Iranian Journal of Chemical Engineering

Journal Homepage: www.ijche.com

pISSN: 1735-5397
eISSN: 2008-2355

Application of Artificial Neural Networks and Machine Learning in the Production of Syngas via Hydrocarbon Reforming

Mohadeseh Zareei [1], Jamshid Khorshidi [2*], Hamed Hemmati [1]

[1] Faculty of Chemical and Petroleum Engineering, University of Hormozgan, Bandar Abbas, Iran
[2] Department of Mechanical Engineering, Faculty of Engineering, University of Hormozgan, Bandar Abbas, Iran

ARTICLE INFO

Article type:
Review article

Article history:
Received: 2026-08-29
Revised: 2026-09-08
Accepted: 2026-09-24
Available online: 2026-09-26

Keywords:
Machine learning,
Artificial neural networks,
Hydrocarbon reforming,
Dry methane reforming,
Syngas production

ABSTRACT

Hydrocarbon reforming processes are important routes for the production of synthesis gas and hydrogen. However, the complex interactions among operating conditions, reaction kinetics, transport phenomena, and catalyst properties make process prediction challenging. This review examines the application of machine learning (ML) and artificial neural networks (ANNs) in the modeling, prediction, and optimization of hydrocarbon reforming processes, with emphasis on methane reforming. Studies employing ANN, Random Forest, Support Vector Machine, Gradient Boosting, XGBoost, and CatBoost were reviewed and compared based on the statistical metrics, validation strategies, size and quality of datasets, overfitting, and generalizability of the model. The results indicate that the performance of the model depends on the reforming process, target output, data structure, and validation strategy, and no single algorithm is universally superior. Although MLP has shown better statistical performance than RBF in some studies on dry methane reforming, independent test datasets may reveal limitations in the generalizability of ANN. Ensemble methods, including Random Forest, CatBoost, and Gradient Boosting, have demonstrated competitive performance, although the best-performing model varies depending on the predicted output. Furthermore, SHAP and Partial Dependence Plots can identify influential factors such as the reaction temperature, feed ratio, and metal loading. Overall, rigorous validation, independent testing, overfitting control, and interpretability are essential for the reliable ML application in reforming processes.

DOI: 10.22034/ijche.2026.599855.1612 URL: https://www.ijche.com/article_253854.html

[*]Corresponding author: khorshidi@hormozgan.ac.ir

1. Introduction

Syngas, mainly composed of hydrogen and carbon monoxide, serves as a key intermediate in numerous chemical processes and can be generated from a wide range of feedstocks, such as natural gas [1]. One of the major routes for syngas production is hydrocarbon reforming, which can be carried out through different pathways depending on the reactant employed and the underlying chemical route. Several processes have been developed for the production of syngas from natural gas, including steam methane reforming (SMR), dry methane reforming (DRM), partial oxidation (POX), autothermal reforming (ATR), and tri-reforming of methane (TMR) [1].

In steam reforming, hydrocarbons are catalytically converted into syngas in the presence of steam through an endothermic process. In the case of methane, the resulting syngas is characterized by a relatively high H_2/CO ratio [2]. In contrast, dry reforming uses CO_2 and CH_4 as reactants, resulting in syngas with an H_2/CO ratio close to one. The partial oxidation of methane uses oxygen as the oxidizing agent and is an exothermic process, unlike steam and dry reforming, which are endothermic processes [3]. By combining the endothermic reforming reactions with the exothermic partial oxidation reaction, autothermal reforming allows part of the energy required for the process to be supplied by oxidation reactions [2, 3]. Tri-reforming, also referred to as methane tri-reforming, has also been proposed through the integration of steam reforming, dry reforming, and partial oxidation pathways. This process allows the production of syngas with an H_2/CO ratio ranging from 1.5 to 2 by adjusting the reactant composition.[1]. Therefore, different reforming pathways exhibit distinct thermal characteristics, reactant requirements, and syngas compositions. Accordingly, the choice of a specific reforming route is primarily dictated by the target syngas composition and H_2/CO ratio required for the intended downstream process.[1].

Why is ML particularly relevant to reforming? The strong coupling of the kinetics, thermodynamic equilibrium, heat and mass transfer, catalyst deactivation, and multiple operational variables of the reaction makes it difficult to represent reforming systems with a single simple empirical equation. ML methods are therefore relevant because they can learn nonlinear input–output relationships from experimental, operational, or simulation data. Their usefulness, however, depends on whether the available data adequately represent the operational domain and whether the model is validated against observations that were not used during the development of the model.

The importance of reforming processes is largely associated with the role of the resulting syngas as a valuable intermediate in the chemical and energy industries. Primarily consisting of hydrogen and carbon monoxide, syngas can serve as a feedstock for producing a wide range of fuels and chemicals, including olefins, gasoline, oxygenates, aldehydes, and alcohols. In addition, various products can be obtained from syngas through processes such as Fischer–Tropsch synthesis, methanol-to-gasoline, and oxo synthesis [3]. In addition, hydrogen and syngas are essential intermediates in major industrial processes, including the synthesis of ammonia and methanol, the production of synthetic fuels and oils via the Fischer–Tropsch process, and the hydrogenation reactions in the chemical and petrochemical industries [4]. However, the utility of syngas is not limited to its production volume; rather its composition, particularly the H_2/CO ratio, plays a crucial role in determining its potential applications. Therefore, the selection of the reforming

process should be consistent with the required specifications of the syngas [5]. For example, steam reforming, owing to its production of hydrogen-rich syngas, is suitable for applications where higher hydrogen production is desired, including ammonia production. Autothermal reforming and partial oxidation, in contrast, are suitable for the production of syngas for the synthesis of liquid hydrocarbon fuel because the H_2/CO ratio can be adjusted based on the requirements of downstream processes [2]. Furthermore, dry reforming and tri-reforming, which employ CO_2 as one of the feedstocks, provide an opportunity to convert this compound into syngas and subsequently into value-added products [1]. Consequently, controlling the yield and composition of the resulting syngas is of particular importance from an industrial perspective, highlighting the need for a thorough understanding of the factors governing the performance of these reforming processes.

Table 1.
Comparison of different hydrocarbon reforming methods for the production of syngas

Reforming Method	Main Reactants	Thermal Nature	H_2/CO Ratio	Typical Industrial Application
Steam Methane Reforming (SMR)	Steam + Methane (H_2O + CH_4)	Endothermic	Relatively high	Hydrogen-rich production; suitable for ammonia synthesis
Dry Reforming of Methane (DRM)	Carbon Dioxide + Methane (CO_2 + CH_4)	Endothermic	Close to 1	CO_2 utilization; lower H_2/CO ratio syngas
Partial Oxidation (POX)	Oxygen + Methane (O_2 + CH_4)	Exothermic	Typically, around 2 for ideal CH_4 partial oxidation; actual values vary with operating conditions and feed composition.	—
Autothermal Reforming (ATR)	Combination of steam reforming and partial oxidation	Combined (endothermic & exothermic)	Adjustable based on process needs	Liquid hydrocarbon fuels
Tri-Reforming (TMR)	Combination of steam reforming, dry reforming, and partial oxidation	Combined	About 1.5 to 2	Precise tuning of syngas ratio

Despite the widespread application of reforming processes, predicting their behavior remains challenging because of the complex and interdependent nature of the underlying physical and chemical phenomena. The behavior of an actual reactor is determined not only by thermodynamic equilibrium but also by the chemical reaction kinetics, heat transfer, mass transfer, and species diffusion within the bed and catalyst particles, all of which significantly influence the performance of the reactor. In particular, in steam methane reforming, the interactions among reaction kinetics, multicomponent diffusion, and various heat-transfer mechanisms complicate accurate process modeling [6]. The simultaneous interaction of these phenomena means that changes in operating conditions can significantly affect the behavior of the reactor and product composition, highlighting the

need for reliable models to describe and predict the performance of the system [7]. Reforming process design, analysis, and optimization have therefore relied on modeling and simulation tools, which make it possible to examine how operational and design variables influence the performance of the reactor [6]. Equilibrium models, kinetic models, and numerical approaches such as computational fluid dynamics (CFD) are among the tools available for modeling reforming processes, each capturing a distinct aspect of the behavior of the process [8]. Kinetic models, by accounting for chemical reactions and transport resistances, can predict process conditions and product composition, whereas numerical modeling enables the inclusion of details such as reactor geometry, flow hydrodynamics, and transport mechanisms to be incorporated into the analysis [9]. Modeling therefore offers a means of examining the temperature and flow distributions, reaction rates, and reactant conversion, along with demonstrating how changes in variables such as the temperature, pressure, feed composition, and reactor characteristics affect the process — providing a foundation for parametric studies and the identification of favorable operating conditions [8]. The development of accurate models can consequently improve the ability to predict the behavior of the reactor and facilitate the design, control, and optimization of the process, as well as the assessment of suitable conditions for the production of hydrogen and syngas [7].

Another key factor determining reforming performance is the behavior of the catalyst and its interaction with operating conditions. In dry methane reforming, the high temperatures required for the reaction can promote phenomena such as carbon deposition, metal particle sintering, and catalyst poisoning, ultimately leading to reductions in the activity and stability of the catalyst [10]. Similarly, catalyst deactivation is a major operational challenge in steam methane reforming and may be accompanied by the decreased hydrogen yield and conversion, increased energy consumption, a greater need for maintenance, and disruptions to the continuous operation of the unit [11]. The performance of the process, however, is not determined solely by the nature of the catalyst, as a combination of catalyst properties and operating conditions can simultaneously influence the reactant conversion and product composition. Variables such as the reaction temperature, gas hourly space velocity, reaction time, nickel loading, and catalyst structural characteristics have been employed to predict the conversion of methane and carbon dioxide and the H_2/CO ratio [12]. On the other hand, the wide diversity in catalyst compositions, structures, and operating conditions makes it difficult to evaluate all possible options solely through conventional experimental methods based on trial-and-error. Consequently, computational approaches and predictive methods have attracted attention for screening potential options and predicting the performance of catalysts [13]. The integration of machine learning methods with kinetic models and experimental data has also emerged as a promising approach for predicting performance-related properties and accelerating the evaluation and selection of catalysts [13]. Therefore, the complexity of the relationships among the characteristics of catalysts, operating conditions, and process outputs increases the need for methods capable of extracting these multivariable and nonlinear relationships from available data and predicting the performance of the system with adequate accuracy.

