

Review

Overview on the Development of Intelligent Methods for Mineral Resource Prediction under the Background of Geological Big Data

Shi Li ¹, Jianping Chen ^{2,3,*} and Chang Liu ^{2,3}

¹ Department of Big Data, School of Information, Beijing Wuzi University, Beijing 101149, China; lishi@bwu.edu.cn

² Department of Remote Sensing and Geo-Information, School of Earth Sciences and Resources, China University of Geosciences, Beijing 100083, China; 3001190108@cugb.edu.cn

³ Land Resources Information Development and Research Key Laboratory of Beijing, Beijing 100083, China

* Correspondence: 3s@cugb.edu.cn

Abstract: In the age of big data, the prediction and evaluation of geological mineral resources have gradually entered a new stage, intelligent prospecting. This review briefly summarizes the research development of textual data mining and spatial data mining. It is considered that the current research on mineral resource prediction has integrated logical reasoning, theoretical models, computational simulations, and other scientific research models, and has gradually advanced toward a new model. This type of new model has tried to mine unknown and effective knowledge from big data by intelligent analysis methods. However, many challenges have come forward, including four aspects: (i) discovery of prospecting big data based on geological knowledge system; (ii) construction of the conceptual prospecting model by intelligent text mining; (iii) mineral prediction by intelligent spatial big data mining; (iv) sharing and visualization of the mineral prediction data. By extending the geological analysis in the process of prospecting prediction to the logical rules associated with expert knowledge points, the theory and methods of intelligent mineral prediction were preliminarily established based on geological big data. The core of the theory is to promote the flow, invocation, circulation, and optimization of the three key factors of “knowledge”, “model”, and “data”, and to preliminarily constitute the prototype of intelligent linkage mechanisms. It could be divided into four parts: intelligent datamation, intelligent informatization, intelligent knowledgeization, and intelligent servitization.

Keywords: the fourth paradigm; geological big data; prospecting prediction; intelligent algorithm



Citation: Li, S.; Chen, J.; Liu, C. Overview on the Development of Intelligent Methods for Mineral Resource Prediction under the Background of Geological Big Data. *Minerals* **2022**, *12*, 616. <https://doi.org/10.3390/min12050616>

Academic Editors: Xiancheng Mao, Chengbin Wang, Zhankun Liu and Yosoon Choi

Received: 20 March 2022

Accepted: 8 May 2022

Published: 12 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 global informatization entering an advanced stage, the big data era is being gradually embraced. With the rapid development of emerging technologies, such as cloud computing and artificial intelligence, big data technologies and applications have seen breakthroughs in diverse fields. Recently, journals, such as *Nature* and *Science*, have successively published special issues on big data, putting forward relevant concepts including “data is gold” and “data is science”. McKinsey, a globally renowned consulting company, has stated the following: “Data has entered every industry and business scope, becoming an essential productive factor. The discovery, mining, and application of big data will contribute to new productivity growth and technological innovation” [1]. Besides the 4V characteristics of big data [2], geological big data have the following features: direction, space–time, diversity, polymorphism, causality, multisource, variability, uniqueness, combination, heterogeneity, nonlinearity, sampling, randomness, fuzziness, scarcity, association, and zonality. As a key component of geoscientific big data, geological big data

contain unstructured, structured, and semistructured information from numerous subdisciplines including stratigraphy, sedimentology, geomorphology, geotectonics, paleontology, mineralogy, mathematical geology, geochronology, cartography, petroleum geology, hydrogeology, and geothermal research. As a cross discipline between geological science and mathematics, mathematical geology can be considered the “data science” aspect of geological science.

In China, the applied research field of mathematical geology is the quantitative prediction and evaluation of mineral resources, with the longest history, the widest aspects, the most participants, and the most prominent achievements [3]. As the material basis for human survival, mineral resources are not only the pillar for the development of an information society, but also the premise for industrialization construction, playing a significant and strategic role in promoting the development of the national economy. With increasing demand for mineral resources, there are increasingly fewer surface-exposed and shallow ores, adding to the challenges of prospecting. Countries around the world have started exploring deep underground ore deposits, the success rate of which depends on the in-depth study of metallogenic regularity and scientific research on the theory and methods of mineral resource prediction and evaluation. Zhao reported that the introduction of big data in the field of geosciences is an inevitable development trend, so as to leap from “mathematical geology” to “digital geology”, and in order to conduct scientific quantitative evaluation and analysis of prospecting prediction [4]. With the sharp growth in the volume and complexity of geological big data, the current new challenge in intelligent prospecting is how to expand the application of intelligent prospecting means from digital data to textual, 2D, and 3D spatial data, based on geological expertise and experience, to ensure the efficiency and precision of prospecting prediction. So far, few scholars have proposed the theory and method of intelligent prospecting prediction that is based on geological big data, and the technological process of intelligent prediction that is based on big data has not yet been established.

Hence, from the perspective of textual and spatial big data mining, this study intended to thoroughly and meticulously analyze problems and solutions of big data intelligent prospecting prediction, to provide certain theoretical and technical support for the development of intelligent prospecting prediction under the background of big data.

2. Development of Geological Big Data

As information technology is constantly changing in the big data age, the application of geological big data and artificial intelligence has become a research hotspot in recent years, which means that the ideas, technologies, and methods of big data in the information age are gradually being successfully applied in the geological field. Currently, geological archives, geological survey database, and geological document databases of various countries have provided multilevel and all-round geological data, including textual and spatial data, and mining models for unstructured, semistructured, and structured data. Geological big data are a type of spatiotemporal big data, which are mainly generated in the investigation, exploration, and corresponding geological scientific research processes in the fields of basic geology, mineral geology, hydrogeology, engineering geology, environmental geology, and hazardous geology; the data also come from the exploitation and utilization of energy and mineral resources, the monitoring, prevention, and control processes of the environment and disasters, and various space-based and air-based ground remote-sensing observation activities. The approaches to obtain data include geophysics, geochemistry, drilling, remote sensing, as well as various applications, such as mapping compiles, analytical calculation, analog simulation, prediction evaluation, and intelligent control, which exist in multiple forms, such as texts, graphics, acoustic images, and specimens [5]. Hence, how to effectively acquire, screen, sort, and excavate effective geological information from massive and complicated big data to better serve geological construction is one of the major research directions of many domestic scholars. With this research topic as the focus, some

achievements have been made in the platform construction and data processing methods for geological big data.

Studies on geological big data strongly depend on the construction of big data platforms. Thus, the organization and implementation of geological and spatiotemporal big data relationships between various topics largely depend on the solutions and overall architecture of big data platforms, the secure storage of big data, indexing, scheduling mechanisms, engine methods and technologies of big data, management function monitoring, as well as the supercomputing of early warning and control data links. Countries around the world are vigorously promoting the construction and development of geological big data platforms. For instance, in 2007, the European Union (EU) released the “Act of Infrastructure for Spatial Information in the European Community”, enabling environmental spatial data to be shared across the EU. Subsequently, The National Science Foundation of the United States officially launched the “Earth Cube” project in 2012, which successfully integrated a large amount of geological data, as well as relevant information and knowledge. Characterized by openness, transparency, and strong comprehensiveness, the “Earth Cube” project substantially improves the flow of knowledge and communication [6]. Peter Baumann et al. from the University of Bremen in Germany successfully developed EarthServer for an earth data analysis engine in 2015, which includes functions such as data access, data analysis, data retrieval, and information extraction, and achieved an open, cross-platform, and easy-to-operate geological data access mechanism [7]. Currently, many countries have established their own geological big-data-sharing platforms (such as Geo Connections in Canada and Geospatial One-Stop in the United States), and have realized effective search, access, and reception of geospatial information, thus assisting government sectors of national, provincial, and other levels in decision-making. In 2012, the Geological Information Technology Laboratory of Ministry of Land and Resources successfully established GeoBDA, a technical experimental platform for studying geological big data technologies [8]. In 2013, Hadoop-distributed infrastructure was gradually upgraded to standard storage architecture for geological big data [9]. Simultaneously, developers employed the MapReduce programming model to significantly enhance the computing efficiency of massive geological datasets through parallel computing [10–12]. In 2014, China’s first independently developed geological big data platform began trial operation at Tibet Geological Survey Academy. Subsequently, China Geological Survey released an intelligent spatial data platform in 2016. In 2018, the Chinese Academy of Sciences launched a strategic scientific research project known as CASEarth, which is committed to supporting major scientific discoveries and macro decisions. In 2019, the Deep-time Digital Earth (DDE), led by Chinese academics Chengshan et al., which was integrated with the ideas provided by 254 scientists from 24 countries, was officially launched. Its five scientific objectives include reconstructing earth science knowledge system in big data, building the largest multisource heterogeneous data system for earth science, creating the world’s leading one-stop processing analysis platform for earth science, solving major scientific problems of earth science driven by big data, and forming a deep communication integration platform of the digital earth, so as to create the first “Geological Google”.

Regarding the intelligent analysis and deep mining methods for geological big data, Chinese researchers have made remarkable achievements in the acquisition, organization and management, preprocessing, analysis, and visualization fields; the most representative ones include geological anomaly extraction based on mining algorithms; research and development into geological big data integration and platform architecture; an intelligent retrieval system and virtual visualization technology for geological big data. When combined with the service demands of practical applications, the geological big data analysis technology has been extensively applied in fields such as basic geological survey [13], exploration and evaluation of oil and gas resources [14,15], intelligent cities [16], land and resource administration [17], ocean exploration (seabed prospecting) [18], remote sensing of big data [19], and mineral exploration and evaluation [20]. In particular, the intelligent analysis algorithm for big data mining has made breakthroughs in diverse fields, such

as geochemical anomaly extraction, gravity and magnetic anomaly extraction, logging lithology identification, spatial prediction for seismic and sedimentary facies, earthquake time prediction, and exploration and evaluation of mineral resources, bringing about a profound impact on the field of geology. Mineral resources are one of the essential material bases supporting the development of the national economy. Therefore, the application of this intelligent analysis algorithm to the exploration and evaluation of mineral resources is an inevitable trend in the development of mathematical geology; this is also the first step in realizing intelligent prospecting. The combined application of the new theory and the new technology is of great significance for breakthroughs in prospecting and in enhancing the assurance degree of mineral resources.

