Hindawi
Journal of Electrical and Computer Engineering
Volume 2022, Article ID 9802512, 1 page
https://doi.org/10.1155/2022/9802512

Retraction

Retracted: Prediction of Coal Mining Subsidence Based on Machine Learning Probability Theory

Journal of Electrical and Computer Engineering

Received 22 November 2022; Accepted 22 November 2022; Published 18 December 2022

Journal of Electrical and Computer Engineering has retracted the article titled "Prediction of Coal Mining Subsidence Based on Machine Learning Probability Theory" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the editorial board.

References

[1] X. Tian, X. Jin, and X. He, "Prediction of Coal Mining Subsidence Based on Machine Learning Probability Theory," *Journal of Electrical and Computer Engineering*, vol. 2022, Article ID 9772539, 10 pages, 2022.

[2] L. Ferguson, "Advancing research integrity collaboratively and with vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Journal of Electrical and Computer Engineering
Volume 2022, Article ID 9772539, 10 pages
https://doi.org/10.1155/2022/9772539

Research Article

Prediction of Coal Mining Subsidence Based on Machine Learning Probability Theory

Xiaohong Tian,[1] Xinyuan Jin ,[2] and Xinwei He[3]

[1]*Coal and Chemical Industry College, Shaanxi Energy Institute, Xianyang 712000, Shaanxi, China*
[2]*Wenzhou University of Technology, Wenzhou 325035, Zhejiang, China*
[3]*Research and Production Department, Xi'an AiSheng Technology Group Company, Xi'an 710065, Shaanxi, China*

Correspondence should be addressed to Xinyuan Jin; jinxy@wzu.edu.cn

Received 7 March 2022; Revised 11 April 2022; Accepted 22 April 2022; Published 21 May 2022

Academic Editor: Xuefeng Shao

Geological disasters such as subsidence caused by mining have been continuously affecting people's production and life. Therefore, how to predict the occurrence of geological disasters such as mining subsidence is an urgent technical problem. The study of mining subsidence prediction can effectively guide the safe production of mining area, and it is of great practical significance. Probability theory is a discipline used to express uncertainty. It provides methods and axioms to quantify and deduce uncertainty, which allows us to reason when uncertainty exists. This paper mainly studies the prediction process of mining subsidence based on machine learning probability theory. In this paper, the mining subsidence problem is studied. The main research contents are as follows: through the research and analysis of the mining subsidence prediction method, the probability integral method is determined as the theoretical basis for the study. The mining parameters are obtained from the three-dimensional geological body model of the mining area, and the prediction parameters are calculated by the least square fitting. This paper makes a detailed plan for the overall design of mining subsidence prediction module, adopts the form of independent research and development and joint development with the existing software, designs five submodules according to the design objectives, principles, and functional needs of the module, and studies the application of mining subsidence prediction module. In this study, it is found that the predicted data of the maximum settlement value appear at the 23rd point, and the maximum value is 1567 mm, which is completely consistent with the change trend of the measured value. The settlement value is calculated from the edge position of the working face to the position where the chariot gradually increases, which conforms to the change law of surface movement and settlement.

1. Introduction

Machine learning is an interdisciplinary specialty, covering probability theory knowledge, statistical knowledge, approximate theory knowledge, and complex algorithm knowledge. It uses computer as a tool, is committed to simulating human learning methods in real time, and divides the existing content into knowledge structures to effectively improve learning efficiency. Coal is China's main energy source, accounting for more than 70% of the composition of disposable energy. With the rapid development of China's economy, the demand for coal has increased sharply. The large-scale exploitation of underground resources has caused surface movement and deformation and damage to surface buildings and other facilities, causing land collapse and damage; the quality of the soil is reduced, and the effective land use area is reduced. The problem will also cause groundwater system and ecological environment damage and geological disasters such as landslides and mudslides [1, 2]. According to the survey, the area of forests directly damaged by mining in China has reached 1.06 million hectares and the area of destroyed grassland is 263,000 hectares. The country has occupied a total of about 5 million hectares of land, about 1.57 million hectares of damaged land, and still has 40,000 hectares per year. The speed of the increase is only 10% [3–5].

According to estimates, for every 10,000 tons of coal in China, the average collapse is 0.2 hectares; in the dense plain mining area of the village, about 10,000 people need to migrate for every 10 million tons of coal [6]. It can be seen that the impact of the disaster caused by mining subsidence on the environment is continuous, multifaceted, and often severely destructive, which has a serious impact on the economic development of mining enterprises and the environmental protection of mining areas, and also threatens the safety of people's lives and property around the mining area restricting the sustainable development of enterprises [7, 8]. Mining subsidence is a time and space development process. The movement of surface and rock layers not only is accompanied by the mining process but also has a certain persistence to the damage of the surface and buildings [9]. Mining subsidence is expected to be the core content of mining subsidence disciplines. Mining subsidence research is also of great significance for the safe production of mining areas and the development of national economy [10]. It is necessary to carry out in-depth study on the mining subsidence of the mining area [11, 12].

