

PREDICTION OF HYDROGEN YIELD AND METHANE CONVERSION IN METHANE PYROLYSIS USING A CONVOLUTIONAL NEURAL NETWORK

Subhan Azeem^{*1}, Nadeem Hassan², Muhammad Ashraf³

^{1,2}NFC Institute of Engineering & Technology, Multan

³Bahauddin Zakriya University, Multan

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

DOI: <https://doi.org/10.5281/zenodo.19629457>

Keywords

Methane pyrolysis, Hydrogen yield prediction, Methane conversion, Convolutional neural network, Low-carbon hydrogen

Article History

Received: 05 January 2026

Accepted: 10 February 2026

Published: 01 March 2026

Copyright @Author

Corresponding Author: *

Subhan Azeem

Abstract

Methane pyrolysis has emerged as a promising low-carbon route for hydrogen production because it generates hydrogen without direct carbon dioxide emissions in the main reaction pathway while producing solid carbon as a by-product. However, the prediction of hydrogen yield and methane conversion remains challenging due to the complex nonlinear interactions among operating conditions and catalyst composition variables. In this study, a convolutional neural network (CNN) framework is proposed to simultaneously predict hydrogen yield and methane conversion in methane pyrolysis using structured process and catalyst data. The developed model was trained on methane pyrolysis input variables, including temperature, gas hourly space velocity, CH₄ concentration, reaction time, calcined temperature, and catalyst composition descriptors. The CNN was designed as a multi-output regression model to capture hidden nonlinear relationships between the input variables and the two target responses. Experimental evaluation demonstrated stable convergence during training, with training loss decreasing from nearly 0.18 to 0.009 and validation loss dropping from about 0.20 to 0.016 over 60 epochs, indicating effective learning and limited overfitting. Comparative analysis showed that the proposed CNN outperformed benchmark models, including Linear Regression, Support Vector Regression, Random Forest, multilayer perceptron, and CatBoost, achieving the highest predictive accuracy with an R² of approximately 0.95 and the lowest RMSE of nearly 4.4. Feature-importance analysis further revealed that temperature, gas hourly space velocity, reaction time, and methane concentration were the dominant predictors, together accounting for nearly 60% of the overall importance. These findings confirm that CNN-based regression provides a robust and accurate framework for methane pyrolysis modeling and offers strong potential for predictive analysis, process screening, and future optimization of hydrogen production systems.

INTRODUCTION

The global energy transition has intensified interest in hydrogen as a clean energy carrier capable of supporting decarbonization across industrial production, power systems, and

transportation. Yet the environmental value of hydrogen depends strongly on how it is produced. Conventional fossil-based pathways remain dominant because of their technical

maturity and established infrastructure, but they are associated with substantial carbon emissions unless paired with carbon capture systems [1]. In this context, methane pyrolysis has emerged as an increasingly attractive alternative because it can generate hydrogen while avoiding direct carbon dioxide formation in the main reaction pathway, instead yielding solid carbon as a co-product [2]. This feature has positioned methane pyrolysis as a strategically important route within the broader landscape of low-carbon hydrogen technologies. Recent reviews and assessments have highlighted its promise for cleaner hydrogen generation, especially where the management or valorization of solid carbon is technically and economically feasible.

Methane pyrolysis is conceptually simple, but its practical optimization is far from straightforward. Hydrogen yield and methane conversion are governed by a tightly coupled set of reaction and design variables, including reaction temperature, residence time, methane concentration, gas hourly space velocity, catalyst composition, catalyst support, reactor configuration, and catalyst deactivation behavior [3]. Even small changes in these factors can alter reaction pathways, carbon deposition patterns, catalyst activity, and overall process efficiency. Recent studies have shown that high conversion and favorable hydrogen productivity are achievable under carefully selected operating windows, but they have also emphasized that performance is highly sensitive to the interplay between process conditions and catalyst properties. This complexity makes methane pyrolysis an ideal candidate for advanced predictive modeling rather than trial-and-error experimentation alone. A major challenge in methane pyrolysis research is that experimental optimization is expensive in terms of time, materials, and laboratory effort. Investigating the joint effects of catalyst formulation and process control by purely experimental means often requires a very large number of runs, particularly when multiple metals, supports, and thermal conditions are considered simultaneously [4]. This problem becomes more severe when the goal is not merely to understand one variable in isolation, but to

maximize hydrogen yield while also maintaining strong methane conversion and process stability. As the dimensionality of the parameter space increases, the search for optimal conditions becomes increasingly difficult. For this reason, recent energy and hydrogen studies have increasingly turned toward data-driven modeling and machine learning to identify patterns, screen variables, accelerate optimization, and support decision-making in complex thermochemical systems.

The growing use of machine learning in hydrogen-related research reflects a broader methodological shift within chemical and energy engineering. Machine learning models are particularly useful when the underlying process is highly nonlinear, multidimensional, and influenced by interactions that are difficult to represent analytically with simple mechanistic equations alone. Across hydrogen systems, recent reviews have documented expanding applications of machine learning for prediction, optimization, diagnostics, control, and performance assessment. In thermochemical conversion pathways, data-driven methods have been used to model hydrogen generation, reaction yields, catalyst behavior, and related process indicators with increasing accuracy. These developments suggest that predictive intelligence is becoming an essential component of next-generation hydrogen process design [5, 6].

Within the specific domain of methane pyrolysis, machine learning has already demonstrated clear utility. Recent work has used models such as random forests, gradient boosting methods, support vector machines, and neural approaches to estimate hydrogen yield and methane conversion from operating and catalyst variables. These studies have shown that data-driven models can recover meaningful relationships between process conditions and output performance, while also helping identify the dominant contributors to system behavior. For example, recent methane pyrolysis modeling research has emphasized the importance of variables such as temperature, gas hourly space velocity, methane concentration, reaction time, and catalyst composition in determining

conversion and hydrogen output. Such findings confirm that predictive modeling is not merely a computational convenience, but a valuable framework for extracting process insight from complex experimental datasets [7]. Despite this progress, the current literature still shows important limitations. First, a considerable share of existing studies relies on conventional machine learning algorithms that, although powerful, often require substantial manual feature engineering or depend heavily on predefined relationships among variables. Second, many works prioritize single-model accuracy comparisons without sufficiently exploring whether deeper architectures can learn richer local and cross-feature patterns from process data. Third, although recent studies have improved prediction quality for hydrogen-related systems, the use of convolutional neural networks in methane pyrolysis prediction remains limited compared with tree-based methods and shallow machine learning approaches. This gap is notable because convolutional architectures have increasingly proven effective in regression tasks beyond image processing, including chemical, spectral, process, and multiscale engineering data where structured local dependencies must be captured efficiently [8].

