Artificial Intelligence in Geosciences 3 (2022) 226–236

Contents lists available at ScienceDirect

Artificial Intelligence in Geosciences

journal homepage: www.keaipublishing.com/en/journals/artificial-intelligence-in-geosciences

# A study on geological structure prediction based on random forest method

Zhen Chen [a,*], Qingsong Wu [b], Sipeng Han [b], Jungui Zhang [b], Peng Yang [b], Xingwu Liu [b]

[a] School of Earth Sciences and Resources, China University of Geosciences (Beijing), Beijing, 100083, China
[b] Applied Geology Research Center, China Geological Survey, Chengdu, 610036, China

ARTICLE INFO

*Keywords:*
Geological big data
Basic rock
Tectonic environment discrimination
Random forest

ABSTRACT

The Xingmeng orogenic belt is located in the eastern section of the Central Asian orogenic belt, which is one of the key areas to study the formation and evolution of the Central Asian orogenic belt. At present, there is a huge controversy over the closure time of the Paleo-Asian Ocean in the Xingmeng orogenic belt. One of the reasons is that the genetic tectonic setting of the Carboniferous volcanic rocks is not clear. Due to the diversity of volcanic rock geochemical characteristics and its related interpretations, there are two different views on the tectonic setting of Carboniferous volcanic rocks in the Xingmeng orogenic belt: island arc and continental rift. In recent years, it is one of the important development directions in the application of geological big data technology to analyze geochemical data based on machine learning methods and further infer the tectonic background of basalt. This paper systematically collects Carboniferous basic rock data from Dongwuqi area of Inner Mongolia, Keyouzhongqi area of Inner Mongolia and Beishan area in the southern section of the Central Asian Orogenic Belt. Random forest algorithm is used for training sets of major elements and trace elements in global island arc basalt and rift basalt, and then the trained model is used to predict the tectonic setting of the Carboniferous magmatic rock samples in the Xingmeng orogenic belt. The prediction results shows that the island arc probability of most of the research samples is between 0.65 and 1, which indicates that the island arc tectonic setting is more credible. In this paper, it is concluded that magmatism in the Beishan area of the southern part of the Central Asian Orogenic belt in the Early Carboniferous may have formed in the heyday of subduction, while the Xingmeng orogenic belt in the Late Carboniferous may have been in the late subduction stage to the collision or even the early rifting stage. This temporal and spatial evolution shows that the subduction of the Paleo-Asian Ocean is different from west to east. Therefore, the research results of this paper show that the subduction of the Xingmeng orogenic belt in the Carboniferous has not ended yet.

## 1. Introduction

Central Asian Orogenic Belt is a huge accretionary orogenic belt located between the Tarim Plate, the Siberian Plate and the North China Plate, which is one of the most intensely accretionary regions in the Phanerozoic in the world (Xiao et al., 2003). The Xingmeng orogenic belt belongs to the eastern part of the Central Asian orogenic belt. It is an important suture zone between the North China Plate and the Siberian Plate, mainly located in Inner Mongolia and Northeast China. Since the early Paleozoic, it has undergone a series of evolutionary processes such as the subduction, closure and collision orogeny of the Paleo-Asian Ocean. Among them, the Carboniferous was the period when the tectonic activity in this area was relatively active; therefore, constraining the geological tectonic process in this period is of great significance for understanding the evolution of the Xingmeng orogenic belt. However, there are fierce debates about the geological tectonic evolution in the Carboniferous period. There are two main viewpoints: one is the island-arc tectonic setting, that is, the development of the trench - arc - basin system formed by the two-way subduction of the Paleo-Asian Ocean in the north-south direction (Liu et al., 2006; Xiao et al., 2003; Xiao et al., 2009; Xiao et al., 2015; Liu et al., 2017); on the contrary, the other view holds that the Paleo-Asian ocean closed in the Middle and Late Devonian, and the area was in the continental rifting tectonic setting during the Carboniferous, and was in an extensional tectonic setting (He and Shao, 1983; Shao, 1989; Tang, 1990; Shao, 1991; Tang, 1992; Xu and Chen, 1997; Xu et al., 2013; Xu et al., 2014; Zhao et al., 2015; Zhao et al., 2016; Zhao et al., 2017). Therefore, sorting out the closure time of the Paleo-Asian Ocean is the key to solve this problem.

Both the island arc view and the continental rift view are mainly based on the geochemical indicators of volcanic rocks. However, these

* Corresponding author.
*E-mail address:* chenzhen@cugb.edu.cn (Z. Chen).

https://doi.org/10.1016/j.aiig.2023.01.004
Received 4 June 2022; Received in revised form 20 January 2023; Accepted 23 January 2023
Available online 30 January 2023

geochemical indicators often have multiple solutions, because they are often affected by the nature of the magma source area and partial melting degree, and the influence of magma evolution processes such as crustal contamination. Previous studies have usually explored the origin of basalt according to one or several specific traditional geochemical indicators, and these studies often reached different conclusions on the analysis of tectonic background. It is not possible to invert regional tectonic evolution only by using volcanic rock geochemical indicators. Therefore, it is necessary to develop a new method to infer the tectonic setting of basalt more comprehensively.

Machine learning methods have been widely used in earth sciences (Ueki et al., 2017; Zhao et al., 2019). Compared with the traditional geological method, the machine learning method is to establish a training model based on a global database, and then use the model to predict the tectonic setting of the predicted samples in the study area. This method not only considers more variables, but also uses the quantitative tectonic setting probability to measure the qualitative structural environment discrimination problem; so it has more advantages over the traditional method.

In order to define the geological and tectonic evolution of the Xingmeng orogen in carboniferous, this paper collected a total of 6554 basalt samples from two types of tectonic settings, island arcs and rifts from all over the world, and established a training model. Then 31 basic carboniferous rock samples from Xingmeng orogen in Inner Mongolia were collected to predict the tectonic setting. The prediction results show that most of the samples are of island arc tectonic setting, and the predicted probability distribution of island arc is between 0.65 and 1. This result provides new evidence for the existence of a ditch - arc - basin system in the Carboniferous of Xingmeng orogenic belt.

The results of this paper demonstrate that the machine learning method can be used as an effective tool for processing high-dimensional geochemical data, and the machine learning methods can contribute to the study of geological processes.

## 2. Geological background

In terms of tectonic location, the Xingmeng orogenic belt is located in the southeastern part of the Central Asian Orogenic Belt between the North China Plate and the Siberian Plate, which is an important part of the Central Asian Orogenic Belt. Since the Early Paleozoic, the region has experienced multiple periods of oceanic slab subduction and collision, until the Paleo-Asian Ocean closed (Xu et al., 2018). There are many ophiolitic melange belts developed in the Xingmeng orogenic belt, from west to east, from south to north, the order is Suolunshan-Mandula,

