scientific reports

OPEN

Comparative study of ten machine learning algorithms for short-term forecasting in gas warning systems

Robert M. X. Wu[1,6]✉, Niusha Shafiabady[2,5,6], Huan Zhang[3], Haiyan Lu[1], Ergun Gide[4], Jinrong Liu[1] & Clement Franck Benoit Charbonnier[1]

This research aims to explore more efficient machine learning (ML) algorithms with better performance for short-term forecasting. Up-to-date literature shows a lack of research on selecting practical ML algorithms for short-term forecasting in real-time industrial applications. This research uses a quantitative and qualitative mixed method combining two rounds of literature reviews, a case study, and a comparative analysis. Ten widely used ML algorithms are selected to conduct a comparative study of gas warning systems in a case study mine. We propose a new assessment visualization tool: a 2D space-based quadrant diagram can be used to visually map prediction error assessment and predictive performance assessment for tested algorithms. Overall, this visualization tool indicates that LR, RF, and SVM are more efficient ML algorithms with overall prediction performance for short-term forecasting. This research indicates ten tested algorithms can be visually mapped onto optimal (LR, RF, and SVM), efficient (ARIMA), suboptimal (BP-SOG, KNN, and Perceptron), and inefficient algorithms (RNN, BP_Resilient, and LSTM). The case study finds results that differ from previous studies regarding the ML efficiency of ARIMA, KNN, LR, LSTM, and SVM. This study finds different views on the prediction performance of a few paired algorithms compared with previous studies, including RF and LR, SVM and RF, KNN and ARIMA, KNN and SVM, RNN and ARIMA, and LSTM and SVM. This study also suggests that ARIMA, KNN, LR, and LSTM should be investigated further with additional prediction error assessments. Overall, no single algorithm can fit all applications. This study raises 20 valuable questions for further research.

Keywords Machine learning algorithms, Short-term forecasting, Gas warning systems, Case study, Assessment visualization tool

As the world's largest coal producer and the fourth-largest coal reserve, China's coal mine industry accounted for approximately 46% of global coal production in 2020[1,2]. China has a significant number (3284) of coal mines with high gas content at outburst-prone risk levels across almost all 26 central coal mining provinces in China[3]. Most coal seams are now deep and require underground coal mining, which accounts for approximately 60% of the world's coal production[4]. Almost 60% of coal mine accidents were caused by methane gas (called gas in this paper) in China[5]. Gas explosion or ignition in underground mines remains an ever-present risk[6].

Therefore, the State Administration of China Coal Safety Prevention Regulations for Coal and Gas Outbursts were updated on October 1, 2019[7], requiring coal mines to deploy a gas monitoring system[8]. Many techniques and methods have been used to reduce coal mine risks, such as monitoring acoustic emission signals, electric radiation, gas emission, and micro-seismic effects on the physical properties of sound, electricity, magnetism, and thermal[9]. The existing gas monitoring systems mainly monitor the gas data, which will alarm the safety-responsive team if the gas concentration reaches the threshold limit value (TLV)[10]. However, gas accidents are associated with the complex elements of underground gas mines, which require more robust early warning systems to improve coal mining safety[11]. Machine learning (ML) (including deep learning) approaches have been widely used to explore a vast number of predictor variables in prediction ability[12,13].

[1]Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia. [2]Department of Information Technology, Peter Faber Business School, Australian Catholic University, North Sydney, Australia. [3]School of Geography, Shanxi Normal University, Taiyuan, China. [4]School of Engineering and Technology, Central Queensland University, Sydney, Australia. [5]Faculty of Science and Technology (Sydney Campus), Charles Darwin University, Sydney, Australia. [6]These authors contributed equally: Robert M. X. Wu and Niusha Shafiabady. ✉email: mingxuan.wu@uts.edu.au; robert_wumx@hotmail.com

The literature shows that ML algorithms have been used to build prediction models to avoid exceeding the gas concentration's threshold limit value (TLV)[14]. When the models predict that the gas data outputs reach the TLV, the gas monitoring system alerts the mine's safety-responsive team. However, choosing the appropriate feature selection method for a specific scenario is not trivial[15]. Based on the time scale, forecasting can be classified into four categories: very short-term forecasting (a few seconds to 30 min ahead), short-term forecasting (30 min to six hours ahead), medium-term forecasting (six hours to one day ahead), and long-term forecasting (one-day to one-week ahead)[16]. The current literature lacks research on selecting practical ML algorithms for short-term forecasting in real-time industrial applications.

This research aims to explore more efficient ML algorithms with better performance for short-term forecasting. This research uses two rounds of literature reviews, a case study, and a comparative analysis of mixed methods. The first round of the literature review focuses on top-tier publications on ML algorithms used in China's industrial applications. The second round of the literature review focuses on Q1 publications related to the performance measurement of ML algorithms. A case study method is applied to compare the ML algorithm's prediction error and predictive performance assessments. A comparative analysis is then conducted to understand research outcomes better. The following sections include literature, methodology, case study, discussions, conclusions, findings, further research, implications, and contributions.

Literature

This study conducts the first literature review of prediction error and predictive performance assessments, widely used to assess ML algorithms. The second round of the literature review focuses on understanding practical ML algorithms used in real-time industrial applications.

First round of review focusing on widely used ML algorithms

China has become a world leader in ML publications and patents[17]. Reviews of China's research on ML algorithms used in industrial applications will assist researchers and practitioners in understanding the current situation of ML approaches. The first round of the literature review focuses on the top-tier publications in both Scopus and China's most significant scientific database—CNKI—on ML algorithms used in China's industrial applications between 2016 and 2020.

Twenty-nine algorithms are found in 347 industrial applications. They include Back-Propagation (BP) (27.38%, 95 out of 347), Support Vector Machine (SVM)(24.50%, 85 out of 347), Linear Regression (LR) (8.65%, 30 out of 347), Perceptron (5.19%, 18 out of 347), Recurrent Neural Networks (RNN) (4.90%, 17 out of 347), Random Forest (RF) (3.75%, 13 out of 347), Convolutional Neural Networks (CNN) (3.17%, 11 out of 347), K-means (3.17%, 11 out of 347), AdaBoost (2.88%, 10 out of 347), Bayesian Network (2.59%, 9 out of 347), K-Nearest Neighbour (KNN) (2.02%, 7 out of 347), Stepwise Regression (1.44%, 5 out of 347), Naive Bayes (1.44%, 5 out of 347), Self-Organizing Map (SOM) (1.15%, 4 out of 347), Partial Least Squares Regression (PLSR) (1.15%, 4 out of 347), Logistic Regression (1.15%, 4 out of 347), Learning Vector Quantization (LVQ) (0.86%, 3 out of 347), Classification And Regression Tree (CART) (0.86%, 3 out of 347), Hierarchical Clustering(0.58%, 2 out of 347), C4.5 (0.58%, 2 out of 347), Radial Basis Function Networks (RBFN) (0.29%, 1 out of 347), Locally Weighted Learning (LWL) (0.29%, 1 out of 347), Projection pursuit (0.29%, 1 out of 347), Principal Component Regression (PCR) (0.29%, 1 out of 347), Partial least squares discriminant analysis (PLS) (0.29%, 1 out of 347), Linear Discriminant Analysis (LDA) (0.29%, 1 out of 347), Gradient Boosted Regression Trees (GBRT) (0.29%, 1 out of 347), Expectation Maximization (0.29%, 1 out of 347), and Ridge Regression (0.29%, 1 out of 347) (see Appendix 1).

Among the above algorithms, nine have been discussed in more than ten publications, including AdaBoost, BP, CNN, K-means, LR, Perceptron, RNN, RF, and SVM (see Appendix 2). AdaBoost is used for classification and regression tasks[18]. The classification method needs a proper training mechanism to be well applied for prediction tasks. CNN uses a convolutional layer to detect patterns in input data for classification or prediction[19]. It is usually used for image processing applications. The k-means algorithm partitions the data into clusters defined by centroids and starts with initial estimates for the centroids. These estimates are randomly generated from the datasets[20]. Therefore, AdaBoost, CNN, and K-means algorithms are unsuitable for application to gas warning systems. They will not be tested in this study.

BP, initially developed for networks of neuron-like units, is currently one of the most widely used neural networks[21]. Because of its simple structure, BP can effectively solve the approximation problem of nonlinear objective functions, such as system simulation, function fitting, pattern recognition, and other fields[22]. BP_Resilient has relatively high accuracy, robustness, and convergence speed[23]. However, when a significant network topology is selected, the standard BP algorithms have problems, such as being trapped in a local minimum and slow convergence due to the gradients with atomic magnitude[24]. Therefore, this study accepts BP_Resilient and Second Order Gradient BP (BP_SOG) as testing algorithms.

KNN is used only by a few industrial applications in China (2.02%, 7 out of 347). This study tested KNN because it is simplistic in its workings and calculations[25]. KNN can bypass the complex equation-solving process with computational efficiency[26,27] and efficiently work on forecasting accuracy in a wider variety of datasets[25,27]-sometimes without any loss of accuracy[28]. As a non-parametric and supervised learning classifier, KNN uses proximity to make classifications or predictions about the grouping of an individual data point[29] and focuses on the correlation using raw data characteristics[26]. It has been widely used in forecasting applications in economics, finance, production, and natural systems[27].

In addition to the above ML algorithms, Autoregressive Integrated Moving Average (ARIMA) is another widely used algorithm in research[30]. Although ARIMA is not a typical ML algorithm and cannot effectively capture all the details in very short-term forecasting[31], research highlights that ARIMA can account for underlying trends, autocorrelation, and seasonality and allows for flexible modeling of different types of impacts[32].

Many studies have primarily used it in identification applications and as a common approach used for addressing short-term prediction problems[27,33]. For example, ARIMA has successfully produced good short-term forecasts as the mainstay of financial forecasting[34].

Long Short-Term Memory (LSTM) is another algorithm tested in this research. Althougth it was initially observed only in China's industrial applications after this study began, LSTM stands out as a specialized type of RNN with a different structure[35–37]. LSTM is found to be more frequently used in forecasting tasks than other algorithms[38]. It may overcome the exploding/vanishing gradient problems that typically arise when learning long-term dependencies, even when the minimal time lags are very long[37,39].

Thus, ten algorithms for short-term forecasting, including ARIMA, BP_Resilient, BP_SOG, KNN, LR, LSTM, Perceptron, RF, RNN, and SVM, have been identified and tested to determine their performance.

Second round of review focusing on prediction assessments

As science becomes increasingly cross-disciplinary and scientific models become increasingly cross-coupled, standardized practices of model evaluation are more important than ever[40]. The prediction error and predictive performance assessments of the employed ML algorithms were measured using different statistical indicators[41–43]. Most studies use computational time to measure predictive performance assessment. However, it is challenging for most researchers to select suitable efficiency criteria to calculate prediction error[44].

