Explainable Artificial Intelligence and Advanced Feature Selection Methods for Predicting Gas Concentration in Longwall Mining

Haoqian Chang[a], Xiangqian Wang[a,**], Alexandra I. Cristea[b,*], Xiangrui Meng[a], Zuxiang Hu[a], Ziqi Pan[b]

[a]*Anhui University of Science and Technology, School of Economics and Management, Huainan, 232001, Anhui, China*
[b]*Durham University, Department of Computer Science, Durham, DH1 3LE, UK*

Abstract

Accurate prediction of gas concentrations at longwall mining faces is critical for safety production, yet current methods still face challenges in interpretability and reliability. This study aims to enhance prediction accuracy and model interpretability by employing advanced feature selection techniques. We integrate Shapley Additive Explanations (SHAP) into feature selection process to identify and quantify the contributions of multivariate features to gas concentration variations. The effectiveness of SHAP-based feature selection is systematically evaluated alongside Principal Component Analysis, Dynamic Time Warping, and unfiltered features, across four baseline predictive models chosen based on their structural characteristics: Long Short-Term Memory, Gated Recurrent Unit, Transformer and Graph Neural Network. Using public dataset from the Upper Silesian coal basin in Poland, we demonstrate that models trained with SHAP-selected features outperform baseline models, particularly in terms of accuracy and reliability for long-term predictions. By identifying the most relevant features and clarifying their interactions, this study enhances predictive performance and provides deeper insights into the dynamics governing gas concentrations, emphasising the value of advanced, interpretable feature selection techniques in developing robust models for industrial applications in mining.

Keywords: Explainable Artificial Intelligence (XAI), Multivariate Time Series Prediction, Shapley Additive Explanations (SHAP), Longwall Mining Safety, Gas Concentration Modelling

1. Introduction

Predicting gas concentrations at longwall mining faces is essential for ensuring mining safety Fan et al. (2023). In the semi-enclosed tunnel environment, underground monitoring data often display intricate coupling patterns Diaz et al. (2022) and spatiotemporal correlations Palka et al. (2023). While the interactions between various factors influencing gas concentration are apparent, accurately quantifying and representing these interactions remains a challenge Barnewold and Lottermoser (2020). To ensure precision in prediction, it is critical to identify the most relevant and influential features from the extensive array of underground sensors Liu et al. (2023), as these key features directly impact the prediction targets, thereby providing a robust foundation for reliable prediction in complex mining environments Liang et al. (2023).

However, determining the most influential factors is challenging Dougherty et al. (2023), particularly due to the linear and nonlinear coupling relationships between monitoring data Zhao et al. (2023), which makes it more complex to identify the features that have the greatest impact on gas concentration predictions Wen et al. (2023). For instance, temperature variations influence gas behaviour by altering solubility, while changes in airflow affect the diffusion rates of gases Nie et al. (2024). These coupled factors intricately influence gas dynamics in underground environments, exacerbated by the dynamic nature of mining operations, underscores the need for advanced methods to improve gas concentration predictions and ensure safety Chaturvedi (2023). Refining feature selection in longwall mining operations necessitates the adoption of advanced methodologies to accurately quantify and prioritise the most influential factors affecting gas concentration, thereby enhancing prediction accuracy amidst complex coupling challenges.

Feature selection is crucial for refining gas concentration predictive models, particularly in mining environments where complex, non-linear relationships between variables are prevalent Zhang and Wang (2023). This process involves identifying the most relevant variables from extensive datasets to enhance model accuracy and interpretability Hassija et al. (2024); Angelov et al. (2021). Techniques such as principal component analysis (PCA) Maćkiewicz and Ratajczak (1993), dynamic time warping (DTW) Müller (2007), and Pearson correlation are commonly employed to reduce dimensionality and highlight influential features Masini et al. (2023). However, these methods often focus on statistical correlations, potentially overlooking nuanced interactions and leading to a loss of critical information Zamanzadeh Darban et al. (2024). To address these shortcomings, advanced methods such as explainable artificial intelligence (XAI) offer a promising alternative Ahmed et al. (2022). XAI techniques are capable of uncovering and elucidating the complex dependencies influencing gas concentration variations with greater clarity. Notably, XAI has

*Corresponding author: alexandra.i.cristea@durham.ac.uk
**Principal corresponding author: xiqwang@aust.edu.cn

been successfully applied in environmental domains, including gully erosion Gholami et al. (2023), land subsidence Gholami et al. (2024b), wildfire susceptibility prediction Abdollahi and Pradhan (2023), and dust emission assessments Gholami et al. (2024a). Integrating these interpretative methods into gas concentration prediction frameworks could significantly enhance transparency, reliability, and practical applicability within mining contexts.

The advent of explainable artificial intelligence Minh et al. (2022) further enhances the feature selection process. Explainable AI techniques, such as SHapley Additive exPlanations (SHAP) Lundberg and Lee (2017); Parsa et al. (2020) and Local Interpretable Model-agnostic Explanations (LIME) Ribeiro et al. (2016), provide insights into the contribution of each feature to the model's predictions. These methods not only improve the transparency of deep learning models but also facilitate a deeper understanding of the underlying data interactions. By leveraging explainable AI, researchers and engineers can ensure that the selected features align with domain knowledge and operational realities, thereby enhancing the reliability and trustworthiness of the predictive models.

To advance this research, a comprehensive feature selection study is proposed, focused on predicting gas concentrations specifically at the upper corner of the coal mining face. The study employ four distinct feature selection methodologies—Principal Component Analysis (PCA) Maćkiewicz and Ratajczak (1993), Dynamic Time Warping (DTW) Müller (2007), SHapley Additive exPlanations (SHAP) Lundberg and Lee (2017), and an unfiltered entire dataset to identify the most pertinent sensor data influencing gas concentration. The identified features will be evaluated across four baseline multivariate time series prediction models, selected based on their structural architectures to assess the effectiveness of the feature selection method under different model frameworks: Long Short-Term Memory (LSTM) Hochreiter and Schmidhuber (1997), Gated Recurrent Unit (GRU) Cho et al. (2014), Transformer Vaswani et al. (2017), and Graph Neural Network (GNN) Scarselli et al. (2008). Each model will be run three times with varying sliding window sizes and different random seeds to minimise experimental error and enhance result reliability.

This approach enables a rigorous comparison of the effectiveness of each feature selection method in enhancing the accuracy of gas concentration predictions. Furthermore, it provides insights into the underlying patterns within longwall mining face data, yielding interpretable results that can inform safety measures.

The principal contributions of this paper are as follows:

- To the best of our knowledge, this is the first study to apply the SHAP local explanation method to investigate the coupled relationships of gas concentration features at the longwall mining face.

- Using real-world data from longwall mining face, we comprehensively evaluate methods for selecting gas concentration characteristics at the mining face and rigorously validate their effectiveness.

- Our study offers insights into the dynamics of gas concentration at longwall mining face, contributing to the development of robust gas management strategies and enhanced safety measures in mining operations.

2. Related Work

Effective feature selection enhances model performance and interpretability by reducing dimensionality and focusing on the most impactful variables Arrieta et al. (2020). Those echniques include Principal Component Analysis (PCA) Maćkiewicz and Ratajczak (1993), Mutual Information, and Recursive Feature Elimination (RFE) Yoon et al. (2005). Advanced methods such as SHapley Additive exPlanations (SHAP) Aldrees et al. (2024); Song et al. (2023) and Local Interpretable Model-agnostic Explanations (LIME) are also employed. These techniques facilitate understanding of complex interactions among variables and improve predictive accuracy in various applications.

Feature selection methods, while widely applied across various industrial sectors Chushig-Muzo et al. (2024) Hooker et al. (2021), remain relatively traditional within the mining industry and have not yet reached the same level of diversity or advancement. Liu et al. Liu et al. (2020) introduced a hybrid feature selection model for coal and gas outbursts, utilising random forest to identify the most relevant features. Zhou et al. Zhou et al. (2022) used four basic feature selection methods to identify optimal features and built predictive models with classical machine learning algorithms. Miao et al. Miao et al. (2024) developed a coal mine rock burst risk prediction model using standard machine learning and feature selection algorithms to identify key indicators. Chen et al. Chen and Dong (2020) implemented a sequential approach for water inflow prediction in coal mines, combining conventional feature selection and optimisation techniques. Huang et al. De et al. (2021) employed a multi-objective feature selection model to remove redundant and irrelevant features, enhancing fault diagnosis performance.

While these methods have improved safety and operational efficiency, the mining industry still lags in adopting sophisticated, explainable AI (XAI) techniques. Integrating advanced XAI methods for feature selection is essential to enhance model interpretability and trust.

Table 1: Characteristics and Measurement Units of Sensors

Sensor	Characteristics
MM	Methane(Gas concentration)(%CH4)
AN	Anemometer(m/s)
TP	Temperature(°C)
RH	Humidity(%RH)
BA	Barometer(hPa)
CM	High concentration methane meter(%CH4)
AM/DM	Mining equipment(A)

