applied sciences

MDPI

Article

Occupational Risk Prediction for Miners Based on Stacking Health Data Fusion

Xuhui Zhang [1,2,*], Wenyu Yang [1], Wenjuan Yang [1,2], Benxin Huang [1], Zeyao Wang [1] and Sihao Tian [1]

[1] College of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an 710054, China; yangwy2024@163.com (W.Y.); yangwenjuan@xust.edu.cn (W.Y.); m15958727698@163.com (B.H.); 23205016031@stu.xust.edu.cn (Z.W.); 23205224164@stu.xust.edu.cn (S.T.)

[2] Shananxi Key Laboratory of Mine Electromechanical Equipment Intelligent Monitoring, Xi'an University of Science and Technology, Xi'an 710054, China

* Correspondence: zhangxh@xust.edu.cn

Abstract: Occupational health risk prediction of miners is a core issue to ensure the safety of high-risk operations. Current risk assessment methodologies face critical limitations, as conventional unimodal prediction systems frequently demonstrate limited efficacy in capturing the multifactorial nature of occupational health deterioration. This study presents a novel stacked ensemble architecture employing dual-phase algorithmic optimization to address these muti-parametric interactions. The proposed framework implements a hierarchical modeling paradigm: (1) a primary predictive layer employing heterogeneous base learners (Random Forest and Logistic Regression classifiers) to establish foundational decision boundaries, and (2) a meta-modeling stratum utilizing regularized logistic regression with hyperparameter optimization via grid search-assisted k-fold cross-validation. Empirical validation through comparative analysis reveals the enhanced ensemble achieves a mean accuracy of 90%. Receiver operating characteristic analysis confirms superior discriminative capacity (AUC = 0.89), surpassing conventional ensemble methods by 23.3 percentile points. The model's capacity to quantify nonlinear exposure–response relationships while maintaining computational tractability suggests significant utility in occupational health surveillance systems. These findings substantiate that the proposed dual-layer optimization framework substantially advances predictive capabilities in occupational health epidemiology, particularly in addressing the complex synergies between environmental hazards and physiological responses in confined industrial environments.

Keywords: miner safety; miners; machine learning; safety analytics; digital twins; stacking

check for updates

Academic Editor: Alexandre Carvalho

Received: 15 January 2025
Revised: 10 March 2025
Accepted: 12 March 2025
Published: 13 March 2025

Citation: Zhang, X.; Yang, W.; Yang, W.; Huang, B.; Wang, Z.; Tian, S. Occupational Risk Prediction for Miners Based on Stacking Health Data Fusion. *Appl. Sci.* **2025**, *15*, 3129. https://doi.org/10.3390/app15063129

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

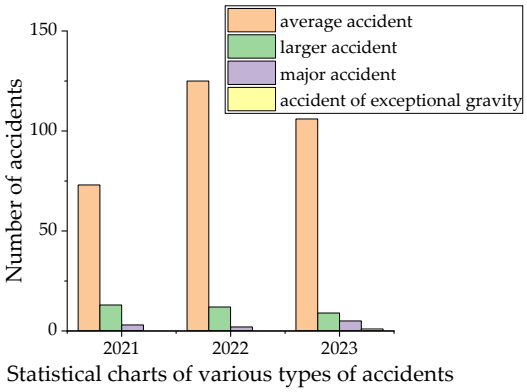
1. Introduction

With the continuous advancement of modern science, sensor technology and machine learning have witnessed significant progress and been extensively applied in human health monitoring. This is particularly evident in high-risk environments such as mines and construction sites, where these technologies have attracted considerable attention in recent years [1]. However, the mechanized excavation processes in mining operations are subject to multiple influencing factors, which can be categorized into environmental (e.g., mining conditions and geological features), technical (e.g., mechanical equipment and operational expertise), and organizational (e.g., safety measures and management practices). With the rapid growth of the global economy, both mining operations and the number of miners have significantly increased. Workers are frequently exposed to high

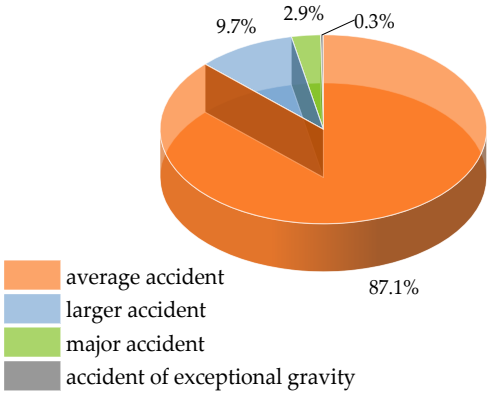
levels of dust, extreme temperatures, and high humidity, which have significantly impacted on their health, leading to a marked increase in health-related issues among miners [2].

Miners, as high-risk workers, are exposed to harsh environments that pose significant threats to their lives. Traditional safety monitoring methods primarily focus on environmental and operational conditions but often neglect the real-time monitoring of human physiological states. This oversight results in inadequate early warnings of potential hazards. In recent years, although many researchers have used sensors to monitor human vital signs, these efforts have primarily focused on disease diagnosis and the exploration of pathogenic factors [3]. Additionally, numerous experts have focused on the impact of environmental factors on miners, such as evaluating gas concentrations and assessing safe zones to identify potentially hazardous areas [4]. However, few studies analyze vital sign data in relation to workplace safety risks. Accurately assessing the health status and safety risks of personnel remains a significant challenge. Fa et al. [5] analyzed mining accidents in China and concluded that unsafe behaviors were the primary factors influencing safety. Wang et al. [6] analyzed mining accidents from 2013 to 2023. The majority of accidents were caused by unsafe behaviors of personnel, accounting for 74%. The second most common cause included machine states and environmental factors. Li Guowei et al. [7] demonstrated that addressing unsafe individual states significantly reduces the likelihood of unsafe behaviors among miners. A total of 349 safety incidents occurred between 2021 and 2023, including 89 incidents with 191 casualties in 2021, 139 incidents with 215 casualties in 2022, and 121 incidents with 281 casualties in 2023. Figure 1 shows the accident statistics in China from 2021 to 2023. Serious accidents are often caused by multiple factors, with unsafe personnel behavior being a key contributor. By predicting and assessing personnel insecurity factors, our system can reduce the likelihood of accidents at the source.

Machine learning is widely applied across various domains, with vital sign data, including physiological signals such as blood pressure, heart rate, and body temperature, collected by sensors, playing a crucial role in assessing individual health and predicting potential risks. Currently, many monitoring systems rely on data from a single sensor to assess health conditions, limiting the benefits of multi-sensor fusion for more comprehensive analysis. In this context, machine learning provides a powerful approach for risk prediction. Traditional classification methods, including support vector machines (SVM), decision trees, and random forests, have been applied in health data analysis. However, these methods typically rely on a single-model approach, which may be insufficient for effectively handling multi-dimensional and complex datasets.

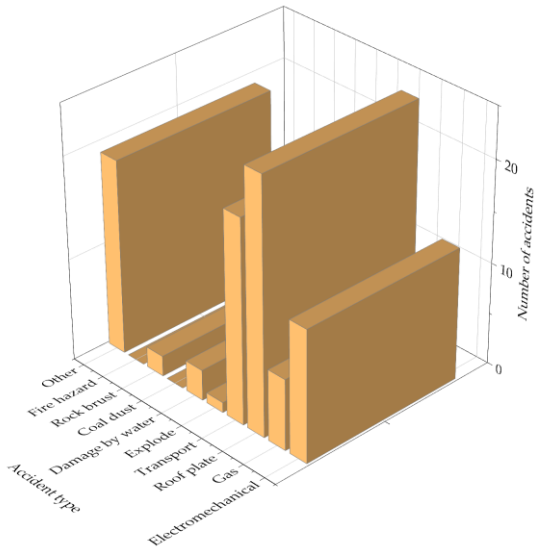


(**a**) Statistical charts of various types of accidents

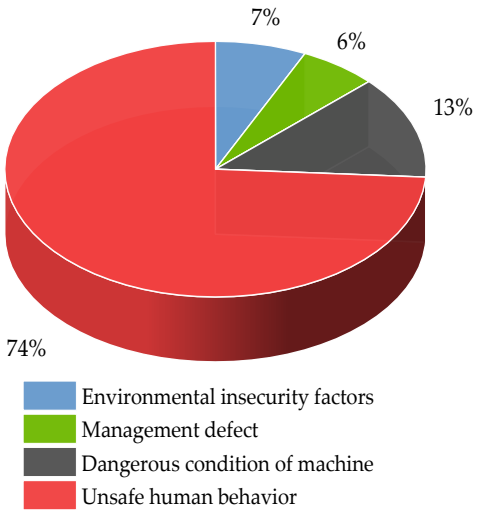


(**b**) Total accident analysis chart

Figure 1. *Cont.*



(**c**) Statistical map of various types of accidents in 2021



(**d**) Analysis diagram of different accident causes

Figure 1. Statistics of miner accidents by level in China from 2021 to 2023.

