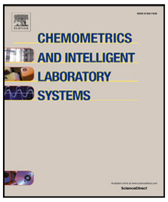
Chemometrics and Intelligent Laboratory Systems 267 (2025) 105553

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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



Advancing chemical manufacturing processes through data-driven approaches: A survey

Yellam Naidu Kottavalasa *, Lauro Snidaro

Department of Mathematics, Computer Science and Physics, University of Udine, Udine, 33100, Italy

ARTICLE INFO

Keywords:
Chemical industry
Fault detection
Process optimization
Data-driven models
Neural networks
Fusion models

ABSTRACT

The chemical industry is the backbone of global manufacturing, driving innovation across multiple sectors. Since chemical processes are complex and dynamic in nature, it is still difficult to maintain efficiency, consistency in product, and optimize process parameters. Traditional approaches often fall short in handling these complexities, prompting manufacturers to adopt data-driven methodologies, including statistical models, machine learning techniques, and deep learning architectures. This survey discusses how these models help in fault detection, process optimization, and quality control. We examine the role of statistical models in capturing process variation, machine learning models in detecting patterns and anomalies, and neural networks in predictive maintenance and real-time monitoring. Additionally, we explore fusion-based architectures, including hybrid statistical, machine learning, and deep learning methods, that facilitate better fault detection and parameter estimation. The survey also highlights how data-driven approaches support sustainable chemical manufacturing by enabling real-time decisions, adaptive control, and effective process monitoring.

Contents

---

* Corresponding author.
E-mail addresses: kottavalasa.yellamnaidu@spes.uniud.it (Y.N. Kottavalasa), lauro.snidaro@uniud.it (L. Snidaro).

https://doi.org/10.1016/j.chemolab.2025.105553
Received 1 April 2025; Received in revised form 9 October 2025; Accepted 20 October 2025

1. Introduction

The chemical industry plays a crucial role in modern society. It serves as the foundation for many essential manufacturing processes like healthcare, agriculture, electronics, and consumer goods that we rely on in our daily lives [1,2]. It includes everything from life-saving medications to sophisticated materials like plastics and polymers, relying on complicated chemical reactions and process control techniques to deliver quality and efficient products. Chemical manufacturing involves the process of converting feedstocks, including natural gas, petroleum, minerals, and plant-based materials into a variety of chemical products through intricate processes. It is a complex process that requires multiple steps and reactions in order to meet the desired specifications. Furthermore, chemical products may involve non-linear dynamics, multiple interacting variables, and restrictions associated with executing the process in real-time [3,4].

To meet increasing demands for efficiency, sustainability, and product quality, the chemical industry has implemented many advanced technologies, including Machine Learning (ML), Deep Learning (DL), and Neural Networks (NN). These new technologies allow chemical plants to sharply transition from traditional empirical methods to data-driven decision-making processes elevating operational efficiency and promoting environmental sustainability [5,6]. Among these technologies, neural networks, especially deep learning architectures, have been particularly innovative since they provide accurate forecasting of system behavior, optimize process parameters, and facilitate real-time monitoring for fault detection and diagnosis [7,8].

The integration of intelligent computational approaches has led to significant development in two major areas; Fault Detection and Diagnosis (FDD), Optimization and Estimation of Process Parameters. FDD has become a critical aspect of chemical manufacturing due to the advantages of early detection of process anomalies that avert equipment failure, reduce downtime, and enhance operational safety [9,10]. Although traditional approaches possessed some degree of success, they struggled to address the complexity and high-dimensional data from contemporary chemical plants. Predictive analytics models, however, provide a more viable solution to FDD since they can not only avoid downtime and bad outcomes, but they can also continuously monitor disturbances in real-time and enable rapid detection of problems and root analysis [11].

Similarly, the Optimization and Estimation of Process Parameters have also benefited from advanced computational-based methodologies. Accurate estimation of critical process parameters, such as temperature, pressure, and flow rates, is important for product quality assurance as well as for reducing costs and increasing productivity. These adaptive learning systems can dynamically adjust process conditions based on actual data, improving yield while maintaining sustainability standards [12,13].

A particularly significant advancement is the rise of fusion models, which integrate statistical approaches, ML, and DL techniques. They effectively utilize the complementary advantages of these techniques to predict with greater accuracy, lower false alarms, and provide actionable insights in process optimization and quality assurance. Fusion frameworks have proven especially effective in managing multi-modal datasets that combine sensor readings, spectral data, and prior process information to offer a comprehensive view of industrial operations.

In this paper, we explore the relationship between technological concepts and the optimization of manufacturing processes within the chemical industry. Section 2, explains the challenges that chemical sectors face, such as fault detection, and optimization of chemical processes. Section 3 investigates the critical function of FDD in ensuring the safety, reliability, and efficiency of chemical engineering processes. This section explores various approaches, such as statistical, machine learning, and neural network models, with particular attention to fusion models. Fused methods take advantage of multi-source data, which has been demonstrated to increase the accuracy of fault detection while providing a more comprehensive understanding of system behavior. Both hybrid statistical fusion approaches and deep learning-driven fusion frameworks are discussed, highlighting their ability to address complex, nonlinear process dynamics. Section 4 focuses on the optimization and estimation of process parameters, covering strategies to fine-tune operational conditions, improve manufacturing performance, and ensure product quality. The discussion includes advanced methodologies for process parameter estimation, real-time process control, and quality enhancement. Section 5 provides a discussion of the reviewed works, along with some thoughts on future research directions. Finally, Section 6 presents the conclusion.

Fig. 1. Key challenges in chemical industries explored in this survey.

2. Challenges faced in chemical industries

The chemical sector is one of the largest contributors to global production, sustaining numerous downstream industries. It provides essential intermediate compounds that drive industrial, agricultural, and technological development worldwide [2,14,15]. Despite the significance, the chemical manufacturing process is highly complex, involving multiple stages, precise conditions, and strict quality standards. This complexity often leads to several operational challenges that can affect productivity, efficiency, and product quality.

Key challenges include ensuring process stability, maintaining consistent product quality, minimizing production downtime, and following environmental and safety regulations. Variations in raw material properties, equipment malfunctions, and slight process anomalies can lead to significant disruptions, resulting in increased costs and compromised output. Moreover, traditional process monitoring and control systems often fall short in detecting early signs of faults, making it difficult to prevent issues before they get worse. To overcome these challenges and achieve higher operational efficiency, it is essential to integrate ML, DL, and NN models into chemical manufacturing processes [16–18]. These technologies offer real-time monitoring, predictive maintenance, and process optimization, enabling industries to proactively address issues and maintain high-quality production standards.

In this survey paper, we focus on two prominent challenges encountered in chemical industries: Fault Detection and Diagnosis (FDD), Optimization and Estimation of Process Parameters, as shown in Fig. 1. FDD involves identifying abnormal conditions and determining their root causes, allowing for timely intervention and preventing potential failures. On the other hand, process parameter optimization ensures that critical variables such as temperature, pressure, and flow rates, among others, are maintained at optimal levels to achieve consistent product quality while minimizing energy consumption and waste.

These challenges not only highlight the need for intelligent monitoring and control but also emphasize the transformative potential of advanced computational methods in chemical manufacturing. By using data-driven modeling techniques and real-time insights, chemical industries can address operational difficulties, improve product consistency, and promote sustainable manufacturing practices.

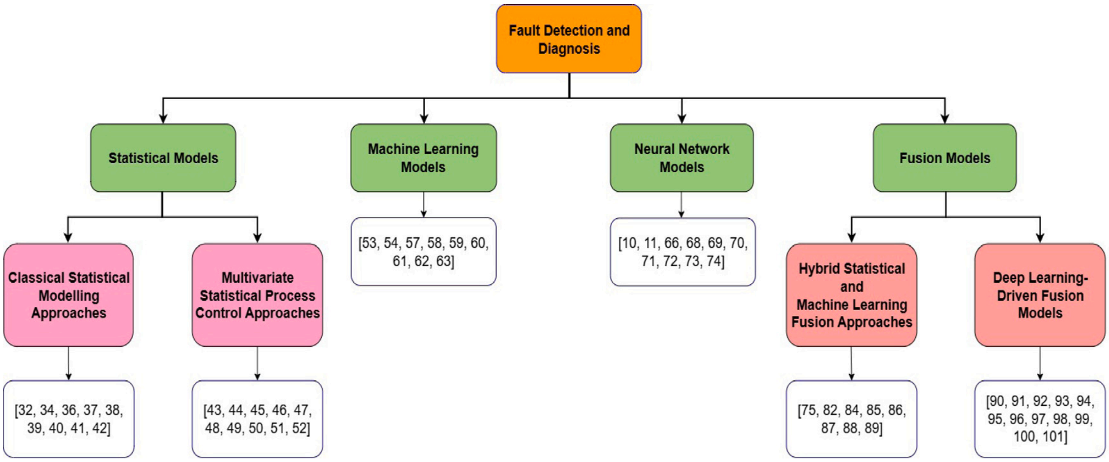


Fig. 2. The diagram presents the hierarchical classification of fault detection and diagnosis approaches discussed in this survey paper.

3. Fault detection and diagnosis

Ensuring reliable and high-quality operations in chemical engineering processes has become increasingly important. Fault detection and diagnosis are fundamental to maintaining operational efficiency. Modern chemical processes are becoming more complex, involving numerous variables and inter-dependencies, making effective fault detection essential [19]. As systems get more complicated problems like catalyst deactivation, valve blockages or compressor breakdowns happen more often, resulting in costly downtime and product quality issues [20]. Early detection and diagnosis of these problems not only improve safety and productivity but also reduces the risk of mechanical failure significantly, making plant operations smoother. Installation of a robust FDD system can result in enormous cost advantages, with potential estimates suggesting process industries could save billions of dollars annually through the prevention of costly downtime [21].

FDD is typically based on continuous monitoring of process and equipment data, gathered through sensors and other instrumentation, which serve as the primary tools for process supervision [22]. Over the years, researchers have developed a wide range of analytical approaches to identify faulty processes, ranging from traditional mathematical models to advanced probabilistic methods [23]. Once a fault is detected, further analysis helps diagnose the root cause, isolate the problem, and determine the most effective corrective actions [24,25].

As FDD research progresses, new advancements are transforming how faults are detected and diagnosed in industrial processes. These advancements are typically categorized into four main types: Statistical models, Machine Learning models, Neural Network models, and Fusion models. Traditional FDD methods have relied on Mathematical models and Expert-based knowledge [26–29], but recent trends indicate a shift toward Data-driven techniques [30,31] utilizing machine learning and deep learning algorithms. Among these, fusion methods combine statistical techniques with deep learning frameworks. They are gaining popularity for their ability to integrate multiple data sources, resulting in more accurate and reliable fault detection and diagnosis. The classification of these approaches is presented in Fig. 2, offering a clear overview of the methodologies employed in modern FDD systems.

3.1. Statistical models

Fault detection in industrial processes relies heavily on statistical models, a cornerstone of chemometrics, due to their ability to handle high-dimensional data efficiently. Several studies have explored different statistical approaches to improve fault detection accuracy and reliability.

3.1.1. Classical statistical modeling approaches

Among the earliest and most widely adopted approaches are classical statistical methods, such as PCA and PLS, which provide dimensionality reduction, noise filtering, and interpretable monitoring statistics. These methods form the basis for more advanced chemometric strategies and continue to play a vital role in fault detection.

Yin et al. [32] present various data-driven approaches for monitoring and diagnosing industrial process faults with the Tennessee Eastman (TE) process as the benchmark process. The study provides a comprehensive review of techniques such as PCA, Dynamic PCA, Independent Component Analysis (ICA), Modified ICA, Fisher Discriminant Analysis (FDA), Partial Least Squares (PLS), Total PLS, Modified PLS, and the Subspace-Aided Approach (SAP). Comparative evaluation indicates that SAP model outperforms other models with improved capabilities to handle process variations in the dynamic process as well as accurate fault detection with no strict assumption about data.

For complex industrial processes like Low Density Polyethylene (LDPE) production, anomaly detection becomes crucial, especially in critical equipment such as hypercompressors. Several methods like mathematical modeling methods [33] and time–frequency analysis methods require high sampling frequency, which poses challenges due to limited storage capacity. To overcome these drawbacks, Park et al. [34] address this issue by introducing a Weighted PCA (WPCA)-based method [35], combining PCA with Slow Feature Analysis (SFA) to improve detection in time-variant systems. This method effectively predicts emergency shutdowns by defining a Principal Component Control Limit, applied successfully to five emergency shutdown instances. Similarly, Sharmin et al. [36] propose a method for LDPE reactor monitoring using PCA and heat balance equations to identify excessive heat generation, a key indicator of decomposition. By analyzing heat balance closure errors, this technique provides early warnings, enabling operators to prevent hazardous incidents.

Beyond PCA-based approaches, alternative statistical methods have also been explored. Zhiwen Chen et al. [37] propose a Canonical Correlation Analysis (CCA) fault detection method applied to an alumina evaporation process. Unlike traditional chemometric methods such as PCA or PLS, CCA focuses on correlations between input and output data, offering more robust residual signal generation. The study also introduces Dynamic CCA (DCCA), improving fault detection in processes with time dependencies. Building on non-parametric approaches, Zheyu Jiang et al. [38] introduce a quantile-based statistical process control (SPC) framework. This method extends multivariate cumulative sum (CUSUM) charts to handle big data streams with customizable false alarm rates. Applied to the TE process dataset, this approach outperforms PCA and SVM techniques in both detection speed and false alarm reduction, demonstrating adaptability to heterogeneous industrial scenarios.

Table 1

Comparison of classical statistical models for fault detection in industrial processes.

Ref.	Methodology	Benchmark process	Fault types detected	Performance metrics	Strengths	Limitations
[32]	PCA, DPCA, ICA, MICA, FDA, PLS, TPLS, MPLS, SAP	Tennessee Eastman	Step, drift, random, sticking faults	SAP has High Fault Detection Rate (FDR): 99.88%, low FAR: 1.25	Handles dynamic changes, computational efficiency	Process-specific calibration required
[34]	WPCA, PCA, SFA	LDPE hyper-compressor	Shutdown anomalies	Detection Rate: 100% (α = 0.6), better than PCA, SFA	Time-variant adaptability	Requires high-frequency data
[36]	PCA, Heat balance equations	LDPE autoclave reactor	Decomposition faults	Reasonable lead time	Early warnings for decomposition, robust heat balance monitoring	Limited to heat-related anomalies
[37]	CCA, DCCA	Alumina evaporation	Sensor and actuator faults, process anomalies	Valve stuck fault, FDR: 90.7% (CCA), 98.7% (DCCA)	Suitable for linear static and dynamic processes	Requires steady-state conditions for DCCA
[38]	Quantile-based SPC	Tennessee Eastman	Composition faults, temperature shifts, reaction kinetics drift	Fastest detection, lowest false alarm rate (0.27%)	Effective for nonparametric, heterogeneous data	Needs robust in-control data categorization
[39]	Multiscale PCA, Kantorovich distance	Continuous stirred-tank reactor	Bias, drift, intermittent faults	FDR: 97.27% (Bias), 95% (Intermittent), 86.09% (Drift)	Robust multiscale representation, Handles noisy data	High computational cost
[41]	FFT, GLR	Continuous stirred-tank reactor	Coolant temperature, feed imbalance	GLR threshold = 99% confidence	Notable for disturbance decoupling	Needs physical models, baseline calibration
[42]	PCA, FDA, KFDA, SQP	Tennessee Eastman, Penicillin fermentation process	Nonlinear faults	KFDA has FDR > 97%, FAR < 2%, MDR < 3%	Kernel-based transformations, suitable for nonlinear systems	KFDA needs more computation

To enhance robustness in noisy environments, Ramakrishna Kini et al. [39] combine Multiscale PCA (MSPCA) with Kantorovich Distance (KD). Using wavelet-based filtering, this approach effectively removes noise while retaining critical fault-related information. Tested on distillation columns and Continuous Stirred-Tank Reactor (CSTR) setups, the KD metric improved sensitivity to drift, bias, and intermittent faults, making it superior to traditional MSPCA methods. Probabilistic models offer another layer of reliability for industrial fault detection. Yanting Zhu et al. [40] study explores probabilistic extensions of PCA, ICA, and CCA, introducing models like Probabilistic PCA and Probabilistic ICA. These techniques handle missing data, non-Gaussian distributions, and outliers, making them well-suited for dynamic, high-dimensional industrial systems. Bayesian learning and Maximum Likelihood Estimation further enhance their robustness.

Pu Du et al. [41] explore functional observation-based fault detection and isolation (FDI) methods in chemical reactors, focusing on disturbance-decoupled residual generation. Using fast Fourier transform (FFT) filtering and generalized likelihood ratio (GLR) analysis for noise suppression, this method detects faults in coolant inlet temperature and feed concentration imbalances without requiring linear approximations. Likewise, Faizan E. Mustafa et al. [42] examine a combination of statistical and optimization methods, comparing PCA, FDA, and Kernel FDA (KFDA) for fault detection. While PCA is widely used, the linear assumptions associated with it limit fault isolation, which FDA and KFDA improve upon. KFDA, in particular, excels in nonlinear fault separation. The study also integrates Sequential Quadratic Programming (SQP) to optimize fault detection thresholds, reducing false alarms and improving accuracy. Tested on TE process and Penicillin Fermentation Process datasets, these methods prove valuable in enhancing industrial fault detection and system reliability.

