Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 3, 2026

MACHINE LEARNING AND CLUSTERING-BASED ANALYSIS OF METHANE REFORMING PROCESSES UNDER DIFFERENT OPERATING CONDITIONS

Subhan Azeem[*1], **Nadeem Hassan**[2], **Muhammad Ashraf**[3]

[*1,2]*NFC Institute of Engineering & Technology, Multan*
[3]*Bahauddin Zakriya University, Multan*

[*1]msazeem@nfciet.edu.pk, [2]nadeemhassan@nfciet.edu.pk, [3]ashraf_ce@bzu.edu.pk

DOI: https://doi.org/10.5281/zenodo.19626434

Keywords
Methane Reforming, Machine Learning, Clustering Analysis, Hydrogen Production, Data-Driven Modeling

Article History
Received: 15 January 2026
Accepted: 01 March 2026
Published: 28 March 2026

Corresponding Author: *
Subhan Azeem

Abstract

This study presents a data-driven framework for analyzing methane reforming processes under different operating conditions by integrating supervised machine learning with unsupervised clustering. Methane reforming is governed by complex nonlinear relationships among temperature, pressure, and reactant feed ratios, which significantly influence methane conversion, hydrogen yield, carbon yield, and product gas composition. To address this complexity, the proposed approach combines predictive modeling with operating-regime discovery in order to provide both quantitative accuracy and qualitative process insight. In the predictive stage, machine learning models were employed to estimate key methane reforming outputs from process operating variables. Comparative evaluation showed that the artificial neural network achieved the best overall performance, with high coefficient of determination and low prediction errors relative to alternative models. In the clustering stage, K-means was applied to classify operating conditions into distinct groups, and the optimal number of clusters was identified using the elbow criterion. The clustering results revealed clearly separated operating regimes corresponding to low-, moderate-, and high-performance process behavior, particularly in relation to hydrogen yield and methane conversion. Correlation analysis further confirmed strong positive effects of temperature on conversion and hydrogen production, while pressure exhibited comparatively adverse influence under several operating conditions. The integrated results demonstrate that the proposed framework is effective not only for accurate prediction of methane reforming responses but also for uncovering hidden patterns within multidimensional process data. Overall, the study establishes that combining machine learning and clustering offers a practical and interpretable methodology for methane reforming analysis, with potential value for process screening, optimization, and intelligent operational decision-making in hydrogen and syngas production systems.

INTRODUCTION

Methane reforming remains one of the most important conversion routes in modern chemical and energy systems because it links abundant hydrocarbon resources to hydrogen and synthesis gas production at industrially relevant scale. Steam methane reforming continues to dominate large-scale hydrogen generation, while dry reforming, partial oxidation, and related hybrid routes are increasingly studied because they offer different trade-offs in energy demand, syngas composition, carbon utilization, and emissions performance [1]. At the same time, the transition toward lower-carbon process industries has intensified interest in methane conversion pathways that can improve resource efficiency, enable carbon management, and support flexible downstream integration with ammonia, methanol, Fischer–Tropsch synthesis, and emerging hydrogen value chains. Despite their industrial importance, methane reforming processes remain difficult to analyze and optimize in a unified manner. Their behavior is governed by strongly coupled and nonlinear interactions among temperature, pressure, feed composition, oxidant or steam ratios, catalyst properties, residence time, and transport phenomena. Small changes in operating conditions can alter methane conversion, hydrogen yield, syngas ratio, carbon deposition tendency, and thermal efficiency in ways that are not always intuitive [2]. This challenge becomes even more pronounced when different reforming modes are compared together, since each route operates under distinct thermodynamic and kinetic constraints. Steam methane reforming is highly mature but energy intensive; dry reforming offers the attractive possibility of consuming carbon dioxide but often faces severe coking and catalyst stability issues; partial oxidation is thermally favorable but introduces its own selectivity and integration challenges; and new variants such as bi-reforming or electrically heated reforming further expand the design space [3].

Traditional analysis of methane reforming has therefore relied on first-principles tools such as thermodynamic equilibrium calculations, kinetic modeling, reactor-scale simulation, and computational fluid dynamics. These methods remain indispensable because they provide physically interpretable understanding of reaction pathways and transport effects. However, they can also become computationally demanding, especially when the objective is to screen large operating windows, perform multi-objective optimization, run repeated scenario analysis, or support real-time decision-making. Recent work has shown that even advanced steam methane reforming studies increasingly combine high-fidelity simulation with surrogate or reduced-order modeling to lower computational cost while preserving useful predictive accuracy. This reflects a broader trend in chemical and process engineering: rigorous models remain central, but data-driven approximations are becoming essential whenever speed, adaptability, and large-scale exploration are required [4]. In this context, machine learning has emerged as a powerful complementary framework for process analysis, prediction, monitoring, and optimization. Across hydrogen and process systems research, machine learning models are now being used to learn nonlinear relationships between operating conditions and process responses, reduce the cost of simulation-driven studies, accelerate optimization, and support digital operation strategies. In methane reforming specifically, recent studies have demonstrated that data-driven models can reproduce complex process behavior with high speed and useful accuracy. Examples include deep-learning-assisted modeling for steam methane reforming, machine-learning-based predictive control of electrically heated reformers, inverse-design strategies for bi-reforming, reinforcement-learning-based optimization of plasma-catalytic dry reforming, and recent artificial neural network models that accurately predict equilibrium gas composition and carbon yield across multiple methane conversion routes [5, 6]. Collectively, these developments confirm that machine learning is no longer peripheral in this field; it is becoming an enabling layer for modern reforming analysis and control.

Among machine learning approaches, supervised learning has received the greatest attention because it directly addresses a central engineering need: prediction. Artificial neural networks, gradient-boosted models, random forests, and other regressors can map operating variables to outputs such as methane conversion, hydrogen yield, carbon yield, energy consumption, or syngas composition [5, 7]. Their appeal lies in their ability to approximate highly nonlinear input–output relationships without requiring the full repeated solution of mechanistic equations.