3. Development of the Quantitative Prediction Theory of Big Data

In the big data age, the prediction and evaluation of geological mineral resources have gradually entered a new stage, transforming from digital prospecting to intelligent prospecting. In terms of big data sharing, multiple big data cloud platforms, such as cloud computing, data cloud, and automatic text reading, have been constructed. The corresponding big data analysis methods also provide some help in the quantitative evaluation of mineral resources. Zhao (2013) stressed the importance of research on digital geology and stated that digital geology in the big data age promotes the development of geological prospecting and that focus should be on digital geology and mineral prediction [4]. In addition, geological big data should be studied and utilized based on the characteristics of the geological data and the geological demands [21]. The theory of prospecting, based on geological big data, has also been systematically summarized [22]. Combining the predictive thinking method in the big data age with the prediction and evaluation of mineral resources, and based on the causality theory, trend theory, and the relation of information change theory, Professor Xiao further explored four theories: mineral prediction model theory, mineral prediction correlation theory, trend analysis method, and differentiation theory [23]. In the context of big data science becoming a new scientific paradigm, Yu put forward a new model-driven method for the quantitative prediction and evaluation of mineral resources; it utilizes the geological theory to guide geological data analysis and computer technology for the data mining of geological big data, as well as for the quantitative prediction and evaluation of mineral resources [24]. After systematically summarizing the application of geological big data in the study of metallogenic regularity, metallogenic series, and metallogenic systems, Wang concluded that geological big data can provide new ideas for metallogenic theory and prediction [25]. Professor Chen believed that only by effectively extracting information according to demands, mining useful knowledge, and applying it to government decision-making, management services, and scientific research can future information wealth be obtained; thus, the author proposed the data cycle theory, based on demand main line and data chain main line [2]. Based on the “double-chain” theory and aiming at the quantitative prediction and evaluation of mineral resources, Xiang proposed a big data prediction and evaluation theory based on these “two main lines”, which has been successfully incorporated into the prediction theory [26].

In summary, the essence of big data is massive data. If there is no clear demand or effective mining means, data will remain data. The core idea of big data is not prediction; rather, the process from the known to the unknown can be called the predictive thinking method, which involves the excavating of potential knowledge through data mining methods. If the predictive thinking of big data were to be applied to mineral resource prediction, we not only need to introduce a variety of data analysis algorithms, but also have to establish a perfect and effective theoretical system of big data prediction methods, so as to advance mineral resource prediction toward intellectualization.

4. Development of Intelligent Mining Methods for Big Data

With the wide application of big data, different data mining processes have been summarized in various fields. The most representative one is the Cross-industry Standard

Process for Data Mining (CRISP-DM) process proposed by the European Union, comprising six steps: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. According to Zhou, the process of data mining comprises three stages: data preparation, data mining, and result expression and interpretation [5]. The data mining process should start with actual requirements and end with solutions to problems. Furthermore, the design of specific processes is based on mining requirements. For the demands of prospecting prediction, the data mining process in this study includes demand definition, data preparation, data mining, model evaluation, and knowledge expression.

Data mining is a complex iterative process. For the massive amounts of big data used, the various artificial intelligence algorithms represented by machine learning provide a broad application space for big data mining. Currently, artificial intelligence, machine learning, and deep learning are the three hot words used in current research, which are sometimes used interchangeably. The Alpha Go program of Google Deep Mind beat world Go champions Shishi and Jie, making deep learning an instant focus of global media. However, deep learning did not suddenly appear. There is a close connection between deep learning, artificial intelligence, big data, and machine learning (Figure 1).

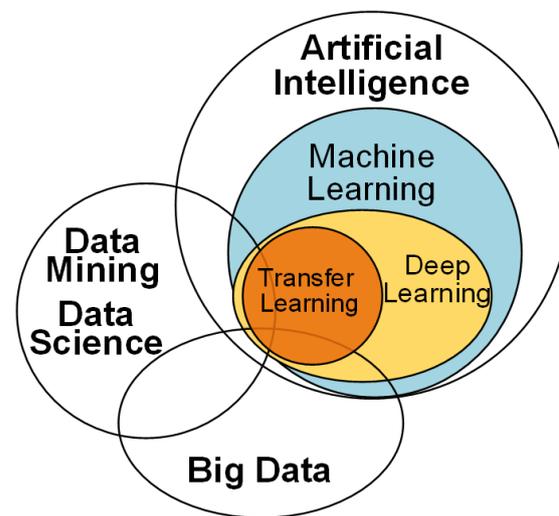


Figure 1. Relationship between deep learning and other related hot topics.

Undoubtedly, the current big data mining in prospecting prediction is far from the level of artificial intelligence. The application of big data mining methods in earth science remains in its infancy. With the increase in research and exploration investment, the massive amounts of heterogeneous geological data in various databases have seen a rapid growth. Conventional analysis methods cannot thoroughly mine the deep correlations that are hidden in various geological data. Through the development of data mining technology, the heterogeneous and spatial geological data can be analyzed. Through spatial, correlation, classification, regression, and cluster analyses of geological data, abnormal extraction, prospecting prediction, and geological background discrimination have been performed, so as to conduct effective production practice activities, such as mineral exploration and prediction [5]. There are numerous means of geological big data mining. Geological big data contain unstructured, semistructured, and structured information, including textual, spatial, multimedia, and other diversified and massive data. In view of the demand for a quantitative prediction of mineral resources, this paper summarizes research progress from two aspects: intelligent mining methods for textual big data and spatial big data.

4.1. Development of Intelligent Mining Methods for Textual Big Data

For geological research in the big data age, it is necessary to comprehensively collect various geological survey data, including semistructured and structured data, as well as unstructured journal literature, technical reports, scientific papers, and geological reports.

As a result of the recent open data initiatives, various government agencies and scientific organizations have started publishing open data online for widespread use [27,28]. Geoscience textual data are a key part of open data. Accumulated over hundreds of years, most of the geological theory and research results are in the form of words and diagrams. The mining of these accumulated data and the extraction of their geological knowledge are of great significance for prospecting. Facing the ever-increasing amount of textual data, how to intelligently classify and organize such geological textual data and understand, analyze, and visualize their semantic associations have become the core problems in transforming geological big data from data to knowledge.

Text mining was first proposed by Feldman in 1995. Table 1 shows the key technical processes of text mining, including dataset acquisition, text preprocessing, data mining, and visualization. For English documents, feature extraction, mining model construction, and mining algorithms are relatively mature techniques, which have been successfully applied in many fields [29–33]. Relevant Chinese researchers have mainly focused on the construction of Chinese language models, Chinese word segmentation and tagging, various algorithm improvements, and professional applications. There is a significant difference between the language models of Chinese and English. Although text mining in Chinese has some theoretical foundations and research results [34–36], many technical difficulties remain. On the one hand, geology majors have high professional requirements for word segmentation. The generalized word segmentation method cannot meet the actual requirements. On the other hand, geological texts have the characteristics of description differences, compilation diversity, and geological expertise. The textual representation of conventional feature vectors cannot intuitively express professional geological semantic information, cannot meet the diverse and grammatical characteristics of geological texts, and cannot dig deep in the prospecting of information in geological texts. The hybrid dictionary-based word segmentation and analysis method proposed by Wang et al., (2018) for geological dictionary segmentation problems can only handle a single geological report [37]. Automated technologies, such as acquisition, preprocessing, mining, and visual analysis of massive geological text data, are difficult to realize.

Table 1. Key technical processes and content of text mining.

Technical Process	Technology/Content
Dataset	Text databases, web pages, documents, e-books, etc.
Text preprocessing	Noise removal, word segmentation, parts-of-speech tagging, stop-word removal, feature representation, feature extraction, etc.
Data mining	Named entity recognition, word frequency analysis, sentiment analysis, automatic summarization, semantic network, similarity analysis, classification, clustering, association rules, intelligent retrieval, regression, trend analysis, etc.
Visualization	Knowledge and conclusions, graphical interface, command line, word cloud, knowledge graph, etc.

Deep learning was proposed by Hinton in 2006, providing a new direction for textual data mining [38–40]. With the development of deep learning, neural network methods have achieved outstanding results in the semantic analysis of natural language processing (NLP), query retrieval, sentence classification modeling, and other conventional methods [41,42]. As one of the most robust deep learning classification algorithms [43], convolutional neural networks (CNNs) have been successfully applied in NLP [44]. Sun completed the multilabel sentiment classification of Weibo through supervised multisentiment classification learning, based on a CNN model [45]. Subsequently, Feng et al., (2018) successfully applied the M-CNN algorithm to multilabel emotion detection [46]. Researchers have made multiple attempts to extract structured information from unstructured networks to build knowledge bases and transform textual information into semantic networks [47,48]. Google’s knowledge graph representation provides a new idea for the visual representation of NLP [49]. A knowledge graph, as an essential semantic network [50], is a data organization form or product that expresses entities, concepts, and their semantic relationships through directed

graphs. They have been widely used in the field of geology [51,52]. Wang successfully applied the knowledge graph to represent key information in geological unstructured text and demonstrated the application potential of NLP and knowledge graph techniques in geoscience research [37]. Morrison et al., (2017) used network analysis methods to represent and mine the information hidden in mine-type, location, and observational data [53]. Wu proposed a construction method for the conceptual prospecting model, based on metallogenetic theory, using geoscience-based textual big data discovery and mining technologies, and verified this method through a typical study area research [54].

For a quantitative prediction of mineral resources, academic literature on the public-domain network and mineral exploration reports on local area networks involves a multitude of key prospecting information, such as prospecting background, prospecting age, mineralization type, and genetic type. The application of text mining technology in prospecting information mining is a development trend, which not only helps extract prospecting information from diverse, massive, heterogeneous geological textual big data, but also provides new ideas and directions for constructing prospecting models.