To further illustrate the effectiveness of the statistical models discussed in this section, Table 1 provides a comparative overview of the methodologies, benchmark processes, detected fault types, and key performance metrics. The table also highlights the strengths and limitations of each approach, offering insights into their practical applicability for complex industrial fault detection.

3.1.2. Multivariate statistical process control approaches

Multivariate Statistical Process Control (MSPC) has long been a central methodology in chemometrics, offering interpretable and statistically rigorous tools for monitoring high-dimensional industrial data. Unlike univariate control charts, MSPC exploits correlations among process variables to detect, diagnose, and prevent abnormal events. Over the years, numerous extensions have been developed to address challenges such as nonlinearity, dynamics, incipient and intermittent faults, and batch variability. The following contributions illustrate how chemometric advances have expanded MSPC from its classical PCA, PLS foundations into a versatile framework for modern chemical manufacturing.

Villalba et al. [43] illustrated the practical implementation of PCA-based MSPC through a MATLAB graphical interface validated on simulated distillation columns and real LDPE processes. This benchmark tool reinforced the two-phase MSPC structure and emphasized the role of contribution plots in diagnosing abnormal variables, while also highlighting limitations such as covariance inversion under high-dimensional noise. Building on this, Fuentes-García et al. [44] focused on the diagnostic stage of PCA-MSPC, systematically comparing methods such as contribution plots, reconstruction-based contributions, and oMEDA against the proposed Univariate-Squared (U-Squared) method. Their results revealed that univariate diagnosis strategies could mitigate the “smearing effect” common in multivariate approaches, providing more reliable identification of faulty variables in chemometric monitoring frameworks. Alongside these developments, Godoy et al. [45] advanced PLS-based monitoring by introducing a decomposition of the measurement space into complementary subspaces. This structure enabled anomalies to be selectively identified while minimizing overlapping fault signatures. By integrating Hotelling T^2 and Squared Prediction Error (SPE) statistics into a combined index and decomposing variable contributions, their approach improved both detection and attribution of faults, underscoring the diagnostic capability of chemometric PLS-MSPC frameworks.

Sánchez-Fernández et al. [46] proposed a hybrid methodology that integrates time-series modeling with MSPC to explicitly account for the dynamic and nonlinear nature of industrial processes. By combining residuals from ARIMA, support vector regression, and neural networks with traditional Hotelling T^2 and SPE charts. Their approach achieved superior monitoring performance on the TE process and wastewater treatment plants, notably reducing false alarms and detection delays compared to PCA, DPCA, CVA, and Kernel PCA (KPCA). Complementing this work in the batch processing domain, Wang et al. [47] study introduced a Multivariate Functional Kernel PCA (MFKPCA), which employs functional data analysis to represent variable trajectories as smooth functions and kernel mapping to capture nonlinear correlations. Their method effectively handled irregular datasets with unequal batch lengths, variable sampling rates, and missing data. It also demonstrated improved sensitivity over Multiway PCA (MPCA) and MKPCA in both simulation studies and an industrial semiconductor etching process. Together, these contributions highlight the growing role of chemometric methods that incorporate temporal and functional data representations to strengthen MSPC under realistic process conditions.

Special attention has also been paid to incipient and intermittent faults, which are notoriously difficult to detect. Ji et al. [48] introduced the Canonical Variate Residual Statistics Analysis (CVRSA) method, which extracts features from canonical variate residuals to enhance sensitivity to small-magnitude disturbances. The proposed CVRSA approach was evaluated on numerical and CSTR benchmark datasets, achieving fault detection rates above 99% for weak faults that other methods failed to capture. Expanding on the challenge of incipient detection, Bo Chen et al. [49] proposed a Multivariate q-sigma (Mq-sigma) rule, which evaluates canonical variate residuals over moving windows to identify persistent small deviations as early fault signatures. When tested on the TE and multiphase flow processes, the Mq-sigma rule detected incipient variations earlier and with fewer false alarms than Canonical Variate Analysis (CVA), CV Dissimilarity Analysis (CVDA), and Generalized Canonical Correlation Analysis (GCCA), reinforcing its value as a chemometric tool for dynamic fault monitoring. Complementarily, Chen et al. [50] developed a Moving Window-Principal Component Pursuit (MW-PCP) strategy for intermittent faults, combining sparse matrix factorization with a moving window to isolate short-lived events and suppress noise. Experimental validation across multiple case studies confirmed its superiority to PCA and MW-PCA, making it particularly relevant for early intervention in chemical plants.

Beyond detection, the interpretability of latent-variable structures has also been advanced. Borràs-Ferrís et al. [51] proposed Sequential Multi-Block PLS (SMB-PLS), which enforces a sequential block structure that separates correlated and orthogonal variations across process stages. In the context of food manufacturing, SMB-PLS improved diagnostic clarity and prevented disturbance propagation, providing evidence that multiblock approaches enrich MSPC for chemometric monitoring of complex processes. Another contribution was presented by Van Son et al. [52] tackled the often-overlooked issue of sensor delays, proposing clustering-based PLS methods (PLS-CON-LOAD and PLS-SEQ) that reduce time delay estimation errors by grouping correlated sensors. This preprocessing step improves alignment and enhances the reliability of MSPC models in real industrial settings. Taken together, these works reflect the diversity of MSPC developments in recent years. Table 2 further provides a structured comparison of their applications, strengths, and limitations.

3.2. Machine learning models

Machine learning has transformed fault detection by offering smarter, more adaptive solutions. ML models can understand complex relationships within the data, even when the patterns are nonlinear or constantly changing. This flexibility allows to quickly identify faults and adapt as the system evolves. As a result, industries are increasingly turning to machine learning not just for detection but also for faster and more accurate diagnosis.

To address non-linear data processing challenges in fault detection, Wang et al. [53] developed a novel KNN Distance Contribution method. This approach collects distance control indices from each sample and assigns variables based on their contribution to anomalies. It effectively identifies multivariate anomalies and abnormal variables while reducing smearing effects. The validity and reliability of the method were confirmed using a CSTR system, indicating its feasibility for real-time fault detection.

In addition to addressing non-linearity, Ragab et al. [54] focused on ensuring scalability and interpretability by introducing the Logical Analysis of Data (LAD) model. Unlike conventional machine learning models, LAD identifies patterns linked to underlying physical phenomena, providing enhanced transparency in fault detection. The authors evaluated LAD using two case studies: simulated data from the TE process [55] and real data from a Black Liquor Recovery Boiler (BLRB) in a pulp mill [56]. They compared this proposed approach with traditional ML models such as Decision Trees (DT), Random Forest (RF), K-Nearest Neighbors (KNN), Quadratic Discriminant Analysis (QDA), and Support Vector Machines (SVM). LAD demonstrated comparable accuracy while offering superior explanatory power, enabling operators to better understand the root causes of faults.

For more complex fault scenarios, Lamiaa et al. [57] propose an innovative FDD strategy that combines the KNN algorithm with the Fuzzy C-Means (FCM) clustering technique. The integration addresses the computational burden and storage space requirements of traditional KNN-based monitoring methods by clustering data points into centers and analyzing their distances for fault detection. The proposed FCM-KNN approach leverages the robustness of fuzzy clustering to handle uncertainties caused by noise and outliers in the training data. The method was applied to the TE process, achieving outstanding FDD performance, particularly in multimodal and non-linear industrial environments. In addition to clustering-based approaches, researchers have explored feature extraction techniques for fault diagnosis. Lv et al. [58] proposed a digital image processing (DIP)-based fault detection model for reciprocating compressors (RCs), commonly used in chemical and petroleum industries. Their method employed the Hit-or-Miss Transform (HMT) and Vertical Projection Transform to extract detailed features from pressure-volume (p-V) curve diagrams. These features were then fed into a three-layer artificial neural network (ANN), which demonstrated superior fault identification and classification compared to models like SVM, RF, and one-dimensional convolutional neural network (1DCNN).

To further enhance fault detection under noisy conditions, Grbovic et al. [59] introduced Dem-Den Boost, an improved AdaBoost-based algorithm designed to reduce detection delay while handling noisy labels. The method employs an adaptive exponential cost function that prioritizes fault transition periods, dynamically updating sample weights to improve accuracy. When tested on the TE process, Dem-Den Boost achieved a 90% true positive rate (TPR) while reducing detection delay to 13.7 samples, outperforming the baseline AdaBoost method. It showed exceptional performance results for challenging faults, such as Fault 6 (TPR: 98.99%) and Fault 19 (TPR: 71.21%), though complexity increases under extreme noise conditions. Recognizing the limitations of conventional models in handling incomplete data, Yuequn Zhang et al. [60] proposed the Decision Path Random Forest (DPRF) approach. This method introduces reliability scoring to quantify the impact of missing data on decision paths, allowing dynamic prediction adjustments. When applied to the TE process, DPRF outperformed traditional RF, Back Propagation Neural Networks (BPNN), Deep Belief Networks (DBN), and Radial Basis Function (RBF) networks, particularly under conditions of high missing rates (up to 40%). While effective, the models iterative correction process increases computational demands.

To address both feature selection and classification challenges, Hu et al. [61] developed a hybrid XGB-AVSSA-KELM framework, combining XGBoost for feature selection, followed by the Adaptive Variation Sparrow Search Algorithm (AVSSA) for optimizing the Kernel Extreme Learning Machine (KELM) hyperparameters. This integrated approach achieved a 91% fault diagnosis rate across 21 fault types in the TE process, surpassing conventional optimization methods, particularly for nonlinear, high-dimensional datasets. However, the performance of this approach is sensitive to hyperparameter tuning and requires significant computational resources.

In thermocouple networks of smelting furnaces, Diego et al. [62] proposed a fault detection and isolation system using XGBoost, RF,

Table 2
Comparison of MSPC-related methods for fault detection in chemical processes.

Ref.	Methodology	Benchmark process	Fault types detected	Strengths	Limitations
[43]	PCA-based MSPC GUI	Distillation (simulated), LDPE, Pasteurization	Simulated disturbances/PI failures, external abnormals	Benchmark/tutorial tool, contribution plots for diagnosis, teaching aid	PCA-only framework, covariance inversion issues in noisy/high-dimensional data
[44]	U-Squared	Monte Carlo, Saccharomyces cerevisiae process	Root cause localization	Avoids smearing effect, superior in D-statistic	RBC fails for D-statistic with 1 PC, multivariate methods may spread fault effects
[45]	PLS-decomposition with T^2 & SPE integration	Static process, Simulated reactor	Sensor faults, process disturbance	Separates overlapping signatures, strong diagnostic capability	Simulation-only validation, computationally intensive, depends on reliable PLS model
[46]	Time-series+MSPC	Tennessee Eastman, Wastewater treatment	Dynamic & nonlinear faults	Captures dynamics, lower FAR and detection delays	Requires careful model selection
[47]	MFKPCA (FDA + Kernel PCA)	Simulations data, semiconductor etch	Batch faults	Handles unequal length & missing data, captures nonlinear correlations	Requires basis/parameter selection, performance depends on functional model fit
[48]	CVRSA (Canonical Variate Residual Statistics)	Numerical example, CSTR	Incipient faults	High FDR (>99%), sensitive to weak signals	Requires CV feature extraction, depends on window width choice
[49]	Multivariate q-sigma rule	Tennessee Eastman, Multiphase flow	Incipient faults	Earlier detection, fewer false alarms, robust for small deviations	Needs careful window tuning, less effective for large faults
[50]	MW-PCP (sparse matrix factorization)	Numerical simulation, CSTR, Cranfield multiphase flow	Intermittent faults	Exploits sparsity, strong detectability (FDR up to 97%)	Higher complexity than PCA, sensitive to window length choice
[51]	SMB-PLS	Food manufacturing	Multiblock monitoring	Improves interpretability, separates correlated vs. orthogonal variations, prevents disturbance propagation	Assumes predefined sequential block structure, not suited for parallel/recycle flows
[52]	Clustering-based PLS TDE	Chemical plant data (95 sensors)	Delay-induced misalignment	Improves preprocessing robustness, lowers TDE error with many sensors	Preprocessing tool only, not a direct fault detection method

and SVM. Their method employed rolling time window analysis and neighboring sensor redundancy to detect both abrupt and incipient sensor failures. With XGBoost as the final classifier, the system achieved 80% accuracy, while an up-down counter stabilized predictions, reducing false alarms. This approach proved especially effective for unbalanced datasets, though the performance of the system varies under highly dynamic conditions. For real-time process monitoring, Hui Jiang et al. [63] integrated Raman spectroscopy with machine learning to monitor the yeast fermentation process. They compared two variable selection methods, Competitive Adaptive Reweighted Sampling (CARS) and Variable Combination Population Analysis (VCPA) to identify critical wavenumbers for fault diagnosis. Their study demonstrated that VCPA outperformed CARS in selecting relevant spectral features, improving prediction accuracy for key biochemical parameters. This approach highlighted the potential of combining spectral analysis with ML-based monitoring to enhance process stability and detect diagnosis in fermentation conditions at an early stage. Table 3 provides a comparative overview of the machine learning models discussed in this section.

3.3. Neural network models

Neural network architectures have emerged as key enablers for modelling nonlinear relationships and time-varying behaviours in industrial FDD. Their ability to analyze high-dimensional data and learn temporal patterns makes them ideal for fault monitoring in evolving process conditions. This section reviews neural network models for fault detection in chemical and industrial processes, highlighting their strengths, limitations, and operational implementations.

To address the limitations of traditional fault detection methods like PCA and autoencoders, Zhao et al. [10] introduced the Neural Component Analysis (NCA) model. This approach combines nonlinear feature extraction with linear orthogonal transformation to improve fault detection accuracy while reducing overfitting. NCA uses a neural encoder to extract features and a linear decoder to reconstruct the original data space. The effectiveness of NCA was evaluated using the TE process simulator dataset, where it achieved better fault detection accuracy than PCA, KPCA [64,65], and autoencoders. To further optimize the efficiency introduced by NCA, Mingxuan et al. [66] explored deep compression techniques to make fault detection models more efficient without sacrificing accuracy. They addressed the challenges of deploying large ANNs in real-world environments by applying three key compression techniques: pruning, clustering, and quantization. These techniques have the highest compression rate of 91.5% with negligible average accuracy change of −1.8%. This approach demonstrated significant potential for deploying lightweight models in resource-constrained environments.

For more efficient and adaptive deep learning, Zhang Hao et al. [11] proposed the RIC (ResNet-Isqrt-Cov) network, enhancing deep learning-based fault detection by preserving essential features during convolutional processing. The convolutional neural network (Resnet34) architecture extracts edge features, while the square root normalized covariance pool (iSQRT-COV) [67] module constructs covariance matrices for each dimension after convolution. This approach ensures that critical fault-related information is retained throughout the feature extraction process.

Extending Long Short-Term Memory (LSTM) based approaches, Xavier et al. [68] introduced a Recurrent Neural Network (RNN) architecture using LSTM units. This model effectively captured long-term patterns in chemical process data, offering robust fault detection even in complex scenarios. The LSTM-RNN model achieved a 99.62% FDR for major fault types (y1-y7) but struggled with minor faults like y3, where the FDR dropped to 37.04%. Despite this limitation, the model proved highly effective for detecting significant process anomalies. Although CNNs effectively capture spatial features, they often struggle with temporal dependencies. To overcome this, Han

Table 3
Performance evaluation of machine learning models for fault detection, emphasizing strengths, limitations, and industrial applications.

Ref.	Methodology	Dataset	Performance metrics	Strengths	Limitations
[53]	KNN distance contribution	Continuous stirred tank reactor	FDR 96%, Identified abnormal variables with reduced smearing effects	Handles non-Gaussian and nonlinear characteristics; eliminates variable smearing effects	Limited scalability for high-dimensional datasets
[54]	DT, RF, KNN, QDA, SVM, Logical Analysis of Data (LAD)	Black Liquor Recovery Boiler (BLRB), Tennessee Eastman	Accuracy: 86.96% (Gaussian TE), 83.76% (Non-Gaussian TE), 100% (BLRB); F1-score: 97.69% (TE), 100% (BLRB Normal/Abnormal)	Interpretable fault diagnosis, uncovering root causes of faults	Computationally slower compared to DT, RF
[57]	KNN with Fuzzy C-Means Clustering (FCM-KNN)	Tennessee Eastman	FDR: 99.9%; FAR: 0.625%; Precision: 99.87%; Accuracy: 99.5%	Reduces computational complexity with clusters	Performance depends on cluster count
[58]	SVM, RF, 1DCNN, hit-or-miss transform (HMT)+ANN	Reciprocating compressors	F1: 97.45%; Accuracy: 96.80%	HMT+Vertical Projection Transform for detailed feature extraction	Limited scalability to other processes; relies on clean pressure-volume data curves
[59]	Dem-Den Boost	Tennessee Eastman	True positive rate: 90.00%; Detection delay: 13.7 samples (vs. 17.52 for baseline AdaBoost)	Reduces detection delay; handles noisy labels; excellent in fault transitions	Moderate complexity in implementation; reduced performance under extreme noise levels
[60]	DPRF (Improved random forest)	Tennessee Eastman	OACC: 66.07% under 40% missing rate, Robustness Score: 3.25	Improved classification under data loss	Iterative correction needs high computation
[61]	XGB-AVSSA-KELM	Tennessee Eastman	FDR: 91%	Hybrid optimization and classification	Requires high computational resources; Sensitive to tuning
[62]	XGBoost, RF, SVM	Thermocouple networks	Accuracy: 80%	Combines redundancy and rolling time window analysis for fault isolation	Moderate performance on unbalanced datasets
[63]	VCPA-SVM	Yeast fermentation	R^2: 0.979, RMSEP: 0.108	Minimal computational complexity due to variable reduction	Specific to spectral data; limited generalization

et al. [69] developed an optimized LSTM network to enhance fault diagnosis accuracy and reduce correlations. Their approach determines the optimal number of hidden layer nodes for different fault types, utilizing a self-adaptive neural network and deep learning techniques for effective fault diagnosis. The study revealed that the optimized LSTM significantly outperforms conventional LSTM models in handling dynamic fault conditions, achieving higher detection accuracy.