To address these challenges, this study proposes a novel stack-based security threat assessment framework. As an advanced ensemble learning technique, the stack-based algorithm enhances predictive performance by integrating the outputs of multiple heterogeneous base models. This innovative approach effectively mitigates the inherent limitations of individual models, significantly improving the accuracy and robustness. Furthermore, the proposed framework integrates a comprehensive monitoring system that tracks workers' health status in real time through advanced sensor networks, combined with environmental data streams to evaluate potential risks and threats in the workplace. This integrated methodology represents a significant advancement in occupational safety assessment, offering a more comprehensive and reliable approach to risk prediction and management. This method can systematically integrate various physiological indicators, such as heart rate, blood pressure, and body temperature. Compared to existing single classification methods, the innovation lies in enhancing the processing capability of complex, multi-dimensional data by combining the advantages of various classification models. Building on these findings, a new safety assessment framework is proposed, which handles uncertainty and noise in vital sign data. Additionally, an early warning system was developed based on real-time sensor data, which provides timely safety alerts for individuals in high-risk working environments. This system forms a more comprehensive health and safety assessment model, offering new insights for personnel safety management in such environments.

This paper is organized as follows. Section 2 provides a comprehensive literature review, including key studies on human health threat assessment and the application of machine learning techniques in this field. Section 3 elaborates on the methodological framework, elaborates on the key principles of the stacking algorithm, and provides an overview of the base models and the systematic approach to model construction. This section further explains the data acquisition protocols, preprocessing methods, and the design of the personnel risk assessment model. Section 4 presents a comparative analysis of conventional approaches and the proposed method, supported by empirical validation and experimental results that highlight the performance improvements. Finally, Section 5 discusses the limitations of the proposed model and suggests potential directions for future research to address these constraints and advance the field of health threat assessment.

2. Literature Review

With the rapid development of artificial intelligence, occupational disease prediction has become a research priority in the field of occupational health and safety. Many scholars have researched workplace safety and health risks, focusing on the following areas.

Yin et al. [8], in their preliminary study on the design of a safety monitoring system for miner operations, found numerous safety hazards in various operational positions and areas on the miner's digging face. Yang et al. [4] characterized high-stress conditions and the operating environment in the coal industry, which had a significant effect on the psychological state of the workers. Zhao et al. [9] conducted a comprehensive study of coal mine pneumoconiosis (CMDLD) and thyroid nodules and found that workplace conditions are key contributors to these occupational health problems. The occurrence of safety accidents in mining operations is primarily determined by factors such as the production system, work environment, personnel shifts, and operational conditions. Coal mining operations are a complex production system. It is vital that attention is paid to the management of operator health and that threat assessments are carried out to ensure the smooth running of production activities. Wang et al. [10] found uncertainty in the construction process of urban underground projects. They proposed a dynamic analysis method based on a fixed index (FST) and batch normalization (BN) to establish a dynamic safety risk model. Wang Dandan et al. [11] used the occupational health survey method, occupational health testing method, occupational health monitoring analysis and checklist analysis method to evaluate the effectiveness of occupational disease hazard control by integrating occupational health management and testing results.

Liu et al. [12] proposed a novel framework for a comprehensive Occupational Health and Safety Risk Assessment (OHSRA) model by integrating multiple sources of data such as physiological indicators, medical history and workplace environmental parameters. This model facilitates the dynamic evaluation of workplace hazard levels, thereby providing a scientific basis for targeted risk prevention and control measures. Some researchers have employed machine learning algorithms to develop risk assessment models for coronary heart disease (CHD), which can be utilized for personalized risk classification by identifying key risk factors related to both CHD and hypertension. Furthermore, this methodology can be used to reduce the occurrence of occupational disease incidents among miners. Hassanien et al. [13] has optimized contingency plans for occupational diseases and enhanced workplace safety by analyzing the characteristics and patterns of occupational disease incidents in mining environments.

Mestanza-Ramón et al. [14] proposed the risks associated with prolonged exposure to specific chemicals in the coal mining environment, which can cause damage to workers' DNA. Yang et al. [4] proposed a methodology that combines fault tree analysis with fuzzy polymorphic Bayesian networks to classify and quantify the explosion risk to reduce the risk of gas explosions to improve the safety of miners. The stacking algorithm is primarily used to address classification problems, as well as certain regression tasks. Hao et al. [15] applied the algorithm for sales prediction, while Gen et al. [16] utilized the model for classifying jujube varieties and identifying specific varieties.

Zhao et al. [17] proposed a predictive model for musculoskeletal disorder risk in miners. They used univariate and multivariate analyses to identify predictors and compared them with job characteristics using ROC curves and decision curve analysis (DCA). Yang et al. [3] showed that high altitude affected the respiratory, pulmonary, and cardiorespiratory health of miners. It also impacted their resting metabolic rate, heart rate variability, hematological health, and neurological health. Liao et al. [18] applied fuzzy set theory to evaluate the impact of the urban water environment on human health.

Risk assessment in the coal mining industry is crucial for identifying potential hazards, effectively managing them, and distinguishing between various types of risks in the workplace [19]. Miners' safety can be ensured by reducing the risk of accidents and injuries through safety assessment models [20]. Machine learning has become a key analytical tool for analyzing large datasets, detecting anomalous behaviors in various scenarios, and predicting potential hazards [13]. It facilitates the early detection of risks and offers valuable insights for field operations [21].

Due to the limited predictive capability and robustness of single models, Pan and Gong [22] proposed an approach to integrate the Stacking model for assessing the risk of electricity cost recovery among variable users in electricity utilities. The variable user data are processed, constructed, and filtered to optimize the model's generalization performance concerning sample distribution and feature attributes. Prusty et al. [23] employed machine learning techniques to assess the reliability of potential risks in power systems. Seungil Ahn et al. [24] employed various machine learning techniques, including the Stacking algorithm, to predict the occurrence of building fires. They developed a more accurate fire risk prediction model to overcome the limitations of traditional fire risk assessments. Jia and Han et al. [25] proposed a machine learning model based on Stacking multi-model fusion for predicting the lifespan of Insulated Gate Bipolar Transistors (IGBTs), which effectively improves prediction accuracy and efficiency. Liu et al. [26] applied a semi-supervised learning method and the Stacking algorithm to analyze collected samples of structural defects and established a safety assessment model to classify different safety levels. Li et al. [27] studied the safety attitudes and behaviors of coal mining operators. They found that educating miners on safety training and knowledge significantly improved unsafe behaviors. Seo et al. [28] developed an SEM model to explore the relationship between individuals and the whole in safety behavior. Risk assessment is a key step in identifying potential hazards and managing them effectively in the coal mining industry. It also helps differentiate the types of work risks [29]. Miner safety assessment models have been developed to enable mining companies to reduce the risk of accidents and injuries and to ensure the safety of miners [30].

Health threat assessment is a critical aspect of occupational safety and health. Its primary objective is to systematically evaluate the health risk factors to which miners may be exposed in complex underground environments. Scientific methods are employed to identify, quantify, and predict potential health threats. This approach provides a robust theoretical foundation for developing targeted preventive measures and intervention strategies. This study constructs a health dataset using physiological and environmental data from miners. It develops an assessment model based on fusion algorithms and fine-tunes parameters to enhance prediction accuracy. The methodology offers a comprehensive evaluation of miners' health status, analyzes potential health threats, mitigates occupational diseases, and ensures the long-term health and safety of miners.

3. Introduction to Methods and Data Processing

3.1. Database

The dataset, referred to as DATA, used in this study comprises 300 samples of human vital signs and historical data from the mining environment. The wearable device is employed to collect data on human vital signs, with blood pressure measured using the optical sensor integrated into the device. The blood vessels in the human body contract and expand with each heartbeat, resulting in variations in the absorption and reflection of light on the skin's surface. These variations are detected by the sensor and converted into electrical signals. Pulse waves, blood flow, and other physiological information are extracted through signal analysis. Body temperature is measured using a thermistor sensor,

which exhibits a well-established relationship between temperature and resistance. As the body temperature changes, the resistance of the thermistor varies accordingly. The resistance change is monitored, and a built-in algorithm converts the thermistor's resistance into a body temperature value. The smart bracelet continuously monitors the wearer's heart rate in real time using photoplethysmography (PPG) technology, based on the photoelectric volume pulse wave. The system uses a light-emitting diode (LED) as the light source to emit light, and a light sensor measures the reflection changes from blood vessels, capturing the pulse fluctuations of each heartbeat. The heart rate is then calculated using a signal-processing algorithm.

This study utilized diagnostic equipment (e.g., sphygmomanometer, thermometer) from the Affiliated Hospital of Hubei University for Nationalities to collect vital sign data, including blood pressure, heart rate, and body temperature, from 300 participants wearing smart bracelets, as shown in Table 1. Additionally, underground gas, temperature, and humidity sensors were installed to collect mining environment data. The model construction process is illustrated in Figure 2.

Table 1. Dataset data.