In this context, the application of experimental approaches and machine learning algorithms for the modeling, prediction, and optimization of reforming processes has expanded in recent

years. These methods are particularly useful for systems, of which the behavior is simultaneously influenced by multiple operational variables and nonlinear relationships between inputs and outputs, as they enable the extraction of patterns from data and the development of predictive models. In dry methane reforming, various regression models have been employed to simultaneously predict the conversion rates of methane and carbon dioxide, as well as the yields of hydrogen and carbon monoxide. The results of these studies have demonstrated the influence of variables such as the reaction temperature, CH_4/CO_2 feed ratio, and metal loading on the performance of the process [14]. In steam methane reforming, the integration of computational fluid dynamics simulations with deep neural networks has been employed to predict output variables and investigate the effects of parameters such as the heat flux, reformer length, steam-to-carbon ratio, and flow rate. In one study, the developed model achieved an accuracy exceeding 99% for both training and testing datasets and was subsequently used to determine the optimal conditions for hydrogen production [15]. Beyond offering predictive capability, the development of explainable machine learning methods has enabled the investigation of the relative importance of variables and their effects on the performance of the process. For example, in a biogas tri-reforming study employing a Random Forest model and the SHAP analysis, the reaction temperature was identified as one of the important variables affecting methane conversion [16]. Moreover, combining thermodynamic models and process simulations with machine learning algorithms can facilitate the prediction of hydrogen production while revealing the individual and interactive effects of operational parameters, thereby supporting process optimization [17]. Thus, data-driven models can be employed alongside experimental investigations and conventional modeling approaches to predict the behavior of the process, evaluate the effects of process variables, and identify favorable operating conditions.

Among data-driven approaches, artificial neural networks (ANNs) have attracted considerable attention for modeling complex systems because of their ability to approximate nonlinear relationships between input and output variables [18]. These networks consist of interconnected neurons whose connections are characterized by synaptic weights. Neurons, weighted connections, activation functions, and learning algorithms are considered the principal components of artificial neural networks [19]. In a conventional feed-forward architecture, input data are transferred to the neurons of the hidden layer after being multiplied by the corresponding weights. After being processed by a nonlinear activation function, the resulting information is transmitted to the output layer through weighted connections [20]. Training networks is based on adjusting their parameters to minimize the difference between predicted and target values. For this purpose, an error function such as the mean squared error (MSE) can be minimized, while network weights are adjusted using the backpropagation algorithm and gradient-based methods [18]. Once trained, the network can establish a mapping between sets of input and output variables and approximate nonlinear relationships that are difficult to express using conventional explicit mathematical equations [21]. However, the predictive capability of such models depends on the data used for training, and the reliability of the model should therefore be evaluated within the range of data and conditions under investigation [21]. This ability to learn multivariable and nonlinear relationships has led to the application of

neural networks in modeling engineering processes, specifically in studies related to methane reforming.

Modern gradient-based optimizers can also be used during ANN training. In addition to conventional backpropagation, optimization algorithms such as Adam adapt the learning rate during weight updates and are widely used for efficient gradient-based training. Adam can therefore be used as the optimizer within a backpropagation-based training procedure.

Terminology. In this review, artificial intelligence (AI) is treated as the broader field, machine learning (ML) as a subset of AI based on data-driven learning, and artificial neural networks (ANNs) as one family of ML models. Deep learning is reserved for neural-network models with multiple hidden layers. Accordingly, AI, ML, ANN, and deep learning are not used interchangeably.

In recent years, the application of artificial neural networks has been extended to the modeling of methane thermochemical conversion processes, including steam reforming, dry reforming, pyrolysis, and partial oxidation. The performance of these processes is governed by interactions among thermodynamics, reaction kinetics, and transport phenomena [22]. Reported studies on modelling thermodynamic equilibrium, reaction kinetics, and reactor performance have employed architectures such as multilayer perceptions (MLPs) and radial basis function (RBF) networks, demonstrating the capability of these models to predict the behavior of the process [22]. In a study on steam methane reforming, a deep neural network was developed using a dataset comprising the experimental data from kinetic and equilibrium regimes, interpolated data, and the theoretical data generated from mathematical models. This approach was intended to reduce the limitations of models covering only one of the kinetic or equilibrium regimes. The optimized model achieved a Pearson correlation coefficient of 0.963, a mean squared error of 0.000840, and a coefficient of determination (R^2) of 0.921 for predicting the composition of the post-reaction mixture [23]. In another study, an ANN-based surrogate model was developed to reduce the computational cost of a detailed mechanistic model of a steam methane reforming reactor. The surrogate model predicted the behavior of the reactor with satisfactory accuracy while reducing the average simulation time by 93.8% [24]. The model was subsequently incorporated into a multi-objective optimization framework aimed at increasing methane conversion and hydrogen production while reducing carbon dioxide production, and a multi-criteria decision-making approach was employed to identify suitable operating points [24]. Related studies have also reported the application of ANNs in dry methane reforming and the development of data-driven models to predict products and the performance of the process [22, 25]. The literature includes independent studies on improving steam reforming kinetic models using ANNs, predicting products from sorption-enhanced steam reforming, ANN-based modeling of dry methane reforming, and integrating experimental and numerical data for reforming simulations. Overall, these studies demonstrate that neural networks can be used to establish relationships between operating conditions and process responses, develop computationally efficient surrogate models, and support prediction and optimization frameworks for reforming processes [22-24]. Accordingly, the application of artificial neural networks provides an appropriate framework for investigating the relationships between influential reforming parameters and developing predictive models in the present review[25].

Despite the extensive development of machine

learning and neural network methods for reforming process modeling, comparisons among studies indicate that the performance of these methods is strongly dependent on data characteristics, model architecture, input variables, and validation strategy[14]. In early studies, neural networks were primarily employed to approximate nonlinear relationships between operating conditions and process outputs. For example, in dry methane reforming, an MLP model demonstrated greater accuracy than an RBF model in predicting the yields of H_2 and CO and reactant conversion, highlighting the importance of network architecture in predictive accuracy[14]. More recent studies have employed multiple algorithms, including Random Forest, CatBoost, Gradient Boosting, SVR, and MLP, enabling more comprehensive comparisons of the performance od models. However, even models exhibiting high R^2 values do not necessarily possess equivalent generalization capabilities, because factors such as the limited dataset size, unbalanced coverage of the variable space, and potential overfitting can affect predictive reliability[13, 14]. For example, in a study on dry reforming, 15 regression models were compared using only 27 experiments. CatBoost and Random Forest were among the best-performing models; nevertheless, the study emphasized the importance of parameter tuning and the use of multiple evaluation metrics to control overfitting[14].

On the other hand, the transition from purely predictive models toward interpretable methods and hybrid frameworks has enabled more in-depth analysis of the behavior of the process. Studies based on Random Forest and SHAP have investigated not only the performance of the process but also the effects of variables such as the reaction temperature, feed composition, and catalyst characteristics on process outputs[26, 27]. However, a bias in literature-derived datasets toward high-performing catalysts may limit the generalizability and interpretability of models in regions with limited information. In addition, recent studies have employed neural networks as surrogate models in combination with CFD, genetic algorithms, multi-objective optimization, and reinforcement learning to reduce computational costs while enabling the determination of optimal process conditions [24]. For example, an ANN-based surrogate model for steam methane reforming reduced the average simulation time by 93.8%. Therefore, the evaluation of the machine learning methods for reforming processes should not be based solely on the highest R^2 value. Instead, metrics such as MSE, RMSE, and MAE, together with the validation strategy, size and quality of datasets, model generalizability, interpretability, and ability to integrate the model with optimization methods, should be considered simultaneously[14, 24, 26, 27].

Accordingly, the objective of the present article is to systematically review and examine studies on the application of machine learning and artificial neural network methods in reforming processes, with a particular emphasis on methane reforming. In this context, the different approaches and models employed in previous studies are reviewed and compared in terms of their applications, predictive capabilities, and roles in the analysis and modeling of the process. As a review study, this work focuses on collecting, analyzing, and comparing the results and methodologies reported in previous research and does not aim to introduce new experimental data or modeling results. Rather, it seeks to provide an integrated perspective on existing methods and clarify the capabilities, differences, and applications of machine learning and artificial neural networks in reforming processes.

Although the Introduction briefly presents SMR, DRM, POX, ATR, and tri-reforming to establish the broader context of syngas production, the detailed machine-learning analysis in this review is concentrated primarily on methane-reforming systems, particularly DRM and SMR, because these pathways account for the most of the studies identified in ML. POX, ATR, and tri-reforming are discussed where ML-related evidence is available, but they are not treated as evidence bases with equal populations. This distinction is important to avoid implying that the depth of evidence is uniform across all hydrocarbon-reforming pathways.

Based on the above considerations, despite the growing number of experimental studies in the field of reforming, significant uncertainties remain regarding the relative superiority of different algorithms, the influence of validation strategies on reported results, and the appropriate criteria for the fair comparison of models. Many studies have focused on a single model and employed limited datasets or heterogeneous evaluation approaches, making it difficult to directly compare them or draw general conclusions regarding the performance of machine learning models in reforming processes. Furthermore, the recent shift toward interpretable methods and hybrid frameworks has highlighted the need to reconsider model evaluation criteria and to simultaneously consider the predictive accuracy, generalizability, and interpretability.

In view of this research gap, the primary question addressed in this article is: Among the different machine learning and artificial neural network methods applied in reforming processes, which algorithms provide better performance in predicting key outputs such as the reactant conversion, yield of hydrogen, and H_2/CO ratio, and based on what reliable evaluation criteria can their superiority be assessed?

The contribution of this review is not the introduction of a new ML algorithm. Rather, it is the synthesis of reforming-specific evidence under a common evaluation framework that simultaneously considers the predictive metrics, validation strategy, size and origin of datasets, overfitting/generalizability, feature engineering, interpretability, and integration with reaction-process models and optimization. This perspective allows apparently conflicting results—such as a high training R^2 value versus poor performance on the independent test—to be interpreted in relation to the data and validation protocol. The review therefore emphasizes the methodological comparability and physical relevance of ML findings rather than compiling algorithm descriptions alone.

