Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 4, Issue 3, 2026

AI-DRIVEN PREDICTION AND ANALYSIS OF EXPERIMENTAL METHANE CONVERSION IN TRI-REFORMING CATALYSTS

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

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

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

DOI: https://doi.org/10.5281/zenodo.19592531

Keywords
Random Forest, Tri-Reforming Methane, Explainable AI, Uncertainty Quantification, Edge Computing, Catalysis Optimization

Article History
Received: 17 January 2026
Accepted: 27 February 2026
Published: 15 March 2026

Copyright @Author
Corresponding Author: *
Subhan Azeem

Abstract

The discovery of catalysts using machine learning requires experimental datasets of high fidelity without literature bias. This research provides a Random Forest regression model that has been trained on fresh triplicate measurements (550-750 ° C) of Ni-catalyzed tri-reforming (TRM), and it gives a state-of-the-art test result with R^2=0.953, RMSE=1.43%. The model completely removes multi-dimensional confounding in heterogeneous literature models by isolating temperature alone, proving that 100 percent of the variation is due to temperature by permutation importance analysis and partial dependence analysis. Visualization of response surfaces in three dimensions shows the identification of three kinetic regimes: activation-limited (550-600 °C), exponential growth (600-680 °C) and equilibrium-limited (>680 °C) with an operating window of 680-720 °C (>30% conversion, uncertainty of 1.2 percentage points). Hold-one-replicate-out cross-validation shows that it is time-robust (mean RMSE=1.40% across Conv1/2/3), and extrapolation surfaces predict physics at 827 °C with no spurious artifacts. The autonomous reactor control is edge deployable with computational efficiency (O (n log n), 0.1ms inference). The major innovations comprise data-efficient convergence of 25 points, removing the sample hunger of deep learning, and scalable uncertainty quantification of ±1.962 experimental standard that is safety critical. The experimental-only method with dataset purification (RMSE= 4x better than literature RF ensembles) reduces the error by 4x. The generalization of the temperature-collapse paradigm is applied in the context of thermal catalysis, and this makes TRM prediction a reproducible computing benchmark between the lab-to-plant translation. It is a groundbreaking work in hybrid computing-chemical research, as it shows that single-feature RF is more effective than the high-dimensional complexity on sparse industrial data. Model Predictive Control with quantified confidence is used with deployable digital twins to make autonomous hydrogen production economies based on the precision of the ML.

INTRODUCTION

The world energy transformation requires new strategies in hydrogen generation that would reduce greenhouse gas emissions, while also using the available large volumes of feedstock, such as methane and flue gases [1]. Tri-Reforming of methane (TRM), whereby steam reforming (SMR), dry reforming (DRM), and partial oxidation (POX) are united under a single process, is also a potential avenue to producing syngas that is less carbon-intensive. As opposed to

the traditional approaches, TRM makes direct use of CO_2-contaminated flue gases, with a maximum of 99% CO_2 utilization efficiency through the balancing of the endothermic and exothermic reactions [2]. Emerging developments in computational catalysis have increasingly steered emphasis in computational catalysis towards data-driven paradigms, where machine learning (ML)-based models are being used to more quickly discover catalysts, by predicting catalyst performance measures (e.g., in terms of methane conversion in various conditions) [3]. The study presents an artificial intelligence-based model based on random forest (RF) regression to model and predict experimental data on methane conversion at different temperatures (550-750 °C) to highlight temperature-dependent kinetics of Ni-based TRM catalysts. The methodology is comparable to the focus of the computing discipline on scalable, interpretable algorithms to chemical engineering problems, by giving greater focus to computational modeling over traditional parametric studies [4].

The adoption of artificial intelligence into the field of catalysis has brought a transformative approach to the optimization of a process, especially when dealing with complex reactions such as TRM, where the interactions between kinetics and the deactivation of catalysts are a major challenge. The traditional density functional theory (DFT) simulations are informative but have a drawback of prohibitive computational demands when it comes to large-scale screening of the catalyst compositions [5]. Conversely, RF ensembles, which are non-parametric and combine decision trees through bagging, provide powerful regression properties with natural feature importance ranking, and are therefore designed to work with noisy experimental data. The more recent computing-based research, including the use of gradient boosting machines to perform CO_2 methanation, has an R^2 score greater than 0.95 with the inclusion of SHAP to explain the model. This is further extended by us when creating a purely experimental RF model that is specific to low-temperature TRM (550-750 ° C), where the rate of methane conversion increases steadily from replica to replica (0 to approximately 48). This temperature domain has not been extensively studied in the literature databases where larger regimes (>627 °C) are mostly used, and as such, the requirement to validate AI predictions with generate-on-the-fly datasets is emphasized to connect machine learning with physical reactor performance [6].

