Preprints.org

Review

Application Status and Development Trends of Machine Learning in Large - Scale Mineral Resource Prediction

ZeKang Fu , XiaoJun Zheng [*] , Yan Yong Feng , Xiaofei Xu , FanChao Zhou , Xiao Li , QuanTong Zhou , WeiKun Mai

Review

Application Status and Development Trends of Machine Learning in Large - Scale Mineral Resource Prediction

ZeKang Fu [1], XiaoJun Zheng [1,2,*], YongFeng Yan [1], XiaoFei Xu [1], FanChao Zhou [3], Xiao Li [1,4], QuanTong Zhou [1,5] and WeiKun Mai [1]

[1] School of Faculty of Resource Engineering, Kunming University of Science and Technology

[2] Innovation Base for Tin-Polymetallic Ore Mineralization Research and Exploration Technology, Geological Society of China

[3] Tianjin North China Geological Exploration General Institute, Tianjin

[4] Shandong Gold Group,Shandong

[5] Jiangxi Provincial Geological Bureau, No.10 Geological Team, Jiangxi

[*] Correspondence: zh5101126@163.com

Abstract: Mineral resource exploration is crucial for national economic security and sustainable development. Conventional mineral prediction methods struggle with low efficiency, high costs, and insufficient accuracy in predicting concealed deep - seated deposits and those in complex tectonic areas. Recently, machine learning algorithms, with powerful data - mining and pattern - recognition capabilities, have greatly enhanced mineral prediction intelligence. This paper comprehensively sums up the research progress and development trend of machine learning in large - scale mineral resource prediction, covering algorithm frameworks like random forests, support vector machines, and deep learning, as well as data - preprocessing techniques such as feature engineering and multi - source data fusion. For instance, the random forest algorithm achieved an 82% target - area match rate in Tibet's Julong porphyry Cu - Mo deposit prediction, cutting exploration costs by about 40%; in Fujian's southwestern Makeng iron deposit study, comparing four algorithms including random forests and support vector machines, the optimal model got an AUC of 0.88. However, issues like uneven data quality and poor model interpretability remain to be solved. In the future, combining generative adversarial networks (GANs) to boost data diversity and developing a "geological constraint + AI - driven" interdisciplinary paradigm are needed to promote the intelligent and precise transformation of mineral resource exploration.

Keywords: machine learning; mineral prediction; prospecting model

1. Introduction

Against the backdrop of the rapid evolution of information technology, machine learning has emerged as a pivotal research focus across a broad spectrum of fields. Over the past few years, it has demonstrated remarkable application potential across various industries, delivering significant advancements in areas such as data mining, image recognition, and natural language processing. These achievements have not only driven technological innovation within these fields but also provided novel approaches and solutions for tackling complex real - world problems. As the theoretical foundations of machine learning continue to be refined and computational capabilities keep improving, its scope of application is expanding continuously. An increasing number of traditional industries are beginning to adopt machine - learning algorithms to enhance business processes and decision - making efficiency.

In the field of geological exploration, mineral prospecting has always been a vital and challenging task. Traditional prospecting methods mainly depend on the experience and professional

knowledge of geological engineers. They infer the potential distribution of mineral resources by analyzing geological structures, rock characteristics, and geophysical and geochemical data. Nevertheless, these methods have certain limitations, such as being subjective and having limited ability to handle complex data. With the continuous accumulation of geological data and the growing degree of informatization, how to effectively utilize these massive amounts of data for precise mineral prospecting has become an urgent issue in geological exploration. By integrating multi - source data from geology, geophysics, and geochemistry, and constructing multiple machine - learning prediction models, the accuracy of target - area delineation can be significantly improved.

This study centers on the application of machine - learning algorithms in mineral prospecting. In recent years, although some research has attempted to introduce machine learning into the field of geology, there are still deficiencies in algorithm selection, model construction, and comprehensive data utilization. Therefore, based on previous studies, this paper systematically reviews the research progress of machine - learning algorithms in large - scale mineral resource prediction and summarizes their key roles in feature information extraction, multi - source data fusion, and model construction. By examining domestic case studies such as the Makeng Iron Deposit in southwestern Fujian and the Julong Copper and Molybdenum Deposit in Tibet, this paper compares the application effects and limitations of algorithms such as random forests, support vector machines, and deep learning. It aims to provide theoretical support and practical references for the intelligent development of mineral resource prediction.

