International Journal of Hydrogen Energy 158 (2025) 150314

Contents lists available at ScienceDirect

International Journal of Hydrogen Energy

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

A dataset for classifying operational states in dry reforming of biogas processes

Renan Akira Nascimento Garcia Escribano [a,b,*], Marcos Antonio Schreiner [a,b], Luiz Eduardo Soares de Oliveira [b], Guilherme Tamanho [a], Julio Cezar da Silva Ferreira [a], Izadora Costa da Silva [a], Paola Cavalheiro Ponciano [a], Helton José Alves [a]

[a] Renewable Materials and Energy Laboratory (LABMATER), Federal University of Paraná, Rua Pioneiro, 2153, Palotina, 85950-000, Paraná, Brazil
[b] Department of Computer Science, Polytechnic Center, Federal University of Paraná—UFPR, Rua Cel. Francisco Heráclito dos Santos, 100, Centro Politécnico – Jardim das Américas, Curitiba, 81531-990, Paraná, Brazil

ARTICLE INFO

Keywords:
Interpretable machine learning
Dry reforming of biogas
Operational states
Labeled dataset

ABSTRACT

Dry reforming of biogas (DR) converts methane and carbon dioxide into syngas, offering a sustainable solution for hydrogen production and greenhouse gas reduction. This study uses operational data from DR reactor sensors to predict process states: Activation, Reaction, and Irregularity. Nine reaction-specific datasets were analyzed via 11-fold cross-validation, ensuring test data independence. Machine learning (ML) models — k-nearest neighbors (KNN), Quadratic Discriminant Analysis (QDA), Support Vector Machine (SVM), and Random Forest (RF) — were evaluated, with RF performing best (88.40% accuracy, 89.04% F1-score for Irregularity). ML enables efficient monitoring by capturing complex variable relationships and responding to operational changes. Explainability analysis (SHAP and PDP) identified key variables, including record count, humidity, and pressure. The study provides a robust dataset and methodology for predicting DR states using operational data, supporting future research in fault prediction and process optimization. This approach enhances DR reactor control, advancing reliable and sustainable hydrogen production.

1. Introduction

Research and development in the field of energy sources that utilize renewable fuels and reduce pollutant emissions have progressed significantly, aiming to advance sustainable energy supply [1]. The use of hydrogen (H_2) as an energy carrier can provide diversification in the energy sector, particularly in the production of heat and electricity. It can be produced through various methods such as steam methane reforming, dry reforming (DR), partial oxidation, water electrolysis, and biomass gasification, each offering different trade-offs in terms of efficiency, cost, and environmental impact [2]. H_2 also plays a key role in decarbonizing the transportation sector — through fuel cell vehicles or engines that burn H_2 directly or blended with other fuels — as well as in industrial processes where it serves as a raw material [3,4].

Among renewable feedstocks for hydrogen production, biogas stands out due to its favorable composition and sustainability profile. Produced via anaerobic digestion of plant and organic waste, biogas typically contains 50%–75% methane (CH_4), 25%–50% carbon dioxide (CO_2), and small amounts of Nitrogen (N_2) and hydrogen sulfide (H_2S) [5,6]. This composition makes biogas particularly suitable for hydrogen production through dry reforming, as it naturally provides both CH_4 and CO_2.

Dry reforming is a catalytic and highly endothermic process that typically operates at temperatures between 600 °C and 900 °C. In this reaction, methane and carbon dioxide — both greenhouse gases — react to produce hydrogen and carbon monoxide, collectively referred to as synthesis gas or syngas [7–9]. This dual greenhouse gas utilization gives DR notable environmental relevance.

One of the primary technical challenges in DR is catalyst deactivation due to coke formation, which can block the reactor bed and increase pressure drop. This issue is exacerbated by the presence of excess water, which favors undesired side reactions such as methanation—where CO_2 or CO reacts with H_2 to regenerate CH_4 and H_2O [10,11]. To minimize coke formation and optimize conversion, it is necessary to

* Correspondence to: Federal University of Paraná, Rua Pioneiro, 2153, Palotina, 85950-000, Paraná, Brazil.
E-mail addresses: renan.akira@ufpr.br (R.A.N.G. Escribano), marcosantonio@ufpr.br (M.A. Schreiner), luiz.oliveira@ufpr.br (L.E.S. de Oliveira), sgttamanho@gmail.com (G. Tamanho), juliosilva@ufpr.br (J.C. da Silva Ferreira), izadoracostasilva@gmail.com (I.C. da Silva), paolaponciano@ufpr.br (P.C. Ponciano), helton.alves@ufpr.br (H.J. Alves).

https://doi.org/10.1016/j.ijhydene.2025.150314
Received 13 December 2024; Received in revised form 13 May 2025; Accepted 3 July 2025
Available online 19 July 2025

Abbreviation	Description
ML	Machine Learning
DR	Dry Reforming
SVM	Support Vector Machine
RF	Random Forest
k-NN	k-Nearest Neighbors
QDA	Quadratic Discriminant Analysis
SHAP	SHapley Additive exPlanations
PDP	Partial Dependence Plot
ICE	Individual Conditional Expectation
ANN	Artificial Neural Network
MLP	Multilayer Perceptron
RBF	Radial Basis Function
MWR	Minutes Without Records

control key parameters such as temperature, flow rate, support material, and catalyst activation conditions. Recent research has focused on developing catalysts that are more resistant to deactivation [11–13].

Given these operational challenges, recent studies have turned to advanced modeling techniques to optimize DR performance. Future prospects for hydrogen production from biogas highlight promising advances, particularly with the integration of artificial intelligence in process optimization and predictive modeling. Artificial Intelligence offers the potential to accelerate the identification of optimal reaction conditions, enable the direct conversion of methane into liquid fuels, and improve the efficiency of hybrid systems, such as the integration of reforming with Fischer–Tropsch synthesis [14].

Despite these advancements, practical challenges persist in monitoring DR processes, especially for small-scale laboratories with limited access to advanced characterization techniques. Fault diagnosis, prognostics, and maintenance of complex systems remain prominent research topics [15,16]. Understanding the importance of early fault detection for process optimization, we classified the DR process into three states: (i) Catalyst Activation (Activation): Before reaction operations, granular catalysts are activated *in situ* with H_2 [17]; (ii) Dry Reforming (Reaction): Reforming of CH_4 to produce synthesis gas [7,18]; and (iii) Irregularity: Includes all situations deviating from the ideal, such as gas line clogging due to coke, furnace resistance failures, or preventive catalyst maintenance.

The identified states were incorporated into a database previously constructed by an operational data collection system of the DR equipment. Part of this system was implemented in a microcontroller, which communicates with the meters and transducers and also sends the data to a server, where it is stored in a database. In recent studies, the data for the DR process used in ML were obtained from a compilation of articles containing reactional data from the process [19,20]. Unlike previous works that relied on reactional data extracted from published articles, this study presents a novel dataset built directly from real-time sensor and transducer measurements in a laboratory-scale reforming reactor. This enables a practical evaluation of operational states without requiring detailed chemical or structural catalyst information.

Our contribution lies not only in the release of a new dataset but also in the development and validation of a methodology for detecting reactor states based on straightforward operational variables. This approach makes intelligent monitoring feasible in environments with limited access to advanced characterization techniques. Differently from previous studies focused on catalyst-specific performance metrics or synthesis gas composition predictions, our models aim to classify the reactor's operational states (Activation, Reaction, Irregularity) using variables that are not catalyst-dependent. This enables broader applicability across different experimental conditions and catalyst types.

The collected data were analyzed, and through this paper, we present a dataset produced from experiments conducted in a reforming reactor [1]. This work aims to provide a robust dataset and methodology to evaluate the predictive power of ML models using straightforward operational data, contributing to future studies and practical applications such as failure prediction contained in the irregularity state.

The application of ML techniques to the collected operational data stems from the growing need for more efficient and adaptive monitoring and control systems. Traditional approaches to managing DR reactors often rely on fixed models or manual oversight, which can struggle to address the complex relationships between the variables involved. ML techniques can offer the ability to analyze operational data, uncovering patterns and insights that might go unnoticed with conventional methods. By leveraging these capabilities, predictive models can improve process efficiency, reduce operational costs, and enhance safety by providing early warnings of potential irregularities, such as system blockages. Therefore, integrating ML into DR processes aligns with the broader goal of advancing sustainable hydrogen production through innovative and data-driven technologies.

Additionally, the explainability analysis using SHAP and PDP highlighted key variables — such as data count, humidity, and pressure — for distinguishing operational states. The Random Forest (RF) model relied heavily on these variables to accurately predict Activation, Reaction, and Irregularity states. The Partial Dependence Plot (PDP) revealed a direct correlation between prolonged operation time (data count) and Irregularity detection. This analysis elucidated the factors driving both correct and misclassified predictions, thereby enhancing operational validation.

In summary, this work presents three main contributions: (i) a novel dataset collected directly from operational sensors in a laboratory-scale DR reactor, (ii) a machine learning-based methodology capable of classifying operational states using simple and catalyst-independent variables, and (iii) a model explainability analysis that clarifies the relevance of key features, guiding process monitoring and optimization. These aspects represent a practical and accessible advancement compared to existing approaches, particularly for research environments with limited analytical infrastructure.

The remainder of this article is organized as follows: Section 2 provides a review of models and operational data in DR processes. Section 3 presents dataset characterization and analysis. Section 4 discusses experiments and model explainability. Finally, Section 5 concludes the work.

2. Models and operational data in dry reforming processes

Recent advances in modeling DR processes have demonstrated significant progress through the integration of ML techniques and experimental data analysis. Bilgiç et al. [21] conducted a comprehensive review of Artificial Neural Network (ANN) applications in hydrogen production systems, revealing that ANN models achieved superior performance ($R^2 = 0.95$) compared to traditional RSM methods ($R^2 = 0.87$) for catalytic reforming optimization.