4.2. Development of Intelligent Mining Algorithms for Spatial Big Data

As global informatization enters an advanced stage, the big data age is imminent [22,55]. Geological data, as the basis in professional fields, are consistent with the characteristics of big data. The main development trends in this field include the combination of big data knowledge discovery technology and deep mining methods for mineral intelligence prediction evaluation, as well as enhancing the effectiveness and accuracy of prospecting prediction [1]. In addition, it is necessary to emphasize the joint application of new theories and methods [23,24,56]. Faced with increasingly difficult prospecting challenges, the quantitative prediction of mineral resources has gradually shifted from easily identifiable, shallow deposits to hidden, deep deposits. The conventional 2D prospecting prediction method cannot meet the demands of mineral exploration. Three-dimensional mineral resource prediction implies the growth of data dimensions and the enhancement of data relevance. In other words, machine learning is being increasingly integrated with mineral resource prediction. The acquired ground features and geological and geochemical information are derived from the spatial information collected from the same objects using different monitoring methods. This series of spatial information also includes prospecting information, underground mineral information, and geological evolution information. Using deep learning methods to mine spatial information in the field of geosciences is of great value for prospecting prediction. With the core of combining mineral deposit theory with intelligent mining methods, this approach can address the developmental challenges in this field.

With the continuous improvement in the prospecting theory in recent years, big data machine learning methods have achieved preliminary success in geoscience information processing, prospecting anomaly information extraction, and comprehensive information prospecting prediction [57]. As an important branch of artificial intelligence, machine learning provides an effective means for processing a large number of evidence feature layers related to prospecting prediction, big data analysis, pattern recognition, and prediction, as well as technical support for constructing intelligent prospecting models based on big data. Since machine learning can extract potential nonlinear relationships between known deposits and feature layers, it has a strong comprehensive prediction capability [58]. Due to the complex prospecting geological conditions, geological data often exhibit nonlinear characteristics. Hence, machine learning can better represent the complex, nonlinear relationship between mineralization and evidence elements [59,60]. Currently, machine learning algorithms that are applied to mineral resource evaluation mainly include artificial neural networks [61], support vector machines [60,62–64], random forest [65–67], Boltzmann machines [68], extreme learning [69], and maximum entropy models [70,71]. In 2017, Zuo and Xiong identified and extracted multivariate geochemical anomalies in Fe polymetallic deposits found in southwestern Fujian Province with the support of machine learning analysis and geological big data [72]. Furthermore, scholars have successfully

applied machine learning to extract abnormal information from geological big data, thus realizing a quantitative prediction of mineral resources [73–80].

However, 3D deep prospecting information has multisource and multimodal characteristics, which pose challenges for its classification and prediction. Therefore, the conventional machine learning algorithms lack the necessary performance [81,82]. It is appropriate and important to use deep learning in prospecting prediction, which is also a meaningful exploration experiment for the application of big data intelligent algorithms in geological research [83–85]. Deep learning is a special machine learning method, which includes CNNs [45,86], recurrent neural networks (RNNs) [87], stacked automatic coding [39,88], deep networks with constrained Boltzmann machines at the core (deep belief networks and deep Boltzmann machines) [89], as well as multilayer feedback recurrent neural networks and full CNNs [90]. Deep learning has also made breakthroughs in logging lithology identification, seismic fault identification [91], and seismic time prediction [92], bringing a profound impact to the geoscience community. Through the combination of mineral deposit prospecting theory and deep learning method, the problem of prospecting prediction can be solved [93]. Currently, deep learning algorithms have achieved good application effects in 2D prospecting anomaly extraction, such as in the prospecting of geochemical anomaly extraction [72,94–96], the prospecting of gravity and magnetic anomaly extraction [97], and comprehensive prospecting prediction [98–102], particularly through using the CNN algorithm.

However, a 2D intelligent mining algorithm can only extract surface features; it ignores the 3D spatial features when prospecting geological bodies located underground. With the quantitative prediction of mineral resources gradually entering the era of 3D prediction [103–108], existing prospecting prediction studies have the following problems: (1) the emphasis is on the quantitative extraction of anomalies; the importance of spatially distributed features, and the correlation between different prospecting conditions is ignored. (2) In the past, the utilization rate of boreholes depended on whether ore was found. In addition, the weight-of-evidence method only operates on positive examples. How do we make reasonable use of negative samples? (3) The previous approach involved focusing on important ore-controlling factors, while ignoring other seemingly unimportant factors, even in conventional machine learning algorithms. However, this approach is not rigorous. Therefore, it is necessary to set up an intelligent prospecting system on the basis of both geological background and big data, so as to improve the intelligence of prospecting prediction. In addition to extracting the relationship between spatial positions in an image as done by 2D CNNs, 3D convolutional neural networks (3D-CNN) can additionally extract the relevant time-domain information in continuous images [109], with wide usage in medical diagnosis [110–112], sign language action recognition [113,114], facial expression recognition [115], brain signal recognition [116], as well as many other fields [117–122]. For prospecting prediction, the spatial information in the data from the fields of geology, geochemistry, and geophysical research obtained by different monitoring methods contains many ore-forming anomalies. Therefore, the application of the 3D-CNN in underground space feature recognition, underground anomaly extraction, and comprehensive prediction is of great significance.

How to extract effective information from such multiple, massive, heterogeneous information for the analysis of prospecting geological conditions? How to extract time-series features and spatial correlation based on a 3D geological model? How to improve the intelligence of the forecasting process? How to realize the combination of geological expertise and intelligent algorithms? These are the challenges in the intelligent mining of spatial prospecting data. To address asymmetric data information in prospecting prediction, the application prospects of transfer learning are also worth considering.

5. Challenges in Intelligent Prospecting Prediction

Many scholars have conducted useful explorations in this field. However, the current data mining in prospecting prediction has remained in the stage of digital “artificial

prospecting". There is a certain gap in the realization of intelligent prospecting. With the advent of the big data age, there is no big-data-prospecting prediction theory guided by the knowledge of geological experts, with "geological line", "technical line", and "data line" as the main line, and the promotion of prospecting prediction intelligence as the demand. Therefore, a series of intelligent analysis technologies is required to support big data discovery, big data update storage, big data feature analysis, and comprehensive quantitative prediction, thereby promoting the flow and optimization of three core elements of "data", "model", and "knowledge". As for the transformation from the prospecting of data and informatization to intelligence, the main challenges are as follows.

5.1. Discovering of Big Data for Prospecting Prediction Based on Knowledge Systems

For data mining in prospecting prediction, mineral expert knowledge bases and mineral databases are the bases for research. With the rapid development of information technology, China has accumulated a large amount of data and experience in the field of mineral exploration. To fully discover the underlying data and in order to perform classification, storage, statistics, and mining, there is an urgent need to realize the integration of new technologies and new methods. Effective big data discovery methods can provide not only basic data for data mining in study areas, but also solutions for constructing large mineral databases. For new data discovery technologies, four major difficulties should be overcome: regionalism, comprehensiveness, professionalism, and multilingualism. Regarding regionalism, how to obtain effective data from 233 countries and regions around the world? Regarding comprehensiveness, the conventional database query method with string matching, such as using keywords and subject words, cannot accurately and comprehensively query the massive mineral-related data. How to obtain relevant data more comprehensively and systematically according to the demand? Regarding professionalism, how to accurately obtain the target field data list from the various academic conference websites, geological collection institutions, scientific research institutions, search engines, large publishers, geoscience journals, geoscience publishers, and e-book publishers, and then realize real-time updates? Regarding multilingualism, how to break the language barriers to maximize the integration between "big science" and "big data", as well as in the sharing of cross-border mineral data? The current application of the big data discovery technology in the field of prospecting prediction lacks real solutions and research models. There is no theoretical system or technical method for big data discovery that truly integrates expert knowledge; this limits its application in the inter-disciplinary fields of earth science, mathematics, and information science to a certain extent.

5.2. Construction of a Conceptual Prospecting Model Based on Text Mining

The construction of a conceptual model for regional prospecting is the key link in the quantitative prediction of mineral resources. The conceptual prospecting model refers to the textual representation of a mineral deposit model based on various geological prospecting theories with ore-controlling characteristics, signs, and data combinations (excluding genesis and assumptions), including deposit models, regional prospecting models, and local prospecting models, which can be used to guide mineral resource prediction. The conceptual model for regional prospecting mentioned in this paper is for a specific study area. Prospecting prediction requires a comprehensive consideration of various factors, such as expert opinions, geological background, and prospecting type, through a certain set of rules. Among them, the most influential factor of prospecting prediction is the experience and knowledge of the prospector. The research achievements made by mineral experts mainly correspond to large amounts of unstructured textual big data, such as technical reports, exploration reports, and news websites. Textual big data contain important knowledge and experience. In the face of various NLP analysis methods, the following problems need to be solved: how to apply such technology in order to intelligently classify, store, and extract useful prospecting information for massive textual data, how to realize an in-depth analysis, understanding, and visualization of

the semantic correlations in geological text, and how to construct a conceptual model for regional prospecting.

Methods that can extract important information, such as the prospecting age, prospecting background, genetic type, and mineralization type, not only involve simple text processing, but also mineral experts' knowledge and effective NLP means. How to integrate the predictive thinking of big data and mining methods guided by the knowledge system into the quantitative prediction of mineral resources? How to use previous knowledge to optimize the intelligent mining model? How to construct a regional conceptual prospecting model of the study area and realize the automatic expansion of the deposit model base? Addressing these questions is the only way to advance intelligent prospecting prediction. However, the current method system requires further improvement.