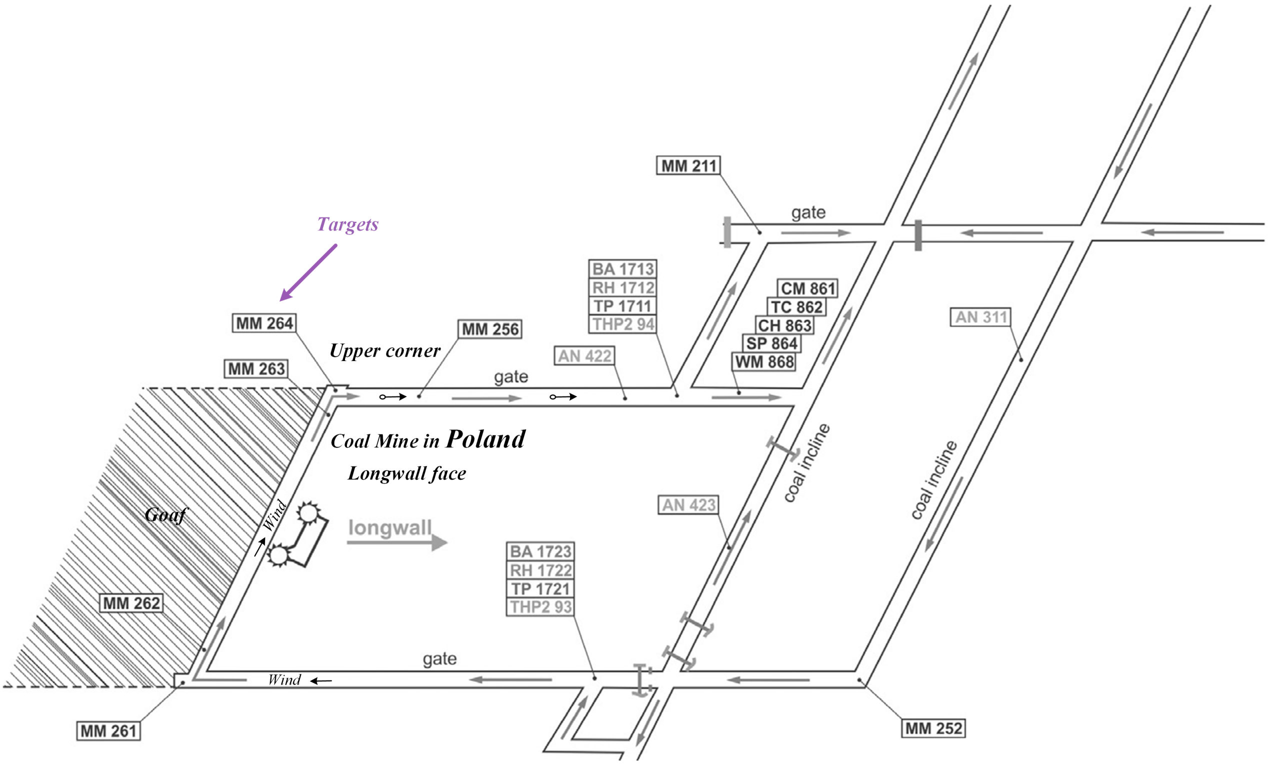

Figure 1: Structure of the Longwall Mining Face and Naïve Pearson Correlation Analysis of Gas Concentration

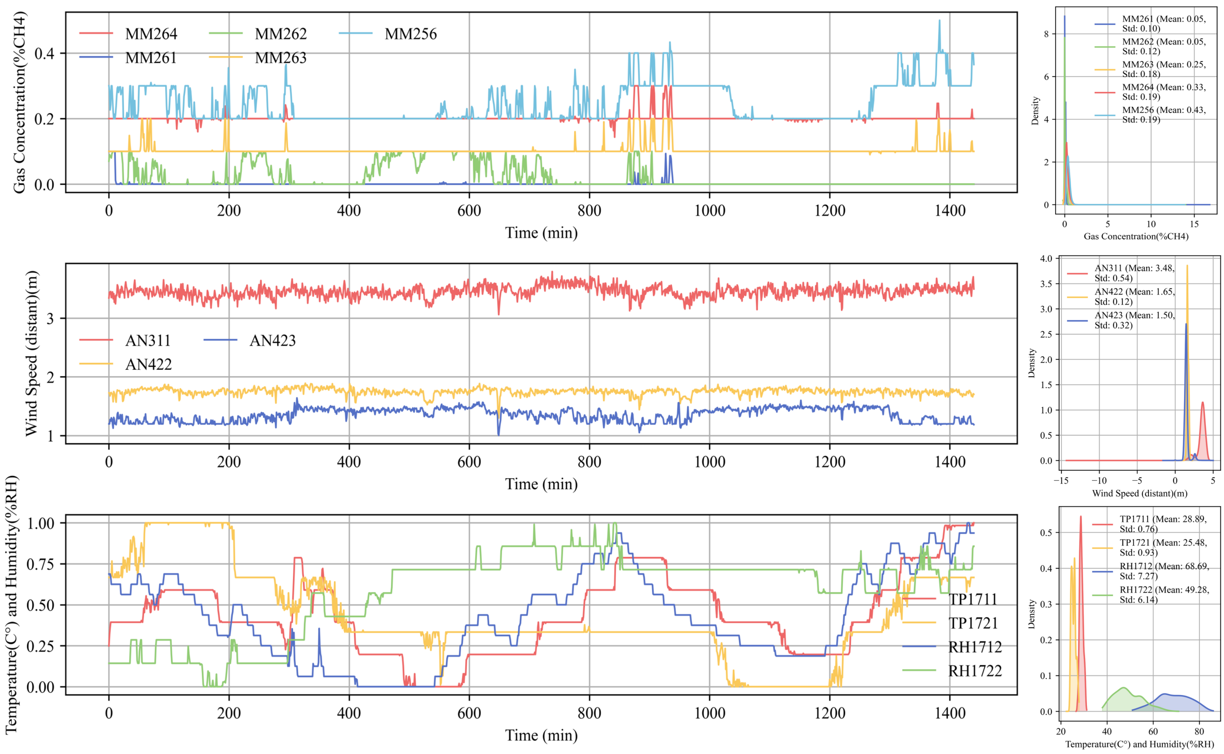

Figure 2: Time Series and Statistical Histogram of Data from Longwall Mining Face

3. Background

In this section, we present some background about longwall mining face, highlight the challenges through an initial dataset analysis, and define the research problem under investigation.

3.1. Dataset

The publicly available dataset used in this study was obtained from the Upper Silesian coal basin in Poland Ślęzak et al. (2018). Real-time data were collected from the underground environment of the longwall mining face, encompasses key environmental parameters such as Methane (Gas concentration), Anemometer, Temperature, Humidity, Barometer etc., as detailed in Table. 1, resulting in a high-frequency multidimensional time series. The dataset includes temporal sensory readings from two distinct sensor arrays, captured from March 2 to June 16, 2014. This repository comprises 9, 199, 930 data instances, each detailed with a timestamp and measurements from 28 sensors. Fig. 1 shows the mining face layout from which the dataset was derived, where fresh airflow is introduced from the intake airway, sweeping across the mining face and expelling exhaust through the return airway. This trend is visualised in Fig. 2, where observing the gas concentration variations over a 24-hour period (1440 minutes), it is a clear temporal dependency, and the data exhibit an apparent co-movement pattern across the time series. Similar temporal trends are also evident in the wind speed and temperature variables. However, although the gas concentration migrates along the airflow within the working face, a straightforward application of correlation fails to capture this relationship.

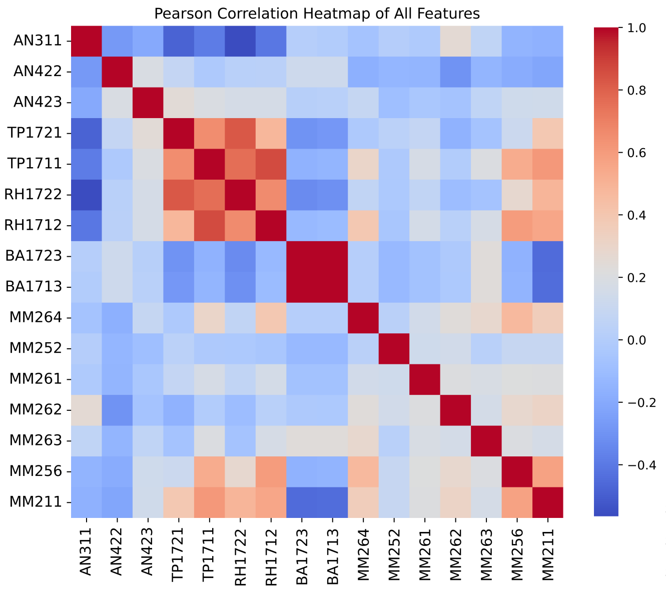

Figure 3: Pearson Correlation Heatmap of All Sensor Data in the Longwall Mining Face

Sensor point MM264 measures the gas concentration at the upper corner of the mining face, located at the intersection of the return airway and the roof. This region is a critical accumulation zone where methane from the goaf and emissions generated during mining operations converge due to ventilation-induced airflow, resulting in elevated localised concentrations. Given its significance in monitoring gas dynamics, this point is selected as the multivariate prediction target, while data from other sensors are classified as either endogenous or exogenous inputs. Endogenous data include sensor readings directly associated with MM264, such as those from MM263 and MM256, which are located downstream along the gas flow path. During this process, gas concentration gradually propagates from upstream positions, such as MM261, to downstream sensors like MM256, creating a chain of dynamic gas concentration distribution. In contrast, exogenous data utilise additional information from peripheral areas of the mining face, including the physical properties of the coal seam (e.g., gas content, gas permeability) and environmental factors (e.g., face temperature, humidity, and atmospheric pressure). Exogenous data provide critical external context for understanding the spatio-temporal variations in gas concentration patterns.

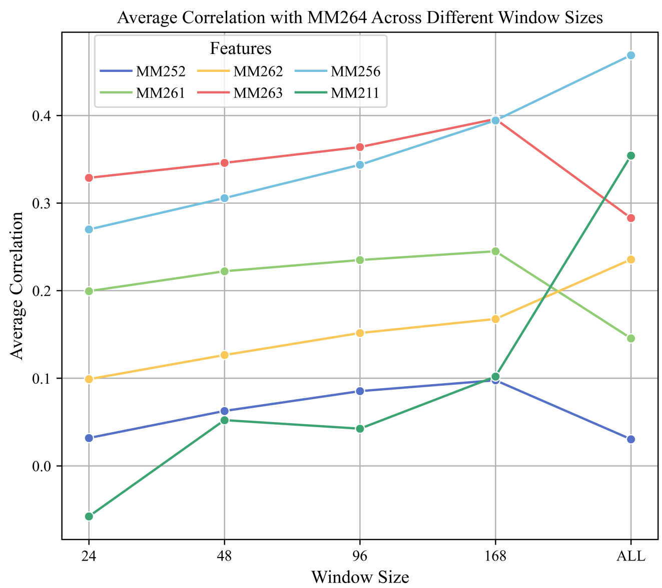

Figure 4: Pearson Correlation of MM264 with Various Features Across Different Window Sizes in the Longwall Mining Face

Direct correlation analysis reveals minimal linear association between nearby sensors (e.g., MM263 and MM264), as shown in Fig. 3. The generally low correlation values across the dataset underscore the complexity and nonlinearity of factors influencing gas concentration within the longwall mining face. Nonetheless, these findings offer valuable prior insights. For instance, an analysis of average correlation across different time windows (as shown in Fig. 4) illustrates the influence of window size on sensor correlation. As the window size increases, all sensor pairs display greater trend consistency at larger window sizes even when the overall correlation remains low. This observation suggests that by selecting an appropriate time window, it is possible to more effectively capture underlying patterns in gas concentration fluctuations and identify latent associations between sensors.