Number	Age (Years)	Blood Pressure (mmHg)	Heart Rate (Beats/min)	Temperature (°C)
1	27	130	142	36.5
2	26	118	144	35.5
3	40	118	162	35.1
4	39	138	152	37.5
5	30	122	144	37
6	41	120	100	36.7
7	33	126	126	35.8
8	22	130	157	37.2
9	20	120	140	36.2
10	31	138	143	36

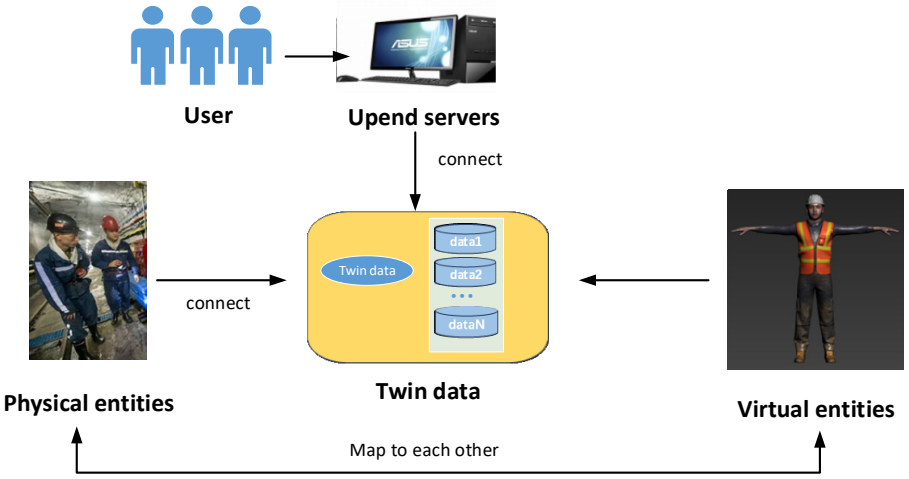


Figure 2. Digital twin (DT) model building.

Before constructing the threat assessment model, the collected dataset, referred to as 'data', must undergo preprocessing. The initial set of human physiological data may contain outliers caused by sensor errors, excessive noise, or other anomalies. Outliers can distort the results of data analysis and modeling. Therefore, it is necessary to remove them. A common method for outlier removal, based on prior conditions, is to utilize the mean and standard deviation. To address both global and local anomalies and achieve a better balance among different types of abnormal data, the LOF algorithm is employed for secondary data

pre-processing. This enhances the robustness and accuracy of data processing, ensuring that the outlier detection method is effective across various data distributions.

The data used in this study include human physiological data and work environment monitoring data, as shown in Figure 3. The raw data underwent preprocessing and feature engineering before being split into a training set, representing 70% of the total dataset, and a test set, representing 30%. The model was built using two algorithms: logistic regression and random forest. The predictions from both models were combined using the Stacking method. Parameter optimization was conducted through cross-validation and grid search to identify the optimal parameters, resulting in the development of the best-performing health threat assessment model.

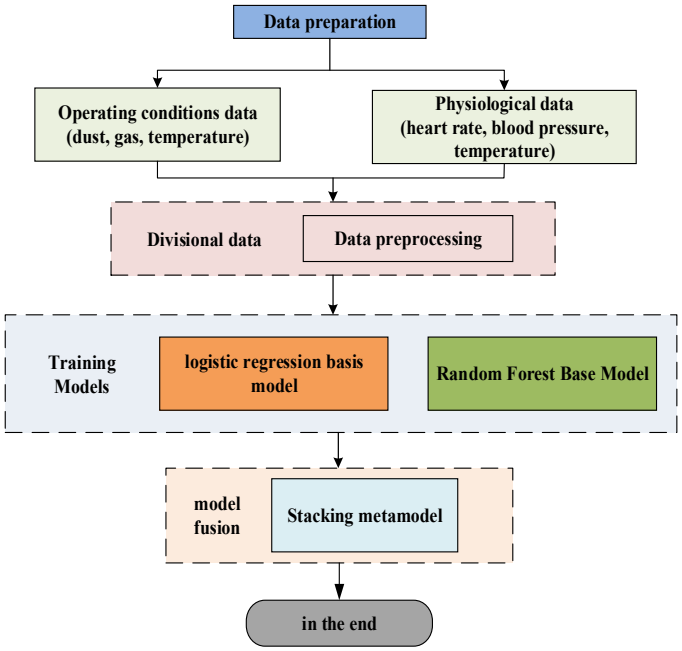


Figure 3. Health threat assessment model for miner operators.

Logistic regression and random forest algorithms were used to train the models separately. Before training, feature selection was performed to identify the most relevant features for the prediction outcomes. For this study, the key features are the data columns that indicate whether an individual's health is at risk.

3.1.1. Removal of Outliers Based on a Priori Conditions

One approach to outlier removal based on prior conditions involves using the mean and standard deviation. This method assumes that the data follow a normal distribution and compares individual data points to the mean. Outliers can be defined by the standard deviation method. Low outliers are defined as values less than $\mu - 3\sigma$, and high outliers are defined as values greater than $\mu + 3\sigma$, where μ is the mean and standard deviation of the data. If the data point deviates from the average value by more than 3 times the standard deviation, the data point will be marked as abnormal and removed due to noise, sensor failure or other atypical factors.

Suppose X is a data matrix of n rows and m columns, where n denotes the number of samples and m denotes the number of features. We use x_{ij} to denote the data in i rows and j columns of X. For each feature j, first, calculate its mean μ_j and standard deviation σ_j.

The mean, computed as the summation of all observed values in a dataset divided by the sample size, serves as a fundamental measure of central tendency in statistical analysis. This parameter is extensively employed to characterize the central position of

data distributions and quantify their overall trends, providing critical insights for data interpretation and hypothesis testing in empirical research.

$$\mu_j = \frac{1}{n}\sum_i^n = 1^{x_{ij}} \tag{1}$$

Standard deviation refers to the average deviation between each data point in the dataset and the average value. It is used to measure the degree of data dispersion.

$$\sigma_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\mu_j)^2} \tag{2}$$

Standardization is to transform data into a distribution with a mean value of 0 and a standard deviation of 1. Normalize the data matrix to obtain the normalized data matrix Z:

$$z_{ij} = \frac{x_{ij}-\mu_j}{\sigma_j} \tag{3}$$

Next, for each feature j, it is assumed to follow a normal distribution N (μ_j, ${\sigma_j}^2$). Based on the 3σ criterion, a threshold $T_j = 3\sigma_j$ is defined, where data points exceeding this threshold will be considered outliers. Finally, for each sample i, it is determined whether it is an outlier. For each feature j, the standardized value z_{ij} is checked to see if it lies within the interval $[-T_j, T_j]$. If, for all the features j, z_{ij} falls within this interval, the sample i is considered normal; otherwise, the sample i is considered abnormal.

3.1.2. Remove Outliers Based on LOF (Local Outlier Factor)

The LOF algorithm is commonly used for outlier detection, evaluating the degree of anomaly of a data point based on differences in local density. The fundamental principle of LOF is that if a point in a local neighborhood exhibits a significantly lower density than its neighbors, it is likely to be considered an outlier. LOF is particularly effective for high-dimensional and complex datasets, as it can identify outliers within local structures. The fundamental principle of the LOF algorithm is to compare the local density of a data point with that of its neighbors. The LOF algorithm evaluates the anomaly of each data point by comparing its local density to that of its neighbors. If a data point exhibits a significantly lower density than its neighbors, it is considered a potential outlier. The LOF algorithm utilizes "k-Nearest Neighbor" (KNN) to determine the local density of each data point. The density is calculated based on the distance between the data point and its k nearest neighbors. The LOF value is computed as follows:

Reachability distance: For a data point p, its reach distance (p, o) is the distance between p and its neighbor o. If the distance between the two is very close, the reachable distance is small; otherwise, it is larger. The formula for the accessible distance is:

$$reachability_distance(p,o) = max(k_distance(o), d(p,o)) \tag{4}$$

Locally accessible density:

$$\mathrm{lrd}(p) = \frac{1}{\sum\limits_{o\in N_k(p)} \text{reachability distance}(p,o)} \tag{5}$$

where $N_k(p)$ denotes the set of k nearest neighbors of the data point p. The LOF algorithm computes the local anomaly factor for each data point, which reflects the degree of

abnormality relative to its neighbors. Specifically, the LOF value is computed using the following formula:

$$\mathrm{LOF}(p) = \frac{\sum_{o \in N_k(p)} \frac{\mathrm{lrd}(o)}{\mathrm{lrd}(p)}}{|N_k(p)|} \tag{6}$$

where $N_k(p)$ is the k nearest points of data point p, and lrd(p) and lrd(o) are the locally accessible densities of data points p and o, respectively. When the LOF value is close to one, it indicates that the density of the data point is similar to that of its neighbors, suggesting normality. A LOF value greater than one indicates that the point is an anomaly, exhibiting a significantly lower density than its neighbors. Conversely, a LOF value less than one suggests that the point has a higher density than its neighbors, potentially classifying it as a local density core point.

