Journal of Applied Geology, vol. 10(1), 2025, pp. 25–41
DOI: http://dx.doi.org/10.22146/jag.104595

Unsupervised Machine Learning for Determining Exploration Areas of Valuable Elements and Potential Toxicology Elements: A Case Study of the Bowen Basin Coal, Australia

Fajri Zakka Addintamma[1], Adzani Nareswari Amaranggana[1], Anindya Ayu Kusuma[1], Aviliana[1], Mochammad Wildanun Solikh[1], Aulia Agus Patria[2], and Ferian Anggara*[1,2]

[1]*Department of Geological Engineering, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[2]*Unconventional Georesources Research Group, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia*

Received: February 10, 2025 | Accepted: August 11, 2025 | Published online: August 31, 2025

ABSTRACT. Global coal production and demand have increased annually increased. In addition to its potential as an alternative source of critical elements, coal also has environmental risks through toxicology elements. Australia is the world's second-largest producer of Rare Earth Elements (REEs) and critical elements, making coal exploration a key focus of the country's mining strategy. Unsupervised Machine learning algorithm (CCA, PCC, PCA, two-ways HCA) and was applied to 56 coal samples from three pits in Bowen Basin, e.g., Blake Central Pit, Blake West Pit, and Bowen No. 2 Pit, to correlate trace elements with the geochemical characteristics of coal, such as proximate and major oxides. These methods are aimed at finding the factors that control the geochemistry of coal in the study area, especially to characterize the coal based on its potential valuable and toxic elements and determining the trend of element distribution (IDW). Blake West Pit is enriched in Ba, Br, and Sr, which associated with inherent moisture and phosphor (P), extending SE-trend. Blake Central Pit and Blake West Pit are enriched in Hf, Mo, Ta, Th, Y, and REY, which are associated with ash and major elements such as Si, Al, Ti, and K, with a trend of potential exploration towards N-NW. However, both pits show the risk of contamination from the toxic element Zn, which is associated with volatile matter, and major elements e.g., Fe, Mg, and Mn, with a trend of distribution towards S-SW. Correlation analysis and regional geology, suggest trace element enrichment in Bowen Basin is controlled by two main factors: 1) the transgressive phase during Early-Late Permian, which enriched inherent moisture, P, Ba, Br, and Sr, and 2) volcanic activity during Early Permian, which enriched silicate minerals and elements such as Hf, Ta, Th, W, and REY. Unsupervised machine learning has proven effective for preliminary coal characterization to support further exploration, such as identifying the sources of elements and the geological factors that control the coal characteristics. The results were obtained from PCC (to find correlations between each pair of variables), PCA (to identify the components that contribute to the correlations), and two-way HCA (to characterize coal based on variable values and sample location).

Keywords: Bowen Basin · Coal geochemistry · Trace elements · Unsupervised machine learning.

*Corresponding author: F. ANGGARA, Department of Geological Engineering, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia. Jl. Grafika No. 2 Yogyakarta 55281. E-mail: ferian@ugm.ac.id

1 INTRODUCTION

Coal as an alternative source for non-conventional elements such as critical elements, has shown a consistent increase in production and global demand annually (Hu *et al.*, 2021). However, the use of coal is still associated with negative health impacts, including endemic diseases such as fluorosis, arsenosis, selenosis, and lung cancer, caused by toxic elements like arsenic (As), cobalt (Co), chromium (Cr), nickel (Ni), selenium (Se), and zinc (Zn) found in coal composition (Dai *et al.*, 2012).

Australia, as the world's second-largest producer of Rare Earth Elements (REEs) and critical elements, has identified 24 elements as critical, focusing on these in exploration activities (Queensland Government-Department of Resources, 2021). Coal exploration, particularly in Queensland's Bowen Basin, extending from Collinsville to New South Wales, plays a key role in the country's mining strategy and policy. Bowen Basin not only produces coal for metallurgical and thermal purposes, but also exhibits significant variations in characteristics, age, and geological conditions (Queensland Government-Department of Resources, 2021).

However, to date, in-depth studies on the relationship between valuable elements and toxicological elements with characteristics and distribution of coal remain limited. Thus, this study aims to identify the geochemical characteristics of coal through the use of Unsupervised Machine Learning. Therefore, this study aims to investigate the geochemical characteristics of coal in the Bowen Basin by applying unsupervised machine learning. The goal is to identify the controlling factors and spatial distribution of both valuable and toxic elements in the coal. Additionally, spatial distribution patterns of these geochemical characteristics, as well as geological modelling to generate targeted exploration areas. Such a context, preventive strategy advice to address environmental impacts around the Bowen Basin, Australia and to support more sustainable coal resource management as well as minimizing negative environmental impacts. Such a context highlights the need for providing preventive strategic recommendations to address environmental impacts in the Bowen Basin, Australia. The findings of this study are expected to support more sustainable coal resource management and contribute to the mitigation of negative environmental effects in the region.

2 GEOLOGICAL BACKGROUND

The Bowen Basin is located in Queensland, Australia (Figure 1), with a complex geological history. This basin contains sediments from the Permian to Triassic periods, with thicknesses reaching up to 9,000 meters in certain locations, such as the Taroom and Denison Troughs (Figure 2). Intense tectonic activity has created various geological features, including the Dawson Fold Zone, Nebo Syncline, and several major faults such as the Hutton-Wallumbilla Fault. Sedimentation in this basin began in the Early Permian, characterized by fluvial, lacustrine deposits, and coal cycles. Deposition continued until the Late Triassic. In the Late Permian, tectonic loading accelerated the deposition of thick marine sediments, which are highly significant for Coal Seam Gas (CSG) exploration. The final stage of deposition occurred in the Late Triassic, followed by extensive erosion, paving the way for sedimentation from the Surat Basin (Queensland Government *et al.*, 2021).

The formation environment of the Blake Seam is in a swampy area with a high ash content, low sulphur content (Blake West 0.24%, Blake Central 0.6%), and minimal sulphide minerals such as pyrite. Based on previous sedimentological studies, the Blake Seam is interpreted as a result of fluvial deposition (Martini and Johnson, 1987). Meanwhile, the Bowen Seam has a lower ash yield compared to the Blake Seam, indicating a lower mineral content. Additionally, this seam exhibits higher sulfur content, particularly in proximity to the intrusive body, when compared to the Blake Seam. From a depositional environment perspective, the Bowen Seam formed under conditions similar to a swamp environment, where repeated cycles of drying and oxidation occurred (Brakel *et al.*, 2009). The absence of sedimentary rock layers indicates that no significant sedimentation processes took place, supporting the conclusion that the coal in this seam has a lower mineral content. The lack of sedimentary rock layers in this section indicates limited depositional activity, which is

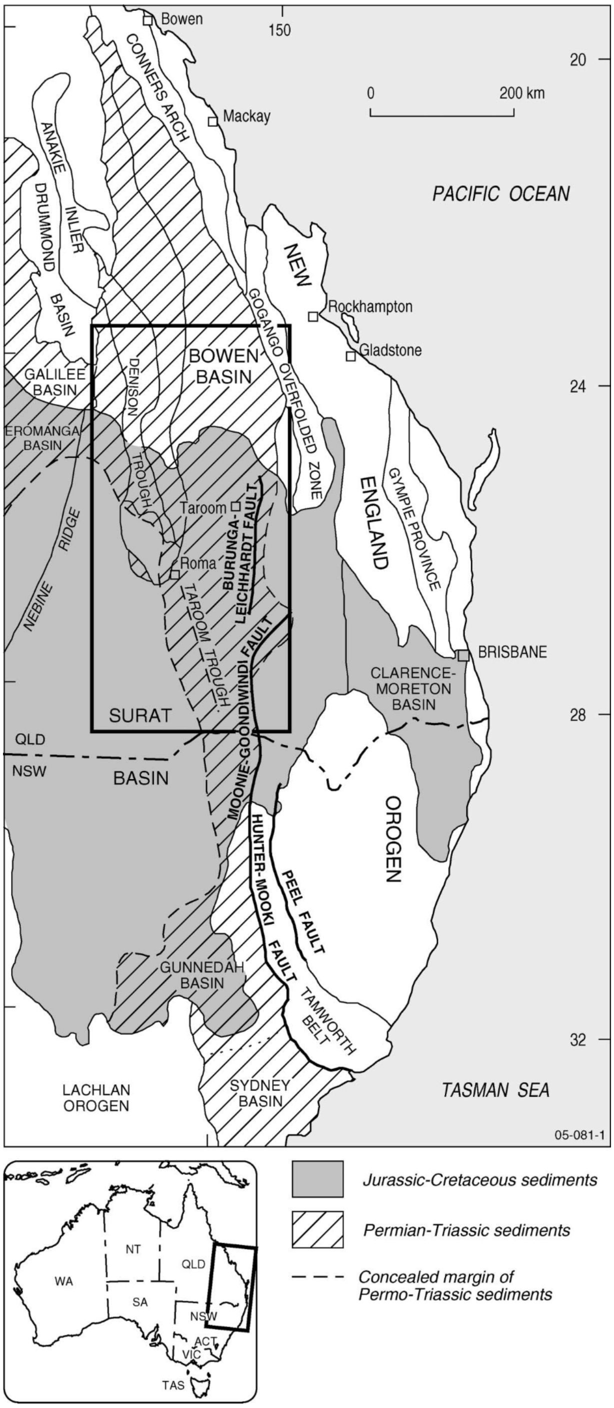

FIGURE 1. Location map of the Bowen Basin in Queensland (Brakel *et al.*, 2009).

potentially linked to the lower mineral content identified in the associated coal seam.