Recognizing the need for enhanced fault detection under dynamic operating conditions, Shuaiyu et al. [70] developed the Adaptive CNN with Enhanced Highway LSTM (ACEL) architecture. This advanced model integrates multiscale convolutional channels with a highway LSTM to capture both short and long-term dependencies. The ACEL model employs an efficient channel attention mechanism to prioritize relevant features while suppressing noise. The authors tested on the TE process and CSTR datasets, where it achieved an F1 score of 0.98, outperforming other advanced methods like Deep convolutional neural network (DCNN), Bidirectional-LSTM, and Multiscale CNN. The models ability to generalize across different datasets highlights the potential for real-world industrial monitoring systems.

To further improve fault detection under noisy conditions, Wu et al. [71] introduced a DCNN for fault diagnosis in chemical processes. Their approach tested 12 different DCNN architectures, varying convolutional layers, pooling layers, and fully connected (FC) layers. Among them, Model-7 (Conv(64), Conv(64), Pool, Conv(128), Pool(2 × 1), FC(300), FC(21)) proved the most effective for detecting faults in the TE process. This model was specifically adapted to illustrate the trade-off between false alarm rate and fault diagnosis rate, while also reducing the occurrence of false alarms. Building on CNN based advancements, Han et al. [72] introduced an innovative framework combining Orthogonal Self-attention Convolutional Autoencoders (OSCAE) with CNNs. The OSCAE enhances fault detection by utilizing orthogonal attention mechanisms, while the CNN component improves classification accuracy. Reconstruction error metrics like SPE and Hotelling T^2 statistics are used to identify anomalies. They tested the approach on the TE process, achieving 98% fault detection accuracy and 99.95% fault identification accuracy, highlighting the benefits of integrating attention mechanisms with deep learning.

To enhance interpretability in industrial fault detection, Silin et al. [73] introduced the Dense Temporal Feature Convolutional Network (DTFCN), integrating Multi-Pattern Representation and Multi-Head Self-Attention for advanced feature extraction while ensuring transparency through Shapley Additive Explanations (SHAP). Subsequently, their eMixDTFCN framework, combining semi-supervised learning with MixMatch, demonstrated high diagnostic accuracy across multiple datasets, including the Acetylsalicylic Acid Crystallization Process (ASACP) and the TE process. This approach effectively handled both labeled and unlabeled data. Similarly, Wei et al. [74] conducted a comparative study of ANN, CNN, and LSTM network with batch normalization (LSTM-BN) for fault detection in the TE process. Their findings showed that LSTM-BN consistently outperformed ANN and CNN, achieving 70% accuracy across three different datasets, while CNN accuracy dropped significantly as dataset complexity increased. The integration of batch normalization enabled LSTM to overcome gradient vanishing and exploding issues, ensuring consistent performance across varying fault conditions, and making it well-suited for complex industrial applications.

Table 4 summarizes the different neural network methods used for fault detection, showing the datasets, performance, strengths, and limitations of each approach. It provides a clear comparison to help understand which methods work best for detecting faults in industrial processes.

3.4. Fusion models

A growing research direction focuses on fusion models that unify multiple data-driven techniques to improve transparency, accuracy, and diagnostic robustness in complex industrial systems. This section reviews prominent fusion approaches, which are categorized into

Table 4
Summary of neural network-based fault detection and diagnosis methods.

Ref.	Model	Dataset	Performance metrics	Strengths	Limitations	Train samples	Test samples
[10]	NCA	Tennessee Eastman	MDR: Best in 11/21 cases; FAR: 0%–5%	Combines nonlinear extraction with linear decoding; reduces overfitting	Limited generalizability	500 (Normal)	960 (Per fault)
[66]	Deep compression techniques	Tennessee Eastman	Compression: 91.5%, Accuracy: >94%, Accuracy drop: −1.8%	Efficient size reduction while maintaining accuracy	Computationally intensive during compression	–	–
[11]	RIC (ResNet-Isqrt-Cov)	Tennessee Eastman	Accuracy: 93.2%	Combines ResNet features with covariance pooling	High computational overhead	12,000	3000
[68]	LSTM	Tennessee Eastman	FDR (y1–y7): 99.62%, FDR (y3): 37.04%	Handles dynamic temporal behavior effectively	Struggles with minor faults like y3	5733	2458
[69]	LSTM (Optimized)	Tennessee Eastman	Single-fault accuracy: 100%(fault 6), Multi-fault error rate: 0.0126(fault 6, 8)	Adaptive to dynamic process changes	High dependency on sequence length	3840	960
[70]	ACEL (Adaptive CNN+Highway LSTM)	Continuous stirred tank reactor, Tennessee Eastman	CSTR-FDR, Precision, F1: 1.00; TE-FDR, Precision, F1: 0.98	Robust feature extraction for multiscale variations	High computation cost, few-shot benchmark bias	TE: 5M, CSTR: –	TE: 9.6M, CSTR: –
[71]	DCNN	Tennessee Eastman	FDR: 98.6% (Train), 88.2% (Test); FPR: 0.1% (Train), 0.5% (Test)	Captures spatial and temporal features	Limited scalability for large datasets	–	–
[72]	OSCAE-CNN	Tennessee Eastman	FDR = 99.01% (SPE), 98.12% (T^2), Identification: 99.95%	Handles multivariate noisy data effectively	High computational complexity	8160	2040
[73]	Dense Temporal Feature CNN (DTFCN)	Acetylsalicylic Acid Crystallization Process (ASACP), Tennessee Eastman	ASACP: mFDR = 99.7%, FDR > 99%; TE: mFDR = 97.1%, FDR > 90%	Robust feature handling with SHAP interpretability	Computationally intensive, requires labeled and unlabeled data	ASACP: 36,931, TE: 32,532	ASACP: 9233, TE: 8133
[74]	LSTM-BN	Tennessee Eastman	Dataset 1 Accuracy: >96%, Dataset 2: 90%, Dataset 3: 70%	Handles complex faults; avoids gradient issues	Drops accuracy for all faults	3990	2390

two subsections: Hybrid Statistical and Machine Learning Fusion Approaches, and Deep Learning-Driven Fusion Models. These models integrate complementary strengths to improve detection accuracy, reduce false alarms, and provide actionable insights for process optimization.

3.4.1. Hybrid statistical and machine learning fusion approaches

The integration of statistical methods with machine learning has led to significant advancements in fault detection systems. These approaches combine traditional process monitoring techniques with modern algorithms to handle nonlinearities, improve interpretability, and enhance fault diagnosis accuracy.

One of the earlier contributions in this domain was proposed by Jiang et al. [75] who introduced the Parallel PCA-KPCA model for complex chemical processes involving both linear and nonlinear relationships. This method employed Randomized Algorithms (RA) [76,77] and Genetic Algorithms (GA) [78] for variable selection, ensuring that PCA captures linear dependencies while KPCA addresses nonlinearities. They applied on both a mathematical case study and a CSTR process, the method significantly improved fault detection accuracy compared to traditional techniques such as PCA [79], KPCA [80], and Serial PCA (SPCA) [81]. The results exhibited enhanced fault detection rates, particularly in distinguishing faults affecting linear and nonlinear process components, making it a more effective solution for nonlinear process monitoring.

To further improve multi-process fault detection, Handayani et al. [82] developed a Markov-based model using MSPC. This method combined the Forward-Backward Hidden Semi-Markov Model (HSMM) [83] with PCA to monitor complex systems operating under multiple conditions. The HSMM utilized forward and backward algorithms for state estimation, while PCA transformed correlated variables into uncorrelated components, enabling more accurate fault detection without the need for process modeling.

Recognizing the limitations of single-scale techniques, Husnain Ali et al. [84] introduced a hybrid framework integrating PCA, Shannon Information Entropy, Wavelet Transformations (WT), and Signed Directed Graphs (SDG). This multiscale approach facilitated fault detection, root-cause identification, and fault propagation path diagnostics. Evaluated on the TE Process, the framework achieved high fault detection rates with low false alarms, offering an effective solution for managing modern industrial complexities. Building on this foundation, Husnain et al. [85] advanced wavelet-based fault detection by integrating Distributed Canonical Correlation Analysis (DCCA) with CUSUM and Exponentially Weighted Moving Average (EWMA) control charts. This approach effectively detected small variations in process statistics, distinguishing deterministic and stochastic features from multiscale data. Their methodology was experimentally validated on the CSTR dataset, where it achieved a 100% FDR and 99.8% precision. Low-frequency monitoring further enhanced detection sensitivity for early fault detection.

Raeisi et al. [86] proposed an optimized hybrid methodology that combined the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) for feature selection with metaheuristic based clustering algorithms. NSGA-II, along with k-means clustering, identifies optimal features while minimizing information loss. The t-distributed Stochastic Neighbor Embedding (t-SNE) further enhances fault separability through nonlinear dimensionality reduction. The approach is validated on the TE process and a four-tank water process, showcasing superior performance in clustering accuracy and fault classification compared to conventional methods. This study highlighted how preprocessing steps, such as feature selection and dimensionality reduction, enhance fault separability and improve classification accuracy.

In the realm of pharmaceutical manufacturing, Jul-Jørgensen et al. [87] introduced a data fusion framework that combined Multivariate Statistical Process Control with Raman spectroscopy and process sensors such as potential of hydrogen (pH), temperature, and turbidity. This approach involved low-level data fusion, which fused raw spectral data with process variables, and mid-level data fusion, which integrated PCA-extracted features. They experimented on two case studies: a simulated industrial-scale penicillin fermentation process and a lab-scale protein crystallization system. In both cases, the framework achieved FDR exceeding 90% with low false alarm rates (FAR), demonstrating potential for Industry 4.0 applications.

Wang et al. [88] approach by integrating Linear Discriminant Analysis (LDA) with a Deep Transfer Network (DTN) for domain adaptation in industrial chemical processes. This LDA-based DTN addresses the challenge of domain adaptation by assigning weighted contributions to feature variables based on their ability to distinguish between source and target data distributions. By embedding a weighted Maximum Mean Discrepancy into the DTNs loss function, the method ensures effective feature alignment between domains while preserving fault-related discriminative information. Using the TE process and hydrocracking processes datasets, the LDA-based DTN achieved classification accuracies of 93.54% and 96.77%, respectively, particularly excelling in unbalanced fault scenarios.

To address challenges like missing data, data redundancy, and feature interaction, Yadong et al. [89] introduced the Fault Diagnosis Method based on Multi-Model Fusion (FDMMF). This framework integrated FunkSVD matrix decomposition for data imputation, XGBoost for key feature selection, and xDeepFM to extract three types of fault interaction features: linear, implicit high-order, and explicit high-order interactions. Extensive experiments on the TE process and Fluidized Catalytic Cracker (FCC) fractionation unit datasets showed significant performance improvements. FDMMF achieved higher precision and recall compared to baseline models like CNN-LSTM-SVM and DCNN, highlighting effectiveness in complex industrial fault diagnosis. Table 5 provides a summary of the hybrid statistical and machine learning fusion approaches discussed in this section. It outlines the methodologies, datasets used, strengths, and limitations of each study, offering a clear overview of their applicability to industrial fault diagnosis.

Table 5
Hybrid statistical and machine learning fusion approaches for fault detection and diagnosis in chemical processes.

Ref.	Proposed methodology	Dataset	Strengths and Limitations
[75]	Parallel PCA-KPCA + GA	Continuous stirred tank reactor, Mathematical case study	Effectively captures both linear and nonlinear dependencies; Computationally intensive
[82]	HSMM + PCA	Industrial Process	Enhances fault detection using historical data; Fixed threshold may impact flexibility
[84]	PCA + Shannon Entropy + WT + SDG	Tennessee Eastman	High fault detection, root-cause analysis; Limited scalability for high-dimensional data
[85]	WT + DCCA + CUSUM + EWMA	Continuous stirred tank reactor	Detects small variations; Requires low-frequency monitoring setup
[86]	NSGAII + t-SNE + Clustering	Tennessee Eastman, Four-tank water process	Automated clustering, enhanced separability; High computational need
[87]	Data Fusion + MSPC	Penicillin fermentation, Protein crystallization	Integrates spectral and process sensor data; Complexity in data fusion setup
[88]	LDA + DTN	Tennessee Eastman, Hydrocracking	Feature alignment for domain adaptation; Limited generalization outside dataset
[89]	FDMMF (FunkSVD + XGBoost + xDeepFM)	Tennessee Eastman, FCC fractionation unit	Addresses missing data and feature interaction learning; Requires fine-tuning feature selection

3.4.2. Deep learning-driven fusion models

Deep learning-driven fusion models integrate advanced neural architectures to extract spatial and temporal features, enabling early fault detection and precise diagnosis. This section presents a review of significant contributions in this domain, summarizing their methodologies, strengths, and real-world applicability in Table 6.

Byeong et al. [90] proposed the Autoencoder Self-Organizing Map (AE-SOM) method for fault detection in the ethylene vinyl acetate (EVA) autoclave reactor. In this approach, an Autoencoder (AE) first extracts features from process operation data, which are then analyzed using a Self-Organizing Map (SOM) [102] to determine operation grades at each level. The best matching unit (BMU) identifies faulty variable propagations, allowing early detection and intervention before drastic shutdowns. This method further utilized the Granger causality test to trace fault propagation paths, confirming the effectiveness of the approach in fault prevention.

To address sequence relevance and feature redundancy, Deng et al. [91] advanced fault diagnosis by integrating Genetic Algorithms (GA) for feature selection with Dynamic CNN (DCNN). This hybrid approach employed GA to eliminate redundant features and determine the optimal sequence for maximizing the diagnostic performance. The methodology transformed time-domain data into two-dimensional matrices, enhancing CNNs ability to capture spatial and temporal patterns. The framework achieved an average fault diagnosis rate of 89.72% with a significant improvement of 3.3% after feature sequence optimization on the TE process dataset, outperforming conventional models like LSTM-CNN and Weighted Cascade Forest (WCForest). Further, to enhance fault diagnosis through feature fusion, Lei et al. [92] proposed a hybrid approach that combined deep learning with statistical feature fusion. Their methodology integrated time and frequency-domain statistical indicators with CNN and LSTM networks for fault detection. Information entropy was applied to identify relevant indicators, while convolutional autoencoders extracted compressed features. These were fused with raw data through a CNN-LSTM network, enabling the model to capture both temporal and spatial dependencies.

Wang et al. [93] introduced the Stacked Supervised Auto-Encoder (SSAE) methodology to overcome the limitations of traditional Stacked Auto-Encoders (SAE), which rely solely on unsupervised self-reconstruction. Unlike conventional SAEs, the SSAE employed supervised pretraining, ensuring fault-relevant feature extraction by leveraging fault labels during training. This approach progressively refined deep, discriminative features through hierarchical supervised autoencoders. The study evaluated the methodology using the TE process and a real hydrocracking industrial process, achieving classification accuracies of 91.52% and 97.62%, respectively. SSAE outperformed other models like SAE, SVM, and LDA, while the supervised pretraining facilitated faster convergence and improved initialization for deep networks.

Chen et al. [94] proposed the CNN-LSTM model (CS-IMLSTM), which integrated Squeeze-and-Excitation (SE) attention mechanisms with Improved Long Short-Term Memory (IMLSTM) networks. This model employed an extended sliding window mechanism to capture small-scale changes in fault features, enhancing the detection of evolving anomalies. The SE module prioritized fault-relevant features, while the IMLSTM gated historical and temporal information. The CS-IMLSTM achieved better performance on two datasets of the TE

Table 6
Summary of deep learning-driven fusion models.

