



Assessment of coal spontaneous combustion potential using statistical and machine learning approaches

by H. Ghasemzadeh¹, H. Madani¹, F. Sereshki², S. Afraei¹

Affiliation:

¹ Department of Mining Engineering,
Amirkabir University of Technology, Tehran,
Iran

² Faculty of Mining, Petroleum and
Geophysics, Shahrood University of
Technology, Shahrood, Iran

Correspondence to:

H. Madani

Email:

madani.hasan.aut@gmail.com

Dates:

Received: 21 Apr. 2026

Revised: 26 Jun. 2025

Accepted: 14 Aug. 2026

Published: Sep. 2026

How to cite:

Ghasemzadeh, H., Madani, H., Sereshki, F.,
Afraei, S. 2026. Assessment of Coal
Spontaneous Combustion Potential Using
Statistical and Machine Learning Approaches.
*Journal of the Southern African Institute of
Mining and Metallurgy*, vol. 126, no. 9,
pp. 541 – 558

DOI ID:

<https://doi.org/10.17159/2411-9717/3974/2026>

ORCID:

H. Ghasemzadeh

<https://orcid.org/0009-0009-6308-7872>

H. Madani

<https://orcid.org/0009-0002-6496-1807>

F. Sereshki

<https://orcid.org/0000-0002-2626-6865>

S. Afraei

<https://orcid.org/0000-0002-8476-6358>

Abstract

Spontaneous combustion of coal remains one of the most critical safety challenges in underground coal mining, posing significant risks to personnel, operational continuity, and the environment. Reliable prediction of coal self-heating potential is therefore essential for effective risk management throughout the mine life cycle. This study presents a data-driven framework for evaluating coal spontaneous combustion potential based on intrinsic coal properties derived from proximate and maceral analyses. A comprehensive dataset comprising 165 coal samples was compiled from published experimental studies, representing a wide range of coal characteristics. The crossing point temperature was adopted as a quantitative indicator of spontaneous combustion tendency. To investigate the complex relationships between coal properties and crossing point temperature, a combination of statistical and intelligent modelling approaches was employed, including nonlinear regression, artificial neural networks (multilayer perceptron and radial basis function), and a decision tree classifier. The results demonstrate that the relationship between intrinsic coal parameters and spontaneous combustion potential is inherently nonlinear and multivariate. While nonlinear regression achieved acceptable predictive performance ($R^2 = 0.82$), its mathematical complexity limits practical applicability. Among the developed intelligent models, the radial basis function neural network achieved the lowest prediction error (approximately 5.7%). In parallel, the decision tree model achieved a classification error of about 8% while providing a transparent and interpretable structure for practical decision-making. A key contribution of this study is the development of an integrated and interpretable classification framework based on readily available proximate and maceral analysis parameters, addressing the practical limitations associated with the lack of detailed elemental data in many mining operations. The developed framework was further evaluated using independent field samples from the Parvadeh coal mines (Tabas coalfield, Iran), confirming its applicability under real mining conditions. Unlike previous studies that mainly focused on individual algorithms or laboratory-scale datasets, this study establishes a comparative modelling framework by integrating multiple statistical and intelligent approaches using a comprehensive database of 165 coal samples and independent field validation. The proposed approach provides a practical and interpretable tool for assessing coal spontaneous combustion potential, with potential applications in mine design, operational planning, and post-closure risk management.

Keywords

Coal spontaneous combustion potential, crossing point temperature, intrinsic coal properties, classical statistical models, intelligent modelling techniques

Introduction

Coal remains one of the most widely utilised fossil fuels worldwide and is inherently prone to combustion. As a result, underground coal mining environments are characterised by complex operational conditions associated with a wide spectrum of safety hazards that must be systematically identified and managed from the early stages of mine design through extraction, operation, and continuous monitoring. Accordingly, the development of effective strategies for hazard prediction, prevention, and control is essential to ensure safe and sustainable mining practices. Among the major hazards encountered in underground coal mining—such as roadway collapse, gas explosions, surface subsidence, and rock bursts—coal spontaneous combustion is widely recognised as one of the most critical due to its direct role in initiating mine fires and causing severe operational disruptions (Akgün, Arisoy, 1994; Nugroho et al., 2000; Smith, Glasser, 2005; Beamish, Arisoy, 2008; Saffari et al., 2017).

Coal spontaneous combustion is a complex physicochemical process primarily governed by low-temperature oxidation reactions. This process may occur over a broad range of ambient conditions and

Assessment of coal spontaneous combustion potential using statistical and machine

can progressively lead to self-heating and eventual ignition in the absence of any external heat source (Demir et al., 2020). The global distribution of coal fire zones and regions affected by spontaneous combustion is illustrated in Figure 1.

The oxidation of coal is an exothermic process that results in a gradual temperature rise due to changes in its intrinsic thermal and chemical properties, which may ultimately lead to combustion. The fundamental mechanisms underlying coal oxidation are schematically presented in Figure 2. Despite extensive research efforts, a comprehensive understanding of this phenomenon remains challenging due to the complex interactions among multiple influencing parameters (Wang et al., 2003).

The impacts of coal spontaneous combustion are far-reaching and multifaceted, including loss of human life, environmental

degradation, greenhouse gas emissions, land subsidence, mine closure, infrastructure damage, production interruptions, and increased operational and remediation costs. These consequences underscore the importance of adopting responsible coal mining practices that align environmental, economic, and social considerations within a sustainable development framework. A conceptual overview of responsible coal mining is presented in Figure 3 (Michaylov, 2002).

Coal spontaneous combustion can occur in a variety of settings, including underground and surface mines, abandoned workings, coal stockpiles, waste dumps, and transportation systems such as maritime shipping (Michaylov, 2002). Consequently, continuous risk assessment is required throughout the entire mining life cycle, encompassing mine design, operation, closure, and post-mining reclamation. Such assessments typically begin with evaluating the spontaneous combustion potential of coal and identifying the key parameters influencing this behaviour (Lang, Fu-Bao, 2010; Wang, Tong, 2020; Ahamed et al., 2016).

Accurate evaluation of coal spontaneous combustion potential requires reliable and robust methods capable of handling large datasets, minimising prediction uncertainty, and maintaining applicability under real mining conditions. Although numerous studies have explored the influence of various parameters on spontaneous combustion, many existing models suffer from important limitations, including reliance on limited datasets, dependence on complex or less accessible input parameters, reduced interpretability, and limited applicability in practical mining environments.

In real-world mining operations, access to detailed coal characterisation data is often constrained, particularly when rapid and cost-effective assessments are required. In this context, proximate analysis parameters are generally more accessible and routinely measured compared to detailed elemental data. Despite their practical advantages, the systematic use of these parameters for developing accurate, reliable, and interpretable predictive models has not been sufficiently addressed in the existing literature.

To overcome these limitations, this study proposes a data-driven and practically applicable framework for evaluating coal spontaneous combustion potential based on intrinsic coal properties derived primarily from proximate and maceral analyses. A comprehensive dataset comprising 165 coal samples was compiled from published experimental studies to enhance model robustness and generalisability. The crossing point temperature (CPT), a widely accepted laboratory-based index for assessing spontaneous combustion tendency, was adopted as the primary evaluation criterion in this study (Misra, Singh, 1994). The classification of coal based on CPT values is presented in Table 1.



Figure 1—Countries and fire zones that have experienced spontaneous combustion problems (Demir et al., 2020)

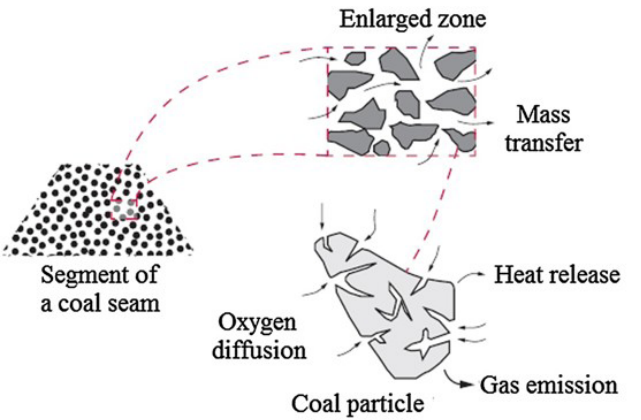


Figure 2—Fundamental phenomena occurring in coal oxidation process (Wang et al., 2003)

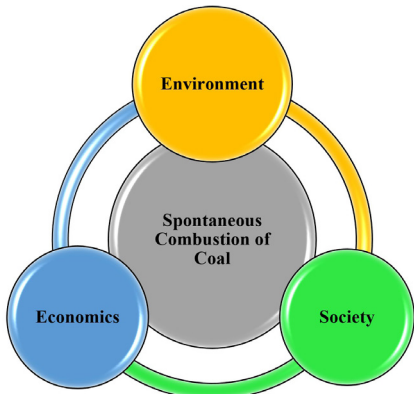


Figure 3—Responsible mining of coal

Table 1 Classification of CPT values and the status of spontaneous combustion of coal potential (Misra, Singh, 1994)		
The spontaneous combustion of coal potential	CPT value (Celsius)	Class
low	>160	1
medium	140-160	2
high	120-140	3

Assessment of coal spontaneous combustion potential using statistical and machine

A combination of classical statistical techniques and advanced intelligent modelling approaches was employed, including nonlinear regression, artificial neural networks (multilayer perceptron and radial basis function), and a decision tree classifier. This integrated methodology enables both high predictive accuracy and improved model interpretability.

The main contributions of this study can be summarised as follows: (1) development of a comprehensive database containing 165 coal samples by integrating published experimental CPT data and intrinsic coal characteristics; (2) systematic comparison of classical statistical models and intelligent approaches, including regression models, artificial neural network (ANN)-based methods, radial basis function (RBF) networks, and decision tree algorithms, within a unified modelling framework; (3) development of an interpretable spontaneous combustion classification decision tree (SCCDT) model that provides a balance between predictive accuracy, simplicity, and practical usability; and (4) validation of the proposed framework using independent coal samples from the Parvadeh Tabas coal mines to evaluate its applicability under real mining conditions.

The proposed framework provides a reliable, accurate, and interpretable decision-support tool for assessing coal spontaneous combustion potential and supports improved safety management and risk mitigation strategies in underground coal mining operations.

Literature review

Over the past century, extensive research has been conducted on the spontaneous combustion of coal due to its critical implications for mine safety, economic performance, and environmental sustainability. Early foundational studies, such as those by Parr and Hilgard (1925), identified pyrite oxidation as a key contributor to self-heating, thereby establishing the fundamental chemical mechanisms governing spontaneous combustion. Subsequent investigations demonstrated that coal composition plays a decisive role in determining susceptibility to self-heating, with different coal ranks exhibiting significantly different reactivity under similar environmental conditions (Guney, 1968; Feng et al., 1973). These findings positioned intrinsic coal properties as the central focus of spontaneous combustion research across the entire mining lifecycle, including post-mining conditions (Şensöğüt, 2011).

