Acta Geophysica (2023) 71:681–696
https://doi.org/10.1007/s11600-022-00937-x

RESEARCH ARTICLE - APPLIED GEOPHYSICS

A Machine learning approach for the magnetic data interpretation of 2-D dipping dike

Sh Bronson Aimol[1] · G. Srinivasa Rao[1] · Thinesh Kumar[1] · Rama Chandrudu Arasada[1]

Received: 12 April 2022 / Accepted: 19 September 2022 / Published online: 14 October 2022

Abstract

2-D dipping dike model is often used in the magnetic anomaly interpretations of mineral exploration and regional geodynamic studies. However, the conventional interpretation techniques used for modeling the dike parameters are quite challenging and time-consuming. In this study, a fast and efficient inversion algorithm based on machine learning (ML) techniques such as K-Nearest Neighbors (KNN), Random Forest (RF), and XGBoost is developed to interpret the magnetic anomalies produced by the 2-D dike body. The model parameters estimated from these methods include the depth to the top of the dike (z), half-width (d), Amplitude coefficient (K), index angle (α), and origin (x_0). Initially, ML models are trained with optimized hyper-parameters on simulated datasets, and their performance is evaluated using Mean absolute error (MAE), Root means squared error (RMSE), and Squared correlation (R2). The applicability of the ML algorithms was demonstrated on the synthetic data, including the effect of noise and nearby geological structures. The results obtained for synthetic data showed good agreement with the true model parameters. On the noise-free synthetic data, XGBoost better predicts the model parameters of dike than KNN and RF. In comparison, its performance decreases with increasing the percentage of noise and geological complexity. Further, the validity of the ML algorithms was also tested on the four field examples: (i) Mundiyawas-Khera Copper deposit, Alwar Basin, (ii) Pranhita–Godavari (P-G) basin, India, (iii) Pima Copper deposit of Arizona, USA, and (iv) Iron deposit, Western Gansu province China. The obtained results also agree well with the previous studies and drill-hole data.

Keywords Magnetic anomalies · 2-D dipping dike · Machine learning · Inversion

Introduction

The 2-D dipping dike model is widely used in exploration and crustal studies to interpret magnetic anomalies over geological structures and mineralized bodies. Several workers (Gay 1963; Kara 1997; McGrath and Hood 1970; Rao and Babu 1983) have used the curve matching techniques to interpret the magnetic anomalies of the dike. In these methods, dike's model parameters (depth and width) were obtained by matching the theoretical curves with the observed anomalies following a trial-and-error approach. Although these techniques are simple to use, the main drawback is that it is time-consuming to fit the field magnetic anomaly curves. Other methods include Hilbert transform (Sundararajan et al. 1985) and several automatic interpretation techniques such as Werner deconvolution (Ku and Sharp 1983), Euler deconvolution (Reid et al. 1990; Thompson 1982), and analytic signal (Bastani and Pedersen 2001; Roest et al. 1992) were also developed to interpret the magnetic dike anomalies. Further numerical methods based on the least-square window (Abdelrahman et al. 2007), steepest descent, and Levenberg–Marquardt were also vividly used in the magnetic dike interpretation (Atchuta Rao et al. 1985; Beiki and Pedersen 2012; Radhakrishna Murthy et al. 1980). However, these methods are highly subjective and require the initial model parameters to be very close to the true model parameters. This can lead to considerable errors in estimating the model parameters of the dike. On the other hand, global optimization methods such as Particle swarm optimization,

Edited by Prof. Sanyi Yuan (ASSOCIATE EDITOR) / Prof. Gabriela Fernández Viejo (CO-EDITOR-IN-CHIEF).

✉ G. Srinivasa Rao
vasugeos@gmail.com; gsrao@iitism.ac.in

[1] Department of Applied Geophysics, IIT (ISM), Dhanbad, Jharkhand 826004, India

Bat algorithm, simulated and very simulated annealing, and higher-order horizontal derivative have also widely been used to solve the above mentioned problems (Biswas 2018; Biswas et al. 2017; Ekinci et al. 2016; Essa and Diab 2022; Essa and Elhussein 2017, 2019).

Over recent years, machine learning (ML), a data-driven method, has gained popularity in geophysics, mainly due to advanced computation power (Sakrikar and Deshpande 2020). It has been used in seismic studies for lithofacies analysis and reservoir characterization (Bhattacharya et al. 2016; Huang et al. 2017; Liu et al. 2021; Wrona et al. 2018; Xu et al. 2021; Yuan et al. 2018, 2022), geophysical well logging (Bressan et al. 2020; Kitzig et al. 2017; Schmitt et al. 2013; Sun et al. 2020; Wang and Zhang 2008; Xie et al. 2017; Xu et al. 2021; Zhou and O'Brien 2016), electrical resistivity surveys (Liu et al. 2020), magnetotelluric time-series analysis (Manoj and Nagarajan 2003), earthquake data analysis (DeVries et al. 2018), and for determining salt structure using gravity data (Chen et al. 2020). Despite all these studies of ML applications in other geophysical fields, very few studies (Al-Garni 2015) in the literature have applied ML techniques in magnetic data interpretation. Al-Garni (2015) has used the modular neural network to interpret the magnetic anomalies due to a 2D dipping dike. Although this method provides an excellent global optimization method that can accept a wide range of input starting models, the computation speed increases as the networks are not connected.

In this study, we applied three supervised ML algorithms, viz. K-Nearest Neighbors, Random Forest, and XGBoost on magnetic anomaly data to interpret model parameters of the 2-D dipping dike: half-width, index angle, amplitude coefficient, depth, and the origin of the dike. Here, we first test the applicability of ML algorithms on the synthetic data, in the presence of nearby geological structures, with and without adding the noise, and the obtained results were analyzed using various evaluation metrics. Subsequently, ML algorithms are also implemented on the four field examples: (i) Mundiyawas-Khera Copper deposit, Alwar Basin, India, (ii) Pranhita–Godavari (P-G) basin, India, (iii) Pima Copper deposit of Arizona, USA, and (iv) Iron deposit, Western Gansu province, China.

Methodology

Forward modeling of magnetic anomalies over a 2-D dipping dike

The total field magnetic anomaly due to a 2-D dipping dike having uniform magnetization and infinite depth and strike

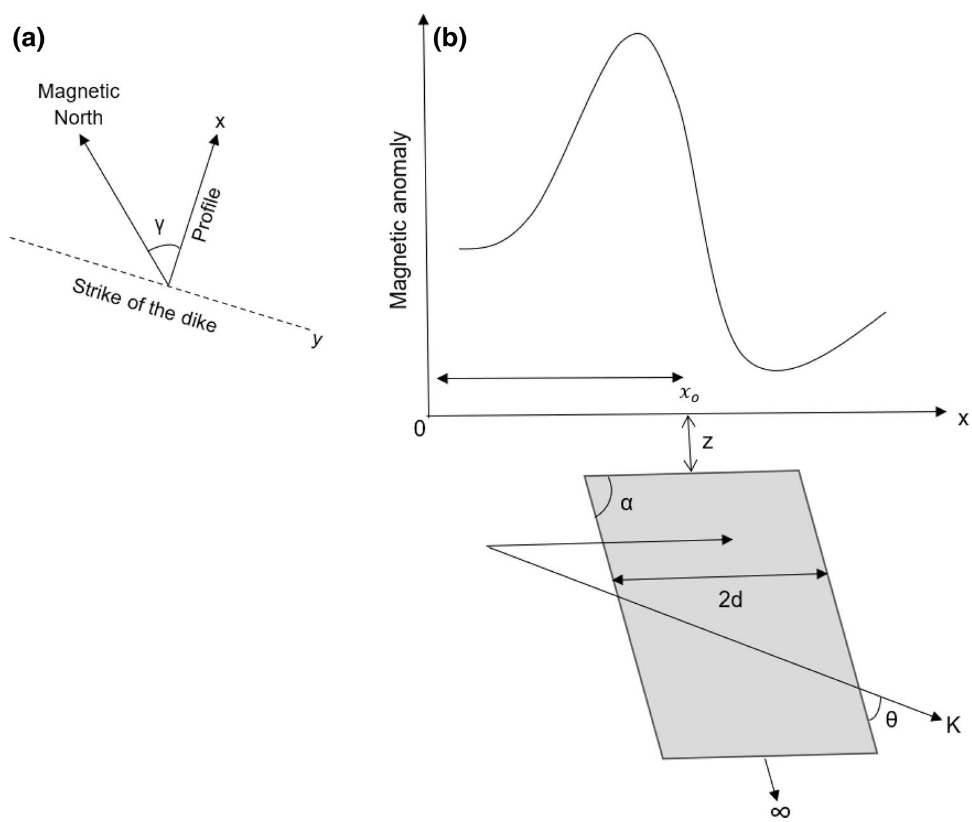

Fig. 1 **a** Top view and **b** Schematic representation of a magnetic anomaly due to a 2D dipping dike (modified from Kara et al. 1996)

extent (Fig. 1) can be represented by the following expression (Hood 1964; Kara et al. 1996; McGrath and Hood 1970):

$$\Delta T = K\left[\sin\theta\left(\tan^{-1}\left(\frac{x+d}{z}\right) - \tan^{-1}\left(\frac{x-d}{z}\right)\right) - \frac{\cos\theta}{2}\ln\left(\frac{(x+d)^2+z^2}{(x-d)^2+z^2}\right)\right], \tag{1}$$

where K is the amplitude coefficient; x is the profile distance; z is the depth to the top; d is the half-width of the dike.