The second round of the literature review focuses on Q1 publications related to the prediction error assessment of ML algorithms between 2020 and 2023. 45 performance criteria are found (see Appendix 3), including absolute average deviation (AAD), average absolute error (AAE), the area under the curve (AUC), commission Error (CE), cross-entropy, coefficient of variance (CoefVar), dice coefficient (DC), developed discrepancy ratio (DDR), Durbin–Watson statistic (DW), error improving rate (EIR), generalization ability (GA), Gain rate criterion (GRC), Gini index (GI), interquartile range and range (IRR), index of agreement (IoA), Kling-Gupta efficiency (KGE), mean absolute error (MAE), mean absolute deviation (MAD), mean absolute percentage error (MAPE), average bias error (MBE), median absolute percentage error (MdAPE), median percentage error (MdPE), mean error (ME), mean square error (MSE), Nash–Sutcliffe efficiency (NSE), omission error (OE), out of bag (OOB) error, overall accuracy (OA), coefficient of determination (R^2), relative absolute error (RAE), ranking mean (RM), root mean square of the successive differences (RMS), root mean squared error (RMSE), receiver operating characteristic (ROC) curve, RR variance (RR), sum of absolute errors (SAE), Se/Sy, standard error of prediction (SEP), symmetric MAPE (SMAPE), scatter index (SI), sum of squared errors (SSE), transition matrix features (TMF), t-statistic test (Tstat). It shows that no single or standard error evaluation criteria are adopted as the expected performance method for evaluating the error characteristics of ML algorithms. The reason should be that different error metrics have been used to check the effectiveness of the proposed forecasting model[31].

Appendix 3 indicates that RMSE (60%, 27 out of 45), MAE (53.33%, 24 out of 45), R2 (48.89%, 22 out of 45), and MSE (37.78%, 17 out of 45) are the most used metrics for evaluating ML algorithms between 2020 and 2023. The results are supported by other studies, including Yaseen[44] who stated MAE, RMSE, and R^2 to be the significant metrics used for the prediction evaluation, and Alhakamy et al.[45] who highlighted MAE, MSE, and RMSE as the primary metrics used to evaluate the performance criteria.

Although SAE is only used by a few researchers[44], this research believes that SAE may provide a different view of summarising all errors to evaluate the algorithms' quality. Thus, SAE is selected to test the algorithms used in this research. In addition to MAE, MSE, and RMSE, R^2, commonly known as the coefficient of determination, is a widely used metric in regression analysis that quantifies the proportion of variance in the dependent variable (response variable) that is explained by the independent variables (predictor variables) in a regression model[46], which indicates the percentage of variability in the actual values that can be explained by the variance in the estimated values[47]. However, R^2 also has bias and is highly variable for bivariate non-normal data[48]. Another reason R^2 is inadequate to assess the predictive power of models is that R^2 can be low for an accurate model, whereas an inaccurate model can yield a high R^{249}. On the other hand, R^2 is oversensitive to extreme values and insensitive to the proportional difference between "actual and predicted values"[44]. Therefore, R2 is not used in this study.

Thus, four metrics (MAE, MSE, RMSE, and SAE) are used to test the prediction error assessment of the above ten ML algorithms. Their advantages and disadvantages can be summarized as follows:

Advantages and disadvantages of MAE

MAE is one of the most prominent criteria in training neural networks[50] and is widely used because of its ease of use and simplicity[51]. It has been accepted as a crucial measure of a model's predictive accuracy[52] and the preferred measure of average model error[53]. This approach assesses the magnitude of the mean error by calculating the absolute difference between the target value and the model's predicted value[51,52,54]. In its calculation process, the MAE is derived by modeling the average of the absolute values between the original calculated and estimated values, assuming that each error has an equivalent weight[55] (see Eq. (1)).

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\left(\sum_{\mathrm{i}=1}^{\mathrm{N}}\left|\mathrm{d_i}-\mathrm{y_i}\right|\right). \tag{1}$$

The advantages of MAE are its intuitive nature and flexibility. MAE is the most straightforward measure to understand and is commonly used to interpret linear algorithms[45]. Compared with other error measures, MAE quantifies the mean error on the basis of absolute values, making it easier to understand and more interpretable[53]. MAE is more suitable in scenarios where the expected error distribution is Laplace distributed[56]. The main disadvantage of MAE is that it cannot determine the severity of an error[51]. Another disadvantage is that MAE is more limited in reflecting these distributional characteristics in the shape of the error distribution, such as the skewed, long tail, and non-standard shapes, because it is insensitive to significant differences in these distributional characteristics[57].

Advantages and disadvantages of MSE
MSE has been widely used as an ideal measure of model performance for data that follow a normal distribution because of its ease of use, mathematical simplicity, and validity[40,51]. Its value shows the difference between the predicted and observed values of a model[51]: if it is zero, it indicates that the model's prediction is perfect; if the model's error increases, its MSE value increases accordingly[45] (see Eq. (2))[58].

$$\mathrm{MSE} = \frac{1}{\mathrm{m}} \sum_{\mathrm{i}=1}^{\mathrm{m}} (|\mathrm{X_i} - \mathrm{Y_i}|)^2. \tag{2}$$

The advantage of MSE is that it is instrumental when outliers in the data need to be identified. If the model produces poor predictions, MSE helps to highlight and identify outliers by emphasizing these errors through the squared term in its function, thus assigning greater weight to these points[59]. MSE has several drawbacks when assessing model performance. As a sum-of-square approach, MSE may face more difficulties in interpreting error statistics[53]. The second drawback is that the effect of outliers may be over-amplified during the application of MSE, resulting in an inappropriate adjustment of the model for misclassified systematic errors or poor model tuning due to an overemphasis on outliers[53]. Another drawback is that MSE is limited in the scope of being appropriate for symmetric distributions[56]. For normally distributed data, MSE provides little insight into which aspects of model performance are "good" or "bad."[40]. The fourth drawback is that MSE cannot determine the severity of an error[51].

Advantages and disadvantages of RMSE
RMSE is a commonly used error function in the objective function of most optimization techniques and is a more accurate measure of accuracy[51]. It has been used as the primary metric recommended to measure the concentration of the data in the optimal fit when analyzing model performance[45]. RMSE is obtained by calculating the square root of MSE between the actual results and the expected quantity[55,56] (see Eq. (3)), which represents the average distance of the data points from the fitted line to the measurement vertical line in absolute terms[47,54]. Smaller values for all error types are considered favorable[51,55].

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{yi} - \mathrm{yi})^2}{\mathrm{n}}}. \tag{3}$$

There are at least two advantages of RMSE. The interpretability of RMSE is enhanced by square root construction[45]. RMSE may reach optimality when the errors follow a normal distribution[56]. There are at least two disadvantages. RMSE may perform less effectively (or more) in dealing with error distributions that deviate from the normal distribution[56]. RMSE does not outperform MAE in measuring the accuracy of an average model in most situations[60].

Advantages and disadvantages of SAE
SAE is used to evaluate the fitting error[61]. It is determined based on differences between the experimental and predicted data due to its ease of use and simplicity[51]. The smaller the error magnitude, the better the model's fitness[62]. The estimate is more accurate when the SAE value is closer to zero[63]. A smaller SAE indicates a better performance of the tested algorithm. However, SAE cannot determine the severity of an error, similar to MAE and MSE[44,51].

$$\mathrm{SAE} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \left|(\mathrm{yp} - \mathrm{yc})\right|. \tag{4}$$

Thus, the following sections will focus on the research method, case study, and comparative analysis. A research flowchart is developed to demonstrate the research processes. A case study method is applied for using the four metrics discussed above (MAE, MSE, RMSE, and SAE) to measure the prediction error and predictive performance assessments for the above ten ML algorithms, including ARIMA, BP_Resilient, BP_SOG, KNN, LR, LSTM, Perceptron, RF, RNN, and SVM. A comparative analysis is then conducted to understand research outcomes better.

Methodology
This study uses a five-step process to find an efficient ML algorithm with better prediction assessments for short-term forecasting (see Fig. 1). This process includes data collection and data preparation, prediction error assessment, predictive performance assessment, validation tests, and comparative analysis as follows:

Step 1: data collection and preprocessing
Data will be directly obtained from the gas monitoring system. Data pre-processing is necessary before data analysis since the raw data gathered in most industrial processes usually come with many dataset issues, such as out-of-range values, outliers, missing values, etc.

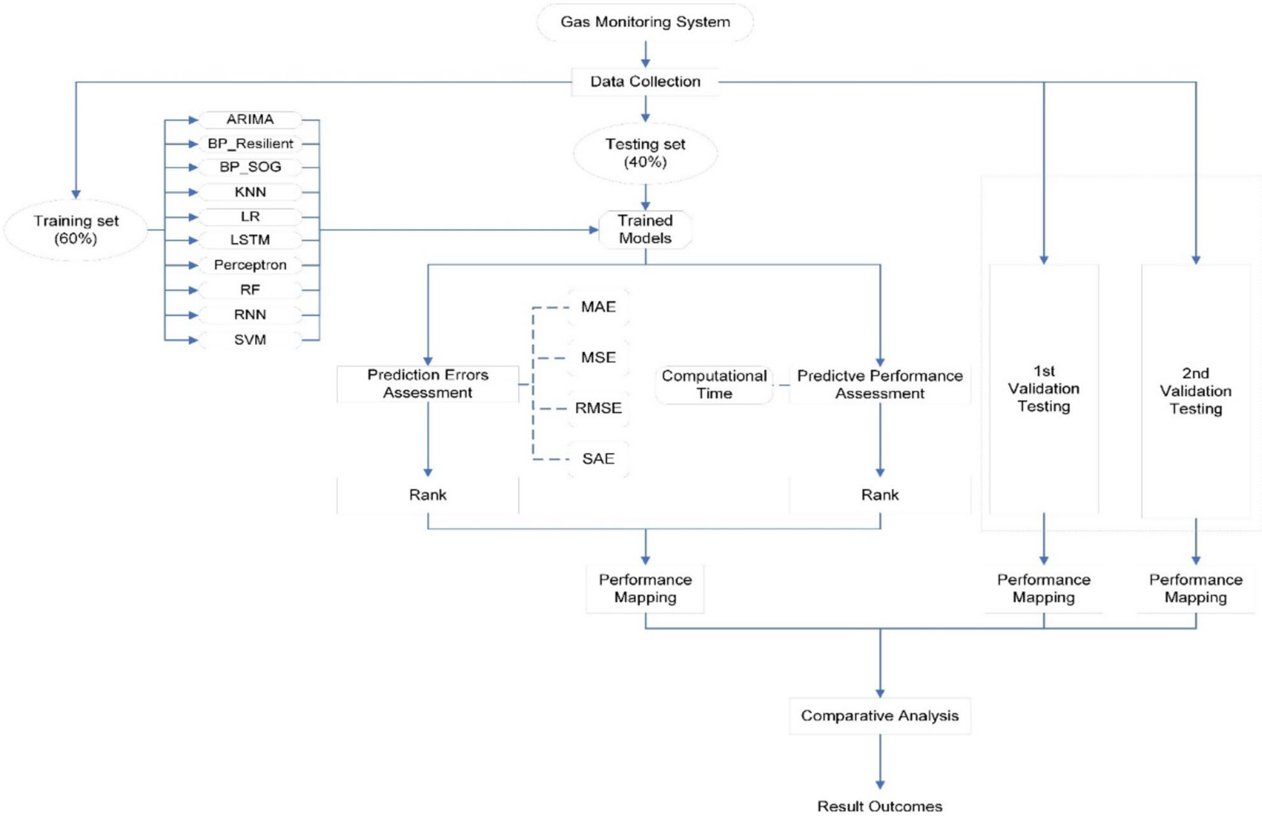

Figure 1. Research Flowchart.