The characteristics of land subsidence disasters are wide distribution, long duration, and high intensity, causing serious damage to surface buildings, underground pipelines, and aquifers. Land subsidence is a loss caused by the gradual and cumulative decline of land elevation in a large area. The deformation is slow, measured in millimeters and centimeters. The disaster is obviously regional, mainly in plains and basins. Once a disaster occurs, it often forms a disaster with large disaster area, serious losses, and difficult to manage. Therefore, it is of great significance to carry out ground settlement stability evaluation and settlement deformation prediction research. Lei combined the experimental data of the settlement area and obtained the final settlement deformation value of all the monitoring points by the nonlinear curve fitting of the monitoring data by the inverse tangent function model. The model proposed by him can predict the trend of settlement deformation at the monitoring point. The correlation coefficients are all above 0.937, indicating that the prediction model has strong reliability. He processed the final settlement profile by processing the final settlement prediction values of 60 monitoring points and using the Kriging interpolation method. Kriging interpolation, also known as spatial local interpolation, is based on variogram theory and structural analysis. The map can predict the overall distribution characteristics of land subsidence deformation in the study area. He then obtained risk partitions by combining the sedimentation rate and residual sedimentation deformation of the study area, but the practicability is not strong [13, 14]. Through geological surveys, geophysical exploration, and theoretical analysis, Sun studies the formation mechanism, related influencing factors, and distribution laws of mining subsidence. He introduced the preliminary assessment and prediction of land subsidence caused by the collapse of goaf. His experimental results show that mining is the main cause of land subsidence, and groundwater infiltration accelerates this process. He proposed a three-zone model to analyze and evaluate the stability of the goaf. Based on the model, he concluded that the formations in the goaf and the surrounding nonmined areas are unstable, and once disturbed, surface subsidence and ground fissures may occur [15, 16]. Suh takes the collapse hazard map near an abandoned coal mine in a geographic information system (GIS) environment as an example. Geospatial databases were constructed using mine drift maps, topographic maps, land use maps, road maps, building plans, borehole data, and settlement checklists showing past settlement events. It includes six mesh type factor layers (i.e., use multiple mine drifts and estimated minefield plates, land use, distance from the nearest rail, distance from the nearest road, and slope to calculate the impact area instability (IAI) layer). The database identifies the relationship between past subsidence events and these factors. There are two IAI factors that combine the complex effects of the ground IAI and calculate the depth of each subsurface and its extraction angle. Six CV layers (one for each factor) were linearly combined to generate a subsidence hazard map, which represents the relative vulnerability of the subsidence in the study area. Then, by comparing the estimated sensitivity level over the entire grid cell range with the actual settlement occurrence, the area under the cumulative frequency map technique is used to verify the predicted subsidence risk. The accuracy of his proposed GIS analysis model in settlement prediction is 91.09%. In addition, through field investigations, he found that in areas with high subsidence hazards, the damage associated with subsidence was severe (the National Coal Council stated that its damage rating was 4 or 5). He determined the factor (slope angle) that was negatively correlated with the subsidence prediction by sensitivity analysis [17, 18]. The main feature of the differential SAR radar interferometer is its high spatial resolution and high accuracy over a wide coverage area. Due to its unique advantages, this technology is widely used to monitor surface deformation. However, in coal mining areas, large-scale collapse of the surface in a short period of time results in inaccurate D-InSAR results and limits its ability to monitor mining subsidence [19, 20]. Diao proposes a data processing method that overcomes these shortcomings by combining D-InSAR with a probabilistic integration method for predicting mining subsidence. Five RadarSat-2 images from Fengfeng Coal Mine in China were used to demonstrate the proposed method and evaluate its effectiveness. He uses this method to monitor surface deformation over thousands of square kilometers and to identify more than 50 areas affected by settlement [21–23]. The relevant contents of the above research on coal mining subsidence prediction are analyzed in detail. It is undeniable that these studies have greatly promoted the development of corresponding fields. We can learn a lot from methods and data analysis. However, there is relatively little research on coal mining subsidence prediction based on machine learning probability theory. It is necessary to fully apply these algorithms to the research in this field.

The innovations of this paper are as follows: (1) The feature selection problem and related technologies are introduced in general. The related technologies and main problems of the feature selection algorithm of Filter mode are analyzed and discussed. The defects and shortcomings of the current algorithm are pointed out through specific examples. The discussion and analysis of these contents is the basis for the next step. (2) A feature evaluation and selection algorithm CoFS based on Banzhaf rights index is proposed. Firstly, using the Banzhaf power index method of interest distribution in the alliance game theory, the number of winning (dependent) alliances possessed by each feature is statistically analyzed and then weighted. The weights will have the characteristics of dependent characteristics. Separated from other features, the winning alliance's judgment depends on the number of features in the alliance that are dependent on the feature to be evaluated; then, by combining mRMR evaluation criteria, features with high correlation, internal dependence, and low redundancy are obtained. The subset improves the generalization ability and classification performance of the learning algorithm. (3) A dynamic weighting-based feature selection algorithm DWFS is proposed. The main feature of the algorithm is that the candidate features are assigned dynamic weights according to their relationship with the selected features (redundancy or dependence). The weights of the candidate features are dynamically adjusted each time the new features are selected, wherein the candidate features are adjusted. The weight is based on the ratio of the mutual information between the mutual information and the original information after the newly selected feature is a known condition. The DWFS algorithm can select highly correlated, at lower time complexity, by continuously increasing the weights of candidate features that have dependencies on selected subsets of features and reducing the weights of features with which they have redundant relationships.

2. Proposed Method

2.1. Machine Learning Theory. The possible realization of machine learning combines the problems of inductiveness, clear expression of probabilities, empirical basis of deductive reasoning, and learnability. Machine learning is a subject that makes model assumptions for research problems, uses computers to learn model parameters from training data, and finally forecasts and analyzes the data. The presupposition of machine learning is that there is a general-purpose machine with enough logic to simulate all the laws of logic. On this basis, the machine learning system can be realized through two complementary ways: deductive reasoning learning and inductive statistical learning. The deductive reasoning model, because its inference rules are quasi-empirical, can input hypotheses and empirical data into the system and obtain new knowledge or improve system performance through the abduction method. The inductive statistical model reinvents the inductive problem, and the probabilistic and statistical associations, combined with the principle of simplicity, enable the machine to determine a hypothesis from the complex hypothesis space and make the learnability problem clear. Hence, we can acquire new knowledge.