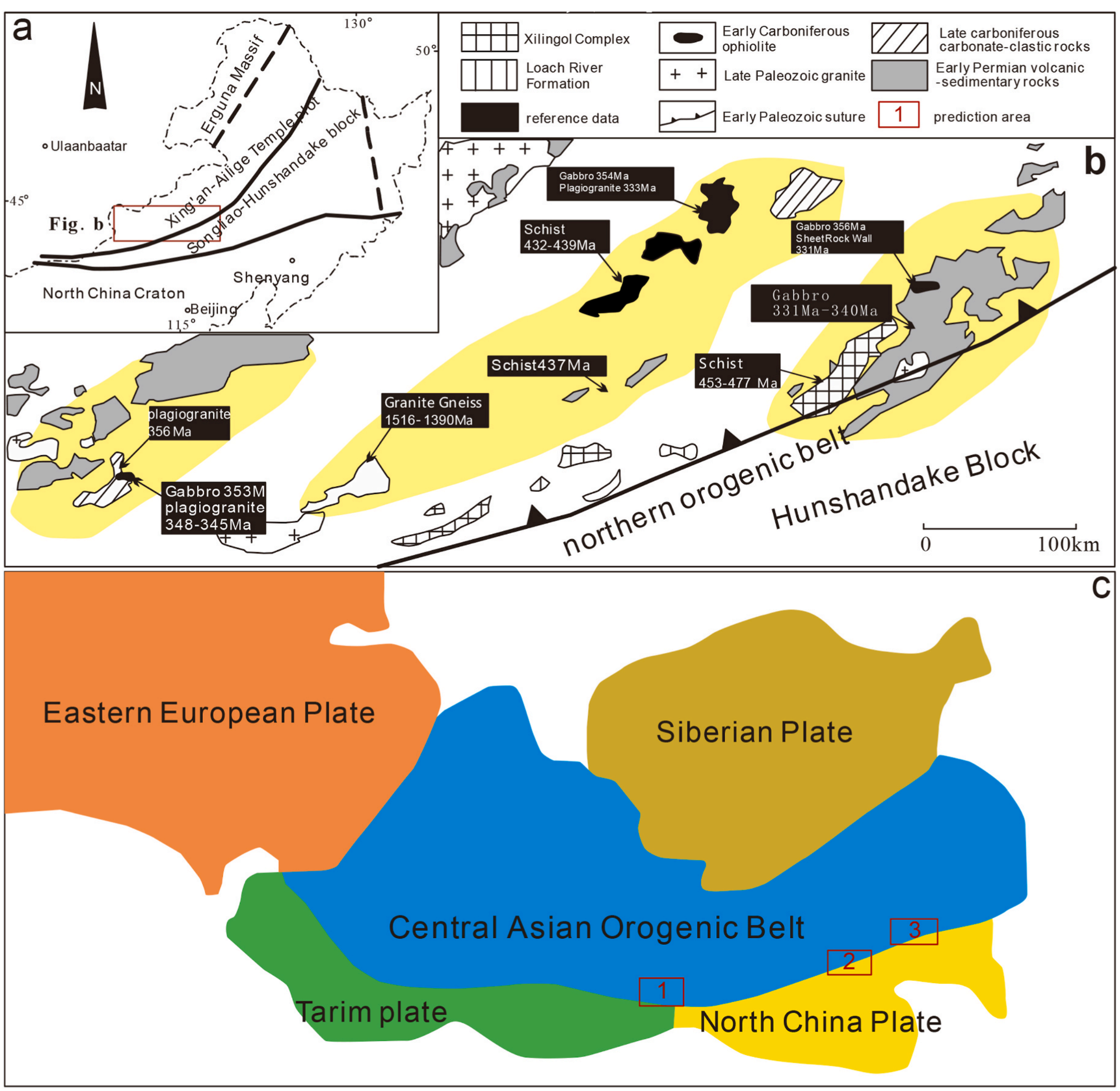


Fig. 1. Sketch geological map of the Xingmeng orogenic belt (Xu et al., 2018).

Wenduermiao-Tulinkai, Kedanshan-Xiaoweitang, Sunitezuoqi-Xilinhot and Erenhot-Hegenshan etc. ophiolite belts (Li et al., 2012; Li et al., 2013; Li et al., 2015; Liu et al., 2006).

The Carboniferous magmatic activity in the Xingmeng orogenic belt is relatively strong, but it is mainly moderately acidic (Cheng et al., 2012; He et al., 2013; Li et al., 2014, 2015). In order to study the carboniferous tectonic setting of the Xingmeng orogenic belt, this paper systematically collects the geochemical data of basic rocks in 3 areas, respectively: 1) Dongwuqi area of Inner Mongolia; 2) Keyouzhongqi area of Inner Mongolia; 3) Beishan area in the southern section of the Central Asian orogenic belt. Each geological profile is as follows (Fig. 1):

1) Dongwuqi region, Inner Mongolia

Located in the Bayandulan area in the west of Dongwuqi, the tectonic location is on the north side of the Erlian - Hegenshan tectonic junction and the southeastern margin of the Siberian plate (Fig. 2). The area is covered by grassland, with many Quaternary systems, and the continuous exposure of strata is poor, but the Paleozoic and Mesozoic are distributed to varying degrees. The late Paleozoic tectonic magmatism is very active in the area. The magmatic rock belt extends westward through Erlianhot into Mongolia intermittently, and extends eastward into the Greater Khingan Mountains area. The exposed strata in the area include the middle and lower devonian Niqiuhe Formation, whose lithology is mainly gray-green and light gray metamorphic siltstone and yellow-gray and gray-green metamorphic argillaceous siltstone; the upper Carboniferous Baoligaomiao Formation, which is mainly composed of continental volcanic rocks - clastic rocks, mainly purple-brown andesite and andesite pyroclastic rocks; the lower Jurassic Hongqi Formation is mainly composed of polymictic conglomerates, gravel-bearing coarse sandstone, and feldspar lithic sandstone, which is angularly unconformable over late Paleozoic rock strata. The intrusive rocks in the study area are mainly composed of late Carboniferous-Early Permian flesh-red medium-fine-grained alkaline granite, medium-fine-grained quartz syenite, and gray-white medium-fine-grained monzonite, which intruded into the middle and lower devonian Niqiuhe Formation. and the Baoligaomiao Formation. This paper mainly focuses on the amphibolites in the Suihe Chagan Ural basic complexes, which are banded inclusions located in the early Permian gray-white medium-fine-grained monzogranites. The zircon U -P b age of the amphibolite dating samples obtained by previous studies is 310 ± 1 Ma, indicating that the pluton was formed in the Late Carboniferous. This paper collects data on amphibolites in the Suihe Chagan Ural basic complexes for machine learning and tectonic context prediction.

2) Keyouzhongqi area, Inner Mongolia

The exposed strata in this area includes the Lower Permian Dashizhai Formation, the Upper Permian Linxi Formation, the Upper Jurassic Manketou Ebo Formation, the Upper Jurassic Manitu Formation, and the Lower Cretaceous Baiyingaolao Formation, etc.. In addition, Carboniferous tectonic melanges and Triassic, Jurassic and Cretaceous granites are also developed in this area. The fault structures in the area are mainly NE - trending and NNE -trending, followed by NW - trending and NNW -trending. The Paleozoic strata in the study area are generally superimposed with different degrees of ductile deformation.

The research objects in this paper are the Duerji mafic rocks and the Jiahada basic rocks, which are respectively exposed in the town of Duerji in the north of the study area and the Jiahada in the south, both of which are located in the tectonic melange belt. In the melange belt, dense cleavage, catatization and mylonitization of different intensities generally occur (Fig. 3). The outcrop area of the Duerji mafic rock is extremely limited, and it is only found on both sides of artificial mining pits and highway walls. The rock types are mainly basalt and diabase, which are intruded by Late Triassic granite. The Jiahada basic rock is exposed in the mining pit on the north side of the northern highway of Jiahada Gacha; the east side is in contact with the metamorphic siltstone fault, and the west side is in contact with the Late Jurassic volcanic rock fault. Previous studies have shown that the basic lithology of Jiahada is mainly basalt, and the zircon U–Pb age is 317.6 ± 3.0 Ma (Jin et al., 2022), indicating that it was formed in the Late Carboniferous. This paper collects Jiahada basalt data for machine learning and tectonic setting prediction.

3) Beishan area, southern section of the Central Asian Orogenic Belt

The geotectonic structure of the study area belongs to the northern margin of the Tarim plate and the southeastern margin of the Kazakhstan plate. The strata in the area are exposed from the Archaean to Cenozoic, mainly including the Neoarchean-Paleoproterozoic Dunhuang rock group, the Changchengian Gukejing group, the Jixian Jipingtoushan Formation, the Qingbaikou Period Daguoluoshan Formation, and the Nanhua Sinian Xichangjing Group, Cambrian Shuangyingshan Formation and Xishuangyingshan Formation, Ordovician Luoyachushan Formation, Huaniushan Group and Xilin Kebo Formation, Baiyunshan Formation, Silurian Heijianshan Formation and