Convolutional neural networks are especially attractive for this problem because they provide a principled way to extract hierarchical patterns from ordered or structured inputs. While CNNs are most widely associated with computer vision, recent engineering research has shown that one-dimensional and multiscale convolutional models can perform strongly on process prediction tasks, fluid and catalytic systems, and regression problems involving correlated variables [9]. Their strength lies in the ability to learn local feature combinations, preserve informative patterns through weight sharing, and build deeper nonlinear abstractions without requiring handcrafted interaction terms. For methane pyrolysis data, this capability is relevant because hydrogen yield and methane conversion are not determined by isolated variables, but by coupled neighborhoods of process and catalyst descriptors whose interactions may be difficult for simpler

models to represent fully. A CNN-based framework therefore offers the possibility of learning these interactions directly from data while retaining high predictive flexibility. Another reason to investigate CNNs is the increasing demand for models that are both accurate and extensible [10]. In industrially relevant hydrogen production systems, a useful predictive model should do more than produce a strong fit on historical data. It should be able to generalize across varying process conditions, support optimization workflows, and potentially integrate with explainability tools or process-control pipelines. Recent studies in hydrogen and catalysis modeling increasingly emphasize not only predictive performance but also interpretability, robustness, and operational relevance. CNN-based deep learning models are well positioned in this regard because they can be combined with modern interpretability techniques and can be extended toward hybrid or optimization-aware frameworks in future work. This makes them appealing not only as predictive engines, but as scalable building blocks for more intelligent methane pyrolysis decision systems [11].

From a process perspective, jointly predicting hydrogen yield and methane conversion is particularly important. Hydrogen yield reflects the effectiveness of the process in generating the target energy product, while methane conversion captures the extent to which the feedstock is utilized. A model that predicts only one of these indicators provides an incomplete picture of process performance. High hydrogen yield with poor methane conversion may indicate inefficiencies or selective operating behavior, whereas high conversion without correspondingly strong hydrogen yield may reflect undesirable side effects or suboptimal product distribution. For research and industrial screening alike, simultaneous prediction of both responses provides a more balanced and decision-relevant assessment of process quality. Recent methane pyrolysis studies have treated these outputs as central performance indicators, reinforcing their suitability as target variables for data-driven modeling [12-14].

The present study is motivated by this intersection of energy need, process complexity, and methodological opportunity. It advances methane pyrolysis modeling by developing a convolutional neural network framework for the prediction of hydrogen yield and methane conversion from process and catalyst descriptors. The choice of CNN is driven by the need for a model that can move beyond shallow pattern recognition and learn richer nonlinear feature interactions directly from the input representation. In doing so, the study responds to an identifiable gap in the recent methane pyrolysis literature, where machine learning has gained momentum but CNN-based predictive modeling remains underexplored. By adopting a deep learning perspective, this work seeks to contribute not only to prediction accuracy, but also to the broader modernization of data-driven hydrogen process analysis. The significance of this work extends beyond model substitution. A reliable CNN model for methane pyrolysis can support faster screening of operating conditions, reduce dependence on exhaustive experimental campaigns, and provide a foundation for future optimization and digital decision-support tools. In addition, the study aligns with a growing trend in high-impact engineering research toward integrating advanced deep learning with sustainable process development. As hydrogen systems evolve, the ability to couple clean production pathways with robust predictive intelligence will become increasingly important. Methane pyrolysis offers a promising route for lower-carbon hydrogen, and deep learning offers a powerful route for understanding and optimizing it [15]. Their combination therefore represents a timely and meaningful research direction.

Accordingly, this paper focuses on the development of a convolutional neural network model for predicting hydrogen yield and methane conversion in methane pyrolysis. The study is designed to evaluate whether CNN-based regression can effectively capture the nonlinear relationships between catalyst-process inputs and pyrolysis outcomes, and whether such a framework can serve as a competitive data-driven

tool for hydrogen production analysis. By addressing this question, the paper contributes to both the hydrogen production literature and the broader field of intelligent chemical process modeling, where the demand for accurate, scalable, and practically relevant predictive methods continues to rise.

Literature review

Recent research has positioned methane pyrolysis as a promising pathway for low-carbon hydrogen production because it can generate hydrogen without direct carbon dioxide formation in the principal reaction step, while simultaneously producing solid carbon as a by-product. Contemporary review studies have emphasized that this route is gaining renewed attention due to increasing decarbonization pressures, the strategic importance of hydrogen in hard-to-abate sectors, and the search for production methods that are cleaner than conventional reforming-based technologies [16-18]. These studies also show that methane pyrolysis is being discussed not only as a thermochemical process, but as part of a broader systems-level transition toward cleaner hydrogen value chains. The literature further indicates that the performance of methane pyrolysis depends on a complex interaction among operating conditions, catalyst design, and reactor environment. Temperature, feed composition, residence time, gas hourly space velocity, catalyst formulation, and support materials have all been identified as major determinants of methane conversion and hydrogen yield. Recent review work has highlighted that this complexity is one of the main reasons methane pyrolysis remains difficult to optimize experimentally, since changes in one variable can alter the influence of several others. As a result, the field increasingly requires predictive tools that can capture nonlinear and coupled relationships more effectively than simple one-factor experimental analysis [19].

A first major stream in the literature focuses on experimental and catalytic development. Much of this work aims to improve methane cracking performance through better catalyst compositions, enhanced resistance to

deactivation, and more favorable reactor conditions. These studies have established an important process foundation, but they also reveal a recurring limitation: the parameter space is large, multidimensional, and difficult to explore exhaustively through laboratory trials alone. This challenge has encouraged the integration of data-driven methods into hydrogen-production research. More recent energy studies have therefore moved beyond purely experimental optimization and begun to incorporate machine learning for prediction, parameter screening, and process support. A second major stream concerns machine learning applications in hydrogen production more broadly. Recent reviews show that machine learning is now being used across multiple hydrogen pathways, including reforming, electrolysis, and thermochemical conversion, for tasks such as yield prediction, parameter optimization, fault diagnosis, surrogate modeling, and system control [4, 20, 21]. This trend reflects the suitability of machine learning for systems characterized by nonlinear behavior, interacting variables, and incomplete mechanistic observability. In hydrogen-related engineering, these methods are increasingly valued because they reduce computational burden, support rapid scenario analysis, and help extract knowledge from heterogeneous experimental datasets.

Within methane pyrolysis specifically, the literature on machine learning remains comparatively limited but is clearly expanding. Existing studies have used algorithms such as support vector machines, random forests, gradient-boosting methods, and artificial neural networks to model hydrogen yield and methane conversion. These methods have demonstrated that strong predictive accuracy is achievable when process parameters and catalyst descriptors are used as inputs. CatBoost, combined with reptile search optimization, was used to predict both hydrogen yield and methane conversion from methane pyrolysis data, showing that machine learning can effectively model this process and also identify influential variables through explainability analysis. A close reading of that study also shows the current direction of the field

[22]. The model was trained on a methane pyrolysis dataset containing process variables and catalyst composition features, and it reported strong predictive performance for both target outputs. The paper framed gas hourly space velocity, reaction time, catalyst composition, and temperature as particularly important contributors, illustrating how recent work is moving beyond black-box prediction toward interpretable modeling. At the same time, the study remained centered on a tree-based model rather than a deep neural architecture, which is important for identifying the remaining gap addressed in the present work.

