Engineering Applications of Artificial Intelligence 121 (2023) 105988

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Survey paper

The role of artificial intelligence-driven soft sensors in advanced sustainable process industries: A critical review

Yasith S. Perera [a,b], D.A.A.C. Ratnaweera [c], Chamila H. Dasanayaka [d], Chamil Abeykoon [a,*]

[a] Northwest Composites Centre and Aerospace Research Institute, Department of Materials, Faculty of Science and Engineering, The University of Manchester, Oxford Road, M13 9PL, Manchester, UK
[b] Department of Textile and Apparel Engineering, Faculty of Engineering, University of Moratuwa, Sri Lanka
[c] Department of Mechanical Engineering, University of Peradeniya, KY 20400, Sri Lanka
[d] Institute of Business Industry and Leadership, University of Cumbria, Paternoster Row, Bowerham Rd, Lancaster, LA1 3JD, UK

ARTICLE INFO

Keywords:
Soft sensor
Virtual sensor
Sustainable development
Process monitoring
Artificial intelligence
Machine learning
Process industry
Data-driven modeling

ABSTRACT

With the predicted depletion of natural resources and alarming environmental issues, sustainable development has become a popular as well as a much-needed concept in modern process industries. Hence, manufacturers are quite keen on adopting novel process monitoring techniques to enhance product quality and process efficiency while minimizing possible adverse environmental impacts. Hardware sensors are employed in process industries to aid process monitoring and control, but they are associated with many limitations such as disturbances to the process flow, measurement delays, frequent need for maintenance, and high capital costs. As a result, soft sensors have become an attractive alternative for predicting quality-related parameters that are 'hard-to-measure' using hardware sensors. Due to their promising features over hardware counterparts, they have been employed across different process industries. This article attempts to explore the state-of-the-art artificial intelligence (Al)-driven soft sensors designed for process industries and their role in achieving the goal of sustainable development. First, a general introduction is given to soft sensors, their applications in different process industries, and their significance in achieving sustainable development goals. AI-based soft sensing algorithms are then introduced. Next, a discussion on how AI-driven soft sensors contribute toward different sustainable manufacturing strategies of process industries is provided. This is followed by a critical review of the most recent state-of-the-art AI-based soft sensors reported in the literature. Here, the use of powerful AI-based algorithms for addressing the limitations of traditional algorithms, that restrict the soft sensor performance is discussed. Finally, the challenges and limitations associated with the current soft sensor design, application, and maintenance aspects are discussed with possible future directions for designing more intelligent and smart soft sensing technologies to cater the future industrial needs.

1. Introduction

Most of the modern process industries are highly energy intensive and hence they are responsible for consuming a significant proportion of the annual global energy production. Therefore, process industries significantly contribute to global warming through the excessive burning of fossil fuels. Moreover, the waste, by-products, and toxic gases generated by these industries lead to environmental pollution. Consequently, global authorities have been imposing tighter environmental rules and regulations over the past decade, and all process industries across the world are expected to adhere to them. Non-compliance to these requirements will lead to legal actions being taken against process industries, an increase in production costs, and reduced consumer demand for products. These environmental concerns, diminishing non-renewable resources, legal requirements, high energy costs, and consumer demand for environmentally friendly products, are driving the modern process industries toward sustainable development (Giret et al., 2015). The concepts of circular economy and cleaner production have emerged as strategies that aim to achieve sustainable development through energy conservation, emission reduction, and improved production efficiency (Henao-Hernández et al., 2019; Hens et al., 2018). These requirements have increased the importance of the process monitoring aspect (and hence the demand for them as well) in the process industries. In any industrial application, process monitoring plays a crucial role in monitoring the process's health to ensure that the process functions within the desired limits (Abeykoon, 2018). This allows material wastage, environmental pollution, and

* Corresponding author.
E-mail addresses: yasith.perera@manchester.ac.uk, yasiths@uom.lk (Y.S. Perera), asanga@eng.pdn.ac.lk (D.A.A.C. Ratnaweera), chamila.dasanayaka@uni.cumbria.ac.uk (C.H. Dasanayaka), chamil.abeykoon@manchester.ac.uk (C. Abeykoon).

https://doi.org/10.1016/j.engappai.2023.105988
Received 17 June 2022; Received in revised form 24 January 2023; Accepted 9 February 2023
Available online 11 March 2023

energy consumption to be controlled while achieving a product with the desired quality.

To enable process monitoring and control, the most common practice is to employ hardware sensors at desired locations of the process (Kadlec et al., 2009). However, hardware sensors may not be suitable for certain applications because of many factors including harsh/hostile working environments which lead to frequent service requirements, disturbances to the process flow and product quality, measurement delays, access requirements, and high costs. This makes it difficult or impossible to measure certain process variables directly using hardware sensors. These 'hard-to-measure' process variables are mostly related to product quality and due to the above-mentioned limitations, they are normally determined by offline laboratory analyses (Warne et al., 2004). Such offline measurements can introduce significant measurement delays and discontinuity. Hence, it is impossible to carry out real-time adjustments to maintain the product within the desired quality constraints, and this may lead to increased material wastage, excessive energy usage, environmental pollution, process degradation, and so forth. To address these issues, some researchers have investigated the possibility of using mathematical models of processes to estimate these 'hard-to-measure' quality parameters, so that hardware sensors can be replaced by such models. This has led to the concept of software-based sensors or soft sensors.

Early soft sensor applications utilized first-principles models or statistical and/or traditional machine learning techniques to estimate the 'hard-to-measure' quality parameters (Kadlec et al., 2009). These traditional models were able to provide real-time predictions of the desired parameters while addressing the limitations of their hardware counterparts but suffered from serious limitations. For example, first-principles models were derived based on numerous assumptions and consequently, they failed to capture actual dynamics in industrial processes. Traditional statistical and machine learning techniques failed to effectively capture spatial and temporal variations in the process data, leading to poor predictive performance. It would not be possible to monitor the process's health and make real-time quality control decisions if the measurements provided by the sensors are not accurate, and this would act as a barrier against the process industries moving toward sustainable development. Hence, researchers were forced to investigate more advanced state-of-the-art algorithms to be used in soft sensor development. Consequently, with the advancements in artificial intelligence (AI), researchers have shifted their focus from traditional algorithms to more advanced AI algorithms to improve the predictive performance of soft sensors to better aid process monitoring and control in process industries.

1.1. Related literature review

First, the previous review works on soft sensor applications are investigated. Abeykoon (2018) provided a comprehensive review of soft sensor applications in the polymer processing industry, covering polymer extrusion, polymerization, and a few other processes. Similarly, Al-Jamimi et al. (2018) reviewed machine learning-based soft sensing solutions for the desulfurization of oil products. However, both studies only reviewed traditional modeling techniques and did not discuss the latest AI-based techniques published in recent years. Zhu et al. (2020) discussed data-driven soft sensors used in industrial fermentation processes. Here, they reviewed some of the latest AI-based techniques such as deep neural networks, fuzzy logic, and evolutionary algorithms, in addition to the traditional techniques. Kadlec et al. (2009) provided a comprehensive review of data-driven soft sensors employed in the process industry. The authors discussed soft sensor design aspects, applications, and algorithms. In a later study, Kadlec et al. (2011) investigated different adaptation mechanisms reported in the literature, for constructing adaptive soft sensors. Although these review articles are more than a decade old, they provide valuable insights that are still relevant to today's process industries. However, they only present traditional statistical and machine learning algorithms that were widely used for soft sensor construction at the time. The reviews by Khatibisepehr et al. (2013), Souza et al. (2015), Liu and Xie (2020), Curreri et al. (2020), and Jiang et al. (2021) mainly discuss the advances in the soft sensor design aspects and their focus is not on the use of the latest AI-based techniques. Kong et al. (2022) critically analyzed the conventional as well as deep learning-based techniques for extracting latent features from process data and proposed a novel framework that combines the excellent model capacity of the deep learning-based techniques with the model interpretability and efficiency of the conventional techniques. However, focus of this review is limited to latent feature extraction methods, and do not discuss other challenges associated with soft sensor design.

Based on the investigation of previously published review articles on soft sensors, the following key limitations or gaps can be identified, which call for a further review to be conducted.

- The prior reviews discuss only traditional modeling techniques and lack focus on the latest AI-based techniques.
- Some reviews only focus on applications from a single industry.
- None of the prior reviews discuss the role of soft sensors in achieving sustainable development nor the contribution of advanced AI-based soft computing algorithms in achieving this goal.

Kadlec et al. (2009) and Souza et al. (2015) discussed some characteristics of process data that pose challenges for soft sensor developers. Problems such as missing data, varying sampling rates, drifting data, and outliers adversely affect the predictive performance of soft sensors. These reviews discussed some of the solutions in the literature to address these challenges, but they were only limited to traditional techniques. However, modern process industries require sensors with high robustness and precision, and solutions based on traditional modeling techniques may not be sufficient. The most recent research works investigated various AI-based techniques to address these limitations with the aim of producing soft sensors with high prediction accuracy. Another limitation of some of the past reviews is that they only focus on soft sensor applications from a single industry (Abeykoon, 2018; Al-Jamimi et al., 2018; Zhu et al., 2020). It is quite reasonable to review work related to only one type of process industry as soft sensors are generally developed according to the requirements of that industry. However, some of the problems/challenges in one industry may be common to other industries as well. Hence, it should be useful to investigate solutions from multiple industries. With the current trend for achieving sustainable development, process industries need more sophisticated process monitoring techniques. Hence, soft sensors play a key role in implementing sustainable manufacturing strategies. The existing studies do not investigate the involvement of soft sensors in this aspect, and how the advanced powerful AI tools can enhance the contribution of soft sensors toward sustainable development.

Therefore, there is a need for a critical review that looks at the latest research on soft sensors that are based on state-of-the-art AI algorithms, and their role in guiding the process industries toward sustainable development. Hence, this study aims to contribute to the existing literature by filling these gaps and providing directions for future researchers and soft sensor designers. This paper is organized as follows:

Section 1 of this paper discusses the importance of process monitoring in achieving sustainable development goals of process industries. Moreover, the need for a critical review is discussed followed by the methodology followed in conducting this study. Section 2 introduces soft sensors and the process industries that use them. In Section 3, the commonly used classes of AI algorithms for soft sensing applications are discussed. Section 4 discusses how modern soft sensors based on the latest AI algorithms contribute toward sustainable development in process industries. A critical review of the state-of-the-art AI-driven soft sensors employed in process industries is presented in Section 5. This section provides an overview of some of the different classes of problems/challenges in developing soft sensors (i.e., missing data, varying sampling rates, small datasets, dimensionality reduction, adapting to varying process conditions, extracting temporal and spatial features, and model interpretability). Then, the most recent AI-based solutions for tackling these problems are discussed in detail and compared with the existing traditional solutions. Section 6 discusses the existing challenges in soft sensor design, application, and maintenance aspects and the future trends of soft sensor applications in process industries. Finally, Section 7 concludes the paper with a set of conclusions drawn from the review conducted in this study.

1.2. Method

An online search was carried out to identify the most relevant articles on state-of-the-art AI-based soft sensing solutions. Google Scholar was used as the primary search database as it provides extensive coverage of scholarly articles including journal articles, conference proceedings, and dissertations. IEEE Xplore, ScienceDirect, and Wiley Online Library were also used to ensure that the search was more comprehensive.

Due to the large volume of articles on soft sensors available in the literature, the scope of Section 5 of this study was restricted to soft sensor applications from three process industries. Polymer processing, petroleum refining, and pharmaceutical industries were chosen as they dominate the soft sensor applications reported in the literature as evident from previous reviews (Abeykoon, 2018; Al-Jamimi et al., 2018; Zhu et al., 2020). Hence, relevant combinations of keywords were used to search for articles related to these industries. *'polymer processing' AND 'soft sensor'*, *'oil refinery' OR 'petroleum refining industry' AND 'soft sensor'*, and *'pharmaceutical' AND 'soft sensor'* constitute the different keyword combinations used to search for the relevant articles. Then, the final inclusion and exclusion criteria were used to further refine the search.

- Is the year of publication within the period 2018–2022?
- Is the soft sensor designed for online prediction tasks?
- What is the AI-based technique introduced?

As per the above criteria, only the articles that were published within the period 2018–2022 were first selected, to obtain the most recent state-of-the-art studies. According to Kadlec et al. (2009), soft sensing applications can be divided into three main categories: 'online prediction', 'process monitoring and process fault detection', and 'sensor fault detection and reconstruction'. Among these, online prediction can be identified as the most popular application, which involves the real-time estimation of quality-related parameters. Hence, articles on online prediction soft sensors were selected to further narrow down the scope of this paper. Finally, the AI-based techniques introduced in the articles were assessed using their abstracts. Articles based on traditional algorithms were excluded and only the articles based on AI techniques were chosen (Section 3.1 describes the classes of AI-based algorithms discussed in this study). Papers were chosen such that no two papers with the same technique were selected for each class of problem discussed in Section 5. This was done to enhance the diversity of different AI-based solutions discussed in this paper.

2. Soft sensors in process industries

A soft sensor is a widely used term to describe software-based sensors, which is a technique of estimating 'hard-to-measure' quality-related parameters in industrial processes. They are also known as virtual sensors, inferential estimators/models, (Abeykoon, 2018; Fortuna et al., 2007), and observer-based sensors (Goodwin, 2000). Soft sensors are based on mathematical or empirical models that map a set of input process variables to a quality parameter so that 'hard-to-measure' quality parameters in process industries can be accurately estimated using a set of 'easy-to-measure' input process variables. Fig. 1 illustrates the conceptual arrangement of a soft sensor used in online prediction applications.

There are mainly two soft sensor categories: model-driven and data-driven soft sensors (Kadlec et al., 2009). Model-driven soft sensors are designed using equations derived from physical or chemical principles related to the process under consideration, hence are also known as first-principles models or mechanistic models. Since the process background or the internal mechanisms of the first-principles models are known, the term 'white-box model' is also used. They are highly complex in nature and they model the physical or chemical behaviors of the processes and are mainly designed for steady-state operating conditions. However, most of the processes in process industries have a highly dynamic behavior and quite often they deviate from the steady state. Hence model-driven soft sensors may not be suitable for predicting parameters related to such dynamic processes. Furthermore, the designing of model-driven soft sensors is time-consuming and requires expert knowledge of the process.

Data-driven soft sensors, which are based on empirical models built using real process data, provide a solution to the limitations associated with their model-driven counterparts. Since they are designed using real process data, data-driven soft sensors are capable of accurately modeling the process dynamics, which leads to better predictive performance compared to model-driven soft sensors. However, these data-driven models may also operate mainly within the processing window that the data was collected to train those models and if a process is running outside that processing window, it is difficult to guarantee the performance of these models as well. Unlike the first-principles models, the internal mechanisms of data-driven models are hidden, hence they are known as 'black-box' models. In addition to the model-driven and data-driven approaches, hybrid models which combine both these components into their model structure can also be found in the literature, and they are known as 'gray-box' models (Kadlec et al., 2009). The benefits and limitations of model-driven, data-driven, and gray-box soft sensors can be found in the work by Lahiri (2017).

Today, soft sensors are widely being employed in process industries, due to the various advantages that they can offer. The ability to operate in harsh working environments where hardware sensors are unsuitable to be employed, the ability to make accurate real-time estimations of the quality parameters without measurement delays, the ability to be implemented on existing hardware without any additional investments, the ability to replace expensive hardware sensors at a less cost, and the ease of maintenance are some of the benefits of soft sensors which make them attractive alternatives to the conventional process monitoring techniques used in process industries (Abeykoon, 2016b, 2018). Table 1 summarizes some of the latest works reported on a variety of soft sensing applications across different process industries. The ability of soft sensors to make accurate real-time estimations of quality parameters allows them to be incorporated into feedback control systems to achieve real-time quality control of the processes (see Fig. 2).

3. AI-based algorithms used in soft sensors

3.1. Introduction to soft sensing algorithms

Obviously, soft computing algorithms play a crucial part in the soft sensor development process. The key role of such algorithms is to establish the model associated with a soft sensor, which maps the 'easy-to-measure' input process variables to the 'hard-to-measure' quality parameters. As explained in Section 2, white-box models employ physics-based mathematical models derived using the first-principles knowledge about the process under consideration. In contrast, data-driven soft sensors use a wide range of computing algorithms. This section discusses the algorithms employed in data-driven soft sensors with a focus on the latest AI algorithms.

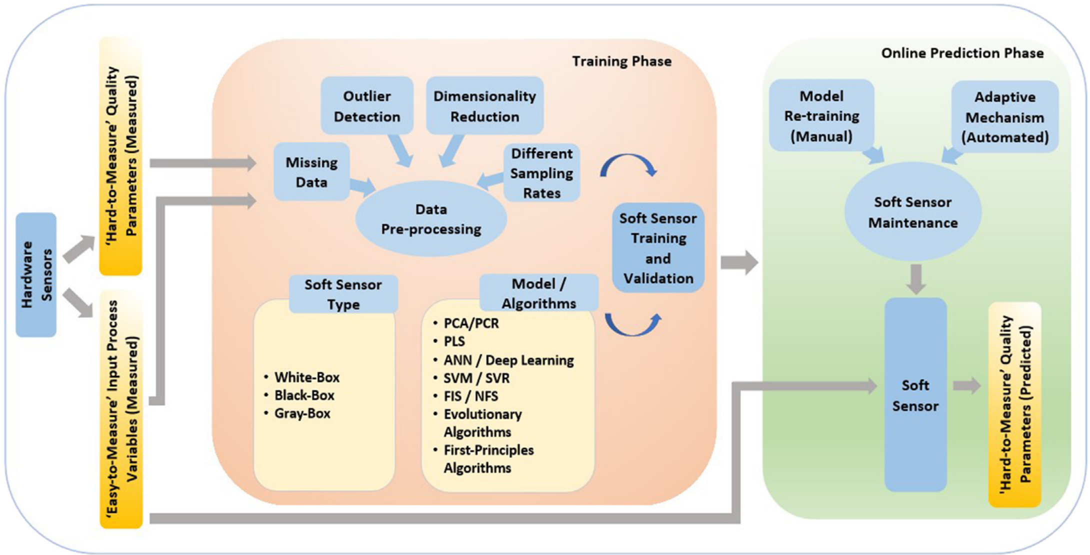

Fig. 1. Conceptual arrangement of an online prediction soft sensor.

Table 1
Soft sensor applications in process industries.