For all sensor configuration, as detailed in Table. 2, includes the minimum, maximum, mean, standard deviation, median, and data length for each sensor, providing a comprehensive overview of their distributions and variability. A variety of sensor types work together to deliver holistic environmental monitoring, capturing factors that impact the longwall environment. This multi-sensor data enables continuous assessment and adjustment, supporting accurate upper corner monitoring and promoting safe production.

Table 2: Statistical Characteristics of Sensor Data

Sensor	Min	Max	Mean	Std. Dev.	Median	Lengths
MM252	-0.1	30	0.038	0.121	0	9199930
MM261	0	30	0.049	0.125	0	9199930
MM262	-0.2	30	0.051	0.136	0	9199930
MM263	-2	30	0.248	0.197	0.2	9199930
MM264	-2	40	0.327	0.206	0.3	9199930
MM256	0	30	0.43	0.204	0.4	9199930
MM211	-2	30	0.7	0.151	0.7	9199930
AN311	-266	5	3.484	0.611	3.6	9199930
AN422	0	2.4	1.655	0.128	1.6	9199930
AN423	-2.4	5.3	1.498	0.33	1.4	9199930
TP1721	0	27.9	25.477	0.932	25.4	9199930
TP1711	0	31.2	28.894	0.757	28.8	9199930
RH1722	0	71	49.283	6.143	48	9199930
RH1712	0	86	68.687	7.268	69	9199930
BA1723	0	1131.7	1106.161	7.625	1105.9	9199930
BA1713	0	1130.9	1105.597	7.617	1105.3	9199930
CM861	-0.2	67.7	32.92	21.395	43.7	9199930
CR863	-8	258	75.081	55.161	78	9199930
P_864	0	435.4	86.967	29.158	94.2	9199930
TC862	0	40.5	29.898	9.898	32.9	9199930
WM868	0	6.39	1.803	1.32	2.2	9199930
AMP1_IR	-255	988	5.854	24.413	0	9199930
AMP2_IR	-255	1009	5.741	24.25	0	9199930
DMP3_IR	-255	216	4.201	17.342	0	9199930
DMP4_IR	-255	198	3.97	17.313	0	9199930
AMP5_IR	-255	121	0.414	10.966	0	9199930
V	0	100	1.347	5.997	0	9199930

3.2. *Problem Definition*

The objective of multivariate gas concentration prediction is to estimate future trends in gas concentration across the longwall mining face. This is achieved using multidimensional time-series data collected from multiple sensors, encompassing various influencing factors. Let $\mathbf{X} = \{\mathbf{x}_t\}_{t=1}^{T}$ represent the multivariate time series data collected by a network of sensors, where each $\mathbf{x}_t \in \mathbb{R}^d$ is a d-dimensional observation vector at time t. Specifically, $\mathbf{x}_t = [x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(d)}]^\top$, with $x_t^{(i)}$ indicating the reading from the i-th sensor at time t, such as gas concentration, wind speed, or temperature. The target variable, denoted as $\mathbf{Y} = \{y_t\}_{t=1}^{T}$, represents the observed gas concentration at each time t.

Based on the above definitions, the multivariate gas concentration prediction problem can be formulated as a time-series regression task, aiming to learn a non-linear mapping function $f(\cdot)$ such that:

$$\hat{y}_{t+\tau} = f(\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_{t-n+1}; \theta) \tag{1}$$

where θ represents the model parameters, n is the input time series window length, τ is the prediction horizon, and $\hat{y}_{t+\tau}$ is the model's predicted gas concentration at future time $t+\tau$.

To capture the dynamic characteristics of gas concentration, we employ a fixed-length time window of size n, utilising observations from the past n time steps to forecast the gas concentration at future time $t+\tau$. Specifically, the input data matrix $\mathbf{X}_t \in \mathbb{R}^{n\times d}$ is defined as:

$$\mathbf{X}_t = [\mathbf{x}_{t-n+1}, \mathbf{x}_{t-n+2}, \ldots, \mathbf{x}_t]^\top \tag{2}$$

The model's objective is to minimise the loss function $\mathcal{L}$ between the predicted values $\hat{y}_{t+\tau}$ and the true values $y_{t+\tau}$, typically using the mean squared error (MSE):

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N}(y_{t_i+\tau} - \hat{y}_{t_i+\tau})^2 \tag{3}$$

where N is the number of training samples.

Based on existing studies (e.g., Zhou et al. (2021); Vaswani et al. (2017)), the time window length n is commonly set to 24, 48, 96 and 168 time steps to capture periodic fluctuations and trends in the data. In this study, the time window length n is set to [24, 48, 96, 168], and the prediction horizon τ is set to 12. The prediction can thus be expressed as $f(\cdot)$, enabling it to utilise multivariate observations from the past n time steps to accurately forecast the gas concentration τ time steps ahead:

$$\hat{y}_{t+\tau} = f(\mathbf{X}_t; \theta) \tag{4}$$

By minimising the loss function $\mathcal{L}(\theta)$, we optimise the model parameters θ to improve prediction accuracy.

4. Methodology

This section provides a detailed description of the framework for multivariate gas concentration prediction in the longwall mining face using feature selection techniques. Followed by a discussion of the various feature selection methods employed, and the baseline models used for multivariate time series prediction.

4.1. *Multivariate Time Series Feature Selection Methods*

Given the extensive dataset obtained from the Upper Silesian coal basin, comprising over nine million data instances collected from 28 sensors, identifying the most relevant features is paramount for effective model training and prediction. We plan to employ four different methods for feature selection and compare their efficacy: SHAP (SHapley Additive exPlanations), PCA (Principal Component Analysis) with 95% confidence, DTW (Dynamic Time Warping), and a baseline approach which is no feature selection. SHAP offer a robust and interpretable way to determine the importance of each feature, providing clear insights into their contributions to the prediction accuracy. PCA with 95% confidence effectively reduces dimensionality by capturing the majority of the data's variance, simplifying the dataset while retaining the most significant information. DTW measures temporal similarity between time

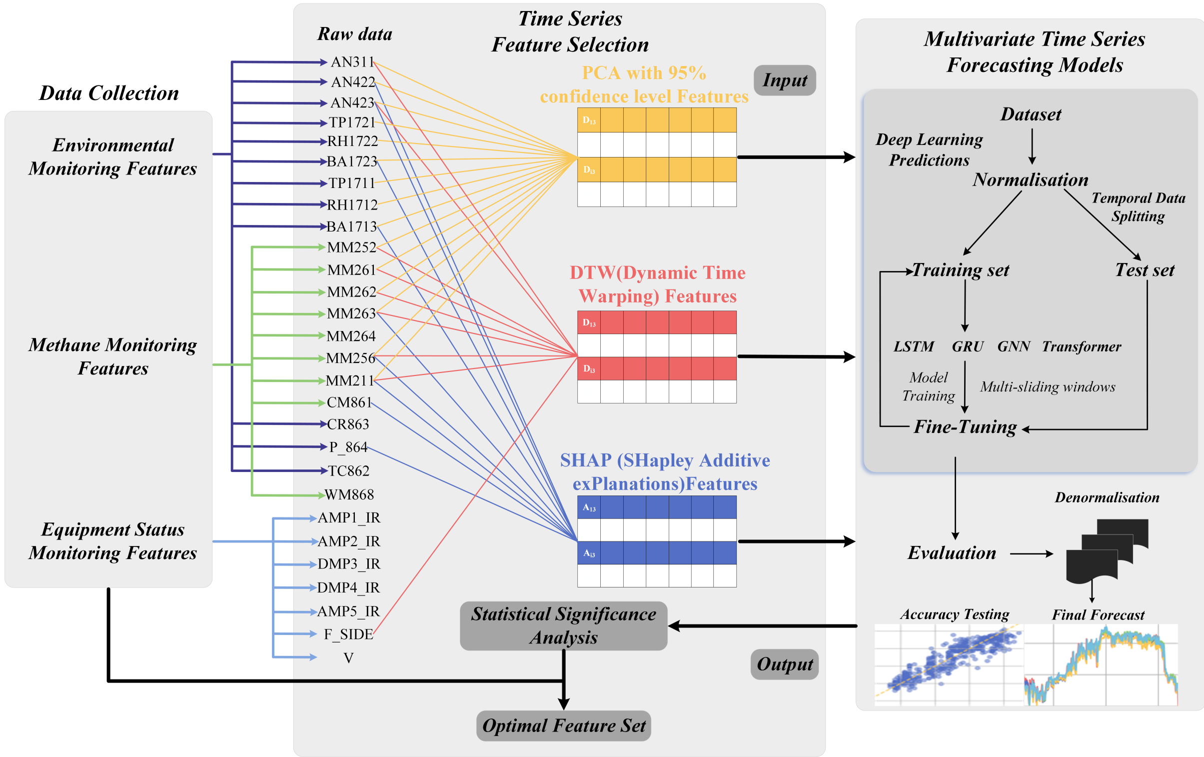

Figure 5: Framework for Multivariate Gas Concentration Prediction in Longwall Mining Face Using Feature Selection Techniques

series, making it ideal for aligning features with similar temporal patterns to the target variable, even if they are shifted in time. And by using all features serves as a control to evaluate the raw data's predictive power and the overall impact of feature selection methods.

This study compares and evaluates the effectiveness of SHAP, PCA, DTW, and a baseline approach in selecting the most relevant features, as illustrated in Fig. 5, covering data collection, feature selection, and multivariate time series prediction. This comparison seeks to identify the optimal method for enhancing the accuracy and interpretability of gas concentration predictions in longwall mining operations.