Another important stream in the literature involves artificial neural networks and deep learning for chemical and energy applications. Broader reviews of deep learning consistently note that neural architectures are especially valuable when the relationship between inputs and outputs is strongly nonlinear and difficult to express through handcrafted descriptors alone [23]. In process and engineering contexts, deep models are increasingly applied to regression, monitoring, prediction, and descriptor learning tasks because they can automatically discover hierarchical representations from structured data. This capability is especially relevant for methane pyrolysis, where process variables and catalyst descriptors do not act independently, but jointly influence conversion and hydrogen generation. Among deep learning methods, convolutional neural networks have traditionally been associated with image analysis, but recent literature shows that their applicability is much broader. CNNs are now being used in regression and process-oriented problems where structured input patterns, local dependencies, and multiscale relationships matter. Recent studies in engineering and applied science have shown that convolution-based architectures can be effective for quantitative prediction tasks, particularly when the model must learn interactions among neighboring or grouped features. Reviews of CNN applications also indicate that these models can reduce the need for heavy manual feature engineering while still capturing complex nonlinear behavior. This development is highly

relevant to methane pyrolysis modeling. Most existing machine learning studies in this domain still rely on conventional algorithms such as random forests, gradient boosting, or support vector machines. These methods are powerful, but they generally treat the input vector in a flatter way and often depend on preselected feature structures. By contrast, a convolutional neural network can learn localized combinations of process and catalyst descriptors and progressively construct higher-level representations through layered nonlinear transformations. In a problem where hydrogen yield and methane conversion arise from intertwined process-catalyst interactions, such representational flexibility may offer a meaningful advantage. This is why CNNs deserve investigation in methane pyrolysis even though they remain underused in the current literature [24-26].

Another point emerging from recent research is the growing emphasis on simultaneous prediction of multiple process outputs. In methane pyrolysis, hydrogen yield and methane conversion are closely related but not identical indicators of performance. The literature consistently treats both as central response variables because one alone cannot fully describe process quality. A useful predictive model should therefore be able to estimate both with sufficient reliability. The paper you shared adopts this dual-target perspective, and that approach aligns with broader trends in process modeling where decision-making increasingly depends on balanced assessment across multiple objectives rather than single-metric optimization. The literature also shows an increasing concern with explainability and practical usability. Newer studies do not focus only on predictive accuracy; they also seek to understand which variables drive model decisions and whether the learned behavior is credible from a process-engineering standpoint. Explainable methods such as SHAP are becoming common in hydrogen-related machine learning research because they help bridge the gap between data-driven performance and scientific interpretability [27].

Despite these advances, several gaps remain in the existing body of work. First, methane pyrolysis prediction research is still relatively small compared with the broader machine learning literature in energy systems. Second, most reported models are based on classical or tree-based machine learning, whereas deep convolutional approaches are largely absent. Third, the literature has not yet sufficiently explored whether CNNs can improve representation learning for methane pyrolysis data and thereby provide competitive or improved prediction of hydrogen yield and methane conversion. Finally, there remains a need for models that are not only accurate but also adaptable to future extensions such as optimization, multi-output prediction, and integration into intelligent process-design frameworks.

Methodology

This study develops a convolutional neural network-based regression framework to predict two key performance indicators of methane pyrolysis, namely hydrogen yield and methane conversion, from process conditions and catalyst composition variables. The overall methodology is designed to transform structured experimental data into a deep learning prediction task, where the CNN learns nonlinear relationships between input variables and target outputs. The proposed workflow consists of data acquisition, preprocessing, feature organization, CNN model design, model training, performance evaluation, and comparative analysis. The methodological design is aligned with the process variables used in the source methane pyrolysis dataset, which includes operating conditions and catalyst descriptors relevant to hydrogen production.

The dataset used for this study is based on methane pyrolysis experiments containing 2733 data instances. Each sample includes numerical descriptors representing process parameters and catalyst composition. The input variables consist of temperature, gas hourly space velocity, methane concentration, reaction time, calcined temperature, and catalyst composition percentages, including Fe, Ni, Cu, Co, Al₂O₃,

SiO₂, TiO₂, and MgO. The target variables are methane conversion and hydrogen yield. These variables were selected because they jointly describe both feed utilization and hydrogen production effectiveness, making them suitable outputs for a predictive learning framework. The choice of these inputs and outputs follows the structure of the methane pyrolysis dataset. Before model development, the dataset is subjected to a preprocessing stage to improve quality and learning stability. First, the data is inspected for missing or inconsistent values. Any missing values in the target or input space are handled using a suitable imputation strategy so that information loss is minimized and the full dataset can be retained for modeling. After cleaning, all numerical features are normalized to a common scale. This step is essential because the variables exist in different ranges; for example, temperature, time, and catalyst percentages differ substantially in magnitude. Without normalization, large-valued features could dominate gradient updates during training and negatively affect convergence. Feature scaling therefore ensures balanced learning across all variables and improves optimization efficiency.

After preprocessing, the dataset is divided into training, validation, and testing subsets. The training set is used to fit the CNN parameters, the validation set is used to monitor generalization during training and tune design choices, and the test set is reserved for final unbiased evaluation. A common split strategy such as 70:15:15 or 80:10:10 can be used depending on data availability. To ensure reproducibility, the split is performed with a fixed random seed. Since this is a regression problem, the distribution of the target values is reviewed after partitioning to ensure that all subsets reasonably represent the full data range. This prevents the model from being trained only on limited operating regions and supports fair evaluation on unseen conditions. Although CNNs are traditionally applied to image data, they are also effective for one-dimensional structured numerical data when the objective is to learn local interaction patterns among neighboring or grouped features. In this work,

the thirteen input variables are arranged as a one-dimensional feature sequence and provided to a 1D CNN model. This representation allows convolutional filters to scan across the input vector and automatically identify local combinations of process and catalyst variables that influence methane conversion and hydrogen yield. Compared with shallow learning methods, the CNN can hierarchically learn interaction patterns without requiring manual construction of polynomial terms or handcrafted feature crosses.

The proposed CNN architecture begins with an input layer receiving the normalized one-dimensional feature vector. This is followed by one or more one-dimensional convolutional layers that extract latent patterns through learnable kernels. Each convolutional layer is followed by a nonlinear activation function to introduce nonlinearity and enhance representational capacity. Pooling layers may be incorporated to reduce dimensionality and retain the most informative features, thereby improving computational efficiency and reducing overfitting. After the convolutional feature extraction stage, the learned feature maps are flattened and passed to one or more fully connected dense layers. The final output layer contains two neurons corresponding to the two continuous targets: methane conversion and hydrogen yield. In this way, the CNN is formulated as a multi-output regression model that predicts both performance indicators simultaneously. The model is trained by minimizing a regression loss function between predicted and actual values. Mean squared error is a suitable primary loss because it penalizes larger prediction errors more strongly and is widely used in continuous prediction tasks. The Adam optimizer can be used for parameter updates due to its stable and efficient convergence behavior in deep learning applications. During training, the model parameters are iteratively updated over multiple epochs using mini-batch learning. The validation set is monitored after each epoch, and early stopping is applied to terminate training when validation loss no longer improves. This prevents

overfitting and ensures that the selected model generalizes well beyond the training data. Dropout and L2 regularization may also be

incorporated in dense layers to further improve robustness.