A recent study highlights that although three ratios of 50:50, 60:40, and 70:30 have been used to measure the performance of models, no single ratio shows its best contribution for generating the best performance for all models by the evaluation parameters[64]. This research will split each dataset into training and testing subsets with a 60%:40% ratio. The test data will be used to examine the transferability and predictive capability of the tested algorithms on new data[36]. More testing subsets may provide sufficient records for testing the system's eventual performance, which is expected to improve the verification of test results.

Step 2: prediction error assessment
Four prediction error metrics—MAE, MSE, RMSE, and SAE—measure the prediction error of the employed modeling. The smaller the calculated metrics, the better the assessment of the tested algorithm.

Step 3: predictive performance assessment
Predictive performance assessment is another critical aspect of evaluating the computational effectiveness of ML algorithms. Computational time is used to measure the predictive performance assessment in this study. The smaller the computation time (the calculated value), the better the performance of the tested algorithm[36].

Step 4: validation tests
Two more tests are followed to validate the above outcomes. The tests use data obtained from the same sensors for two different periods.

Step 5: comparative analysis
A comparative analysis is then conducted to better understand the above outcomes.

Case study
Research background of the case study mine
Shanxi Fenxi Mining ZhongXing Coal Industry Co. Ltd (ZhongXing) is wholly owned by Shanxi Coking Coal Group Co. Ltd, a 485th in the 2020 Fortune Global 500 company located in China[65]. ZhongXing has employed a gas monitoring system that monitors data obtained from methane gas (called gas in this paper) sensors, temperature sensors, wind sensors, dust sensors, O2 sensors, CO sensors, and CO2 sensors. ZhongXing sponsors this industry-engaged research to seek a more responsive ML algorithm for short-term forecasting to predict gas concentration to avoid reaching the TLV[14]. It requests using the three-hour dataset to predict up to one hour ahead of the dataset.

Data collection and preparation

Datasets are collected from a gas sensor T050401 through the real-time gas monitoring system in the Case Study mine. The raw data gathered in most industrial processes usually comes with many quality issues, such as out-of-range values and outliers[66]. Other data quality issues—such as errors in measurement, noise, missing values, etc.- might be impacted by hardware relocation, sensor removal, added detectors, and not in-used sensors[67]. The dataset used in this research is directly obtained from the gas monitoring system in the case study mine. The sensor T050401 and its monitoring system have been reviewed and upgraded. The gas monitoring system in the case study mine does not report errors in measurement and missing values. The above data quality issues are not involved. More details about data preparation have been reported in previous studies[65].

Datasets are collected initially every 15 s from a real-time gas monitoring system between April 16 at 0:00:00 and May 16, 2022 at 23:59:59. The gas monitoring system produced four data points per minute, 240 per hour, and 5,760 dailies. A total of 28,697 valuable datasets were acquired after eliminating out-of-range values and outliers. The datasets are divided into two subsets: 60% for training and 40% for testing. All experiments of ML evaluation are conducted using a standard computer with a CPU (11th Gen Intel i7-1165G7 @ 2.80GHZ 2.80GHZ), RAM (16.0 GB), and a 64-bit operating system.

Data analysis

Prediction error assessment

Four metrics (MAE, MSE, RMSE, and SAE) are tested to measure the prediction error assessment of the employed modeling for both the training and testing datasets (see Table 1). Modeling relations between inputs and outputs is conducted using the above ten algorithms, which use the three-hour dataset to predict up to one hour ahead of the dataset (see Appendix 4).

Error assessment of each algorithm on criteria with MAE, MSE, RMSE, and SAE to the training dataset shows ARIMA with 0.0043215, 0.018554, 0.13621, and 74.408, BP_Resilient with 0.14471, 0.88631, 0.94144, and 42,900,000, BP_SOG with 0.071226, 0.81667, 0.9037, and 21,100,000, KNN with 0.017083, 0.093023, 0.305, and 294.13, LSTM with 0.056083, 0.057971, 0.24077, and 2383.1, LR with 0.0043219, 0.018554, 0.13621, and 74.414, Perceptron with 0.8956, 1.1427, 1.069, and 17,218, RF with 0.002815, 0.007159, 0.08461, and 48.464, RNN with 0.067478, 0.83028, 0.9112, and 2533.5, and SVM with 0.004578, 0.018769, 0.137, and 78.823.

Error assessment of each algorithm on criteria with MAE, MSE, RMSE, and SAE to the testing dataset shows ARIMA with 0.00151, 0.000009, 0.003048, and 17.329, BP_Resilient with 0.060858, 0.006978, 0.083532, and 8,020,000, BP_SOG with 0.030222, 0.001891, 0.043483, and 3,980,000, KNN with 0.025468, 0.16305, 0.4038, and 292.35, LSTM with 0.029636, 0.018449, 0.13583, and 1526.4, LR with 0.00151, 0.000009, 0.003048, and 17.333, Perceptron with 0.87563, 0.76756, 0.87611, and 11,479, RF with 0.001944, 0.000376, 0.01939, and 22.312, RNN with 0.067697, 0.005384, 0.073375, and 2178.5, and SVM with 0.002069, 0.000011, 0.0032586, and 23.75.

Table 1 indicates that both ARIMA and LR have the lowest error metrics in the testing dataset compared with the other algorithms in MAE (0.00151), MSE (0.000009), and RMSE (0.003048). They also have a similar outcome in SAE, such as ARIMA, with the lowest error metric (17.329), and LR, with the second lowest error (17.333). RF and SVM have higher error metrics than ARIMA and LR but lower than others. They have similar

Model	Datasets	MAE	MSE	RMSE	SAE
ARIMA	Training	0.004322	0.018554	0.136210	74.408
	Testing	0.001510	0.000009	0.003048	17.329
BP_Resilient	Training	0.144710	0.886310	0.941440	42,900,000.000
	Testing	0.060858	0.006978	0.083532	8,020,000.000
BP_SOG	Training	0.071226	0.816670	0.903700	21,100,000.000
	Testing	0.030222	0.001891	0.043483	3,980,000.000
KNN	Training	0.017083	0.093023	0.305000	294.130
	Testing	0.025468	0.163050	0.403800	292.350
LR	Training	0.004322	0.018554	0.136210	74.414
	Testing	0.001510	0.000009	0.003048	17.333
LSTM	Training	0.056083	0.057971	0.240770	2383.100
	Testing	0.029636	0.018449	0.135830	1526.400
Perceptron	Training	0.895600	1.142700	1.069000	17,218.000
	Testing	0.875630	0.767560	0.876110	11,479.000
RF	Training	0.002815	0.007159	0.084610	48.464
	Testing	0.001944	0.000376	0.019390	22.312
RNN	Training	0.067478	0.830280	0.911200	2533.500
	Testing	0.067697	0.005384	0.073375	2178.500
SVM	Training	0.004578	0.018769	0.137000	78.823
	Testing	0.002069	0.000011	0.003259	23.750

Table 1. ML Prediction Error Assessment of Datasets between 16 April and 16 May 2022.

error metrics in MAE and SAE. RF (0.001944, 22.312) has better error metrics than SVM (0.002069, 23.75). However, RF (0.000376, 0.01939) has worse error metrics than SVM (0.000011, 0.003259). BP_SOG, KNN, and LSTM have similar error metrics in MAE (0.030222, 0.025468, and 0.029636). BP_Resilient and RNN have similar error metrics in MAE (0.060858, 0.067697), MSE (0.006978, 0.005384), and RMSE (0.083532, 0.073375). BP_SOG (0.001891, 0.043483), LSTM (0.018449, 0.13583), and KNN (0.16305, 0.4038) have worse error metrics in MSE and RMSE. KNN (292.35), LSTM (1526.4), and RNN (2178.5) have significantly worse error metrics in SAE. BP_SOG (3,980,000) and BP_Resilient (8,020,000) have significantly the worst error metrics in SAE compared to other algorithms. Perceptron has the worst error metrics in MAE (0.87563), MSE (0.76756), and RMSE (0.87611) among all testing algorithms and has worse outcomes in SAE (11,479).

Table 2 shows the overall average ranks of MAE, MSE, RMSE, and SAE among ten algorithms. The results show that ARIMA has the top average rank (1) by combining MAE ranked 1, MSE ranked 1, RMSE ranked 1, and SAE ranked 1. LR has the second-top average level (1.3), combining MAE ranked one as the same as ARIMA, MSE ranked one as the same as ARIMA, RMSE ranked one as the same as ARIMA, and SAE ranked 2. RF, SVM, BP_SOG, KNN, LSTM, RNN, and BP_Resilient are followed. Perceptron has the lowest average rank (9.5). Thus, based on prediction error assessment, ARIMA and LR are the top-ranked algorithms. RF and SVM are followed. BP_SOG, KNN, LSTM, RNN, and BP_Resilient are ranked from 5 to 9, respectively. Perceptron is the last-ranked algorithm (10).

Table 2 also demonstrates that all algorithms have the same rank of prediction error between MSE and RMSE. The reason should be that the mathematical definition of RMSE is the square root of MSE[68–70]. MSE measures the relative error for a prediction[33]. In contrast, RMSE is a metric that places a relatively high weight on significant mistakes, thus making it a valuable indicator of large errors[71]. Therefore, taking root does not affect the relative ranks of models that yield a metric with the same units as the data[56]. The suggestion is thus provided that further research does not need to test both MSE and RMSE together.

Predictive performance assessment

A predictive performance assessment is followed using computational time testing. The total training and testing data are used to calculate the time required for each ML algorithm. Table 3 shows that KNN is the best algorithm with the shortest computational time (0.41683 s). Other algorithms are then followed, including RF (1.3503 s), LR (1.749 s), SVM (1.889 s), Perceptron (2.4813 s), BP_SOG (2.5108 s), BP_Resilient (2.8363 s), ARIMA (6.799 s), and RNN (34.933 s). LSTM is the worst algorithm with the longest computational time (145.19 s).

Performance mapping

To better understand the overall prediction performance of the tested models in this research, a scatter plot is developed to map the relations between prediction error assessment and predictive performance assessment (see Fig. 2). It uses the vertical axis to represent the performance rank (measuring prediction error) (see Table 2) and uses the horizontal axis to represent the computational time (measuring predictive performance) (see Table 3). Figure 2 shows that ARIMA, LR, RF, and SVM have better outcomes of prediction error assessment in all tests. Perceptron is the worst algorithm based on prediction error assessment. KNN has the best predictive performance and has the shortest computational time. LSTM has the worst predictive performance with the longest

Model	MAE	Rank	MSE	Rank	RMSE	Rank	SAE	Rank	Average Rank	Rank_Prediction Errors
ARIMA	0.001510	1	0.000009	1	0.003048	1	17.329	1	1.0	1
LR	0.001510	1	0.000009	1	0.003048	1	17.333	2	1.3	2
RF	0.001944	3	0.000376	4	0.019390	4	22.312	3	3.5	3
SVM	0.002069	4	0.000011	3	0.003259	3	23.750	4	3.5	3
BP_SOG	0.030222	7	0.001891	5	0.043483	5	3,980,000.000	9	6.5	5
KNN	0.025468	5	0.163050	9	0.403800	9	292.350	5	7.0	6
LSTM	0.029636	6	0.018449	8	0.135830	8	1526.400	6	7.0	6
RNN	0.067697	9	0.005384	6	0.073375	6	2178.500	7	7.0	6
BP_Resilient	0.060858	8	0.006978	7	0.083532	7	8,020,000.000	10	8.0	9
Perceptron	0.875630	10	0.767560	10	0.876110	10	11,479.000	8	9.5	10

Table 2. Rank of ML Models based on Prediction Error Assessment Using Datasets Obtained between 16 April and 16 May 2022.