After applying the LOF algorithm for outlier detection, the results indicate that certain data points are identified as anomalies. Figure 4 presents the LOF detection outcomes, with blue points representing normal data and red points indicating abnormal data. This algorithm accurately identifies outliers that deviate from their neighboring data points. As a local density-based outlier detection method, the LOF algorithm effectively detects observations that significantly differ from their neighbors. Its flexibility and unsupervised nature enable it to adapt to the requirements of various datasets, particularly in processing complex physiological features such as blood pressure, heart rate, and body temperature. By distinguishing the density of physiological data points within their local neighborhood, it becomes possible to more accurately identify outliers that deviate from neighboring data, while avoiding the exclusion of abnormalities that may hold clinical significance. Abnormal data points, potentially caused by sensor failure or other external factors, can be identified and excluded. In this study, outliers in physiological data are effectively detected and visually represented through chart visualization, offering more reliable input for the subsequent health assessment model. Thus, a more accurate and reliable database is provided for subsequent analysis and modeling.

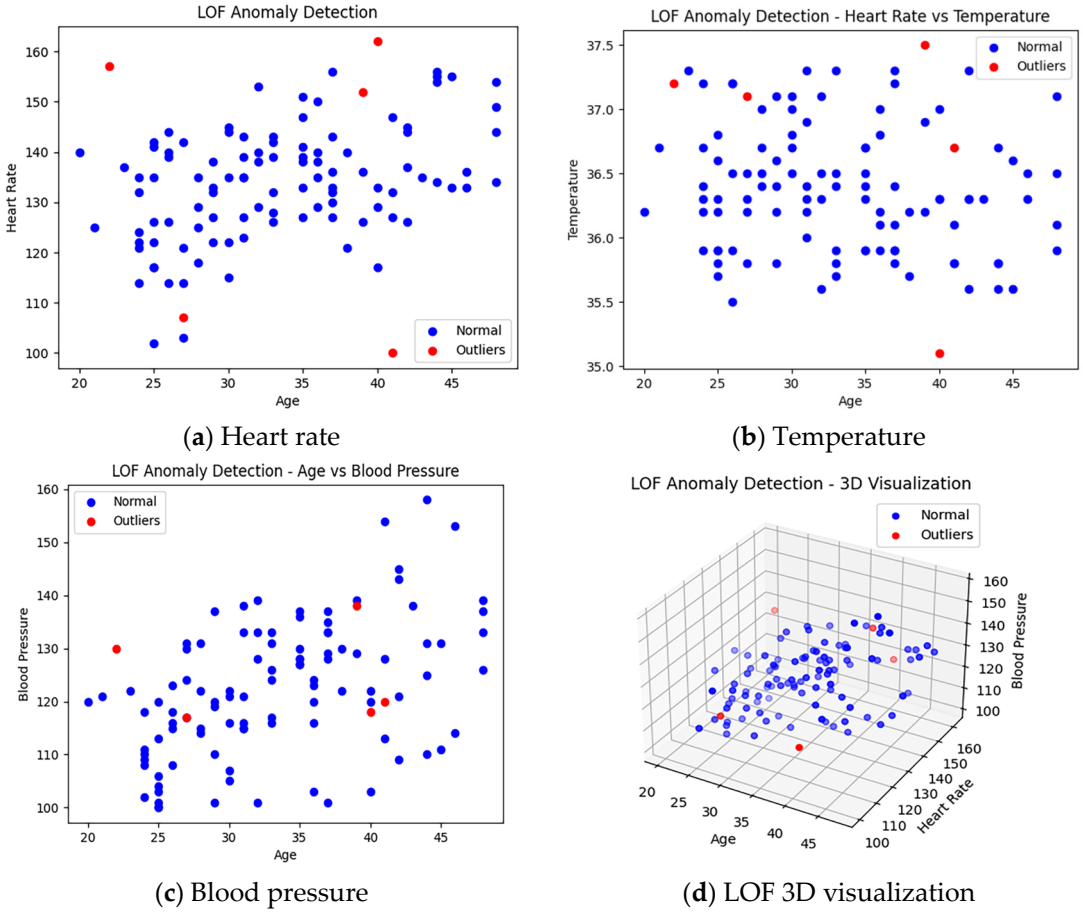
3.1.3. Standardized Processing

The original data in columns 2, 3, 4, and 5 are standardized, a common data preprocessing method. This transformation improves data reliability by converting it into a standard normal distribution with a mean of zero and a standard deviation of one. Z-score standardization (also referred to as standardization or z-transformation) is a technique that transforms data into a standard normal distribution with a mean of 0 and a standard deviation of 1. This method eliminates differences between features, which is particularly important for certain machine learning algorithms, especially distance-based algorithms such as k-nearest neighbors (KNN) and support vector machines (SVM).

$$x' = \frac{x - \mu}{\sigma} \tag{7}$$

where x' is the standardized data, μ is the mean of the data column, and σ is the standard deviation of the data column.

Outliers are removed from the original data using the standard deviation approach to form the data standardization matrix. A new data table set is obtained after standardization and the results are stored in Xdata, as shown in Table 2.



(**a**) Heart rate (**b**) Temperature

(**c**) Blood pressure (**d**) LOF 3D visualization

Figure 4. LOF algorithm outlier detection diagram.

Table 2. Dataset Xdata.

Serial Number	Age (Years)	Blood Pressure (mmHg)	Heart Rate (Beats/min)	Temperature (°C)
1	−0.964951	0.235731	1.498175	−0.037941
2	−1.070332	−0.353498	1.590002	−1.201818
3	0.492144	−0.353498	2.416447	−1.667369
4	0.40449	0.628551	1.957311	1.125937
5	−0.585565	−0.157089	1.590002	0.543998
6	0.621123	−0.255293	−0.430197	0.194835
7	−0.293255	0.039321	0.763557	−0.852655
8	−1.483436	0.235731	2.186879	0.776774
9	−1.664321	−0.255293	1.406347	−0.387104
10	−0.468687	0.628551	1.544088	0.194835

3.2. Model Evaluation

In stacking algorithms, model selection and training are critical, as they directly affect the performance and predictive accuracy of the final ensemble. Multiple base learners are trained on preprocessed data. To mitigate overfitting, techniques like cross-validation can be employed. Additionally, the weight of each base learner is determined using performance metrics such as accuracy, recall, and others. This paper selects the random forest and logistic regression algorithms as models to assess the health risks of personnel working as underground miners.

The stacking algorithm, also referred to as cascading generalization, is an ensemble learning method that combines the predictions of multiple models to train a meta-model, resulting in more accurate predictions. In the stacking algorithm, the data are initially split

into a training set and a test set, with the training set employed to train multiple distinct models. In the stacking algorithm, models are constructed using different algorithms or parameter configurations, which can be either classification or regression models. Next, the models generate predictions using the test set, resulting in multiple outputs. These outputs are then used as input features for the meta-model, which are combined to produce the final prediction. The Stacking algorithm can be utilized in health threat assessments for miner operators to predict factors such as health status and safety risks. The core concept of the Stacking algorithm is to train a meta-model that learns to optimally integrate the outputs of base models, as illustrated in Figure 5. The algorithm generally comprises two phases; the first phase involves training the base models, and the second phase focuses on training the meta-model.

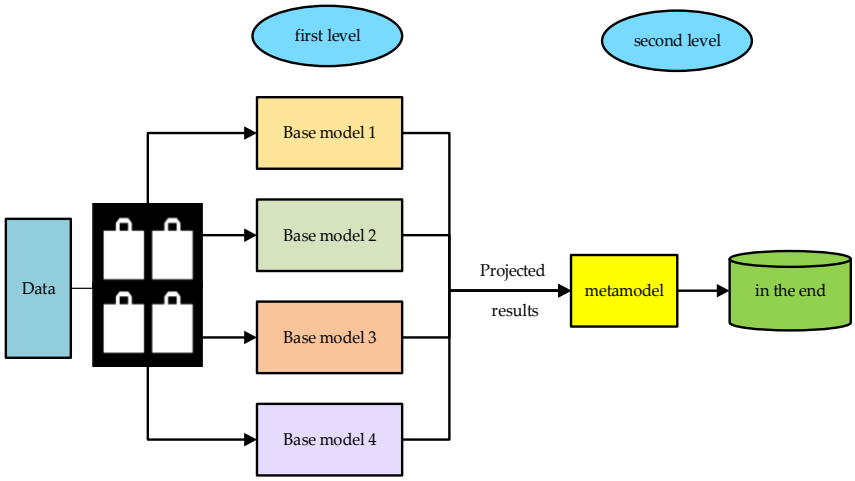


Figure 5. Stacking model fusion principle.

Random forest is an ensemble learning technique that builds multiple decision trees and combines their predictions to produce a final result. It can be applied to both classification and regression tasks, including health threat assessment for miner operators. The model is as follows:

$$y_{pred} = mode(\{y_1, y_2, \cdots, y_n\}) \tag{8}$$

where y_{pred} represents the prediction of the random forest model, y_n denotes the predictions from the n individual decision trees, and mode is the function that computes the plurality, selecting the category with the highest frequency as the final prediction.