4.1.1. SHAP

SHAP (Shapley Additive exPlanations) Lundberg and Lee (2017), which harnesses the Shapley values from cooperative game theory Shapley et al. (1953), provides a solid theoretical foundation for assigning contributions to individual features within predictive models. The explanation model can be simplified as Eq.5.

$$g(x') = \phi_0 + \sum_{i=1}^{M} \phi_i x'_i \tag{5}$$

where $x' \in \{0,1\}^M$, $\phi_i \in \mathbb{R}$ and M symbolises the quantity of simplified input features, as also proposed in LIME Ribeiro et al. (2016). Portraying an explanatory model g as a linear function wherein binary variables signify the inclusion or exclusion of input features from the original model f. In LIME, the contribution of each feature ϕ_i is represented through a linear summation of the model's predictions, presuming the independence of features. Contrastingly, SHAP, which is an instantiation of an additive feature attribution method, employs Shapley values to apportion the contribution of each feature to the model's prediction, which can be defined as in Eq.6

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \left[f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S) \right] \tag{6}$$

where F represents the ensemble of features, S denotes a subset of F excluding feature i, and ϕ_i indicates the predictive outcome of model f employing solely the feature set S. The function $f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)$ gives the prediction with feature i, while $f_S(x_S)$ gives the prediction without it. Each feature's influence is deduced by assessing the model's prediction in both scenarios: the inclusion and exclusion of the feature, averaged over all conceivable subsets. Consequently, SHAP transcends the basic linear model, emerging as an intricate explanatory framework that accounts for the interdependencies and interactions amongst features.

Compared with LIME, SHAP's approach is founded on theoretical underpinnings of cooperative game theory, ensuring equitable and consistent distribution of attributions amongst features. Distinct from LIME's penchant for localised approximations—prone to yielding interpretations that are accurate within a narrow context, but may falter on a universal scale—SHAP considers the full spectrum of the dataset in evaluating feature significance. And unlike PI (Permutation Importance) Breiman (2001), SHAP calculates the imputation of predictions rather than model performance, which makes it easy to interpret. The aggregate of SHAP values for all features precisely equates to the deviation of the model's prediction from a predetermined baseline. This principle resonates with the logical presumption that the sum of contributions from all features should correspond with the variation in output. Such a quality is especially beneficial for elucidating a clear and coherent delineation of feature contributions, bolstering the intelligibility and transparency of the model's interpretative process. This is the main reason why the SHAP method was chosen for this research.

4.1.2. PCA (Principal Component Analysis) with 95% Confidence

Principal component analysis (PCA) Maćkiewicz and Ratajczak (1993), a well-established dimensionality reduction technique, addresses the challenge of high-dimensional data by transforming the original features into a new set of mutually orthogonal principal components through linear transformation. These components are ranked by the variance they explain, with the first few typically capturing the majority of the data's informational content. In this study, we selected the principal components that explain 95% of the variance to reduce data dimensionality. Although PCA is an unsupervised feature selection method that does not rely on the target variable and can simplify the data by reducing the number of features, it is employed here primarily for comparison with other methods. While PCA can lower computational complexity and retain essential data information, its utility in this context is as a benchmark against more targeted approaches.

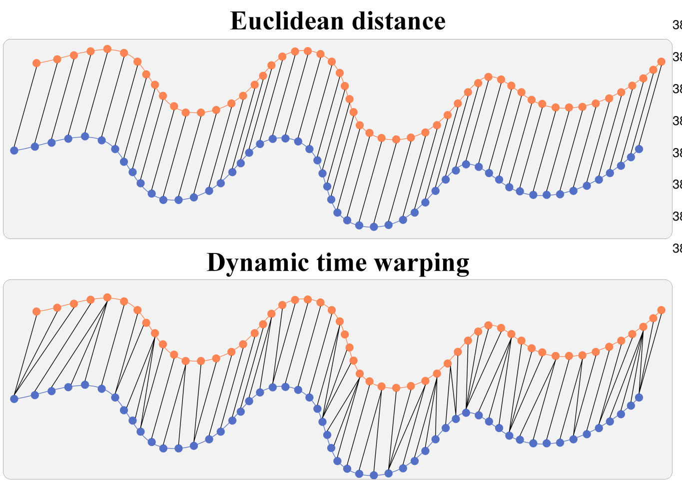

Figure 6: Euclidean Distance and Dynamic Time Warping.

4.1.3. DTW (Dynamic Time Warping)

Dynamic Time Warping (DTW) Müller (2007) offers a solution to this limitation by allowing for adjustments along the time axis, aligning two sequences to effectively identify similar patterns, especially when there are temporal offsets or delays. For assessing the similarity between two temporal sequences, regardless of their alignment or length differences. In the coal mining context, accurately tracking gas movement from one sensor to another is complicated by wind-induced timing variances. Fig. 6 illustrates a comparison between DTW and Euclidean distances.

DTW achieves this by flexibly aligning the sequences, addressing challenges associated with temporal offsets and scaling. For two sensors, $Sensor_1 = \{x_1, x_2, ..., x_n\}$ and $Sensor_2 = \{y_1, y_2, ..., y_m\}$, the DTW distance, denoted as $D[n][m]$.

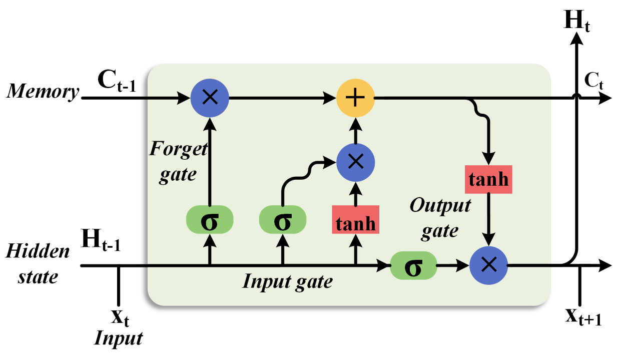

Figure 7: LSTM Units

4.2. Multivariate Time Series Prediction Models

In this study we selected four time series prediction models, which are Long short-term memory (LSTM) as baseline model Hochreiter and Schmidhuber (1997) by introducing a gating mechanism, it successfully addresses the common issues of gradient vanishing and exploding in long time series data, making it widely used in time series prediction. LSTM units control the flow of information through input gates, forget gates, and output gates as shown in Fig. 7, which allowing the LSTM to retain dependencies over long time spans. This gating mechanism enables LSTMs to selectively retain or forget information, adapting to different time series patterns.

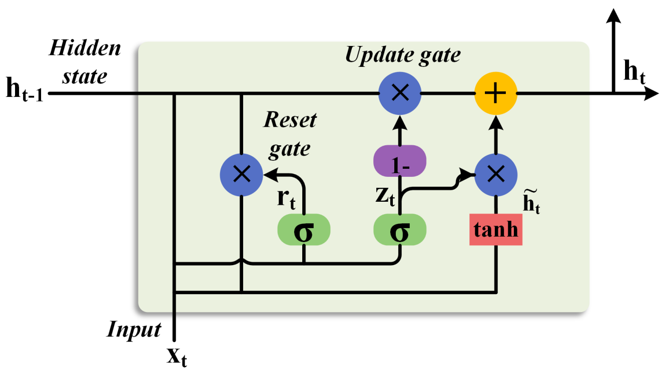

Figure 8: GRU Units

Gated recurrent units (GRU) Cho et al. (2014) as shown in Fig. 8 is a simplified variant of LSTM, designed to reduce model complexity while maintaining the ability to handle long-term dependencies. GRU merges the input gate and forget gate into a single update gate, simplifying the computation process. Additionally, GRU uses reset gates and update gates to control the updating and resetting of information flow, balancing computational efficiency and performance.

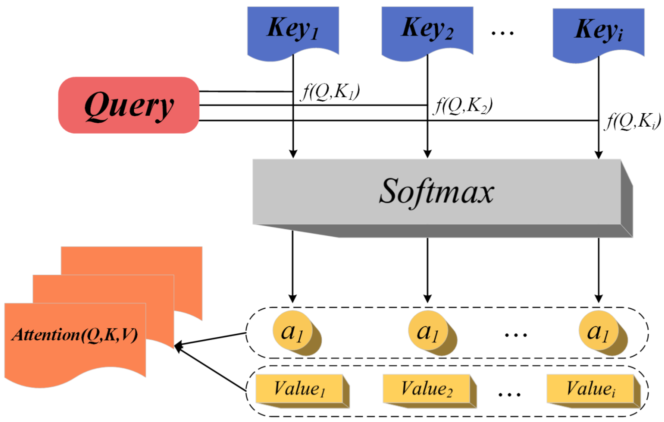

Figure 9: Attention Mechanism

The Transformer mechanism as shown in Fig. 9, by introducing a self-attention mechanism, significantly enhances the capability of time series prediction, particularly in capturing long-term dependencies. Compared to traditional RNN models, Transformers Vaswani et al. (2017) are better at capturing global dependencies in long time series and offer higher parallel efficiency during training. This makes Transformers especially suitable for applications that require handling high-dimensional, multivariate time series data, such as comprehensive environmental data analysis in coal mine safety monitoring.

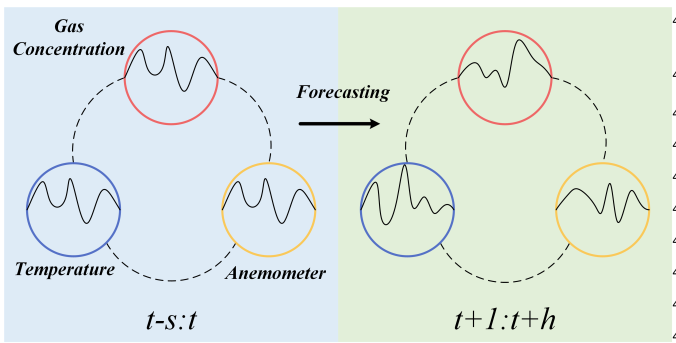

Figure 10: Graph Neural Networks Prediction

Despite the commendable performance of models such as LSTM, GRU and Transformers in handling time-series data, commonly applied in industrial settings such as real-time gas concentration monitoring in longwall mining faces Liu and Meidani (2024), these approaches inherently struggle with capturing the non-Euclidean spatial dependencies underlying complex gas diffusion patterns. As a result, interpreting interdependencies among various monitoring points remains challenging, limiting the overall accuracy of spatiotemporal predictions.