Algorithm	KNN	RF	LR	SVM	Perceptron	BP_ SOG	BP_Resilient	ARIMA	RNN	LSTM
Time (s)	0.41683	1.3503	1.749	1.889	2.4813	2.5108	2.8363	6.799	34.933	145.19
Rank_Time	Best									Worst

Table 3. ML Predictive Performance Assessment of Datasets Obtained between 16 April and 16 May 2022.

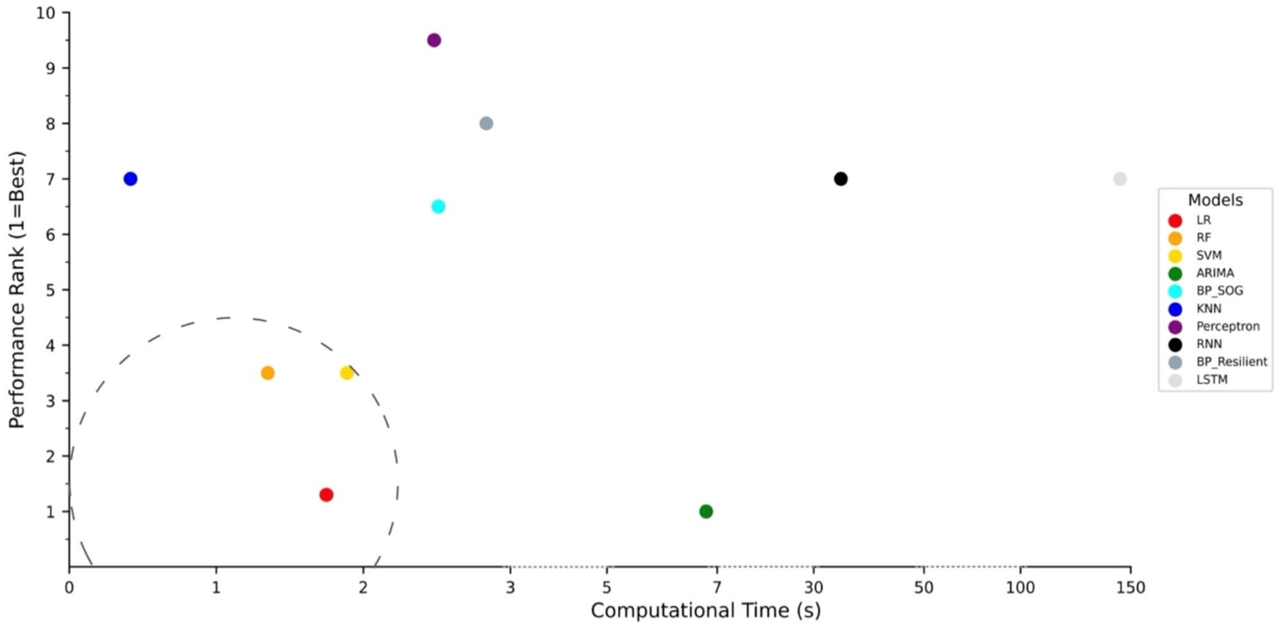

Figure 2. Performance Mapping for Datasets Obtained between 16 April and 16 May 2022.

computational time among the ten algorithms for short-term forecasting. Overall, LR, RF, and SVM are more efficient ML algorithms with better performance for short-term forecasting than the others.

Validation Testing

Two more tests are followed to validate the above outcomes. The tests use data from the same sensors (T050401) for two periods.

First validation testing

The first testing uses data obtained between December 4 and 5, 2021. A total of 11,504 valuable datasets are fed for testing after cleaning the data. The results are shown in Table 4. All ten algorithms are then ranked on the overall average based on the outcomes in Table 4 (see Table 5). The overall average rank shows that RF is the

Model	Datasets	MAE	MSE	RMSE	SAE
ARIMA	Training	0.006468	0.000118	0.010843	44.623
	Testing	0.005011	0.000087	0.009341	23.050
BP_Resilient	Training	0.061892	0.005864	0.076576	2,945,800.000
	Testing	0.056660	0.005659	0.075229	1,198,900.000
BP_SOG	Training	0.062194	0.005902	0.076825	2,960,200.000
	Testing	0.056770	0.005685	0.075398	1,201,300.000
KNN	Training	0.009222	0.000180	0.013402	63.620
	Testing	0.009780	0.000166	0.012870	44.990
LR	Training	0.006468	0.000118	0.010843	44.623
	Testing	0.005011	0.000087	0.009341	23.050
LSTM	Training	0.061419	0.020345	0.142640	915.280
	Testing	0.055999	0.016717	0.129290	545.580
Perceptron	Training	0.863870	0.749280	0.865610	6899.000
	Testing	0.878530	0.774710	0.880180	4600.000
RF	Training	0.006355	0.000115	0.010726	43.844
	Testing	0.004849	0.000087	0.009320	22.304
RNN	Training	0.062088	0.005880	0.076681	938.840
	Testing	0.056621	0.005646	0.075140	559.080
SVM	Training	0.007712	0.000122	0.011049	53.205
	Testing	0.006651	0.000094	0.009719	30.594

Table 4. ML Models based on Prediction Error Assessment Using Datasets on 4 and 5 Dec 2021.

Model	MAE	Rank	MSE	Rank	RMSE	Rank	SAE	Rank	Average of Rank	Rank_Prediction Errors
RF	0.004849	1	0.000087	1	0.009320	1	22.304	1	1	1
ARIMA	0.005011	2	0.000087	1	0.009341	2	23.05	2	1.75	2
LR	0.005011	2	0.000087	1	0.009341	2	23.05	2	1.75	2
SVM	0.006651	4	0.000094	4	0.009719	4	30.594	4	4	4
KNN	0.009780	5	0.000166	5	0.012870	5	44.99	5	5	5
RNN	0.056621	7	0.005646	6	0.075140	6	559.08	7	6.5	6
LSTM	0.055999	6	0.016717	9	0.129290	9	545.58	6	7.5	7
BP_Resilient	0.056660	8	0.005659	7	0.075229	7	1,198,900	9	7.75	8
BP_SOG	0.056770	9	0.005685	8	0.075398	8	1,201,300	10	8.75	9
Perceptron	0.878530	10	0.774710	10	0.880180	10	4600	8	9.5	10

Table 5. Rank of ML Models based on Prediction Error Assessment Using Datasets Obtained between 4 and 5 December 2021.

top-ranked algorithm with prediction error assessment. ARIMA, LR, and SVM are followed. KNN, RNN, LSTM, BP_Resilient, and BP_SOG are ranked from 5 to 9, respectively. Perceptron is the worst algorithm.

A predictive performance assessment is then performed to test the computational time. Table 6 shows that KNN is the best algorithm with the shortest computational time (0.50026 s). Other algorithms are followed, including Perceptron (0.70809 s), BP_SOG (0.83664 s), SVM (1.3899), RF (1.425), LR (1.9711), RNN (3.0593 s), ARIMA (3.9244 s), and BP_Resilient (5.5003). LSTM is the worst algorithm with the longest computational time (42.698 s).

Figure 3 uses a scatter plot to map the relations for tested algorithms between prediction error assessment and predictive performance assessment for datasets obtained between December 4 and 5, 2021. It uses the vertical axis to represent the performance rank (measuring prediction error assessment) (see Table 5) and the horizontal axis to represent the computational time (measuring predictive performance assessment) (see Table 6). Figure 3 shows that ARIMA, LR, RF, and SVM have better outcomes of prediction error assessment in all tests. Perceptron is the worst algorithm for prediction error assessment. KNN has the best predictive performance assessment and the shortest computational time. LSTM has the worst predictive performance assessments with the longest

Algorithm	KNN	Perceptron	BP_SOG	SVM	RF	LR	RNN	ARIMA	BP_Resilient	LSTM
Time (s)	0.50026	0.70809	0.83664	1.3899	1.425	1.9711	3.0593	3.9244	5.5003	42.698
Rank_Time	Best									Worst

Table 6. ML Predictive Performance Assessment of Datasets Obtained between 4 and 5 December 2021.

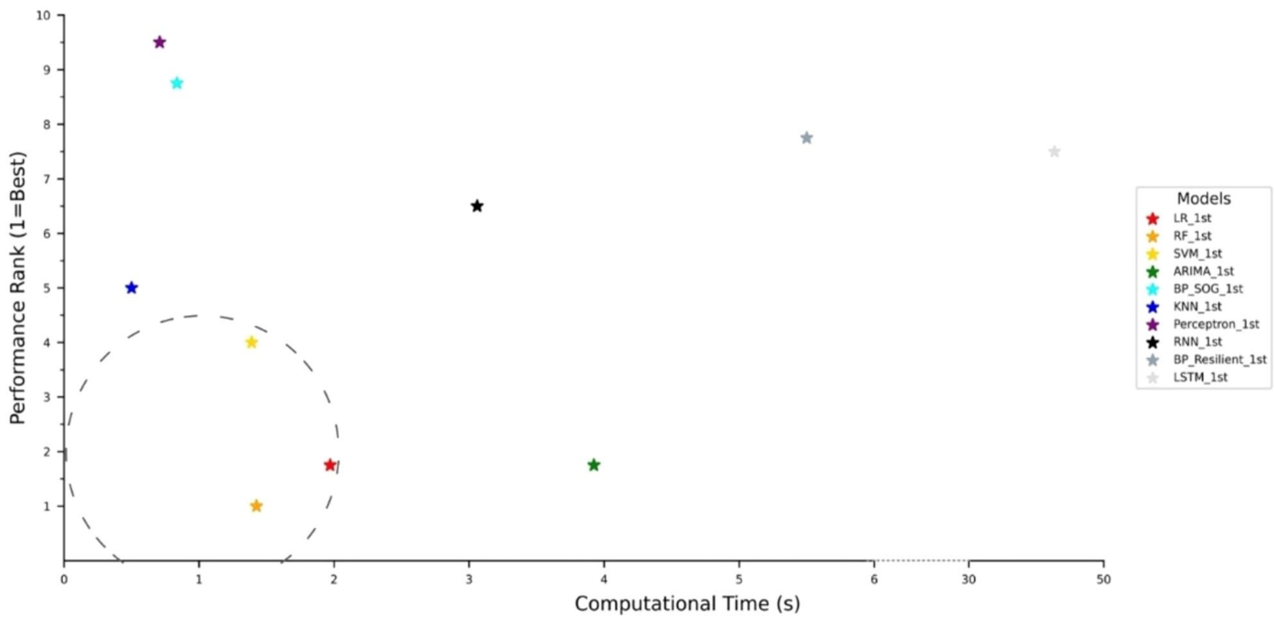

Figure 3. Performance Mapping for Datasets Obtained between 4 and 5 December 2021.

computational time among the ten algorithms for short-term forecasting. Overall, LR, RF, and SVM are more efficient ML algorithms with better performance for short-term forecasting than the others.