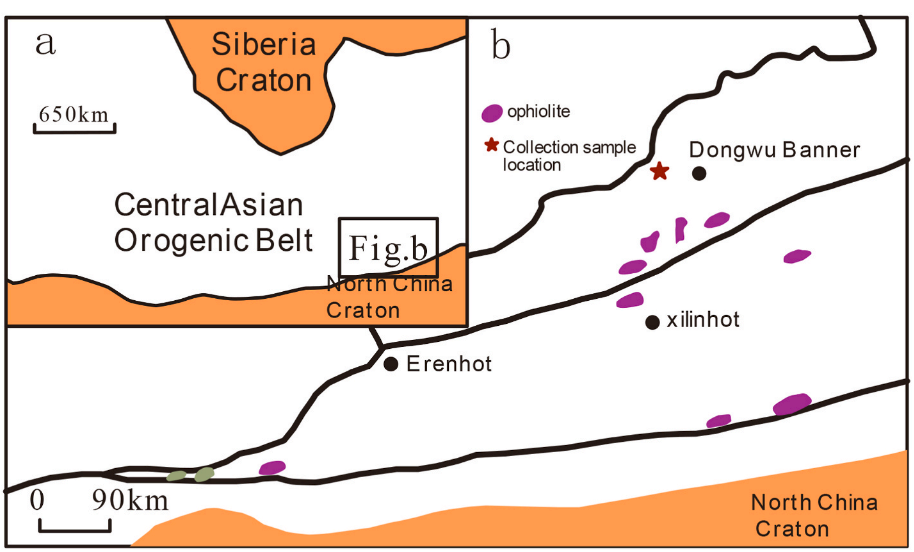


**Fig. 2.** Sketch geological map of the Dongwuqi area, Inner Mongolia (Qian et al., 2020).

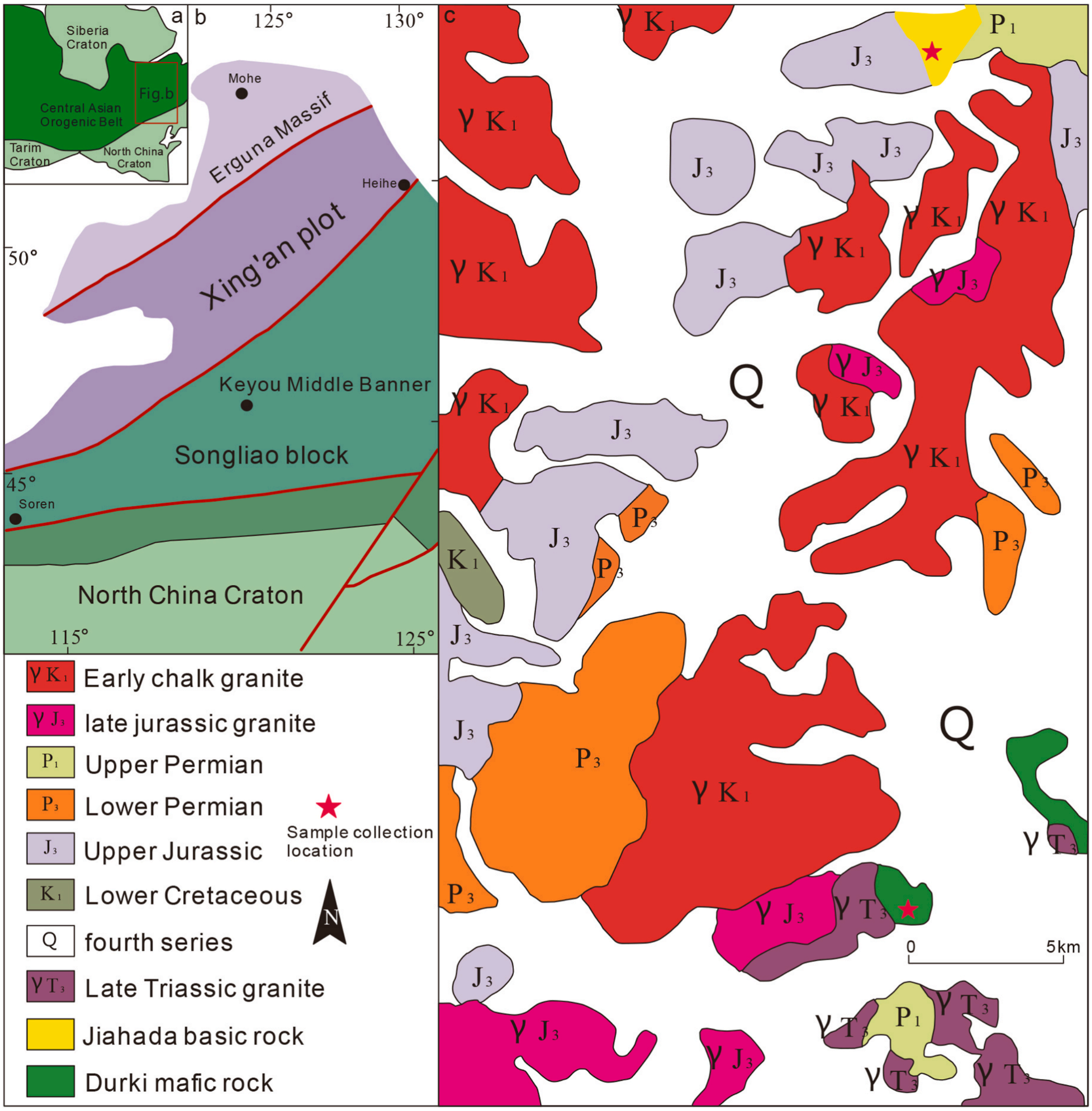


**Fig. 3.** Sketch geological map of the Bainaimiao area, Inner Mongolia (Qian et al., 2020).

Gongpoquan Formation group, Devonian three well groups and Dundunshan group, Carboniferous Hongliuyuan Formation, Permian Shuangbaotang Formation and Hongyanjing Formation, Triassic Erduanjing Formation, Neogene Kuquan Formation and Quaternary (Fig. 4).

The exposed range of Carboniferous strata is limited, and only a small area of Hongliuyuan Formation is exposed in the southeast corner of the study area. Its lithology is mainly a set of normally deposited clastic rocks intercalated with carbonate rocks and volcanic rocks. The lower part is dominated by volcanic rocks with a small amount of clastic rocks, and the lithology is gray-black basalt, gray-green almond-shaped spilite, light gray-light gray-green rhyolite mixed with gray thin-layered feldspar sandstone and siltstone; the upper part is a a rhythmic deposition composed of gravel-coarse sandstone, sandstone, siltstone, etc. Among them, the LA-ICP-MS zircon U–Pb dating of the lower spilite shows its formation age to be 359.9 ± 1.4 Ma, indicating that it was formed in the Early Carboniferous (Pan et al., 2017). This paper mainly collects spilite data for machine learning and tectonic setting prediction.

## 3. Data and methods

### 3.1. Data sources

The whole rock composition data used in this paper to train the Random Forest (RF) and Deep Neural Network (DNN) models are extracted from basalt data downloaded from the GEOROC Geochemical Data website http://georoc.mpch-mainz.gwdg.de/georoc/. A total of 28,366 samples of global island arc basalt and rift basalt were obtained. The predicted samples in the study area are all samples of Carboniferous basic rocks (including amphibole, basalt and spilite).

This paper uses Matlab programming to delete the elements with more than 20,000 missing elements in the sample data, leaving 45 elements, but the data still has the following problems: (1) there are still many elements missing in the sample data; (2) full iron (FEOT) needs to be calculated by formula $0.9*Fe_2O_3+FeO=FeOT$; (3) the geochemical data elements downloaded from the reference are limited, and the element types of the training data must be consistent with the predicted data (geochemical data downloaded from the reference), so heavy data cleaning work is needed to obtain valid geochemical data. According to

