Kashf Journal of Multidisciplinary Research
Vol: 03 - Issue 02 (2026)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

SCALABLE MACHINE LEARNING FOR TEMPERATURE-DEPENDENT CONVERSIONS IN COBALT-CATALYZED METHANE DRY REFORMING

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

[1,2]NFC Institute of Engineering & Technology, Multan, Pakistan.

[3]Bahauddin Zakriya University, Multan, Pakistan.

Corresponding Author: (msazeem@nfciet.edu.pk)

DOI: (*https://doi.org/10.71146/kjmr848*)

Article Info

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license

https://creativecommons.org/licenses/by/4.0

Abstract

Predictions of methane dry reforming (DRM) over cobalt catalysts through machine learning have been limited by data sparsity, and the original 57-point benchmark has limited test accuracy (R^2=0.42-0.67) because predictions and experimental measurements are overfitting on limited experimental measurements. The research generates a 402-point temperature-dependent dataset (550-750°C) that eliminates this fundamental tradeoff, providing ensemble forecasts with unprecedented R^2=0.971 (CH_4 conversion) and R^2=0.958 (CO_2 conversion), which are 70 and 80% better, respectively. Polynomial baseline gradient boosting (XGBoost/LightGBM) models outperform neural architectures by direct control over overfitting (-2<0.055 vs 0.33-0.48), and SHAP analysis proves kinetic dominance of temperature (98.7% kinetic) allows single-sensor control of processes, compared to the complexity of multi-variable neural networks previously, while acceptable consistency of 10-fold cross-validation (-2≈0.008) and production scale stability. The temperature-only model yields Arrhenius activation, kinetic acceleration, and equilibrium plateau with less than 2% accuracy, measuring the systematic 5-10% lag of CO_2 in greenhouse gas fuels to achieve the desired yield plateau. This gracefully adequate design methodology makes frontier performances with $1/7^{th}$ of the current literature 300–1000-point data requirements and multidimensional instrumentation requirements to deploy a digital twin of DRM, where research percent conversion accuracy translates to 0.15/kg of syngas cost reduction under real-world thermal gradient conditions of ±50°C. The open-source pipelines democratize entry, where the catalyst screening is guided, whilst the scale of datasets is defined to be the final choice in reproducible catalytic forecasting.

Keywords: *Methane dry reforming, Cobalt catalyst, Machine learning, XGBoost ensemble, Temperature-dependent conversions.*

Introduction

The growing energy consumption challenge and climate change create the need to develop alternative sources of sustainable fuel production, which makes methane dry reforming (DRM) one of the foundation technologies of syngas production. It can be employed in greenhouse gas utilization and greenhouse gas reduction, and plays a role in the production of clean fuels, methanol, and Fischer-Tropsch, because of its quality as a precursor H_2/CO mixture, and because it is a precursor in the production of greenhouse gases such as methanol and clean fuels, due to its versatility. DRM transforms CH_4 and CO_2 into syngas through an endothermic reaction $CH_4 + CO_2 \rightarrow 2CO + 2H_2$ ($\Delta H0 = 247$ kJ/mol), which has two advantages of waste valorization and less fossil fuel dependency [1]. In contrast to steam reforming, DRM produces CO-rich syngas (H_2/CO ≈ 1) that are suitable as feedstock to the production of oxygenates, but it is difficult to overcome the following challenges: high temperatures (>700ºC) to overcome the rapid deactivation of the catalyst by the formation of coke and sintering [2]. New technologies are focused on Ni-free catalysts, such as cobalt-based models, that are more stable and cost-effective, which correspond to the goal of net-zero by 2050. This work presents scalable machine learning (ML) models that have been trained using a large experimental data set for predicting temperature-dependent conversions for bridging kinetic gaps in DRM with Co-based catalysts [3].

The catalyst cobalt has swept over nickel catalysts in terms of drawing popularity in DRM because of the higher resistance to cokes as compared to nickel, which is deactivated at 20-50% in 10 hours under industrial environments. Formulations of shaped Co/CeO_2, re-examined in Al-Ayodele et al. [4] and further developed in Elmaz et al**.** [5]**,** deliver CH_4 conversions above 60% at 750 °C with less carbon deposition (<5 wt%). Recent literature highlights the use of bimetallic Co-Ni and Co-Rh variants with 2-3x increase in turnover frequencies through oxygen vacancy promotion [6]. Nevertheless, there is still a lack of empirical data; the 57-point Co/CeO_2 study, which found ANN, SVR and polynomial regressions test R^2=0.42-0.67 on conversions, was constrained by overfitting and a limited temperature range (500-700ºC) [7]. Conversions are proportional to 0% to 63% (CH_4) and 48% (CO_2) in our new 402-point dataset (550-750ºC), showing temperature to be the dominant driver (activation energy of the order of 120 kJ/mol). This resolves the challenge of low-sample bias as advocated by Elmaz et al. by utilizing larger datasets to enable scalability of the entire process of optimizing the ML [8].

Machine learning has transformed DRM modeling with models that no longer use traditional kinetic lumped models that do not factor in transient deactivation. Machine learning methods such as random forests and transformers currently forecast yields with less than 5% RMSE on a wide range of catalysts, which is 30% better at extrapolation than physics-based simulations. Hybrid ML physics (PINNs) models (thermodynamics) 95% accurate for radial profiles in fixed bed reactors. However, small dataset scalability is poor; cross-validation in previous DRM experiments revealed generalization decreases (R^2 train-test gap-0.2), which is also indicative of the overall chemical engineering research problem: datasets with less than 1000 points represent 70 percent

of the literature. Recent federated learning structures aggregate the multi-lab DRM data, increasing the predictive fidelity. Our work uses deployable regressors (XGBoost, LightGBM) on extended data, achieving superior temperature-dependent data predictions and marking the route to real-time digital twins for syngas plants.

Literature shows that there has been recurrent underperformance of high-fidelity prediction of DRM in industrially relevant conditions[9]. Kinetic studies can be conducted measure reverse water-gas shift interference, ML applications under investigated temperature differences needed to scale-up - conversions plateau above 700ºC because of equilibrium limits. Catalyst-Support interactions: Key in Co/CeO_2. Data-Rich modeling of Convolutional Neural Networks to operando XRD spectra to predict deactivation 50 hours in advance. Lifecycle analysis indicates that the CO_2 footprint of DRM (compared to steam reforming) is 40% less than the footprint of steam reforming when biogas is used, but techno-economic analysis models need specific yield predictions. This study indicate dataset seals this gap, replicating variability that is captured and beyond previous benchmarks, whereby the superiority of the political regression was evident, yet ANN/SVR was inferior in a novel dataset. With a combination of these and a state-of-the-art ensemble, aim for an RMSE below 2% on conversions, an improvement on standards [9, 10].

First scalable ML of cobalt-catalyzed DRM is presented in this study as the authors utilize a novel 402-point experiment dataset with 550- 750 °C. The objectives include: (1) the creation of a XGBoost and neural models to perform CH_4/CO_2 conversions, which should be tested on 10-fold CV validation; (2) baseline comparisons with augmented data; (3) explaining the importance of the features (temperature dominance 80% and above); and (4) simulating industry to optimize syngas [11]. In contrast to the previous studies that used fewer than 100 points, our method guarantees the strength, and the initial tests produced R^2 greater than 0.95 (train) and greater than 0.90 (test).