3 RELEVANT THEORY

3.1 Valuable and toxicology elements in coal

Coal is utilized as an alternative source for critical elements (Seredin and Dai, 2012), with some elements considered critical such as neodymium (Nd), terbium (Tb), dysprosium (Dy), yttrium (Y), europium (Eu), and erbium (Er). In contrast, elements like lanthanum (La), praseodymium (Pr), samarium (Sm), and gadolinium (Gd) are considered non-critical, while elements such as cerium (Ce), holmium (Ho), thulium (Tm), ytterbium (Yb), and lutetium (Lu) are deemed excessive. One country advancing coal as a source of both conventional and non-conventional resources is Australia, the world's second-largest producer of Rare Earth Elements (REE). According to Australia's 2019 policy, the strategy to meet the demand for critical elements is supported by coal extraction.

With the increasing scarcity and high demand for critical metals, coal has emerged as a potential source for these elements. Historically, uranium was extracted from coal, and currently, germanium, selenium, and vanadium are mined from coal ash. Pilot plants and laboratory technologies have been developed for the extraction of gallium, aluminum, and rare earth elements, among others. Coal ash also holds promise as a source of valuable elements such as PGE, gold, silver, lithium, and others, with significant potential for various applications (Dai & Finkelman, 2018).

The extraction of elements from coal can provide benefits in terms of resources, economy, and the environment (Hu *et al.*, 2021). However, it has been noted that health hazards from exposure to coal dust, which is rich in elements that can be harmful to human health, are significant (Dai *et al.*, 2012). The ash produced from coal combustion can lead to diseases such as endemic fluorosis, arsenosis, selenosis, and lung cancer. The toxicology elements in coal present include As, Se, F, Co, Cr, Cu, Zn, V, and Ni (Dai *et al.*, 2012). According to World Health Organization (WHO) guidelines, the toxic threshold limits for drinking water are as follows: As - 0.01 ppm, Se - 0.04 ppm, F - 1.5 ppm, Co - 2 ppm, Cr - 0.05 ppm, Cu - 2 ppm, Zn - 3 ppm, and Ni - 0.07 ppm.

3.2 Evaluation Concentration

The utilization prospect of Rare Earth Elements and Yttrium (REY) in coal and coal ash can be assessed through several parameters, including metal resource potential, extraction rate, envi-

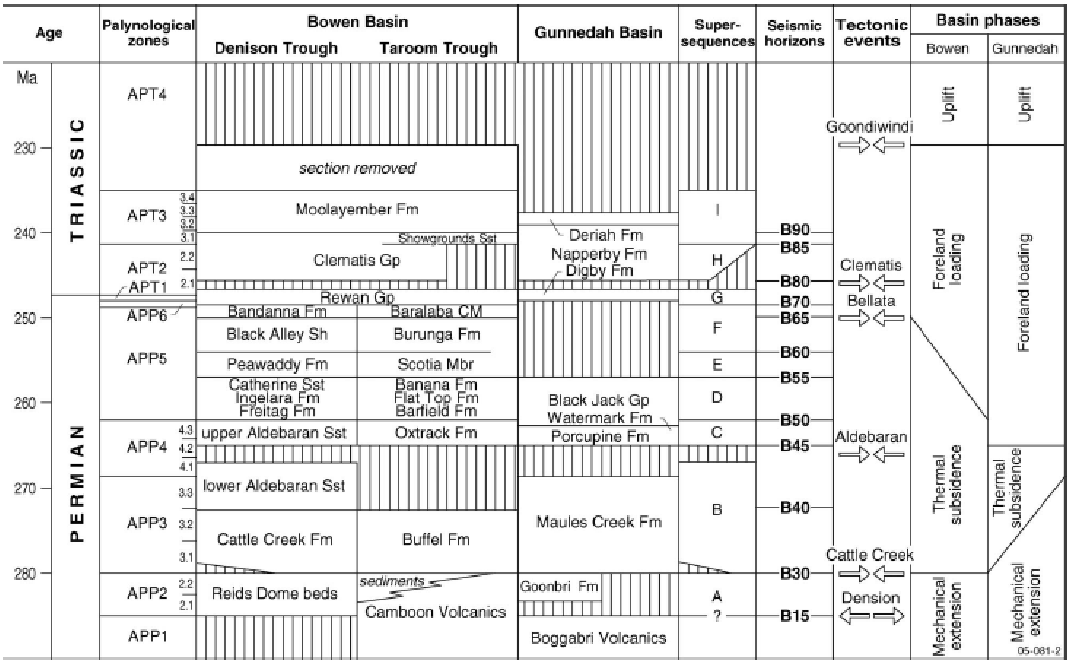

FIGURE 2. Stratigraphy of the Bowen Basin (Brakel *et al.*, 2009).

ronmental and human health impact, and radioactive properties that influence these aspects (Seredin and Dai, 2012). The evaluation of element concentration in coal can be performed using the Concentration Coefficient (CC). The CC represents the ratio of the element's concentration in the investigated coal to its average concentration in world hard coals. Based on this CC value (Dai *et al.*, 2015), the enrichment level of an element can be classified into six categories: depleted (CC < 0.5), normal (0.5 $\leq$ CC < 2), slightly enriched (2 $\leq$ CC < 5), enriched (5 $\leq$ CC < 10), significantly enriched (10 $\leq$ CC < 100), and anomalously enriched (CC $\geq$ 100).

3.3 Unsupervised Machine Learning

Unsupervised learning is a type of machine learning where the algorithm learns from unlabeled data without any predefined outputs or target variables. This means that the data does not have any pre-existing labels or categories. The goal of unsupervised learning is to discover patterns and relationships in the data without any explicit guidance. It involves training a machine using information that is neither classified nor labeled, allowing the algorithm to act on that information without guidance (Wu *et al.*, 2021).

Unsupervised machine learning includes various techniques for discovering patterns and structures in data without predefined labels. Key methods include clustering (e.g., K-Means, DBSCAN, Hierarchical Clustering), dimensionality reduction (e.g., PCA, t-SNE, Autoencoders), and Canonical Correlation Analysis (CCA), which explores relationships between two sets of variables. Other techniques such as association rule learning, anomaly detection, Self-Organizing Maps (SOMs), and generative models (e.g., GANs, VAEs) are used for pattern recognition, anomaly identification, and generating new data (Usama *et al.*, 2019).

4 METHODOLOGY

The methodology consists of eight steps, initiated by data collection, followed by data cleaning (eliminating non-significant values), data pre-processing (standardizing the data), four separate machine learning analyses, and a final step involving spatial distribution analysis using Inverse Distance Weighting (IDW). While each individual machine learning method is capable of providing correlations, this study integrated four different approaches to obtain more meaningful insights. The workflow is summarized in Figure 3.

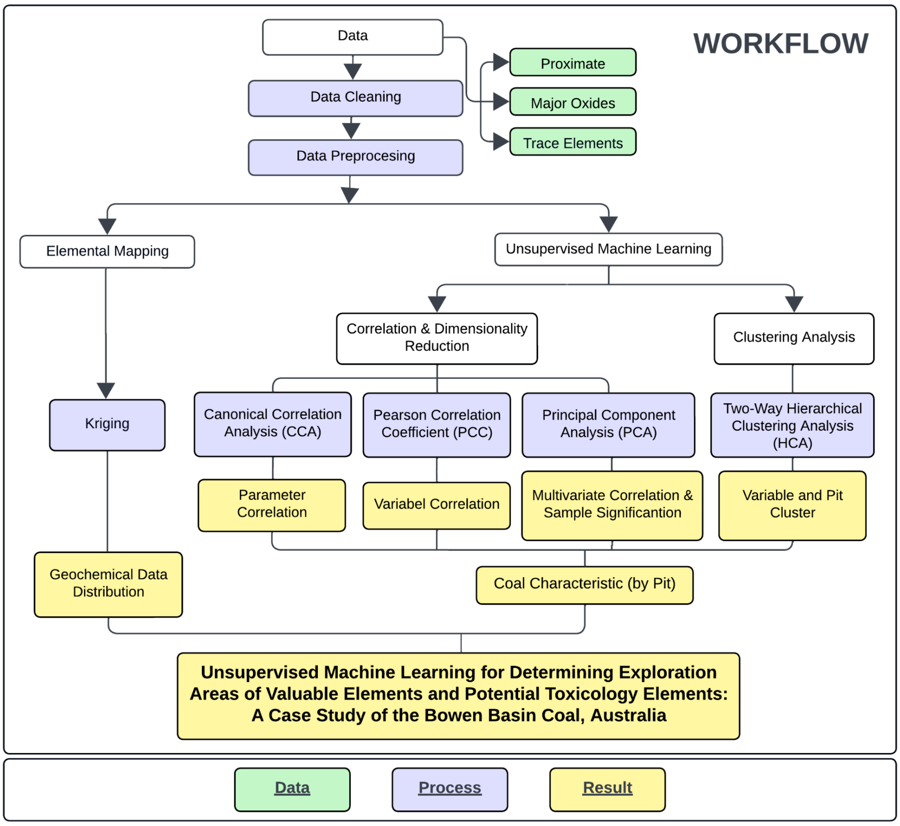

FIGURE 3. Workflow.