$\theta = 2I - \alpha$, is the index angle; $I = tan^{-1}\left(\frac{tani}{cos\gamma}\right)$.

i is the inclination of the geomagnetic field; α is the dip of the dike.

γ is the azimuth of the profile with reference to the magnetic north.

If we include the origin x_o term in the above expression,

$$\Delta T = K\left[\sin\theta\left(\tan^{-1}\left(\frac{x-x_o+d}{z}\right) - \tan^{-1}\left(\frac{x-x_o-d}{z}\right)\right) - \frac{\cos\theta}{2}\ln\left(\frac{\left(x-x_o+d\right)^2+z^2}{\left(x-x_o-d\right)^2+z^2}\right)\right]. \tag{2}$$

In the present study, forward modeling of magnetic anomalies of 2-D dipping dike is computed using Eq. (2), with the model parameters being the amplitude coefficient (K), the origin of the dike (x_o), the half-width (d), the depth to the top of the dike (z), and the index angle (θ).

Machine learning (ML) algorithms

Several workers have vividly discussed the mathematical background of the ML algorithms (Altman 1992; Breiman 2001; Chen and Guestrin 2016) used in the study. Therefore, a brief account of these algorithms is discussed below.

K-Nearest Neighbors (KNN)

The KNN is a nonparametric supervised ML algorithm used for classification and regression (Altman 1992). It is a lazy-learner algorithm that is the algorithm does not learn while being trained but instead it stores the data set and calculates the output value of the new input data by simply identifying the similarity with the training input data. The 'K' in KNN refers to the number of nearest neighbors to consider when calculating the similarity between the data points. Its value is calculated based on the Euclidean or Manhattan distance which can be represented as:

$$\mathrm{D(x, y)} = \left(\sum_{i=1}^{n}\left|x_i - y_i\right|^p\right)^{1/p}, \tag{3}$$

where n is the number of features, $p=1$ for Manhattan distance (L1-norm) and $p=2$ for Euclidean distance (L2-norm). The performance of KNN depends on the value 'K', therefore the optimum value of k needs to be determined by tuning the parameter over a defined search range (Thanh Noi and Kappas 2018).

Random forest

Random forest is an ensembled supervised ML algorithm that can be used for regression and classification problems. This algorithm uses the average output from multiple decision trees to obtain a more accurate and stable prediction model (Breiman 2001). For a given input vector x, the output of the Random forest algorithm after building the K number of decision trees $T(x)$ is represented as:

$$f_{RF}^{K}(x) = \frac{1}{K}\sum_{k=1}^{K} T(x). \tag{4}$$

In general, the individual decision trees tend to over-fit the training data and show high variance. To improve the performance of the model, Random forest uses Bagging or Bootstrap aggregating, in which decision trees are trained on a sample subset of training data through a replacement (Breiman 2001). Additionally, features are selected randomly to limit the number of features of the growing tree which helps in making the decision trees more diverse and less correlatable (Breiman 2001).

XGBoost

XGBoost is a supervised machine learning algorithm that generates an ensemble of regression trees iteratively based on the principle of gradient boosting algorithm (Chen and Guestrin 2016). In comparison with gradient descent, which optimizes the model parameter, gradient boosting optimizes the loss function of the predicted model (Chen and Guestrin 2016; Friedman 2001; Sun et al. 2020). In order to prevent the overfitting issues and penalize the complexity of the problem, Chen and Guestrin (2016) defined the objective function of the XGBoost as follows:

$$l(\varphi) = \sum_{i=1}^{n} L\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{K} \Omega f(k), \quad (5)$$

where n is the number of training samples, and K represents the total number of decision trees. L is the loss function that measures the fit between the predicted $\left(\hat{y}_i\right)$ and actual $\left(y_i\right)$ values. Ω is the regularisation term that deals with the overfitting and the complexity of the problem. This regularization term is given as follows:

$$\Omega \mathrm{f(k)} = \gamma T + \frac{1}{2}\lambda||w||^2. \quad (6)$$

Here $\gamma and \lambda$ are the penalty coefficients, respectively. T and w indicate the leaf number and Weight, respectively. To minimize the objective function, XGBoost uses the Newton–Raphson method and defines the gradient of the loss function $\frac{\partial L}{\partial G(x)}$ and Hessian as $\frac{\partial^2 L}{\partial G(x)^2}$.

$$l(\varphi) = \sum_{i=1}^{n} L\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{K} \Omega f(k) \quad (7)$$

Hyper-parameter tuning and ML model performance evaluation

In machine learning, hyper-parameter tuning is a crucial step in selecting the optimum hyper-parameter values of the learning algorithms (Hall 2016; Haykin 2009). Grid search or Randomized cross-validation techniques are used to determine optimum hyper-parameters for the ML algorithms. In the present case, we have used the grid search technique to tune the hyper-parameters for obtaining the ideal ML model performance. In order to evaluate the performance of the ML algorithms, the models are tested using three evaluation metrics: Mean absolute error (MAE), Root means squared error (RMSE), and Squared correlation (R2) (Goyal 2021).

Mean Absolute Error (MAE) is the absolute measure of the error between observed and predicted magnetic anomalies in the test data. It is given as:

$$MAE = \frac{1}{N}\sum_{i}^{N}\left|y_i - \hat{y}_i\right|, \quad (8)$$

where N is the total number of points, $\hat{y}_i$ is the predicted output and y_i is the real output value.

Root mean squared error (RMSE) is the relative measure of the error between observed and predicted magnetic anomalies in the test data. It is represented as:

Table 1 Hyper-parameter values considered for KNN, Random Forest and XGBoost

Algorithm	Hyper-parameter	Optimum value
KNN	n_neighbors	6
	weights	distance
Random forest	n_estimators	200
	max_features	auto
	min_samples_split	3
	min_samples_leaf	1
XGBoost	n_estimators	1500
	learning_rate	0.5
	reg_alpha	1
	reg_lambda	10

$$RMSE = \sqrt{\frac{1}{N}\sum_{i}^{N}\left(y_i - \hat{y}_i\right)^2}. \quad (9)$$

The squared correlation (R2) is given as:

$$R2\ squared = 1 - \frac{MSE_{model}}{MSE_{base}} \quad (10)$$

$$where\ MSE_{base} = \frac{1}{N}\sum_{i}^{N}\left(y_i - \overline{y}\right)^2, \quad (11)$$

where $\overline{y}$ is the mean of the output training data. The value of R2 ranges from 0 to 1 only. The closer the R2 squared value to 1, the better the model performance.

ML model training, validation, and field examples

Magnetic anomaly datasets imitating field examples and comprising different scenarios such as the presence and absence of noise and nearby geological structures are prepared to train the ML models of KNN, RF, and XGBoost algorithms. Trained ML models are then tested on simulated magnetic datasets incorporating the above situations to verify their efficacy and robustness.

Synthetic examples

In this study, the applicability of three ML algorithms was demonstrated with the help of synthetic magnetic anomaly data obtained using the forward modeling Eq. 2. To construct the synthetic magnetic anomaly due to the dipping dike, we have chosen the profile distance of 120 units, and

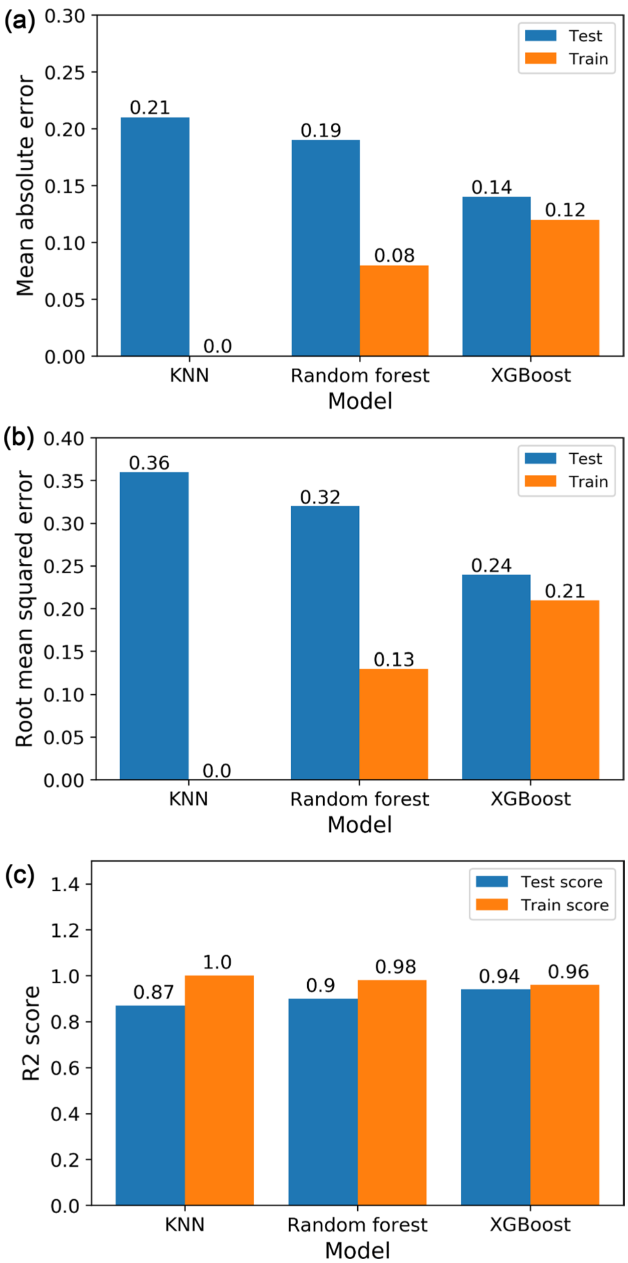

Fig. 2 **a** Mean Absolute Error (MAE), **b** Root Mean Square (RMS), and **c** R2 score of the designed ML models on the training and testing datasets