Ref.	Proposed methodology	Dataset	Strengths and Limitations
[90]	AE-SOM	EVA production process	Identifies faulty propagations before shutdowns; Needs large labeled datasets
[91]	GA + DCNN	Tennessee Eastman	Optimized feature selection via GA; High computation for large-scale processes
[92]	CNN-LSTM + Feature Fusion	Tennessee Eastman	Enhances accuracy with statistical feature fusion; Increased complexity compared to standalone CNN/LSTM
[93]	Stacked Supervised Auto-Encoder (SSAE)	Tennessee Eastman, Hydrocracking	Extracts fault-relevant features using supervised pertaining; Low interpretability of learned features
[94]	CS-IMLSTM (CNN + SE + LSTM)	Tennessee Eastman	Balances spatial–temporal features efficiently; Requires extensive parameter tuning
[95]	CNN + Transfer Learning (TL)	Tennessee Eastman	Adapts to changing control strategies; Accuracy loss for large data shifts
[96]	Siamese Autoencoder (SAE)	Catalytic reforming process	Optimizes local and global features; High complexity
[97]	3DCNN + LSTM + Attention	Coke furnace	Effectively handles spatial–temporal dependencies; Requires extensive dataset preprocessing
[98]	TDLN-Trees (BLSTM + LSTM + ET)	Tennessee Eastman	Captures forward and backward dependencies; High model complexity
[99]	PG-STF (Graph-Based LSTM)	Tennessee Eastman	Graph-based modeling enhances interpretability; Requires domain knowledge for adjacency matrix
[100]	LSTM-LAE (LSTM + Ladder Autoencoder)	Tennessee Eastman, Continuous stirred tank reactor	Utilizes unlabeled data for fault diagnosis; Higher computational cost
[101]	MOLA (Multi-Block OLAE + Bayesian Fusion)	Tennessee Eastman	Orthogonality constraints improve feature representation; High computational cost

process, with average accuracies of 98.29% and 97.74%, surpassing traditional CNN-LSTM and transformer-based models. Building on CNN-based architectures, Souza et al. [95] extended deep learning fusion approaches by developing an FDD framework based on CNNs and Transfer Learning (TL). This approach emphasized fully convolutional architectures, eliminating pooling and fully connected layers to streamline feature extraction while maintaining performance. To ensure robust training, the authors generated new TE process datasets with realistic fault dynamics, enabling models to adapt to changing control strategies without extensive retraining. Fine-tuning experiments demonstrated high precision (79.8%) and recall (76.6%), even with limited data, establishing TL effectiveness for fault detection under dynamic conditions.

Ji et al. [96] further advanced deep learning-based fault diagnosis with the Siamese Autoencoder (SAE). This model combined Siamese neural networks with autoencoders for feature extraction and reconstruction, ensuring optimal separation between normal and fault-related features using a conditional contrastive loss function. The authors evaluated the SAE on a catalytic reforming process with a heat exchanger unit, where it detected faults at least 22 min earlier than conventional approaches, achieving a FAR of only 2.13%. Compared to PCA, AE, and LSTM-AE, the SAE exhibited high generalization and fault detection performance, underscoring the value of hybridized neural network architectures for industrial monitoring.

Addressing the challenges of long temporal sequences, Chen et al. [97] hybrid architecture that integrates 3D Convolutional Neural Networks (3DCNN), LSTM, and a Multi-Head Attention mechanism. The 3DCNN is employed to enhance the extraction of both spatial and temporal features from industrial time-series data, while the LSTM is optimized with attention to focus on critical temporal dependencies. This architecture addresses the limitations of conventional methods in handling complex correlations and long temporal sequences in industrial process data. The proposed model was tested on a coke furnace dataset, achieving a micro-average accuracy of 98.05% and a macro-average accuracy of 98.21%. It outperformed baseline models such as CNN-BiLSTM and CNN-GRU. Additionally, the integration of noise-resilient mechanisms ensured strong fault detection performance even under varying noise levels.

For multi-modal feature extraction, Lu et al. [98] introduced the Three-layer Deep Learning Network Random Trees (TDLN-trees) framework, specifically designed for chemical production processes. This hybrid model integrated Bidirectional LSTM and LSTM networks to capture forward and backward temporal dependencies, enhancing the representation of complex time-series data. The Fully Connected Neural Network (FCNN) layer further transformed temporal features into higher-level representations, while the Extra Trees (ET) algorithm employed the Gini index for optimal node splitting.

Fengzhen et al. [99] introduced the Physical Graph-Based Spatiotemporal Fusion (PG-STF) approach, leveraging Graph Convolutional Networks (GCNs) to extract spatial features from adjacency matrices constructed using process knowledge and Pearson correlation coefficients. Temporal dependencies were captured using an LSTM-based spatiotemporal fusion module, while a dual-supervision training strategy ensured stable convergence. PG-STF outperforms conventional approaches in most fault scenarios. The node masking method further enhanced model explainability, pinpointing critical fault-related variables for actionable insights.

Qiu et al. [100] proposed the LSTM-Ladder Autoencoder (LSTM-LAE), a semi-supervised fault diagnosis approach that combines LSTM-based temporal feature extraction with Ladder Autoencoders (LAE). This approach enhances fault diagnosis by utilizing both labeled and unlabeled data, reducing the need for extensive labeled datasets. A key innovation was the computation of interpretable variable importance, enabling fault localization. The approach was tested on the TE process and a CSTR dataset, achieving an FDR of 95.9%, outperforming conventional supervised methods. Advancing these developments, Ma et al. [101] proposed the MOLA framework, a multi-block orthogonal LSTM autoencoder designed for large-scale industrial fault detection. This approach employed Orthogonal LSTM Autoencoders (OLAE) to extract dynamic and non-redundant features, while the multi-block monitoring strategy segmented process variables into blocks, each monitored by a local OLAE. The framework integrated a quantile-based CUSUM method to capture subtle changes in feature distributions, complemented by an adaptive weight-based Bayesian fusion (W-BF) strategy for block-level aggregation.

4. Optimization and estimation of process parameters

Precise estimation and optimization of process parameters are critical for ensuring high-quality production in chemical manufacturing. Proper tuning of these parameters enhances product quality, minimizes material waste, reduces production costs, and supports environmentally sustainable practices [103,104]. This section explores how advanced ML techniques and neural networks have revolutionized process parameter estimation, transforming traditional approaches into more adaptive and precise methodologies.

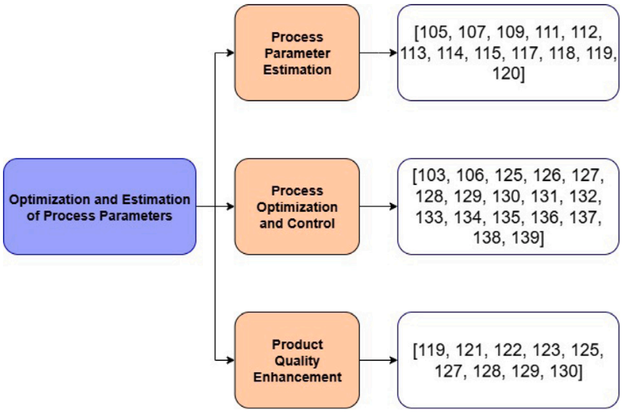


Fig. 3. The diagram shows the core categories associated with the optimization and estimation of process parameters.

Chemical manufacturing involves complex processes with multiple interacting variables, making it challenging to optimize parameters using conventional linear models. Recent advancements in deep learning and ensemble methods have addressed these challenges by capturing intricate, nonlinear relationships within the data [105,106]. These advanced models enable real-time monitoring and precise control of critical parameters, leading to consistent product quality and increased production efficiency.

Beyond optimization, ML and neural networks also improve process control, enhancing system robustness and operational stability. In this section, we review various case studies where these technologies have been successfully applied to improve chemical manufacturing processes. As illustrated in Fig. 3, the approaches are classified into three key areas: process parameter estimation, process optimization and control, and enhancement of product quality. Each category highlights the methodologies employed in the present chemical manufacturing.

4.1. Process parameter estimation

Accurate estimation of key process parameters, such as melt index (MI), ethylene index (EIX), and process conditions, is integral to next-generation chemical manufacturing. While accurate, traditional laboratory-based measurements are time-consuming, costly, and unable to provide real-time insights. To address these challenges, researchers have explored advanced modeling approaches, including machine learning, soft sensors, and hybrid frameworks. These innovations enable continuous monitoring, facilitate process optimization, and ensure consistent product quality.

One prominent approach in this domain is the Ensemble Deep Kernel Learning (EDKL) model introduced by Yi Liu et al. [107]. This model addresses the data-rich-but-information-poor (DRIP) challenge commonly encountered in polymer industries. Conventional data-driven models often struggle to make full use of large volumes of unlabeled process data effectively. The EDKL framework overcomes this by integrating Deep Belief Networks with kernel learning (KL), ensuring accurate MI prediction through multi-layer feature extraction [108]. Unlike traditional linear techniques, such as PCA, the DBN architecture employs Restricted Boltzmann machines (RBMs) to capture nonlinear relationships within the data. The kernel learning regression further refines predictions, while a bagging-based ensemble strategy enhances robustness and prevents overfitting. Furthermore, Min Jun et al. [109] proposed an LSTM-based soft sensor for MI prediction in industrial polymerization. LSTM networks, known for their ability to model temporal dependencies, were particularly well-suited to the dynamic nature of chemical processes. The authors employed the ADAM optimizer and backpropagation through time for training, while PCA preprocessing minimized multicollinearity and computational complexity. Compared to traditional approaches such as PLS, SVM, and Gaussian process regression (GPR) [110], the LSTM model consistently outperformed in both accuracy and generalization. Lei Zhang et al. [111] outlined a Spatial–Temporal Attention Temporal Convolutional Network (STA-TCN) to improve parameter prediction in dynamic chemical processes. The model integrates temporal convolution for sequence learning with an attention mechanism that draws attention to the most relevant variables across time. This structure allows it to capture both short-term fluctuations and long-range dependencies more effectively than recurrent models. This STA-TCN evidence higher accuracy and better generalization compared to LSTM and SVR, showing strong potential for real-time process monitoring and soft sensing.

Advancing this line of research, Song et al. [112] introduced a hybrid modeling approach that combined mechanistic modeling with machine learning. This method utilized a white-box submodel to predict polymerization related variables based on reaction kinetics, while a black-box submodel employed machine learning for MI estimation using both process measurements and white-box outputs. This hybrid approach ensured robust and reliable results across varying operational conditions. Wang et al. [113] formulated a just-in-time hybrid factors Gaussian process latent variable regression model (W-JGPLVR) tailored for extrusion based polymer processes. The method embeds both process variables and screw-configuration factors into a unified latent space, while the Wasserstein distance guides sample selection in the just-in-time learning framework. In this application, the model was used to predict Mean Residence Time (MRT) and Outlet Temperature (OTemp) in a twin-screw extrusion process. This structure enables the model to appropriately capture uncertainty and distributional shifts in dynamic operating environments.

Turning attention to process-specific factors, Tan et al. [114] explored how extrusion parameters affect MI estimation in polypropylene (PP) production. Their study employed multiple neural network architectures, including ANN, stacked neural networks (SNN), and DNN. Among these, the DNN model, which uses the Rectified Linear Unit (ReLU) activation function, emerged as the most effective. By overcoming vanishing gradient and overfitting issues, the model achieved precise MI predictions using complex input variables such as extruder feed rate, peroxide flow rate, motor parameters, and die pressure.

While melt index estimation remains a primary focus, researchers have also explored other critical parameters, such as the ethylene index (EIX), particularly in high-density polyethylene (HDPE) production. Accurate EIX estimation is crucial for ensuring process efficiency and product quality, given the influence on polymer density, melt flow rate, and mechanical properties. Maleki et al. [115] addressed this challenge by employing advanced ANN models, including Multi-layer Perceptron (MLP), Radial Basis Function (RBF), Cascade Feed-Forward (CF), and Generalized Regression Neural Networks (GRNN) [120]. The MLP architecture, configured as 6-3-1, was identified as the most effective, achieving the lowest mean squared error (MSE) and the highest R^2 value of 0.894. The study emphasized the importance of careful input variable selection, including temperature, pressure, and flow parameters, while utilizing the Levenberg–Marquardt optimization algorithm for efficient model training.

Jumari et al. [116] developed neural network-based soft sensors to estimate the Melt Flow Index (MFI) in polypropylene loop reactors, addressing the limitations of traditional laboratory testing. They proposed three models: ANN, a Serial Hybrid Neural Network (HNN), and SNN. The HNN model, integrated first-principle residue data, achieved the lowest RMSE of 0.010848 for overall MFI prediction. Meanwhile, the SNN model, utilizing a two-level architecture, where Level-0 comprised multiple ANN models trained on independent datasets and Level-1 refined predictions based on the outputs of the lower-level networks, outperformed grade-specific estimations. This study highlights the effectiveness of hybrid and ensemble approaches for continuous process

Table 7
Overview of process parameter estimation methods in chemical industries, focusing on modeling approaches, performance metrics, and industrial applications.

Ref.	Modeling approach	Output variables	Performance metrics	Industry application
[107]	Ensemble deep kernel learning	Melt index	RMSE: 0.015, Accuracy: 98.5%	Polymer manufacturing and quality control
[109]	LSTM-based soft sensor	Melt index	RMSE: 0.2583, R^2: 0.8314, MAPE: 2.179	Industrial polymerization monitoring
[111]	Spatial–temporal attention TCN	Quality variables (C4 concentration, Moisture content)	R^2: 0.996 (Debutanizer), R^2: 0.948 (Dryer), lowest RMSE/MAE	Debutanizer column, Tobacco cylinder dryer
[112]	Hybrid mechanistic-ML model	Melt index	R^2: 0.91, MAE: 0.008	Real-time polymer production optimization
[113]	Wasserstein just-in-time Gaussian process latent variable regression	Extrusion quality variables: MRT, OTemp	RMSE: 1.616 (MRT), 5.059 (OTemp); RPV: 2.138 (MRT), 2.240 (OTemp)	Twin-screw extrusion process (Polymer manufacturing)
[114]	DNN with ReLU activation	Melt index	Accuracy: 99.1%, No overfitting	Polypropylene extrusion process
[115]	Multi-layer perceptron	Ethylene index	R^2: 0.894, MSE: 0.02217, AARD%: 0.4213	HDPE production monitoring
[116]	Hybrid neural network	Melt flow index	RMSE: 0.010848, R^2: 99.43%	Polypropylene loop reactor quality estimation
[117]	ANN with infrared spectroscopy	Melt flow index	R^2: 0.95, MAE: 0.002	Real-time MFI monitoring in manufacturing
[118]	CNN+Transformer with feature-driven PLS	Distillate oil properties	Classification accuracy: 99.51%; high R^2, low RMSE/rRMSE for property prediction	Refinery distillate oils
[119]	Symbolic regression with Arrhenius model	Melt flow index and Shear viscosity	R^2: 0.99	Recycled polymer processing
[105]	ChemNet DNN	Cure kinetics parameters	MSE: 5.58×10^{-6}, MAE: 1×10^{-3}	Composite manufacturing for aerospace and automotive

monitoring in polymer production. Anita Rácz et al. [117] further enhanced MFI prediction by integrating infrared spectroscopy (IR) with multivariate modeling. Their approach, combining ATR-IR with PCA, Correlation and Regression Trees (CART), facilitated MFI prediction using PLS and ANN models. The ANN model, trained on transmission IR spectra, achieved an impressive R^2 value of 0.95, highlighting the potential for real-time quality control.

Yifan Wang et al. [118] worked on a CenFormer-PLS framework that fuses convolutional and transformer modules with PLS regression to enable rapid classification and property prediction of distillate oils. This approach, capturing both localized spectral patterns and broader sequential patterns, with a classification accuracy of 99.5% while maintaining interpretability through PLS latent structures. This fusion of deep learning with chemometric regression illustrates how modern architectures can accelerate calibration and improve reliability, offering a practical pathway for real-time soft sensing in petrochemical industries.

In the realm of recycled polymer processing, Lukas Seifert et al. [119] introduced predictive models for MFI and shear viscosity in polypropylene recyclates, addressing variability challenges in recycled polymer processing. Their study employed symbolic regression alongside traditional Arrhenius and Cragoe models, achieving high prediction accuracy. The authors further enhanced model robustness and overfitting through empirical noise injection and variable swapping, ensuring reliable predictions even under varying process conditions. The approach improved the quality and consistency of recycled PP, offering a reliable tool for optimizing compounding processes and supporting sustainable practices in the polymer industry.

Goli et al. [105] introduced ChemNet, a deep neural network designed to optimize cure kinetics parameters for fiber-reinforced polymer–matrix composites (FRPCs), widely used in aerospace, marine, automotive, and energy industries [121]. Leveraging frontal polymerization (FP), a rapid and energy-efficient alternative to traditional composite manufacturing, ChemNet addressed the challenge of real-time process control by predicting key parameters, such as activation energy and reaction enthalpy. The model comprises a nine-layer feed-forward neural network trained on one million examples, using front characteristics like polymerization speed and maximum temperature as inputs. This approach builds on steady-state modeling advancements, where 1-D homogenized models were validated through experiments for reliable front characteristic predictions without complex simulations [122–124]. ChemNet achieved exceptional accuracy, with a MSE of 5.58×10^{-6} and a mean absolute error (MAE) of 1×10^{-3} on unseen test data. The model performance experimented through 10-fold cross-validation, demonstrated robustness and reliability, making it a promising tool for enhancing smart manufacturing strategies in advanced composite production. A detailed summary of the discussed methodologies, along with their performance metrics and area of applications, is provided in Table 7.