4.1 Dataset

Data used in this study include trace elements and geochemical parameters of coal from "Trace Elements in Coal from Collinsville, Bowen Basin, Australia – in-Ground Mode of Occurrence and Behaviour During Utilisation" by Boyd (2004). The dataset consists of a total of 57 samples, comprising 24 from Blake Central Pit, 27 from Blake West Pit, and 5 from Bowen No. 2 Pit (Figure 4). The samples from each pit were vertically distributed and collected using ply-by-ply sampling method.

The data were categorized into four datasets (parameter): proximate, major oxides, and two groups of trace elements—classified as toxicological and valuable elements. Proximate parameters consist of 4 for aspects i.e., ash, fixed carbon, inherent moisture, and volatile matter. Major oxides include 11 elements as its variables, i.e., Si, Al, FeA (for Fe_2O_2), FeB (for FeO), Ca, Mg, Na, K, Ti, Mn, and P. While, the trace elements consisting of six toxicology elements, i.e., As, Co, Cr, Ni, Se, Zn (Dai *et al.*, 2012) and 15 valuable elements include i.e., Au, Ba, Br, Cs, Hf, Hg, Mo, Rb, Sb, Sr, Ta, Th, U, W, Yb and total REY.

4.2 Data cleaning

Data cleaning aimed to eliminate variables before entering the main data process. This step eliminates some trace element variables that are concentrated below the world's average of trace elements in brown coal (Ketris and Yudovich, 2009), i.e., Au, Cs, Hg, Sb, U, and As. Ca as a trace element with concentration below the detection limit are also eliminated. Therefore, the elements included in the discussion are:

- **Valuable elements**: Ba, Br, Hf, Mo, Rb, REY, Sr, Ta, Th, and U
- **Toxic elements**: Co, Cr, Ni, Se, and Zn
- **Major oxides**: Al, FeA, FeB, K, Mg, Mn, Na, P, Si, dan Ti

In addition, all variables from the proximate analysis are included in the next step of the analysis.

Outlier identification is also included as one of data cleaning techniques to avoid outlier

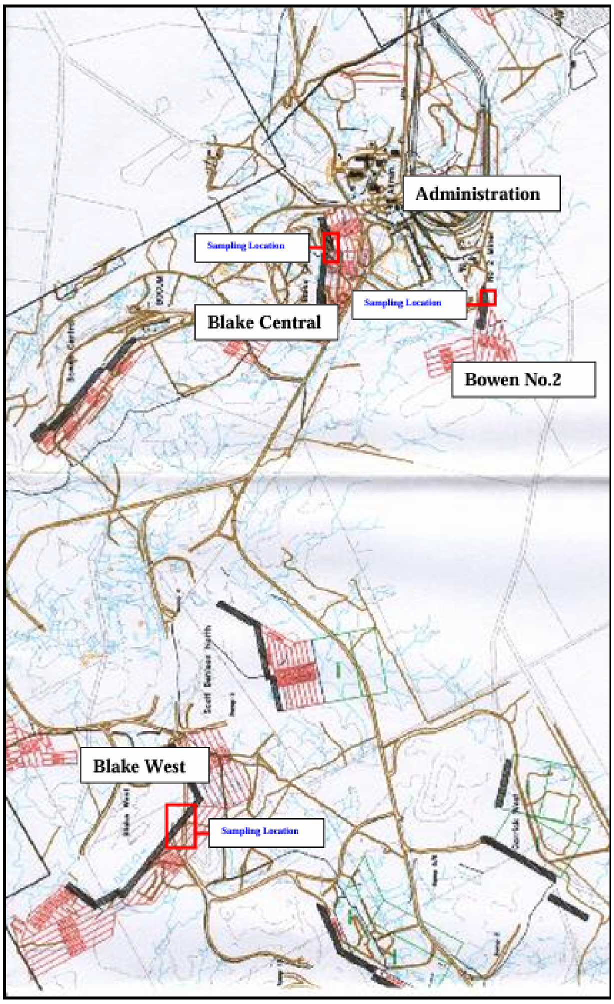

FIGURE 4. Relative location of the pits sampled for this study (Boyd, 2004).

skewing in CCA estimation (Wang *et al.*, 2020; Gelman and Hill, 2007) and recognising the anomalous value of the variables.

4.3 Data pre-processing

Main data processing through machine learning algorithms includes four separate analyses. Data pre-processing is needed in some of the analyses. The Z-score is defined for each variable to enhance domain interpretability before performing CCA (Wang *et al.*, 2020) (Figure 5) and HCA (Jiang *et al.*, 2015). Standardization with z-score is used to eliminate the impact of differences in numerical feature scales which transform feature X into standardized value Z (Li *et al.*, 2024).

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

Where μ is mean of the data and σ is standard deviation of feature X.

4.4 Canonical correlation analysis (CCA)

Canonical Correlation Analysis (CCA) was first introduced by Hotelling (1936). CCA is a statistical method aimed at determining the maximum correlation between two sets of data (multidimensional variables) (Abdi *et al.*, 2018). One common way to visualize CCA results is through a scatter plot with a trendline and the R^2 value (Canonical Correlation Coefficient) (Figure 6). The CCA value ranges from 0 to 1, where 0 indicates no relationship between the two data sets, while 1 indicates a perfect relationship between the data sets.

The use of this method aims to compare each pair of parameter groups, providing a general overview of the relationships between the analysed parameters before conducting a more in-depth correlation analysis. However, Canonical Correlation Analysis (CCA) can only demonstrate the relationship between the parameters as a whole, making it unable to reveal the relationships of individual variables within those parameters.

4.5 Pearson correlation coefficient (PCC)

The Pearson Correlation Coefficient (PCC) is a statistical method used to measure the linear correlation between two data variables. There are three types of correlation that PCC can indicate: 1) Negative correlation ($-1 < R < 0$); 2) No correlation ($R = 0$); and 3) Positive correlation ($0 < R < 1$).

PCC can be visualized in various diagrams, one of which is the heatmap. This visualization is also used in Li *et al.* (2024) and Xia *et al.* (2023). This diagram presents the PCC values (r) using a color scale. Negative correlations are represented by red, while positive correlations are shown in blue. The stronger the correlation, the more intense the color displayed. PCC (r) value calculated from equation below.

$$r_{xy} = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2] - [n\Sigma y^2 - (\Sigma y)^2]}} \tag{2}$$

Where:

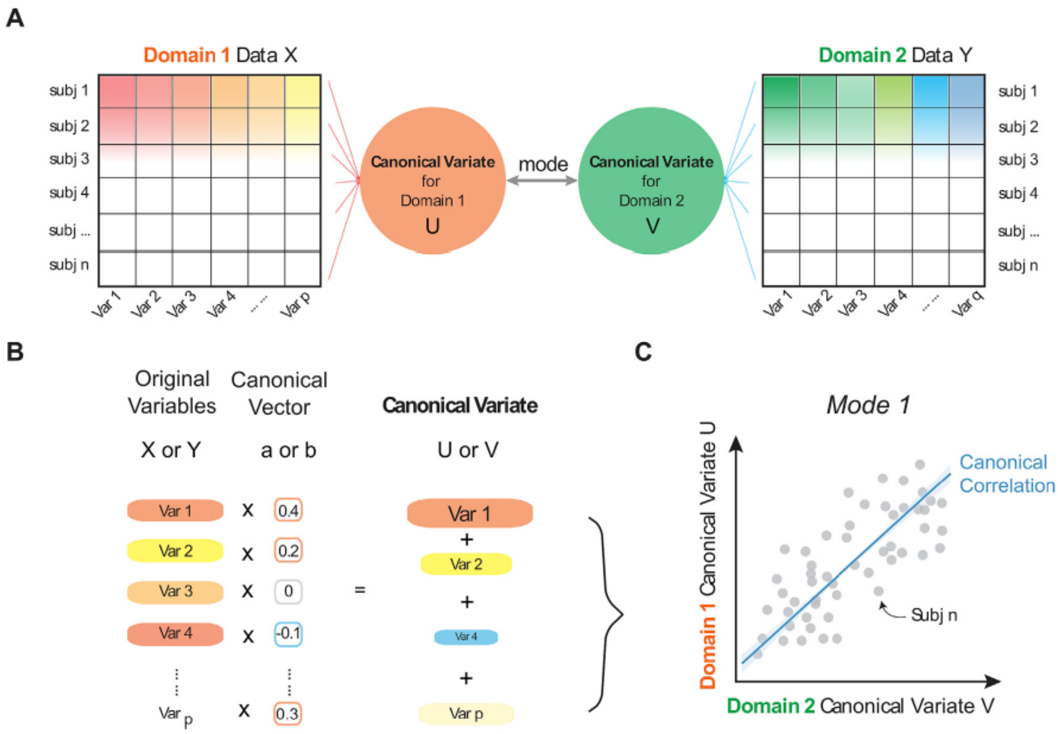

FIGURE 5. General schematic for Canonical Correlation Analysis (Wang *et al.*, 2020).

r_{xy} = pearson correlation coefficient
n = number of sample
Σxy = sum of the cross-products of each pair values
Σx = sum of the x values
Σy = sum of the y values
Σx^2 = sum of the squared x values
Σy^2 = sum of the squared y values

This method is used after determining the correlation relationships between parameter groups to review the correlation of each variable within those parameters. However, the Pearson Correlation Coefficient (PCC) can only indicate the relationships between variables within each parameter and cannot represent the distribution of samples that affect these relationships.