Second validation testing

The second test uses data from the same sensor (T050401) between June 16 and 17, 2022. After cleaning the data, 11,504 valuable datasets are fed for testing.

Table 7 shows each algorithm's error assessment on criteria with MAE, MSE, RMSE, and SAE for the training and testing datasets. All ten algorithms are then ranked on an overall average based on the outcomes in Table 7 (see Table 8). The overall average rank shows that ARIMA and LR are the top-ranked algorithms based on prediction error assessment. RF and SVM are followed. KNN, RNN, BP_SOG, LSTM, and BP_Resilient are ranked 5 to 9, respectively. Perceptron is the worst algorithm .

Table 9 shows that KNN is the best algorithm with the shortest computational time (0.48229 s). Other algorithms are followed, including Perceptron (0.95972 s), BP_SOG (1.1219 s), RNN (1.2993 s), SVM (1.4908 s), RF (1.5924 s), LR (2.4908 s), ARIMA (3.4701 s), and BP_Resilient (6.4682 s). LSTM is the worst algorithm with the longest computational time (43.779 s).

Figure 4 uses a scatter plot to map the relations for tested algorithms between prediction error assessment and predictive performance assessment for datasets obtained between June 16 and 17, 2022. ARIMA, LR, RF, and SVM have better outcomes of prediction error assessment in all tests. Perceptron is the worst algorithm for prediction error assessment. KNN has the best predictive performance assessment and the shortest computational time. LSTM has the worst predictive performance assessments with the longest computational time among the

Model	Datasets	MAE	MSE	RMSE	SAE	TIME
ARIMA	Training	0.003033	0.000040	0.006341	20.933	3.4701
	Testing	0.002191	0.000024	0.004858	10.080	
BP_Resilient	Training	0.072598	0.008426	0.091792	3,458,400.000	6.4682
	Testing	0.048282	0.003549	0.059571	1,022,100.000	
BP_SOG	Training	0.072567	0.008424	0.091782	3,456,900.000	1.1219
	Testing	0.047507	0.003444	0.058682	1,005,700.000	
KNN	Training	0.007731	0.000111	0.010528	53.360	0.48229
	Testing	0.007748	0.000104	0.010219	35.650	
LR	Training	0.003033	0.000040	0.006341	20.933	2.4908
	Testing	0.002191	0.000024	0.004858	10.080	
LSTM	Training	0.070842	0.052564	0.229270	1524.100	43.779
	Testing	0.045438	0.031038	0.176180	791.710	
Perceptron	Training	0.777920	0.609390	0.780630	6902.000	0.95972
	Testing	0.829480	0.689740	0.830510	4601.000	
RF	Training	0.006211	0.000039	0.003172	21.896	1.5924
	Testing	0.002669	0.000025	0.005012	12.280	
RNN	Training	0.072598	0.008430	0.091817	1532.500	1.2993
	Testing	0.047337	0.003429	0.058554	784.090	
SVM	Training	0.004607	0.000046	0.006777	31.800	1.4908
	Testing	0.004554	0.000034	0.005794	20.954	

Table 7. ML Models based on Prediction Error Assessment Using Datasets Obtained on 16 -17 Jun 2022.

Model	MAE	Rank	MSE	Rank	RMSE	Rank	SAE	Rank	Average of Rank	Rank_Prediction Errors
ARIMA	0.002191	1	0.000024	1	0.004858	1	10.08	1	1	1
LR	0.002191	1	0.000024	1	0.004858	1	10.08	1	1	1
RF	0.002669	3	0.000025	3	0.005012	3	12.28	3	3	3
SVM	0.004554	4	0.000034	4	0.005794	4	20.954	4	4	4
KNN	0.007748	5	0.000104	5	0.010219	5	35.65	5	5	5
RNN	0.047337	7	0.003429	6	0.058554	6	784.09	6	6.25	6
BP_SOG	0.047507	8	0.003444	7	0.058682	7	1,005,700	9	7.75	7
LSTM	0.045438	6	0.031038	9	0.176180	9	791.71	7	7.75	7
BP_Resilient	0.048282	9	0.003549	8	0.059571	8	1,022,100	10	8.75	9
Perceptron	0.829480	10	0.689740	10	0.830510	10	4601	8	9.5	10

Table 8. Rank of ML Models based on Prediction Error Assessment Using Datasets Obtained on 16 and 17 June 2022.

Algorithm	KNN	Perceptron	BP_SOG	RNN	SVM	RF	LR	ARIMA	BP_Resilient	LSTM
Time (s)	0.48229	0.95972	1.1219	1.2993	1.4908	1.5924	2.4908	3.4701	6.4682	43.779
Rank_Time	Best									Worst

Table 9. ML Predictive Performance Assessment of Datasets Obtained on 16 and 17 June 2022.

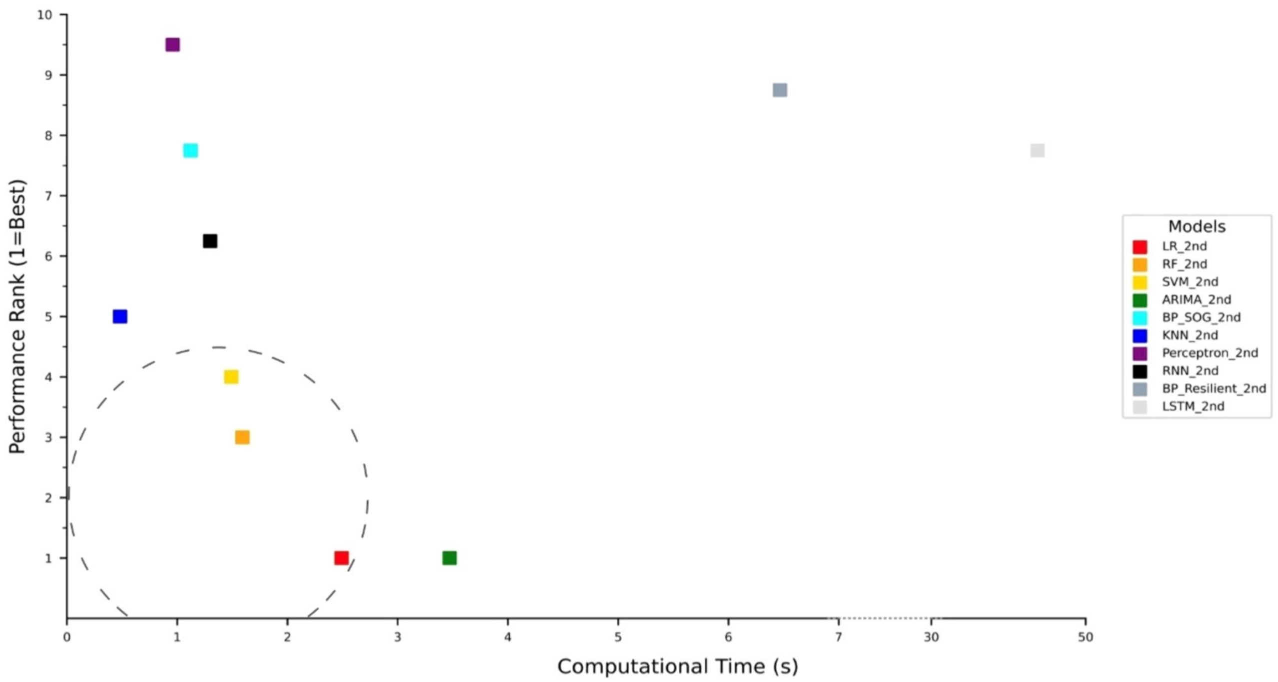

Figure 4. Performance Mapping for Datasets Obtained on 16 and 17 June 2022.

ten algorithms for short-term forecasting. Overall, LR, RF, and SVM are more efficient ML algorithms with better performance for short-term forecasting than the others.

Comparative analysis

A comparative analysis is then conducted to better understand the above outcomes. We propose a new assessment visualization tool for performing comparative analysis to measure ML algorithms' prediction performance: a 2D space-based quadrant diagram (see Fig. 5). This newly developed assessment visualization tool combines all the above tests' outcomes (see Figs. 2, 3, and 4) to visually map prediction error assessment and predictive performance assessment for ten tested algorithms. It uses the vertical axis to represent the performance rank (measuring prediction error assessment) and the horizontal axis to represent the computational time (measuring predictive performance assessment).

This newly developed assessment visualization tool indicates that ARIMA, LR, RF, and SVM have better outcomes of prediction error assessment in all tests. Perceptron is the worst algorithm for prediction error assessment. KNN has the best predictive performance assessment and the shortest computational time. LSTM has the worst predictive performance assessments with the longest computational time among the ten algorithms for short-term forecasting. Overall, LR, RF, and SVM are more efficient ML algorithms with better performance for short-term forecasting than the others.

Through using this assessment visualization tool, ten tested algorithms can be mapped onto four distinct quadrants covering four categories, including optimal, efficient, suboptimal, and inefficient algorithms, as follows:

- Quadrant one (QI) is named optimal and is located at the bottom left fourth of the quadrant diagram. An optimal algorithm is used in an application that measures both prediction error assessment and predictive performance assessments at a satisfied level. LR, RF, and SVM are optimal algorithms.
- Quadrant two (QII) is called efficient and is located at the bottom right-left fourth. An efficient algorithm is deemed an algorithm used in an application that measures prediction error assessment at a satisfied level and predictive performance assessment below a satisfied level. ARIMA is an efficient algorithm.
- Quadrant three (QIII) is titled "suboptimal" and is located at the top left fourth. A suboptimal algorithm is accepted as an algorithm used in an application with measures of prediction error assessment below a satisfied level and predictive performance assessment at a satisfied level. The suboptimal algorithms include BP-SOG, KNN, and Perceptron.