There are fields of application such as medicine and automation. Machine learning primarily uses known data to

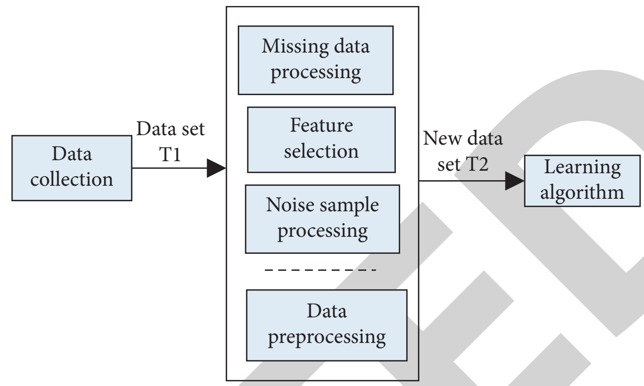

Figure 1: Machine learning steps.

learn and reason about the important properties of unknown, potential probability distributions, revealing the relationships between variables (or features) in data samples. One of the most important factors affecting machine learning performance is the quality of the data provided to the system. With the rapid development and wide application of computer technology, high-dimensional large-scale data is constantly emerging and accumulating. There are a large number of redundant and irrelevant features in these high-dimensional data, which puts higher requirements on existing machine learning algorithms, bringing great challenges. All machine learning problems are eventually transformed into an optimization problem, such as minimizing the mean-square error or maximizing the likelihood function. Feature selection is one of the important research topics in the fields of machine learning, pattern recognition, and statistics and is an important and commonly used means of data preprocessing. Feature selection is to select an optimal feature subset from the original feature space according to the distribution characteristics of the samples and based on some evaluation criteria. The selected feature subset has similar or better classification performance than the original feature space. The feature selection algorithm can effectively eliminate redundant features and irrelevant features, can improve the generalization performance and operational efficiency of machine learning algorithms, and has been widely promoted in practical applications. Feature selection algorithms are mainly divided into three categories: filter, wrapper, and embedded mode. Filter mode is favored because of its fast speed and versatility. However, the existing filter feature selection algorithm has the following problems: either select the most distinguishing features as the optimal feature subset, or select some features with higher discrimination ability and no redundancy between each other as the most excellent feature subsets, and in reality, partially redundant features have mutual support and dependency characteristics. Treating such features as redundant culling will reduce the performance of the learning algorithm. Usually, the feature selection process is completed before the machine learning algorithm is trained and belongs to the data preprocessing category.

The machine learning steps are shown in Figure 1. The pros and cons of the selected subset directly determine the

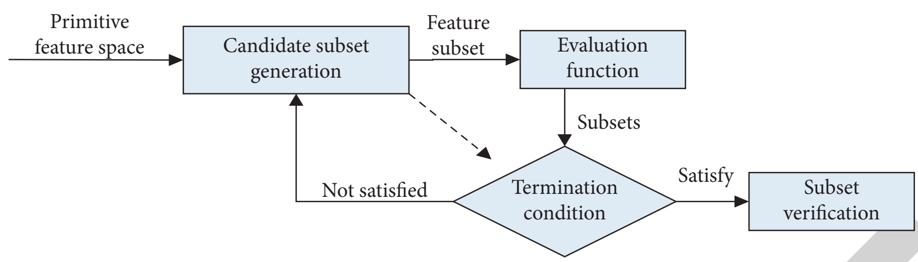

FIGURE 2: Feature selection process.

final performance of the established learning model. A good feature subset can significantly improve the efficiency of the learning algorithm and the performance of the model and, to some extent, improve the generalization ability of the classification model to avoid overfitting.

The feature selection process consists of four steps: candidate subset generation, evaluation function, termination condition, and subset verification, as shown in Figure 2.

2.2. Prediction of Discontinuous Medium Theory of Mining Subsidence

(1) The most effective noncontinuous medium theory for mining subsidence calculation is the stochastic medium theory, which has developed into a probability integral method in China. After nearly half a century of research on mining subsidence workers in China, it has become one of the most widely used mining subsidence calculation methods in China.

(2) The rock mass is regarded as a loose medium which is composed of elliptical medium points and has compression and shear resistance, while the shear resistance is relatively small. When solving the problem of double-layer rock mass, the subsidence section equation of gently inclined coal seam mining is obtained.

(3) The rock mass is assumed to be a discontinuous medium divided by the original joints and cracks caused by mining damage, which can be described by a fragment model; the rock mass is anisotropic, heterogeneous, discontinuous "superposition principle." Based on the analytical method of stochastic medium theory, the Weibull distribution method for the calculation of surface movement deformation is obtained, which can describe the asymmetric mobile basin.

(4) The study of fuzzy probability theory applied to mining subsidence. The problem of ground subsidence caused by mining, i.e., mining subsidence, is a comprehensive manifestation of various geological mining factors, and each factor affects the mining subsidence of the surface in different degrees and ways. Therefore, the problem of surface subsidence caused by underground mining can be regarded as a fuzzy event, which is solved by fuzzy probability theory in fuzzy mathematics.

(5) Noncontinuous deformation analysis based on numerical manifold method. Limited coverage techniques are often used in manifold analysis and are rarely used in numerical calculations. This method is the first to use modern mathematics-manifold in numerical analysis. It uses a manifold covering technique to establish a new unified computing method including finite element, discontinuous deformation analysis method and analytical method. The far-reaching development prospects and application value are known as the new generation of methods in the twenty-first century, which has attracted the attention and welcome of the theoretical and engineering circles. This method can solve the problem of continuous deformation of mining subsidence. However, the relevant literature on applying this method to the field of mining subsidence has not been seen yet. This paper uses the technical route shown in Figure 3 to carry out research work.

2.3. Introduction to the Probability Integration Method. The probability integration method is named for the probability integrals contained in the motion and deformation prediction formulas used. Since the basis of this method is stochastic medium theory, it is also called stochastic medium theory. The theory of stochastic medium uses discontinuous medium as the medium mechanics of granular medium to study the problem of rock stratum and surface movement. It is believed that the law of rock stratum and surface movement caused by mining is similar to the law described by granular medium model of random medium. The granular medium of the random medium is made up of medium particles such as sand grains or relatively small rock blocks. The particles are completely out of contact and can move relative to each other. The movement of the granular medium is characterized by the random movement of the particles and the movement of a large number of granular media is considered a random process. Therefore, from a statistical point of view, if the whole mining process is divided into infinite and infinitesimal unit mining, the influence of the whole mining face on the rock and surface can be regarded as the sum of the influence of infinite and infinitesimal unit mining on the rock and surface.