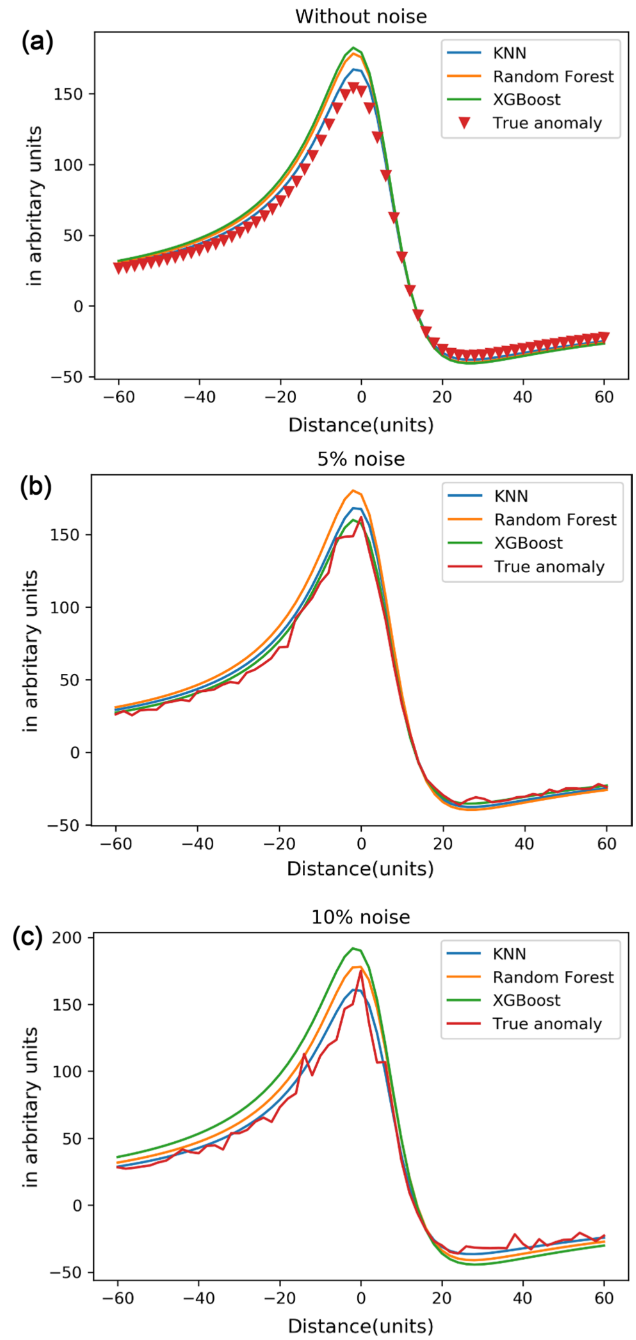

Fig. 3 Predictions of ML algorithms over synthetic magnetic anomaly due to dipping dike **a** noise-free, **b** 5%, and **c** 10% random noise

61 samples are generated with two units interval. The target dike body is assumed to have model parameters $z=10$ units, $d=5$ units, $K=200$ units, $x_o=4$ units, and $\theta=40$ units. The simulated dataset is partitioned into a ratio of 80% for training and 20% for testing. The splitted data is then used for selecting the optimum hyper-parameters of ML algorithms based on the Grid search cross-validation technique. Table 1 shows the optimum value of hyper-parameters chosen for each ML algorithm.

As discussed earlier, MAE, RMSE, and R2 score were computed to study the performance of ML algorithms. The bar plot of MAE, RMSE, and R2 scores for all three algorithms is shown in Fig. 2. It is noticed that KNN provides

Table 2 Predicted model parameters of a 2-D dipping dike (K=200 units, z=10 units, d=5 units, θ=40°, and x_o=4 units) without and with 5% and 10% random noise using KNN, Random Forest, and XGBoost

Algorithms	Model parameters	Predicted values		
		Without noise	5% noise	10% noise
KNN	K	276.04	295.55	303.43
	z	10.29	10.28	10.59
	d	4.07	3.8	3.74
	θ	39.66	39.76	39.96
	x_o	4.14	4.18	4.15
	RMS error	7.51	8.59	9.64
Random forest	K	270.94	278.8	310.71
	z	10.21	10.15	10.66
	d	4.18	4.22	3.82
	θ	39.97	40.15	39.61
	x_o	3.85	3.89	4.52
	RMS error	8.42	12.23	12.27
XGBoost	K	268.16	236.95	376.18
	z	10.31	10.13	11.10
	d	4.48	4.35	3.57
	θ	40.28	40.44	39.12
	x_o	3.83	3.84	4.37
	RMS error	12.3	3.63	19.89

the least MAE and RMS error with the best R2 score of 1 on the training data set (Fig. 2). For the test data, XGBoost (MAE = 0.14, RMS = 0.24, and R2 score = 0.94) give the best performance, followed by Random Forest (MAE = 0.19, RMS = 0.32, and R2 score = 0.90). Whereas KNN gives a poor performance on the test data with an MAE of 0.21, RMS of 0.36, and R2 score of 0.87 (Fig. 2).

Noise-free data

We have applied trained ML models on noise-free data to understand their capability in grasping the basic pattern of anomaly and predicting five model parameters (z, d, K, x_o, and θ) of the dike. Figure 3a shows the comparison between the observed and predicted anomalies from all three ML algorithms. Table 2 shows the RMS error of each ML algorithm. It is noticed that all the ML algorithms well predicted the desired target model parameters of the dike with RMS error varying from 7.51 to 12.3. Unlike other traditional techniques, we do not have to estimate the origin of the dike to proceed ahead, as the proposed ML algorithms can calculate the origin of the dike. The average predicted depth (z) and half-width (d) are 10.27 and 4.24, which show an error of 2.7% and 15.2% from their true values, respectively (Table 2). Whereas the average values of predicted x_o and θ are 3.94 and 39.97, which show an error of 1.33% and 0.07% from their true values, respectively. In comparison, the estimated K values from the ML algorithm show a large error (Table 2). The average predicted value of K is 271.71and shows an error of 35.85% from its true value.

Effect of noise

In order to estimate the robustness of the trained ML models, random noise of 5% and 10% has been added to the magnetic anomaly of the dike. Figure 3 shows the plot of the inverted results of the various algorithms and the observed anomaly curve before and after adding 5% and 10% random noise, respectively. It is noticed that RMS error increases with an increase in the percentage of noise in the case of KNN and RF algorithms (Table 2). Whereas XGBoost does not show consistent results in the addition of noise. For the synthetic anomaly with the addition of 5% random noise, XGBoost provides a better fit compared to the KNN and RF algorithms (Fig. 3b). However, at 10% Gaussian noise (Fig. 3c), XGBoost shows a poor fit between the observed and predicted anomaly, which is in agreement with the RMS error shown in Table 2. At 10% noise data, XGBoost shows the highest RMS error of 19.89 among the three ML algorithms. Thus, it is suggested that KNN and Random Forest are stable even on noisy data and provide better prediction of model parameters of the dike.

Interference from nearby structures

A composite magnetic anomaly consisting of both vertical and inclined dyke bodies is constructed along a profile length of 60 m using the forward modeling Eq. 2. The vertical dike body is assumed to have model parameters z = 4 m, θ = 45°, K = 2500 nT and x_o = 40 m. Whereas the parameters of dipping dike are assumed to be z = 9 m, x_o = 0 m, θ = 40°, K = 1500 nT, and d = 4 m. All three ML algorithms were applied to the composite magnetic anomaly data to investigate the effect of nearby structures in predicting the model parameter of the target body. Figure 4 shows the comparison between the observed and predicted anomalies from the ML algorithm of the composite dike model. It is observed that the ML algorithms have recovered all the five model parameters of the dike with good accuracy, and RMS error varies from 0.72 to 2.97 (Table 3). The error between the average predicted depth (z) from ML algorithms and the true depth of the dike is 11.74%. The predicted half-width (d) varies