4.2. Process optimization and control

Process optimization and control are indispensable in modern chemical manufacturing, especially in environments where conditions are constantly changing and reactions are highly complex. Traditional control approaches often face limitations in managing transient behavior and system uncertainties. But with advancements in ML, hybrid modeling, and chemometric frameworks have reshaped process optimization. These approaches now support real-time decision-making and adaptive control, leading to better overall process performance.

One significant breakthrough in this domain is the optimization of thermoforming process parameters for thermoplastic composites. Long Bin et al. [103] introduced a novel methodology combining Proximal Policy Optimization-Artificial Neural Network (PPO-ANN) with image analysis models to refine the thermoforming process. In composite manufacturing, parameters such as blank temperature, forming pressure, and spring-in angle critically influence the final products mechanical properties. To predict and optimize these parameters, the authors integrated finite element analysis (FEA) with ANNs, achieving precise control while reducing production time and material waste.

Davide et al. [125] further advanced the optimization of thermoforming parameters by developing a machine learning-based methodology to predict the temperature field during the different stages of the thermoforming process, including heating, transporting, forming, and cooling. They integrated FEA with ANN to accurately predict blank temperatures and optimize process parameters like blank forming temperature, mold tool temperature, and consolidation compaction ratio. Their results highlighted how ANN models outperformed traditional approaches in predicting flexural strength, derived production cycle time, and degree of crystallinity. This work marked a shift towards integrating data-driven predictive modeling in composite forming.

Furthermore, Shokry et al. [126] addressed real-time process control challenges through a Multi-Parametric Programming (MPP) approach. Traditional optimization methods often overlook uncertain parameters, leading to inefficiencies when process conditions fluctuate. To overcome this, the authors employed Gaussian process regression and Kriging models, developing surrogate models capable of predicting optimal solutions under varying conditions. This approach reduced optimization time by over 67%, ensuring consistent product quality while enhancing operational efficiency.

Another important stream of work focuses on chemometrics based latent-variable model inversion frameworks. These approaches have played a central role in linking predictive modeling with actionable process optimization and quality assurance of raw materials. MacGregor et al. [127] provided an early systematic formulation, introducing a PLS-based framework for setting simultaneous specifications on multiple raw materials. They modeled the joint effects of material properties on product quality and used Monte Carlo simulations to assess risk. The study showed how latent-variable inversion could define probabilistic specification regions, enabling more reliable supplier acceptance decisions. Building on this, Ruiz et al. [128] reframed PLS inversion as a multi-objective optimization problem within the Quality by Design (QbD) and Process Analytical Technology (PAT) paradigms, using Pareto-optimal fronts to balance correlated quality attributes and define feasible design spaces in LDPE production.

Subsequent works have extended this foundation toward Industry 4.0-oriented process optimization. Borràs-Ferrís et al. [133] developed a probabilistic inversion methodology that defined High-Confidence Design Spaces (HC-DS). By explicitly incorporating uncertainty, their approach partitions the input space into high-confidence, warning, and low-confidence regions, thereby strengthening supplier batch acceptance strategies in industrial case studies. In a related contribution, Borràs-Ferrís et al. [134] implemented the Latent Space-based Multivariate Capability Index (LSb-MCpk), which assesses supplier capability directly in the latent variable space. The method connects the HC-DS with the Raw Material Operating Space (RMOS) to provide a probabilistic evaluation of raw material lots prior to manufacturing. Arnese-Feffin et al. [135] tackled a key limitation of direct inversion methods by introducing Regularized Direct Inversion (RDI), an algebraic formulation that retains all correlated quality variables while regularizing singular matrices, ensuring robust product design under noisy conditions. Paris et al. [136] extended this line of research by introducing Multivariate Specifications with Optimization (MVSO). The approach combines SMB-PLS based specification regions with the optimization of manipulated variables. In this way, raw material lots that would normally be rejected under nominal conditions can be accepted through targeted process adjustments. Acceptance is only granted when the adjustments are economically justified, which widens the specification regions while balancing raw material cost and operational expenses.

Parallel to these inversion approaches, SMB-PLS has emerged as a powerful chemometric tool for multiblock process data, with applications extending to raw material specification. Borràs-Ferrís et al. [137] applied SMB-PLS inversion to define multivariate raw material specifications, separating correlated and uncorrelated variations to build HC-DS that preserve product quality while increasing raw material flexibility. Lauzon-Gauthier et al. [138] showed that SMB-PLS offers superior interpretability compared to Multi-block PLS and Sequential orthogonal PLS, particularly in complex datasets such as polymer film blowing and carbon anode production. Their results highlighted its value for diagnosing feedforward and feedback relationships across process stages. Finally, Stocchero [139] expanded the role of PLS for designed experiments by adapting it to orthogonal and non-orthogonal factorial designs, enabling clearer decomposition of factor effects on multivariate responses. This development bridges experimental design with chemometric regression, offering a pathway to more interpretable and optimization-ready latent-variable models in chemical manufacturing.

Wong et al. [106] introduced a Deep Optimization Prior (DOP) framework for terahertz (THz) model parameter estimation, tackling challenges like low signal-to-noise ratio (SNR) and shot noise. Instead of relying on precise initialization, their approach reparameterizes THz model parameters through a neural network, enabling accurate estimation without complex setups. Tested on real and synthetic datasets, including MetalPCB [140], SynthUSAF (a dataset with ground truth THz model parameters that are synthetically generated, and raw THz data is synthesized), and SynthObj (This dataset includes synthetically simulated ground truth parameters from a 3D object from [141]), the DOP method consistently outperformed traditional optimizers like LBFGS (Limited memory Broyden–Fletcher–Goldfarb–Shanno), AdamW, and a per-pixel autoencoder, achieving higher accuracy and lower loss.

McLean et al. [129] proposed an orthogonalization algorithm combined with a mean squared error criterion to optimize parameter selection in complex nonlinear chemical models. In large-scale chemical systems, estimating all parameters can lead to overfitting, increased computational complexity, and reduced clarity in model explanations. To address these challenges, the authors employed an MSE-based ranking method that identifies the most influential parameters while discarding less significant ones. Applied to multi-response batch reactor models, this approach ensured accurate predictions with reduced computational overhead. Moreover, they proposed a Monte Carlo-based sensitivity analysis to test the robustness of the selected parameter subsets, enhancing reliability across varying initial conditions.

Further advancing process optimization, Maffi et al. [130] proposed a hybrid modeling approach for optimizing styrene polymerization processes, combining first-principle kinetic models with neural networks. Traditional polymerization models often struggle to capture complex plant dynamics, especially during the post-reaction devolatilization stage. To address this, the authors replaced the devolatizer module with a black-box neural network, significantly enhancing both prediction accuracy and computational efficiency. Additionally, they developed a secondary neural network, trained on hybrid model outputs, enabling real-time prediction of tailor-made polymer grades. Moreover, Li et al. [131] presented the application of Digital Annealing Units (DAU) for optimizing chemical reaction conditions. Traditional methods for determining optimal reaction parameters often face computational limitations when exploring large chemical spaces. To overcome this challenge, the authors employed Quadratic Unconstrained Binary Optimization (QUBO) models, utilizing the quantum-inspired capabilities of DAU. By combining machine learning with digital annealing, they significantly accelerated the screening of billions of reaction conditions, achieving speed millions of times faster than conventional methods. This approach facilitated active learning, where the model iteratively refined predictions based on new experimental data, ensuring continuous improvement.

Hong et al. [132] tackled parameter identification and process optimization by developing an Orthogonal Forward Regression (OFR) algorithm with nested optimal regularization. This approach addressed the limitations of conventional RBF networks, which often suffer from overfitting and poor generalization. By iteratively optimizing kernel widths and regularization parameters, the authors produced sparse, high-performing models that ensured robust, real-time process control. Their method balanced model complexity and accuracy, making it well-suited for practical industrial applications. An overview of these optimization strategies, the parameters or specifications they target, and their real-world implementations is presented in Table 8.

Table 8
Summary of process optimization and control methods in chemical manufacturing, presenting key parameters and optimization strategies.

Ref.	Approach	Optimized parameters/specifications	Industry application
Machine learning and Hybrid modeling approaches			
[103]	ANN-based thermoforming optimization	Slip-path length, Shear angle	Thermoplastic composite manufacturing
[125]	ML with FEA integration	Blank forming temperature, Mould tool temperature, Consolidation compaction ratio	Thermoplastic composite manufacturing
[126]	ML-based multi-parametric programming	Optimal process set-points under uncertainty	Chemical process optimization
[106]	Deep Optimization Prior (DOP)	THz model parameters ($\hat{\varepsilon}$, μ, σ, ϕ)	THz Spectroscopy and Quality control
[129]	Orthogonalization with MSE criterion	Chemical reaction parameters (Rate constants, Activation energies)	Chemical batch reactor modeling
[130]	Hybrid NN and Kinetic model	Devolatizer stage parameters (Temperature, Pressure, Flow rate)	Styrene polymerization industry
[131]	Digital annealing units with QUBO	Catalyst, Solvent, Temperature, and Additive conditions	Chemical synthesis and Yield optimization
[132]	Orthogonal forward regression	RBF kernel width, Regularization parameters	Dynamic manufacturing environments
Chemometric and Model inversion approaches			
[127]	PLS-based model inversion	Raw material specifications via latent variables	Food manufacturing
[128]	Optimization-based PLS inversion	Pareto-optimal operating conditions (QbD/PAT)	LDPE production
[133]	Probabilistic PLS inversion	High-Confidence Design Space (HC-DS)	Cereal extraction, Petrochemicals, Film blowing
[134]	Latent space-based multivariate capability index	Supplier capability assessment via HC-DS and RMOS	Maize cereal extraction process
[135]	Regularized Direct Inversion (RDI)	Product design with correlated outputs	Fermentation, Penicillin production
[136]	MVSO framework (SMB-PLS+Optimization)	Widened specification regions with process adjustments	Grinding-flotation plant
[137]	SMB-PLS inversion	Multivariate raw material specifications, Expanded DS	Polymer extrusion film blowing process
[138]	Sequential Multi-Block PLS (SMB-PLS)	Feedforward and Feedback control relations	Polymer extrusion film blowing, Carbon anode manufacturing
[139]	PLS for Designed Experiments (DoE)	Factor-response relationships in multivariate systems	Metabolomics (wine ‘terroir’, radicchio storage), Simulated DoE

4.3. *Product quality enhancement*

Ensuring high product quality remains a primary objective in advanced manufacturing, particularly in polymer production and injection molding. With increasing process complexity, manufacturers now rely on data-driven models to monitor operations and anticipate quality deviations before they occur.

In injection molding, multiple studies have used ML algorithms to predict product quality based on process parameters and material properties. Ogorodnyk et al. [142] employed a combination of MLP and J48 decision tree [151] algorithms to predict part quality, using key parameters such as injection speed, holding pressure, screw speed, and cushion after holding pressure. Through feature selection, they reduced the dataset from 41 to 18 relevant parameters, achieving high prediction accuracy, with the MLP model reaching 99.37% and the J48 model 97.5%. The study further highlighted how real-time monitoring of process parameters can inform machine operators when adjustments are needed to prevent defects.

Similarly, Jung et al. [143] explored ML techniques in predicting the quality of injection molding products, emphasizing their impact on sustainable manufacturing. The study compared multiple ML models, including tree-based algorithms, regression methods, and autoencoders, using a dataset from an injection molding company. The results revealed that the autoencoder model outperformed other algorithms in terms of accuracy, precision, recall, and F1 score, making it highly suitable for real-time quality monitoring. Key findings indicated that factors like molding temperature, hopper temperature, injection time, and cycle time significantly influence product quality. Further advancing injection molding quality prediction, Ke et al. [144] proposed an MLP neural network model integrated with in-mold pressure sensors, focusing on real-time geometric quality assessment. Their approach extracted eleven quality indices from cavity pressure profiles, achieving an impressive prediction accuracy of over 92% for geometric widths, thereby demonstrating the feasibility of intelligent manufacturing systems for quality control.

Silva et al. [145] introduced a robust quality prediction methodology for thermoplastics injection molding using ANN, SVM, and an ensemble method. Their study illustrated how combining multiple classifiers could improve defect detection accuracy while maintaining computational efficiency. The ensemble method achieved an impressive 99% accuracy in identifying non-conforming parts, including burr and filling defects, based on key parameters such as injection time, plastification time, cycle time, cushion, and maximum injection pressure. By incorporating a window-based approach [152] that considered previous and subsequent injection cycles, the authors further enhanced model performance, reducing classification errors in transitional production phases.

Chathura et al. [146] introduced an ML predictive model for Automated Fiber Placement (AFP) unidirectional composite laminates. To address the challenge of small data availability in AFP manufacturing,

Table 9
Overview of product quality enhancement methods, highlighting optimized parameters and performance metrics.

Ref.	Methodology	Optimized parameters	Performance metrics	Industry application
[142]	MLP and J48 decision tree	Injection speed, Holding pressure, Screw speed, Cooling time	MLP accuracy: 99.37%, J48 accuracy: 97.5%	Thermoplastic injection molding
[143]	Autoencoder-based quality prediction	Molding temperature, Hopper temperature, Injection time, Cycle time	Accuracy: 98.9%, Precision: 96.4%, F1-score: 97.2%	Injection molding industry
[144]	MLP with in-mold sensors	First-stage holding pressure, Peak pressure, Residual pressure drop, Pressure integral	Accuracy: 92%	Thermoplastic injection molding
[145]	ANN, SVM, Ensemble method	Injection time, Plastification time, Cycle time, Cushion, Maximum injection pressure	Accuracy: 99%, Recall: 98.7%, F1-score: 98.8%	Thermoplastic injection molding
[146]	BPNN with virtual sample generation	HGT temperature, Consolidation force, Nip-point temperature	MSE improvements of 67.9% for $y1$, 63.8% for $y2$, 62.7% for $y3$, and 71.1% for $y4$	Thermoplastic laminates manufacturing
[147]	Inverse BPNN and VSG	HGT temperature, Nip-point temperature, Consolidation force	MSE improvements of 72.4% for $y1$, 68.5% for $y2$, 64.3% for $y3$, and 73.2% for $y4$	Thermoplastic composite manufacturing
[148]	KNN with Bayesian optimization	Chlorine concentration, Mechanical properties	Accuracy: 97%, Precision: 92%, Recall: 91%	Polymer and Chemical industry
[149]	Deep learning with solvent/nonsolvent selection	Polymer-solvent compatibility	Accuracy: 93.8%, ROC-AUC: 0.98	Polymer dissolution and recycling
[150]	gpHSP Bayesian approach	Hansen solubility parameters (δ_d, δ_p, δ_h)	δ_d: MAE: 0.68, R^2: 0.83; δ_p: MAE: 1.45, R^2: 0.67; δ_h: MAE: 1.93, R^2: 0.56	Pharmaceuticals, Organic semiconductors, Polymer blends

the authors employed Virtual Sample Generation (VSG) [153], enhancing the training dataset for more robust learning. The study used Back Propagation Neural Networks to predict critical laminate properties, including elastic modulus, inter-laminar shear strength (ILSS), maximum flexural stress, and strain. Critical processing parameters such as hot gas torch (HGT) temperature, consolidation force, nip-point temperature, and deposition rate were identified as significant contributors to laminate quality. By integrating VSG with BPNN, the model achieved high prediction accuracy with errors below 16% across all outputs, demonstrating how intelligent modeling can guide process adjustments, ensuring consistent laminate quality while minimizing production variability. Building on this research, Wanigasekara et al. [147] introduced an inverse predictive model to identify optimal AFP processing conditions for achieving desired laminate properties. Using BPNN combined with VSG, the model effectively predicted key outputs like elastic modulus, ILSS, and maximum flexural stress, while minimizing the need for trial-and-error experimentation.

Beyond injection molding, ML has been instrumental in enhancing material properties in polymer production. Zhang et al. [148] applied AI-driven approaches to evaluate the performance of chlorinated polyethylene, using the KNN and Bayesian optimization for precise characterization. Their study highlighted how AI can effectively search through vast databases of known compounds and predict performance characteristics, optimizing the balance between mechanical properties and environmental resilience.

Chandrasekaran et al. [149] introduced a deep learning paradigm for solvent selection in polymers, addressing a critical challenge in polymer dissolution and recycling. Their model, trained on a dataset of over 4500 polymers and corresponding solvents, mapped high-dimensional features into chemically relevant latent spaces. This approach achieved a classification accuracy that exceeded 93%, greatly outperforming traditional methods based on solubility parameters. Moreover, Sanchez et al. [150] proposed a Bayesian approach for predicting solubility parameters (gpHSP) using Gaussian processes. Their model incorporated molecular descriptors, COSMOtherm simulations, and quantum chemistry calculations to provide accurate predictions with uncertainty bounds. This method proved effective across various applications, including organic semiconductors and drug compounds, demonstrating the power of probabilistic machine learning in solvent selection. The Table 9 summarizes the core methodologies, optimized parameters, and their effectiveness across various industrial contexts.

5. Discussion

This review presents a comprehensive analysis of various methodologies applied to fault detection, process optimization, and parameter estimation within chemical manufacturing. Each approach offers unique advantages depending on industrial constraints such as data availability, computational resources, real-time requirements, and scalability.