To address this question, the main objective of the present article extends beyond a simple literature review to provide a systematic and analytical comparison of studies conducted to date in which machine learning algorithms—including artificial neural networks, Random Forest, Support Vector Machine, XGBoost, CatBoost, and ensemble methods—have been employed for modeling reforming processes. With an emphasis on establishing a consistent basis for comparison, the article examines the statistical metrics (R^2, MSE, RMSE, and MAE), validation strategies, size and quality of the datasets, degree of overfitting, and generalizability of the model. In addition, this review seeks to provide a clear assessment of the strengths, limitations, and existing gaps in the experimental studies of reforming by compiling and critically analyzing the findings of previous research.

To make the literature review more transparent and reproducible, the screening procedure should be reported explicitly. This review was structured around four stages: literature identification, title/abstract screening, full-text eligibility assessment, and qualitative /

quantitative data extraction. For each eligible study, information was extracted on the reforming pathway, catalyst (when reported), data source, size of datasets, input and output variables, preprocessing/feature engineering, type of the model, train/test strategy, cross-validation procedure, reported performance metrics, and stated limitations. Importantly, only the information explicitly reported by the reviewed studies was treated as evidence; when a study did not report a particular item, it was classified as "not reported" rather than inferred.

The remainder of this article is organized as follows. Section 2 discusses the evaluation and validation criteria for machine learning models. Section 3 introduces the different data-driven methods employed in reforming processes. Section 4 focuses on the sensitivity analysis, comparison of the performance of models, and effects of feature engineering. Finally, Section 5 presents the conclusion and recommendations for future research.

To strengthen the synthesis and reduce the textbook-style repetition, the detailed algorithm descriptions in Section 3 are intended only to establish the reforming-specific relevance, strengths, limitations, and validation requirements of each model. The central analytical contribution is developed in Section 4, where the performance of the model is interpreted in relation to dataset characteristics, target outputs, validation strategy, feature engineering, and physical consistency.

2. Evaluation Criteria and Comparison of the Performance of the Machine Learning Model

The evaluation of the performance of the machine learning model in reforming processes cannot be based solely on the degree of agreement between the model and the training data. The use of appropriate statistical metrics, together with model evaluation on test data, plays a critical role in assessing predictive capability and generalizability [28]. Various metrics, including the coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE), have been employed to evaluate predictive models in different studies. In general, higher R^2 values indicate better agreement between predicted and actual values, whereas lower MSE, RMSE, and MAE values indicate lower prediction errors. Therefore, the simultaneous consideration of these metrics provides a more comprehensive assessment of the accuracy of models. In a study on reforming, MSE, MAE, and R^2 were specifically selected as evaluation metrics, with lower MSE and MAE values and R^2 values closer to one indicating lower prediction errors and better model fit respectively [29].

The importance of using multiple statistical metrics simultaneously is also evident in comparative studies. For example, in a study on the prediction of the performance of dry methane reforming, the performance of the model was evaluated for four outputs, including CH_4 conversion, CO_2 conversion, the yield of H_2, and the yield of CO, using several statistical metrics. For CH_4 conversion, the CatBoost model applied to the original features achieved $R^2 = 0.91$ and MSE = 12.19, whereas Gradient Boosting with polynomial features resulted in $R^2 = 0.94$ and MSE = 7.58 [14]. For CO_2 conversion, CatBoost with polynomial features achieved $R^2 = 1.00$ and MSE = 0.57. The MAE and RMSE results reported in the same study were also consistent with the trends observed for R^2 and MSE. For instance, in predicting CH_4 conversion, the MAE and RMSE values obtained using CatBoost with the original features were 2.52 and 3.50 respectively, whereas the corresponding values for linear regression

were 9.20 and 10.11. Based on the combined R^2, MSE, MAE, and RMSE results, the authors concluded that CatBoost and Gradient Boosting exhibited overall suitable predictive performance among the evaluated models [12].

Limitations of statistical metrics. R^2 describes the fraction of variance represented by the model relative to a baseline, but it does not directly quantify the magnitude of the prediction error and can be misleading when the dataset is narrow or predictions are systematically biased. MSE strongly penalizes large errors, whereas RMSE expresses this error on the same scale as the predicted variable. MAE provides a more directly interpretable average absolute error and is less dominated by individual large errors than MSE. None of these metrics independently establishes generalizability, physical plausibility, or causal validity. Their interpretation should therefore be combined with the performance of the independent test, cross-validation, residual analysis, dataset coverage, and the practical scale of the target variable.

However, a high R^2 value or low error during the training stage alone cannot indicate the actual performance of a model. The manner in which the data are divided into training and test sets and the validation strategy must also be considered in model evaluation. This issue was specifically emphasized in the study by Elmaz et al. on dry methane reforming [14, 28]. In this work, ANN, Support Vector Regression (SVR), and polynomial regression were evaluated using 10-fold cross-validation to assess the generalizability of the model and reduce bias in comparisons among the methods [28]. The study also demonstrated that highly satisfactory performance on training data can be misleading, as some models exhibited a considerable decline in performance when applied to unseen data. The authors attributed this behavior to the limited number of observations and the insufficient generalizability of the models, emphasizing that training performance alone is not a reliable criterion for evaluating machine learning models [26].

This issue is particularly important when critically evaluating previous studies. In some earlier investigations, ANN models were developed using a very limited number of experimental results, or all available data were used for model training and the performance was evaluated on the same dataset [28]. Under such conditions, the ability of the model to generalize to new data cannot be properly determined [28]. Similarly, in another study, although the data were divided into training, testing, and validation sets, only training performance was reported, while insufficient information was provided regarding test and validation performance and the occurrence of overfitting. In another example, despite the low training error of a hybrid ANN–genetic algorithm model, a discrepancy of up to 28% between laboratory results and predicted values was reported, demonstrating the limitations of the model for practical applications [14, 28].

Hold-out validation provides a simple separation between development and evaluation data, but it can be unstable for small datasets because the result depends strongly on one split. k-fold cross-validation uses repeated training/validation partitions and can provide a more stable estimate when the dataset is limited, while repeated k-fold cross-validation further reduces dependence on a single partition. An independent test dataset, when genuinely external to model development, provides stronger evidence of generalization. However, even cross-validation does not demonstrate extrapolation beyond the distribution represented in the available data. Consequently, the strongest evidence is

obtained when model development uses appropriate cross-validation and the final model is evaluated on an independent dataset covering relevant unseen observations.

In addition to error metrics, the type and volume of data used are important factors when interpreting the performance of the model [30]. For example, in a study by Lee et al. on the energy optimization of the SMR process, an ANN model was developed using 485,710 sets of actual operational data. Following data preprocessing, outlier removal, noise filtering, and network variable tuning, an average R^2 of 0.9987 was reported for the prediction of six process variables [30]. In contrast, Elmaz et al. used only 57 observations and attributed the decline in the performance of some models during the testing stage to the limited amount of data [28]. Therefore, when comparing results from different studies, the value of an individual statistical metric such as R^2 should not be interpreted independently of the size of the dataset, data generation or collection method, and model validation strategy [14, 28, 30].

The reviewed literature spans markedly different dataset sizes. For example, the study on SMR optimization by Lee et al. used 485,710 operational-data records, the DRM study by Elmaz et al. used 57 observations, and the XGBoost SE-SMR study used 400 simulated cases. These datasets should not be treated as equivalent to sources of evidence: large operational datasets may provide broad coverage but can contain temporal correlation or repeated operating states, whereas small experimental datasets can contain richer controlled information but are more vulnerable to overfitting and unstable validation. The size of datasets should therefore be interpreted together with the independence, coverage of the operating space, noise level, and validation design.

Another example highlighting the importance of comparing models under identical conditions can be found in studies concerning sorption-enhanced steam methane reforming (SE-SMR) (SE-SMR) [28, 31]. In this study, ANN and Random Forest models were developed as soft sensors for predicting process variables. Both models achieved R^2 values above 98%; however, Random Forest consistently yielded higher R^2 values and lower MAE values. The reported MAE range was 0.002–0.014 for RF and 0.005–0.024 for ANN [31]. Since both models were evaluated using the same information and for the same prediction problem, such a comparison provides a more direct assessment of differences in the performance of the algorithm [28].

In addition to the size of the dataset, the reliability of an ML result depends on how observations are generated, cleaned, split, and reused during model development. The potential sources of optimistic performance estimate include the presence of duplicate observations in both the training and test sets, highly correlated measurements derived from the same experimental run, preprocessing performed before data splitting, feature selection using the entire dataset, and hyperparameter tuning based on the test set. For small reforming datasets, these issues can be especially important because a model may learn the structure of the available experiments without demonstrating reliable predictions at unseen operating conditions. Therefore, interpolation within the sampled operating domain should be distinguished from extrapolation beyond the experimental range. In cases where the reviewed papers did not provide sufficient information to assess these issues, this limitation is explicitly recognized rather than assuming an absence of leakage.

In explainable machine learning studies, model evaluation has not been limited to error

metrics, and the feature-importance analysis has also been employed alongside statistical measures [12, 16]. In a study by Valeripour et al. on catalytic dry methane reforming, the performance of the model was evaluated using RMSE and R^2, followed by the application of methods such as Gini importance, permutation importance, SHAP, and Partial Dependence Plots (PDPs) to interpret the effects of input variables [12]. In this study, the CatBoost model demonstrated the best performance among the evaluated tree-based models for predicting the H_2/CO ratio, achieving an R^2 value of 0.91 [12, 16].

Overall, based on the reviewed studies, R^2, MSE, RMSE, and MAE, together with the performance of the model on test data, validation strategy, the differences between the training and testing performance, size and of the dataset, and assessment of overfitting represent important criteria for comparing machine learning models in reforming processes. Furthermore, for explainable models, the feature-importance analysis and methods such as SHAP and PDP provide opportunities to investigate the roles of process and catalyst variables in model outputs. Therefore, comparing models based solely on a single statistical metric cannot provide a complete picture of their performance. Instead, the error metrics, validation strategy, generalizability, and model interpretability should be evaluated collectively [12, 28].

To avoid treating reported numerical metrics as they are directly comparable across unrelated studies, the reviewed evidence is interpreted using a structured framework comprising: (i) reforming pathway and catalyst system, (ii) source and size of the dataset, (iii) input and output definitions, (iv) preprocessing and feature engineering, (v) training/testing and cross-validation strategy, (vi) performance metrics, (vii) evidence of overfitting or data leakage, (viii) external or independent testing, and (ix) interpretability and physical consistency. Thus, a high R^2 value is considered a strong indicator of a good fit only within the specific dataset and validation protocol from which it was obtained, rather than as a universal measure of algorithmic superiority.