Publication	Industry	Application
Zhao et al. (2021)	Cement industry	Prediction of the content of free calcium oxide in a cement clinker
Farahani et al. (2021)	Power plant	Prediction of the active power and fuel flow
Wang et al. (2019b)	Steel industry	Coke dry quenching operation prediction
Yan et al. (2020)	Wastewater treatment plant	Prediction of the total Kjeldahl nitrogen
Phatwong and Koolpiruck (2019)	Pulp paper industry	Kappa number prediction of a pulp digester
Sun and Ge (2019)	Ammonia synthesis process	Prediction of the CO_2 concentration in a CO_2 absorption column
Liu et al. (2021a)	Polymer processing industry	Prediction of the melt flow index (MFI) in a polypropylene polymerization process
Guo et al. (2020b)	Petrochemical industry	Prediction of the butane content in a debutanizer column of an oil refinery
Qiu et al. (2021)	Pharmaceutical industry	Prediction of the penicillin concentration in a penicillin fermentation process
Meng et al. (2019)	Food processing industry	Prediction of the mother liquid purity and supersaturation in a cane sugar crystallization process

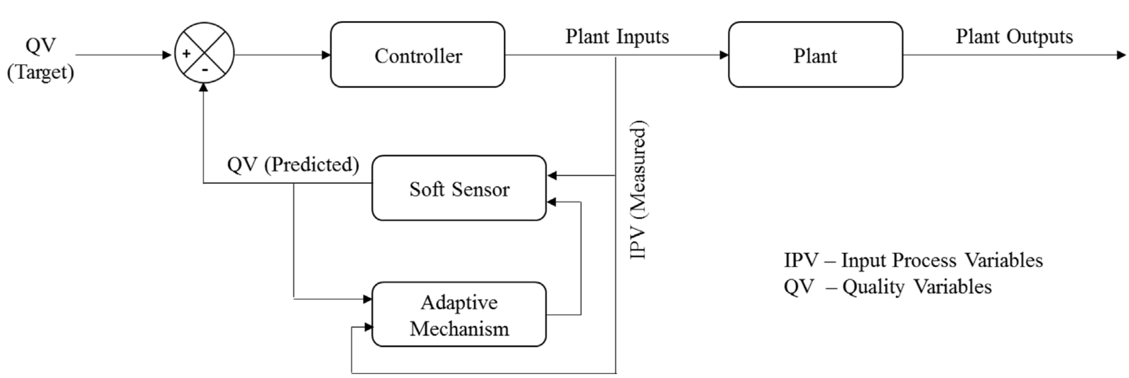

Fig. 2. A soft sensor used in a feedback control system for real-time process control.

Multiple linear regression (MLR) is a simple and useful technique for predicting the behavior of response variables based on a set of independent variables (Tobias, 1995). However, the complex nature of the processes in process industries involves a large number of independent variables (i.e., 'easy-to-measure' input process variables) which are highly redundant (i.e., collinear) and their relationship to the response variables (i.e., 'hard-to-measure' quality parameters) may not be well understood. MLR is not capable of handling such highly correlated data. Furthermore, due to the limitations such as the long measurement delays of hardware sensors that are used to measure the quality parameters, only a limited number of observations are made compared to the large number of input process variables. If MLR is used in designing soft sensors under these conditions, it is highly likely that the predictive performance of those sensors will be poor and will lead to overfitting problems (Tobias, 1995). Overfitting occurs when a soft sensor makes good estimations on the data that was used to train it but fails to make accurate predictions on a new set of data that it had not seen before. Partial least squares (PLS) is a widely used statistical algorithm to address the issue of collinearity. The PLS transforms the correlated variables into a new set of variables (i.e., latent space) that are uncorrelated or orthogonal to each other. This transformation reduces the dimensions of the input space, without losing much of the information in the original data. Then it constructs the mapping between the latent spaces of both independent and dependent variables. Today, researchers are investigating more advanced algorithms for such regression problems to achieve better accuracy. A wide range of state-of-the-art soft computing algorithms based on statistical methods, machine learning, and AI techniques are used in designing soft sensors for process industries. Algorithms such as the principal component analysis (PCA), PLS, artificial neural network (ANN) and deep learning, support vector machine/regression (SVM/SVR), fuzzy inference system (FIS) and neuro fuzzy system (NFS), and evolutionary algorithms are heavily used in soft sensor development.

As this study focuses on AI algorithms, it is important to identify the distinction between AI algorithms and traditional machine learning algorithms. Arthur Samuel defined machine learning as the field of study that allows computers to learn from experience, without being explicitly programmed (Samuel, 1959). Later, Tom Mitchell provided a more formal definition which states that: *"a machine learns with respect to a particular task T, performance metric P, and type of experience E, if the system reliably improves its performance P at task T, following experience E"*. (Mitchell, 1997). In contrast, AI is considered as the ability of a computer system to mimic human cognitive functions (Microsoft, 2022). Therefore, machine learning is generally considered as an application or a subfield of AI (Cioffi et al., 2020). This makes it difficult to make a clear distinction between machine learning and AI, and consequently, these terms are used interchangeably in the literature, which may cause some confusion.

With the rise of deep learning, more advanced machine learning algorithms have come into existence, and are increasingly being used in AI applications. Soft sensor designers are also shifting from traditional machine learning algorithms such as the PCA, PLS, and SVR, to more sophisticated algorithms such as deep neural networks. Unlike traditional algorithms, these advanced deep neural networks can effectively mimic the cognitive functions of humans, leading to excellent performance in a wide range of AI applications. Hence, in this study, the focus is given to the most commonly used advanced neural network algorithms in soft sensor applications. In addition to this, fuzzy-based algorithms and evolutionary algorithms are also discussed in brief as these algorithms also mimic the intelligent behaviors of humans and animals. The aim of this section is not to provide a theoretical mathematical foundation of the algorithms, but to familiarize the reader with the different AI algorithms used in soft sensing applications and discuss how these algorithms have succeeded in developing soft sensors with excellent performance while addressing the challenges in modeling industrial processes.

3.2. Neural networks

An ANN is a computing algorithm that mimics the way that the human brain analyzes and processes information. ANNs were inspired by the network of neurons in the human brain (Agatonovic-Kustrin and Beresford, 2000). In an ANN, the biological neurons are modeled by a set of processing units known as artificial neurons. The ability of ANNs to capture nonlinear relationships present in the data has made them an attractive solution in soft sensor design. However, despite their wide use, soft sensors based on ANNs suffer from several drawbacks during the soft sensor design stage. During the training phase, ANNs tend to get trapped in local minima, which makes the soft sensor development process challenging. Optimization of the neural network architecture has also been another challenge and it is usually carried out through an ad-hoc approach. Consequently, most of the feedforward ANNs are limited to shallow structures which limits the ability of the soft sensor to make accurate predictions. Increasing the structure complexity can lead to the overfitting problem, which reduces the generalization ability of the soft sensor (Kadlec et al., 2009).

Early works on soft sensors involving ANNs were focused on shallow feedforward architectures such as the multilayer perceptron (MLP) and radial basis function (RBF) neural networks (Sliskovic et al., 2004). Huang et al. (2006) introduced the extreme learning machine (ELM), which exhibited better performance and faster learning ability using the Moore inverse learning algorithm, compared to the traditional MLPs based on the backpropagation algorithm. The ELM is a feedforward neural network with a single hidden layer. The functional link neural network (FLNN) is another shallow neural network architecture, which simplifies the structure of backpropagation neural networks (Pao, 1989). An FLNN has no hidden layers and contains only an input layer and an output layer. Although these traditional shallow neural network architectures may perform poorly on industrial data, some of their modified and improved versions have successfully been used in recent soft sensor applications due to their computational efficiency. He et al. (2015) used a double parallel ELM in developing soft sensors for modeling complex processes in the chemical industry due to its ability to provide accurate results with a fast response time, compared to the backpropagation neural networks. Geng et al. (2017) introduced a novel self-organizing ELM based on the biological neuron-glia interaction principle, to improve the generalization performance and stability of the traditional ELM, and this novel algorithm was employed in a soft sensor for modeling a purified terephthalic acid (PTA) process. In another work, Tian et al. (2020) introduced a soft sensor with a regularization-based FLNN and tested its performance on a PTA process. Considering the better generalization performance and learning speed of the Moore inverse algorithm used in ELMs, the authors used the same algorithm with the FLNN.

With the rise of deep learning, there has been a trend of using different deeper neural network architectures to overcome the limitations of shallow neural networks. Some recent works on soft sensor applications focused on deep neural network architectures such as the recurrent neural network (RNN) (Kataria and Singh, 2018), convolutional neural network (CNN) (Yuan et al., 2020c), generative adversarial network (GAN) (He et al., 2020a; Wang and Liu, 2020), and stacked autoencoder (SAE) (Yuan et al., 2020d) can be found in the literature. RNNs are feedforward networks with feedback connections. They are designed to process a sequence of data that varies with time. The feedback connections enable them to retain useful information from previous time steps which are then processed with the data from the current time step. Consequently, they can extract and learn temporal patterns from data, which might not be highly effective with conventional feedforward network architectures. This makes them suitable candidates for soft sensing applications in process industries as these processes are quite dynamic and time-dependent in nature. However, RNNs are not that capable of handling long-term dependencies because of the problem of vanishing or exploding gradients during training (Bengio et al., 1994). To overcome this issue, variants of the RNN such as the long short-term memory (LSTM) neural network, the gated recurrent unit (GRU), and the echo state network (ESN) have been introduced.

LSTMs introduce a memory cell, which can process data with time lags. These memory cells are composed of three gates (i.e., input, output, and forget gates), which can control how information is remembered and forgotten. This mechanism, which is absent in an RNN enables the LSTM to learn long-term dependencies. This property is quite useful for implementing soft sensors as process data is highly time-dependent. For example, in industrial processes, when a process control variable is changed, the quality parameter may not change instantly. Instead, it may show a delayed response. An RNN may fail to capture such dynamics due to its inability to handle long-term dependencies. Hence, the LSTM networks provide a promising alternative. Some recent research works associated with LSTM-based soft sensors can also be found in the literature (Pisa et al., 2019; Shen et al., 2020; Sun et al., 2021). A GRU is a simplified version of the LSTM, which is computationally less expensive than the LSTM (Cho et al., 2014). This high computational efficiency has been achieved by using only two gates known as the reset gate and the update gate. The update gate is designed by combining the forget and input gates of the LSTM. This simpler internal structure of the GRUs compared to the LSTMs, with the ability to learn long-term dependencies has made GRUs an attractive solution in soft sensor applications that require faster computation (Guo and Liu, 2021). An ESN is another variant of the RNN that consists of an input layer, a hidden layer (i.e., a reservoir), and an output layer. The reservoir has a nonlinear recurrent structure and has highly sparse connections. This nonlinear recurrent structure enables the dynamic characteristics of industrial process data to be stored in this reservoir, and hence ESNs are a favorable candidate for soft sensor development (He et al., 2020b). In ESNs only the weights of the output layer are trainable, and hence they are quite fast during the training phase compared to RNNs and LSTMs. However, hyperparameter tuning is more challenging, as there are a large number of hyperparameters that significantly affect the model performance (Lemos et al., 2021).

A transformer is a state-of-the-art AI algorithm, which is a novel neural network architecture introduced recently based on an attention mechanism (Vaswani et al., 2017) and is revolutionizing the field of natural language processing. Transformers overcome the issue of gradient vanishing of RNNs and can take advantage of graphics processing units (GPUs) for faster computation. Hence, they have been reported in the latest studies for the real-time modeling of complex time-varying industrial processes (Geng et al., 2022; Wang et al., 2022).

Some soft sensor applications may require spatial information in the process data to be effectively captured, as spatial correlations may exist among process variables (Yuan et al., 2020c). Conventional feedforward neural networks and other traditional machine learning algorithms often fail to capture such information. The CNN has proven to be quite effective in achieving this task. It is a deep learning algorithm inspired by the network of neurons in the visual cortex of the human brain (Lecun et al., 1998). CNNs can successfully extract both spatial and temporal dependencies in the input data by applying filters based on the convolution operation. Moreover, the complexity of the CNN architecture is significantly reduced through the sharing of parameters and pooling operations. Hence, CNNs provide excellent soft sensing solutions for modeling industrial processes (Wang et al., 2019a; Yuan et al., 2020c).

The GAN, which was first introduced by Goodfellow et al. (2014), involves the training of two models simultaneously. These two models are termed the generator and the discriminator. The generator captures the distribution of training data and generates new data samples, while the discriminator determines how close the generated samples are to the real training data. The generator and the discriminator are trained simultaneously until the generator becomes capable of generating new data that are very close to the real data. GANs and their extensions are quite useful in soft sensor applications with a significant amount of missing data since the GANs can be used as a data imputation method (Yao and Zhao, 2021). GANs can also be used as a sample generation technique to increase the number of data samples, for industrial processes with small datasets caused by the slow sampling rates of quality variables (Zhu et al., 2021c).

As discussed in Section 3.1, industrial process data are typically high-dimensional and highly correlated in nature. Traditionally, dimensional reduction techniques such as PCA and PLS have been used to address the issue of collinearity (Kadlec et al., 2009). However, the conventional PCA and PLS algorithms can handle only the linear relationships between the variables, and hence they might fail in effectively extracting the nonlinear relationships. Several nonlinear extensions of the conventional PCA and PLS algorithms such as the kernel PCA (Cui et al., 2008), nonlinear PCA (Jia et al., 2000), and kernel PLS (Zhang et al., 2008) have been proposed in the literature to address this limitation. However, with the development of deep learning, researchers have investigated more advanced techniques to reduce the input space dimensions. Neural networks based on unsupervised learning algorithms have widely been used for this purpose.

An autoencoder (AE) is an unsupervised learning algorithm with a fully connected shallow neural network architecture. An AE consists of two components: an encoder and a decoder. The encoder is responsible for extracting important features from the unlabeled input data, by mapping the input variables to a lower dimensional latent space. The task of the decoder is to reconstruct the original data from the latent features extracted by the encoder, such that the reconstruction loss is minimized. Once the AE is successfully trained, the decoder part is discarded, and the encoder can be used to map the input data to a lower dimensional latent space. Hence, AEs can be used as an effective means of dimensionality reduction for industrial process data (Zhu et al., 2021a). A suitable regression algorithm can then be used to map the latent space to the output space to make accurate predictions. More advanced variants of the conventional AE such as the SAE (Yuan et al., 2020d), variational autoencoder (VAE) (Zhu et al., 2021a), and the recurrent Kalman VAE (Zhang et al., 2020) have also been widely employed in soft sensor applications to address the shortcomings of the traditional AE. VAEs are based on a probabilistic framework compared to the deterministic AE algorithm. They encode the inputs as distributions instead of points, resulting in better feature extraction. The ability of VAEs to reduce the original input space to a multivariate Gaussian distribution has been effective in reducing the misdetection of process faults, and hence they have widely been used in soft sensors constructed for process monitoring and process fault detection applications (Cheng et al., 2019; Lee et al., 2019; Zhang et al., 2019a).

A deep-belief network (DBN) is another deep learning algorithm based on unsupervised learning. It is a multi-layer neural network constructed by stacking multiple individual restricted Boltzmann machines (RBMs). Each RBM can extract nonlinear features from the data through an unsupervised learning approach (Liu et al., 2018). After this initial unsupervised learning phase, a supervised learning stage can be implemented to map the extracted features to the output data. A Kohonen map or a Self-Organizing Map (SOM) is another type of ANN that deals with unsupervised learning problems. An SOM maps the input space to a lower dimensional space, called the 'map' while maintaining the underlying structure of the input space. Hence SOMs can be used as an effective dimensional reduction technique for high-dimensional industrial process data (Ramachandran et al., 2019).

It should be noted that the classes of ANN algorithms are not limited to the ones discussed in this section. Only the main classes of ANN algorithms used in soft sensor applications were discussed, and different extensions, as well as combinations of these algorithms, can be used for constructing powerful soft sensing solutions.

3.3. Fuzzy Inference Systems (FIS) and neuro-Fuzzy Systems (NFS)

An FIS is a knowledge-based system that involves human-like reasoning. Fuzzy inferencing is based on a rule base that includes a set of IF-THEN type of rules (Deb, 2011). The ability of the FIS to represent complex processes has contributed to its wide use in soft sensor applications in advanced industrial processes. Although the concept of fuzzy logic-based soft sensors is not new, the most common classes of fuzzy logic-based models are discussed in brief as they are still used in soft sensor construction. Further details on the theory and applications of fuzzy sets and fuzzy logic can be found in the literature (Abeykoon, 2014a, 2016a; Klir and Yuan, 1995).

Mamdani and Takagi–Sugeno (T–S) are the most common classes of FISs. Pani and Mohanta (2016) compared these two models with conventional neural networks to be used for a soft sensor for predicting eight quality parameters of a cement clinker. The T–S model showed the best performance and both T–S and Mamdani models have exhibited better performance compared to backpropagation and RBF neural network models in terms of generalization ability as well as computational complexity. Angelov et al. (2008) reported a soft sensor with an evolving Takagi–Sugeno (eTS) model, which can detect data drifts and update its structure and parameters accordingly, to account for those changes.

An NFS is a hybrid intelligent model which combines the low-level learning ability of neural networks and the high-level, human-like reasoning ability of fuzzy logic systems (Wang, 1999). The evolving versions of NFSs are suitable for process industries as they can capture the dynamic behavior of the processes and the soft sensors based on such models may adapt to varying process conditions easily by changing their structure (Kadlec et al., 2009). Wu et al. (2009) developed a soft sensor based on a fuzzy neural network (FNN) and its parameters were optimized using a particle swarm optimization (PSO) algorithm. This soft sensor was trained using real data from a cement factory to model the raw material blending process. An adaptive neuro-fuzzy inference system (ANFIS) based soft sensor was proposed for estimating the top and bottom compositions in a benzene toluene distillation column of an oil refinery (Jalee and Aparna, 2016). The ANFIS model was used due to its ability to capture process nonlinearities and adapt to varying process conditions. In another study, an ANFIS model was adopted for estimating the fineness of cement in a cement grinding process (Pani and Mohanta, 2014). The ANFIS model showed superior performance compared to traditional SVR, Mamdani and, T–S models.

The FISs as well as the more advanced NFSs have shown excellent predictive performance while adapting to data drifts in many soft sensor applications. These models have also outperformed traditional algorithms in many cases, and hence, the fuzzy logic-based models are still used in soft sensor construction and process control applications in process industries.

3.4. Evolutionary algorithms

Evolutionary algorithms are a class of algorithms that mimic the genetic improvements of human beings or the natural behavior of animals (Deb, 2011). Although there are a variety of evolutionary algorithms, all of them are based on the concept that, when the individuals of a population compete for a limited amount of resources, only the fittest individuals in the population survive. This concept can be applied to an optimization problem, where an objective function needs to be maximized or minimized. Generally, in soft sensor design, evolutionary algorithms are used for training and hyperparameter tuning of soft sensor models (Jiang et al., 2012).