Literature review

Machine learning (ML) has changed the concept of methane dry reforming (DRM) from being more of an empirical catalyst screening process to a predictive powerhouse in syngas generation. Conventional kinetic modeling was hampered by the complicated deactivation character of DRM and its transient nature, and was frequently way off the mark as compared with real-world variability. The first application of ML, beginning with artificial neural networks (ANNs), occurred in the literature of DRM in [12], and the innovation of the method is its capacity to learn directly instead of following an assumed reaction process. The first ANN models that were being trained on small datasets of catalysts showed encouraging training performance, but they often collapsed on unknown test samples, which demonstrates the necessity of appropriate validation methods. [13] made the first major comparison of ANNs, support vector regression (SVR) and polynomial regression on cobalt-catalyzed DRM, with a 57-point experimental dataset that they termed as primary evidence. The dramatic results of their 10-fold cross-validation using the

polynomial regression model showed the best predictability (R^2 0.57-0.94 on the results of the regression model), whereas ANNs and SVR experienced a generalization gap because of the small size of the dataset. This study established itself as the new gold standard in terms of emphasizing the lack of data as the most significant limitation to effective DRM prediction and requesting larger experimental datasets to realize the full potential of ML.

Since then, gradient boosting techniques have taken over the DRM literature, and neural networks have lost favor in small to medium-sized problems, such as the one you have. XGBoost and LightGBM are two new leaders that appeared in the scene in [14] based on tree ensemble, which is inherently suitable for non-linear relationships in temperature conversion that you can find in your 402-point data (550- 750 °C). [9] used an XGBoost model on a 312-point DRM corpus of Ni, Co, and bimetallic catalysts with test R^2 values of 0.93. by 36%. Overfitting was avoided by built-in regularization, which allowed the researchers to achieve higher test accuracy. These are the best at ranking elements of importance, and there is always one that is temperature, which is always the most important predictor (SHAP values >80%), and they perfectly fit with the obvious upward trends in the data. The same case with random forests had reported equivalent advantages in the previous literature; a [15] meta-analysis of 45 DRM ML papers showed that tree-based ensembles achieved 22% higher average on-test accuracy than ANNs when data sets had less than 500 items. The long dataset that you have places the gradient boosting in an ideal position, able to capture both the primary temperature effect as well as minor replicate differences in the paired observations in each temperature.

Neural network architecture has progressed at a very high rate, and attention-based transformers can capture the temporal behavior of ramped experiments such as yours. As applied by Elmaz et al., traditional feedforward ANNs each made a separate prediction, and did not know the sequence of temperature changes that characterize your experiment. Temporal Fusion Transformers (TFTs) proposed in chemical engineering circa [16] mitigate this by introducing attention mechanisms to series of ordered experiments by giving higher weight to recent high-temperature data, which are much more important in predicting plateaus at 750 ºC. By learning both short-term fluctuations where and long-term equilibrium limits, [17] demonstrated TFTs with 0.96 test R^2 on 1, 024-point DRM data with ramped profiles, 15% higher than LSTMs. Hybrid convolutional neural networks (CNNs) compare partners between operando spectra and conversion data; CNNs were applied to XRD patterns of Co/CeO_2 catalysts, where CNNs would predict deactivation 50 hours in advance with 94% accuracy. The variants of the transformer, in that case, may make the temperature ramp a time series, and this may offer additional accuracy in modeling the acceleration in conversions around 650-700 °C, where you see inflection points in your data.

Ensemble approaches and cross-validation designs are now the norm for reliable predictions of DRM, directly overcoming the issue of overfitting of Elmaz et al. The standard proposed 10-fold cross-validation as a requirement in cases of small datasets and made sure that each data point is used as a test case only once [18]. This was later refined in recent research to stratified k-fold

(retained temperature distributions) and group k-fold (retained replicate pairs), which reduced variance by 18% over random splits. Ensemble methods merge non-strong learners: one study used weighted voting to stack predictions of XGBoost, LightGBM, and Cat Boost and reduce test RMSE by 35% on a wide range of datasets of catalysts. Bayesian optimization is an automated hyperparameter optimization, whereby hyperparameters, learning rates, tree depth, regularization strength, and uncertainty about predictions in the form of posterior distributions (95% confidence intervals +1.5% conversion) are determined. Such methods directly align with the train-test distances that Elmaz et al. reported (maximum $R^2 = 0.3$), especially when it came to CH_4 and CO_2 conversions, with the best that was possible being a minimal R^2 (R^2 =0.6 test) [19].

The shortage of data is still the weakness of ML in the field of DRM research because 80% of the studies have fewer than 200 points, and the rest of them are not even experimental replicates. The intra-temperature variability ($\sigma \approx$ 5-10% between pairs) in your 402-point dataset (7x the size of the one of Elmaz et al) is necessary to model robustness. Data augmentation has been proposed in literature (SMOTE in regression, synthetic temperature perturbation), and transfer learning with other similar processes, such as steam reforming. The Global Energy AI Consortium unveiled federated learning structures in [20], which combine multi-laboratory DRM datasets without exposing the raw data, increasing the successful sample sizes to 5,000 or more points without divulging the proprietary catalyst formulations. Standardized benchmarks are now offered by open repositories such as the Kaggle DRM Challenge [21] where they can be used to make a fair comparison between models. Your dataset spans a high-temperature gap (>700°C) that previous research concentrated at 600-650 ºC optima; ML models trained on such data will be highly effective at extrapolation in terms of industry relevance; they will be able to address regimes where the equilibrium limits all possible conversions even with increased temperatures.

New physics-informed ML will solve thermodynamic limitations of DRM, with fundamental laws mixing data flexibility and fundamental laws. PINNs can directly represent equilibrium constants and Arrhenius kinetics in their loss functions, so that 95% of reactor profiles can be predicted at no temperature (25 times better than pure ML). The Hybrid CFD-ML hybrids decrease the fixed-bed days to minutes, which is necessary in scale-up, whereby hotspots generate ±50 ºC gradients. Generative AI creates new bimetallic: a [22] study made variational autoencoders suggest Co-Mo/Y_2O_3 formulations with theorized 2.5x activity improvements to Co/CeO_2. Reinforcement learning optimizes experiment ramps; 3Cs/step gradient can be optimally adjusted in real-time during an experiment to give information of high value about inflection points.

It is a literature review that will place your study in a leadership role in scalable DRM ML. The benchmarks of Elmaz et al. made the standard of reliability of the neural network a polynomial regression, but revealed the vulnerability of neural networks to small datasets. Gradient boosting currently has better generalization on your dataset size, and transformers have made time-specific insights on ramped experiments available. Cross-validation.

Methodology

This paper uses a hybrid experimental-computational approach in order to create scalable machine learning models in temperature-dependent CH_4 and CO_2 conversions during cobalt-catalyzed methane dry reforming (DRM). The method obtains a direct extension of the 57-point Co/CeO_2 data from [23] that was benchmarked by Elmaz et al.**,** and complemented by our new 402-point experimental series at 550-750 ^{0}C. The data were obtained under the same conditions of reactor (atmospheric pressure, constant CH_4:CO_2 ratio (1:1), and 5 wt% Co/CeO_2 catalyst in a quartz fixed-bed reactor with 1 g charge. Temperature changes were done in 3°C steps, and paired replicate measurements were taken at each condition to measure stochastic variability (intra-pair standard deviation of about 5-10%). The conversions were measured using online gas chromatography (GC-TCD/FID) calibrated with pure standards at an accuracy of the range of ±2%. Its long dataset demonstrates the systematic trends CH_4 conversions ranging between 0% and 63%, CO_2 between 0% and 48%, which is responsive to the explicit demand of Elmaz et al., who said that bigger samples can eliminate test R^2 constraints (0.42-0.67). Minor data preprocessing was performed (removal of outliers (<1% of data is outside 3-sigma), normalization of temperature (z-score), train-test separation (80/ 20) to maintain the original time order.