The kinetic model developed by Nakajima et al. [22] using Langmuir-Hinshelwood mechanisms [23,24] successfully simulated fixed-bed DR reactors under specific conditions (599.85–699.85 °C, 0.3–0.4 L/min flow rates, 1 bar pressure, and CH_4/CO_2 = 1:1 ratio). This model accounts for reactant adsorption, syngas formation, and product desorption phenomena.

Complementing this approach, Hossain et al. [25], utilized ANN models, specifically Multilayer Perceptron (ANN-MLP) and Radial Basis Function (ANN-RBF), to predict the H_2 yield, CO yield, CH_4 conversion, and CO conversion in DR processes using $NiCaFe_2O_4$ catalysts. The input variables for these models included catalyst metal loading (5–15 wt%), feed ratio (0.4–1.0), reaction temperature (700–800° C), and gas

[1] Dataset available at: https://doi.org/10.5281/zenodo.15799058

composition data, which were obtained through gas chromatography. Of the 27 experiments conducted, 70% of the data were used for training, while the remaining 30% were equally divided between testing and validation. The models demonstrated strong predictive performance, with coefficients of determination (R^2) for the ANN-MLP model of 0.9726, 0.8597, 0.9638, and 0.9394, and for the ANN-RBF model of 0.9218, 0.7759, 0.8307, and 0.7425, for H_2 yield, CO yield, CH_4 conversion, and CO conversion, respectively.

Recent work by Kumbhat et al. [5] introduced innovative ML applications for nickel catalyst deactivation prediction during biogas reforming. Their RF models achieved mean $R^2 = 0.979$ in forecasting product distributions (H_2, CO, CH_4, CO_2) and conversion rates, surpassing ANN performance across all metrics. The models' robustness was confirmed through validation with unseen experimental data with $R^2 >$ 0.9. Bilgiç et al. [21] noted that ANN models trained with thermodynamic parameters (CO_2/CH_4 partial pressures) could indirectly detect coke formation — an approach previously demonstrated by Ahmed et al. [26] through mass spectrometry analysis of reactor outputs.

For H_2/CO ratio optimization, Vellayappan et al. [19] developed a CatBoost model analyzing 1,637 data points from 221 studies. SHAP value analysis identified temperature and catalyst particle size as critical control parameters. Similarly, Roh et al. [20] created an interpretable ML framework for catalyst selection, incorporating 6,067 data points from 132 publications to predict CH_4 conversion based on catalyst composition and pretreatment conditions.

While these modeling approaches have advanced DR understanding, Bilgiç et al. [21] identified persistent gaps in real-time catalyst deactivation monitoring and renewable process integration. Experimental studies by Schwengber et al. [12] revealed operational challenges, showing that increased gas flow rates reduce CH_4 conversion in Ni/Al_2O_3 systems, while elevated temperatures risk undesirable methanation reactions. Zain and Mohamed [10] additionally demonstrated moisture-related catalyst degradation mechanisms in copper-based systems, emphasizing the need for comprehensive operational parameter monitoring.

Current research continues to address these limitations through embedded data acquisition systems and advanced ML architectures. The collected experimental data and modeling approaches presented in subsequent sections aim to bridge these gaps, particularly in predicting DR failure modes and identification operational states under variable conditions.

3. Dataset description

The methodological path for data collection and analysis progresses through the following stages:

1. Data acquisition and selection.
2. Data visualization using scatter plots.
3. Analysis of scatter plots in collaboration with an expert to label the operational states of the reaction process.
4. Descriptive statistics of the dataset.
5. Analysis of data dispersion using boxplot [27].
6. Shapiro–Wilk normality test [28,29].
7. Calculation of kurtosis and skewness [30].

The presented dataset consists of 9 reaction processes from DR equipment located in the Laboratory of Materials and Renewable Energies (LABMATER). Data collection was conducted between January 2024 and October 2024 during experiments carried out by laboratory researchers. These experiments utilized DR equipment comprising gas lines that connect the components of the synthetic biogas reforming process, as depicted in Fig. 1.

The dry reforming process operates continuously with the input of CH_4 and CO_2 gases in a 1:1 ratio (1). The input gas is a synthetic mixture of biogas, with CO_2 and CH_4 already in the specified proportion. Initially, the gases are mixed in a gas mixer (2). Then, the gases pass through two heating stages: one in the pre-heating furnace or pre-reactor(3) at 650° C and another in the main furnace or main reactor(4) at 800° C. Inside the main furnace, a steel reactor (5) contains the granular catalyst where the reaction process occurs, producing syngas. After the reaction, the syngas passes through additional gas lines (6) and is monitored for humidity and outlet temperature (d). The equipment also includes thermocouples for temperature monitoring (b and c) and a pressure transducer (a). For control and safety, there is a safety control panel (7). Samples can be collected at the gas outlet for composition analysis using a gas chromatograph.

Table 1
Statistics for all datasets include the feature names, their mean, minimum, and maximum values, as well as the total number of samples: 26066.

Feature	Mean	Min.	Max.
Pressure (bar)	0.58	0.34	2.58
Reactor temperature (°C)	707.20	18	806
Pre-reactor temperature (°C)	573.95	20	654
Humidity (%)	58.37	12.4	99.0
Outlet temperature (°C)	27.93	22.4	43.4
Number data	2187.28	0	7032
Label	–	0	2

Data collected from the reactor represents operational values recorded by sensors and transducers controlled by microcontrollers. Once the DR reactor is activated, approximately 3 data points per minute are automatically recorded, including operating line pressure, pre-furnace temperature, furnace temperature, humidity, outlet gas temperature, record number, and a label identifying the current operational state of the equipment (Table 1).

During the data collection period, seven datasets were excluded due to a lack of main reactor bed heating, indicating equipment maintenance. The raw dataset initially contained 16 variables, including nine operational variables, three sequential identifiers related to the reaction, data collection, and timestamp marking, two alerts for furnace resistance failures (which were not triggered), and two redundant sensor values. Four operational variables were removed due to missing data. The redundant sensors showed no significant variation, and the unused alerts were discarded. The identifiers used solely to track or index reactions in sequence — functioning essentially as row IDs in the DataFrame — were also removed, retaining only the "number data" variable to represent the sequential order of records within each reaction process and also indicates the duration of the reaction process

The variable "number data" is a sequential numeric identifier assigned to each data point collected by the equipment. It represents the chronological order of samples recorded during the reaction process, effectively serving as a temporal reference. Although it does not correspond to an absolute timestamp, its sequential nature allows for the reconstruction of the reaction's progression over time. This makes it particularly useful for analyzing the evolution of operational variables throughout each reaction cycle.

3.1. Data per minute and missing values

To better understand the temporal aspect of the records, the number of records per minute in the datasets was analyzed, as shown in Table 2.

In Table 2, *count* represents the number of minutes of available data. Some datasets are smaller, while others are larger, depending on the duration of reactor usage. The *mean* column indicates the average number of records per minute. Intervals without records (column *min*) indicate system failures, listed as *Minutes Without Records* (MWR). These gaps are attributed to technical issues such as power surges or communication failures. However, the average number of records per minute across the datasets is approximately 2.5, with the median (50%) being 3 in most cases, as well as the 3rd quartile (75%) and the maximum, which suggests that, the system collects 3 records per minute in the majority of cases.