3. Machine Learning and Neural Network Methods Used in Reforming Processes

The complexity of reforming processes and the simultaneous dependence of their performance on multiple operational variables have increased the interest in empirical approaches for modeling and predicting these processes [12, 30]. A combination of operational and feed-related parameters, including the temperature, pressure, steam-to-carbon ratio, feed composition and flow rate, as well as reactor conditions, can collectively influence feed conversion, hydrogen production, and the resulting syngas composition [16, 31]. Accordingly, various machine learning algorithms and neural network methods have been employed to extract relationships between process inputs and outputs and to develop predictive models [14, 28]. These approaches range from artificial neural networks and regression models to tree-based algorithms and ensemble learning methods. In some studies, predictive models have also been integrated with optimization algorithms [30, 32].

Across the reviewed studies on reforming, commonly used inputs include the reaction temperature, pressure, feed composition, CH_4/CO_2 ratio, steam-to-carbon ratio, oxygen-to-carbon ratio, gas flow rate or space velocity, catalyst composition/loading, and catalyst structural descriptors. Common outputs include the CH_4 conversion, CO_2 conversion, yield of H_2, yield of CO, H_2/CO ratio, product-gas composition, and selected reactor-performance variables. The actual variables

differ among studies; therefore, these categories are used to organize the literature rather than to imply that every model uses the same feature set.

3.1. Artificial Neural Networks (ANNs)

Among the empirical modeling approaches applied to reforming processes, artificial neural networks (ANNs) have been extensively employed [29]. An artificial neural network is composed of interconnected processing units that establish the relationship between input and output variables through the iterative adjustment of connection weights during model training [14, 31]. This capability allows the effective representation of complex and nonlinear relationships between operating conditions and process responses to be [29, 30]. In studies on reforming, ANNs have been employed to predict variables such as the feed conversion, yield of hydrogen, and product composition, and their application has not been limited to a specific type of reforming processes [12, 31].

Among ANN architectures, the multilayer perceptron (MLP) has been employed in studies on reforming [29, 31]. The network is composed of an input layer, one or more hidden layers, and an output layer. During training, the weights associated with the connections between neurons are modified to capture the relationship between operational parameters and process responses [28]. Studies on the production of hydrogen and syngas have demonstrated that ANN/MLP models can be used to capture nonlinear relationships present in process data, with their performance evaluated using statistical model-performance metrics [29, 31].

In addition to MLP, other neural network architectures have been investigated in the empirical studies of chemical and reforming processes [14]. Variations in network architecture, including the number of layers and neurons, activation functions, and training methods, may influence the overall performance of the resulting model. Therefore, selecting an appropriate network architecture and tuning its parameters constitute important aspects of developing an ANN model for predicting the behavior of the process [28, 30, 31].

ANN/MLP is well suited to reforming datasets containing nonlinear interactions among the temperature, feed composition, pressure, catalyst properties, and process responses. Its flexibility, however, increases the need for adequate data coverage, regularization, scaling, architecture selection, and independent validation. RF/RFR can be relatively robust against nonlinear relationships and variable interactions and less dependent on feature scaling, but tree ensembles may be less reliable when predictions require smooth extrapolation beyond the training domain. SVR can perform well on small nonlinear datasets, but its kernel and regularization parameters require careful tuning and its performance can be sensitive to feature scaling. Boosting methods such as XGBoost and CatBoost can capture complex interactions and often perform strongly on tabular data, but their flexibility can also lead to overfitting if hyperparameters and validation are not controlled. Thus, the selection of algorithms should be linked to the structure and size of the reforming dataset rather than to generic algorithm popularity.

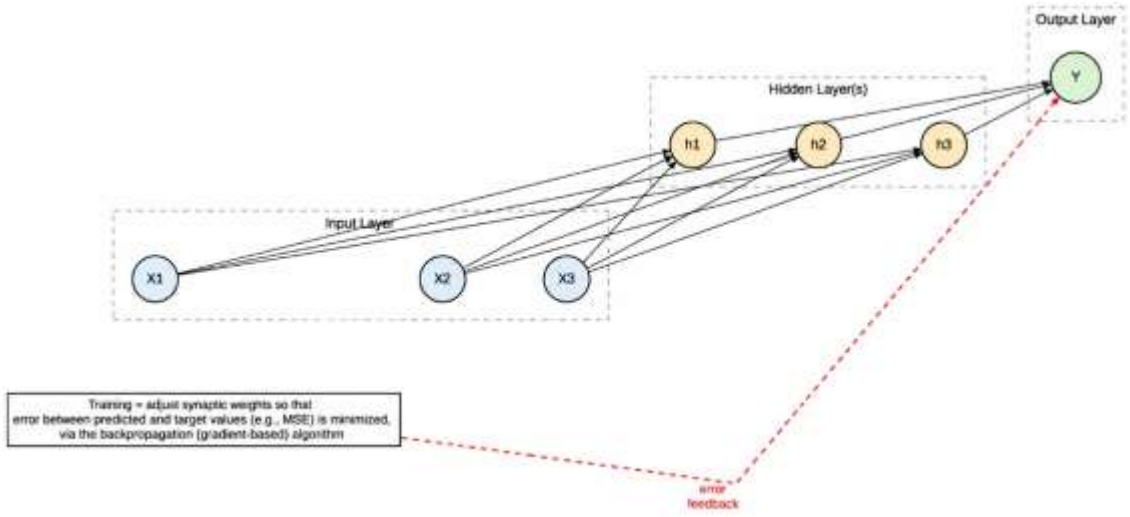

Figure 1. General structure of an artificial neural network (MLP), showing the input, hidden, and output layers and the iterative weight-update process used during training.

3.2. Random Forest (RF/RFR)

Another machine learning approach for modeling complex processes is the Random Forest method [12, 31]. In this approach, multiple decision trees are combined to form an ensemble, and the model produces its final output by aggregating the predictions generated by the individual trees [32]. In regression problems, the method is referred to as Random Forest Regression (RFR) [32]. The use of bootstrap sampling and random feature selection during tree construction introduces diversity among the trees, while aggregating their outputs is employed to reduce prediction variance [12].

In the study by Bishnu et al., RFR was employed to model an SE-SMR reactor and was subsequently coupled with an optimization algorithm [32]. The model was evaluated using R^2, RMSE, and MSE, with R^2 values close to one and error values close to zero reported for the model outputs [31, 32].

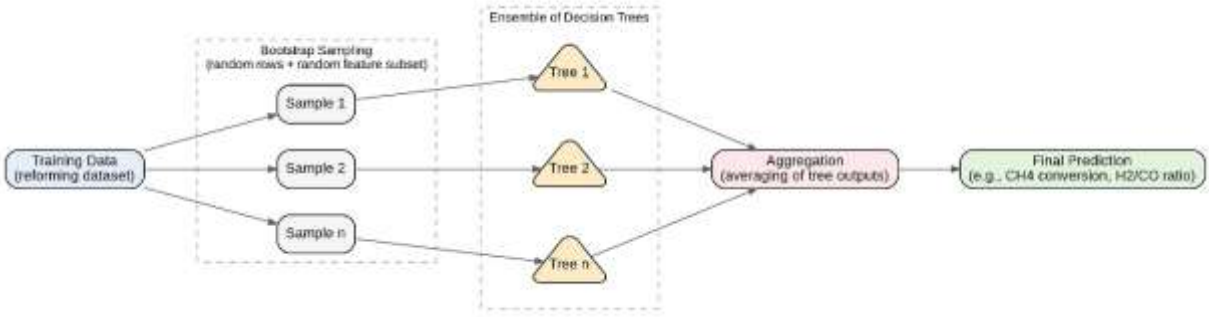

Figure 2. Structure of the Random Forest regression framework, illustrating bootstrap sampling, decision-tree construction, and aggregating tree predictions for estimating the process output.

3.3. Support Vector Regression (SVR)

Support Vector Regression (SVR) represents an additional algorithm that has been applied to the empirical modeling of chemical processes [14, 28]. Under this framework, the prediction problem is treated as a regression task, with kernel functions being used to

capture nonlinear relationships between the input and output variables [14]. In comparative studies, SVR has been evaluated alongside other machine learning models, with its performance assessed using metrics such as R^2 and prediction error indices. Therefore, SVR can be considered one of the models of interest when comparing empirical algorithms applied to hydrogen and syngas production processes [28].

Figure 3. General structure of Support Vector Regression (SVR) for nonlinear mapping between reforming process inputs and predicted outputs using a kernel function.

3.4. Ensemble Learning Methods

In addition to individual models, ensemble learning methods have also been employed in studies related to reforming processes and syngas production [12, 33]. These approaches generate the final predictive model by integrating the outputs produced by multiple learners [16, 33]. AdaBoost, Gradient Boosting, XGBoost, and CatBoost are among the algorithms in this category, each uses a distinct mechanism to enhance the prediction accuracy of the model [12, 16].

In a study on predicting the composition of syngas derived from bio-oil steam reforming, the use of ANN and AdaBoost was investigated, and product composition was estimated based on operational parameters including temperature, pressure, and the steam-to-feed ratio. The performance of the model was assessed through several statistical metrics, namely MSE, RMSE, MAE, and R^2 [33]. Such studies highlight the importance of directly comparing different models using the same dataset and consistent evaluation criteria, an issue that will be further investigated in the model performance comparison section [28].

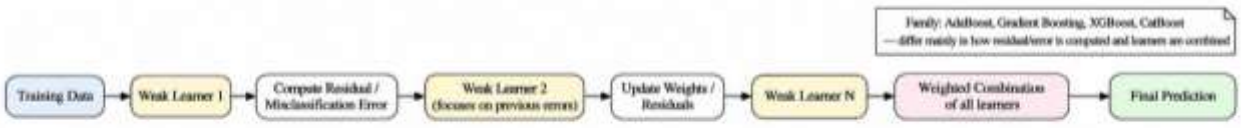

Figure 4. General structure of boosting and ensemble learning methods, including AdaBoost, Gradient Boosting, XGBoost, and CatBoost, for predicting the reforming process.