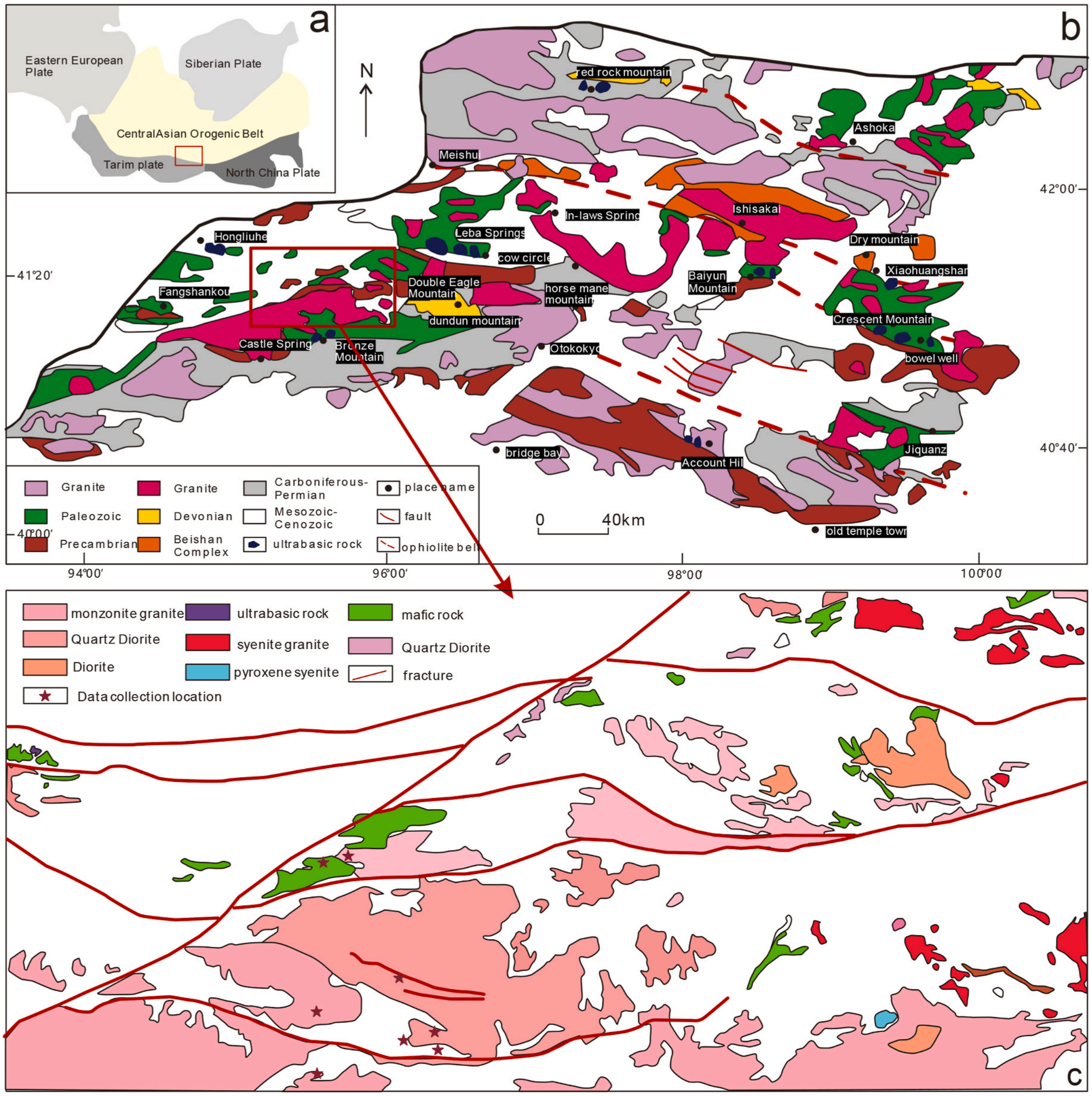


**Fig. 4.** Sketch geological map of the Bainaimiao area, Inner Mongolia (Qian et al., 2020).

the above problems, comprehensively considering the types of data that can be downloaded from the references, on the basis of 28366 total samples, 27 kinds of elements (Table 1) are selected in a targeted manner. Under the condition that all the data samples of each element are not missing, a total of 6554 basalt samples are obtained from two types of tectonic settings, island arc and rift, including 5367 samples from island arc (convergent margin) and 1187 samples from rift volcanics. The 27 elements including SiO2, TiO2, Al203, FEOT, MnO, MgO, CaO, Na20, K20, P2O5, Rb, Y, Nb, Ba, La, Ce, Nd, Sm, Eu, Tb, Yb, Lu, Hf, Ta, Th, U etc. are selected as model variables, using random forest algorithm for model training.

**Table 1**
The roc curve parameter table.

| Algorithm | AUC area | Standard error | Progressive significance | Progressive 95% confidence interval | |
|---|---|---|---|---|---|
| | | | | Lower limit | Upper limit |
| RF | 0.978 | 0.003 | 0.000 | 0.971 | 0.985 |
| DNN | 0.838 | 0.008 | 0.000 | 0.822 | 0.854 |

On the basis of model training, this paper extracts prediction samples from relevant references. The prediction samples are based on 31 Carboniferous basic rocks, including: 6 amphibolite samples in Dongwuqi area, Inner Mongolia; there are 11 basalt samples in the Keyouzhongqi area; 14 spilite samples in the Beishan area in the southern section of the Central Asian Orogenic Belt.The distribution of predicted data in the research area is shown in Fig. 5.

### 3.2. Methods

This paper uses matlab2016 software programming to implement data cleaning, uses python combined with sklearn third-party library to jointly implement the random forest (RF) model and Deep neural network(DNN) model, and then predicts the tectonic background of 31 prediction basalt samples in the Xingmeng orogenic belt area.

#### 3.2.1. Random forest method

In this paper, in the process of random forest (RF) model, we firstly train the model, then use the trained model to predict, and realize the classification of basic rock tectonic setting of test samples, and calculate

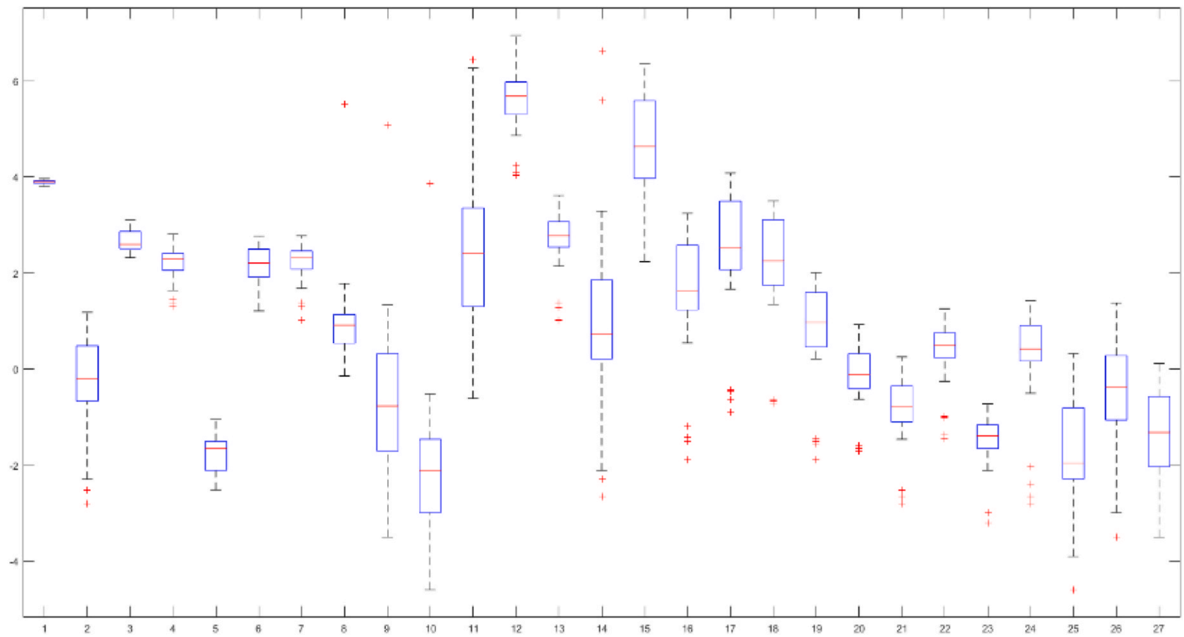

**Fig. 5.** Box map of distribution of geochemical element data
(1.SiO2; 2.TiO2; 3.Al2O3; 4.FeOT; 5.MnO; 6.MgO; 7.CaO; 8.Na2O; 9.K2O; 10.P2O5; 11.Rb; 12.Sr; 13.Y; 14.Nb; 15.Ba; 16.La; 17.Ce; 18.Nd; 19.Sm; 20.Eu; 21.Tb; 22. Yb; 23.Lu; 24.Hf; 25.Ta; 26.Th; 27.U).

the importance of different elements. Random forest(Athey et al., 2019) belongs to an ensemble learning algorithm(Choi et al., 2020). The basic idea of the algorithm is to use the Bootstrap sampling method to perform a sampling operation with replacement from the original data set, build a decision tree from the sampled original data subset, and combine multiple decision trees into a random forest. And the mean value of the constructed decision tree is finally used as the result of the random forest regression prediction. The specific steps of the algorithm are that each decision tree in the random forest algorithm model (Fig. 6) contains a tree-like sequence of decision nodes. Based on this sequence, the tree is split into various branches until it reaches the end (leaf) of the tree. The prediction results of each decision tree are output through leaf nodes, and finally, the outputs of multiple decision trees are combined for prediction. The random forest algorithm has the advantages of fast training speed and avoiding overfitting etc.

### 3.2.2. Deep neural network method

Deep neural network (DNN) is a framework of deep learning. It is a neural network with at least one hidden layer. Similar to shallow neural networks, deep neural networks can also provide modeling for complex nonlinear systems, but the extra layers provide a higher level of abstraction for the model, thus improving the ability of the model. A deep neural network is a discriminant model that can be trained using a back propagation algorithm.

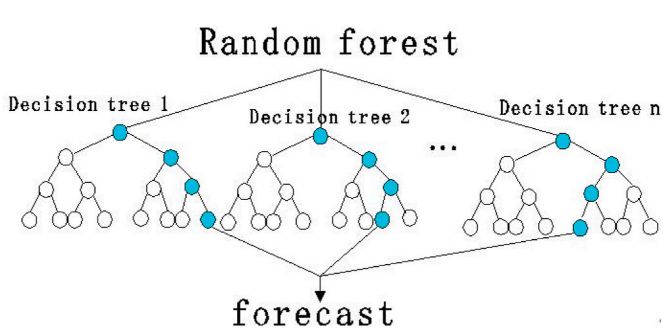


**Fig. 6.** Random Forest model.

The whole rock composition data used in this paper to train Deep Neural Network (DNN) models are extracted from basalt data downloaded from the GEOROC Geochemical Data website http://georoc.mpch-mainz.gwdg.de/georoc/. The training sets and prediction sets of deep neural network model and random forest model are consistent.

