UDC 378
DOI: 10.15587/1729-4061.2025.320573

REAL-TIME PREDICTION OF HIGHER HEATING VALUE OF COAL IN COAL-FIRED POWER PLANTS USING OPERATING PARAMETERS

Enrico Gultom
Master's Candidate in Mechanical Engineering*
Dimas Angga Fakhri Muzhoffar
Corresponding author
Doctoral*
E-mail: dimas.anggafm@ui.ac.id
Muhammad Arif Budiyanto
Doctoral*
Achmad Riadi
Doctoral*
Andy Rivai
Master's Candidate in Chemical Engineering
Department of Chemical Engineering
Institut Teknologi Bandung
Ganesa str., 10, Coblong, Bandung,
West Java, Indonesia, 40132
*Department of Mechanical Engineering
Universitas Indonesia
Margonda Raya str., Pondok Cina, Kecamatan Beji,
Depok, West Java, Indonesia, 16424

This study introduces a novel approach to estimate the higher heating value of coal using real-time operational data from coal-fired power plants, addressing a significant gap in conventional methodologies. Traditionally, coal quality assessments involve extensive laboratory testing, which is impractical for real-time applications. This research develops a practical alternative by leveraging operational parameters such as main steam pressure, temperature, load, condensate flow, and coal flow as indicators of coal's calorific value.

The model developed in this study bypasses the time-consuming processes associated with traditional methods, enabling real-time estimation of coal's higher heating value. Empirical validation shows the model's high predictive accuracy, evidenced by an R^2 value of 0.9666, indicating that it accounts for approximately 96.66 % of the variance in higher heating value. These results are supported by low mean square error and root mean square error values, underscoring superior performance compared to conventional methods.

The effective use of operational data not only addresses the challenge of real-time higher heating value estimation but also optimizes the combustion process and enhances power plant efficiency. The practical application of these findings is pivotal for real-time coal quality control and plant performance management, providing a crucial tool for optimizing energy management.

In conclusion, this research successfully develops and validates a data-driven approach for the real-time prediction of coal's calorific value. This approach holds potential for widespread application, thereby improving energy management and operational efficiency in an industry that remains a major global energy provider

Keywords: predictive modeling, coal-fired power plant, higher heating value, real-time predictions

Received 04.12.2024
Received in revised form 16.01.2025
Accepted 12.02.2025
Published 28.02.2025

How to Cite: *Gultom, E., Muzhoffar, D. A. F., Budiyanto, M. A., Riadi, A., Rivai, A. (2025). Real-time prediction of higher heating value of coal in coal-fired power plants using operating parameters. Eastern-European Journal of Enterprise Technologies, 1 (8 (133)), 68–78. https://doi.org/10.15587/1729-4061.2025.320573*

1. Introduction

The historical importance of coal in facilitating the initial phases of industrialization is thoroughly recorded, establishing it as a fundamental element of contemporary industrial progress. Despite progress in renewable energy sources, coal continues to be a vital worldwide energy source, primarily because of its significant calorific value [1]. The gross calorific value (GCV), or higher heating value (HHV), of coal is the total heat energy released during the complete combustion of a specific quantity of coal. The value indicates the maximum thermal energy obtainable from coal combustion [2]. The HHV can be determined experimentally using an adiabatic oxygen bomb calorimeter, which measures the change in enthalpy between reactants and products before and after combustion. This calorimeter ensures that no heat is lost to the surroundings, accurately assessing the fuel's energy content [3]. Accurate determination of coal's calorific value is essential for several reasons: classification, precise evaluation of energy potential, identification of productive applications, and accurate valuation in the commodity market. The heating value of coal is a critical factor in the design and efficient operation of coal-fired equipment [4]. Recent changes in energy policies and the global emphasis on sustainability prompt inquiries regarding the ongoing significance of coal as an energy source. Nonetheless, the imperative for effective and eco-friendly exploitation of current coal deposits persists as a significant issue, especially in developing nations where alternative energy frameworks are not yet completely developed. This highlights the necessity of improving our comprehension of coal's calorific characteristics using sophisticated techniques. Conventional experimental methods for determining the HHV are costly and labor-intensive [5]. Therefore, the advancement and enhancement of predictive models by machine learning present interesting alternatives. These models aim to lower operational costs and improve the accuracy of calorific value evaluations, thereby improving coal utilization in energy production while reducing its environmental impact [6]. Therefore, predictive models for determining the calorific value of coal are highly relevant.

2. Literature review and problem statement

This study [6] reported research findings on the assessment of the HHV of coal using machine learning methodologies, specifically, random forest (RF) and artificial neural network (ANN). This research emphasizes the use of hyperparameter tuning to enhance predictive models in the context of constrained and heterogeneous data sources. The results indicated that both RF and ANN could accurately predict HHV using few data, such as carbon content, attaining R^2 values of 0.968. This methodology was corroborated across diverse coal ranks and characteristics from many nations, using both proximate and analysis outcomes. Nonetheless, lingering concerns remain regarding the generalization capabilities of these models, especially in operational contexts. The principal problem is the models' reliance on static, pre-existing datasets, which may not accurately represent real-time variations in coal qualities. This constraint is critical in operating settings where swift and dynamic modifications are essential for efficient energy management.

The paper [7] reports findings on the GCV of coal, which is an essential metric for evaluating coal quality. Diverse regression analysis techniques, including an innovative cubist regression model, have been used to forecast GCV. The model was refined by feature selection employing correlation analysis and recursive feature reduction, pinpointing the three best variable sets for regression models: proximal analysis variables, elemental analysis variables, and comprehensive index variables. Comparative analyses using several regression models indicate that the cubist model exhibits enhanced accuracy and efficiency. Nonetheless, there remain unanswered concerns regarding the practical implementation of these models in real-time operational environments. The main obstacles arise from the substantial data requirements and computational demands associated with feature selection and model training. These constraints may render the use of complex regression models difficult in operational settings that require swift and real-time forecasting. The reliance on extensive chemical tests, which may not be practical in all operating settings, underscores a notable deficiency in existing prediction approaches. One method to address these challenges may involve the creation of predictive models that leverage real-time operational data, which are readily accessible during power plant operations. This methodology has been investigated in several situations; nonetheless, its use in forecasting coal's calorific value is still constrained.

The paper [8] reports the findings of a study on the HHV of biochar, which is an essential characteristic for evaluating the quality of biochar as biofuel. This research established two empirical correlations derived from proximate and ultimate studies to forecast HHV by employing various linear and nonlinear regression techniques. The correlations exhibited remarkable precision, characterized by a low average absolute error and bias, and were corroborated with existing data, establishing their universality in predicting biochar HHV. Nonetheless, persistent concerns about the broad applicability of these models to alternative biofuels, including coal, remain unresolved. The principal problem lies in the unique characteristics of biochar relative to other fuels, potentially constraining the direct applicability of the established correlations to coal. This is exacerbated by the dependence on comprehensive laboratory tests, which may be impractical in operational environments that require swift and real-time predictions.

