ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940www.ijsem.org

Vol 7, Issuse.4 Oct2019

A Nonlinear Regression Application for Processing Geomagnetic Data Using Machine Learning Methods

[1]*Ch.DIVYA,*
[2]*A Shiva Rama Krishana,*
[3]*Gunde Chandeeshwar ,*

Abstract—

When it comes to near-surface exploration, finding unexploded ordnance, and other applications that rely on geomagnetic data, the accuracy and reliability of such data are key considerations. Based on machine learning methods, this research proposes a geomagnetic data reconstruction method for undersampled geomagnetic data. When compared to the conventional linear interpolation methods, the suggested methodology is more time efficient and lower in labour cost. The support vector machine, random forests, and gradient boosting models were all developed in this study. Recurrent neural network, a deep learning method, was also used to boost training performance. A continuous regression hyperplane was generated from a training dataset using the suggested learning methods. Using the provided regression hyperplane, the relationship between the mock-up missing data and the rest of the data is mapped out. Finally, the hyperplanes that were trained were utilised to rebuild the missing geomagnetic traces for validation, and they may be used to reconstruct further gathered fresh field data. In the end, numerical experiments were developed. When compared to the standard linear technique, our methods produced better results, with a reconstruction accuracy that was improved by 10% to 20%.

Index Terms— Modeling and reconstruction using a deep neural network with geomagnetic data

INTRODUCTION

At magnetic observations, continuous measurements of the geomagnetic field are made with a typical time period ranging from seconds to decades. It is possible that the reliability of geomagnetic data is not always guaranteed, particularly in the event of system errors [2]. If the data are under- or missing-sampled, the accuracy of the interpretation of the geomagnetic data is jeopardised [3], necessitating work on data reconstruction. Numerous approaches for reconstructing geomagnetic data have been devised to this point. [4] and [5] used numerical simulations to examine the data assimilation approach to try to forecast unknown geomagnetic field changes

.

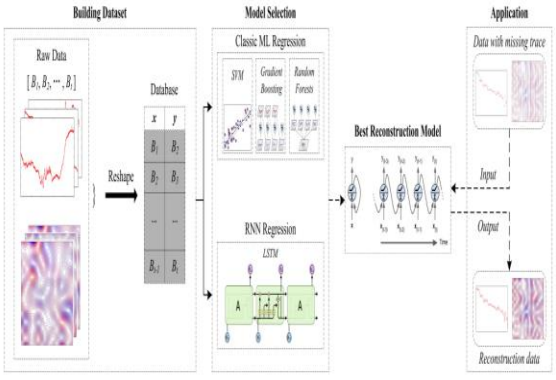
Assistant Professor

Assistant Professor

Assistant Professor

Department of CSE Engineering,
Pallavi Engineering College Hyderabad, Telangana 501505

Despite the positive outcomes of the simulation, the results remained in theoretical simulation form and had not yet been applied to a real-world situation. Verification of the performance is still pending. Other techniques have been used in real-world situations. For the assessment and observation of changes in the geomagnetic field, many models of global geomagnetic data have been suggested in [6–9], and their predictions (e.g., CALS3K) may be used as a baseline. The spherical harmonic approach [11] and multiple model fusion [12] are both well established methods. Good reconstruction results from these approaches, on the other hand, are contingent on a number of assumptions. There should be a restricted number of linear events in geomagnetic records, for example; in addition, reconstructed data should be sparser than the actual data with missing traces. Nondestructive testing [13], objective detection [14], and classification and linear regression [15] are some of the scientific domains where machine learning is rapidly being employed because it can automatically investigate the hidden characteristics or correlations in the data set. In many sectors, this is a viable solution to reducing the amount of physical labour required. Linear regression [16], decision trees [17], support vector machine (SVM) [18], artificial neural networks [19], and instance-based learning [20] are the basic machine learning technologies. Regression, classification, clustering, and other models are some of the most common ones conducted by machine learning. To summarise, this is an excellent product because of its high quality.