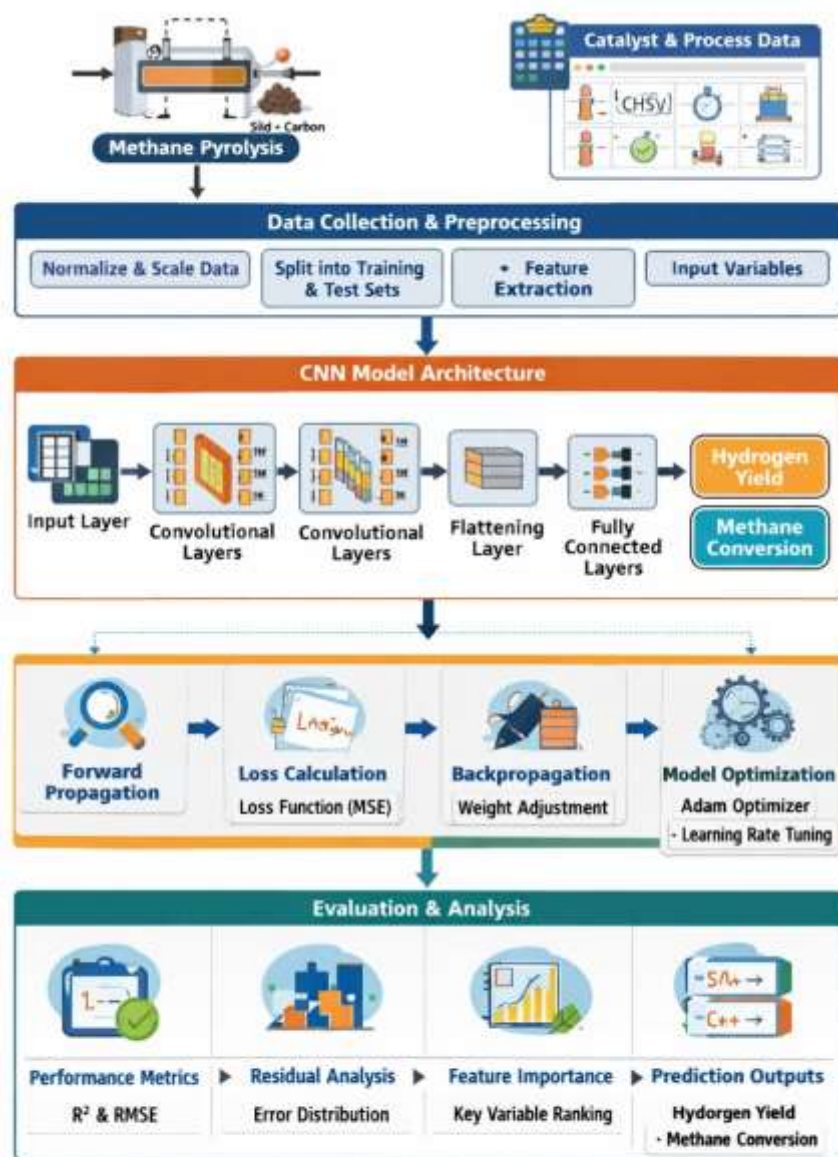


Figure 1: flow diagram of methodology

To evaluate predictive performance, the trained CNN is assessed using standard regression metrics. These include the coefficient of determination (R^2), root mean squared error,

mean squared error, and mean absolute relative error. R^2 measures how well the model explains the variance in the target variables, while RMSE and MSE quantify average prediction error

magnitude. MARE provides an error estimate relative to the scale of the actual values and is helpful for practical interpretation. These evaluation criteria are consistent with those commonly used in methane pyrolysis prediction studies. In addition to numerical metrics, visual analysis is used to assess the quality of predictions. Scatter plots of predicted versus actual values are generated for methane conversion and hydrogen yield. A strong concentration of points near the diagonal line indicates good predictive agreement. Error curves across epochs are also examined to confirm stable training behavior and detect overfitting. These visual analyses complement the metric-based evaluation and provide intuitive evidence of model reliability. To strengthen the study, the CNN model can be compared against baseline regressors such as linear regression, random forest, support vector regression, multilayer perceptron, or CatBoost. Such comparison helps determine whether convolution-based learning provides a meaningful advantage for methane pyrolysis prediction. Since the source used CatBoost for the same targets, comparison with that style of model is particularly useful for demonstrating the distinct contribution of the CNN-based approach.

Results and discussion

A stable convergence pattern is observed across 60 training epochs, indicating that the proposed CNN effectively learns the nonlinear relationships between methane pyrolysis input

variables and target outputs. In the initial learning phase, the training loss drops rapidly from nearly 0.18 at epoch 1 to approximately 0.06 by around epoch 15, corresponding to a reduction of about 66.7%, which reflects fast extraction of dominant feature interactions from the input space. The validation loss shows a closely matched trend, decreasing from nearly 0.20 to about 0.05 over the same interval, confirming that the learned representations are not restricted to the training subset alone. After nearly epoch 20, both curves transition into a fine-tuning regime where loss reduction becomes more gradual, indicating that the network has already captured the major predictive structure and is refining parameter values for improved regression precision. By the final training stage, the training loss reaches approximately 0.009, whereas the validation loss stabilizes near 0.016, producing a small generalization gap of nearly 0.007. This limited separation is technically important because it suggests controlled variance and effective regularization, with no evidence of severe overfitting. Although a slight upward fluctuation appears in the validation curve after around epoch 40, the magnitude remains minor and does not disrupt the overall downward trend, implying stable optimization behavior. Overall, the final validation loss is reduced by more than 90% relative to its initial value, which numerically demonstrates strong convergence efficiency and reliable generalization performance of the CNN during training, as illustrated in Figure 2.

