EPJ Web of Conferences **354**, 02001 (2026) https://doi.org/10.1051/epjconf/202635402001
GCMM 2025

Machine Learning Surrogated Coal Blend Optimisation for Inventory-Aware Coking

Abhay Nimbalkar[1], *Nishchal Kashyap*[1], and *Dega Nagaraju*[2*]
[1] School of Mechanical Engineering, Vellore Institute of Technology, Vellore-632014, India
[2] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore-632014, India

Abstract. The disruptions triggered by the SARS COVID-19 pandemic, followed by the Russia–Ukraine conflict and American sanctions resulted in reduced accessibility of Russian coal to Indian steelmakers. This decline in availability forced producers to undertake extensive trials and experiments hunting viable alternatives. This study argues that optimized blend compositions formulated exclusively from the available coal inventory can effectively address such disruptions thereby ensuring production and scheduling remain independent of external supply fluctuations. To achieve this, multiple machine learning models including multivariate regression, decision trees, partial least squares regression, random forests and neural networks are developed from data collected from an integrated steel plant to predict coke quality from blend data. The most accurate model is adopted as a surrogate objective function for a genetic algorithm that reallocates blend proportions under inventory constraints. Results demonstrate that coke quality can be maintained without introducing new coal sources, enabling resilient, data-driven adaptation to uncertain supply chains.

1. INTRODUCTION

The global steel industry has increasingly grappled with raw material volatility, revealing the fragility of conventional coal sourcing and blending strategies. The COVID-19 pandemic, compounded by geopolitical disruptions such as the Russia–Ukraine conflict and American sanctions, severely impacted coal trade flows—particularly the availability of Russian coal to Indian steelmakers. These disruptions forced producers into costly, trial-and-error substitution programs to sustain blast furnace operations [1]. In this context, maintaining metallurgical coke quality becomes paramount. Coke properties—especially mechanical strength (M_{40}, M_{10}), reactivity (CRI), and strength after reaction (CSR)—must be precisely controlled, even as procurement costs rise and inventory constraints tighten [2].

---

[*] Corresponding Author: deganagarajulc@gmail.com

Traditional empirical and weighted-average blending methods fall short, as key coking phenomena—such as thermoplastic behaviour, fluidity, and coking pressure—are inherently non-additive across coal types [3]. Microstructural interactions, maceral-level differences, and petrographic texture features drive nonlinear responses in coke performance, necessitating predictive models that are both high-fidelity and physically grounded. Recent advancements in machine learning (ML) offer promising solutions. Convolutional neural networks, interpretable deep learning models, and ensemble methods have demonstrated strong predictive capabilities for CRI/CSR and related indices. Moreover, interpretability techniques such as SHAP and CDRE align model outputs with domain knowledge, enhancing trust and transparency. Simultaneously, multi-objective optimization frameworks—ranging from enhanced genetic algorithms to Q-learning–augmented NSGA-III—have achieved measurable cost savings while meeting stringent coke quality constraints, paving the way for resilient, inventory-aware blending strategies.

To ensure industrial relevance, this study utilizes two months of operational data from an integrated steel plant. Each batch was manually tested in laboratories, with corresponding blend data extracted from Auto Proportioning Devices (APDs). Figures 1 and 2 illustrate the variations in coal blends and coke properties, respectively.

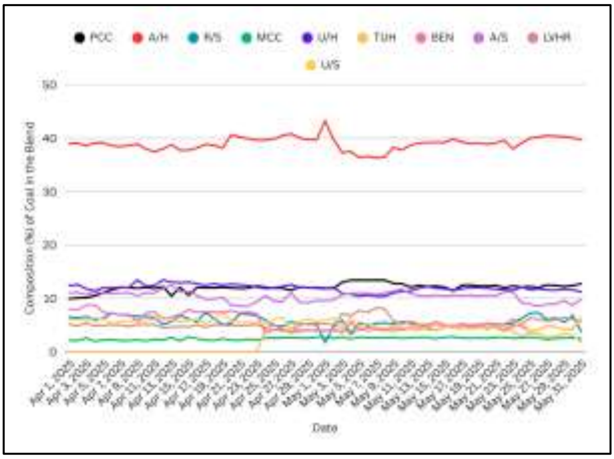

Figure 1: Variations observed in coal blends

Initially, coke oven gases and size parameters were considered. However, since coke oven gases are byproducts and not directly tied to operational objectives, their relevance was limited to downstream decision-making (e.g., merchant mills) and thus excluded. Similarly, size parameters were omitted due to their dependence on post-production handling.

This study proposes a unified, data-driven framework that:

1. Trains and selects high-performing surrogate ML models for coke quality prediction,
2. Synthesizes model performance using fuzzy CRITIC–WASPAS,
3. Maps metallurgical influence via perturbation-based sensitivity analysis, and
4. Executes multi-objective optimization under inventory and quality constraints.

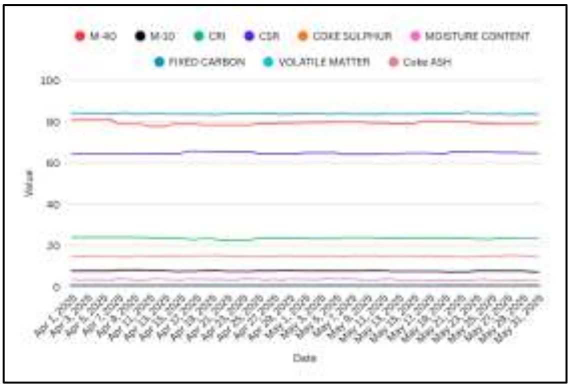

Figure 2: Variations observed in coke properties

2. LITERATURE REVIEW

The prediction and optimization of coke quality from coal blends have been an active area of research, leveraging both experimental studies and advanced computational methods. Early investigations highlighted the non-linear interactions in coal blends. [5] demonstrated that blending coals of different rank leads to non-additive fluidity behaviour due to plastic material transfer and volatile matter adsorption, emphasizing the inadequacy of weighted-average models. Similarly, [6], [7] used fractography to reveal microstructural failure mechanisms in cokes from single and binary coal blends, showing that interface porosity and pore wall collapse critically influence compressive strength, reinforcing the need for high-resolution petrographic features in predictive modelling.

The significance of coal petrography and intermediate properties in determining final coke quality has been widely reported. [8] integrated Automated Coal Petrography (ACP) into statistical models, improving prediction accuracy for M metrics by quantifying maceral distribution and associations. [9] emphasized particle-level fusible content as a determinant of mechanical strength, while [10] established the predictive value of Optical Texture (OT) as an intermediate feature bridging coal inputs and CRI/CSR outputs. Structural characteristics of coals, such as aromaticity, condensation rings, and aliphatic hydrogen content, have also been highlighted as key features influencing fluidity and swelling indices [11]; [12].