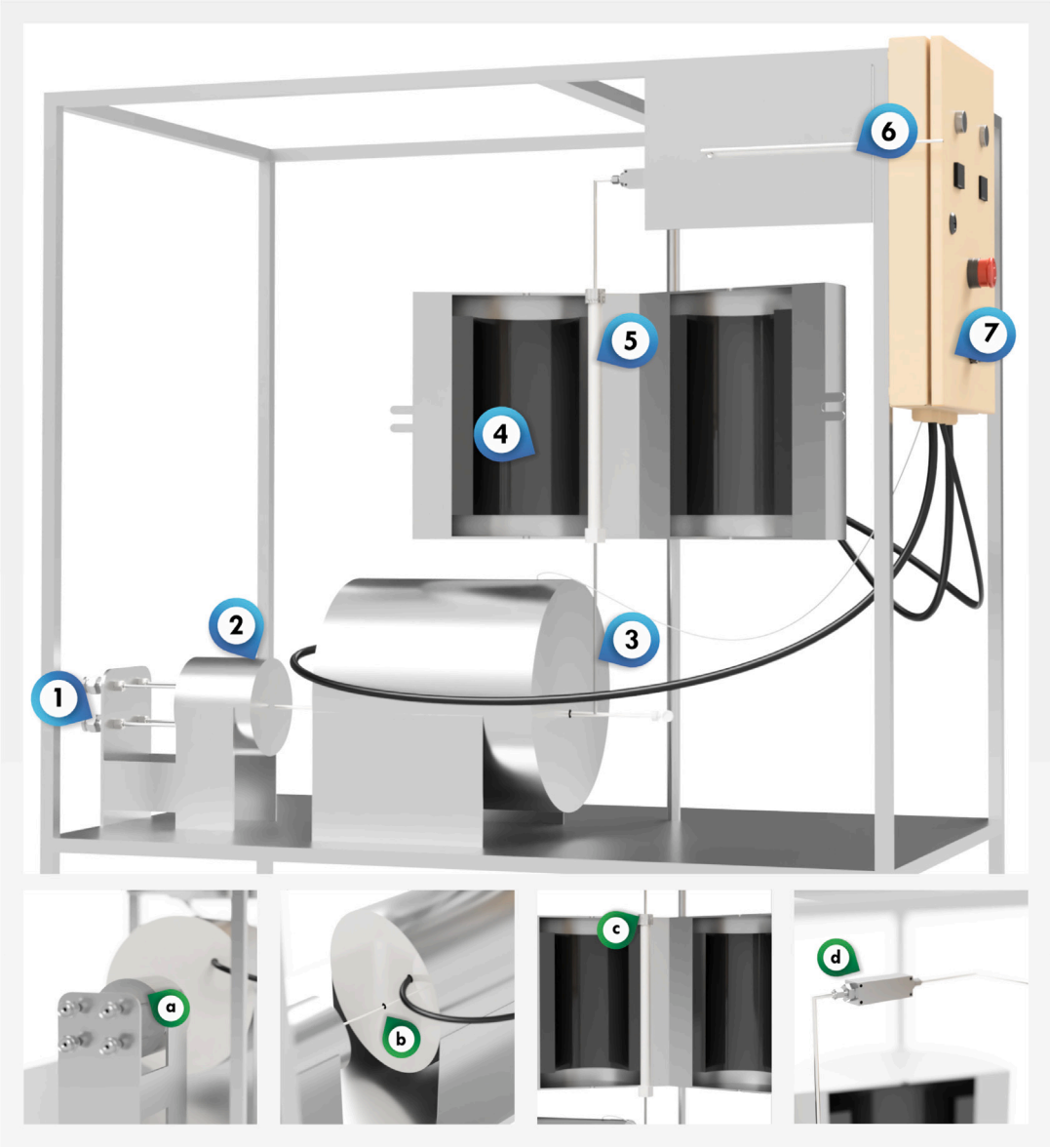


Fig. 1. Diagram of the DR reactor system: (1) gas inlet; (2) gas mixer; (3) pre-heating furnace; (4) main furnace; (5) steel reactor; (6) gas outlet; (7) safety control panel. (a) pressure transducer; (b and c) thermocouples; (d) outlet humidity and temperature sensor.

Table 2
Statistics for the number of records per minute, including the total count, mean, minimum, maximum values, and minutes without records (MWR), as well as the first, second, and third most frequent values.

Count	Mean	Min	MWR	1°	2°	3°	Max
556	2.51	0	13	2	3	3	3
563	2.38	0	31	2	3	3	3
511	2.22	0	53	2	2	3	3
1846	2.33	0	155	2	3	3	3
2223	2.54	0	5	2	3	3	3
2766	2.54	0	4	2	3	3	3
584	2.54	0	1	2	3	3	3
660	2.54	0	1	2	3	3	3
815	2.53	1	0	2	3	3	3

3.2. Labeling of DR process states

Labeling is the process of assigning labels or categories to data samples, transforming raw information into structured and useful data for building supervised learning models. In supervised learning, labels serve as "truths" that guide the model during training, allowing it to learn consistent patterns in the data. Each data input must be linked to a corresponding label, representing the expected output or behavior, which is essential for the model's accuracy and reliability.

In the context of this work, the operation process was characterized by three possible states during a DR reaction. In the dataset, these states are numerically represented as 0, 1, and 2:

0. Catalyst activation (Activation): Before the reaction operations, the granular catalysts are activated *in situ* with H_2.
1. Dry reforming (Reaction): The process of reforming CH_4 to produce syngas.
2. Irregularity: Situations that deviate from the expected, such as gas line blockage due to coke, furnace resistance failure, or preventive catalyst maintenance.

To label the different operational states, scatter plots were created for each of the 9 datasets (Fig. 2). The data presented includes Pressure,

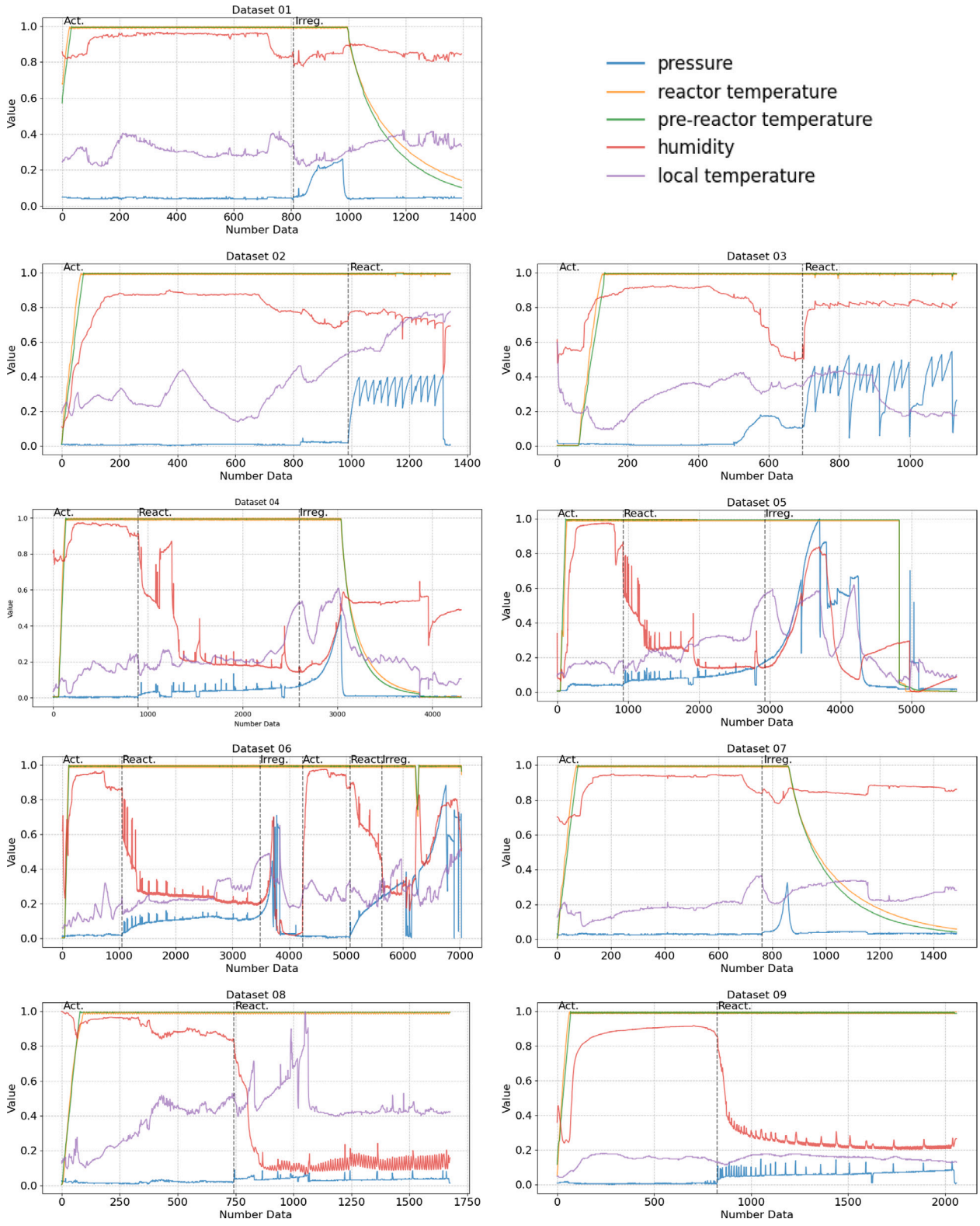


Fig. 2. Scatter plots for the 9 datasets, with vertical dashed lines separating the different labeled classes, along with the behavior of each variable during an operation.

Reactor Temperature, Pre-Reactor Temperature, Humidity, and Outlet Gas Temperature. The data from different datasets were jointly normalized, considering variations in the variables that may occur depending on whether irregularities are present or not.

With the collaboration of the experts who operate the equipment, it was identified that the activation state is characterized by an increase in furnace temperature, humidity, and low pressure with little variation. The reaction state, in turn, is marked by the stabilization of furnace temperatures, a gradual decrease in humidity, and a slight increase in pressure. Small pressure and humidity fluctuations are evident, identified as points where gas is collected from the line for analysis. Irregularities refer to situations that deviate from this pattern; reactions

Table 3
Statistics for the Activation class include feature names, mean, minimum, maximum values, and the total samples: 8546.

Feature	Mean	Min	Max
Pressure (bar)	0.39	0.34	0.74
Reactor Temperature (°C)	755.66	20.00	805.00
Pre-Reactor Temperature (°C)	612.51	20.00	654.00
Humidity (%)	86.99	16.70	99.00
Outlet Temperature (°C)	27.06	22.90	34.90
Number Data	848.96	0.00	5069.00

Table 4
Statistics for the Reaction class include feature names, mean, minimum, maximum values, and the total samples: 9642.

Feature	Mean	Min	Max
Pressure (bar)	0.56	0.34	1.56
Reactor Temperature (°C)	799.15	772.00	806.00
Pre-Reactor Temperature (°C)	649.89	641.00	653.00
Humidity (%)	40.07	18.00	91.50
Outlet Temperature (°C)	28.24	24.80	43.40
Number Data	1974.73	696.00	5627.00

Table 5
Statistics for the Irregularity class include feature names, mean, minimum, maximum values, and the total samples: 7878.

Feature	Mean	Min	Max
Pressure (bar)	0.82	0.34	2.58
Reactor Temperature (°C)	542.10	18.00	805.00
Pre-Reactor Temperature (°C)	439.17	21.00	654.00
Humidity (%)	49.72	12.40	90.60
Outlet Temperature (°C)	28.49	22.40	36.10
Number Data	3899.23	762.00	7032.00

where the bed became clogged showed a gradual increase in humidity and pressure until the reactor was shut down.