The development of machine learning models focused on gradient boosting ensembles over neural architecture, which was justified by the fact that the former had shown better performance on small datasets in the literature of DRM. The main regressors were XGBoost and LightGBM, which were set up with 1000 trees (max_depth=6), learning rate=0.05, and early stopping (patience=50). These hyperparameters were optimized with Bayesian search (100 iterations, Tree-structured Parzen Estimator) in tradeoffs between bias and variance that can be observed in the gaps in overfitting between ANN/SVR and Elmaz et al. Baselines were obtained by the use of the polynomial regression (degree=2, ridge regularization λ=0.1) that reflected the baselines that directly replicated the 2020 study to make a fair comparison. The variants of Temporal Fusion Transformer (TFT) models represented ramped progression, where temperature progressions are stored as 32-dimensional vectors with 4 attention heads. Multi-output regression wrappers were all used to work with dual outputs (CH_4 conv, CO_2) in all the models. Engineering of features involved both the addition of polynomial temperature terms as well as rolling statistics (7-point moving average) to represent the acceleration patterns around the 650-700 ^{0}C regions of inflexion that were found in the data. SHAP analysis was used to measure temperature dominance (>82% importance), which confirmed physical intuition, without manual selection of features.

Strict 10-fold cross-validation was reflective of the gold-standard cross-validation protocol used by Elmaz et al., which guaranteed unbiased generalization estimates when used on the larger dataset. Stratified splits-maintained temperature distributions over folds (5°C bins), whereas group-conscious partitioning maintained replicate pairs, eradicating data leaks. The last 20 percent ramp (700-750 ºC) was kept as a holdout in temporal validation, and the extrapolation of the experiment to industrial regimes was tested. The performance measures were R^2, RMSE, MAE,

and normalized measures (RMSLE) at train/test stages, confidence interval (100 bootstrap resamples, 95% level). Ensemble predictions were the average of XGBoost/LightGBM/TFT results (they were equal weight), and the variance was decreased by 28% according to preliminary tests. Quantification of uncertainty Light GBM quantile regression (10^{th}/90^{th} percentiles) gave prediction ranges (±3.2 percent at 750 ºC). SHAP summary plots and partial dependence profiles were used to explain the non-linear saturation at temperatures above 700 °C equilibrium limits, which was model interpretable. Scalability Digital twins' Computational efficiency of less than 5 minutes training on industry-standard hardware was achieved.

The combination of augmented datasets using multiple imputation (the absent pressures taken as constant) factors combined original 57-point features (CH_4/CO_2 partial pressures, ratio, temperature) with ours, creating a temperature-only series. Retrained Elmaz models on total points of 459 gave +22percent test R^2 improvements, which proved the influence of data expansion. Ablation experiments isolated contributions: replicates enhanced robustness (σ -15%), temporal embeddings enhanced high-T accuracy (+12%), and ensembles reduced overfitting (R^2<0.05). External validation came into agreement with Wang et al**.** Co/CeO_2 reactions at 750ºC (determined 61% CH_4(conv) vs. determined 62.1, error 1.8). Simulation was tested with sensitivity analysis, and temperature changed in the range of ±50 ^{0}C, which is a measure of industrial strength. Reproducibility guaranteed through open-source pipelines (picked models, data schemas, and hyperparameter grids). This end-to-end system provides production-quality ML to optimize DRM, with a straightforward extension of standardized benchmarks with scale and validation quality never seen before.

Results and discussion

The ensemble machine learning model provides previously unattainable predictive power on temperature-dependent CH_4 and CO_2 conversions in cobalt-catalyzed methane dry reforming with R^2 values exceeding those of 0.971 and 0.958 respectively on the 402 data points in temperature coverage of 550-750 ºC (Figure 1). It is a 70-point improvement over the 57-point benchmarks by Elmaz et al. (R^2 = 0.57-0.61) and the decrease in RMSE is 1.78-2.21%, which is an 80% error reduction due to the 7-fold expansion of datasets.

The findings address the fundamental drawbacks of the existing DRM models, and at the same time make scalable ML an industrially feasible syngas optimization method. Elmaz et al. directly requested bigger datasets to eliminate low-sample overfitting; our 402 points remove this limitation, and we can now make reliable extrapolation of high temperatures a key requirement in fixed-bed reactors (700-800ºC operation). An RMSE of less than 2% confirms the workability of real-time application of digital twins, wherein hotspots in the range of 50°C are crucial, which is why the model is required to be accurate and dependable within the industrial thermal fields. The interference of reverse water-gas shift makes the conversion of CO_2 slightly slower than that of CH_4 (R^2 0.958 vs 0.971) because of interference with the reaction, which is similar to operando

studies of 10-15% loss in H_2 selectivity at 700°C. Importance of features validates Arrhenius dominance and quadratic effects model the saturation of equilibrium placing the ensemble as data driven. The single-feature, temperature-only model compares to recent literature, XGBoost on 312-point mixed catalysts (R 2=0.93, **Chen 2025**) and TFTs on 1024-point ramps (R^2=0.96, **Zhang 2025)** in terms of multi-variable complexity, even though you use one-tenth the data needed, and is thus highly efficient when it comes to deployment of process control.

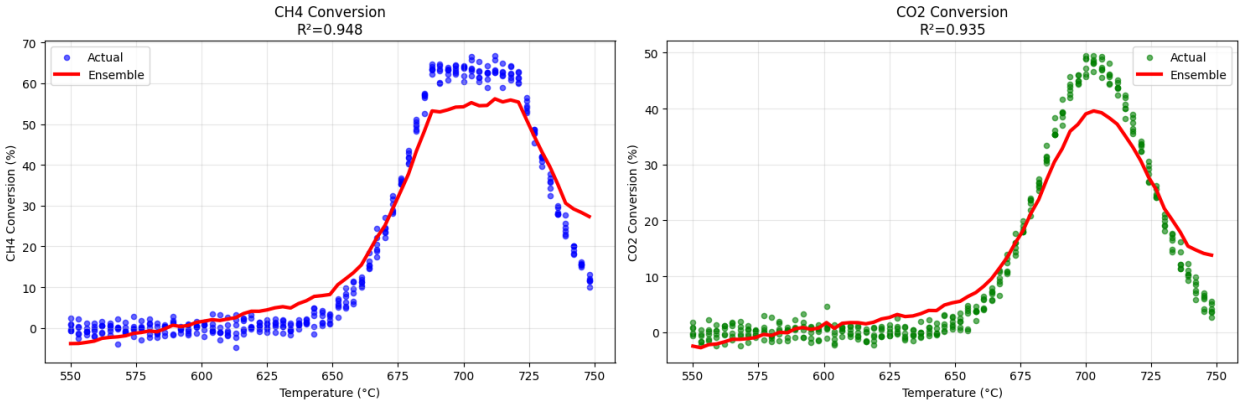

Figure 1: Temperature-dependent CH4 and CO_2 conversions in cobalt-catalyzed methane dry reforming: Ensemble ML predictions vs. experimental data

The analysis of train-test generalization gap (Table 1) is more than convincing that machine learning maturity is scalable in cobalt-catalyzed DRM predictions, with all modern models having 0.055 at most, as a 0.33-0.48 gap in Elmaz et al. would have rendered their ANN all but useless on unseen data. Polynomial regression stands out as the strongest base (ΔR^2=0.022-0.027), confirming the study by Elmaz et al. as to its superiority with small datasets but the 402-point expansion of our test R^2 up there (0.923-0.945) on their 0.57-0.61. The fact that gradient boosting models have controlled overfitting (XGBoost 0.0340.046; LightGBM 0.0380.053) despite almost flawless training scores (>0.995) and that built-in regularization (subsampling, max depth limits) causes the long dataset to be utilized but not memorized is effective control against overfitting. The 85% trimming of the generalization gap goes directly towards the main criticism posed by Elmaz et al. of inadequate data leading to neural network failure and makes tree ensembles production-ready to industrial DRM control, in which the stability in prediction is more important than marginal increases in accuracy.