5.3. Realizing Intelligent Mining Technology Based on Spatial Big Data

With the advent of the big data age, the evaluation of mineral resource prediction is gradually shifting from digital prospecting to intelligent prospecting. Despite the endless number of theoretical methods for big data intelligent prospecting, most of them have not realized technical innovations. Currently, intelligent prospecting based on big data remains in its infancy. How to integrate the idea of big data intelligent prospecting into the demand line of intelligent prospecting? How to construct a prediction theory and technical framework that promotes the process of intelligent prospecting? How to expand the connotation of geological big data from numbers into unstructured, semistructured, and structured data, such as geoscience textual data, 2D image data, and even 3D model data, particularly in the intelligent mining of spatial data? These are the major technical difficulties in big data prospecting.

In addition, prospecting prediction requires a comprehensive consideration of various factors, such as expert opinion, geological background, and prospecting type, through a certain set of rules. How to combine the knowledge of geological experts with artificial intelligence to realize an organic combination of data-driven, model-driven, and knowledge-driven patterns, a systematic evolution path incorporated with prior scientific knowledge, and the leap from artificial intelligence to precise artificial intelligence? How to monitor the automatic acquisition and updating of geological big data, and to improve system decision-making through the self-optimization mechanism of the model? These are the core aspects to realize in the transformation of prospecting prediction models under the fourth scientific paradigm.

5.4. Sharing and Visualizing the Prospecting Prediction Big Data

Generally, the sharing and visualization of big data is the terminal link to realize auxiliary decision-making through big data mining. The effective use of geological big data can solve scientific problems such as prospecting prediction, geological law research, and resource evaluation. This ultimately requires some methods to express it and serve the user community. The core of big data lies in the flow of data. Therefore, the sharing of data plays a vital role. The textual, spatial, process, and result data obtained in the big-data-prospecting prediction process often have complex storage structures and large amounts of information. In addition to the diversity, mass, and heterogeneity of general big data, this type of data has unique characteristics such as space–time, variability, causality, and correlation. Currently, the mature Virtual Reality Platform (VRP) multivariate data integration technology can only realize offline sharing. A platform that can maximize the sharing and re-datafication of predictive big data other than confidential data is lacking; this is one of the main unresolved problems.

The data generated in intelligent prospecting prediction based on geological big data not only have a simple data layer, but also include an information layer, a knowledge layer, and even a decision-making layer, such as an expert knowledge graph, network structure tree, and intelligent prediction model. For intelligent prospecting, these inference data and knowledge networks have storage value. Currently, there is no such storage

system to effectively store and manage all the intermediate data, including basic data, algorithm models, expert knowledge, and common software. Therefore, it is meaningful and challenging to establish an intelligent mineral prediction platform that integrates digital expression, quantitative correlation, virtual simulation, and quantitative prediction with a data base (DB), model base (MB), and knowledge base (KB) as the bases, and with spatiotemporal cloud storage and a spatiotemporal data base, knowledge base, and knowledge graph, as well as high-performance computing, virtual visualization, and application development as the support.

6. Theory and Method of Intelligent Mineral Prediction

Characterized by their massive scale, heterogeneity, multimodality, complex correlations, and dynamic emergence, big data require efficient computing models and methods [84]. Moreover, earth science research remains a “multiparadigm” mixture as a whole, where the main body remains in the first paradigm of “logical reasoning”, while some have entered the second paradigm of “theoretical modeling”. A few disciplines have explored the third paradigm of “computational simulation”, which is gradually advancing toward the fourth paradigm of “data analysis”; this requires combining the specific characteristics of earth science and proposing a new theory for the prospecting prediction of geological big data. The intelligent prospecting prototype theory and methods of geological big data proposed in this study (Figure 2) attempt to transform the “logical reasoning” in the first paradigm to knowledge graph association rules, and transform the “theoretical modeling” in the second paradigm to the reasoning rules in the algorithm model, which successfully combines geological knowledge and “artificial intelligence” in the “fourth paradigm”.

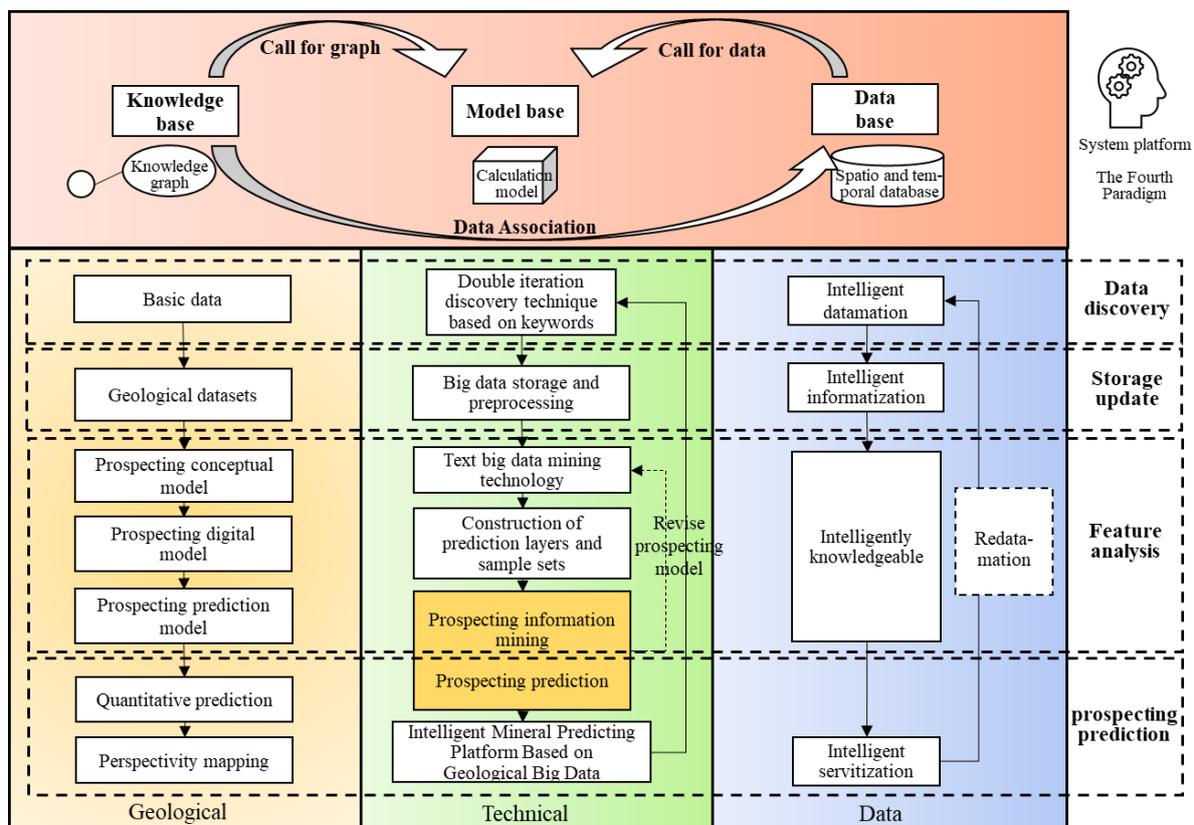


Figure 2. Theoretical method framework of intelligent prospecting based on geological big data.

Specifically, the intelligent prospecting theory and method of geological big data implies the consideration of “geological”, “technicality”, and “data” as the three main lines. Unlike the conventional quantitative prediction method, it adopts the hybrid-driven pattern for prospecting prediction under the big data. The hybrid pattern is composed

of a knowledge-driven pattern based on a geological, model-driven pattern, which in turn is based on technical, and data-driven pattern based on data. The core of the theory is to promote the flow, invocation, circulation, and optimization of the three key factors of “knowledge”, “model”, and “data”, and preliminarily constitute a prototype of the intelligent linkage mechanism, which advances the data-based and information-based prospecting stage to the smart and intelligent stage. Aiming at the quantitative prediction based on geological big data, the data line can be divided into the following four parts.

6.1. Intelligent Datamation

The theoretical method of intelligent discovery and acquisition of geological big data provides a theoretical basis for the process of intelligent datamation. The comprehensiveness, professionalism, and effectiveness of the relevant data required for prospecting prediction directly determine the accuracy of intelligent prospecting prediction results in the big data age, which is also the basis for conducting intelligent prediction. The previous prediction has been often based on existing geological data in geological departments; however, the advent of the big data age has tremendously expanded our data sources. Hence, it is necessary to establish a perfect intelligent data discovery and update system based on the unique characteristics of geological big data, so as to effectively screen and extract semistructured, structured, and unstructured data. Through the intelligent discovery, intelligent screening, intelligent matching, and intelligent updating of data, the data can be automatically classified and stored, providing detailed textual and spatial data for the informatization and knowledgeization processes.

6.2. Intelligent Informatization

The theoretical methods of integration, storage, and automatic updating of multivariate, massive big data not only provide theoretical and methodological support for data informatization, but also provide powerful background support for intelligent servitization and intelligent redatamation. The normal operation of an intelligent system is inseparable from the intelligent flow of geological data. Currently, the integration of geological data typically revolves around the integration of spatial location, ignoring the significance and value of the data itself. Therefore, based on the theoretical and methodological framework of spatial data integration (Figure 3), including data source integration, geometric integration, and band abstract integration, this study fully utilizes constraints related to various geological meanings and knowledge to establish a multivariate data integration and management system based on a unified basic geographic space. Accordingly, the spatial data (such as geology, structure, mineralized point, geophysics, geochemistry, and remote sensing), textual data (such as exploration reports, references, news, and information), algorithm model (such as intelligent mining algorithm models), and expert knowledge (such as prospecting concept model bases, knowledge graphs, and logical trees) can be unified and integrated into the system platform. Thus, the data exchange and sharing can be achieved through the intelligent prospecting prediction platform, ultimately serving the social demands.