Machine learning has spread rapidly across multiple fields, suggesting that it is likely to lead the next wave of innovation in geophysics [21]. The proposed geomagnetic data reconstruction framework includes four main modules: data set building, classic machine learning regression, RNN regression, and the selected model for regression and reconstruction performance. How can we use machine learning to analyse geomagnetic data? Despite the above-mentioned uses, geomagnetic exploration is still in need of enhancement. Techniques comparable to this have been used to characterise reservoir properties at the preliminary stage[22], [23]. Reconstructing or interpolating geomagnetic data may be seen of as a regression issue for continuous output. An approximation function (a continuous hyperplane) may be generated using machine learning to recreate the missing traces. The three primary phases of the suggested technique are as follows: The process of creating a training data set is the initial stage. Because there are no missing traces in the training data, the model built from it is simply a mapping from the prior values, x(t), to the new values, x(t + 1). Training, validation, and testing data are randomly divided from the data set. A regression model (classic and RNN) is then trained on the training data to fit a hyperplane. Finally, the geomagnetic traces of the mock-up defective data set must be reconstructed. Validation and testing of the rebuilt data follows, and the most optimistic model is determined based on the precision of the results. Only the original data features are used in the proposed machine learning technique to address some of the disadvantages of linear events and sparsity of reconstruction data, for example. Aside from breaking the prior limits, this also demonstrates a greater capacity to adapt to diverse types of data sets. Using this technology, scientists can get all of Earth's magnetic field information for near-surface exploration and identification of magnetic minerals.

THEORY

Machine learning approaches have not yet been employed to recover the geomagnetic data that has been undersampled. Algorithms 1 and 2 provide a high-level overview of the training and inference algorithms in the Geomagnetic Data Reconstruction (GDR) framework, which is based on machine learning. Figure 1 depicts the framework's four components: 2) classic machine learning regression module, which is used to extract pattern representations from time series data and establish

deterministic models (refer to Algorithm 1); 3) recurrent neural network (RNN) regression module, which manipulates data from time series and is utilised for the generation of candidate memory models (refer to Algorithm 1); 4) neural network (RNN) regression model (refer to Figure 5) Reconstruction of geomagnetic data may be regarded as a regression issue from the standpoint of statistics. In the following descriptions, the rationale of regression is discussed. Let's assume that the training data set contains n pairs of data as xi, yi, where xi is the feature vector for each pair of data. In a feature vector, the values of the points around a target sample point are stored as an array of data. When testing or reconstructing, the value of the target sample point is derived from the feature vector. The associated label for xi is yi (the real value of the sample point), which is used to compare with the computed target sample point value to arrive at an accuracy. It's important to keep in mind that yi is a non-negative integer. F (x) must be constructed as an approximation of f (x) in order to solve regression problems. The dependent variable yi should be predicted from the independent variable xi. If there are missing geomagnetic data traces, they are used as input into the model training process as part of the vector xi. Because of this, the uncompleted magnetic field strength is used to determine the "ground truth" value of yi.

Classic Machine Learning Methods

The regression issue was solved using SVM, gradient boosting, and random forests, three standard machine learning models. The following is a succinct description of each model. a) Vector machine: b) vector machine In terms of its regression method, SVM incorporates all the essential traits that define the maximum margin algorithm, despite its origins as a classification system that maximises the margin of decision boundary: It is possible to train a nonlinear function using a linear learning machine that maps it into high-dimensional kernel induced feature space. The system's capacity is determined by parameters that are independent of the feature space's dimension. Using a fixed mapping strategy, the SVM regression model first maps the input variables x = [x1], [x2], [x3], [x4], [x5] into an n-dimensional space. This n-dimensional space is then used to build a linear regression model [18]. For this reason, gi(x) stands for a collection of nonlinear transformations, and the anticipated "bias" value, yi, may be expressed using mathematical notation as: SVM regression employs a loss function that is -insensitive for training purposes. In [24], you'll find more information. It is possible to create an ensemble regression model using numerous weak models using gradient boosting [26]. Additionally, it builds up new models in a forward-looking manner, optimising the differential of loss functions Approximation function f (x) mapping x to y is obtained using a training sample (xi, yi)n 1 of known (x, y) values. The loss function L(y, f (x)) is also minimised across the joint distribution of (x, y) values. A variety of common loss functions are used depending on the kind of issue being addressed, including the squared-error L(y, f (x)) and the absolute error |y f (x)|. The loss function in this study is the squared-error function. A regression tree is fitted to the negative gradient in each training step for a specified loss function. As a result, the approximation value may be anticipated using the vector of x with known values and matching values of the vector of y. A weighted sum of the function hi(x) may be approximated in the regression model of gradient boosting, which is known as base learners.

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \hat{y}_i g_i(\boldsymbol{x}) + b$$

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \gamma_i h_i(\boldsymbol{x}) + b$$

A separate optimum multiplier, L(yi, f (xi) +), may be found via line search [27]. It is a well-known machine learning technique that is built using a large number of decision trees. A regression model's projected value isn't dependent on any one tree's predictions, but rather on the average of all of the trees' predictions. Overfitting may be avoided by using a large number of decision trees. It's also possible to increase the accuracy of the forecasted value. When training a random forest model, the same bagging procedures are used for all tree learners [28]. Given a training variable called x1 and the ground truth called y1 and y2, the following equation

may be used:

Algorithm 1 Reconstruction Based on Classic Machine Learning

Result: Geomagnetic data reconstruction.
Input : Experimental geomagnetic data x, missing geomagnetic data y.
Output: Predicted geomagnetic data.