The Tables 3, 4, and 5 present the descriptive statistics of the variables for the Activation, Reaction, and Irregularity classes, respectively. The analysis of these tables reveals important variations in the operational conditions associated with each class. For example, in the Irregularity class (Table 5), there is a significant increase in the average and maximum pressure values compared to the other classes, suggesting that high-pressure conditions may be associated with failures or operational deviations, making it a relevant variable for system monitoring and control.

For reactor and pre-reactor temperatures, a larger range is observed in the Activation and Irregularity classes, with extremely low minimum values, indicating heating when the furnace is turned on and cooling when turned off. In the Reaction class, temperatures tend to be higher and more stable, indicating greater thermal stability during reactions.

Humidity varies significantly across the phases. In the Activation class, the average is higher (86.99), while in the Reaction class, the average drops to 40.07. This suggests that the activation process occurs in a more humid environment, while the reactions require a drier environment. In the Irregularity class, the average humidity is intermediate, possibly contributing to irregularities or being a consequence of them.

Outlet temperature increases slightly across the phases, with more fluctuation in the Reaction class, indicating wider fluctuations during reactions and greater stability during irregularities. This highlights the stability of the outlet temperature as a relevant factor for avoiding irregularities.

The moment when the state changes was identified in the 9 datasets, but it is important to note that the transition between states is gradual, and the number of records in the transition range is variable, depending on the experiment being conducted. To accurately identify this transition range, new studies including reaction data beyond the scope of this work are needed.

Table 6
Skewness and kurtosis values for each variable, combining all datasets.

Feature	Skewness	Kurtosis
Pressure (bar)	2.66	7.25
Reactor temperature (°C)	−2.29	3.47
Pre-reactor temperature (°C)	−2.25	3.27
Humidity (%)	0.01	−1.63
Outlet temperature (°C)	1.09	1.32
Number data	0.83	−0.38

3.3. Analysis of data dispersion using boxplot

The boxplot analysis was conducted to examine variable distributions and identify outliers. This graphical approach provides clear visualization of data dispersion, median values, quartiles, and extreme values (Fig. 3), essentials for detecting system deviations. Importantly, all identified outliers were preserved in our dataset as they correspond to genuine physical phenomena in the DR process rather than measurement noise. These abrupt variations reflect critical operational events including catalyst activation procedures, system failures, or process restart conditions.

The analysis revealed distinct patterns across variables. Reactor and pre-reactor temperatures predominantly cluster at high values near upper limits, with lower-value outliers corresponding to physically meaningful scenarios such as furnace heating/cooling cycles or equipment failures. Pressure measurements show concentration at lower levels but include significant high-value outliers representing important operational events like sudden pressure surges during coke formation. Humidity displays relatively stable variation within a 20–100 range without extreme outliers, indicating its more consistent behavior. Outlet temperature primarily ranges between 25–35 with occasional higher values marking rare operational conditions, while data count distribution shows consistent sampling density between 1000–3500 records.

This intentional preservation of outliers ensures machine learning models capture the complete dynamic range of authentic process behavior, including critical transition events between operational states. The robust handling of these characteristic data features enhances model capability to distinguish normal from atypical operation while maintaining the physical significance of all measured process parameters. The approach acknowledges that in catalytic reforming processes, abrupt variable changes often carry more diagnostic value than steady-state measurements for operational state classification.

3.4. Analysis of the normality of the data

The Shapiro–Wilk normality test rejected the hypothesis of normality for all variables ($p < 0.001$ in all cases), even after applying the Box–Cox transformation. This indicates that the operational variables do not follow a normal distribution. Subsequently, additional analyses of skewness and kurtosis were performed to better understand the characteristics of these distributions.

Analyzing Table 6, the skewness revealed that pressure exhibits strong positive skewness (2.66), while the reactor and pre-reactor temperatures show negative skewness (−2.29 and −2.25), indicating long tails to the left. The kurtosis for these variables showed high values (7.25 for pressure), suggesting more concentrated distributions with higher peaks than a normal distribution. In contrast, humidity exhibited a flatter and more symmetric distribution, with negative kurtosis, while the output temperature showed a slightly more concentrated distribution around the mean values.

These results reinforce the rejection of normality and help to understand the nature of the variable distributions, which has significant implications for the choice of ML models. Parametric models that

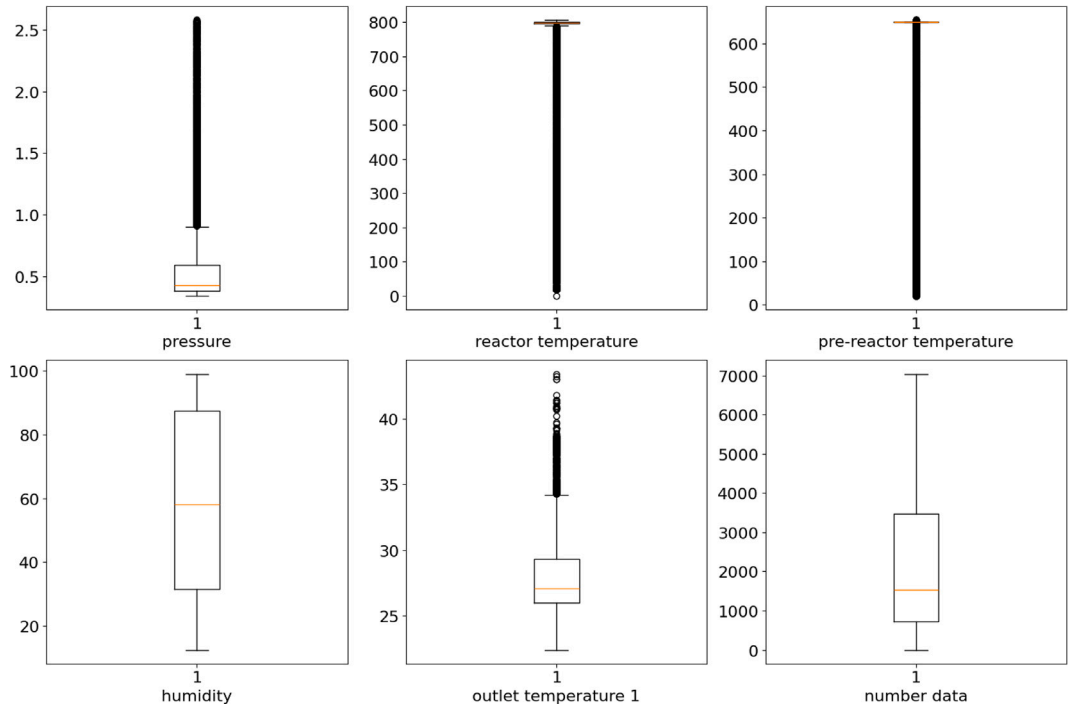


Fig. 3. Box plots of the variables combining data from all datasets, showing the distribution, median, quartiles, and potential outliers.

assume data normality, such as Linear Regression and Linear Discriminant Analysis, may be less effective due to the presence of skewed and heavy-tailed distributions. Conversely, non-parametric algorithms, such as Decision Trees, Random Forests, and Support Vector Machines, may be more suitable as they are more robust to non-normality.

4. Analysis of the experiments

This section presents the experimental protocol adopted for conducting the initial experiments, aiming to evaluate the performance of various ML models. The importance assigned to the features is also discussed, using SHAP values for the model that demonstrated the best performance. Finally, an individual analysis is conducted using the techniques ICE (Individual Conditional Expectation) and PDP (Partial Dependence Plot) to explore the most complex situation identified in the analyzed datasets.

4.1. Protocol

To evaluate the performance of machine learning models, the nine available datasets were split into different training and testing combinations. The main criterion for the split was that a dataset used in training should not be used in testing, ensuring independence between the sets. Additionally, due to the absence of all three operational states (Activation, Reaction, and Irregularity) in some datasets, it was necessary to create specific combinations. Pairs of datasets containing the Activation and Reaction states or Activation and Irregularity states were considered, allowing for a robust analysis even with limited data. Table 7 presents these combinations, detailing the training and testing sets for each case.

Each combination was identified by a unique code C_n, where n is a sequential number from 1 to 11, as presented in the first column of Table 7. This notation simplifies the reference to combinations throughout the study, facilitating comparative analysis between them.

For the ML tests, four different classifiers were selected to test the data sample: k-nearest neighbors (KNN), quadratic discriminant analysis (QDA), support vector machines (SVM), and random forests (RF) of decision trees. KNN is an instance-based learning method that stores all available training data and classifies test samples based on a similarity measure, such as Euclidean distance. QDA was chosen for this study because it is a parametric model that assumes data measurements follow a normal distribution. This model allows for the verification of normality assumptions using the Shapiro–Wilk test and explores the relationship with other models that also assume normal data distribution. Due to this characteristic, QDA is expected to perform worse in scenarios where this assumption does not hold adequately. SVM, a classification algorithm, constructs a hyperplane in a high-dimensional space, separating classes with a maximum margin. This approach makes SVM particularly suitable for complex classification problems. RF, is an ensemble approach that combines multiple decision tree predictors. The principle behind ensemble methods is that a group of weak learners (in this case, decision trees) can come together to form a strong learner. One of the advantages of RF is its ability to handle imbalanced data efficiently and its speed in training and classification processes.

These classifiers were chosen due to the number of samples in the dataset and their different approaches and characteristics, allowing for a comprehensive evaluation of the data sample and a better understanding of its behavior and performance under different conditions [31].

Table 7
Different combinations of datasets for training and testing.