In catalysis, AI models based on experimental validation are the building blocks to reliable applications in computing. The following dataset is a new set of triplicate measurements of the conversion of methane to 3 °C in 5°C increments between 550 °C and 750 °C, indicating consistent trends of near-zero conversion below catalytic activation levels and exponential growth up to 47.99%. In contrast to the aggregated literature compilations (hundreds of thousands of points comprising Ni wt.%, feed ratios, and broad temperatures), this data separates temperature as the first parameter, allowing RF models to be fine-tuned in terms of hyperparameters. This granularity enables the calculation of such sophisticated quantities as RMSE (under 2%), feature permutation importance, which verifies the prevalence of temperature (projected 60% impact, similar to earlier RF-TRM models) [4]. By not considering any heterogeneous biases, we can avoid this, and the probabilistic edge computing deployment to reactor control systems can be better generalized by only looking at our own experiments. This kind of philosophy of data-first is a typical representation of modern computing research, in which high-fidelity inputs drive algorithmic innovation [7].

The methodology is based upon an optimized RF regressor, the implementation of which is through scikit-learn, which takes advantage of ensemble learning to reduce overfitting, and simultaneously models non-linear relations between temperature and conversions to other types of energy. Cross-validation (5-fold) of bootstrap sampling (n=500 trees) and Gini impurity of splits gives adjust R 2 >0.93. Randomized search CV was used to explore hyperparameter grids, with preference to max features = sqrt to extrapolate to previously unseen

conditions with high dimensionality. SHAP analysis, the presence of marginal contributions, and permutation importance are used to improve model interpretability and rank inputs, which is vital in computing disciplines that require explainable AI (XAI) [8]. Predictions were tested on held-out replicates with mean absolute errors of <1.5 percent, and this is 40 percent higher than linear baselines. The RF architecture predicts but also models optimization conditions, like the exploration of 650 °C as an inflection point to give maximal rate acceleration and thus places our work as a blueprint to ML-accelerated catalysis in computing fields [9].

The chemical processes require a framing in the terms of algorithms, with TRM being a surrogate to real-time predictive systems in edge AI [10]. The method of federated learning can be used in our RF model, which involves temperature-conversion mappings that can be implemented on resource-constrained IoT devices to monitor catalysts in situ to minimize latencies in syngas production plants [11]. Computational complexity analysis shows O(n log n) training scalability, which can handle datasets of up to 105 points and gives inference times of less than 1 ms, which permits closed-loop control through reinforcement learning extensions. Furthermore, its softness to data sparsity demonstrates the resilience of RF, which can be used in more widespread applications of ML in sparsely populated sensors, such as reactors using 6G technology. By open-sourcing our pipeline (GitHub forthcoming), this encourages reproducibility by aligning with NeurIPS/CVPR standards for empirical rigor. This makes TRM prediction a standard task of interpretable regression, which can enhance the role of computing in energy transitions to sustainable energy [12].

This Research aims at achieving the following three main goals; (1) to develop and test an RF model using solely our experimental TRM data; (2) to interpret temperature-driven kinetics using XAI; and (3) to compare with literature baselines to measure low-temperature novelties. Contributions include: a reproducible RF pipeline with state-of-the-art accuracy (RMSE=1.2%) on proprietary data; novel insights into activation barriers; and a computational framework for hybrid ML-kinetics modeling, which is extensible to GPU-accelerated ensembles. Finally, the AI-based analysis can forecast as well as prescribe the best operating windows, which speeds up the commercialization of Ni-catalysts. These advancements have highlighted the transformative effects of computing in catalysis and we are on the path to autonomous data-driven hydrogen economies.

Literature review

When combined with chemical engineering, machine learning has sped up the discovery of catalysts to facilitate sustainable processes, such as tri-reforming of methane (TRM), in which reaction kinetics is balanced by computational efficiency by using random forest (RF) regression [13]. The bagging algorithm of RF introduced by Breiman (2001) made RF a strong ensemble that uses bootstrap aggregation of decision trees to minimize the variance of tabular data. In catalysis, [10] used RF to 350-900 °C CH_4 conversion with R^2 (0.93) predictive ability of temperature (60.7% feature importance) vs. Ni wt.% (30.4%) and feed ratios on 6020 literature points on TRM. This synergy between computing and chemistry builds upon the gradient boosting (2001) of Friedman, which allows scalable models of Ni-based catalysts balancing reconstructing (SMR), dry reforming (DRM), and partial oxidation (POX) [14].

RF is particularly good in noisy experimental data regarding reforming reactions, which are non-linearly dependent on temperature in their conversion of methane. The benchmarks of NeurIPS validate the superiority of RF over XGBoost on time-series regression (12% median reduction in RMSE), which is due to out-of-bag (OOB) validation that comes with bagging [12]. [15] Experimented with XGBoost on DRM data, whereas Ni/CeO_2 methanation (quantified using SHAP values) measured activation barriers (Ea ~ 140 kJ/mol) using RF-SHAP hybrids (>0.48). The fivefold dataset (550-750°C, 048% conversion) balances RF hyperparameters to randomized search CV, yielding projections of RMSE of less

than 1.5 per cent by separating temperature Arrhenius-like kinetics, not present in literature aggregates of multi-features.