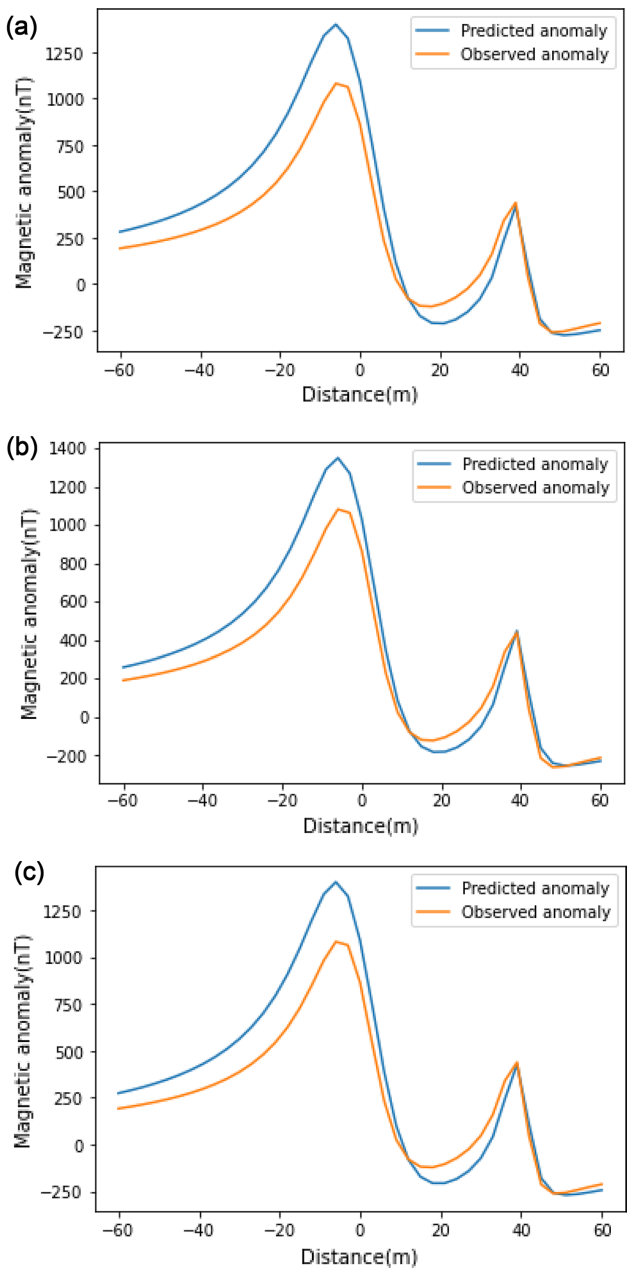

Fig. 4 Comparison of observed and ML model predicted magnetic anomalies of a composite dike model: **a** KNN, **b** Random Forest, and **c** XGBoost

from 5.75 to 6.15 m which shows an average error of 49.25% with respect to true half-width. In comparison with the associated errors with K, x_o, and θ are minimum (Table 3).

Sensitivity analysis

Sensitivity analysis provides insight into how the input parameters influence the model's output (Tunkiel et al. 2020). Several data-driven and statistical based methods such as Sobol' indices, shapely effect, meta-models, and partial derivatives (PaD) have been proposed by the earlier worker to analyze the sensitivity of the machine learning regression models (Radaideh et al. 2019; Simpson et al. 2001; Sobol 1993; Tunkiel et al. 2020). In the present study, we have used the partial derivatives (PaD) method proposed by Tunkiel et al. (2020) to conduct the sensitivity analysis as it is suited for ML models for predicting multiple outputs. To better understand this method, consider a model described as $Y = f(X)$, where Y is the output and X is the model's input described by a function f. The sensitivity index can be calculated from the following equation, which can be considered a partial derivative (Tunkiel et al. 2020):

$$SI_{YX} = \frac{\left[Y\left(X_0 + \Delta X\right) - Y\left(X_0 - \Delta X\right)\right]}{2\Delta X}, \tag{12}$$

where SI_{YX} denotes sensitivity index for an output variable Y per unit change in the input X from its base value X_o. ΔX is the change applied to the input.

Figure 5 shows the results of sensitivity analysis for each machine learning model. The SI value of all three ML models falls in the range of − 10 to 15 (Fig. 5). For KNN and Random Forest models, the index angle (θ) shows the highest sensitivity value, followed by the amplitude coefficient (K), depth (z), and half-width (d). Whereas in the case of XGBoost, both the amplitude coefficient and index angle (θ) show a higher SI value than the other model parameters. It is noticed that origin (x_o) has the least SI value compared to the other model parameters in the case of KNN and Random Forest models (Fig. 5). Whereas for XGBoost, both origin (x_o) and half-width (d) show a similar range of SI values (Fig. 5).

Table 3 Predicted results for a synthetic data composed of a 2-D inclined dike (with $z=9$ m, $xo=0$ m, $\theta=40°$, $K=1500$ nT, and $d=4$ m), and a vertical dike (with $z=4$ m, $\theta=45°$, $K=2500$ nT, and $xo=40$ m) using KNN, Random Forest and XGBoost

Model parameters	True parameters	Predicted parameters		
		KNN	Random forest	XGBoost
K (nT)	1500	1502.23	1489.27	1500.55
z (m)	9	10.19	9.99	9.99
d (m)	4	6.15	5.75	6.01
θ (°)	40	39.47	41.02	40.03
x_o (m)	0	0.13	−0.37	−0.02
RMS error	1.25	2.97	0.72	

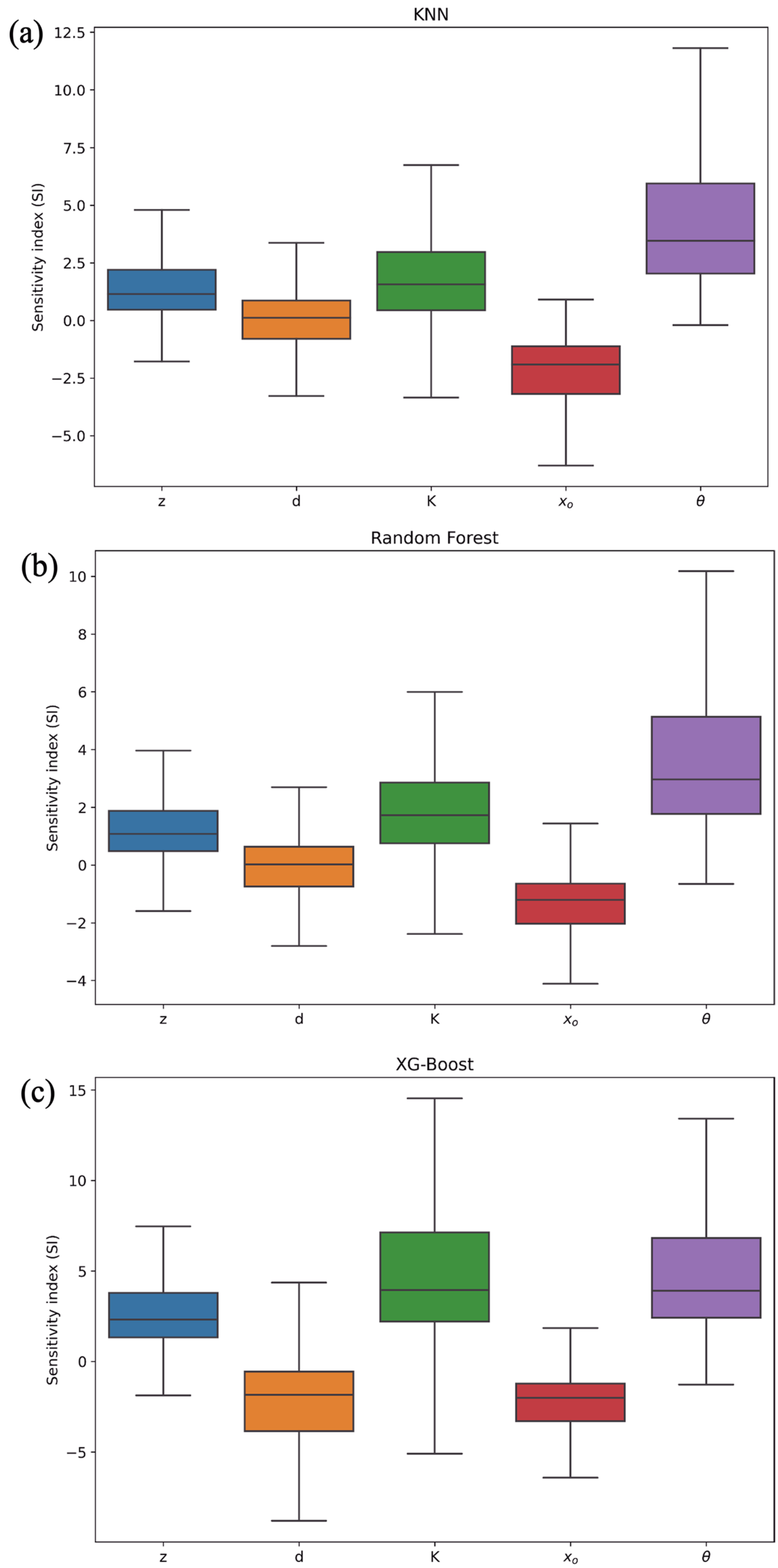

Fig. 5 Results of sensitivity analysis obtained for: **a** KNN, **b** Random Forest, and **c** XGBoost

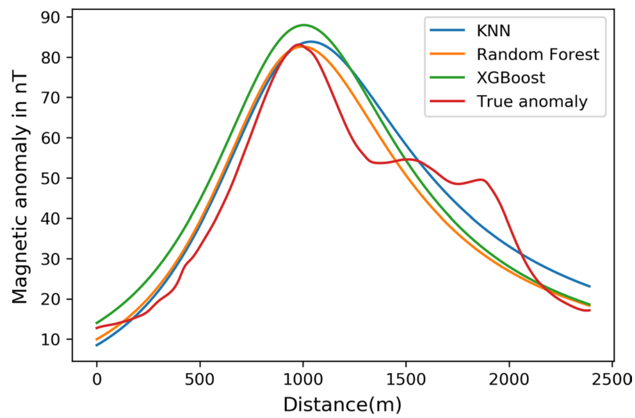

Fig. 6 Comparison of observed and ML model predicted (KNN, Random Forest, and XGBoost) magnetic anomaly in Mundiyawas-Khera Copper deposit, Alwar Basin (Rao et al. 2019)

Field examples

Several studies have pointed out that the proposed inversion algorithms must be tested on field data to illustrate the efficiency and validity of the algorithms in obtaining the different model parameters of the dike (Al-Garni 2015; Biswas and Rao 2021; Essa and Elhussein 2018; Mehanee 2014; Rao and Biswas 2021). For this purpose, we have chosen four field examples from the published literature, which include: (i) Mundiyawas-Khera Copper deposit, Alwar Basin (Rao et al. 2019), (ii) Pranhita–Godavari (P-G) basin, India (Radhakrishna Murthy and Bangaru Babu 2009), (iii) Pima Copper deposit of Arizona, USA (Asfahani and Tlas 2004), and (iv) Iron deposit, Western Gansu province China (Essa and Elhssein 2017). We have also compared the obtained results from ML algorithms with the previous studies and drill-hole data.