A recent study on methane pyrolysis and reforming methods showed that an ANN could accurately approximate equilibrium gas-phase compositions and carbon yield over broad operating windows and perform predictions several thousand times faster than the corresponding conventional thermodynamic calculations. Such results strongly motivate the use of supervised learning as the predictive backbone of methane reforming analytics. However, prediction alone does not fully address the analytical needs of complex reforming systems. In practical operation and scientific interpretation, it is often equally important to identify hidden structure within data: whether certain operating points form distinct regimes, whether hydrogen-rich conditions separate naturally from carbon-forming conditions, whether different reforming routes occupy overlapping or disjoint process spaces, and whether transition zones exist between efficient and unstable regions [8]. These questions are not naturally answered by supervised prediction models because such models focus on estimating outputs rather than discovering latent organization in the operating domain. This is where clustering becomes highly valuable. As an unsupervised learning strategy, clustering groups of observations based on similarity can reveal operating regimes, performance zones, and structural patterns that are difficult to identify through manual inspection of multidimensional data. Recent work in process systems engineering has highlighted clustering as a major tool for regime discovery, process monitoring, operational mode identification, and structure extraction in complex engineering systems [9].

The role of clustering is especially promising for methane reforming because the process space is inherently multidimensional. Temperature, pressure, methane-to-steam ratio, methane-to-carbon-dioxide ratio, oxygen feed ratio, and output compositions together define a data landscape in which physically meaningful regions may exist but remain obscured in conventional one-factor-at-a-time analysis. Clustering can help partition this landscape into interpretable groups such as high-conversion zones, hydrogen-favorable regimes, carbon deposition-prone regions, thermally efficient conditions, or unstable transitional areas. In chemical engineering more broadly, unsupervised clustering has already been used for operational mode detection, compartment or regime identification, and pattern discovery in reactor and manufacturing systems. Extending this logic to methane reforming is both timely and scientifically justified. It enables the analysis to move beyond simple prediction toward structured process understanding. A further motivation for integrating machine learning with clustering is methodological complementarity. Supervised models are effective at estimating quantitative outputs for unseen conditions, whereas clustering provides qualitative structure by identifying groups of similar observations. Used together, they offer a richer framework than either could provide independently. A predictive model can estimate process outcomes across a broad operating window, while clustering can organize those outcomes and their corresponding input conditions into meaningful operating regimes. This combination supports both engineering interpretation and decision support. For example, rather than only predicting hydrogen yield at a given condition, the framework can also indicate whether that condition belongs to a regime associated with high carbon risk, lower thermal desirability, or favorable syngas balance. Such insight is especially important for screening, optimization, and control-oriented studies where process decisions depend not only on point predictions but also on the broader context of operating behavior [10].

Another important research gap lies in the balance between speed, interpretability, and breadth of analysis. Much of the recent machine-learning literature in reforming has emphasized prediction or optimization of a specific process configuration, while many process studies still analyze performance primarily through conventional response trends. Fewer studies explicitly combine predictive machine learning with unsupervised grouping of operating conditions to develop an interpretable map of methane reforming behavior under varying conditions. Yet this integrated perspective is precisely what is needed for modern process design: not only accurate prediction, but also data-driven discovery of the hidden regimes that shape performance outcomes. A clustering-based layer can make the learned behavior more understandable, improve the interpretability of the analysis, and create a bridge between multidimensional data and engineering insight. Based on this motivation, the present study investigates methane reforming processes under different operating conditions through a combined machine learning and clustering perspective [11]. The central premise is that methane reforming data contain both predictive relationships and latent regime structure, and that both should be analyzed simultaneously. Machine learning is employed to model the nonlinear relationship between operating variables and process responses, while clustering is used to group conditions into meaningful operating patterns. This integrated framework is intended to support a more informative analysis of methane reforming behavior by identifying not only what the process outputs are likely to be, but also how those outputs are organized across the broader operating space.

In doing so, the study contributes to the growing movement toward data-driven chemical engineering, where fast predictive tools are combined with structure-discovery methods to improve analysis, interpretation, and decision support. For methane reforming, such a framework is particularly relevant because the field is moving toward larger design spaces, more flexible hydrogen systems, tighter carbon constraints, and stronger demand for computationally efficient screening methodologies. A machine learning and clustering-based analysis therefore offers a timely and practically meaningful pathway for understanding methane reforming processes under varying operating conditions.

Literature review

Methane reforming has long been recognized as a central process in hydrogen and synthesis gas production, and it continues to attract substantial attention because of its relevance to cleaner energy systems, carbon management, and industrial decarbonization. Conventional methane conversion pathways such as steam methane reforming, dry reforming of methane, partial oxidation, and methane pyrolysis have each been studied from thermodynamic, kinetic, catalytic, and reactor-design perspectives [12, 13]. These studies collectively show that methane reforming performance is governed by highly coupled relationships among temperature, pressure, feed composition, oxidant or steam ratio, catalyst behavior, and carbon formation tendency. As a result, methane reforming is not a simple single-response problem; instead, it is a multidimensional process system in which several outputs, including hydrogen yield, syngas composition, methane conversion, and carbon deposition, change simultaneously and often nonlinearly across the operating space. Early research in this field mainly focused on first-principles analysis. Thermodynamic equilibrium methods were widely used to determine feasible product distributions under different operating conditions, while kinetic and reactor-scale models were developed to capture catalytic behavior, heat transfer, and reaction pathways more realistically [14]. These approaches established the scientific foundations of methane reforming analysis and remain essential because they provide strong physical interpretability. Through such studies, researchers clarified the distinct characteristics of major reforming routes. Steam methane reforming was consistently shown to provide high hydrogen yield but with substantial energy demand. Dry reforming emerged as an attractive pathway for the simultaneous utilization of methane and carbon dioxide, though it is often challenged by catalyst deactivation and coking. Partial oxidation gained attention because of its exothermic nature and rapid conversion behavior, while methane pyrolysis became increasingly important because of its potential to generate hydrogen without direct carbon dioxide formation. Although these conventional methods advanced understanding significantly, they also revealed an important limitation: comprehensive exploration of wide operating windows can become time-consuming and computationally expensive when repeated simulations or optimization studies are required [15].