The existing literature [9] extensively explores the use of various artificial intelligence methods to estimate the HHV of biomass fuels, as demonstrated in recent studies employing techniques such as multilayer perceptron artificial neural networks (MLP-ANN), least-squares support vector machines (LSSVM), and genetic algorithms (GA-RBF). These approaches, used on a varied dataset of 535 biomass samples, have demonstrated encouraging outcomes in forecasting HHV with a high level of precision. The creation of models such as MNR and GA-RBF, as confirmed by considerable empirical data, demonstrates the promise of sophisticated modeling techniques in the energy sector. Nonetheless, persistent unresolved challenges regarding the direct implementation of these AI-based models for coal HHV prediction in operational power plant environments remain. The principal problem resides in the static characteristics of the data used in these models, which may fail to reflect the real-time fluctuations in coal qualities that influence the HHV. This constraint is critical in operating settings when rapid modifications to combustion processes are necessary to enhance energy output and sustain efficiency. Moreover, the dependence on pre-collected, laboratory-derived data for model training and validation results in delays and potential mistakes stemming from the time-sensitive nature of coal quality fluctuations.

The study [10] reports research findings on the HHV of coal by employing multilinear regression (MLR) techniques based on proximate and ultimate analysis data from a varied collection of coal samples, including hard coals, lignites, and anthracites from multiple nations. This study clearly illustrates the substantial importance of elemental analysis compared to proximate analysis characteristics in forecasting HHV, with carbon content emerging as a particularly influential predictor. This was validated using extensive statistical analysis, including Analysis of Variance (ANOVA), which demonstrated the reliability of the created prediction models. Nonetheless, unresolved challenges remain regarding the scalability and real-time implementation of these MLR models in operational power plant environments. The principal problem resides in the static characteristic of the MLR methodology, which predominantly depends on historical data sets that may not adequately represent the dynamic alterations in coal characteristics influencing the HHV in real-time operational contexts. This constraint is especially significant in operating settings where rapid modifications to combustion processes are necessary to enhance energy production and sustain efficiency. The dependence on substantial dataset collection and preprocessing, which is often time-consuming and expensive, hinders the practical implementation of these models in real-world contexts.

The research [11] demonstrates notable progress in predicting the GCV of coal using decision tree-based algorithms and machine learning methods, including Extra Trees, Bagging, Decision Tree, and Adaptive Boosting. This study examines the efficacy of established empirical models and compares them with newly developed algorithms, emphasizing the application of a comprehensive grid search technique for model optimization. The application of statistical indices such as explained variance, mean absolute error, and coefficient of determination highlights the efficacy of these models in forecasting GCV based on proximate and ultimate analysis data obtained from comprehensive databases like the U.S. Geological Survey Coal Quality database. Notwithstanding these developments, persistent challenges remain, especially regarding the implementation of these models in real-time operational contexts within power plants. The principal problem lies in the models' reliance on static, pre-compiled datasets, which may not adequately represent the dynamic fluctuations in coal characteristics that affect the GCV in real-time operating contexts. This constraint is critical in operating settings where swift and dynamic modifications are essential for successful energy management and sustaining efficiency.

The paper [12] proposes an innovative method for predicting the calorific value of coal gangue by integrating support vector regression (SVR) with a hybrid kernel function and evolutionary algorithms. This approach leverages the nonlinear correlations between coal gangue properties and its calorific value, thereby improving the precision and efficacy of forecasts. The use of hybrid kernel functions, integrating linear and Gaussian kernels, has exhibited enhanced forecasting efficacy relative to models employing a singular kernel function. The incorporation of genetic algorithms to optimize the important parameters of SVR represents a notable advancement in predictive modeling in the energy sector. Nevertheless, unresolved challenges remain with the direct implementation of these advanced statistical models in operational contexts at power plants. The principal problem lies in the model's reliance on historical data sets, which may not adequately reflect the dynamic fluctuations in coal characteristics that affect the calorific value in real-time situations. This constraint is critical in operational settings where rapid modifications to combustion processes are essential for optimizing energy production and sustaining efficiency.

The study [13] investigates GCV prediction of coal using proximate analysis data, employing MLR and ANN models, including multilayer perceptron (MLP), general regression neural network (GRNN), and radial basis function neural network (RBFNN). This research, notable for using the largest dataset to date with 6520 coal samples, introduces the use of GRNN and RBFNN in GCV prediction based on proximate analysis. The findings indicate that moisture and ash content are key discriminative predictors of GCV, with RBFNN models showing the highest accuracy. Performance rankings of the models were RBFNN, GRNN, MLP, and MLR. Comparative analysis with prior studies using similar datasets shows this study's RBFNN model outperforms existing SVR and RF models in both accuracy and precision, evidenced by superior R^2 and lower MAPE values. This research advances GCV predictive modeling and highlights challenges in generalizing models across different coal types and operational settings due to varying coal properties and the reliance on extensive laboratory testing for model validation.

The study [14] investigates the forecasting of coal's GCV utilizing data from 3,344 samples in the U.S. geological survey coal quality database. It utilizes four machine learning models – random forest regression (RFR), support vector machine (SVM), gradient boosting regression tree (GBRT), and extreme gradient boosting (XGB) – surpassing conventional empirical techniques, with the XGB model attaining an R^2 of 0.9908. To improve model transparency, explainable artificial intelligence (XAI) approaches such as LIME, ALE, and ICE are incorporated, clarifying the influence of significant factors on GCV. This integration enhances forecast accuracy and improves interpretability, which is essential for industrial applications. The study recognizes limits with regional data specificity and proposes future expansion to various coal kinds and hybrid models to improve precision and interpretability. This research enhances energy resource management by offering a data-driven framework for informed decision-making in coal usage and policy formulation.

This research presents an innovative approach that utilizes real-time operational data from power plants to forecast the calorific value of coal, employing key operational parameters such as main steam pressure, temperature, load, condensate flow, and coal flow as indicators of coal's calorific value. Although similar approaches have been explored in many contexts, their specific application to predict coal's HHV is still constrained. This gap highlights the suggested method's potential as a practical and accelerated alternative to conventional, time-intensive laboratory investigations, with the objective of improving real-time monitoring and control of coal quality. This research is essential due to a thorough literature assessment that highlights ongoing difficulties in the precise and prompt prediction of coal's calorific value. Contemporary techniques predominantly depend on comprehensive laboratory analyses and static data models that inadequately address the dynamic fluctuations in coal characteristics observed in operational settings. This constraint is vital in power plant operations, where fast modifications to coal intake according to its energy content are necessary for enhancing combustion processes and overall plant efficiency. The assessment emphasizes a considerable dependence on historical data sets, which may not capture real-time fluctuations, as well as the laborious data collecting and preprocessing necessitated by conventional models. These variables lead to delays and possible mistakes in coal quality evaluation, directly affecting energy management and environmental compliance. The suggested study seeks to create an innovative predictive model by tackling the highlighted issues, utilizing sophisticated methodologies such ANN and response surface methodology (RSM). This model aims to offer a reliable instrument for the real-time prediction of coal's calorific value using publicly accessible operational data, thus avoiding the constraints of traditional approaches. All this allows us to assert that it is expedient to conduct a study on the real-time prediction of coal's calorific value using operational data, as it promises to significantly enhance the accuracy, efficiency, and responsiveness of coal quality assessments in power plant operations. This approach not only fills a critical research gap but also aligns with the ongoing need for improved energy management and environmental sustainability in the energy production sector.