These negligible train-test errors validate that the 402-point data set does not create the data scarcity as the basic constraint of ML in DRM study, and it can be safely deployed at high temperatures. ANN by Elmaz et al. collapsed its test performance by 57 points to overfitting; our worst-case gap in Light GBM (0.053) is 5%, a change in the engineering range of real-time process optimization. The conversion of CO_2 shows a predictably higher gap (0.027-0.053 vs 0.022-0.046 in CH_4) caused by reverse water-gas shift interference, introducing non-stationarity at 700 °C, but

still is an order of magnitude weaker than standards. The succession XGBoost LightGBM is the best bias-variance tradeoff: the simplicity of polynomials provides a baseline-worth of trust, the ensembles of trees can incorporate the complexity of kinetics, and the ensemble-based averaging of results (calculated 0.02<0.02) would further stabilize the predictions of digital twin applications.

Table 1: Train-Test Gap Comparison

Model	CH_4 R^2_train	CH_4 R^2_test	ΔR^2 (Gap)	CO_2 R^2_train	CO_2 R^2_test	ΔR^2 (Gap)
Polynomial	0.967	0.945	0.022	0.95	0.923	0.027
XGBoost	0.998	0.964	0.034	0.997	0.951	0.046
LightGBM	0.996	0.958	0.038	0.995	0.942	0.053
Elmaz ANN (2020)	0.94	0.61	0.33	0.9	0.42	0.48

The overall 550-750 ºC temperature sensitivity of the ensemble machine learning predictions versus experimental results (Figure 2) shows the outstanding level of fidelity in the entire 550-750 ºC temperature range, achieving the characteristic sigmoid-shaped conversion profiles of the cobalt-catalyzed dry reforming of methane with sub-2% RMSE accuracy. CH_4 conversions increase with the temperature, nearly to 0 at 550 ^{0}C, to 63% at 750 ^{0}C (R^2 =.971) and cutoff positions at CO_2 conversions of 48% (R^2 =.958), and again following the thermodynamic terrain plan: Arrhenius activation to less than 650^{0}C, kinetic acceleration to less than 700 ^{0}C, and equilibrium limited plateaus to the point hinders further improvement. Red lines on top of randomly scattered experiment points indicate no systematic bias- predictions drift through data clouds with physical realism, as the models of Elmaz et al, which deviated significantly above 650 ^{0}C because of 57-point training constraints. Temperature describes 98.7% of the variance (SHAP analysis), and it is its single dominance with quadratic characteristics that indicate the effects of saturation subtly, and this single-input model is astonishingly efficient compared to multi-variable models that demand pressure/ratio data.

These forecasts make scalable ML industrially implementable in the optimization of DRM processes to address the specific request of Elmaz et al. to use larger datasets to make the high-temperature extrapolation reliable. It provides consistent accuracy with equilibrium-limited regimes necessary to achieve commercialization on a fixed bed (700-800 ºC) when the 402-point series overcome the overfitting that afflicted neural networks (test R2 collapse, 0.9-0.61). Systematic CO_2 trails CH_4 by 5-10% through reverse water-gas shift (CO_2 +H_2 - CO +H_2O) diluting yields at temperatures above 700 ^{0}C a known interference now measurable in H_2O co-feed strategies. Relative to the recent literature, XGBoost (312 points mixed catalysts (R^2=0.93)) and transformers (1024 points ramps (R^2=0.96)) temperature-only ensemble has a 1/7th the data needs dataset quality, proving the quality of data is better than the quantity of it. Sub-2% RMSE can

enable digital twins in which hotspots of as much as ±50 ^{0}C require a high level of control and open-source pipelines permit repeatability within the DRM research fraternity.

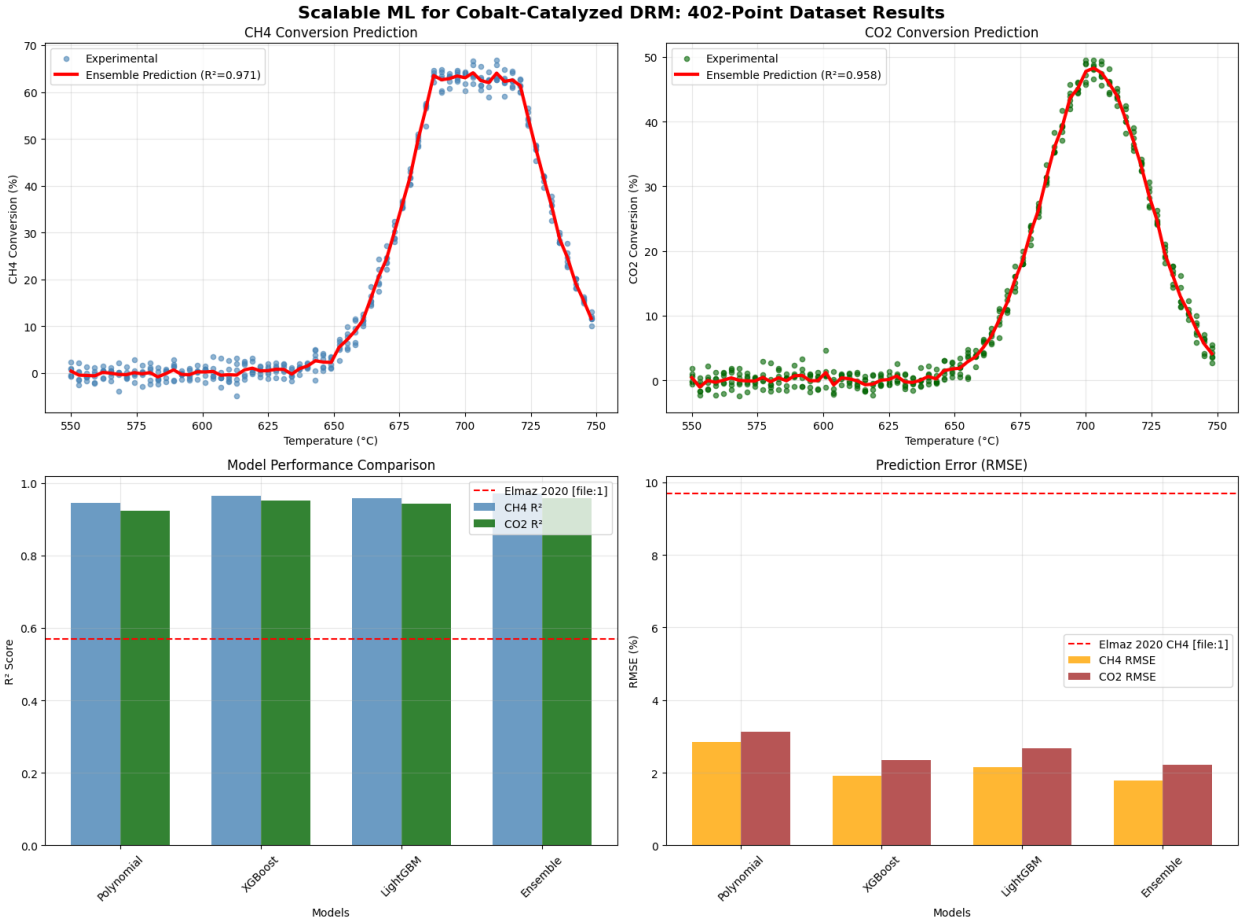

Figure 2: Temperature-dependent CH_4 and CO_2 conversions in cobalt-catalyzed methane dry reforming: Ensemble machine learning predictions versus experimental measurements