3.5. XGBoost

XGBoost is a tree-based ensemble learning approach that has recently been applied in studies on reforming for both the predictive and surrogate modeling [16, 34]. Wang et al. initially established a CFD-based reaction model for sorption-enhanced steam methane reforming (SE-SMR) and subsequently constructed a database by varying temperature, pressure, gas velocity, and the steam-to-carbon ratio. XGBoost was then developed using these four variables as inputs to predict the yield of H_2 [34].

The dataset comprised 400 cases, which were partitioned into training and testing sets containing 70% and 30% of the data

respectively. During model development, feature standardization was performed, and K-fold cross-validation was also applied [34]. The XGBoost model consists of an ensemble of decision trees, with each tree based on the CART (Classification and Regression Tree) algorithm. The outputs of the individual trees are then combined to generate the final prediction. [16]. Moreover, a regularization term was added to the objective function to limit the complexity of the model and reduce the risk of overfitting [12].

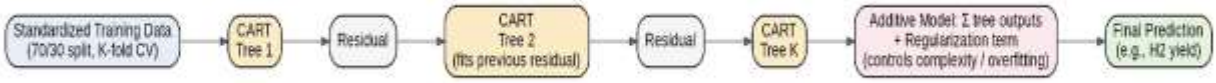

Figure 5. Structure of XGBoost for predicting the reforming process, showing additive CART and regularization to control the complexity of the model.

Preprocessing should be considered part of the modeling methodology rather than a secondary technical step. Common procedures include handling missing observations, detecting influential outliers, normalization or standardization, and feature selection or transformation. Scaling is particularly important for distance- or gradient-based approaches such as ANN and SVR, whereas tree-based models such as RF and boosting methods are generally less sensitive to the scale of features. To prevent information leakage, any transformation of which parameters are estimated from the data should be fitted using the training portion and then applied to validation/test data. The reviewed studies are therefore interpreted according to the preprocessing procedures they actually report.

3.6. CatBoost and Explainable Machine Learning Models

Alongside models that primarily focus on prediction accuracy, the use of explainable machine learning (XML) methods has increased in studies on reforming in recent years. These approaches are intended not only to achieve accurate predictions, but also to assess the contribution of input variables and elucidate the relationships between process conditions, catalyst properties, and process responses [12]. In studies related to dry reforming of methane, boosting-based models, including CatBoost, have been employed in conjunction with model interpretation tools to investigate the importance of process and catalyst parameters. Such approaches provide a basis for conducting the sensitivity analysis and identifying the variables that most strongly influence the performance of the process [16].

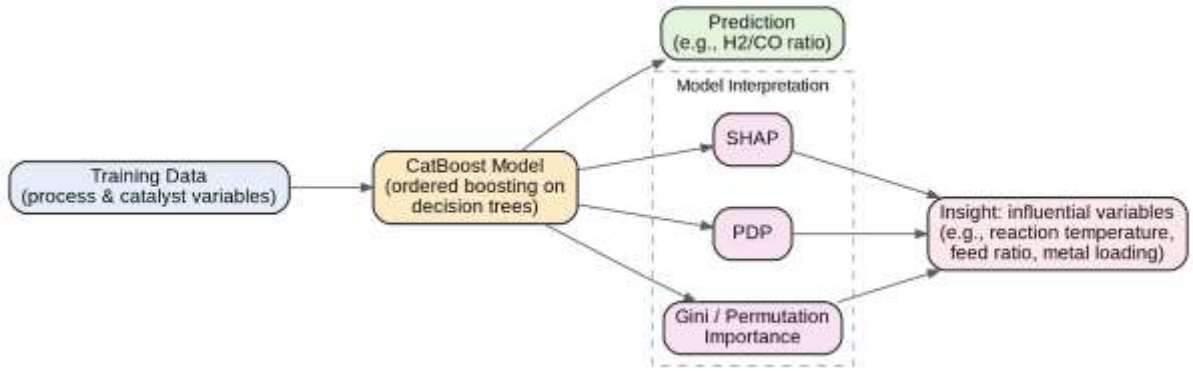

Figure 6. CatBoost prediction framework combined with explainable ML tools (SHAP, PDP, and feature importance) for evaluating the statistical contribution of reforming-process inputs.

3.7. Surrogate Models and Hybrid Frameworks

The application of machine learning models as surrogate models has emerged as an important research direction in recent studies [24, 34]. In this framework, the necessary data are initially obtained from experiments, simulations, or reaction models, after which a machine learning model is trained to capture the relationship between the input and output variables [24, 32]. Once trained, the model can be used instead of repeatedly executing the original model during the optimization process.

A representative example is provided by Wang et al., who integrated a CFD model for data generation, XGBoost for predicting the yield of hydrogen , and WOA for optimization within a unified framework [34]. The authors reported that the XGBoost model was capable of predicting the yield of H_2 in less than 4 s, while the optimization framework determined the operating conditions corresponding to the maximum yield of hydrogen in less than 1 min [24].

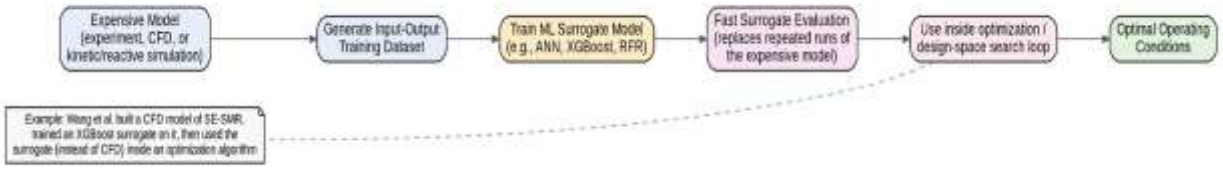

Figure 7. General surrogate-model framework in which experimental or simulation data train an ML model that replaces repeated high-cost process calculations during analysis or optimization.

3.8. Optimization Algorithms Combined with Machine Learning Models

The application of machine learning algorithms in studies on reforming is not limited to predicting process outputs. In some studies, trained machine learning models have been integrated with optimization algorithms [32, 34]. In such frameworks, the surrogate model predicts the behavior of the process, while the optimization algorithm searches the input-variable space to determine the conditions that satisfy the desired objective function [24, 30].

One of the optimization algorithms employed in this context is the Whale Optimization Algorithm (WOA). Wang et al. proposed a three-stage framework consisting of a CFD-based reaction model, an XGBoost surrogate model, and WOA to optimize the yield of hydrogen in SE-SMR [34]. Within this framework, the XGBoost model was used to evaluate the fitness function of the WOA algorithm. The optimized results were compared with those obtained from the original model for validation, yielding a percentage error of only 0.79%. Moreover, under the optimal operating conditions, the dry molar fraction of hydrogen was determined to be 95.96% [34].

Another optimization algorithm employed in this field is Simulated Annealing (SA). In 2025, Bishnu et al. employed an RFR model to characterize the behavior of an SE-SMR reactor and subsequently integrated it with SA [32]. In this framework, the RFR model served as the surrogate model, while SA searched for operating conditions that maximized CH_4 conversion by varying the input variables within the specified ranges [24, 32].

In this study, SA explores the solution space using a probability-based search strategy, in which less favorable solutions may be accepted during the initial stages. With the gradual reduction in temperature, the search progressively shifts toward more promising regions of the solution space. This behavior allows the algorithm to overcome local optima and conduct a broader exploration of the solution space. Ultimately, the integrated RFR–SA framework achieved a methane conversion of 0.99979 under the optimal conditions [24, 32].

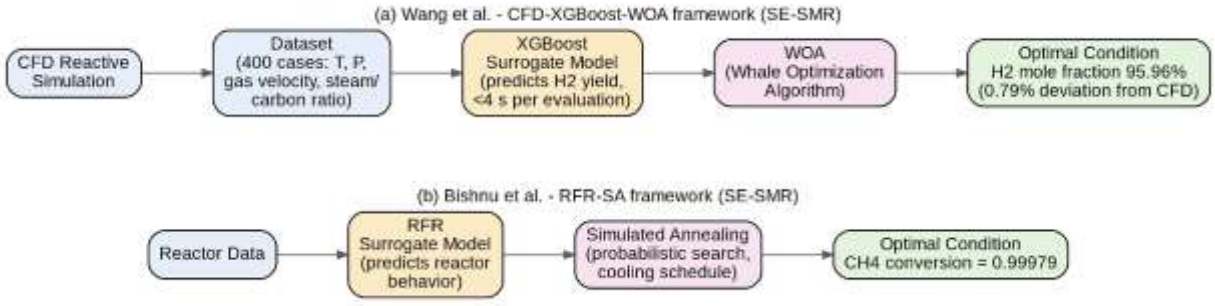

Figure 8. Optimization frameworks for the hybrid reforming-process: (a) CFD–XGBoost–WOA for optimizing the yield of H_2 and (b) RFR–SA for optimizing CH_4-conversion.

4. Sensitivity Analysis and Comparing the Performance of Machine Learning Models in Reforming Processes

A comparison of machine learning models used in studies on reforming shows that the performance of a given algorithm should be evaluated in relation to the type of the process, output variable, size and of data, and validation strategy, rather than considering its superiority independently [28, 29]. For example, in a study by Hossain et al. on the dry reforming of methane, two ANN architectures, namely MLP and RBF, were compared. For the MLP model, the R^2 values obtained for the yield of H_2, yield of CO, CH_4 conversion, and CO_2 conversion were 0.9726, 0.8597, 0.9638, and 0.9394 respectively. In comparison, the corresponding RBF model produced the R^2 values of 0.9218, 0.7759, 0.8307, and 0.7425. In addition, the standard error of estimation was reported to be lower for MLP than the same for RBF across all output variables. Accordingly, the MLP model demonstrated

better predictive performance than the RBF model in that study [29].

However, the comparison of these results with those of more recent studies indicates that the model evaluation strategy plays a decisive role in interpreting the performance of the model [14, 28]. In 2025, Devasahayam et al. evaluated 15 regression models using the same DRM dataset. For a consistent comparison, they also reimplemented the ANN-MLP and ANN-RBF models presented in the study by Hossain et al., applying a 70:30 training-to-testing data split. In this re-evaluation, the test-set R^2 values for ANN-MLP were reported as −0.99 and −1.03 for the yields of H_2 and CO respectively, and −0.12 and −0.08 for the conversions of CH_4 and CO_2 respectively [14]. The corresponding values for ANN-RBF were even lower. The authors interpreted these findings as the evidence of the limited generalization capability of the previously developed models on unseen test data and suggested that the high R^2 values reported in the earlier study were likely associated with the training data or the entire dataset. This comparison demonstrates that a high R^2 value, without clearly specifying the training–testing data separation and validation procedure, is insufficient for making a definitive judgment regarding the quality of the model [14, 28, 29].