Earlier soft sensor applications employed popular traditional evolutionary algorithms such as genetic algorithm (GA), PSO, ant colony optimization (ACO), and differential evolution (DE) (Chen and Yu, 2005; Lahiri and Khalfe, 2009; Li and Liu, 2011; Shakil et al., 2009). However, later studies reported several novel evolutionary algorithms for soft sensor design. Wang and Chen (2017) used an immune evolutionary algorithm (IEA) for optimizing the parameters of a least squares SVM (LS-SVM) based soft sensor for predicting the burning zone temperature of a rotary kiln. The cuckoo optimization algorithm (COA) is another novel evolutionary algorithm inspired by the cuckoo's strategy of survival. Behnasr and Jazayeri-Rad (2015) employed the COA for optimizing the hyperparameters of a soft sensor based on an iteratively weighted least squares SVR model for predicting the butane content in a debutanizer column of an oil refinery. The authors claimed that the COA algorithm showed faster convergence and had the ability to achieve a better global minimum compared to other evolutionary algorithms. The Fruit-fly optimization algorithm (FOA) is a novel swarm intelligence algorithm developed based on the food-finding behavior of the fruit-fly swarm. Wang and Liu (2015) developed an adaptive extension of the FOA, called the adaptive mutation FOA (AM-FOA), and used it to optimize an LS-SVM based soft sensor model for predicting the MFI in a propylene polymerization process. The AM-FOA can escape the local minima during the optimization process, which is a possible problematic issue associated with the FOA. Beetle Antennae Search (BAS) algorithm is another evolutionary algorithm inspired by the searching behavior of longhorn beetles. Gao et al. (2020b) used the BAS algorithm to optimize an Elman neural network incorporated in a soft sensor for predicting the conversion rate of the vinyl chloride monomer.

The evolutionary algorithms discussed in this section are some of the most commonly used algorithms in soft sensor design. Other evolutionary algorithms as well as extensions of the discussed algorithms may also be used in constructing powerful soft sensing solutions.

It is obvious that the availability of a wide range of intelligent algorithms has provided soft sensor designers with the flexibility to choose the most suitable algorithm/s for a given application. Also, the reported works suggest that these intelligent AI-based algorithms have led to the development of soft sensors with superior performance which was not possible with traditional machine learning algorithms. The following section investigates in more detail, how the soft sensors based on these powerful AI-based algorithms can contribute toward the sustainable development of process industries.

4. The role of AI-driven soft sensors toward sustainable development

The Brundtland Report (World Commission on Environment and Development (WCED), 1987) introduced in 1987 defines sustainable development as the "development that meets the needs of the present without compromising the ability of future generations to meet their own needs". The concept of sustainable development derives from the triple bottom line concept (Elkington, 1994, 1999, 2018, 2019), which aims to achieve a balance among the three pillars of sustainability: namely environmental, social, and economic sustainability (Klarin, 2018). Process industries aim to achieve this goal through the implementation of various sustainable manufacturing strategies. Rashid et al. (2008) identified, compared, and contrasted four primary sustainable manufacturing strategies available in the literature, namely: waste minimization, material efficiency, resource efficiency, and eco-efficiency. Waste minimization refers to the reduction and prevention of waste generation, reduction of the hazardousness of the waste, and encouragement of reuse, recycling, and recovery. Material efficiency is closely related to the concept of dematerialization, which is the reduction of the quantity of material used to achieve a functional performance. Resource efficiency refers to the efficiency with which energy and materials are used throughout the economy. Eco-efficiency aims to prevent waste generation, improve resource efficiency, and improve the quality of life, while ensuring minimal impact on the environment, without exceeding the Earth's limits. Here, waste minimization and material efficiency are simpler strategies that are easier to implement and measure, but they have a very limited scope. On the other hand, resource efficiency and eco-efficiency are more complex strategies, where eco-efficiency is the broadest of the four strategies, hence more difficult to implement and measure. Different industries may implement different strategies based on their preference and policies. The implementation of soft sensors in the process industry may contribute to one or more of these sustainable manufacturing strategies, depending on which strategy is implemented in a given industry. The following example can be used to further elaborate on this. The spent catalyst of the fluid catalytic cracking unit (FCCU) constitutes a large proportion of the non-hazardous waste generated in an oil refinery (González, 2015). Soft sensors can be used to predict the catalyst saturation levels in real-time, and these predictions can then be used to optimize the use of the catalyst within the FCCU. This results in a higher product yield with less catalyst consumption, which in turn improves the efficiency of the material usage while minimizing waste generation. Hence, in this case, the soft sensor contributes to waste minimization, material efficiency, or eco-efficiency strategy, depending on which of these sustainable manufacturing strategies is implemented by the industry. This can result in an overlapping of the different strategies, but ultimately, they all lead toward the common goal of sustainable development.

This section looks at how the use of AI-based techniques in soft sensor development contributes toward these different sustainable manufacturing strategies to achieve the ultimate goal of sustainable development. A key issue in industrial polymer extrusion processes is used as a case study to explain how state-of-the-art AI-based modeling techniques can be used to overcome the limitations in traditional data-driven soft sensors, to contribute toward sustainable manufacturing strategies.

In polymer extrusion, melt temperature is considered a key indicator of melt quality. Past studies have revealed that the melt temperature is not homogeneous across the melt flow and that a radial melt temperature profile exists at the extruder discharge (i.e., at the die entry). The thermal homogeneity (indicated by the degree of flatness of the radial melt temperature profile) is significantly affected by the process settings such as screw speed and barrel set temperatures (Kelly et al., 2005, 2006, 2008). It has been shown that operating the extruder at high screw speeds leads to poor thermal homogeneity (i.e., temperature variations across the melt flow increases), while low screw speeds result in better thermal homogeneity (i.e., a flatter radial melt temperature profile). However, the specific energy consumption of single and twin-screw extruders was found to decrease with increasing screw speed (Abeykoon et al., 2020). This shows that there is an opposing behavior between the melt thermal stability and the specific energy consumption of the extruder, with the screw speed. Operating the extruder at the optimum operating point would enable the right balance between the melt thermal stability and the extruder-specific energy consumption to be maintained. However, this requires precise control of the extrusion process in real-time. To achieve this, it should be possible to monitor the radial melt temperature profile in real-time. However, the existing hardware sensors employed in the polymer processing industry suffer from limitations such as the inability to detect temperature variations across the melt flow, limited durability, and disturbances to the melt flow (Abeykoon et al., 2012; Bur et al., 2001; Kelly et al., 2008). Consequently, industrial polymer extrusion processes are carried out at less-than-optimal operating conditions. This results in inefficient utilization of energy, which is not sustainable. Significant energy savings could be achieved if the extrusion process is optimized. In the absence of suitable hardware sensing techniques, inferential approaches would be a better alternative for monitoring the melt temperature profile in real-time.

As a step toward this, Abeykoon et al. (2011) proposed a static nonlinear polynomial model optimized using a fast recursive algorithm (FRA), for monitoring the melt temperature profile in a single screw extrusion process. This model was then used to optimize the process settings (i.e., screw speed and barrel set temperatures) of the extruder. It was reported that melt temperature variations were reduced by 3.1%–60.9% while achieving the desired average melt temperature over 10 different process conditions. Later, Abeykoon (2014b) developed a dynamic soft sensor to predict the radial melt temperature profile based on a nonlinear polynomial model optimized with the FRA. This soft sensor reported good accuracy with normalized prediction error (NPE) values in the range of 1.33–2.89 for each radial position of the temperature profile. It was subsequently incorporated into a process control strategy to optimize the extruder process settings such that melt temperature variations are minimized (Abeykoon, 2014a). To the best of the knowledge of the authors, no other studies have been reported in the literature for the real-time prediction of the melt temperature profile using soft sensors. Although these models can make real-time melt temperature profile predictions, their prediction accuracy is not sufficient for an industrial setting. Furthermore, these soft sensors have only been tested on a single material processed with a single screw type. In industrial polymer extrusion processes, a wide range of polymeric materials are processed using different screw geometries. Hence, it is crucial that the soft sensor can make accurate predictions under these varying process conditions.

To optimize extrusion processes, robust soft sensors with a high degree of precision are required, and it is clear that the existing soft sensors fail to meet these requirements. These soft sensors were based on traditional data-driven modeling techniques, and no attempt has been made to address the existing limitations using state-of-the-art AI algorithms. Extrusion processes are highly dynamic in nature, and hence, advanced deep learning algorithms such as RNNs and their extensions, transformers, and CNNs (discussed in Section 3.2) could be used to effectively extract temporal and spatial features in the extrusion data. The soft sensors developed by Abeykoon et al. (2011) and Abeykoon (2014b) used five input process parameters (i.e., screw speed and four barrel set temperatures) to predict the melt temperature profile, and this constitutes a high dimensional input space. Unsupervised learning techniques such as the AE and its extensions discussed in Section 3 could be useful in reducing the dimensions and collinearity to enhance the predictive performance of the soft sensor. Moreover, adaptation mechanisms based on the latest AI-driven techniques should be useful in making the soft sensor adaptive to different polymeric materials and screw types. The existing AI algorithms can further be modified (as will be evident by some of the applications discussed in Section 5) to improve the computational efficiency of the soft sensor as well. The enhanced prediction accuracy and computational efficiency will ensure accurate monitoring of the radial melt temperature profile in real-time. This would enable real-time quality control strategies to be implemented to optimize the extrusion process parameters to improve melt thermal stability while reducing the specific energy consumption of the extruder. The reduction in the specific energy consumption of the extruder would contribute to the resource efficiency strategy or the eco-efficiency strategy, depending on which sustainable manufacturing strategy is implemented in the industry.

Similar to the case study discussed, soft sensor applications in other process industries can also contribute toward sustainable development through different sustainable manufacturing strategies. A few examples are discussed here. Soft sensors can be used for predicting melt viscosity, which is another key quality indicator in extrusion processes (Deng et al., 2014; Liu et al., 2012; McAfee and Thompson, 2007). These soft sensors can aid in reducing the production of low-quality products by optimizing the extrusion process, preventing the generation of waste products. The use of soft sensors for estimating the H_2S concentration in real-time has enabled oil refineries to continuously monitor and maintain the emissions within the desired limits, and this has a direct impact on environmental sustainability. In the pulp paper industry, soft sensors can be designed to monitor the AOX content in the bleaching wastewater, which contains carcinogenic compounds (Ma et al., 2020). The use of soft sensors for predicting the chemical oxygen demand (COD) in the pulp paper industry enables achieving a good degree of washing of paper with less amount of chemicals leading to economic gains and material efficiency (Soares et al., 2011). Soft sensors used in thermal power plants and wastewater treatment plants play a significant role in reducing harmful emissions and improving effluent quality (Fernandez de Canete et al., 2021; Sun et al., 2019).

Fig. 3 illustrates the contribution of different soft sensor applications toward sustainable manufacturing strategies. However, the limitations in traditional data-driven modeling techniques hinder the potential of soft sensors in successfully implementing these strategies. Here, the key is to improve the prediction accuracy and the computational efficiency of these soft sensors by addressing the limitations of traditional modeling techniques. The latest AI-based algorithms have enabled soft sensors to gain excellent prediction accuracy and computational efficiency compared to traditional statistical and machine learning-based

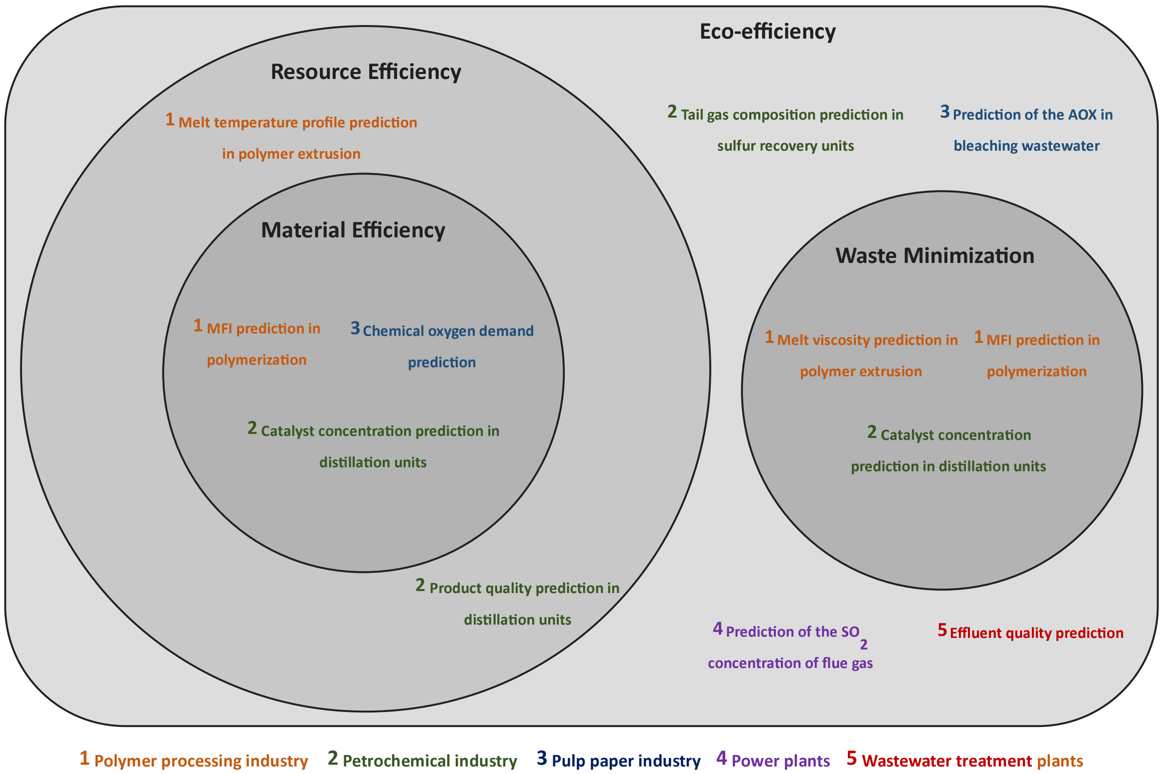

Fig. 3. The contribution of different soft sensor applications toward sustainable manufacturing strategies.

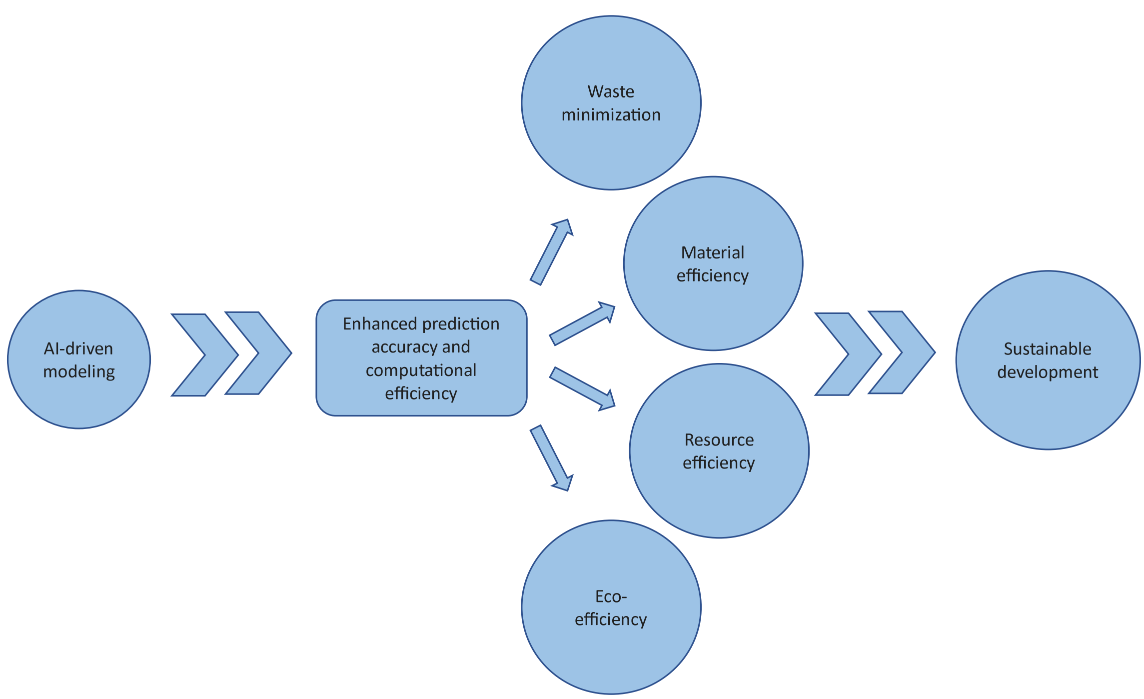

Fig. 4. The role of AI-driven modeling toward sustainable development of process industries.

models. This enables more effective process monitoring and control to be carried out resulting in less material and energy wastage, reduced emissions, and so forth. This paves the path toward achieving the ultimate goal of sustainable development (see Fig. 4). The following section provides a critical review of the effectiveness of state-of-the-art AI-based modeling techniques in overcoming the limitations associated with traditional data-driven soft sensors to improve prediction accuracy and computational efficiency of soft sensors.

5. AI-driven soft sensor applications in process industries

This section provides a critical review of the state-of-the-art AI-based solutions reported in the literature, for addressing some of the common challenges faced by soft sensor designers when constructing soft sensors for the process industry. As explained in Section 1.1, different classes of problems (i.e., missing data, varying sampling rates, small datasets, dimensionality reduction, adapting to varying process conditions, extracting temporal and spatial features, and model interpretability) are identified and the latest AI-based solutions used to tackle these problems are discussed. Although the focus is only limited to soft sensing solutions from three process industries (i.e., polymer processing, petroleum refining, and pharmaceutical), the problems and solutions discussed here are applicable to other process industries as well.

5.1. Missing data, varying sampling rates, and small datasets

Missing data is a common challenge faced by soft sensor designers. This is mainly caused by frequent hardware sensor failures. For example, in the polymer processing industry, harsh conditions in polymerization reactors can cause hardware sensors to malfunction frequently. Traditionally, this issue was addressed through simple solutions such as removing data points that contained missing values or replacing the missing values with the mean values of the affected variables. However, these techniques were not regarded as optimal solutions as they could affect the model performance. (Kadlec et al., 2009). For example, removing data points with missing values could reduce the number of usable data points for training a soft sensor model, and risk the loss of valuable dynamic information. As a solution to this problem, Xie et al. (2020b) presented a novel approach based on deep variational AEs (DVAEs). Two sub-models based on DVAEs were constructed. The first sub-model which is a supervised DVAE (SDVAE) learns the distribution information of the latent features, and this is followed by the second sub-model which is a modified unsupervised DVAE (MUDVAE). The soft sensor was constructed by combining the encoder of the SDVAE with the decoder of the MUDVAE. The performance of the soft sensor was assessed on a simulated dataset with three levels of missing data: light, medium, and heavy. The proposed model surpassed three conventional missing data imputation methods (i.e., deletion, mean imputation, and PCA imputation) for all three levels of missing data. It is the superior data reconstruction ability of this novel extension of the AE, which benefits from both supervised and unsupervised learning methods, that was able to provide a viable solution to the missing data problem compared to the traditional methods.

Varying sampling rates are another common issue in soft sensor construction. In industrial polymerization processes, the MFI of the polymer product is measured offline every several hours, while process variables such as temperature and pressure can be measured online every few seconds. In bioprocesses also, the quality variables have a lower sampling rate than the input variables due to the slow speed of the process. Consequently, out of all the collected data, the amount of labeled data (i.e., contains both input and output data) represents only a small proportion, while most of the data is unlabeled (i.e., contains only the input data). Traditional soft sensors based on supervised learning algorithms utilize only the labeled data, and they fail to extract the information hidden in the unlabeled data. As a result, useful information hidden in the unlabeled data is discarded without being utilized, and this could adversely affect the predictive performance of the soft sensor. Additionally, this results in a smaller dataset for training the model.

