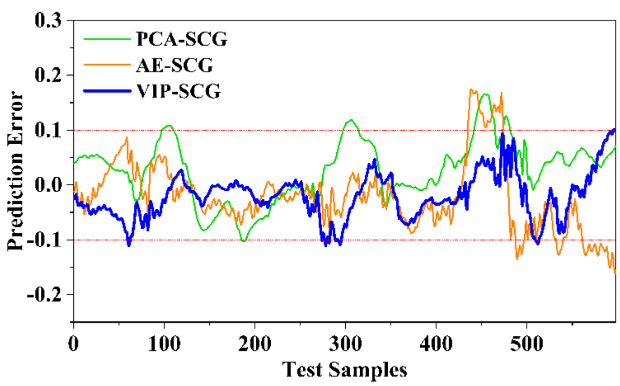


Fig. 8 The prediction errors of PCA-SCG, AE-SCG and VIP-SCG.

*(3) Performance of soft sensors with or without ATL:* To verify the effectiveness of ATL, Table VI depicts the prediction results of soft sensors with or without ATL in terms of RMSE, R and time consumption. Obviously, by performing the ATL, the prediction results of all soft sensors have a large improvement in terms of RMSE values being reduced by 43.57%, 28.81% and 43.83% than without ATL. This is mainly because the historical data can be continually transferred and optimized during the adversarial transfer learning process, which will be beneficial to the instruction of model with higher prediction performance. It is worth noting that since the methodology makes the model become more complex, the time consumption will be increasing, but it is still acceptable.

TABLE VI

The prediction results in terms of RMSE, R and time consumption of different soft sensors with and without ATL.

| Method | RMSE | R | Time(s) | Method | RMSE | R | Time(s) |
|---|---|---|---|---|---|---|---|
| PCA-SCG | 0.0038 | 99.32% | 1.86 | PCA-SCG-ATL | 0.0027 | 99.42% | 3.6 |
| AE-SCG | 0.0044 | 99.18% | 1.97 | AE-SCG-ATL | 0.0034 | 99.19% | 3.75 |
| VIP-SCG | 0.0022 | 99.56% | 1.74 | VIP-SCG-ATL | 0.0015 | 99.63% | 3.57 |

*(4) Performance of different soft sensors:* Like the first case study, we compared the prediction results of VIP-SCG-ATL and some other state-of-the-art regression methods-based soft sensors, VIP-LSTM and VIP-SAE. As profiled in Fig. 9, even though VIP-LSTM and VIP-SAE basically fit the profiles of true values, there occurs some prediction deviation in the peak and valley, so leading to that their overall prediction accuracies are inferior to VIP-SCG-ATL. This is mainly because the training data, especially resembling to under the peak and valley locations, can be augmented by using ATL. In addition, it can be seen from Table VII that the time consumption of VIP-SCG-ATL is lowest. This demonstrates that VIP-SCG-ATL can not only achieve best prediction results, but also have the most highly computational efficiency.

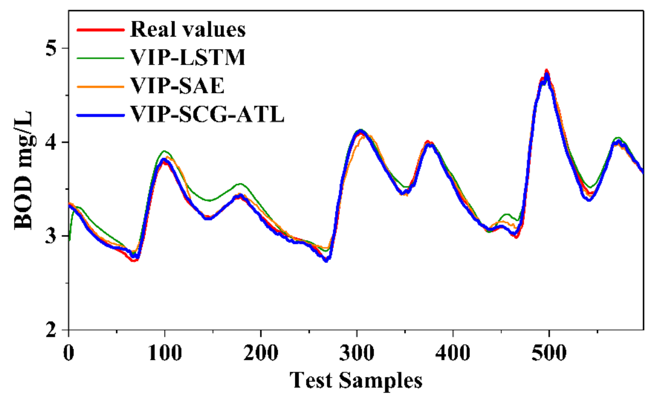


Fig. 9 The fitting profile between the prediction and true values of VIP-LSTM, VIP-SAE and VIP-SCG-ATL.