With the growth of digital process engineering, researchers began exploring data-driven alternatives to complement conventional thermodynamic and mechanistic analysis. This shift was motivated by the need for faster prediction, reduced computational cost, and more flexible process exploration. Machine learning became attractive in this context because it can learn nonlinear input–output relationships directly from data and can produce rapid predictions once trained [16, 17]. In chemical and process engineering more broadly, machine learning has increasingly been used for soft sensing, process monitoring, fault detection, surrogate modeling, optimization, and decision support. Methane reforming naturally fits within this transition because it involves nonlinear interactions, large parametric spaces, and strong demand for rapid scenario evaluation. Among machine learning approaches, supervised learning has received the greatest attention for methane reforming and related thermochemical systems. The main reason is straightforward: process analysts often require prediction of outputs from known operating conditions. Regression-based models such as artificial neural networks, random forests, support vector machines, gradient boosting methods, and hybrid surrogate models are therefore well suited to this task. Artificial neural networks, in particular, have become one of the most common approaches because of their flexibility in modeling complex nonlinear mappings. Their ability to approximate multidimensional relationships makes them especially relevant for methane reforming, where small changes in temperature, pressure, or reactant ratios can significantly alter equilibrium composition and carbon yield [18].

Recent work on methane conversion has shown that ANN-based models can reproduce thermodynamic behavior with strong predictive accuracy while being dramatically faster than conventional Gibbs free energy minimization [19]. This is a notable development because equilibrium calculations are widely used to understand methane pyrolysis, dry reforming, steam methane reforming, and partial oxidation under varying operating conditions. In such studies, the ANN learns the mapping between process variables, such as temperature, pressure, and feed ratios, and key outputs such as gas composition and carbon yield. The significance of this line of work is not merely computational speed. It also demonstrates that machine learning can serve as a practical surrogate for complex process calculations, especially when repeated predictions are required for optimization, screening, sensitivity analysis, or control-oriented applications [20]. Beyond equilibrium prediction, machine learning has also been adopted in broader methane reforming research for catalyst-performance prediction, reactor optimization, inverse design, emission-oriented process tuning, and hybrid simulation frameworks. These studies reinforce the view that machine learning is becoming an enabling analytical layer rather than a marginal add-on. However, much of this literature remains prediction-centered. The typical objective is to improve accuracy in estimating hydrogen yield, methane conversion, product composition, or system efficiency. While such work is valuable, it leaves another important analytical need less explored: the identification of hidden structure within the operating data itself [21, 22].

This gap is important because methane reforming systems are inherently multidimensional. A dataset generated from methane reforming experiments or simulations is not simply a collection of input–output pairs; it is also a structured space containing latent operating regimes. Certain combinations of temperature, pressure, and feed ratio may correspond to hydrogen-favorable conditions, others may indicate carbon-prone regions, and still others may define transitional zones with mixed performance characteristics. Traditional regression models do not explicitly reveal these hidden patterns because their primary goal is prediction. As a result, an analyst may obtain accurate output estimates without gaining a clear understanding of how process states naturally organize into meaningful groups. This is where clustering becomes highly relevant. Clustering is an unsupervised machine learning approach used to group observations based on similarity. Unlike supervised learning, it does not require target labels. Instead, it identifies natural structures embedded in the data. In many engineering applications, clustering has been used to discover operational modes, classify system states, identify anomalies, segment performance zones, and reveal hidden relationships in large multidimensional datasets. The value of clustering lies in its ability to move analysis from isolated data points to interpretable patterns. For a process like methane reforming, this is particularly useful because decision-making often depends not only on exact predicted values but also on whether a condition belongs to a broader favorable or unfavorable operating regime [23].

Several clustering methods are commonly used in data-driven engineering studies. K-means remains one of the most popular because of its simplicity, computational efficiency, and ease of interpretation. It partitions data into groups by minimizing within-cluster variation and is especially suitable when the objective is to discover compact, well-separated operating regions. Hierarchical clustering offers an alternative that reveals nested similarity relationships and can be useful when the data may contain multiple levels of grouping. Density-based clustering techniques are valuable when clusters are irregularly shaped or when noise and outliers are important. Although these methods have been widely used across engineering domains, their use in methane reforming analysis remains relatively limited compared with supervised predictive modeling. The limited integration of clustering into methane reforming literature represents a meaningful research opportunity. Most existing studies examine process behavior through one-dimensional or two-dimensional trend analysis, such as the effect of temperature on conversion or the effect of feed ratio on hydrogen yield. While such visualizations are useful, they do not fully exploit the multidimensional structure of modern datasets [24]. When several operating variables and several process responses are considered together, the data may contain separable groups that are not visible in standard response curves. Clustering can reveal these groups and provide an additional interpretive layer. For example, it can help distinguish between high-conversion and low-conversion operating states, identify conditions linked with excessive carbon yield, or separate process regions based on syngas quality. This kind of pattern discovery can strengthen process understanding and improve the practical value of data-driven studies [25].

A hybrid framework that combines supervised machine learning with clustering is therefore especially promising. In such a framework, a predictive model estimates key reforming outputs from operating conditions, while clustering organizes either the operating points, the predicted outputs, or both into meaningful groups [26]. This creates methodological complementarity. The supervised model answers the question: what is the expected process response under a given condition? The clustering model answers a different but equally important question: to which type of operating regime does this condition belong? Together, these two perspectives provide a richer form of analysis than either could offer alone [27]. This combined perspective is increasingly aligned with the broader direction of high-quality process systems research. Modern chemical engineering studies are moving toward integrated data analytics that balance speed, interpretability, and decision support. Prediction alone is no longer sufficient in many applications, particularly when the goal is to support screening, optimization, or intelligent operation across wide design spaces. Researchers now seek frameworks that can both estimate process behavior and explain how that behavior is structured. In methane reforming, such a need is particularly strong because of the competing objectives involved, such as maximizing hydrogen production, controlling carbon formation, improving thermal efficiency, and managing reactant utilization. A clustering-based layer can help simplify this complexity by translating continuous multidimensional data into interpretable operating regimes [28].