The role of fluidity and thermoplastic properties has been extensively studied. [13] demonstrated that coal fluidity and interaction between vitrinites (IVR) govern coking pressure and microstructural integrity, establishing critical constraints for optimization. [14] and [15] developed additive and parabolic reconstruction models for blend plastograms, enabling accurate prediction of Gieseler fluidity (logMF, PR), though these approaches do not fully capture non-linear effects.

Machine learning has emerged as a transformative tool for predicting coke quality indices. developed a Back-Propagation Neural Network (BP-NN) to predict ignition properties of coal blends, achieving high accuracy and demonstrating the economic benefits of blend optimization. [16] used Boost for predicting caking ability (Roga Index), confirming the non-additive nature of blends and highlighting moisture, volatile matter, and vitrinite content as critical features. [17] combined SHAP and XGBoost to predict Free Swelling Index and Maximum Fluidity, offering interpretability for industrial decision-making.[18], [19] advanced this further by employing CNN-RF hybrids, interpretable deep learning models, and Adaptive Graph Neural Networks to accurately predict CRI and CSR, even under missing data conditions, providing guidance for feature selection, ranking, and trustworthiness of predictions.

Hybrid modelling strategies integrating mechanistic and AI approaches have also been proposed. [20] demonstrated superior prediction of H_2 and CH_4 concentrations in coke oven gas using Light calibrated with Aspen Plus simulations, enabling optimization of by-product economics. [21] used GA-LSTM hybrids to simulate energy consumption in coking, quantifying savings linked to coal properties and validating energy efficiency objectives. Some studies applied BP-MLPNN for syngas prediction with SHAP analysis, confirming the importance of operational parameters alongside coal characteristics [22].

Optimization frameworks leveraging machine learning predictions have been widely explored. [23] integrated Ga-XGBoost-SVR models with Modified Particle Swarm Optimization (MPSO) to optimize coal blending cost while meeting CSR/CRI constraints. [24] combined Improved Genetic Algorithms with residual prediction models to achieve 6% cost reduction. [25] implemented Multi-Objective Coal Blending (MOCB) using BPNNs and QNSGA-III, simultaneously optimizing seven objectives, including cost and all major coke quality metrics. [26] demonstrated ANN-MOGA and ensemble learning-based approaches for multi-objective optimization in co-pyrolysis, supporting methodological robustness for MOO tasks in coke production. [27] applied metaheuristic-tuned Random Forests to predict Abrasive Index, integrating cost minimization by reducing equipment wear, directly applicable to coke-making optimization.

Coal structural and chemical characteristics continue to provide critical inputs for predictive modelling. [28] introduced novel parameters capturing plastic layer behaviour, while [29] correlated bi reflectance with CRI/CSR. [30] explored the “coking component” of Indian coal, demonstrating potential for cost-effective additive use. [31] and [2] emphasized the influence of biomass and ash composition on reactivity and CRI, informing blend constraints and secondary features for predictive models. The predictive framework is further strengthened by studies linking swelling, thermoplastic properties, and coking pressure to operational constraints [32]; [11]; [33].

Finally, reinforcement learning and surrogate modelling offer paths for next-generation operational control. [34] demonstrated DRL-based heat and pyrolysis control, while [35] provided interpretable neural network strategies for prediction aligned with domain knowledge. These studies confirm the potential for real-time control, bridging predictive modelling and operational execution. Three persistent gaps emerge. First, end-to-end integration remains rare: high-accuracy prediction, interpretable feature design, inventory-aware constraints, and multi-objective optimization are often studied in isolation rather than as a unified industrial system ([36]; [4]). Second, generalization across geological diversity and plant contexts is limited, requiring adaptive architectures and transfer learning to preserve accuracy across regions and campaigns. Third, real-time control integration is nascent; while DRL and surrogate modelling show promise, coupling closed-loop control with inventory-aware optimization and online quality prediction is an open opportunity for resilient, autonomous coke making [37].

3. METHODOLOGY

The proposed framework integrates predictive modelling, fuzzy multi-criteria decision analysis, and multi-objective optimization to achieve metallurgical equivalence in coke properties while minimizing overall blending costs. This enables greater flexibility in raw material sourcing and enhances operational resilience under uncertain supply conditions. Two primary datasets collected from the industry were employed—coal blend composition and charging coal characteristics—corresponding to a set of coke quality parameters.

1) The coal blend dataset contained ten distinct coal types, each represented by its blending percentage in the production mix.
2) Charging coal characteristics included proximate and petrographic properties such as ash content, volatile matter, fixed carbon, free swelling index, sulphur content, and mean maximum reflectance.
3) Coke quality was described through parameters including mechanical strength (M_40, M_10), reactivity index (CRI), coke strength after reaction (CSR), and compositional metrics such as sulphur, moisture, fixed carbon, volatile matter, and ash.

Total 61 (Batches) observations representing industrial campaigns from a functioning integrated steel plant were analysed. These were used as the training and validation basis for subsequent predictive modelling. To predict coke quality responses from both coal blend and charging coal characteristics, a multi-model regression framework was developed. Five machine learning algorithms—multiple linear regression (MLR), partial least squares regression (PLSR), decision tress (DT), random forest regression (RFR), and feed-forward neural networks (FNN)—were trained independently on both datasets. Each model was trained using 70% of the data for learning, 15% for validation, and 15% for testing.

Multiple Linear Regression (MLR) was implemented using ordinary least squares estimation without regularization. No hyperparameters were tuned, as the model assumes a linear relationship between coal blend proportions (or charging characteristics) and coke quality indices. The closed-form solution minimizes the sum of squared residuals, providing a baseline for comparison with nonlinear models. Intercept was included by default. PLSR used SIMPLS with 10 components (blend data) or 6 (charging data). Inputs were centered but not scaled. DT applied CART with MSE impurity, min leaf size = 1, and no pruning. Surrogate splits handled potential missing values. RF trained 100 trees via bagging (100% row sampling). Base DT had min leaf size = 5; feature bagging was enabled. The Feedforward Neural Network (FNN) consisted of one hidden layer with 15 neurons, selected via preliminary grid search to balance model capacity and convergence speed. The Levenberg–Marquardt algorithm (trainlm) was used for training due to its efficiency on medium-sized datasets. Hyperbolic tangent sigmoid (tansig) was applied as the hidden layer activation function, and linear (purelin) for the output layer to allow unbounded coke quality predictions. The dataset was split internally into 70% training, 15% validation, and 15% test sets using MATLAB's dividerand with default random seeding.

Model performance was evaluated using four statistical indicators: root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2). A fuzzy-based multi-criteria evaluation system was then employed to rank all candidate models. Specifically, a fuzzy CRITIC–WASPAS integration framework was developed, where:

1) The CRITIC method assigned relative importance weights to each performance metric based on its discriminating power.
2) The weighted aggregated sum product assessment (WASPAS) method consolidated these criteria into a composite performance index.