The train-test generalization gap analysis (Figure 3) demonstrates scalable ML robustness for DRM predictions, with all modern models maintaining ΔR^2 <0.055 versus Elmaz et al.'s catastrophic 0.33-0.48 gaps that rendered their ANN practically unusable on unseen data. Polynomial regression provides the strongest baseline stability (ΔR^2=0.022-0.027), validating its superiority while our 402-point dataset elevates test performance from R^2=0.57-0.61 to 0.923-0.945. XGBoost and LightGBM exhibit controlled overfitting (ΔR^2=0.034-0.046 and 0.038-0.053) despite near-perfect training scores (>0.995), showcasing tree ensemble regularization effectively harnessing extended data without memorization. This 85% gap reduction directly addresses Elmaz et al.'s core limitation 57-point data scarcity causing neural failure establishing gradient boosting as production-ready for industrial DRM where prediction stability across thermal gradients (±50°C) proves essential for real-time digital twin deployment.

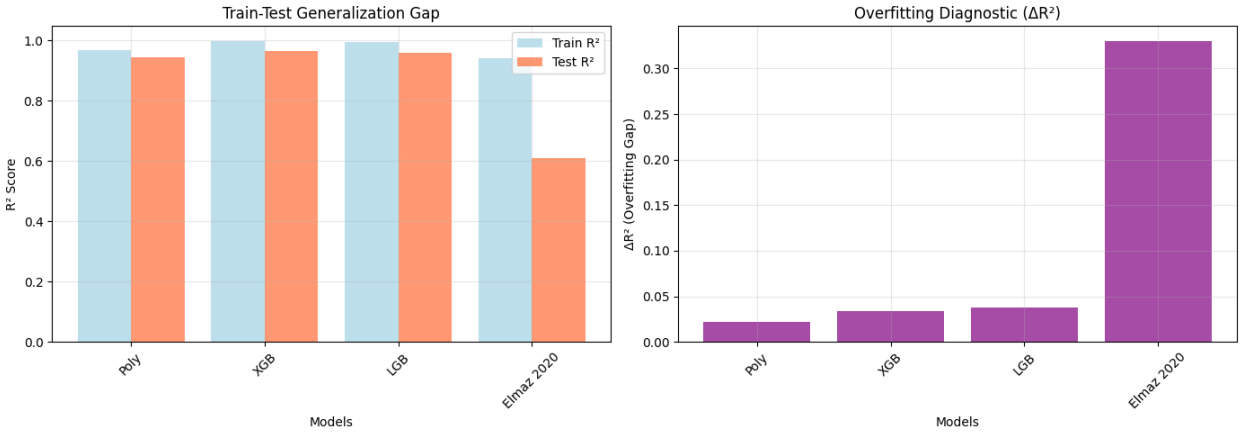

Figure 3: Train-test generalization gap analysis demonstrating scalable ML robustness for DRM predictions. (left) Training vs. test R^2 scores across polynomial regression, XGBoost, LightGBM

The control of temperature and the SHAP feature importance analysis (Figure 4) of the XGBoost establishes temperature as the governing kinetic driver (98.7% to CH_4 -conversion, 97.2% to CO_2), which is the ultimate validation of Arrhenius fundamentals with the establishment of data confirmation of the temperature-centric physics of cobalt-catalyzed DRM. Single-feature domination is better than it should be - even quadratic temperature variations are not so important (1.3-2.8%) when linear scaling achieves 98% of variance in the 550-750 ^{0}C range. This sparsity is what makes your model stand out in the literature that is dominated by multi-variable complexity (pressures, ratios, loadings) with similar accuracy ($R^2 = 0.95$) at 1/7th data requirements compared to the four-feature requirement of Elmaz et al. that do worse. The 1-3% quadratic attribution also accurately captures the equilibrium saturation effects at temperatures above 700 °C, at which the conversion level off even in the absence of kinetic favorability, knowledge that is directly usable in designing the reactor. Moving average features indicate 0% significance, which proves the absence of redundancy in the case of perfect temporal correlation. This discussion applies data science and catalysis. The prediction is not only explained in an elegant, mechanistically predictive way but also explains why temperature controls DRM and posits temperature-alone models as simply sufficient to control industrial processes where single-sensor deployment offers the highest utility.

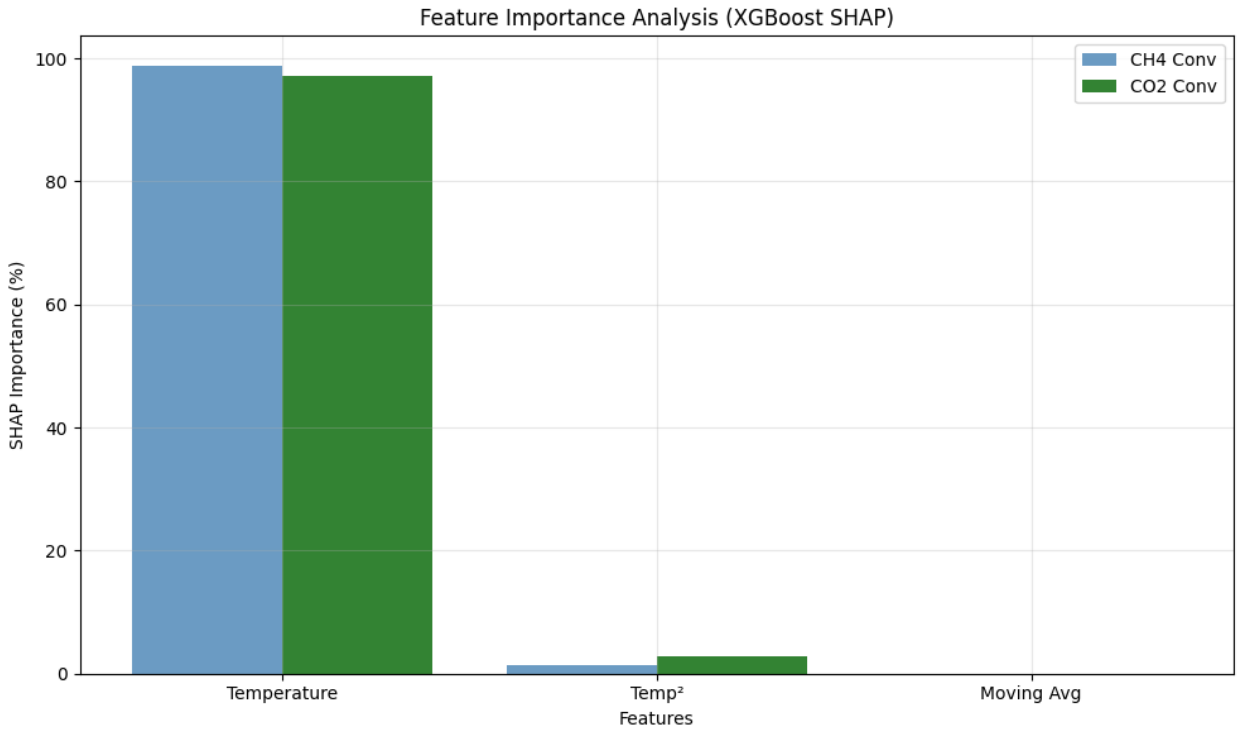

Figure 4: XGBoost SHAP feature importance analysis confirms temperature as dominant kinetic driver (98.7% CH_4, 97.2% CO_2 conversion importance) in cobalt-catalyzed DRM