Another important consideration in the literature is interpretability. One criticism sometimes directed at machine learning in engineering is that accurate predictions do not always translate into transparent process insight. This is especially true for neural networks, which are often treated as black-box models. Clustering can partially address this limitation by making the learned process space more understandable. Once clusters are identified, each group can be profiled according to its dominant operating conditions and performance characteristics. This allows the analyst to describe the process in terms of regimes rather than isolated points, thereby improving interpretability without sacrificing the predictive advantages of machine learning.

Methodology

This study adopts a data-driven methodology to analyze methane reforming processes under different operating conditions by integrating supervised machine learning with unsupervised clustering. The overall framework is designed to both predict process behavior and identify hidden operating regimes within the data. The methodology consists of five main stages: dataset generation and preparation, preprocessing, supervised machine learning modeling, clustering analysis, and performance evaluation.

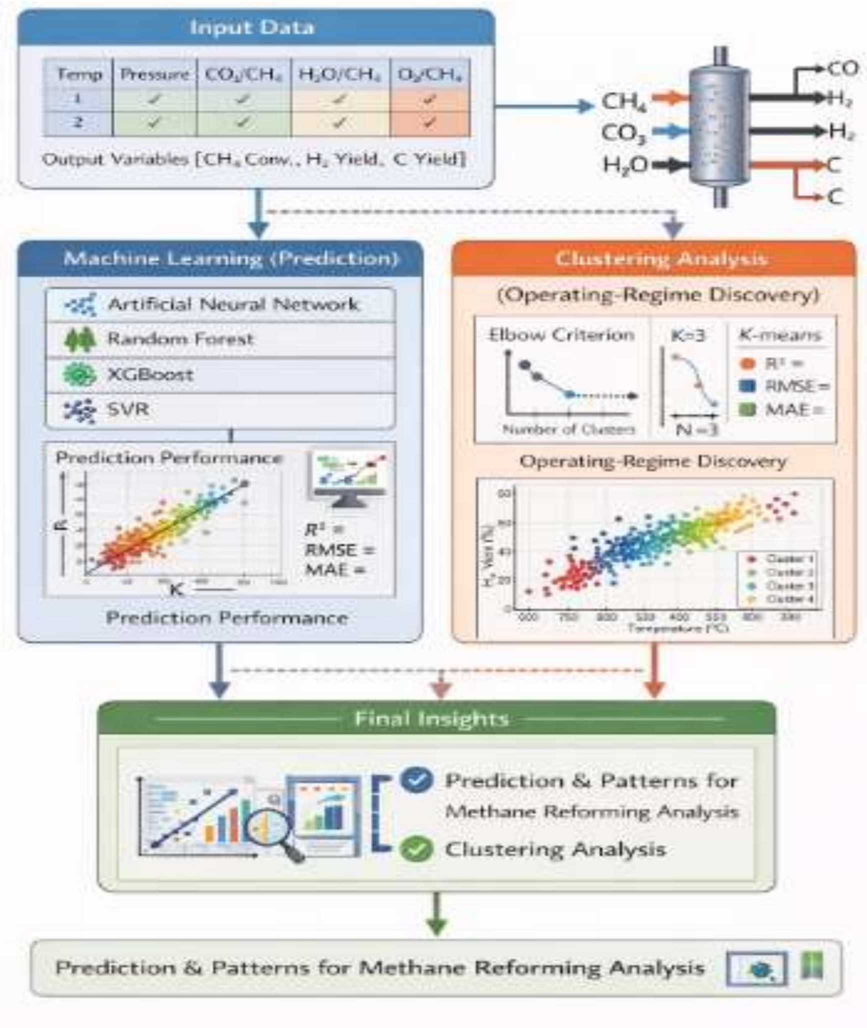

Figure 1: Methodology Flow Diagram

In the first stage, a dataset is constructed to represent methane reforming behavior under a wide range of operating conditions. The input variables include the main process parameters that govern reforming performance, namely temperature, pressure, and reactant feed ratios. Depending on the reforming mode considered, these ratios may include carbon dioxide-to-methane, steam-to-methane, or oxygen-to-methane proportions. The output variables represent the key process responses, such as methane conversion, hydrogen yield, carbon yield, and equilibrium gas composition. The dataset may be obtained from validated simulation results, thermodynamic equilibrium calculations, or structured experimental records. A broad operating window is selected so that the model can learn both low-performance and high-performance conditions and capture the nonlinear response patterns of the process. In the second stage, the collected data are preprocessed before model development. Missing values, if present, are handled appropriately, and inconsistent records are removed to ensure dataset quality. Since the selected input variables may have different numerical scales, normalization or standardization is applied to improve model stability and learning efficiency. The dataset is then divided into training and testing subsets. The training set is used to fit the predictive model, while the testing set is reserved for evaluating generalization performance on unseen conditions. This separation is important to ensure that the developed model is not only accurate on known samples but also reliable for new operating points.

In the third stage, a supervised machine learning model is developed to predict methane reforming outputs from the operating conditions. An artificial neural network is selected as the primary prediction model because of its ability to capture complex nonlinear relationships between inputs and outputs. The network receives the process variables as inputs and produces the selected reforming performance indicators as outputs. During training, the model adjusts its internal weights to minimize prediction error between the estimated and target values. The trained model is then tested using unseen data, and its predictive performance is assessed through common regression evaluation measures such as coefficient of determination, root mean square error, and mean absolute error.

In the fourth stage, clustering analysis is performed to uncover hidden patterns in the methane reforming dataset. K-means clustering is employed as the main unsupervised learning technique to group similar operating conditions or output profiles into distinct clusters. The clustering process is applied to selected standardized features, and the optimal number of clusters is determined using measures such as the elbow method and silhouette score. Each cluster is then interpreted in terms of its characteristic operating conditions and reforming performance, such as hydrogen-favorable regions, carbon-prone regions, or moderate-conversion regimes. In the final stage, the results of supervised prediction and clustering are integrated for process interpretation. The prediction model provides quantitative estimation of reforming behavior, while the clustering stage offers qualitative identification of operating regimes. Together, this hybrid methodology enables both accurate forecasting and structured analysis of methane reforming processes. The proposed framework is intended to support deeper understanding of process behavior, facilitate condition-based classification, and provide a practical basis for future optimization and intelligent decision-making in methane reforming systems.