The model attaining the highest fuzzy performance index was selected as the optimal predictive model. This model was then used for subsequent perturbation and optimization stages. To assess the responsiveness of coke quality parameters to small deviations in individual coal proportions, a perturbation-based sensitivity analysis was conducted on the selected predictive model. Each input variable (representing a coal type) was perturbed by ±5% and ±10% around its mean value, while all others were held constant. The resulting percentage change in each coke quality output was quantified, forming a sensitivity matrix that identified the most influential blend constituents. This analysis provided crucial metallurgical insights—highlighting which coals governed strength indices (M_{40}, CSR) versus those that predominantly affected reactivity and ash behaviour—thereby guiding the optimization bounds and constraints in the subsequent stage. A Multi-Objective Genetic Algorithm (MOGA) was employed to solve the coal blend optimization problem where all coke-quality variables are expressed in consistent units (percentage).

Let,

Symbol	***Description***
x_i	Proportion (%) of coal *i* in the blend
C_i	Unit cost of coal *i* (₹/ton) (refer Table 1)
Q_j^{pred} $(x_1, \ldots, x_n)$	Predicted value of coke quality parameter *j* from the selected ML model
Q_j^{ref}	Reference (benchmark) value of coke quality parameter *j*
w_j	Importance weight of coke quality parameter *j* (normalized)
n	Total number of coal types in the blend
m	Total number of coke quality parameters considered

The algorithm generates a set of Pareto-optimal solutions that represent trade-offs between minimizing blend cost and preserving coke quality. Each chromosome in the population encodes a unique coal blend vector

$$\mathrm{x} = [x_1, x_2, \ldots, x_n]$$

Where $x_i$denotes the proportion of coal $i$in the blend. The selected machine learning model acts as an embedded evaluator, predicting coke quality parameters $Q_j^{\mathrm{pred}}(x_1, \ldots, x_n)$for each candidate solution during fitness evaluation. The optimization simultaneously minimizes two objective functions:

1) Coke Quality Preservation – minimizing the weighted squared deviation from benchmark coke quality values.

$$min\, f_1(x) \;=\; \sum_{\mathrm{j}} = 1^{\mathrm{m}}\, w_{\mathrm{j}} \,*\, ((Q_{\mathrm{j}}{}^{\mathrm{pred}}\,(x_1, \ldots, x_{\mathrm{n}}) \,-\, Q_{\mathrm{j}}{}^{\mathrm{ref}}) \,/\, Q_{\mathrm{j}}{}^{\mathrm{ref}})\ ^2$$

2) Cost Minimization – minimizing the total cost of the blend based on coal prices and proportions.

$$min\, f_2(x) = \sum_{\mathrm{i}} = 1^{\mathrm{n}}\, C_{\mathrm{i}} * x_{\mathrm{i}}$$

The optimization is subject to constraints including:

1) The total proportion of all coal types in the blend must sum to 100%:

$$\sum (i = 1\, to\, n)\, x_{\mathrm{i}} = 100$$

2) Each coal type must lie within its practical blending limits:

$$x_{\mathrm{i}}\_min \leq x_{\mathrm{i}} \leq x_{\mathrm{i}}\_max, for\ i = 1, 2, \ldots, n$$

3) For any coal temporarily unavailable due to logistical or geopolitical constraints:

$$x_{\mathrm{k}} = 0$$

4) To ensure minimum mechanical and metallurgical performance (Optional):

$$Q_{\mathrm{j}}\_pred(x_1, \ldots, x_{\mathrm{n}}) \geq Q_{\mathrm{j}}\_min$$

The Multi-Objective Genetic Algorithm was implemented using MATLAB's gamultiobj solver based on NSGA-II. The population size was set to 200 individuals, and evolution ran for a maximum of 200 generations or until 50 consecutive generations showed no improvement in the Pareto front (stall criterion). Crossover fraction was 0.8 (default), using intermediate recombination to generate offspring within parental bounds. Mutation followed the adaptive feasible method with a rate of approximately 1/nVar (~0.1), ensuring constraint satisfaction. Tournament selection (size = 4, default) was used to promote elite solutions. The NN served as a surrogate fitness evaluator, and parallel evaluation was disabled for reproducibility. Russian Soft coal proportion was hard constrained to zero via lower and upper bounds, sum(x) = 100 and $0 \leq x_{\mathrm{i}} \leq 100$ were the applied constraints.

Table 1: Fixed (per ton) costs noted during the time of study.

Symbol	**Unit Cost (₹/ton)**
PCC	6,000
MCC	3,000
A_H	22,140
U_H	21,600
A_S	16,072
U_S	19,680
R_S	12,350
LVHR	9,430
BEN	7,200
TUH	16,000

The machine-learning models and multi-objective optimisation were implemented in MATLAB R2024b, using the Neural Network Toolbox, Statistics & Machine Learning Toolbox, and the Global Optimization Toolbox. All computations were performed on a workstation equipped with an AMD Ryzen 5 5600H processor (6 cores / 12 logical threads, 3.3 GHz), 8 GB RAM, running Windows 11 (64-bit). For the MOGA (NSGA-II) optimisation, 4 logical processors were allocated for parallel evaluation. The industrial dataset used for modelling originates from a working coke-making unit employing 7-metre slot-type coke ovens. This ensures the model is reproducible with affordable hardware can tested on opensource software as well, both in academia and in the industry.

This approach enables resilient and cost-effective coal blending strategies that maintain metallurgical performance under dynamic supply conditions.

The proposed framework thus establishes a closed-loop coal blending model that integrates:

1) Predictive Modelling (machine learning regression for coke quality estimation),
2) Performance Synthesis (fuzzy CRITIC–WASPAS for model selection),
3) Sensitivity Diagnostics (input perturbation for metallurgical influence mapping), and
4) Optimization (multi-objective programming for resilient blend planning).

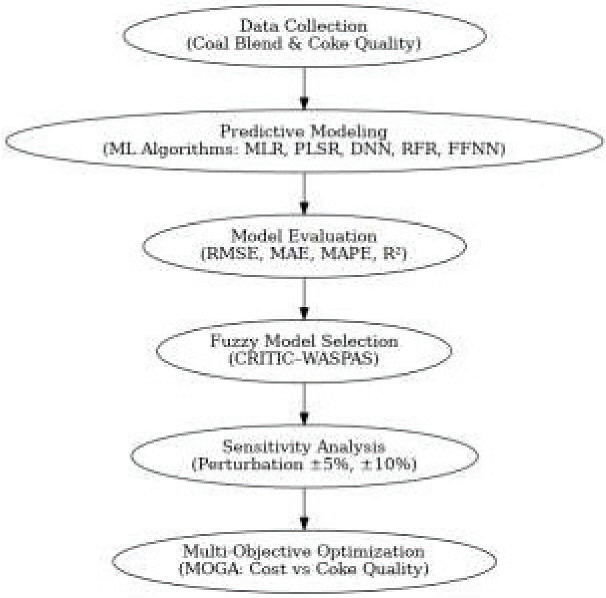

Figure 3: Methodology flowchart

This integrated methodology enables the formulation of adaptive coal blending strategies that maintain consistent coke quality despite variations in supply, cost, or regional availability. By coupling machine learning with fuzzy logic and evolutionary optimization, the system effectively transforms historical plant data into an intelligent decision-support tool for metallurgical operations planning.