Between the 1970s and early 2000s, research expanded to incorporate both intrinsic and operational factors influencing spontaneous combustion. Early studies examined the effects of moisture content (Hodges, 1963), coalification degree (Chandra, Prasad, 1990), maceral composition (Cotterell, 1997), and ventilation-induced heat accumulation (Mitchell, 2003; Yuan, Smith, 2012), reinforcing the dominant role of intrinsic coal properties in governing oxidation behaviour. Concurrently, several investigations addressed mining-related and geological parameters—including gob characteristics, air leakage pathways, and longwall face advance rate (Zhang, Qin, 2009; Taraba, Michalec, 2011; Xia et al., 2015)—demonstrating that operational conditions can significantly amplify or mitigate spontaneous combustion risk. A summary of these influencing factors is presented in Table 1.

Building on these foundations, more systematic investigations over the past two decades evaluated spontaneous combustion potential across various mining environments, including underground mines, open-pit operations, and coal stockpiles. Empirical and semi-analytical relationships between coal properties

and self-heating indices were developed, and structured risk assessment frameworks—often based on historical fire data and hierarchical decision models—were introduced, particularly in Chinese coal mining practices (Lang, Fu-Bao, 2010; Banerjee, 1985; Smith, Lazzara, 1987; Kaymakci, Didari, 2002).

From approximately 2010 onward, the literature reflects a transition toward quantitative and mechanistic modelling approaches. Beamish and Beamish (2008) and Beamish and Sainsbury (2010) proposed empirical self-heating rate equations to quantify oxidation kinetics, Liu and Qin (2017) developed multi-physics coupling models integrating oxidation reactions, heat transfer, and gas migration processes in longwall gob environments. Thermodynamic analyses further highlighted the catalytic role of pyrite in accelerating oxidation reactions and heat generation (Yang et al., 2019). A conceptual representation of these coupled processes is illustrated in Figure 2.

Between 2018 and 2023, rapid methodological advancements emerged with the adoption of data-driven and intelligent modelling techniques. Studies by Onifade and Genc (2018a; 2020), Saffari et al. (2019), and Gbadamosi et al. (2021) demonstrated strong correlations between proximate analysis parameters—including fixed carbon (FC), ash content (AC), moisture content (MoC), volatile matter (VM), and pyritic content (PC)—and spontaneous combustion indices. Additional studies investigated the influence of external factors such as drying methods (Zhang et al., 2018; Zhao et al., 2016), particle size distribution (Więckowski et al., 2020), and moisture migration mechanisms (Wu et al., 2020), further refining the understanding of coal reactivity.

Recent research has increasingly incorporated advanced computational tools and digital technologies. These include computational fluid dynamics (CFD) modelling for simulating airflow and heat transfer (Taraba, Michalec, 2011), computer-based mining system simulations (Meshkov, 2019), and internet of things (IoT)-based classification systems for real-time monitoring (Mishra, Gupta, 2024). Furthermore, machine learning approaches—including neuro-fuzzy systems (Sahu, Mahapatra, 2013), logistic regression models (Kursunoglu, Gogebakan, 2022), support vector machine-based algorithms such as SASVM (Deng et al., 2021), and multi-parameter optimisation frameworks (Xu, Zhang, 2023)—have significantly enhanced prediction accuracy and enabled automated risk classification. A comparative overview of selected predictive models is presented in Table 2.

More recent studies have expanded the research scope to include integrated analyses of coal mine fires, their environmental impacts, and preventive strategies. Accurate prediction of spontaneous combustion potential has been identified as a critical component of fire prevention systems. Consequently, advanced data-driven techniques have gained increasing attention. For instance, random forest algorithms have been applied to predict spontaneous combustion potential with high accuracy in large datasets. Additionally, several studies have emphasised the critical role of intrinsic coal properties—particularly moisture content—in controlling oxidation behaviour. Comprehensive analyses have shown that increases in fixed carbon, moisture, hydrogen, volatile matter, nitrogen, and pyrite content, combined with lower ash content, tend to enhance spontaneous combustion susceptibility (Saffari et al., 2017, 2020; Song, Kuenzer, 2014; Tosun, 2017; Lei et al., 2018; Onifade, Genc, 2018b).

In parallel, methodological developments have led to a wide range of approaches for assessing spontaneous combustion

Assessment of coal spontaneous combustion potential using statistical and machine

potential, including laboratory-based experimental techniques, mathematical modelling, probabilistic risk assessment methods, and artificial intelligence-based frameworks. Comparative studies indicate that machine learning techniques often outperform traditional empirical models; for example, random forest algorithms have demonstrated superior predictive performance compared to multilayer perceptron (MLP) networks when applied to large coal datasets (Onifade, Genc, 2020; Said et al., 2021; Deng et al., 2021; Zhang et al., 2022).

More recently, emerging technologies have been integrated into comprehensive risk assessment frameworks. IoT-based systems have been proposed for classifying coal seam self-heating behaviour using intrinsic coal properties as input parameters. In addition, risk-based analytical techniques such as fault tree analysis (FTA) have been applied to evaluate spontaneous combustion risk under complex mining conditions (Mishra, Gupta, 2024; Yetkin et al., 2024; Wang et al., 2023).

Based on the reviewed studies, it is evident that coal spontaneous combustion is a complex, multi-factor phenomenon governed by a wide range of interacting variables. These influencing factors can be broadly categorised into three main groups: (i) intrinsic coal properties, (ii) geological characteristics of coal seams, and (iii) mining and operational parameters. Among these, intrinsic properties—including moisture content, ash yield, volatile matter, fixed carbon, pyrite content, and petrographic composition—have consistently been identified as the most influential determinants of spontaneous combustion potential across different spatial and temporal scales.

As summarised in Table 2, the parameters influencing coal spontaneous combustion potential can be systematically categorised into three major groups based on the direction and nature of their effects: intrinsic coal properties, geological characteristics, and mining-related factors. This classification provides a structured framework for understanding whether each parameter contributes positively or negatively to spontaneous combustion tendency.

From the perspective of intrinsic coal properties, coal rank plays a pivotal role in determining reactivity. Lower-rank coals typically exhibit higher porosity and greater oxygen accessibility, thereby increasing their susceptibility to spontaneous combustion. Pyrite content has been widely recognised as a catalytic factor that accelerates oxidation reactions and enhances heat generation. Moisture content presents a dual and competing influence: while it promotes oxidation and heat accumulation under certain conditions, moisture evaporation can simultaneously absorb heat and locally suppress combustion processes. Oxygen availability,

whether through natural permeability or ventilation leakage, directly intensifies oxidation reactions and increases combustion risk. Petrographic composition further governs coal reactivity, with vitrinite- and liptinite-rich coals exhibiting higher susceptibility compared to inertinite-dominated coals. In contrast, ash content generally reduces spontaneous combustion potential by acting as a thermal sink that absorbs oxidation heat. Methane presence tends to inhibit oxygen diffusion within the coal matrix, thereby reducing oxidation intensity. Coal brittleness increases the exposed reactive surface area, which enhances oxidation kinetics.

Additionally, inorganic constituents exhibit contrasting roles, where certain salts suppress oxidation, while alkali elements such as sodium and potassium promote it. Volatile matter content is consistently associated with increased spontaneous combustion potential.

From a geological standpoint, increasing seam depth results in higher in situ stresses, leading to coal fracturing and enhanced permeability, both of which facilitate oxygen ingress and oxidation. Greater seam thickness increases the exposed coal surface area, thereby intensifying oxidation reactions. Steeper seam dip angles often reduce recovery efficiency and increase residual coal, indirectly elevating combustion risk.

Geological discontinuities such as faults, clastic dikes, clay veins, and intrusive bodies create pathways for air leakage and further increase exposure surfaces. Multi-seam mining operations can induce stress redistribution and fracturing in adjacent seams, thereby amplifying spontaneous combustion susceptibility. Additionally, the presence of adjacent oil shales or geothermal heat sources can elevate seam temperature and accelerate self-heating processes.

Mining-related parameters also play a critical role in controlling spontaneous combustion risk. Mining methods that achieve higher recovery rates and minimise residual coal generally reduce combustion potential. However, mining-induced pressure differentials may promote air leakage into goaf areas, thereby increasing oxidation rates. Lower face advance rates prolong coal exposure to oxygen, which significantly increases combustion risk. Ventilation systems exhibit a dual influence: insufficient airflow limits heat dissipation, while excessive airflow enhances oxygen supply and accelerates oxidation. Air leakage into caved zones is one of the most critical contributors to spontaneous combustion in underground mines. Furthermore, higher extraction ratios reduce the amount of residual coal and consequently decrease oxidation potential.

It is important to note that inconsistencies in previous studies

Table 2 The parameters affecting coal spontaneous combustion			
	Parameter type	Parameters with a direct relationship	Parameters with inverse relationship
1	Intrinsic factors	Pyrite, moisture, oxygen, vitrinite, liptinite, coal brittleness, potassium and sodium, and volatiles.	Degree of charring, inertinite, ash, gassiness, and salt family.
2	Geological factors	Depth, thickness, layer slope, fault presence, and multi-layer extraction.	
3	Exploitation factors	Direction of advance, intensity of ventilation air flow, and air leakage.	Complete extraction, advance speed, type of side ventilation system, and coal extraction coefficient.

Assessment of coal spontaneous combustion potential using statistical and machine

often arise from the isolated evaluation of individual parameters without considering their interactions. Such approaches fail to capture the complex, nonlinear, and overlapping relationships among influencing factors. This limitation highlights the necessity for integrated modelling frameworks capable of simultaneously incorporating multiple parameters and their interdependencies.

Given the wide range of influencing factors and their complex interactions, evaluating spontaneous combustion potential using single-parameter or traditional empirical approaches is insufficient. Many existing studies are limited by small datasets, site-specific conditions, or simplified modelling assumptions, which restrict their general applicability. Therefore, there is a clear need for robust, data-driven, and integrated modelling approaches that can handle multiple variables, capture nonlinear relationships, and provide reliable predictions across different mining conditions.

Accordingly, this study aims to develop a comprehensive predictive framework for assessing coal spontaneous combustion potential based on intrinsic coal properties. The approach integrates classical statistical methods, including linear and nonlinear regression models, with advanced intelligent techniques such as artificial neural networks (ANN), radial basis function (RBF) models, and decision tree algorithms. The crossing point temperature (CPT) is used as a quantitative indicator for classification and validation. By systematically comparing model performance in terms of accuracy, robustness, and practical applicability, this study seeks to provide a reliable decision-support tool for improving safety management in underground coal mining operations.

The remainder of this paper is organised as follows. A presentation of the data analysis and evaluation procedures is followed by a description of the development of classical statistical and intelligent models. The section that follows provides model validation results and finally, the findings are discussed the main conclusions presented.

Analysis and evaluation of data affecting the spontaneous combustion of coal potential

The analysis and evaluation of parameters influencing coal spontaneous combustion potential were carried out through a series of systematic and sequential steps. Based on the literature review, intrinsic coal properties have received the greatest research attention, with 32 studies focusing on intrinsic characteristics, compared to only 2 studies on geological characteristics and 12 studies addressing mining-related factors. This distribution highlights not only the dominant role of intrinsic coal properties but also the unequal consideration of different influencing parameters. Table 3 presents the distribution of studies on coal spontaneous combustion potential from 1980 to the present.

