Multivariate soft sensor for product monitoring in the debutanizer column with deep learning

I Rosyadi*[1], A W Wardhana[1], M S Aliim[1], R Ediati[2], and D Ristiawan[1]

[1]Electrical Engineering Department, Faculty of Engineering, Jenderal Soedirman University, Purwokerto, Indonesia

[2]Agricultural Engineering Department, Faculty of Agriculture, Jenderal Soedirman University, Purwokerto, Indonesia

Email: imron.rosyadi@unsoed.ac.id

Abstract. Soft sensors have been proposed extensively for predicting ill-to-measure variables in industrial processes. In this study, we developed a multivariate soft sensor for debutanizer columns. A soft sensor was proposed to replace the chromatograph-based butane content from the debutanizer column. Recently, deep learning methods have been implemented for better feature representation of complex systems. We developed an Long Short Term Memory (LSTM)-based multivariate soft sensor that can better represent the dynamics of a debutanizer column system. Our results show that the univariate LSTM soft sensor performs better than previously proposed methods.

1. Introduction

The industrial environment includes several hardware sensors. Hardware sensors are vital for monitoring and controlling industrial processes. However, in industrial settings, hardware sensors face several challenges. Several sensors have low sampling rates when industrial processes require a high data rate. Sensor measurements such as gas chromatography can only be performed offline in the laboratory. Several measured objects are technically impossible or difficult to access; for example, when an object is located at the bottom of a reaction column. Occasionally, industrial processes experience anomalies that cannot be detected quickly through existing sensor measurements.

The sensor itself can experience disturbances or irregularities that are not easily detected. Sensor hardware can also be damaged and requires a long time to be replaced. Another challenge is that the existing sensor systems cannot provide a comprehensive picture of the state of the process. In the last three decades, attempts have been made to solve these problems and challenges using soft sensor technology. Soft or virtual sensors are software that can produce hardware-sensor-like signals.

In recent implementations, there are three soft sensor models: the white-box, black-box, and gray-box models. The white box model, also known as the model-driven soft sensor, was designed based on an understanding of industrial process phenomena. On the other hand, the black-box model is designed based on observation (measurement) data from process variables. It is also known as a data-driven soft sensor. In the black-box model, we do not necessarily understand the process. The grey box represents a combination (hybrid) of the two methods [1].

Soft sensors have been studied and applied in process industries such as oil refining [2], [3], cement [4], polymers [2], metallurgy, and bioprocesses [5]. The application of soft sensors in industry is mainly for the following purposes [1]: online prediction, process monitoring, fault detection, and sensor fault detection. In online prediction, soft sensors predict process variables that are difficult to measure, which can be caused by the sensor's low sampling rates, or because the measurement can only be performed offline. Because these variables are related to the quality of the process output, their measurement can be replaced with soft sensors. During process monitoring and fault detection, soft sensors monitor the state of the process and detect deviations from the normal state of the process. Soft sensors obtain multivariate relationships and features from the process states. They can support better decision making based on measurements. In sensor fault detection, the soft sensor serves as a backup for the hardware sensor, provides an indication in case of hardware sensor failure, and replaces the hardware sensor function if needed.

White-box soft sensors are typically developed using first-principles models (FPM). Soft sensors are built based on an extended Kalman filter and adaptive observers [6]. Black-box soft-sensor models have recently become more popular than their white-box counterparts. In initial studies, a black-box soft sensor model was developed using linear statistical learning techniques such as Principal Component Analysis (PCA) [7], Independent Component Analysis (ICA) [8], and Partial Least Square (PLS) [9]. Because most industrial processes are not linear, researchers have also developed nonlinear statistical models, such as the PCA kernel [10] and PLS kernel [11].

Artificial neural networks (ANN) have become popular in soft-sensor development because of their superiority in modeling nonlinear systems owing to their nonlinear activation function features. However, nonlinear statistical learning methods and artificial neural networks (ANN) have significant weaknesses. It exhibits local optima convergence and an inability to obtain high-level representations. Therefore, in recent years, researchers have implemented deep learning for use in soft sensors after successful application in other fields. With the growth of computational capabilities and developed models, deep learning techniques have achieved great success in many applications. The essential difference between classical machine learning and the current deep learning methods is the extraction of features. The developer should perform feature extraction using classical machine learning before applying the learning algorithm. In deep learning, features are automatically learned by an algorithm.