2.3.1. Calculation Formula for Semi-Infinite Mining Movement Deformation. The so-called semi-infinite mining is

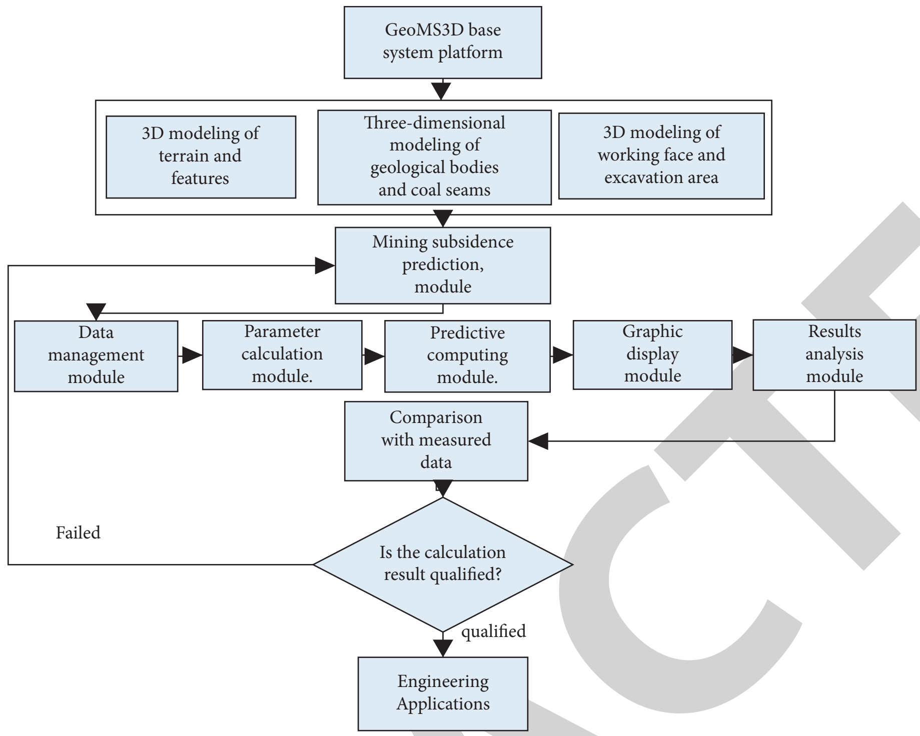

Figure 3: Technical route.

when the mining width increases to a considerable extent and then continues to increase the mining width has little effect on the movement and deformation of the surface above the stop line. It can be considered that the coal seam on one side of the stop line is not disturbed, while the other side the coal seams are all mined out.

When the coal seam is horizontal or gently inclined coal seam (coal inclination angle $\alpha \leq 15^{\circ}$), the calculation formula of the moving deformation value of the main section is

$$\begin{cases} W(x) = \dfrac{W_0}{2}\left[erf\left(\dfrac{\sqrt{\pi}}{r}x\right)+1\right] \\ i(x) = \dfrac{dW(x)}{dx} = \dfrac{W_0}{r}e^{-\pi\frac{x^2}{r^2}} \\ k(x) = \dfrac{d^2W(x)}{d^2x} = 2\pi\dfrac{W_0}{r^2}\left(-\dfrac{x}{r}\right)e^{-\pi\frac{x^2}{r^2}} \\ U(x) = b\cdot r\cdot i(x) = bW_0e^{-\pi\frac{x^2}{r^2}} \\ \varepsilon(x) = b\cdot r\cdot k(x) = 2\pi\, d\dfrac{W_0}{r}\left(-\dfrac{x}{r}\right)e^{-\pi\frac{x^2}{r^2}} \end{cases}, \quad (1)$$

where x is the abscissa of the point on the main section; the origin of the coordinate is directly above the boundary of the working plane (i.e., the coal wall); the x-axis points to the gob along the surface; $\mathrm{erf}(x) = 2/\sqrt{\pi}\int_0^x e^{-\lambda x}d\lambda$ is the probability integral function; r is the influence radius (m), which can be calculated by the formula $r = H_0/\tan\beta$, where H_0 is the average mining depth of the coal seam (m) and $\tan\beta$ is the main influence angle tangent; W_0 is the maximum sinking value (mm); and b is the horizontal movement coefficient.

The calculation formula of the movement deformation along the main section of the inclination is the same as the formula of the strike, but it is necessary to replace x in the formula with y and replace r with r1 and r2. The origin of the coordinate system is set at the surface projection of the lower mountain boundary after considering the offset of the inflection point, and the y-axis is along the surface (set to the horizontal ground), pointing to the direction of the mountain. r1 and r2 are the influence radius (m) of the downhill and the uphill, respectively, which can be calculated by $r_1 = H_1/\tan\beta$ and $r_2 = H_2/\tan\beta$. H1 and H2 are the downhill and uphill depth (m), respectively, and θ is the main influence propagation angle.