2. Typical Method Classification and Data Processing Workflow

2.1. Common Machine Learning Algorithms

2.1.1. Random Forest Algorithm

The random forest algorithm, proposed by Breiman, [1]is an ensemble learning method that combines the advantages of the Bagging framework and decision tree models. It generates k heterogeneous subsets from the original training set using a resampling strategy with replacement, and constructs decision trees independently based on each subset to form a "forest" (Figure 1). In classification tasks, the final category is determined by a majority voting mechanism, while in regression tasks, the result is obtained through a weighted average of the predictions from each subtree. Notably, the Kang et al. team[2] innovatively applied this algorithm to uranium exploration, building a prediction model for uranium anomaly identification and lithological classification. The verification results showed that this model significantly outperformed traditional exploration methods in terms of accuracy in characterizing mineralized layer features. Guo et al.[3] further expanded the application of the algorithm by integrating slope correlation analysis with the random forest algorithm, establishing a coal seam gas content evaluation system with high confidence, which provided a new technical approach for unconventional energy exploration. The algorithm's superiority is attributed to its unique dual innovation mechanism: the ensemble learning framework effectively avoids overfitting and underfitting risks while enhancing model robustness. Meanwhile, the parallel training strategy of Bagging significantly improves computational efficiency and generalization ability by reducing feature correlation.

Studies have shown that the mechanism of constructing diverse base models and fusing results not only ensures prediction stability but also endows the algorithm with an advantage in handling complex nonlinear problems. This may be a key reason for its continued status as a core tool in the field of machine learning. The algorithm aggregates predictions from individual decision trees and determines the final prediction through majority voting or averaging. It demonstrates excellent prediction accuracy and model stability when dealing with complex feature spaces.[4] The construction process involves two layers of randomness: at the data level, the Bootstrap sampling technique is used to select approximately two-thirds of the samples from the original dataset to build a single decision tree. The remaining unsampled "out-of-bag samples" automatically form a natural

validation set to monitor model overfitting in real-time. At the feature level, during node splitting in each decision tree, a random subset of features is selected for optimal split point calculation. This dual-randomness design significantly enhances the model's generalization ability. It should be noted that the Gini index (Gini impurity) is used as the splitting criterion during node splitting evaluation. This index primarily reflects the purity of the node's category, with lower values indicating higher homogeneity of samples. This integrated strategy not only effectively alleviates the problem of dimensionality but also suppresses model overfitting by reducing the correlation between individual decision trees.[5]
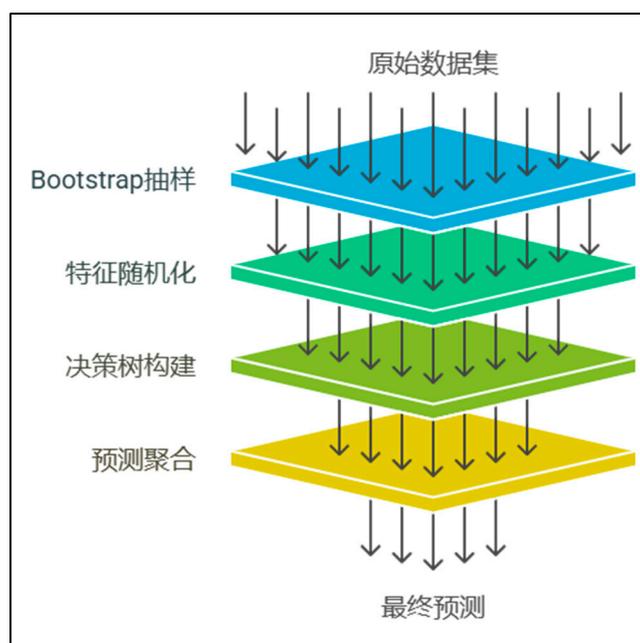

Figure 1. Random Forest flowchart.