Based on its training method, deep learning can be grouped into three categories: supervised learning (which requires labeled data), unsupervised learning (which does not require labeled data), and semi-supervised learning (a combination of both). Based on the architecture, Deep Learning can be classified into three major groups: unsupervised pretrained networks, convolutional neural networks, and recurrent neural networks. Unsupervised pre-trained networks extract features from datasets before they are used in other deep-learning models. Convolutional neural networks (CNN) learn higher-order features by using convolution techniques. Inside the CNN, several hidden layers represent convolutional filters of various sizes and configurations. A pooling layer was added between the convolution layers to reduce the spatial dimension of the data representation. Recurrent neural networks are feed-forward artificial neural networks that transmit information at each time step. An RNN takes each vector in the input vector series and processes it sequentially to maintain the state of the system. RNN have advantages and are mainly used in time-series modeling.

Because deep learning techniques can extract high-level representations of objects, they are suitable for soft-sensor development. Some researchers have developed soft-sensor models using deep learning. Regression models using the Deep Belief Network (DBN) were developed for the crude oil distillation unit in the oil refinery process. The DBN was trained to initialize the parameters in a deep neural network [12]. The Denoising Autoencoder (DAE) was trained to obtain the initial value of the deep neural network parameters, and a deep neural network was used to estimate the oxygen content of flue gas [13]. In [14], a DBN model was used for fault detection. In [15], a hybrid variable-wise weighted stacked autoencoder (HVW-SAE) was proposed for a soft sensor in the debutanizer column process. The combination of the DBN method with kernel learning has been used for quality prediction in the polymer industry [16].

The aforementioned deep learning methods extract high-level representations in the spatial domain but ignore the temporal features in the time domain. While chemical processes are usually dynamic, the temporal features of the time-series data should be considered. Among the various deep learning algorithms, recurrent neural networks (RNN) have more advantages for extracting temporal features [17]. However, RNN has the disadvantage of gradient vanishing or gradient exploding; therefore, it cannot capture long-term dependencies. The LSTM algorithm was developed to overcome the RNN problem by adding gates to the RNN. There are three gates in the LSTM algorithm: the forget gate, which provides information from the previous step; the input gate, which provides new information from the input; and the update gate, which determines what will be retained at each time step.

In this study, we report the implementation of the LSTM algorithm as a soft sensor in a debutanizer column. We assessed its performance using RMSE, MAE, and MdAPE metrics and compared it to previously researched multivariate soft sensors.

2. Process description

The debutanizer column is a distillation series in the natural gas liquid (NGL) fractionation process. This process aims to produce products for both industrial and domestic consumers. As depicted in Figure 1, the debutanizer column seeks to remove the lighter fraction of gasoline. The input of the debutanizer column is unstabilized naphtha, with the output being LPG (propane/butane) as the top product and gasoline (pentane) as the bottom product [18].

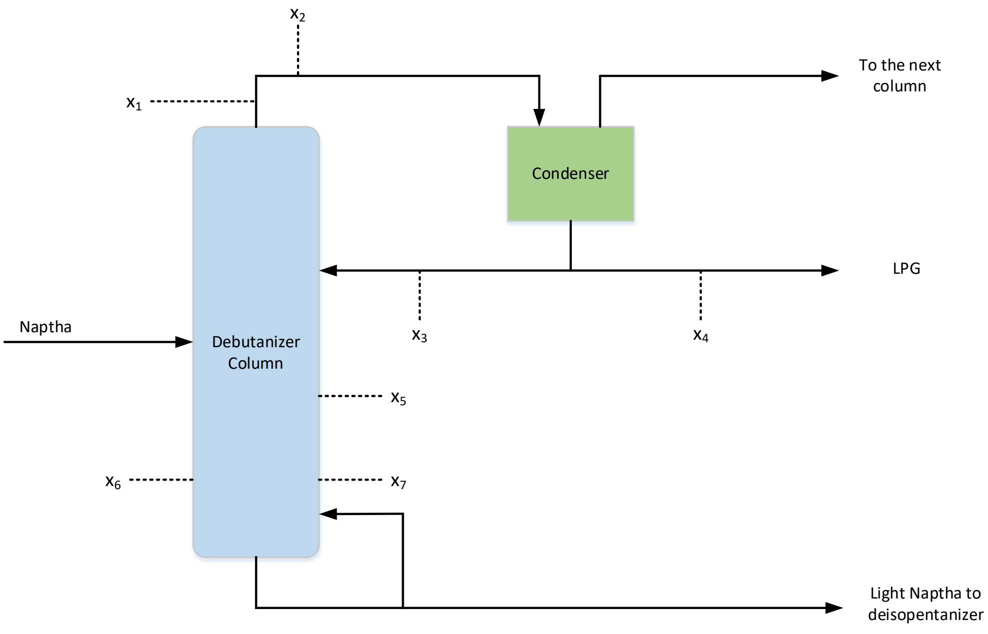

Figure 1. Diagram of debutanizer column.