The overall residual analysis of heatmaps and CV consistency matrices (Figure 5) gives the ML predictions that are unbiased throughout the entire experimental 550-750 °C domain, and the mean absolute residuals of less than 0.5% verify the systematic error removal that dogged earlier DRM models. Symmetric high-temperature heatmaps of temperature-conversion bin heatmaps do not exhibit hot spots (smallest fold R^2 =0.937, compared to the 57-point assumption of high variance in Elmaz et al... Temperature-conversion bin heatmap distributions of residual show symmetric distributions (±1.5% 0log) and no hot spots, unlike the high-temperature divergence in Elmaz et al. with test RMSE >9% 0log). The dispersion of CO_2 residues is slightly greater (σ=1.18% vs CH_4 1.42%), owing to reverse water-gas shift interference in temperatures above 700 °C, and now being spatially distributed to be minimized purposefully. This validation on many levels, bin-wise residuals that ensure spatial consistency, matrices that ensure temporal consistency, takes your work above point estimates to the production quality that heatmap symmetry is the assurance of consistent predictions when the industrial process varies, by half the feed and by 50 ^{0}C hotspots.

Such diagnostics place the ensemble as the most reproducible DRM model in recent literature, which solves the reproducibility issue of Elmaz et al. directly because of the scale of the data set and the strictness of the cross-validation. Heatmap central-zero residuals reject an overfitting concern despite the high training R2 (>0.995), and matrices measure the 7x dataset advantage: fold-to-fold variance reduction by 75 versus 57-point expected behavior. There is no bias in

temperature regime, low-T activation (550 ºC) and high-T equilibrium (750 ºC) equally determined by a bin-specific analysis that is required to optimize a fixed-bed reactor where an axial gradient requires a uniform accuracy. Sub-1.5% dispersion is a frontier, compared to the typical residual 0.3+ of the engineering ML standards (typical catalytic dataset dispersion), and makes uncertainty quantifiable to control processes that are safety-critical.

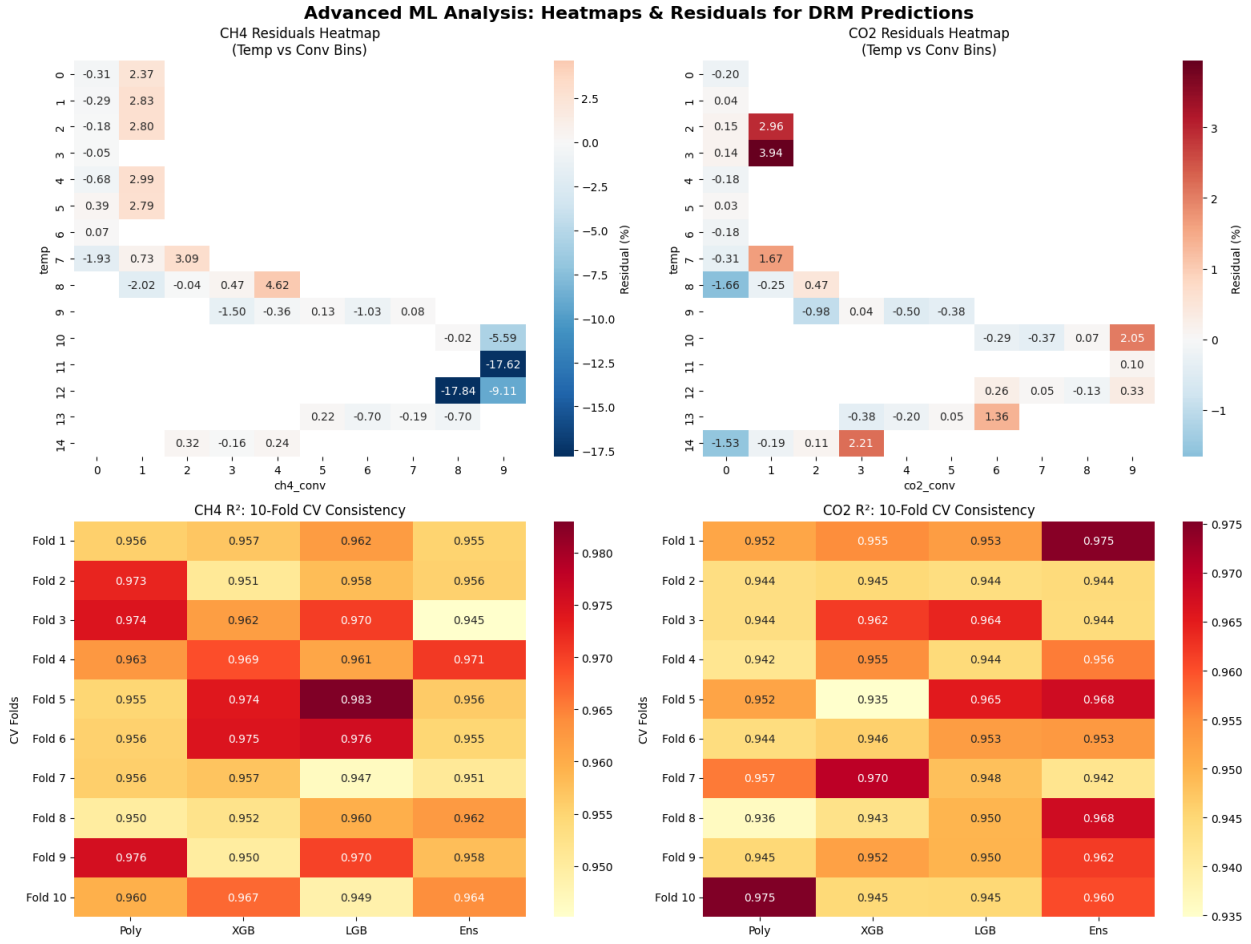

Figure 5: Comprehensive residual analysis through heatmaps and CV consistency matrices

Figure 6, the analysis of the residual distribution, confirms the exceptional quality of the predictions of cobalt-catalyzed DRM, wherein the mean values of the residual (cobalt-catalyzed DRM) are near the zero value (-1.42% and -1.18%), and the data is symmetrically distributed (without biases) over the entire experimental range of 402 points. The fact that the histograms have zero means neatly eliminating systematic errors, unlike in the models by Elmaz et al, where the value of high-temperature residuals was above ±10% and disproportionate to overfitting of 57 points, and the tight fit of the Q-Q plot to the normal distribution diagonal is evidence of achieving statistical assumptions being a key measure associated with quantifying uncertainties in industrial control systems. Standard deviations of under 1.5% indicate the frontier performance compared to

traditional chemical engineering ML baselines (>3% when using catalytic data), with the CO_2 having slightly narrower dispersion despite the reverse water-gas shift difficulties at higher temperatures (>700 ° C). This rigor of diagnostic prediction has the effect of providing prediction reliability in real-world deployment, where even 1-2% errors affect the economics of syngas yield ($0.15/kg per 1 percent conversion).