For fault detection, traditional statistical models like PCA, ICA, and FDA remain popular due to their transparency and low computational demands. These models are well-suited to structured datasets with linear dependencies. Techniques such as SAP and DCCA further improve fault detection in dynamic scenarios by enhancing residual signal quality [32,37]. Recent advances in MSPC, including dynamic, kernel, and multi-block variants, have increased responsiveness to incipient and intermittent disturbances while enabling clearer diagnostic insight in complex process environments [46,47,50]. However, their effectiveness tends to decline in the presence of nonlinear relationships or nonstationary process behaviors, which are common in modern chemical production environments. To overcome these challenges, machine learning techniques such as KNN Distance Contribution, LAD, and Decision Path Random Forest have proven more effective at capturing complex, nonlinear patterns and managing issues like missing data or subtle anomalies [53,54,60]. These models are more adaptive and flexible than traditional methods, and they offer better generalization in multi-fault scenarios. However, they often require more extensive tuning and may lack interpretability, especially when applied to high-dimensional datasets.

Deep learning methods, including LSTM networks, CNN architectures, and Dense Temporal Convolutional models, further enhance fault detection by learning temporal and spatial patterns [68,70,73]. These

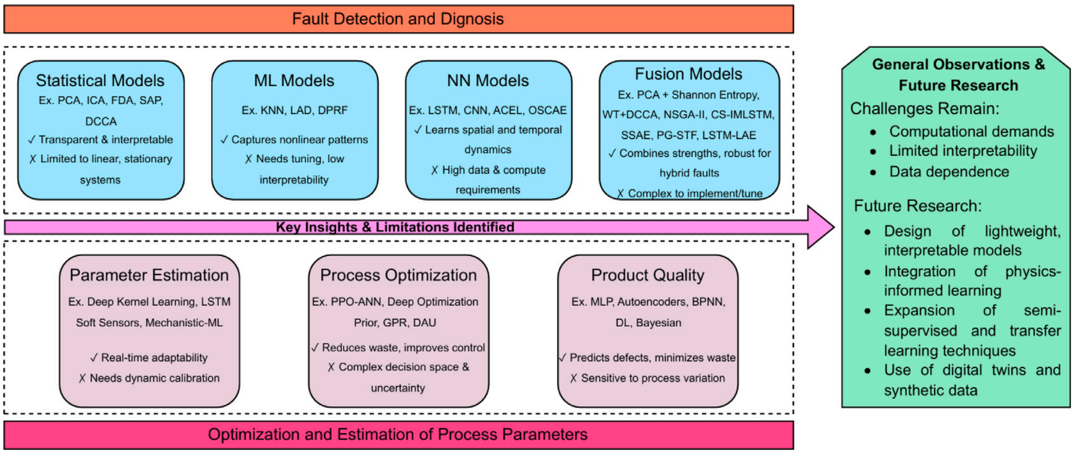


Fig. 4. Summary of fault detection and process optimization models along with key limitations and future research directions.

models are especially effective for processes with strong dynamics or multivariate interactions. Optimized LSTM variants and attention-based networks such as ACEL and OSCAE-CNN have shown excellent performance in real-world datasets [70,72]. Yet, their dependency on large labeled datasets, along with high resources and training costs, continues to pose limitations for industrial deployment in time-sensitive or resource-constrained environments.

Fusion-based approaches offer a compelling middle ground. By combining the strengths of statistical, machine learning, and deep learning methods. These models aim to enhance fault detection performance while maintaining a reasonable trade-off between explainability and resource efficiency. By integrating approaches like PCA with Shannon Information Entropy and wavelet-based DCCA, these models are capable of capturing fault characteristics across multiple dimensions and timescales [84,85]. Such hybrid frameworks are particularly useful in industrial settings where process behavior is governed by both linear and nonlinear dynamics. Models like Parallel PCA-KPCA and HSMM offer robust solutions for systems operating under varying modes, while multi-objective optimization techniques, such as NSGA-II combined with clustering, enhance fault separation and classification [75,82,86]. These methods also support feature reduction and improve fault interpretability, making them suitable for applications that require root-cause analysis and fault propagation tracking. Despite their strengths, these models may require detailed preprocessing and careful tuning of feature selection strategies, which can add complexity to implementation in real-time systems.

In parallel, deep learning-driven fusion models have shown promising results in addressing the limitations of single-model architectures. These methods combine convolutional, recurrent, and attention-based mechanisms with statistical feature extraction or domain adaptation layers, enabling both early fault detection and high classification accuracy. For instance, CS-IMLSTM and SSAE incorporate spatial and temporal data representations while preserving explainability and performance across multiple datasets [93,94]. Other models, such as PG-STF and LSTM-LAE, and MOLA, demonstrate the advantages of integrating graph-based learning, semi-supervised architectures, multi-block monitoring frameworks, especially in scenarios with limited labeled data or high-dimensional sensor streams [99–101]. These models are also capable of handling heterogeneous sensor data and can generalize across different process environments, making them attractive for Industry 4.0 applications.

In terms of parameter estimation, hybrid and ensemble learning frameworks have shown strong results in predicting variables like melt index. Models like Deep Kernel Learning and LSTM-based soft sensors have outperformed traditional regression techniques, adapting better to dynamic process variations [107,109,116]. Mechanistic-ML hybrids further enhance robustness and accuracy across diverse operational settings [112]. For process optimization, techniques like PPO-ANN and Deep Optimization Prior (DOP) refine control strategies while reducing waste and improving product properties [103,106]. Multi-parametric programming using Gaussian Process Regression and quantum-inspired models like Digital Annealing Units (DAU) also accelerate optimization under uncertainty, enabling faster and more accurate decision-making [126,131]. Chemometric model inversion approaches such as PLS and SMB-PLS based frameworks provide interpretable pathways for raw material specification and quality-by-design initiatives [127,133,136,137]. While not as flexible in handling highly nonlinear dynamics, they remain valuable for defining feasible operating spaces and ensuring reliable supplier acceptance, thereby linking optimization strategies directly to consistent product quality outcomes.

In product quality enhancement, MLP and autoencoder methods have been particularly effective in injection molding and composite manufacturing. These models accurately predict defects and optimize material properties while minimizing waste [142,143]. Methods like virtual sample generation have further improved defect classification and product consistency, reducing waste and manual inspection efforts [146]. In parallel, advanced ML models for solvent compatibility and polymer performance have supported more sustainable practices in chemical recycling and dissolution [149,150].

5.1. General observations and future directions

This review has highlighted the growing reliance on data-driven approaches for fault detection and process optimization in chemical manufacturing. Although traditional models remain useful in low-variability systems, modern manufacturing increasingly demands adaptive predictive solutions. Despite progress, several practical challenges remain:

- Computational demands: Deep learning and fusion models, though effective, often require significant resources. This makes real-time implementation difficult in edge or legacy systems.
- Limited interpretability: High-performing models lack transparency, making it difficult for plant operators to understand fault causes or take corrective action.
- Data dependence: Many advanced methods rely on clean and labeled data, which is rare in practice. Techniques like transfer learning and semi-supervised methods help, but they still face domain adaptation challenges.

Future research should focus on the following:

- Design of lightweight and interpretable models that enable deployment in constrained environments;

- Integration of physics-informed learning to embed domain constraints and improve robustness;
- Expansion of semi-supervised and transfer learning techniques to minimize the reliance on labeled data;
- Utilization of digital twins and synthetic data for training and validation when real data is limited;

These research directions, if pursued, have the potential to significantly advance the reliability, transparency, and scalability of intelligent chemical manufacturing systems. An integrated summary of models, insights, limitations, and future directions is presented in Fig. 4

6. Conclusions

This survey presents a detailed exploration of how data-driven approaches are reshaping chemical manufacturing by improving fault detection, process optimization, and parameter estimation. From statistical models and machine learning techniques to advanced neural network architectures and hybrid frameworks, the survey highlights how these methods are increasingly being adopted to address the complexity and nonlinearity of modern industrial processes. By using real-time data, these models offer greater accuracy in detecting anomalies, fine-tuning process conditions, and predicting product quality, ultimately leading to more efficient, reliable, and sustainable operations. Furthermore, the integration of fusion models and hybrid systems shows strong potential in bridging the gap between model transparency and predictive power, especially in dynamic and multi-variable environments. As industries continue transitioning towards smart manufacturing, the insights presented in this survey emphasize the importance of combining domain knowledge with adaptable, scalable, and data-efficient models to meet the evolving challenges of chemical production.

CRediT authorship contribution statement

Yellam Naidu Kottavalasa: Writing – review & editing, Writing – original draft, Resources, Investigation. **Lauro Snidaro:** Writing – review & editing, Supervision, Conceptualization.

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.

Acknowledgments

This work was funded by Versalis S.p.A. and by the European Union - NextGenerationEU, National Recovery and Resilience Plan (NRRP) M4C2 Inv. 3.3 D.M. 117/2023. The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

Data availability

No data was used for the research described in the article.

References

[1] M. AlMa'adeed, D. Ponnamma, A. El-Samak, Polymers to Improve the World and Lifestyle: Physical, Mechanical, and Chemical Needs, Elsevier, 2020, pp. 1–19, http://dx.doi.org/10.1016/B978-0-12-816808-0.00001-9.

[2] J. Clomburg, A. Crumbley, R. Gonzalez, Industrial biomanufacturing: The future of chemical production, Science 355 (2017) aag0804, http://dx.doi.org/10.1126/science.aag0804.

[3] R. Sheldon, M. Norton, Green chemistry and the plastic pollution challenge: Towards a circular economy, Green Chem. 22 (2020) 6310–6322, http://dx.doi.org/10.1039/D0GC02630A.

[4] G. Towler, R. Sinnott, Chemical Engineering Design: Principles, Practice and Economics of Plant and Process Design, Butterworth-Heinemann, 2021.

[5] M. Pirdashti, S. Curteanu, M. Kamangar, M. Hassim, M. Khatami, Artificial neural networks: Applications in chemical engineering, Rev. Chem. Eng. 24 (2013) 205–239, http://dx.doi.org/10.1515/revce-2013-0013.

[6] M.-L. Tseng, T. Tran, H. Ha, T.D. Bui, M.K. Lim, Sustainable industrial and operation engineering trends and challenges toward industry 4.0: a data driven analysis, J. Ind. Prod. Eng. 38 (2021) 1–18, http://dx.doi.org/10.1080/21681015.2021.1950227.

[7] Y.D. Boon, S. Joshi, S. Bhudolia, G. Gohel, Recent advances on the design automation for performance-optimized fiber reinforced polymer composite components, J. Compos. Sci. 4 (2020) 61, http://dx.doi.org/10.3390/jcs4020061.

[8] J. Nagy, J. Oláh, E. Erdei, D. Máté, J. Popp, The role and impact of industry 4.0 and the internet of things on the business strategy of the value chain—the case of hungary, Sustainability 10 (2018) http://dx.doi.org/10.3390/su10103491.

[9] J.B. Gomm, M. Weerasinghe, D. Williams, Diagnosis of process faults with neural networks and principal component analysis, Proc. Inst. Mech. Eng. Part E: J. Process. Mech. Eng. 214 (2000) 131–143, http://dx.doi.org/10.1243/0954408001530164.

[10] H. Zhao, Neural component analysis for fault detection, Chemometr. Intell. Lab. Syst. 176 (2017) http://dx.doi.org/10.1016/j.chemolab.2018.02.001.

[11] H. Zhang, H. Zong, J. Wang, Industrial fault diagnosis model with covariance information, in: 2023 International Conference on the Cognitive Computing and Complex Data, ICCD, 2023, pp. 231–235, http://dx.doi.org/10.1109/ICCD59681.2023.10420702.

[12] W. Sun, A. Paiva, P. Xu, A. Sundaram, R. Braatz, Fault detection and identification using bayesian recurrent neural networks, Comput. Chem. Eng. 141 (2020) 106991, http://dx.doi.org/10.1016/j.compchemeng.2020.106991.

[13] G. Berning, M. Brandenburg, K. Gürsoy, J. Kussi, V. Mehta, F.-J. Tölle, Integrating collaborative planning and supply chain optimization for the chemical process industry (i) - methodology, Comput. Chem. Eng. 28 (2004) 913–927, http://dx.doi.org/10.1016/j.compchemeng.2003.09.004.

[14] L. Zhang, C. Yan, Q. Guo, J. Zhang, J. Ruiz-Menjivar, The impact of agricultural chemical inputs on environment: Global evidence from informetrics analysis and visualization, Int. J. Low-Carbon Technol. 13 (2018) 338–352, http://dx.doi.org/10.1093/ijlct/cty039.

[15] M. Timmer, E. Dietzenbacher, B. Los, R. Stehrer, G. de Vries, An illustrated user guide to the world input–output database: The case of global automotive production, Rev. Int. Econ. 23 (2015) http://dx.doi.org/10.1111/roie.12178.

[16] J. Mohd Ali, M. Hussain, M. Tade, J. Zhang, Artificial intelligence techniques applied as estimator in chemical process systems – a literature survey, Expert Syst. Appl. 42 (2015) 5915–5931, http://dx.doi.org/10.1016/j.eswa.2015.03.023.

[17] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, H. Oza, Application of machine learning and artificial intelligence in oil and gas industry, Pet. Res. 6 (2021) http://dx.doi.org/10.1016/j.ptlrs.2021.05.009.

[18] Y.N. Kottavalasa, A. Battaglia, G. Bevilacqua, G. Marchetti, L. Snidaro, Fusion-based lstm-gru-attention model for time series forecasting of fouling factor in polymer production reactor, in: 2025 28th International Conference on Information Fusion, FUSION, 2025, pp. 1–7, http://dx.doi.org/10.23919/FUSION65864.2025.11123885.

[19] M. Ayoubi, R. Isermann, Neuro-fuzzy systems for diagnosis, Fuzzy Sets and Systems 89 (1997) 289–307, http://dx.doi.org/10.1016/S0165-0114(97)00011-0, Application of Neuro-Fuzzy Systems.

[20] K. Severson, P. Chaiwatanodom, R. Braatz, Perspective on process monitoring of industrial systems, Annu. Rev. Control. 42 (2016) http://dx.doi.org/10.1016/j.arcontrol.2016.09.001.

[21] H. Wang, T.-Y. Chai, J. Ding, M. Brown, Data driven fault diagnosis and fault tolerant control: Some advances and possible new directions, Acta Automat. Sinica 35 (2009) 739–747, http://dx.doi.org/10.1016/S1874-1029(08)60093-2.

[22] R. Isermann, Fault-Diagnosis Systems from Fault Detection to Fault Tolerance, Vol. 28, 2006, http://dx.doi.org/10.1007/3-540-30368-5.

[23] B. Goodlin, D. Boning, H. Sawin, B. Wise, Simultaneous fault detection and classification for semiconductor manufacturing tools, J. Electrochem. Soc. 150 (2002) http://dx.doi.org/10.1149/1.1623772.

[24] K. Salahshoor, M. Kordestani, M. Khoshro, Fault detection and diagnosis of an industrial steam turbine using fusion of svm (support vector machine) and anfis (adaptive neuro-fuzzy inference system) classifiers, Energy 35 (2010) 5472–5482, http://dx.doi.org/10.1016/j.energy.2010.06.001.

[25] H. Chang, D. Song, P. Kim, J. Choi, Spatiotemporal pattern modeling for fault detection and classification in semiconductor manufacturing, IEEE Trans. Semicond. Manuf. - IEEE Trans. Semiconduct. Manuf. 25 (2012) 72–82, http://dx.doi.org/10.1109/TSM.2011.2172469.

[26] N. Laouti, N. Sheibat-Othman, S. Othman, Support vector machines for fault detection in wind turbines, IFAC Proc. Vol. (IFAC-PapersOnline) 18 (2011) http://dx.doi.org/10.3182/20110828-6-IT-1002.02560.

[27] V. Venkatasubramanian, R. Rengaswamy, S. Kavuri, K. Yin, A review of process fault detection and diagnosis part iii: Process history based methods, Comput. Chem. Eng. 27 (2003) 327–346, http://dx.doi.org/10.1016/S0098-1354(02)00162-X.

[28] H. Chen, J. Cen, Z. Yang, W. Si, H. Cheng, Fault diagnosis of the dynamic chemical process based on the optimized cnn-lstm network, ACS Omega 7 (2022) http://dx.doi.org/10.1021/acsomega.2c04017.

[29] Z. Ge, Z. Song, F. Gao, Review of recent research on data-based process monitoring, Ind. Eng. Chem. Res. 52 (2013) 3543–3562, http://dx.doi.org/10.1021/ie302069q.

[30] K. Tidriri, N. Chatti, S. Verron, T. Tiplica, Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges, Annu. Rev. Control. 42 (2016) http://dx.doi.org/10.1016/j.arcontrol.2016.09.008.

[31] S.A.A. Taqvi, H. Zabiri, L.D. Tufa, F. Uddin, A. Fatima, A. Maulud, A review on data-driven learning approaches for fault detection and diagnosis in chemical processes, ChemBioEng Rev. 8 (2021) 1–22, http://dx.doi.org/10.1002/cben.202000027.