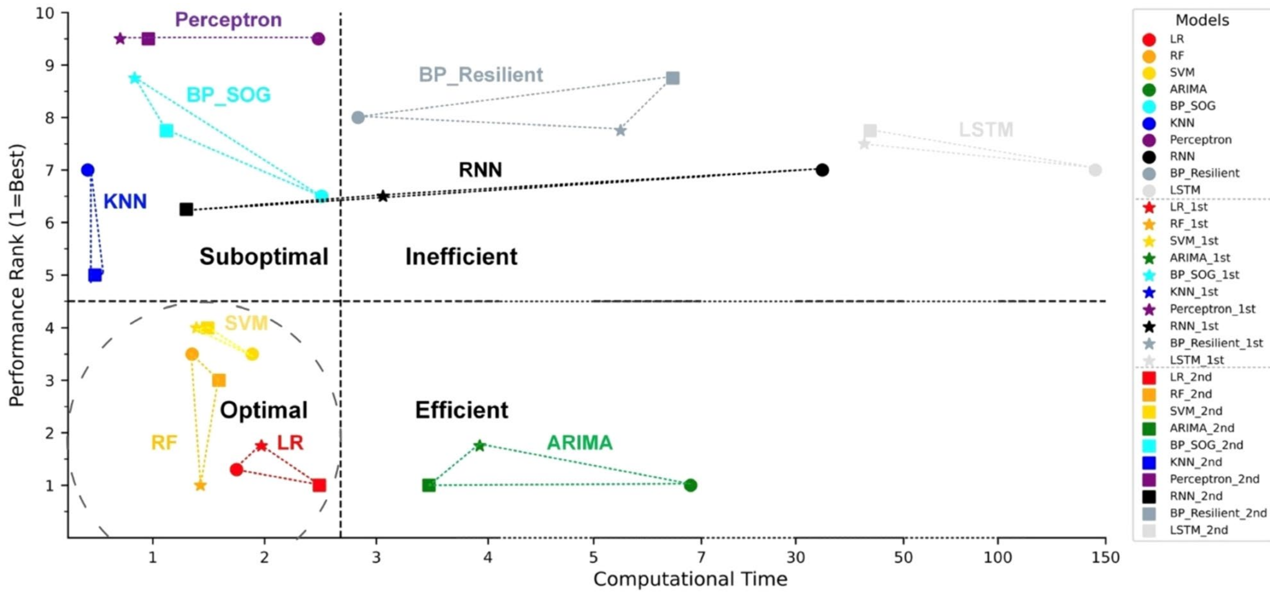

Figure 5. An Assessment Visualization Tool for Measuring ML Algorithms' Performance.

- Quadrant four (QIV) is named inefficient and is located at the top right fourth. An inefficient algorithm is used in an application that measures both prediction error assessment and predictive performance assessments below a satisfied level. Among the three inefficient algorithms (BP_Resilient, RNN, and LSTM), RNN has a worse prediction error assessment. The computational time is based on the number of datasets. With increasing data sampling frequency, RNN requires more computational time because more computations with more data points are needed[72]. LSTM has the worst predictive performance assessments and the longest computational time among the ten algorithms for short-term forecasting.

Discussions

This section focuses on each category (optimal, efficient, suboptimal, and inefficient algorithms) and discusses the research findings compared with those of previous studies.

Optimal algorithms

LR

LR is the optimal algorithm. This research finds that LR is one of the most efficient ML algorithms with better performance for short-term forecasting than other algorithms. However, it is against previous studies that LR performs poorly[73] and yields unreliable predictions due to its low flexibility[74]. This research thus raises a different view on the performance of LR among various studies. Further research is required to understand the prediction performance of LR.

RF

RF is indicated as another optimal algorithm. RF frequently shows a statistically lower error performance[75] and achieves the highest prediction accuracy[76]. This research finds that RF has a better assessment than KNN in MAE (0.001944, 0.025468), MSE (0.000376, 0.163050), and RMSE (0.019390, 0.403800), which supports Pakzad, Roshan & Ghalehnovi[68].

This study finds diffent research outcomes between RF and LR based on prediction error assessment compared to other studies. This research indicates that LR performs better in prediction error assessment than RF in MAE (0.001510, 0.001944), MSE (0.000009, 0.000376), and RMSE (0.003048, 0.019390), which supports another research by Ustebay et al.[77] that LR performs better than RF. However, it is against the earlier studies that RF has higher discrimination performance and calibrated probabilities than LR, such as in MAE, MSE, and RMSE[68,69,78,79]. There is a need to investigate more prediction performance measures between RF and LR.

SVM

This research indicates that SVM is another efficient algorithm. This study finds that SVM is acceptable on the computational time compared with previous studies. This study finds that the SVM performs well and has a shorter computational time. However, Sharma, Kim & Gupta[80] highlight that SVM has the shortest training time and prediction speed. Another study states that although SVM may take numerical inputs and work well on small datasets, it will require too much training time as the dataset size increases[81]. It may be argued that no single algorithm can be used to fit all applications. Thus, further investigation of SVM's computational time is needed in various applications.

This study finds different research outcomes between SVM and RF based on prediction error assessment among various studies. This research finds that RF has a better prediction of achieving MAE (0.001944) and SAE (22.312) than SVM (0.0020690, 23.750), which supports previous studies by Šušteršič et al.[69] and Kasbekar et al.[75]. However, the first validation testing, the second validation testing, and several previous studies indicate that RF has a better-predicting outcome than SVM in terms of all criteria[78,82, 83]. The results also indicate that SVM has significantly better prediction, achieving MSE (0.000011) and RMSE (0.003259) than RF (0.000376 and 0.019390). Further research is required to investigate additional measures of prediction error assessment between SVM and RF.

Efficient algorithms

ARIMA is an efficient algorithm. This result finds a different view of ARIMA performance, contrary to a previous study that ARIMA may produce worse results with the extensive data in the algorithms generated[38]. Further research is required to verify the prediction error assessment of ARIMA using extensive data.

Suboptimal algorithms

Suboptimal algorithms include BP_SOG, Perceptron, and KNN. BP_SOG and Perceptron should be discussed further in the literature. There is a need to investigate the limitations of BP_SOG and Perceptron, which may lead to less use in industrial applications.

KNN has the best predictive performance assessment with the shortest computational time in all testing and validation tests among ten short-term forecasting algorithms. However, KNN has poor prediction error assessment in all testing in this research. The literature states that a KNN performs poorly if the training set is large[28,73]. However, a KNN has a disadvantage because of the enormous computing requirement for classifying an object, as the distance for all neighbors in the training dataset must be calculated[81]. It is valuable to conduct further research to test how large datasets will impact the performance of KNN.

This study finds a different view of KNN and ARIMA compared with previous studies. It finds that KNN is worse than ARIMA in all tests. However, an earlier study states that ARIMA performs marginally better than KNN for the complete set of all-time series[27]. Thus, further research is needed to conduct more tests on the prediction error assessment between KNN and ARIMA.

A previous study had a different view on prediction error assessment between KNN and LR. This research indicates that LR has a better prediction error assessment than KNN. However, another study argues that KNN (4.648) is better than LR (5.317) in MAE[68]. Further research is needed to investigate why there are different results between KNN and LR in MAE.

This study also finds a different view of the performance between KNN and SVM compared with previous studies. The research outcome indicates that SVM has a better prediction error assessment than KNN. However, this contradicts another previous study that KNN outperforms SVM on most datasets[84]. Recent studies have assumed that KNN may be outperformed by more exotic techniques such as SVM[28]. Thus, further research is required on the prediction error assessment between KNN and SVM.

Inefficient algorithm

BP_Resilient, RNN, and LSTM are inefficient compared with the other algorithms.

BP_Resilient

The literature does not discuss BP-Resilient much. It is necessary to investigate its limitations, which have led to its low use in industrial applications.

RNN

RNN is another inefficient algorithm with a worse prediction error assessment. This research has a different view of prediction error assessments between RNN and ARIMA compared with previous studies. This research indicates that RNN has significantly worse performance outcomes in prediction error assessment than ARIMA in all tests. Previous research has demonstrated the superiority of RNN over the traditionally used ARIMA[85]. Therefore, conducting further research to verify the prediction error assessments between RNN and ARIMA is valuable.

LSTM

As an inefficient algorithm, this study highlights that LSTM has the worst predictive performance assessments with the longest computational time among the ten algorithms for short-term forecasting for all tests. There are different views on LSTM. This research indicates that LSTM does not perform well in all tests. Kasbekar et al.[75] state that the statistical comparison results for absolute errors (AE) confirm that LSTM does not perform well on lower errors. Other studies state that LSTM may produce better predictions of modeling time series data[35,36, 38, 71, 86]. Thus, it will be valuable to investigate the prediction error assessment of LSTM in future research, including AE.

This study finds a different view of prediction error assessment between LSTM and ARIMA compared with previous studies. This research indicates that LSTM is worse than ARIMA in all tests. However, an earlier study has claimed that LSTM outperforms ARIMA with a large quantity of data in MAE and RMSE criteria[38]. Thus, further research is required to investigate the prediction error assessment in MAE and RMSE criteria between LSTM and ARIMA.

Another different view of prediction error assessment has been discussed between LSTM and SVM. This study indicates that SVM performs better with overall prediction error assessment than LSTM in all tests. It is against another previous study that LSTM outperforms SVM[71]. Thus, further research is needed on the prediction error assessment between LSTM and SVM.

Conclusions

Conclusion

This study aims to explore more efficient ML algorithms with better performance for short-term forecasting. This research uses a quantitative and qualitative mixed method combining two rounds of literature reviews, a case study, and a comparative analysis. The first round of the literature review focuses on top-tier publications on ML algorithms used in China's industrial applications. Twenty-nine algorithms have been found in 347 industrial applications (see Appendix 1). Among them, ten short-term forecasting methods are identified and tested to determine their performance for short-term forecasting, including ARIMA, BP_Resilient, BP_SOG, KNN, LR, LSTM, Perceptron, RF, RNN, and SVM. This research conducts the second round of literature review on Q1 publications related to the prediction error assessment of ML algorithms between 2020 and 2023. Forty-five performance criteria were identified.

Four metrics (MAE, MSE, RMSE, and SAE) have been widely discussed and used to test the prediction error assessment of the above ten ML algorithms. Computational time is used to measure predictive performance assessment. The case study indicates that no single or standard error evaluation criteria can be adopted as the expected performance method for evaluating the error characteristics of ML algorithms (see Appendix 3). This research also finds that MSE and RMSE have the same prediction error assessment (see Table 2), and further search does not need to test MSE and RMSE together.

A comparative analysis is then conducted to better understand the above outcomes. We propose a new assessment visualization tool for performing comparative analysis to measure ML algorithms' prediction performance: a 2D space-based quadrant diagram (see Fig. 5). This newly developed assessment visualization tool combines all the above tests' outcomes (see Figs. 2, 3 and 4) to visually map prediction error assessment and predictive performance assessment for ten tested algorithms. It uses the vertical axis to represent the performance rank (measuring prediction error assessment) and the horizontal axis to represent the computational time (measuring predictive performance assessment). This newly developed assessment visualization tool indicates that ARIMA, LR, RF, and SVM have better outcomes of prediction error assessment in all tests. Perceptron is the worst algorithm for prediction error assessment. KNN has the best predictive performance assessment and the shortest computational time. LSTM has the worst predictive performance assessments with the longest computational time among the ten algorithms for short-term forecasting. Overall, LR, RF, and SVM are more efficient ML algorithms with better performance for short-term forecasting than the others.

All tested algorithms can be visually mapped onto four distinct quadrants covering four categories, including optimal (LR, RF, and SVM), efficient (ARIMA), suboptimal (BP-SOG, KNN, and Perceptron), and inefficient algorithms (RNN, BP_Resilient, and LSTM) (see Fig. 5). As a results, LR, RF, and SVM are more efficient ML algorithms with overall prediction performance for short-term forecasting. LSTM is the worst algorithm for short-term forecasting. Overall, no single algorithm can fit all applications. This study raises 20 valuable questions for further research.