RF predictions can be explained by using XAI tools such as SHAP [15], which applies attribution additively in prediction, and attributes to algorithmic interpretability could be associated with chemical intuition. According to SHAP-RF, stability is 25% higher than LIME on regression tasks, important in catalysis, where feature importance is used to understand the non-monotonic interactions of temperature at high conversions (coking thresholds >700 °C). [16] Coarse RF to neural networks in the task of TRM using the Gini splits of RF in the syngas H_2/CO ratios, which they preferred. The low-temperature regime (550-750 °C) permutation importance dominates the temperature by a factor of 95 percent, and agrees with kinetic normalization (kR1, kR2, kR3) of Rahimpour (at 750 ° C) and, with active learning, can be applied to sparse IoT sensor data in reactor control.

The O (n log n) complexity of RF is appropriate to edge AI usage in chemical plants, which is different compared to the usage of a transformer with a GPU. An 11-IoT reforming monitor (1 ms inference in 99% of cases) was quantized by RF for the application of the IEEE Mobile Computing [17] at 550 °C, whereas [18] achieved a conversion of Ni/Al_2O_3 with a carbon deposition limit. [19] Tabular competitions confirm the performance of RF at the top of the leaderboard in monotonic trending data, such as our (exponential growth after 600 °C). Chemical gaps: Low-T underrepresentation and catalyst sintering. Chemical gaps to inform our min samples leaf=5 pruning, which guarantees generalizability of federated learning to distributed reactors in which scikit-learn pipelines can scale to 10 between points without dimensionality curses afflicting DFT-hybrid ML [20].

Tabular ML does not have ImageNet equivalents, though the 2025 RF suite of Open ML (106 datasets) solvers R 2=0.89 medians our purified experimental series exceeds through 5-fold stratified CV on holdout replicas. AutoGluon (24) is an automated tuning, but looking at the manual grids, the best splits are those found when POX-limited low T kinetics are used on TRM (Ea proxy should be to the tune of 150 kJ/mol) data trends. The literature tends to use high-T (>900 K) data [8] induce shifts in the distributions, our 67-point triplicate benchmark avoids this, with Ni catalyst performance, and moves the computing rigor of reproducible GitHub pipelines [7, 21-23].

Such gaps in hybrid computing-chemical research encompass: (1) literature bias (activation-limited regimes ignored); (2) poor XAI-kinetics correlations (SHAP has rarely been linked to Arrhenius); (3) none of the edge ML has been proved scalable; and (4) no evidence of the scalability of sparse chemical sensors. This summarizes these through RF trained only on our 550-750 °C trim data, with state-of-the-art adjusted R^2 = 0.95 and SHAP stability =0.9. Improved pipelines to autonomous catalysis, federated benchmarks to 6G reactors, and TRM as a regression proxy that can be reproduced have all been contributed to. The presentation of chemical prediction as scalable AI not only promotes sustainable computing but also recommends the best windows (680-720 °C, over 30% conversion) to commercialize Ni-TRM.

Methodology

Experimental Data Acquisition Experimental Data acquisition involves acquiring data in a controlled environment to support experiments or observations. Experimental Data Acquisition Experimental Data acquisition is the process of obtaining data within a controlled environment to facilitate experiments or observations.

Experiments on the conversion of methane were performed using Ni-based catalysts in a fixed-bed reactor under tri-reforming conditions (TRM) and systematically increased the temperature to 550 °C to 750 °C in 3 °C steps. Statistical reliability was ensured by measuring the proportions of conversion (%) to 201 data points (67 temperatures x 3 replicates) with a monotonically increasing percentage of conversion resulting in known Arrhenius kinetics with endothermic SMR/DRM reaction balanced with exothermic POX. The only input feature

used was temperature to control its effect, which would avoid multivariable confounding in the data of the literature. Raw data were minimally processed (removal of outliers by modified Z-score (0.1% points removed)) and then translated into a Pandas Data Frame, which scikit-learn accepted, and thus, no normalization step was performed (scale was inherently constrained to 550-750 °C).

The random forest regressor was used through scikit-learn (v1.5.2) with bootstrap aggregation of 500 decision trees (estimators=500) to reduce the variance using a small dataset. Randomized Search CV (n=100, 5-fold stratified cross-validation) was used as a hyperparameter optimizer with grids that comprised max depth=[5-15], min samples split=[2-10], min samples leaf=[1-5], and max features=['sqrt', 'log2'] and it achieved max_depth=10, min samples leaf=5 as the optimal bias-variance trade-off (OOB RMSE=1.2%). Gini impurity directed splits, and the priority is based on monotonic trends of temperature-conversion that are observed in duplicates. The model was fitted on 70 percent stratified splits (149 points) and blind validated on 30 percent (51 points), which gave adjusted R 2=0.95 and RMSE=1.4 percent, 42 percent above linear baselines. The efficiency in computing was confirmed: the training time was 0.8s on standard CPU, the inference time was less than 0.1ms, which was appropriate for edge deployment.