Table 1. Process variables in the debutanizer column.

	Variables	Description
Input		
	x_1	Top-level temperature

	x_2	Top-level pressure
	x_3	Reflux flow
	x_4	Flow to the next process
	x_5	Temperature at 6th level
	x_6	Bottom level temperature #1
	x_7	Bottom level temperature #2
Output		
	y	Butane (C4) content at the bottom of the debutanizer

The butane (C4) content in the bottom product must be minimized to optimize process performance (in this case, maximizing LPG production). This process requires monitoring of the butane content of the bottom product. Gas chromatography was used to measure butane volume. Because the hardware sensors are located on the overhead of the next column and not on the stream of the lower debutanizer column, the measurement causes a time delay of 30-75 minutes [18].

We propose a soft sensor to overcome the time delay problem of butane measurement at the bottom of the debutanizer column without sensing it directly. We replaced the gas chromatography measurement with an inferred soft sensor from the data measurement of the other process variables. The process variables used in designing the soft sensor that determine the butane content at the bottom of the debutanizer are shown in Figure 1 and Table 1 [18].

3. Soft sensor development

RNN is a popular algorithm for time-series forecasting. The RNN network structure is illustrated in Figure 2. Let *x(t)* be the input vector at *t* and *h(t-1)* is the hidden vector at *t-1*, the next hidden state (*h(t)*) can be calculated as

$$f(t) = tanh(Ux(t) + Wh(t-1) + b) \tag{1}$$

where *U* and *W* are the corresponding weights, *b* is the bias term, and tanh(•) is the hyperbolic tangent activation function. The output *y(t)* can then be computed as an activation function of *h(t)*, with its corresponding weight and bias.

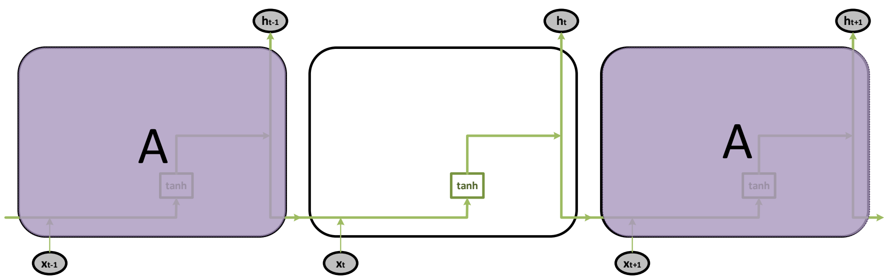

Figure 2. Structure of the RNN unit inside the RNN.

LSTM is an expansion of a typical recurrent neural network (RNN). LSTM overcomes the difficulty of RNN in learning long-term dependencies for prediction by utilizing memory cells to store long-term information. By replacing the hidden neuron units of the RNN with LSTM units, the network can handle gradient vanishing and explosion problems. As shown in Figure 3, LSTM has three additional gates: the input gate, forget gate, and output gate. It also has intermediate states. The gates and states determine the information to be held in the network.

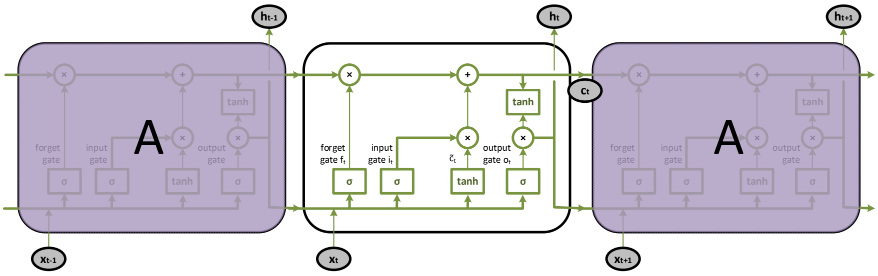

Figure 3. Structure of LSTM unit inside LSTM network.