### 3.2.3. Comparative study of stochastic forest model and deep neural network model

In this paper, ROC curves were made based on the prediction results of random forest model and deep neural network model. The AUC area of the random forest model was 0.978 and that of the deep neural network model was 0.838. According to the AUC area, the random forest model had high accuracy(Fig. 7 and Table .1).

Through F1 value and confusion matrix analysis(Table .2, Table .3, Table .4), it can be seen that the random forest model is more accurate than the deep neural network model.

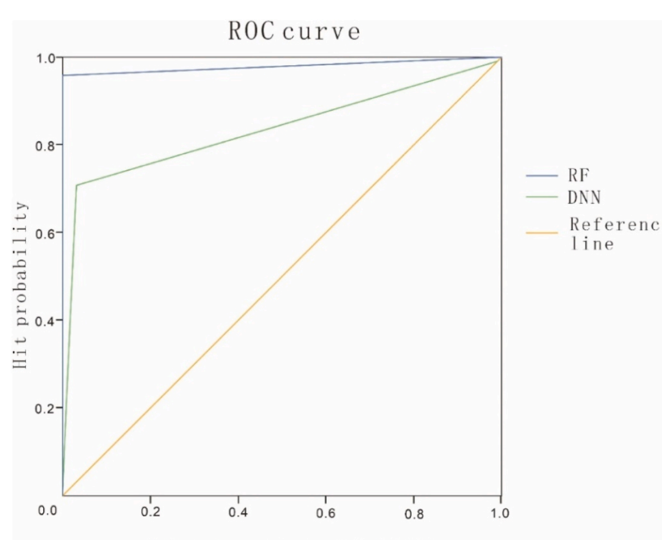


**Fig. 7.** The roc curves of RF and DNN.

**Table 2**
The table of F1 values for RF and DNN models.

| Algorithm | Category | Precise | Recall | F1 |
|---|---|---|---|---|
| DNN | Island arc | 0.9379 | 0.9648 | 0.9511 |
| | Continental rift | 0.8170 | 0.7110 | 0.7604 |
| RF | Island arc | 0.9906 | 0.9991 | 0.9948 |
| | Continental rift | 0.9956 | 0.9570 | 0.9759 |

**Table 3**
The confusion matrix of DNN model.

| category | Island arc | Continental rift |
|---|---|---|
| Island arc | 5178 | 189 |
| Continental rift | 343 | 844 |

**Table 4**
The confusion matrix of RF model.

| category | Island arc | Continental rift |
|---|---|---|
| Island arc | 5362 | 5 |
| Continental rift | 51 | 1136 |

According to the AUC area of the roc curves of RF and DNN models, F1 value and confusion matrix of ROC curve of DNN and RF models, it can be seen that RF model has higher accuracy than DNN and is more suitable for the prediction of geological structure background. Therefore, RF model is adopted in this paper to predict the geological structure background of the study area.

## 4. Results

In this paper, random forest and principal component analysis are used to clean the data, analyze the characteristics of the preprocessed island arc and ocean island basalt, and then calculate the importance of each category of features, and finally obtain tectonic setting probability of the island arc and rift of each predicted sample, and take the one with the highest probability as the tectonic setting of the predicted sample.

In this paper, the above 5043 basalt samples (27 elements for each sample) are trained, and the training accuracy calculation adopts the cross-validation method, that is, the training data is divided into K groups (K-Fold), and each subset data is verified once set, and the rest of K-1 subsets of data are used as training sets, so that K models will be obtained. The K models are evaluated in the validation set respectively, and the final error MSE (Mean Squared Error) is added and averaged to obtain the cross-validation error. In this paper, K is set as 5, and the training accuracy of the random forest model reaches 81%. Model training converges faster, as shown in Fig. 8.

In this paper, the hyperparameters tuning range of random forest model are shown as follows. 1.imators:[200,300,600,1200,2400] (Number of trees to grow). 2.max_features: [2,4,6,8] (Number of variables randomly sampled as candidates at each split). 3.min_samples_leaf: [1,3,6] (Minimum size of terminal nodes). After grid search, the best parameters were obtained. 1. n_estimators = 1200; 2. max_depth = 21; 3.max_features = 8; 4.min_samples_leaf = 1. The training accuracy reached 0.9312.

The overall importance of the geochemical elements is ranked as follows (Fig. 9): TiO2, Ta, Nb, FeOT, SiO2, La, Eu, Al2 O3, CaO, MnO, Sr, Ce, Ba, Nd, Na2 O, Sm, Hf, Y, Rb, K2 O, Th, U, Yb, P2 O5, Lu, MgO, Tb.

The importance of geochemical elements in the island arc tectonic setting (as shown in Fig. 10): $SiO_2$ (+), $TiO_2$ (-), $Al_2O_3$ (+), FeOT (+), MnO (-), MgO (+), CaO (+), $Na_2O$ (+), $K_2O$ (-), $P_2O_5$ (-), Rb (+), Sr (+), Y (+), Nb (+), Ba (+), La (+), Ce (+), Nd (+), Sm (+), Eu (-), Tb (-), Yb (-), Lu (-), Hf (-), Ta (-), Th (-), U (-),' + 'and' – 'signs represent positive and negative principal component loading values; respectively, positive values represent greater importance to island arcs, and negative values represent less importance to island arcs, the same below.

The importance of geochemical elements in the rift tectonic setting (Fig. 11): $SiO_2$ (+), $TiO_2$ (-), $Al_2O_3$ (+), FeOT (+), MnO (-), MgO (+), CaO (+), $Na_2O$ (+), $K_2O$ (-), $P_2O_5$ (-), Rb (+), Sr (+), Y (+), Nb (+), Ba (+), La (+), Ce (+), Nd (+), Sm (+), Eu (-), Tb (-), Yb (-), Lu (-), Hf (-), Ta (-), Th (-), U (-). It can be seen that the positive and negative component loading values of the geochemical elements in the rift tectonic setting are similar to those of the island arc tectonic setting, but the degree is different. The training results of these models show that different elements play different roles in the process of judging the tectonic environment of basalt. For example, Nb is an important element in distinguishing the tectonic settings of island arc and rift valley (Figs. 10 and 11).

The number of decision trees in the random forest classification method used in this paper is set as 200. The final vote of the decision tree is to select the category with the highest average probability. The predicted probability of each sample class represents the probability of each sample classification in the random forest. After establishing the training model based on random forest algorithm, this paper collects the forecast samples from the references and the forecast results are as follows (Table 1): (1) 6 forecast samples in the Dongwuqi area of Inner Mongolia; all the forecast results are island arcs, and most predicted probabilities are distributed between 0.65 and 0.995. (2) 11 prediction samples in Keyouzhongqi area of Inner Mongolia; all the prediction results are island arcs, and the prediction probability is between 0.83 and 1. (3) 14 prediction samples in the Beishan area in the southern section of the Central Asian Orogenic Belt; all the prediction results are island arcs, and the prediction probability is between 0.84 and 1. The total forecast results are shown in Table 1.

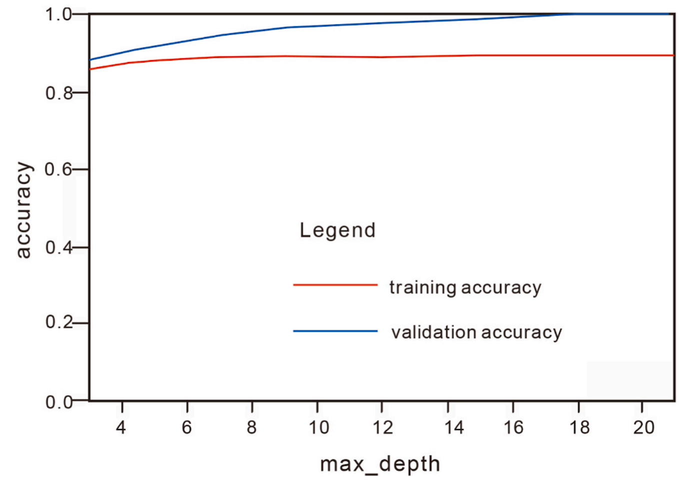


**Fig. 8.** Random Forest model training curve.

## 5. Discussion

In recent years, scholars have made a lot of progress and achievements in the research on the evolution of the Xingmeng orogenic belt. However, the evolution of geological structures has become extremely complex in the Xingmeng orogenic belt, due to the possible influence of the superposition and transformation of the Paleo-Asian Ocean, the Mongolian-Okhotsk Ocean and the Paleo-Pacific tectonic domain. Among them, the most controversial issue is the time limit for the closure of the Paleo-Asian Ocean. There are mainly two different views. The first view is that the Paleo-Asian Ocean developed two branches, the south and the north. The north branch subducted NW from the Early Paleozoic to the Late Carboniferous, and then closed gradually; the southern branch developed bidirectional subduction from the Early Paleozoic to the end of the Late Paleozoic and finally closed. (Liu et al.,