[32] S. Yin, S. Ding, A. Haghani, H. Hao, P. Zhang, A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process, J. Process Control 22 (2012) 1567–1581, http://dx.doi.org/10.1016/j.jprocont.2012.06.009.

[33] M. Farzaneh-Gord, H. Khoshnazar, Valve fault detection for single-stage reciprocating compressors, J. Nat. Gas Sci. Eng. 35 (2016) http://dx.doi.org/10.1016/j.jngse.2016.09.062.

[34] B. Park, J. Kim, J.-K. Lee, I.-B. Lee, Anomaly detection in a hyper-compressor in low-density polyethylene manufacturing processes using wpca-based principal component control limit, Korean J. Chem. Eng. 37 (2020) 11–18, http://dx.doi.org/10.1007/s11814-019-0403-y.

[35] H. Zhang, T. Liu, S. Yi, L. Gu, M. Zhou, Targeting mycn ires in mycn-amplified neuroblastoma with mir-375 inhibits tumor growth and sensitizes tumor cells to radiation, Mol. Oncol. 9 (2015) http://dx.doi.org/10.1016/j.molonc.2015.03.005.

[36] R. Sharmin, S. Shah, U. Sundararaj, A pca based fault detection scheme for an industrial high pressure polyethylene reactor, Macromol. React. Eng. 2 (2008) 12–30, http://dx.doi.org/10.1002/mren.200700023.

[37] Z. Chen, S.X. Ding, K. Zhang, Z. Li, Z. Hu, Canonical correlation analysis-based fault detection methods with application to alumina evaporation process, Control Eng. Pract. 46 (2016) 51–58, http://dx.doi.org/10.1016/j.conengprac.2015.10.006.

[38] Z. Jiang, Online monitoring and robust, reliable fault detection of chemical process systems, in: A.C. Kokossis, M.C. Georgiadis, E. Pistikopoulos (Eds.), 33rd European Symposium on Computer Aided Process Engineering, in: Computer Aided Chemical Engineering, vol. 52, Elsevier, 2023, pp. 1623–1628, http://dx.doi.org/10.1016/B978-0-443-15274-0.50258-4, URL: https://www.sciencedirect.com/science/article/pii/B9780443152740502584.

[39] K.R. Kini, M. Madakyaru, F. Harrou, A. Vatti, Y. Sun, Robust fault detection in monitoring chemical processes using multi-scale pca with kd approach, ChemEngineering 8 (2024) 45, http://dx.doi.org/10.3390/chemengineering8030045.

[40] Y. Zhu, S. Zhao, Y. Zhang, C. Zhang, J. Wu, A review of statistical-based fault detection and diagnosis with probabilistic models, Symmetry 16 (2024) 455, http://dx.doi.org/10.3390/sym16040455.

[41] P. Du, B. Wilhite, C. Kravaris, Model-based fault diagnosis for safety-critical chemical reactors: An experimental study, AIChE J. 70 (2024) http://dx.doi.org/10.1002/aic.18565.

[42] F.E. Mustafa, A.Q. Khan, A. Samee, I. Ahmed, M. Abid, M.M. Alqahtani, M. Khalid, Advanced statistical and meta-heuristic based optimization fault diagnosis techniques in complex industrial processes: A comparative analysis, IEEE Access 11 (2023) 104373–104391, http://dx.doi.org/10.1109/ACCESS.2023.3317516.

[43] P. Villalba, J. Sanchis, A. Ferrer, A graphical user interface for pca-based mspc: A benchmark software for multivariate statistical process control in matlab, Chemometr. Intell. Lab. Syst. 185 (2019) 135–152, http://dx.doi.org/10.1016/j.chemolab.2018.12.004.

[44] M. Fuentes-García, G. Maciá-Fernández, J. Camacho, Evaluation of diagnosis methods in pca-based multivariate statistical process control, Chemometr. Intell. Lab. Syst. 172 (2018) 194–210, http://dx.doi.org/10.1016/j.chemolab.2017.12.008.

[45] J.L. Godoy, J.R. Vega, J.L. Marchetti, A fault detection and diagnosis technique for multivariate processes using a pls-decomposition of the measurement space, Chemometr. Intell. Lab. Syst. 128 (2013) 25–36, http://dx.doi.org/10.1016/j.chemolab.2013.07.006.

[46] A. Sánchez-Fernández, F. Baldán, G. Sainz-Palmero, J. Benítez, M. Fuente, Fault detection based on time series modeling and multivariate statistical process control, Chemometr. Intell. Lab. Syst. 182 (2018) 57–69, http://dx.doi.org/10.1016/j.chemolab.2018.08.003.

[47] H. Wang, M. Yao, Fault detection of batch processes based on multivariate functional kernel principal component analysis, Chemometr. Intell. Lab. Syst. 149 (2015) 78–89, http://dx.doi.org/10.1016/j.chemolab.2015.09.018.

[48] H. Ji, Q. Hou, Y. Shao, Y. Zhang, Incipient fault detection for dynamic processes with canonical variate residual statistics analysis, Chemometr. Intell. Lab. Syst. 252 (2024) 105189, http://dx.doi.org/10.1016/j.chemolab.2024.105189.

[49] B. Chen, X.-L. Luo, Novel multivariate q-sigma rule focusing on process variation for incipient fault detection in dynamic processes, Chemometr. Intell. Lab. Syst. 206 (2020) 104149, http://dx.doi.org/10.1016/j.chemolab.2020.104149.

[50] J. Chen, Z. He, J. Wei, J. Wang, X. Zhou, An effective mw-pcp-based intermittent fault detection method via sparse matrix factorization and noise reduction, Chemometr. Intell. Lab. Syst. 264 (2025) 105439, http://dx.doi.org/10.1016/j.chemolab.2025.105439.

[51] J. Borràs-Ferrís, C. Duchesne, A. Ferrer, Multivariate spc via sequential multiblock-pls, Chemometr. Intell. Lab. Syst. 254 (2024) 105236, http://dx.doi.org/10.1016/j.chemolab.2024.105236.

[52] B.M. van Son, T. Offermans, C.G. Bertinetto, J.J. Jansen, Improved multivariate sensor delay estimation using a hierarchical clustering-based approach, Chemometr. Intell. Lab. Syst. 257 (2025) 105306, http://dx.doi.org/10.1016/j.chemolab.2024.105306.

[53] G. Wang, Z. Du, Y. Hu, Y. Li, Fault diagnosis of chemical processes based on k-nn distance contribution analysis method, in: 2019 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes, SAFEPROCESS, 2019, pp. 356–361, http://dx.doi.org/10.1109/SAFEPROCESS45799.2019.9213324.

[54] A. Ragab, M. El Koujok, B. Poulin, M. Amazouz, S. Yacout, Fault diagnosis in industrial chemical processes using interpretable patterns based on logical analysis of data, Expert Syst. Appl. 95 (2017) http://dx.doi.org/10.1016/j.eswa.2017.11.045.

[55] A. Ragab, M. El-koujok, M. Amazouz, S. Yacout, Fault detection and diagnosis in the tennessee eastman process using interpretable knowledge discovery, in: 2017 Annual Reliability and Maintainability Symposium, RAMS, 2017, pp. 1–7, http://dx.doi.org/10.1109/RAM.2017.7889650.

[56] H. Tran, E.K. Vakkilainnen, The kraft chemical recovery process, Tappi Kraft Pulping Short Course (2016) 1–8.

[57] L.M. Elshenawy, C. Chakour, T.A. Mahmoud, Fault detection and diagnosis strategy based on k-nearest neighbors and fuzzy c-means clustering algorithm for industrial processes, J. Franklin Inst. 359 (2022) 7115–7139, http://dx.doi.org/10.1016/j.jfranklin.2022.06.022.

[58] Q. Lv, L. Cai, X. Yu, H. Ma, Y. Li, Y. Shu, An automatic fault diagnosis method for the reciprocating compressor based on hmt and ann, Appl. Sci. 12 (2022) 5182, http://dx.doi.org/10.3390/app12105182.

[59] M. Grbovic, S. Vucetic, W. Li, P. Xu, A.K. Usadi, A boosting method for process fault detection with detection delay reduction and label denoising, in: Proceedings of the First International Workshop on Data Mining for Service and Maintenance, KDD4Service '11, Association for Computing Machinery, New York, NY, USA, 2011, pp. 7–11, http://dx.doi.org/10.1145/2018673.2018675.

[60] Y. Zhang, L. Luo, X. Ji, Y. Dai, Improved random forest algorithm based on decision paths for fault diagnosis of chemical process with incomplete data, Sensors 21 (2021) http://dx.doi.org/10.3390/s21206715.

[61] M. Hu, X. Hu, Z. Deng, B. Tu, Fault diagnosis of tennessee eastman process with xgb-avssa-kelm algorithm, Energies 15 (2022) http://dx.doi.org/10.3390/en15093198.

[62] D.A. Velandia Cárdenas, E.J.L. Pulgarín, J.I. Sofrony, Data-driven fault detection in a thermocouple network using neighboring redundancy, xgboost classifier, and up–down counter, IEEE Sens. J. 24 (2024) 5215–5223, http://dx.doi.org/10.1109/JSEN.2023.3342614.

[63] H. Jiang, W. Xu, Y. Ding, Q. Chen, Quantitative analysis of yeast fermentation process using raman spectroscopy: Comparison of cars and vcpa for variable selection, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 228 (2020) 117781, http://dx.doi.org/10.1016/j.saa.2019.117781.

[64] M. Mansouri, M. Nounou, H. Nounou, N. Karim, Kernel pca-based glrt for nonlinear fault detection of chemical processes, J. Loss Prev. Process. Ind. 40 (2016) http://dx.doi.org/10.1016/j.jlp.2016.01.011.

[65] L. Jiang, Z. Song, Z. Ge, J. Chen, Robust self-supervised model and its application for fault detection, Ind. Eng. Chem. Res. 56 (2017) http://dx.doi.org/10.1021/acs.iecr.7b00949.

[66] M. Li, Y. Shao, Deep compression of neural networks for fault detection on tennessee eastman chemical processes, in: 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering, ICBAIE, 2021, pp. 476–481, http://dx.doi.org/10.1109/ICBAIE52039.2021.9389826.

[67] P. Li, J. Xie, Q. Wang, Z. Gao, Towards faster training of global covariance pooling networks by iterative matrix square root normalization, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 947–955, http://dx.doi.org/10.1109/CVPR.2018.00105.

[68] G.M. Xavier, J.M. de Seixas, Fault detection and diagnosis in a chemical process using long short-term memory recurrent neural network, in: 2018 International Joint Conference on Neural Networks, IJCNN, 2018, pp. 1–8, http://dx.doi.org/10.1109/IJCNN.2018.8489385.

[69] Y. Han, N. Ding, Z. Geng, Z. Wang, C. Chu, An optimized long short-term memory network based fault diagnosis model for chemical processes, J. Process Control 92 (2020) 161–168, http://dx.doi.org/10.1016/j.jprocont.2020.06.005.

[70] S. Zhao, Y. Duan, N. Roy, B. Zhang, A deep learning methodology based on adaptive multiscale cnn and enhanced highway lstm for industrial process fault diagnosis, Reliab. Eng. Syst. Saf. 249 (2024) 110208, http://dx.doi.org/10.1016/j.ress.2024.110208.

[71] H. Wu, J. Zhao, Deep convolutional neural network model based chemical process fault diagnosis, Comput. Chem. Eng. 115 (2018) http://dx.doi.org/10.1016/j.compchemeng.2018.04.009.

[72] S. Han, L. Yang, D. Duan, L. Yao, K. Gao, Q. Zhang, Y. Xiao, W. Wu, J. Yang, W. Liu, X. Gao, A novel fault detection and identification method for complex chemical processes based on oscae and cnn, Process. Saf. Environ. Prot. 190 (2024) 322–334, http://dx.doi.org/10.1016/j.psep.2024.08.055.

[73] S. Rao, J. Wang, A comprehensive fault detection and diagnosis method for chemical processes, Chem. Eng. Sci. 300 (2024) 120565, http://dx.doi.org/10.1016/j.ces.2024.120565.

[74] K. Wei, H. Gu, X. Li, B. Liu, Research on the application of chemical process fault diagnosis methods based on neural network, in: Proceedings of the 2024 3rd International Conference on Cryptography, Network Security and Communication Technology, CNSCT '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 209–213, http://dx.doi.org/10.1145/3673277.3673314.

[75] Q. Jiang, X. Yan, Parallel pca–kpca for nonlinear process monitoring, Control Eng. Pract. 80 (2018) 17–25, http://dx.doi.org/10.1016/j.conengprac.2018.07.012.

[76] Z. Chen, S. Ding, T. Peng, C. Yang, W. Gui, Fault detection for non-gaussian processes using generalized canonical correlation analysis and randomized algorithms, IEEE Trans. Ind. Electron. PP (2017) 1, http://dx.doi.org/10.1109/TIE.2017.2733501.

[77] M. Krueger, T. Koenings, Y. Liu, S. Ding, J. Saijai, L. Li, Randomized algorithm based fault detection system design for uncertain lti systems, IFAC-PapersOnLine 50 (2017) 3648–3653, http://dx.doi.org/10.1016/j.ifacol.2017.08.711, 20th IFAC World Congress.

[78] Q. Jiang, S. Ding, Y. Wang, X. Yan, Data-driven distributed local fault detection for large-scale processes based on ga-regularized canonical correlation analysis, IEEE Trans. Ind. Electron. 64 (2017) 8148–8157, http://dx.doi.org/10.1109/TIE.2017.2698422.

[79] Q. Jiang, X. Yan, B. Huang, Performance-driven distributed pca process monitoring based on fault-relevant variable selection and bayesian inference, IEEE Trans. Ind. Electron. 63 (2015) 1, http://dx.doi.org/10.1109/TIE.2015.2466557.

[80] Y. Zhang, W. du, X. Li, Observation and detection for a class of industrial systems, IEEE Trans. Ind. Electron. PP (2017) 1, http://dx.doi.org/10.1109/TIE.2017.2708034.

[81] X. Deng, X. Tian, S. Chen, C. Harris, Nonlinear process fault diagnosis based on serial principal component analysis, IEEE Trans. Neural Netw. Learn. Syst. PP (2016) http://dx.doi.org/10.1109/TNNLS.2016.2635111.

[82] H. Handayani, V. Vrignat, K. Kratz, Improved fault detection method based on hsmm: Application to a chemical process, Int. J. Progn. Heal. Manag. 15 (2024) http://dx.doi.org/10.36001/ijphm.2024.v15i1.3585.

[83] L. Zhijiang, Y. Wang, Multimode continuous processes monitoring based on hidden semi-markov model and principle component analysis, Ind. Eng. Chem. Res. 56 (2017) http://dx.doi.org/10.1021/acs.iecr.7b01721.

[84] H. Ali, Z. Zhang, F. Gao, Multiscale monitoring of industrial chemical process using wavelet-entropy aided machine learning approach, Process. Saf. Environ. Prot. 180 (2023) 1053–1075, http://dx.doi.org/10.1016/j.psep.2023.10.066.

[85] H. Ali, R. Safdar, Y. Zhou, Y. Yao, L. Yao, Z. Zhang, M. Hammad Rasool, F. Gao, Robust statistical industrial fault monitoring: A machine learning-based distributed cca and low frequency control charts, Chem. Eng. Sci. 299 (2024) 120460, http://dx.doi.org/10.1016/j.ces.2024.120460.

[86] N.R. Ardali, R. Zarghami, R.S. Gharebagh, Optimized data driven fault detection and diagnosis in chemical processes, Comput. Chem. Eng. 186 (2024) 108712, http://dx.doi.org/10.1016/j.compchemeng.2024.108712.

[87] I. Jul-Jørgensen, P. Facco, K. Gernaey, M. Barolo, C. Hundahl, Data fusion of raman spectra in mspc for fault detection and diagnosis in pharmaceutical manufacturing, Comput. Chem. Eng. 184 (2024) 108647, http://dx.doi.org/10.1016/j.compchemeng.2024.108647.

[88] Y. Wang, D. Wu, X. Yuan, Lda-based deep transfer learning for fault diagnosis in industrial chemical processes, Comput. Chem. Eng. 140 (2020) 106964, http://dx.doi.org/10.1016/j.compchemeng.2020.106964.

[89] H. Yadong, Y. Zhe, W. Dong, G. Chengdong, L. Chuankun, G. Yian, A fault diagnosis method for complex chemical process based on multi-model fusion, Chem. Eng. Res. Des. 184 (2022) 662–677, http://dx.doi.org/10.1016/j.cherd.2022.06.029.

[90] P. Byeong Eon, J. Yumi, S. Ye Seul, L. Kyu-Hwang, L. Ho Kyung, Fault detection and diagnosis for eva production processes using ae-som, Korean Chem. Eng. Res. 58 (2020) 408–415, http://dx.doi.org/10.9713/kcer.2020.58.3.408.

[91] L. Deng, Y. Zhang, Y. Dai, X. Ji, L. Zhou, Y. Dang, Integrating feature optimization using a dynamic convolutional neural network for chemical process supervised fault classification, Process. Saf. Environ. Prot. 155 (2021) 473–485, http://dx.doi.org/10.1016/j.psep.2021.09.032.