The random forest algorithm is employed as a model for hazard assessment of underground miners, predicting the risk status of personnel. The model is constructed using personnel life characteristic data as input variables. Labels are assigned to each data point in both sets to indicate whether the personnel's health is at risk. The dataset consists of human body data, where X represents the feature set and Y represents the corresponding label set. A random seed is set to ensure result consistency. An instance of the random forest classifier model is created, and the training data along with corresponding labels are used to train the model. Classification predictions are made on the test data, and the results are stored. The model's accuracy is calculated by comparing the predicted outcomes with the true labels, counting the number of correct predictions, as illustrated in Figure 6.

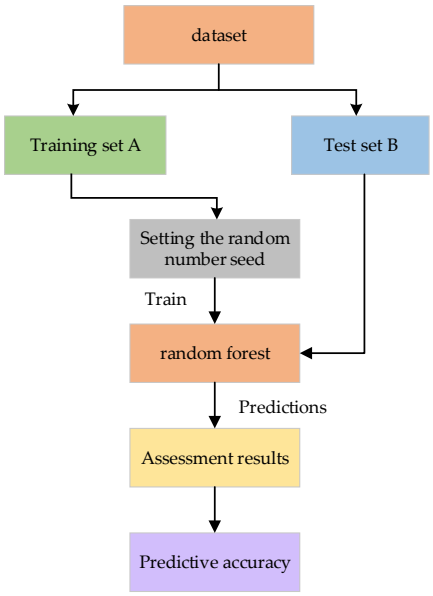


Figure 6. Random forest algorithm-based threat assessment model.

Single Models

Logistic regression can be employed as a model for analyzing health threats to miner operators in health risk assessments. The dataset used in this analysis includes vital signs, environmental factors, and a target variable indicating the presence or absence of a health threat. The logistic regression model is applied to predict the probability of a health threat. The model is used for assessing health risks to miner operators:

$$p(y = 1 \mid x) = \frac{1}{1 + e - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)} \quad (9)$$

where $p(y = 1 \mid x)$ is the probability of occurrence of a health threat given feature x; β_n represents the model parameters, which are fitted using the training dataset; x_n represents the input features, such as vital sign data and environmental factors. The model uses the train_test_split function to divide the dataset X_{data} and the associated labels Y into a training set and a test set. It then uses a logistic regression model to make classification predictions.

The logistic regression algorithm is employed as the second model for hazard assessment of underground miners, helping predict their risk levels. The model is constructed using data on personnel's life characteristics and underground environmental factors as inputs. Using Scikit-learn, an instance of the logistic regression model is created and trained on the training dataset. Prediction accuracy is computed by comparing the predicted labels with the true labels. For each correct prediction, the count is incremented by 1. The final accuracy is obtained by dividing the count by the total number of instances in the test set, as shown in Figure 7.

This study selects four evaluation models—KNN [31], random forest [32], XGBoost [33], and logistic regression [34]—as candidate base models (PyCharm Community Edition 2014.1.4). By comparing AUC, ROC, and accuracy metrics, algorithms with relatively higher accuracy were chosen as the base models, as shown in Table 3. Accuracy, ROC curve, and AUC were used to evaluate the model, as shown in Figure 8. The AUC represents the area under the ROC curve and is commonly used to assess model performance. A higher AUC indicates better model performance.

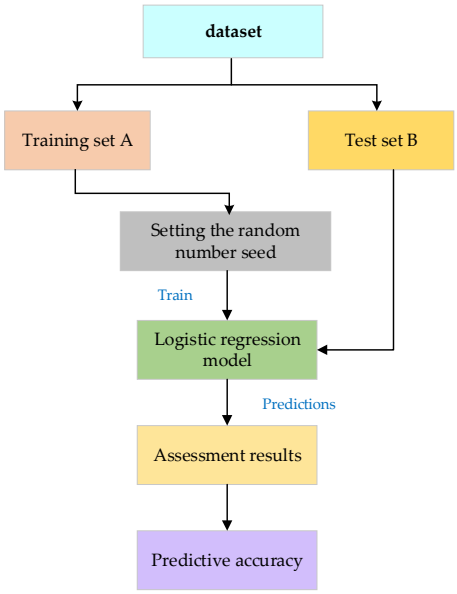
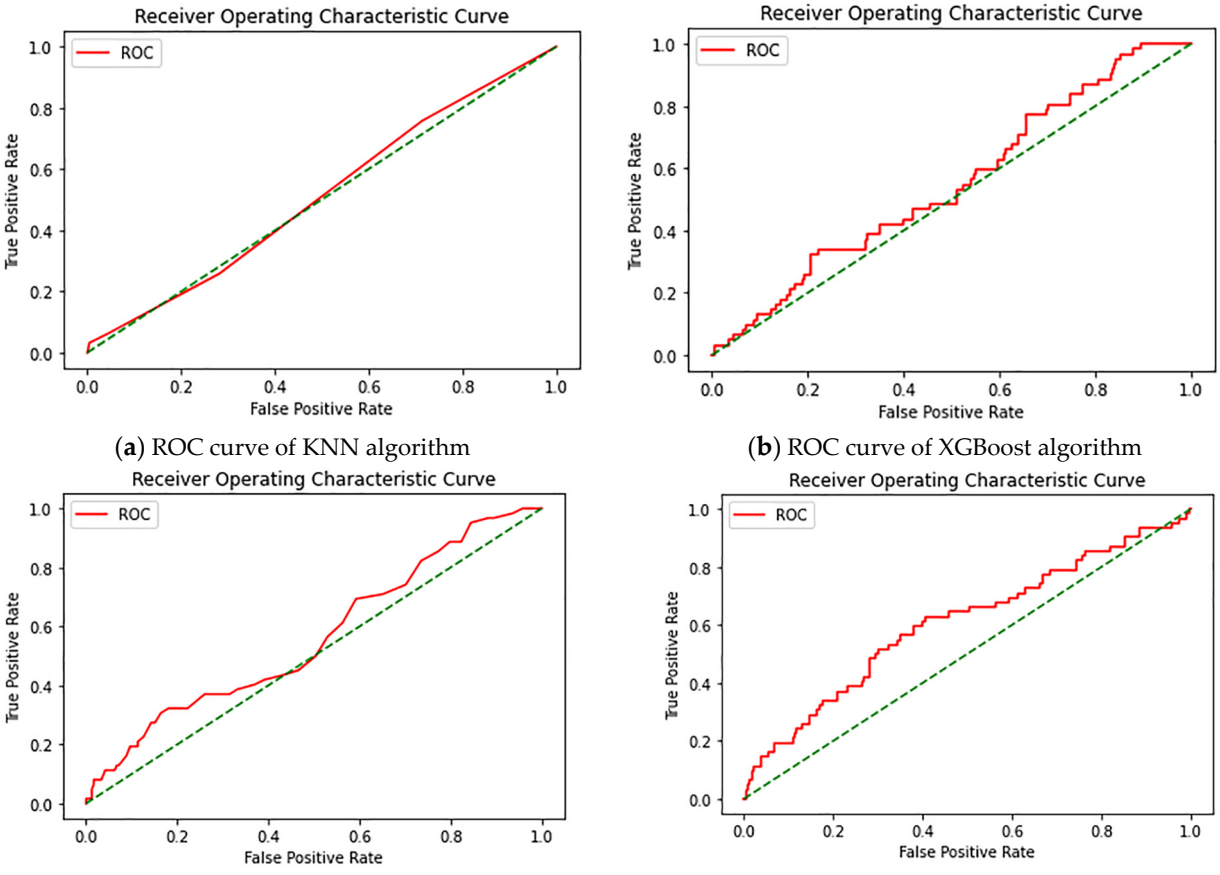


Figure 7. Threat assessment model based on logistic regression algorithm.

Table 3. Evaluation accuracy of different algorithms.

Assessment Model	KNN	XGboost	Random Forest	Logistic Regression
Accuracy	0.77	0.73	0.78	0.79



(**a**) ROC curve of KNN algorithm

(**b**) ROC curve of XGBoost algorithm

(**c**) Random forest algorithm ROC curve

(**d**) Logistic regression algorithm ROC curve

Figure 8. Comparison of ROC curves for different algorithms.

The ideal ROC curve is closer to the upper left corner. If the curve is near this corner, the model has a high True Positive Rate (TPR) and a low False Positive Rate (FPR), indicating strong classification performance. Based on the analysis of prediction accuracy and ROC curves in Table 1 and Figure 2, it is evident that the random forest and logistic regression models significantly outperform the others, indicating higher reliability. Consequently, logistic regression and random forest are selected as the final models in this study.

To train the Stacking meta-model, two base models, logistic regression and random forest, are first defined. The prediction results from these models are then used as input features to train the Stacking meta-model, with logistic regression serving as the meta-model. x_i denotes the feature vector of the first i sample and y_i denotes the label of the first i sample. Suppose the constructed Stacking model contains z-model and a meta-model (logistic regression), where the j-th model is $f_j(x)$ and the meta-model is $g(x)$. The meta-model is shown:

$$g(x) = \frac{1}{1 + exp(-\theta^T x)} \tag{10}$$

where θ parameters of the meta-model are represented. Set W_j to represent the weight of the j basic model, and the predicted output of the Stacking model is as follows:

$$y_i = g(\sum_{j=1}^{z} w_j f_j(x_i)) \tag{11}$$

The optimal model parameters are identified through the selection and adjustment of parameters, ensuring the model's generalization and predictive performance are maximized, thereby enhancing its overall effectiveness.