In contrast, Graph Neural Networks (GNNs) Scarselli et al. (2008) explicitly model spatial relationships by representing monitoring points as graph nodes and their interactions as edges Cheng et al. (2022); Xu et al. (2023). As shown in Fig. 10, variables such as gas concentration, temperature, and airflow velocity within the historical interval $[t - S, t]$ form the input nodes, with spatial or physical connections defined as edges. This structure leverages both spatial and temporal dimensions, facilitating forecasts from $t + 1$ to $t + h$ that more accurately represent diffusion pathways and delayed propagation effects across multiple monitoring locations.

4.3. Evaluation Criterion

We use mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) as performance metrics to evaluate the models, defined in Eq. 7, Eq. 8, Eq. 9, and Eq. 10 respectively.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{7}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{9}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{10}$$

5. Experiments and Results

5.1. Feature Selection Results

The final results of the feature selection methods as shown in Fig. 11 including Principal Component Analysis (PCA), SHAP, and Dynamic Time Warping (DTW)—distinguished by different colors. The figure illustrates the layout of sensors in a coal mining face and the application of the feature selection methods to each sensor, excluding the pipeline extraction of gas P864. The primary target is the MM264 sensor at the upper corner, highlighted with a purple arrow to indicate it as the targets for feature selection. Sensors marked with blue borders represent features selected using the SHAP method; red borders indicate those selected using DTW; and yellow borders denote features selected using PCA95. Some sensors, such as MM256, MM263, and MM211, are consistently selected by multiple methods, highlighting their significance in the feature selection process. The baseline approach utilizes all 28 sensors for model training and prediction. In contrast, PCA, used as an unsupervised method, selected fifteen features. The supervised SHAP method identified the nine most relevant features,

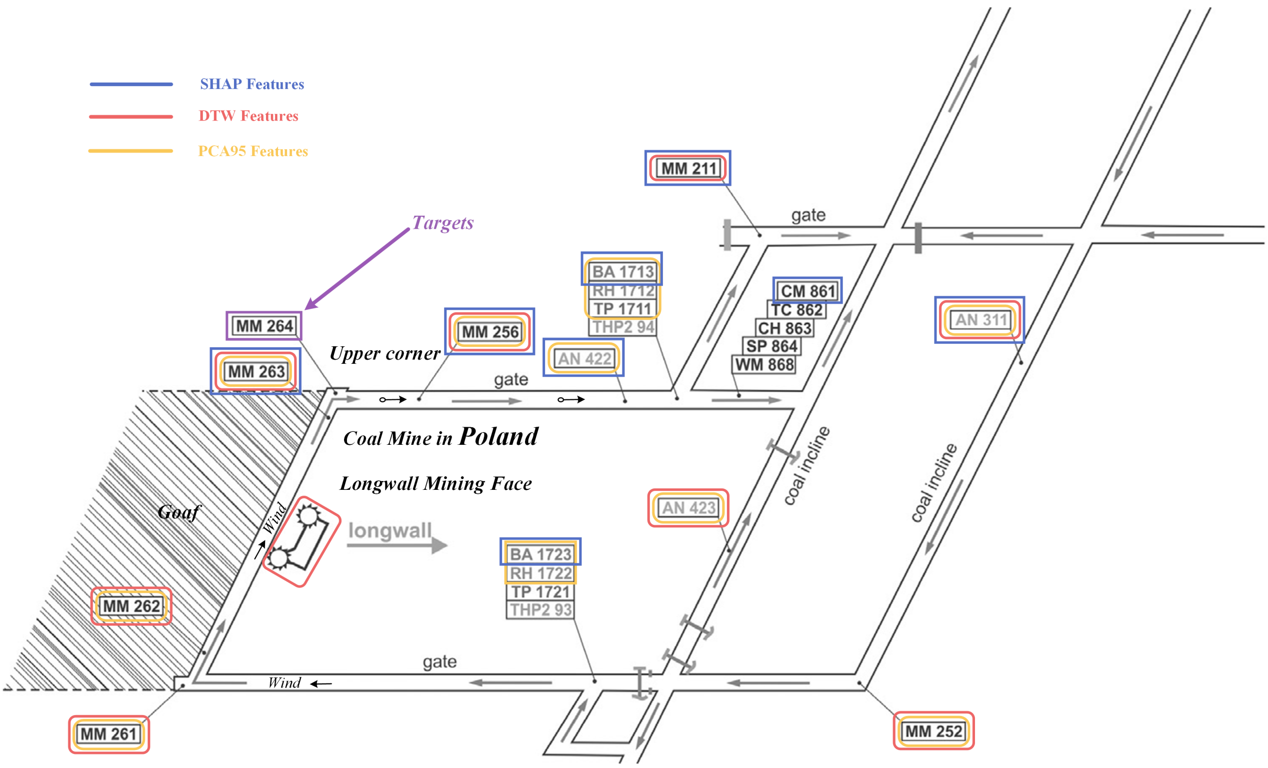

Figure 11: Comparison of Features Selected for Gas Concentration Prediction.

and DTW also selected nine features for comparison. The diagram also shows the wind direction and the layout of the mine passages.

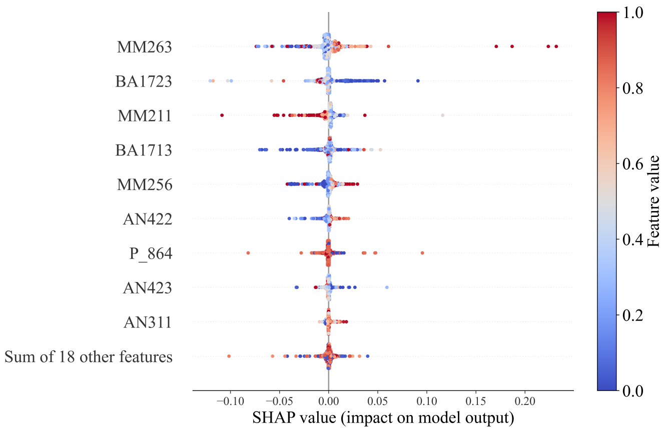

Figure 12: SHAP Summary Plot of Feature Importance and Distribution

A summary plot of SHAP values illustrates the importance of each feature in the gas concentration prediction model and the direction of their impact, as shown in Fig. 12. Where the vertical axis lists features that significantly affect prediction performance, including MM263, BA1723, and MM211. Positive or negative SHAP values indicate whether each feature has a positive or negative influence on the model output, while the color reflects the magnitude of the feature values. This representation reveals the importance of key features, facilitating the identification and analysis of the main driving factors behind gas concentration changes in longwall mining environments, thereby providing a reference for feature selection and model optimization.

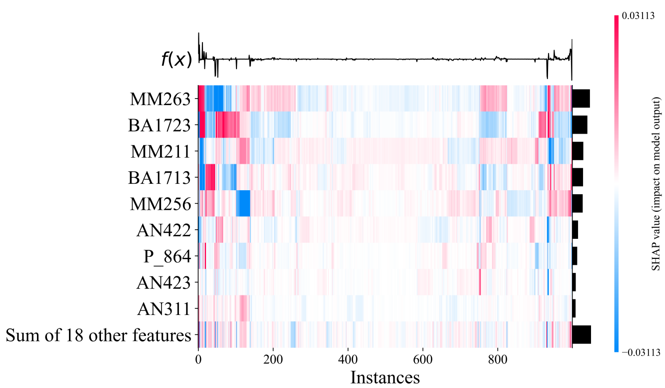

Figure 13: SHAP Values for Feature Impact Analysis on Model Output

To demonstrate how the impact of each feature on the model output varies across different instances, a SHAP value heatmap is utilized in Fig. 13. The horizontal axis represents different instances, while the vertical axis lists a set of key features. Colors ranging from blue (negative impact) to red (positive impact) depict the dynamic contribution of each feature to the prediction output. The model output trend at the top provides context for the feature impacts, making it easier to observe the relationship between feature variations and output changes. This indicates that the influence of key features dynamically changes under different operating conditions, revealing complex relationships within the predictive model and aiding in understanding the actual role of features in gas concentration prediction.

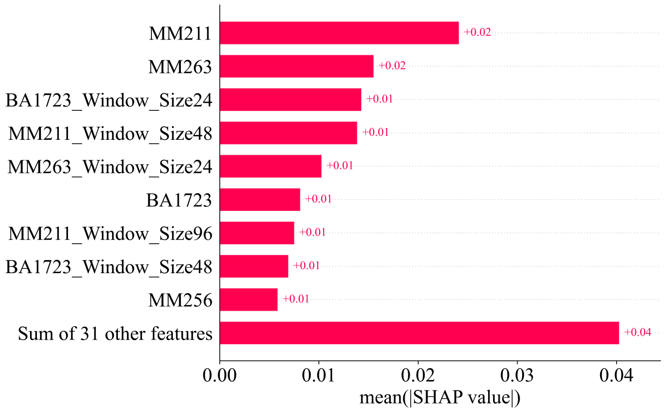

Figure 14: Mean Absolute SHAP Value Ranking of Key Features across Different Window Sizes

Furthermore, the ranking of average absolute SHAP values of key features under different time windows quantifies the influence of each feature on gas concentration prediction. As in Fig. 14, the horizontal axis represents the average SHAP value of the features, where larger values indicate a more important impact on the model output. Features MM211 and MM263 exhibit the highest average SHAP values, indicating they play a major role in the prediction. Additionally, features like BA1723 and MM211 have a high impact under different time windows, reflecting the dynamic influence of time window selection on feature importance.

In Fig. 15, an analysis of the average SHAP values of features MM211 and MM263 under different time windows further evaluates the impact of the time window on feature importance. The distribution of average SHAP values for the MM211 feature at different window sizes (24, 48, 96, and 168) shows that its influence on the prediction is significantly higher at a window size of 48 compared to other settings. Similarly, the average SHAP value distribution for the MM263 feature indicates that its influence is most prominent at a window size of 24, gradually diminishing as the window size increases.

5.2. Experimental Details

In this study, we evaluated four deep learning architectures for multivariate gas concentration prediction: Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRU), Transformers, and Graph Neural Networks (GNN), for a more detailed architecture see Table. 3. The dataset was partitioned with an 80:20 ratio, reserving 80% for training and 20% for testing. Each model was trained and tested on identical datasets with input sequence lengths of $[24, 48, 96, 168]$ time steps and a prediction horizon of 12 time steps.