4. RESULTS AND DISCUSSION

This section presents a comprehensive evaluation of the predictive performance of five machine learning models—Multiple Linear Regression (MLR), Decision Tree (DT), Random Forest (RF), Partial Least Squares Regression (PLS), and Feedforward Neural Network (NN)—across eight key coke quality parameters. The models were assessed using four statistical metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2).

Tables 2 through 10 summarize the performance of each model across the coke quality parameters: M_{40}, M_{10}, CRI, CSR, sulphur content, moisture, fixed carbon, volatile matter, and ash. Mechanical Strength (M_{40}, M_{10}): The Decision Tree (DT) model outperformed others, achieving the lowest RMSE and MAE values for both M_{40} and M_{10} (Table 2 & Table 3). For instance, DT achieved an R^2 of 0.9024 for M_{40} and 0.8075 for M_{10}, indicating strong predictive accuracy. Reactivity and Strength After Reaction (CRI, CSR): DT again demonstrated superior performance, with the highest R^2 values of 0.7974 for CRI and 0.8359 for CSR (Tables 4 & 5). This suggests its robustness in capturing the nonlinear dependencies between coal blends and coke reactivity. Compositional Parameters (Sulphur, Moisture, Fixed Carbon, Volatile Matter, Ash): The Neural Network (NN) model showed better generalization for volatile and compositional parameters, particularly in predicting sulphur and moisture content (Tables 6, 7, 8, 9 & 10). This is attributed to its ability to model complex nonlinear relationships. MLR and PLS consistently underperformed across most parameters, with lower R^2 values and higher error metrics. This indicates their limited capacity to capture the intricate interactions among coal constituents and coke properties.

Table 2: Performance metrics of ML models for predicting M_{40}

Output Var	Model	RMSE	MAE	MAPE	R²
M_40	MLR	0.5293	0.3828	0.48%	0.6282
	DT	0.2712	0.188	0.24%	0.9024
	RF	0.457	0.332	0.42%	0.7229
	PLS	0.5293	0.3828	0.48%	0.6282
	NN	0.3596	0.2513	0.32%	0.8284

Table 3: Performance metrics of ML models for predicting M_{10}

Output Var	Model	RMSE	MAE	MAPE	R²
M_10	MLR	0.1504	0.1239	1.64%	0.4786
	DT	0.0914	0.0575	0.77%	0.8075
	RF	0.118	0.0903	1.20%	0.6794
	PLS	0.1504	0.1239	1.64%	0.4786
	NN	0.1652	0.1311	1.74%	0.3715

Table 4: Performance metrics of ML models for predicting CRI

Output Var	Model	RMSE	MAE	MAPE	R²
CRI	MLR	0.2845	0.2278	0.98%	0.4298
	DT	0.1696	0.1125	0.48%	0.7974
	RF	0.2084	0.1637	0.70%	0.6941
	PLS	0.2845	0.2278	0.98%	0.4298

Table 5: Performance metrics of ML models for predicting CSR

Output Var	Model	RMSE	MAE	MAPE	R²
CSR	MLR	0.3444	0.2812	0.43%	0.4197
	DT	0.1831	0.1121	0.17%	0.8359
	RF	0.2427	0.2047	0.32%	0.7118
	PLS	0.3444	0.2812	0.43%	0.4197
	NN	0.3219	0.237	0.36%	0.4929

Table 6: Performance metrics of ML models for Predicting Sulphur

Output Var	Model	RMSE	MAE	MAPE	R²
Coke Sulphur	MLR	0.0753	0.0422	6.11%	0.1626
	DT	0.0212	0.0166	1.20%	0.9337
	RF	0.0632	0.0299	2.00%	0.4108
	PLS	0.0753	0.0422	3.12%	0.1626
	NN	0.0795	0.0537	1.60%	0.0682

Table 7: Performance metrics of ML models for Predicting Moisture

Output Var	Model	RMSE	MAE	MAPE	R²
Moisture Content	MLR	0.2145	0.1798	5.34%	0.2713
	DT	0.1301	0.1006	3.00%	0.732
	RF	0.1614	0.1282	3.78%	0.5872
	PLS	0.2145	0.1798	5.34%	0.2713
	NN	0.215	0.165	4.84%	0.2679

Table 8: Performance metrics of ML models for predicting Fixed Carbon

Output Var	Model	RMSE	MAE	MAPE	R²
Fixed Carbon	MLR	0.2115	0.1688	0.20%	0.2421
	DT	0.1196	0.0834	0.10%	0.7579
	RF	0.1639	0.1332	0.16%	0.545
	PLS	0.2115	0.1688	0.20%	0.2421
	NN	0.2181	0.1744	0.21%	0.1941

Table 9: Performance metrics of ML models for predicting Volatile Material

Output Var	Model	RMSE	MAE	MAPE	R²
Volatile Matter	MLR	0.0146	0.013	1.38%	0.3206
	DT	0.0082	0.0041	0.43%	0.784
	RF	0.011	0.0096	1.02%	0.6113
	PLS	0.0146	0.013	1.38%	0.3206
	NN	0.0228	0.0186	1.96%	0.6691

Table 10: Performance metrics of ML models for predicting Ash

Output Var	Model	RMSE	MAE	MAPE	R²
Ash	MLR	0.2115	0.1688	0.20%	0.2421
	DT	0.1196	0.0834	0.10%	0.7579
	RF	0.1639	0.1332	0.16%	0.545
	PLS	0.2115	0.1688	0.20%	0.2421
	NN	0.2181	0.1744	0.21%	0.1941

Fuzzy CRITIC+WASPAS was applied on the obtained results to select the best performing model (Table 11 and 12).

Table 11: WASPAS Rank for Performance Metrics

Metric	Weight
MAE	0.1517
MAPE	0.0019
RMSE	0.1992
R²	0.6472

Table 12: WASPAS Rank for all models

Participant	WSM	WPM	WASPAS Q	Rank
NN	0.8303	0.8212	0.8258	1
RF	0.8237	0.8231	0.8234	2
DT	0.8247	0.7848	0.8047	3
MLR	0.8033	0.791	0.7972	4
PLS	0.8033	0.791	0.7972	5

NN emerged as a clear winner. The research here on, was focused on neural networks. A perturbation-based sensitivity analysis was conducted on the selected NN model by varying each coal type's proportion by ±10% around its mean. The resulting changes in coke quality parameters were used to construct a sensitivity matrix (Figure 4).