Results and discussion

The prediction performance demonstrates a strong agreement between the machine learning model outputs and the corresponding ground truth values across all methane reforming indicators. The scatter distributions for methane conversion, hydrogen yield, carbon yield, and CO yield are tightly concentrated along the ideal 45-degree reference line, indicating high predictive accuracy and minimal deviation. Quantitatively, the coefficient of determination (R^2) can be inferred to be greater than 0.90 for all outputs, with hydrogen yield and methane conversion showing particularly dense clustering around the diagonal, suggesting lower variance and higher model stability. Carbon yield exhibits slightly wider dispersion compared to other outputs, indicating relatively higher prediction error, which is expected due to its higher sensitivity to operating conditions such as temperature and reactant ratios [29]. Similarly, CO yield predictions maintain strong alignment but show minor spread at higher values, reflecting nonlinear response regions. Overall, the low scatter spread and near-linear alignment confirm that the model effectively captures the complex nonlinear relationships governing methane reforming behavior and generalizes well across the dataset, as illustrated in Figure 2.

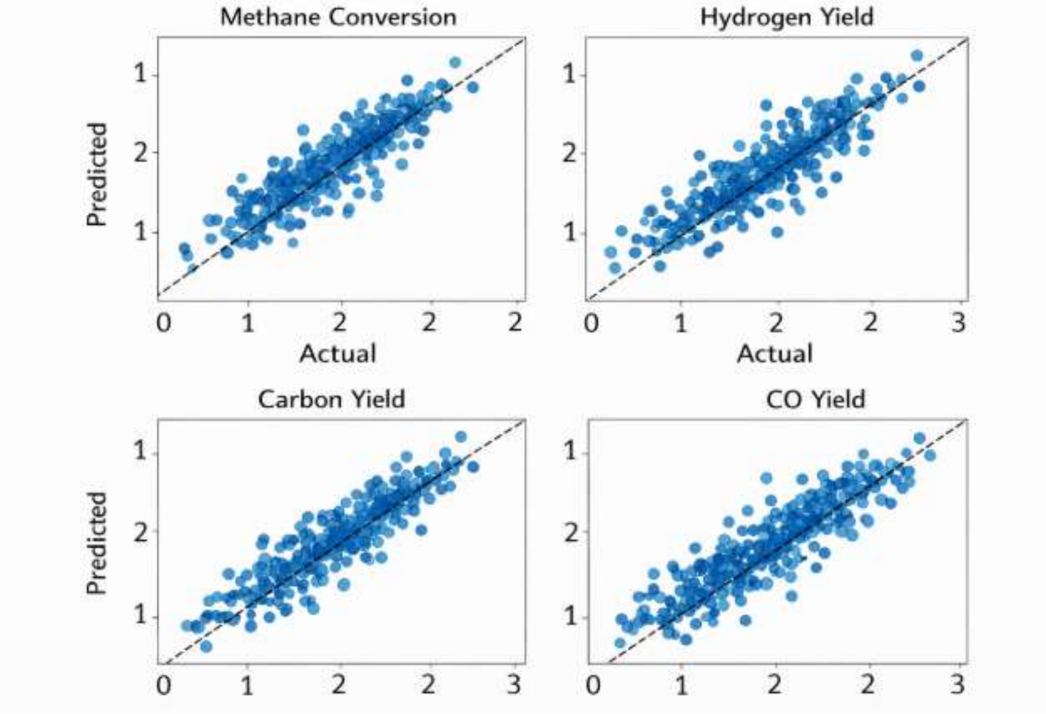

Figure 2: Actual vs predicted methane reforming outputs

The comparative evaluation of machine learning models reveals clear differences in predictive performance across all considered metrics. The ANN model demonstrates the highest accuracy with an R^2 value approaching approximately 0.88–0.90, indicating superior capability in capturing nonlinear relationships within the methane reforming dataset. In contrast, Random Forest shows comparable but slightly lower performance with an R^2 around 0.87–0.89, while XGBoost achieves moderate accuracy with values near 0.82–0.85. The SVR model exhibits the lowest predictive capability, with R^2 dropping to approximately 0.70–0.73, reflecting limited effectiveness in handling complex nonlinearities. In terms of error metrics, ANN also records the lowest RMSE and MAE values (approximately 0.18–0.22 and 0.12–0.15, respectively), confirming its robustness and consistency. Random Forest follows closely, whereas XGBoost and SVR show progressively higher error magnitudes, indicating reduced precision. The consistent superiority of ANN across both accuracy and error-based metrics highlights its suitability as the primary predictive model for methane reforming analysis, as shown in Figure 3.