4.6 Principal component analysis (PCA)

Principal Component Analysis (PCA) is a multivariate statistical analysis method used to identify correlations between datasets which was developed by Hotelling (1933). The principle of PCA is to simplify the data structure by reducing its dimensionality (Jiang *et al.*, 2015). The original parameters are then re-organized into several comprehensive, uncorrelated factors, without losing significant information (Jiang *et al.*, 2015; Brown and Brown, 1998). In addition to showing correlations, Principal Component Analysis (PCA) also exhibits the distribution of sample groups, allowing for sample identification and controlled parameters.

4.7 Heatmap – two ways hierarchical clustering analysis (HCA)

Hierarchical Clustering Analysis (HCA) is a statistical method for grouping data based on similarity of samples and variables. One effective way to visualize HCA results is by combining a heatmap and a dendrogram as a two-way hierarchical clustering as used in Ma *et al.* (2020). The heatmap displays the values or concentrations of each variable represented by z-score, while the dendrogram illustrates the data clusters based on the similarity of analyzed variables.

The objective of clustering is to identify groups of data points based on their similarities and to determine the samples associated with each group.

4.8 Inverse-distance weighting (IDW)

The Inverse-Distance Weighting (IDW) method, as described by Wong (2017), is a widely used spatial interpolation technique in Geographic Information System (GIS) that estimates values at unmeasured locations by using the spatial dependence of neighbouring data points. The weight λ_i assigned to each sampled point s_i is inversely proportional to its distance d_{i0} from the unsampled location s_0, raised to a power defined by the distance-decay parameter α. The predicted value $\hat{a}(s_0)$ is calculated as:

$$\hat{a}(s_0) = \sum_{i=1}^{n} \lambda_i a(s_i) \quad (3)$$

Where:

$$\lambda_i = \frac{d_{i0}^{-\alpha}}{\sum_{i=1}^{n} d_{i0}^{-\alpha}} \quad (4)$$

Here, $a(s_i)$ represents the observed value at sampled location s_i, and n is the number of sampled points included in the calculation. The parameter α controls the rate of distance-decay: higher α values prioritize nearby points, while lower values allow distant points to contribute more. IDW is an exact interpolator, ensuring $\hat{a}(s_i) = a(s_i)$ at sampled locations.

The accuracy of IDW depends on two key parameters: α and the neighbourhood size. Neighbourhoods can be defined by a fixed radius around s_0 or by selecting the k-nearest sampled points. Smaller neighborhoods or higher α values produce localized estimates sensitive to nearby variations, whereas larger neighbourhoods or lower α values yield smoother surfaces by incorporating more distant observations. The method assumes spatial autocorrelation, where closer locations exhibit greater similarity, but cannot extrapolate beyond the spatial range of sampled data (Wong, 2017).

5 RESULT AND DISCUSSION

Based on the studies conducted by Ketris and Yudovich (2009) and Dai (2012), the data used for further analysis were classified into several main groups based on their characteristics, namely proximate data, major elements, and trace elements. A summary of the data used is presented in Figure 6.

TABLE 1. Data excluded from the analysis.

Criteria	Parameters
Values below the detection limit	Ca
Values below the global coal standard	Au, Cs, Hg, Sb, U, As

However, not all data were used in the analysis. Data were excluded if they fell below the instrument's detection limit or were considered depleted relative to global coal standards, which is defined by a concentration coefficient (cc) value of less than 1. The dataset is summarized in Table 1.

Proximate data were used to understand the physical and chemical properties and its association to major and trace elements, in terms of toxicological–valuable elements abundances and geological control.

Valuable elements, such as REY, Sr, and W, highlight their potential for technological or industrial applications. Conversely, toxicology elements such as Co, Cr, and Zn were analyzed to evaluate their potential environmental impacts. Data excluded from the analysis, such as Ca, Au, Cs, and others, were disregarded as they were deemed analytically insignificant in the context of this study.

5.1 Data correlation and dimensionality reduction

5.1.1 CCA

Based on the results of the correlation analysis between parameters using Canonical Correlation Analysis (CCA) as shown in Figure 7, the relationships between different parameter groups were categorized as either strong or weak correlations (Table 2).

A strong correlation observed between proximate parameters, major oxides, and valuable elements indicate significant interdependence in their geochemical characteristics. For instance, the relationship between proximate parameters and major oxides may reflect the mineralogical influences on the physical properties of the material, while the association with valuable elements highlights their potential co-occurrence in industrially valuable compounds.

Conversely, the weak correlations between

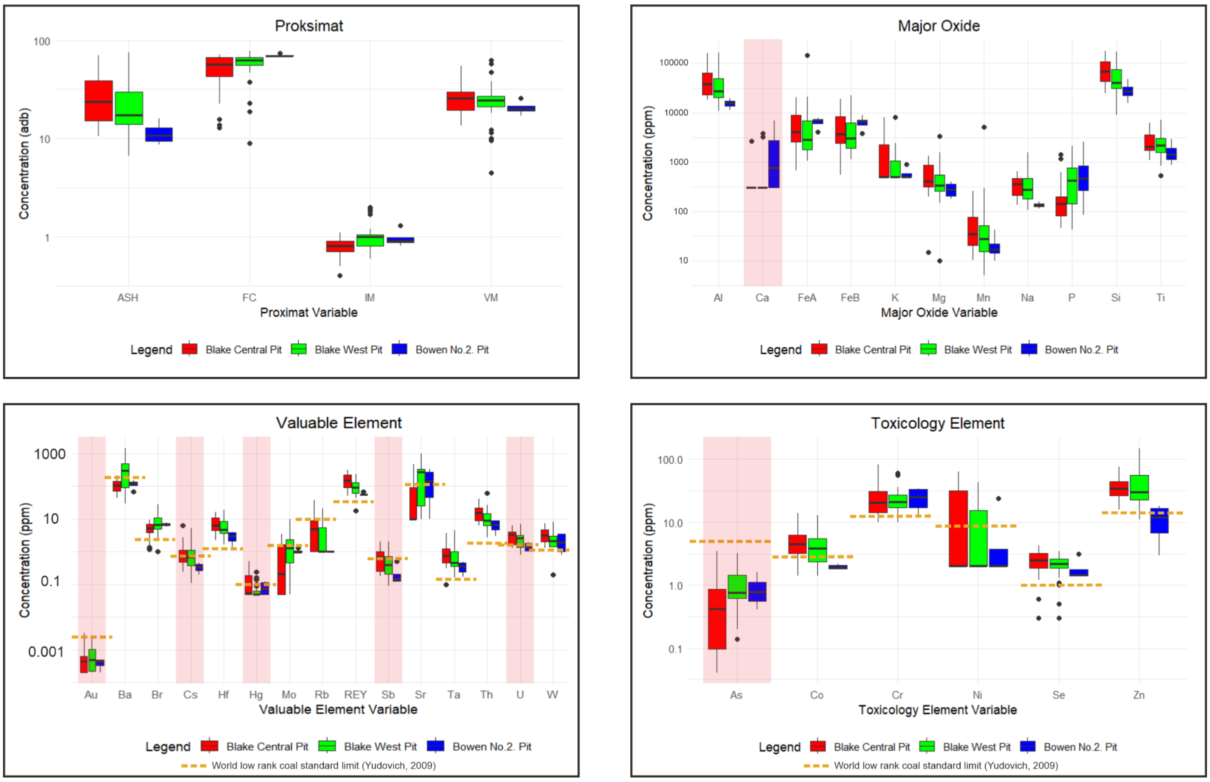

FIGURE 6. Data used for the analysis. The figure displays the full suite of variables considered for analysis, categorized into Proximate, Major Elements, and Trace Elements. The variables highlighted by the pink shaded areas were excluded from the final statistical analysis. Exclusion was based on two criteria: (1) values falling below the instrument's detection limit, or (2) values considered depleted, as defined by a concentration coefficient (cc) of less than 1. The classification of trace elements into valuable and toxicology groups is adapted from Ketris & Yudovich (2009) and Dai (2012).

TABLE 2. Correlation strength between parameter groups.

Parameter Groups	Correlation Strength
Proximate - Major Oxides	Strong
Proximate - Valuable Elements	Strong
Major Oxides - Valuable Elements	Strong
Major Oxides - Toxicology Elements	Strong
Proximate - Toxicology Elements	Weak
Valuable Elements - Toxicology Elements	Weak

proximate parameters and toxicology elements, as well as between valuable and toxicology elements, suggest limited interaction or co-dependence. This indicates that toxicology elements may have independent geochemical behaviour relative to other parameter groups. These findings provide critical insights into the geochemical processes influencing the studied material and its potential environmental and industrial implications.