3. The aim and objectives of the study

The aim of this research is to develop a novel approach for forecasting the calorific value of coal using real-time operational data from power plants. This innovative method suggests using essential operating parameters, specifically, the main steam pressure, temperature, load, condensate flow, and coal flow as possible indications of coal's energy content. The proposed approach is a more pragmatic and rapid alternative to conventional laboratory investigations for evaluating coal quality.

To achieve this aim, the following objectives are accomplished:

– to evaluate the relationship between operational factors and coal's calorific value using statistical analysis and modeling to confirm their effectiveness as predictors and ensure scientific validity;

– to develop a predictive model using ANN and RSM to estimate coal's calorific value from real-time power plant data, including steam pressure, temperature, load, condensate, and coal flow;

– to evaluate the practicality and effectiveness of the proposed ANN and RSM model in real-time settings, this study examines its capability to deliver accurate and timely forecasts of coal quality, thereby facilitating enhanced energy management in coal-fired power plants.

4. Materials and methods

4. 1. Object and hypothesis of the study

The object of the study is the heating value of coal. It was necessary to develop and validate a predictive model that employs real-time operational data from power plants to assess the calorific value of coal. This approach seeks to utilize operational characteristics, including main steam pressure, temperature, load, condensate flow, and coal flow, as indications of coal's energy content. The emphasis is on improving the precision and promptness of coal quality evaluations in operating environments, thus enabling more efficient and eco-friendly power plant operations.

The primary hypothesis of this work posits that the calorific value of coal can be precisely forecasted utilizing real-time operational data available in power plant systems. It asserts that particular operational parameters are strongly associated with the HHV of coal, and that models created with these parameters can accurately forecast coal's calorific value in real-time, thereby eliminating the necessity for conventional, labor-intensive laboratory techniques.

This study is based on numerous fundamental assumptions that are crucial for the validity of its results. It presupposes consistent operational circumstances. It also presumes negligible impact from unmeasured confounders, indicating that all relevant determinants of coal's calorific value are either incorporated in the model or their impacts are small. This is essential for guaranteeing the model's outputs are impartial and precise. The reliability of historical operational data utilized for model training and validation is presumed, which is essential to the integrity of the model building process, as erroneous or imprecise data may result in underperforming predictive models.

To manage the complexity of the modeling process and focus on the core objectives, the study adopts several simplifications. It concentrates on a select set of operational parameters that are most readily available and likely to have significant predictive power, based on preliminary analyses. This focus on key variables helps streamline the analytical process. The study also deliberately excludes external economic, environmental, or regulatory factors that might affect power plant operations but are outside the scope of operational data. Furthermore, it employs well-established statistical and machine learning techniques, such as ANN and RSM, to ensure the research remains focused on evaluating the efficacy of using operational data for prediction rather than on developing new predictive methodologies.

These strategic choices ensure that the study remains manageable and focused while adequately addressing the research hypothesis and objectives. They provide a structured approach to exploring the predictive capabilities of operational data in estimating the calorific value of coal, which is crucial for enhancing the operational efficiency and environmental sustainability of power plants.

4. 2. Data collection

The architecture of a coal-fired power station, as depicted in Fig. 1, involves a sophisticated sequence of energy generation procedures.

In coal-fired power plant operations, the combustion process involves igniting coal with a designated calorific value. This method transforms water into steam at elevated temperatures and pressures, subsequently producing electricity [15]. The principal operating factors that define this process include the main steam flow, main steam temperature, load, coal flow, and condensate flow. This process conventionally employs a bottom-up methodology, where the calorific value of the coal serves as an input, and the resulting steam and electricity outputs are the outputs. In this study, the technique was reversed to a top-down paradigm. The operational characteristics are used to forecast the calorific value of coal combusted in real time. The top-down approach uses operational data from a power plant to assess the calorific value of coal. This approach employs characteristics like main steam flow, main steam temperature, load, coal flow, and condensate flow to forecast the calorific value of coal. By analyzing these factors, the model is able to infer the amount of energy produced and, consequently, ascertain the calorific value of the combusted coal.

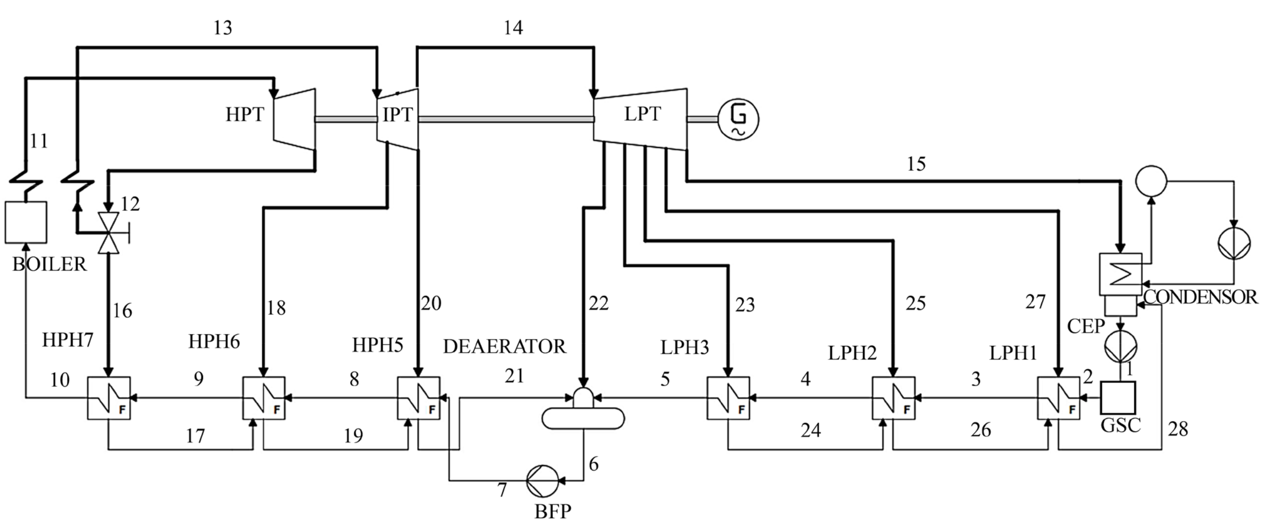


Fig. 1. Schematic of coal-fired power station

The dataset used in this study was derived from historical performance test data collected over four years. The main aim of performance testing is to assess the overall efficiency, output, and adherence to contractual and technical criteria for coal-fired power plants. This entails quantifying several critical parameters to evaluate the plant's performance under standard operating conditions. ASME Performance Test Code 6 (PTC 6) is the global standard for steam turbine acceptance testing, offering a uniform methodology for assessing the performance of existing, retrofitted, and new steam turbines [16].

Data is gathered and evaluated during the performance evaluation to ascertain the adjusted load, heat rate, and other critical parameters like main steam pressure, load, main steam flow, total coal flow, and condensate flow. These indicators are essential for evaluating the plant's efficacy and productivity [17]. Furthermore, coal samples are collected before entering the boiler and undergo proximate and ultimate testing to verify compliance with the established quality criteria. This study uses historical data from performance examinations conducted over four years. This dataset comprises input characteristics, including the main steam pressure, load, main steam flow, coal flow, and condensate flow. The study seeks to enhance the plant's performance by examining past data points and identifying any discrepancies or concerns that could impact its efficiency and output.