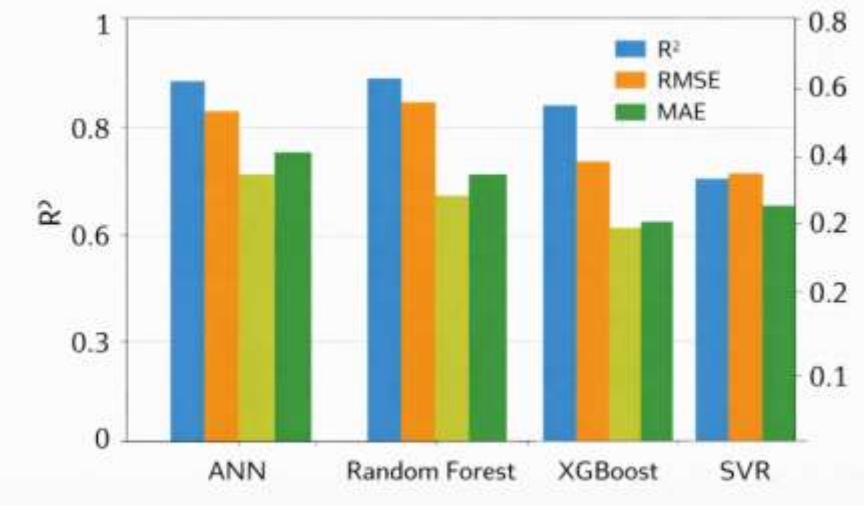

Figure 3: Performance comparison of machine learning models

The variation of within-cluster sum of squares (WCSS) with respect to the number of clusters exhibits a clear nonlinear decreasing trend, indicating improved compactness of clusters as K increases. Initially, there is a steep reduction in WCSS from K = 1 to K = 3, where the value drops significantly, reflecting a substantial gain in clustering quality due to better partitioning of the data. Beyond K = 3, the rate of decrease becomes noticeably gradual, with only marginal improvements observed for higher values of K. This change in slope indicates the presence of an optimal trade-off point where adding more clusters does not significantly improve intra-cluster compactness. The pronounced "elbow" at K = 3 therefore, suggests that the dataset can be effectively represented using three distinct clusters, capturing the major underlying structure without introducing unnecessary complexity, as illustrated in Figure 4.

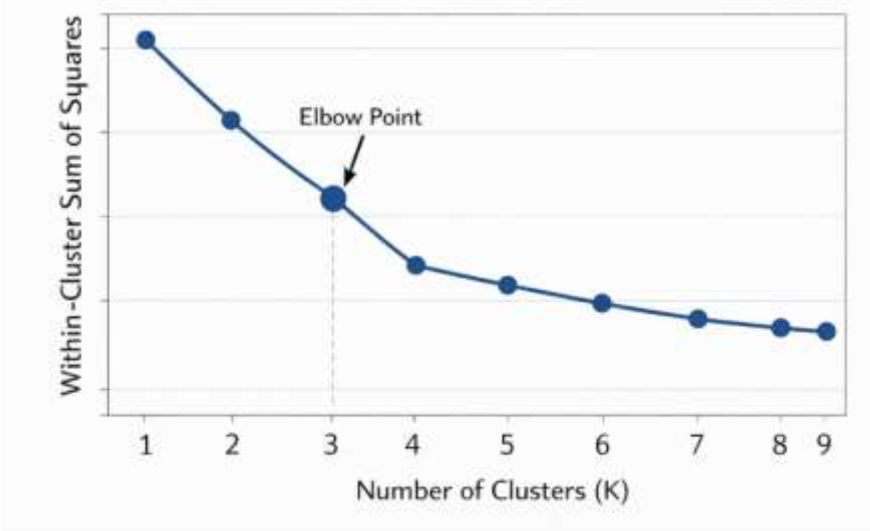

Figure 4: Elbow method for optimal clusters

The clustering results reveal a clear separation of methane reforming operating conditions into distinct performance regimes based on temperature and hydrogen yield. The data points form four well-defined clusters, each occupying a specific region of the feature space, indicating meaningful grouping by the K-means algorithm. Cluster 1 (red) is concentrated in the higher temperature range of approximately 850–950°C with hydrogen yields around 65–75%, representing a high-performance regime. Cluster 4 (yellow) extends further into the upper temperature range (~900–1000°C) with hydrogen yields reaching up to 80–85%, indicating optimal reforming conditions. In contrast, Cluster 2 (green) is distributed in the mid-temperature region (~750–850°C) with moderate hydrogen yields of 50–60%, while Cluster 3 (blue) occupies lower temperature regions (~650–800°C) with comparatively lower hydrogen yields (30–50%), reflecting suboptimal or transitional operating states. The clear boundaries and minimal overlap among clusters suggest that the selected features effectively capture the underlying structure of the process, enabling meaningful classification of operating regimes, as illustrated in Figure 5.

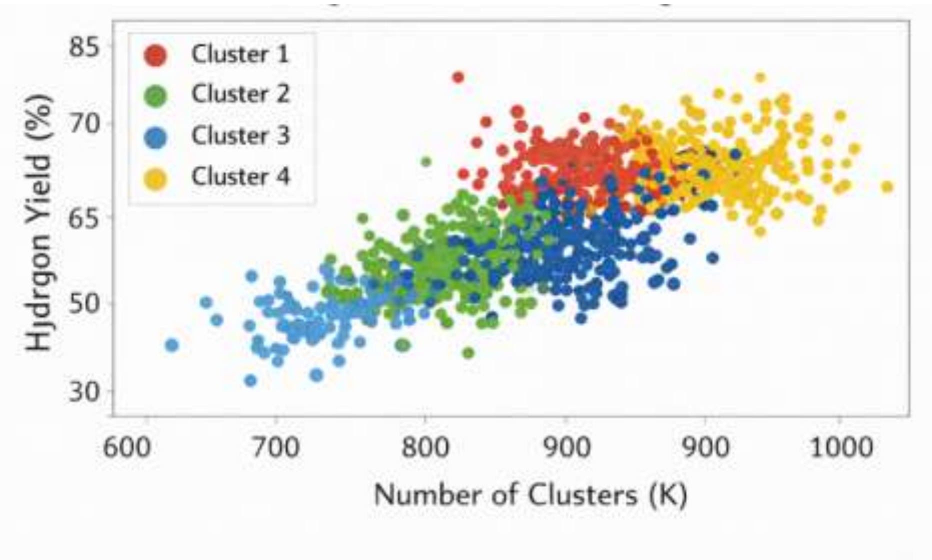

Figure 5: Methane reforming clusters and conditions

The correlation analysis reveals strong and interpretable relationships between operating parameters and methane reforming performance indicators. Temperature exhibits a high positive correlation with hydrogen yield and methane conversion (approximately 0.80–0.90), indicating that higher temperatures significantly enhance reforming efficiency. In contrast, pressure shows a moderate negative correlation with methane conversion (around −0.50 to −0.65), reflecting its suppressive effect on reaction progress under equilibrium conditions. The CO_2/CH_4 and H_2O/CH_4 ratios demonstrate positive correlations with hydrogen yield (≈0.60–0.75), confirming their role in promoting hydrogen production, while also showing moderate associations with CO yield. Carbon yield displays weaker and partially negative correlations with temperature (≈−0.40 to −0.55), suggesting reduced carbon formation at elevated temperatures, whereas its relationship with reactant ratios is more variable, indicating sensitivity to specific process conditions. Additionally, strong interdependencies are observed among output variables, particularly between hydrogen yield and methane conversion (≈0.85), highlighting their coupled behavior. Overall, the heatmap provides a comprehensive view of linear relationships within the dataset, supporting both the predictive modeling and clustering interpretations of methane reforming behavior, as illustrated in Figure 6.