5.1.2 PCC

By using the Pearson Correlation Coefficient (PCC), the relationships between parameters were identified, providing insights into the associations among elements. The results of the analysis are presented in Figure 8.

For this study, Pearson Correlation Coefficient (R) were interpreted as follows: strong ($R \geq 0.6$), moderate ($0.4 \leq R < 0.6$), and weak ($R < 0.4$) (Evans, 1996). The results show valuable elements can be categorized into two main categories based on their correlation patterns:

- **Group X (Ba, Br, Sr)**: with a strong relationship with inherent moisture (IM) and phosphorus (P).
- **Group Y (Hf, Mo, Ta, Th, W)**: with a strong correlation with ash yield and major ox-

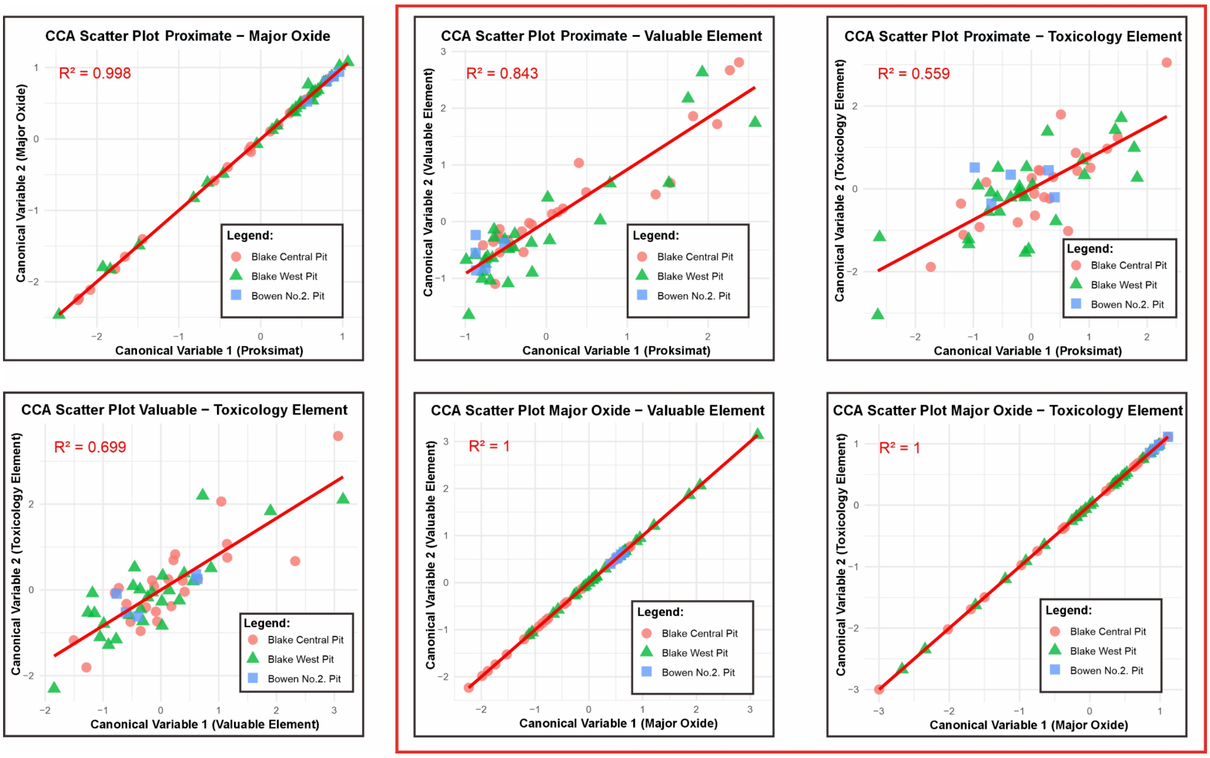

FIGURE 7. Result of Canonical Correlation Analysis (CCA). The analysis illustrates the correlation strength among parameter groups, showing strong relationships between Proximate-Major Oxides, Proximate-Valuable Elements, Major Oxides-Valuable Elements, and Major Oxides-Toxicology Elements. In contrast, weak correlations are observed between Proximate-Toxicology Elements and Valuable Elements-Toxicology Elements, highlighting the central role of Major Oxides as a key linking factor.

ides such as silicon (Si), aluminum (Al), titanium (Ti), and potassium (K).

Primary toxicological element identified is zinc (Zn), which exhibits significant correlations with volatile matter (VM) and ash, as well as with Fe_2O_3 (FeA), magnesium (Mg), and manganese (Mn). No significant correlation was found between valuable elements and toxicological elements, suggesting that they behave independently in terms of geochemical interactions.

5.1.3 *PCA*

The results of the multivariate correlation using Principal Component Analysis (PCA) show the distribution of samples based on pit locations, which influences the relationships between variables. Based on this analysis, the following findings were obtained regarding the relationships between valuable elements and toxicology elements, as summarized in Figure 9.

For valuable elements, two main groups were identified. Group X, consisting of Ba, Br, and Sr, is associated with IM (Inherent Moisture) and P, and is dominated by samples from Blake West Pit. Group Y, which includes Hf, Mo, Ta, Th, and W, is associated with ash and elements such as Si, Al, Ti, and K, and is dominated by samples from Blake Central Pit, with some samples also coming from Blake West Pit.

Additionally, samples from Bowen No. 2 Pit accumulate separately and are associated with high FC (Fixed Carbon) values, indicating that the coal from this pit is predominantly clean coal. An outlier sample was also observed from Blake West Pit, suggesting greater variability in the material composition from this location.

Regarding the relationship between valuable elements and toxicology elements, no significant correlation was found, indicating that these two element groups do not strongly influence each other in the context of this study.

5.2 Clustering analysis (HCA)

Overall, six groups were identified, with one outlier group based on the data analysis. Each group exhibits unique characteristics related to the sample origin and the associated indicators.

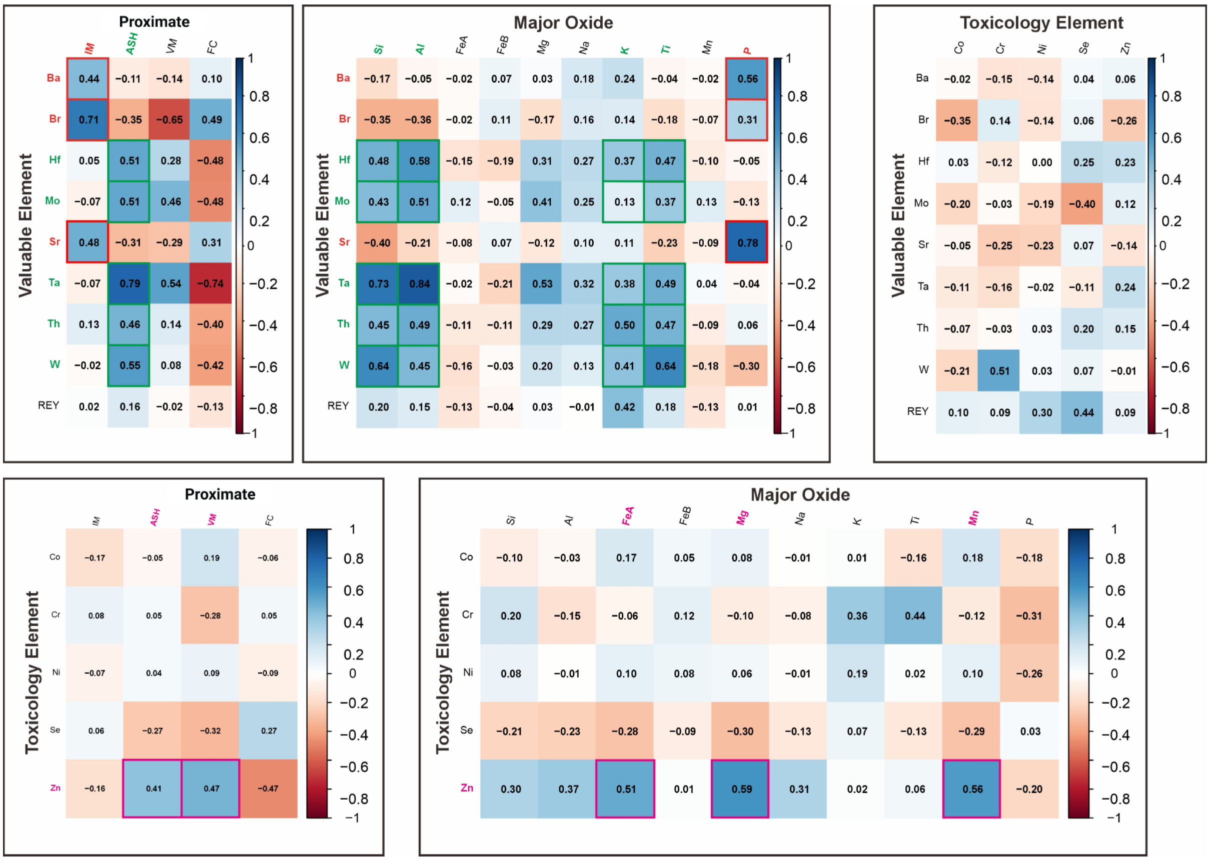

FIGURE 8. Result of Pearson Correlation Coefficient (PCC). The figure highlights significant correlations between multiple variables. Ba, Br, and Sr show positive correlations with inherent moisture (IM) and phosphorus (P), while Hf, Mo, Ta, Th, and W are correlated with ash content and major oxides such as Si, Al, Ti, and K. For toxicological elements, Zn shows strong correlation with volatile matter (VM), ash, and oxides including Fe2O3, Mg, and Mn. No significant correlations are observed between valuable and toxicological elements.