The analysis revealed nonlinear and asymmetric sensitivities, consistent with real-world coking behaviour. The correlation analysis revealed clear relationships between coal proportions and coke-quality outputs. A_S exhibited the strongest positive correlation with CRI (+0.57) and the strongest negative correlation with CSR (−0.46), making it the most sensitive contributor to reactivity-based indices. PCC showed moderate positive correlation with volatile matter (+0.51) and mild positive correlation with CSR, supporting its beneficial influence. U_H displayed the largest negative correlation with M_{40} (−0.39), indicating a weakening effect on mechanical strength, while U_S had the highest positive correlation with M_{10} (+0.37) and moisture (+0.36), reflecting its sensitivity toward degradation of size and moisture characteristics. A_H and MCC showed mixed but moderate correlations across outputs, while LVHR and BEN showed relatively uniform and low-magnitude correlations, indicating stable **behaviour.**

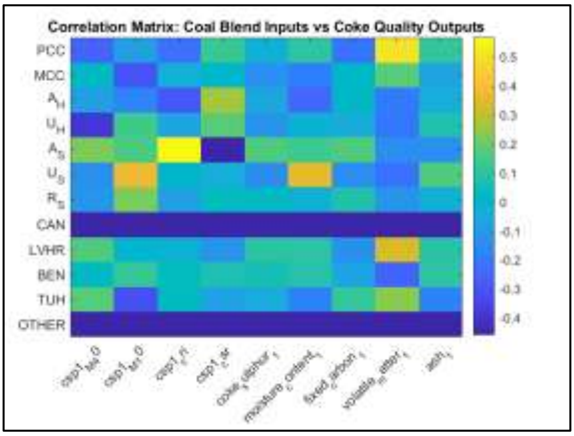

Figure 4: *Correlation Matrix obtained for Neural Networks*

The sensitivity analysis (Figure 5 to 14) reveals that increasing U_S causes increases in M_{10} (up to +1.41) and Moisture (+3.41), while reducing M_{40} by as much as −0.61, indicating deterioration in mechanical strength at higher proportions. A similar effect is observed for TUH, where decreases in M_{10} (down to −1.45) and substantial increases in Moisture (+1.84 to +3.97) highlight its sensitivity to moisture-driven behaviour. In contrast, A_H and A_S primarily influence mechanical indices and reactivity. A_H increases M_{10} and reduces M_{40}, while A_S elevates CRI and moisture, indicating their role in reducing coke integrity. Coals such as BEN and LVHR exhibit minimal impact on most outputs, inducing only small changes (<±0.30), confirming their compositional stability and limited metallurgical sensitivity and MCC show moderate but targeted effects, with PCC raising volatile matter (+0.24 to +0.33) and MCC affecting fixed carbon and ash content. Their influence remains contained relative to U_S and TUH. Finally, U_H produces asymmetric effects, decreasing M_{40} (−0.19 to −0.32) while sharply increasing Moisture (+1.39), revealing trade-offs between strength and moisture sensitivity.

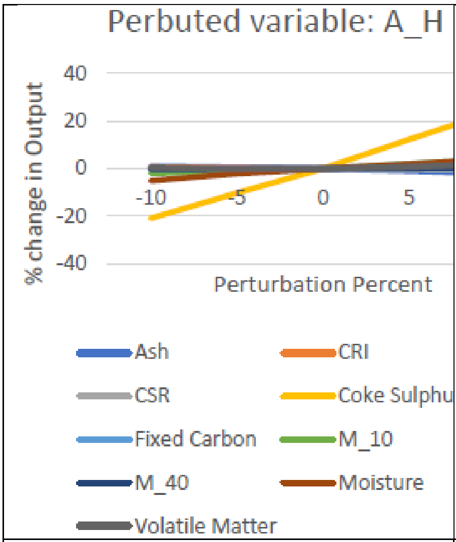

Figure 5: Sensitivity of coke quality parameters to A_H variation.

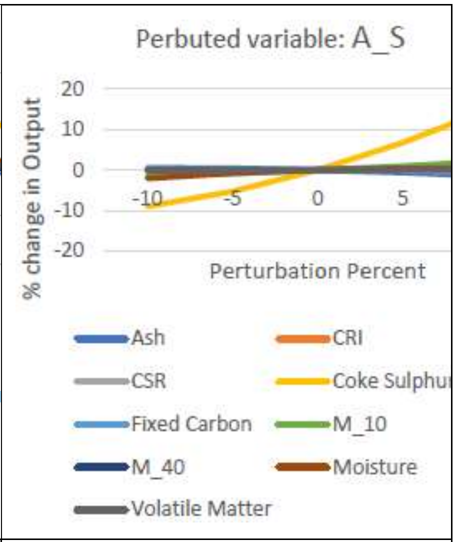

Figure 6: Sensitivity of coke quality parameters to A_S variation.

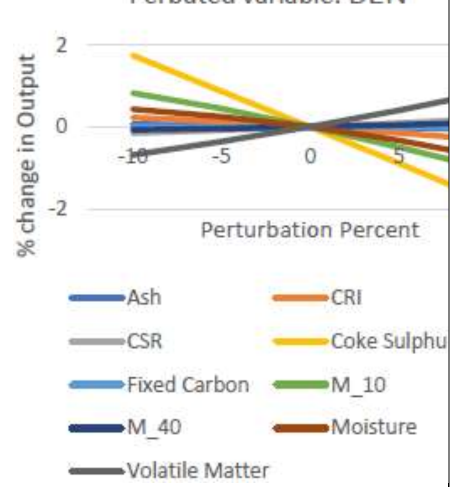

Figure 7: Sensitivity of coke quality parameters to Benga variation.

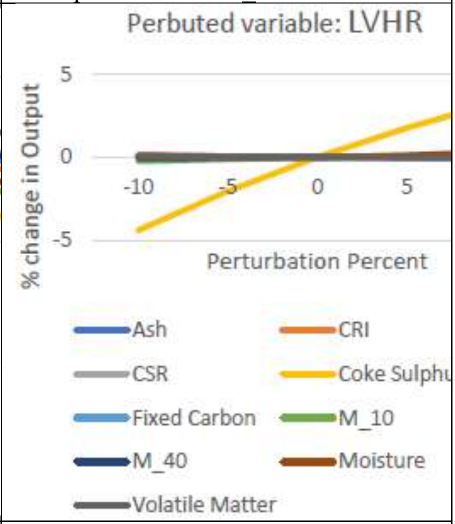

Figure 8: Sensitivity of coke quality parameters to LVHR variation.