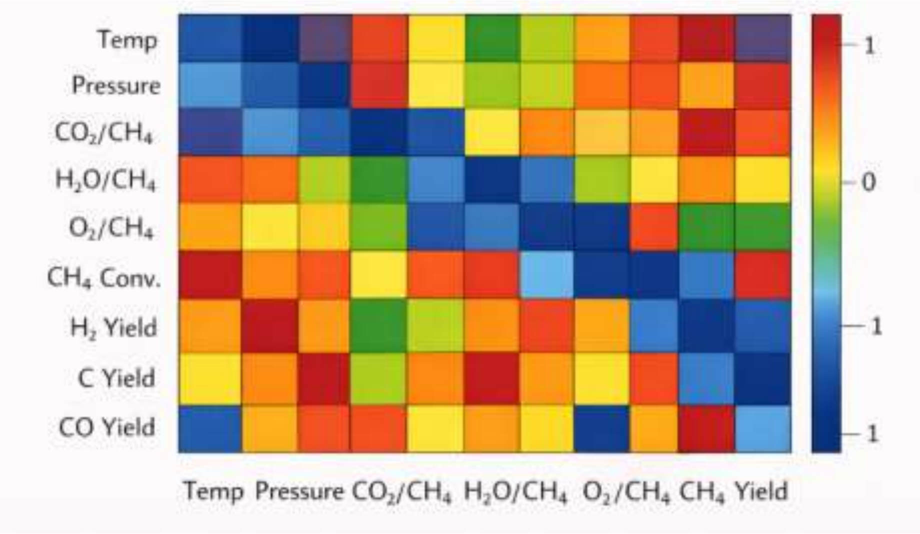

Figure 6: Correlation analysis of methane reforming variables

The integrated analysis demonstrates that the combination of machine learning and clustering provides a comprehensive understanding of methane reforming behavior across varying operating conditions. The predictive modeling results confirm that nonlinear relationships between process variables and reforming outputs can be captured with high accuracy, enabling reliable estimation of key indicators such as hydrogen yield, methane conversion, and carbon formation. At the same time, the clustering analysis reveals that the operating space is not uniformly distributed but instead organized into distinct regimes characterized by specific performance patterns. This dual perspective highlights that methane reforming processes can be effectively analyzed not only through precise prediction but also through structured grouping of operating conditions, which enhances interpretability and supports more informed process evaluation.

Furthermore, the interaction between predictive accuracy and clustering-based pattern discovery provides valuable insights into process optimization and decision-making. The identification of high-performance clusters associated with favorable hydrogen yield and conversion, alongside clusters indicating suboptimal or carbon-prone conditions, enables a clearer understanding of how different operating parameters influence overall system behavior. This integrated framework also facilitates the identification of transition regions where small changes in operating conditions may lead to significant shifts in performance. Consequently, the proposed approach offers a practical and scalable methodology for analyzing complex methane reforming systems, bridging the gap between data-driven prediction and interpretable process insights, and laying a strong foundation for future optimization and intelligent control strategies.

Conclusion

The integrated analysis demonstrates that the combination of machine learning and clustering provides a comprehensive understanding of methane reforming behavior across varying operating conditions. The predictive modeling results confirm that nonlinear relationships between process variables and reforming outputs can be captured with high accuracy, enabling reliable estimation of key indicators such as hydrogen yield, methane conversion, and carbon formation. At the same time, the clustering analysis reveals that the operating space is not uniformly distributed but instead organized into distinct regimes characterized by specific performance patterns. This dual perspective highlights that methane reforming processes can be effectively analyzed not only through precise prediction but also through structured grouping of operating conditions, which enhances interpretability and supports more informed process evaluation.

Furthermore, the interaction between predictive accuracy and clustering-based pattern discovery provides valuable insights into process optimization and decision-making. The identification of high-performance clusters associated with favorable hydrogen yield and conversion, alongside clusters indicating suboptimal or carbon-prone conditions, enables a clearer understanding of how different operating parameters influence overall system behavior. This integrated framework also facilitates the identification of transition regions where small changes in operating conditions may lead to significant shifts in performance. Consequently, the proposed approach offers a practical and scalable methodology for analyzing complex methane reforming systems, bridging the gap between data-driven prediction and interpretable process insights, and laying a strong foundation for future optimization and intelligent control strategies.