Most of the AI-based solutions proposed to address this issue are based on semi-supervised learning techniques designed to extract features from unlabeled data in addition to labeled data. Liu et al. (2018) constructed a semi-supervised soft sensor based on an ensemble deep kernel learning model. The modeling framework incorporates a DBN to extract useful information from unlabeled data through an unsupervised learning stage. This stage is then followed by a supervised learning stage, which utilizes a regression model based on a kernel learning strategy to map the extracted features to the output variable. The proposed soft sensor achieved a root mean square error (RMSE) of 2.74, outperforming soft sensors based on SVR, PLS, and single DBN models. Yuan et al. (2020b) proposed a pre-training strategy based on a semi-supervised stacked autoencoder (SS-SAE). Soft sensors developed according to the proposed methodology were employed to predict the final boiling point of aviation kerosene in a hydrocracking process and the butane concentration at the bottom of a debutanizer column. The performance of these soft sensors was compared against soft sensors based on ANN, DBN, and basic SAE models. The ANN model does not incorporate a pre-training strategy, hence showed the worst performance compared to the other models. The DBN and basic SAE models incorporate only an unsupervised pre-training strategy, while the SS-SAE model incorporates a semi-supervised pre-training strategy and consequently, the SS-SAE model outperformed the other models. Gopakumar et al. (2018) used an SOM to extract features from unlabeled data, during an unsupervised pre-training phase. This was followed by a supervised learning phase using the backpropagation algorithm. The soft sensor was tested on streptokinase and penicillin fermentation processes. In terms of the RMSE, the soft sensor clearly outperformed traditional SVR and deep neural network models trained with supervised learning techniques only. It is clear that the semi-supervised learning strategies based on DBN, SAE, and SOM algorithms employed in these soft sensors were capable of extracting useful information from the unlabeled data in addition to the labeled data, resulting in superior performance compared to traditional supervised learning techniques.

To address the problem of small datasets, techniques such as bootstrap aggregation and noise injection were used traditionally (Di Bella et al., 2007; Fortuna et al., 2009; Noor and Ahmad, 2011). Moreover, techniques such as the SVR have also been quite useful in training small datasets (Kadlec et al., 2009). Although these traditional techniques were able to improve the predictive performance of soft sensors trained with small datasets, researchers have investigated better approaches to further improve their accuracy. In an attempt to address this issue, Zhu et al. (2021c) introduced a novel virtual sample generation (VSG) technique to increase the size of small datasets. A conditional GAN (CGAN), which is an improved conditional version of the traditional GAN was proposed under this technique. The performance of the proposed soft sensor was evaluated on a dataset from a high-density polyethylene (HDPE) polymerization process and compared with three other conventional VSG methods: bootstrap, mega-trend diffusion, and tree-based trend diffusion. The CGAN soft sensor showed the lowest RMSE value (0.4995) among all the VSG methods, indicating its superiority over the traditional methods. The CGAN searches for the sparse areas of the provided small dataset and generates new virtual data points in the sparse areas. This expands the dataset leading to better-quality data. In another study, Chou et al. (2020) proposed a sequence-to-sequence neural network model based on GRUs to handle the problem of small datasets. The encoder of the model predicts the future dynamics of the process variables. The trained encoder was then used to train the decoder using the limited amount of data related to the distillate impurity and the bottom product impurity of a distillation column of an oil refinery. The proposed soft sensor outperformed a conventional ANN in terms of prediction accuracy.

5.2. Dimensionality reduction

Data collinearity is another challenge in designing soft sensors for the process industry (Kadlec et al., 2009). Data collinearity results in redundant variables and this could unnecessarily increase the complexity of the soft sensor model. This in turn could negatively affect the performance of the soft sensor. Hence, dimensionality reduction techniques are usually incorporated to eliminate redundant variables. Traditionally, this has been achieved using algorithms such as PCA and PLS, which transform the original high-dimensional input space, into a low-dimensional latent space with less collinearity (Warne et al., 2004). However, the latest studies have introduced AI-based solutions to reduce the input space dimensions while extracting important features into a latent space. Soft sensors developed for the debutanizer column in oil refineries have largely benefited from these dimensionality reduction techniques. Generally, seven input process variables are used to predict the butane content in debutanizer columns (Siddharth et al., 2019), and this results in a high-dimensional input space. The latest AI-based solutions have mostly been based on AEs and in most cases, the conventional AE algorithm has been modified using different strategies to improve its ability to extract important features. Some of these strategies are discussed here with their impact on the predictive performance of the soft sensors.

The standard AE uses an unsupervised learning technique that tries to minimize the reconstruction input error. Hence, it can learn features that are a good representation of the raw input data, but it does not guarantee that the learned features are quality relevant. To address this problem, a stacked quality-driven autoencoder (SQAE) network was proposed by Yuan et al. (2020e). During the pretraining stage of the SQAE, in each layer high-level features are learned from the adjacent low-level features, subject to the additional constraint that the learned features can predict the quality variables accurately. The learned features can then be used in a regression model to predict the quality variable. A hybrid variable-wise weighted SAE (HVW-SAE) is another technique reported in the literature for extracting features from the input space that are more relevant to the quality parameter to be predicted (Yuan et al., 2020a). Pearson and Spearman correlation coefficients were used to calculate the linear and nonlinear correlations respectively, between each input variable and the quality variable. Then the two coefficients were combined to create a hybrid coefficient, which was subsequently used to construct a weighted reconstruction objective function for each AE. This approach enabled the SAE to extract both linear and nonlinear features that are more relevant to the quality parameter to be predicted while minimizing the information loss at each AE. Yuan et al. (2020d) proposed a stacked isomorphic autoencoder (SIAE) based soft sensor for predicting the tail gas composition in a sulfur recovery unit (SRU). In the traditional SAE, the reconstruction error accumulates from the lower layers to the higher layers. Unlike the SAE, at each layer, the SIAE extracts features from the previous layer and attempts to reconstruct the original raw input data as accurately as possible. This ensures that the information loss is not accumulated from one layer to the next, leading to a better latent representation of the original raw data.

Another limitation of the standard AE is that it ignores the intrinsic data structure information in the process data. Recent studies have introduced more advanced variants of SAEs to address this limitations with the aim of further enhancing the feature extraction and dimensional reduction capability. Motivated by the concept of neighborhood preserving embedding in manifold learning, Liu et al. (2021d) proposed a stacked neighborhood preserving autoencoder (S-NPAE). This study utilized neighborhood preserving embedding regularization to enable the AE to learn neighborhood-preserving features in the process data. Unlike the conventional SAE which neglects the neighboring data structure during feature extraction, the proposed S-NPAE improves the generalization performance by preserving the local neighborhood structure of the process data while minimizing the reconstruction error. In another study, Liu et al. (2021b) introduced a spatiotemporal neighborhood preserving SAE (STNP-SAE), which not only captures the spatial neighborhood structure information, but also the temporal neighborhood structures. This is achieved by constructing spatial and temporal adjacent graphs. This novel algorithm was incorporated into a soft sensor for predicting the initial and final distillation temperatures of heavy naphtha in a hydrocracking process, and the soft sensor reported respective RMSE values of 0.0698 and 0.1179, outperforming the conventional SAE.

Gao et al. (2021) proposed a teacher student stacked sparse recurrent AE (TS-SSRAE) model for constructing a soft sensor for predicting the penicillin concentration in a penicillin fermentation process. An LSTM network was used for constructing the AE in order to extract the autocorrelation and cross-correlation characteristics of the process variables. A stacked version of this AE was proposed in order to extract the features more effectively. Sparsity was introduced to the model with the aim of discarding redundant information and retaining the more important ones. This SSRAE model which was termed the 'teacher' model was trained first, and then the hidden layer information of this model was transferred to a simpler two-layer network which was termed the 'student' model. A knowledge distillation compression framework was used to achieve the transfer. Finally, an output layer was added after the hidden layer of the trained student model and its parameters were fine-tuned. This soft sensor showed excellent predictive performance with an RMSE of 0.032 surpassing SAE, SSAE, and a teacher student stacked recurrent autoencoder (TS-SRAE).

The modifications to the traditional AE algorithm discussed here such as the SQAE, HVW-SAE, and SIAE ensure that only the features which are relevant to the quality parameter to be predicted are extracted while minimizing the reconstruction error. Moreover, S-NPAE, STNP-SAE, and TS-SSRAE models ensure that both spatial and temporal characteristics of process data are also considered when constructing the latent space. Hence, these models have proven to be better in performance compared to the standard AE.

In another study, a different technique based on a DBN has been attempted to reduce the input space dimensions (Graziani and Xibilia, 2019). First, a set of input regressors were chosen through a cross-correlation analysis, which was then used as the input features for the unsupervised phase of DBN training. Next, a set of latent features were extracted from the DBNs. Furthermore, the authors proposed a method for estimating the measurement delay, which involved performing a cross-correlation between the extracted latent features and the plant output. Finally, fine-tuning of the DBNs was carried out after selecting the target plant output values based on the estimated measurement delay.

In the work by Hikosaka et al. (2020), the authors proposed a novel feature and dynamics selection method for designing soft sensors. The GA-based process variables and dynamics selection (GAVDS) method introduced in the study attempts to identify both the important process variables as well as their time delays simultaneously. The GAVDS is an extension of the GA-based variable selection combined with PLS. The authors extended the GAVDS to an ensemble GAVDS (EGAVDS), which attempts to minimize the uncertainty caused by using a single dataset in GAVDS, by splitting the training data into several subsets and then determining the variables and time delays from these subsets. The proposed method was used in predicting the tail gas composition in an SRU as well as the butane content in the bottom flow of a debutanizer column. Here, the use of an evolutionary algorithm was the key to selecting the most relevant process variables and their time delays, and this has resulted in a soft sensor with enhanced performance, surpassing a conventional PLS model.

5.3. Adapting to varying process conditions

After the implementation of a soft sensor in the process industry, its performance can deteriorate during long-term use, due to drifts in the process data. These drifts can occur due to internal or external changes in the process operating conditions. Performance deterioration due to changes in external environmental conditions such as weather or seasonal variations is a critical issue in certain industries such as oil refineries (Zhang et al., 2019b) and these variations cannot be predicted at the soft sensor design stage. Historical data used to train soft sensors do not represent all possible future states and consequently, the soft sensor performance deteriorates as data drifts occur. Once the performance reaches an unacceptable level, the soft sensor needs to be re-trained or re-developed from scratch. A solution to this is provided by equipping the soft sensor with adaptive capabilities. Such sensors are known as adaptive soft sensors.

Kadlec et al. (2011) reviewed the commonly used traditional adaptive mechanisms for soft sensors. The authors categorized these adaptation mechanisms into three categories: moving window-based, recursive, and ensemble-based methods. Traditional adaptive soft sensors were mostly based on PCA and PLS algorithms as they can easily be combined with window-based and recursive adaptation techniques. Consequently, the Recursive PCA, Moving Window PCA, Recursive PLS, and Moving Window PLS algorithms have been quite popular among soft sensor designers (He and Yang, 2008; Liu et al., 2010; Qin et al., 1999; Wang et al., 2003). Recursive techniques update related models by adding new samples without discarding the old ones resulting in an increasing dataset. This offers efficient computation but at the expense of the speed of adaptation. On the other hand, moving window-based techniques create a new process model by adding the newest sample while discarding the oldest one. It provides a constant speed of adaptation. With small window sizes, increased speeds of computation can be achieved, but the temporal variations in process data may not be effectively captured. To overcome these limitations, researchers have investigated more advanced algorithms based on AI techniques.

In recent studies, local models based on a Just-In-Time Learning (JITL) strategy, have widely been used over global models for developing adaptive soft sensors, due to the inability of the global models to adapt to varying process conditions. Liu et al. (2020) proposed a JITL framework with local models for developing a soft sensor. In the proposed framework, a local dataset is selected from the feature space, by evaluating the similarity of an incoming query sample to that of each historical sample in the database. The Euclidean distance was used to measure the similarity. Then, a nonlinear local model training algorithm called nonlinear Bayesian weighted regression (NBWR) was used to construct a local model to map the selected local dataset to the output space. One of the benefits of using the NBWR approach is that it develops probabilistic local models that can handle uncertainties, unlike deterministic models. The proposed soft sensor was employed to predict the butane content at the bottom of the debutanizer column of an oil refinery. As the JITL strategy ensures that the closest samples to the query sample are selected from the historical dataset, the sensor can effectively adapt to varying process conditions. It reported an RMSE of 0.0405 on an unseen dataset, showing superior performance than soft sensors based on global models including PLS, least squares SVR, Deep ELM, and a few other machine learning algorithms. In another study, Zheng et al. (2021) introduced a Mahalanobis distance-based JITL soft sensing solution, to solve the multiphase issue of the penicillin fermentation process. Three operating phases can be identified in the penicillin fermentation process, which can cause global soft sensor models to perform poorly. A multiway Mahalanobis distance-based metric learning was introduced as the similarity measurement method in this JITL-based local modeling approach. The proposed solution allows data samples from different operating phases to be identified without an additional phase identification step. Based on the similarity measurements, the most relevant data from the historical dataset is selected for creating a local LSTM model to predict the penicillin concentration.

The use of an appropriate similarity measurement technique is another important aspect of JITL model construction. Distance-based metrics such as the Euclidean distance and Mahalanobis distance have widely been used in the reported literature (Liu et al., 2020; Shen et al., 2020; Zheng et al., 2021). Guo et al. (2020b) claimed that traditional deterministic similarity measurement techniques such as distance-based, angle-based, and correlation-based techniques, do not consider the uncertainties in the query samples of JITL models. Hence, Kullback–Leibler (KL) divergence and symmetric KL divergence methods were proposed by Guo et al. (2020a,b), to measure the similarity between the historical samples and the query sample.

Transfer learning is another effective adaptation technique that can be used to share useful information extracted from data belonging to one operating mode with other operating modes that have limited data. Due to the advancements in deep learning, transfer learning-based techniques are increasingly being used in developing adaptive soft sensors for the process industry (Curreri et al., 2021). In the polymer processing industry, grade changeovers take place frequently which shifts the process from one operating mode to another. Soft sensors are generally trained on data gathered under a single operating mode and they fail to perform with the same level of accuracy when the operating mode is shifted. As a solution to this problem, Liu et al. (2019) proposed a domain adaptive transfer learning soft sensor. The authors employed a domain adaptation ELM (DAELM) as a transfer learning method and the soft sensor was evaluated on three different grades. The performance of the DAELM approach was found to be superior to that of a regularized ELM (RELM) model, in terms of the relative prediction error. In another study, Zhu et al. (2021d) developed a transfer learning-based adaptive soft sensor to predict the Pichia pastoris cell concentration in a fermentation process with multiple operating conditions. The authors compared three different transfer learning methods: transfer component analysis, joint distributed adaptation (JDA), and balanced distributed adaptation (BDA). The difference in the three methods is related to how the marginal and conditional probability distribution between the source and target domains are adapted. As JDA and BDA are applied to classification problems, the authors introduced an improved JDA, and an improved BDA (IBDA) integrated with fuzzy sets to make them suitable for the regression problem under consideration. The three transfer learning methods were combined with a RELM. Out of the three transfer learning methods, the IBDA showed the best predictive performance.

In addition to the widely used JITL and transfer learning-based adaptation techniques, a few other novel approaches reported in the literature are discussed here. Liu et al. (2021c) introduced a novel adaptive soft sensor based on a stacked multi-manifold autoencoder (S-MMAE) to estimate the final distillation temperature of heavy naphtha and aviation kerosene in a hydrocracking process of an oil refinery. This novel soft sensor was capable of extracting features within each operating condition (i.e., manifold) as well as the interconnections among different operating conditions. This was achieved by combining a within-manifold adjacent graph and a between-manifold adjacent graph to construct a multi-manifold regularization method, which was in turn used to train an SAE at the pre-training stage. The effectiveness of the proposed soft sensor was evaluated by comparing its performance against ANN, SAE, and Laplacian regularization AE (LAE) models. The soft sensor outperformed all the other models with RMSE values of 0.0721 and 0.0626 for predicting the distillation temperature of heavy naphtha and aviation kerosene respectively.

Sun et al. (2020) developed an adaptive soft sensor based on an output recursive wavelet neural network (ORWNN) and Gaussian process regression (GPR) to predict the total sugar content in a chlortetracycline fermentation process. This adaptive soft sensor was based on an online learning technique, where a cumulative update training method was used to update the historical dataset with new input and output data, each time a prediction is made. The GPR model could make accurate predictions during the early stages of online learning where only a small amount of data was available for training, while the ORWNN improved the accuracy as the training data accumulated. Hence, the ORWNN-GPR model exhibited good predictive performance at both the initial and latter stages of online learning, compared to the individual models.

5.4. Extracting temporal and spatial features

Industrial processes are highly time-dependent in nature. Traditional static models such as SVR and PLS cannot extract useful dynamic information from the process data. Hence, as discussed in Section 3.2, dynamic neural network models such as RNNs and their variants have widely been used in the most recent AI-based soft sensing solutions, to extract temporal features in the process data effectively. Similarly, CNNs have been employed to extract both temporal and spatial features. Recent studies have attempted to modify the structure of some of these dynamic neural networks to further enhance their performance by addressing their shortcomings or limitations. For example, a GRU was modified by developing a two-stream λ GRU algorithm to overcome the issue of the linear coupling constraint that exists in the basic GRU algorithm (Xie et al., 2020a). The modified algorithm outperformed soft sensors based on SVR, PLS, SAE, and LSTM algorithms. He et al. (2020b) used singular value decomposition (SVD) to calculate the weights between the reserve and output layers to tackle the collinearity issue in the outputs of the reserve layer of the basic ESN algorithm. In another study, an Orthogonal ESN (OESN) was developed to eliminate the issue of collinearity, and the parameters of this model were optimized using an improved DE algorithm (Zhang et al., 2021).

Zhu et al. (2021b) used a bidirectional LSTM (BiLSTM) instead of the traditional unidirectional LSTM, to better exploit the dynamic information in the process data by processing the data in both forward and backward directions. Furthermore, the authors modified the LSTM structure by replacing the forget gates and the output gates with converted input gates, and the resulting model was named a converted gates LSTM (CG-LSTM). This modification was introduced to reduce the computational complexity of the traditional LSTM model. An eXtreme Gradient Boosting (Xgboost) algorithm was implemented for input variable selection and the selected input variables were fed to the Bidirectional CG-LSTM (BiCG-LSTM) model. A Self-attention (SEA) strategy was also introduced to reduce the overfitting of the model. This novel soft sensor showed superior performance compared to state-of-the-art models based on two-stream λ GRU and SAE as well as traditional PLS and SVR models.