4. Experimental Results

4.1. Threat Assessment Results

This paper evaluates the model's performance using accuracy, recall, precision, and F1-score, as presented in Table 4. The results indicate that the stacked model achieves an accuracy of 89% on the test set, which significantly outperforms the logistic regression and random forest models. Furthermore, the stacked model demonstrates strong performance across the precision, recall, and F1-score metrics. The ROC curve also reveals a high level of discrimination for the stacked model (AUC = 0.87), as illustrated in Figures 9 and 10.

Table 4. Algorithmic model evaluation accuracy.

Assessment Model	Stacking	Random Forest	Logistic Regression
Accuracy	0.89	0.87	0.82

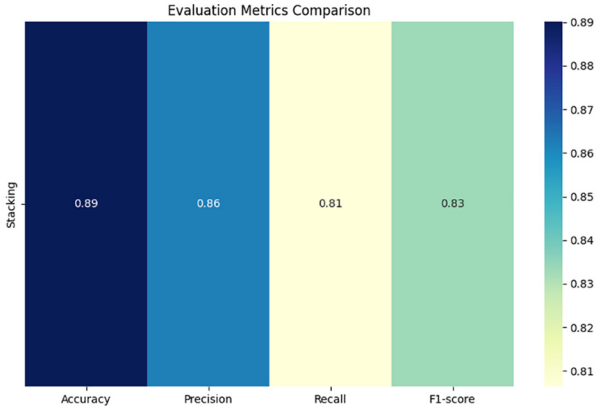


Figure 9. Algorithmic model evaluation of accuracy, precision, recall value, F1 value.

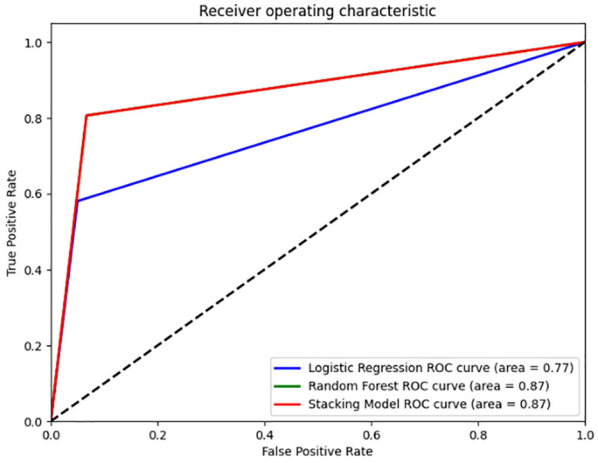


Figure 10. Algorithmic model to evaluate ROC-AUC values.

Based on industry safety norms and historical data statistics, if certain key physiological indicators exceed the normal range, the person is judged to be in danger. The specific judgment criteria include:

A heart rate consistently higher than 150 beats per minute is considered abnormal.

A body temperature greater than 39 °C or less than 35.5 °C is considered abnormal.

A blood pressure reading higher than 150 mmHg is considered abnormal.

The collected data are utilized to construct test datasets for validating the health monitoring and anomaly alarm functionalities of miners. First, a threat assessment experiment is performed on miner operators. The test dataset serves as input to the threat assessment model, and the results are compared with the actual threat outcomes. As shown in Table 5, life characteristics data from 10 personnel were randomly selected from the test set for evaluation. The actual threat outcomes were highly consistent with the threat assessment predictions, with the prediction probability exceeding 0.84. This confirms the validity and reliability of the threat assessment model for miner operators. A value of 1 indicates that the health of miner operators is at risk, while 0 indicates no risk.

Table 5. Results of threat assessment for miner operators.

Age	Blood Pressure (mmHg)	Heart Rate (Beats/min)	Threatening Real Results	Threat Assessment Results	Predictive Probability
44	133	152	1	1	0.8761
48	121	146	0	0	0.8638
35	138	182	1	1	0.8555
54	110	158	0	0	0.8547
42	120	194	1	1	0.8521
62	128	156	1	1	0.8512
37	120	145	0	0	0.8507
56	125	154	1	1	0.8496
43	132	147	1	1	0.8418

4.2. Improvement of Stacking Fusion Algorithm Based on the Network Search Algorithm

In machine learning, hyperparameters must be tuned to enhance model performance and generalization. Commonly employed traditional methods for hyperparameter optimization include grid search and random search. However, their efficiency and performance are often constrained by the dimensionality and scale of the search space. To enhance both efficiency and performance, optimization can be conducted using a network search algorithm. This algorithm is a graph-based optimization technique designed to identify the

optimal solution within the hyperparameter space, thereby maximizing a given evaluation metric. This algorithm identifies the optimal hyperparameter combination by automatically constructing a graph-based representation of the search space and performing the search within this structure. It is formally defined as follows. Let the hyperparameter space be denoted as H, where each hyperparameter combination is represented as h ∈ H, and the corresponding evaluation metric is denoted as f(h). The basic procedure of the search is outlined as follows:

- Defining the search space

The hyperparameter space H usually consists of a combination of several hyperparameters, $h = (h_1, h_2, \ldots, h_k)$, where h_i denotes the value of i hyperparameter. Thus, the search space can be expressed as follows:

$$H = \{(h_1, h_2, \ldots, h_k) \mid h_i \in H_i, i = 1, 2, \ldots, k\} \frac{1}{2} \tag{12}$$

where H_i indicates the range of values for the hyperparameter i.

- Constructing a search map

To construct the search graph automatically, it is essential to define the representation of the graph's nodes and edges. Typically, the nodes in a search graph correspond to hyperparameter combinations, while the edges represent the transformation relationships between these hyperparameters. Let the nodes u and v represent the hyperparameter combinations hu and hv, respectively. The edges connecting them can then be represented as:

$$E(u, v) = \{(h_u, h_v) \mid f(h_v) \geq f(h_u)\} \tag{13}$$

where $f(h)$ represents the evaluation metrics for the hyperparameter combination h.

- Evaluation Nodes

In the search graph, each node represents a set of hyperparameter values, and thus, each node needs to be evaluated to calculate its evaluation metric $f(h)$. The evaluation of the nodes can be performed using methods such as cross-validation for model training and assessment.

Termination conditions can be defined based on factors such as the number of evaluated hyperparameter combinations, the value of the objective function, and other performance metrics. Therefore, the network search algorithm must define termination criteria to determine when to halt the search process.

In this study, we conduct hyperparameter tuning and cross-validation (CV) on the Stacking fusion algorithm model developed in previous work, using the Grid Search CV function. The objective is to identify the optimal hyperparameters, train the model on the training set, and evaluate its performance on the test set to enhance model accuracy. First, logistic regression and Random Forest are used as models to construct the Stacking model. Subsequently, hyperparameter tuning and cross-validation are performed on the meta-model (logistic regression) to identify the optimal hyperparameters. The model is then trained on the training set and evaluated on the test set. K-fold cross-validation is employed, where the training set is divided into K subsets. Each time, one subset is used as the validation set, and the remaining subsets are used as the training set. Subsequently, the performance metrics of the model are computed on the validation set. The cross-validation procedure is outlined below:

$$CV = -\frac{1}{k} \sum_{k=1}^{K} \sum_{i \in V_k} logp(y_i \mid x_i, \theta_k) \tag{14}$$

where $\sum_{i \in Vk}$ denotes the validation set for the k-th fold, and $p(yi \mid xi, \theta_k)$ denotes the model predictions on the kth fold. The parameter θ_k denotes the hyperparameters of the model on the k-th fold. The cross-validation performance under different combinations of hyperparameters is obtained through cross-validation to find the optimal hyperparameters. The hyperparameters of the meta-model are then searched using the grid search technique as follows:

$$\lambda^* = argmin_{\lambda \in \Lambda} CV(\lambda) \tag{15}$$

where λ denotes the range of hyperparameter search and $CV(\lambda)$ denotes the cross-validation performance under the hyperparameter λ. Finally, using the found optimal hyperparameters λ*, the Stacking model is fitted on the training set and predicted on the test set. As demonstrated in Table 6 and Figure 11, the accuracy of the optimized Stacking model algorithm has increased from 89% to 90%. Additionally, the precision, recall, and F1-score show significant improvement compared to the pre-optimization period. As illustrated in Figure 12, the ROC curve and AUC value have also improved from 0.87 to 0.89, further enhancing the reliability of the model evaluation.

Table 6. Evaluation accuracy of the optimized algorithmic model.