Mundiyawas-Khera copper deposit, Alwar Basin

Mundiyawas-Khera area, located in the Alwar basin of India, is well known for Copper-Au rich mineralization hosted within the dolomite and felsic metavolcanic rocks (Khan et al. 2015; Rao et al. 2019). We have considered a magnetic profile (Fig. 6) of 2390 m across the anomalous zone from Rao et al. (2019). It shows NNW-SSE orientation with references to the dominant lithologies of the area. Previous studies (Khan et al. 2015; Rao et al. 2019) interpreted that sulfide mineralization occurs in this region in the form of massive pyrrhotite, extending from a shallow depth of 50–100 m to a deeper depth of < 300 m with a dip toward the west. Further, the width of the anomalous bodies varies from 30 to 80 m (Rao et al. 2019).

The predicted dike parameter from each ML algorithm and its comparison with the previous studies are shown in Table 4. The plots of the observed and predicted anomalies from KNN, RF, and XGBoost are shown in Fig. 6. It is noticed that the predicted results of all the ML algorithms are in agreement with each other (Table 4). Although the predicted half-width obtained from ML shows good agreement with their results, ML algorithms predict higher depth values than the previous study (Rao et al. 2019) (Table 4). It is relevant to note here that the previous study (Rao et al. 2019) only predicts the two parameters of the dike, whereas the proposed ML algorithms are able to predict five parameters of the dike with reasonable accuracy.

RMSE and MAE values (Table 4) indicate that all ML algorithms show reasonably a good fit between predicted and observed anomaly curves (Fig. 6). However, KNN and RF show the least RMSE and MAE values than XGBoost. The scatter plots between the predicted and observed anomalies for three ML algorithms are plotted in Fig. 7 to illustrate the goodness of fit. Although the variance of the predicted anomaly for all the three algorithms is quite less, KNN

Table 4 Comparison predicted model parameters of Mundiyawas-Khera Copper deposit, Alwar Basin, India using the present method (KNN, Random Forest, and XGBoost) and previous studies

Model parameters	Present method			Drilling result, Khan et al. (2015)	Rao et al. (2019)
	KNN	Random Forest	XGBoost		
K (nT)	418.09	410.2	407.01	–	–
z (m)	595	548.41	566.26	40–80	50–250
d (m)	62.28	56.81	62.3	–	30–80
θ (°)	112.74	107.98	103.61	–	–
x_o (m)	918.06	910.65	936.87	–	–
RMSE	7.1	7.6	9.53	–	–
MAE	5.62	5.95	8.29	–	–

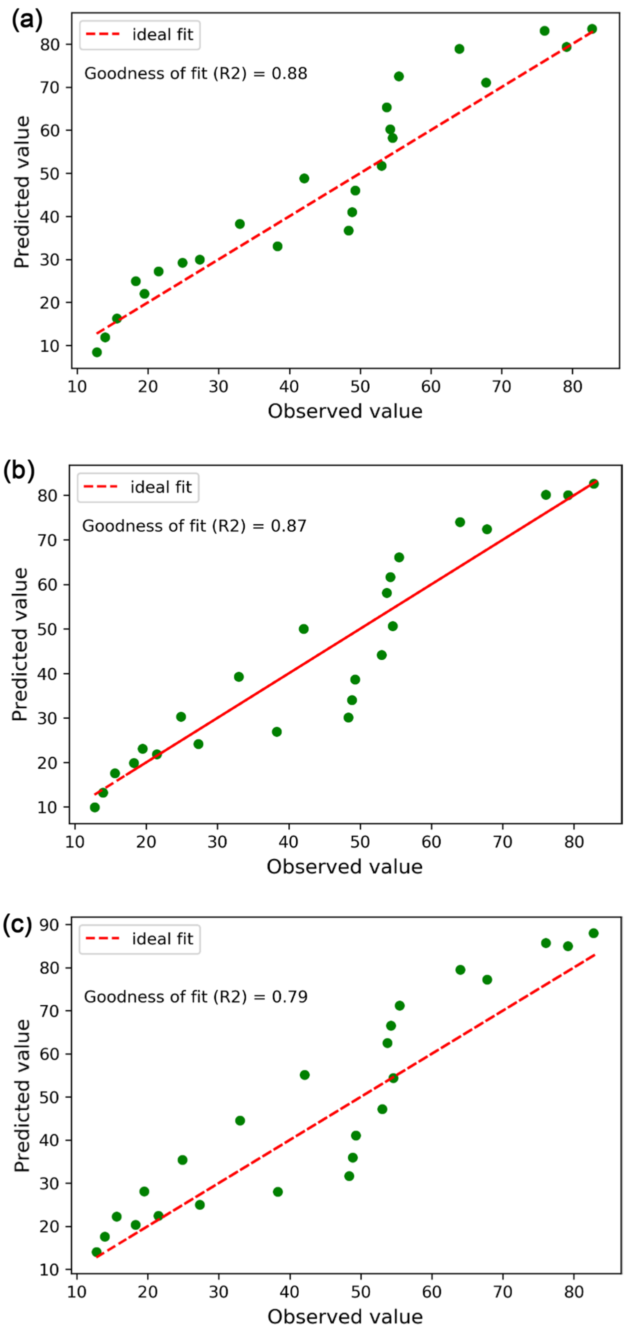

Fig. 7 Correlation between the observed and ML model predicted anomaly in Mundiyawas-Khera Copper deposit, Alwar Basin. **a** KNN, **b** Random Forest, and **c** XGBoost

and RF show the highest R2 score compared to the than XGBoost.

Pranhita–Godavari (P-G) Basin, India

The aeromagnetic anomaly profile of 60 km length (Fig. 8) is constructed across the Pranhita–Godavari (P-G) basin, India (Radhakrishna Murthy and Bangaru Babu 2009). The

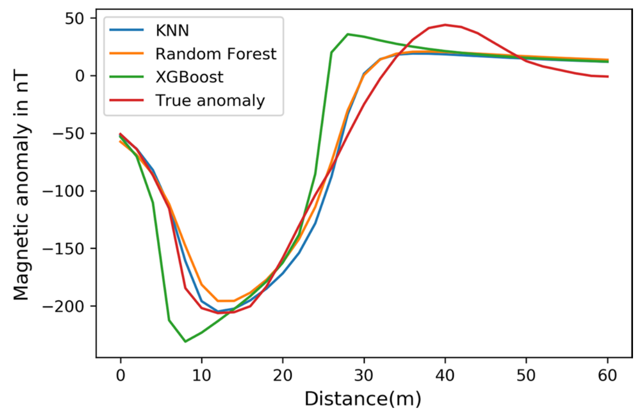

Fig. 8 Comparison of observed and ML model predicted (KNN, Random Forest, and XGBoost) magnetic anomaly in Pranhita–Godavari (P-G) basin, India (Radhakrishna Murthy and Bangaru Babu 2009)

anomaly curve is sampled at an interval of 2 km, and a total of 31 points are obtained along the profile. Earlier workers (Mishra et al. 1987; Radhakrishna Murthy and Bangaru Babu 2009) attributed the magnetic anomalies in the area are due to the emplacement of dolerite dike intrusive into the basement of the P-G basin. Therefore, we have re-modeled these anomalies for the dike model using the KNN, RF, and XGBoost. Based on Marquardt's optimization technique, Radhakrishna Murthy and Bangaru Babu (2009) also estimated the depth and half-width of the dike as 8 km and 5.5 km, respectively.

The predicted value of the depth and half-width from all the ML algorithms show good agreement with previous studies (Table 5). The plot of the predicted anomaly and the observed anomaly of each ML algorithm is shown in Fig. 8. It is noticed that KNN and RF give a very good fit and also show less RMSE and MAE error compared to XGBoost (Table 5). The prediction error plot of each algorithm shown in Fig. 9 indicates that all the algorithms give the same value of goodness of fit (R2 score) of 0.97.

Pima copper deposit, Arizona, USA

The magnetic anomaly profile of length 750 m (Fig. 10) over Pima copper deposit, Arizona, USA is compiled from Gay (1963). The magnetic data along this profile is digitized at an interval of 15 m. Several earlier workers (Abdelrahman and Essa 2015; Abdelrahman et al. 2003; Asfahani and Tlas 2004, 2007; Biswas et al. 2017; Gay 1963; Mehanee et al. 2021; Tlas and Asfahani 2015) have interpreted this magnetic anomaly data using different techniques by considering a thin-dyke model. This study has re-modeled this data using the KNN, RF, and XGBoost algorithms.