Zhang et al. (2019b) introduced a deep Weighted Auto Regressive LSTM (WAR-LSTM) model that can extract high-level representations from multivariables in the spatial domain. The soft sensor reported a mean square error (MSE) of 0.80 ± 0.08, outperforming a conventional deep LSTM model. Increasing the number of feedback outputs of the WAR-LSTM structure resulted in improved performance but at the expense of the model simplicity. The use of a scaling factor normalization instead of the conventional min–max normalization for data pre-processing and a recurrent denoising AE for feature extraction, which can retain temporal dependence of the extracted features unlike the basic denoising AE, further contributed to the superior performance of the deep WAR-LSTM based soft sensor.

Yi et al. (2020) developed a soft sensor based on an ensemble deep-learning strategy for the online estimation of the fraction yields of crude oil. A CNN and a nearest neighbor regression (NNR) model were used as component learners, while a random vector functional link (RVFL) network was used as a meta-learner to combine the two component learners. Nuclear magnetic resonance spectrum data, which were transformed into 2D form to feed the CNN, were used as the input data to estimate the fraction yield. According to the results, the model reported an RMSE value of 0.2524, outperforming PLS, MLP neural network, DBN, and NNR models. A study by Wu et al. (2021) introduced a Dilated CNN (DCNN) to predict the melt index of a PP production process. The dilated convolution increases the receptive field of the soft sensor without losing the feature resolution, which improves the predictive performance of the model. The DCNN model exhibited an RMSE of 0.0289, outperforming a traditional CNN model and an ELM model.

Hu et al. (2021) proposed a novel spatio-temporal attention network (STAN), which incorporates a temporal attention module and a spatial attention module to extract temporal and spatial features in the data, respectively. The extracted features are then merged using a spatio-temporal fusion module. Moreover, a highway network is implemented to make the model give more weight to the current state of the input variables. The merged spatio-temporal features and the highway features are then fed to the output layer. The STAN model showed an RMSE of 0.0661, surpassing ELM, backpropagation neural network, LSTM, and convolutional LSTM (CNN + LSTM) models.

5.5. Model interpretation

As discussed in Section 2, one of the main limitations of data-driven soft sensors is the inability to interpret the model due to the black-box nature of the model. Although first-principles models are easily interpretable, they suffer from poor predictive performance due to the numerous assumptions made when deriving them. Gray-box modeling techniques have been introduced to obtain interpretable models with good predictive performance. Ahmad et al. (2020) provide a comprehensive review of the design aspects and applications of recent gray-box soft sensors. Here, the gray-box soft sensors were classified into three categories: serial, parallel, and combined gray-box models. These categories were defined based on the role of the data-driven component in the gray-box model. In serial gray-box models, the data-driven component is used to estimate unknown parameters of the first-principles model, while in the parallel gray-box models, the data-driven component is used to compensate for the error of the first-principles model. The combined gray-box soft sensors combine both serial and parallel configurations into their model structure. Since the focus of this study is only on data-driven soft sensors, gray-box models are not extensively discussed here. However, a recent study that attempted to improve the model interpretability of a data-driven model is reviewed due to its state-of-the-art nature.

To integrate process knowledge into model construction and to enable the development of interpretable models for predicting the butane content in debutanizer columns, Chen and Ge (2021) introduced a novel soft sensor framework called Graph mining, convolution, and explanation stacked target-related autoencoder (GMCE-STAE) In this approach a spatial self-attention mechanism was used to extract process knowledge based on available historical data. Then, the extracted knowledge along with the human experience was used to extract features using a graph convolution layer. These extracted features were then input into the soft sensor model to predict the butane content. An SAE was used as the soft sensor model, which was trained by two phases: layer-wise pre-training and fine-tuning. Finally, a graph neural network explainer was used to explain how the process knowledge was utilized by the model for predicting the butane content. Moreover, it was shown that the knowledge discovered by the spatial self-attention mechanism was consistent with prior knowledge. However, the prior knowledge consideration has slightly increased the computational complexity of the model as well. This state-of-the-art approach in soft sensor construction has opened a new avenue for soft sensor designers to develop data-driven soft sensors by incorporating process knowledge as well as to interpret the model predictions.

The AI-based techniques for solving the problems discussed in this section have proven to be highly effective, compared to traditional methods. The enhanced performance of the soft sensors can be observed through reduced prediction errors and computational times. This has mainly been achieved by combining two or more AI-based algorithms together or modifying the internal architecture of the basic AI-based algorithms. The enhanced predictive performance is the key to accurately monitoring industrial processes in real-time which enables maintaining the process within desired limits. This ultimately helps in reducing waste generation, optimizing material and energy utilization, reducing harmful emissions, and so forth.

6. Discussion

Over the years, soft sensors have been used in a wide range of process industries to replace or to work in parallel with hardware sensors. This paper discussed the state-of-the-art AI-based algorithms employed in modern soft sensors, their contribution toward sustainable development, and their latest applications in the process industry. The key limitations presented in Section 1.1, which were identified from previously published review articles on soft sensors were addressed, effectively bridging the existing research gaps. Unlike the previous

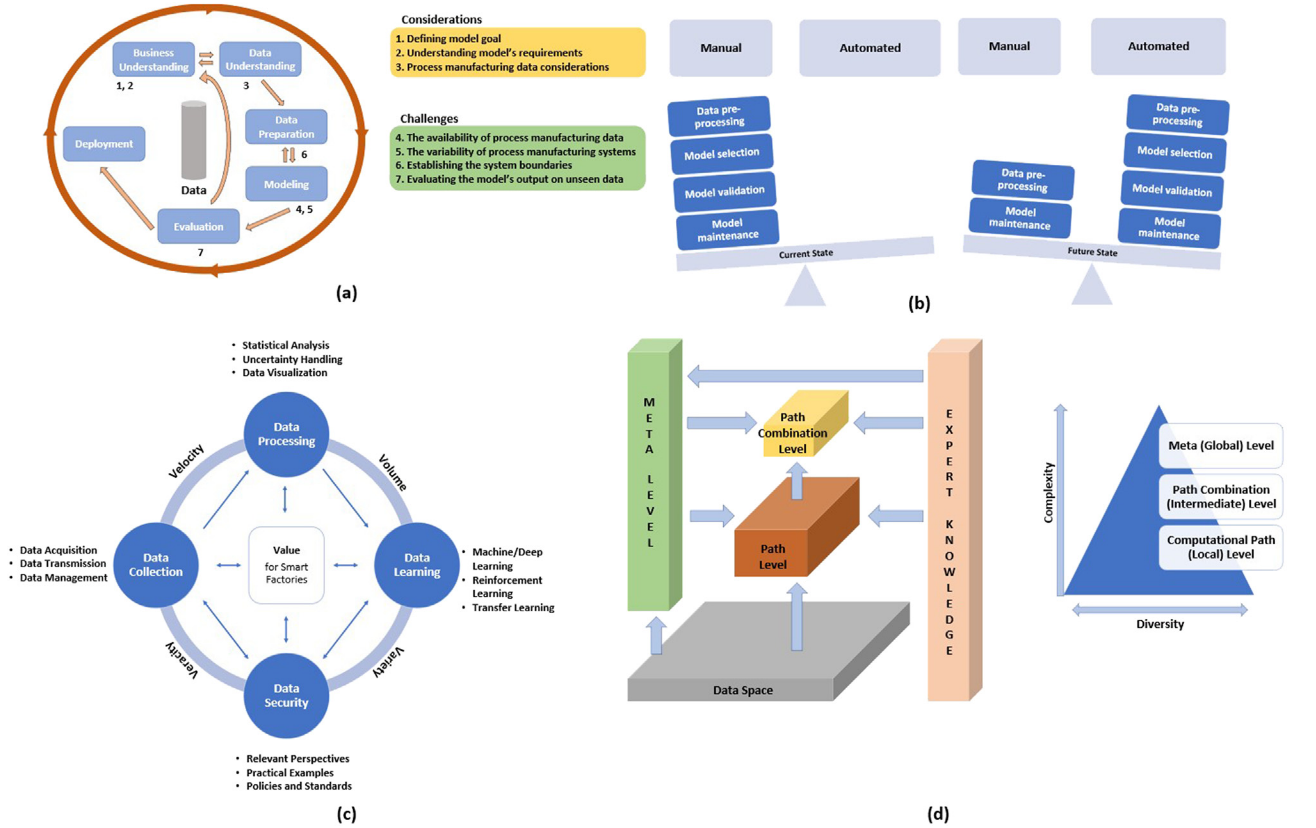

Fig. 5. Driving factors for soft sensors of the future (a) CRISP-DM framework for developing data-driven models (Fisher et al., 2020) (b) Current and future states of soft sensors (Kadlec and Gabrys, 2009) (c) Relationships among key elements in big data analytics and smart factories (Gao et al., 2020a) (d) Model development architecture for future soft sensors proposed by Kadlec and Gabrys (2009).

reviews, this study entirely focused on AI-based soft sensing algorithms, instead of traditional modeling techniques. Instead of reviewing studies related to a single industrial process, this paper investigated soft sensing solutions provided across multiple industries, which enabled the identification of a wide range of solutions provided to address a given problem/challenge. Moreover, this study discussed the role of the latest AI-based soft sensors in achieving sustainable development goals and provided a critical review of some of the latest works that investigated these AI-based techniques for addressing common challenges associated with soft sensor development.

A summary of the AI-driven soft sensors reviewed in Section 5 is provided in Table 2. This should be useful for the readers to compare the performance of advanced AI-based soft sensors with traditional soft sensing solutions. Table 3 presents a summary of suitable soft sensing solutions for addressing the problems/challenges discussed in Section 5. It can be used as a guide by soft sensor designers to select the most favorable option, out of the wide range of soft computing algorithms. However, it should be noted that the soft sensing solutions provided in Table 3 are not exhaustive.

6.1. Current challenges in soft sensor design, application, and maintenance

As explained in Section 4, the prediction accuracy and computational efficiency of soft sensors play a key role in ensuring the sustainability of the process under consideration. Problems/challenges that occur during different stages of the life cycle of a soft sensor can inhibit their prediction accuracy as well as computational efficiency, and this could hinder the implementation of sustainable manufacturing strategies in process industries. Hence, it is important to identify these challenges and investigate solutions to address them. Some of the current challenges in soft sensor design, application, and maintenance stages are discussed here.

Soft sensors are mainly developed based on the historical data collected from hardware sensors installed in process industries. Therefore, the quality of data collected by these hardware sensors plays a key role in the soft sensor development process, as the final predictive performance of the soft sensor is directly dependent upon the data used in developing the soft sensor. Data quality or veracity is one of the key aspects that constitute the 5Vs (i.e., volume, velocity, variety, veracity, and value) of the big data paradigm (see Fig. 5(c)) (Gao et al., 2020a). The use of high-quality hardware sensors for data collection is a key requirement for ensuring the veracity of the data used for training and validating soft sensors. Abeykoon (2018) recommended using high-quality hardware sensors instead of using filtering techniques to remove noise from data. Filtering might treat fluctuations in data caused by temporal variations as noise and filter them out, and this could hurt the ability of the soft sensor to predict such fluctuations. Furthermore, hardware sensors must be properly calibrated from time to time and their functionality should be monitored regularly.

Capturing a set of data that is representative of the entire range of operating conditions in which the manufacturing system operates, is another challenge faced by soft sensor developers. Manufacturing processes are highly dynamic in nature and show numerous temporal variations; hence there could be fluctuations in the collected data that show deviations from the steady-state behavior of the process. If these variations are not captured in the dataset, the predictive performance of the soft sensor would be adversely affected. Therefore, data collection should be carried out for a sufficient period of time, such that the temporal variations in the system are captured.

As discussed in Section 1.1, characteristics of process data such as missing values, outliers, collinearity, and varying sampling rates also have a direct impact on the veracity of data (Kadlec et al., 2009; Souza et al., 2015). It is crucial to address these issues using appropriate data pre-processing techniques, to obtain an accurate dataset before training

Table 2
A summary of the soft sensor applications reviewed in this study.

Publication	Year	Industry	Models/ Algorithms Used	Application	Accuracy/Performance/Comments
Xie et al. (2020a)	2020	Polymer Processing	Two-stream λ GRU	Estimating the MFI in a polyester polymerization process	Reported an MSE of 0.00238 on an unseen dataset, outperforming soft sensors based on SVR, PLS, SAE and LSTM, with respective MSE values of 0.00890, 0.08, 0.00972, and 0.00887.
Xie et al. (2020b)	2020	Polymer Processing	DVAEs	Estimating the MFI in a polyester polymerization process	Reported an MSE of 0.0319 on an unseen dataset, outperforming soft sensors based on MLP neural network, AE, and VAE, with respective MSE values of 0.1262, 0.0821, and 0.0683. The model showed superior data reconstruction ability compared to the traditional deletion, mean imputation, and PCA imputation methods.
He et al. (2020b)	2020	Polymer Processing	SVD-ESN	Estimating the melt density index in an HDPE polymerization process	Reported an RMSE of 0.0599 on an unseen dataset, outperforming ESN, ELM, and LSTM models with respective RMSE values of 0.2311, 0.1539, and 0.2600.
Wu et al. (2021)	2021	Polymer Processing	DCNN	Estimating the melt index in a propylene polymerization process	Reported an RMSE of 0.0289, surpassing CNN and ELM models with respective RMSE values of 0.0865 and 0.0945.
Zhu et al. (2021c)	2021	Polymer Processing	CGAN	Estimating the MFI in an HDPE polymerization process	Reported an RMSE of 0.4995, surpassing conventional VSG methods based on bootstrap, mega trend diffusion, and tree-based trend diffusion, with respective RMSE values of 0.5298, 0.5574, and 0.6217.
Zhang et al. (2021)	2021	Polymer Processing	OESN	Estimating the melt index in a propylene polymerization process	Reported an RMSE of 0.0060.
Hu et al. (2021)	2021	Polymer Processing	STAN	Estimating the melt index in a propylene polymerization process	Reported an RMSE of 0.0661, surpassing ELM, backpropagation neural network, LSTM, and convolutional LSTM models with respective RMSE values of 0.1437, 0.1449, 0.1380, and 0.1132.
Liu et al. (2018)	2018	Polymer Processing	Ensemble Deep Kernel Learning	Estimating the MFI in a polyethylene polymerization process	Reported an RMSE of 2.74, outperforming SVR, PLS, and single DBN models with RMSE values of 4.70, 5.91, and 3.51.
Zhu et al. (2021b)	2021	Polymer Processing	Xgboost-BiCG-LSTM-SEA	Estimating the melt intrinsic viscosity of a polyester polymerization process.	Reported an MSE of 0.0018, outperforming soft sensor models based on two-stream λ GRU, SAE, PLS, and SVR, with respective MSE values of 0.00401, 0.0091, 0.0112, and 0.0098.
Liu et al. (2019)	2019	Polymer Processing	DAELM	Estimating the MFI in a polyethylene polymerization process	Reported lower relative prediction errors compared to an RELM for three different material grades. The model is adaptive to the grade changeovers.
Zhang et al. (2021)	2021	Polymer Processing	GOESN	Estimating the MFI in a polypropylene polymerization process	Reported an RMSE of 0.0077, outperforming orthogonal ESN and conventional ESN models with RMSE values of 0.0079 and 0.2132.
Yuan et al. (2020e)	2020	Petroleum Refining	SQAE	Estimating the butane content in a debutanizer column	Reported an RMSE of 0.0303, outperforming SAE, SDAE, and SSAE models with respective RMSE values of 0.0438, 0.0427, and 0.0412.
Yuan et al. (2020a)	2020	Petroleum Refining	HVW-SAE	Estimating the butane content in a debutanizer column	Reported an RMSE of 0.0308, outperforming MLP neural network, SAE, and variable-wise weighted SAE (using only the Pearson correlation coefficient) models with respective RMSE values of 0.0542, 0.0450, and 0.0389.
Liu et al. (2021d)	2021	Petroleum Refining	S-NPAE	Estimating the initial and final distillation temperatures of heavy naphtha in a hydrocracking process	Reported RMSE values of 0.0905 and 0.0782 for predicting initial and final distillation temperatures respectively, outperforming an SAE model with respective RMSE values of 0.1373 and 0.1099.
Liu et al. (2021b)	2021	Petroleum Refining	STNP-SAE	Estimating the initial and final distillation temperatures of heavy naphtha in a hydrocracking process	Reported RMSE values of 0.0698 and 0.1179 for predicting initial and final distillation temperatures respectively, outperforming an SAE model with respective RMSE values of 0.0838 and 0.1451.
Liu et al. (2020)	2020	Petroleum Refining	AE + NBWR	Estimating the butane content in a debutanizer column	Reported an RMSE of 0.0405, outperforming PLS, least squares SVR, and Deep ELM models with RMSE values of 0.1414, 0.0591, and 0.0565 respectively. The model is adaptive to varying process conditions.

(*continued on next page*)

Table 2 (*continued*).

Guo et al. (2020b)	2020	Petroleum Refining	VAE + GPR	Estimating the butane content in a debutanizer column	Reported an RMSE of 0.1014, while reducing the input variables to three latent variables. The model is adaptive to varying process conditions.
Guo et al. (2020a)	2020	Petroleum Refining	VAE + GPR	Estimating the butane content in a debutanizer column	Reported RMSE values in the range 0.0779-0.1195 for different missing data levels between 10% and 50%. The model is adaptive to varying process conditions and able to handle missing data.
Chen and Ge (2021)	2021	Petroleum Refining	GMCE-STAE	Estimating the butane content in a debutanizer column	Reported an RMSE of 0.021 ± 0.007, outperforming soft sensors based on deep neural network, SAE, stacked target AE, variable-wise weighted SAE, dynamic CNN, and supervised LSTM with respective RMSE values of 0.033 ± 0.005, 0.028 ± 0.010, 0.027 ± 0.002, 0.025 ± 0.010, 0.050 ± 0.011, and 0.049 ± 0.008. The model is interpretable but is not adaptive to varying process conditions.
Graziani and Xibilia (2019)	2019	Petroleum Refining	DBN	Estimating the butane content in a debutanizer column	An RMSE of 2.07 and a correlation coefficient of 0.74 were reported.
Yuan et al. (2020d)	2020	Petroleum Refining	SIAE	Estimating the SO_2 concentration in the tail gas of an SRU	Reported an RMSE of 0.0279, outperforming SAE, MLP neural network, and shallow two-layer neural network models with respective RMSE values of 0.0290, 0.0297, and 0.0340.
Hikosaka et al. (2020)	2020	Petroleum Refining	EGAVDS	Estimating the H_2S concentration in the tail gas of an SRU and the butane content in a debutanizer column.	Reported an RMSE of 0.031 and 0.102 for the SRU and the debutanizer column respectively.
Chou et al. (2020)	2020	Petroleum Refining	GRU	Estimating the distillate impurity and bottom product impurity of a distillation column.	The soft sensor outperformed a conventional ANN in terms of prediction accuracy as well as consistent physical interpretations.
Liu et al. (2021c)	2021	Petroleum Refining	S-MMAE	Estimating the final distillation temperature of heavy naphtha and aviation kerosene in a hydrocracking process.	Reported RMSE values of 0.0721 and 0.0626 for predicting the distillation temperature of heavy naphtha and aviation kerosene respectively. Outperformed ANN, SAE, and LAE models with respective RMSE values of 0.1111, 0.0946, and 0.0854 for heavy naphtha, and 0.1001, 0.0960, and 0.0824 for aviation kerosene.
Zhang et al. (2019b)	2019	Petroleum Refining	WAR-LSTM	Estimating the product flowrate and the product yields of an FCCU.	Reported an MSE of 0.80 ± 0.08, outperforming a deep LSTM with an MSE value of 1.58 ± 0.47.
Yi et al. (2020)	2020	Petroleum Refining	CNN+NNR+RVFL	Estimating the fraction yields of crude oil	A multiple ensemble of the proposed model reported an RMSE of 0.2524, outperforming a single ensemble of the proposed model, a PLS, an MLP neural network, a DBN, and an NNR with RMSE values of 0.2618, 0.8412, 0.8930, 0.5470, and 0.7209 respectively.
Yuan et al. (2020b)	2020	Petroleum Refining	SS-SAE	Estimating the final boiling point of aviation kerosene in a hydrocracking process and the butane content in a debutanizer column.	The soft sensor outperformed ANN, DBN and SAE models in the presence of different levels of unlabeled data.
Zhu et al. (2021d)	2021	Pharmaceutical	IBDA-RELM	Estimating the Pichia pastoris cell concentration in a fermentation process.	Reported RMSE values in the range 1.4932-2.3803 under different operating conditions. The model is adaptive to multiple operating conditions.
Gopakumar et al. (2018)	2018	Pharmaceutical	SOM	Estimating the streptokinase and biomass concentrations in a streptokinase fermentation process.	Reported RMSE values of 0.0565 and 0.0091 for streptokinase and biomass concentration prediction respectively, outperforming SVR and supervised deep neural network models.
				Estimating the penicillin, biomass, and substrate concentrations in a penicillin fermentation process.	Reported RMSE values of 0.0274, 0.0069, and 0.00063 for predicting the penicillin, biomass, and substrate concentrations respectively, outperforming SVR and supervised deep neural network models. Performance increased significantly when the number of unlabeled datapoints increased.