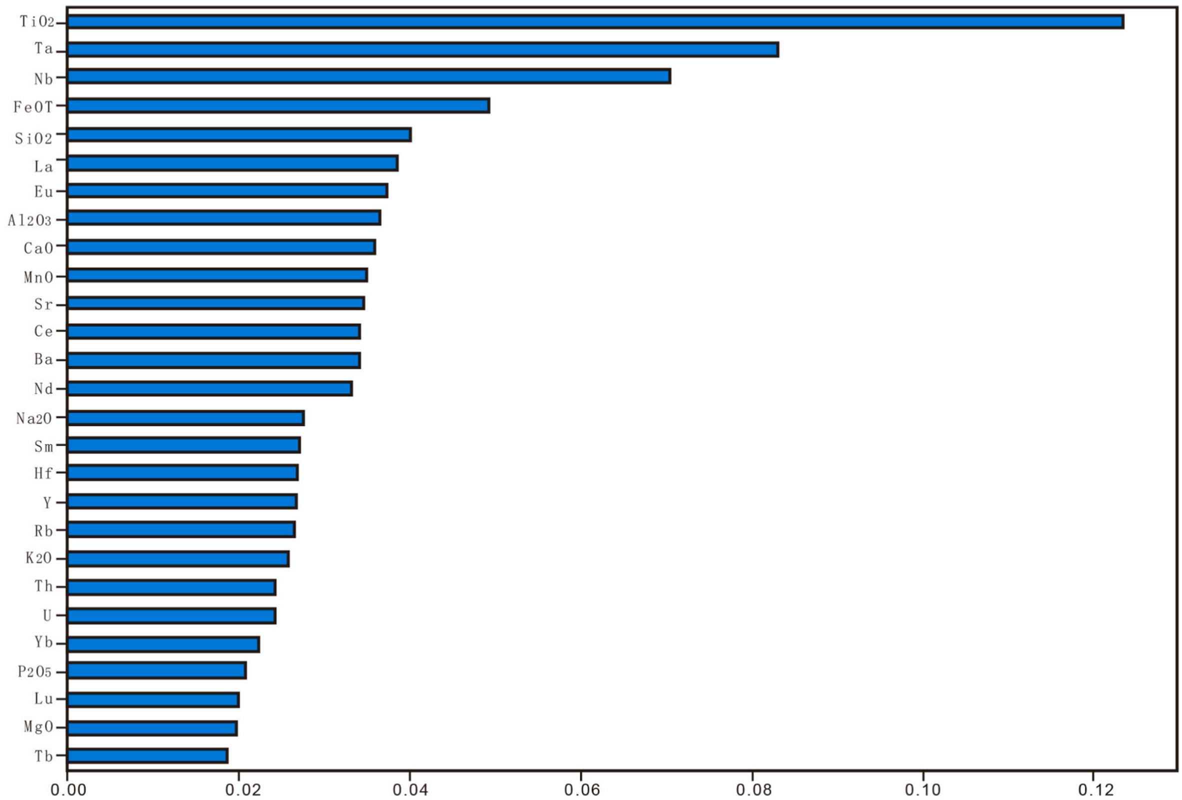


**Fig. 9.** Element importance graph.

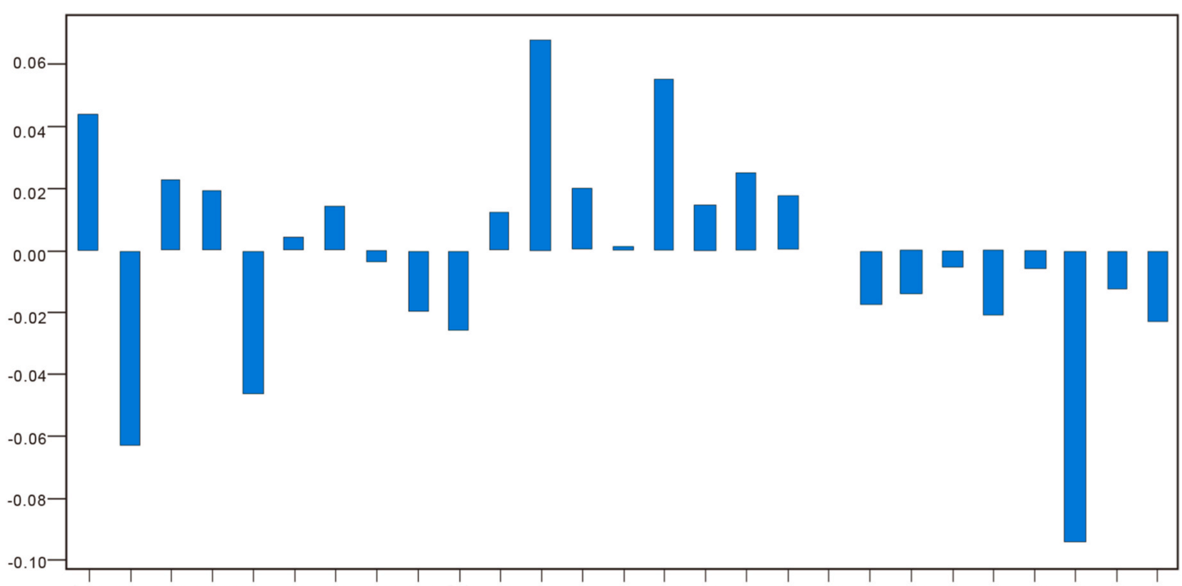


**Fig. 10.** Importance graph of geochemical elements in island arcs.

2017). Some scholars believe that the Paleo-Asian Ocean developed continuously from south to north subduction to the end of the early Triassic and finally formed the Sauron - Sylamulun suture belt (Liu et al., 2006; Xiao et al., 2003; Xiao et al., 2009; Xiao et al., 2015). Although there are disputes over the subduction model, both views show that a trench - arc - basin system existed in the Paleo-Asian Ocean during the Carboniferous period.

In contrast, another view shows that the continental rift tectonic setting, that is, the Paleo-Asian Ocean had subducted and disappeared during this period, and entered the orogenic belt stage; that is, it was in an extensional tectonic environment (He and Shao, 1983; Shao, 1989; Tang, 1990; Shao, 1991; Tang, 1992; Xu et al., 2014; Zhao et al., 2015; Zhao et al., 2016; Zhao et al., 2017). In the newly compiled 1:5000000 International Geological Map of Asia, Ren Jishun et al. Ren Jishun proposed that the Carboniferous Permian volcanic rocks in Central Asia were formed in an extensional rift environment rather than an island arc environment. Therefore, the key to the above debate is whether the ocean was closed, and whether there was still subduction during the

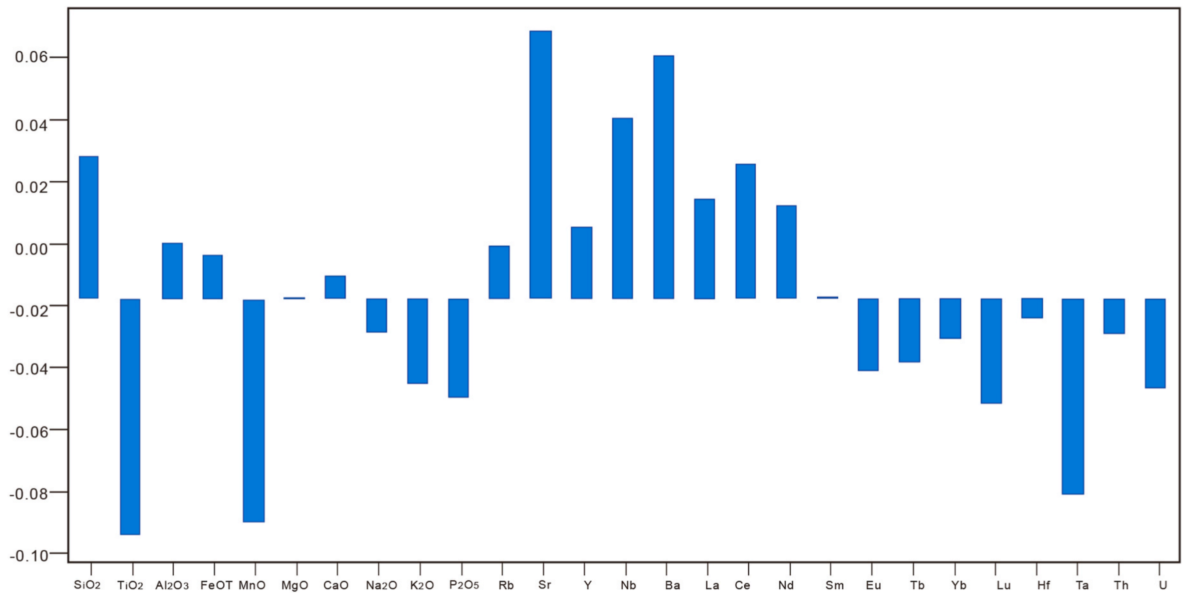


**Fig. 11.** Importance graph of geochemical elements in rift valleys.

Carboniferous period.