6.3. Intelligently Knowledgeable

The theoretical method of metallogenic feature analysis under big data provides a theoretical basis for the knowledgeization of big data in intelligent prospecting decision making, which aims to construct a conceptual model and a prediction model for prospecting. The analysis and extraction of metallogenic features under geological big data involve not only the extraction of key metallogenic information from text corpus, but also the analysis and extraction of metallogenic features from multidimensional spatial data distribution. Based on the textual big data feature extraction, the conceptual prospecting model is constructed, so as to guide the construction of the prospecting prediction model based on the feature analysis of the spatial data. The analysis of metallogenic features involves multiple structured, semistructured, and unstructured information; analysis methods include the NLP

method, artificial intelligence mining method, geostatistics method, nonlinear analysis method, and multidimensional space analysis method. Prospecting prediction models can be divided into knowledge-driven, data-driven, and hybrid-driven models [123–125]. The primary task of metallogenic prediction is to develop a mineral resources prediction technology rooted in metallogenic regularity. In this regard, big data intelligent mining methods, such as deep learning, are data-driven models, which can extract useful features from massive data and form a predictive-thinking pattern from the known to the unknown. However, these have drawbacks in that they only focus on correlations rather than causality and only emphasize results rather than processes. Hence, how to better integrate the geological knowledge into metallogenic regularity or metallogenic models with the intelligent mining method is undoubtedly the key for intelligent metallogenic prediction under the big data background.

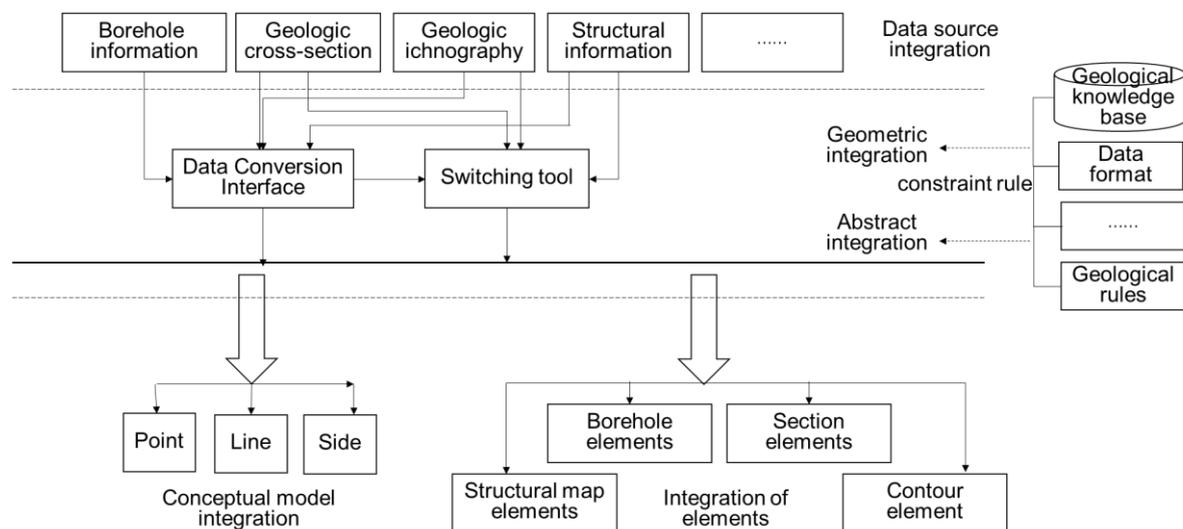


Figure 3. Theoretical method framework of multisource data integration based on geological knowledge [5].

Intelligent knowledgeization refers to the process of forming unknown and potential knowledge through the intelligent deep mining of textual and spatial data, based on hybrid-driven patterns. This progress not only combines the knowledge graphs and knowledge logical structure of geological experts with conceptual prospecting models of nearly hundreds of ore deposits in guiding data-driven mining, but also ensures the flow, invocation, circulation, and optimization of “knowledge” carriers. “Knowledge” carriers here refer to new knowledge nodes, new logical structures, and new conceptual prospecting models from the mining process. Data mining methods refer to the process of searching and excavating currently unknown but valuable potential information from massive data by combining the modern big data technology and artificial intelligence technology with conventional data analysis methods [126]. Data mining involves a variety of algorithms, such as Naive Bayes, neural network, K neighborhood, decision tree, and other machine learning algorithms, as well as the association analysis, clustering analysis, classification regression tree, and support vector machine algorithm based on statistical learning theory. Data mining is closely associated with various information technologies such as statistics, artificial intelligence, expert systems, and database systems. Currently, many researchers have extensively applied multiple technologies to data-intensive industries or fields. The concept of knowledge discovery was first proposed in the 11th International Joint Conference on Artificial Intelligence (1989), where the most important core link was data mining. The data mining knowledge generally includes regularities, concepts, patterns, and rules [20].

Based on the extracted key prospecting for textual information and key spatial vector information, this study conducted in-depth mining to obtain the following two aspects of knowledge: (i) to clarify the metallogenic background, metallogenic period, genetic type, and mineralization type, in order to establish the metallogenic regularity and metallogenic model, so as to determine the conceptual prospecting model of the study area; (ii) to train the reasoning rules of the prospecting prediction model for the self-optimization of the model and accurate artificial intelligence, incorporated with prior scientific knowledge.

On the one hand, to establish the conceptual prospecting model, owing to over 300 years of results accumulated in geology research, the development of earth science information systems should be based on the existing geological theories and research results. Its innovation is to integrate the “discipline accumulation” recorded in the form of texts and graphs into geological big data analysis. A metallogenic system is a process of controlling all the geological elements and mineralization of the formation, change, and preservation of ore deposits in a certain geological spatiotemporal domain, as well as the integrity within ore deposit series and mineralization anomaly series, which is a natural system with a metallogenic function. Certain “metallogenic background, metallogenic age, genetic type and mineralization types” determine certain “metallogenic conditions, ore-controlling factors, prospecting signs, and mineralization anomalies”, which further determine a certain “prospecting model”. Therefore, combined with the series of prospecting model achievements in the big data discovery of the earth system, this study extracted knowledge of metallogenic regularity and metallogenic models based on multiple data mining methods, such as the NLP technology, data classification technology, and machine (deep) learning technology, and constructed a prospecting model of the study area, which is another key factor in realizing intelligent prospecting theories.

On the other hand, aiming at intelligent prospecting, this paper proposes extracting the core prediction data of the study area based on the vector layer base and 3D geological model data, to determine the characteristics of the geological spatial data and the relationship between geological bodies. To extract the metallogenic anomaly features from spatial vector data, this study combined the conventional quantitative feature extraction methods, including the various extraction methods for useful metallogenic information, such as structural development features, structural zone features, structural ore-guiding features, and rock mass features. Meanwhile, based on the super computing power and the method of artificial intelligence, the automatic metallogenic predictive features of the geological big data were studied, verifying the correlation between multiple geological variables and the ore bodies. By comparing the metallogenic prediction ability of different prediction methods in the random combination of different ore-controlling factors, this study conducted the intelligent optimization and evaluation of the metallogenic prediction method, so as to determine the optimal prediction model and factor combination suitable for the study area, thereby increasing the prospecting accuracy.

The metallogenic features that were difficult to be extracted due to asymmetric information were processed with geological knowledge transfer via transfer learning, which combines geological expert knowledge and artificial intelligence. It realizes the system evolution path incorporated with prior scientific knowledge, characterized with stochastic and nonlinear relationships, and transforms deep learning, based on the adjustable parameters, into operational learning, based on the system, forming a systematic algorithm model base, so as to leap from artificial intelligence to precise artificial intelligence. Consequently, the entire process of information extraction, knowledge integration, knowledge reasoning, and intelligent predictive analysis is realized.

6.4. Intelligent Servitization

The theory and method of big-data-prospecting prediction not only provides theoretical and methodological support for the servitization link of intelligent prospecting, but is also the final stage of intelligent prospecting prediction. Therefore, studies must prioritize the comprehensive analysis of the results of metallogenic feature extraction and ultimately

serve mineral exploration. This will help achieve two objectives: (1) accomplishing the task of intelligent prospecting and improving its accuracy and efficiency; (2) forming friendly knowledge expression and transmission, this study established an intelligent mineral prediction platform based on geological big data, with the decision-making system as the core, and the DB, MB, and KB as the bases. Moreover, supported by the five engines of spatiotemporal cloud storage and spatiotemporal data bases, knowledge bases and knowledge graphs, high-performance computing, virtual visualization, and application development, and with prospecting prediction and geological data sharing as the practical application, this platform integrates digital expression, quantitative correlation, virtual simulation, and quantitative prediction.

On the one hand, to improve the efficiency and accuracy of the mineral resource prediction, this paper proposes not only the intelligent prediction method incorporated with the prior knowledge of experts, but also a series of model evaluation methods to ensure the stability and effectiveness of the prediction model. In the process, the algorithm automatically learns the logical rules and establishes the mechanism of random combination of the ore-controlling factors, to gradually approach the optimal combination and achieve the ore targets. Considering that the random combination quantity can reach hundreds of thousands or even millions, this paper proposes determining the optimal parameter method based on the fractal theory, which shortens the learning time of the computer and accelerates the convergence speed of the model, enabling the prediction model to quickly determine the optimal solution and achieve self-optimization. This method is useful in intelligent spatial data mining as well as in the decision-making analysis.

On the other hand, for friendly knowledge expression and transmission, this study conducted integrated processing for the metallogenic knowledge acquired in the process of prediction and is expressed in a visual manner, whereby the results can be easily perceived. With the decision-making system as the core and prospecting prediction and geological data sharing as the practical application, this paper proposes and establishes an intelligent mineral prediction platform based on geological big data, which integrates digital expression, quantitative correlation, virtual simulation, and quantitative prediction. The metallogenic knowledge obtained by text mining is expressed by a knowledge graph, which is helpful for mineral resource prediction and is easier to understand. For the knowledge acquired in the process of spatial data mining, this research visualizes not only the convolution kernel of the training model, but also the digital model and prediction results obtained using the Web3D method. This is helpful for better understanding the spatial characteristics of underground metallogenic space and the spatial relationship between underground and surface, so as to provide a reasonable basis for mineral exploration.