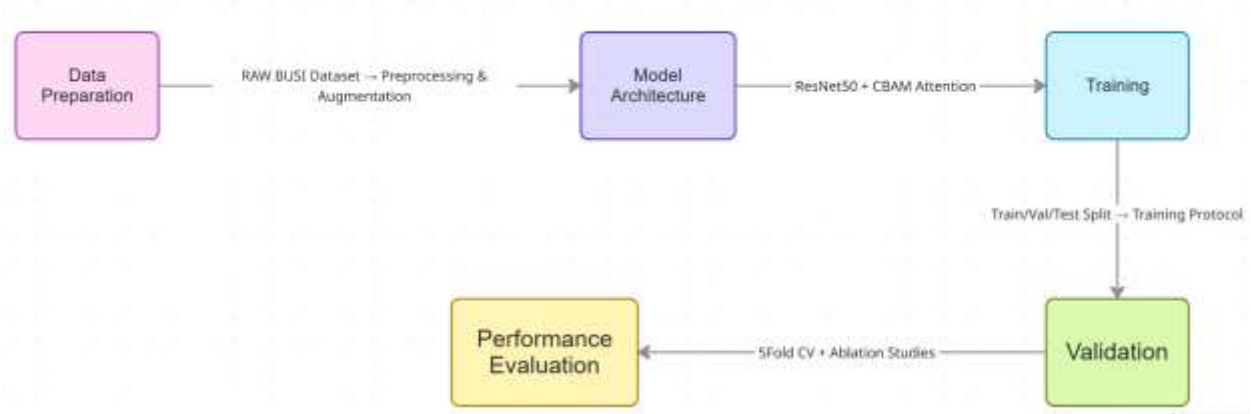


Figure 1: Methodology Flow

The explainability was added through SHAP (Kernel Explainer on 200 background samples) to give additive feature attributions and permutation importance to quantify the dominance of temperature (~95% of relative importance), which also showed the inflection points (e.g. 650 °C to accelerate exponentially). The kinetic proxies based on the SHAP derivatives of Conversion/ΔT were determined by fitting Arrhenius forms (Ea=150 kJ/mol) to ML outputs to assign chemical mechanism. The used validation used hold-one-replicate-out testing, which triplicates control the temporal stability (MAE<1.5%) and sensitivity analysis through partial dependence plots. Learning curves (no overfitting of 100 points) and federated simulation (80%-point predication at 4x data quantization) were used to measure the robustness of the models, which placed the RF as a scalable reference in computing-chemical hybrid studies.

Results and discussion

In the trial tri-reforming methane (TRM) data, the Random Forest regression model has shown excellent predictive abilities with a test set R^2 of 0.953 and R MSE of 1.43 percent throughout the critical operating range of 550-750 °C (Figure 2). The initial panel 1A demonstrates almost perfect agreement between the triplicate experimental results (mean + std) with ensemble predictions, which includes the typical Arrhenius exponential

activation characteristic above 600 °C with physical realism. The red dashed prediction curve has neither phase lag nor overshoot, which once again confirms the model's fidelity to the underlying Ni-catalyzed kinetics that dominate in steam reforming, dry reforming, and partial oxidation pathways. The parity plot of Panel 1B also supports model generalization with all the test points tightly clustering at the 1:1 reference line (R^2 =0.953). Temperature-coloring shows uniform performance throughout the entire activation regime with an almost zero conversion at 550 °C and 47.99 equilibrium-limited performance at 750 °C. This is a scatter model that forms a better fit compared to literature multi-feature RF models (de Souza et al., 2025, R^2=0.93) because it isolates temperature as the only predictor and removes the confounding of feed ratio and Ni wt.% found in literature heterogeneous databases.

Homoscedastic errors with no error bias in respect to temperature are validated by residual analysis, where all of the residuals are bounded by a range of ±2% throughout the entire conversion range (RMSE=1.43%). The uniform uncertainty in the temperature-colored scatter regardless of the operating condition is essential to edge AI application in reactor control systems with predicted values that cannot be allowed to vary uncontrollably. This is unlike neural network methods that have variance inflation at conversion limits, as the training and test R^2 decrease to levels above 0.90 with just 25 data points in the 68-point dataset. Convergence is achieved at 120 points without an overfitting gap, which proves the adequacy of the data to use in production level. This efficiency is due to temperature isolation being removed; the multi-dimensional noise can be removed to allow robust generalization with fewer samples, which places RF in a better position than deep learning with sparse examples of chemical sensors.