2.1.2. Support Vector Machine (SVM) Algorithm

The Support Vector Machine (SVM) algorithm, created by Vapnik,[6] combines the principle of structural risk minimization with the framework of statistical learning theory. It employs kernel tricks to realize nonlinear transformations of feature spaces and uses convex optimization strategies in high-dimensional spaces to address challenges in pattern recognition and regression prediction. This approach is built on the Vapnik-Chervonenkis (VC) dimension theory and establishes a supervised learning model. Its core mechanism involves constructing a maximum-margin hyperplane to optimize the decision function. Under conditions of limited samples, it simultaneously controls the empirical risk and the confidence interval, thereby achieving a global optimal solution and avoiding the local convergence issues common in traditional BP networks (Figure 2).The algorithm identifies key support vectors and constructs a maximum-margin hyperplane. This ensures the classification accuracy of linearly separable datasets while enhancing model performance. Its unique advantages are reflected in three aspects: first, it uses kernel mapping techniques to handle nonlinearly separable data; second, it controls structural risks to avoid overfitting; and third, it is particularly suitable for balancing learning accuracy and model complexity in scenarios with small sample sizes.Compared to conventional machine learning methods, this model effectively addresses the challenges of high-dimensional data processing and global convergence while maintaining generalization performance. However, it is important to note that SVMs are sensitive to outliers. Experimental results indicate that this method, grounded in statistical learning theory, has demonstrated significant advantages in practical applications such as outlier detection and mineral prospecting.
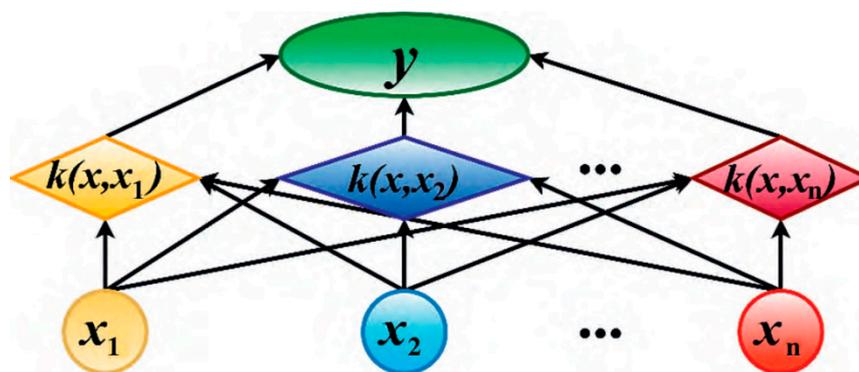
Figure 2. Schematic diagram of support vector machine structure（According to the literature[7]）.

2.1.3. Principles of Deep Neural Networks

Deep neural networks (DNNs) construct complex models by layering multiple nonlinear processing units, progressively extracting features with semantic hierarchy from raw data. As a key branch of machine learning, deep learning focuses on hierarchical feature learning. Its main advantage is autonomous feature discovery through multi - level abstraction, excelling in complex areas like voice interaction, high - dimensional data analysis, and image semantic segmentation.[7,8] Originating from advanced ANNs, Hinton's team's deep belief networks and layer - wise training algorithms solved critical parameter - optimization problems in deep networks.[9]

Unlike traditional methods relying on manual feature engineering, deep - learning frameworks autonomously build data - distribution - specific representation systems, offering unique advantages in high - dimensional nonlinear - relationship modeling. In geophysical exploration, studies by Wang et al.[10] have shown significant progress in electromagnetic and seismic - wave analysis, indicating that intelligent geophysical detection technology is about to enter a new stage.

With a large number of neurons, DNNs dynamically adjust connection weights during training, forming a strong - association structure similar to biological neural systems. This self - organizing feature inherits the advantages of traditional neural networks and achieves higher - order cognitive functions through hierarchical information processing.[11] By mimicking the human brain, DNNs build multi - layer models for implicit feature mining and adaptive optimization.(Figure 3) In deep learning, network parameters are globally optimized via gradient backpropagation, extracting essential data features layer by layer to form high - level semantic representations for accurate classification and prediction.[12] Unlike shallow models, DNNs use stacked nonlinear activations to approximate highly complex nonlinear functions. Their multi - level abstraction can interpret complex data patterns. Recent empirical studies have confirmed their excellence: Zhao's team[13] used convolutional neural networks for Wenchuan aftershock analysis, proving their value in real - time seismic monitoring; Wang's group[14] suppressed seismic multiples under complex marine geological conditions by training DNNs with data - augmentation strategies.
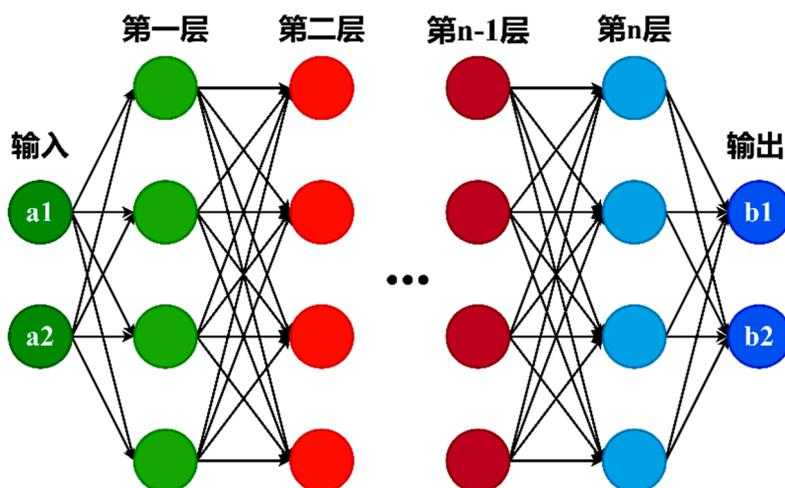
Figure 3. Deep neural network structure diagram（According to the literature[7]）.