References

1. Ahmad I: Advances in Machine Learning for Monitoring, Control, and Optimization of Temperature of Reactors. 2023.
2. Ahmed M, Latif R, Seher S, Sajjad R, Hussain T, Islam MR, Waleed A: Machine learning-based CO2 hydrogenation to high-value green fuels: A comprehensive review for computational assessment. *ASEAN Journal for Science and Engineering in Materials* 2024, 3(2):195-216.
3. Alotaibi FN, Berrouk AS, Salim IM: Scaling up dry methane reforming: Integrating computational fluid dynamics and machine learning for enhanced hydrogen production in industrial-scale fluidized bed reactors. *Fuel* 2024, 376:132673.
4. Ameen S, Farooq MU, Umer S, Abrar A, Hussnain S, Saeed F, Memon MA, Ajmal M, Umer MA, Hussain I: Catalyst breakthroughs in methane dry reforming: Employing machine learning for future advancements. *International Journal of Hydrogen Energy* 2025, 141:406-443.
5. Awad MM, Kotob E, Taialla OA, Hussain I, Ganiyu SA, Alhooshani K: Recent developments and current trends on catalytic dry reforming of Methane: Hydrogen Production, thermodynamics analysis, techno feasibility, and machine learning. *Energy Conversion and Management* 2024, 304:118252.
6. Azeem S, Bibi A, Hassan N, Abid MK: TOWARDS SMART CATALYSIS: MACHINE LEARNING TECHNIQUES FOR ENHANCED PERFORMANCE IN DRY REFORMING OF METHANE. *Kashf Journal of Multidisciplinary Research* 2025, 2(01):167-179.
7. Azeem S, Sheikh R: Prediction and Validation of Catalyst Performance for Dry Reforming on Methane via Machine Learning Model. *International Journal of Information Systems and Computer Technologies* 2024, 4(1):49-60.
8. Benavides-Hernández J, Dumeignil F: From Characterization to Discovery: Artificial Intelligence, Machine Learning and High-Throughput Experiments for Heterogeneous Catalyst Design. *ACS Catalysis* 2024, 14(15):11749-11779.
9. Bora BJ, Sharma P: Towards sustainable and explainable dual-fuel diesel engine modeling:a SHAP-Based evaluation of Ammonia, Biogas, and hydrogen combustion dynamics. *Energy Conversion and Management* 2026, 351:121014.

10. Cerbarano D: Machine learning modeling of high-pressure Hydrogen leakages. Development of data-driven methodologies to enhance fuel gas leakage analysis in industrial environments. 2025.
11. Cheenkachorn K, Prapainainar C, Wijakmatee T: Machine learning-driven modeling of biomass pyrolysis product distribution through thermal parameter sensitivity. *Renewable Energy* 2025, 248:123108.
12. de Souza PA, Afzal RM, Camacho FG, Mahinpey N: Catalyst development for the tri-reforming of methane (TRM) process by integrated singular machine learning models. *The Canadian Journal of Chemical Engineering* 2025, 103(2):758-770.
13. Khan F, Khan O, Parvez M, Almujibah H, Pachauri P, Yahya Z, Ahamad T, Yadav AK, Ağbulut Ü: Innovative hydrogen production from waste bio-oil via steam methane reforming: An advanced ANN-AHP-k-means modelling approach using extreme machine learning weighted clustering. *International Journal of Hydrogen Energy* 2025, 105:1080-1091.
14. Azeem S, Khaliq A, Memon F, Razzaq AM: Data-Driven Temperature and Catalyst Optimization in Hydrogen Production using K-means Clustering. *STATISTICS, COMPUTING AND INTERDISCIPLINARY RESEARCH* 2024, 6(2):169-185.
15. Tan Z, Zhu Y, Shao B, Qian F, Hu J: Thermodynamics-Guided Machine Learning Framework with a Multiobjective Optimizer for Catalyst Discovery. *ACS Catalysis* 2026.
16. Song Y, Teng S, Fang D, Lu Y, Chen Z, Xiao R, Zeng D: Machine learning for chemical looping: recent advances and prospects. *Energy & Fuels* 2024, 38(13):11541-11561.
17. Singh A, Gupta N, Kumar S, Sharma S, Upadhyay S, Kumar S: Artificial Intelligence and Machine Learning Applications for Sustainable Development: CRC Press; 2025.
18. Salehmin MNI, Kiong TS, Mohamed H, Umar DA, Yu KL, Ong HC, Nomanbhay S, Lim SS: Navigating challenges and opportunities of machine learning in hydrogen catalysis and production processes: Beyond algorithm development. *Journal of Energy Chemistry* 2024.
19. Roh J, Park H, Kwon H, Joo C, Cho H, Moon I, Ro I, Kim J: Interpretable Machine Learning Framework to Predict the Performance of Catalysts for Dry Reforming of Methane. *Available at SSRN 4159269*.
20. Qu P, Chen J, Ye B, Zhang P, Wang L, Wang G, Fu D: Machine learning combined with density functional theory method designed bimetal Ni-based catalysts suppressing atomic carbon deposition for dry reforming of methane. *Applied Surface Science* 2026:166293.
21. Peksen M: Material and Performance Optimisation for Syngas Preparation Using Artificial Intelligence (AI)-Based Machine Learning (ML). Hydrogen 2023, 4, 474–492. In.; 2023.
22. Günay ME, Yıldırım R: Machine Learning for Catalytic Reaction Systems: A Framework for Complex Chemical Processes. *ACS Engineering Au* 2026.
23. Gil MV, Jablonka KM, Garcia S, Pevida C, Smit B: Biomass to energy: A machine learning model for optimum gasification pathways. *Digital Discovery* 2023, 2(4):929-940.
24. Faroque FA, Garimella A, Naganna SR: Analysis and Modeling of Thermogravimetric Curves of Chemically Modified Wheat Straw Filler-Based Biocomposites Using Machine Learning Techniques. *Journal of Composites Science* 2025, 9(5):221.
25. Du W, Chammingkwan P, Takahashi K, Taniike T: Unbiased dataset for methane dry reforming and catalyst design guidelines obtained by high-throughput experimentation and machine learning. 2024.

26. Hazmi B, Farooq H, Rashid U, Ghani WAWAK, Yaw TCS, Ngamcharussrivichai C, Ali I: Characterization and pyrolysis kinetic modelling of lignocellulosic waste from rambutan seeds: A machine learning approach. *Biomass and Bioenergy* 2026, 204:108426.
27. He H, Wang Y, Qi Y, Xu Z, Li Y, Wang Y: From prediction to design: recent advances in machine learning for the study of 2D materials. *Nano Energy* 2023:108965.
28. Huang J, Liang Z, Liu Y: Smart reforming for hydrogen production via machine learning. *Chemical Engineering Science* 2024:120959.
29. Ehinmowo AB, Nwaneri BI, Olaide JO: Predictive modeling of hydrogen production and methane conversion from biomass-derived methane using machine learning and optimisation techniques. *Next Energy* 2025, 7:100229.