4. 3. Response surface methodology

The RSM is an approach that combines mathematical and statistical tools to develop, enhance, and optimize processes. The RSM has demonstrated significant advantages in situations where multiple input variables have varying effects on the outputs of a particular process [18]. The data set was analyzed using Design-Expert software version 13, which facilitates user regression analysis to develop equations that determine the polynomial degree of the output variables. Experimental Design (Design of Experiment, DoE) was employed to determine the causal relationships among various experimental parameters, ranging from dependencies to interactions between these parameters [19]. The RSM seeks to model and understand the relationships between input variables (e.g., $X_1, X_2, ..., X_k$) and response variable (Y). The relationships are expressed as (1):

$$Y = f\left(X_1, X_2, ..., X_k\right) + \in. \tag{1}$$

In the model, $F(x)$ denotes the function delineating the link between the input variables and the output, while $\in$ indicates the residual error. For analytical simplification, these interactions are generally modeled using polynomial approximations, either linear (first-order) or quadratic (second-order) [20]. The First-Order polynomial model is used to examine the principal impacts of input variables on the response variable, excluding interaction effects and quadratic correlations. This method is especially beneficial in the preliminary phases of research, where the aim is to comprehend the direct impact of each independent variable on the dependent variable, thereby establishing a fundamental understanding prior to investigating more intricate interactions, as shown in the formula (2) [21]:

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \in. \tag{2}$$

In regression analysis, Y represents the response variable predicted using the input variables X_i. The model incorporates an intercept β_0, denoting the baseline value of Y when all X_i are zero, along with regression coefficients β_i for each predictor, signifying the anticipated change in Y for each unit change in X_i. The symbol ϵ denotes the residual error, which encapsulates the variance in Y that remains unexplained by the model.

4. 4. Artificial neural network

ANN streamline intricate processes related to identifying correlations among experimental data [22]. In this study, Bayesian regularization was used to train the ANN. Bayesian regularization improves a model's generalizability by regulating its complexity. The proposed approach combines regularization principles with Bayesian theory to determine the appropriate probabilistic distribution of the ANN weights depending on the existing data. Bayesian regularization mitigates overfitting by including a probability distribution over the weights [23]. The fundamental ANN has an input layer, one or more hidden layers, and an output layer. Each layer is linked to a node or neuron that computes the weighted total of the inputs and transmits it through a transfer function to serve as inputs for the subsequent layer [24]. The output function of an ANN can be mathematically expressed in formula (3):

$$Y = f\left(W_x + b\right), \tag{3}$$

where x represents the input, W represents the weights, b indicates the bias, and f is the activation function that introduces nonlinearity, such as the sigmoid function, ReLU (Rectified Linear Unit), or tanh [25]. The weights and biases are modified to minimize the combination of squared errors and weight size, thereby preventing overfitting and enhancing generalizability [23].

4. 5. Performance assessment of the proposed model

The constructed models will be assessed according to their performance metrics to identify the most appropriate machine-learning model for monitoring exhaust emissions and boiler efficiency. The statistical analysis metrics, as specified in formulas (4)–(6), comprise the correlation coefficient (R), mean square error (MSE), and root mean square error ($RMSE$) [26]. These indices are frequently used to evaluate the performance of models implementing diverse machine learning algorithms, including ANN:

$$R^2 = 1 - \frac{\sum_{j-i}^{k}\left(Y_j^{actual} - Y_j^{pred}\right)}{\sum_{j-i}^{k}\left(Y_j^{exp} - Y_j^{ave}\right)}, \tag{4}$$

$$MSE = \frac{1}{K}\sum_{j=1}^{k}\left(Y_j^{actual} - Y_j^{pred}\right)^2, \tag{5}$$

$$RMSE = \left(\frac{1}{K}\sum_{j=1}^{k}\left(Y_j^{actual} - Y_j^{pred}\right)^2\right)^{\frac{1}{2}}, \tag{6}$$

Y^{pred} and Y^{ave} denote the predicted and average values of Y, respectively, and K indicates the number of datasets.

4. 6. Model development

The model was created using a comprehensive data collection approach. The input parameters were derived from the operational data collected between 2020 and 2024, particularly during performance evaluations. The primary input parameters are the main steam pressure, load, main steam flow, coal flow, and condensate flow, and the output is the HHV of the fuel. These extensive datasets are essential for analyzing the operational dynamics and efficiency of power plants.

The subsequent stage involves determining the ideal number of neurons in the neural network model [27]. This was achieved using a systematic trial and error methodology in which the number of neurons was adjusted from 1 to 25. Each configuration was assessed according to performance, particularly by quantifying the RMSE. The model configuration with the lowest RMSE value was selected as the optimal parameter. This iterative procedure ensures that the model is refined for precise HHV prediction based on the input parameters.

Because of the substantial disparities in magnitude between the input and output variables, data preparation is required. All input and output data were normalized using the min-max scaling method to resolve this. This method standardizes the data to a consistent range of 0 to 1, guaranteeing that all features contribute equally to the model's predictions and preventing any single feature from overshadowing others due to variations in magnitude [28]. The min-max scaling formula is implemented as the formula (7):

$$X = \frac{X - X_{\min}}{X_{\max} - X_{\min}}. \quad (7)$$

Here, X represents the initial feature value, $X_{\min}$ denotes the least feature value, and $X_{\max}$ represents the maximum feature value in the dataset [29]. This normalization method is especially beneficial when the estimated upper and lower limits of the dataset are identified and the data distribution is roughly uniform throughout the range.

5. Results of developing response surface methodology and artificial neural network models for real-time coal calorific value estimation

5. 1. Evaluation of the relationship between operational factors and higher heating value

The initial phase of modeling entails determining the correlation between the dependent and independent factors essential for predicting the HHV of coal using real-time operational data [30]. Fig. 2 illustrates the correlation between the excellent heating value and the input parameters for all datasets.

The collection of three-dimensional charts demonstrates the correlations between different operational characteristics and the HHV of fuel in a power plant context, each emphasizing the effect of variations in two particular operational parameters on HHV. Fig. 2, *a* analyzes the relationship between main steam flow (kg/h) and load (MW), distinctly categorizing into zones of high (red), medium (green), and low (blue) HHV. This clear distinction indicates substantial threshold effects where HHV varies markedly at particular steam flow and load levels, signifying optimal operational parameters for optimizing HHV. Fig. 2, *b* illustrates the correlation between condensate flow (kg/h) and main steam flow (kg/h), revealing a consistent pattern of distinct zones that signify substantial interactions between these variables influencing HHV. Fig. 2, *c* shows the connection between coal flow (ton/hour) and load (MW) is predominantly linear, characterized by an extensive region of elevated HHV (red) and a limited region of reduced HHV (blue), indicating a positive association whereby increases in coal flow and load result in increased HHV. Fig. 2, *d*, illustrating the correlation between main steam pressure (kg/cm^2) and load (MW), exhibits a complicated gradient from elevated to diminished HHV, signifying a persistent and substantial link wherein HHV escalates with increased steam pressure and load.

These graphs illustrate that particular combinations of operational factors can improve the HHV, which is essential for increasing the power plant's efficiency, improving fuel efficacy, and minimizing operational expenses. Each plot enhances the overall understanding of the interactions between various operational parameters. With this enhanced understanding, we can proceed with the development of the model.