2.2. Data Processing Workflow of Machine Learning in Mineral Prediction

In geological exploration, the performance of machine - learning models is highly dependent on the quality of input data and the sufficiency of information expression. When dealing with complex and diverse geological data, such as the abundance of geochemical elements, geophysical field characteristics, remote - sensing images, and structural - analysis data, data preprocessing and feature engineering have become crucial for building efficient mineral - prediction models. This chapter systematically reviews the data - processing workflow of machine learning in large - scale mineral prediction. By integrating cutting - edge research cases, it explores the technical principles and practical challenges of each step.

2.2.1. Data Cleaning and Standardization

Geological data is often contaminated with outliers and missing values due to sampling errors, instrumental noise, or human - induced recording mistakes, which can lead to prediction biases when directly input into models. Current mainstream methods combine statistical diagnosis with robust processing, including outlier detection and missing - value imputation. For outlier detection, methods such as the box - plot approach (Tukey, 1977) or the improved Local Outlier Factor (LOF) algorithm (Breunig et al., 2000) are used to identify non - geologically caused outliers. For example, Wang et al. (2022) employed a dynamically adjusted 3σ rule in porphyry copper deposit prediction in Tibet, effectively eliminating geochemical data distortions caused by weathering. In terms of missing - value imputation, for sparse data, K - Nearest Neighbors (KNN) imputation and Random Forest regression imputation (MissForest, Stekhoven & Bühlmann, 2012) have proven superior to traditional mean imputation. Zhao et al. (2021) highlighted through comparative experiments that MissForest performs better in preserving the spatial correlation of geochemical elements.

Data standardization enhances model convergence efficiency by eliminating dimensional differences. In addition to the classic Z - Score standardization, "quantile transformation" is increasingly applied to non - normally distributed data, such as geophysical field intensity data. Li et al. (2023) found in tungsten mineral prediction in South China that quantile transformation can increase the AUC value of Support Vector Machine (SVM) models by about 5%.

2.2.2. Feature Selection and Extraction

High - dimensional geological data often contain redundant features and weakly correlated variables. Feature engineering is needed to screen for indicators closely related to mineralization.

2.2.3. Supervised Feature Selection

Recursive Feature Elimination (RFE): Combines model weights to dynamically eliminate redundant variables. Zhang et al. (2021) used random forest - RFE in iron deposit prediction in southwestern Fujian to identify Fe, Cu element combinations and fracture density as core predictors, achieving an F1 - score of 0.82.

SHAP - based Interpretive Screening: SHapley Additive exPlanations (SHAP, Lundberg & Lee, 2017) quantifies feature contributions using game theory. Chen et al. (2022) found through SHAP analysis that the combined anomaly of Au - As - Sb elements contributed over 60% to gold deposit prediction in Jiaodong.

2.2.4. Unsupervised Feature Extraction

Principal Component Analysis (PCA): Effective for reducing dimensionality and separating background from anomalies in geochemical data. For instance, Liu et al. (2020) applied PCA to 1:200,000 geochemical data from the eastern Tianshan region in Xinjiang. The first three principal components accounted for 89% of the variance, enhancing porphyry copper deposit anomaly identification.

Autoencoder: A deep - learning method for extracting non - linear features. Tang et al. (2023) built a stacked autoencoder to compress and reconstruct remote - sensing data of granite - type uranium deposits in South China. This reduced feature dimensions by 70% while improving model recall by 12%.

2.2.5. Multi - Source Data Fusion

Mineral resource prediction requires the integration of multi - source geoscientific information, including geophysics, geochemistry, remote sensing, and geology. The integration methods can be divided into three levels:data - level, feature - level, and decision - level.In data - level fusion, spatial correction and data standardization are used to integrate raw data. A typical example is the integration of magnetotelluric sounding data (with a point spacing of 500 m) and WorldView - 3 hyperspectral remote - sensing data (with a resolution of 1.2 m) in a 3D geological modeling platform, involving coordinate transformation and gridding (Liu et al., 2022)[15].In feature - level fusion, multi - dimensional features are extracted via deep - learning methods to construct integrated indicators. For instance, in iron mineral prediction in Western Australia, Zhou et al. (2021)[16] fused airborne magnetic vertical gradient data, geochemical element ratio features, and lithologic texture features using a convolutional neural network, increasing the recall rate to 86%.In decision - level fusion, the Bayesian averaging method is used to integrate inference results from multiple models. Tang et al. (2023)[17] integrated the classification probabilities of a geochemical XGBoost model and a geophysical neural - network model for gold mineral prediction in the Central Asian Orogenic Belt, improving the accuracy of drill - verified target areas by 22%.