Assessment Model	Stacking	Random Forest	Logistic Regression
Accuracy	0.90	0.89	0.82

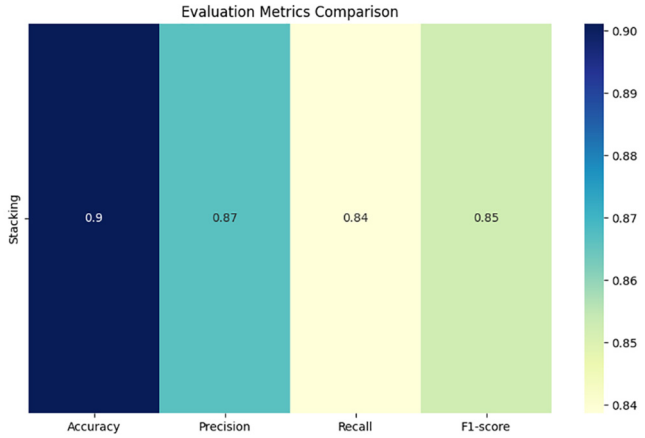


Figure 11. Optimized algorithmic model evaluation accuracy, precision, recall value, F1 value.

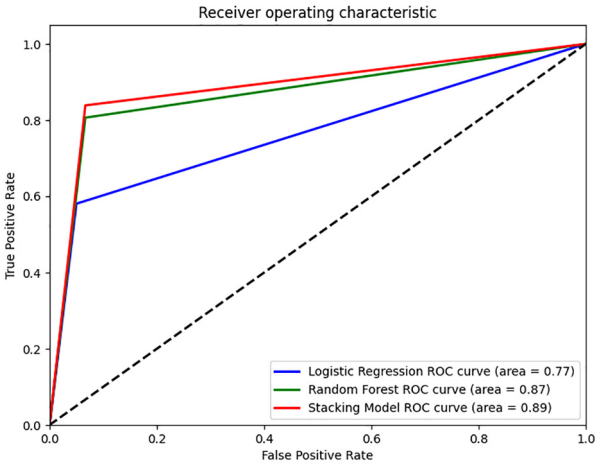


Figure 12. Optimized algorithmic model evaluating ROC-AUC values.

4.3. Safety Assessment Platform for Miner Personnel Based on DT

Digital twin (DT) technology, a novel approach, was first introduced by the National Aeronautics and Space Administration (NASA) and enables interactive mapping between physical and virtual models. The core concept involves using data to generate a virtual representation of a physical system. DT technology can simulate and control physical objects in the real world, with applications in real-time monitoring, action prediction, virtual decision making, and other domains. The development of DT enables simulations across various scales and real-world scenarios, meeting the requirements of diverse applications. This technology enhances system robustness.

The digital platform is developed and tested using Unity3D software (version 2022.3.34f1c1), a widely used virtual engine primarily designed for game development and interactive applications across personal computers (PCs), mobile devices, game consoles, and virtual reality platforms. The goal of establishing the DT platform is to create a simulation system that ensures rapid response times and closed-loop data processing. This system facilitates real-time data perception, efficient analysis, and high-speed operational responses, thereby improving overall system performance.

First, for model creation, data such as size, scale, and other relevant attributes are collected for the entities. Using 3dsMax (version 2022), 3D rendering is performed, and appropriate textures are applied to enhance the accuracy of the model's representation of solid objects. Once the file is saved in FilmBox (FBX) format, the virtual model is imported into Unity3D, where a mapping is established between the physical entities and their virtual counterparts, as shown in Figure 13. Next, data collection is carried out; the physiological information of miners working in mining faces is gathered through wearable sensors, which monitor vital signs such as blood pressure, body temperature, and heart rate.



Figure 13. A model diagram of the miner.

Miner personnel collect physiological data through the MKB0908 wearable device designed for human vital sign information collection. This device consists primarily of five chipsets: the YK1801 pulse sensor chip, the HR6707 pulse chip, the HR6816 gain chip, the SFB9712 algorithm chip, and the WD3703 temperature sensor. It is used to collect human pulse information through analog front-end chip HR607 + HR6816 and algorithm chip SFB9712 to output serial port signals such as blood pressure and heart rate. The WD3703 temperature sensor outputs temperature data via the GPIO interface and the SFB9712 algorithm chip. These data are displayed on the system's man–machine interface. The interface between the DT platform and users allows for monitoring, prediction, analysis, and optimization of personnel within the platform. It displays and analyzes data via the

Message Queue Telemetry Transfer (MQTT) communication protocol and DT platform, using a user interface (UI) specifically designed for risk assessment.

The interface displays health assessment results and facilitates human–computer interaction within the system. Managers can monitor mine operators through the DT platform and set thresholds for health parameters such as blood pressure. Currently, the miner Xiao Ming is detected to have a heart rate of 124 beats/min, blood pressure of 159 mmHg, body temperature of 36.6 °C, age of 30, and is in a healthy state, as shown in Figure 14. If blood pressure exceeds a defined threshold, an alarm is triggered to notify the manager. As shown in Figure 15, the worker Zhao Guang has a heart rate of 143 beats/min, blood pressure of 153 mmHg, body temperature of 37.1 °C, and is 38 years old. At present, his health status is in danger, and an alarm will be issued on the platform to remind the management staff to make emergency plans in time. Additionally, the platform can monitor real-time changes in environmental parameters, including methane concentration, carbon monoxide concentration, dust concentration, ambient temperature, and ambient pressure.



Figure 14. Personnel health the DT page is normal.



Figure 15. The DT platform interface where people's health is at risk.

Methane, a common and hazardous gas in coal mining, can cause explosions when its concentration becomes excessive due to its flammability and explosiveness. Therefore, monitoring methane levels is essential for safety management. Carbon monoxide, a colorless and odorless gas with toxic properties, presents a significant danger, as elevated concentrations can lead to unconsciousness or even death, resulting in large-scale accidents. Additionally, long-term exposure to high concentrations of dust can severely affect the respiratory system of workers, leading to pneumoconiosis. Thus, controlling dust levels is essential for ensuring the safety and health of the personnel.

5. Discussion

The results presented in this paper demonstrate that the proposed Stacking method for safety assessment of underground miners can effectively predict whether miners' health is at risk after training the model. The following findings were observed.

The experimental results demonstrate that the model exhibits strong accuracy and generalization capabilities. By integrating vital sign data (e.g., heart rate, blood pressure, body temperature) from multiple sensors with environmental data (e.g., gas concentration, temperature, and humidity), potential health threats can be effectively identified, particularly in mining environments with challenging conditions.

The accuracy of the evaluation model can be significantly improved by using industrial and mining data, along with vital sign data (e.g., heart rate, blood pressure, body temperature, age), as model inputs, and incorporating environmental data as auxiliary variables. This multi-dimensional data fusion method significantly improves the accuracy of health threat assessment and provides more comprehensive support for the safety management of miners.

The model can predict common health issues (e.g., abnormal heart rate, abnormal body temperature, hypertension) and identify abnormal vital signs and environmental factors to detect potential safety risks in advance. For example, if the model detects that a miner's heart rate and blood pressure are above or below predefined thresholds, combined with data on harmful gas concentrations and temperature in the mining area, it can indicate that the miner may be at risk of health threats due to environmental factors, triggering an early warning alarm.

At the same time, through the equipment positioning module worn by personnel, the position information of the threatened personnel can be obtained and handled in time.

The practical application prospect of this model is broad, especially in high-risk working environments such as mines and construction sites. It is very important to find the health problems of workers in time and take effective intervention. Applying this model to a health monitoring system can provide more personalized and timely health assessment for miners, thus reducing the incidence of occupational diseases and accidents.

The operating environment of a coal mine is complex. There is dust, high temperature, high humidity and other factors, which may affect the accuracy and stability of sensor data. In order to reduce the interference brought by environmental factors, deep learning models may be limited by computing resources, affecting real-time performance. Discussion of optimized computing performance can be added, such as the adoption of lightweight network architectures, edge computing architectures, and optimization strategies based on cloud–edge collaborative computing to ensure that systems can operate efficiently in resource-constrained environments. The acceptance degree of the intelligent evaluation system and the ease of use of the system will directly affect its promotion and application effect. Factors such as user training, interface friendliness and data privacy protection can be considered in the design of human–computer interaction systems to improve workers' willingness to use and trust the system.

To sum up, the stacking algorithm model proposed in this paper has important application value in the health threat assessment of miners. Although there are still some challenges, with the continuous optimization of data quality and model algorithm, this model is expected to become an effective tool for future miner safety management and workers' health monitoring.

6. Conclusions

This paper proposes a threat assessment model for underground miners based on an improved Stacking fusion algorithm. A health threat assessment model for mine operators is developed using a machine learning approach. The random forest and logistic regression models are employed, and model fusion is performed using Stacking to enhance the model's predictive performance. The final results demonstrate that the Stacking model proposed in this paper outperforms a single model, achieving a prediction accuracy of 90% and an ROC-AUC value of 0.87. This verifies the model's assessment performance and reliability, enabling pre-operation predictions for miner personnel and allowing for early safety evaluations of operators. Furthermore, the model improves operational efficiency and provides valuable data support for the health monitoring and safety management of miner operators.