As shown in Table 3, and as discussed in the literature review section, intrinsic coal characteristics have been investigated far

more extensively than geological and mining-related factors in the assessment of spontaneous combustion potential. Therefore, this study primarily focuses on intrinsic coal properties as the main input variables for model development.

The selection of input variables was performed based on a combination of previous research findings, physicochemical understanding of coal spontaneous combustion mechanisms, and practical data availability from operating mines. In this study, intrinsic coal parameters obtained from proximate analysis and petrographic characterisation were selected because these parameters directly influence oxygen accessibility, oxidation reactivity, and heat generation during the self-heating process. Although elemental analysis and other advanced characterisation techniques can provide additional information, such data are not consistently available for practical mine applications. Therefore, the selected variables represent a balance between scientific relevance and operational applicability. Feature reduction techniques such as recursive feature elimination or principal component analysis were not applied because the selected parameters were already established as influential factors in previous studies and were required to maintain physical interpretability of the developed models.

Statistical analysis of intrinsic characteristics of coal

Previous studies have applied a wide range of approaches to evaluate and predict coal spontaneous combustion potential, ranging from conventional statistical techniques—such as linear and nonlinear regression models—to advanced intelligent methods, including artificial neural networks (ANN), radial basis function (RBF) networks, decision trees, random forests, and hybrid machine learning models. While classical statistical models offer simplicity and interpretability in analysing parameter relationships, they are often insufficient for capturing complex nonlinear interactions.

To evaluate coal spontaneous combustion potential, a comprehensive statistical analysis of intrinsic coal properties was conducted to establish relationships between these characteristics and crossing point temperature (CPT) values. For this purpose, a total of 165 datasets were collected from previously published studies covering various underground coal mines. In all referenced studies, moisture, ash, volatile matter, and fixed carbon contents were determined using proximate analysis in accordance with ASTM standards.

Table 4 summarises the results of proximate analysis, petrographic maceral composition, and CPT values for the collected coal samples. The frequency distribution of CPT values is illustrated in Figure 4, with CPT values presented in the final column of Table 4. As shown, the dataset consists of 165 samples, with CPT values ranging from approximately 70 °C to 200 °C. The mean CPT value is 153 °C, with the first and third quartiles equal to 140 °C and 174 °C, respectively.

Table 3 Distribution of studies on the spontaneous combustion of coal potential from 1980 to the present			
Characteristics affecting the spontaneous combustion of coal potential	Intrinsic characteristics	Geological characteristics	Mining characteristics of coal
Number of studies	32	2	12

Assessment of coal spontaneous combustion potential using statistical and machine

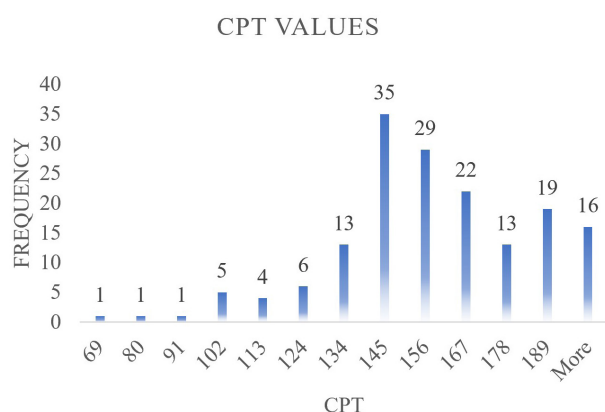


Figure 4—Frequency distribution chart of CPT values

Based on the CPT frequency distribution, approximately 31 samples are classified as having high spontaneous combustion potential, 86 samples fall within the medium category, and 48 samples are categorised as having low spontaneous combustion potential.

Table 4 presents descriptive statistical measures, including mean, standard deviation, first quartile, and third quartile, for the eight primary intrinsic parameters considered in this study. Furthermore, Figure 5 illustrates the frequency distribution of intrinsic coal characteristics across the dataset, confirming the total number of 165 samples.

Before model development, 25 samples were randomly excluded from the original dataset and completely reserved as an independent internal testing dataset. These samples were not involved in model training, parameter adjustment, or model selection. The remaining samples were used for developing the predictive models. Furthermore, independent field data collected from the Parvadeh Tabas coal mines were used exclusively for external validation and were not included in any stage of model development or evaluation.

Cross-correlation of intrinsic coal characteristics and experimental CPT values

In this section, the relationship between intrinsic coal characteristics and experimentally measured CPT values was investigated using univariate regression analysis. Both linear and nonlinear regression models were examined for each intrinsic parameter, and the model with the highest coefficient of determination (R^2) was selected as the representative relationship.

The results indicate that none of the intrinsic coal parameters exhibit a purely linear relationship with CPT, which is widely used as an indicator of coal spontaneous combustion potential. This nonlinearity can be attributed to the simultaneous and interacting effects of multiple intrinsic parameters governing coal oxidation behaviour.

As illustrated in Figure 6, moisture content exhibits a non-monotonic relationship with CPT. Specifically, spontaneous combustion potential decreases with increasing moisture content up to approximately 10%, after which further increases in moisture lead to higher combustion potential. A similar trend is observed for ash content, where CPT values increase when ash exceeds approximately 21%. Volatile matter demonstrates a threshold behaviour around 18%, while fixed carbon shows a turning point near 40%. Pyrite content initially reduces CPT values up to approximately 5%, followed by an increase in spontaneous combustion potential at higher concentrations.

Petrographic components also display nonlinear relationships with CPT. Vitrinite content increases combustion potential up to approximately 30%, beyond which higher proportions reduce susceptibility. Liptinite follows a similar pattern with a turning point around 4%, while inertinite content increases combustion potential up to approximately 70%, after which the tendency declines.

The univariate regression results further indicate that, when considered independently, pyrite content exhibits the strongest correlation with CPT ($R^2 \approx 0.59$), whereas ash content and fixed carbon show the weakest relationships ($R^2 \approx 0.04$). As demonstrated in Figure 6, none of the intrinsic parameters alone can adequately explain the variability in CPT values, and the predictive capability of single-variable models remains limited.

Therefore, to achieve a more accurate and realistic

Table 4

Results of proximate analysis, petrographic maceral analysis, and CPT values of coal data

Statistical Variable	CPT	Intrinsic characteristics- %							
		Inertinite	Liptinite	Vitrinite	Pyrite	Fixed carbon	Volatile matter	Ash	Moisture
Mean	153	36.5	7.6	55.9	2	44	23.3	25.6	5.1
Median	149	35.3	7.9	57.4	1.8	44.3	23.2	23	3.2
Mode	180	0	11.5	72.1	0	38.1	18.5	41.2	2.2
Standard deviation	26.6	21.1	4.0	18.2	1.9	11.1	6.5	13.9	5.3
Sample variance	714	4.5	0.2	3.3	0.03	1.3	0.4	2	0.3
Range of variation	131	90	17	84	12	60	42	77	31
The minimum value	68.9	0	0.6	2.4	0	7.5	2	2.2	0.5
The maximum value	200	90.4	17.9	86.6	12.1	67.4	44	79.3	31.9
the first quartile	140	24.5	4.2	47.1	0.5	38	18.5	15.3	2
the third quartile	174	48.8	9.6	65.2	2.7	51.6	28	32.5	6.3
Number of values	165	165	165	165	165	165	165	165	165

Assessment of coal spontaneous combustion potential using statistical and machine

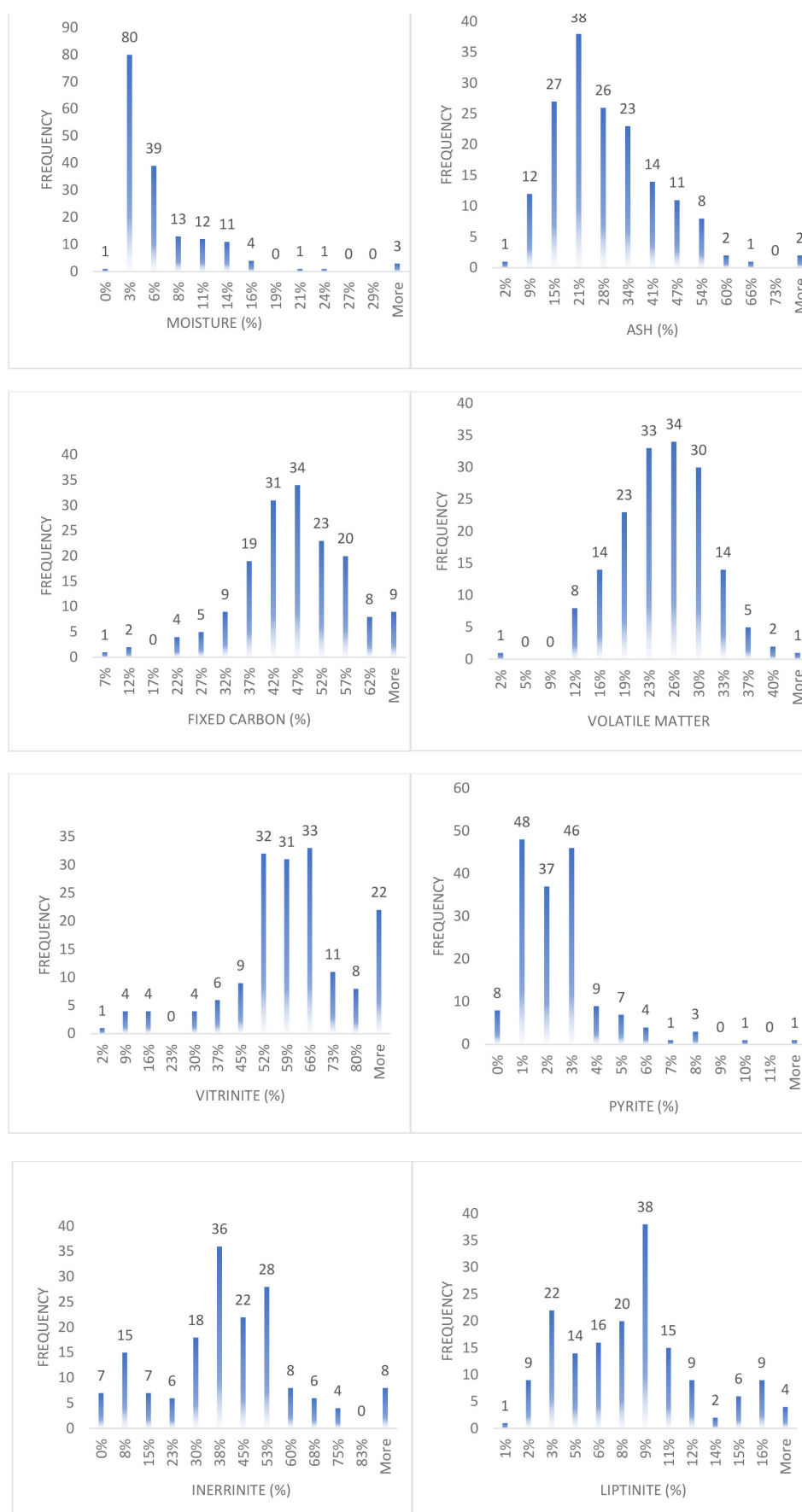


Figure 5—Frequency distribution chart of intrinsic characteristics values of coal

Assessment of coal spontaneous combustion potential using statistical and machine