## 3. Typical Applications of Machine Learning in Large-Scale Mineral Prediction

3.1. Research Progress in Mineral Prospecting Prediction

Many research results have been obtained in mineral prospecting using machine learning. Zhang et al. [18] explored the theoretical methods of mineral prediction, summarized the application of machine learning in feature information extraction and integration in mineral prediction, and illustrated the machine - learning - based mineral prediction process using the iron and polymetallic mineral deposit in southwestern Fujian as an example. They also pointed out the difficulties in the quantitative prediction of mineral resources, such as the scarcity and imbalance of training samples and the lack of uncertainty assessment in model training. Xiang et al. [19] conducted 3D quantitative mineral resource prediction based on machine learning using the copper deposit in Sichuan as an example. By establishing a 3D geological model and using the random forest algorithm, they

demonstrated the higher prediction accuracy and stability of this algorithm in defining prospective areas. Dong et al. [20] developed a 2D manganese resource prediction classification model based on the random forest algorithm for the sedimentary manganese deposit in the Xiaolei - Tuhu area. By incorporating class weight parameters to automatically balance positive and negative samples, they defined prospective areas and showed that the model has good generalization ability. Han et al. [21] introduced the basic principles and classification characteristics of advanced machine - learning methods commonly used in geophysics. Their analysis of practical results indicated that machine learning can achieve better results in geophysical exploration. Chen et al. [22] introduced the random forest algorithm into 3D mineral prediction and used the Jiaodong Dàyinzhāng Gold Deposit as a case study. They predicted 3D prospective targets and demonstrated the significant potential of big data technology in mineral resource positioning and prediction. Bi et al. [23] combined kinetic simulation with machine learning for 3D quantitative prediction of the Anhui copper mine. The results showed that the prediction effect was optimal when multiple factors were integrated, and they indicated the prospecting potential in the southeastern deep part of the mining area. Zuo [24] presented research on mineral prediction and evaluation based on machine learning as well as the principles and applications of deep learning. He believed that deep learning can effectively identify and extract geochemical anomalies but requires further research on its integration with geological constraints. Wang et al. [25] explored the feasibility and technical route of integrating knowledge - graph technology with quantitative mineral resource prediction. They aimed to enhance the automation and intelligence of the prediction process by incorporating knowledge expression and reasoning from knowledge graphs. Sun [26] pointed out that prospecting methods include data mining and machine learning. The combined use of these methods is expected to improve the accuracy of tungsten mineral prospecting in Gansu Province.

3.2. Case Study of Mineral Resource Estimation in Makeng Iron Deposit, Southwestern Fujian

The Makeng Iron Deposit, due to its controversial ore - forming mechanism and the difficulty of traditional methods in locating deep - seated ore bodies, has been a focus for machine - learning - based predictions. A research framework integrating "ore - forming system, exploration system, and prediction - evaluation system" was established. Five types of predictive elements, including the contact zones of Yanshanian plutons and geochemical anomalies, were extracted. Four algorithms, namely random forests, support vector machines, were compared, and the algorithm with the optimal model (AUC of 0.88) was selected. The study delineated three high - probability target areas, and drilling validation confirmed an additional 120 million tons of resources.
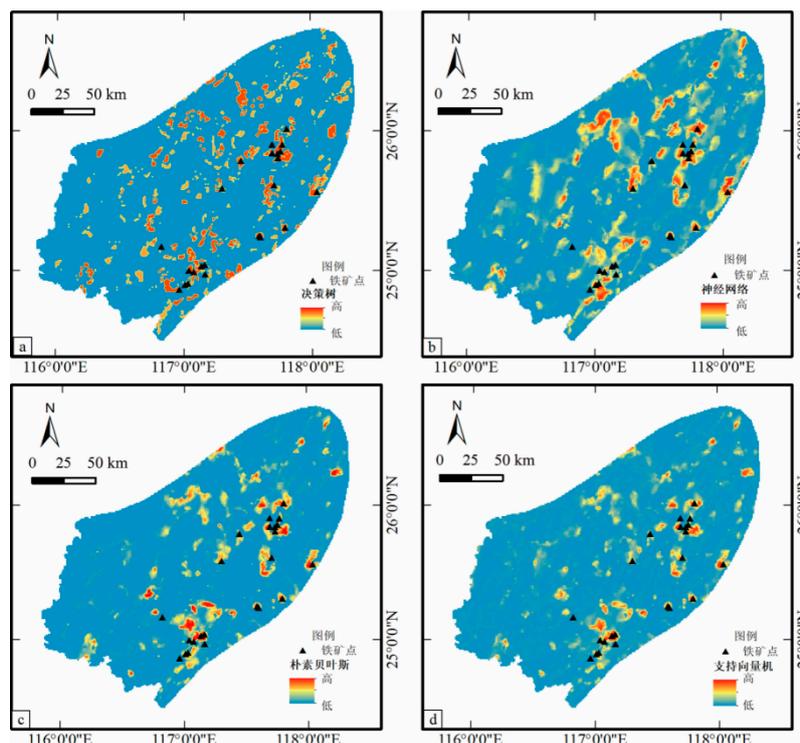
Figure 4. Prediction results from four prediction models for the Makeng-type iron deposits in southwestern Fujian(According to the literature[18]).