Table 5 Comparison of predicted parameters of the dike in Pranhita–Godavari (P-G) basin, India using the present method (KNN, Random Forest, and XGBoost) and previous studies

Model parameters	Present method			Radhakrishna Murthy and Bangaru Babu (2009)
	KNN	Random Forest	XGBoost	
K (gammas)	77.67	80.9	72.66	–
z (km)	2.75	3.49	1.08	5.5
d (km)	9.8	9.42	9.36	8
θ (°)	247.33	243.58	247.07	–
x_o (km)	17.82	18.16	15.28	–
RMSE	14.36	14.68	35.03	
MAE	11.48	12.27	22.55	

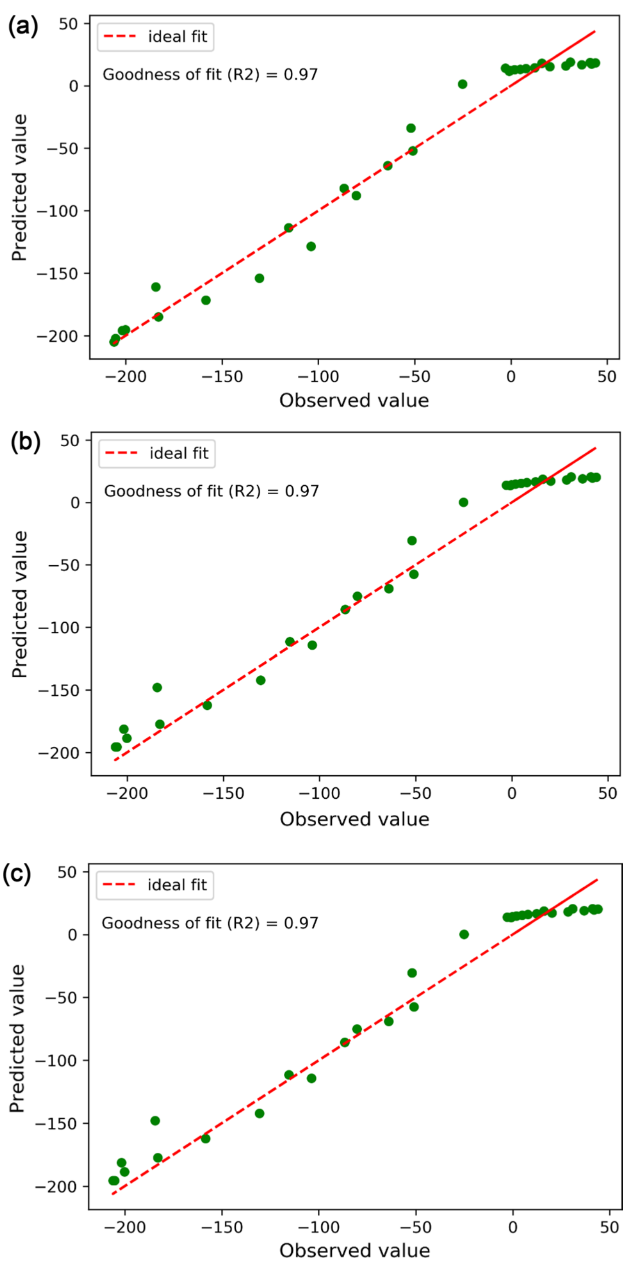

Fig. 9 Correlation between the observed and ML model predicted anomaly in Pranhita–Godavari (P-G) basin, India. **a** KNN, **b** Random Forest, and **c** XGBoost

It is noticed that the predicted parameters of dike using our present technique show a good agreement with previous studies and are also comparable with each other. Further, most of the earlier methods predict a maximum of four parameters (*K*, *z*, α, and xo) of the dike, whereas the proposed ML algorithms are able to predict five parameters, i.e., all the above four parameters, including the half-width of the dike (Table 6). The predicted anomaly (Fig. 10) from all three ML algorithms shows a good fit with the observed anomaly and shows small RMSE and MAE errors (Table 6). Among the three ML algorithms, KNN shows relatively small RMSE (14.98) and MAE (10.01) errors compared to the RF (RMSE = 17.01; MAE = 11.58) and XGBoost (RMSE = 16.43; MAE = 13.59) (Table 6). The scatter plot of the prediction error for the three ML algorithms (Fig. 11) also shows a high R2 score of 0.99.

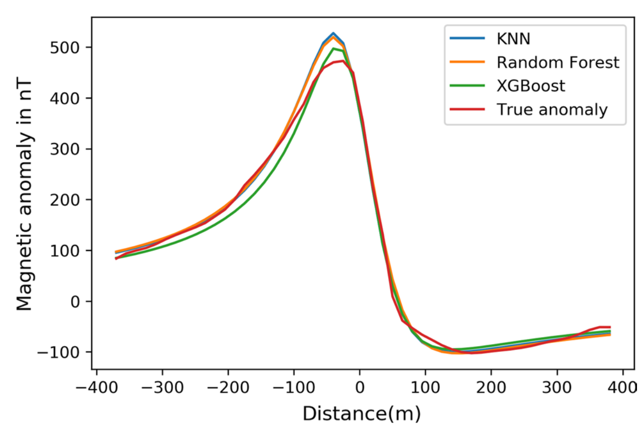

Fig. 10 Comparison of observed and ML model predicted (KNN, Random Forest, and XGBoost) magnetic anomaly in Pima Copper deposit of Arizona, USA (Asfahani and Tlas 2004)

Table 6 Comparison of predicted parameters of the Pima Copper deposit of Arizona, USA using the present method (KNN, Random Forest, and XGBoost) and previous studies

Model parameters	Present method			Gay (1963)	Abdelrahman et al. (2003)	Asfahani and Tlas (2004)	Asfahani and Tlas (2007)	Abdelrahman and Essa (2015)	Tlas and Asfahani (2015)	Biswas et al. (2017)	Mehanee et al. (2021)
	KNN	Random Forest	XGBoost								
K (nT)	778.08	689.2	801.83	–	1611	577.61	577.6	–	–	613.0	46,424.38
z (m)	64.33	65.54	61.48	70	64	71.5	71.5	60	64.1	68.0	71.0
d (m)	27.02	31.5	23.62	–	–	–	–	–	–	–	–
θ (°)	−43.67	−42.9	−43.52	−50	−52	−50.46	−50.5	–	−44.7	–	−55.11
x_o (m)	−7.67	−5.72	−4.23	–	–	–	–	–	–	−4.30	4.72
RMSE	14.98	17.01	16.43	–	–	–	–	–	–	–	–
MAE	10.01	11.58	13.59	–	–	–	–	–	–	–	–

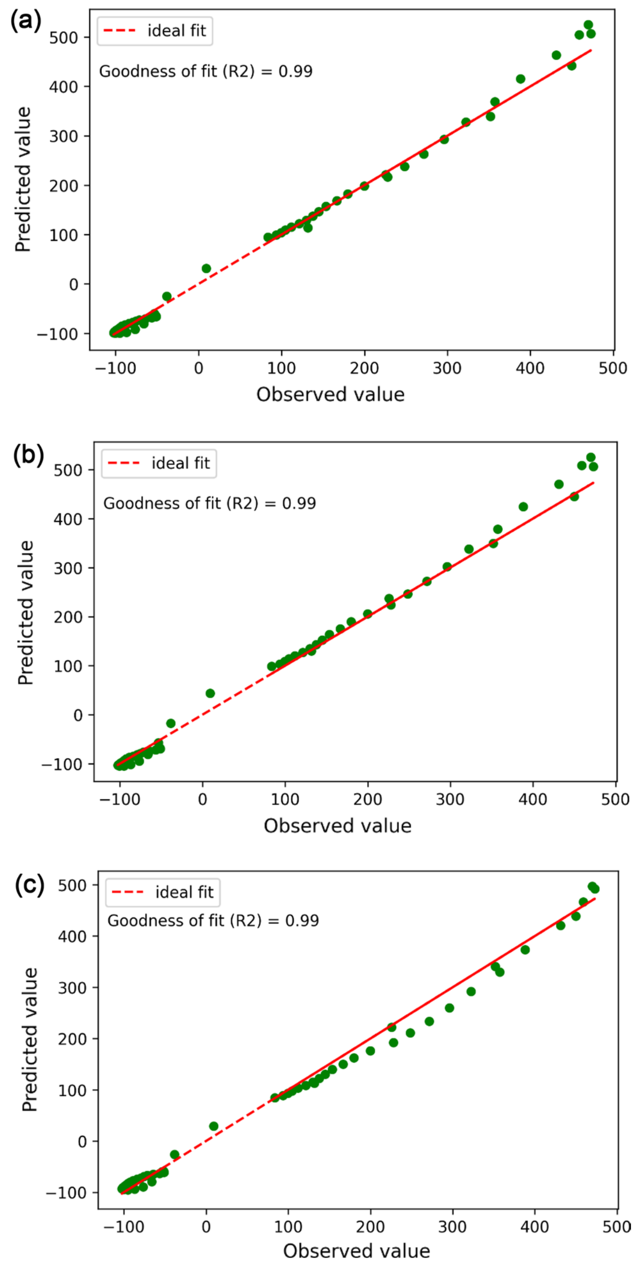

Fig. 11 Correlation between the observed and ML model predicted anomaly in Pima Copper deposit of Arizona, USA. **a** KNN, **b** Random Forest, and **c** XGBoost

Magnetite iron deposit, China

This field data are taken from the magnetite iron deposit in western Gansu Province, China (Guo et al. 1998). The profile length is 222.5 m, and it is digitized with a sampling interval of 10 m. The predicted dike model parameter using the ML techniques is shown in Table 7, including the results from the previous studies. It is observed that the predicted results are in good agreement with each other and also with the earlier studies (Essa and Elhussein