Combinations	Training	Testing
C_1	1, 2, 3, 5, 6, 7, 8, 9	4
C_2	1, 2, 4, 3, 6, 7, 8, 9	5
C_3	1, 2, 4, 3, 5, 7, 8, 9	6
C_4	4, 3, 5, 6, 7, 8, 9	1, 2
C_5	2, 4, 5, 6, 7, 8, 9	1, 3
C_6	2, 4, 3, 5, 6, 7, 9	1, 8
C_7	2, 4, 3, 5, 6, 7, 8	1, 9
C_8	1, 4, 3, 5, 6, 8, 9	7, 2
C_9	1, 2, 4, 5, 6, 8, 9	7, 3
C_{10}	1, 2, 4, 3, 5, 6, 9	7, 8
C_{11}	1, 2, 4, 3, 5, 6, 8	7, 9

Table 8
Default scikit-learn hyperparameters of the classifiers used for training.

Model	Hyperparameter	Default value
RF	n_estimators	100
	criterion	"gini"
	max_depth	None
	min_samples_split	2
	min_samples_leaf	1
	max_features	"sqrt"
	bootstrap	True
QDA	reg_param	0.0
	store_covariance	False
	tol	1e−4
SVM	C	1.0
	kernel	"rbf"
	degree	3
	gamma	"scale"
	shrinking	True
	tol	1e−3
	max_iter	−1
K-NN	n_neighbors	5
	weights	"uniform"
	algorithm	"auto"
	leaf_size	30
	p	2
	metric	"minkowski"

Table 9
Accuracy for each combination, with the best model underlined. The last row shows the average Accuracy per model across all combinations.

Combination	KNN	RF	QDA	SVM
C_1	89.4	89.0	83.1	88.1
C_2	82.8	92.3	77.9	87.5
C_3	67.2	65.5	76.0	70.1
C_4	83.2	88.9	73.1	82.0
C_5	83.8	83.9	59.9	73.8
C_6	93.3	93.3	88.5	89.6
C_7	97.8	94.4	72.6	93.9
C_8	85.9	91.2	75.8	94.7
C_9	83.1	81.2	71.5	75.3
C_{10}	95.3	95.0	92.6	94.5
C_{11}	98.7	97.7	78.0	97.6
Average	**87.3**	88.4	**77.2**	**86.1**

All classifiers were used with their default hyperparameters (Table 8), setting only the random state to 42 to ensure the reproducibility of results. The models were implemented in pipelines along with StandardScaler [32], to standardize the data by adjusting them to zero mean and unit standard deviation. The experiments and results to be presented were carried out using scikit-learn [33] and SHapley Additive exPlanations (SHAP) [34], open-source machine learning libraries in Python.

To evaluate the performance of the models, the chosen metrics are **accuracy** and **F1 score**. While accuracy is commonly used to measure the overall correctness of predictions, it can be misleading in the presence of class imbalance, as it does not distinguish between types of errors. To address this, the F1 score is employed as a complementary metric that balances the trade-off between *precision* and *recall*.

According to Powers (2020) [35], precision measures the proportion of true positives among all predicted positives:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

while recall (or sensitivity) measures the proportion of actual positives that were correctly identified:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

These two metrics often vary inversely, and optimizing one can reduce the other. The F1-score, defined as the harmonic mean of precision and recall,

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

provides a more informative evaluation in scenarios where positive class detection is critical. Powers (2020) [35] highlights that although the F1 score does not account for true negatives, it remains a robust metric, especially for imbalanced classification problems where accuracy alone may fail to represent the model's effectiveness.

Table 10
F1-Score for each combination, with the best model underlined. The last row shows the average F1-Score per model across all combinations.

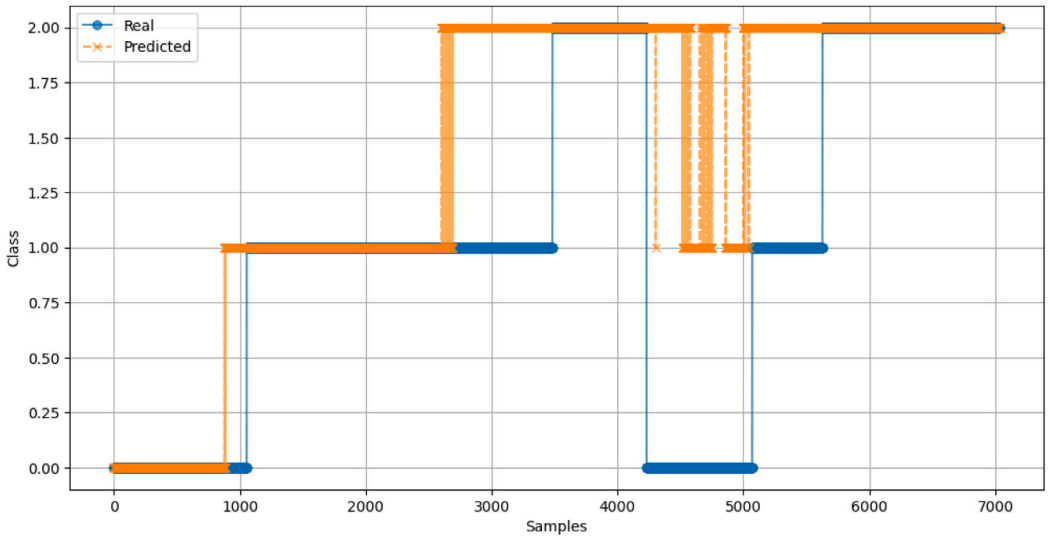
Combin.	Class	KNN	RF	QDA	SVM
C_1	Activation (0)	92.7	93.1	85.2	89.6
	Reaction (1)	87.3	86.9	83.6	86.5
	Irregularity (2)	89.9	89.0	81.2	88.8
C_2	Activation (0)	88.7	93.2	85.9	77.5
	Reaction (1)	76.6	88.7	77.3	94.8
	Irregularity (2)	84.2	94.4	75.6	86.4
C_3	Activation (0)	67.4	63.5	66.2	69.0
	Reaction (1)	69.0	62.3	85.1	64.1
	Irregularity (2)	65.9	69.1	71.7	76.5
C_4	Activation (0)	93.7	92.4	88.2	89.8
	Reaction (1)	43.4	73.4	38.8	39.7
	Irregularity (2)	79.0	91.6	58.1	80.2
C_5	Activation (0)	94.2	93.8	79.6	91.8
	Reaction (1)	52.4	60.0	23.8	3.1
	Irregularity (2)	78.7	78.9	52.8	64.3
C_6	Activation (0)	96.7	96.7	94.6	93.9
	Reaction (1)	91.0	91.6	87.0	88.6
	Irregularity (2)	87.7	86.8	71.9	77.7
C_7	Activation (0)	99.4	97.1	89.5	96.9
	Reaction (1)	97.4	94.2	64.2	95.1
	Irregularity (2)	93.8	87.2	47.4	80.2
C_8	Activation (0)	94.3	92.8	86.5	95.7
	Reaction (1)	46.4	77.4	43.1	89.3
	Irregularity (2)	85.5	96.4	71.8	94.8
C_9	Activation (0)	93.3	93.7	87.1	92.2
	Reaction (1)	50.7	35.4	32.4	2.4
	Irregularity (2)	79.6	79.0	65.7	69.4
C_{10}	Activation (0)	97.3	95.1	95.2	94.9
	Reaction (1)	92.8	94.0	94.4	94.2
	Irregularity (2)	94.0	96.2	83.3	94.2
C_{11}	Activation (0)	99.2	97.7	92.2	97.4
	Reaction (1)	99.0	98.3	70.2	99.3
	Irregularity (2)	97.2	96.8	60.8	95.0
Average	Activation (0)	**93.0**	**92.2**	**87.6**	**91.7**
	Reaction (1)	**77.8**	**83.0**	**67.9**	**79.9**
	Irregularity (2)	**84.8**	**89.0**	**66.5**	**83.7**

4.2. Results

Tables 9 and 10 presents the accuracy and F1 score of the selected ML models for each of the 11 training and testing combinations described earlier.

The results presented in Table 9 indicate that the RF model achieved the highest average accuracy (88.4%), standing out as the most consistent across the evaluated combinations. KNN also performed well, especially in combinations C_7 and C_{11}, where it achieved the best results, with 97.8% and 98.7%, respectively. These results suggest that both models have good generalization capabilities in most evaluated scenarios.

Table 10, which presents the F1-scores, reinforces these findings by detailing the balance between precision and recall. RF stands out again, particularly in the Irregularity class, where it achieved the highest average F1-score (89.0%), indicating effectiveness in correctly capturing the most critical class for the study. In contrast, KNN achieved its



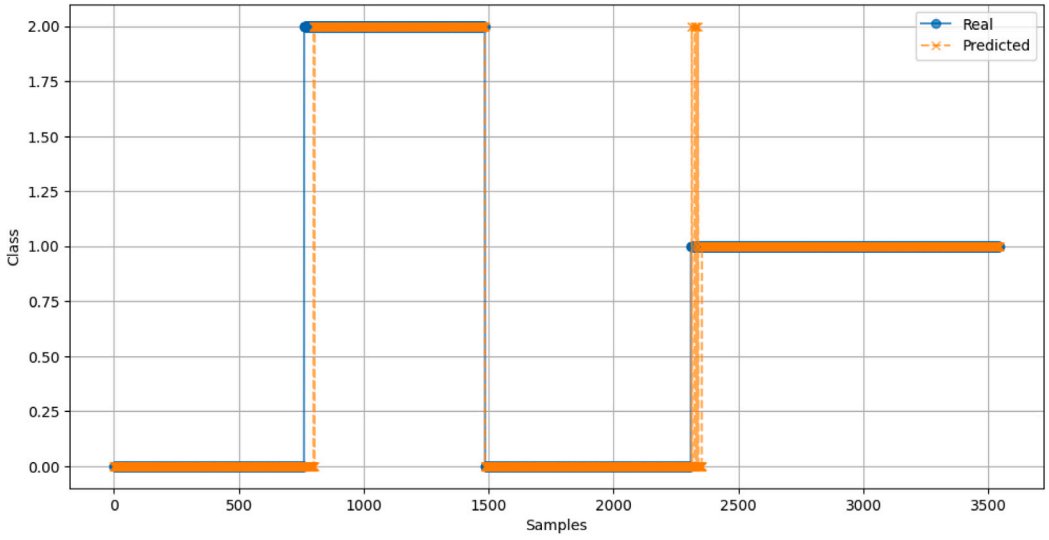


Fig. 4. Classification error comparison showing the plots with the predictions of the RF model and the true values for the combinations C_3 and C_{11}.

best performance in the Activation class, with an average of 92.9%, demonstrating strength in scenarios with more well-defined patterns.