A summary of the group identification results is presented in Figure 10.

Group A, originating from Blake Central Pit and Blake West Pit, shows high VM and Ash values and is associated with elements such as Hf, Mo, Ta, Th, W, Si, Al, Ti, and K. This group stands out in terms of material composition, which suggests potential for industrial or geotechnological applications.

Group B, found in Blake Central Pit, Blake West Pit, and Bowen No. 2 Pit, exhibits high IM and FC values and is associated with elements such as Ba, Br, Sr, and P. These characteristics indicate a close relationship with mineral compositions that may be important for material applications.

Group C, also from Blake Central Pit and Blake West Pit, is associated with valuable elements such as Hf, Ta, Th, W, and REY, as well as toxic elements like Ni, Co, Se, and major oxides like Ti and K. This group shows high potential value but also highlights potential toxic impacts that need to be considered.

Group D, from Blake Central Pit and Blake West Pit, is associated with toxic elements such as Co, Ni, Se, and valuable REY elements. This indicates potential environmental risks that should be addressed if the material is to be used.

Group E, consisting of only one sample from Blake West Pit and Blake Central Pit, shows relatively high IM and FC values and is associated with toxic elements such as Se, Co, Ni, Zn, and valuable REY elements. This highlights the potential use of the material but also raises concerns about its environmental and health impacts.

Group F, found in Blake Central Pit, Blake West Pit, and Bowen No. 2 Pit, shows high IM and FC values and is associated with the toxic

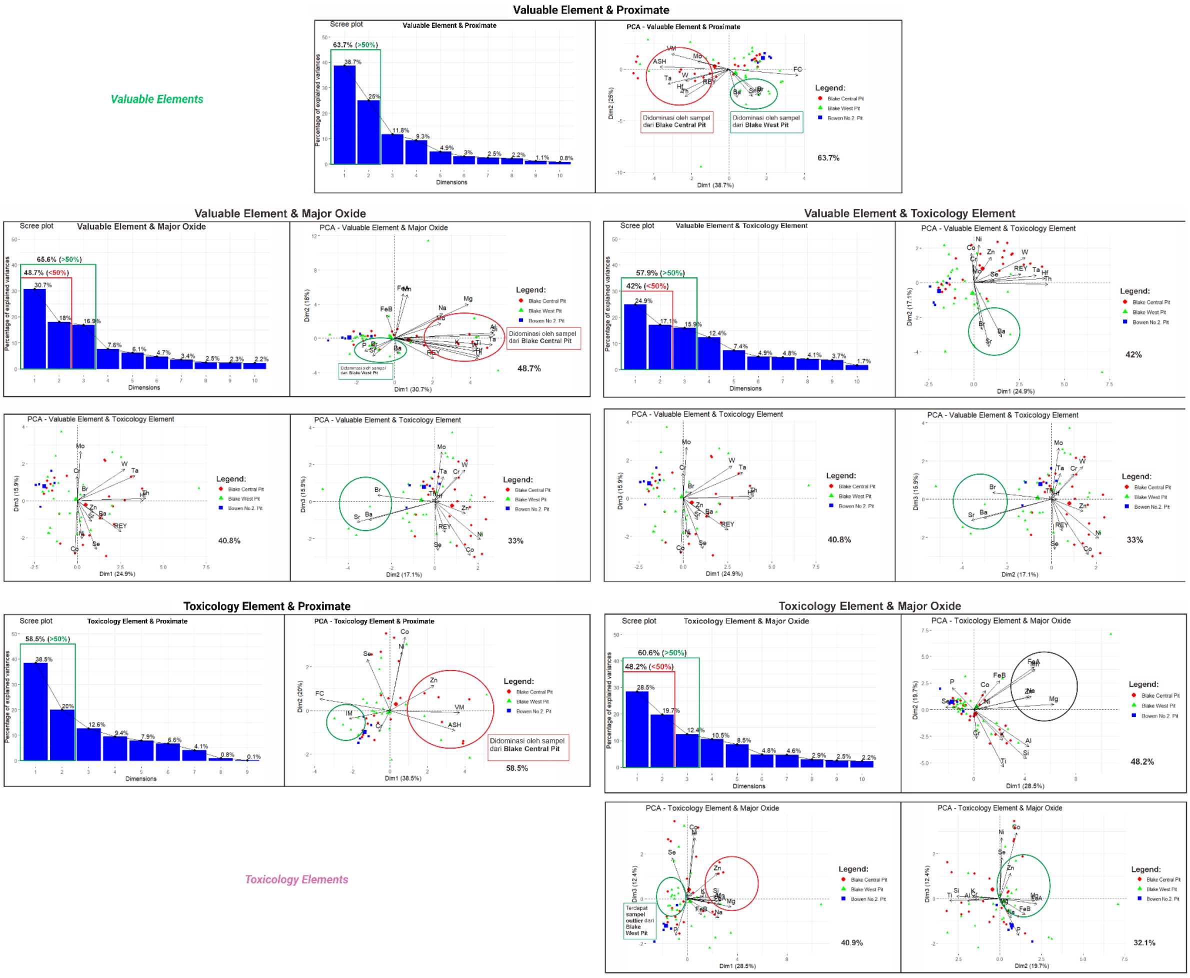

FIGURE 9. Sample distributions based on pit and relationships between variables. The plot reveals a distinct separation of samples based on their origin. Samples from Bowen No.2 Pit are characterized by high Fixed Carbon (FC), while Blake Central and Blake West pits are associated with different suites of valuable and toxic elements. Specifically, Blake Central samples correlate with Hf, Mo, Ta, and Zn, whereas Blake West samples correlate with Ba, Br, and Sr. The analysis also identifies an outlier from Blake West Pit.

element Cr. This group suggests potential risks that need to be considered in processing and usage.

5.3 Elemental distribution

Bowen coals are enriched in some critical elements (e.g., Hf, Ta, Th, and W). This association was first identified by the Pearson Correlation Coefficient (PCC) and confirmed by the Principal Component Analysis (PCA), which shows these elements grouping with Si, Al, Ti, and K, as reflected in the high Ash content of this coal group. The high Ash values indicate that these concentrated elements are contaminants carried by the mineral matter.

In terms of spatial distribution, as shown in Figure 11 and Table 3, the distribution of Ash, Si, Al, Ti, K, Hf, Ta, Th, and W shows similar patterns, increasing toward Blake Central Pit. Therefore, exploration for the extraction of these concentrated elements can be focused on this area.

Additionally, REY is a valuable element concentrated in Bowen coals. The analysis further shows that REY has a significant, though weak, association with K. The spatial pattern of REY also shows an increasing concentration toward the NW or Blake Central Pit.

Zn, which is a toxicology element, is also concentrated in Bowen coals. This element is associated with FeA, Mn, and Mg, where the VM and Ash values of this coal group are relatively high. The spatial distribution of Zn, along with VM, Ash, FeA, Mn, and Mg, shows similar pat-