(*continued on next page*)

Table 2 (*continued*).

Sun et al. (2020)	2020	Pharmaceutical	ORWNN-GPR	Estimating the total sugar content in a chlortetracycline fermentation process.	The model showed good predictive performance throughout the online learning process with respective RMSE values of 0.22 and 0.21 at 50% and 100% training data due to the integration of ORWNN and GPR models. The soft sensor has adaptive capabilities due to the online learning technique used.
Gao et al. (2021)	2021	Pharmaceutical	TS-SSRAE	Estimating the penicillin concentration in a penicillin fermentation process.	Reported an RMSE of 0.032, outperforming SAE, SSAE, and TS-SRAE models with respective RMSE values of 0.042, 0.039, and 0.035.
Zheng et al. (2021)	2021	Pharmaceutical	JITL-LSTM	Estimating the penicillin concentration in a penicillin fermentation process.	Reported an RMSE of 0.0820. The model is adaptive to different operating phases in the penicillin fermentation process but is not applicable when the batch lengths are not even.

Table 3
Suitable AI-driven soft sensing solutions for different classes of problems (based on the works reviewed in this study).

Class of problems	Comments
Addressing the issue of missing data	GANs, AEs, and their extensions can be used. These algorithms possess superior data reconstruction ability, which is ideal for handling missing data. Hence, they outperform traditional missing data treatment techniques such as deletion and mean imputation.
Utilizing unlabeled data to improve prediction accuracy (addressing the issue of varying sampling rates)	Unsupervised and semi-supervised learning techniques can be used. AE and its extensions, SOM, and DBN models have widely been used for predicting quality variables with low sampling rates. These techniques utilize both labeled and unlabeled data, resulting in improved predictive performance compared to the traditional supervised learning techniques that utilize labeled data only.
Improving the predictive performance of models trained using small datasets	VSG techniques based on the GAN and its extensions can be used. They create virtual data points in the sparse areas of the dataset to expand the dataset size. This approach has the potential to provide better results compared to more traditional techniques such as bootstrap aggregation and noise injection
Input regressor selection and dimensionality reduction	Unsupervised and semi-supervised learning techniques can be used. AEs and DBNs can extract nonlinear features from the input data while reducing the dimensions of the input space. Extensions of the traditional AE algorithm can extract only the features that are relevant to the quality parameter to be predicted while reducing the dimensions. This results in improved predictive performance.
Addressing the issue of process drifts (i.e., varying process conditions)	JITL methods, transfer learning, and online learning techniques can be used.
Extracting temporal and spatial features	RNNs and their extensions such as the LSTM, GRU, and ESN as well as CNN and its extensions have widely been used.
Model interpretation	Gray-box modeling approaches are widely used
Model structure optimization	The use of evolutionary algorithms is less time consuming and less tedious than the trial-and-error based approaches and almost always produce better results.
Soft sensing solutions with faster computation	ELM, GRU, and ESN algorithms generally provide faster computation due to the simpler architecture and less trainable parameters.

the soft sensor. Some of the latest solutions proposed for addressing some of these issues have been discussed in Section 5.

In addition to the difficulties associated with collecting a high-quality dataset, data pre-processing, model selection, validation, and maintenance aspects of soft sensors present various challenges due to the ad-hoc manner in which these tasks are carried out. Kadlec and Gabrys (2009) predicted that these stages of soft sensor development would be more automated in the future (see Fig. 5(b)). However, even after a decade since their study, data pre-processing, model selection, and validation are still mostly carried out manually. Performing these tasks manually takes up a lot of time and effort and may not result in an optimum solution. However, due to the increased use of AI-based modeling techniques, the model maintenance aspect has become slightly more automated at present. JITL models, online learning, and transfer learning techniques have become more popular in constructing adaptive soft sensors, and these sensors can adapt to varying process conditions without needing to be retrained manually (Liu et al., 2019; Sun et al., 2020; Zheng et al., 2021) However, most of the soft sensing solutions reported in the literature are still limited in terms of adaptive capabilities.

The current challenges associated with soft sensor design, application, and maintenance hinder the development of high-performance robust soft sensors for process industries. Hence, these challenges should be addressed with a systematic approach, so that future process industries could benefit from increased soft sensor applications and their contribution toward sustainability. In the following section, the means by which the future soft sensors can be improved are discussed in detail.

6.2. Soft sensors in future process industries

In the past, applications of data-driven models were limited by the lack of data availability and lack of computational power (Ge, 2017). Today, computers are equipped with high-performance processors and GPUs that are capable of processing large amounts of data in less time. Furthermore, the use of cloud computing has given process industries access to high processing power. The adoption of the concept of the internet of things (IoT) has led to increased availability and easy access to large quantities of data (Fisher et al., 2020). This has fueled the use of data-driven soft sensors in current process industries. With the advancement of AI technologies, cloud computing, IoT, and Industry 4.0, the use of data-driven soft sensors is expected to grow in number in the future.

It is expected that the development of future soft sensors will be focused on overcoming the current challenges associated with the design, application, and maintenance aspects. Fisher et al. (2020) identified poor generalization and overfitting as the main sources of error in data-driven models. In soft sensor development, these can be caused by the difficulties associated with capturing high-quality process data. Due to measurement delays and missing data, the amount of available data may be limited. Captured data may not represent the temporal variations associated with the process dynamics. Fisher et al. (2020) suggested adhering to a structured methodology when developing data-driven models which enable the developers to avoid such mistakes that lead to poor generalization and overfitting. They recommended the Cross-Industry Standard Process for Data Mining (CRISP-DM) (Shearer, 2000) methodology as it is the most widely used methodology for developing data-driven models (see Fig. 5(a)).

As discussed in Section 6.1, during the design stage of soft sensors, data pre-processing, model selection, validation, and maintenance steps are performed manually. Kadlec and Gabrys (2009) proposed a framework for automating the soft sensor development process through a systematic approach. The proposed framework has an architecture of three hierarchical levels of information processing which incorporate local learning and meta-learning concepts into the soft sensor development process (see Fig. 5(d)). Furthermore, this framework allows the incorporation of expert knowledge and adaptive capabilities at all three hierarchical levels. Although this study is more than a decade old, the proposed framework is still relevant today, and an increase in the incorporation of expert knowledge and adaptive capabilities in the most recent studies can be observed. Future soft sensor designers can use it as a guide for automating different stages of the soft sensor development process with the aid of powerful AI-based algorithms available.

The lack of model interpretability is one of the major limitations in data-driven soft sensors, due to the black-box nature of the model structure. This affects the reliability of the features extracted by the soft sensor and its final output. As discussed in Section 5.5, gray-box models that combine the benefits of both first-principles and black-box models have widely been used to address this issue (Ahmad et al., 2020). Despite the increase in the use of gray-box models, model interpretability still remains a challenge in the soft sensor development process. One of the most recent studies investigated a deep neural network based on attention mechanisms to address this issue (Guo et al., 2022). Here, the weight coefficients calculated by the attention mechanisms were used to interpret the soft sensor data selection. It is clear that future soft sensors will be heavily based on such novel AI-based approaches with the aim of making them more robust and reliable.

The digital twin technology is an emerging concept in the IoT era. A digital twin is a digital representation of a physical system. Digital twins can be employed to simulate manufacturing processes, where the twin simulates the physical system in real-time, based on the inputs gathered from sensors (He et al., 2019). Digital twins are a promising approach for data-driven modeling and in the future, soft sensors will be used as digital twins in process industries.

Web-based sensor networks equipped with communication units and measurement devices (Fukatsu et al., 2011) are another flexible data processing approach, from which future soft sensors can benefit. Despite the numerous advantages they offer such as the possibility of remote access and control, and increased flexibility, several risks could also be involved due to the use of the internet. With the increased use of big data, concerns regarding data security should be properly addressed. The risks involved with the use of big data could prevent the manufacturers from utilizing them in soft sensor applications. Traditional approaches such as encryption, network segmentation, and virtual local area networks as well as state-of-the-art technologies such as cryptographic hashing and digital certificates can be incorporated within the manufacturing facility to ensure data security (Gao et al., 2020a).

Finally, it should be noted that most of the soft sensing solutions reported in the literature have been limited to laboratory-scale experiments. In many cases, data collection and soft sensor testing have been carried out through simulation, and the implementation of soft sensors in a real industrial setting for process monitoring and control is also limited. Hence, future soft sensor designers should consider scaling up the soft sensor development process from the laboratory to actual industrial scenarios and should develop control systems that incorporate these soft sensors for real-time quality control mechanisms.

7. Conclusions

The current trend toward sustainable development drives process industries to adopt energy conservation and emission reduction strategies to reduce the global carbon footprint and adverse environmental issues. Consequently, process monitoring has become vital to ensure that the processes operate within their desired boundaries, via advanced process optimization and control strategies, so that material and energy wastage and environmental pollution can be minimized while improving production efficiency. Today, soft sensors are widely used across process industries for online prediction, process monitoring, and fault detection applications with the aim of achieving the said goals. The rise in IoT, Industry 4.0, AI, and big data concepts as well as the increased processing power enabled by cloud computing technologies, have fueled the growth of soft sensors and the numerous benefits that they can offer have led them to be suitable candidates even to replace some hardware sensors. The latest AI-based modeling techniques have the potential to enhance the prediction accuracy and the computational efficiency of soft sensors which enables accurate and continuous monitoring of industrial processes. Furthermore, with the recent advancements in mechatronics, soft sensors can be incorporated into feedback control systems, which enables the implementation of real-time low-cost quality control strategies. These robust process monitoring and control strategies ultimately help in achieving the sustainable development goals of process industries.

Despite their advantages, the design, application, and maintenance of soft sensors present numerous challenges. All stages of soft sensor development, comprising the collection of process data, data pre-processing, model selection and validation, and soft sensor maintenance are associated with various limitations and challenges which inhibit the growth of soft sensor applications in process industries. The goal of the next generation soft sensor development should be to adopt more automated strategies rather than manual methods during the soft sensor design and application stages and their long-term predictive performance should be ensured by equipping them with adaptive capabilities. This goal should be achieved using more systematic approaches through the incorporation of data-driven modeling and soft sensor development frameworks. Implementation of these future directions will allow the soft sensor developers to overcome the current challenges associated with soft sensor applications and the future process industries will thrive on more advanced applications of soft sensors. Eventually, these approaches should be invaluable in enabling a sustainable energy future with a cleaner/greener environment for the future generation.

Abbreviations

ACO Ant Colony Optimization

AE Autoencoder

AI Artificial Intelligence

AM-FOA Adaptive Mutation Fruit-fly Optimization Algorithm

ANFIS Adaptive Neuro-Fuzzy Inference System

ANN Artificial Neural Network

BAS Beetle Antennae Search

BDA Balanced Distributed Adaptation

BiCG-LSTM Bidirectional Converted Gates Long Short-Term Memory

BiLSTM Bidirectional Long Short-Term Memory

CGAN Conditional Generative Adversarial Network

CG-LSTM Converted Gates Long Short-Term Memory

CNN Convolutional Neural Network

COA Cuckoo Optimization Algorithm

COD Chemical Oxygen Demand

CRISP-DM Cross-Industry Standard Process for Data Mining

DAELM Domain Adaptation Extreme Learning Machine

DBN Deep-Belief Network

DCNN Dilated Convolutional Neural Network

DE Differential Evolution

DVAE Deep Variational Autoencoder

EGAVDS Ensemble Genetic Algorithm-based process Variables and Dynamics Selection

ELM Extreme Learning Machine

ESN Echo State Network

eTS Evolving Takagi–Sugeno

FCCU Fluid Catalytic Cracking Unit

FIS Fuzzy Inference System

FLNN Functional Link Neural Network

FNN Fuzzy Neural Network

FOA Fruit-fly Optimization Algorithm

FRA Fast Recursive Algorithm

GA Genetic Algorithm

GAN Generative Adversarial Network

GAVDS Genetic Algorithm-based process Variables and Dynamics Selection

GMCE-STAE Graph Mining, Convolution, and Explanation Stacked Target-related Autoencoder

GPR Gaussian Process Regression

GPU Graphics Processing Unit

GRU Gated Recurrent Unit

HDPE High-Density Polyethylene

HVW-SAE Hybrid Variable-wise Weighted Stacked Autoencoder

IBDA Improved Balanced Distributed Adaptation

IEA Immune Evolutionary Algorithm

IoT Internet of Things

JDA Joint Distributed Adaptation

JITL Just-In-Time Learning

KL Kullback–Leibler

LAE Laplacian Regularization Autoencoder

LS-SVM Least Squares Support Vector Machine

LSTM Long Short-Term Memory

MFI Melt Flow Index

MLP Multilayer Perceptron

MLR Multiple Linear Regression

MSE Mean Squre Error

MUDVAE Modified Unsupervised Deep Variational Autoencoder

NBWR Nonlinear Bayesian Weighted Regression

NFS Neuro Fuzzy System

NNR Nearest Neighbor Regression

NPE Normalized Prediction Error

OESN Orthogonal Echo State Network

ORWNN Output Recursive Wavelet Neural Network

PCA Principal Component Analysis

PLS Partial Least Squares

PSO Particle Swarm Optimization

PTA Purified Terephthalic Acid

RBF Radial Basis Function

RBM Restricted Boltzmann Machine

RELM Regularized Extreme Learning Machine

RMSE Root Mean Square Error

RNN Recurrent Neural Network

RVFL Random Vector Functional Link

SAE Stacked Autoencoder

SDVAE Supervised Deep Variational Autoencoder

SEA Self-Attention

SIAE Stacked Isomorphic Autoencoder

S-MMAE Stacked Multi-Manifold Autoencoder

S-NPAE Stacked Neighborhood Preserving Autoencoder

SOM Self-Organizing Map

SQAE Stacked Quality-driven Autoencoder

SRU Sulfur Recovery Unit

SS-SAE Semi-Supervised Stacked Autoencoder

STAN Spatio-Temporal Attention Network

STNP-SAE Spatiotemporal Neighborhood Preserving Stacked Autoencoder

SVD Singular Value Decomposition

SVM/SVR Support Vector Machine/Regression

TS-SRAE Teacher Student Stacked Recurrent Autoencoder

TS-SSRAE Teacher Student Stacked Sparse Recurrent Autoencoder

T–S Takagi–Sugeno

VAE Variational Autoencoder

VSG Virtual Sample Generation

WAR-LSTM Weighted Auto Regressive Long Short-Term Memory

Xgboost eXtreme Gradient Boosting

CRediT authorship contribution statement

Yasith S. Perera: Writing (original draft), Revision, Formatting, Analysis, Planning. **D.A.A.C. Ratnaweera:** Writing (review & editing), Planning, Supervision. **Chamila H. Dasanayaka:** Writing (review & editing), Formatting, Analysis, Planning. **Chamil Abeykoon:** Coordination, Writing (review & editing), Planning, Supervision, Analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No datasets were generated or analyzed in this study.

Acknowledgments

We would like to acknowledge the funding support provided by Engineering and Physical Sciences Research Council (EPSRC), UK under the grant number EP/T517823/1.

References

Abeykoon, C., 2014a. A novel model-based controller for polymer extrusion. IEEE Trans. Fuzzy Syst. 22 (6), 1413–1430. http://dx.doi.org/10.1109/TFUZZ.2013.2293348.

Abeykoon, C., 2014b. A novel soft sensor for real-time monitoring of the die melt temperature profile in polymer extrusion. IEEE Trans. Ind. Electron. 61 (12), 7113–7123. http://dx.doi.org/10.1109/TIE.2014.2321345.

Abeykoon, C., 2016a. Single screw extrusion control: A comprehensive review and directions for improvements. Control Eng. Pract. 51, 69–80. http://dx.doi.org/10.1016/j.conengprac.2016.03.008.

Abeykoon, C., 2016b. Soft sensing of melt temperature in polymer extrusion. In: 2016 European Control Conference. ECC, Aalborg, Denmark, pp. 340–345. http://dx.doi.org/10.1109/ECC.2016.7810308.

Abeykoon, C., 2018. Design and applications of soft sensors in polymer processing: A review. IEEE Sens. J. 19 (8), 2801–2813. http://dx.doi.org/10.1109/JSEN.2018.2885609.

Abeykoon, C., Li, K., McAfee, M., Martin, P.J., Niu, Q., Kelly, A.L., Deng, J., 2011. A new model based approach for the prediction and optimisation of thermal homogeneity in single screw extrusion. Control. Eng. Pract. 19 (8), 862–874. http://dx.doi.org/10.1016/j.conengprac.2011.04.015.

Abeykoon, C., Martin, P.J., Kelly, A.L., Brown, E.C., 2012. A review and evaluation of melt temperature sensors for polymer extrusion. Sensors Actuators A 182, 16–27. http://dx.doi.org/10.1016/j.sna.2012.04.026.

Abeykoon, C., Pérez, P., Kelly, A.L., 2020. The effect of materials' rheology on process energy consumption and melt thermal quality in polymer extrusion. Polym. Eng. Sci. 60 (6), 1244–1265. http://dx.doi.org/10.1002/pen.25377.

Agatonovic-Kustrin, S., Beresford, R., 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J. Pharm. Biomed. Anal. 22 (5), 717–727. http://dx.doi.org/10.1016/S0731-7085(99)00272-1.