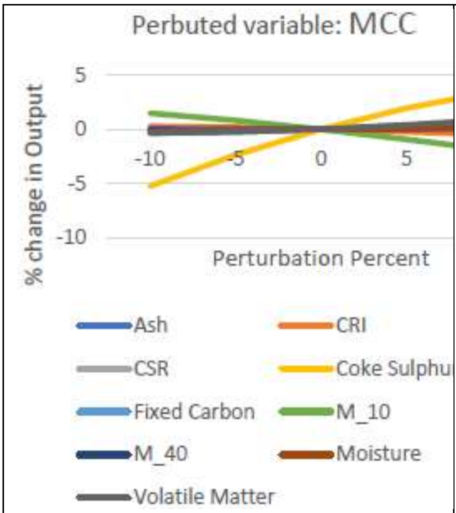

Figure 9: Sensitivity of coke quality parameters to MCC variation.

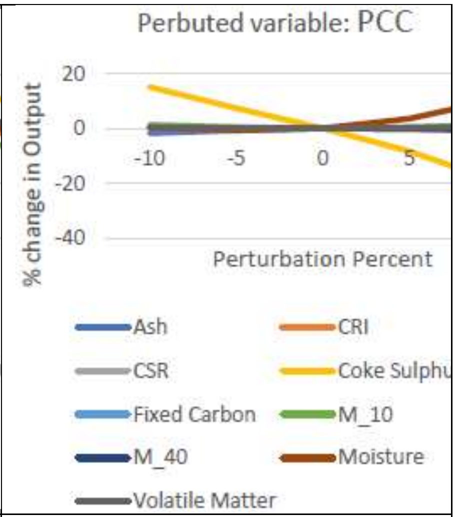

Figure 10: Sensitivity of coke quality parameters to PCC variation.

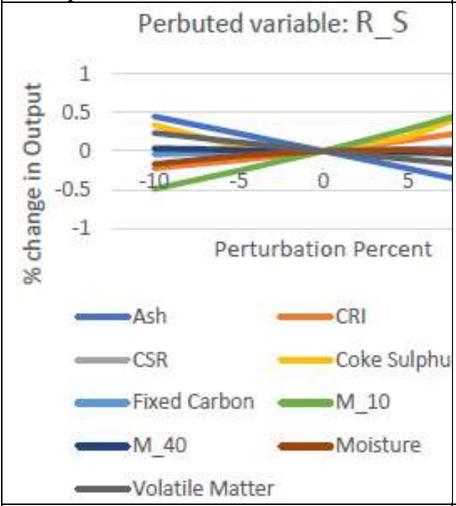

Figure 11: Sensitivity of coke quality parameters to R_S variation.

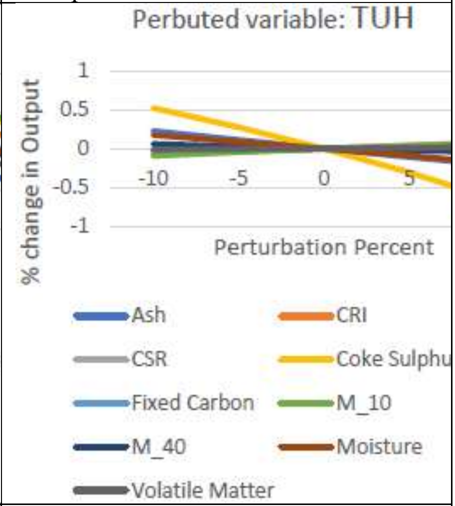

Figure 12: Sensitivity of coke quality parameters to TUH variation.

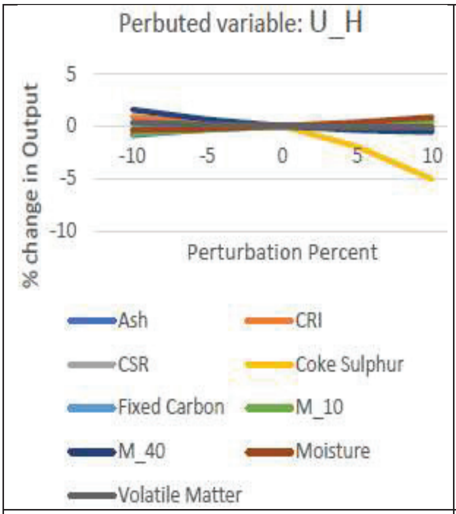

Figure 13: Sensitivity of coke quality parameters to U_H variation.

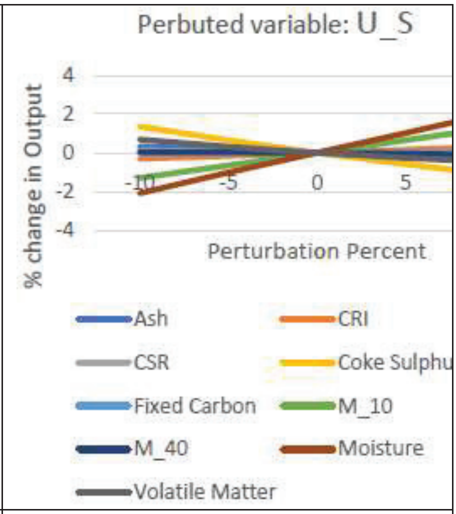

Figure 14: Sensitivity of coke quality parameters to U_S variation.

The sensitivity analysis helps us understand the exact scale of the challenge, for this study, assessment of relation between Russia soft coal was quintessential for upcoming step i.e. MOGA where its percentage in charging blend is suppressed to zero. The Neural Network model was embedded within a Multi-Objective Genetic Algorithm (MOGA) to optimize coal blend proportions under inventory and quality constraints. The optimization aimed to minimize total blend cost while preserving coke quality, particularly M_{40} and CSR.

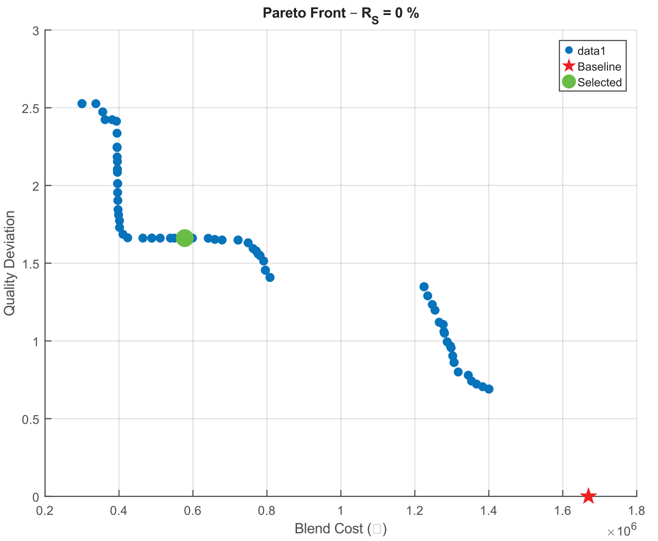

Figure 15: Pareto Front Visualisation for Quality Deviation VS Cost Trade-off

Table 13: Baseline and optimized blend compositions with max and min observations from training data.