TABLE VII

The prediction results in terms of RMSE, R and time consumption of VIP-LSTM, VIP-SAE and VIP-SCG-ATL.

| Method | RMSE | R | Time(s) |
|---|---|---|---|
| VIP-LSTM | 0.0066 | 98.89% | 21.71 |
| VIP-SAE | 0.0031 | 99.40% | 12.35 |
| VIP-SCG-ATL | 0.0015 | 99.63% | 3.57 |

*(5) Result of GCA-based rationale analyzer:* Fig. 10 provides an intuitive explanation for the corresponding model and prediction results. We can find that the first-highest and second highest causalities to BOD-e are DO-e and SNH-e, respectively. In the full-scale OD wastewater treatment plant, the larger DO and SNH are, the more intense the biochemical reaction, the higher BOD is. Obviously, the analysis result is consistent with the biochemical reaction process. Also, the improved GCA-based rationale analyzer can accurately analyze the causality among input variables simultaneously. As shown in Fig. 10, we can determine the first-highest and second-highest causalities for each input variable. And the analysis result also follows the biochemical reaction process. For example, after the wastewater is adequately reacted in the first aerated section, the reaction will be weakened in the second

aerated section, so OUR-r2 is the first-highest causality to OUR-r4. Besides, the nitrification and denitrification reaction processes will directly affect the concentration of nitrogen, thus OUR-r4 is the first-highest causality to TKN-e.

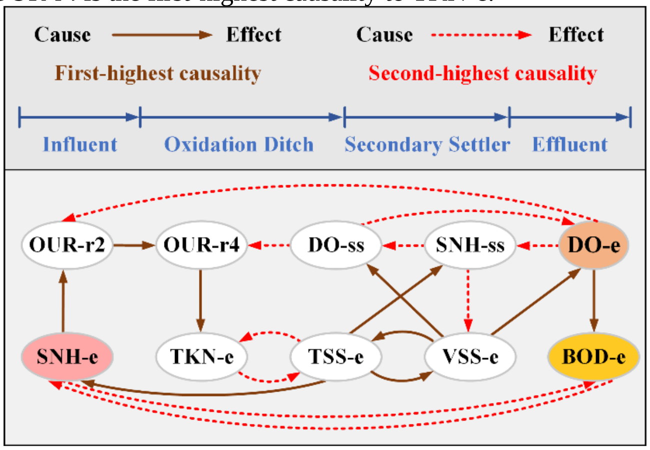


Fig. 10 The first-highest and second highest causalities among variables.

## V. Discussions

To improve the prediction performance of soft sensor for the quality-related but hard-to-measure variables in industrial systems, this paper proposed an ATL-based soft sensor. Firstly, a hierarchical transfer learning algorithm integrating VIP-based feature extraction with SCG-NN transfer learning is proposed and used to refine the useful hidden information from historical data. Then, we design a novel adversarial learning network to prevent deterioration of transferred results, which consists of a SCG-NN-based generator and a cross-entropy error-based discriminator. Finally, a GCA-based rationale analyzer is added to provide a visual explanation for the corresponding model and prediction results.

In this paper, two case studies are provided to validate the effectiveness of ATL-based soft sensor and the explanation ability of GCA-based rationale analyzer. The data set of the first case study belongs to a large-scale data set, and these data are collected from a simulation platform, whereas the second is a small-scale data set, and these data are collected from a full-scale OD wastewater treatment plant. In two case studies, their sampling periods are different, which are 15 mins (in the first case study) and 200 sets one day (in the second case study), respectively. The first data set can be used to represent a well instrumented plant in the city, whereas the second one can represent a plant in the rural areas with less sensors and instrumentations. It is important to notice that the data set from the second case study have larger fluctuations than the first case study, which will increase the difficulty of prediction and the complexity of model. Overall, both the proposed ATL-based soft sensor and the GCA-based rationale analyzer can achieve satisfactory prediction and analysis results in two case studies. This demonstrates that the ATL-based soft sensor and the GCA-based rationale analyzer have the wide applicability.