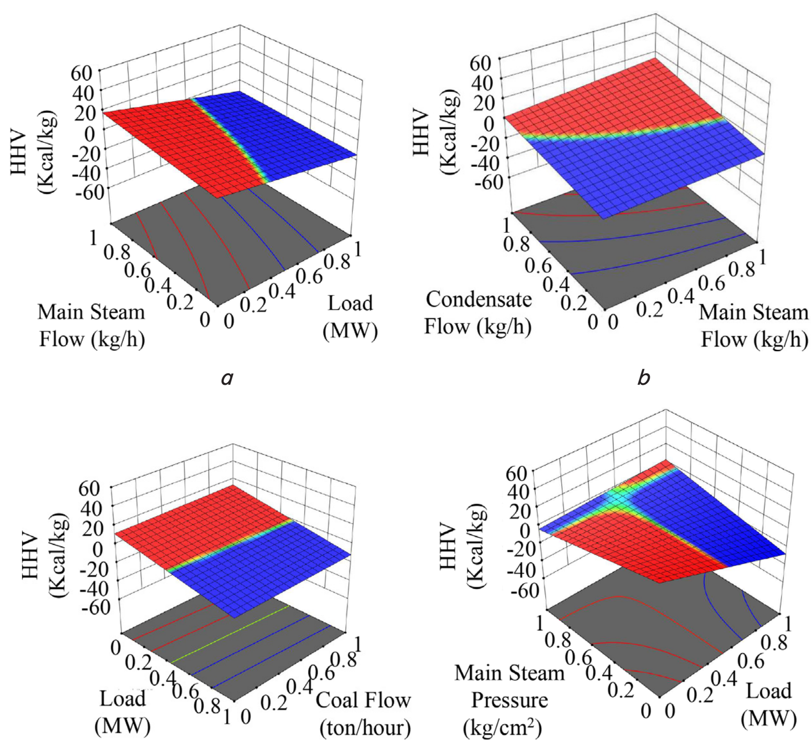


Fig. 2. 3D surface contour plots: *a* — higher heating value interaction with main steam flow and load; *b* — higher heating value dependency on main steam and condensate flows; *c* — higher heating value correlation with coal flow and electrical load; *d* — higher heating value effects from main steam pressure and load

5. 2. Development of the predictive model

After analyzing the correlation between operational factors and calorific value, the RSM and ANN models were developed. The established RSM model, as represented by formula (8), quantitatively predicts the HHV based on multiple operational parameters, which are critical for optimizing the performance and efficiency of coal-fired power plants. In the model, the variables A, B, C, D, and E represent main steam pressure, load, main steam flow, coal flow, and condensate flow, respectively. Each of these variables is assigned a coefficient in the equation, reflecting its relative influence on the HHV:

$$HHV = 0.304953\times\left(0.119909\times A\right)+ \\ +\left(0.421305\times B\right)+\left(-0.0155923\times C\right)\times \\ \times\left(-0.881042\times D\right)+\left(0.226002\times E\right). \qquad (8)$$

The coefficients indicate the sensitivity of the HHV to changes in each operational parameter. For instance, the positive coefficients of A, B, and E suggest that increases in main steam pressure, load, and condensate flow respectively contribute to a higher HHV, indicating more efficient energy production. Conversely, the negative coefficients associated with C (main steam flow) and D (coal flow) imply that increases in these parameters may lead to a reduction in HHV, possibly due to inefficiencies or dilution effects in the combustion process.

The development of an efficient neural network model for predicting coal calorific value begins by determining the optimal architecture. This subsection details the methodological approach used to determine the appropriate number of neurons in a single hidden layer of the neural network. A trial-and-error strategy was used to determine the appropriate number of neurons in the architecture of a neural network model designed to forecast the calorific value of coal based on operational parameters in a power plant. This method involved testing several neuron configurations (ranging from 1 to 25) within a single hidden neural network layer, as shown in Fig. 3. The aim of this study was to determine the configuration that reduced the RMSE, which is a conventional statistic for assessing the precision of regression models.

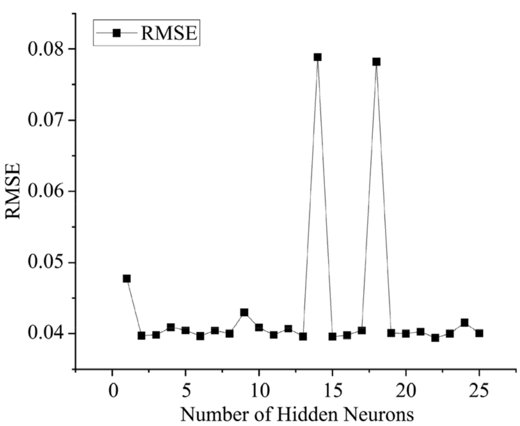


Fig. 3. Root mean square error for different numbers of hidden neurons

Based on RMSE results from studies with 1–25 neurons, Fig. 4 shows the ideal ANN architecture. It presents the layout of the suggested ANN, which is meant to predict the HHV by utilizing a number of input factors related to the operational parameters of a power plant. The input layer, hidden layer, and output layer are the three layers that make up the network. With an RMSE of 0.0394, the testing results demonstrated that the architecture with 22 neurons in the hidden layer generated the best results. This indicates that the neural network's predictions of the coal's calorific value are highly accurate, suggesting that the 22-neuron model is skillfully adjusted to the complexities of the dataset and the operational dynamics of the power plant.

Fig. 5 illustrates the training process of the ANN used for regression analysis, highlighting its performance across training, testing, and combined datasets. Fig. 5, *a* illustrates the ANN's training data regression analysis. Fig. 5, *b* displays the testing data regression analysis. Fig. 5, *c* combines both training and testing data, reflecting a consistent overall performance. Additionally, Fig. 5, *d* details the model's error distribution, showing most errors clustering near zero, which indicates minor deviations between the predicted outputs and actual targets.