Findings from different views and further research

The case study finds results that differ from previous studies regarding the ML prediction efficiency of ARIMA, BP_SOG, BP_Resilient, KNN, LR, LSTM, Perceptron, and SVM. The following research questions (RQs) need to be investigated further:

- RQ1: prediction performance of LR.
- RQ2: computational time of SVM in different applications.
- RQ3: prediction error assessment of ARIMA using extensive data.
- RQ4: limitations of BP_SOG, BP_Resilient, and Perceptron for industrial applications.
- RQ5: how large datasets will impact the performance of the KNN.
- RQ6: prediction error assessment of LSTM in further research, including AE.

This study finds different views on the prediction performance of a few paired algorithms compared with previous studies, including RF and LR, SVM and RF, KNN and SVM, RNN and ARIMA, and LSTM and SVM. There is a need to investigate the following RQs for additional measures of prediction error assessment:

- RQ7: between RF and LR.
- RQ8: between SVM and RF.
- RQ9: between KNN and ARIMA.
- RQ10: between KNN and SVM.
- RQ11: between RNN and ARIMA.
- RQ12: between LSTM and SVM.

This study also suggests that ARIMA, KNN, LR, and LSTM should be investigated with additional prediction error assessments in further research as follows:

- RQ13: MAE between KNN and LR.
- RQ14: MAE and RMSE between LSTM and ARIMA.

Limitations and further research

The main limitation of this research is that it aims to find the most suitable ML Algorithms for prediction systems rather than discuss the features of ML Algorithms. Further research is required to investigate the impact of these algorithms' advantages and limitations on predicting warning systems (RQ15). Another limitation is that this research uses data from a gas warning system in a Case Study mine to test ten algorithms to predict gas concentration. Further investigation must test the research outcomes in different industry cases (RQ16). The third limitation is that this research only focuses on limited prediction error assessments (MAE, MSE, RMSE, and SAE). It is valuable for testing other prediction error criteria (see Appendix 3) (RQ17).

Other further research

The following RQs also need to be addressed further:

- RQ18: conducting research for very short-term, medium-term, and long-term forecasting.
- RQ19: Study of other performance assessments not included in this research, such as accuracy, precision, recall, F1 score, sensitivity (SN), specificity (SP), balanced accuracy (BA), geometric mean (GM), Cohen's kappa (CK), and Matthew's correlation coefficient (MCC).
- RQ20: The first round of the literature review mainly focuses on the popularity of ML algorithms within China's industrial applications, which may only partially represent the most appropriate choice for the specific application of gas warning systems. There is a need to conduct literature on global studies to gain a better understanding of the appropriate choice of ML algorithms in different industrial applications.

Implications

The research outcomes of the Ten ML algorithms for short-term forecasting should add value to higher education institutions in developing up-to-date teaching contexts for ML courses. The research outcomes also implicate that the coal mining industry deploying an efficient ML algorithm with better performance for short-term forecasting may effectively reduce the risk of accidents such as gas explosions, safeguard workers, and enhance the ability to prevent and mitigate disasters so that economic losses might be reduced[87].

Contributions

The main contributions of this study can be highlighted as follows:

- Proposing a new assessment visualization tool for measuring ML algorithms' prediction performance.
- Clarifying that no single prediction error assessment can be used as the expected performance measure for evaluating the error characteristics of ML algorithms, and
- Exploring significantly different research outcomes that violate the results of previous studies on the performance of ten short-term ML algorithms.

Data availability

The data supporting the study's findings are available in the public domain Figshare with license CC BY4.0 from CC BY4.0 from https://doi.org/10.6084/m9.figshare.24083076.v2.

Received: 12 September 2023; Accepted: 9 July 2024
Published online: 20 September 2024

References

1. Hutzler. China's Economic Recovery will be Powered by Coal', Viewed 08 Jan 2021, https://www.powermag.com/chinas-economic-recovery-will-be-powered-by-coal/ (2020).
2. IEA. Coal 2020 analysis and forecast to 2025. Viewed 7 Jan 2021, https://www.iea.org/reports/coal-2020/supply (2020).
3. Zou, Q. *et al.* Rationality evaluation of production deployment of outburst-prone coal mines: A case study of Nantong coal mine in Chongqing, China. *Saf. Sci.* **122**, 1–16. https://doi.org/10.1016/j.ssci.2019.104515 (2020).
4. Priyadarsini, V. *et al.* Labview based real time monitoring system for coal mine worker. *I-Manag. S J. Digit. Signal Process.* **6**(4), 1–6 (2018).
5. Zhang, Y., Guo, H., Lu, Z., Zhan, L. & Hung, P. C. K. Distributed gas concentration prediction with intelligent edge devices in coal mine. *Eng. Appl. Artif. Intell.* **92**, 1–11 (2020).
6. Tutak, M. & Brodny, J. Predicting methane concentration in Longwall regions using artificial neural networks. *Int. J. Environ. Res. Public Health* **16**(8), 1–21 (2019).
7. China Coal Safety. The prevention regulations of coal and gas outbursts. Viewed 11 Nov 2020. https://www.chinacoal-safety.gov.cn/zfxxgk/fdzdgknr/tzgg/201908/t20190821_349184.shtml (2019).
8. Xia, X., Chen, Z. & Wei, W. Research on monitoring and prewarning system of accident in the coal mine based on big data. *Sci. Program.* **2018**, 1–11 (2018).
9. Zhao, X. *et al.* Applications of online integrated system for coal and gas outburst prediction: A case study of Xinjing mine in Shanxi, China. *Energy Sci. Eng.* **8**(6), 1980–1996 (2020).
10. Wu, R. M. X. *et al.* A comparative analysis of the principal component analysis and entropy weight methods to establish the indexing measurement. *PLoS One* **17**(1), 1–26 (2022).
11. Jo, B. W., Khan, R. M. A. & Javaid, O. Arduino-based intelligent gases monitoring and information sharing Internet-of-Things system for underground coal mines. *AIS* **11**(2), 183–194 (2019).
12. Arango, M. I., Aristizábal, E. & Gómez, F. Morphometrical analysis of torrential flows-prone catchments in tropical and mountainous terrain of the Colombian Andes by machine learning techniques. *Nat. Hazards* **105**, 983–1012 (2021).
13. Féret, J.-B. *et al.* Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: Potential and limitations of physical modeling and machine learning. *Remote Sens. Environ.* **231**, 1–14 (2019).

14. Wu, R. M. X. *et al.* Using multi-focus group method in systems analysis and design: A case study. *PLoS One* **8**(3), 1–16 (2023).
15. Afrash, M. R., Mirbagheri, E., Mashoufi, M. & Kazemi-Arpanahi, H. Optimizing prognostic factors of five-year survival in gastric cancer patients using feature selection techniques with machine learning algorithms: A comparative study. *BMC Med. Inform. Decis. Mak.* **23**(1), 54 (2023).
16. Soman, S. S. *et al. A Review of Wind Power and Wind Speed Forecasting Methods with Different Time Horizons* 1–8 (IEEE, 2010).
17. Li, D.T., Tong, T.W. & Xiao, Y.G. Is China Emerging as the Global Leader in AI? *Harvard Business Review* (2021).
18. El Bilali, A., Taleb, A., Bahlaoui, M. A. & Brouziyne, Y. An integrated approach based on Gaussian noises-based data augmentation method and AdaBoost model to predict faecal coliforms in rivers with small dataset. *J. Hydrol.* **599**, 1–11 (2021).
19. Readshaw, J. & Giani, S. Using company-specific headlines and convolutional neural networks to predict stock fluctuations. *Neural Comput. Appl.* **33**, 17353–17367 (2021).
20. Srikanth, K., Ul Huq, S. Z. & Kumar, A. P. Big data based analytic model to predict and classify breast cancer using improved fractional rough fuzzy K-means clustering and labeled ensemble classifier algorithm. *Concurr. Computat.-Pract. Exp.* **34**(10), 1–21 (2022).
21. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**(9), 533–536 (1986).
22. Huang, J.-C., Ko, K.-M., Shu, M.-H. & Hsu, B.-M. Application and comparison of several machine learning algorithms and their integration models in regression problems. *Neural Comput. Appl.* **32**(10), 5461–5469 (2020).
23. Sui Kim, I. T. *et al.* Fenugreek seeds and okra for the treatment of palm oil mill effluent (POME)–Characterization studies and modeling with backpropagation feedforward neural network (BFNN). *J. Water Process Eng.* **37**, 1–16 (2020).
24. Erkaymaz, O. Resilient back-propagation approach in small-world feed-forward neural network topology based on Newman-Watts algorithm. *Neural Comput. Appl.* **32**(20), 16279–16289 (2020).
25. Uddin, S., Haque, I., Lu, H., Moni, M. A. & Gide, E. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Sci. Rep.* **12**(1), 1–11 (2022).
26. Dong, Y., Ma, X. & Fu, T. Electrical load forecasting: A deep learning approach based on K-nearest neighbors. *Appl. Soft Comput.* **99**, 1–15 (2021).
27. Kück, M. & Freitag, M. Forecasting of customer demands for production planning by local k-nearest neighbor models. *Int. J. Prod. Econ.* **231**, 1–22 (2021).
28. Cunningham, P. & Delany, S. J. k-nearest neighbour classifiers—A tutorial. *ACM Comput. Surv.* **54**(6), 1–25 (2021).
29. Dritsas, E. & Trigka, M. Data-driven machine-learning methods for diabetes risk prediction. *Sensors* **22**(14), 1–18 (2022).
30. Brownlee, J. A Gentle Introduction to Exponential Smoothing for Time Series Forecasting in Python. Viewed 12 Apr 2020, https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python/ (2018).
31. Aasim, S. N. S. & Mohapatra, A. Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting. *Renew. Energy* **136**, 758–768 (2019).
32. Schaffer, A. L., Dobbins, T. A. & Pearson, S.-A. Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: A guide for evaluating large-scale health interventions. *BMC Med. Res. Methodol.* **21**(1), 1–12 (2021).
33. Pu, Z. *et al.* Road surface friction prediction using long short-term memory neural network based on historical data. *J. Intell. Transp. Syst.* **26**(1), 34–45 (2021).
34. Swathi, T., Kasiviswanath, N. & Rao, A. A. An optimal deep learning-based LSTM for stock price prediction using twitter sentiment analysis. *Appl. Intell.* **52**(12), 13675–13688. https://doi.org/10.1007/s10489-022-03175-2 (2022).
35. Butt, U. A. *et al.* Cloud-based email phishing attack using machine and deep learning algorithm. *Complex Intell. Syst.* **9**(3), 3043–3070 (2023).
36. Mahmoud, N., Abdel-Aty, M., Cai, Q. & Yuan, J. Estimating cycle-level real-time traffic movements at signalized intersections. *J. Intell. Transp. Syst.* **26**(4), 400–419 (2022).
37. Van Houdt, G., Mosquera, C. & Nápoles, G. A review on the long short-term memory model. *Artif. Intell. Rev.* **53**(8), 5929–5955 (2020).
38. Elsaraiti, M. & Merabet, A. A comparative analysis of the arima and lstm predictive models and their effectiveness for predicting wind speed. *Energies* **14**(20), 1–16 (2021).
39. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D* **404**, 1–28 (2020).
40. Hodson, T. O., Over, T. M. & Foks, S. S. Mean squared error, deconstructed. *J. Adv. Model. Earth Syst.* **13**(12), 1–10 (2021).
41. Ameer, K. *et al.* A hybrid RSM-ANN-GA approach on optimization of ultrasound-assisted extraction conditions for bioactive component-rich Stevia rebaudiana (Bertoni) leaves extract. *Foods* **11**(6), 1–24 (2022).
42. Sukawutthiya, P., Sathirapatya, T. & Vongpaisarnsin, K. A minimal number CpGs of ELOVL2 gene for a chronological age estimation using pyrosequencing. *Forensic Sci. Int.* **318**, 1–6 (2021).
43. Verhaeghe, J. *et al.* Development and evaluation of uncertainty quantifying machine learning models to predict piperacillin plasma concentrations in critically ill patients. *BMC Med. Inform. Decis. Mak.* **22**(1), 1–17 (2022).
44. Yaseen, Z. M. An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals: Review, challenges and solutions. *Chemosphere* **277**, 1–22 (2021).
45. Alhakamy, A., Alhowaity, A., Alatawi, A. A. & Alsaadi, H. Are used cars more sustainable? Price prediction based on linear regression. *Sustain. Sci. Pract. Policy* **15**(2), 1–17 (2023).
46. Zhang, D. Coefficients of determination for mixed-effects models. *J. Agric. Biol. Environ. Stat.* **27**(4), 674–689 (2022).
47. Cabeza-Ramírez, L. J., Rey-Carmona, F. J., Del Carmen Cano-Vicente, M. & Solano-Sánchez, M. Á. Analysis of the coexistence of gaming and viewing activities in Twitch users and their relationship with pathological gaming: A multilayer perceptron approach. *Sci. Rep.* **12**(1), 1–18 (2022).
48. Barber, C., Lamontagne, J. R. & Vogel, R. M. Improved estimators of correlation and R2 for skewed hydrologic data. *Hydrol. Sci. J.* **65**(1), 87–101 (2020).
49. Onyutha, C. From R-squared to coefficient of model accuracy for assessing 'goodness-of-fits'. *Geoscientific Model Development Discussions* 1–25 (2020).
50. Heravi, A. R. & Abed Hodtani, G. A new correntropy-based conjugate gradient backpropagation algorithm for improving training in neural networks. *IEEE Trans. Neural Netw. Learn Syst.* **29**(12), 6252–6263 (2018).
51. Kiraz, A., Canpolat, O., Erkan, E. F. & Özer, Ç. Artificial neural networks modeling for the prediction of Pb(II) adsorption. *Int. J. Environ. Sci. Technol.* **16**(9), 5079–5086 (2019).
52. Joseph, R. V. *et al.* A hybrid deep learning framework with CNN and Bi-directional LSTM for store item demand forecasting. *Comput. Electr. Eng.* **103**, 1–14 (2022).
53. Robeson, S. M. & Willmott, C. J. Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. *PLoS One* **18**(2), 1–8 (2023).
54. Pham, A.-D., Ngo, N.-T. & Nguyen, T.-K. Machine learning for predicting long-term deflections in reinforce concrete flexural structures. *Finite Elem. Anal. Des.* **7**(1), 95–106 (2020).
55. Bhardwaj, R. & Bangia, A. Data driven estimation of novel COVID-19 transmission risks through hybrid soft-computing techniques. *Chaos Solitons Fractals* **140**, 1–16 (2020).