When the coal seam is a medium inclined coal seam ($15^{\circ} < \alpha < 55^{\circ}$), the calculation formula of the trending displacement deformation is the same as that of the horizontal coal seam. The calculation formula of the sinking, inclination, and curvature of the main section is also the same as the horizontal coal seam tendency formula. The horizontal movement and horizontal deformation are calculated by

$$u_y = W_0\left(be^{-x\left(\frac{y}{r_{1,2}}\right)} \pm \cot\theta\cdot\frac{1}{\pi}\int_{-\sqrt{\pi}}^{\infty}\frac{y}{r_{1,2}}e^{-\lambda^2}d\lambda\right), \quad (2)$$

$$\varepsilon_y = -\frac{W_0}{r_{1,2}} e^{-\pi\left(\frac{y}{r_{1,2}}\right)^2}\left(2\pi b\frac{y}{r_{1,2}} \pm \cot\theta\right). \tag{3}$$

2.3.2. Calculation Formula for Limited Mining Movement Deformation. When the size of the goaf is small, the mining work cannot fully exploit the surface. This mining situation is called limited mining. The calculation of the movement deformation at this time is regarded as the difference between the two semi-infinite mining. The calculation formula for the movement deformation of the main section of the horizontal or gently inclined coal seam is

$$\begin{cases} W_0(x) = \dfrac{W_0}{2}\left\{\left[erf\left(\dfrac{\sqrt{\pi}}{r}x\right)+1\right]-\left[1+erf\left(\dfrac{\sqrt{\pi}}{r}(x-1)\right)\right]\right\} = W(x)-W(x-1), \\ i^0(x) = \dfrac{dW_0(x)}{dx} = \dfrac{W_0}{r}\left[e^{-\pi x^2/r^2} - e^{-\pi(x-1)^2/r^2}\right] = i(x)-i(x-1), \\ k^0(x) = \dfrac{d^2W_0(x)}{d^3x} = 2\pi\dfrac{W_0}{r^2}\left[\left(-\dfrac{x}{r}\right)e^{-\pi x^2/r^2} - \dfrac{x-1}{r} - e^{-\pi(x-1)^2/r^2}\right] = k(x)-k(x-1), \\ U_0(x) = bri(x) = bW_0\left[e^{-\pi x^2/r^2} - e^{-\pi(x-1)^2/r^2}\right] = U(x)-U(x-1), \\ \varepsilon^0(x) = brk(x) = 2\pi\, d\dfrac{W_0}{r}\left[\left(\dfrac{x}{r}\right)e^{-\pi x^2/r^2} - \dfrac{x-1}{r} - e^{-\pi(x-1)^2/r^2}\right] = \varepsilon(x)-\varepsilon(x-1), \end{cases} \tag{4}$$

where l is the calculated boundary length (m) and L is the tendency to calculate the boundary length (m), which can be calculated by the following formula:

$$\begin{cases} l = l_0 - S_{01} - S_{02}, \\ L = \dfrac{(L_0 - S_1 - S_2)\cdot \sin(180-\theta-\alpha)}{\sin\theta}. \end{cases} \tag{5}$$

In the formula, S01, S02, S1, and S2 are the offset points (m) of the inflection point, the stop line and the downhill and the uphill, respectively. The calculation formula of the main section moving deformation value of the horizontal or gently inclined coal seam is the same as that of the strike. It is only necessary to replace x in the formula with y, and r is replaced by r1 and r2, respectively. The calculation formula of the moving deformation value of the main section of the inclined coal seam is the same as the calculation formula of the horizontal coal seam. The formula for sinking, tilting, and curvature of the main section moving deformation value is the same as that of the horizontal coal seam, and the horizontal and horizontal deformations are calculated by

$$u_y = W_0\left(b\left(e^{-\pi(y/r_1)} - e^{-\pi(y-l/r_2)}\right)\right) + \cot\theta\left(\frac{1}{\sqrt{\pi}}\int_{-\sqrt{\pi}}^{\infty}\frac{y}{\frac{1}{r}}e^{-\lambda^2}d\lambda - \frac{1}{\sqrt{\pi}}\int_{-\sqrt{\pi}}^{\infty}\frac{y-l}{r_2}e^{-\lambda}d\lambda\right), \tag{6}$$

$$\varepsilon_y = W_o\left(-2\pi b\left(\frac{y}{r_1}e^{-\pi(y/r_1)^2} - \frac{y-l}{r_2}e^{-\pi(y-l/r_2)^2}\right) + \cot\theta\left(\frac{1}{r_1}e^{-\pi(y/r_1)^2} - \frac{1}{r_2}e^{-\pi(y-l/r_2)^2}\right)\right). \tag{7}$$

Under the limited mining conditions, it should be noted that when the trend is full mining and the tendency is not fully mining, it is necessary to calculate the propulsion coefficient C_y and then multiply the C_y by the moving deformation value, i.e., (4), multiply with C_y; when the trend is not fully mining, the tendency is to fully exploit and

Table 1: East measured and expected sinking value (mm).

Point name	19	20	21	22	23	24	25	26	27	28
Predictive value	1322	1419	1503	1567	1566	1507	1527	1478	1077	998
Measured value	882	973	1239	1549	1567	1536	1501	1327	1159	1056

calculate the mining coefficient C_x of the strike and C_x is multiplied by the trending displacement deformation value; when the trend and the tendency are all insufficiently mining. At the same time, C_y and C_x are calculated and then multiplied by the trending and trending deformation values, respectively, and the calculated result is finally the final value of the moving deformation value of the finite mining main section.

The mining coefficient C_x can be calculated by the sinking value calculation formula in (4), where x in the formula is replaced by l/2 and then the calculated W_x^0 is divided by W_0. The C_y calculation method of C_x is the same, y is replaced by $(L/2 - S_1)\sin(\theta + \alpha)/\sin\theta$ and substituted into (4), and the value of the propensity coefficient C_y is obtained.

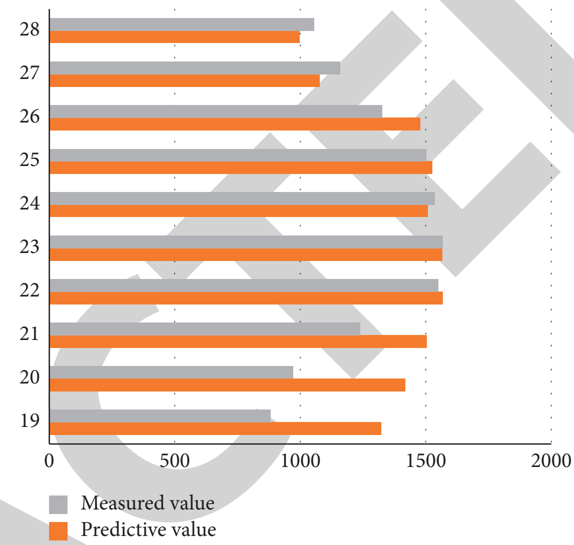

Figure 4: Contrast curve between predicted and measured values.