However, there are still some unresolved issues. Firstly, when suffering from diverse samples, the hierarchical transfer learning algorithm are sensitive in the training process, thus they could be substituted by other similar methods or models when necessary [41, 42]. In addition, the process of adversarial learning is time-consuming due to the evaluation and adjustment at each iteration stage. To improve the computational efficiency, it will become a good research direction to replace the adversarial learning process with online deep learning in the future research [43]. Finally, the GCA-based rationale analyzer is only devoted to analyzing the causality among the input variables and between input and output variables, lacking the analysis and discussion regarding the internal structure of models [44].

## VI. Conclusion

In this paper, an ATL-based soft sensor and a GCA-based rationale analyzer are proposed to address the problems of prediction and model explanation. Firstly, the useful hidden information from historical data is refined by using the adversarial transfer learning network, resulting in the enrichment of training data and the improvement of prediction performance. Then, the GCA-based rationale analyzer provides an intuitive and reasonable explanation for the corresponding model and prediction results, which can greatly increase the credibility of soft sensors and prediction results. Finally, their effectiveness is validated in a simulated BSM2 plant and a full-scale OD wastewater treatment plant. The quality-related variables can be predicted effectively with RMSE and R being 0.35 and 99.18% (in the first case study) and 0.0015 and 99.63% (in the second case study), respectively. In the future research, we will focus on more efficient feature extraction methods and regression algorithms for a industrial system with multiple working conditions. Moreover, recursive deep learning networks could be a potential solution to improve the adaptive ability of soft sensors.

## References

[1] Y. Sun, W. Qin, Z. Zhuang, and H. Xu. “An adaptive fault detection and root-cause analysis scheme for complex industrial processes using moving window KPCA and information geometric causal inference,” Journal of Intelligent Manufacturing. Vol. 32, pp. 2007-2021, 2021.
[2] Q. Jiang, S. Yan, H. Cheng, and X. Yan, "Local-global modeling and distributed computing framework for nonlinear plant-wide process monitoring with industrial big data,” IEEE transactions on neural networks and learning systems. vol. 32, no. 8, pp. 3355-3365, 2020.
[3] C. Lui, Y. Liu, and M. Xie, "A Supervised Bidirectional Long Short-Term Memory Network for Data-Driven Dynamic Soft Sensor Modeling," IEEE Transactions on Instrumentation and Measurement. Vol. 71, pp. 1-13, 2022.
[4] Z. J. C. Ge, "Review on data-driven modeling and monitoring for plant-wide industrial processes," Chemometrics and Intelligent Laboratory Systems, vol. 171, pp. 16-25, 2017.
[5] D. Li, D. Huang, and Y. Liu, "A novel two-step adaptive multioutput semisupervised soft sensor with applications in wastewater treatment," Environmental Science and Pollution Research, pp. 1-15, 2021.
[6] X. Wu, H. Han, and J. Qiao, "Data-Driven Intelligent Warning Method for Membrane Fouling," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 8, pp. 3318-3329, 2021.
[7] X. Yuan, S. Qi, Y. Wang, K. Wang, C. Yang, and L. Ye, "Quality Variable Prediction for Nonlinear Dynamic Industrial Processes Based on Temporal Convolutional Networks," IEEE Sensors Journal, vol. 21, no. 18, pp. 20493-20503, 2021.
[8] D. Li, Y. Liu, D. Huang, and C. Xu, " A Semi-supervised Soft-sensor of Just-in-time Learning With Structure Entropy Clustering and Applications for Industrial Processes Monitoring," IEEE Transactions on Artificial Intelligence, vol. 1, no. 1, pp. 1-12, 2022.
[9] W. Yan, D. Tang, and Y. Lin, "A data-driven soft sensor modeling method based on deep learning and its application," IEEE Transactions on Industrial Electronics, vol. 64, no. 5, pp. 4237-4245, 2016.