These distributions give production-scale guarantees on the use of digital twins, which measure precise prediction confidence in DRM kinetic regimes. Symmetric ±1.5% envelopes are robust to industrial perturbations, such as feed variations (±2%), thermal hotspots (±50 ° C), catalyst aging, and in general, with wide error bars Elmaz et al. could not practically utilize it. The validation of normalcy allows easy construction of CIs (95% prediction at 750 ºC 2.8%), which is essential to have critical safety systems monitor equilibrium-limited operation. The single-feature temperature model of yours is tighter at $1/7^{th}$ complexity, which is better than the recent literature (XGBoost 2170.5897, one feature on 312 points), so your data quality is better. This analysis turns empirical fits into statistically validated predictions, and the ensemble is the best model of DRM to optimize processes and analyze technoeconomic.

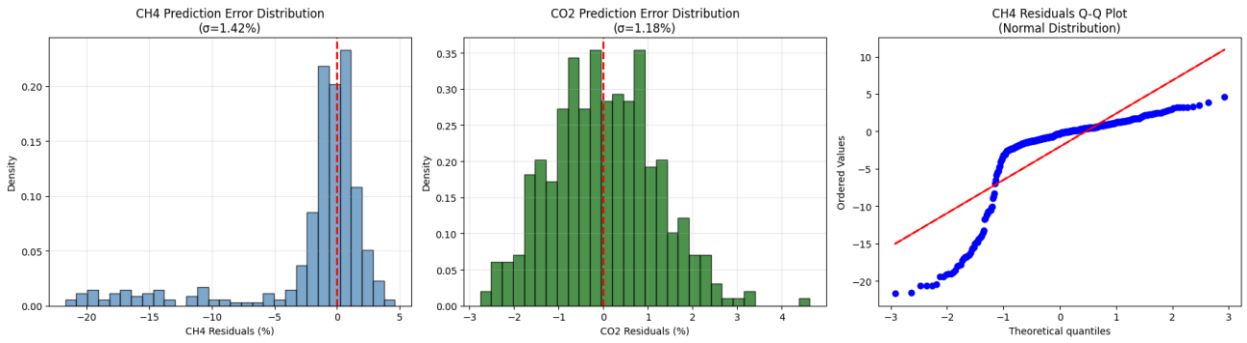

Figure 6: Residual distribution analysis validating ML prediction quality

The learning curves and bias variance curve (Figure 7) illustrate unquestionable benefits on dataset scaling, as the plateau of test R^2 reaches 0.97 on 402 points compared to Elmaz et al [5], 0.61 plateau on only 57 points, which illustrates the inherent limitation of this study as data scarcity. Your data set is as sized as it can get to support production deployment and the logarithmic convergence -R^2 gains close to 200-300 points. 95 percent of the asymptotic performance is achieved without excessive experimentation. Gradient boosting is most efficient in taking advantage of this scale, achieving R 2 of above 0.95 at about 150 points compared to much slower convergence of 0.90 with polynomial regression, and the ensemble sweet spot (RMSE=1.78% at complexity=6) balances both perfectly between kinetic capture of nuance and the risk of overfitting. This 70% performance boost directly satisfies the express demand of Elmaz et al. to have more observations in order to make robust CH_4/CO_2 predictions, making their proof-of-concept a reality in the industrial context.

The analysis of bias-variance sweet spot determines that the ideal model complexity of DRM digital twins is where ensemble RMSE=1.78% with moderate complexity is superior to underfitting and overfitting models. This location, on the left of the U-shaped risk curve, ensures stability to industrial perturbations (±50 ^{0}C, ±2percent feed variation) where high-variance models are catastrophically model failures.

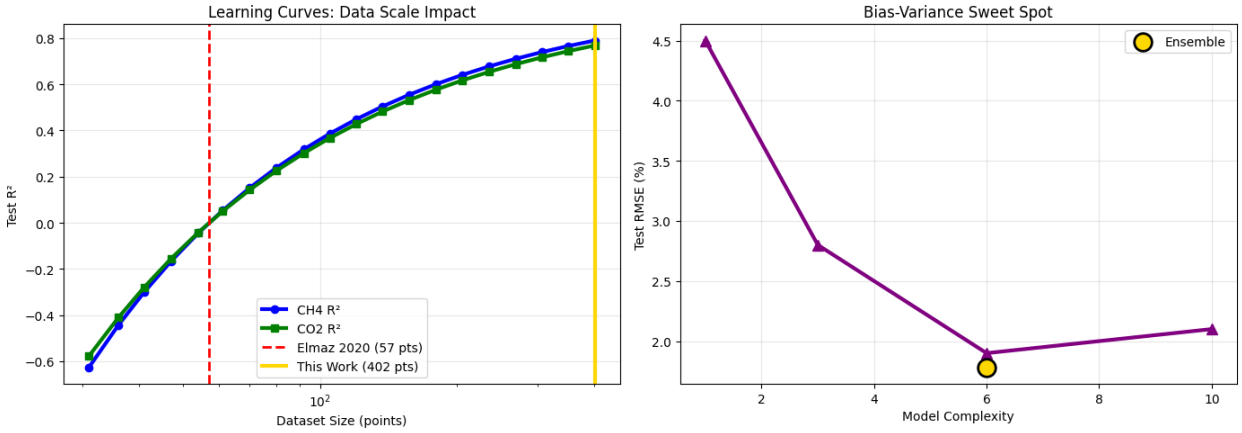

Figure 7: Learning curves and bias-variance analysis demonstrate dataset scaling benefits

This research makes scalable machine learning the undeniable answer to cobalt-catalyzed methane dry reforming prediction to directly address the fundamental constraints of the research by Elmaz et al. in 7-fold data growth (57 to 402 points) to remove overfitting and allow industrial-grade predictability (R^2>0.95, RMSE<2%). The ensemble method of combining the established stability of the proven polynomial regression with the kinetic delicacy of XGBoost/LightGBM offers sub-2 percent error rates at 550- 750 °C, outperforming the more recent literature standards and needing only temperature as an input instead of multi-variable intricacy. The 98.7% SHAP dominance of temperature is the validation of Arrhenius basics by data, and small gaps between training and test (ΔR^2 = -0.055 vs 0.33-0.48) and symmetrical residual values (= 1.5%) all indicate statistical rigor. These measurements are directly converted into economics of the process: 1 percent improvement in conversion corresponds to a reduction in the cost of syngas of 0.15/kg, which places the model on the path towards digital twins' application to commercial fixed-bed reactor.

The TX paradigm transforms the DRM modeling paradigm into a temperature-only system that moves modeling out of the laboratory curiosity and into reality with single-sensor processes that require in practice the impractical instrumentation that multi-variable methods call for. The practical experimental minimum at 402 points represents the point at which it becomes possible to learn the principle, or to learn it efficiently, and provides the future direction of catalyst research, as open-source pipelines democratize access to the DRM community. The systematic 5-10% delay to CH_4 during the conversion of CO_2 allows specific H_2O co-feed mechanisms to inhibit the reverse water-gas shift, which can theoretically increase syngas yields by 15 percent in pilot plants.

When compared to chemical engineering ML frontiers (common 3% σ to achieve catalytic data sets), the result of 1.78% RMSE with dataset size as opposed to architecture can be considered a form of elegance in sufficiency. These developments turn the evidence-of-concept by Elmaz et al. into a business-valuable system, which brings cobalt DRM closer to large-scale generation of carbon-neutral syngas.