[92] L. Zhang, H. Xia, B. Cha, Deep learning based fault diagnosis for chemical process with statistical feature fusion, in: 2023 42nd Chinese Control Conference, CCC, 2023, pp. 8515–8520, http://dx.doi.org/10.23919/CCC58697.2023.10240609.

[93] Y. Wang, H. Yang, X. Yuan, Y.A. Shardt, C. Yang, W. Gui, Deep learning for fault-relevant feature extraction and fault classification with stacked supervised auto-encoder, J. Process Control 92 (2020) 79–89, http://dx.doi.org/10.1016/j.jprocont.2020.05.015.

[94] H. Chen, J. Cen, Z. Yang, W. Si, H. Cheng, Fault diagnosis of the dynamic chemical process based on the optimized cnn-lstm network, ACS Omega 7 (2022) 34389–34400, http://dx.doi.org/10.1021/acsomega.2c04017.

[95] A. Souza, M. de Souza Jr., F. Silva, Enhancing fault detection and diagnosis systems for a chemical process: a study on convolutional neural networks and transfer learning, Evol. Syst. 15 (2023) 1–23, http://dx.doi.org/10.1007/s12530-023-09523-y.

[96] C. Ji, F. Ma, J. Wang, W. Sun, Fault detection for industrial chemical production using siamese autoencoder, IFAC-PapersOnLine 58 (2024) 823–828, http://dx.doi.org/10.1016/j.ifacol.2024.08.439, 12th IFAC Symposium on Advanced Control of Chemical Processes ADCHEM 2024.

[97] Y. Chen, R. Zhang, F. Gao, Fault diagnosis of industrial process using attention mechanism with 3dcnn-lstm, Chem. Eng. Sci. 293 (2024) 120059, http://dx.doi.org/10.1016/j.ces.2024.120059.

[98] M. Lu, Z. Gao, Y. Zou, Z. Chen, P. Li, Three-layer deep learning network random trees for fault detection in chemical production process, Can. J. Chem. Eng. (2024) http://dx.doi.org/10.1002/cjce.25465.

[99] F. Zhang, Q. Jin, D. Li, Y. Zhang, Q. Zhu, Physical graph-based spatiotemporal fusion approach for process fault diagnosis, ACS Omega 9 (2024) 9486–9502, http://dx.doi.org/10.1021/acsomega.3c09122.

[100] S. Zhang, T. Qiu, Semi-supervised lstm ladder autoencoder for chemical process fault diagnosis and localization, Chem. Eng. Sci. 251 (2022) 117467, http://dx.doi.org/10.1016/j.ces.2022.117467.

[101] F. Ma, C. Ji, J. Wang, W. Sun, X. Tang, Z. Jiang, Mola: Enhancing industrial process monitoring using a multi-block orthogonal long short-term memory autoencoder, Processes 12 (2024) http://dx.doi.org/10.3390/pr12122824.

[102] J. Tian, M.H. Azarian, M. Pecht, Anomaly detection using self-organizing maps-based k-nearest neighbor algorithm, in: Proceedings of the European Conference of the Prognostics and Health Management Society, Vol. 2, 2014, URL: https://doi.org/10.36001/phme.2014.v2i1.1554.

[103] L.B. Tan, N.D.P. Nhat, Prediction and optimization of process parameters for composite thermoforming using a machine learning approach, Polymers 14 (2022) http://dx.doi.org/10.3390/polym14142838.

[104] J. Gong, F. You, A new superstructure optimization paradigm for process synthesis with product distribution optimization: Application to an integrated shale gas processing and chemical manufacturing process, AIChE J. 64 (2018) 123–143, http://dx.doi.org/10.1002/aic.15882.

[105] E. Goli, S. Vyas, S. Koric, N. Sobh, P. Geubelle, Chemnet: A deep neural network for advanced composites manufacturing, J. Phys. Chem. B (2020) http://dx.doi.org/10.1021/acs.jpcb.0c03328.

[106] T.M. Wong, H. Bauermeister, M. Kahl, P.H. Bolívar, M. Möller, A. Kolb, Deep optimization prior for thz model parameter estimation, in: 2022 IEEE/CVF Winter Conference on Applications of Computer Vision, WACV, 2022, pp. 4049–4058, http://dx.doi.org/10.1109/WACV51458.2022.00410.

[107] Y. Liu, C. Yang, Z. Gao, Y. Yao, Ensemble deep kernel learning with application to quality prediction in industrial polymerization processes, Chemometr. Intell. Lab. Syst. 174 (2018) http://dx.doi.org/10.1016/j.chemolab.2018.01.008.

[108] J.A.K. Suykens, Deep restricted kernel machines using conjugate feature duality, Neural Comput. 29 (2017) 2123–2163, http://dx.doi.org/10.1162/neco_a_00984.

[109] M.J. Song, S. Kim, S.H. Oh, P.S. Jo, J.M. Lee, Soft sensor for melt index prediction based on long short-term memory network, IFAC-PapersOnLine 55 (2022) 857–862, http://dx.doi.org/10.1016/j.ifacol.2022.07.552, 13th IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems DYCOPS 2022.

[110] L. Chan, J. Chen, Melt index prediction with a mixture of gaussian process regression with embedded clustering and variable selections, J. Appl. Polym. Sci. 134 (2017) 45237, http://dx.doi.org/10.1002/app.45237.

[111] L. Zhang, G. Ren, S. Li, J. Du, D. Xu, Y. Li, A novel soft sensor approach for industrial quality prediction based tcn with spatial and temporal attention, Chemometr. Intell. Lab. Syst. 257 (2025) 105272, http://dx.doi.org/10.1016/j.chemolab.2024.105272.

[112] M.J. Song, S.H. Ju, S. Kim, S.H. Oh, J.M. Lee, Hybrid modeling approach for polymer melt index prediction, J. Appl. Polym. Sci. 139 (2022) e52987, http://dx.doi.org/10.1002/app.52987.

[113] Y. Dai, C. Yang, Z. Gu, Y. Yao, Y. Liu, Hybrid factors latent gaussian process modeling with wasserstein distance for soft sensing of extruder processes, Chemometr. Intell. Lab. Syst. 261 (2025) 105387, http://dx.doi.org/10.1016/j.chemolab.2025.105387.

[114] C. Tan, K. Yusof, S.R. Wan Alwi, Quality prediction for polypropylene extrusion based on neural networks, IOP Conf. Ser.: Mater. Sci. Eng. 1257 (2022) 012034, http://dx.doi.org/10.1088/1757-899X/1257/1/012034.

[115] A. Maleki, M.S. Shadloo, A. Rahmat, Application of artificial neural networks for producing an estimation of high-density polyethylene, Polymers 12 (2020) 2319, http://dx.doi.org/10.3390/polym12102319.

[116] N. Jumari, K. Mohd-Yusof, Comparison of product quality estimation of propylene polymerization in loop reactors using artificial neural network models, J. Teknol. 78 (2016) http://dx.doi.org/10.11113/jt.v78.9279.

[117] A. Rácz, D. Tátraaljai, S. Klébert, Determination of melt flow index and polymer additives in polyethylene based on ir spectra and multivariate modeling, Mater. Today Chem. 33 (2023) 101671, http://dx.doi.org/10.1016/j.mtchem.2023.101671.

[118] Y. Wang, X. Chen, L. Jiang, Y. Hu, An integrated framework combining cenformer and pls regression for rapid distillate oil classification and property prediction, Chemometr. Intell. Lab. Syst. 267 (2025) 105530, http://dx.doi.org/10.1016/j.chemolab.2025.105530.

[119] L. Seifert, L. Leuchtenberger-Engel, C. Hopmann, Enhancing the quality of polypropylene recyclates: Predictive modelling of the melt flow rate and shear viscosity, Polymers 16 (2024) http://dx.doi.org/10.3390/polym16162326.

[120] H. Ghritlahre, R. Prasad, Exergetic performance prediction of solar air heater using mlp, grnn and rbf models of artificial neural network technique, J. Environ. Manag. 223 (2018) 566–575, http://dx.doi.org/10.1016/j.jenvman.2018.06.033.

[121] M.I. Alam, K.M. Maraz, R.A. Khan, A review on the application of high-performance fiber-reinforced polymer composite materials, GSC Adv. Res. Rev. 10 (2022) 020–036.

[122] E. Goli, N. Parikh, M. Yourdkhani, N. Hibbard, J. Moore, N. Sottos, P. Geubelle, Frontal polymerization of unidirectional carbon-fiber-reinforced composites, Compos. Part A: Appl. Sci. Manuf. 130 (2019) 105689, http://dx.doi.org/10.1016/j.compositesa.2019.105689.

[123] S. Vyas, E. Goli, X. Zhang, P. Geubelle, Manufacturing of unidirectional glass-fiber-reinforced composites via frontal polymerization: A numerical study, Compos. Sci. Technol. 184 (2019) 107832, http://dx.doi.org/10.1016/j.compscitech.2019.107832.

[124] E. Goli, N.A. Parikh, S.K. Vyas, X. Zhang, N.R. Sottos, J.S. Moore, P.H. Geubelle, Manufacturing of carbon-and glass-fiber composites using frontal polymerization, in: International Conference on Composite Materials (22nd: 2019: Melboune, VIC.), Engineers Australia, Melbourne, VIC, 2019, pp. 1509–1517.

[125] D. Nardi, J. Sinke, Design analysis for thermoforming of thermoplastic composites: prediction and machine learning-based optimization, Compos. Part C: Open Access 5 (2021) 100126, http://dx.doi.org/10.1016/j.jcomc.2021.100126.

[126] A. Shokry, S. Medina-González, P. Baraldi, E. Zio, E. Moulines, A. Espuña, A machine learning-based methodology for multi-parametric solution of chemical processes operation optimization under uncertainty, Chem. Eng. J. 425 (2021) 131632, http://dx.doi.org/10.1016/j.cej.2021.131632.

[127] J.F. MacGregor, Z. Liu, M.-J. Bruwer, B. Polsky, G. Visscher, Setting simultaneous specifications on multiple raw materials to ensure product quality and minimize risk, Chemometr. Intell. Lab. Syst. 157 (2016) 96–103, http://dx.doi.org/10.1016/j.chemolab.2016.06.021.

[128] S. Ruiz, M. Ortiz, L. Sarabia, M. Sánchez, A computational approach to partial least squares model inversion in the framework of the process analytical technology and quality by design initiatives, Chemometr. Intell. Lab. Syst. 182 (2018) 70–78, http://dx.doi.org/10.1016/j.chemolab.2018.08.014.

[129] K. McLean, S. Wu, K. McAuley, Mean-squared-error methods for selecting optimal parameter subsets for estimation, Ind. Eng. Chem. Res. 51 (2012) 6105–6115, http://dx.doi.org/10.1021/ie202352f.

[130] J. Maffi, J. Sosa, D. Estenoz, Development of hybrid (ai/first principle) models for the optimization of styrene polymerization processes, Chem. Eng. Sci. 290 (2024) 119875, http://dx.doi.org/10.1016/j.ces.2024.119875.

[131] S.-C. Li, P.-H. Wang, J.-W. Su, W.-Y. Chiang, S.-H. Huang, Y.-C. Lin, C.-H. Ou, C.-Y. Chen, Application of the digital annealer unit in optimizing chemical reaction conditions for enhanced production yields, 2024, ArXiv abs/2407.17485.

[132] X. Hong, S. Chen, J. Gao, C.J. Harris, Nonlinear identification using orthogonal forward regression with nested optimal regularization, IEEE Trans. Cybern. 45 (2015) 2925–2936, http://dx.doi.org/10.1109/TCYB.2015.2389524.

[133] J. Borràs-Ferrís, D. Palací-López, C. Duchesne, A. Ferrer, Defining multivariate raw material specifications in industry 4.0, Chemometr. Intell. Lab. Syst. 225 (2022) 104563, http://dx.doi.org/10.1016/j.chemolab.2022.104563.

[134] J. Borràs-Ferrís, C. Duchesne, A. Ferrer, A latent space-based multivariate capability index: A new paradigm for raw material supplier selection in industry 4.0, Chemometr. Intell. Lab. Syst. 258 (2025) 105339, http://dx.doi.org/10.1016/j.chemolab.2025.105339.

[135] E. Arnese-Feffin, P. Facco, F. Bezzo, M. Barolo, Digital design of new products: accounting for output correlation via a novel algebraic formulation of the latent-variable model inversion problem, Chemometr. Intell. Lab. Syst. 227 (2022) 104610, http://dx.doi.org/10.1016/j.chemolab.2022.104610.

[136] A. Paris, C. Duchesne, Éric Poulin, Adjusting plant operating conditions to widen multivariate specification regions for incoming raw materials – an optimization framework, Chemometr. Intell. Lab. Syst. 244 (2024) 104991, http://dx.doi.org/10.1016/j.chemolab.2023.104991.

[137] J. Borràs-Ferrís, C. Duchesne, A. Ferrer, Defining multivariate raw material specifications via smb-pls, Chemometr. Intell. Lab. Syst. 240 (2023) 104912, http://dx.doi.org/10.1016/j.chemolab.2023.104912.

[138] J. Lauzon-Gauthier, P. Manolescu, C. Duchesne, The sequential multi-block pls algorithm (smb-pls): Comparison of performance and interpretability, Chemometr. Intell. Lab. Syst. 180 (2018) 72–83, http://dx.doi.org/10.1016/j.chemolab.2018.07.005.

[139] M. Stocchero, Pls for designed experiments, Chemometr. Intell. Lab. Syst. 240 (2023) 104928, http://dx.doi.org/10.1016/j.chemolab.2023.104928.

[140] T. Wong, M. Kahl, P. Haring Bolívar, A. Kolb, M. Möller, Training auto-encoder-based optimizers for terahertz image reconstruction, 2019, pp. 93–106, http://dx.doi.org/10.1007/978-3-030-33676-9_7.

[141] B. Drost, M. Ulrich, P. Bergmann, P. Härtinger, C. Steger, Introducing mvtec itodd — a dataset for 3d object recognition in industry, in: 2017 IEEE International Conference on Computer Vision Workshops, ICCVW, 2017, pp. 2200–2208, http://dx.doi.org/10.1109/ICCVW.2017.257.

[142] O. Ogorodnyk, O. Lyngstad, M. Larsen, K. Martinsen, Application of machine learning methods for prediction of parts quality in thermoplastics injection molding, 2019, pp. 237–244, http://dx.doi.org/10.1007/978-981-13-2375-1_30.

[143] H. Jung, J. Jeon, D. Choi, J.-Y. Park, Application of machine learning techniques in injection molding quality prediction: Implications on sustainable manufacturing industry, Sustainability 13 (2021) http://dx.doi.org/10.3390/su13084120.

[144] K.C. Ke, M.-S. Huang, Quality prediction for injection molding by using a multilayer perceptron neural network, Polymers 12 (2020) 1812, http://dx.doi.org/10.3390/polym12081812.

[145] B. Silva, J. Sousa, G. Alenya, Machine learning methods for quality prediction in thermoplastics injection molding, in: 2021 International Conference on Electrical, Computer and Energy Technologies, ICECET, 2021, pp. 1–6, http://dx.doi.org/10.1109/ICECET52533.2021.9698455.

[146] C. Wanigasekara, E. Oromiehie, A. Swain, G. Prusty, S. Nguang, Machine learning based predictive model for afp based unidirectional composite laminates, IEEE Trans. Ind. Inform. (2019) http://dx.doi.org/10.1109/TII.2019.2932398.

[147] C. Wanigasekara, E. Oromiehie, A. Swain, G. Prusty, S. Nguang, Machine learning-based inverse predictive model for afp based thermoplastic composites, J. Ind. Inf. Integr. (2020) http://dx.doi.org/10.1016/j.jii.2020.100197.

[148] H. Zhang, L. Zhou, Evaluation of performance of chlorinated polyethylene using wireless network and artificial intelligence technology, Wirel. Commun. Mob. Comput. 2022 (2022) 1–8.

[149] A. Chandrasekaran, C. Kim, S. Venkatram, R. Ramprasad, A deep learning solvent-selection paradigm powered by a massive solvent/nonsolvent database for polymers, Macromolecules 53 (2020) http://dx.doi.org/10.1021/acs.macromol.0c00251.

[150] B. Sanchez, L. Roch, J. Perea, S. Langner, C. Brabec, A. Aspuru-Guzik, A bayesian approach to predict solubility parameters, Adv. Theory Simul. 2 (2018) http://dx.doi.org/10.1002/adts.201800069.

[151] A. Brunello, E. Marzano, A. Montanari, G. Sciavicco, J48ss: A novel decision tree approach for the handling of sequential and time series data, Computers 8 (2019) 21, http://dx.doi.org/10.3390/computers8010021.

[152] A. Ahad, Vote-based: Ensemble approach, Sak. Univ. J. Sci. 25 (2021) http://dx.doi.org/10.16984/saufenbilder.901960.

[153] D.-C. Li, I.-H. Wen, A genetic algorithm-based virtual sample generation technique to improve small data set learning, Neurocomputing 143 (2014) 222–230, http://dx.doi.org/10.1016/j.neucom.2014.06.004.