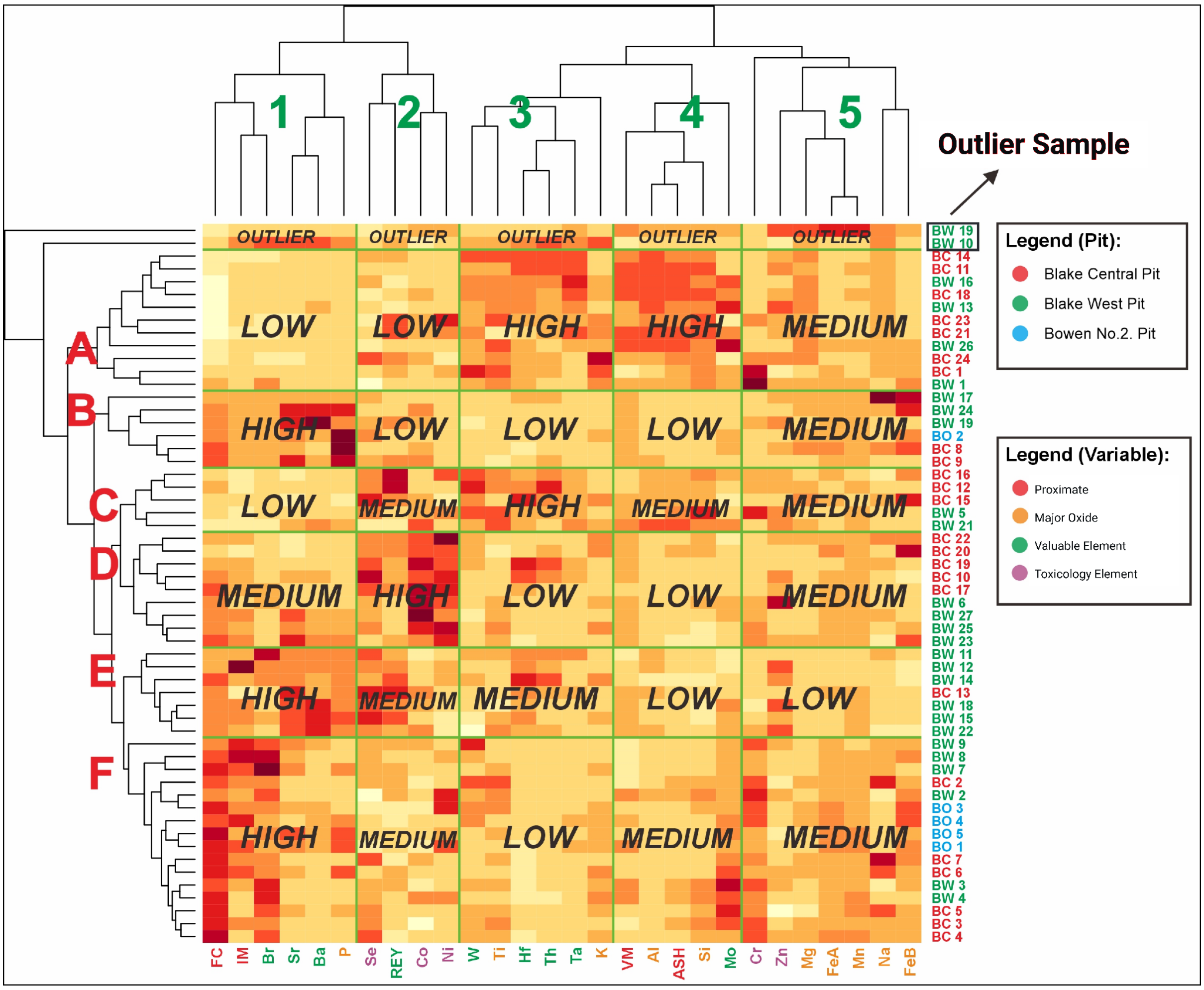

FIGURE 10. Hierarchical Cluster Analysis (HCA) dendrogram of the coal samples. The clustering was performed on the standardized geochemical data. The y-axis represents the linkage distance, indicating the degree of dissimilarity at which samples and clusters are merged. The analysis reveals the presence of six primary clusters (labeled A through F) and two outlier samples, each delineated by a different color shade for clarity.

terns, increasing toward the west. For Zn and major oxides, the spatial distribution points toward Blake West Pit, indicating that coal from this area has higher concentrations of the toxicology element Zn compared to other pits.

The spatial distribution of IM in Bowen Coals shows a trend of increasing concentrations toward the south or Blake West Pit, while FC shows a trend of increasing concentrations toward the southeast or Bowen No. 2 Pit. This is consistent with previous analysis, which stated that samples from Bowen No. 2 Pit are clean coal.

Finally, the distribution of P increases toward Bowen No. 2 Pit and is associated with valuable elements Ba, Br, and Sr. However, these three elements are not concentrated in the Bowen coal samples.

These findings suggest that certain areas, such as Blake Central Pit and Blake West Pit, have significant potential for the extraction of valuable elements, while the presence of toxic elements should also be considered in future exploration planning.

5.4 Geological controls on elemental distribution

The Valuable Elements group, including Ba, Br, and Sr, which are found in carbonates, organic material, sulfates, phosphates, and silicates, shows a strong correlation with the major element P. This indicates that the coal formation process involved marine water infiltration during the transgressive phase (rising sea

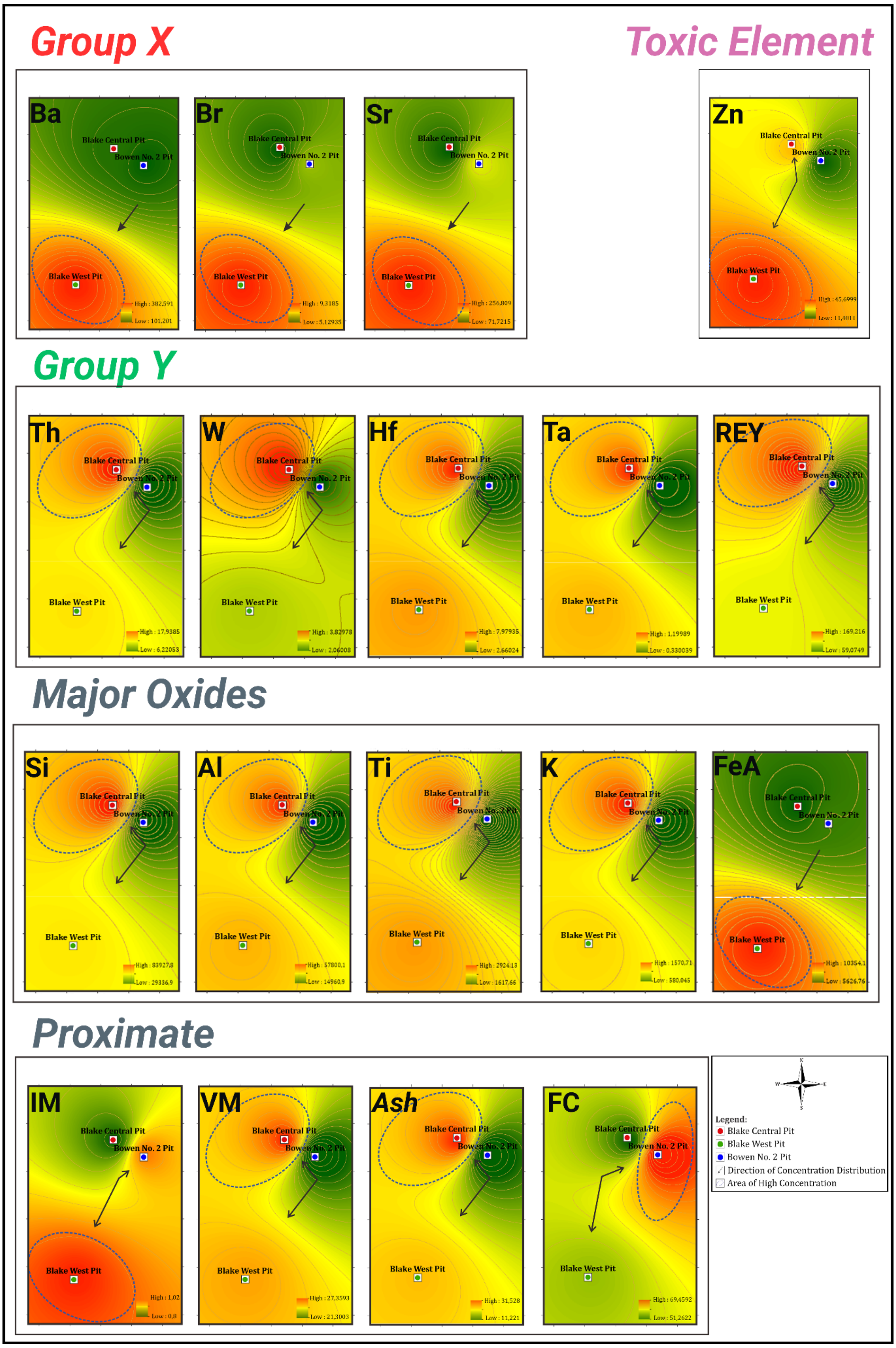

FIGURE 11. Spatial distribution maps confirming the result of multivariate analysis for some parameters and elements. Concentration levels are indicated by color gradient (green: low, red: high). The maps visually demonstrate the element groupings identified previously. A large suite of variables, including Group Y and Major Oxides, shows a clear enrichment trend towards Blake Central Pit. Conversely, Group X elements, Toxic Element, and Inherent Moisture (IM) are enriched towards Blake West Pit. This provides compelling visual evidence for the distinct geochemical associations operating within the study area.

TABLE 3. Distribution patterns of elements and their associations in Bowen Coals.

Element	Association	Distribution Pattern
Hf, Ta, Th, W	Si, Al, Ti, K	Increases toward Blake Central Pit
REY	K	Increases toward NW or Blake Central Pit
Zn	FeA, Mn, Mg	Increases toward the west, higher in Blake West Pit
IM	-	Increases toward the south or Blake West Pit
FC	-	Increases toward the southeast or Bowen No. 2 Pit
P	Ba, Br, Sr	Increases toward Bowen No. 2 Pit, not concentrated

levels) in the Early-Late Permian period (Baker *et al.*, 1993), when organic material from marine environments became trapped in the coal. This association highlights the influence of marine environments on coal formation during this period (Table 4).

The Group Y Elements, consisting of Hf, Mo, Ta, and W, are strongly associated with Ash as well as major elements such as Si, Al, Ti, and K. This association suggests that these elements come from silicate and clay minerals, which were formed by volcanic activity around the Bowen Basin during the Early Permian (Fielding *et al.*, 1997). The presence of zircon, containing Hf, and silicate minerals confirms that volcanic activity in the region played a significant role in the formation of minerals in coal. Additionally, elements associated with sulfides, such as Mo and W, are linked to marine fluid infiltration during the transgressive phase of the Early-Late Permian.