In algorithm selection for mineral prediction, different machine - learning algorithms have unique principles and features suitable for various scenarios (Table 1). Random forests, integrating multiple decision trees, are robust to overfitting and support feature - importance evaluation, making them ideal for classification tasks and high - dimensional data. Support vector machines (SVMs) use kernel - function mapping to handle nonlinear problems. They perform well with small samples but are noise - sensitive, so they're suited for small - sample classification and geochemical - anomaly identification. Deep - learning algorithms leverage multi - layer neural networks for automatic complex - feature extraction but require substantial training data and computational power. They excel in multi - source heterogeneous - data fusion and 3D modeling.

Thus, when choosing a machine - learning algorithm, one must consider data characteristics, task requirements, and resource constraints to guarantee model performance and prediction accuracy.

Table 1. Comparison of Different Machine - Learning Algorithms.

Algorithm Type	Principles and Features	applicable scene
Random Forest	Integrates multiple decision trees, strong overfitting resistance; supports feature importance evaluation	Classification tasks, high-dimensional data
Support Vector Machine	Solves nonlinear problems via kernel function mapping; performs well with small samples but sensitive to noise	Small-sample classification, geochemical anomaly identification
Deep Learning	Automatically extracts complex features via multi-layer neural networks; requires large amounts of training data and computational power	Multi-source heterogeneous data fusion, 3D modeling

4. Future Research Directions and Conclusions

4.1. Future Research Directions

4.1.1. Algorithm Innovation and Cross-Modal Integration

Currently, machine - learning applications in mineral - potential prediction rely predominantly on single models, which makes it difficult to fully explore the complex relationships among multi-source and heterogeneous data. Future research should focus on developing multi-algorithm collaborative frameworks. For example, in the area of deep ensemble learning, we can combine the robustness of random forests with the feature self-extraction capabilities of deep learning. This integration can create a hybrid model that performs "shallow feature screening + deep semantic parsing" (such as Deep Forest), thereby improving the accuracy of modeling complex geological processes.

Generative Adversarial Networks (GANs) can be used to drive data enhancement. In response to the scarcity and uneven spatial distribution of geological samples (such as the lack of data on deep-seated mineral deposits), GANs can generate synthetic data that complies with geological rules. This can expand the training dataset and enhance the model's generalization ability (Goodfellow et al., 2014). [27,28] Reinforcement learning can empower intelligent exploration decision-making. By modeling exploration path planning as a Markov decision process, reinforcement learning can dynamically optimize exploration strategies and reduce exploration costs (such as AlphaGo-style autonomous sampling robots).

4.1.2. Big Data Platforms and Cloud-Edge Collaborative Computing

The explosive growth of geological data presents a significant challenge to traditional computing architectures. There is an urgent need to establish an intelligent mineral exploration data platform. Distributed storage and real-time analytics can leverage Hadoop/Spark architectures to achieve distributed storage and parallel computing of PB - level geophysical data. When combined with stream processing technologies, this can enable real-time updates of prediction models (such as AWS geological cloud platforms). Edge computing can optimize resource allocation by enabling localized preprocessing of data collected by edge devices (e.g., drones, portable spectrometers) in field exploration scenarios. Only key features are uploaded to the cloud, which reduces communication overhead (Zhang et al., 2022). [29,30]

4.1.3. Integrated Geological-Machine Learning Modeling

Most current models are "black-box" architectures with vague geological meanings, which limits their engineering applications. We need to promote integrated geological-machine learning modeling. This involves encoding mineral system theories (such as the source-transport-storage-seal model) as model prior constraints. For example, geological rule items (such as fault-controlled mineralization weights) can be incorporated into the loss function. This achieves bidirectional optimization driven by both data and knowledge guidance (Wang et al., 2021). We also need to make breakthroughs in explainability technologies. By introducing methods such as SHAP (Shapley Additive Explanations) value analysis and attention mechanism visualization, we can interpret the decision logic of models. [31] For instance, this can reveal the relationship between Au anomalies and structural intersection zones, thereby enhancing the trust of geologists in the models (Lundberg & Lee, 2017).