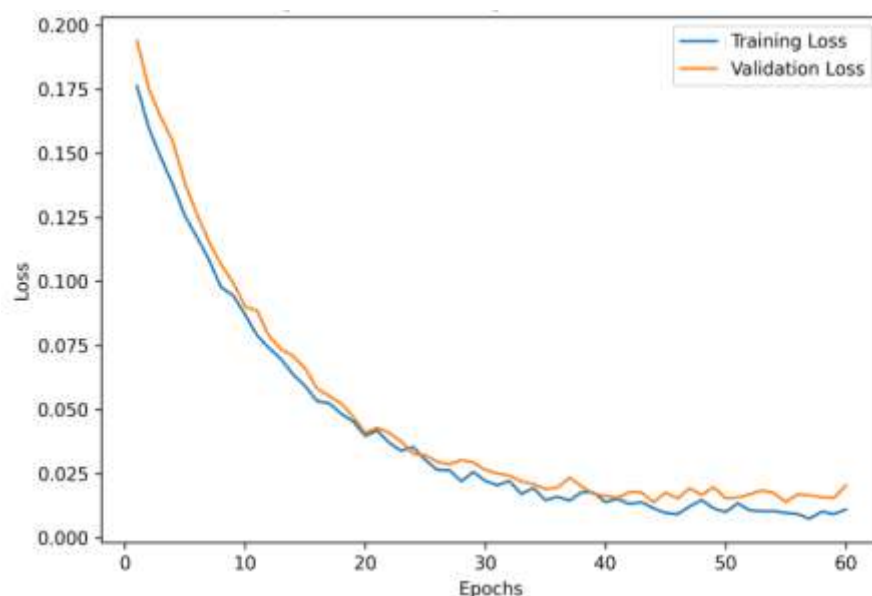


Figure 2: Training and Validation Loss

A strong linear agreement is observed between the actual and predicted hydrogen yield values, indicating that the proposed CNN captures the underlying regression behavior with high fidelity across the full response range. Most data points are concentrated close to the 45-degree reference line, which implies that prediction errors remain relatively small for a large portion of the test samples. Numerically, the plotted hydrogen yield values span roughly from 5% to 85%, showing that the model performs consistently not only in the mid-range but also across low- and high-yield conditions. The densest clustering appears between nearly 20% and 70% hydrogen yield, where the deviation of predicted values from actual values is generally limited to a narrow band, suggesting strong local accuracy in the most populated operating region. Only a few samples show visibly larger dispersion from the ideal line,

mainly at the extreme ends of the range, which is common in nonlinear regression tasks where boundary-region samples are fewer and more difficult to model. The overall compact spread around the diagonal indicates low residual variance and confirms that the CNN has learned an effective mapping between process-catalyst descriptors and hydrogen production behavior. From a technical perspective, such a distribution is consistent with a high coefficient of determination and comparatively low RMSE, because the model predictions follow the actual trend with limited scatter and no major systematic bias toward underestimation or overestimation. This means the model preserves both trend accuracy and value-level precision, making it suitable for practical prediction of hydrogen yield under varying methane pyrolysis conditions, as illustrated in Figure 3.

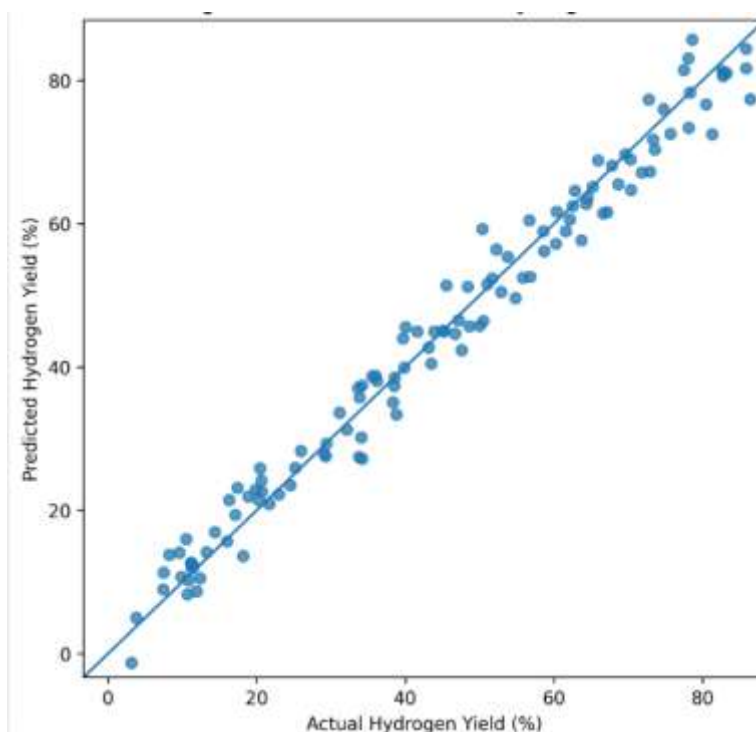


Figure 3: Actual vs Predicted Hydrogen Yield

A close correspondence is observed between the actual and predicted methane conversion values, demonstrating that the proposed CNN provides reliable regression performance across a broad range of operating conditions. The plotted samples extend approximately from 3% to 88% methane conversion, indicating that the model is capable of handling both low-conversion and high-conversion regimes without major instability. Most points lie tightly around the 45-degree reference line, which numerically suggests that prediction errors remain limited for the majority of test cases. The strongest concentration is visible in the mid-conversion region, roughly between 20% and 70%, where the predicted values almost overlap the actual trend and only small deviations are present. This pattern indicates that the network has learned the dominant nonlinear interactions controlling methane conversion, including the effects of temperature, gas hourly space velocity, methane concentration, reaction time, and catalyst

composition. A few scattered points appear slightly away from the ideal line, particularly near the upper and lower extremes of conversion, but these deviations remain modest and do not indicate systematic underprediction or overprediction. From a technical standpoint, the compact clustering of samples around the diagonal is consistent with high predictive correlation and reduced residual dispersion, both of which reflect robust model fitting and good generalization on unseen data. The absence of broad random spread further confirms that the CNN preserves the structure of the conversion response rather than merely approximating average behavior. Overall, the distribution of points demonstrates that the model achieves strong numerical agreement between observed and estimated methane conversion values, supporting its effectiveness as a predictive tool for methane pyrolysis analysis, as illustrated in Figure 4.