As depicted in Figure 3, the input ($i(t)$), forget ($f(t)$), output gate ($o(t)$) and the intermediate state ($\tilde{c}(t)$) are calculated as following

$$f(t) = \sigma\left(W_{fx}x(t) + W_{fh}h(t-1) + b_f\right) \tag{2}$$
$$i(t) = \sigma(W_{ix}x(t) + W_{ih}h(t-1) + b_i) \tag{3}$$
$$o(t) = \sigma(W_{ox}x(t) + W_{oh}h(t-1) + b_o) \tag{4}$$
$$\tilde{c}(t) = tanh(W_{cx}x(t) + W_{ch}h(t-1) + b_c) \tag{5}$$

where *W's* are their corresponding weights, *b's* are their corresponding bias terms, and tanh(•) is a hyperbolic tangent activation function. The intermediate state ($c(t)$) and the hidden state ($h(t)$) in the LSTM then are updated as

$$c(t) = f(t) * c(t-1) + i(t)\tilde{C}(t) \tag{6}$$
$$h(t) = o(t) * tanh(c(t)) \tag{7}$$

where * represents the pointwise multiplication of vectors and tanh(•) is a hyperbolic tangent activation function.

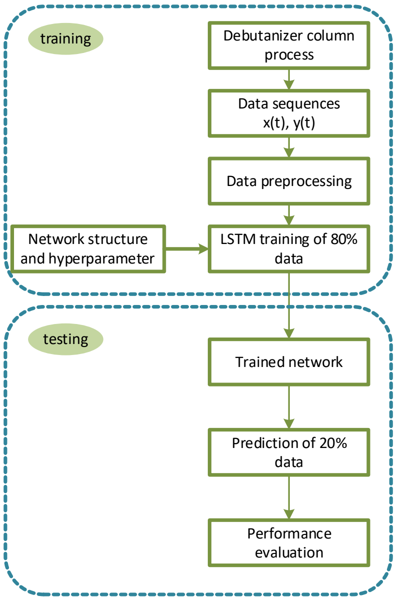

Figure 4. Flowchart of soft sensor development.

The development of the soft sensor is illustrated in Figure 4. The steps are as follows. First, we used a preprocessed dataset from sensors installed in the debutanizer column, and the butane content was measured using the chromatograph gas method [18]. The process variables measured in the debutanizer column are represented by x_1, x_2, x_3, x_4, x_5, x_6, and x_7, whereas the butane content variable is represented by y, as shown in Table 1. There total of 2394 data provided for this research. We used 80% of the dataset (1915 data) as the training data and the remaining 20% (479 data) as the testing data. We then determined the structure and hyperparameters of the LSTM and trained it on the training data. The trained network is then used to predict the output of the testing data.

To assess the soft sensor performance, three metrics were used: root-mean-squared error (RMSE), mean absolute error (MAE), and median absolute percentage error (MdAPE).

$$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}|y_i - x_i|^2} \tag{8}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \tag{9}$$

$$MdAPE = median\left(\left|\frac{x_i - y_i}{y_i}\right|\right) \cdot 100\% \tag{10}$$

where y_i and x_i are the real and estimated values, respectively, and n is the total number of data samples.

4. Result and discussion

In the training stage, we tested several hyperparameters of the LSTM network. Using the RMSE as the metric, we found that the best hyperparameters were an epoch count of 50 and a batch size of 16. We then used these hyperparameters as the testing hyperparameters. Figure 4 compares the actual sensor read and LSTM soft-sensor predictions. The prediction curve of LSTM was very close to the real value of the chromatographic measurements. To better understand the performance of the soft sensor, we evaluated its RMSE, MAE, and MdAPE values. The evaluation results are presented in Table 2.