In summary, in addition to the above four parts, the intelligent prospecting prototype theory and method of geological big data also provide the theoretical and methodological basis for the redatamation process of “knowledge”, “model”, and “data”. Big data not only contain predictive-thinking patterns, but also have dynamic features; this is reflected in the redatamation process and is also the core of the intelligent prospecting theory. To avoid forming “data islands”, this paper presents a data flow model of intelligent redatamation, with KB, DB, and MB as the bases. The automatic matching and update mechanism of the core database can be achieved by constantly incorporating operational data, the obtained “new knowledge”, “new model”, and “new data” from users, via the cloud storage architecture, into the three bases.

7. Conclusions

With the rapid development of earth science, manual data analysis cannot in a timely manner “digest” the rapid growth in the volume and complexity of data. Characterized by multisource heterogeneity, different datums, multiple features, and complex semantics, geological big data typically have a complex relationship with mineralization. The “dimension disaster” encountered in high-dimensional space statistics severely affects the statistical modeling and results from multisource observational data; it may reduce the

reliability of the prospecting prediction results. Moreover, the potential ore-controlling factors are likely to be neglected. Therefore, when combined with the predictive modeling of big data and the reasoning of artificial intelligence, establishing an efficient, unified, intelligent prospecting prediction process is necessary for mineral resource prediction in the big data age. This paper reviewed the intelligent prospecting theory and method of geological big data, aiming to create a big-data-prospecting prediction system combined with knowledge-driven, model-driven, and data-driven patterns, so as to make the prediction process smart and intelligent in the big data age.

Author Contributions: S.L. and J.C. designed the review; J.C. organized the framework of the paper; S.L. and C.L. wrote and revised the paper carefully. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by the project “Research on Key Technologies of Geological Text Big Data Discovery and Mining” (2022XJQN18).

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

References

1. Bristol, R.S.; Euliss, N.H., Jr.; Booth, N.L.; Burkardt, N.; Diffendorfer, J.E.; Gesch, D.B.; McCallum, B.E.; Miller, D.M.; Morman, S.A.; Poore, B.S. *Science Strategy for Core Science Systems in the US Geological Survey, 2013–2023*; US Geological Survey: Reston, VA, USA, 2012.
2. Chen, J.P.; Li, J.; Xie, S.; Liu, J.; Hu, B. China geological big data research status. *J. Geol.* **2017**, *41*, 353–366.
3. Zhao, P.D. Digital geology and mineral resources evaluation. *J. Geol.* **2012**, *36*, 225–228.
4. Zhao, P.D. Digital geology research needs to be emphasized in the era of big data. *China Land Research Daily*, 14 March 2013; p. 6.
5. Zhou, Y.Z.; Zhang, L.J.; Zhang, A.D. *Big Data Mining and Learning in Geoscience*; Sun YAT-SEN University Press: Guangzhou, China, 2018; pp. 1–21.
6. Earth Cube Guidance for the Community. Available online: <http://www.nsf.gov/pubs/2011/nsfl085/nsfl085.pdf> (accessed on 19 March 2022).
7. Baumann, P.; Mazzetti, P.; Ungar, J.; Barbera, R.; Barboni, D.; Beccati, A.; Bigagli, L.; Boldrini, E.; Bruno, R.; Calanducci, A. Big data analytics for earth sciences: The EarthServer approach. *Int. J. Digit. Earth* **2016**, *9*, 3–29. [[CrossRef](#)]
8. Miao, J.L.; Qu, H.G.; Xu, Z.; Yin, T.F. The Experimental Platform of Technical Study of the Big Data of Geoscience: GeoBDA. *Geomat. World* **2014**, *21*, 48–52.
9. Wei, Q.; Wan, L. Hadoop-based massive xml index query solution. *Comput. Cd Softw. Appl.* **2013**, *16*, 69–73.
10. Shen, D.R.; Yu, G.; Wang, X.T.; Nie, T.Z.; Kou, Y. Survey on NoSQL for Management of Big Data. *J. Softw.* **2013**, *24*, 1803. [[CrossRef](#)]
11. Yan, J.H. Parallel Implementation of PageRank. *J. Taiyuan Norm. Univ. (Nat. Sci. Ed.)* **2014**, *13*, 53–57, 78.
12. Eldawy, A.; Mokbel, M.F. Spatialhadoop: A mapreduce framework for spatial data. In Proceedings of the 2015 IEEE 31st International Conference on Data Engineering, Seoul, Korea, 13–16 April 2015.
13. Li, C.L.; Li, J.Q.; Zhang, H.C.; Gong, A.H.; Wei, D.Q. Big data application architecture and key technologies of intelligent geological survey. *Geol. Bull. China* **2015**, *34*, 1288–1299.
14. Sun, Y.Z. Important achievement and advance of natural gas geology and geochemical exploration in China. *Acta Geol. Sin. Engl. Ed.* **2015**, *4*, 1411.
15. Hao, Z.G.; Fei, H.C.; Hao, Q.Q.; Turner, S. Reserves of Natural Gas Hydrates Equivalent to 100–150 Billion m³ Natural Gas Has Been Discovered in the Pearl Mouth Basin of the South China Sea. *Acta Geol. Sin.-Engl. Ed.* **2014**, *88*, 361. [[CrossRef](#)]
16. Li, Y.B.; Dai, W.Y.; Ming, Z.; Qiu, M.K. Privacy protection for preventing data over-collection in smart city. *IEEE Trans. Comput.* **2015**, *65*, 1339–1350. [[CrossRef](#)]
17. Zang, S.Y.; Ye, M.L. Human Resource Management in the Era of Big Data. *J. Hum. Resour. Sustain. Stud.* **2015**, *3*, 41–45. [[CrossRef](#)]
18. Ren, M.Y.; Chen, J.P.; Shao, K.; Zhang, S. Metallogenic information extraction and quantitative prediction process of seafloor massive sulfide resources in the Southwest Indian Ocean. *Ore Geol. Rev.* **2016**, *76*, 108–121. [[CrossRef](#)]
19. Ma, Y.; Wu, H.P.; Wang, L.Z.; Huang, B.; Ranjan, R.; Zomaya, A.; Jie, W. Remote sensing big data computing: Challenges and opportunities. *Future Gener. Comput. Syst.* **2015**, *51*, 47–60. [[CrossRef](#)]
20. Han, J.W.; Tung, A.K.; He, J. SPARC: Spatial association rule-based classification. In *Data Mining for Scientific and Engineering Applications*; Springer: New York, NY, USA, 2001; pp. 461–485.
21. Zhao, P.D. Characteristics and rational utilization of geological big data. *Earth Sci. Front.* **2019**, *26*, 1–5.
22. Zhao, P.D. Digital mineral exploration and quantitative evaluation in the big data age. *Geol. Bull. China* **2015**, *34*, 1255–1259.
23. Xiao, K.Y.; Li, N.; Wang, K.; Sun, L.; Fan, J.F.; Ding, J.H. Mineral resources assessment under the thought of big data. *Geol. Bull. China* **2015**, *34*, 1266–1272.
24. Yu, P.P.; Chen, J.P.; Chai, F.H.; Zheng, X.; Yu, M.; Xu, B. Research on model-driven quantitative prediction and evaluation of mineral resources based on geological big data concept. *Geol. Bull. China* **2015**, *34*, 1333–1343.

