Machine Learning for Soft Sensing and Process Estimation in Industrial Systems: Advances, Challenges, and Future Prospects

Dr. Chirasree Roy Chaudhuri [1], Dr. Sandip Kumar Lahiri[2]
Associate Professor[1], Associate Professor[2]
[1]Electronics & Telecommunication Engineering, [2]Chemical Engineering
[1]Indian Institute of Engineering Science & Technology (IIEST), Shibpur, [2]National Institute of Technology (NIT), Durgapur
Email: *chirasree.chaudhuri123@gmail.com[1], sandip.lahiri789@yahoo.co.in[2]*

ABSTRACT

Soft sensing and process estimation are critical techniques in modern industrial automation and process control systems. Conventional measurement approaches often rely on physical sensors, which can be expensive, prone to drift, or impractical for certain process variables. Machine learning (ML) has emerged as a powerful alternative for developing data-driven soft sensors that infer unmeasured or difficult-to-measure variables from readily available process data. This paper presents a comprehensive review of machine learning-based soft sensing and process estimation approaches, highlighting recent advancements, common algorithms, applications across various industries, associated challenges, and future research directions. Emphasis is placed on the advantages of ML models over traditional statistical methods, the integration of real-time monitoring with predictive control, and the potential of hybrid modeling approaches that combine first-principles knowledge with data-driven insights.

KEYWORDS: *Soft sensing, Process estimation, Machine learning, Industrial automation, Predictive modeling, Data-driven control, Process monitoring.*

INTRODUCTION

Industrial processes in chemical, manufacturing, energy, and material production sectors often involve complex dynamics, nonlinearities, and constraints that make real-time measurement of certain critical variables challenging. Traditional instrumentation may not always provide timely or accurate data due to sensor limitations, environmental interference, or prohibitive installation costs. Soft sensing, also known as virtual sensing, provides an effective solution by estimating these unmeasured process variables using models derived from available sensor data.

Machine learning, a branch of artificial intelligence, enables the construction of highly accurate soft sensors by learning complex nonlinear relationships from historical and real-time process data. Unlike classical regression or model-based approaches, ML can automatically adapt to evolving process conditions and extract meaningful patterns from large datasets. This capability is crucial for modern industrial applications where efficiency, safety, and predictive maintenance are key objectives.

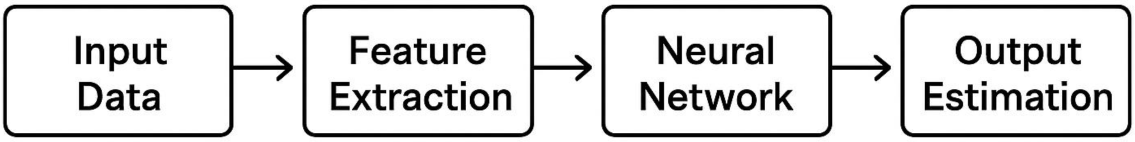

Figure 1: Soft Sensing Workflow

LITERATURE REVIEW

Soft Sensing Concepts

Soft sensors are computational models that estimate inaccessible or expensive-to-measure process variables based on available measurements. These sensors can be categorized as follows:

1. **Model-driven soft sensors**: Utilize first-principles equations of the physical process, such as mass or energy balances, combined with estimation algorithms.
2. **Data-driven soft sensors**: Depend entirely on historical and real-time process data to infer unknown variables using machine learning or statistical methods.
3. **Hybrid soft sensors**: Combine model-driven and data-driven approaches to leverage both physical knowledge and data patterns, enhancing robustness and generalization.