Table 7 Comparison of predicted parameters of the Iron deposit in Western Gansu province, China using the present method (KNN, Random Forest, and XGBoost) and previous studies

Model parameters	Present method			Essa and Elhussein (2017)	Drilling data, Gou et al. (1998)
	KNN	Random Forest	XGBoost		
K (nT)	5358.99	5773.16	6363.29	10,380.62	–
z (m)	20.1	20.4	22	26.56	22–25
d (m)	8.49	7.92	10.37	9.035	9
θ (°)	56.79	60.83	56.65	58.97	–
x_o (m)	−0.6	−0.09	−1.615	–	–
RMSE	387.62	381.36	524.02	–	–
MAE	283.15	296.4	407.01	–	–

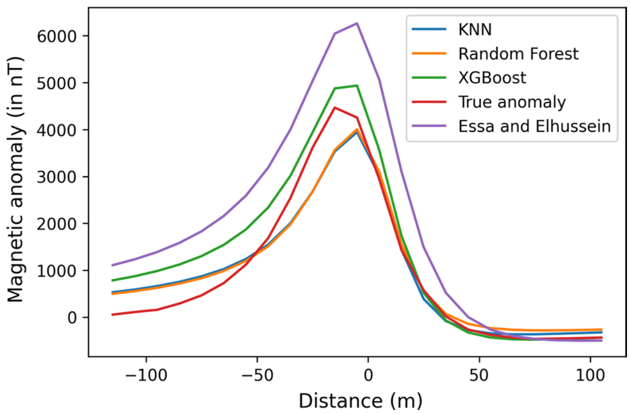

Fig. 12 Comparison of observed and ML model predicted (KNN, Random Forest, and XGBoost) magnetic anomaly in Iron deposit, Western Gansu province China (Essa and Elhssein 2017)

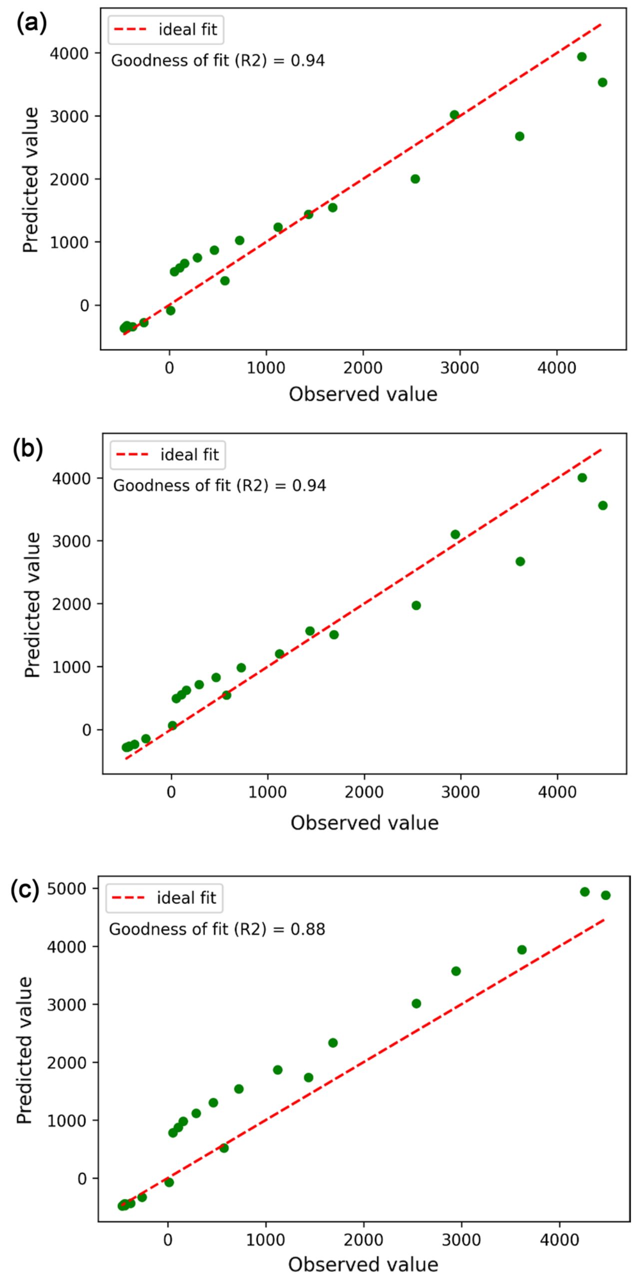

Fig. 13 Correlation between the observed and ML model predicted anomaly in Iron deposit, Western Gansu province China. **a** KNN, **b** Random Forest, and **c** XGBoost

2017). The predicted model parameters viz. depth to the top of the dike (*z*) and the half-width (*d*) also agree with the Drilling data (Guo et al. 1998). The predicted anomaly and the observed anomaly also show a similar trend and are in good agreement, as shown in Fig. 12. The RMSE and MAE errors (Table 7) from algorithms seem higher than the previous field examples discussed in this study. This is due to the high amplitude of the field anomaly data. The prediction error plot of the three algorithms is given in Fig. 13. The goodness of fit (R2) score of KNN and Random Forest shows the same value (R2 = 0.94), whereas XGBoost gives a relatively lesser R2 score of 0.88 and shows higher variance.

Conclusions

In the present study, an attempt was made to investigate the performance of three Machine learning algorithms, such as KNN, Random Forest (RF), and XGBoost, in predicting the model parameters of the dike. The major conclusions drawn in this study are summarized below:

- The results on synthetic and field examples indicate that KNN, RF, and XGBoost perform well in obtaining all five model parameters of the dike, which are depth to the top of the dike (z), half-width (d), Amplitude coefficient (K), index angle (α), and origin (xo). However, they show different prediction power, depending on the anomaly complexity.
- KNN and RF are less sensitive to noise or anomaly complexity and give comparable results in both cases. On the other hand, XGBoost performs well only on noise-free data, whereas its performance drops drastically with increasing the percentage of noise or the complexity of the magnetic anomaly.
- The effect of interference from nearby structures on the ML algorithms was also tested, and it was found that all the ML algorithms are affected very little by this interference.
- The field examples demonstrate that KNN and RF have less RMSE and MAE values than XGBoost, suggesting that KNN and RF have the highest prediction power. In most of the field examples, the R2 score of KNN and RF is found to be 0.89–0.99, which is better than XGBoost.

Acknowledgments GSR sincerely expresses his gratitude to the Department of Science and Technology (DST-FIST/197/2018-19/580), and Science and Engineering Research Board (ECR/2016/001860 and CRG/2021/006513), the Government of India.

Author contributions All authors contributed to the study conception and design. Algorithm development, and data analysis were performed by Sh Bronson Aimol and G Srinivasa Rao. The first draft of the manuscript was written by Sh Bronson Aimol and G Srinivasa Rao. Thinesh Kumar and Rama Chandrudu Arasada commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the Department of Science & Technology (DST-FIST/197/2018–19/580) and the Science and Engineering Research Board (ECR/2016/001860 and CRG/2021/006513), Government of India.

Declarations

Conflict of interest Authors declare they have no financial interests.

References

Abdelrahman EM, Essa KS (2015) A new method for depth and shape determinations from magnetic data. Pure Appl Geophys 172:439–460

Abdelrahman EM, El-Arby HM, El-Arby TM, Essa KS (2003) A least-squares minimization approach to depth determination from magnetic data. Pure Appl Geophys 160(7):1259–1271

Abdelrahman EM, Abo-Ezz ER, Soliman KS, El-Araby TM, Essa KS (2007) A least-squares window curves method for interpretation of magnetic anomalies caused by dipping dikes. Pure Appl Geophys 164:1027–1044

Al-Garni MA (2015) Interpretation of magnetic anomalies due to dipping dikes using neural network inversion. Arab J Geosci 8:8721–8729

Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 46:175–185

Asfahani J, Tlas M (2004) Nonlinearly constrained optimization theory to interpret magnetic anomalies due to vertical faults and thin dikes. Pure Appl Geophys 161:203–219

Asfahani J, Tlas M (2007) A robust non-linear inversion for the interpretation of magnetic anomalies caused by faults, thin dikes and spheres like structure using stochastic algorithms. Pure Appl Geophys 164:2023–2042

Atchuta Rao D, Ram Babu HV, Venkata Raju DC (1985) Inversion of gravity and magnetic anomalies over some bodies of simple geometric shape. Pure Appl Geophys 123:239–249

Bastani M, Pedersen LB (2001) Automatic interpretation of magnetic dike parameters using the analytical signal technique. Geophysics 66:551–561

Beiki M, Pedersen LB (2012) Estimating magnetic dike parameters using a non-linear constrained inversion technique: an example from the Särna area, west central Sweden. Geophys Prospect 60:526–538

Bhattacharya S, Carr TR, Pal M (2016) Comparison of supervised and unsupervised approaches for mudstone lithofacies classification: Case studies from the Bakken and Mahantango-Marcellus Shale, USA. J Nat Gas Sci Eng 33:1119–1133

Biswas A (2018) Inversion of source parameters from magnetic anomalies for mineral/ore deposits exploration using global optimization technique and analysis of uncertainty. Nat Resour Res 27(1):77–107

Biswas A (2021) Rao K (2021) Interpretation of magnetic anomalies over 2D fault and sheet-type mineralized structures using very fast simulated annealing global optimization: an understanding of uncertainty and geological implications. Lithosphere 2021(Special 6):2964057