3. Experiments

3.1. Experimental Design. Three typical feature selection algorithms were selected as simulation experiments in the simulation experiments, namely, mRMR, ReliefF, and IG. The mRMR algorithm is the basic algorithm supported by the proposed CoFS algorithm. mRMR algorithm is mainly to solve the problem that the best m features obtained by maximizing the correlation measurement between features and target variables do not necessarily obtain the best prediction accuracy, because these m features have redundant features (meaning that the information contained in this feature can be deduced from other features). Therefore, compared with the original mRMR feature selection algorithm, the effectiveness of the proposed algorithm is better and the mRMR algorithm is widely praised by the industry for information feature-based feature selection algorithm, with a high representation. The ReliefF algorithm is an excellent feature selection algorithm based on Euclidean distance. Euclidean distance refers to the true distance between two points in m-dimensional space or the natural length of the vector (that is, the distance from the point to the origin). According to Robnik-Sikonja and Kononenko, the parameters Neighbors and Instances are set to 5 and 30, respectively. Information Gain (IG) is also a well-known feature selection algorithm based on information entropy. The value of the constraint factor ω in the CoFS algorithm is set to 3. To fully validate the validity of the proposed feature selection algorithm, the simulation experiment used 15 test data sets from different UCI machine learning repositories. These datasets are often used to compare the performance of learning algorithms or feature selection algorithms in the field of machine learning and data mining. From the diversity of these datasets, they can verify the performance of the feature selection algorithm under different conditions to some extent. In this experiment, the experimental platform uses the well-known machine learning integration software Weka and the relevant parameters of each learning algorithm are set to Weka's default values.

3.2. Experimental Data Collection. In order to obtain more reliable classification performance, the simulation experiment used 10 times of 10-fold cross-validation. The 10-fold cross-validation is to divide the dataset into 10 parts and take 9 of them as training data and 1 part as test data in turn. Each simulation experiment will yield the corresponding classification accuracy, and the average of the 10 results will be used as an estimate of the accuracy of the algorithm. In order to better analyze the performance of the feature selection algorithm, we performed a statistical significance test on the experimental results to describe whether the selected feature subsets of multiple feature selection algorithms have significant performance gaps under the same classifier. The statistical significance test method used is the t-test, which indicates that the performance of a feature selection algorithm is significantly higher (or lower) than that of other algorithms; if and only if the corresponding statistical P value is less than 0.05, that is, the confidence is greater than 95%.

4. Discussion

4.1. Analysis of Sinking Data in Coal Mining. From the display data of Table 1 and the curve comparison of Figure 4, the following conclusions can be found: (1) The maximum subsidence value of the measured data in this paper is at the 23rd point, with a value of 1567 mm, and the maximum subsidence value of the predicted data is also at the 23rd

Table 2: East measured and predicted tilt value (mm· m^{-1}).

Point name	19	20	21	22	23	24	25	26	27	28
Predictive value	5.63	4.72	3.21	1.05	-1.89	-3.59	-5.9	-7.23	-7.21	-7.02
Measured value	8.14	6.82	7.07	0.23	0.13	-0.17	-2.16	-5.39	-6.53	-7.22

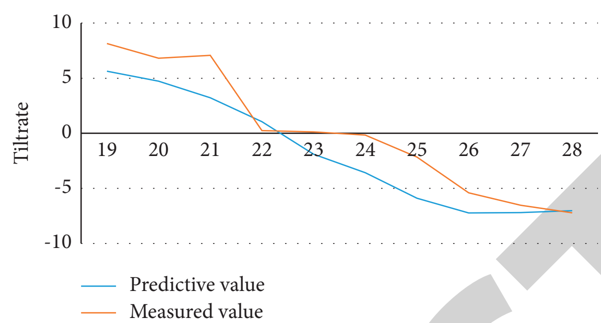

Figure 5: Tilt value comparison graph.

Table 3: East measured and predicted curvature value (mm m^{-2}).

Point name	19	20	21	22	23	24	25	26	27	28
Predictive value	-0.11	-0.02	-0.08	-0.08	-0.08	-0.07	0.01	0.08	0.01	0.02
Measured value	0.07	-0.04	0.01	0.01	-0.01	-0.06	-0.06	-0.11	-0.05	0.03

point, with a value of 1566 mm. The difference between the two is very small, and there is only one maximum subsidence point, which indicates that the 1176 east working face is not completely submerged. (2) The predicted value and the measured value are completely consistent, and the sinking value is from the edge position of the working surface to the middle position, the position gradually increases, and it conforms to the law of the change of the sinking value of the surface movement. (3) The difference between the predicted data and the measured data is relatively large between the 18 and 23 points. The maximum difference is 367 mm at 18th point, the phase error is 49%, and the point with the smallest prediction error is at point 23, only 1 mm.

4.2. Analysis of Coal Mining Slope Data. From the data of Table 2 and the curve comparison, Figure 5 can lead to the following conclusions: (1) The measured slope values are 22, 23, with positive and negative maxima at points 19 and 28, respectively For 8.15 mm m^{-1} and -7.94 mm m^{-1}, the positive and negative maximum values of the predicted value data are 6.47 mm m^{-1} at the 18th point and -7.18 mm·m^{-1} at the 27th point. (2) Because the measured data is lost, the curve is not smooth and the prediction curve is smooth, but the predicted value and the measured value are also consistent. From the edge position to the working surface, the tilt value is gradually increased to reach the positive maximum. After the value is gradually reduced, the tilt value is also reduced to 0 in the gob area, then becomes a negative value, then reaches a negative maximum value, and then gradually decreases to 0. (3) Tilt value prediction result and measured value are obtained. There is a gap in the point, but on the other hand, it can be seen in the layout of the east observation point that the 23rd point is in the middle of the goaf, the position is in the direction of the downhill, and the prediction curve in the curve comparison chart is inclined. The point where the value reaches 0 is just between the 22 and 23 points, indicating that the sinking basin is in the middle of the 23rd and is also downhill offset value is consistent with the variation of the inclination of the inclined face.