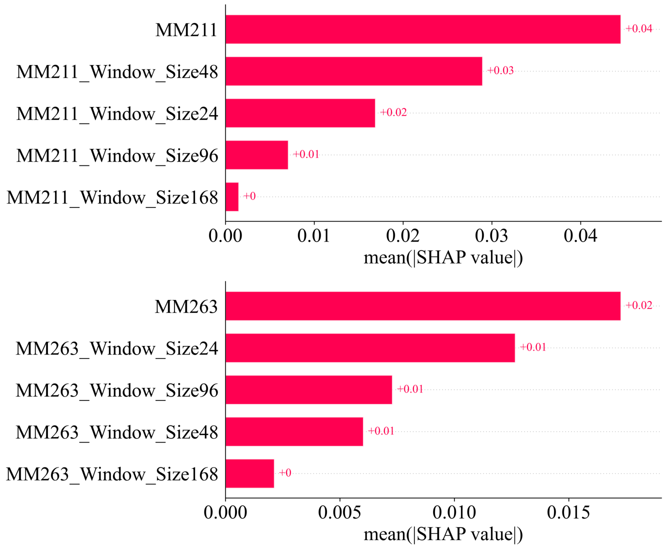

Figure 15: Mean SHAP Value Analysis of Feature Importance Across Different Time Windows for MM211 and MM263

Experimental was conducted on AMD EPYC 7542 32-Core Processor with NVIDIA GeForce RTX 3090 GPUs, running on Ubuntu 22.04.1 LTS Jellyfish. The software setup included Python 3.7 and PyTorch with CUDA 11.2.

All models were trained using consistent hyperparameters and training configurations to ensure a fair comparison. Common settings included Adam optimizer with an initial learning rate of 1×10^{-4}, which decayed by a factor of 0.5 after each epoch to facilitate convergence Zhou et al. (2021). Training was conducted for up to 100 epochs with early stopping based on validation loss to prevent overfitting. A batch size of 64 was used across all experiments and a teacher forcing ratio of 0.5 was applied during training.

5.3. Gas Concentration Prediction Performance

We evaluated multiple models across various sliding time windows, employing sequence lengths of 24, 48, 96, and 168. This approach captures temporal dependencies at multiple scales. The dataset was split into an 80%:20% ratio for training and testing, respectively, and we compared four baseline models—Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Transformer, and Graph Neural Network (GNN)—alongside their feature selection-enhanced counterparts: SHAP, PCA95, and DTW-improved versions. The results are presented in Tables 4 and 5. We report MSE, MAE, RMSE, and MAPE to provide a comprehensive assessment of predictive accuracy and error characteristics. MSE and RMSE are more sensitive to larger errors, MAE captures the average absolute error, and MAPE reflects the relative percentage error with respect to actual values.

Among the evaluated models, the SHAP GNN consistently demonstrates low prediction errors across all sequence lengths. For example, with a sequence length of 24, the SHAP GNN attains an MSE of 0.0406 and MAE of 0.1318, coupled with

Table 3: Summary of Model Hyperparameters

Method	**Hyperparameters**	**Value**
LSTM	Hidden Dimension	512
	Number of Layers	3 layers
	Activation Function	Tanh and Sigmoid
	Sequence Length	[24, 48, 96, 168]
	Batch Size	64
	Optimizer	Adam; Initial learning rate of $1e^{-4}$, decaying two times smaller every epoch
	Dropout Rate	0.05
	Number of Runs	3 times with random seed
	Epochs	100
	Early Stopping	Yes
GRU	Hidden Dimension	512
	Number of Layers	3 layers
	Activation Function	Tanh and Sigmoid
	Sequence Length	[24, 48, 96, 168]
	Batch Size	64
	Optimizer	Adam; Initial learning rate of $1e^{-4}$, decaying two times smaller every epoch
	Dropout Rate	0.05
	Number of Runs	3 times with random seed
	Epochs	100
	Early Stopping	Yes
Transformer	Hidden Dimension	512
	Number of Layers	4 Encoder/Decoder layers
	Activation Function	ReLU
	Sequence Length	[24, 48, 96, 168]
	Batch Size	64
	Optimizer	Adam; Initial learning rate of $1e^{-4}$, decaying two times smaller every epoch
	Dropout Rate	0.05
	Number of Runs	3 times with random seed
	Epochs	100
	Other Parameters	Attention Heads = 8
	Early Stopping	Yes
GNN	Hidden Dimension	512
	Number of Layers	Determined by Block Size = 3
	Activation Function	ReLU
	Sequence Length	[24, 48, 96, 168]
	Batch Size	64
	Optimizer	Adam; Initial learning rate of $1e^{-4}$, decaying two times smaller every epoch
	Dropout Rate	0.05
	Number of Runs	3 times with random seed
	Epochs	100
	Early Stopping	Yes

Table 4: Baseline and Feature Selection Enhancements for Multivariate Long-Sequence Time-Series Prediction Results (MSE and MAE)

Methods	**24**		**48**		**96**		**168**	
	MSE	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**
LSTM Hochreiter and Schmidhuber (1997)	0.0825	0.1544	0.1292	0.1896	0.0940	0.1697	0.1023	0.1729
GRU Cho et al. (2014)	0.0821	0.0945	0.1048	0.1843	0.0893	0.1571	0.0869	0.1462
Transformer Vaswani et al. (2017)	0.0774	0.1245	0.0863	0.1118	0.0815	0.1161	0.0872	0.1158
GNN Scarselli et al. (2008)	0.0429	0.1404	0.0428	0.1403	0.0427	0.1398	0.0443	0.1392
SHAP LSTM	0.0720	0.1124	0.0721	0.1085	0.0737	0.1067	0.0671	0.1057
SHAP GRU	0.0746	0.1024	0.0750	0.1006	0.0773	0.1053	0.0745	0.1009
SHAP Transformer	0.0751	0.1003	0.0701	0.1000	0.0884	0.1112	0.0826	0.1048
SHAP GNN	0.0406	0.1318	0.0418	0.1332	0.0402	0.1262	0.0410	0.1301
PCA95 LSTM	0.0924	0.1507	0.1024	0.1804	0.0960	0.1574	0.1023	0.1692
PCA95 GRU	0.0912	0.1497	0.1020	0.1737	0.1001	0.1666	0.0976	0.1596
PCA95 Transformer	0.0803	0.1208	0.0869	0.1210	0.0870	0.1223	0.0868	0.1138
PCA95 GNN	0.04285	0.1402	0.0428	0.1403	0.0470	0.1459	0.0426	0.1392
DTW LSTM	0.0774	0.1113	0.0788	0.1178	0.0772	0.1143	0.0802	0.1148
DTW GRU	0.0832	0.1134	0.0818	0.1084	0.0796	0.1070	0.0828	0.1061
DTW Transformer	0.0917	0.1170	0.0894	0.1137	0.0865	0.1008	0.0776	0.1013
DTW GNN	0.0428	0.1393	0.0430	0.1405	0.0430	0.1398	0.0427	0.1403

an RMSE of 0.2015 and MAPE of 0.4834. These values indicate that the SHAP GNN not only maintains high accuracy but also effectively handles instances where larger prediction errors might occur. Its stability is evident even at a sequence length of 168, where MSE remains at 0.0410 and MAE at 0.1301, suggesting that the model adapts well to longer input sequences without significant performance deterioration.

Comparatively, the original GNN model also shows good performance but with slightly higher error metrics than the SHAP GNN. For example, with a sequence length of 96, the original GNN has an MSE of 0.0427, whereas the SHAP GNN's MSE is 0.0402. This difference may be attributed to the effectiveness of the SHAP feature selection method in extracting important features, thereby reducing the MSE and RMSE, which enhances predictive accuracy, especially for larger errors. Baseline models such as LSTM and GRU exhibit relatively high error metrics without feature selection. For instance, the original LSTM with a sequence length of 48 has an MSE of 0.1292, MAE of 0.1896, RMSE of 0.3595, and MAPE of 0.6346, indicating larger average prediction errors and less accuracy in predicting instances with larger errors. Incorporating SHAP feature selection reduces the error metrics of the LSTM model; at the same sequence length, the SHAP LSTM reduces the MSE to 0.0721, MAE to 0.1085, RMSE to 0.2685, and MAPE to 0.4925, demonstrating that the feature selection method effectively reduces the average absolute prediction error and relative error.

The original GNN model also yields robust results, though its error metrics are slightly higher than those of the SHAP GNN. For instance, at a sequence length of 96, the original GNN has an MSE of 0.0427, whereas the SHAP GNN reduces it to 0.0402. This improvement highlights the effectiveness of SHAP-based feature selection, which appears to preserve and emphasise critical features more effectively than methods like PCA95. In contrast, baseline models such as LSTM and GRU without feature selection generally report higher error metrics. For example, an LSTM at a sequence length of 48 exhibits an MSE of 0.1292 and MAE of 0.1896, values that are notably reduced when SHAP feature selection is applied (MSE = 0.0721, MAE = 0.1085). The incorporation of SHAP thus aids these models in better identifying and focusing on key input features, subsequently lowering both absolute and relative prediction errors.