Variable	*Baseline (%)*	*Optimized (%)*	*Max Observed (%)*	*Min Observed (%)*
PCC	12.04	21.92	13.4	6
MCC	2.46	3.97	3.8	1.9
A_H	39.02	7.86	43.3	36.3
U_H	11.92	2.02	13.4	10.4
A_S	10.32	4.87	12.4	6.5
U_S	5.26	23.26	17.1	3.7
R_S	5.59	0.00	7.4	1.7
LVHR	5.37	3.67	8.2	3.5
BEN	5.53	6.82	8.7	3.4
TUH	2.49	25.61	15.1	0
Cost/Ton (₹)	16,695.94	13,906.05	-	-

Table 14: Comparison between benchmark output and output for the optimised solution.

Property	**Baseline**	**Optimized**
M_40	79.17	80.05
M_10	7.59	7.69
CRI	23.33	23.42
CSR	64.93	64.82
Coke Sulphur	0.578	0.574
Moisture	3.38	3.60
Fixed Carbon	83.68	83.48
Volatile Matter	0.93	0.96
Ash	14.89	15.21

The final optimization (Table 13 and 14) achieved 16.7% cost savings while preserving coke quality, demonstrating the practical viability of the proposed framework under real-world inventory constraints.

5. Conclusion:

This study developed a machine learning–based optimization framework for inventory-aware coal blending in coke production. Among five models evaluated, Decision Trees and Neural Networks outperformed traditional linear methods. The NN model, selected via fuzzy CRITIC–WASPAS analysis, showed strong generalization across coke quality parameters, making it ideal for optimization tasks attaining both the goals of neatly predicting outputs and then optimising the blends. The multi-objective genetic algorithm achieved a 16.7% cost reduction while maintaining coke quality. It eliminated Russian Soft coal, increased PCC usage, and replaced expensive A_H with TUH and U_S, demonstrating the framework's ability to adapt to supply constraints and optimize costs effectively.

Future work could extend this framework to dry quenching systems for environmental benefits, integrate real-time sensor data for dynamic control, and explore alternative feedstocks like biomass. These enhancements would support sustainable, autonomous coke production across diverse steelmaking contexts. Due to the limited dataset size (61 rows), a 70/15/15 train–validation–test split was used. k-fold cross-validation was not applied to avoid excessively small fold sizes, which can lead to unstable estimates. The random split implemented in MATLAB's trainlm and divider and ensures that validation and test partitions remain independent, though future work may incorporate repeated or k-fold validation as additional data becomes available. MAPE has known limitations when applied to variables with narrow operating ranges or observations close to zero. This affects metrics such as M_{40} and CSR—which typically fluctuate within tight operational bands as well as volatile matter, where small absolute changes can produce disproportionately large percentage errors. Future work can be based around MASE instead. Similarly, NMSE can be an alternative to RMSE owing to better contextualization former adds to the study. Seasonal cost drifts can also be incorporated by adding a month variable (Classification) nesting any algorithm inside to include seasonality provided a larger data is collected.

References:

[1] S. Mittal, S. Pathak, H. Dhawan, and S. Upadhyayula, "A machine learning approach to improve ignition properties of high-ash Indian coals by solvent extraction and coal blending," *Chemical Engineering Journal*, vol. 413, Jun. 2021, doi: 10.1016/j.cej.2020.127385.

[2] B. Brooks, S. K. Rish, H. Lomas, A. Jayasekara, and A. Tahmasebi, "Advances in low carbon cokemaking – Influence of alternative raw materials and coal properties on coke quality," Aug. 01, 2023, *Elsevier B.V.* doi: 10.1016/j.jaap.2023.106083.

[3] R. Sakurovs, "Interactions between coking coals in blends q," 2002. [Online]. Available: www.fuelfirst.com

[4] L. Meng, W. Zhu, N. Xu, and J. C. Hower, "Explainable AI for coal Hardgrove grindability index prediction: a genetic programming and linear regression approach," *Fuel*, vol. 404, Jan. 2026, doi: 10.1016/j.fuel.2025.136072.

[5] R. Sakurovs, "Interactions between coking coals in blends q," 2002. [Online]. Available: www.fuelfirst.com

[6] R. Roest, H. Lomas, and M. R. Mahoney, "Fractographic approach to metallurgical coke failure analysis. Part 2: Cokes from binary coal blends," *Fuel*, vol. 180, pp. 794–802, Sep. 2016, doi: 10.1016/j.fuel.2016.04.022.

[7] R. Roest, H. Lomas, K. Hockings, and M. R. Mahoney, "Fractographic approach to metallurgical coke failure analysis. Part 1: Cokes of single coal origin," *Fuel*, vol. 180, pp. 785–793, Sep. 2016, doi: 10.1016/j.fuel.2016.01.058.

[8] S. Gupta, F. Shen, W. J. Lee, and G. O'Brien, "Improving coke strength prediction using automated coal petrography," *Fuel*, vol. 94, pp. 368–373, Apr. 2012, doi: 10.1016/j.fuel.2011.09.045.

[9] K. Warren, G. Krahenbuhl, M. Mahoney, G. O'Brien, and P. Hapugoda, "Estimating the fusible content of individual coal grains and its application in coke making," *Int J Coal Geol*, vol. 152, pp. 3–9, Nov. 2015, doi: 10.1016/j.coal.2015.09.012.

[10] B. D. Flores *et al.*, "How coke optical texture became a relevant tool for understanding coal blending and coke quality," *Fuel Processing Technology*, vol. 164, pp. 13–23, 2017, doi: 10.1016/j.fuproc.2017.04.015.

[11] B. Cui, B. Wu, M. Wang, X. Jin, Y. Shen, and L. Chang, "A preliminary study on the quality evaluation of coking coal from its structure thermal transformation:

Applications of fluidity and swelling indices," *Fuel*, vol. 355, Jan. 2024, doi: 10.1016/j.fuel.2023.129418.

[12] J. Guo *et al.*, "Impact of chemical structure of coal on coke quality produced by coals in the similar category," *J Anal Appl Pyrolysis*, vol. 162, Mar. 2022, doi: 10.1016/j.jaap.2022.105432.

[13] J. J. Duffy, M. R. Mahoney, and K. M. Steel, "Influence of coal thermoplastic properties on coking pressure generation: Part 2 - A study of binary coal blends and specific additives," *Fuel*, vol. 89, no. 7, pp. 1600–1615, Jul. 2010, doi: 10.1016/j.fuel.2009.08.035.