25. Wang, D.H.; Liu, X.X.; Liu, L.J. Characteristics of Big Geodata and its application to study of minerogenetic regularity and minerogenetic series. *Miner. Depos.* **2015**, *34*, 1143–1154.
26. Xiang, J. Quantitative Prediction and Evaluation of Mineral Resources Based on Geological Big Data. Master's Thesis, China University of Geosciences, Beijing, China, 2018.
27. Cernuzzi, L.; Pane, J. Toward open government in Paraguay. *IT Prof.* **2014**, *16*, 62–64. [[CrossRef](#)]
28. Ma, X.G. Linked Geoscience Data in practice: Where W3C standards meet domain knowledge, data visualization and OGC standards. *Earth Sci. Inform.* **2017**, *10*, 429–441. [[CrossRef](#)]
29. Lent, B.; Agrawal, R.; Srikant, R. Discovering Trends in Text Databases. In Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, Newport Beach, CA, USA, 14–17 August 1997.
30. Fung, G.P.C.; Yu, J.X.; Lam, W. Stock prediction: Integrating text mining approach using real-time news. In Proceedings of the 2003 IEEE International Conference on Computational Intelligence for Financial Engineering, Hong Kong, China, 20–23 March 2003.
31. Patrick, J. The scamseek project—Text mining for financial scams on the internet. In *Data Mining*; Williams, G.J., Simoff, S.J., Eds.; Springer: Cham, Switzerland, 2008; pp. 295–302.
32. Porter, A.; Rafols, I. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics* **2009**, *81*, 719–745. [[CrossRef](#)]
33. Byrd, R.J.; Steinhubl, S.R.; Sun, J.; Ebadollahi, S.; Stewart, W.F. Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records. *Int. J. Med. Inform.* **2014**, *83*, 983–992. [[CrossRef](#)] [[PubMed](#)]
34. Lin, H.F.; Gong, D.Y.; Zhang, Y.; Yao, T.S. Visualization Model for Chinese Text Mining. *Comput. Sci.* **2000**, *27*, 37–41.
35. Chen, Z.Q.; Zhang, G.X. Study on the Text Mining and Chinese Text Mining Framework. *Inf. Sci.* **2007**, *25*, 1046–1051.
36. Deng, T.L. Analysis and Implementation of Text Mining for Call Centers with Big Data. Master's Thesis, Beijing University of Posts and Telecommunications, Beijing, China, 2015.
37. Wang, C.B.; Ma, X.G.; Chen, J.G.; Chen, J.W. Information extraction and knowledge graph construction from geoscience literature. *Comput. Geosci.* **2018**, *112*, 112–120. [[CrossRef](#)]
38. Hinton, G.E.; Osindero, S.; Teh, Y. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)]
39. Bengio, Y.; Lamblin, P.; Popovici, D.; Larochelle, H. Greedy layer-wise training of deep networks. In *Advances in Neural Information Processing Systems 19 (NIPS 2006)*; Schölkopf, B., Platt, J., Hoffman, T., Eds.; MIT Press: Cambridge, UK, 2006; p. 1672.
40. Guimaraes, R.G.; Rosa, R.L.; De Gaetano, D.; Rodriguez, D.Z.; Bressan, G. Age groups classification in social network using deep learning. *IEEE Access* **2017**, *5*, 10805–10816. [[CrossRef](#)]
41. Manning, C.; Schütze, H. *Foundations of Statistical Natural Language Processing*; MIT Press: Cambridge, UK, 1999; p. 680.
42. Kalchbrenner, N.; Grefenstette, E.; Blunsom, P. A convolutional neural network for modelling sentences. *arXiv* **2014**, arXiv:1404.2188.
43. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
44. Lei, T.; Barzilay, R.; Jaakkola, T. Molding CNNs for text: Non-linear, non-consecutive convolutions. *Indiana Univ. Math. J.* **2015**, *58*, 1151–1186.
45. Sun, S.T.; He, Y.X. Multi-label Emotion Classification for Microblog Based on CNN Feature Space. *J. Sichuan Univ. (Eng. Sci. Ed.)* **2017**, *49*, 162–169.
46. Feng, S.; Wang, Y.Q.; Song, K.S.; Wang, D.L.; Yu, G. Detecting multiple coexisting emotions in microblogs with convolutional neural networks. *Cogn. Comput.* **2018**, *10*, 136–155. [[CrossRef](#)]
47. Ehrlich, K.; Lin, C.; Griffiths-Fisher, V. Searching for experts in the enterprise: Combining text and social network analysis. In Proceedings of the 2007 International ACM Conference on Supporting Group Work, Sanibel Island, FL, USA, 4–7 November 2007.
48. Paranyushkin, D. Identifying the pathways for meaning circulation using text network analysis. *Nodus Labs* **2011**, *26*, 1–26.
49. Schuhmacher, M.; Ponzetto, S.P. Knowledge-based graph document modeling. In Proceedings of the 7th ACM International Conference on Web Search and Data Mining, New York, NY, USA, 24–28 February 2014.
50. Lu, F.; Yu, L.; Qiu, P.Y. On Geographic Knowledge Graph. *J. Geo-Inf. Sci.* **2017**, *19*, 723–734.
51. Xu, J.; Pei, T.; Yao, Y.H. Conceptual Framework and Representation of Geographic Knowledge Map. *Geo-Inf. Sci.* **2010**, *12*, 496–502. [[CrossRef](#)]
52. Hou, Z.W.; Zhu, Y.Q.; Gao, Y.; Song, J.; Qin, C.Z. Geologic Time Scale Ontology and Its Applications in Semantic Retrieval. *J. Geo-Inf. Sci.* **2018**, *20*, 17–27.
53. Morrison, S.M.; Liu, C.; Eleish, A.; Prabhu, A.; Li, C.R.; Ralph, J.; Downs, R.T.; Golden, J.J.; Fox, P.; Hummer, D.R. Network analysis of mineralogical systems. *Am. Mineral. J. Earth Planet. Mater.* **2017**, *102*, 1588–1596. [[CrossRef](#)]
54. Wu, Y.L. Research on the Key Technology of Geological Text Big Data Discovery and Mining. Master's Thesis, China University of Geosciences, Beijing, China, 2018.
55. The Age of Big Data. Available online: <https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html> (accessed on 19 March 2022).
56. Zhao, P.D.; Chen, J.P.; Zhang, S.T. New advances in “triplex” mineralization prediction. *Earth Sci. Front.* **2003**, *10*, 455–463.
57. Chen, S.M. Research of Multiple Geoscience Information Prospecting Prediction in Xikuangshan Antimony Ore Field. Ph.D. Dissertation, China University of Geosciences, Beijing, China, 2012.

58. Li, S.; Chen, J.P.; Xiang, J.; Zhang, Z.P.; Zhang, Y. Two-dimensional ore finding prediction based on AlexNet network: A Case of sedimentary manganese ore in Songtao-Huayuan area. *Geol. Bull. China* **2019**, *38*, 2022–2032.
59. Brown, W.M.; Gedeon, T.D.; Groves, D.I.; Barnes, R.G. Artificial neural networks: A new method for mineral prospectivity mapping. *Aust. J. Earth Sci.* **2000**, *47*, 757–770. [[CrossRef](#)]
60. Abedi, M.; Norouzi, G.; Bahroudi, A. Support vector machine for multi-classification of mineral prospectivity areas. *Comput. Geosci.* **2012**, *46*, 272–283. [[CrossRef](#)]
61. Leite, E.P.; de Souza Filho, C.R. Artificial neural networks applied to mineral potential mapping for copper-gold mineralizations in the Carajás Mineral Province, Brazil. *Geophys. Prospect.* **2009**, *57*, 1049–1065. [[CrossRef](#)]
62. Zuo, R.; Carranza, E.J.M. Support vector machine: A tool for mapping mineral prospectivity. *Comput. Geosci.* **2011**, *37*, 1967–1975. [[CrossRef](#)]
63. Chang, C.Y.; Chen, S.; Tsai, M. Application of support-vector-machine-based method for feature selection and classification of thyroid nodules in ultrasound images. *Pattern Recognit.* **2010**, *43*, 3494–3506. [[CrossRef](#)]
64. Li, D. Comparative Study on Metallogenic Prediction Models Based on BP Neural Network and SVM: A Case Study of Gold Ore in HATU Baobao Area. Master's Thesis, University of Chinese Academy of Sciences, Beijing, China, 2014.
65. Rodriguez-Galiano, V.F.; Chica-Olmo, M.; Chica-Rivas, M. Predictive modelling of gold potential with the integration of multisource information based on random forest: A case study on the Rodalquilar area, Southern Spain. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1336–1354. [[CrossRef](#)]
66. Hariharan, S.; Tirodkar, S.; Porwal, A.; Bhattacharya, A.; Joly, A. Random forest-based prospectivity modelling of greenfield terrains using sparse deposit data: An example from the Tanami Region, Western Australia. *Nat. Resour. Res.* **2017**, *26*, 489–507. [[CrossRef](#)]
67. Phillips, S.J.; Dudík, M. Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography* **2008**, *31*, 161–175. [[CrossRef](#)]
68. Chen, Y.L.; Zhou, B.; Li, X.B. Mineral target prediction based on Boltzmann machines. *Prog. Geophys.* **2012**, *27*, 179–185.
69. Chen, Y.L.; Wu, W. Mapping mineral prospectivity using an extreme learning machine regression. *Ore Geol. Rev.* **2017**, *80*, 200–213. [[CrossRef](#)]
70. Liu, Y.; Zhou, K.F.; Xia, Q.L. A MaxEnt model for mineral prospectivity mapping. *Nat. Resour. Res.* **2018**, *27*, 299–313. [[CrossRef](#)]
71. Liu, Y.; Zhou, K.F.; Zhang, N.N.; Wang, J.L. Maximum entropy modeling for orogenic gold prospectivity mapping in the Tangbale-Hatu belt, western Junggar, China. *Ore Geol. Rev.* **2018**, *100*, 133–147. [[CrossRef](#)]
72. Zuo, R.G.; Xiong, Y.H. Big data analytics of identifying geochemical anomalies supported by machine learning methods. *Nat. Resour. Res.* **2018**, *27*, 5–13. [[CrossRef](#)]
73. Twarakavi, N.K.; Misra, D.; Bandopadhyay, S. Prediction of arsenic in bedrock derived stream sediments at a gold mine site under conditions of sparse data. *Nat. Resour. Res.* **2006**, *15*, 15–26. [[CrossRef](#)]
74. Chen, M.; Mao, S.W.; Liu, Y.H. Big data: A survey. *Mob. Netw. Appl.* **2014**, *19*, 171–209. [[CrossRef](#)]
75. O'Brien, J.J.; Spry, P.G.; Nettleton, D.; Xu, R.; Teale, G.S. Using Random Forests to distinguish gahnite compositions as an exploration guide to Broken Hill-type Pb–Zn–Ag deposits in the Broken Hill domain, Australia. *J. Geochem. Explor.* **2015**, *149*, 74–86. [[CrossRef](#)]
76. Gonbadi, A.M.; Tabatabaei, S.H.; Carranza, E.J.M. Supervised geochemical anomaly detection by pattern recognition. *J. Geochem. Explor.* **2015**, *157*, 81–91. [[CrossRef](#)]
77. Kirkwood, C.; Cave, M.; Beamish, D.; Grebby, S.; Ferreira, A. A machine learning approach to geochemical mapping. *J. Geochem. Explor.* **2016**, *167*, 49–61. [[CrossRef](#)]
78. Zhao, J.N.; Chen, S.Y.; Zuo, R.G. Identifying geochemical anomalies associated with Au–Cu mineralization using multifractal and artificial neural network models in the Ningqiang district, Shaanxi, China. *J. Geochem. Explor.* **2016**, *164*, 54–64. [[CrossRef](#)]
79. Xiong, Y.H.; Zuo, R.U. Recognition of geochemical anomalies using a deep autoencoder network. *Comput. Geosci.* **2016**, *86*, 75–82. [[CrossRef](#)]
80. Nie, H.; Zhu, Y.Q.; Chang, L.H.; Yan, D. Research on construction method of data-driven minerals prediction model. *China Min. Mag.* **2018**, *27*, 82–87.
81. Lin, N. Study on the Metallogenic Prediction Models Based on Remote Sensing Geology and Geochemical Information: A Case Study of Lalingzaohuo Region in Qinghai Province. Ph.D. Dissertation, Jilin University, Changchun, China, 2015.
82. Yan, G.S.; Xue, Q.W.; Xiao, K.Y.; Chen, J.P.; Miao, J.L.; Yu, H.L. An analysis of major problems in geological survey big data. *Geol. Bull. China* **2015**, *34*, 1273–1279.
83. Zhang, Q.; Zhou, Y.Z. Big data will lead to a profound revolution in the field of geological science. *Chin. J. Geol.* **2017**, *52*, 637–648.
84. Zhou, Y.Z.; Li, P.X.; Wang, S.G.; Xiao, F.; Li, J.Z.; Gao, L. Research Progress on Big Data and Intelligent Modelling of Mineral Deposits. *Bull. Mineral. Petrol. Geochem.* **2017**, *36*, 327–331, 344.
85. Zhou, Y.Z.; Wang, J.; Zuo, R.G.; Xiao, F.; Shen, W.J.; Wang, S.G. Machine learning, deep learning and Python language in field of geology. *Acta Petrol. Sin.* **2018**, *34*, 3173–3178.
86. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
87. Martens, J.; Sutskever, I. *Learning Recurrent Neural Networks with Hessian-Free Optimization*; ACM Press: New York, NY, USA, 2011; pp. 1033–1040.