Biswas A, Parija MP, Kumar S (2017) Global non-linear optimization for the interpretation of source parameters from total gradient of gravity and magnetic anomalies caused by thin dyke. Ann Geophys 60:G0218–G0218

Breiman L (2001) Random forests. Mach Learn 45:5–32

Bressan TS, de Souza MK, Girelli TJ, Junior FC (2020) Evaluation of machine learning methods for lithology classification using geophysical data. Comput Geosci 139:104475

Chen T, Guestrin C (2016) XG-Boost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794

Chen J, Schiek-Stewart C, Lu L, Witte S, Eres Guardia K, Menapace F, Devarakota P, Sidahmed M (2020) Machine learning method to determine salt structures from gravity data. In SPE annual technical conference and exhibition. OnePetro

DeVries PM, Viégas F, Wattenberg M, Meade BJ (2018) Deep learning of aftershock patterns following large earthquakes. Nature 560:632–634

Ekinci YL, Balkaya Ç, Göktürkler G, Turan S (2016) Model parameter estimations from residual gravity anomalies due to simple-shaped sources using differential evolution algorithm. J Appl Geophys 129:133–147

Essa KS, Diab ZE (2022) Magnetic data interpretation for 2D dikes by the metaheuristic bat algorithm: sustainable development cases. Sci Rep 12(1):1–29. https://doi.org/10.1038/s41598-022-18334-1

Essa KS, Elhussein M (2017) A new approach for the interpretation of magnetic data by a 2-D dipping dike. J Appl Geophys 136:431–443

Essa KS, Elhussein M (2018) PSO (particle swarm optimization) for interpretation of magnetic anomalies caused by simple geometrical structures. Pure Appl Geophys 175:3539–3553
Essa KS, Elhussein M (2019) Magnetic interpretation utilizing a new inverse algorithm for assessing the parameters of buried inclined dike-like geological structure. Acta Geophys 67(2):533–544
Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29:1189–1232
Gay P (1963) Standard curves for interpretation of magnetic anomalies over long tabular bodies. Geophysics 28:161–200
Goyal S (2021) Evaluation metrics for regression models, analytics Vidhya. https://medium.com/analytics-vidhya/evaluation-metrics-for-regression-models91c65d73af
Guo W, Dentith MC, Li Z, Powell CM (1998) Self demagnetisation corrections in magnetic modelling: some examples. Explor Geophys 29(4):396–401
Hall B (2016) Facies classification using machine learning. Lead Edge 35(10):906–909
Haykin S (2009) Neural networks and learning machines. Prentice Hall, New York, p 938
Hood P (1964) The Königsberger ratio and the dipping-dyke equation. Geophys Prospect 12:440–456
Huang L, Dong X, Clee TE (2017) A scalable deep learning platform for identifying geologic features from seismic attributes. Lead Edge 36:249–256
Kara I (1997) Magnetic interpretation of two-dimensional dikes using integration-nomograms. J Appl Geophys 36:175–180
Kara I, Özdemir M, Ahmet Yüksel F (1996) Interpretation of magnetic anomalies of dikes using correlation factors. Pure Appl Geophys 147:777–788
Khan I, Sahoo PR, Rai DK (2015) Geological set up of low-grade copper-gold mineralization at Mundiyawas-Khera area, Alwar district, Rajasthan. In: Golani PR (ed) Recent developments in metallogeny and mineral exploration in Rajasthan. Geol Soc Spec Publ 101:43–58
Kitzig MC, Kepic A, Kieu DT (2017) Testing cluster analysis on combined petrophysical and geochemical data for rock mass classification. Explor Geophys 48:344–352
Ku CC, Sharp JA (1983) Werner deconvolution for automated magnetic interpretation and its refinement using Marquardt's inverse modelling. Geophysics 48:754–774
Liu B, Guo Q, Li S, Liu B, Ren Y, Pang Y, Guo X, Liu L, Jiang P (2020) Deep learning inversion of electrical resistivity data. IEEE Trans Geosci Remote Sens 58:5715–5728
Liu X, Ge Q, Chen X, Li J, Chen Y (2021) Extreme learning machine for multivariate reservoir characterization. J Pet Sci Eng 205:108869
Manoj C, Nagarajan N (2003) The application of artificial neural networks to magnetotelluric time-series analysis. Geophys J Int 153:409–423
McGrath PH, Hood PJ (1970) The dipping dike case: A computer curve-matching method of magnetic interpretation. Geophysics 35:831–848
Mehanee SA (2014) Accurate and efficient regularized inversion approach for the interpretation of isolated gravity anomalies. Pure Appl Geophys 171:1897–1937
Mehanee S, Essa KS, Diab ZE (2021) Magnetic data interpretation using a new R-parameter imaging method with application to mineral exploration. Nat Resour Res 30(1):77–95
Mishra DC, Gupta SB, Rao MV, Venkatarayudu M, Laxman G (1987) Godavari basin-a geophysical study. J Geol Soc India 30:469–476
Radaideh MI, Surani S, O'Grady D, Kozlowski T (2019) Shapley effect application for variance-based sensitivity analysis of the few-group cross-sections. Ann Nucl Energy 129:264–279
Radhakrishna Murthy IV, Bangaru Babu S (2009) Magnetic anomalies across Bastar craton and Pranhita-Godavari basin in south of central India. J Earth Syst Sci 118:81–87
Radhakrishna Murthy IV, Visweswara Rao C, Krishna GG (1980) A gradient method for interpreting magnetic anomalies due to horizontal circular cylinders, infinite dykes and vertical steps. Proc Indian Acad Sci-Earth Planet Sci 89:31–42
Rao DA, Babu HR (1983) Quantitative interpretation of self-potential anomalies due to two-dimensional sheet-like bodies. Geophysics 48:1659–1664
Rao K, Biswas A (2021) Modeling and uncertainty estimation of gravity anomaly over 2D fault using very fast simulated annealing global optimization. Acta Geophys 69:1735–1751
Rao GS, Arasada RC, Sahoo PR, Khan I (2019) Integrated geophysical investigations in the Mudiyawas-Khera block of the Alwar basin of North Delhi Fold Belt (NDBF): implications on copper and associated mineralization. J Earth Syst Sci 128:1–13
Reid AB, Allsop JM, Granser H, Millett AJ, Somerton IW (1990) Magnetic interpretation in three dimensions using Euler deconvolution. Geophysics 55:80–91
Roest WR, Verhoef J, Pilkington M (1992) Magnetic interpretation using the 3-D analytic signal. Geophysics 57(1):116–125
Sakrikar C, Deshpande K (2020) Use of machine learning and artificial intelligence in earth science. In: ICSITS–2020 Conference proceedings ICSITS. Int J Eng Res Technol (IJERT), 8(05)
Schmitt P, Veronez MR, Tognoli FMW, Todt V, Lopes RC, Silva CAU (2013) Electrofacies modelling and lithological classification of coals and mudbearing ingrained siliciclastic rocks based on neural networks. Earth Sci Res 2:193–208
Simpson TW, Poplinski JD, Koch PN, Allen JK (2001) Metamodels for computer-based engineering design: survey and recommendations. Eng Comput 17(2):129–150
Sobol IYM (1990) On sensitivity estimation for nonlinear mathematical models. Matematicheskoe Model 2(1):112–118
Sun Z, Jiang B, Li X, Li J, Xiao K (2020) A data driven approach for lithology identification based on parameter-optimized ensemble learning. Energies 13(15):3903
Sundararajan N, Mohan NL, Raghava MS, Rao SV (1985) Hilbert transform in the interpretation of magnetic anomalies of various components due to a thin infinite dike. Pure Appl Geophys 123:557–566
Thanh Noi P, Kappas M (2018) Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. Sensors 18:18
Thompson DT (1982) EULDPH: A new technique for making computer-assisted depth estimates from magnetic data. Geophysics 47:31–37
Tlas M, Asfahani J (2015) The simplex algorithm for best-estimate of magnetic parameters related to simple geometric-shaped structures. Math Geosci 47:301–316
Tunkiel AT, Sui D, Wiktorski T (2020) Data-driven sensitivity analysis of complex machine learning models: a case study of directional drilling. J Pet Sci Eng 195:107630
Wang K, Zhang L (2008) Predicting formation lithology from log data by using a neural network. Pet Sci 5(3):242–246
Wrona T, Pan I, Gawthorpe RL, Fossen H (2018) Seismic facies analysis using machine learning. Geophysics 83:O83–O95
Xie F, Xiao C, Liu R, Zhang L (2017) Multi-threshold de-noising of electrical imaging logging data based on the wavelet packet transform. J Geophys Eng 14(4):900–908
Xu J, Li Y, Ren C, Wang S, Vanapalli SK, Chen G (2021) Influence of freeze-thaw cycles on microstructure and hydraulic conductivity of saline intact loess. Cold Reg Sci Technol 181:103183
Yuan S, Liu J, Wang S, Wang T, Shi P (2018) Seismic waveform classification and first-break picking using convolution neural networks. IEEE Geosci Remote Sens Lett 15(2):272–276

Yuan S, Jiao X, Luo Y, Sang W, Wang S (2022) Double-scale supervised inversion with a data-driven forward model for low-frequency impedance recovery. Geophysics 87(2):R165–R181

Zhou B, O'Brien G (2016) Improving coal quality estimation through multiple geophysical log analysis. Int J Coal Geol 167:75–92