56. Hodson, T. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev.* **15**(14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022 (2022).
57. Pernot, P., Huang, B. & Savin, A. Impact of non-normal error distributions on the benchmarking and ranking of quantum machine learning models. *Mach. Learn. Sci. Technol.* **1**(3), 1–14 (2020).
58. Chicco, D., Warrens, M. J. & Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **7**, 1–24 (2021).
59. Barashid, K., Munshi, A. & Alhindi, A. Wind farm power prediction considering layout and wake effect: Case study of Saudi Arabia. *Energies* **16**(2), 1–22 (2023).
60. Qi, J., Du, J., Siniscalchi, S. M., Ma, X. & Lee, C.-H. On mean absolute error for deep neural network based vector-to-vector regression. *IEEE Signal Process. Lett.* **27**, 1485–1489 (2020).
61. Ali, F. *et al.* Parameter extraction of photovoltaic models using atomic orbital search algorithm on a decent basis for novel accurate RMSE calculation. *Energy Convers. Manag.* **277**, 1–15 (2023).
62. Khamparia, S. & Jaspal, D. K. Xanthium strumarium L. seed hull as a zero cost alternative for Rhodamine B dye removal. *J. Environ. Manag.* **197**, 498–506 (2017).
63. Lee, E., Jang, D. & Kim, J. A two-step methodology for free rider mitigation with an improved settlement algorithm: Regression in CBL estimation and new incentive payment rule in residential demand response. *Energies* **11**(2), 1–17 (2018).
64. Nurwatik, N., Ummah, M. H., Cahyono, A. B., Darminto, M. R. & Hong, J.-H. A comparison study of landslide susceptibility spatial modeling using machine learning. *ISPRS Int. J. Geo-Inf.* **11**(12), 1–21 (2022).
65. Wu, R. M. X. *et al.* A correlational research on developing an innovative integrated gas warning system: a case study in ZhongXing, China. *Geomat. Nat. Hazards Risk* **12**(1), 3175–3204 (2021).
66. Moghadasi, M., Ozgoli, H. A. & Farhani, F. Steam consumption prediction of a gas sweetening process with methyldiethanolamine solvent using machine learning approaches. *Int. J. Energy Res.* **45**(1), 879–893 (2021).
67. Wu, R. M. X. *et al.* An FSV analysis approach to verify the robustness of the triple-correlation analysis theoretical framework. *Sci. Rep.* **13**(1), 1–20 (2023).
68. Pakzad, S. S., Roshan, N. & Ghalehnovi, M. Comparison of various machine learning algorithms used for compressive strength prediction of steel fiber-reinforced concrete. *Sci. Rep.* **13**(1), 1–15 (2023).
69. Šušteršič, T. *et al.* The effect of machine learning algorithms on the prediction of layer-by-layer coating properties. *ACS Omega* **8**(5), 4677–4686 (2023).
70. Tabbussum, R. & Dar, A. Q. Comparative analysis of neural network training algorithms for the flood forecast modelling of an alluvial Himalayan river. *J. Flood Risk Manag.* **13**(4), 1–18 (2020).
71. Essam, Y. *et al.* Predicting streamflow in Peninsular Malaysia using support vector machine and deep learning algorithms. *Sci. Rep.* **12**(1), 1–13 (2022).
72. Pang, Z., Niu, F. & O'Neill, Z. Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. *Renew. Energy* **156**, 279–289 (2020).
73. Patel, S., Wang, M., Guo, J., Smith, G. & Chen, C. A study of R-R interval transition matrix features for machine learning algorithms in AFib detection. *Sensors* **23**(7), 1–27 (2023).
74. Al-Swaidani, A. M., Khwies, W. T., Al-Baly, M. & Lala, T. Development of multiple linear regression, artificial neural networks and fuzzy logic models to predict the efficiency factor and durability indicator of nano natural pozzolana as cement additive. *J. Build. Eng.* **52**, 1–27 (2022).
75. Kasbekar, R. S., Ji, S., Clancy, E. A. & Goel, A. Optimizing the input feature sets and machine learning algorithms for reliable and accurate estimation of continuous, cuffless blood pressure. *Sci. Rep.* **13**(1), 1–13 (2023).
76. Shafiabady, N. *et al.* eXplainable Artificial Intelligence (XAI) for improving organisational regility. *PLoS One* **19**, 1–21 (2024).
77. Ustebay, S., Sarmis, A., Kaya, G. K. & Sujan, M. A comparison of machine learning algorithms in predicting COVID-19 prognostics. *Intern. Emerg. Med.* **18**(1), 229–239 (2023).
78. Castonguay, A. C. *et al.* Predicting functional outcome using 24-hour post-treatment characteristics: Application of machine learning algorithms in the STRATIS *Registry*. *Ann. Neurol.* **93**(1), 40–49 (2023).
79. Mulugeta, G., Zewotir, T., Tegegne, A. S., Juhar, L. H. & Muleta, M. B. Classification of imbalanced data using machine learning algorithms to predict the risk of renal graft failures in Ethiopia. *BMC Med. Inform. Decis. Mak.* **23**(1), 1–17 (2023).
80. Sharma, R., Kim, M. & Gupta, A. Motor imagery classification in brain-machine interface with machine learning algorithms: Classical approach to multi-layer perceptron model. *Biomed. Signal Process. Control* **71**, 1–10 (2022).
81. Panesar, A. Machine learning and AI for healthcare. *Apress* https://doi.org/10.1007/978-1-4842-6537-6_4 (2021).
82. Ahmadi, M., Nopour, R. & Nasiri, S. Developing a prediction model for successful aging among the elderly using machine learning algorithms. *Digit. Health* **9**, 1–22 (2023).
83. Hassanzadeh, R., Farhadian, M. & Rafieemehr, H. Hospital mortality prediction in traumatic injuries patients: Comparing different SMOTE-based machine learning algorithms. *BMC Med. Res. Methodol.* **23**(1), 1–15 (2023).
84. Mailagaha Kumbure, M., Luukka, P. & Collan, M. A new fuzzy k-nearest neighbor classifier based on the Bonferroni mean. *Pattern Recognit. Lett.* **140**, 172–178 (2020).
85. Šestanović, T. & Arnerić, J. Can recurrent neural networks predict inflation in euro zone as good as professional forecasters?. *Sci. China Ser. A Math.* **9**(19), 1–14 (2021).
86. Azeem, A. *et al.* Deterioration of electrical load forecasting models in a smart grid environment. *Sensors* **22**(12), 1–28 (2022).
87. Kumari, K. *et al.* UMAP and LSTM based fire status and explosibility prediction for sealed-off area in underground coal mine. *Process Saf. Environ. Prot.* **146**, 837–852 (2021).

Acknowledgements
This research was supported by the Shanxi Coking Coal Project (201809fx03) and the Shanxi Social Science Federation (SSKLZDKT2019053). The authors thank all reviewers for their positive comments and suggestions on the manuscript.

Author contributions
Conceptualization: R.M.X.Wu. Data collection: H.Zhang, R.M.X.Wu. Formal analysis: N. Shafiabady, R.M.X.Wu. Investigation: R.M.X.Wu, N.Shafiabady, H.Zhang Methodology: R.M.X.Wu, N.Shafiabady, H.Y.Lu,E.Gide Project administration: R.M.X.Wu. Resources: R.M.X.Wu. Supervision: R.M.X.Wu. H.Y.Lu Validation: R.M.X.Wu, N.Shafiabady, J.R.Liu, C.F.B.Charbonnier Visualization: R.M.X.Wu, N.Shafiabady, J.R.Liu Writing—original draft: R.M.X.Wu, N.Shafiabady, H.Y.Lu, E.Gide Writing—review and editing: R.M.X.Wu, H.Y.Lu.

Competing interests
The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-67283-4.

Correspondence and requests for materials should be addressed to R.M.X.W.