Author Contributions: Conceptualization, X.Z. and W.Y. (Wenyu Yang); methodology, W.Y. (Wenjuan Yang) and X.Z.; software, B.H. and W.Y. (Wenyu Yang); validation, S.T. and B.H.; formal analysis, B.H.; investigation, S.T. and Z.W.; resources, W.Y. (Wenyu Yang); data curation, Z.W.; writing—original draft preparation, W.Y. (Wenyu Yang); writing—review and editing, W.Y. (Wenyu Yang) and S.T.; visualization, W.Y. (Wenyu Yang), B.H., Z.W. and S.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded through the financial support of The National Natural Science Founds of China (Grant No.52104166) and Key R & D project in Shaanxi (No.2023-YBGY-063).

Data Availability Statement: The data used to support the findings of this paper are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, Y.; Fu, G.; Lyu, Q.; Wu, Y.; Jia, Q.; Yang, X.; Li, X. Reform and development of coal mine safety in China: An analysis from government supervision, technical equipment, and miner education. *Resour. Policy* **2022**, *77*, 102777. [CrossRef]
2. Liu, F.-D.; Pan, Z.-Q.; Liu, S.-L.; Chen, L.; Chen, L.; Wang, C.-H. The estimation of the number of underground coal miners and normalization collective dose at present in China. *Radiat. Prot. Dosim.* **2017**, *174*, 302–307. [CrossRef]
3. Yang, S.; Tian, C.; Yang, F.; Chen, Q.; Geng, R.; Liu, C.; Wu, X.; Lam, W.-K. Cardiorespiratory function, resting metabolic rate and heart rate variability in coal miners exposed to hypobaric hypoxia in highland workplace. *PeerJ* **2022**, *10*, e13899. [CrossRef] [PubMed]
4. Yang, J.; Zhao, J.; Shao, L. Risk Assessment of Coal Mine Gas Explosion Based on Fault Tree Analysis and Fuzzy Polymorphic Bayesian Network: A Case Study of Wang zhuang Coal Mine. *Processes* **2023**, *11*, 2619. [CrossRef]
5. Fa, Z.; Li, X.; Liu, Q.; Qiu, Z.; Zhai, Z. Correlation in causality: A progressive study of hierarchical relations within human and organizational factors in coal mine accidents. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5020. [CrossRef] [PubMed]
6. Wang, H.; Qi, Q.; Ling, Y.; Qi, Q.; Liu, Y.; Sun, Z. Statistical analysis and countermeasures of major accidents in coal mines in China. *China Saf. Sci. J.* **2024**, *34*, 9–18.
7. Guowei, L.; Shuicheng, T. Study on intelligent detection system for unsafe state of pre-job personnel in coal mine: Taking Hong liulin Coal Mine as example. *J. Saf. Sci. Technol* **2023**, *2*, 106–113. [CrossRef]
8. Yin, J.; Shi, L.; Liu, Z.; Lu, W.; Pan, X.; Zhuang, Z.; Jiao, L.; Kong, B. Study on the variation laws and fractal characteristics of acoustic emission during coal spontaneous combustion. *Processes* **2023**, *11*, 786. [CrossRef]

9. Zhao, F.; Zhang, H.; Ren, D.; Li, C.-M.; Gu, Y.; Wang, Y.; Lu, D.; Zhang, Z.; Lu, Q.; Shi, X.; et al. Association of coal mine dust lung disease with Nodular thyroid disease in coal miners: A retrospective observational study in China. *Front. Public Health* **2022**, *10*, 1005721. [CrossRef]
10. Wang, Q.; Zhang, J.; Zhu, K.; Guo, P.; Shen, C.; Xiong, Z. The safety risk assessment of mine metro tunnel construction based on fuzzy Bayesian network. *Buildings* **2023**, *13*, 1605. [CrossRef]
11. Wang, D.-D.; Hu, T.; Zhai, H. Evaluation on control effect of occupational hazards in a coal mine resource integration project in Shaanxi Province. *Occup. Health* **2022**, *38*, 2133–2136.
12. Liu, H.-C.; Wang, J.-H.; Zhang, L.; Chen, Q.-Y. An integrated model for occupational health and safety risk assessment based on probabilistic linguistic information and social network consensus analysis. *J. Oper. Res. Soc.* **2024**, *75*, 1308–1324. [CrossRef]
13. Hassanien, A.E.; Darwish, A.; Abdelghafar, S. Machine learning in telemetry data mining of space mission: Basics, challenging and future directions. *Artif. Intell. Rev.* **2020**, *53*, 3201–3230. [CrossRef]
14. Mestanza-Ramón, C.; Jiménez-Oyola, S.; Montoya, A.V.G.; Vizuete, D.D.C.; D'orio, G.; Cedeño-Laje, J.; Straface, S. Assessment of Hg pollution in stream waters and human health risk in areas impacted by mining activities in the Ecuadorian Amazon. *Environ. Geochem. Health* **2023**, *45*, 7183–7197. [CrossRef] [PubMed]
15. Hao, X.; Chen, Z.; Yi, S.; Liu, J. Application of improved Stacking ensemble learning in NIR spectral modeling of corn seed germination rate. *Chemom. Intell. Lab. Syst.* **2023**, *243*, 105020. [CrossRef]
16. Gen, L.; Ma, M.; Xiao, Z.; Liu, Y. Jujube classification based on a convolution neural network with multi-channel weighting and information aggregation. *Food Sci. Technol. Res.* **2019**, *25*, 647–656.
17. Zhao, H.L.; Dou, H.; Yong, X.T.; Liu, W. Construction and validation of a musculoskeletal disease risk prediction model for underground coal miners. *Front. Public Health* **2023**, *10*, 9175. [CrossRef]
18. Liao, D.; Jin, Y.; Zhang, X. Environmental health risk assessment of urban water sources based on fuzzy set theory. *Open Geosci.* **2023**, *15*, 20220565. [CrossRef]
19. Zhang, G.; Wang, E.; Zhang, C.; Li, Z.; Wang, D. A comprehensive risk assessment method for coal and gas outburst in underground coal mines based on variable weight theory and uncertainty analysis. *Process Saf. Environ. Prot.* **2022**, *167*, 97–111. [CrossRef]
20. Chai, N.; Zhou, W. Evaluating operational risk for train control system using a revised risk matrix and FD-FAHP-Cloud model: A case in China. *Eng. Fail. Anal.* **2022**, *137*, 106268. [CrossRef]
21. Ayvaz, S.; Alpay, K. Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time. *Expert Syst. Appl.* **2021**, *173*, 114598. [CrossRef]
22. Pan, G.; Gong, M. Stacking Model Fusion Based Risk Identification Method for Electricity Recovery of Dedicated Transformer Customers. *Electr. Power Autom. Equip.* **2021**, *41*, 152–160.
23. Prusty, B.R.; Krishna, S.M.; Bingi, K.; Gupta, N. Risk-based reliability assessment of modern power systems using machine learning and probability theory. In Proceedings of the 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), Bangalore, India, 21–22 April 2023; IEEE: Piscataway, NJ, USA, 2023.
24. Ahn, S.; Won, J.; Lee, J.; Choi, C. Comprehensive Building Fire Risk Prediction Using Machine Learning and Stacking Ensemble Methods. *Fire* **2024**, *7*, 336. [CrossRef]
25. Jia, J.; Han, D. Fault Diagnosis Algorithm for Pumping Unit Based on Stacking Model Fusion. *Oil-Gasfield Surf. Eng.* **2023**, *10*, 6896.
26. Liu, D.; Zhang, W.; Dai, Q.; Chen, J.; Duan, K.; Li, M. Safety evaluation method for operational shield tunnels based on semi-supervised learning and a Stacking algorithm. *Tunn. Undergr. Space Technol.* **2024**, *153*, 106027. [CrossRef]
27. Li, Y.; Wu, X.; Luo, X.; Gao, J.; Yin, W. Impact of safety attitude on the safety behavior of coal miners in China. *Sustainability* **2019**, *11*, 6382. [CrossRef]
28. Seo, H.-C.; Lee, Y.-S.; Kim, J.-J.; Jee, N.-Y. Analyzing safety behaviors of temporary construction workers using structural equation modeling. *Saf. Sci.* **2015**, *77*, 160–168. [CrossRef]
29. Tong, R.; Yang, Y.; Ma, X.; Zhang, Y.; Li, S.; Yang, H. Risk assessment of Miners' unsafe behaviors: A case study of gas explosion accidents in coal mine, china. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1765. [CrossRef]
30. Kharzi, R.; Chaib, R.; Verzea, I.; Akni, A. A Safe and Sustainable Development in a hygiene and healthy company using decision matrix risk assessment technique: A case study. *J. Min. Environ.* **2020**, *11*, 363–373.
31. Bzdok, D.; Krzywinski, M.; Altman, N. Machine learning: Supervised methods. *Nat. Methods* **2018**, *15*, 5. [CrossRef]
32. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; IEEE: Piscataway, NJ, USA,, 1995; Volume 1.

33. Ribeiro, M.H.D.M.; dos Santos Coelho, L. Ensemble approach based on bagging, boosting and Stacking for short-term prediction in agribusiness time series. *Appl. Soft Comput.* **2020**, *86*, 105837. [CrossRef]
34. Nti, I.K.; Adekoya, A.F.; Weyori, B.A. A comprehensive evaluation of ensemble learning for stock-market prediction. *J. Big Data* **2020**, *7*, 20. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.