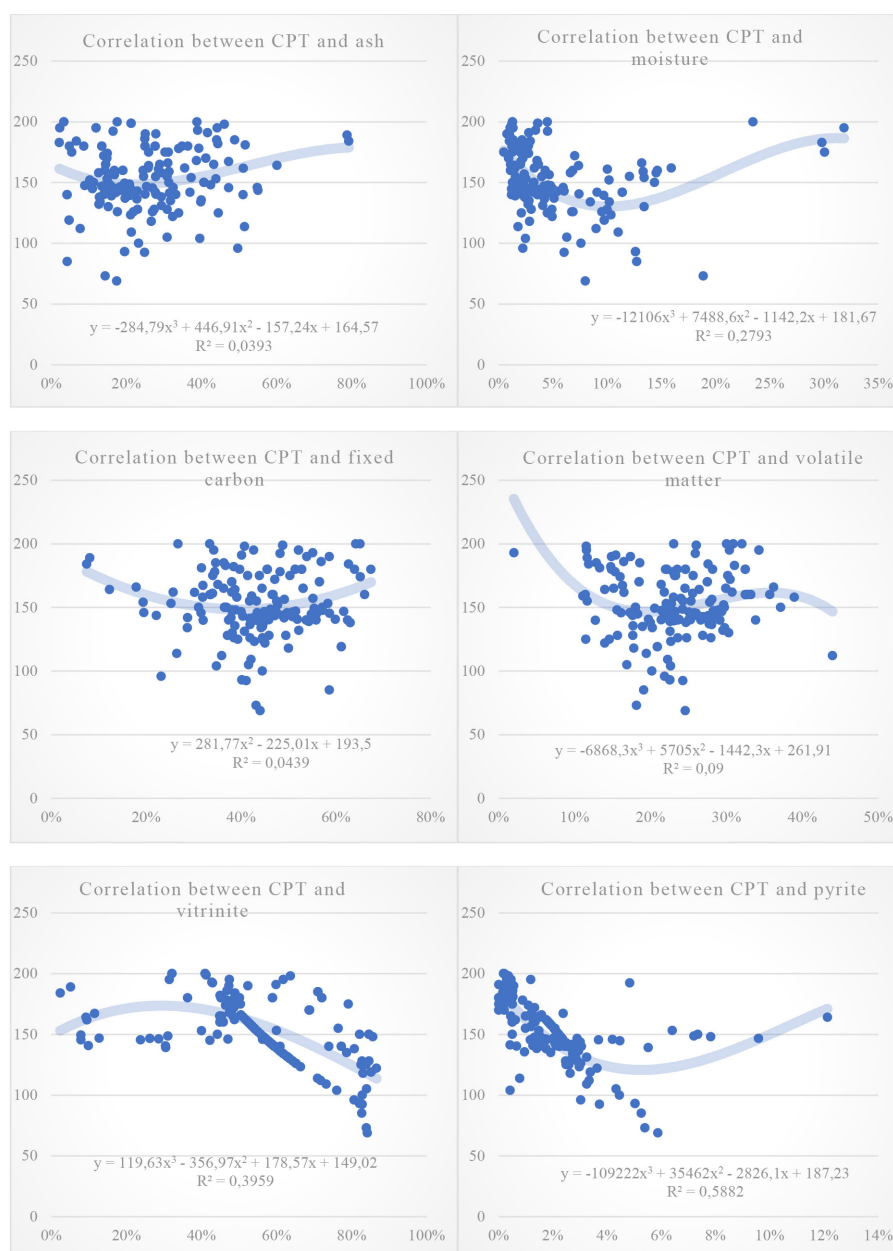


Figure 6—Relationship between intrinsic coal characteristics and CPT values

understanding of coal spontaneous combustion behaviour, and considering the complex and multifactorial nature of coal self-heating, multivariate regression analysis was adopted in the next stage of this study. This approach allows for capturing the combined effects and interactions among multiple parameters, thereby providing a more robust foundation for subsequent modelling using advanced intelligent techniques.

Implementation and development of models

In this section, a comprehensive set of statistical and intelligent modelling approaches was developed and systematically evaluated to identify a robust and reliable framework for predicting coal spontaneous combustion potential. The applied methods were categorised into two main groups: classical statistical approaches and artificial intelligence-based techniques.

The selection of modelling approaches was based on the specific characteristics of coal spontaneous combustion prediction,

including nonlinear relationships among variables, limited availability of experimental data, and the need for both predictive accuracy and practical interpretability. Classical regression models were initially considered because they provide a quantitative understanding of relationships between intrinsic coal properties and CPT values and serve as baseline models for evaluating predictive improvements.

Artificial neural network (ANN)-based models were selected due to their strong capability in approximating nonlinear and complex interactions among multiple input variables without requiring predefined mathematical relationships. In particular, multilayer perceptron (MLP) networks were considered as a general nonlinear approximation approach, whereas radial basis function (RBF) networks were selected because of their efficient learning capability and suitability for problems involving nonlinear mapping with limited datasets.

In addition, decision tree modelling was employed because

Table 5

5 Results of implementing multiple linear regression

Parameter	Unstandardised	coefficients	Standardised	coefficients				VIF
	Beta	Std. Error	Beta		t	Sig	Tolerance	
Constant	192.834	10.233			18.844	0.000		
Moisture	-31.247	25.669	-0.063		-1.217	0.225	0.804	1.244
Volatile matter	-4.052	20.332	-0.010		-0.199	0.842	0.878	1.139
Fix carbon	-22.016	12.072	-0.092		-1.824	0.070	0.836	1.197
Pyrite	-520.916	73.943	-0.362		-7.045	0.000	0.807	1.239
Liptinite	-324.848	52.006	-0.491		-6.246	0.000	0.345	2.900
Inertinite	19.354	9.393	0.153		2.061	0.041	0.386	2.592

Table 6

Results of implementing nonlinear multiple regression

Parameter	Model grade
Constant value	-
Moisture	3
Ash	3
Volatile matter	3
Fix carbon	2
Pyrite	3
Vitrinite	3
Liptinite	3
Inertinite	3

spontaneous combustion assessment in mining applications requires not only accurate prediction but also transparent decision-making. Unlike black-box models, decision trees generate explicit classification rules that can be directly interpreted by mining engineers. Therefore, the selected models provide complementary advantages, including statistical interpretability, nonlinear prediction capability, and practical applicability for mine safety decision support.

After selecting the appropriate modelling approaches, different model configurations and parameter settings were evaluated through an iterative testing procedure to determine the most suitable structures and parameters for prediction. For regression-based approaches, different mathematical formulations and model structures were examined, and the optimal formulations were selected based on the highest prediction performance and the lowest error values. In nonlinear regression analysis, different polynomial degrees were tested for the input parameters, and the final model was selected based on the best agreement between predicted and experimental CPT values.

For artificial intelligence models, several network architectures with different numbers of hidden layers and neurons were developed and evaluated. The final ANN, MLP, and RBF structures reported in this study correspond to the configurations that provided the best predictive performance among the evaluated alternatives. Similarly, different decision tree structures were

examined, and the final model was selected considering both classification accuracy and interpretability. Due to the large number of evaluated configurations, only the optimal model structures and their corresponding performance results are presented to maintain the clarity and conciseness of the manuscript.

Linear and nonlinear multiple regression

Initially, multiple linear regression (MLR) was implemented using selected intrinsic coal parameters. The results are presented in Table 5, where the developed model achieved a coefficient of determination of $R^2 = 0.633$, indicating moderate predictive capability.

Despite the acceptable performance, the univariate analyses presented in the aforementioned demonstrated that the relationships between intrinsic parameters and CPT are predominantly nonlinear. This discrepancy highlights an inherent limitation of linear regression when applied to complex physicochemical processes such as coal spontaneous combustion, where multiple variables interact simultaneously.

Although most regression coefficients were statistically significant ($\text{Sig.} < 0.05$) and no severe multicollinearity was observed (as indicated by tolerance and VIF values in Table 5), further improvement using purely linear formulations remained limited. Additional linear models with different parameter combinations yielded R^2 values of up to approximately 0.73.

To overcome these limitations, nonlinear multiple regression was subsequently employed. Based on the relationships identified earlier, polynomial degrees of 3 were assigned to most variables, while fixed carbon was modelled using a second-degree polynomial. The resulting nonlinear regression model, summarised in Table 6, significantly improved predictive performance, achieving $R^2 = 0.824$, corresponding to an approximate 20% increase compared to the linear model.

However, this improvement was accompanied by increased model complexity due to the large number of polynomial terms, which may reduce interpretability and limit practical implementation.

To address dimensionality and multicollinearity concerns, principal component analysis (PCA) was integrated with regression modelling. Several PCA-based scenarios were evaluated. Linear regression using three principal components derived from the

Assessment of coal spontaneous combustion potential using statistical and machine

original variables resulted in $R^2 = 0.424$, while nonlinear regression based on PCA-transformed variables achieved $R^2 = 0.763$. However, applying PCA to nonlinear features resulted in a reduction in predictive accuracy compared to the full nonlinear regression model.

Overall, while regression-based approaches demonstrated acceptable to good performance, their limitations in capturing complex nonlinear interactions motivated the adoption of intelligent modelling techniques.

Artificial intelligence methods

Given the inherently nonlinear and multivariate nature of coal spontaneous combustion, artificial intelligence (AI) techniques were employed to improve prediction accuracy and capture complex interactions among intrinsic coal parameters.

Model configuration and hyper parameter selection

The configuration of the developed statistical and intelligent models was determined through an iterative model development and evaluation procedure. Due to the complexity of coal spontaneous combustion behaviour and the limited size of the available dataset, different model configurations were examined, and the final models were selected based on their predictive performance and generalisation capability.

For artificial neural network models, different network architectures were evaluated by varying the number of hidden layers and neurons. Several combinations of network structures were trained and tested, and the architecture providing the lowest prediction error and the best agreement between predicted and experimental CPT values was selected as the final configuration. The detailed results of all tested architectures are not presented to avoid excessive repetition; however, the selected optimal structures are reported in the following sections.

Similarly, for regression-based models, different mathematical formulations and parameter combinations were investigated. In nonlinear regression analysis, different polynomial orders were

evaluated for each input variable, and the final formulation was selected based on the highest coefficient of determination and the lowest prediction error. For the principal component analysis (PCA)-based models, different numbers of components were also examined to determine the optimal representation of the original variables.

For the decision tree model, different splitting criteria and tree structures were evaluated, and the final model was selected based on classification accuracy while maintaining an interpretable structure suitable for practical mining applications.

Therefore, the reported models represent the optimal configurations obtained after systematic evaluation of alternative structures rather than arbitrary model settings.

Hybrid linear–nonlinear artificial neural network

In the first step, a hybrid linear–nonlinear artificial neural network was developed using intrinsic coal parameters as input variables. This approach combines linear transformations with nonlinear activation functions to enhance model flexibility while maintaining numerical stability.

As illustrated in Figure 7, the developed model achieved a mean squared error (MSE) of 69, a root mean squared error (RMSE) of 8.3, a mean error close to zero (-2×10^{-6}), and a mean error standard deviation of 8.33.