The effect of feature selection methods varies by model. While SHAP consistently lowers the errors across different model architectures—most notably with GNN—PCA95 sometimes increases error metrics, likely due to the loss of critical feature information through dimensionality reduction. For example, the PCA95 LSTM at a sequence length of 24 records an MSE of 0.0924, exceeding the original LSTM's MSE of 0.0825. Similarly, the Transformer model performs well at shorter sequence lengths but does not show a clear advantage at longer intervals, even after applying SHAP. For example, though the SHAP Transformer improves upon the original Transformer at a sequence length of 48 (MSE from 0.0863 down to 0.0701), its performance does not continue to improve

Table 5: Baseline and Feature Selection Enhancements for Multivariate Long-Sequence Time-Series Prediction Results (RMSE and MAPE)

Methods	**24**		**48**		**96**		**168**	
	RMSE	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**
LSTM Hochreiter and Schmidhuber (1997)	0.2873	0.6054	0.3595	0.6346	0.3067	0.5620	0.3199	0.6374
GRU Cho et al. (2014)	0.2713	0.6031	0.3238	0.8112	0.2988	0.7405	0.2948	0.6822
Transformer Vaswani et al. (2017)	0.2783	0.5243	0.2939	0.5018	0.2855	0.5209	0.2954	0.5525
GNN Scarselli et al. (2008)	0.2070	0.5032	0.2069	0.4963	0.2066	0.4831	0.2105	0.5122
SHAP LSTM	0.2683	0.4979	0.2685	0.4925	0.2716	0.4945	0.2591	0.4949
SHAP GRU	0.2732	0.4961	0.2739	0.5033	0.2781	0.5045	0.2729	0.4918
SHAP Transformer	0.2740	0.4956	0.2649	0.4890	0.2973	0.5295	0.2874	0.5019
SHAP GNN	0.2015	0.4834	0.2045	0.4719	0.2004	0.4632	0.2024	0.4873
PCA95 LSTM	0.3040	0.5709	0.3200	0.6174	0.3099	0.5402	0.3198	0.5889
PCA95 GRU	0.3019	0.7403	0.3195	0.8383	0.3164	0.8430	0.3124	0.6867
PCA95 Transformer	0.2834	0.4899	0.2948	0.5417	0.2949	0.5064	0.2946	0.5068
PCA95 GNN	0.2090	0.5178	0.2070	0.5026	0.2165	0.5068	0.2064	0.5126
DTW LSTM	0.2783	0.4931	0.2808	0.4971	0.2779	0.4944	0.2832	0.5152
DTW GRU	0.2885	0.5077	0.2860	0.5001	0.2821	0.4898	0.2878	0.4883
DTW Transformer	0.3029	0.5037	0.2991	0.5291	0.2941	0.5094	0.2786	0.4933
DTW GNN	0.2069	0.4989	0.2074	0.5015	0.2073	0.4921	0.2067	0.4986

at longer sequence lengths, indicating that the Transformer's capacity for modelling extended input sequences may not fully align with these feature selection strategies.

In summary, the SHAP GNN model demonstrates superior overall performance, consistently delivering low MSE, MAE, RMSE, and MAPE values across diverse sequence lengths. This finding underscores the potential of SHAP to highlight critical variables more effectively than PCA or DTW. Although other models also benefit from SHAP-based feature selection to varying degrees, the gains are most pronounced for the GNN, suggesting that integrating topological structures with judiciously selected features is especially beneficial for accurate and stable gas concentration predictions.

5.4. Final Predictions

The final predictions are displayed in Fig. 16, based on the results summarised in Table. 4 and Table. 5. In particular, Fig. 16 (a)-(d) present the predicted time-series curves of gas concentration accros four different models. Fig. 16 (e)-(h) illustrate scatter plots of prediction accuracy, with the 45-degree diagonal line representing ideal agreement between predicted and target values; the closer the points are to this line, the higher the model's prediction accuracy. Fig. 16 (i)-(l) further provide a magnified view of specific regions from the second layer of scatter plots, highlighting the effects of different feature selection methods on local prediction accuracy in gas concentration prediction.

The performance of different models—including LSTM, GRU, Transformer, and GNN—exhibits notable variations in gas concentration prediction. As illustrated in Fig. 16. It can be observed that LSTM and GRU exhibit larger prediction errors in certain intervals, while the Transformer and GNN models (particularly the SHAP-based variants) produce predictions that are closer to the target values, as shown in Fig. 16 (c) and (d). In handling outliers, the SHAP-based Transformer and GNN models (in blue) outperform models using PCA and DTW.

In Fig. 16 (e)–(h), the comparison between the models' predicted values and the actual values is illustrated through scatter plots, where GNN model (Fig. 16 (h)) is generally denser and closer to the 45-degree line, especially with the assistance of the SHAP feature selection method, which further enhances prediction accuracy. Moreover, in magnified local regions Fig. 16 (i)-(l), the impact of SHAP feature selection on enhancing model accuracy becomes more pronounced, rendering the model more sensitive in selecting critical features. The effects of different feature selection methods on fine-grained predictions are further demonstrated. Notably, the prediction results using the SHAP feature selection method (marked with blue points) are denser and closer to the diagonal line, indicating the efficacy of SHAP in refining feature selection. This leads to higher prediction accuracy in local regions and reduces the deviation of outlier points. This outcome is particularly evident in the GNN model (Fig. 16 (l)); in regions where all four feature selection methods perform well, the SHAP method predicts even more accurately, causing the blue scatter points to almost adhere to

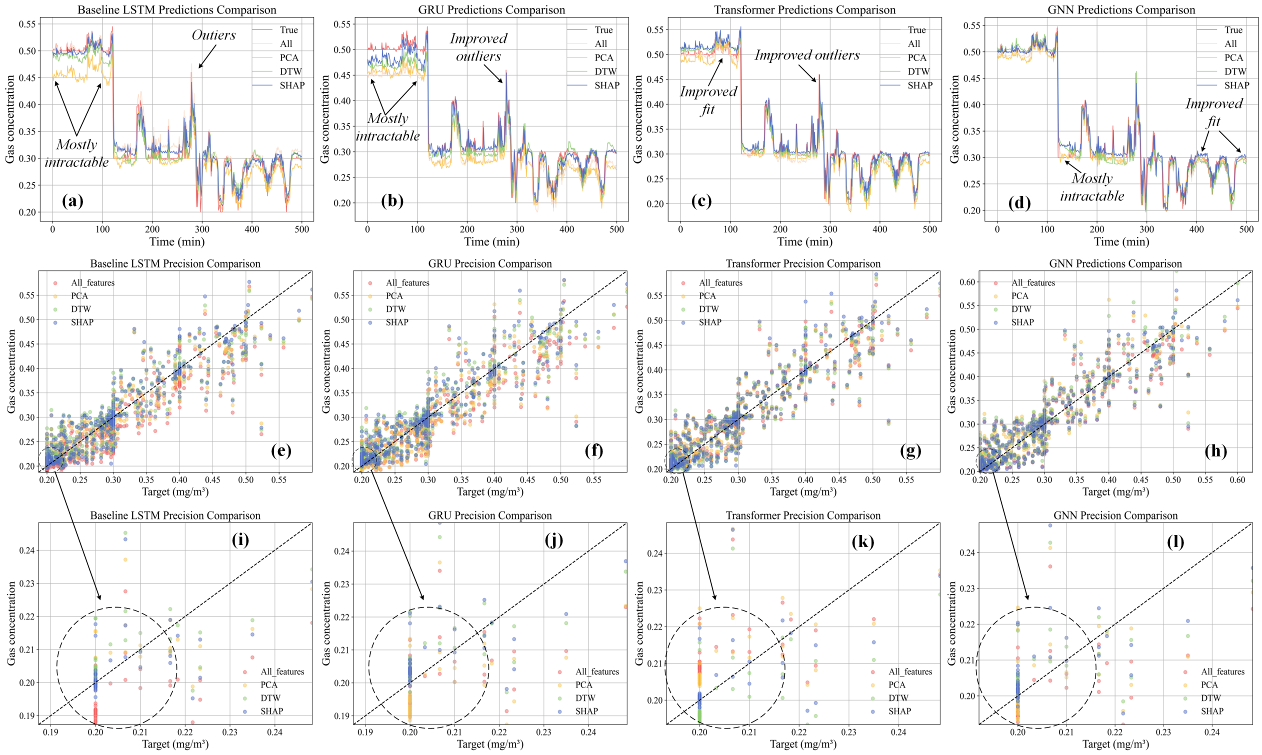

Figure 16: Performance comparison of feature selection methods across different prediction models.

the diagonal line, demonstrating strong predictive consistency.

The experimental results show that SHAP-based feature selection outperforms traditional methods like PCA and DTW in multivariate time series prediction. SHAP more effectively captures complex patterns, improves accuracy, and enhances model stability by reducing residuals and managing outliers.

Fig. 17 illustrates the computational time per epoch (in seconds) for various predictive models—LSTM, GRU, Transformer, and GNN—and their enhanced versions using SHAP, PCA95, and DTW feature selection methods. Models are evaluated across prediction horizons of 24, 48, and 168 time steps, represented by red, green, and blue bars, respectively. The top graph compares the original feature set with SHAP-selected features, showing that SHAP-enhanced models, particularly the Transformer and GNN, generally reduce computational time per epoch. The bottom graph contrasts PCA95 and DTW feature selection methods, indicating that PCA95-enhanced models exhibit consistent computational times across different horizons, while DTW-selected features may increase computational time in GRU and Transformer models due to added complexity.

6. Conclusion

This study underscores the crucial role of advanced feature selection in predicting gas concentrations at longwall mining faces. By applying and comparing four feature selection techniques—SHAP, PCA, DTW, and an unfiltered baseline across multiple prediction models, SHAP emerged as the most effective method for enhancing both model accuracy and inter-

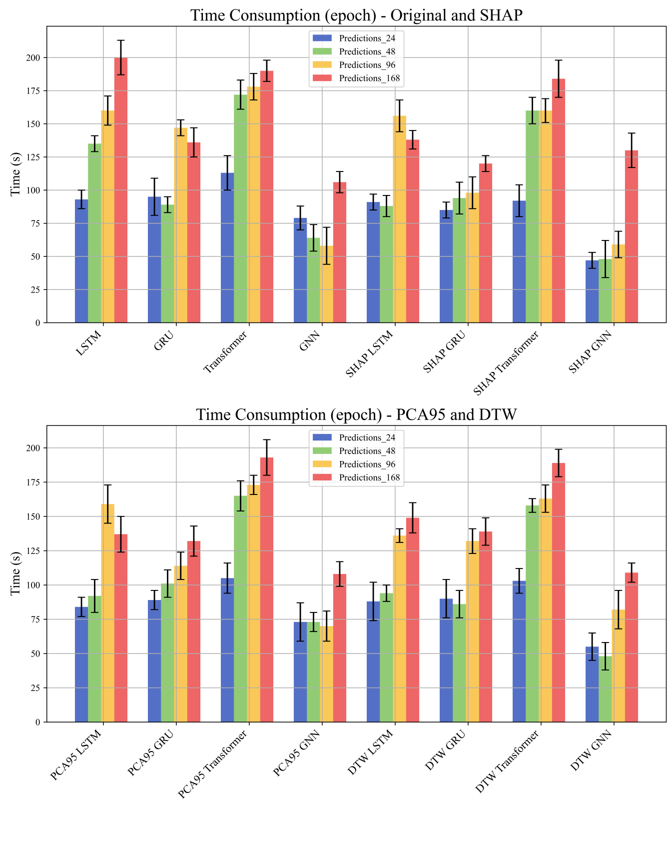

Figure 17: Time Consumption

pretability. The SHAP-based approach delivered more precise predictions while offering critical insights into the interdependencies among key variables, thereby deepening the understanding of gas concentration dynamics. These results highlight the importance of sophisticated feature selection in developing robust models, especially within complex, high-dimensional datasets typical of industrial environments. However, the study's generalisability is limited, as the dataset may not fully reflect the variability across diverse mining environments, and the SHAP-based interpretability may not sufficiently explain anomalies in model outputs, which are critical for early warning. Our future work will focus on validating these methods across diverse datasets and developing advanced interpretability techniques to better capture and explain outliers. Aims to support more robust early-warning systems, refine feature selection processes, and strengthen model reliability in practical applications.