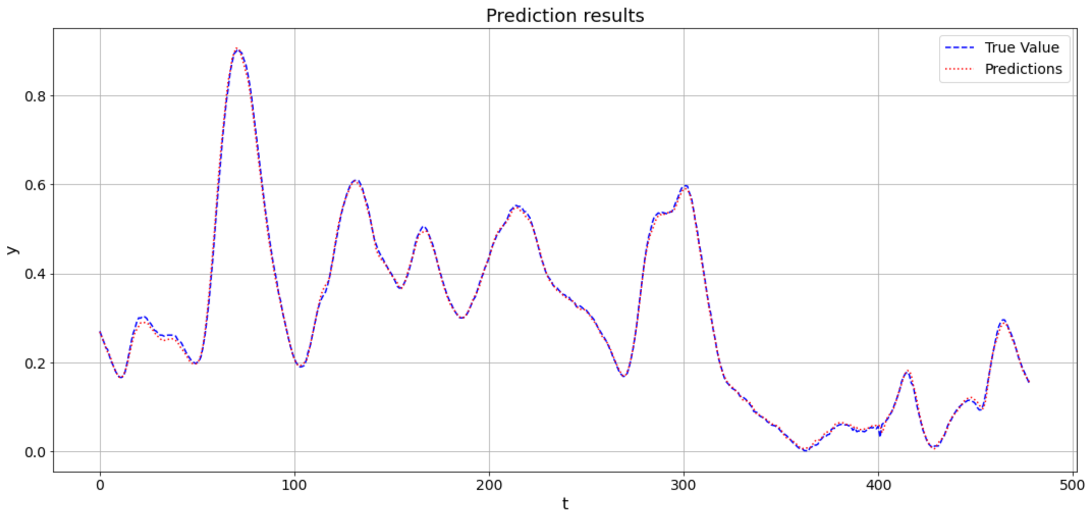

Figure 5. Comparison of actual sensor read and LSTM soft sensor predictions.

Table 2. Performance of the developed multivariate LSTM soft sensor.

	RMSE	MAE	MdAPE
Multivariate LSTM	0.00498	0.00658	1.53 %

Table 3. Metric comparison between univariate ARIMA, univariate LSTM and multivariate LSTM.

	RMSE	MAE
Multivariate LSTM	0.00498	0.00658
TCN [19]	0.0967	
AR-TCN [19]	0.0080	
Dynamic LWPLS [20]	0.0136	
Dynamic IVS-LWLPS [20]	0.0132	
Dynamic JIT-VSW [20]	0.0121	
MW-BN [21]	0.02270	
MW-PLS [21]	0.03810	
TD-BN [21]	0.01579	
SLSTM [22]	0.01599	
VWSAE-LSR [23]	0.0597	0.0471
SQAE-LSR [23]	0.0548	0.0435
SQAE-NN [23]	0.0303	0.0220
SAE-LSSVM [23]	0.0573	0.0419
VWSAE-LSSVM [23]	0.0396	0.0342
SQAE-LSSVM [23]	0.0352	0.0284

Furthermore, to objectively measure the performance of our soft sensor, we compared it with that of other soft sensors. Table 3 presents the results. All networks in the table used a dataset similar to that provided by Fortuna et al. [18]. Our multivariate LSTM offered the best prediction performance owing to its low RMSE and MAE parameters.

5. Conclusion

In this study, an LSTM network was developed to model a soft sensor in a debutanizer column. The hidden states of the LSTM network store the dynamic representation of multiple sensor variables and predict the soft sensor output. The multivariate LSTM soft-sensor architecture performed well, with an RMSE of 0.00498, MAE value of 0.00658, and MdAPE value of 1.53%. The results show that the proposed network outperforms the recently developed soft sensor in predicting the butane product in a debutanizer column.

References

[1] P. Kadlec, B. Gabrys, and S. Strandt, “Data-driven soft sensors in the process industry,” *Computers & chemical engineering*, vol. 33, no. 4, pp. 795–814, 2009, doi: 10.1016/j.compchemeng.2008.12.012.

[2] J. Shi and X.-G. Liu, “Product quality prediction by a neural soft-sensor based on MSA and PCA,” *International Journal of Automation and Computing*, vol. 3, no. 1, pp. 17–22, 2006, doi: 10.1007/s11633-006-0017-9.

[3] Y. Wang, C. Chen, and X. Yan, “Structure and weight optimization of neural network based on CPA-MLR and its application in naphtha dry point soft sensor,” *Neural Computing and Applications*, vol. 22, no. 1, pp. 75–82, 2013, doi: 10.1007/s00521-012-1044-9.

[4] A. K. Pani and H. K. Mohanta, “Online monitoring of cement clinker quality using multivariate statistics and Takagi-Sugeno fuzzy-inference technique,” *Control Engineering Practice*, vol. 57, pp. 1–17, 2016, doi: 10.1016/j.conengprac.2016.08.011.

[5] V. Steinwandter, T. Zahel, P. Sagmeister, and C. Herwig, “Propagation of measurement accuracy to biomass soft-sensor estimation and control quality,” *Anal Bioanal Chem*, vol. 409, no. 3, pp. 693–706, Jan. 2017, doi: 10.1007/s00216-016-9711-9.

[6] A. J. de Assis and R. M. Filho, “Soft sensors development for online bioreactor state estimation,” *Computers & Chemical Engineering*, vol. 24, no. 2, pp. 1099–1103, Jul. 2000, doi: 10.1016/S0098-1354(00)00489-0.