The model was trained using different polynomial orders, and the results indicated that a configuration with polynomial order 8 yielded the lowest prediction error. The final model demonstrated a prediction error of approximately 8.3%, confirming the effectiveness of hybrid ANN structures in modelling coal spontaneous combustion potential.

Multilayer perceptron neural network

In the next step, a multilayer perceptron (MLP) neural network was implemented to model the nonlinear relationships between intrinsic coal parameters and CPT values.

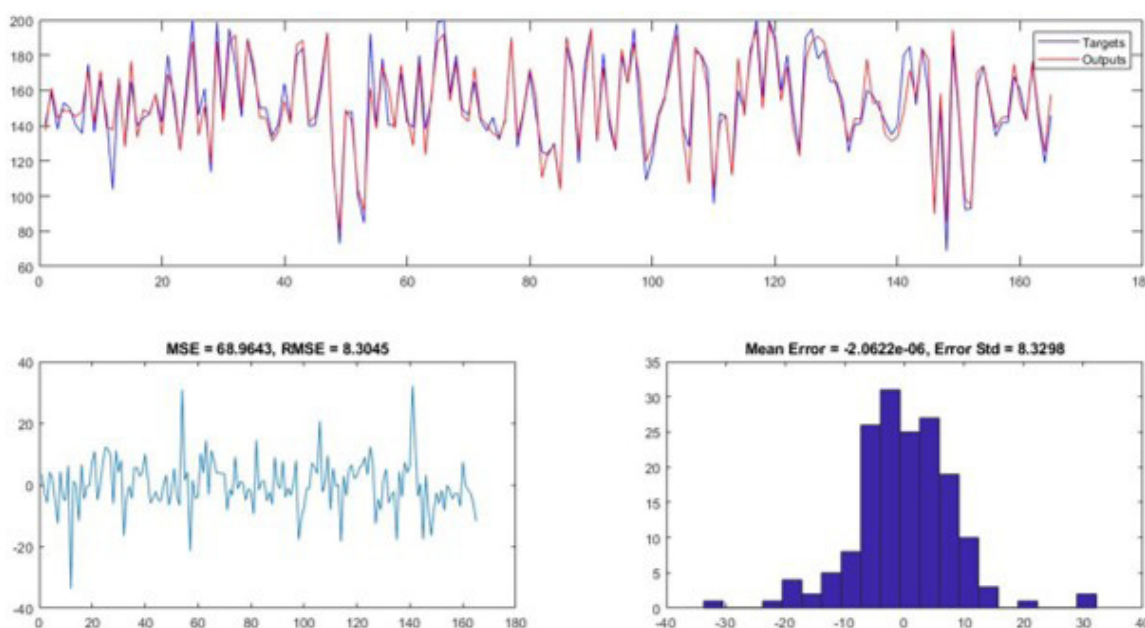


Figure 7—Implementation of a combined linear and nonlinear neural network method

Assessment of coal spontaneous combustion potential using statistical and machine

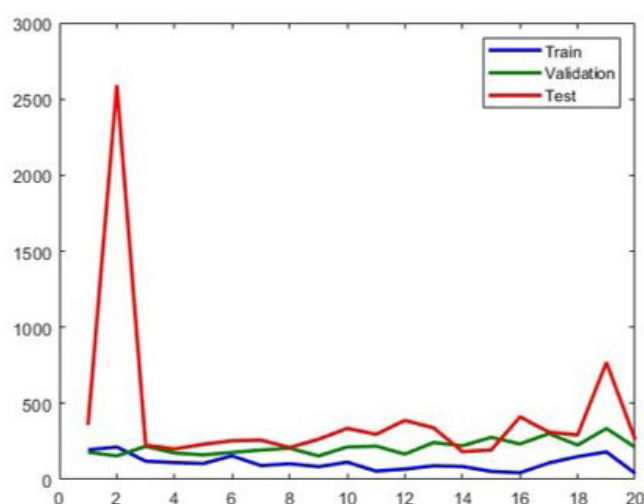


Figure 8—Implementing regression using the MLP neural network method

As shown in Figure 8, sensitivity analysis was conducted to determine the optimal number of neurons in the hidden layer. The results indicated that the minimum prediction error was achieved with three neurons, which was selected as the optimal configuration.

To further enhance model performance, the architecture was extended to include three hidden layers. The results of this enhanced model are presented in Figure 9, which illustrates regression plots comparing predicted and experimental CPT values for training, validation, testing, and overall datasets.

As observed in Figure 9, the data points are closely distributed around the 45-degree reference line, indicating strong agreement between predicted and measured values. The correlation coefficients were calculated as 0.94 (training), 0.87 (validation), 0.96 (testing), and 0.93 (overall), demonstrating high predictive accuracy.

Moreover, the consistency between training and testing results suggests that the model does not exhibit overfitting and maintains good generalisation performance when applied to unseen data.

Radial basis function neural network

In the third step, a radial basis function (RBF) neural network was developed to further improve prediction accuracy. RBF networks are particularly effective in capturing localised nonlinear patterns due to their radial activation functions.

As illustrated in Figure 10, the RBF model achieved a mean error close to zero (4.35×10^{-13}) and an error standard deviation of 5.72, resulting in an overall prediction error of approximately 5.7%.

Compared to the MLP model, the RBF network demonstrated

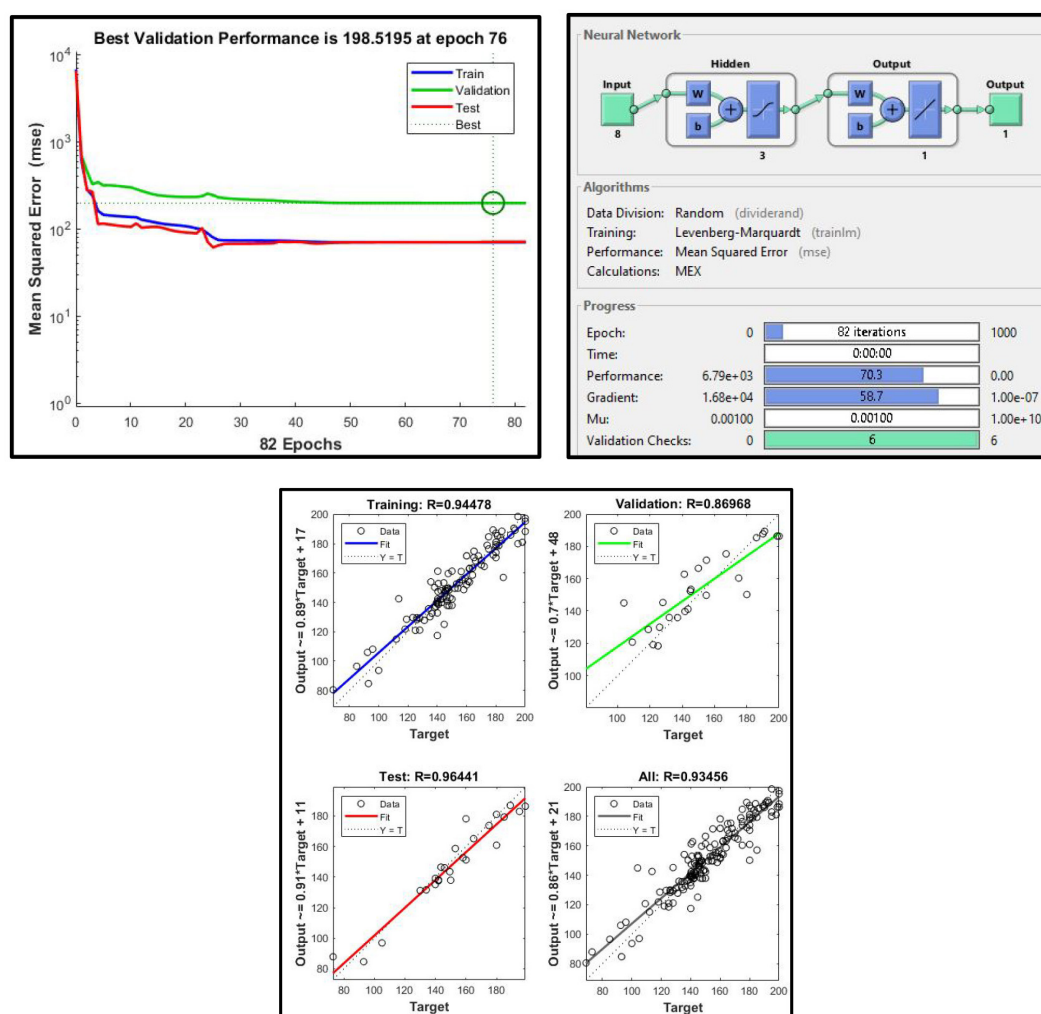


Figure 9—Implementing regression using a neural network method

Assessment of coal spontaneous combustion potential using statistical and machine

improved predictive performance, highlighting its superior capability in modelling complex nonlinear relationships among intrinsic coal parameters.

Decision tree model

Finally, a decision tree (DT) model was developed using hierarchical classification based on intrinsic coal parameters. The model was constructed using ID3 and CART algorithms with a top-down recursive partitioning approach.

The analysis identified pyrite content as the most influential parameter, which was selected as the root node with a threshold value of 0.64%. At the second level, inertinite content was identified as the dominant splitting parameter, with a threshold of 12.36%.

Based on the derived decision rules:

- Samples with pyrite < 0.64% and inertinite > 12.36% show approximately 97% probability of low spontaneous combustion potential.
- Samples with pyrite > 0.64% and inertinite > 31.758% show approximately 74% probability of moderate potential.
- Samples with pyrite > 0.64% and inertinite < 31.758% show approximately 82% probability of high potential.

The final decision tree model is presented in Figure 11, providing a transparent, interpretable, and practically applicable framework for qualitative and semi-quantitative assessment of coal spontaneous combustion potential.

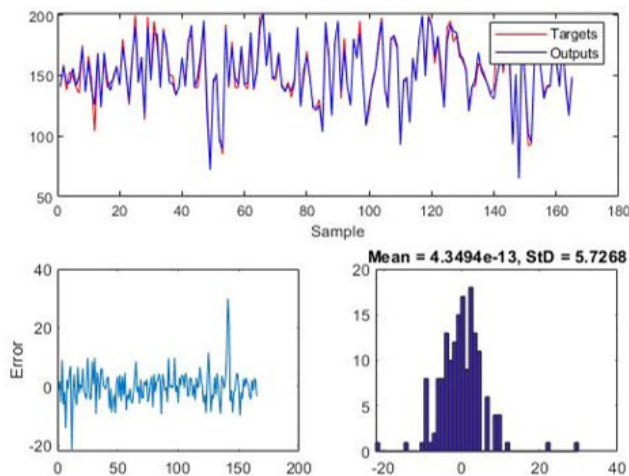


Figure 10—Implementing regression using the RBF neural network method

Quantitative comparison and performance evaluation of developed models

To provide an overall assessment of the developed approaches, the predictive performance of all statistical and intelligent models was quantitatively compared using mean squared error (MSE) and root mean squared error (RMSE). The results of this comparative evaluation are summarised in Table 7. This analysis enables a direct quantitative assessment of the predictive capability of each modelling approach and provides a basis for discussing their advantages and limitations in practical applications.