In general, the Carboniferous magmatic activity in the Xingmeng orogenic belt is relatively strong, which provides a good condition for us to solve the closure time limit of the Paleo-Asian Ocean. Taking the Erlian-Hegenshan and Linxi fault ophiolitic melange belts as the boundaries, the Xingmeng orogenic belt is divided into three magmatic rock belts, showing distinct rock assemblages from east to west: (1) Erlian-East Wuqi magmatic rock belt in the north is dominated by granodiorite and high-potassium calcalkaline monzonite (Cheng et al., 2012; He et al., 2013; Li et al., 2014, 2015) and it is the southern Mongolia granodiorite, calcalkaline gabbro diorite, and granite belt spreading westward, (Hong et al., 1994; Tong et al., 2015). (2) In the central part of Suzuoqi-Xiwuqi magmatic rock belt, there are rock assemblages of quartz diorite, granodiorite, dolomitic diorite, and basalt, mainly distributed in low-potassium puller series and calc-alkaline volcanic rock assemblages (Liu et al., 2009; Shi et al., 2004; Bao et al., 2007; Zhang et al., 2017). (3) The southern part is the late Paleozoic magmatic rock belt in the northern margin of North China, mainly distributed in quartz diorite, monzogranite, granodiorite, diorite, etc., roughly parallel to the northern margin of the North China block, and it is distributed in bands showing an east-west direction (Zhang et al., 2004, 2007; Zhou et al., 2009). It can be seen that although Carboniferous magmatic rocks are widely developed in the Xingmeng orogenic belt, they are dominated by medium-acid lithology, while the distribution of basic rocks is relatively limited. However, compared with intermediate-acid rocks, basic rocks can more effectively constrain their genetic tectonic setting.

In this study, the method of random forest in machine learning was used to train 27 major and trace elements in 5367 basalt samples from island arc tectonic environments and 1187 from continental rift valleys in GEOROC geochemical data. Based on the training model thirty-one basic rock samples from four regions of the Central Asian Orogenic Belt and the Xingmeng Orogenic Belt were used for prediction. Among them, the prediction results in the Dongwuqi area and Keyouzhongqi area of Inner Mongolia show that the island arc probability is relatively high and relatively concentrated, which clearly indicates that these basic rocks have a high probability of forming in the island arc environment. The Beishan area in southern section of Central Asian orogenic belt also shows a high and uniform island arc probability, indicating that the subduction is probably not over yet, and it is also in the tectonic environment of the island arc. The age of the samples in this paper spans the early and late Carboniferous period, and the region spans the Xingmeng orogenic belt and the southern section of the Central Asian orogenic belt. This temporal and spatial distribution indicates that there was still extensive subduction in the Paleo-Asian Ocean during the Carboniferous period. Therefore, the results of this paper support the view that oceanic subduction existed in the Paleo-Asian Carboniferous. This is also supported by the following two aspects: 1. More and more research results have confirmed the existence of Carboniferous- Early Permian ophiolite (Xiao et al., 2009), indicating that the Paleo-Asian Ocean still existed in the Carboniferous to Early Permian (Early Middle Permian); 2. The pattern of the late Carboniferous adakite in the Erlian-Hegenshan suture zone indicates that the Paleo-Asian Oceanic Carboniferous-Early Permian was in the process of oceanic subduction and extinction characterized by intra-ocean subduction (Wang et al., 2021).

According to previous studies, the Paleo-Asian Ocean has the characteristics of “scissor-like” closure from west to east, that is, the closure time of the western segment is relatively early, and the closure time of the eastern segment is relatively late (Xu et al., 2019). The research in this paper has important implications for the subduction age and intensity of the eastern and western segments. The predicted area is from west to east, from the Beishan area in the southern part of the Central Asian Orogenic Belt to the Dongwuqi area in Inner Mongolia. and the island arc probability has a downward trend from the early Carboniferous to the Late Carboniferous. The island arc probability mainly reflects the degree of fluid participation in the source area, while the magma source area of continental rift basalt often has less fluid participation. Therefore, the magmatism in the Beishan area of the southern part of the Central Asian orogenic belt in the Early Carboniferous may have formed in the heyday of subduction, while the Xingmeng orogenic belt in the Late Carboniferous may have been in the late subduction stage to the collision or even the early rifting stage. This temporal and spatial evolution shows that the subduction of the Paleo-Asian Ocean is different from west to east.

## 6. Conclusion

Through machine learning in this paper, the prediction results show that most of the predicted values in the Carboniferous Inner Mongolia Dongwuqi area, Keyouzhongqi area, and Beishan area in the southern section of the Central Asian Orogenic Belt are island arc tectonic settings, and the island arcs probabilities of most of the predicted samples are all between 0.65 and 1. Therefore, it is inferred that the Carboniferous Xingmeng orogenic belt was formed in an active continental margin environment and belongs to the island arc tectonic setting.

Machine learning approaches reduce biases caused by human subjectivity by automatically searching for relationships in large scale Spaces. The results show that the random forest model can obtain tectonic setting information, and the random forest model can also demonstrate the potential application value of machine learning methods in processing high-dimensional geochemical data. In addition, the random forest model can capture key characteristic attributes in the classification process, making models interpretable and helping us to understand geological processes.

The advantage of the new method used in this paper is that it can predict the tectonic setting more accurately, and it can obtain the importance map of each element (chemical composition), which can open up ideas for further research on the origin of magmatic rocks. It can be promoted in the field of geology, and it can promote the further development of geological big data experimental technology.

### Authors' contributions

Zhen Chen conceived the analysis and wrote the manuscript; Qingsong Wu conceptualized and critically revised the manuscript; Sipeng Han provided scientific drawing; Jungui Zhang, Peng Yang, Xingwu Liu provided technical support on the method and critically revised the manuscript.

### Declaration of competing interest

There are no conflicts of interest to declare.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aiig.2023.01.004.

### References


Athey, S., Tibshirani, J., Wager, S., 2019. Gemeralized random forests. Ann. Stat. 47 (2), 1148–1178.

Bao, Q., Zhang, C., Wu, Z., 2007. Zircon SHRIMPU-Pb dating of granitic rocks in the late Paleozoic rift region in southeastern Inner Mongolia and its geological significance. Chinese Geol. 34 (5), 790–798 (in Chinese with English abstract).

Cheng, B., Teng, X., Xin, H., 2012. SHRIMP zircon U-Pb age of the Runmai Wendur granite in Dongwu Banner, Inner Mongolia and its geological significance. J. Rock Mineral. 31 (3), 323–334 (in Chinese with English abstract).

Choi, J., Gu, B., Chin, S., 2020. Machine learning predictive model based on national data for fatal accidents of construction workers. Autom. ConStruct. 110, 102974 https://doi.org/10.1016/j.autcon.2019.102974.

He, F., Xu, J., Gu, X., 2013. Age, genesis and geological significance of the Amuguleng complex granite body in East Ujumqin Banner, Inner Mongolia. Geol. Rev. 59 (6), 1150–1164 (in Chinese with English abstract).

He, G., Shao, J., 1983. Confirmation of the Early Paleozoic Ophiolite Formation along the Xilamulun River in Southeastern Inner Mongolia (Zhaomeng) and its Tectonic Significance. See: Tang Kedong, Editor-In-Chief. Anthology of Plate Tectonics in Northern China. Geological Press, Beijing, pp. 243–249 (in Chinese with English abstract).

Hong, D., Huang, H., Xiao, Y., 1994. Permian alkaline granite in central Inner Mongolia and its geodynamic significance. Chin. J. Geol. 68 (3), 219–230 (in Chinese with English abstract).

Jin, S., Li, Y., Qian, C., 2022. Chronology, geochemistry and tectonic significance of two types of mafic rocks in Keyouzhongqi, Inner Mongolia. Earth Sci. 47 (1), 342–356 (in Chinese with English abstract).

Li, K., Zhang, Z., Feng, Z., 2015a. Two periods of late Paleozoic magmatic activity and its tectonic significance in the northern middle section of the Xingmeng orogenic belt. Chin. J. Geol. 89 (2), 272–288 (in Chinese with English abstract).

Li, K., Zhang, Z., Feng, Z., 2014. Zircon SHRIMPU-Pb dating of Late Carboniferous-Early Permian volcanic rocks in the Bayan-Ula area of central Inner Mongolia and its geological significance. J. Petrol. 30 (7), 2041–2054 (in Chinese with English abstract).