Ahmad, I., Ayub, A., Kano, M., Cheema, I.I., 2020. Gray-box soft sensors in process industry: Current practice, and future prospects in era of big data. Processes 8 (2), http://dx.doi.org/10.3390/pr8020243.

Al-Jamimi, H.A., Al-Azani, S., Saleh, T.A., 2018. Supervised machine learning techniques in the desulfurization of oil products for environmental protection: A review. Process Saf. Environ. Prot. 120, 57–71. http://dx.doi.org/10.1016/j.psep.2018.08.021.

Angelov, P., Kordon, A., Zhou, X., 2008. Evolving fuzzy inferential sensors for process industry. In: 2008 3rd International Workshop on Genetic and Evolving Systems, Vol. 4. Witten-Bommerholz, Germany, pp. 1–46. http://dx.doi.org/10.1109/GEFS.2008.4484565.

Behnasr, M., Jazayeri-Rad, H., 2015. Robust data-driven soft sensor based on iteratively weighted least squares support vector regression optimized by the cuckoo optimization algorithm. J. Nat. Gas Sci. Eng. 22, 35–41. http://dx.doi.org/10.1016/j.jngse.2014.11.017.

Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Netw. 5 (2), 157–166. http://dx.doi.org/10.1109/72.279181.

Bur, A.J., Vangel, M.G., Roth, S.C., 2001. Fluorescence based temperature measurements and applications to real-time polymer processing. Polym. Eng. Sci. 41 (8), 1380–1389. http://dx.doi.org/10.1002/pen.10838.

Chen, Z., Ge, Z., 2021. Knowledge automation through graph mining, convolution and explanation framework: A soft sensor practice. IEEE Trans. Ind. Inform. 18 (9), 6068–6078. http://dx.doi.org/10.1109/TII.2021.3127204.

Chen, G., Yu, J., 2005. Particle swarm optimization neural network and its application in soft-sensing modeling. In: Wang, L., Chen, K., Ong, Y.S. (Eds.), Advances in Natural Computation. Springer, Berlin, pp. 610–617.

Cheng, F., He, Q.P., Zhao, J., 2019. A novel process monitoring approach based on variational recurrent autoencoder. Comput. Chem. Eng. 129 (4), 106515. http://dx.doi.org/10.1016/j.compchemeng.2019.106515.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv:1406.1078 [cs, stat].

Chou, C.-H., Wu, H., Kang, J.-L., Wong, D.S.-H., Yao, Y., Chuang, Y.-C., Jang, S.-S., Ou, J.D.-Y., 2020. Physically consistent soft-sensor development using sequence-to-sequence neural networks. IEEE Trans. Ind. Inf. 16, 2829–2838. http://dx.doi.org/10.1109/TII.2019.2952429.

Cioffi, R., Travaglioni, M., Piscitelli, G., Petrillo, A., De Felice, F., 2020. Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions. Sustainability 12 (2), 492. http://dx.doi.org/10.3390/su12020492.

Cui, P., Li, J., Wang, G., 2008. Improved kernel principal component analysis for fault detection. Expert Syst. Appl. 34 (2), 1210–1219. http://dx.doi.org/10.1016/j.eswa.2006.12.010.

Curreri, F., Fiumara, G., Xibilia, M.G., 2020. Input selection methods for soft sensor design: A survey. Future Internet 12 (6), 97. http://dx.doi.org/10.3390/fi12060097.

Curreri, F., Patanè, L., Xibilia, M.G., 2021. Soft sensor transferability: A survey. Appl. Sci. 11 (16), 7710. http://dx.doi.org/10.3390/app11167710.

Deb, A.K., 2011. Introduction to soft computing techniques: artificial neural networks, fuzzy logic and genetic algorithms. In: Soft Computing in Textile Engineering. Elsevier, pp. 3–24.

Deng, J., Li, K., Harkin-Jones, E., Price, M., Fei, M., Kelly, A., Vera-Sorroche, J., Coates, P., Brown, E., 2014. Low-cost process monitoring for polymer extrusion. Trans. Inst. Meas. 36 (3), 382–390. http://dx.doi.org/10.1177/0142331213502696.

Di Bella, A., Fortuna, L., Graziani, S., Napoli, G., Xibilia, M.G., 2007. Development of a soft sensor for a thermal cracking unit using a small experimental data set. In: 2007 IEEE International Symposium on Intelligent Signal Processing. Alcala de Henares, Spain, pp. 1–6.

Elkington, J., 1994. Towards the sustainable corporation: Win-Win-Win business strategies for sustainable development. Calif. Manage. Rev. 36 (2), 90–100. http://dx.doi.org/10.2307/41165746.

Elkington, J., 1999. Triple bottom line revolution: reporting for the third millennium. Aust. CPA 69 (11), 75–76.

Elkington, J., 2018. 25 Years ago I coined the phrase "Triple Bottom Line". Here's why it's time to rethink it. https://hbr.org/2018/06/25-years-ago-i-coined-the-phrase-triple-bottom-line-heres-why-im-giving-up-on-it. (Accessed 5 April 2022).

Elkington, J., 2019. Beyond the Triple Bottom Line - An interview with John Elkington. https://www.marketingjournal.org/beyond-the-triple-bottom-line-interview-with-john-elkington/. (Accessed 5 April 2022).

Farahani, H.S., Fatehi, A., Nadali, A., Shoorehdeli, M.A., 2021. Domain adversarial neural network regression to design transferable soft sensor in a power plant. Comput. Ind. 132, http://dx.doi.org/10.1016/j.compind.2021.103489.

Fernandez de Canete, J., del Saz-Orozco, P., Gómez-de Gabriel, J., Baratti, R., Ruano, A., Rivas-Blanco, I., 2021. Control and soft sensing strategies for a wastewater treatment plant using a neuro-genetic approach. Comput. Chem. Eng. 144, http://dx.doi.org/10.1016/j.compchemeng.2020.107146.

Fisher, O.J., Watson, N.J., Escrig, J.E., Witt, R., Porcu, L., Bacon, D., Rigley, M., Gomes, R.L., 2020. Considerations, challenges and opportunities when developing data-driven models for process manufacturing systems. Comput. Chem. Eng. 140, http://dx.doi.org/10.1016/j.compchemeng.2020.106881.

Fortuna, L., Graziani, S., Rizzo, A., Xibilia, M.G., 2007. Soft Sensors for Monitoring and Control of Industrial Processes, first ed. Springer, London.

Fortuna, L., Graziani, S., Xibilia, M.G., 2009. Comparison of soft-sensor design methods for industrial plants using small data sets. IEEE Trans. Instrum. Meas. 58 (8), 2444–2451. http://dx.doi.org/10.1109/TIM.2009.2016386.

Fukatsu, T., Kiura, T., Hirafuji, M., 2011. A web-based sensor network system with distributed data processing approach via web application. Comput. Stand. Interfaces 33 (6), 565–573. http://dx.doi.org/10.1016/j.csi.2011.03.002.

Gao, X., Meng, L., Gao, H., Han, H., Qi, Y., 2021. Fermentation process quality prediction using teacher student stacked sparse recurrent autoencoder. Can. J. Chem. Eng. 100 (10), 2907–2917. http://dx.doi.org/10.1002/cjce.24303.

Gao, R.X., Wang, L., Helu, M., Teti, R., 2020a. Big data analytics for smart factories of the future. CIRP Ann. 69 (2), 668–692. http://dx.doi.org/10.1016/j.cirp.2020.05.002.

Gao, S., Zhang, Y., Zhang, Y., Zhang, G., 2020b. Elman neural network soft-sensor model of PVC polymerization process optimized by chaos beetle antennae search algorithm. IEEE Sens. J. 21 (3), 3544–3551. http://dx.doi.org/10.1109/JSEN.2020.3026550.

Ge, Z., 2017. Review on data-driven modeling and monitoring for plant-wide industrial processes. Chemometr. Intell. Lab. Syst. 171, 16–25. http://dx.doi.org/10.1016/j.chemolab.2017.09.021.

Geng, Z., Chen, Z., Meng, Q., Han, Y., 2022. Novel transformer based on gated convolutional neural network for dynamic soft sensor modeling of industrial processes. IEEE Trans. Ind. Inform. 18 (3), 1521–1529. http://dx.doi.org/10.1109/TII.2021.3086798.

Geng, Z., Dong, J., Chen, J., Han, Y., 2017. A new Self-Organizing Extreme Learning Machine soft sensor model and its applications in complicated chemical processes. Eng. Appl. Artif. Intell. 62, 38–50. http://dx.doi.org/10.1016/j.engappai.2017.03.011.

Giret, A., Trentesaux, D., Prabhu, V., 2015. Sustainability in manufacturing operations scheduling: A state of the art review. J. Manuf. Syst. 37 (1), 126–140. http://dx.doi.org/10.1016/j.jmsy.2015.08.002.

González, G., 2015. Sustainability and Waste Management: A Case Study on an UK Oil Refinery During Day-To-Day and Turnaround Operations (MPhil thesis). University of Surrey.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. arXiv:1406.2661 [Cs, Stat].

Goodwin, G.C., 2000. Predicting the performance of soft sensors as a route to low cost automation. Annu. Rev. Control 24, 55–66. http://dx.doi.org/10.1016/S1367-5788(00)90013-0.

Gopakumar, V., Tiwari, S., Rahman, I., 2018. A deep learning based data driven soft sensor for bioprocesses. Biochem. Eng. J. 136, 28–39. http://dx.doi.org/10.1016/j.bej.2018.04.015.

Graziani, S., Xibilia, M.G., 2019. Design of a soft sensor for an industrial plant with unknown delay by using deep learning. In: 2019 IEEE International Instrumentation and Measurement Technology Conference. I2MTC, Auckland, New Zealand.

Guo, F., Bai, W., Huang, B., 2020a. Output-relevant variational autoencoder for just-in-time soft sensor modeling with missing data. J. Process Control 92, 90–97. http://dx.doi.org/10.1016/j.jprocont.2020.05.012.

Guo, R., Liu, H., 2021. A hybrid mechanism- and data-driven soft sensor based on the generative adversarial network and gated recurrent unit. IEEE Sens. J. 21 (22), 25901–25911. http://dx.doi.org/10.1109/JSEN.2021.3117981.

Guo, R., Liu, H., Xie, G., Zhang, Y., Liu, D., 2022. A self-interpretable soft sensor based on deep learning and multiple attention mechanism: From data selection to sensor modeling. IEEE Trans. Ind. Inform. 1–12. http://dx.doi.org/10.1109/TII.2022.3181692.

Guo, F., Xie, R., Huang, B., 2020b. A deep learning just-in-time modeling approach for soft sensor based on variational autoencoder. Chemom. Intell. Lab. Syst. 197, http://dx.doi.org/10.1016/j.chemolab.2019.103922.

He, R., Chen, G., Dong, C., Sun, S., Shen, X., 2019. Data-driven digital twin technology for optimized control in process systems. ISA Trans. 95, 221–234. http://dx.doi.org/10.1016/j.isatra.2019.05.011.

He, Y.-L., Geng, Z.-Q., Zhu, Q.-X., 2015. Data driven soft sensor development for complex chemical processes using extreme learning machine. Chem. Eng. Res. Des. 102, 1–11. http://dx.doi.org/10.1016/j.cherd.2015.06.009.

He, R., Li, X., Chen, G., Chen, G., Liu, Y., 2020a. Generative adversarial network-based semi-supervised learning for real-time risk warning of process industries. Expert Syst. Appl. 150, http://dx.doi.org/10.1016/j.eswa.2020.113244.

He, Y.-L., Tian, Y., Xu, Y., Zhu, Q.-X., 2020b. Novel soft sensor development using echo state network integrated with singular value decomposition: Application to complex chemical processes. Chemom. Intell. Lab. Syst. 200, http://dx.doi.org/10.1016/j.chemolab.2020.103981.

He, X.B., Yang, Y.P., 2008. Variable MWPCA for adaptive process monitoring. Ind. Eng. Chem. Res. 47 (2), 419–427. http://dx.doi.org/10.1021/ie070712z.

Henao-Hernández, I., Solano-Charris, E.L., Muñoz-Villamizar, A., Santos, J., Henríquez-Machado, R., 2019. Control and monitoring for sustainable manufacturing in the Industry 4.0: A literature review. IFAC-PapersOnLine 52 (10), 195–200. http://dx.doi.org/10.1016/j.ifacol.2019.10.022.

Hens, L., Block, C., Cabello-Eras, J.J., Sagastume-Gutierez, A., Garcia-Lorenzo, D., Chamorro, C., Herrera Mendoza, K., Haeseldonckx, D., Vandecasteele, C., 2018. On the evolution of "Cleaner Production" as a concept and a practice. J. Clean. Prod. 172, 3323–3333. http://dx.doi.org/10.1016/j.jclepro.2017.11.082.

Hikosaka, T., Aoshima, S., Miyao, T., Funatsu, K., 2020. Soft sensor modeling for identifying significant process variables with time delays. Ind. Eng. Chem. Res. 59 (26), 12156–12163. http://dx.doi.org/10.1021/acs.iecr.0c01655.

Hu, X., Geng, Z., Han, Y., Huang, W., Chen, K., Xie, F., 2021. Novel soft sensor model based on spatio-temporal attention. In: 2021 International Joint Conference on Neural Networks. IJCNN, IEEE, Shenzhen, China, pp. 1–7. http://dx.doi.org/10.1109/IJCNN52387.2021.9534088.

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006. Extreme learning machine: Theory and applications. Neurocomputing 70 (1–3), 489–501. http://dx.doi.org/10.1016/j.neucom.2005.12.126.

Jalee, E.A., Aparna, K., 2016. Neuro-fuzzy soft sensor estimator for benzene toluene distillation column. Procedia Technol. 25, 92–99. http://dx.doi.org/10.1016/j.protcy.2016.08.085.

Jia, F., Martin, E.B., Morris, A.J., 2000. Non-linear principal components analysis with application to process fault detection. Int. J. Syst. Sci. 31 (11), 1473–1487. http://dx.doi.org/10.1080/00207720050197848.

Jiang, H., Xiao, Y., Li, J., Liu, X., 2012. Prediction of the melt index based on the relevance vector machine with modified particle swarm optimization. Chem. Eng. Technol. 35 (5), 819–826. http://dx.doi.org/10.1002/ceat.201100437.

Jiang, Y., Yin, S., Dong, J., Kaynak, O., 2021. A review on soft sensors for monitoring, control, and optimization of industrial processes. IEEE Sens. J. 21 (11), 12868–12881. http://dx.doi.org/10.1109/JSEN.2020.3033153.

Kadlec, P., Gabrys, B., 2009. Soft sensors: where are we and what are the current and future challenges? IFAC Proc. 42 (19), 572–577. http://dx.doi.org/10.3182/20090921-3-TR-3005.00098.

Kadlec, P., Gabrys, B., Strandt, S., 2009. Data-driven soft sensors in the process industry. Comput. Chem. Eng. 33 (4), 795–814. http://dx.doi.org/10.1016/j.compchemeng.2008.12.012.

Kadlec, P., Grbić, R., Gabrys, B., 2011. Review of adaptation mechanisms for data-driven soft sensors. Comput. Chem. Eng. 35 (1), 1–24. http://dx.doi.org/10.1016/j.compchemeng.2010.07.034.

Kataria, G., Singh, K., 2018. Recurrent neural network based soft sensor for monitoring and controlling a reactive distillation column. Chem. Prod. Process Model. 13 (3), http://dx.doi.org/10.1515/cppm-2017-0044.

Kelly, A.L., Brown, E.C., Coates, P.D., 2005. Melt temperature field measurement: influence of extruder screw and die geometry. Plast. Rubber Compos. 34 (9), 410–416. http://dx.doi.org/10.1179/174328905X72003.

Kelly, A.L., Brown, E.C., Coates, P.D., 2006. The effect of screw geometry on melt temperature profile in single screw extrusion. Polym. Eng. Sci. 46 (12), 1706–1714. http://dx.doi.org/10.1002/pen.20657.

Kelly, A.L., Brown, E.C., Howell, K., Coates, P.D., 2008. Melt temperature field measurements in extrusion using thermocouple meshes. Plast. Rubber Compos. 37 (2–4), 151–157. http://dx.doi.org/10.1179/174328908X283393.

Khatibisepehr, S., Huang, B., Khare, S., 2013. Design of inferential sensors in the process industry: A review of Bayesian methods. J. Process Control 23 (10), 1575–1596. http://dx.doi.org/10.1016/j.jprocont.2013.05.007.

Klarin, T., 2018. The concept of sustainable development: From its beginning to the contemporary issues. Zagreb Int. Rev. Econ. Bus. 21 (1), 67–94. http://dx.doi.org/10.2478/zireb-2018-0005.

Klir, G.J., Yuan, B., 1995. Fuzzy Sets and Fuzzy Logic: Theory and Applications, first ed. Prentice Hall PTR, New Jersey.

Kong, X., Jiang, X., Zhang, B., Yuan, J., Ge, Z., 2022. Latent variable models in the era of industrial big data: Extension and beyond. Annu. Rev. Control 54, 167–199. http://dx.doi.org/10.1016/j.arcontrol.2022.09.005.

Lahiri, S.K., 2017. Soft sensors. In: Multivariable Predictive Control: Applications in Industry. John Wiley & Sons, Ltd., Chichester, pp. 145–165.

Lahiri, S.K., Khalfe, N.M., 2009. Novel soft sensor modeling and process optimization technique for commercial petrochemical plant. Asia-Pac. J. Chem. Eng. 5 (5), 721–731. http://dx.doi.org/10.1002/apj.399.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. In: Proceedings of the IEEE, Vol. 86, No. 11. pp. 2278–2324.

Lee, S., Kwak, M., Tsui, K.-L., Kim, S.B., 2019. Process monitoring using variational autoencoder for high-dimensional nonlinear processes. Eng. Appl. Artif. Intell. 83, 13–27. http://dx.doi.org/10.1016/j.engappai.2019.04.013.

Lemos, T., Campos, L.F., Melo, A., Clavijo, N., Soares, R., Câmara, M., Feital, T., Anzai, T., Pinto, J.C., 2021. Echo state network based soft sensor for monitoring and fault detection of industrial processes. Comput. Chem. Eng. 155, 107512. http://dx.doi.org/10.1016/j.compchemeng.2021.107512.

Li, J., Liu, X., 2011. Melt index prediction by RBF neural network optimized with an adaptive new ant colony optimization algorithm. J. Appl. Polym. Sci. 119 (5), 3093–3100. http://dx.doi.org/10.1002/app.33060.

Liu, J., Chen, D.-S., Shen, J.-F., 2010. Development of self-validating soft sensors using fast moving window partial least squares. Ind. Eng. Chem. Res. 49 (22), 11530–11546. http://dx.doi.org/10.1021/ie101356c.

Liu, J., Hou, J., Chen, J., 2021a. Dual-layer feature extraction based soft sensor methods and applications to industrial polyethylene processes. Comput. Chem. Eng. 154, http://dx.doi.org/10.1016/j.compchemeng.2021.107469.