The Toxicology Elements group, including Zn, which is found in minerals like sphalerite, pyrite, and silicates, is associated with VM, Ash, and major elements such as Fe, Mg, and Mn. The presence of pyrite (FeS_2) as the primary source of these elements indicates the influence of marine fluids that infiltrated the coal during the transgressive phase of the Early-Late Permian period. This association strengthens the understanding that toxic elements in coal may originate from marine environments trapped during coal formation.

6 CONCLUSION

By applying an unsupervised machine learning algorithm, coal characteristics were successfully identified for preliminary exploration activities. Although each resulting correlation remains open to interpretation, the validity of the findings is supported by geological conditions in the study area that are consistent with the analytical results. The main conclusions are as follows:

a. Machine learning has proven effective in characterizing coal based on its geochemical properties. Variables that exhibit correlation are considered significant in distinguishing coal from different locations.

TABLE 4. Occurrence and associations of elements in coal.

Element Groups		Occurrence	Associations
Group X Elements	Ba (Barium)	Carbonates, organic material, sulfates, phosphates, silicates	Strong correlation with major element P
	Br (Bromine)	Carbonates, organic material, silicates	Strong correlation with major element P
	Sr (Strontium)	Carbonates, organic material, phosphates	Strong correlation with major element P
Group Y Elements	Hf (Hafnium)	Zircon, clay	Associated with Ash, Si, Al, Ti, K
	Mo (Molybdenum)	Silicates, sulfides	Associated with Ash, Si, Al, Ti, K
	Ta (Tantalum)	Oxides, silicates	Associated with Ash, Si, Al, Ti, K
	W (Tungsten)	Silicates, sulfides	Associated with Ash, Si, Al, Ti, K
Toxicology Element	Zn (Zinc)	Sphalerite, pyrite, silicates	Associated with VM, Ash, Fe, Mg, Mn

Correlated variables tend to show similar trends in their value increases.

b. Blake Central Pit is characterized by high ash yield, which is associated with inorganic elements such as Si, Al, Ti, K, as well as valuable elements such as Hf, Mo, Ta, Th, W, and REY. This pit has significant potential for valuable element exploration, with a directional trend towards the N-NW.
c. Blake West Pit is characterized by high inherent moisture, which correlates with the presence of P and valuable elements such as Ba, Br, and Sr. However, it also indicates a potential distribution of toxicology elements with a directional trend towards S-SW.
d. Bowen No. 2 Pit has high fixed carbon and low ash yield, indicating a cleaner coal quality compared to the other pits.

These findings highlight the effectiveness of unsupervised machine learning in identifying coal characteristics and their association in valuable and toxicology elements, supporting targeted exploration strategies in Bowen Basin.

Acknowledgements

REFERENCES

Abdi, H., Guillemot, V., & Beaton, D. (2018). Encyclopedia of Social Network Analysis and Mining. Encyclopedia of Social Network Analysis and Mining. DOI: 10.1007/978-1-4614-7163-9.

Baker, J. C., Fielding, C. R., De Caritat, P., And, T. M., & Wilkinson, M. M. (1993). Permian evolution of sandstone composition in a complex back-arc extensional to foreland basin: the Bowen Basin, Wastern Australia. SEPM Journal of Sedimentery Research. DOI: 10.1306/d4267c35-2b26-11d7-8648000102c1865d.

Boyd, R., J. (2004). Trace elements in coal from Collinsville, Bowen Basin, Australia-in-ground mode of occurrence and behaviour during utilisation.

Brakel, A. T., Totterdell, J. M., Wells, A. T., & Nicoll, M. G. (2009). Sequence stratigraphy and fill history of the Bowen Basin, Queensland. Australian Journal of Earth Sciences, 56(3), 401–432. DOI: 10.1080/08120090802698711.

Brown, C. E. (1998). Applied Multivariate Statistics in Geohydrology and Related Sciences. In Sensory Evaluation of Sound. Springer-Verlag.

Dai, S., & Finkelman, R. B. (2018). Coal as a promising source of critical elements: Progress and future prospects. International Journal of Coal Geology, 186(June 2017), 155–164. DOI: 10.1016/j.coal.2017.06.005.

Dai, S., Ren, D., Chou, C. L., Finkelman, R. B., Seredin, V. V., & Zhou, Y. (2012). Geochemistry of trace elements in Chinese coals: A review of abundances, genetic types, impacts on human health, and industrial utilization. International Journal of Coal Geology, 94, 3–21. DOI: 10.1016/j.coal.2011.02.003.

Dai, S., Seredin, V. V., Ward, C. R., Hower, J. C., Xing, Y., Zhang, W., Song, W., & Wang, P. (2015). Enrichment of U-Se-Mo-Re-V in coals preserved within marine carbonate successions: geochemical and mineralogical data from the Late Permian Guiding Coalfield, Guizhou, China. Minerals Deposita, 50(2), 159–186.

Dai, S., Ward, C. R., Graham, I. T., French, D., Hower, J. C., Zhao, L., & Wang, X. (2017). Altered volcanic ashes in coal and coal-bearing sequences: A review of their nature and significance. In Earth-Science Reviews (Vol. 175, hal. 44–74). Elsevier B.V. DOI: 10.1016/j.earscirev.2017.10.005.

Evans, J.D. (1996). Straight forward statistics for the behavoral science. Pacific Grove, Ca: Brooks/Cole Publishing.

Fielding, C. R., Holcombe, R. J., Fielding, C. R., Stephens, C., & Holcombe, R. J. (1997). Permian stratigraphy and palaeogeography of the eastern Bowen Basin, Gogango Overfolded Zone and Strathmuir Synclinorium in the Rockhampton-Mackay region, central Queensland. Geological Society of Australia, 19, 80–95.

Gelman, A., & Hill, J. (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 24(6), 417.

Hotelling, H. (1936). Relation Between Two Sets of Variates. In Biometrika (Vol. 28, hal. 321–377). DOI: 10.1016/B978-0-12-385022-5.00013-0.

Hu, Y., You, M., Liu, G., Dong, Z., Jiao, F., & Meng, Y. (2021). The potential utilizing of critical element from coal and combustion residues. Energies, 14(15). DOI: 10.3390/en14154710.

Jiang, Y., Guo, H., Jia, Y., Cao, Y., & Hu, C. (2015). Principal component analysis and hierarchical cluster analyses of arsenic groundwater geochemistry in the Hetao basin, Inner Mongolia. Chemie der Erde, 75(2), 197–205. DOI: 10.1016/j.chemer.2014.12.002.

Ketris, M. P., & Yudovich, Y. E. (2009). Estimations of Clarkes for Carbonaceous biolithes: World averages for trace element contents in black shales and coals. International Journal of Coal Geology, 78(2), 135–148. DOI: 10.1016/j.coal.2009.01.002.

Li, W., Li, W., Busch, A., Wang, L., Anggara, F., & Yang, S. (2024). Machine Learning Algorithm to Predict Methane Adsorption Capacity of Coal. Energy and Fuels. 10.1021/acs.energyfuels.4c04906.

Ma, L., Li, Y., Abuduwaili, J., Abdyzhapar, S., & Liu, W. (2020). Hydrochemical composition and potentially toxic elements in the Kyrgyzstan portion of the transboundary Chu - Talas river basin , Central Asia. Scientific Reports, 0123456789, 1–15. DOI: 10.1038/s41598-020-71880-4.

Martini, I. P., & Johnson, D. P. (1987). Cold-climate, fluvial to paralic coal-forming environments in the Permian Collinsville Coal Measures, Bowen Basin, Australia. International Journal of Coal Geology, 7(4), 365–388. DOI: 10.1016/0166-5162(87)90054-1

Queensland Government, D. of R., KPGM, GHD, & NSAII. (2021). Bowen Basin Concept Study Final Report Queensland Government-Department of Resources. December.

Seredin, V. V., & Dai, S. (2012). Coal deposits as potential alternative sources for lanthanides and yttrium. International Journal of Coal Geology, 94, 67–93. DOI: 10.1016/j.coal.2011.11.001.

Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K. L. A., Elkhatib, Y., Hussain, A., & Al-Fuqaha, A. (2019). Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges. IEEE Access, 7, 65579–65615. https://doi.org/10.1109/ACCESS.2019.2916648

Wang, H. T., Smallwood, J., Mourao-Miranda, J., Xia, C. H., Satterthwaite, T. D., Bassett, D. S., & Bzdok, D. (2020). Finding the needle in a high-dimensional haystack: Canonical correlation analysis for neuroscientists. NeuroImage, 216(December 2019), 116745. DOI: 10.1016/j.neuroimage.2020.116745

Wong, D. W. S. (2017). Interpolation: Inverse-Distance Weighting. International Encyclopedia of Geography, 1–7. DOI: 10.1002/9781118786352.wbieg0066.

Wu, G., Chen, G., Cheng, Q., Zhang, Z., & Yang, J. (2021). Unsupervised Machine Learning for Lithological Mapping Using Geochemical Data in Covered Areas of Jining, China. Natural Resources Research, 30(2), 1053–1068. DOI: 10.1007/s11053-020-09788-z.

Xia, X., Jiang, C., & Hu, M. (2023). Geochemical characteristics and ecological risks of rare earth elements in river sediments of coal-grain composite area in eastern China. Environmental Monitoring and Assessment, 195(12). DOI: 10.1007/s10661-023-12071-2.