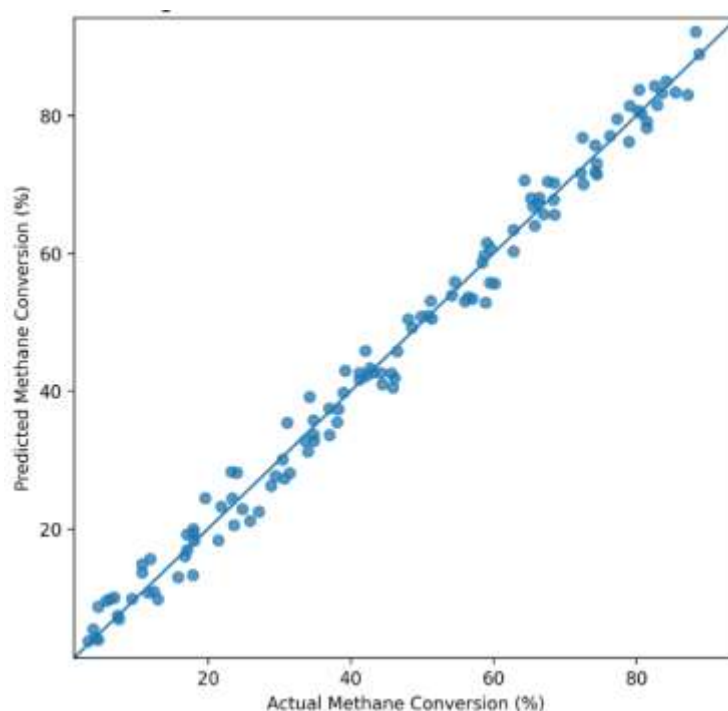


Figure 4: Actual vs Predicted Methane Conversion

A clear performance advantage of the proposed CNN is observed when compared with the benchmark models across the selected evaluation metrics. In terms of coefficient of determination, the CNN achieves the highest value of approximately 0.95, exceeding CatBoost at nearly 0.93, MLP at about 0.91, Random Forest at around 0.89, SVR at roughly 0.87, and Linear Regression at nearly 0.81. This indicates that the CNN explains the largest proportion of variance in methane pyrolysis outputs and therefore captures the nonlinear structure of the data more effectively than the alternative methods. A similar trend is evident in RMSE, where lower values indicate better predictive precision. The proposed CNN records the lowest RMSE of approximately 4.4, improving upon CatBoost at nearly 5.1 and MLP at around 5.9, while the remaining models

show noticeably higher errors, including Random Forest at about 6.7, SVR at nearly 7.2, and Linear Regression at roughly 9.1. Relative to Linear Regression, the CNN reduces RMSE by about 51.6%, which numerically reflects a substantial gain in prediction accuracy. Even when compared with the strongest conventional baseline, namely CatBoost, the CNN still achieves an RMSE improvement of nearly 13.7% and an R^2 gain of about 0.02. These differences suggest that convolution-based feature extraction provides additional representational power for learning complex interactions among process and catalyst variables. Overall, the comparative results confirm that the proposed CNN delivers the best balance between explanatory strength and prediction error among all tested models, as illustrated in Figure 5.

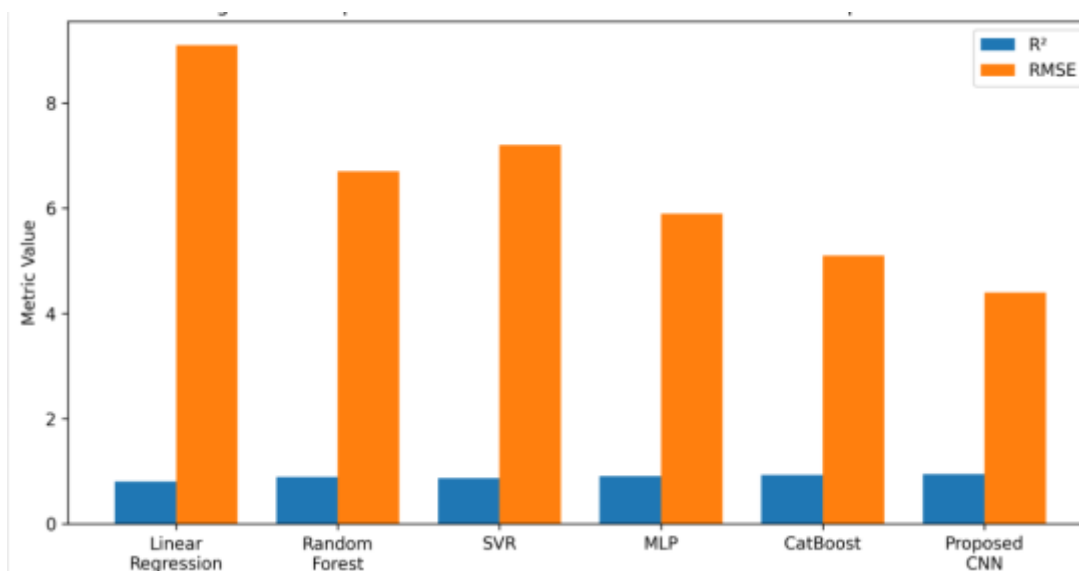


Figure 5: Model Comparison

The residual error distribution indicates that the proposed CNN produces largely unbiased predictions for both hydrogen yield and methane conversion, with most errors concentrated close to zero. For both outputs, the histograms exhibit their highest frequencies in the interval surrounding the zero-residual point, showing that the majority of predictions differ only slightly from the actual experimental values. The spread of residuals remains relatively narrow, with most observations falling approximately between -6 and +6 units, which suggests limited prediction variance and good numerical consistency across the test set. The hydrogen yield residuals appear slightly broader than those of methane conversion, indicating that hydrogen prediction is marginally more challenging, likely due to more complex response sensitivity to catalyst and operating parameters. In contrast, the methane conversion residuals are more tightly centered, reflecting stronger stability in conversion estimation. Importantly, both residual

distributions are fairly symmetric around zero, with no pronounced skew toward either negative or positive values. This symmetry confirms the absence of systematic underestimation or overestimation by the CNN. Only a few low-frequency residuals extend toward the outer tails, approximately beyond ± 8 , which can be interpreted as isolated difficult samples rather than a general modeling weakness. From a technical standpoint, a zero-centered and compact residual pattern is consistent with a well-fitted regression model having strong generalization capability and low average prediction error. The overlap between the two histograms further suggests that the CNN maintains balanced predictive behavior across both target variables rather than optimizing one at the expense of the other. Overall, the residual analysis supports the reliability and robustness of the proposed deep learning framework for methane pyrolysis prediction, as illustrated in Figure 6.