[10] P. Jin, W. Robbins, and G. Bota, "High-temperature corrosion by carboxylic acids and sulfidation under refinery conditions—mechanism, model, and simulation," Industrial & Engineering Chemistry Research, vol. 57, no. 12, pp. 4329-4339, 2018.
[11] P. C. Lopez et al., "Towards a digital twin: a hybrid data-driven and mechanistic digital shadow to forecast the evolution of lignocellulosic fermentation," Biofuels, bioproducts and biorefining, vol. 14, no. 5, pp. 1046-1060, 2020.
[12] Y. Liu and M. Xie, "Rebooting data-driven soft-sensors in process industries: A review of kernel methods," Journal of Process Control, vol. 89, pp. 58-73, 2020.
[13] Y. L. He, Z. Q. Ge, and Q. Zhu, "Data driven soft sensor development for complex chemical processes using extreme learning machine," Chemical Engineering Research and Design, vol. 102, pp. 1-11, 2015.
[14] N. Mohajerin, and S. Waslander, "Multistep prediction of dynamic systems with recurrent neural networks," IEEE transactions on neural networks and learning systems, vol. 30, no. 11, pp. 3370-3383, 2019.
[15] D. Li, Y. Liu, and D. Huang, "Development of semi-supervised multiple-output soft-sensors with Co-training and tri-training MPLS and MRVM," Chemometrics and Intelligent Laboratory Systems, vol. 199, pp. 103970-103981, 2020.
[16] H. Mohamed, A. Negm, M. Zahran, and O. Saavedra, "Bathymetry determination from high resolution satellite imagery using ensemble learning algorithms in Shallow Lakes: Case study El-Burullus Lake," International Journal of Environmental Science and Development, vol. 7, no. 4, p. 295-302, 2016.
[17] Y. Liu, D. Huang, B. Liu, Q. Feng, and B. Cai, "Adaptive ranking based ensemble learning of Gaussian process regression models for quality-related variable prediction in process industries," Applied Soft Computing, vol. 101, pp. 107060-103981, 2021.
[18] K. B. Newhart, R. W. Holloway, A. S. Hering, and T. Cath, "Data-driven performance analyses of wastewater treatment plants: A review," Water research, vol. 157, pp. 498-513, 2019.
[19] G. Wang, J. Qiao, J. Bi, W. Li, and M. Zhou, "TL-GDBN: Growing deep belief network with transfer learning," IEEE Transactions on Automation Science and Engineering, vol. 16, no. 2, pp. 874-885, 2018.
[20] L. Cai, J. Gu, J. Ma, and Z. Jin, "Probabilistic wind power forecasting approach via instance-based transfer learning embedded gradient boosting decision trees," Energies, vol. 12, no. 1, pp. 159-178, 2019.
[21] J. Zhao, S. Shetty, and J. W. Pan, "Feature-based transfer learning for network security," In MILCOM 2017-2017 IEEE Military Communications Conference (MILCOM), pp. 17-22, 2017.
[22] R. Furfaro, R. Linares, and V. Reddy, "Space objects classification via light-curve measurements: deep convolutional neural networks and model-based transfer learning," in AMOS Technologies Conference, Maui Economic Development Board, pp. 1-17, 2018.
[23] R. Ye, Q. Dai, and M. Li, "A hybrid transfer learning algorithm incorporating TrSVM with GASEN," Pattern Recognition, vol. 92, pp. 192-202, 2019.
[24] Y. Du, Y. Liang, J. Jiang, R. J. Berry, and Y. Ozaki, "Spectral regions selection to improve prediction ability of PLS models by changeable size moving window partial least squares and searching combination moving window partial least squares," Analytica chimica acta, vol. 501, no. 2, pp. 183-191, 2004.
[25] Y. Liu, D. Huang, and Y. Li "Development of interval soft sensors using enhanced just-in-time learning and inductive confidence predictor," Industrial & engineering chemistry research, vol. 51, no. 8, pp. 3356-3367, 2012.
[26] D. Li, D. Huang, G. Yu, and Y. Liu, "Learning adaptive semi-supervised multi-output soft-sensors with co-training of heterogeneous models," Ieee Access, vol. 8, pp. 46493-46504, 2020.