4.3. Analysis of Curvature Data of Coal Mining. Table 3 data and curve comparison in Figure 6 can get the following conclusions: (1) The curvature prediction value and the measured data are quite different. (2) The measured curvature change is very unstable, the curve has many break points, the prediction curve is smooth, and the wave is presented. (3) Theoretically, the curvature should reach a negative maximum at the point where the sinking value is the largest, while the measured curve in the figure is obviously not reached; the prediction curve is consistent with the theory, and from the overall trend of the prediction curve, it also conforms to the trend of the theoretical curve of the probability integration method.

4.4. Overall Analysis of Coal Mining. As shown in Figure 7, the integral value comparison curve calculated by the approximate calculation algorithm can be seen from the figure. For the operation efficiency test, a working surface with a trend of 500 m and a tendency of 180 m is divided according to the distance of 20 m × 20 m. The entire working face has a total of 2700 points. If there are multiple working faces, the grid is divided according to the distance of 20 m × 20 m, and

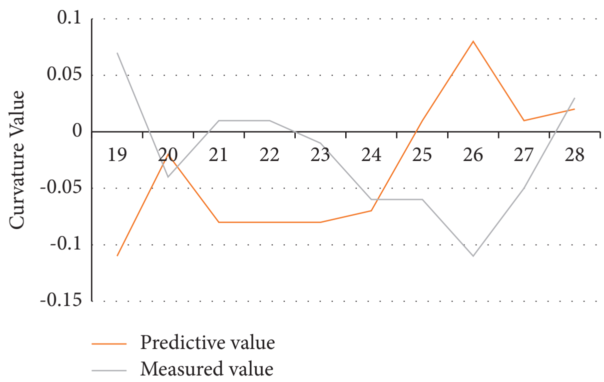

FIGURE 6: Curvature value comparison curve.

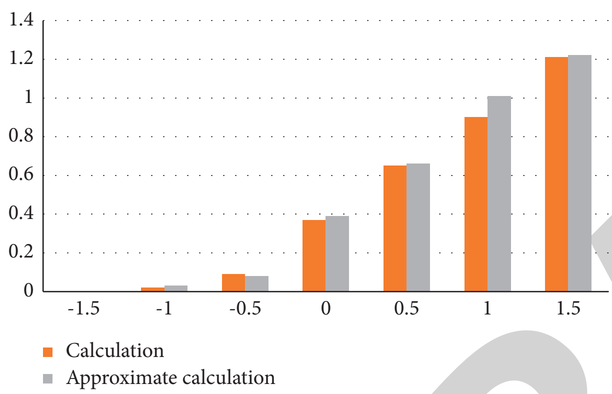

FIGURE 7: Contrast curve with approximate calculated integral value.

the timing is 20 s. If it is 30 m × 30 m, there are 4554 points, and the timing is 8 s. It can be seen from the data that there are some errors in the predicted values and the measured values. It is considered that there are two possible reasons: one is the deficiency of the probability integration method itself. It can be seen from the data table and the curve comparison chart. There are obvious differences between the measured displacement deformation values of some points and the theoretical values. The changes of the values are irregular and abnormal. This may be because the geological structure of the working area is complex, causing severe local movement and deformation, and the probabilistic integration method is not sufficient for the prediction of subsidence under complex geological conditions. The predicted value and the measured value are quite different. The other is that the predicted parameters are inaccurate. The predicted movement deformation curve is consistent with the measured curve in the overall trend, but there is a difference in the value. It may be that the primary selection parameter is not accurate enough. In addition, the influence of the observation error of the actual data is not excluded.

5. Conclusion

Based on a large amount of literature review, this paper studies the mining subsidence and draws the following conclusions: The probability integral method is introduced. The prediction of coal mining subsidence by probability integration method has the characteristics of easy computer programming, simple calculation, and easy access to parameters. It can well meet the practical engineering application. Some mining parameters are obtained from the 3D geological body model and working face model of the coal mine. The method is simple and practical, and the calculation is accurate and fast. Through database technology, the mining and settlement of various data are managed, the security of data is increased, and data sharing can also be realized. Using the approximate calculation method of the probability integral formula, the integral value can be obtained quickly and accurately, and the efficiency of moving deformation calculation can be improved. The generated moving deformation contours are superimposed with the topographic map of the mining area, which facilitates the user's analysis of the moving deformation data. Three typical feature selection algorithms are selected for simulation experiments.

The research results in this paper focus on the following two points. (1) The combination of mining subsidence prediction and three-dimensional model. The GeoMS3D platform provides a three-dimensional model of the mining area. Some parameters of the mining subsidence prediction are automatically acquired from the 3D model, and the model data is fully mined. The prediction results are displayed in the GeoMS3D platform. (2) Complete mining subsidence prediction module. A set of data acquisition, prediction calculation, impact assessment, result analysis, and output data processing flow is designed. The module is simple and convenient to use and can meet the basic requirements of coal mine for mining subsidence prediction.

Through the research in this paper, the authors have done some useful practices and discoveries. However, due to the complexity of the mining subsidence itself and the limited research level of the authors themselves, there are still many problems that have not been solved and further research is needed: mining subsidence prediction for the rectangular working face has a good predictive effect. For the working face of any shape, although the module is also considered, the calculation result is not completely satisfactory and it needs to be improved. The dynamic prediction is insufficient, and the surface movement deformation is the degree of advancement of the working surface changes in real time. Therefore, its time prediction function is very complicated. Although some research work has been done on dynamic prediction in the process of developing the mining subsidence prediction module, the effect is not very satisfactory and needs further improvement.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Z. Chang, J. Wang, M. Chen, Z. Ao, and Q. Yao, "A novel ground surface subsidence prediction model for sub-critical mining in the geological condition of a thick alluvium layer," *Frontiers of Earth Science*, vol. 9, no. 2, pp. 330–341, 2015.