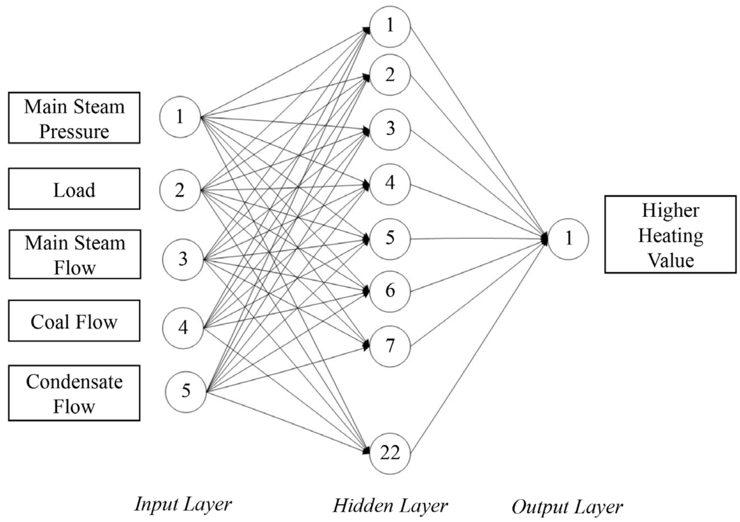


Fig. 4. Artificial neural network architecture

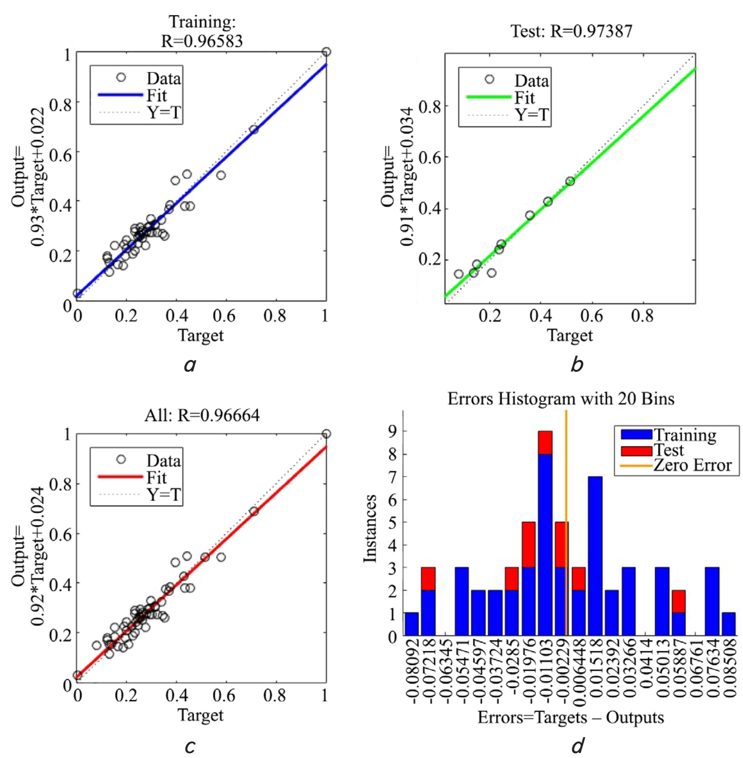


Fig. 5. Artificial neural network data regression results: *a* – training data regression analysis; *b* – test data regression analysis; *c* – combined data regression analysis; *d* – model error distribution

The data presented offer a thorough assessment of a predictive model's efficacy via regression analysis, emphasizing the relationship between anticipated outputs and actual target values during several validation stages. The error histogram comprehensively depicts the prediction errors and is divided into 20 bins to demonstrate the frequency of different error magnitudes throughout the training and testing phases [31]. The clustering of errors near the zero-error line, especially during the testing phase, highlights the model's accuracy, as most predictions diverge slightly from the actual values.

5. 3. Model assessment

After developing the models, the ANN and RSM models were evaluated. Fig. 6 is an essential element in the model evaluation process of RSM, offering insights into the validity and effectiveness of the predictive model employed for calculating the HHV. This picture comprises two sections, each illustrating a distinct facet of the residual analysis.

The model developed by RSM was further evaluated for data sufficiency and anomalies using normal plots for residuals and outliers. In this scenario, a satisfactory model is expected to deviate from any discernible pattern or sequence, with the data points near a linear trajectory.

The graph shown in Fig. 7 thoroughly assesses the two predictive models – the ANN model and the RSM model – in calculating the HHV of coal. The graph illustrates the projected HHV values from both models compared to the actual observed HHV values across a series of data points, numbered from 0 to 57. The predictions from the ANN model are denoted by squares, predictions from the RSM model are denoted by circles, and the actual HHV measurements are denoted by triangles.

Fig. 8 compares two statistical models, the ANN and the RSM, based on two performance metrics: R-squared and RMSE.

In Fig. 8, *a*, R-squared values demonstrate that the ANN model more accurately predicts the variance in the dependent variable than the RSM model, as evidenced by superior R-squared values. Fig. 8, *b* illustrates the RMSE for each model. RMSE quantifies the average size of errors between the anticipated and observed values, with smaller values signifying superior prediction ability. This subsection concludes with a detailed comparison of the ANN and RSM models, providing a clear indication of the enhanced capability of the ANN model in accurately and efficiently predicting the calorific value of coal.

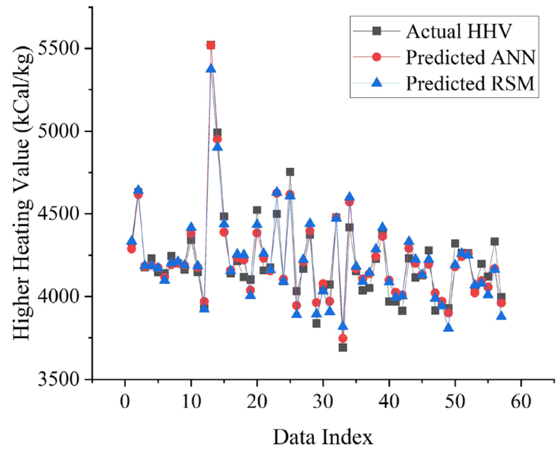
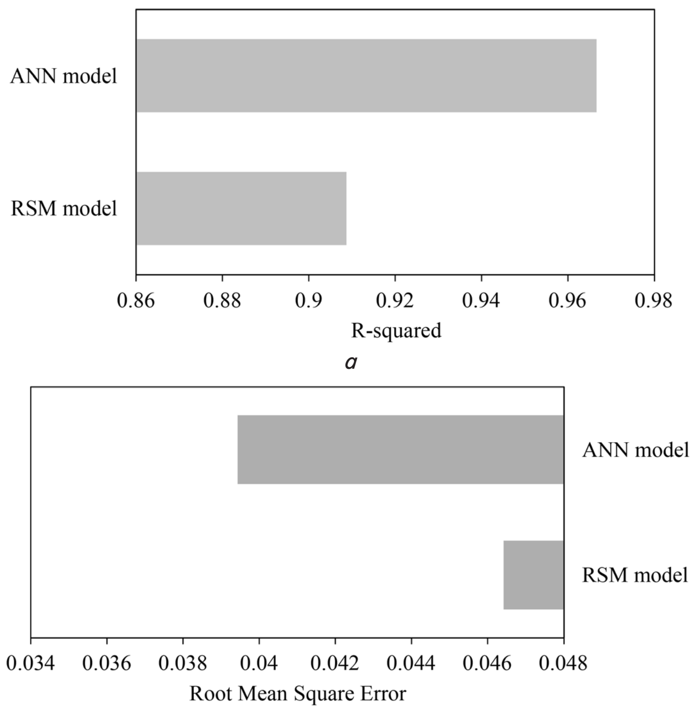


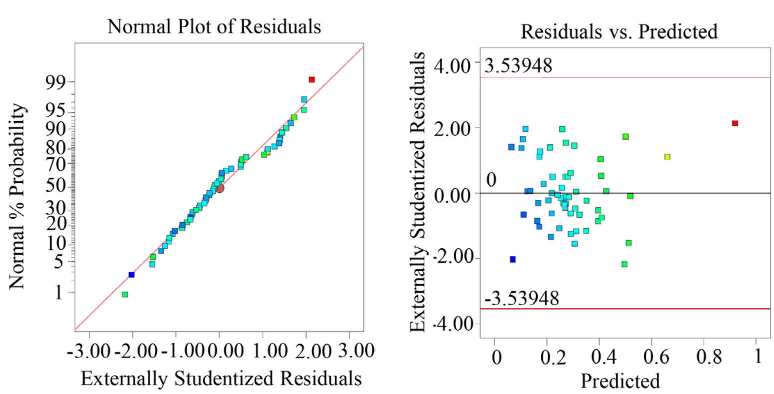
Fig. 7. Comparison between predicted and actual higher heating value



a

b

Fig. 8. Comparison of model performance: *a* – R-squared; *b* – root mean square error for artificial neural network and response surface methodology



a

b

Fig. 6. RSM model evaluation: *a* – normal plot of residuals; *b* – outlier higher heating value