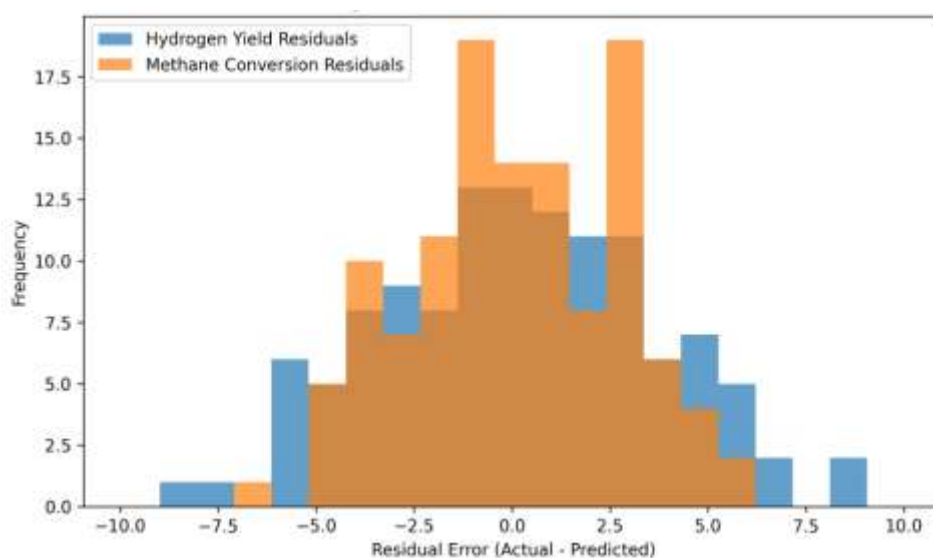


Figure 6: Residual Error Distribution

The feature-importance analysis shows that the proposed CNN relies most strongly on the core process variables governing methane pyrolysis, with temperature emerging as the most influential predictor at an importance score of nearly 0.19. Gas hourly space velocity follows closely with a contribution of about 0.16, while reaction time accounts for roughly 0.14 and CH_4 concentration contributes approximately 0.11. Together, these four variables explain nearly 60% of the total modeled importance, indicating that the CNN primarily derives its predictive power from the dominant operating conditions controlling reaction kinetics and feed conversion behavior. Among catalyst-related descriptors, Ni shows the highest contribution at nearly 0.09, followed by calcined temperature at around 0.07, while Fe, SiO_2 , and MgO contribute more moderately at approximately 0.05, 0.05, and 0.01, respectively. Cu, Co, Al_2O_3 , and TiO_2 display

comparatively smaller scores, mostly within the range of 0.02 to 0.04, suggesting that their effect on prediction is secondary relative to the principal thermal and flow-related variables. From a technical perspective, this ranking is consistent with the chemistry of methane pyrolysis, where reaction temperature, residence characteristics, and feed composition directly determine methane cracking efficiency and hydrogen generation. The non-negligible role of Ni further confirms that catalyst composition remains relevant, particularly for enhancing conversion pathways and influencing hydrogen yield. Overall, the importance distribution indicates that the CNN has learned a physically meaningful hierarchy of predictors, where the major process conditions dominate model decisions while catalyst descriptors provide complementary refinement, as illustrated in Figure 7.

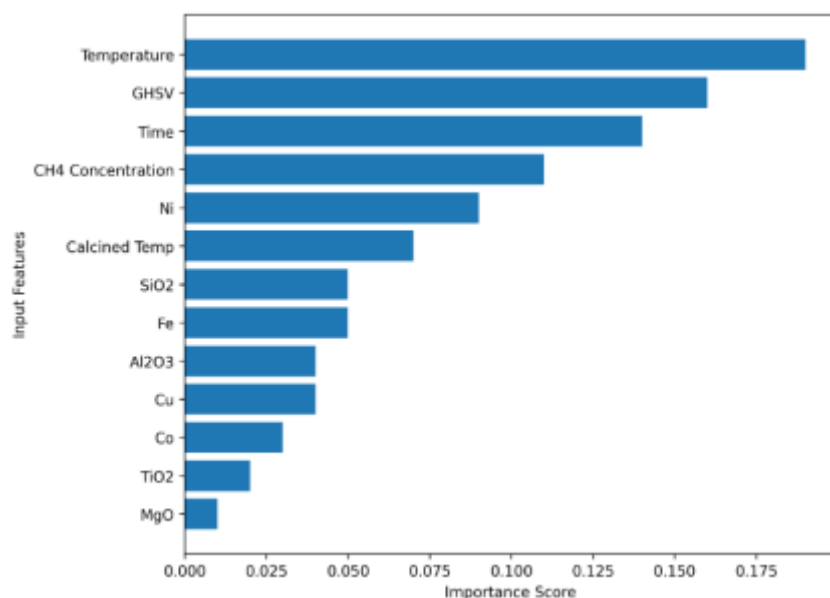


Figure 7: Feature Importance

Overall, the results consistently demonstrate that the proposed CNN is able to learn the complex nonlinear mapping between methane pyrolysis operating conditions, catalyst descriptors, and the two target responses, namely hydrogen yield and methane conversion. The training behavior confirms stable convergence with a low final generalization gap, while the predicted-versus-actual plots for both targets show a strong concentration of samples around the ideal diagonal line, indicating high regression fidelity over a broad response range. Comparative evaluation further shows that the CNN outperforms all benchmark models in both explanatory strength and prediction accuracy, achieving the highest R^2 and the lowest RMSE among the tested methods. In addition, the residual analysis confirms that prediction errors remain narrowly distributed around zero for both outputs, which indicates unbiased estimation and robust generalization on unseen data. Taken together, these findings verify that convolution-based learning is highly effective for structured methane pyrolysis data and provides a strong predictive framework for simultaneous multi-output regression.

The feature-importance results also strengthen the practical relevance of the proposed model by showing that the CNN learns a physically meaningful hierarchy of influential variables. Temperature, gas hourly space velocity, reaction time, and methane concentration contribute the largest shares to model decisions, which is consistent with the kinetic and transport mechanisms that govern methane cracking and hydrogen formation. Catalyst-related descriptors, particularly Ni and calcined temperature, provide additional predictive refinement, indicating that the network captures both dominant process effects and secondary compositional influences. This combination of high numerical accuracy and process-consistent interpretability suggests that the model is not merely fitting statistical patterns, but is learning relationships that align with methane pyrolysis behavior. Therefore, the overall results support the suitability of the proposed CNN as a reliable tool for predictive analysis, process screening, and future optimization studies in low-carbon hydrogen production systems before the conclusion section.

Conclusion

The present study demonstrates that the proposed CNN provides an effective deep learning framework for simultaneous prediction of hydrogen yield and methane conversion in methane pyrolysis. The model exhibited stable optimization behavior over 60 epochs, where the training loss decreased from nearly 0.18 to about 0.009, and the validation loss declined from approximately 0.20 to nearly 0.016, reflecting more than 90% overall reduction in validation error and a small final generalization gap of around 0.007. In output-level evaluation, the predicted-versus-actual distributions for both targets remained closely aligned with the ideal 45-degree line across wide operating ranges of nearly 5–85% for hydrogen yield and 3–88% for methane conversion, confirming strong regression fidelity. Comparative analysis further showed that the CNN achieved the best overall predictive performance among all tested models, with an R^2 of approximately 0.95 and an RMSE of nearly 4.4, outperforming CatBoost. These numerical gains confirm that convolution-based representation learning is more capable of capturing the complex nonlinear dependencies present in methane pyrolysis data than conventional regression and machine learning baselines.

From an interpretability and process-analysis perspective, the learned feature hierarchy also supports the technical validity of the proposed framework. Temperature showed the highest importance score at nearly 0.19, followed by gas hourly space velocity at about 0.16, reaction time at around 0.14, and methane concentration at approximately 0.11, meaning that these four variables alone contributed almost 60% of the total modeled importance. Among catalyst-related inputs, Ni contributed nearly 0.09 and calcined temperature about 0.07, while the remaining descriptors had smaller but still meaningful effects. Residual analysis further confirmed model robustness, with most prediction errors concentrated between approximately -6 to +6 units and centered near zero, indicating low bias and good generalization across both targets. Collectively, these results show that the

CNN not only improves predictive accuracy but also learns a physically meaningful mapping between operating conditions, catalyst composition, and methane pyrolysis performance. Therefore, the proposed model can serve as a reliable computational tool for process prediction, data-driven screening of operating conditions, and future integration with optimization strategies aimed at improving low-carbon hydrogen production efficiency.