Li, Y., Wang, J., Li, H., 2015b. Identification of melautowula ophiolite in Xiwu Banner, Inner Mongolia. Chin. J. Petrol. 31 (5), 1461–1470 (in Chinese with English abstract).

Li, Y., Wang, J., Li, H., 2013. Geochemical characteristics of the Baiyinbao ophiolite in Xiwu Banner, Inner Mongolia. Chin. J. Petrol. 29 (8), 2719–2730 (in Chinese with English abstract).

Li, Yi, Wang, J., Li, H., 2012. Identification of diyanmiao ophiolite in Xiwu Banner, Inner Mongolia. Chin. J. Petrol. 28 (4), 1282–1290 (in Chinese with English abstract).

Liu, J., 2009. Late Paleozoic Magmatism in the Linxi-Dongwuqi Area of Inner Mongolia and its Constraints on Regional Tectonic Evolution. Doctoral thesis. Changchun Gulin University, China, pp. 1–149 (in Chinese with English abstract).

Liu, J., Zhang, T., Xu, L., 2006. Discovery and significance of late Paleozoic ultrabasic-basic rocks in Haoluchang area, Inner Mongolia. Geologic. Surv. Res. 29 (1), 21–29 (in Chinese with English abstract).

Liu, Y., Li, W.M., Feng, Z., Wen, Q., Neubauer, F., Liang, C., 2017. A review of the paleozoic tectonics in the easter part of central asian orogenic belt. Gondwana Res. 43, 123–148. https://doi.org/10.1016/j.gr.2016.03.013.

Qian, C., Lu, L., Wang, Y., 2020. Paleoproterozoic metamorphic basement discovered in Bainaimiao Island Arc, Inner Mongolia Evidence from the age and geochemistry of amphibolites in Shuangsheng area. Geol. Bull. (6) (in Chinese with English abstract).

Shao, J., 1991. Crustal Evolution in the Middle Section of the Northern Margin of the China-DPRK Plate. Peking University Press, Beijing, pp. 105–117 (in Chinese with English abstract).

Shao, J., 1989. Continental crust accretion and tectono-magmatic:activityat the northern margin of the Sino-Korean plate. J. Southeast Asian Earth Sci. 3 (1–4), 57–62. https://doi.org/10.1016/0743-9547(89)90009-3.

Shi, G., Miao, L., Zhang, F., 2004. Age and regional tectonic significance of A-type granites in Xilinhot, Inner Mongolia. Sci. Bull. 49 (4), 384–389 (in Chinese with English abstract).

Tang, K., 1992. Tectonic Evolution and Metallogenic Regularity of the Fold Belt on the North Side of the Sino-DPRK Plate. Peking University Press, Beijing, pp. 1–277 (in Chinese with English abstract).

Tang, K., 1990. Tectonic development of Paleozoic foldbelts athe northmargin of the SinoKorean Craton. Tectonics 9 (2), 249–260. https://doi.org/10.1029/TC009i002p00249.

Tong, Y., Jahn, B., Wang, T., 2015. Permian alkaline granites in the Erenhot-Hegenshan belt, northern InnerMongolia, China Model of generation, time ofemplacement and regional tectonic significance. J. Asian Earth Sci. 97, 320–336. https://doi.org/10.1016/j.jseaes.2014.10.011.

Ueki, K., Hino, H., Kuwatani, T., 2017. Geochemical discrimination and characteristics of magmatic tectonic settings A machine learning-based approach. G-cubed 19, 1327–1347. https://doi.org/10.1029/2017GC007401.

Wang, J., Li, Y., Li, H., Dong, P., 2021. Discovery of late carboniferous adakite in Erlian-Hegenshan suture zone and paleo-asian Intraoceanic subduction. China Geol. 48 (2), 520–535 (in Chinese with English abstract).

Xiao, W., Windley, B., Hao, J., 2003. Accretion leading to collision and the permian solonker suture, InnerMongolia, China teminationofthe CentralAsian orogenic belt. Tectonics 22 (6), 1484–1505. https://doi.org/10.1029/2002TC001484.

Xiao, W., Windley, B., Huang, B., Han, C., Yuan, C., Chen, H., Sun, M., Sun, S., Li, J., 2009. EndHPermian to MidTriassic terminationof the accretionary processes of the souther Altaids Implications forthe geodynamic evolution, Phanerozoic continental growth, and met allogeny of Central Asia. Int. J. Earth Sci. 98 (6), 1189–1217. https://doi.org/10.1007/s00531-008-0407-z.

Xiao, W., Windley, B., Sun, S., Li, J., Huang, B., Han, C., Yuan, C., Sun, M., Chen, H., 2015. A tale of amalgamation of three Permo-Triassiccollage systems in Central Asia: oroclines, sutures, and terminalaccretion. Annu. Rev. Earth Planet Sci. 43 (1), 477–507.

Xu, B., Charvet, J., Chen, Y., Zhao, P., Shi, G., 2013. Middle Paleozoic convergent orogenic belts in western Inner Mongolia (China) Framework, kinematics, geochronology and implications for tectonic evolution of the Central Asian Orogenic Belt. Gondwana Res. 23 (4), 1342–1364. https://doi.org/10.1016/j.gr.2012.05.015.

Xu, B., Chen, B., 1997. Framework and evolutionof the MiddlePaleozoic orogenic belt between Siberian and North China plates innorthern Inner Mongolia. Sci. China 40 (5), 463–469. https://doi.org/10.1007/BF02877610.

Xu, B., Wang, Z., Zhang, L., 2018. Xingmeng Intracontinental orogenic belt. Acta Petrol. Sin. 34 (10), 5–30 (in Chinese with English abstract).

Xu, B., Zhao, P., Bao, Q., Zhou, Y., Wang, Y., Luo, Z., 2014. Preliminary study on the division of pre-Mesozoic tectonic units in the Xingmeng orogenic belt. Acta Petrol. Sin. 30 (7), 1841–1857 (in Chinese with English abstract).

Xu, W., Sun, C., Tang, J., 2019. Basement properties and tectonic evolution of the Xingmeng orogenic belt. Earth Sci. 44 (5), 1620–1646 (in Chinese with English abstract).

Zhang, S., Zhao, Y., Song, B., 2007. Carboniferous granitie plutons from the northem margin of he North Chinablock: implications for a Late Paleozoic active continental margin. J. Geolog.Soc London 164, 451–463. https://doi.org/10.6084/M9.FIGSHARE.3454385.V1.

Zhang, S., Zhao, Y., Song, B., 2004. Late Paleozoic gneissic granodiorite-zircon SHRIMP U-Pb age and its tectonic significance in the Early Precambrian high-grade metamorphic zone in the uplift of northern Hebei. Acta 20 (3), 621–626 (in Chinese with English abstract).

Zhang, Z., Chen, Y., Li, K., 2017. Geochronology and geochemistry of Permian bimodal volcanic rocks from centralInner Mongolia. China Implications for he Late Palacozoic tectonic evolution ofthe south-easterm Cemtral Asianorogenic belt. J. Asian Earth Sci. 135, 370–389. https://doi.org/10.1016/j.jseaes.2017.01.012.

Zhao, P., Faure, M., Chen, Y., Shi, C., Xu, B., 2015. A new Triassic shortening-extrusion tectonic model for Central-Eastern Asia Structural geochronological and

paleomagnetic investigations in the Xilamulun Fault (North China). Earth Planet Sci. Lett. 426, 46–57.
Zhao, P., Jahn, B., Xu, B., Liao, W., Wang, Y., 2016. Geochemistry,geochronology and zircon Hf isotopic study of peralkaline-alkalineintrusions along the northern margin of the North China Craton andits tectonic implication for the southeastern Central Asian OrogenicBelt. Lithos 261, 92–108. https://doi.org/10.1016/j.lithos.2015.12.013.
Zhao, P., Xu, B., Zhang, C., 2017. A rift system in southeastern Central Asian Orogenic Belt Constraint from sedimentological geochronological andgeochemical investigations ofthe Late Carboniferous-Early Permian strata in northern Inner Mongolia (China). Gondwana Res. 47, 342–357. https://doi.org/10.1016/j.gr.2016.06.013.
Zhao, Y., Zhang, Y., Geng, M., Jiang, J., Zou, X., 2019. Involvement of slab derived fluid in the generation of cenozoic basalts in Northeast China inferred from machine learning. Geophys. Res. Lett. 46, 5234–5242. https://doi.org/10.1029/2019GL082322.
Zhou, Z., Zhang, H., Liu, H., 2009. Zircon dating of basic intrusive rocks in Siziwangqi area of central Inner Mongolia and its significance. Chin. J. Petrol. 25 (6), 1519–1528 (in Chinese with English abstract).