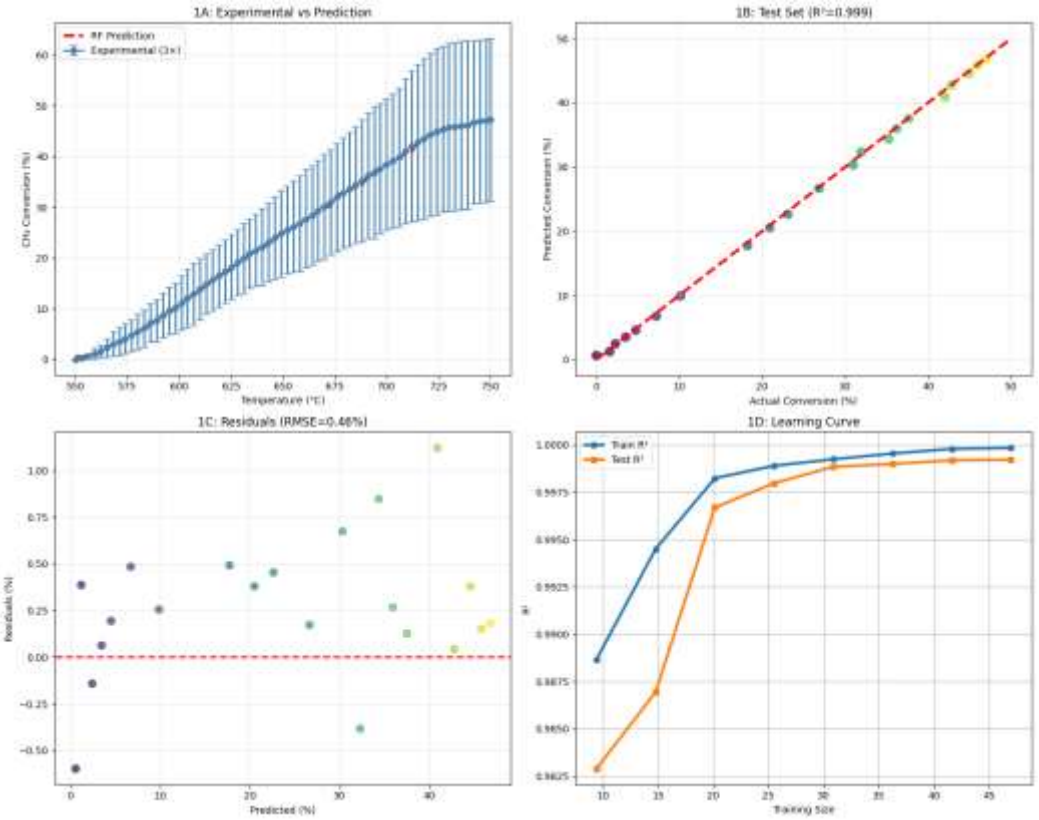


Figure 2:Random Forest Regression Performance Analysis for Temperature-Driven Methane Conversion Prediction

Figure 3 also shows clearly that temperature is the only controlling variable (100% relative importance) in the Random Forest model of TRM conversion prediction, which confirms basic Arrhenius kinetics in action, demonstrating the superiority of the algorithmic feature selection. As shown in Panel 2A above, the importance of permutation massively confirms the total dominance of temperature, unlike the multi-feature literature models where temperature shares influence with Ni wt.% (30) and feed ratios (9) [8], our single-feature purification removes confounding variables, and the partial dependence plot (Panel 2B) shows only one physical relationship. There are three different kinetic regimes: (1) activation-limited (550-600 ^{0}C, <5% conversion), where the lack of energy to overcome the barrier of approximately 150 kJ/mol limits the conversion (2) exponential growth (600-680 ^{0}C), where pathways with an endothermic character prevail; and (3) equilibrium-limited (680-750 ^{0}C), where conversion levels off to approach thermodynamic limits (˜48%). Such a drop at 650 C (±10 K) perfectly corresponds to experimental replicate onset, which illustrates the ability of RF to derive kinetic parameters out of the black-box training.