Table 1: Comparison of Soft Sensing Approaches

Type of Soft Sensor	Methodology	Advantages	Limitations
Model-driven	Uses physical/first-principles models	High interpretability, reliable under known conditions	Requires accurate process knowledge, complex for nonlinear processes
Data-driven	Uses historical/process data with ML algorithms	Can model nonlinearities, adaptive to process changes	Data quality dependent, may lack interpretability
Hybrid	Combines first-principles with ML	Best of both worlds, improves robustness and generalization	More complex implementation, needs expertise in both domains

Machine Learning Techniques for Soft Sensing

Machine learning algorithms have become central to soft sensing due to their flexibility and ability to model complex relationships:

1. **Artificial Neural Networks (ANNs)**: ANNs are widely used for nonlinear process estimation, capable of capturing complex interactions among multiple variables. Variants such as feedforward, recurrent, and convolutional neural networks have been applied to chemical process monitoring, power generation, and manufacturing.
2. **Support Vector Regression (SVR)**: SVR provides a robust regression framework, effective in handling high-dimensional and noisy data, making it suitable for estimating critical quality attributes in industrial processes.
3. **Random Forests (RF) and Gradient Boosting Machines (GBM)**: Ensemble learning methods such as RF and GBM combine multiple decision trees to achieve high predictive accuracy and robustness against overfitting.
4. **Gaussian Process Regression (GPR)**: GPR provides probabilistic predictions with uncertainty estimates, which is valuable for process control decisions and risk assessment.
5. **Reinforcement Learning (RL) for Process Optimization**: RL-based approaches are emerging for adaptive soft sensing, where agents learn optimal estimation policies based on feedback from process performance.

Table 2: Common Machine Learning Algorithms for Soft Sensing

Algorithm	Key Features	Strengths	Typical Applications
Artificial Neural Networks (ANNs)	Nonlinear mapping, multiple layers	Captures complex relationships, flexible	Chemical process monitoring, manufacturing
Support Vector Regression (SVR)	Margin-based regression	Handles noisy/high-dimensional data	Quality prediction, concentration estimation
Random Forests (RF)	Ensemble of decision trees	Robust, less prone to overfitting	Energy systems, environmental monitoring
Gaussian Process Regression (GPR)	Probabilistic predictions	Provides uncertainty estimates	Predictive maintenance, process control
Reinforcement Learning (RL)	Policy learning from feedback	Adaptive, online learning capability	Dynamic process optimization

APPLICATIONS OF MACHINE LEARNING-BASED SOFT SENSORS

Machine learning-based soft sensors have emerged as powerful tools for estimating process variables that are difficult or expensive to measure directly. These soft sensors leverage historical and real-time data from readily available measurements to provide accurate, timely, and actionable information for process control, optimization, and monitoring. Their applications span multiple industrial sectors:

Chemical and Petrochemical Industries

In chemical and petrochemical plants, many critical process variables—such as reactant concentration, pH, viscosity, and temperature profiles—are challenging to measure continuously using conventional sensors due to high cost, slow response, or harsh operating conditions. ML-based soft sensors can infer these variables from easily measured parameters like flow rates, pressures, and temperatures.

Key Benefits:

- Process Optimization: Accurate estimation allows operators to fine-tune reaction conditions, improving yield and reducing waste.

- Quality Control: Continuous monitoring of variables such as product composition ensures consistent product quality, minimizing deviations from specifications.
- Energy Efficiency: By estimating process conditions accurately, ML-based soft sensors enable better control over heating, cooling, and mixing operations, reducing energy consumption.
- Fault Detection: Early identification of anomalies in chemical processes prevents equipment damage and production losses.

Example: In a polymerization process, the concentration of monomers and catalysts is critical but difficult to measure in real time. ML soft sensors can accurately predict these concentrations from temperature and flow data, allowing automated adjustments to maintain optimal polymer quality.

Food and Beverage Industry

In food and beverage production, product quality depends on precise control of fermentation, moisture content, nutrient levels, and other biochemical parameters. Traditional laboratory testing is labor-intensive, time-consuming, and not feasible for real-time control. ML-based soft sensors provide a practical alternative.

Key Benefits:

- Real-time Monitoring: Variables like fermentation progress, sugar content, or acidity levels can be estimated continuously, reducing the need for manual sampling.
- Consistency and Compliance: Ensures uniform product quality across batches, meeting both regulatory standards and consumer expectations.
- Operational Efficiency: Minimizes manual testing, reduces human error, and accelerates process decision-making.

Example: In beer brewing, ML soft sensors can monitor fermentation activity, predicting alcohol concentration and acidity levels, allowing brewers to optimize timing and maintain consistent flavor profiles.