Combinations C_3 and C_5 highlight the greatest challenges for all models. In the case of C_3, both accuracy and F1-score were low, suggesting greater complexity in class separation. In C_5, the low performance, especially in the Reaction class, reflects difficulties in capturing the nuances of this operational state, possibly due to class imbalance or high variability in the data.

Additionally, SVM's performance in combination C_8 was remarkable (94.6% accuracy), corroborated by a high F1-score in the Reaction class (89.3%). This suggests that the model can effectively capture decision boundaries in specific scenarios, although its overall average was behind RF and KNN.

As expected, QDA showed lower performance in both metrics due to its sensitivity to the normality assumption. However, its inclusion was valuable for highlighting the limitations of parametric models in this dataset. These results emphasize the importance of evaluating multiple combinations to identify not only the most consistent models but also scenarios requiring additional adjustments to the data or models.

Table 11
Average confusion matrix for the Random Forest model across the 11 combinations.

Class	0	1	2
Activation (0)	**88.01%**	7.72%	4.27%
Reaction (1)	2.98%	**79.42%**	17.59%
Irregularity (2)	1.59%	5.61%	**92.79%**

To better understand the classification behavior of the RF model, Table 11 presents the average confusion matrix in proportional terms. We do not present the absolute values because there is an imbalance between the classes.

As seen in Table 11, the Irregularity class has the highest accuracy rate with 92.79%, followed by the Activation class, with 88.01%, and finally the Reaction class with 79.42%. This data suggests that the RF classifier has greater accuracy in the Inconsistency class, which among other things represents a failure in the reforming reaction.

Table 12
Confusion matrix from Random Forest models, showing the distribution of classification errors for the combinations C_3 and C_{11} across the three classes.

Combination	Class	0	1	2
C_3	Activation (0)	881	496	518
	Reaction (1)	0	1575	1411
	Irregularity (2)	0	0	2152
C_{11}	Activation (0)	1586	0	0
	Reaction (1)	35	1193	7
	Irregularity (2)	39	0	685

The false negatives in the Irregularity class are mostly associated with the Reaction class (5.61%), likely due to the transition phase from the Reaction state to the Irregularity state (the state transition will be further explored in Section 4.3). However, there are also cases where the transition occurs directly from the Activation state to Irregularity, resulting in 1.59% of false negatives. The false positives for the Irregularity class are predominantly observed during the transition to the Inconsistency class, accounting for 17.59% of the examples from the Reaction class, i.e., false negatives from Reaction. This indicates that during the transition from Reaction to Irregularity, the RF model tends to misclassify inputs as Inconsistency. Such behavior is desirable in the context of failure prediction, where early warnings are preferable.

Table 12 presents the confusion matrices for combinations C_3 and C_{11}, highlighting the distribution of classification errors for the operational states. These results reflect the worst (C_3) and best (C_{11}) performance scenarios of the RF model, providing a clear view of the differences between combinations.

In combination C_3, the errors are more pronounced. In the Reaction class, 1411 instances were incorrectly classified as Irregularity, indicating again that the false negatives of the reaction class are mostly for irregularity. The Activation class also shows a high number of errors, with 496 instances classified as Reaction and 518 as Irregularity. However, in this scenario, all instances of the Irregularity class were classified correctly, which suggests that in this scenario, the Irregularity class overlaps with other classes, i.e., it correctly classifies all instances of the irregularity class, but incorrectly indicates this class for instances of the reaction and activation classes.

On the other hand, combination C_{11} shows much more consistent results. The Activation class was correctly classified in all 1586 instances, while the Reaction and Irregularity classes only showed a few errors, with 35 instances of Reaction classified as Activation and 7 as Irregularity. In addition to 39 Irregularities classified as Activation. In Fig. 4, it is observed that the errors in combination C_{11} occur primarily during the transitions between classes. During these transitions, there is a gradual variation in the features, making it challenging to differentiate between the classes.

These results highlight that the errors observed in combination C_3 are directly related to specific characteristics of the dataset used. In C_3, dataset 6 was tested, characterized by a unique event: the regeneration of the catalyst followed by the restart of the DR process. This distinct situation between the datasets underscores the importance of a detailed analysis of the features (presented in the next section) and their influence on the model, as well as the need for adjustments to mitigate issues of imbalance in the relevance of the variables.

4.3. Interpretations of ML

To interpret in more detail the impact of predictor variables on the performance of the Random Forest (RF) model in the combinations with the best and worst performance, we present the SHAP value plots in Fig. 5. The SHAP-generated plot precisely displays the distribution of SHAP values for each predictor variable, providing an in-depth view of how each variable influences the model's predictions. Each point on the plot represents the impact of a specific observation on the model's prediction, allowing for a granular analysis. The colors, ranging from blue (for lower values) to red (for higher values), indicate the magnitude of the variable's value, while the position of each point on the plot shows the intensity and direction of the impact that this variable has on the model's decision. This methodology is crucial as it allows not only the evaluation of the relative importance of variables, but also a clear and objective understanding of how these variables shape the model's predictions, offering a detailed explanation of the model's behavior in relation to different inputs and experimental conditions.

The variable `number data` stands out as the most relevant in both the best and worst-performing scenarios. This behavior highlights the hypothesis that operational time plays together with the other variables a key role in classifying operational states.

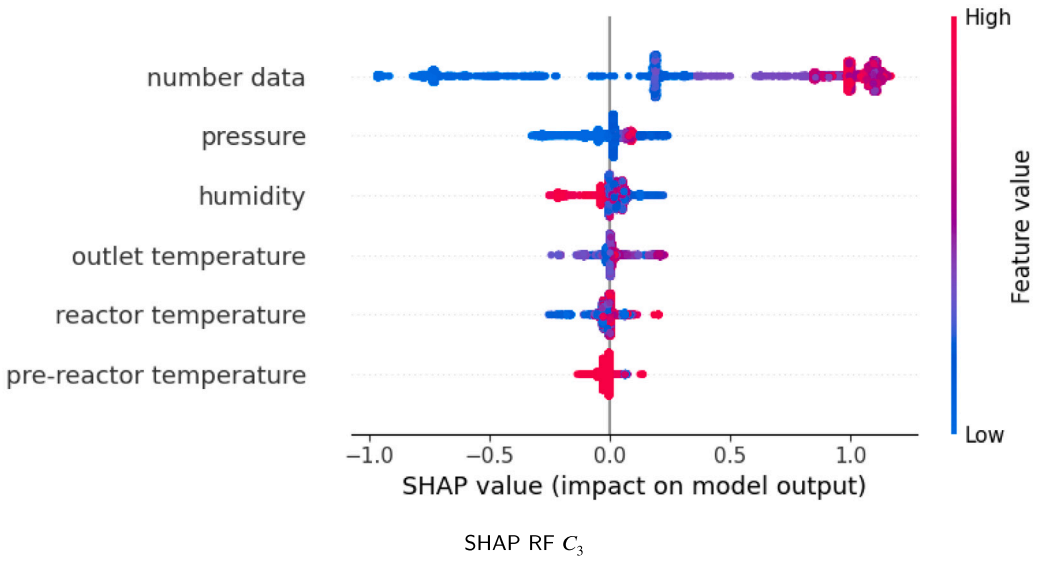
In the best-performing combination, we observe that secondary variables such as `humidity` and `pressure` also have a significant impact, contributing to the model's robustness. This diversity of influences reinforces the model's ability to capture multiple aspects of the experimental conditions, reflecting its capacity to consider a wider range of operational characteristics. The balance in the importance of variables, which does not overly rely on a single factor, is crucial to ensuring correct classification of the operational states, demonstrating the model's ability to understand the complexities associated with the data and experimental context.

On the other hand, in the worst-performing combination, the dominance of the variable `number data` over the others severely compromised the model's performance. This excessive focus on a single variable seems to limit the model's ability to generalize to more complex scenarios, as observed in dataset 6 (Fig. 2), where the reaction process is restarted after a long operational period. This over-reliance may be a reflection of the model's inability to explore other important variables, leading to inadequate and less accurate classifications.