Table 2.

Comparison of R^2 values for ANN models (MLP and RBF) in the dry reforming of methane **-** original reported performance [29]versus retesting on independent test data [14].

Process Output	Original R^2 (MLP - Hossain)	Original R^2 (RBF - Hossain)	Independent Test R^2 (ANN-MLP - Devasahayam)	ANN-RBF (Independent Test)
H_2 yield	0.9726	0.9218	-0.99	Lower than ANN-MLP
CO yield	0.8597	0.7759	-1.03	Lower than ANN-MLP
CH_4 conversion	0.9638	0.8307	-0.12	Lower than ANN-MLP
CO_2 conversion	0.9394	0.7425	-0.08	Lower than ANN-MLP

In the same multi-model DRM study, ensemble algorithms demonstrated notable performance [12, 14]. For CH_4 conversion, CatBoost achieved $R^2 = 0.91$ and MSE = 12.19 using the original features, whereas Gradient Boosting, after incorporating polynomial features, achieved $R^2 = 0.94$ and MSE = 7.58. For CO_2 conversion, CatBoost with polynomial features achieved $R^2 = 1.00$ and MSE = 0.57. For the yield of H_2, Random Forest exhibited the best reported performance, with $R^2 = 0.88$ and MSE = 3.72, followed by CatBoost with $R^2 = 0.84$ and MSE = 5.17. For the prediction of the yield of CO, CatBoost showed the best performance, achieving $R^2 = 0.89$ and MSE = 3.97.

These findings indicate that even when using the same dataset, the best-performing model may vary depending on the target output. Therefore, selecting a single algorithm for all process variables is not necessarily the most effective approach [12, 14].

The numerical performance values reported in different studies should not be interpreted as direct head-to-head rankings unless the models have been evaluated using the same prediction

task, dataset, target definition, and validation protocol. Differences in catalysts, operating ranges, sample sizes, preprocessing, feature engineering, and the way data is split into training and testing sets can change the apparent performance of an algorithm. Accordingly, the comparisons in this review are used primarily to identify methodological patterns and contrasts, while direct algorithmic comparisons are emphasized only when models were evaluated under sufficiently comparable conditions.

Table 3.
Comparison of the statistical performance metrics of several machine learning models for predicting the dry reforming of methane (DRM) outputs [14]

Output	Model	R^2	MSE	MAE	RMSE
CH_4 conversion	CatBoost (original data)	0.91	12.19	2.52	3.50
CH_4 conversion	Gradient Boosting (polynomial features)	0.94	7.58	—	—
CH_4 conversion	Linear Regression	—	—	9.20	10.11
CO_2 conversion	CatBoost (polynomial features)	1.00	0.57	—	—
H_2 yield	Random Forest	0.88	3.72	—	—
H_2 yield	CatBoost	0.84	5.17	—	—
CO yield	CatBoost	0.89	3.97	—	—

The dependence of the performance of the model on the target output has also been observed in bio-oil steam reforming. Adeniyi et al. compared ANN and AdaBoost for predicting syngas composition [28, 33]. When evaluated on the training data, both models achieved $R^2 > 0.999$ for all outputs, with RMSE < 0.007 and MAE < 0.005 in most cases. However, after applying stratified 10-fold cross-validation, the results were not consistent. Under this validation strategy, AdaBoost showed better performance for H_2 and CO_2, whereas ANN performed better for CH_4 and CO, with $R^2 > 0.9$ obtained in the reported cases [33]. Therefore, these findings further indicate that the relative ranking of machine learning models can vary when a more rigorous validation strategy is employed. A direct comparison of ANN and Random Forest in SE-SMR led to the similar conclusion. Ankilikianka et al. employed both models as soft sensors for predicting the concentration of gases and CH_4 conversion [31]. Both algorithms achieved R^2 values above 98%; however, Random Forest consistently yielded higher R^2 values and lower MAE values. The reported MAE ranged from 0.002 to 0.014 for RF and from 0.005 to 0.024 for ANN. Because both models were assessed using an identical dataset and the same output variables, the higher prediction accuracy observed for RF provides a more direct basis for comparison than evaluations conducted across separate studies [28, 31].

4.1. Sensitivity Analysis and the Importance of Input Variables

In addition to comparing the accuracy of the model, the sensitivity analysis is important for determining the role of input variables in the performance of the process [12, 16]. In recent studies, explainable methods such as SHAP, PDP, and feature importance have been employed for this purpose. For biogas tri-reforming, a database, comprising 1,183 records obtained from 116 experiments reported across 29 articles, was assembled. Random Forest was then used to predict CH_4

conversion, CO_2 conversion, and the H_2/CO ratio [16]. SHAP was also used to determine the effects of descriptors on the predicted outputs, while a Decision Tree was employed to extract combined relationships among the variables [16, 28]. The study found that the maximum CO_2 conversion was obtained at a high operating temperature combined with a low calcination temperature and a low O_2 ratio [16]. Additionally, the Decision Tree model revealed the distinct combinations of variables that can produce different H_2/CO ratios appropriate for applications such as dimethyl ether, methanol, and ammonia production. Thus, explainable methods serve a purpose beyond enhancing the prediction accuracy, as they also facilitate the identification and interpretation of relationships between operating conditions and the performance of the process [16, 28].

The ML trends should be interpreted together with the known chemistry of reforming. Temperature affects reaction kinetics and equilibrium and is therefore expected to influence conversion and product distribution strongly. The CH_4/CO_2 feed ratio directly changes reactant availability and can shift the H_2/CO product ratio, while catalyst composition and Ni loading can affect active sites and resistance to deactivation. In DRM, high temperature can promote conversion but can also be accompanied by carbon deposition and catalyst sintering, so a statistically favorable ML prediction should not automatically be interpreted as an experimentally optimal condition. Similarly, constraints related to pressure, heat transfer, mass transfer, and equilibrium can limit the physically accessible operating space. These considerations support the use of ML as a data-driven complement to understanding the reaction mechanism, rather than a substitute for understanding it. The feature-importance analysis in the study of the multi-model DRM further identified the reaction temperature, CH_4/CO_2 feed ratio, and Ni loading as key influential variables [12, 14]. Since different models may exhibit different sensitivities to input features, combining statistical performance metrics with the feature-importance analysis can provide a more comprehensive understanding of the behavior of the model [12, 28]. Exercise caution when interpreting explainable ML results. SHAP values, permutation importance, Gini importance, and PDPs should be interpreted as the model-based/statistical measures of association or sensitivity, not as the direct evidence of causality. When process variables are strongly correlated, importance may be redistributed among correlated features. In addition, limited experimental coverage and dataset bias can influence the resulting explanation. PDPs may also become difficult to interpret when the plotted feature combinations are unrealistic or poorly represented in the original data. Accordingly, the physical or chemical interpretation of an influential variable should be supported by the established knowledge of reaction engineering, thermodynamic, or catalysis rather than merely by explainable outputs.

4.2. Effect of Feature Engineering and the Complexity of the Model

Another factor affecting the performance of the model is the representation of input data. In the study by Devasahayam et al., the use of polynomial features resulted in a substantial improvement in the performance of some models. For example, Gradient Boosting achieved $R^2 = 0.94$ in predicting CH_4 conversion when polynomial features were incorporated, while CatBoost achieved $R^2 = 1.00$ for CO_2 conversion using the same type of features [14]. In contrast, models characterized by high variance, such as SVR

with polynomial features, showed a greater tendency toward overfitting and consequently performed less effectively on the test set. Conversely, because of their lower bias, linear models were not capable of adequately representing the complex and nonlinear relationships contained in the data [14, 28]. The DRM results provide a concrete example of why feature engineering should be evaluated on unseen data. In the reported comparisons, polynomial feature expansion improved the performance of the Gradient Boosting model for CH_4-conversion to R^2 = 0.94 and the performance of the CatBoost model for CO_2-conversion to R^2 = 1.00. However, the same literature also showed that models can lose a substantial portion of their performance when evaluated on unseen observations; therefore, an apparent improvement in R^2 after feature expansion cannot, by itself, demonstrate improved generalizability. In particular, a model may fit interactive sequences that describe a limited experimental sample but fail when operating conditions that differ from those represented during training. Hyperparameters such as ANN architecture, learning rate, number of neurons/hidden layers, regularization parameters, SVR kernel parameters, and tree depth or learning rate can materially affect the performance of the model. Common selection strategies include grid search, random search, Bayesian optimization, and evolutionary/metaheuristic methods. Hyperparameter tuning should be nested within the training/validation procedure; otherwise, repeated optimization against the test set can produce overly optimistic performance estimates. For this reason, the reported performance of the model is interpreted more cautiously when the tuning procedure is not described. These findings suggest that greater complexity in either the input representation or the model itself does not necessarily result in better performance. Instead, an appropriate balance is required between the model's capacity to capture nonlinear relationships and its ability to generalize effectively. This consideration becomes particularly critical in the studies on reforming based on small datasets, where the likelihood of overfitting is typically higher [14].

4.3. Summary of Performance Comparison

Overall, the findings from the reviewed studies suggest that the performance of machine learning models varies across reforming processes and output variables, with no single model consistently demonstrating superior performance in all cases [29]. In the initial study on DRM, MLP demonstrated better statistical performance than RBF; however, re-evaluation of the same architectures on an independent test set revealed their limited generalizability [14]. Ensemble models, including Random Forest, CatBoost, and Gradient Boosting, have demonstrated satisfactory performance across several studies [12, 14], although the best-performing model varied depending on the target output. Similarly, ANN and AdaBoost exhibited different levels of performance depending on the predicted output in bio-oil steam reforming, while in SE-SMR, Random Forest achieved higher predictive accuracy and lower MAE than ANN in a direct comparison [31].