88. Sainath, T.N.; Kingsbury, B.; Ramabhadran, B. Auto-encoder bottleneck features using deep belief networks. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012.
89. Hinton, G. Deep belief nets. In *Encyclopedia of Machine Learning*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2011; pp. 267–269.
90. Cireşan, D.C.; Giusti, A.; Gambardella, L.M.; Schmidhuber, J. Deep neural networks segment neuronal membranes in electron microscopy images. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; Volume 2.
91. Holtzman, B.K.; Paté, A.; Paisley, J.; Waldhauser, F.; Repetto, D. Machine learning reveals cyclic changes in seismic source spectra in Geysers geothermal field. *Sci. Adv.* **2018**, *4*, eaao2929. [[CrossRef](#)]
92. Rouet Leduc, B.; Hulbert, C.; Lubbers, N.; Barros, K.; Humphreys, C.J.; Johnson, P.A. Machine learning predicts laboratory earthquakes. *Geophys. Res. Lett.* **2017**, *44*, 9276–9282. [[CrossRef](#)]
93. Zhou, Y.Z.; Chen, S.; Zhang, Q.; Xiao, F.; Wang, S.G.; Liu, Y.P.; Jiao, S.T. Advances and prospects of big data and mathematical geoscience. *Acta Petrol. Sin.* **2018**, *34*, 255–263.
94. Xiong, Y.H.; Zuo, R.G.; Carranza, E.J.M. Mapping mineral prospectivity through big data analytics and a deep learning algorithm. *Ore Geol. Rev.* **2018**, *102*, 811–817. [[CrossRef](#)]
95. Zuo, R.G. Deep Learning-Based Mining and Integration of Deep-Level Mineralization Information. *Bull. Mineral. Petrol. Geochem.* **2019**, *38*, 53–60.
96. Zuo, R.G. Exploration geochemical data mining and weak geochemical anomalies identification. *Earth Sci. Front.* **2019**, *26*, 67–75.
97. Li, C.; le Jiang, Y.; Hu, M.K.; Jiang, Y.D.; Zheng, C. Study and application of gravity anomaly separation by cellular neural networks. *Comput. Tech. Geophys. Geochem. Explor.* **2015**, *37*, 16–21.
98. Cai, H.H.; Zhu, W.; Li, Z.X.; Liu, Y.Y.; Li, L.B.; Liu, C. Prediction Method of Tungsten-molybdenum Prospecting Target Area based on Deep Learning. *J. Geo-Inf. Sci.* **2019**, *21*, 928–936.
99. Li, S.; Chen, J.P.; Xiang, J. Applications of deep convolutional neural networks in prospecting prediction based on two-dimensional geological big data. *Neural Comput. Appl.* **2020**, *32*, 2037–2053. [[CrossRef](#)]
100. Li, T.; Zuo, R.G.; Xiong, Y.H.; Peng, Y. Random-drop data augmentation of deep convolutional neural network for mineral prospectivity mapping. *Nat. Resour. Res.* **2021**, *30*, 27–38. [[CrossRef](#)]
101. Zuo, R.G. Geodata science-based mineral prospectivity mapping: A review. *Nat. Resour. Res.* **2020**, *29*, 3415–3424. [[CrossRef](#)]
102. Zuo, R.G.; Wang, Z.Y. Effects of random negative training samples on mineral prospectivity mapping. *Nat. Resour. Res.* **2020**, *29*, 3443–3455. [[CrossRef](#)]
103. Mao, X.C.; Zou, Y.H.; Chen, J.; Lai, J.Q.; Peng, S.L.; Shao, Y.J.; Shu, Z.M.; Lu, J.W.; Lu, C.Y. Three-dimensional visual prediction of concealed ore bodies in the deep and marginal parts of crisis mines: A case study of the Fenghnangshan ore field in Tongling, Anlu, China. *Geol. Bull. China* **2010**, *29*, 401–413.
104. Xiao, K.Y.; Li, N.; Sun, L.; Zou, W.; Li, Y. Large scale 3D mineral prediction methods and channels based on 3D information technology. *J. Geol.* **2012**, *36*, 229–236.
105. Chen, J.P.; Wang, C.N.; Shang, B.C.; Shi, R. Three-Dimensional Metallogenic Prediction in Yongmei Region Based on Digital Ore Deposit Model. *Sci. Technol. Manag. Land Resour.* **2012**, *29*, 14–20.
106. Rong, J.H.; Chen, J.P.; Shang, B.C. Three-dimensional Prediction of Blind Orebodies in Gejiu, Yunnan Province Based on the Ore-Search Model. *Geol. Prospect.* **2012**, *48*, 191–198.
107. Yan, Q.; Chen, J.P.; Shang, B.C. The 3D Prediction Model and Division of Targets in Lutangba Study Area of Gaosong Ore Field in Gejiu, Yunnan Province. *Geoscience* **2012**, *26*, 286–293.
108. Xiang, J.; Chen, J.P.; Hu, Q.; An, W.T.; Yin, X.Y. Three-dimensional mineralization prediction based on mineralization series of ore deposits—Tongling mine collection area in Anhui Province as an example. *Geoscience* **2016**, *30*, 230–238.
109. Deng, L. Protein Structure Evaluation Based on 3DCNN. Master’s Thesis, Jilin University, Changchun, China, 2019.
110. Oh, K.H.; Kim, W.; Shen, G.F.; Piao, Y.H.; Kang, N.; Oh, I.; Chung, Y.C. Classification of schizophrenia and normal controls using 3D convolutional neural network and outcome visualization. *Schizophr. Res.* **2019**, *212*, 186–195. [[CrossRef](#)]
111. Sato, R.; Ishida, T. Protein model accuracy estimation based on local structure quality assessment using 3D convolutional neural network. *PLoS ONE* **2019**, *14*, e221347.
112. Torng, W.; Altman, R.B. High precision protein functional site detection using 3D convolutional neural networks. *Bioinformatics* **2019**, *35*, 1503–1512. [[CrossRef](#)]
113. Jiang, S.Y.; Chen, Y.M. Hand gesture recognition by using 3DCNN and LSTM with adam optimizer. In *Advances in Multimedia Information Processing—PCM 2017*; Zeng, B., Huang, Q., El Saddik, A., Li, H., Jiang, S., Fan, X., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 10735, pp. 743–753.
114. Ma, Z.W. Research of Gesture Recognition Based on Densely Connected 3DCNN and Convolutional GRU. Master’s Thesis, Fujian Normal University, Fuzhou, China, 2018.
115. Wu, J.; Min, Y.; Li, C.; Zhang, W.H. A Micro-expression Recognition Algorithm Based on 3 D-CNN. *Telecommun. Eng.* **2019**, *59*, 1115–1120.
116. Luo, J.; Li, Y. Brain fMRI signal recognition method based on three-dimensional convolutional neural network. *Chin. J. Stereol. Image Anal.* **2019**, *25*, 191–198.

117. Xing, W.W.; Li, Y.; Zhang, S.L. View-invariant gait recognition method by three-dimensional convolutional neural network. *J. Electron. Imaging* **2018**, *27*, 248–258. [[CrossRef](#)]
118. Liao, J.G.; Wang, S.L.; Zhang, X.X.; Liu, G.S. 3d convolutional neural networks based speaker identification and authentication. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018.
119. Wang, X.; Xie, W.X.; Song, J.Y. Learning spatiotemporal features with 3DCNN and ConvGRU for video anomaly detection. In Proceedings of the 2018 14th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 12–16 August 2018.
120. Qi, H.F.; Li, J.; Wu, Q.; Wan, W.B.; Sun, J.D. A 3D-CNN based video hashing method. In Proceedings of the Tenth International Conference on Digital Image Processing (ICDIP 2018), Shanghai, China, 11–14 May 2018.
121. Zhao, Y.; Yang, Q.J. Research on hyperspectral remote sensing image classification based on 3 D convolutional neural network. *Inf. Technol. Netw. Secur.* **2019**, *38*, 46–51.
122. Wang, H.Y.; Li, X.F.; Li, Y.B.; Sun, Y.X.; Xu, H.L. Nondestructive detection of multi-quality parameters of apples based on hyperspectral images and 3D-CNN. *J. Nanjing Agric. Univ.* **2020**, *43*, 178–185.
123. Agterberg, F. *Geomathematics: Theoretical Foundations, Applications And Future Developments*; Springer: Cham, Switzerland, 2014.
124. Bonham-Carter, G.F. Geographic information systems for geoscientists-modeling with GIS. *Comput. Methods Geosci.* **1994**, *13*, 398.
125. Carranza, E.J.M.; Hale, M. Logistic regression for geologically constrained mapping of gold potential, Baguio district, Philippines. *Explor. Min. Geol.* **2001**, *10*, 165–175. [[CrossRef](#)]
126. Ji, G.L.; Zhao, B. A Survey of Spatiotemporal Data Mining for Big Data. *J. Nanjing Norm. Univ. (Nat. Sci. Ed.)* **2014**, *37*, 1–7.