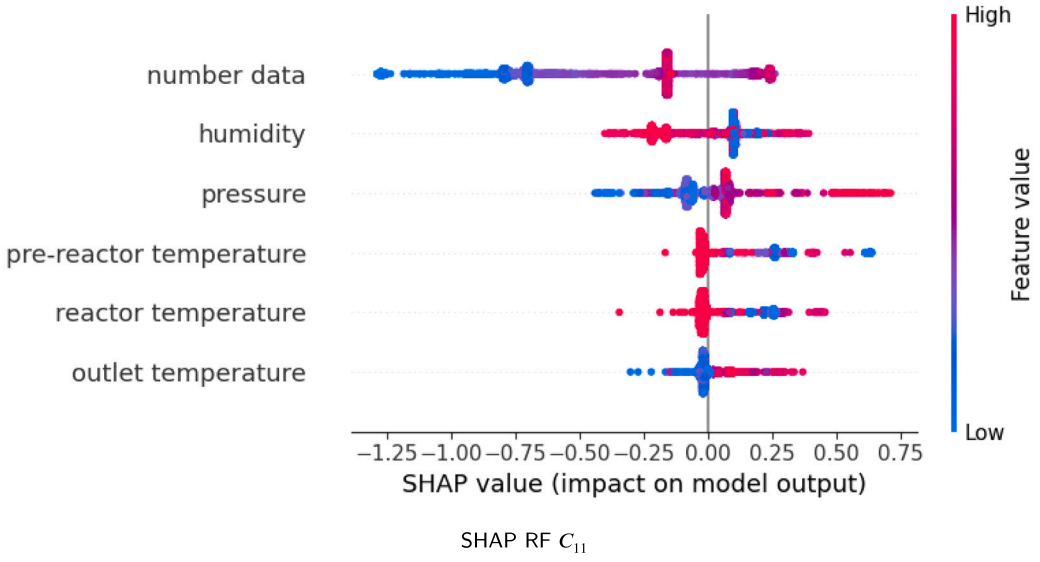
To understand the poor performance in combination C_3, Fig. 6 presents the ICE and PDP graphs, addressing all features in the classification of the Irregularity class (class 2). The choice of the class with the highest number of false negatives is key, as it allows for a focused analysis of the model's behavior and its sensitivity to variations in these features while keeping others constant. This analysis, by exploring interactions between variables, offers a deeper understanding of the reasons behind the model's poor performance, connecting with previous observations and identifying potential limitations of the model in detecting the Irregularity class.

The variable `number data`, which was identified as the most important according to the SHAP method, shows, through the ICE plots, a considerable impact, especially in specific value intervals (above 2500, as shown in Fig. 6), where the probability of classification as Irregularity increases drastically. This suggests a possible temporal association with irregularities, where longer reaction times may correlate with operational failures or disturbances. However, the high sensitivity of this variable may also explain the difficulty in distinguishing Irregularity from other classes. This phenomenon is reflected in the confusion matrix, where 1411 samples from the Reaction class were incorrectly classified as Irregularity. This finding raises the hypothesis that the model is capturing superficial patterns related to `number data`, but it fails to generalize to more complex and varied contexts.

The variable `pressure`, which ranks second in importance, exhibits an interesting nonlinear behavior in the ICE plots. Its influence becomes more pronounced in intermediate value ranges, where the probability of Irregularity increases. This pattern can be interpreted as a reflection of operational instability, a characteristic of the Irregularity class, which may be associated with risks of clogging or failure within systems. In operational reality, elevated pressure is often a sign that the system is nearing a critical point, where increased pressure indicates an impending failure, often linked to material buildup or obstructions in the flow.



SHAP RF C_3



SHAP RF C_{11}

Fig. 5. SHAP graphs showing the importance and impact of variables in the RF models with C_3 and C_{11} combinations.

The variable `humidity` shows an intriguing pattern in the ICE plots, with a more pronounced impact at intermediate levels. This suggests that humidity may be associated with specific scenarios where irregularities are more likely to occur, indicating a correlation between these two factors. However, the individual variations in the ICE plots reveal that the model may be responding to specific sample fluctuations, contributing to classification errors between Activation and Irregularity, both of which are often associated with high humidity levels.

In contrast, the variables `reactor temperature` and `pre-reactor temperature` exhibit a more stable behavior and have minimal influence on the predictions for the Irregularity class. This stability reflects their low relative importance in the SHAP ranking and suggests that these variables capture more general system conditions or characteristics related to other classes, rather than serving as distinguishing factors for identifying irregularities in Dataset 6.

As evidenced by the confusion matrix values (Table 12), along with the ICE and PDP plots, the model tends to associate prolonged reactions with Irregularity states. This results in false predictions of Irregularity for Activation and Reaction states during restart moments. Additionally, the low relevance attributed to variables like `pre-reactor temperature` and `reactor temperature` highlights the model's difficulty in identifying complementary patterns that would be essential for correctly distinguishing operational states.

Therefore, the findings emphasize the need to explore approaches that promote a more balanced contribution of the features in the model, particularly in scenarios with unique characteristics, such as less frequent state transitions. This analysis also highlights the potential of fine-tuning input variables to minimize excessive dependencies on a single factor, fostering more consistent performance across different data combinations and operational conditions.

5. Final considerations

This study demonstrated the feasibility of using basic operational data to develop predictive models for biogas dry reforming processes. Through comparative evaluation of four machine learning algorithms — *Random Forest*, KNN, SVM, and QDA — the approach showed potential for classifying operational states (Activation, Reaction, and Irregularity), with *Random Forest* achieving the best performance (88.40% accuracy). However, the analysis revealed model sensitivity to data variability, particularly regarding temporal dependence expressed by

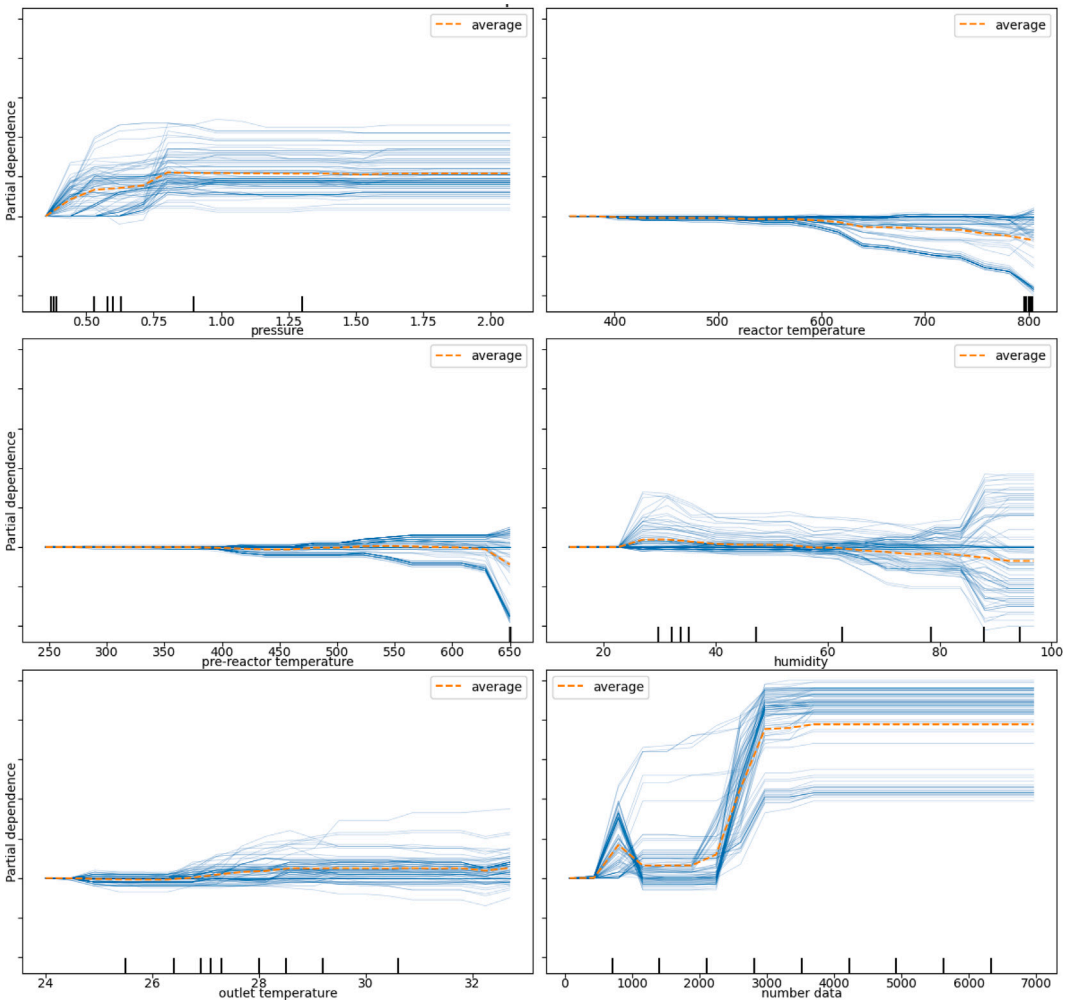


Fig. 6. ICE and PDP representations for all features with respect to the Irregularity class using the RF model with the C_3 combination.

the `number data` variable, identified as a predominant factor through SHAP and PDP techniques.

The obtained results establish foundations for predictive monitoring systems in laboratory settings, yet highlight the need for improvements in complex operational scenarios. The explainability analysis demonstrated that strong correlation with temporal variables may compromise model generalization capability, suggesting the need for strategies incorporating additional process features.

Looking ahead, future research will focus on:

- Reducing variable dependencies and improving model accuracy in extreme operating conditions through advanced feature engineering and hybrid modeling approaches;
- Investigating state transition dynamics using time-series analysis and unsupervised learning to enhance phase-specific performance;
- Incorporation of reaction volumetrics (inlet/outlet molar flows of CH_4 and CO_2, where available) as additional parameters;
- Developing efficiency metrics derived from volumetric and kinetic data to refine real-time process optimization;
- Reproducing the state labeling approach in pilot-scale experiments will be critical to validate the model's industrial applicability and robustness under scaled-up conditions.

This comprehensive approach will enable more precise monitoring and optimization of dry reforming processes across various operational conditions.

CRediT authorship contribution statement

Renan Akira Nascimento Garcia Escribano: Writing – review & editing, Software, Investigation, Writing – original draft, Methodology, Data curation. **Marcos Antonio Schreiner:** Writing – original draft, Investigation, Data curation, Supervision, Formal analysis, Conceptualization. **Luiz Eduardo Soares de Oliveira:** Writing – review & editing, Methodology, Supervision, Investigation. **Guilherme Tamanho:** Visualization, Data curation. **Julio Cezar da Silva Ferreira:** Software, Data curation. **Izadora Costa da Silva:** Software, Data curation. **Paola Cavalheiro Ponciano:** Supervision, Project administration. **Helton José Alves:** Resources, Funding acquisition.