As shown in Table 7, the multiple linear regression model exhibited the highest prediction error, indicating its limited capability in representing the complex nonlinear interactions governing coal spontaneous combustion. The nonlinear regression model improved prediction accuracy; however, its mathematical complexity limits practical implementation. Among intelligent models, the RBF neural network achieved the lowest RMSE (5.73%). The decision tree model showed a competitive prediction error (8%) while providing an interpretable classification framework, which is further elucidated later in the paper.

Model validation

Model validation represents a critical step in assessing the reliability, predictive capability, and practical applicability of the developed

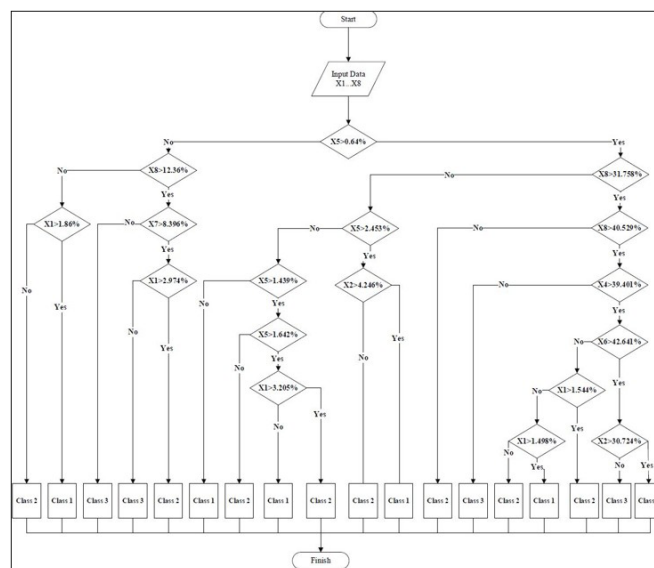


Figure 11—Developed model of SCCDT

Table 7

Quantitative comparison of the developed statistical and intelligent models for predicting coal spontaneous combustion potential

No.	Modelling approach	MSE	RMSE (%)
1	Multiple linear regression (all variables)	239	15.45
2	Multiple nonlinear regression (all variables)	125	11.18
3	Hybrid linear–nonlinear artificial neural network	68.96	8.30
4	Multilayer perceptron (MLP) neural network	67.90	8.38
5	Radial basis function (RBF) neural network	32.80	5.73
6	Decision tree (DT) model	64	8.00

Assessment of coal spontaneous combustion potential using statistical and machine

predictive frameworks for coal spontaneous combustion potential. As previously discussed, the performance of artificial intelligence-based models—including hybrid ANN, MLP, and RBF—was quantitatively evaluated during the modelling phase using standard error metrics.

Among all developed approaches, the decision tree (DT) model demonstrated a highly competitive predictive performance while offering a key advantage in terms of interpretability and transparency, which are essential for practical implementation in mining environments. Therefore, this model was selected for comprehensive validation.

The validation framework was designed using a strict independent testing strategy. Before model development, 25 samples were randomly separated from the original dataset and completely excluded from all training, model configuration, and parameter selection procedures. These samples were used exclusively as an independent internal testing dataset to evaluate the predictive performance of the developed models. In addition, nine independent coal samples collected from the Parvadeh Tabas coal mines were used exclusively for external validation. These field samples were not involved in model training, testing, or parameter adjustment and were only used to assess the generalisation capability of the proposed model under real mining conditions.

Although k-fold cross-validation and repeated cross-validation are commonly used for evaluating machine learning models, the present study adopted an independent hold-out testing strategy combined with external field validation. This approach was selected because the primary objective of this research was to evaluate the capability of the developed models in predicting spontaneous combustion potential for completely unseen coal samples and to assess their applicability under actual mining conditions. Therefore, the reserved testing dataset and independent mine samples provide a more realistic evaluation of model generalisation capability.

Considering the relatively limited size of the available dataset, potential overfitting was carefully considered during model development. To minimise this risk, the independent testing samples were separated before model training and were not used for model configuration or parameter selection. Furthermore, the final model structures were selected based on their predictive performance on unseen data rather than training performance alone. The additional external validation using independent coal samples from the Parvadeh Tabas mines provided further evidence regarding the generalisation capability of the developed framework. However, the authors acknowledge that increasing the number of samples from different coalfields and geological conditions would further improve the reliability and applicability of the proposed models.

The validation results were evaluated through two complementary stages: internal testing using reserved samples and external validation using independent field data. Although the developed models demonstrated satisfactory predictive performance, the uncertainty associated with model predictions should be considered when applying the framework in practical mining environments. The main sources of uncertainty include variability in coal properties, differences in experimental CPT measurements, and limitations related to the available dataset size. Therefore, the validation strategy adopted in this study, including independent testing samples and external field validation, was intended not only to evaluate prediction accuracy but also to examine the reliability of model outputs under unseen conditions.

Internal validation

In the first stage, internal validation was conducted using a subset of the original dataset that had been excluded from model training.

Out of the 25 coal samples reserved for testing, the decision tree model correctly classified 23 samples, with only two misclassifications, resulting in an accuracy of approximately 92%.

This result confirms the robustness and generalisation capability of the proposed decision tree framework, indicating that the model can reliably predict spontaneous combustion potential for unseen data.

External validation using field data

To further evaluate the practical applicability of the model, external validation was performed using independent data collected from the Parvadeh underground coal mines in the Tabas coalfield, Iran.

The Tabas coalfield, located approximately 85 km southeast of Tabas in South Khorasan Province, represents one of the most significant coal reserves in Iran, with an estimated resource of approximately 1.2 billion tonnes. The Parvadeh mining complex consists of six structurally controlled zones separated by major fault systems, as illustrated in Figure 12.

The spatial distribution of the main fault systems is further detailed in Figure 13, which highlights the geological complexity of the mining area.

For validation purposes, intrinsic coal properties from nine samples collected from four Parvadeh mines were used. The dataset includes proximate analysis and petrographic maceral composition, as summarised in Table 8.

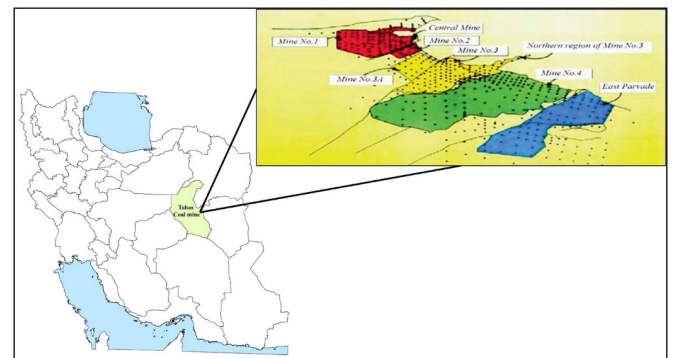


Figure 12—Location of the Parvadeh coal mines and major separating faults

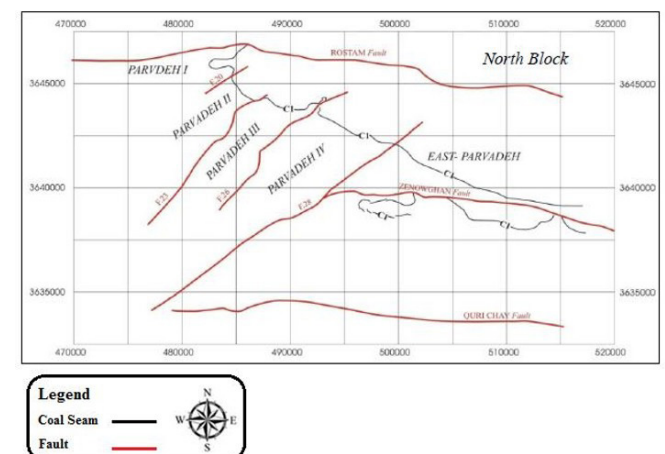


Figure 13—Location of major fault systems in the study area

Assessment of coal spontaneous combustion potential using statistical and machine

Table 8

Intrinsic parameters of nine coal samples from Parvadeh Tabas mines

sample code	moisture (%)	ash (%)	volatile matter (%)	fixed carbon (%)	pyrite (%)	vitrinite (%)	liptinite (%)	inertinite (%)	class of tendency to spon com
S ₁ P ₁ C ₁ W ₃	0.77	27.93	17.38	53.92	0.11	47.31	3.89	48.69	1
S ₂ P ₁ C ₁ E ₃	2.73	36.59	14.94	45.74	0.45	36.17	8.15	55.23	1
S ₃ P ₂ B ₁ B ₂	4.68	33.11	14.97	44.24	1.78	46.22	10.69	41.31	2
S ₄ P ₂ B ₁ B ₃	6.56	32.3	17.65	43.5	4.52	80.23	15.25	0	3
S ₅ P ₂ B ₂ B ₃	3.73	43.95	15.38	36.94	1.53	44.73	5.8	47.94	1
S ₆ P ₃ C ₁ B ₁	7.94	24.38	21.11	46.56	4.66	78.97	16.37	0	3
S ₇ P ₄ C ₁ B ₁	2.56	15.11	23.11	59.21	1.51	39.44	9.25	49.8	2
S ₈ P ₄ B ₂ B ₄	0.87	17.59	24.49	55.99	1.06	39	14	47	2
S ₉ P ₄ C ₁ B ₄	1.57	19.77	25.04	52.26	1.37	39	12	49	3

Table 9

Comparison between experimental and predicted spontaneous combustion classes

Name	The class of tendency to spon com based on the experimental CPT value	The class of tendency to spon com based on the predicted CPT value
S ₁ P ₁ C ₁ W ₃	weak	weak
S ₂ P ₁ C ₁ E ₃	weak	weak
S ₃ P ₂ B ₁ B ₂	moderate	moderate
S ₄ P ₂ B ₁ B ₃	high	high
S ₅ P ₂ B ₂ B ₃	weak	weak
S ₆ P ₃ C ₁ B ₁	high	high
S ₇ P ₄ C ₁ B ₁	moderate	moderate
S ₈ P ₄ B ₂ B ₄	moderate	moderate
S ₉ P ₄ C ₁ B ₄	high	high

The selected samples represent different levels of spontaneous combustion tendency (low, moderate, and high), providing a suitable basis for evaluating model performance under real mining conditions.

Model performance and comparison

The comparison between experimentally determined CPT-based classes and the predicted classes obtained from the decision tree model is presented in Table 9.

As shown in Table 8, the decision tree model successfully predicted the spontaneous combustion tendency class for all nine coal samples, achieving 100% prediction accuracy for the external validation dataset.

This result demonstrates the strong generalisation capability of the model when applied to completely independent field data. Moreover, it confirms that the decision tree model can effectively capture the governing relationships between intrinsic coal properties and spontaneous combustion behaviour.

Practical implications

Beyond its predictive accuracy, the decision tree model offers significant practical advantages. Unlike complex black-box models

such as neural networks, the decision tree provides **explicit decision rules** that can be easily interpreted and applied by mining engineers.

The ability to directly relate parameters such as pyrite and inertinite content to spontaneous combustion risk enables rapid and cost-effective preliminary assessment in underground coal mining operations.