Energy Systems

Power generation and energy distribution systems rely on accurate monitoring of operational variables for efficiency and safety. Many of these parameters, such as boiler temperature, turbine efficiency, or emission levels, are challenging to measure directly under operating conditions. ML-based soft sensors provide predictive insights that enhance performance and reliability.

Key Benefits:

- Predictive Maintenance: Estimation of key operational parameters allows identification of deviations before failure, reducing downtime.
- Load Forecasting: Soft sensors can predict system behavior under variable loads, aiding in grid management and energy planning.
- Energy Optimization: Enables real-time adjustments to maximize efficiency and minimize fuel consumption.
- Emission Control: Estimation of pollutant levels like NOx or CO_2 helps maintain environmental compliance.

Example: In a coal-fired power plant, soft sensors can estimate the heat rate and combustion efficiency based on flue gas temperature and flow measurements, enabling real-time optimization of fuel usage and reducing emissions.

Pharmaceutical Manufacturing

The pharmaceutical industry requires stringent control over production variables to ensure product safety and regulatory compliance. Many critical variables, such as drug concentration, viscosity, and process temperature, are difficult to measure directly in real time. ML-based soft sensors allow accurate estimation, facilitating continuous process monitoring and quality assurance.

Key Benefits:

- Compliance: Continuous estimation of unmeasured variables ensures adherence to Good Manufacturing Practice (GMP) standards.
- Quality Assurance: Soft sensors enable real-time detection of deviations in concentration or process parameters, preventing defective batches.
- Process Control: Supports automated adjustments to maintain optimal operating conditions, improving efficiency and reducing waste.

- Reduced Laboratory Dependence: Minimizes manual testing, lowering operational costs and speeding up production cycles.

Example: In the production of injectable drugs, soft sensors can estimate active ingredient concentrations during mixing and filtration processes, ensuring that each batch meets the required potency and purity standards.

Summary

Machine learning-based soft sensors provide industries with the ability to monitor and control critical process variables in real time, even when direct measurement is impractical. By integrating these sensors, industries achieve improved product quality, energy efficiency, predictive maintenance, and compliance with regulatory standards. The flexibility of ML models allows their application across diverse domains, making them a cornerstone of modern **process automation and digitalization initiatives.**

CHALLENGES IN MACHINE LEARNING FOR SOFT SENSING

Data Quality and Availability

High-quality, representative datasets are essential for training effective ML models. Industrial environments often suffer from missing, noisy, or inconsistent data, which can significantly reduce soft sensor performance.

Model Generalization and Adaptability

ML models trained on historical process data may fail under changing operating conditions or when encountering previously unseen process states. Continuous model adaptation and online learning are critical to maintaining accuracy.

Interpretability and Trust

Many machine learning models, particularly deep learning approaches, operate as black boxes, making it difficult for operators to interpret or trust the predictions. Model explainability is essential for safety-critical applications.

Computational and Integration Constraints

Real-time estimation demands computationally efficient algorithms and seamless integration with existing control systems. Balancing model complexity and computational cost is an ongoing challenge.

SCOPE AND FUTURE DIRECTIONS

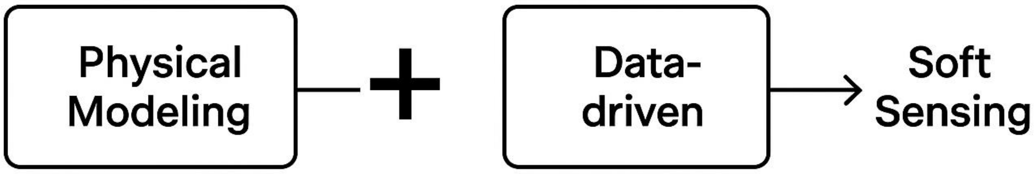

Figure 2: Hybrid Modeling for Soft Sensing

Hybrid Modeling Approaches

Combining first-principles knowledge with machine learning enhances model reliability and interpretability. Such hybrid approaches can leverage physical insights to reduce the amount of training data needed while improving prediction accuracy.

Adaptive and Online Learning Soft Sensors

Implementing adaptive ML algorithms capable of continuous learning allows soft sensors to maintain performance despite changes in process dynamics, raw material properties, or environmental conditions.