Therefore, based on the reviewed literature, the comparison of machine learning models for reforming processes should consider, at a minimum, R^2, MSE, RMSE, and MAE, performance on the test set, validation strategy, the potential for overfitting, the size of dataset, and the sensitivity analysis of input variables. Furthermore, model comparisons are more reliable when multiple algorithms are evaluated on a common dataset using the same validation strategy, because directly

comparing performance metrics reported in studies that utilize different datasets, target outputs, and evaluation procedures may lead to misleading conclusion regarding the absolute superiority of a particular model [12, 28]. Because the studies differ in reforming pathway, catalyst, operating range, the size of dataset, input variables, target outputs, preprocessing, and validation strategy, the evidence does not support a universal ranking of ML algorithms. The practical conclusion is instead conditional: an algorithm should be selected and validated for the specific reforming task and dataset, and claims of superiority should be restricted to comparisons performed under sufficiently comparable conditions.

Table 4.
Brief comparison of ANN with other machine learning algorithms across different reforming processes.

Process / Study	Models Compared	Main Finding
SMR [30]	ANN	R^2=0.9987 (average of 6 process variables); very large dataset (485,710 samples)
DRM [28]	ANN, SVR, Polynomial Regression	Noticeable performance drop under 10-fold cross-validation; small dataset (57 observations)
Bio-oil Steam Reforming [33]	ANN vs AdaBoost	Training: both R^2>0.999; Cross-validation: AdaBoost better for H_2 and CO_2, ANN better for CH_4 and CO (R^2>0.9)
SE-SMR (soft sensor)	ANN vs Random Forest	Both R^2>0.98; RF MAE range 0.002-0.014; ANN MAE range 0.005-0.024
Catalytic DRM [12]	Tree-based models (incl. CatBoost)	CatBoost best among tree-based models; R^2=0.91 for H_2/CO ratio

Table 5. Structured the comparison of selected studies on reforming reviewed in the present work, emphasizing dataset characteristics, variables, validation strategy, reported metrics, and limitations. “NR” indicates that the manuscript does not report sufficient information for that item. The table is intended as a synthesis tool rather than a numerical ranking. Where studies used different datasets, targets, catalysts, or validation procedures, the reported metrics are retained as study-specific evidence and are not interpreted as scores allowing for direct comparison.

Table 5.
Input and output variables, dataset characteristics, and machine learning methods reported in studies on reforming.

Reforming process	Dataset / Data source	Main input variables / descriptors	Predicted output variables	ML method
Dry Methane Reforming (DRM) (29)	24 experimental observations obtained from DRM experiments over $Ni/CaFe_2O_4$ catalyst	Catalyst metal loading (5–15 wt%), feed ratio (0.4–1.0), reaction temperature (700–800 °C)	H_2 yield, CO yield, CH_4 conversion, CO_2 conversion	ANN-MLP and ANN-RBF
Dry Methane Reforming (DRM) (28)	57 experimental observations; dataset derived from DRM over a ceria-supported cobalt catalyst	CH_4 partial pressure, CO_2 partial pressure, CH_4/CO_2 ratio, reaction temperature	H_2 production rate, CO production rate, CH_4 conversion, CO_2 conversion	ANN, Support Vector Regression (SVR), Polynomial Regression; 10-fold cross-validation
Ni-catalyzed DRM – Vellayappan et al.	1,638 data points curated from 221 published studies	CH_4 feed fraction, reaction temperature, Ni loading, reaction time, GHSV, pore size, pore volume, surface	CH_4 conversion, CO_2 conversion, H_2/CO ratio (syngas ratio)	Random Forest, Gradient Boosting Regressor, CatBoost;

		area, H_2-TPR peak temperature, Ni particle size, modifier electronegativity		SHAP and PDP for interpretation
Tri-reforming of Biogas (TRM) (16).	1,183 data entries compiled from 29 articles published between 2004 and 2024; 41 descriptors	Reaction temperature, calcination time, H_2O and O_2 percentages in the reaction stream, W/F ratio, together with catalyst and other process descriptors	CH_4 conversion, CO_2 conversion, H_2/CO ratio	Random Forest regression, SHAP analysis, Decision Tree classification
Sorption-Enhanced Steam Methane Reforming (SE-SMR) – soft-sensor study (31)	Process dataset for SE-SMR reactor/regenerator modeling	Temperature, pressure, steam-to-carbon ratio, sorbent-to-carbon ratio	Gas concentrations in the reformer and regenerator reactors	ANN and Random Forest
Sorption-Enhanced Steam Methane Reforming (SE-SMR) – XGBoost study(34)	400 simulated cases generated using a CFD/reaction model	Temperature, pressure, gas velocity, steam-to-carbon ratio	H_2 yield	XGBoost; compared with AdaBoost and Bagging
Biomass Bio-oil Steam Reforming (33)	Experimental data for biomass bio-oil steam reforming	Temperature, pressure, steam-to-feed ratio	H_2, CO, CO_2 and CH_4 product/syngas composition	ANN and AdaBoost; stratified 10-fold cross-validation
Steam Methane Reforming (SMR) **(30)**	485,710 actual operating data records from an on-site SMR process	Natural-gas feed/fuel flow rate, H_2 flow rate, water flow rate and temperature, air flow rate, SMR inlet temperature and pressure, LTS inlet temperature and other operating variables	Syngas flow rate, CO, CO_2, CH_4 and H_2 compositions, steam temperature	ANN with preprocessing, outlier removal, noise filtering and hyperparameter tuning
DRM – cobalt catalyst carbon-deposition study	170 experimental runs	CH_4/CO_2 ratio, reaction temperature, N_2 flow rate	Carbon deposition per gram of catalyst	MLPNN and RBFNN

5. Conclusion

This review systematically examined and analyzed the application of machine learning (ML) methods and artificial neural networks (ANNs) in the modeling, prediction, analysis, and optimization of hydrocarbon reforming processes, with particular emphasis on methane reforming. The reviewed studies demonstrated that the inherent complexity of reforming processes, arising from the simultaneous interactions among operating variables, chemical reactions, transport phenomena, and catalyst properties, provides a suitable context for the application of data-driven approaches. Artificial neural networks, particularly MLP architectures, have demonstrated a strong ability to capture nonlinear relationships between input and output variables. However, the results of various studies indicate that the high accuracy of the model during training does not necessarily imply satisfactory predictive performance on unseen data.

The comparison of the reviewed studies revealed that no machine learning algorithm can be identified as a universally superior model without considering the specific problem, target output, data structure, and validation strategy. For example, in some studies on the dry reforming of methane, MLP outperformed RBF, whereas subsequent reevaluation of these models using independent test data revealed substantial limitations in their generalization capability. On the other hand, ensemble-based models, including CatBoost, Random Forest, and Gradient Boosting, have demonstrated notable performance across various studies. However, even when evaluated on a common dataset, the best-performing model was not necessarily the same for CH_4 conversion, CO_2 conversion, the yield of H_2, and the yield of CO. This finding indicates that model selection should be

tailored to the target output and the characteristics of the dataset, rather than being based on the popularity or performance of a model in a single study.

One of the most important findings of this review is the critical role of validation strategy in assessing the actual performance of machine learning models. The comparison of the reviewed studies indicates that reporting a high R^2 value on training data alone cannot be considered a reliable criterion for model selection and may provide an unrealistic representation of predictive capability in the presence of overfitting. Therefore, the performance of the model on an independent test set, data-splitting strategy, cross-validation procedure, differences between training and testing performance, and the size and quality of the dataset should be considered simultaneously. Accordingly, directly comparing R^2 or other statistical metrics across studies employing different datasets and validation strategies may lead to misleading conclusion regarding the superiority of a particular algorithm.

From a statistical evaluation perspective, the simultaneous use of R^2, MSE, RMSE, and MAE, rather than relying on a single metric, provides a more comprehensive assessment of the performance of the model. In addition, for explainable machine learning models, approaches such as SHAP, PDP, and the feature-importance analysis can provide insights that go beyond mere prediction by improving the understanding of relationships among operating variables, catalyst characteristics, and process performance. The reviewed studies have shown that variables such as the reaction temperature, CH_4/CO_2 feed ratio, and Ni loading are among the important factors influencing the performance of reforming processes. Therefore, combining predictive accuracy with model interpretability can enhance the scientific and practical values of data-driven models.

The review also demonstrated that feature engineering and the increased complexity of the model do not necessarily lead to improved performance. Although the incorporation of polynomial features has improved the accuracy of models such as Gradient Boosting and CatBoost in some studies, increasing complexity may also increase the risk of overfitting, particularly when dealing with small datasets. Therefore, the development of machine learning models should aim at establishing an appropriate balance between the model's ability to learn nonlinear relationships and its capability to generalize to unseen data.

Furthermore, recent research trends indicate a shift from purely predictive models toward hybrid frameworks and surrogate models. Integrating machine learning models with CFD simulations, explainability methods, and optimization algorithms can reduce computational costs, facilitate the analysis of variable effects, and identify suitable operating conditions. Examples of combining XGBoost with CFD and the WOA algorithm, as well as RFR with the SA algorithm, demonstrate that data-driven models can evolve from simple predictive tools into the components of an integrated framework for the analysis and optimization of the process.

Finally, it can be concluded that the future of machine learning applications in reforming processes lies not in selecting a single algorithm, but in developing reliable, generalizable, and interpretable data-driven frameworks. To achieve this objective, future studies should focus on generating larger and higher-quality datasets, ensuring the balanced coverage of operating and catalyst-variable spaces, using independent datasets for final testing, implementing rigorous validation procedures, and reporting multiple statistical

performance metrics simultaneously. In addition, the development of hybrid models incorporating physical and chemical knowledge, broader application of explainable machine learning methods, and integration of machine learning models with optimization algorithms can provide a promising pathway toward models with high accuracy, robust generalization capability, and physically meaningful interpretations. Within such a framework, machine learning can serve as an effective tool for accelerating the process design, performance prediction, identification of key variables, and optimization of hydrogen and syngas production processes.

Established findings from the reviewed literature indicate that model performance depends strongly on dataset characteristics and validation strategy; training-set accuracy alone is insufficient evidence of generalization; and different target outputs can favor different algorithms. In contrast, larger and more diverse datasets, independent external validation, uncertainty quantification, physics-based or mechanism-constrained ML, and broader explainability present future research avenues rather than the definitive conclusion of the present review.