Discussion

Nonlinear behaviour of coal spontaneous combustion potential

The present study demonstrated that coal spontaneous combustion potential is governed by complex nonlinear interactions among multiple intrinsic coal properties rather than by individual controlling factors. The pairwise regression analyses showed relatively low coefficients of determination ($R^2 \approx 0.04\text{--}0.60$), indicating that no single parameter can sufficiently describe the variation in CPT values. This finding is consistent with the physicochemical nature of coal oxidation, where moisture, mineral matter, maceral composition, and reactive components simultaneously influence oxygen adsorption, heat generation, and

Assessment of coal spontaneous combustion potential using statistical and machine

thermal accumulation.

The limited performance of single-variable models highlights the importance of considering interactions among multiple parameters. Although individual factors such as pyrite content have strong theoretical relevance due to their catalytic effect on oxidation reactions, their influence is strongly dependent on the surrounding coal matrix and environmental conditions.

Comparison of statistical and intelligent modelling approaches

The comparison of developed models indicates that different approaches provide different advantages depending on the intended application. The quantitative comparison presented in Table 7 confirms that increasing model complexity improves the capability to capture nonlinear relationships among intrinsic coal properties and spontaneous combustion potential. Classical statistical models provide valuable insight into parameter–response relationships and are useful for identifying general trends; however, their predictive capability decreases when multiple interacting variables and nonlinear mechanisms are involved.

The nonlinear regression model improved prediction performance compared with linear regression, achieving an R^2 value of approximately 0.82. Nevertheless, the resulting high-order mathematical formulation reduced model transparency and limited its direct application in routine mine safety assessments.

Intelligent models demonstrated higher flexibility in representing complex nonlinear interactions. The RBF neural network achieved the lowest prediction error (approximately 5.7%), indicating its strong approximation capability for coal oxidation-related processes. However, similar to other neural network-based approaches, the internal relationship between input variables and final predictions remains difficult to interpret.

The decision tree model provided a balance between predictive accuracy and practical applicability. Although its numerical error was slightly higher than that of the RBF model, it generated explicit classification rules that can be directly interpreted by mining engineers. Therefore, the selection of an appropriate model depends on the intended objective: RBF networks are preferable when maximum prediction accuracy is required, whereas decision tree models are advantageous for operational decision support where transparency and interpretability are essential.

Physical interpretation of dominant parameters

The decision tree analysis identified pyrite and inertinite contents as the dominant parameters controlling spontaneous combustion classification. This finding agrees with previous experimental observations.

The selected intrinsic parameters represent different physical and chemical mechanisms involved in coal self-heating. Moisture content has a dual influence on spontaneous combustion behaviour: a certain amount of moisture can facilitate oxidation reactions and oxygen adsorption, whereas excessive moisture may consume heat through evaporation and delay temperature accumulation. Ash content generally reduces spontaneous combustion susceptibility by acting as an inert thermal component that absorbs part of the generated oxidation heat. However, the influence of ash depends on its mineral composition and interaction with reactive coal components.

Volatile matter and fixed carbon contents are closely related to coal rank and chemical reactivity. Higher volatile matter generally indicates a greater amount of reactive organic compounds, which

can enhance oxygen adsorption and oxidation rates. In contrast, fixed carbon represents the relatively stable carbon fraction, and its effect on spontaneous combustion is influenced by the balance between combustible carbon content and coal structure. Petrographic components also play a fundamental role in oxidation behaviour. Vitrinite and liptinite are considered more reactive macerals due to their higher oxygen affinity and chemical activity, whereas inertinite generally exhibits lower reactivity because of its more condensed aromatic structure and reduced oxidation potential.

Pyrite is one of the most important mineralogical factors affecting spontaneous combustion because its oxidation reactions are highly exothermic and can accelerate heat generation during the early stages of self-heating. Therefore, the identification of pyrite as a dominant parameter by the decision tree model is consistent with established coal oxidation mechanisms.

Pyrite accelerates coal oxidation through catalytic reactions and promotes heat generation during low-temperature oxidation. However, its effect is influenced by other coal characteristics, explaining why pyrite alone cannot fully determine spontaneous combustion potential.

Inertinite content also plays an important role due to its relatively low reactivity compared with vitrinite and liptinite. The interaction between reactive macerals, mineral components, and oxidation conditions explains why spontaneous combustion behaviour cannot be attributed to a single parameter.

The use of proximate analysis parameters in this study was primarily motivated by their practical availability in mining operations. Although elemental analysis may provide additional chemical information, its application is often limited by data availability from operating mines. Therefore, the proposed framework provides a balance between prediction accuracy and practical accessibility.

Practical applicability of the proposed framework

The practical applicability of a predictive model is a critical consideration for mine safety management. Unlike complex black-box models that require extensive computational resources and specialised interpretation, the proposed SCCDT model provides a simple and transparent framework that can be directly implemented in underground coal mining operations.

The input variables required by the model are limited to intrinsic coal characteristics obtained from routine laboratory analyses, including moisture, ash, volatile matter, fixed carbon, pyrite, and maceral composition. These parameters are commonly measured during coal quality assessment and mine characterisation programmes; therefore, the proposed approach does not require additional costly experimental procedures or advanced monitoring infrastructure.

From a computational perspective, the SCCDT model has low processing requirements and can be implemented using standard computer systems without the need for high-performance computing facilities. Once the decision rules are established, the spontaneous combustion tendency of newly encountered coal seams can be rapidly classified based on their intrinsic characteristics.

The proposed framework can also be integrated with mine monitoring and decision-support systems. For example, laboratory-based coal characterisation data can be periodically updated, while real-time monitoring information such as temperature, oxygen concentration, and ventilation conditions can be incorporated as

Assessment of coal spontaneous combustion potential using statistical and machine

complementary information in future extended versions of the model. Such integration may enable the development of more comprehensive early-warning systems for spontaneous combustion risk management.

It should be emphasised that the SCCDT model is intended as a practical screening and decision-support tool rather than a replacement for continuous mine monitoring systems. Its primary purpose is to provide an initial assessment of spontaneous combustion potential, assist engineers in identifying high-risk coal seams, and support the prioritisation of prevention and control strategies during mine planning and operation.

Limitations and future research directions

Despite the promising results, several limitations should be acknowledged. First, although the compiled database of 165 coal samples represents a substantial effort, the dataset size remains limited compared with the complexity of coal spontaneous combustion phenomena. Therefore, the developed models should not be considered universally applicable under all geological and operational conditions.

In addition, the relatively limited dataset size may increase the risk of overfitting, particularly for complex intelligent models with higher flexibility. To mitigate this issue, the model development process was conducted using independent testing data that were separated before training and were not involved in model configuration or parameter selection. Moreover, the final model evaluation included external validation using independent coal samples from the Parvadeh Tabas mines, providing an additional assessment of model generalisation capability. Nevertheless, further validation using larger datasets covering diverse coal types and geological conditions is recommended to ensure broader applicability of the developed framework.

Second, the external validation dataset consisted of only nine samples from the Parvadeh coal mines. Although the proposed model achieved 100% classification accuracy for these samples, this result should be interpreted cautiously and does not indicate perfect prediction capability. Increasing the number of independent field samples from different mining regions is necessary to provide more comprehensive validation and to evaluate the transferability of the developed framework.

Third, uncertainties may arise from variations in coal properties, experimental CPT measurements, geological conditions, and mining practices. These uncertainties may influence both the input space and the resulting classification boundaries of data-driven models. Therefore, the predicted spontaneous combustion classes should be interpreted as risk estimations rather than absolute deterministic outcomes. Future studies should therefore incorporate uncertainty quantification approaches, probabilistic modelling techniques, and uncertainty-aware machine learning frameworks to estimate prediction confidence and improve the reliability of risk assessments.

Future research should also focus on expanding the database with additional coal types and mining environments, integrating real-time monitoring data from sensor-based systems, and developing hybrid modelling strategies that combine machine learning techniques with mechanistic understanding of coal oxidation processes. Such approaches may further enhance prediction accuracy, reduce uncertainty, and improve practical applicability in underground coal mine safety management.

Conclusions

Coal spontaneous combustion is a complex and nonlinear phenomenon influenced by the combined effects of intrinsic coal properties, geological conditions, and mining-related factors. Among these, intrinsic coal characteristics represent the primary intrinsic factors controlling coal oxidation and self-heating behaviour. In this study, a comprehensive data-driven framework was developed to evaluate and predict coal spontaneous combustion potential using a database consisting of 165 coal samples compiled from published experimental studies and CPT measurements.

A systematic comparison of classical statistical and intelligent modelling approaches demonstrated that the relationship between intrinsic coal properties and spontaneous combustion potential is predominantly nonlinear and multivariate. Univariate and linear regression models showed limited predictive capability, confirming that individual parameters cannot adequately represent the coupled interactions governing coal self-heating processes. Although nonlinear regression and neural network-based models improved prediction accuracy, their mathematical complexity or limited interpretability may restrict their direct application in practical mine safety management.

Among the developed approaches, the proposed spontaneous combustion classification decision tree (SCCDT) model provided an effective balance between predictive performance, simplicity, and interpretability. The model achieved an overall testing error of approximately 8% using 25 independent test samples that were completely excluded from the training process. Furthermore, external validation using independent coal samples from the Parvadeh Tabas coal mines achieved correct classification for all nine samples, demonstrating the applicability of the proposed framework for independent field data.

The scientific contribution of this research is not based on proposing a new machine learning algorithm; rather, it lies in developing an integrated and practical framework that combines comprehensive data compilation, comparative evaluation of multiple statistical and intelligent modelling approaches, and interpretable prediction for mining applications. The SCCDT model identified pyrite and inertinite contents as dominant parameters influencing spontaneous combustion classification, which is consistent with the established physicochemical mechanisms of coal oxidation. The transparent structure of the model provides an important advantage over black-box approaches and enhances its suitability for preliminary risk assessment and decision-support in underground coal mining operations.

Despite the promising results, several limitations should be acknowledged. Although the database compiled in this study represents a substantial effort, the number of available samples remains limited compared with the complexity and variability of coal spontaneous combustion phenomena. Therefore, the developed models should be considered practical predictive tools within the range of coal characteristics represented in the dataset rather than universally applicable models for all geological and mining conditions. In addition, external validation was conducted using nine coal samples from the Parvadeh Tabas coal mines, representing a specific regional geological setting. Although these samples confirmed the applicability of the SCCDT model, additional validation using larger datasets from different coalfields is required to further evaluate model generalisation and minimise potential regional bias. Future studies should focus on expanding the database, incorporating uncertainty quantification methods,

Assessment of coal spontaneous combustion potential using statistical and machine

developing hybrid physics-informed machine learning approaches, and integrating real-time monitoring data to enhance the reliability and operational applicability of spontaneous combustion prediction systems.

Overall, this study demonstrates that combining statistical analysis with interpretable intelligent modelling provides a practical and reliable framework for assessing coal spontaneous combustion potential. The proposed SCCDT approach can serve as an effective decision-support tool during mine planning, operation, and safety management, while providing a foundation for future hybrid and uncertainty-aware predictive models.