Author contributions statement

Haoqian Chang: Conceptualisation, Methodology, Writing original draft; Xiangqian Wang: Data curation, Resources; Alexandra I. Cristea: Validation, Review & editing; Xiangrui Meng: Funding acquisition, Supervision; Zuxiang Hu: Review & editing; Ziqi Pan: Data curation;

Competing interest

We declare that we have no financial or personal relationships with other people or organisations that can inappropriately influence our paper.

Availability of Data and Material

The data used in this paper are available on request from the corresponding author.

Funding

This work was financially supported by the National Natural Science Foundation of China (Grant nos. 52374074,51874003, 51474007).

Acknowledgements

Not applicable.

References

Abdollahi, A., Pradhan, B., 2023. Explainable artificial intelligence (xai) for interpreting the contributing factors feed into the wildfire susceptibility prediction model. Science of the Total Environment 879, 163004.

Ahmed, I., Jeon, G., Piccialli, F., 2022. From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where. IEEE Transactions on Industrial Informatics 18, 5031–5042.

Aldrees, A., Khan, M., Taha, A.T.B., Ali, M., 2024. Evaluation of water quality indexes with novel machine learning and shapley additive explanation (shap) approaches. Journal of Water Process Engineering 58, 104789.

Angelov, P.P., Soares, E.A., Jiang, R., Arnold, N.I., Atkinson, P.M., 2021. Explainable artificial intelligence: an analytical review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 11, e1424.

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al., 2020. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. Information fusion 58, 82–115.

Barnewold, L., Lottermoser, B.G., 2020. Identification of digital technologies and digitalisation trends in the mining industry. International journal of mining science and technology 30, 747–757.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

Chaturvedi, A., 2023. Time series for data sciences: Analysis and forecasting.

Chen, S., Dong, S., 2020. A sequential structure for water inflow forecasting in coal mines integrating feature selection and multi-objective optimization. IEEE Access 8, 183619–183632.

Cheng, L., Li, L., Li, S., Ran, S., Zhang, Z., Zhang, Y., 2022. Prediction of gas concentration evolution with evolutionary attention-based temporal graph convolutional network. Expert Systems with Applications 200, 116944.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 .

Chushig-Muzo, D., Calero-Díaz, H., Lara-Abelenda, F.J., Gómez-Martínez, V., Granja, C., Soguero-Ruiz, C., 2024. Interpretable data-driven approach based on feature selection methods and gan-based models for cardiovascular risk prediction in diabetic patients. IEEE Access .

De, H., Jian, L., Yong, L., Xiangyang, L., Lijun, D., Yonghong, L., Pingping, H., Changshou, H., 2021. Experimental research on combination selection of observation feature of resistance variation fault in mine ventilation. Journal of China Coal Society 46, 3922–3933.

Diaz, J., Agioutantis, Z., Hristopulos, D.T., Schafrik, S., Luxbacher, K., 2022. Time series modeling of methane gas in underground mines. Mining, Metallurgy & Exploration 39, 1961–1982.

Dougherty, H., Watkins, E., Kimutis, R., 2023. A network model analysis of an unconventional gas well breach above an underground coal mine. Mining, Metallurgy & Exploration 40, 2161–2166.

Fan, C., Xu, L., Elsworth, D., Luo, M., Liu, T., Li, S., Zhou, L., Su, W., 2023. Spatial–temporal evolution and countermeasures for coal and gas outbursts represented as a dynamic system. Rock Mechanics and Rock Engineering 56, 6855–6877.

Gholami, H., Mohammadifar, A., Behrooz, R.D., Kaskaoutis, D.G., Li, Y., Song, Y., 2024a. Intrinsic and extrinsic techniques for quantification uncertainty of an interpretable gru deep learning model used to predict atmospheric total suspended particulates (tsp) in zabol, iran during the dusty period of 120-days wind. Environmental Pollution 342, 123082.

Gholami, H., Mohammadifar, A., Golzari, S., Song, Y., Pradhan, B., 2023. Interpretability of simple rnn and gru deep learning models used to map land susceptibility to gully erosion. Science of the Total Environment 904, 166960.

Gholami, H., Mohammadifar, A., Song, Y., Li, Y., Rahmani, P., Kaskaoutis, D.G., Panagos, P., Borrelli, P., 2024b. An assessment of global land susceptibility to wind erosion based on deep-active learning modelling and interpretation techniques. Scientific Reports 14, 18951.

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., Hussain, A., 2024. Interpreting black-box models: a review on explainable artificial intelligence. Cognitive Computation 16, 45–74.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural computation 9, 1735–1780.

Hooker, G., Mentch, L., Zhou, S., 2021. Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance. Statistics and Computing 31, 1–16.

Liang, R., Huang, C., Zhang, C., Li, B., Saydam, S., Canbulat, I., 2023. The fusion of data visualisation and data analytics in the process of mining digitalisation. IEEE Access .

Liu, T., Meidani, H., 2024. End-to-end heterogeneous graph neural networks for traffic assignment. Transportation Research Part C: Emerging Technologies 165, 104695.

Liu, X., Qi, P., Siarry, P., Hua, D., Ma, Z., Guo, X., Kochan, O., Li, Z., 2023. Mining security assessment in an underground environment using a novel face recognition method with improved multiscale neural network. Alexandria Engineering Journal 80, 217–228.

Liu, X., Zhang, G., Zhang, Z., 2020. A novel hybrid feature selection and modified knn prediction model for coal and gas outbursts. Journal of Intelligent & Fuzzy Systems 39, 7671–7691.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30. Curran Associates, Inc., pp. 4765–4774. URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

Maćkiewicz, A., Ratajczak, W., 1993. Principal components analysis (pca). Computers & Geosciences 19, 303–342.

Masini, R.P., Medeiros, M.C., Mendes, E.F., 2023. Machine learning advances for time series forecasting. Journal of economic surveys 37, 76–111.

Miao, D., Yao, K., Wang, W., Liu, L., Sui, X., 2024. Risk prediction of coal mine rock burst based on machine learning and feature selection algorithm. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards , 1–14.

Minh, D., Wang, H.X., Li, Y.F., Nguyen, T.N., 2022. Explainable artificial intelligence: a comprehensive review. Artificial Intelligence Review , 1–66.

Müller, M., 2007. Dynamic time warping. Information retrieval for music and motion , 69–84.

Nie, W., Cai, Y., Wang, L., Liu, Q., Jiang, C., Hua, Y., Cheng, C., Zhang, H., 2024. Coupled diffusion law of windflow-gas-dust in tunnel energy extraction processes and the location of optimal pollution control exhaust duct. Energy , 132145.

Palka, D., Blistan, P., Badura, H., 2023. Forecast of the maximum methane concentration in the longwall outlet and in the ventilation roadway. case study. Management Systems in Production Engineering 31, 398–403.

Parsa, A.B., Movahedi, A., Taghipour, H., Derrible, S., Mohammadian, A.K., 2020. Toward safer highways, application of xgboost and shap for real-time accident detection and feature analysis. Accident Analysis & Prevention 136, 105405.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. " why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1135–1144.

Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G., 2008. The graph neural network model. IEEE transactions on neural networks 20, 61–80.

Shapley, L.S., et al., 1953. A value for n-person games .

Ślęzak, D., Grzegorowski, M., Janusz, A., Kozielski, M., Nguyen, S.H., Sikora, M., Stawicki, S., Wróbel, Ł., 2018. A framework for learning and embedding multi-sensor forecasting models into a decision support system: A case study of methane concentration in coal mines. Information Sciences 451, 112–133.

Song, Y., Wang, Q., Zhang, X., Dong, L., Bai, S., Zeng, D., Zhang, Z., Zhang, H., Xi, Y., 2023. Interpretable machine learning for maximum corrosion depth and influence factor analysis. npj Materials Degradation 7, 9.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems 30.

Wen, H., Yan, L., Jin, Y., Wang, Z., Guo, J., Deng, J., 2023. Coalbed methane concentration prediction and early-warning in fully mechanized mining face based on deep learning. Energy 264, 126208.

Xu, N., Kosma, C., Vazirgiannis, M., 2023. Timegnn: Temporal dynamic graph learning for time series forecasting, in: Complex Networks.

Yoon, H., Yang, K., Shahabi, C., 2005. Feature subset selection and feature ranking for multivariate time series. IEEE transactions on knowledge and data engineering 17, 1186–1198.

Zamanzadeh Darban, Z., Webb, G.I., Pan, S., Aggarwal, C., Salehi, M., 2024. Deep learning for time series anomaly detection: A survey. ACM Computing Surveys 57, 1–42.

Zhang, G., Wang, E., 2023. Risk identification for coal and gas outburst in underground coal mines: A critical review and future directions. Gas Science and Engineering , 205106.

Zhao, D., Su, G., Cheng, G., Wang, P., Chen, W., Yang, Y., 2023. Research on real-time perception method of key targets in the comprehensive excavation working face of coal mine. Measurement Science and Technology 35, 015410.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W., 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting, in: Proceedings of the AAAI conference on artificial intelligence, pp. 11106–11115.

Zhou, J., Lin, H., Jin, H., Li, S., Yan, Z., Huang, S., 2022. Cooperative prediction method of gas emission from mining face based on feature selection and machine learning. International Journal of Coal Science & Technology 9, 51.