Integration with Predictive Maintenance and Digital Twins

Soft sensors integrated with predictive maintenance frameworks and digital twin simulations can enable holistic process monitoring, early fault detection, and intelligent decision-making for process optimization.

Explainable AI (XAI) in Process Estimation

Future research will likely focus on developing interpretable machine learning models, providing transparency in prediction and estimation results, which is essential for regulatory compliance and operator confidence.

CONCLUSION

Machine learning-driven soft sensing and process estimation represent a paradigm shift in industrial automation and process monitoring. By providing accurate, real-time estimates of critical but inaccessible variables, ML-based soft sensors enable enhanced process control, energy efficiency, and safety across diverse industrial sectors. Despite challenges related to data quality, model generalization, and interpretability, ongoing advancements in hybrid modeling, adaptive learning, and explainable AI are expanding the scope and reliability of these approaches. The integration of machine learning with modern industrial systems promises not only operational efficiency but also a pathway toward fully autonomous and intelligent process control environments.

REFERENCES

1. AlRijeb, M. F. S., Othman, M. L., Ishak, A., Hassan, M. K., & Albaker, B. M. (2025). *Machine Learning-Driven Soft Sensor Implementation for Real-Time Fault Detection in CDU of Oil Refinery*. Engineering, Technology & Applied Science Research, 15(1), 20425–20432. https://doi.org/10.48084/etasr.9781 ETASR Mail+1
2. Cao, L., Su, J., Conde, E., Siang, L. C., Cao, Y., & Gopaluni, B. (2025). *A novel automated soft sensor design tool for industrial applications based on machine learning*. Control Engineering Practice, 160, 106322. https://doi.org/10.1016/j.conengprac.2025.106322 ResearchGate+1
3. Grimstad, B., Løvland, K., Imsland, L. S., & Gunnerud, V. (2024). *A deep latent variable model for semi-supervised multi-unit soft sensing in industrial processes*. Applied Soft Computing, 186, 114198. https://doi.org/10.1016/j.asoc.2025.114198 ResearchGate+1
4. Farahani, H. S., Fatehi, A., Nadali, A., & Aliyari Shoorehdeli, M. (2020). *A novel method for designing transferable soft sensors and its application*. arXiv. https://arxiv.org/abs/2008.02186 arXiv+1
5. Takeuchi, S., Nishino, T., Saito, T., & Watanabe, I. (2019). *Semi-supervised approach to soft sensor modeling for fault detection in industrial systems with multiple operation modes*. arXiv. https://arxiv.org/abs/1902.09426 arXiv
6. Sánchez, L. F., Coelho, E. C., Negri, F., Gallo, F., Vallerio, M., Matos, H. A., & Manenti, F. (2025). *Machine Learning-Based Soft Sensor for Hydrogen Sulfide Monitoring in the Gas Treatment Section of an Industrial-Scale Oil Regeneration Plant*. Systems and Control Transactions, 4, 1263–1268. https://doi.org/10.69997/sct.181239 PSE Community
7. Zambonin, G., Altinier, F., Beghi, A., Coelho, L. d. S., Fiorella, N., Girotto, T., … Susto,

G. A. (2019). *Machine learning-based soft sensors for the estimation of laundry moisture content in household dryer appliances*. Energies, 12(20), 3843. https://doi.org/10.3390/en12203843 MDPI

8. Zhang, X., Zou, Y., Li, S., & Xu, S. (2022). *A data-driven semi-supervised soft-sensor method: Application on an industrial cracking furnace*. Frontiers in Chemical Engineering, 4, 899941. https://doi.org/10.3389/fceng.2022.899941 Frontiers
9. Curreri, F. (2021). *Soft Sensor Design, Transferability and Causality through Domain Adversarial Neural Networks*. (Doctoral dissertation). Retrieved from https://iris.unipa.it/handle/10447/xxx Iris
10. Curreri, F., & Patané, L. (2021). *Soft-Sensor Transferability: A Survey*. *Soft Sensor Transferability: A Survey*. Retrieved from Semantic Scholar