[27] P. Khuwaja, S. Khowaja, K. Dev. "Adversarial learning networks for FinTech applications using heterogeneous data sources, " IEEE Internet of Things Journal. Vol. 10, no. 3, pp. 2194-2201, 2021.
[28] A. Asadi, A. Verma, K. Yang, and B. Mejabi, "Wastewater treatment aeration process optimization: A data mining approach," Journal of environmental management, vol. 203, pp. 630-639, 2017.
[29] T. Wang, C. Rudin, F. Doshi-Velez, Y. Liu, E. Klampfl, and P. MacNeille, "A bayesian framework for learning rule sets for interpretable classification," The Journal of Machine Learning Research, vol. 18, no. 1, pp. 2357-2393, 2017.
[30] J. Zeng, B. Ustun, and C. Rudin, "Interpretable classification models for recidivism prediction," Journal of the Royal Statistical Society: Series A (Statistics in Society), vol. 180, no. 3, pp. 689-722, 2017.
[31] B. Kailkhura, B. Gallagher, S. Kim, A. Hiszpanski, and T. Han, "Reliable and explainable machine-learning methods for accelerated material discovery," npj Computational Materials, vol. 5, no. 1, pp. 1-9, 2019.
[32] M. Farrés, S. Platikanov, S. Tsakovski, and R. Tauler, "Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation," Journal of Chemometrics, vol. 29, no. 10, pp. 528-536, 2015.
[33] Y. Liu, Y. Pan, and D. Huang, "Development of a novel adaptive soft-sensor using variational Bayesian PLS with accounting for online identification of key variables," Industrial & Engineering Chemistry Research, vol. 54, no. 1, pp. 338-350, 2015.
[34] A. R. Heravi, and G. Hodtani, "A new correntropy-based conjugate gradient backpropagation algorithm for improving training in neural networks," IEEE transactions on neural networks and learning systems, vol. 29, no. 12, pp. 6252-6263, 2018.
[35] C. B. Khadse, M. A. Chaudhari, and V. Borghate, "Electromagnetic compatibility estimator using scaled conjugate gradient backpropagation based artificial neural network," IEEE Transactions on Industrial Informatics, vol. 13, no. 3, pp. 1036-1045, 2016.
[36] O. Baghirli, "Comparison of Lavenberg-Marquardt, scaled conjugate gradient and Bayesian regularization backpropagation algorithms for multistep ahead wind speed forecasting using multilayer perceptron feedforward neural network," Digitala Vetenskapliga Arkivet, pp. 1-12, 2015.
[37] J. Chen, X. Wu, L. Duan, and S. Gao, "Domain adversarial reinforcement learning for partial domain adaptation," IEEE Transactions on Neural Networks and Learning Systems, pp. 1-15, 2020.
[38] X. Wang, J. Jiang, M. Gao, Z. Liu, and C. Zhao, "Activation ensemble generative adversarial network transfer learning for image classification," Journal of Electronic Imaging, vol. 30, no. 1, pp. 13016-13024, 2021.
[39] Y. Liu, L. Yuan, S. Huang, D. Huang, and B. Liu, "Integrated design of monitoring, analysis and maintenance for filamentous sludge bulking in wastewater treatment," Measurement, vol. 155, pp. 107548-107559, 2020.
[40] L. F. Deng, J. G. Wang, J. R. Su, Y. Yao, and J. L. Liu, "Root Cause Diagnosis of Plant Wide Oscillations using Kernel Granger Causality," in 2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS), pp. 812-816, 2020.
[41] X. Yuan, Y. Wang, C. Yang, Z. Ge, Z. Song, and W. Gui, "Weighted linear dynamic system for feature representation and soft sensor application in nonlinear dynamic industrial processes," IEEE Transactions on Industrial Electronics, vol. 65, no. 2, pp. 1508-1517, 2017.
[42] V. Nourani, G. Elkiran, and S. Abba, "Wastewater treatment plant performance analysis using artificial intelligence–an ensemble approach," Water Science and Technology, vol. 78, no. 10, pp. 2064-2076, 2018.
[43] D. Sahoo, Q. Pham, J. Lu, and S. Hoi, "Online deep learning: Learning deep neural networks on the fly," arXiv preprint, 2017.
[44] R. Moraffah, M. Karami, R. Guo, A. Raglin, and H. Liu, "Causal interpretability for machine learning-problems, methods and evaluation," ACM SIGKDD Explorations Newsletter, vol. 22, no. 1, pp. 18-33, 2020.