Liu, X., Li, K., McAfee, M., Nguyen, B.K., McNally, G.M., 2012. Dynamic gray-box modeling for on-line monitoring of polymer extrusion viscosity. Polym. Eng. Sci. 52 (6), 1332–1341. http://dx.doi.org/10.1002/pen.23080.

Liu, K., Shao, W., Chen, G., 2020. Autoencoder-based nonlinear Bayesian locally weighted regression for soft sensor development. ISA Trans. 103, 143–155. http://dx.doi.org/10.1016/j.isatra.2020.03.011.

Liu, C., Wang, K., Wang, Y., Xie, S., Yang, C., 2021b. Deep nonlinear dynamic feature extraction for quality prediction based on spatiotemporal neighborhood preserving SAE. IEEE Trans. Instrum. Meas. 70, 1–10. http://dx.doi.org/10.1109/TIM.2021.3122187.

Liu, C., Wang, K., Wang, Y., Yuan, X., 2021c. Learning deep multi-manifold structure feature representation for quality prediction with an industrial application. IEEE Access 18 (9), 5849–5858. http://dx.doi.org/10.1109/TII.2021.3130411.

Liu, C., Wang, K., Ye, L., Wang, Y., Yuan, X., 2021d. Deep learning with neighborhood preserving embedding regularization and its application for soft sensor in an industrial hydrocracking process. Inf. Sci. 567, 42–57. http://dx.doi.org/10.1016/j.ins.2021.03.026.

Liu, Y., Xie, M., 2020. Rebooting data-driven soft-sensors in process industries: A review of kernel methods. J. Process Control 89, 58–73. http://dx.doi.org/10.1016/j.jprocont.2020.03.012.

Liu, Y., Yang, C., Gao, Z., Yao, Y., 2018. Ensemble deep kernel learning with application to quality prediction in industrial polymerization processes. Chemom. Intell. Lab. Syst. 174, 15–21. http://dx.doi.org/10.1016/j.chemolab.2018.01.008.

Liu, Y., Yang, C., Liu, K., Chen, B., Yao, Y., 2019. Domain adaptation transfer learning soft sensor for product quality prediction. Chemom. Intell. Lab. Syst. 192, 103813. http://dx.doi.org/10.1016/j.chemolab.2019.103813.

Ma, Z., Zhong, P., Li, J., Yin, Y., 2020. Soft sensor model of adsorbable organic halogen based on bleached pulp quality indices. Bioresources 15 (1), 62–77.

McAfee, M., Thompson, S., 2007. A soft sensor for viscosity control of polymer extrusion. In: 2007 European Control Conference. ECC, Kos, Greece, pp. 5671–5678.

Meng, Y., Lan, Q., Qin, J., Yu, S., Pang, H., Zheng, K., 2019. Data-driven soft sensor modeling based on twin support vector regression for cane sugar crystallization. J. Food Eng. 241, 159–165. http://dx.doi.org/10.1016/j.jfoodeng.2018.07.035.

Microsoft, 2022. Artificial Intelligence (AI) vs Machine Learning (ML). https://azure.microsoft.com/en-gb/overview/artificial-intelligence-ai-vs-machine-learning/#introduction. (Accessed 15 April 2022).

Mitchell, T.M., 1997. Machine Learning, McGraw-Hill Series in Computer Science. McGraw-Hill, New York.

Noor, R., Ahmad, Z., 2011. Neural network based soft sensor for prediction of biopolycaprolactone molecular weight using bootstrap neural network technique. In: 2011 3rd Conference on Data Mining and Optimization. DMO, Putrajaya, Malaysia, pp. 70–73. http://dx.doi.org/10.1515/cppm-2017-0044.

Pani, A.K., Mohanta, H.K., 2014. Soft sensing of particle size in a grinding process: Application of support vector regression, fuzzy inference and adaptive neuro fuzzy inference techniques for online monitoring of cement fineness. Powder Technol. 264, 484–497. http://dx.doi.org/10.1016/j.powtec.2014.05.051.

Pani, A.K., Mohanta, H.K., 2016. Online monitoring of cement clinker quality using multivariate statistics and Takagi–Sugeno fuzzy-inference technique. Control Eng. Pract. 57, 1–17. http://dx.doi.org/10.1016/j.conengprac.2016.08.011.

Pao, Y.-H., 1989. Adaptive Pattern Recognition and Neural Networks, first ed. Addison-Wesley Longman Publishing Co. Inc., Boston.

Phatwong, A., Koolpiruck, D., 2019. Kappa number prediction of pulp digester using LSTM neural network. In: ECTI-CON 2019–16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. Pattaya, Thailand, pp. 151–154. http://dx.doi.org/10.1109/ECTI-CON47248.2019.8955373.

Pisa, I., Santin, I., Morell, A., Vicario, J.L., Vilanova, R., 2019. LSTM-based wastewater treatment plants operation strategies for effluent quality improvement. IEEE Access 7, 159773–159786. http://dx.doi.org/10.1109/ACCESS.2019.2950852.

Qin, S., Li, W., Henry Yue, H., 1999. Recursive PCA for adaptive process monitoring. IFAC Proc. Vol. 32 (2), 6686–6691. http://dx.doi.org/10.1016/S1474-6670(17)57142-6.

Qiu, K., Wang, J., Wang, R., Guo, Y., Zhao, L., 2021. Soft sensor development based on kernel dynamic time warping and a relevant vector machine for unequal-length batch processes. Expert Syst. Appl. 182, http://dx.doi.org/10.1016/j.eswa.2021.115223.

Ramachandran, A., Rustum, R., Adeloye, A.J., 2019. Anaerobic digestion process modeling using Kohonen self-organising maps. Heliyon 5 (4), http://dx.doi.org/10.1016/j.heliyon.2019.e01511.

Rashid, S.H.A., Evans, S., Longhurst, P., 2008. A comparison of four sustainable manufacturing strategies. Int. J. Sustain. Eng. 1 (3), 214–229. http://dx.doi.org/10.1080/19397030802513836.

Samuel, A.L., 1959. Some studies in machine learning using the game of checkers. IBM J. Res. Dev. 3 (3), 210–229. http://dx.doi.org/10.1147/rd.33.0210.

Shakil, M., Elshafei, M., Habib, M.A., Maleki, F.A., 2009. Soft sensor for and using dynamic neural networks. Comput. Electr. Eng. 35 (4), 578–586. http://dx.doi.org/10.1016/j.compeleceng.2008.08.007.

Shearer, C., 2000. The CRISP-DM model; the new blueprint for data mining. J. Data Warehous. 5, 13–22.

Shen, F., Zheng, J., Ye, L., Ma, X., 2020. LSTM soft sensor development of batch processes with multivariate trajectory-based ensemble just-in-time learning. IEEE Access 8, 73855–73864. http://dx.doi.org/10.1109/ACCESS.2020.2988668.

Siddharth, K., Pathak, A., Pani, A.K., 2019. Real-time quality monitoring in debutanizer column with regression tree and ANFIS. J. Ind. Eng. Int 15, 41–51. http://dx.doi.org/10.1007/s40092-018-0276-4.

Sliskovic, D., Nyarko, E.K., Peric, N., 2004. Estimation of difficult-to-measure process variables using neural networks - a comparison of simple MLP and RBF neural network properties. In: Proceedings of the 12th IEEE Mediterranean Electrotechnical Conference. Dubrovnik, Croatia, pp. 387–390. http://dx.doi.org/10.1109/MELCON.2004.1346888.

Soares, S., Araujo, R., Sousa, P., Souza, F., 2011. Design and application of Soft Sensor using Ensemble Methods. In: ETFA2011. Toulouse, France, pp. 1–8. http://dx.doi.org/10.1109/ETFA.2011.6059061.

Souza, F.A., Araújo, R., Mendes, J., 2015. Review of soft sensor methods for regression applications. Chemometr. Intell. Lab. Syst. 152, 69–79. http://dx.doi.org/10.1016/j.chemolab.2015.12.011.

Sun, Q., Ge, Z., 2019. Probabilistic sequential network for deep learning of complex process data and soft sensor application. IEEE Trans. Ind. Inform. 15 (5), 2700–2709. http://dx.doi.org/10.1109/TII.2018.2869899.

Sun, Y., Han, X., Zhang, D., Sun, Q., Chen, X., Yao, M., Huang, S., Ma, D., Zhou, B., 2020. Study on online soft sensor method of total sugar content in chlorotetracycline fermentation tank. Open Chem. 18 (1), 31–38. http://dx.doi.org/10.1515/chem-2020-0004.

Sun, J., Meng, X., Qiao, J., 2021. Prediction of oxygen content using weighted PCA and improved LSTM network in MSWI process. IEEE Trans. Instrum. Meas. 70, 1–12. http://dx.doi.org/10.1109/TIM.2021.3058367.

Sun, K., Wu, X., Xue, J., Ma, F., 2019. Development of a new multi-layer perceptron based soft sensor for SO2 emissions in power plant. J. Process Control 84, 182–191. http://dx.doi.org/10.1016/j.jprocont.2019.10.007.

Tian, Y., He, Y.-L., Zhu, Q.-X., 2020. Soft sensor development using improved whale optimization and regularization-based functional link neural network. Ind. Eng. Chem. Res. 59 (43), 19361–19369. http://dx.doi.org/10.1021/acs.iecr.0c03839.

Tobias, R.D., 1995. An introduction to partial least squares regression. In: Proc. 20th Annu. SAS Users Group Int. Conf. SASInst. Cary, NC, pp. 1250–1257.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is All you Need. In: 31st Conference on Neural Information Processing Systems. NIPS 2017, Long Beach, CA, USA.

Wang, Y., 1999. A self-organizing neural-network-based fuzzy system. Fuzzy Sets Syst. 103 (1), 1–11. http://dx.doi.org/10.1016/S0165-0114(97)00196-6.

Wang, Y., Chen, X., 2017. On temperature soft sensor model of rotary kiln burning zone based on RS-LSSVM. In: 2017 36th Chinese Control Conference. CCC, Dalian, China, pp. 9643–9646. http://dx.doi.org/10.23919/ChiCC.2017.8028894.

Wang, X., Kruger, U., Lennox, B., 2003. Recursive partial least squares algorithms for monitoring complex industrial processes. Control Eng. Pract. 11 (6), 613–632. http://dx.doi.org/10.1016/S0967-0661(02)00096-5.

Wang, W., Liu, X., 2015. Melt index prediction by least squares support vector machines with an adaptive mutation fruit fly optimization algorithm. Chemometr. Intell. Lab. Syst. 141, 79–87. http://dx.doi.org/10.1016/j.chemolab.2014.12.007.

Wang, X., Liu, H., 2020. Data supplement for a soft sensor using a new generative model based on a variational autoencoder and wasserstein GAN. J. Process Control 85, 91–99. http://dx.doi.org/10.1016/j.jprocont.2019.11.004.

Wang, Y., Liu, D., Liu, C., Yuan, X., Wang, K., Yang, C., 2022. Dynamic historical information incorporated attention deep learning model for industrial soft sensor modeling. Adv. Eng. Inf. 52, 101590. http://dx.doi.org/10.1016/j.aei.2022.101590.

Wang, K., Shang, C., Liu, L., Jiang, Y., Huang, D., Yang, F., 2019a. Dynamic soft sensor development based on convolutional neural networks. Ind. Eng. Chem. Res. 58 (26), 11521–11531. http://dx.doi.org/10.1021/acs.iecr.9b02513.

Wang, J.-G., Xie, Z., Yao, Y., Yang, B.-H., Ma, S.-W., Liu, L.-L., 2019b. Soft sensor development for improving economic efficiency of the coke dry quenching process. J. Process Control 77, 20–28. http://dx.doi.org/10.1016/j.jprocont.2019.03.011.

Warne, K., Prasad, G., Rezvani, S., Maguire, L., 2004. Statistical and computational intelligence techniques for inferential model development: a comparative evaluation and a novel proposition for fusion. Eng. Appl. Artif. Intell. 17 (8), 871–885. http://dx.doi.org/10.1016/j.engappai.2004.08.020.

World Commission on Environment and Development (WCED), 1987. Our Common Future. The Brundtland Report, Oxford Univ. Press, Oxford.

Wu, H., Han, Y., Jin, J., Geng, Z., 2021. Novel deep learning based on data fusion integrating correlation analysis for soft sensor modeling. Ind. Eng. Chem. Res. 60 (27), 10001–10010. http://dx.doi.org/10.1021/acs.iecr.1c01131.

Wu, X., Yuan, M., Yu, H., 2009. Soft-sensor modeling of cement raw material blending process based on fuzzy neural networks with particle swarm optimization. In: 2009 International Conference on Computational Intelligence and Natural Computing. Wuhan, China, pp. 158–161. http://dx.doi.org/10.1109/CINC.2009.186.

Xie, R., Hao, K., Huang, B., Chen, L., Cai, X., 2020a. Data-driven modeling based on two-stream λ gated recurrent unit network with soft sensor application. IEEE Trans. Ind. Electron. 67 (8), 7034–7043. http://dx.doi.org/10.1109/TIE.2019.2927197.

Xie, R., Jan, N.M., Hao, K., Chen, L., Huang, B., 2020b. Supervised variational autoencoders for soft sensor modeling with missing data. IEEE Trans. Ind. Inform. 16 (4), 2820–2828. http://dx.doi.org/10.1109/TII.2019.2951622.

Yan, W., Xu, R., Wang, K., Di, T., Jiang, Z., 2020. Soft sensor modeling method based on semisupervised deep learning and its application to wastewater treatment plant. Ind. Eng. Chem. Res. 59 (10), 4589–4601. http://dx.doi.org/10.1021/acs.iecr.9b05087.

Yao, Z., Zhao, C., 2021. FIGAN: A missing industrial data imputation method customized for soft sensor application. IEEE Trans. Automat. Sci. Eng. 19 (4), 1–11. http://dx.doi.org/10.1109/TASE.2021.3132037.

Yi, L., Lu, J., Ding, J., Liu, C., Chai, T., 2020. Soft sensor modeling for fraction yield of crude oil based on ensemble deep learning. Chemometr. Intell. Lab. Syst. 204, 1–14. http://dx.doi.org/10.1016/j.chemolab.2020.104087.

Yuan, X., Ou, C., Wang, Y., Yang, C., Gui, W., 2020a. Deep quality-related feature extraction for soft sensing modeling: A deep learning approach with hybrid VW-SAE. Neurocomputing 396, 375–382. http://dx.doi.org/10.1016/j.neucom.2018.11.107.

Yuan, X., Ou, C., Wang, Y., Yang, C., Gui, W., 2020b. A novel semi-supervised pre-training strategy for deep networks and its application for quality variable prediction in industrial processes. Chem. Eng. Sci. 217, 1–12. http://dx.doi.org/10.1016/j.ces.2020.115509.

Yuan, X., Qi, S., Shardt, Y.A., Wang, Y., Yang, C., Gui, W., 2020c. Soft sensor model for dynamic processes based on multichannel convolutional neural network. Chemometr. Intell. Lab. Syst. 203, http://dx.doi.org/10.1016/j.chemolab.2020.104050.

Yuan, X., Wang, Y., Yang, C., Gui, W., 2020d. Stacked isomorphic autoencoder based soft analyzer and its application to sulfur recovery unit. Inf. Sci. 534, 72–84. http://dx.doi.org/10.1016/j.ins.2020.03.018.

Yuan, X., Zhou, J., Huang, B., Wang, Y., Yang, C., Gui, W., 2020e. Hierarchical quality-relevant feature representation for soft sensor modeling: A novel deep learning strategy. IEEE Trans. Ind. Inform. 16 (6), 3721–3730. http://dx.doi.org/10.1109/TII.2019.2938890.

Zhang, Z., Jiang, T., Zhan, C., Yang, Y., 2019a. Gaussian feature learning based on variational autoencoder for improving nonlinear process monitoring. J. Process Control 75, 136–155. http://dx.doi.org/10.1016/j.jprocont.2019.01.008.

Zhang, X., Yan, W., Shao, H., 2008. Nonlinear multivariate quality estimation and prediction based on kernel partial least squares. Ind. Eng. Chem. Res. 47 (4), 1120–1131. http://dx.doi.org/10.1021/ie070741+.

Zhang, B., Zhang, J., Han, Y., Geng, Z., 2021. Dynamic soft sensor modeling method fusing process feature information based on an improved intelligent optimization algorithm. Chemom. Intell. Lab. Syst. 217, 104415. http://dx.doi.org/10.1016/j.chemolab.2021.104415.

Zhang, Z., Zhu, J., liu, Y., Ge, Z., 2020. Industrial process modeling and fault detection with recurrent Kalman variational autoencoder. In: 2020 IEEE 9th Data Driven Control and Learning Systems Conference. DDCLS, Liuzhou, China, pp. 1370–1376. http://dx.doi.org/10.1109/DDCLS49620.2020.9275274.

Zhang, X., Zou, Y., Li, S., Xu, S., 2019b. A weighted auto regressive LSTM based approach for chemical processes modeling. Neurocomputing 367, 64–74. http://dx.doi.org/10.1016/j.neucom.2019.08.006.

Zhao, Y., Ding, B., Zhang, Y., Yang, L., Hao, X., 2021. Online cement clinker quality monitoring: A soft sensor model based on multivariate time series analysis and CNN. ISA Trans. 117, 180–195. http://dx.doi.org/10.1016/j.isatra.2021.01.058.

Zheng, J., Shen, F., Ye, L., 2021. Improved Mahalanobis distance based JITL-LSTM soft sensor for multiphase batch processes. IEEE Access 9, 72172–72182. http://dx.doi.org/10.1109/ACCESS.2021.3079184.

Zhu, X., Damarla, S., Hao, K., Huang, B., 2021a. Parallel interaction spatiotemporal constrained variational autoencoder for soft sensor modeling. IEEE Trans. Ind. Inform. 18 (8), 5190–5198. http://dx.doi.org/10.1109/TII.2021.3110197.

Zhu, X., Hao, K., Xie, R., Huang, B., 2021b. Soft sensor based on extreme gradient boosting and bidirectional converted gates long short-term memory self-attention network. Neurocomputing 434, 126–136. http://dx.doi.org/10.1016/j.neucom.2020.12.028.

Zhu, Q.-X., Hou, K.-R., Chen, Z.-S., Gao, Z.-S., Xu, Y., He, Y.-L., 2021c. Novel virtual sample generation using conditional GAN for developing soft sensor with small data. Eng. Appl. Artif. Intell. 106, 104497. http://dx.doi.org/10.1016/j.engappai.2021.104497.

Zhu, X., Liu, W., Wang, B., Wang, W., 2021d. A soft sensor model of *Pichia pastoris* cell concentration based on IBDA-RELM. Prep. Biochem. Biotechnol. 52 (6), 618–626. http://dx.doi.org/10.1080/10826068.2021.1980799.

Zhu, X., Rehman, K.U., Wang, B., Shahzad, M., 2020. Modern soft-sensing modeling methods for fermentation processes. Sensors 20 (6), 1771. http://dx.doi.org/10.3390/s20061771.