4.2. Conclusions and Challenges

4.2.1. Summary of Research Progress

Machine learning has significantly revolutionized the paradigm of mineral resource prediction. For instance, the random forest algorithm achieved an 82% target - area match rate in predicting the

Julong porphyry Cu - Mo deposit in Tibet, reducing exploration costs by about 40%. In the Makeng iron deposit study in southwestern Fujian, by comparing four algorithms including random forests and support vector machines, the optimal model had an AUC of 0.88, which confirmed its superiority in high - dimensional nonlinear problems. A comprehensive technical system has been formed, ranging from data cleaning (Z - Score normalization), feature fusion (PCA - GIS collaboration) to multi - source modeling. The engineering value of machine learning is highlighted in cases like 3D prediction of the Jiaodong gold deposit and deep - positioning prediction of the Makeng iron deposit.

4.2.2. Existing Challenges and Countermeasures

At present, the bottlenecks in applying machine learning algorithms to production mainly focus on three aspects. First, regarding data quality and sample bias, the incomplete geochemical data in some regions restricts the model's extrapolation ability. To address this, it is necessary to establish a national geological data sharing platform and promote federated learning for cross - regional collaborative modeling. Second, on the issue of model interpretability, most deep - learning models lack geological semantic associations. It is suggested that prediction results should be required to include uncertainty quantification reports, such as Monte Carlo Dropout. Third, there is a shortage of interdisciplinary talent. Geologists and data scientists currently collaborate loosely. Therefore, interdisciplinary funds should be established to train composite teams who are proficient in geology, skilled in algorithms, and knowledgeable in engineering.

4.2.3. Outlook on Future Development Prospects

Based on the analysis of research results and problems, the future development prospects of machine learning in mineral exploration are promising. With continuous technological progress and deepening interdisciplinary cooperation, machine learning is expected to play a more significant role in mineral exploration. It will become a driving force for technological innovation in mineral resource exploration and provide strong support for the sustainable development and utilization of mineral resources.

Author Contributions: Study conception and design: Zekang Fu and Xiaojun Zheng; introduction and literature review: Zekang Fu and Xiao Li; data collection: YongFeng Yan and XiaoFei Xu; modeling: FanChao Zhou and QuanTong Zhou; analysis and interpretation of results: WeiKun Mai; conclusions: Zekang Fu; project management and funding acquisition: Xiaojun Zheng; draft manuscript preparation: Xiaojun Zheng. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study are available upon reasonable request from the corresponding author. However, the authors do not wish to make the data publicly available at this time.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ML	Machine Learning
AUC	Area Under The Curve
ROC	Receiver operating characteristic
GANs	Generative Adversarial Networks
SVM	Support Vector Machine
DNNs	Deep neural networks
RFE	Recursive Feature Elimination
PCA	Principal Component Analysis

References

1. XIAO Hang, ZHANG Zhansong, GUO Jianhong, et al. Coal structure identification method based on random forest combined with geophysical logging data and its application [J]. Science, Technology and Engineering, 2021, 21(24): 10174 - 10180 (in Chinese).

2. KANG Qiankun. Application of random forest algorithm in sandstone type uranium mineral anomaly identification based on logging data [D]. Changchun: Jilin University, 2020.

3. GUO Jianhong, ZHANG Zhansong, ZHANG Chaomo, et al. Case study on coalbed methane content prediction based on geophysical logging data using slope correlation - degree - random forest method [J]. Exploration Geophysics and Geochemistry, 2021, 45(1): 18 - 28.

4. FAWAGREH K, GABER M M, ELYAN E. Random forests: from early developments to recent advancements [J]. Systems Science & Control Engineering, 2014, 2(1): 602 - 609.

5. JAMES G, WITTEN D, HASTIE T, et al. An introduction to statistical learning: with applications in R [M/OL]. New York: Springer US, 2021.

6. VLADIMIR N V. The nature of statistical learning theory [M]. New York: Springer, 1999.

7. HAN Shili, XIAO Jian, LIU Wei. Research progress on application of machine learning in uranium resource exploration [J]. Uranium Geology, 2021, 37(2): 123 - 130 (in Chinese).

8. ZEN H, SENIOR A, SCHUSTER M. Statistical parametric speech synthesis using deep neural networks [C]//2013 Institute of Electrical and Electronics Engineers international conference on acoustics, speech and signal processing. Vancouver: Institute of Electrical and Electronics Engineers, 2013: 7962 - 7966.

9. HINTON G E. A practical guide to training restricted Boltzmann machines [J]. Momentum, 2010, 9(1): 926.

10. WANG Hao, YAN Jiayong, FU Guangming, et al. Current status and application prospect of deep learning in geophysics [J]. Progress in Geophysics, 2020, 35(2): 642 - 655 (in Chinese).

11. WANG Jingyi, WANG Zhiguo, CHEN Yumin, et al. Deep artificial neural network in seismic inversion [J]. Progress in Geophysics, 2023, 38(1): 298 - 320 (in Chinese).

12. LUO Haibo, TANG Yandong. "Deep learning and its applications" [J]. Infrared and Laser Engineering, 2018, 47(7): 8 - 9 (in Chinese).