6. Discussion of the comparative efficacy of artificial neural networks and response surface methodology models for real-time coal calorific value estimation

The scatter graphs illustrate the model's outputs for the targets for training, testing, and overall data, each exhibiting a strong correlation coefficient: 0.96583 for training, 0.97387 for testing, and 0.96664 for the combined data, as shown in Fig. 5, *a–c*. These coefficients demonstrate robust linear correlation, indicating that the model successfully assimilated the training data and efficiently generalized to novel, unseen data. The congruence of the data points with the optimal line ($Y=T$) in these plots underscores the model's predictive accuracy. The error distribution reflects the model's robustness, which validates its ability to generate dependable and accurate predictions across diverse datasets. These visualizations confirm the model's excellent prediction accuracy and stability, underscoring its appropriateness for analogous predictive applications in several domains and demonstrating its potential for extensive applicability and efficacy in real-world contexts, as shown in Fig. 5, *d*. The graph in Fig. 7 illustrates the comparative predicted accuracy of the ANN and RSM models for the HHV of coal. The ANN model exhibits a closer correlation with the real HHV values, a conclusion supported by the statistical metrics in the accompanying table, which includes R^2, MSE, and RMSE for each model. The ANN model is superior with an (R^2) value of 0.9666, signifying that approximately 96.66 % of the variance in HHV is accounted for by this model. The elevated (R^2) value is supported by markedly reduced MSE (0.001555) and RMSE (0.039435) values, indicating that the ANN model forecasts the HHV of coal with considerable accuracy and negligible error fluctuation. Conversely, the RSM model demonstrated a diminished (R^2) value of 0.9087, coupled with elevated MSE (0.002155) and RMSE (0.046418) values, indicating less predictive accuracy and increased error variability, as depicted in Fig. 8.

The improved performance of the ANN model can be ascribed to its capacity to accurately represent intricate nonlinear correlations and interactions among variables, which are typical in assessing coal's HHV. Although the RSM demonstrated adequate predictive potential, it failed to encapsulate the inherent complexities of the dataset. This comparative research highlights the benefits of using advanced modeling techniques, such as ANN, for predicting HHV, stressing the necessity of choosing suitable modeling methods according to the data characteristics and specific predictive goals. The literature presents a thorough examination of diverse methodologies for estimating HHV based on the chemical composition of coal, with distinct models yielding different levels of predictive accuracy and methodologies [7–11, 32]. The papers [13, 14] examine the prediction of the HHV of coal based on proximate analysis, emphasizing the continued dependence on laborious laboratory assessments. This method, although based on conventional empirical techniques, requires comprehensive laboratory analysis to ascertain critical factors including moisture, volatile matter, ash content, and fixed carbon. This study enhances the current knowledge base by demonstrating the possibility of real-time operational data as a predictive instrument for coal's calorific value, thereby addressing a notable deficiency in real-time, data-driven HHV estimation.

This research has limitations. Although effective in this context, the implementation of the ANN model may lack generalizability across various coal types or operational environments without retraining or modification. The dependence on large historical data for model training may restrict the model's application in contexts where such data are unavailable. Future studies may investigate the incorporation of real-time data collecting technologies and the creation of adaptive learning models capable of adjusting their parameters in reaction to variations in coal quality or operational conditions. This methodology may encounter hurdles, including the necessity to create algorithms that can learn and rectify errors in dynamic settings. In conclusion, although this study enhances our understanding of predictive modeling for coal's HHVs using ANN and RSM, additional research is required to improve the adaptability and precision of these models, especially in diverse operational scenarios. This may ultimately facilitate more efficient and environmentally sustainable energy generation from coal-fired power plants.

7. Conclusions

1. The investigation substantiated the strong association between operational parameters and the calorific value of coal, affirming its efficacy as a reliable predictor. The series of three-dimensional surface contour plots reveal how specific combinations of operational factors such as main steam flow, load, condensate flow, coal flow, and main steam pressure interact to influence the HHV. Notably, these interactions vary, with some showing distinct threshold effects where HHV changes markedly at certain operational levels, while others display more linear relationships.

2. The proposed prediction model effectively employs artificial ANN and RSM to estimate the calorific value of coal using real-time operational data from power plants. The ANN model exhibited exceptional performance, accounting for approximately 96.66 % of the variance in the HHV of coal. This indicates a strong nonlinear correlation between the operating parameters and the calorific value, underscoring the effectiveness of using real-time data for swift and precise energy content assessment. The ANN's ability to represent intricate relationships among variables highlights its potential to improve forecast accuracy in dynamic energy management settings.

3. The practical application of the ANN model in real-time operating settings was evaluated, demonstrating its capacity to deliver timely and accurate predictions of coal quality. The comparative analysis demonstrated the ANN model's superior performance relative to the RSM model, as indicated by the higher coefficient of determination (R^2) and MSE and RMSE. This superiority is ascribed to the ANN's ability to precisely model the dataset. This research underscores the necessity of choosing suitable modeling techniques according to data attributes, promoting sophisticated methods, such as ANNs, for improved forecast precision. The RSM demonstrated sufficient potential although its shortcomings in encapsulating the intricacies of the dataset were obvious. This research advances the field by illustrating the benefits of real-time data-driven approaches for coal calorific value prediction, thereby addressing the critical gap in operational efficiency and energy management.

Conflict of interest

The authors declare that they have no conflicts of interest in relation to this research, whether financial, personal, authorship, or otherwise, that could affect the research and its results presented in this paper.

Financing

This research was supported by PT PLN Persero, a state-owned Indonesian enterprise.

Data availability

Data cannot be made available for reasons disclosed in the data availability statement.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