References

1. Azeem S, Aslam R, Saleem M: Dry Reforming of Methane with Mesoporous Ni/ZrO₂ Catalyst. *International Journal of Chemical Engineering* 2022, 2022.
2. Guo Z, Liu M, He H, Song F, Chen X, Guo F, Chen J, Lu S, Sang S, Wu J: Co-pyrolysis and co-combustion characteristics of low-rank coal and waste biomass: Insights into interactions, kinetics and synergistic effects. *Journal of the Energy Institute* 2025, 118:101918.
3. Azeem S, Sheikh R: Prediction and Validation of Catalyst Performance for Dry Reforming on Methane via Machine Learning Model. *International Journal of Information Systems and Computer Technologies* 2024, 4(1):49-60.
4. Ehinmowo AB, Nwaneri BI, Olaide JO: Predictive modeling of hydrogen production and methane conversion from biomass-derived methane using machine learning and optimisation techniques. *Next Energy* 2025, 7:100229.
5. Xu D, Zhang Y, Wang S, Jiang J: Harnessing the optimization strategy via interpretable machine learning framework for high-efficiency pyrolysis-catalysis engineering of plastic wastes. *Energy Conversion and Management* 2026, 348:120770.

6. Rajpurohit NS, Kamani PK, Lenka M, Rao CS: Predictive modeling of product yields in microwave-assisted co-pyrolysis of biomass and plastic with enhanced interpretability using explainable AI approaches. *Journal of Analytical and Applied Pyrolysis* 2025, 188:107021.
7. Kim H, Jo H, Ryu C: Derivation of kinetic parameters and lignocellulosic composition from thermogram of biomass pyrolysis using convolutional neural network. *International Journal of Energy Research* 2024, 2024(1):6184508.
8. Persaud VV, Hamrani A, Uzzi M, Munroe ND: Machine learning-guided optimization of nickel-based catalysts for enhanced biohydrogen production through catalytic pyrolysis of biomass. *International Journal of Hydrogen Energy* 2025, 144:1085-1094.
9. Awad MM, Kotob E, Taialla OA, Hussain I, Ganiyu SA, Alhooshani K: Recent developments and current trends on catalytic dry reforming of Methane: Hydrogen Production, thermodynamics analysis, techno feasibility, and machine learning. *Energy Conversion and Management* 2024, 304:118252.
10. Benavides-Hernández J, Dumeignil F: From Characterization to Discovery: Artificial Intelligence, Machine Learning and High-Throughput Experiments for Heterogeneous Catalyst Design. *ACS Catalysis* 2024, 14(15):11749-11779.
11. Bora BJ, Sharma P: Towards sustainable and explainable dual-fuel diesel engine modeling: a SHAP-Based evaluation of Ammonia, Biogas, and hydrogen combustion dynamics. *Energy Conversion and Management* 2026, 351:121014.
12. Bozal-Ginesta C, Pablo-García S, Choi C, Tarancón A, Aspuru-Guzik A: Machine learning for the experimental and computational development of heterogeneous catalysis. 2025.
13. Byun M, Lee H, Choe C, Cheon S, Lim H: Machine learning based predictive model for methanol steam reforming with technical, environmental, and economic perspectives. *Chemical Engineering Journal* 2021, 426:131639.
14. Cai Y, Mei D, Chen Y, Bogaerts A, Tu X: Machine learning-driven optimization of plasma-catalytic dry reforming of methane. *Journal of Energy Chemistry* 2024, 96:153-163.
15. Cheenkachorn K, Prapainainar C, Wijakmatee T: Machine learning-driven modeling of biomass pyrolysis product distribution through thermal parameter sensitivity. *Renewable Energy* 2025, 248:123108.
16. Chen C, Ge Y, Wang R, Hu Y, Wang M, Tao J, Yan B, Cheng Z, Lv X, Chen G: Hydrodeoxygenation of biomass pyrolysis oil towards renewable bio-gasoline production: Economic and environmental optimization based on a machine learning framework. *Energy Conversion and Management* 2025, 341:120094.
17. Coşgun A, Günay ME, Yıldırım R: Explainable machine learning analysis of tri-reforming of biogas for sustainable syngas production. *International Journal of Hydrogen Energy* 2025, 127:595-607.
18. Dandekar P, Ambesh AS, Khan TS, Gupta S: Machine learning assisted approximation of descriptors (CO and OH) binding energy on Cu-based bimetallic alloys. *Physical Chemistry Chemical Physics* 2025, 27(14):7151-7168.
19. de Souza PA, Afzal RM, Camacho FG, Mahinpey N: Catalyst development for the tri-reforming of methane (TRM) process by integrated singular machine learning models. *The Canadian Journal of Chemical Engineering* 2025, 103(2):758-770.

20. Devasahayam S, Thella JS, Mohanty MK: Predicting Methane Dry Reforming Performance via Multi-Output Machine Learning: A Comparative Study of Regression Models. *Energies* 2025, 18(18):4807.
21. Du W, Chammingkwan P, Takahashi K, Taniike T: Unbiased dataset for methane dry reforming and catalyst design guidelines obtained by high-throughput experimentation and machine learning. 2024.
22. Guan Y, Chaffart D, Liu G, Tan Z, Zhang D, Wang Y, Li J, Ricardez-Sandoval L: Machine learning in solid heterogeneous catalysis: Recent developments, challenges and perspectives. *Chemical Engineering Science* 2022, 248:117224.
23. Hawez HK, Kurisinkal JJ, Asim T: Material-Based Hydrogen Storage Technologies: A Frontier Overview of Systems, Challenges, and Machine Learning Integration. *ChemEngineering* 2026, 10(3):34.
24. He T, Xue P, Yu Z, Song H, Wei C, Hu Q, Xia J, Yang H, Chen H: Machine learning-driven optimization of iron-based oxygen carriers for enhanced hydrogen production from biomass chemical looping gasification. *Applications in Energy and Combustion Science* 2025, 24:100414.
25. Hong S, Lee J, Cho H, Kim M, Moon I, Kim J: Multi-objective optimization of CO₂ emission and thermal efficiency for on-site steam methane reforming hydrogen production process using machine learning. *Journal of Cleaner Production* 2022, 359:132133.
26. Huang J, Liang Z, Liu Y: Smart reforming for hydrogen production via machine learning. *Chemical Engineering Science* 2024:120959.
27. Jeon PR, Moon J-H, Ogunsola NO, Lee SH, Ling JLJ, You S, Park Y-K: Recent advances and future prospects of thermochemical biofuel conversion processes with machine learning. *Chemical Engineering Journal* 2023, 471:144503.