[2] X. Zhu, G. Guo, J. Zha, T. Chen, Q. Fang, and X. Yang, "Surface dynamic subsidence prediction model of solid backfill mining," *Environmental Earth Sciences*, vol. 75, no. 12, p. 1007, 2016.

[3] P. Polanin, "Application of two parameter groups of the Knothe-Budryk theory in subsidence prediction," *Journal of Sustainable Mining*, vol. 14, no. 2, pp. 67–75, 2015.

[4] B. Zhang, L. Zhang, H. Yang, Z. Zhang, and J. Tao, "Subsidence prediction and susceptibility zonation for collapse above goaf with thick alluvial cover: a case study of the Yongcheng coalfield, Henan Province, China," *Bulletin of Engineering Geology and the Environment*, vol. 75, no. 3, pp. 1117–1132, 2016.

[5] Q. Yi, P. Sun, S. Niu, and Y. Kim, "Potential for sediment phosphorus release in coal mine subsidence lakes in China: perspectives from fractionation of phosphorous, iron and aluminum," *Biogeochemistry*, vol. 126, no. 3, pp. 315–327, 2015.

[6] D. Xuan, J. Xu, B. Wang, and H. Teng, "Borehole investigation of the effectiveness of grout injection technology on coal mine subsidence control," *Rock Mechanics and Rock Engineering*, vol. 48, no. 6, pp. 2435–2445, 2015.

[7] Q. Guo, G. Guo, L. Xin, W. Zhang, and S. Qin, "Research on surface subsidence prediction of dense solid backfilling mining based on the coupling of continuous and discrete medium," *Journal of Central South University*, vol. 48, no. 9, pp. 2491–2497, 2017.

[8] T. Hao, J. Xu, D. Xuan, and B. Wang, "Surface subsidence characteristics of grout injection into overburden: case study of Yuandian No. 2 coalmine, China," *Environmental Earth Sciences*, vol. 75, no. 6, pp. 1–11, 2016.

[9] I. A. Wright, B. Mccarthy, N. Belmer, and P. Price, "Subsidence from an underground coal mine and mine wastewater discharge causing water pollution and degradation of aquatic ecosystems," *Water, Air, & Soil Pollution*, vol. 226, no. 10, p. 348, 2015.

[10] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, "A review of relational machine learning for knowledge graphs," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 11–33, 2015.

[11] X.-m. Guo, T.-q. Zhao, W.-k. Chang, C.-y. Xiao, and Y.-x. He, "Evaluating the effect of coal mining subsidence on the agricultural soil quality using principal component analysis," *Chilean Journal of Agricultural Research*, vol. 78, no. 2, pp. 173–182, 2018.

[12] A. Holzinger, "Interactive machine learning for health informatics: when do we need the human-in-the-loop?" *Brain Informatics*, vol. 3, no. 2, pp. 119–131, 2016.

[13] N. Lei, H. Wang, and X. Yan, "Application of the arctangent function model in the prediction of ground mining subsidence deformation: a case study from Fushun City, Liaoning Province, China," *Bulletin of Engineering Geology and the Environment*, vol. 76, no. 4, pp. 1–16, 2016.

[14] R. Malhotra, "A systematic review of machine learning techniques for software fault prediction," *Applied Soft Computing*, vol. 27, no. C, pp. 504–518, 2015.

[15] Y. Sun, X. Zhang, W. Mao, and L. Xu, "Mechanism and stability evaluation of goaf ground subsidence in the third mining area in Gong Changling District, China," *Arabian Journal of Geosciences*, vol. 8, no. 2, pp. 639–646, 2015.

[16] Z. Julian, T. Marcin, B. Adam, M. Maciej, B. Subhadip, and P. Dariusz, "Multi-level machine learning prediction of protein-protein interactions in *Saccharomyces cerevisiae*," *PeerJ*, vol. 3, no. 1, e1041 pages, 2015.

[17] J. Suh, Y. Choi, and H. D. Park, "GIS-based evaluation of mining-induced subsidence susceptibility considering 3D multiple mine drifts and estimated mined panels," *Environmental Earth Sciences*, vol. 75, no. 10, pp. 1–19, 2016.

[18] G. Li, N. Cao, P. Zhu et al., "Towards smart transportation system," *Journal of Organizational and End User Computing*, vol. 33, no. 3, pp. 35–49, 2021, https://doi.org/10.4018/JOEUC.20210501.oa3.

[19] Q. Jia, R. Xie, T. Huang, J. Liu, and Y. Liu, "Caching resource sharing for network slicing in 5G core network," *Journal of Organizational and End User Computing*, vol. 31, no. 4, pp. 1–18, 2019, https://doi.org/10.4018/JOEUC.2019100101.

[20] X. Yang, H. Li, L. Ni, and T. Li, "Application of artificial intelligence in precision marketing," *Journal of Organizational and End User Computing*, vol. 33, no. 4, pp. 209–219, 2021, https://doi.org/10.4018/JOEUC.20210701.oa10.

[21] A. Singh, B. Ganapathysubramanian, A. K. Singh, and S. Sarkar, "Machine learning for high-throughput stress phenotyping in plants," *Trends in Plant Science*, vol. 21, no. 2, pp. 110–124, 2016.

[22] X. Diao, K. Wu, D. Zhou, and L. Li, "Integrating the probability integral method for subsidence prediction and differential synthetic aperture radar interferometry for monitoring mining subsidence in Fengfeng, China," *Journal of Applied Remote Sensing*, vol. 10, no. 1, 016028 pages, 2016.

[23] A. Giusti, J. Guzzi, C. C. Dan, H. Fang-Lin, P. R. Juan, and F. Flavio, "A machine learning approach to visual perception of forest trails for mobile robots," *IEEE Robotics and Automation Letters*, vol. 1, no. 2, 1 page, 2017.