[14] N. Guelton, "The prediction of the Gieseler characteristics of coal blends," *Fuel*, vol. 209, pp. 661–673, Dec. 2017, doi: 10.1016/j.fuel.2017.07.017.

[15] F. Gayo, R. García, and M. A. Diez, "Modelling the Gieseler fluidity of coking coals modified by multicomponent plastic wastes," *Fuel*, vol. 165, pp. 134–144, Feb. 2016, doi: 10.1016/j.fuel.2015.10.053.

[16] M. Rzychoń, A. Żogała, and L. Róg, "Experimental study and extreme gradient boosting (XGBoost) based prediction of caking ability of coal blends," *J Anal Appl Pyrolysis*, vol. 156, Jun. 2021, doi: 10.1016/j.jaap.2021.105020.

[17] A. Homafar, H. Nasiri, and S. C. Chelgani, "Modeling coking coal indexes by SHAP-XGBoost: Explainable artificial intelligence method," *Fuel Communications*, vol. 13, p. 100078, Dec. 2022, doi: 10.1016/j.jfueco.2022.100078.

[18] Y. Qiu *et al.*, "A novel image expression-driven modeling strategy for coke quality prediction in the smart cokemaking process," *Energy*, vol. 294, May 2024, doi: 10.1016/j.energy.2024.130866.

[19] Y. Qiu *et al.*, "The interpretable deep learning-based coke quality prediction strategy aligned with domain knowledge for sustainable smart cokemaking applications," *Fuel*, vol. 398, Oct. 2025, doi: 10.1016/j.fuel.2025.135533.

[20] Y. Lei, Y. Chen, J. Chen, X. Liu, X. Wu, and Y. Chen, "A novel modeling strategy for the prediction on the concentration of H2 and CH4 in raw coke oven gas," *Energy*, vol. 273, Jun. 2023, doi: 10.1016/j.energy.2023.127126.

[21] Y. Ji, G. Sheng, Q. Zhang, and J. Shen, "Energy consumption simulation and diagnosis in the coking process: A mechanism and data-driven approach," *Energy for Sustainable Development*, vol. 88, Oct. 2025, doi: 10.1016/j.esd.2025.101763.

[22] Y. jie Zhao, J. cheng Wang, and Q. Yi, "Bridging uncertainty gaps with artificial intelligence-assisted syngas precise prediction in coal gasification," *Chem Eng Sci*, vol. 301, Jan. 2025, doi: 10.1016/j.ces.2024.120734.

[23] Y. Yuan, Q. Qu, L. Chen, and M. Wu, "Modeling and optimization of coal blending and coking costs using coal petrography," *Inf Sci (N Y)*, vol. 522, pp. 49–68, Jun. 2020, doi: 10.1016/j.ins.2020.02.072.

[24] M. Liu, Z. Yu, B. Li, Q. Wang, H. Ren, and D. Xu, "Coal allocation optimization based on a hybrid residual prediction model with an improved genetic algorithm," *Eng Appl Artif Intell*, vol. 137, Nov. 2024, doi: 10.1016/j.engappai.2024.109072.

[25] Z. Li, L. Liu, Z. Zhao, H. Jin, and Y. Zhuo, "Optimizing coal blending for metallurgical coke production using a reinforcement learning-enhanced evolutionary algorithm," *Fuel*, vol. 406, Feb. 2026, doi: 10.1016/j.fuel.2025.136777.

[26] Q. Yang, D. Rong, Q. Guo, R. Bao, and D. Zhang, "Ensemble learning-driven multi-objective optimization of the co-pyrolysis process of biomass and coal for high economic and environmental performance," *Chin J Chem Eng*, Aug. 2025, doi: 10.1016/j.cjche.2025.06.001.

[27] H. Qi *et al.*, “An optimized machine learning framework for prediction of coal abrasive index: Leveraging supervised learning, metaheuristic optimization, and interpretability analysis,” *Fuel*, vol. 403, Jan. 2026, doi: 10.1016/j.fuel.2025.136065.

[28] W. jia Hu *et al.*, “Relevance between various phenomena during coking coal carbonization. Part 3: Understanding the properties of the plastic layer during coal carbonization,” *Fuel*, vol. 292, May 2021, doi: 10.1016/j.fuel.2021.120371.

[29] S. Pusz and R. Buszko, “Reflectance parameters of cokes in relation to their reactivity index (CRI) and the strength after reaction (CSR), from coals of the Upper Silesian Coal Basin, Poland,” *Int J Coal Geol*, vol. 90–91, pp. 43–49, Feb. 2012, doi: 10.1016/j.coal.2011.10.008.

[30] A. Mohanty, S. Chakladar, S. Mallick, and S. Chakravarty, “Structural characterization of coking component of an Indian coking coal,” *Fuel*, vol. 249, pp. 411–417, Aug. 2019, doi: 10.1016/j.fuel.2019.03.108.

[31] M. A. Diez and A. G. Borrego, “Evaluation of CO2-reactivity patterns in cokes from coal and woody biomass blends,” *Fuel*, vol. 113, pp. 59–68, 2013, doi: 10.1016/j.fuel.2013.05.056.

[32] Y. Mochizuki, Y. Wang, S. Konno, T. Shishido, and N. Tsubouchi, “Analysis of swelling in coke formed by carbonization of low-quality coal in a pressurized atmosphere,” *Carbon Resources Conversion*, 2025, doi: 10.1016/j.crcon.2025.100370.

[33] Z. Klika, J. Serenčíšová, I. Kolomazník, L. Bartoňová, and P. Baran, “Prediction of CRI and CSR of cokes by two-step correction models for stamp-charged coals – Statistical analysis,” *Fuel*, vol. 262, Feb. 2020, doi: 10.1016/j.fuel.2019.116623.

[34] Y. jie Zhao, J. cheng Wang, and Q. Yi, “Bridging uncertainty gaps with artificial intelligence-assisted syngas precise prediction in coal gasification,” *Chem Eng Sci*, vol. 301, Jan. 2025, doi: 10.1016/j.ces.2024.120734.

[35] Y. Qiu *et al.*, “The novel adaptive graph neural network-based coke quality prediction for coal samples with missing properties in sustainable smart cokemaking applications,” *Fuel*, vol. 397, Oct. 2025, doi: 10.1016/j.fuel.2025.135377.

[36] L. North, K. Blackmore, K. Nesbitt, and M. R. Mahoney, “Models of coke quality prediction and the relationships to input variables: A review,” May 01, 2018, *Elsevier Ltd*. doi: 10.1016/j.fuel.2018.01.062.

[37] P. Zhao *et al.*, “Surrogate model-integrated deep reinforcement learning for temperature and pyrolysis behavior control in the coal coking procs,” *Fuel*, vol. 405, Feb 2026, doi: 10.1016/j.fuel