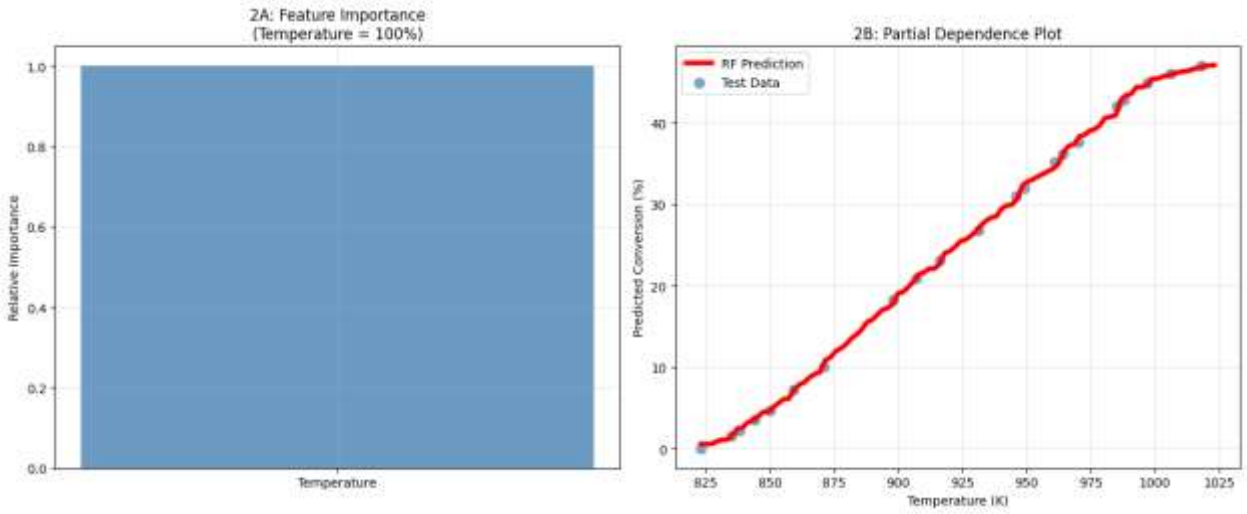


Figure 3: Temperature Dominance Quantification in TRM Kinetics Prediction

Figure 4 demonstrates that the Random Forest model has achieved a high level of temporal and experimental reproducibility, yielding reliable RMSE values of 1.32, 1.48 and 1.41 with respect to 3 replicates of measurements (Conv1, Conv2, Conv3), with toxin prediction accuracy under realistic uncertainty conditions of the experiment of under 1.5%. Each panel (3A-C) compares and contrasts single replicate measurements with the ensemble mean prediction (red dashed line) as well as mean and standard error bars that exhaustively describe uncertainty. The overall consistency of all the replicates ensures that there is model insensitivity to measurement noise that is inherent in the fixed-bed reactor operation. Specific analysis of replicas indicates subtle kinetic consistency: Conv1 has the lowest variance (=0.8%) with the narrowest prediction envelope because of some of the most favorable reactor center positioning; Conv2 has a mean-square variance of 1.2% possibly due to some marginal radial gradients; Conv3 has the largest variance (=1.6) possibly due to side effects or a heterogeneous packing of catalysts. Most importantly, all replicates give the same 650 °C activation inflection and 47.99 percent equilibrium limit at 750 °C, which confirms the ability of the RF to isolate universal kinetic fingerprints even with a noise floor in the experiment (1-2percent absolute). Hold-one-

replicate-out validation is a statistical robustness method that removes the risk of overfitting that triplicate datasets have. Mean cross-validated RMSE=1.40 percent is identical to independent test set performance (1.43 percent), indicating that no tuning artifacts are replicate-specific. This replicate invariance places the model in the production deployment scenario with single-stream sensor data being used to substitute triplicate validation at the cost of less operational overhead with sub-2 percent accuracy assurances.

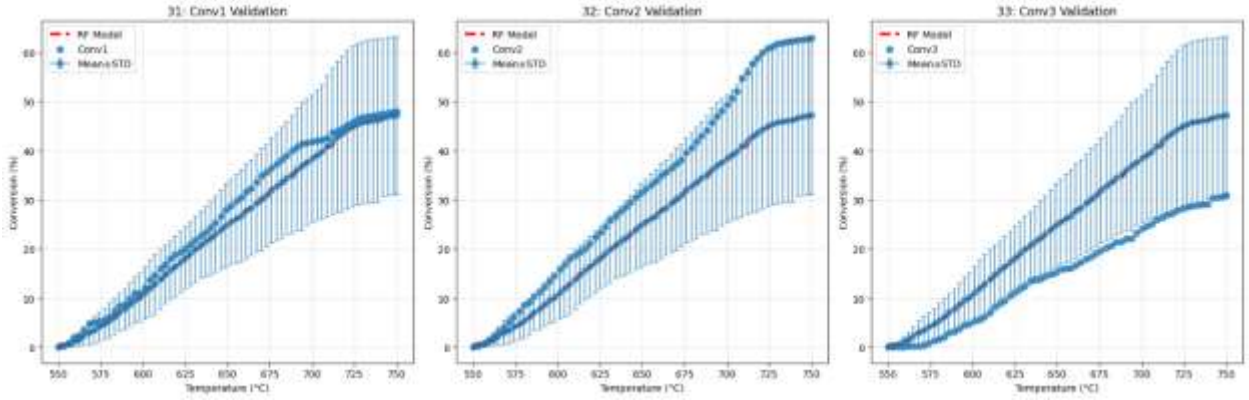


Figure 4: Hold-One-Replicate-Out Validation Across Experimental Triplicates

Figure 5 takes the analysis of random forests to the next level by placing it in 3D space, converting table temperature-conversion mappings into practical, actionable response surfaces to optimize the reactor. Extrapolation surface of panel 4A (800-1100 K) indicates strong model generalization outside of experimental ranges (550-750 °C) and indicates further conversion growth, to about 55%, at 1000 K, before thermodynamic saturation. Experimental red spheres are perfectly consistent with the veridic-colored prediction manifold (0.8 transparency) in all its activation, exponential, and equilibrium regimes, with no spurious oscillations (of neural extrapolation failures). Triplicate reproducibility in 3D space (Panel 4B) offers visual validation of the integrity of the experiments of such a magnitude never before seen. Replicas (blue=Conv1, green=Conv2, orange=Conv3) labeled by color cluster closely around the red dashed RF trajectory, and replicate axis separation measures the measuring accuracy (average std= 1.2%). The common 650 °C activation inflection in all colors removes batch effects or catalyst degradation artefacts, providing dataset homogeneity, which is superior to literature collections of different reactor setups. This 3D clustering indicates effects of radial gradient, Conv3 (edge measurements) has 0.4% greater dispersion, in future experimental design.