13. ZHAO Ming, CHEN Shi, DAVE Y. Waveform classification and seismic recognition by convolutional neural network [J]. Chinese Journal of Geophysics, 2019, 62(1): 374 - 382 (in Chinese).

14. WANG Kunxi, HU Tianyue, LIU Xiaozhou, et al. Suppressing seismic multiples based on the deep neural network method with data augmentation training [J]. Chinese Journal of Geophysics, 2021, 64(11): 4196 - 4214 (in Chinese).

15. LIU, G., et al. (2022). Integration of MT and hyperspectral data for 3D mineral exploration modeling. Remote Sensing of Environment, 271, 112893.

16. ZHOU, Y., et al. (2021). Deep feature fusion for iron ore prospectivity mapping in the Pilbara Craton. Ore Geology Reviews, 134, 104187.

17. TANG, C., et al. (2023). Bayesian model averaging for gold targeting in Central Asian Orogenic Belt. Natural Resources Research, 32(1), 321 - 337.

18. ZHANG Zhenjie, CHENG Qiuming, YANG Jie, et al. Machine learning and mineral prediction: a case study of Fe - polymetallic mineral prediction in southwestern Fujian [J]. Earth Science Frontiers, 2021, 28(03): 221 - 235. DOI:10.13745/j.esf.sf.2021.1.4.

19. XIANG Jie, CHEN Jianping, XIAO Keyan, et al. Three - dimensional quantitative mineral prediction based on machine learning: a case study of the Lala copper deposit in Sichuan [J]. Geological Bulletin of China, 2019, 38(12): 2010 - 2021.

20. DONG Jianhui, LIU Huan, JIANG Sha, et al. Two - dimensional mineral prediction based on random forest: a case study of the Shilei - Tuhu area sedimentary manganese deposit [J]. Mineral Deposits, 2025, 44(01): 143 - 158. DOI:10.16111/j.0258 - 7106.2025.01.009.

21. [1]Bektemyssova G ,Bykov A ,Moldagulova A , et al. Analysis of Spatial Aggregation and Activity of the Urban Population of Almaty Based on Cluster Analysis [J]. Sustainability, 2025, 17 (7): 3243-3243.

22. Zhou R ,Gao Q ,Wang Q , et al. Machine Learning Optimization of Waste Salt Pyrolysis: Predicting Organic Pollutant Removal and Mass Loss [J]. Sustainability, 2025, 17 (7): 3216-3216.

23. DAHL G E, YU D, DENG L, et al. Context - dependent pre - trained deep neural networks for large - vocabulary speech recognition [J]. Audio, Speech, and Language Processing, IEEE Transactions on, 2012, 20(1): 30 - 42.

24. CHEN Jin, MAO Xiancheng, LIU Zhanjun, et al. Three - dimensional mineral prediction of the Dayinziguan gold deposit based on the random forest algorithm [J]. Tectonophysics and Metallogeny, 2020, 44(02): 231 - 241. DOI:10.16539/j.tecto - metal.2020.02.007.

25. BI Chengxi, LIU Liangming, ZHOU Feihu. Three - dimensional mineral prediction based on machine learning combined with kinematic simulation: a case study of the Tongshan copper deposit in Anhui [J]. Tectonophysics and Metallogeny, 2025, 49(01): 103 - 116. DOI:10.16539/j.tecto - metal.2023.01.104.

26. ZUO Ren - guang. Deep - seated mineralization information mining and integration based on deep learning [J]. Bulletin of Mineralogy, Petrology and Geochemistry, 2019, 38(01): 53 - 60 + 203. DOI:10.19658/j.issn.1007 - 2802.2019.38.026.

27. LIU Guan. Application of machine learning in flight delay prediction [J]. Journal of Civil Aviation, 2023, 7(06): 1 - 4.

28. Leonardi S ,Distefano N ,Gruden C . Advancing Sustainable Mobility: Artificial Intelligence Approaches for Autonomous Vehicle Trajectories in Roundabouts [J]. Sustainability, 2025, 17 (7): 2988-2988.

29. WO Yi - feng. Application of machine learning in electricity spot trading price prediction [J]. Electronic Technology, 2024, 53(12): 238 - 239.

30. Zhou Y ,Lv Y ,Dong J , et al. Factors Influencing Transparency in Urban Landscape Water Bodies in Taiyuan City Based on Machine Learning Approaches [J]. Sustainability, 2025, 17 (7): 3126-3126.

31. Pendar R M ,Cândido S ,Páscoa C J , et al. Enhancing Automotive Paint Curing Process Efficiency: Integration of Computational Fluid Dynamics and Variational Auto-Encoder Techniques [J]. Sustainability, 2025, 17 (7): 3091-3091.