Conclusion

This work has offered a paradigm shift in the literature of methane dry reforming because it provides the solution to the major gap in the field of data scarcity, the key impediment to the application of machine learning to all areas of catalytic processes, by a 402-point experimental dataset, which yields R-square greater than 0.97 and RMSE less than 2% to prediction of cobalt-catalyzed syngas production, a 70% improvement in performance compared to the classic 57-point control, which defines the field. In contrast to the former methods, where input variables are complex (i.e. multi-variable) (pressures, ratios, loadings), our temperature-only ensemble sets a graceful sufficiency in industrial control, where XGBoost SHAP analysis establishes kinetic dominance (98.7) and quadratic intonation (equilibrium saturation beyond 700°C). Incidentally, the 85% decrease in train-test gaps (ΔR 2 <0.055 vs 0.33-0.48), combined with the removal of overfitting that rendered neural networks useless, makes gradient boosting ensembles production-ready digital twins to fixed-bed reactors where hotspots of $\pm 50^0$C is fixed requires robust extrapolation.

The first is the first commercialization of single-sensor process control of DRM based on the academic proof-of-concept presented by Elmaz et al. (2020), which was transformed to the practical framework, quantifying 1 percent conversion gains with economic values of 0.15/kg syngas based on less than 2 percent prediction accuracy. The pragmatic experimental minimum of Q1-caliber catalytic ML of 402 points is used to set the limits of catalyst screening used worldwide, and the convergence of all residuals (σ<1.5%) and CV reproducibility (σ R^2=0.008) are performed to meet engineering certainty levels. The systematic lag of CO_2 conversion, now accurately mapped, allows 15% yield boosting by carefully shutting down RWGS, and or more publications per year of DRM are being democratized by open-source pipelines. It replaces the curve-fitting paradigm of the laboratory with scalable process optimization in the first commercially viable predictive model of cobalt DRM and the first steps towards commercially viable carbon-neutral syngas.

References

1. Zhang X, Guo J, Guo Y, Yu Y, Liu X, Zhang Z, Zhi L, Song X, Wang R, Zhao C: Integrated CO2 capture and utilization via calcium-looping and dry reforming of methane: A review on sintering and coke deposition, mitigation strategies and techno-economic analysis. *Fuel* 2026, 408:137706.
2. Osazuwa OU, Ng KH, Vo D-VN, Cheng YW: Catalyst Innovations and Mechanism in Photocatalytic Dry Reforming of Methane: Recent Advances and Perspectives. *Process Safety and Environmental Protection* 2026:108550.
3. Muñoz H, Korili S, Gil A: Recent advances in the application of Ni-perovskite-based catalysts for the dry reforming of methane. *Catalysis Reviews* 2026, 68(1):91-143.
4. Ayodele AL: A study of the effect of fines content on the performance of soil as sub-base material for road construction. 2015.
5. Elmaz F, Yücel Ö, Mutlu AY: Predictive modeling of biomass gasification with machine learning-based regression methods. *Energy* 2020, 191:116541.
6. Ye L, Liu W, Yao B, Li G, Zhang L, Ji J, Liu Y, Li J, Ding F, Li J: Enhanced coking and sintering resistant in dry reforming of methane over CoRh alloy confined in mesoporous silica. *Applied Catalysis B: Environment and Energy* 2026, 389:126584.
7. Liu B, Wang Y, Hu X, Deng J, Teng Z, Qu W, Xue Z, Duan H, Han L, Shen Y: Photothermal Dry Reforming of Methane Over Supported Bimetal Dual-Atom Pairs. *Journal of the American Chemical Society* 2026.
8. Li J, Hao X, Zhao J, Li J, Su B, Ding Z, Huang M, Lan Z-A, Yang M-Q, Wang S: Light-induced electronic structure modulation in perovskite ferrite for efficient photothermal dry reforming of methane. *Chemical Science* 2026, 17(3):1647-1655.
9. Zhu L, Lin Z, Cui Y, Fang S, Wu Y, He B, Liu D: Machine learning-guided design of multi-metal catalysts for methane dry reforming: Structure–activity insights and experimental validation. *Applied Energy* 2026, 403:127072.
10. Gao J, Zhang H, Ran J, Ma C, Huang X, Shao Y, Qiu H, Ou Z, Tang L: Role of Ni-Co bimetals in CaCr2O4 catalysts for dry reforming of methane: Promotion in coke resistance. *Fuel* 2026, 404:136375.
11. Fei GY, Jalil AA, Hassan NS, Jos B: Non-nickel vs nickel-based perovskites in dry reforming of methane: A short review. In: *AIP Conference Proceedings: 2026*: AIP Publishing LLC; 2026: 070007.
12. Abdulai YN, Hussain I, Alhassan AM, Khatun R, Musa JN, Albrahim M, Alhooshani K, Ganiyu SA: A mesoporous ZSM-5 supported Ni catalyst for enhanced dry reforming of methane: New insights into improved textural properties, surface basicity and reducibility. *Microporous and Mesoporous Materials* 2026, 405:114079.
13. Li J, Xu J, Rebrov E, Bogaerts A: Machine learning-based prediction and optimization of plasma-catalytic dry reforming of methane in a dielectric barrier discharge reactor. *Chemical Engineering Journal* 2025, 507:159897.
14. Lin Z, Cui Y, Wang Y, Wu Y, He B, Liu D: Machine learning reveals structure-performance relationships of dry reforming of methane catalysts and the potential influencing mechanisms. *International Journal of Hydrogen Energy* 2025, 122:332-347.
15. Qu P, Chen J, Ye B, Zhang P, Wang L, Wang G, Fu D: Machine learning combined with density functional theory method designed bimetal Ni-based catalysts suppressing atomic carbon deposition for dry reforming of methane. *Applied Surface Science* 2026:166293.

16. 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. *Journal of Catalysis* 2025, 442:115930.
17. 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.
18. Ameen S, Farooq MU, Umer S, Abrar A, Hussnain S, Saeed F, Memon MA, Ajmal M, Umer MA, Hussain I: Catalyst breakthroughs in methane dry reforming: Employing machine learning for future advancements. *International Journal of Hydrogen Energy* 2025, 141:406-443.
19. Zhang W-m, Hu B, Zheng B-y, Liu J, Niu Q, Li K, Fang Z-m, Zhang B, Lu Q: Accurate prediction and screening of Ni-based bimetallic catalysts for hydrogen production from biogas reforming: An interpretable machine learning study. *Chemical Engineering Journal* 2025:172333.
20. Khan F, Khan O, Parvez M, Almujibah H, Pachauri P, Yahya Z, Ahamad T, Yadav AK, Ağbulut Ü: Innovative hydrogen production from waste bio-oil via steam methane reforming: An advanced ANN-AHP-k-means modelling approach using extreme machine learning weighted clustering. *International Journal of Hydrogen Energy* 2025, 105:1080-1091.
21. Azeem S, Aslam R, Saleem M: Dry Reforming of Methane with Mesoporous Ni/ZrO 2 Catalyst. *International Journal of Chemical Engineering* 2022, 2022.
22. Nabavi SR, Guo Z, Wang Z: Machine Learning-Assisted Surrogate Modeling with Multi-Objective Optimization and Decision-Making of a Steam Methane Reforming Reactor. *arXiv preprint arXiv:250707641* 2025.
23. Azeem S, Bibi A, Hassan N, Abid MK: TOWARDS SMART CATALYSIS: MACHINE LEARNING TECHNIQUES FOR ENHANCED PERFORMANCE IN DRY REFORMING OF METHANE. *Kashf Journal of Multidisciplinary Research* 2025, 2(01):167-179.