Data availability

The datasets generated and/or analysed during the current study are not publicly available because they are part of ongoing research and may be used in future publications. However, the data are available from the corresponding author upon reasonable request.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Ahamed, S., Monir, M.U., Biswas, P.K., Khan, A.A. 2016. Investigation the risk of spontaneous combustion in Barapukuria coal mine, Dinajpur, Bangladesh. *Journal of Geoscience and Environment Protection*, vol. 4, no.4, pp.74–79.
- Akgün, F., Arisoy, A. 1994. Effect of particle size on the spontaneous heating of a coal stockpile. *Combustion and Flame*, vol. 99, no.1, pp.137–146.
- Banerjee, S.C. 1985. Spontaneous combustion of coal and mine fires.
- Beamish, B., Beamish, R. 2010. Benchmarking moist coal adiabatic oven testing. Aziz, N.(ed.), pp.264–268.
- Beamish, B.B., Arisoy, A. 2008. Effect of mineral matter on coal self-heating rate. *Fuel*, vol. 87, no. 1, pp.125–130.
- Beamish, B.B., Sainsbury, W. 2008, February. Development of a site specific self-heating rate prediction equation for a high volatile bituminous coal. In *Proceedings 8th Coal Operator's Conference* pp. 161–165.
- Chandra, D., Prasad, Y.V.S. 1990. Effect of coalification on spontaneous combustion of coals. *International Journal of Coal Geology*, vol 16, no. 1-3, pp.225–229.
- Cotterell, M.E. 1997. The effect of coal maceral composition on spontaneous combustion (Doctoral dissertation, University of Queensland).
- Deng, J., Chen, W., Wang, C., Wang, W. 2021. Prediction model for coal spontaneous combustion based on SA-SVM. *ACS omega*, vol. 6, no. 17, pp.11307–11318.
- Demir, B., Ören, Ö., Şensöğüt, C. 2020. Investigation of the effect of reactor size on spontaneous combustion properties of coals with different coalification degrees. *Arabian Journal of Geosciences*, vol. 13, no. 15, pp.730.
- Feng, K.K., Chakravorty, R.N., Cochrane, T.S. 1973. Spontaneous combustion-coal mining hazard. *CIM BULLETIN*, vol. 66, 738, pp.75–84.
- Gbadamosi, A.R., Onifade, M., Genc, B., Rupprecht, S. 2021. Spontaneous combustion liability indices of coal. *Combustion Science and Technology*, vol. 193, no. 15, pp.2659–2671.
- Guney, M. 1968. Oxidation and spontaneous combustion of coal: Review of individual factors. *Colliery Guardian*, vol. 216, no. 105-110, pp.137–143.
- Hodges, D.J. 1963. The influence of moisture on the spontaneous heating of coal (Doctoral dissertation, University of Nottingham).
- Kaymakci, E., Didari, V. 2002. Relations between coal properties and spontaneous combustion parameters. *Turkish Journal of Engineering and Environmental Sciences*, vol. 26, no.1, pp.59–64.
- Kursunoglu, N., Gogebakan, M. 2022. Prediction of spontaneous coal combustion tendency using multinomial logistic regression. *International Journal of Occupational Safety and Ergonomics*, vol. 28, no. 4, pp.2000–2009.
- Lang, L., Fu-Bao, Z. 2010. A comprehensive hazard evaluation system for spontaneous combustion of coal in underground mining. *International journal of coal geology*, vol. 82, no. 1-2, pp.27–36.
- Lei, C., Deng, J., Cao, K., Ma, L., Xiao, Y., Ren, L. 2018. A random forest approach for predicting coal spontaneous combustion. *Fuel*, 223, pp.63–73.
- Liu, W., Qin, Y. 2017. Multi-physics coupling model of coal spontaneous combustion in longwall gob area based on moving coordinates. *Fuel*, vol. 188, pp.553–566.
- Meshkov, S. 2019. Determination of the Parameters of Technological Schemes for the Development of Coal Seams Prone to Spontaneous Combustion using Computer Simulation. In *E3S Web of Conferences* vol. 105, pp. 01016. EDP Sciences.
- Michaylov, M. 2002. Expert system for assessment of risk from spontaneous combustion. *Mining and Mineral Processing*, pp.44–45.
- Mishra, A., Gupta, S.K. 2024. Intelligent classification of coal seams using spontaneous combustion susceptibility in IoT paradigm. *International Journal of Coal Preparation and Utilization*, vol. 44, no. 7, pp.757–779.
- Mitchell, P. 2003. Controlling and reducing heat on long-wall faces. *coal*, pp.234–245.
- Nugroho, Y.S., McIntosh, A., Gibbs, B.M. 2000. Low-temperature oxidation of single and blended coals. *Fuel*, vol 79, no. 15, pp.1951–1961.
- Onifade, M., Genc, B. 2018a. Modelling spontaneous combustion liability of carbonaceous materials. *International Journal of Coal Science & Technology*, vol. 5, no. 2, pp.191–212.
- Onifade, M., Genc, B. 2018b. Spontaneous combustion of coals and coal-shales. *International Journal of Mining Science and Technology*, vol. 28, no. 6, pp.933–940.
- Onifade, M., Genc, B. 2020. A review of research on spontaneous combustion of coal. *International Journal of Mining Science and Technology*, vol. 30, no. 3, pp.303–311.
- Parr, S.W., Hilgard, E.R. 1925. Oxidation of Sulfur as a Factor in Coal Storage. *Industrial & Engineering Chemistry*, vol. 17, no. 2, pp.117–118.

Assessment of coal spontaneous combustion potential using statistical and machine

- Saffari, A., Sereshki, F., Ataei, M., Ghanbari, K. 2017. Presenting an engineering classification system for coal spontaneous combustion potential. *International Journal of Coal Science & Technology*, vol. 4, no. 2, pp.110–128.
- Saffari, A., Ataei, M., Sereshki, F. 2019. Evaluation of the spontaneous combustion of coal (SCC) by using the R70 test method based on the correlation among intrinsic coal properties (Case study: Tabas Parvadeh coal mines, Iran). *Rudarsko-geološko-naftni zbornik*, vol. 34, no. 3.
- Saffari, A., Sereshki, F., Ataei, M. 2020. A comprehensive study on the effect of moisture content on coal spontaneous combustion tendency. *Iranian Journal of Earth Sciences*, vol. 12, no. 3, pp.194–204.
- Sahu, H.B., Mahapatra, S.S. 2013. Forecasting spontaneous heating susceptibility of Indian coals using neuro fuzzy system. *Geotechnical and Geological Engineering*, vol. 31, no. 2, pp.683–697.
- Said, K.O., Onifade, M., Lawal, A.I. and Githiria, J.M., 2021. An artificial intelligence-based model for the prediction of spontaneous combustion liability of coal based on its proximate analysis. *Combustion Science and Technology*, vol. 193, no. 13, pp.2350–2367.
- Şensöğüt, C. 2011. Spontaneous combustion related fire ratios. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, vol. 5, no. 1.
- Smith, M.A., Glasser, D. 2005. Spontaneous combustion of carbonaceous stockpiles. Part II. *Factors affecting the rate of the low-temperature oxidation reaction*. *Fuel*, vol. 84, no. 9, pp.1161–1170.
- Smith, A.C., Lazzara, C.P. 1987. Spontaneous combustion studies of US coals (Vol. 9079). US Department of the Interior, Bureau of Mines.
- Song, Z., Kuenzer, C. 2014. Coal fires in China over the last decade: A comprehensive review. *International Journal of Coal Geology*, vol. 133, pp.72–99.
- Taraba, B., Michalec, Z. 2011. Effect of longwall face advance rate on spontaneous heating process in the gob area—CFD modelling. *Fuel*, vol. 90, no. 8, pp.2790–2797.
- Tosun, A. 2017. Development of a technology to prevent spontaneous combustion of coal in underground coal mining. *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 117, no. 12, pp.1133–1138.
- Misra, B.K., Singh, B.D. 1994. Susceptibility to spontaneous combustion of Indian coals and lignites: an organic petrographic autopsy. *International Journal of Coal Geology*, vol. 25, no. 3–4, pp.265–286.
- Wang, C., Du, Y., Deng, Y., Zhang, Y., Deng, J., Zhao, X., Duan, X. 2023. Study on spontaneous combustion characteristics and early warning of coal in a deep mine. *Fire*, vol. 6, no. 10, pp.396.
- Wang, H., Dlugogorski, B.Z., Kennedy, E.M. 2003. Coal oxidation at low temperatures: oxygen consumption, oxidation products, reaction mechanism and kinetic modelling. *Progress in energy and combustion science*, 29, no. 6, pp.487–513.
- Wang, L., Tong, R. 2020, March. An Improved Evaluation of Comprehensive for Mine Fire Risk. In *IOP Conference Series: Materials Science and Engineering* vol. 782, no. 4, pp. 042036. IOP Publishing.
- Więckowski, M., Howaniec, N., Smoliński, A. 2020. Assessment of Heating and Cooling of a Spontaneous Fire Source in Coal Deposits—Effect of Coal Grain Size. *Minerals*, vol. 10, no. 10, pp.907.
- Wu, Y., Zhang, Y., Wang, J., Zhang, X., Wang, J., Zhou, C. 2020. Study on the effect of extraneous moisture on the spontaneous combustion of coal and its mechanism of action. *Energies*, vol. 13, no. 8, pp.1969.
- Xia, T., Wang, X., Zhou, F., Kang, J., Liu, J., Gao, F. 2015. Evolution of coal self-heating processes in longwall gob areas. *International Journal of Heat and Mass Transfer*, vol. 86, pp.861–868.
- Xu, X., Zhang, F. 2023. Evaluation and optimization of multi-parameter prediction index for coal spontaneous combustion combined with temperature programmed experiment. *Fire*, vol. 6, no. 9, pp.368.
- Yang, F., Lai, Y., Song, Y. 2019. Determination of the influence of pyrite on coal spontaneous combustion by thermodynamics analysis. *Process Safety and Environmental Protection*, 129, pp.163–167.
- Yetkin, M.E., Özfirat, M.K., Kun, M., Pamukcu, C. 2024. The prediction of the risks of spontaneous combustion in underground coal mines using a fault tree analysis method. *MethodsX*, vol. 13, pp.102835.
- Yuan, L., Smith, A.C. 2012. The effect of ventilation on spontaneous heating of coal. *Journal of Loss Prevention in the Process Industries*, vol. 25, no. 1, pp.131–137.
- Zhang, L.L., Qin, X.U. 2009. Analysis and key control technologies to prevent spontaneous coal combustion occurring at a fully mechanized caving face with large obliquity in deep mines. *Mining Science and Technology China*, vol. 19, no. 4, pp.446–451.
- Zhang, L., Song, Z., Wu, D., Luo, Z., Zhao, S., Wang, Y., Deng, J. 2022. Prediction of coal self-ignition tendency using machine learning. *Fuel*, 325, p.124832.
- Zhang, Y., Dong, J., Guo, F., Chen, X., Wu, J., Miao, Z. 2018. Experimental study on the effects of drying methods on the stabilities of lignite. *Chinese Journal of Chemical Engineering*, vol. 26, no. 7, pp.1545–1554.
- Zhao, H., Geng, X., Yu, J., Xin, B., Yin, F., Tahmasebi, A. 2016. Effects of drying method on self-heating behavior of lignite during low-temperature oxidation. *Fuel Processing Technology*, vol. 151, pp.11–18.