References

1. Guan, G. (2017). Clean coal technologies in Japan: A review. Chinese Journal of Chemical Engineering, 25 (6), 689–697. https://doi.org/10.1016/j.cjche.2016.12.008
2. Majumder, A., Jain, R., Banerjee, P., Barnwal, J. (2008). Development of a new proximate analysis based correlation to predict calorific value of coal. Fuel, 87 (13-14), 3077–3081. https://doi.org/10.1016/j.fuel.2008.04.008
3. Callejón-Ferre, A. J., Carreño-Sánchez, J., Suárez-Medina, F. J., Pérez-Alonso, J., Velázquez-Martí, B. (2014). Prediction models for higher heating value based on the structural analysis of the biomass of plant remains from the greenhouses of Almería (Spain). Fuel, 116, 377–387. https://doi.org/10.1016/j.fuel.2013.08.023
4. Akkaya, A. V. (2009). Proximate analysis based multiple regression models for higher heating value estimation of low rank coals. Fuel Processing Technology, 90 (2), 165–170. https://doi.org/10.1016/j.fuproc.2008.08.016
5. Akkaya, A. V. (2013). Predicting Coal Heating Values Using Proximate Analysis via a Neural Network Approach. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 35 (3), 253–260. https://doi.org/10.1080/15567036.2010.509090
6. Büyükkanber, K., Haykiri-Acma, H., Yaman, S. (2023). Calorific value prediction of coal and its optimization by machine learning based on limited samples in a wide range. Energy, 277, 127666. https://doi.org/10.1016/j.energy.2023.127666
7. Chen, J., He, Y., Liang, Y., Wang, W., Duan, X. (2024). Estimation of gross calorific value of coal based on the cubist regression model. Scientific Reports, 14 (1). https://doi.org/10.1038/s41598-024-74469-3
8. Qian, C., Li, Q., Zhang, Z., Wang, X., Hu, J., Cao, W. (2020). Prediction of higher heating values of biochar from proximate and ultimate analysis. Fuel, 265, 116925. https://doi.org/10.1016/j.fuel.2019.116925
9. Noushabadi, A. S., Dashti, A., Ahmadijokani, F., Hu, J., Mohammadi, A. H. (2021). Estimation of higher heating values (HHVs) of biomass fuels based on ultimate analysis using machine learning techniques and improved equation. Renewable Energy, 179, 550–562. https://doi.org/10.1016/j.renene.2021.07.003
10. Sözer, M., Haykiri-Acma, H., Yaman, S. (2021). Prediction of Calorific Value of Coal by Multilinear Regression and Analysis of Variance. Journal of Energy Resources Technology, 144 (1). https://doi.org/10.1115/1.4050880
11. Munshi, T. A., Jahan, L. N., Howladar, M. F., Hashan, M. (2024). Prediction of gross calorific value from coal analysis using decision tree-based bagging and boosting techniques. Heliyon, 10 (1), e23395. https://doi.org/10.1016/j.heliyon.2023.e23395
12. Gao, X., Jia, B., Li, G., Ma, X. (2022). Calorific Value Forecasting of Coal Gangue with Hybrid Kernel Function-Support Vector Regression and Genetic Algorithm. Energies, 15 (18), 6718. https://doi.org/10.3390/en15186718
13. Açikkar, M., Sivrikaya, O. (2018). Prediction of gross calorific value of coal based on proximate analysis using multiple linear regression and artificial neural networks. Turkish Journal of Electrical Engineering & Computer Sciences, 26 (5), 2541–2552. https://doi.org/10.3906/elk-1802-50
14. Zhu, W., Xu, N., Hower, J. C. (2025). Unveiling the predictive power of machine learning in coal gross calorific value estimation: An interpretability perspective. Energy, 318, 134781. https://doi.org/10.1016/j.energy.2025.134781
15. Dzikuć, M., Kuryło, P., Dudziak, R., Szufa, S., Dzikuć, M., Godzisz, K. (2020). Selected Aspects of Combustion Optimization of Coal in Power Plants. Energies, 13 (9), 2208. https://doi.org/10.3390/en13092208
16. Reza, A. M. (2020). ASME PTC 6 Steam Turbine. Available at: https://www.researchgate.net/publication/370071825_ASME_PTC_6_Steam_Turbine
17. Bolarinwa, M. A., Udensi, F.-G. (2024). Operational Performance Analysis of Generating Power Plant. European Journal of Energy Research, 4 (2), 44–56. https://doi.org/10.24018/ejenergy.2024.4.2.144
18. Jradi, R., Marvillet, C., Jeday, M. R. (2024). Multi-objective optimization and performance assessment of response surface methodology (RSM), artificial neural network (ANN) and adaptive neuro-fuzzy interfence system (ANFIS) for estimation of fouling in phosphoric acid/steam heat exchanger. Applied Thermal Engineering, 248, 123255. https://doi.org/10.1016/j.applthermaleng.2024.123255
19. Jankovic, A., Chaudhary, G., Goia, F. (2021). Designing the design of experiments (DOE) – An investigation on the influence of different factorial designs on the characterization of complex systems. Energy and Buildings, 250, 111298. https://doi.org/10.1016/j.enbuild.2021.111298
20. Jensen, W. A. (2017). Response Surface Methodology: Process and Product Optimization Using Designed Experiments 4th edition. Journal of Quality Technology, 49 (2), 186–188. https://doi.org/10.1080/00224065.2017.11917988
21. Draper, N., Smith, H. (1998). Applied regression analysis. John Wiley & Sons. https://doi.org/10.1002/9781118625590
22. Rizal, R., Wahyuni, F., Julian, J., Nasruddin, Yulia, F. (2024). Optimal design and modelling of sustainable bio-catalytic enzyme for wastewater treatment using response surface methodology and artificial neural network. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 46 (1), 5254–5273. https://doi.org/10.1080/15567036.2024.2332468
23. Vendittoli, V., Polini, W., Walter, M. S. J., Geißelsöder, S. (2024). Using Bayesian Regularized Artificial Neural Networks to Predict the Tensile Strength of Additively Manufactured Polylactic Acid Parts. Applied Sciences, 14 (8), 3184. https://doi.org/10.3390/app14083184
24. Suphawan, K., Kardkasem, R., Chaisee, K. (2022). A Gaussian Process Regression Model for Forecasting Stock Exchange of Thailand. Trends in Sciences, 19 (6), 3045. https://doi.org/10.48048/tis.2022.3045
25. Haykin, S. (2009). Neural Networks and Learning Machines. Pearson. Available at: https://dai.fmph.uniba.sk/courses/NN/haykin.neural-networks.3ed.2009.pdf

26. Jradi, R., Marvillet, C., Jeday, M. R. (2020). Modeling and comparative study of heat exchangers fouling in phosphoric acid concentration plant using experimental data. Heat and Mass Transfer, 56 (9), 2653–2666. https://doi.org/10.1007/s00231-020-02888-9
27. Yotov, K., Hadzhikolev, E., Hadzhikoleva, S. (2020). Determining the Number of Neurons in Artificial Neural Networks for Approximation, Trained with Algorithms Using the Jacobi Matrix. TEM Journal, 9 (4), 1320–1329. https://doi.org/10.18421/tem94-02
28. Amato, A., Di Lecce, V. (2023). Data preprocessing impact on machine learning algorithm performance. Open Computer Science, 13 (1). https://doi.org/10.1515/comp-2022-0278
29. Gasser, M., Naguib, A., Abdelhafiz, M., Elnekhaily, S., Mahmoud, O. (2023). Artificial Neural Network Model to Predict Filtrate Invasion of Nanoparticle-Based Drilling Fluids. Trends in Sciences, 20 (5), 6736. https://doi.org/10.48048/tis.2023.6736
30. Muhammad, R., Rahman, A., Jamil, M. F. (2024). A Linear Regression Modeling Analysis of the Energy, Water, and Chemical Consumption in the Operating Configuration at 740 MW Priok Combined Cycle Power Plant. Journal of Mechanical Design and Testing, 6 (1), 27. https://doi.org/10.22146/jmdt.97748
31. Padmaja, B., Manikandan, V. M. (2021). A Novel Prediction Error Histogram Shifting-based Reversible Data Hiding Scheme for Medical Image Transmission. 2021 4th International Conference on Security and Privacy (ISEA-ISAP), 1–6. https://doi.org/10.1109/isea-isap54304.2021.9688572
32. García-Nieto, P. J., García-Gonzalo, E., Paredes-Sánchez, J. P. (2024). Estimation of the coal higher heating value for energy systems relied on ultimate analysis with machine learning techniques. Fuel, 357, 130037. https://doi.org/10.1016/j.fuel.2023.130037