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

The authors acknowledge the support of the Coordination for the Improvement of Higher Education Personnel (CAPES), through the Program of Academic Excellence (PROEX), which contributed to the development of this work.

The authors also acknowledge GIZ (Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH), the Brazilian Ministry of Mines and

Energy, and Germany's Federal Ministry for Economic Cooperation and Development (BMZ) for the financial support provided via the iH2Brasil public call, linked to the H2Brasil project (Contract No. 39/2022).

Additionally, the authors gratefully acknowledge the financial support from the National Council for Scientific and Technological Development (CNPq) under grants No. 132088/2024-0 (R. A. N. G. Escribano), 303298/2022-7 (L. E. S. de Oliveira), and 305433/2022-9 (H. J. Alves).

References

[1] Longoria G, Lynch M, Curtis J. Green hydrogen for heating and its impact on the power system. Int J Hydrog Energy 2021;46(53):26725–40. http://dx.doi.org/10.1016/j.ijhydene.2021.05.171.

[2] Hajizadeh A, Mohamadi-Baghmolaei M, Cata Saady NM, Zendehboudi S. Hydrogen production from biomass through integration of anaerobic digestion and biogas dry reforming. Appl Energy 2022;309:118442. http://dx.doi.org/10.1016/j.apenergy.2021.118442, URL: https://www.sciencedirect.com/science/article/pii/S0306261921016676.

[3] Abdin Z, Zafaranloo A, Rafiee A, Mérida W, Lipiński W, Khalilpour KR. Hydrogen as an energy vector. Renew Sustain Energy Rev 2020;120:109620. http://dx.doi.org/10.1016/j.rser.2019.109620.

[4] Alves HJ, Bley Junior C, Niklevicz RR, Frigo EP, Frigo MS, Coimbra-Araújo CH. Overview of hydrogen production technologies from biogas and the applications in fuel cells. Int J Hydrog Energy 2013;38(13):5215–25. http://dx.doi.org/10.1016/j.ijhydene.2013.02.057.

[5] Kumbhat A, Madaan A, Goel R, Appari S, Al-Fatesh AS, Osman AI. Predicting nickel catalyst deactivation in biogas steam and dry reforming for hydrogen production using machine learning. Process Saf Environ Prot 2024;191:1833–46. http://dx.doi.org/10.1016/j.psep.2024.09.020.

[6] Chava R, Purbia D, Roy B, Janardhanan VM, Bahurudeen A, Appari S. Effect of calcination time on the catalytic activity of $Ni/y-Al_2O_3$ cordierite monolith for dry reforming of biogas. Int J Hydrog Energy 2021;46(9):6341–57. http://dx.doi.org/10.1016/j.ijhydene.2020.11.125.

[7] Chen X, et al. Biogas dry reforming for syngas production: catalytic performance of nickel supported on waste-derived SiO2. Catal Sci Technol 2015;5(2):860–8.

[8] Chein R, Yang Z. Experimental study on dry reforming of biogas for syngas production over Ni-based catalysts. ACS Omega 2019;4(25):20911–22. http://dx.doi.org/10.1021/acsomega.9b01784, PMID: 31867481.

[9] Oliveira LG, Machado B, Pereira de Souza L, Gosch Corrêa GC, Polinarski MA, Cavalcanti Trevisan SV, et al. Dry reforming of biogas in a pilot unit: Scale-up of catalyst synthesis and green hydrogen production. Int J Hydrog Energy 2022;47(84):35608–25. http://dx.doi.org/10.1016/j.ijhydene.2022.08.141.

[10] Zain MM, Mohamed AR. An overview on conversion technologies to produce value added products from CH4 and CO2 as major biogas constituents. Renew Sustain Energy Rev 2018;98:56–63. http://dx.doi.org/10.1016/j.rser.2018.09.003.

[11] jun Zhang R, fu Xia G, feng Li M, Wu Y, Nie H, dong Li D. Effect of support on the performance of Ni-based catalyst in methane dry reforming. J Fuel Chem Technol 2015;43(11):1359–65. http://dx.doi.org/10.1016/S1872-5813(15)30040-2.

[12] Schwengber CA, da Silva FA, Schaffner RA, Fernandes-Machado NRC, Ferracin RJ, Bach VR, et al. Methane dry reforming using Ni/Al2O3 catalysts: Evaluation of the effects of temperature, space velocity and reaction time. J Environ Chem Eng 2016;4(3):3688–95. http://dx.doi.org/10.1016/j.jece.2016.07.001.

[13] Ameen S, Farooq MU, Samia, Umer S, Abrar A, Hussnain S, et al. Catalyst breakthroughs in methane dry reforming: Employing machine learning for future advancements. Int J Hydrog Energy 2024. http://dx.doi.org/10.1016/j.ijhydene.2024.08.506.

[14] Kumar R, Kumar A. Recent advances of biogas reforming for hydrogen production: Methods, purification, utility and techno-economics analysis. Int J Hydrog Energy 2024;76:108–40. http://dx.doi.org/10.1016/j.ijhydene.2024.02.143.

[15] Shin S, Choi Y-Y, Sohn Y-J, Kim M, Lim IS, Oh H. Machine learning-based fault diagnosis for various steady conditions of proton exchange membrane fuel cell systems. Int J Hydrog Energy 2024;89:507–17. http://dx.doi.org/10.1016/j.ijhydene.2024.09.227.

[16] Datong L, Yu P, Xiyuan P. Online adaptive status prediction strategy for data-driven fault prognostics of complex systems. In: 2011 prognostics and system health managment confernece. 2011, p. 1–6. http://dx.doi.org/10.1109/PHM.2011.5939530.

[17] Aziz MAA, Setiabudi HD, Teh LP, Annuar NHR, Jalil AA. A review of heterogeneous catalysts for syngas production via dry reforming. J Taiwan Inst Chem Eng 2019;101:139–58. http://dx.doi.org/10.1016/j.jtice.2019.04.047.

[18] Pakhare D, Spivey J. A review of dry (CO_2) reforming of methane over noble metal catalysts. Chem Soc Rev 2014;43(22):7813–37. http://dx.doi.org/10.1039/C3CS60395D.

[19] Vellayappan K, Liu B, Susman MA, Chen X, Kawi S, Wilcox J. Impacts of catalyst and process parameters on Ni-catalyzed methane dry reforming via interpretable machine learning. Appl Catal B: Environ 2023;330:122593. http://dx.doi.org/10.1016/j.apcatb.2023.122593.

[20] Roh J, Park H, Kwon H, Joo C, Moon I, Cho H, et al. Interpretable machine learning framework for catalyst performance prediction and validation with dry reforming of methane. Appl Catal B: Environ 2024;343:123454. http://dx.doi.org/10.1016/j.apcatb.2023.123454.

[21] Bilgiç G, Bendeş E, Öztürk B, Atasever S. Recent advances in artificial neural network research for modeling hydrogen production processes. Int J Hydrog Energy 2023;48(50):18947–77. http://dx.doi.org/10.1016/j.ijhydene.2023.02.002.

[22] Nakajima EA, Oliveira LG, Gasparrini LJ, Souza GEd, Ignacio AA, Alves HJ, et al. Kinetics of dry reforming of methane catalyzed by Ni/Si-MCM-41. Int J Hydrog Energy 2023;48(67):32331–41. http://dx.doi.org/10.1016/j.ijhydene.2023.05.010.

[23] Fogler HS. Elements of chemical reaction engineering. 6th ed.. London: Pearson Education; 2010.

[24] Fogler HS. Cálculo de Reatores: O Essencial da Engenharia de Reações Químicas. 1st ed.. Rio de Janeiro: LTC; 2014.

[25] Hossain MA, Ayodele BV, Cheng CK, Khan MR. Artificial neural network modeling of hydrogen-rich syngas production from methane dry reforming over novel Ni/CaFe2O4 catalysts. Int J Hydrog Energy 2016;41(26):11119–30. http://dx.doi.org/10.1016/j.ijhydene.2016.04.034.

[26] Ahmed R, Sasaki T, Olea M. Original method to predict and monitor carbon deposition on Ni-based catalysts during dry reforming of methane. Front Chem Eng 2020;2. http://dx.doi.org/10.3389/fceng.2020.00009.

[27] Wickham H, Stryjewski L. 40 years of boxplots. Tech. rep., had.co.nz; 2012.

[28] Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika 1965;52(3/4):591–611, URL: http://www.jstor.org/stable/2333709.

[29] Box GEP, Cox DR. An analysis of transformations. J R Stat Soc Ser B Stat Methodol 1964;26(2):211–52, URL: http://www.jstor.org/stable/2984418.

[30] Kokoska S, Zwillinger D. CRC Standard probability and statistics tables and formulae, student edition. 1st ed.. CRC Press; 2000, http://dx.doi.org/10.1201/b16923.

[31] Jiang W, Liang M, Schiebel M, Zaremba S, Drechsler K. Development of machine learning based classifier for the pressure test result prediction of type IV composite overwrapped pressure vessels. Int J Hydrog Energy 2024;58:380–8. http://dx.doi.org/10.1016/j.ijhydene.2024.01.182.

[32] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. 2nd ed.. Springer; 2009, p. 27–30. http://dx.doi.org/10.1007/978-0-387-84858-7.

[33] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in python. J Mach Learn Res 2011;12:2825–30.

[34] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020;2(1):2522–5839.

[35] Powers DMW. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. 2020, arXiv:2010.16061. URL: https://arxiv.org/abs/2010.16061.