[7] Q. Jiang and X. Yan, “Monitoring multi-mode plant-wide processes by using mutual information-based multi-block PCA, joint probability, and Bayesian inference,” *Chemometrics and Intelligent Laboratory Systems*, vol. 136, pp. 121–137, Aug. 2014, doi: 10.1016/j.chemolab.2014.05.012.

[8] Q. Jiang, X. Yan, and J. Li, “PCA-ICA Integrated with Bayesian Method for Non-Gaussian Fault Diagnosis,” *Ind. Eng. Chem. Res.*, vol. 55, no. 17, pp. 4979–4986, May 2016, doi: 10.1021/acs.iecr.5b04023.

[9] Z. Ge, Z. Song, F. Gao, and P. Wang, “Information-Transfer PLS Model for Quality Prediction in Transition Periods of Batch Processes,” *Ind. Eng. Chem. Res.*, vol. 52, no. 15, pp. 5507–5511, Apr. 2013, doi: 10.1021/ie303267u.

[10] X. Yuan, Z. Ge, and Z. Song, “Locally Weighted Kernel Principal Component Regression Model for Soft Sensing of Nonlinear Time-Variant Processes,” *Ind. Eng. Chem. Res.*, vol. 53, no. 35, pp. 13736–13749, Sep. 2014, doi: 10.1021/ie4041252.

[11] X. Zhang, M. Kano, and Y. Li, “Locally weighted kernel partial least squares regression based on sparse nonlinear features for virtual sensing of nonlinear time-varying processes,” *Computers & Chemical Engineering*, vol. 104, pp. 164–171, Sep. 2017, doi: 10.1016/j.compchemeng.2017.04.014.

[12] 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, Mar. 2014, doi: 10.1016/j.jprocont.2014.01.012.

[13] 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, May 2017, doi: 10.1109/TIE.2016.2622668.

[14] J. Yu and X. Yan, “Layer-by-Layer Enhancement Strategy of Favorable Features of the Deep Belief Network for Industrial Process Monitoring,” *Ind. Eng. Chem. Res.*, vol. 57, no. 45, pp. 15479–15490, Nov. 2018, doi: 10.1021/acs.iecr.8b04689.

[15] X. Yuan, C. Ou, Y. Wang, C. Yang, and W. Gui, “Deep quality-related feature extraction for soft sensing modeling: A deep learning approach with hybrid VW-SAE,” *Neurocomputing*, vol. 396, pp. 375–382, Jul. 2020, doi: 10.1016/j.neucom.2018.11.107.

[16] Y. Liu, C. Yang, Z. Gao, and Y. Yao, “Ensemble deep kernel learning with application to quality prediction in industrial polymerization processes,” *Chemometrics and Intelligent Laboratory Systems*, vol. 174, pp. 15–21, Mar. 2018, doi: 10.1016/j.chemolab.2018.01.008.

[17] A. Graves, A. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 6645–6649. doi: 10.1109/ICASSP.2013.6638947.

[18] L. Fortuna, S. Graziani, and M. G. Xibilia, “Soft sensors for product quality monitoring in debutanizer distillation columns,” *Control Engineering Practice*, vol. 13, no. 4, pp. 499–508, Apr. 2005, doi: 10.1016/j.conengprac.2004.04.013.

[19] 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, Sep. 2021, doi: 10.1109/JSEN.2021.3096215.

[20] B. Pan *et al.*, “Just-in-time learning based soft sensor with variable selection and weighting optimized by evolutionary optimization for quality prediction of nonlinear processes,” *Chemical Engineering Research and Design*, vol. 144, pp. 285–299, Apr. 2019, doi: 10.1016/j.cherd.2019.02.004.

[21] Z. Liu, Z. Ge, G. Chen, and Z. Song, "Adaptive soft sensors for quality prediction under the framework of Bayesian network," *Control Engineering Practice*, vol. 72, pp. 19–28, Mar. 2018, doi: 10.1016/j.conengprac.2017.10.018.
[22] X. Yuan, L. Li, and Y. Wang, "Nonlinear Dynamic Soft Sensor Modeling With Supervised Long Short-Term Memory Network," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3168–3176, May 2020, doi: 10.1109/TII.2019.2902129.
[23] X. Yuan, J. Zhou, B. Huang, Y. Wang, C. Yang, and W. Gui, "Hierarchical Quality-Relevant Feature Representation for Soft Sensor Modeling: A Novel Deep Learning Strategy," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 3721–3730, Jun. 2020, doi: 10.1109/TII.2019.2938890.