NOMENCLATURE

AI	Artificial Intelligence
ANN	Artificial Neural Network
ATR	Autothermal Reforming
CART	Classification and Regression Trees
CFD	Computational Fluid Dynamics
DRM	Dry Reforming of Methane
MAE	Mean Absolute Error
ML	Machine Learning
MLP	Multilayer Perceptron
MSE	Mean Squared Error
PDP	Partial Dependence Plot
POX	Partial Oxidation
RBF	Radial Basis Function
RF / RFR	Random Forest / Random Forest Regression
RMSE	Root Mean Squared Error
SA	Simulated Annealing
SE-SMR	Sorption-Enhanced Steam Methane Reforming
SHAP	SHapley Additive exPlanations
SMR	Steam Methane Reforming
SVR	Support Vector Regression
TMR	Tri-Reforming of Methane
WOA	Whale Optimization Algorithm
XGBoost	Extreme Gradient Boosting

References

[1] Kiani P, Rahimpour MR. Syngas production from natural gas by reforming process: Dry, steam, partial oxidation, tri, and autothermal. Advances in Natural Gas: Formation, Processing, and Applications Volume 7: Natural Gas Products and Uses: Elsevier ; .2024p. 95-117.https://doi.org/10.1016/B978-0-443-19227-2.00002-2

[2] Farsi M. Reforming process design and modeling: Steam, dry, and autothermal reforming. Advances in synthesis gas: methods, technologies and applications: Elsevier; 2023. p. 123-40.https://doi.org/10.1016/B978-0-323-91879-4.00003-5

[3] Saad M, Abdurahman N, Yunus RM, Kamil M, Awad OI (2020) An Overview of Reforming Technologies and the Effect of Parameters on the Catalytic Performance of Mesoporous Silica/Alumina Supported Nickel Catalysts for Syngas Production by Methane Dry Reforming. Recent Innovations in Chemical Engineering (Formerly Recent Patents on Chemical Engineering). 13:303-22 https://doi.org/10.2174/2405520413666200313130420.

[4] Boscherini M, Storione A, Minelli M, Miccio F, Doghieri F (2023) New perspectives on catalytic hydrogen production by the reforming, partial oxidation and decomposition of methane and biogas. Energies. 16:6375 https://doi.org/10.3390/en16176375.

[5] Filimonova A, Chichirov A, Chichirova N, Pechenkin A (2023) Reforming of hydrocarbon fuel in electrochemical systems.

Thermal Engineering. 70:615-23 https://doi.org/10.1134/S0040601523080025.
[6] Kaydouh M-N, El Hassan N, Chalhoub E, Bukharin N, El Hassan M (2024) A review of Computational Fluid Dynamics modeling of dry reforming of methane for improved catalyst and reactor design. Fuel. 366:131290 https://doi.org/10.1016/j.fuel.2024.131290.
[7] Elserfy K, Farooqi AS, El-Adawy M, Hamdy M, Shahid MZ, Elserfy Z, et al. (2026) Heat transfer and reaction kinetics in porous reacting media for steam methane reforming: a critical review and thermodynamic analysis. Engineering Applications of Computational Fluid Mechanics. 20:2628975 https://doi.org/10.1080/19942060.2026.2628975.
[8] Sedaghat MH, Rahimpour MR (2023) Computational fluid dynamics simulation of natural gas reformers. Advances in Synthesis Gas: Methods, Technologies and Applications.103-21 https://doi.org/10.1016/B978-0-323-91879-4.00013-8.
[9] Chen Y, Liu X, Liu X, Lu H, Wang Z (2025) Machine Learning Applications in Fuel Reforming for Hydrogen Production in Marine Propulsion Systems. Journal of Marine Science and Engineering. 14:85 https://doi.org/10.3390/jmse14010085.
[10] Yuan B, Zhu T, Han Y, Zhang X, Wang M, Li C (2023) Deactivation mechanism and anti-deactivation measures of metal catalyst in the dry reforming of methane: a review. Atmosphere. 14:770 https://doi.org/10.3390/atmos14050770.
[11] Ge Z, Lu Q, Xia K, Feng B, Lin G, Zhang H (2025) Structural evolution and coke deposition-driven deactivation mechanism of industrial-scale steam methane reforming catalysts. International Journal of Hydrogen Energy. 185:152028 https://doi.org/10.1016/j.ijhydene.2025.152028.
[12] Vellayappan K, Yue Y, Lim KH, Cao K, Tan JY, Cheng S, et al (2023) .Impacts of catalyst and process parameters on Ni-catalyzed methane dry reforming via interpretable machine learning. Applied Catalysis B: Environmental. 330:122593 https://doi.org/10.1016/j.apcatb.2023.122593.
[13] [13] Liu Z, Tian W, Cui Z, Liu B (202 (3 A universal microkinetic-machine learning bimetallic catalyst screening method for steam methane reforming. Separation and Purification Technology. 311:123270 https://doi.org/10.1016/j.seppur.2023.123270.
[14] Devasahayam S, Thella JS, Mohanty MK (202 (5Predicting methane dry reforming performance via multi-output machine learning: a comparative study of regression models. Energies. 18:4807 https://doi.org/10.3390/en18184807.
[15] Jafarizadeh A, Panjepour M, Emami MD (2024) Advanced modelling and optimization of steam methane reforming: From CFD simulation to machine learning-Driven optimization. International Journal of Hydrogen Energy. 96:1262-80 https://doi.org/10.1016/j.ijhydene.2024.11.352.
[16] Coşgun A, Günay ME, Yıldırım R (2025) Explainable machine learning analysis of tri-reforming of biogas for sustainable syngas production. International Journal of Hydrogen Energy. 127:595-607 https://doi.org/10.1016/j.ijhydene.2025.04.213.
[17] Huang J, Liang Z, Liu Y (2025) Smart reforming for hydrogen production via machine learning. Chemical Engineering Science. 304:120959 https://doi.org/10.1016/j.ces.2024.120959.
[18] Cherbański R, Kotkowski T, Molga E (2026) Hydrogen and syngas production from methane via thermochemical conversion: a mini review of artificial neural network applications. Chemical and Process Engineering: New Frontiers. 47:e135-e https://doi.org/10.24425/cpe.2026.158133.
[19] Madhiarasan M, Louzazni M (2022) Analysis of artificial neural network: architecture, types, and forecasting applications. Journal of Electrical and Computer Engineering. 2022:5416722 https://doi.org/10.1155/2022/5416722.
[20] Rahman MH, Biswas M (2024) Modeling of dry reforming of methane using artificial neural networks. Hydrogen. 5:800-18 https://doi.org/10.3390/hydrogen5040042.
[21] Deng H, Guo Y (2022) Artificial neural network model for the prediction of methane Bi-reforming products using CO2 and steam. Processes. 10:1052 https://doi.org/10.3390/pr10061052.
[22] Cherbański R, Kotkowski T, Murgrabia S, Molga E, Stankiewicz A (2026) ANN-based prediction of equilibrium gas composition and carbon yield in methane pyrolysis and reforming methods. Reaction Chemistry &

Engineering. 11:691-706 https://doi.org/10.1039/d5re00322a.

[23] Pizoń Z, Kimijima S ,Brus G (2025) Bridging equilibrium and kinetics prediction with a data-weighted neural network model of methane steam reforming. International Journal of Hydrogen Energy. 175:151367 https://doi.org/10.1016/j.ijhydene.2025.151367.

[24] Nabavi SR, Guo Z ,Wang Z (2026) Machine learning assisted surrogate modeling with multi objective optimization and decision making of a steam methane reforming reactor. Scientific Reports. https://doi.org/10.1038/s41598-026-64776-2.

[25] Alotaibi FN, Berrouk AS, Saeed M (2023) Optimization of yield and conversion rates in methane dry reforming using artificial neural networks and the multiobjective genetic algorithm. Industrial & Engineering Chemistry Research. 62:17084-99 https://doi.org/10.1021/acs.iecr.3c01813.

[26] Devasahayam S (2026) Interpretable SHAP Analysis of Key Operating Parameters in Methane Dry Reforming. Energies. 19:2618 https://doi.org/10.3390/en19112618.

[27] de Souza PA, Muhammad Afzal R, Gomes Camacho F, Mahinpey N (2025) Random forest regression for catalyst performance prediction and validation of tri-reforming of methane (TRM). The Canadian Journal of Chemical Engineering. https://doi.org/10.1002/cjce.70215.

[28] Elmaz F, Yücel Ö, Mutlu AY (2020) Predictive modeling of the syngas production from methane dry reforming over cobalt catalyst with statistical and machine learning based approaches. International Journal of Advances in Engineering and Pure Sciences. 32:8-14 https://doi.org/10.7240/jeps.558373.

[29] Hossain MA, Ayodele BV, Cheng CK , Khan MR (2016) Artificial neural network modeling of hydrogen-rich syngas production from methane dry reforming over novel Ni/CaFe2O4 catalysts. International Journal of Hydrogen Energy. 41:11119-30 https://doi.org/10.1016/j.ijhydene.2016.04.034.

[30] Lee J, Hong S, Cho H, Lyu B, Kim M, Kim J, et al. (2021) Machine learning-based energy optimization for on-site SMR hydrogen production. Energy Conversion and Management. 244:114438 https://doi.org/10.1016/j.enconman.2021.114438.

[31] Nkulikiyinka P, Yan Y ,Güleç F, Manovic V, Clough PT (2020) Prediction of sorption enhanced steam methane reforming products from machine learning based soft-sensor models. Energy and AI. 2:100037 https://doi.org/10.1016/j.egyai.2020.100037.

[32] Bishnu SK, Alnouri SY, Al Mohannadi DM (2025) Stochastic algorithm-based optimization using artificial intelligence/machine learning models for sorption enhanced steam methane reformer reactor. Computers & Chemical Engineering. 196:109060 https://doi.org/10.1016/j.compchemeng.2025.1.09060

[33] Adeniyi AG, Ighalo JO, Marques G (2021) Utilisation of machine learning algorithms for the prediction of syngas composition from biomass bio-oil steam reforming. International Journal of Sustainable Energy. 40:310-25 https://doi.org/10.1080/14786451.2020.18038.62

[34] Wang L, Li H, Du C, Hong W (2024) Optimizing hydrogen yield in sorption-enhanced steam methane reforming: A novel framework integrating chemical reaction model, ensemble learning method, and whale optimization algorithm. Journal of the Energy Institute. 114:101649 https://doi.org/10.1016/j.joei.2024.101649.