**Dong Li** received the B.E. degree in computational mathematics from He'nan Normal University, in 2016. He is currently pursuing the Ph.D. degree in control science and engineering with the Department of Automation, South China University of Technology, Guangzhou, China. His research interests include statistical theory research, machine learning, data-driven soft sensors and semi-supervised learning.

**Yiqi Liu** received B.S. and M.S. degrees in control engineering from the Chemical University of Technology, Beijing, in 2009 and a Ph.D. degree in control engineering from the South China University of Technology, Guangzhou, China in 2013. From 2013 to 2016, he was a lecturer at the South China University of Technology.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3291771



Since 2016, he has been an Associate Professor in the Department of Automation at the South China University of Technology. He is senior IEEE member and Associate Editor of IEEE ACCESS, ISA Transactions. He is the author of more than 60 articles. His research interests include soft sensors, fault diagnosis, and wastewater treatment, and he holds three patents. He was a recipient of Marie Skłodowska-Curie Individual Fellowships in 2020, the Hong Kong Scholar Award in 2016, the Chinese Scholarship Council Award in 2011 and the Deutscher Akademischer Austausch Dienst Award in 2015.

**Daoping Huang** received a B.Eng. degree in chemical automation and instruments and M.Eng. and Ph.D. degrees in automatic control theory and applications from the South China University of Technology (SCUT), Guangzhou, China in 1982, 1986, and 1998, respectively. From 1995 to 1996, he worked at the University of Gent in Belgium as a government sponsored visiting scholar. In 1982, he started his teaching and research career with the Department of Automation at SCUT. Now, as a Full Professor of the School of Automation Science and Engineering, he has authored three academic books and more than 200 conference and journal papers. His research interests cover intelligent detection, control, soft-sensing technology as well as fault diagnosis and accident prediction for industrial processes. Dr. Huang serves as the Vice Director of the Education Committee and a member of the Process Control and Application Committees at the Chinese Association of Automation. He has directed over 10 research projects. He was granted a Third Prize from the National Education Committee in 1992 and a Second Prize from the Guangdong Provincial Government in 2005.

**Chun Fai Lui** is currently pursuing his Ph.D. at City University of Hong Kong, Kowloon, Hong Kong. He graduated as the college representative with B.Eng. degree in Systems Engineering from the same university. He is also a recipient of the Hong Kong PhD Fellowship awarded by the Hong Kong government. His current research interests include data-driven process monitoring and quality improvement in intelligent manufacturing systems.

**Min Xie** (Fellow, IEEE) received the Ph.D. degree in quality technology from Linkoping University, Linkoping, Sweden, in 1987. He is currently a Chair Professor with Department of Advanced Design and Systems Engineering, City University of Hong Kong, Kowloon, Hong Kong. He has authored or coauthored about 300 peer-reviewed journal papers and 8 books on quality and reliability engineering. His research interests include reliability engineering, quality management, software reliability, and applied statistics. Prof. Xie was elected as a member of the European Academy of Sciences and Arts in 2022. He was the recipient of the prestigious Lee Kuan Yew (LKY) Research Fellowship in 1991. He has chaired many international conferences and given keynote speeches. He also serves as an Editor and Associate Editor and on the Editorial Board of many established international journals.