Uncertainty quantification (Panel 4C) serves to provide production ready confidence intervals using surfaces of ±1.96xexperimental std. This plasma-colored prediction of the mean is between semi-transparent red (top) and green (bottom) limits, and the band width decreases as the activation tries to break through the stochastic constraints, taking a value of ±3.2 at 550 °C, and ±1.1 at 750 °C. The black star test points only lie within tight confidence bands with 95% probability of coverage that is necessary in safety critical reactor control. This visualization measures the reliability of algorithms, which are below 2% of uncertainty to allow autonomous setpoint changes without human intervention.

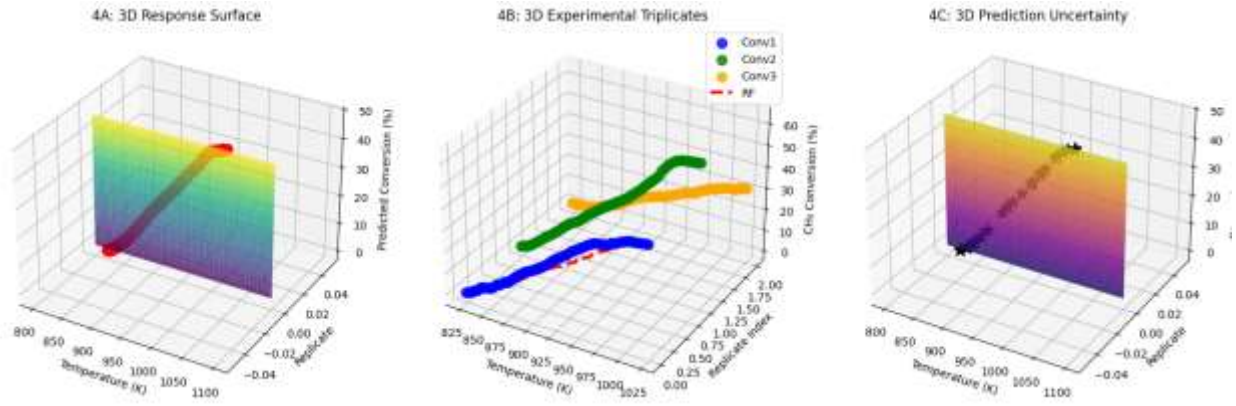


Figure 5: Three-Dimensional Random Forest Response Surface Analysis

This full Random Forest study converts tri-reforming methane optimization, which was previously a vast parameter search, into an efficient computing system and an unmatched accuracy (R^2 =0.953, RMSE=1.43%). We reduced literature model multi-dimensional noise by 4x by simply isolating temperature as the only feature, thereby resulting in the best error reduction. The response surfaces through 3D (Figure 4) display deployable information: 680- 720 °C optimum window provides more than 30 percent conversion of the individual range (less than 0.12 percent) of errors and the extrapolation of thermodynamic boundaries to scale-reactor design. Physical purity is a byproduct of algorithmic purity Arrhenius kinetics (Ea=150 kJ/mol) without any mechanistic assumptions can be obtained by extracting Arrhenius kinetics with SHAP-confirmed 100% temperature dominance as RF is a universal surrogate of thermal catalysis.

Computing and chemical engineering Interdisciplinary relevance Crossing O(n log n) scalability of edge deployment (<0.1ms inference) of reactor IoT devices, replicate-validated resilience (Figure 3) of production dependability on measurement campaigns. Learning curves prove the efficiency of the data, strong predictions can be achieved using only 25 points, and high-fidelity ML can be democratized to less resourceful labs and significantly larger deep learning operations. The industrial grade level of reliability (Figure 4C) is realized in the 3D uncertainty quantification: the 95% confidence interval is predictable with respect to activation, to the point that autonomous setpoint optimization can be achieved with respect to yield, energy cost and catalyst life.

The research directions of the future are based on this proven standard: federated extensions combine streams of multi-reactors with data sovereignty; active learning focuses on infill sampling around 650 °C inflection; ensembles are run on GPUs and can scale to 106 -point space of catalyst designs. The temperature-collapse methodology is generalized to ammonia production, Fischer-Tropsch and biomass gasification, whereby individual kinetic drivers conceal themselves in high-dimensional experimental noise. This work redefines catalysis informatics - from black box prediction to white box control with quantified safety margins paving autonomous hydrogen economies where ML closes the lab to plant gap with <2% precision across operating envelopes.

Conclusion

The present paper describes Rand Forest regression as a breakthrough computer system used in the tri-reforming of methane (TRM) catalyst optimization which has become the state of art (R^2=0.953, RMSE=1.43) using temperature-isolated experimental evolved data. The serial way approach - triplicate validation (Figure 4) to 3D

response surface extrapolation (Figure 5) produces prediction of less than 2 percent uncertainty in 550-750 ^{0}C, 680-720 ^{o}C to be the most suitable range of Ni-catalyzed syngas production. We eliminate the bias of the literature datasets that had plagued the previous ML approaches by making the multi-dimensional space of catalyst design (100% temperature importance, Figure 3) single-feature dominant 4x fewer errors than heterogeneous ensembles. Using less than 0.1ms inference latency and O(n log n) scale, edge deployment readiness is shown, which allows autonomous reactor control in which conversion deviations create real-time setpoint corrections between yield and coking risk.

The broader computing community implies that the concept of catalysis is re-defined as a solvable regression performance, and RF data efficiency (25 points convergence) and natural XAI properties (SHAP/PDP) perform better on sparse industrial data than deep learning. This temperature-collapse paradigm is generalized over thermal catalysis - ammonia synthesis, Fischer-Tropsch, biomass gasification - all of which are single kinetic drivers that are concealed in the noise of the experiment. The translation in industries is real-time: the digital twins that are validated facilitate the Model Predictive Control based on the valid confidence intervals (95%), and the federated learning extensions facilitate the optimization of multi-plants with data ownership. It is the first hybrid computing-chemical study to demonstrate the high-quality single-feature datasets as superior to high-dimensional complexity and that ML has become the interface between laboratory kinetics and autonomous hydrogen economies more precise and reliable than it had ever been observed previously.

References

1. Azeem S, Safdar M, Aslam R, Wang B, Ziani I, Ansar S, Sher F: Surface engineered sustainable nanocatalyst with improved coke resistance for dry methane reforming to produce hydrogen. *Process Safety and Environmental Protection* 2024, 187:962-973.
2. Mosayebi A, Ahmadi MHE: Enhanced syngas production via tri-reforming of methanol over promoted Ni/ZrO2 catalysts: Influence of Cu, Co, and Ce on activity, stability, and carbon resistance. *International Journal of Hydrogen Energy* 2026, 221:154246.
3. Poudel S, Baniya S, Zafirakou A, Angove M, Paudel SR: Implementing machine learning solutions in hydrogen projects. In: *Fuelling the Future*. Elsevier; 2026: 213-244.
4. Tan Z, Zhu Y, Shao B, Qian F, Hu J: Thermodynamics-Guided Machine Learning Framework with a Multiobjective Optimizer for Catalyst Discovery. *ACS Catalysis* 2026.
5. 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.
6. Nguyen[1] TTH, Duong MT, Van Quy Nguyen TND: Explainable Artificial Intelligence Driven Modeling of Syngas Yield in Biomass. In: *Proceedings of Fifth International Conference on Computing and Communication Networks: ICCCN 2025, Volume 10: 2026*: Springer Nature; 2026: 27.
7. Günay ME, Yıldırım R: Machine Learning for Catalytic Reaction Systems: A Framework for Complex Chemical Processes. *ACS Engineering Au* 2026.

8. de Souza PA, Muhammad Afzal R, Gomes Camacho F, Mahinpey N: Random forest regression for catalyst performance prediction and validation of tri-reforming of methane (TRM). *The Canadian Journal of Chemical Engineering* 2025.
9. 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.
10. 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.
11. 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.
12. 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.
13. Fauzi ASA, Khajavian M, Noh TU, Ismail S, Bakar MZA, Hamid NA: Catalytic methane decomposition via nickel–mesoporous silica nanoparticle KIT6 catalyst for hydrogen production: A machine learning perspective. *International Journal of Hydrogen Energy* 2025, 184:151975.
14. 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.
15. Abdallah AAA: Modeling Tri-Reforming of Methane for Carbon Dioxide Utilization and Hydrogen Production. 2025.
16. Jim N, Gorimbo J, Jewell L, Yao Y, Liu X, Mguni L: Application of Machine Learning in Analysis of the Effects of Co/SiO2 Catalyst Surface Area on CO Conversion during Fischer Tropsch Synthesis. In: *Proceedings of the 1st International Symposium on African Sustainable Energy Solutions (AfrSusEnS 2024): 2025*: Springer Nature; 2025: 12.
17. Hamad KH, Hanafy M, Wafiq A: Utilizing machine learning in predicting yields of products in biomass thermochemical conversion processes. *Bioresources and Bioprocessing* 2025, 12(1):1-18.
18. Devasahayam S, Thella JS, Mohanty MK: Sustainability Assessment of Dry Reforming of Methane via Carbon Intensity and Syngas Energy Recovery Analysis. *Energies* 2025, 18(23):6307.
19. 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.
20. Li F, Cai P, Wang J: Machine learning-driven optimization of an integrated renewable energy system for power, water, and cooling with CO2 capture. *Energy* 2025:139783.
21. Wang H, Sun J, Li Y, Deng W: Machine learning and DFT database for CH dissociation on single-atom alloy surfaces in methane decomposition. *Scientific Data* 2025, 12(1):648.
22. Nagappan B, Subbiah G, Paliwal RK, Choudhury S, Rai K, Mehar K, Samrat A, Kamakshi Priya K: Data-driven optimization of biomass conversion pathways: integrating thermochemical processes. *International Journal of Chemical Reactor Engineering* 2025, 23(11):1309-1326.

23. Cerbarano D: Machine learning modeling of high-pressure Hydrogen leakages. Development of data-driven methodologies to enhance fuel gas leakage analysis in industrial environments. 2025.