1 **Training:**
2 A data set is built where x stand for the strength of magnetic field at time (t) and y stands for the strength of magnetic field at time (t + 1);
3 Building different independent regression models:
4 $clf1 = svr()$
5 $clf2 = RandomForestRegression()$
6 $clf3 = GradientBoostingRegression()$
7 Input all the pairs of (x_t, x_{t+1}) to each of the models to train.

Testing

In order to choose the best model, trained models are applied to geomagnetic data with missing traces and the anticipated missing data is obtained.

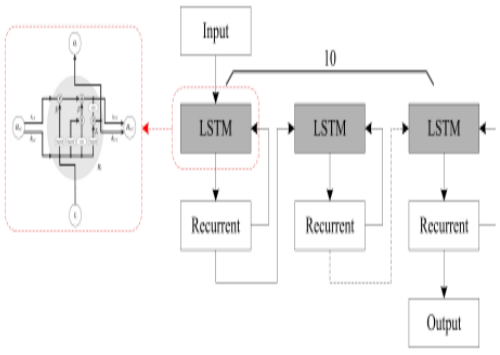
Geomagnetic data x with missing traces is applied to the trained models, resulting in the predicted missing data y, which is evaluated to find the most optimal model.

$$\hat{f} = 1/B \sum_{b=1}^{B} \hat{f}_b(x).$$

In addition, taking the majority vote of all the decision trees can also get the predicted values. The main steps in our reconstruction method using classic machine learning techniques are given in Algorithm 1. Once all the following steps have been successfully completed, the output f (x) is the associative function of each model. In practise, the following are the methods for predicting outcomes. Enter a new vector of geomagnetic data x with missing traces and convert it to a 1-dimensional vector. It records changes in the intensity of the magnetic field over time. Magnetic field strength over time (t)

In addition, the majority of the decision trees may be used to anticipate the values of the trees. Algorithm 1 outlines the basic phases of our machine learning-based reconstruction approach. The output associative function f (x) of each model will be produced after the following procedures have been successfully accomplished. In practise, the following processes are used to make predictions. New geomagnetic data must be entered and then transposed into one-dimensional vectors. It holds the values of magnetic field strength that change over time.. It is possible to predict the intensity of magnetic fields over time using the standard machine learning approaches, such as the Gauss-Boltzmann method. Assumption: All inputs and outputs are assumed to be completely independent of one another. But in the reconstruction work, it would be beneficial to know the previous values in order to forecast the upcoming one. Because RNNs do the same job for every time-series vector, the result is statistically reliant upon the prior computations, they are recurrent. As a result, the values x ti, x ti+1,..., x t1 are well-known and may be used consecutively in the model function to produce an LSTM when RNNs are used. As a result, in addition to x t1, the prior inputs have an effect on x t as well. Furthermore, the network may be used to forecast the next value xt+1, which is generated by x ti, x ti+1, etc. xt and the anticipated value xt. The remaining steps may be completed iteratively in the same way. The more values that are already known, the more accurate it is to forecast new values.Deep

4 model = Sequential() 5 model.add (LST M(hi dden layer, input_shape = 6 (vi siblelayer,look_back))) 7 model. fit(x, y, epochs = 100) 8 The input vi sible layer is set as 1 while the hi dden layer contains 10 LSTM blocks. The out put layer makes value predictions based on the look_back, which it defaulted to 1; epochs stands for the number of overall times for the training vectors; 9 Fit all pairs of (xt , xt+1) to the network, which is trained for a epochs of 100. 10 Testing: 11 Input the missing geomagnetic data x to the trained LSTM network to obtain the predicted missing data y



LSTM RNN architecture is shown in Fig. 4.

The point pairs (variable x and ground truth y) with completed data are known for the training geomagnetic data in Algorithms 1 and 2, and the GDR framework paired with our algorithms is utilised to develop a best practise regression model y = f (x) concealed in these training point pairs. After feeding x into this trained model, it is possible to estimate the value of y, the other missing data in (x, y). We used examples of geomorphological structures that were comparable to the geomagnetic data recreated in the training phase. To train each of the regression models (SVM, Gradient Boosting, Random Forests, and LSTM), we input the training data into the database. Then, four continuous regression models may be constructed, and they can also be preserved for future usage in the form of reconstruction. It is possible to make educated guesses about the missing geomagnetic data at this level by looking at the data that has come before or after it in the timeline. The regression model is run with all of x's missing values.

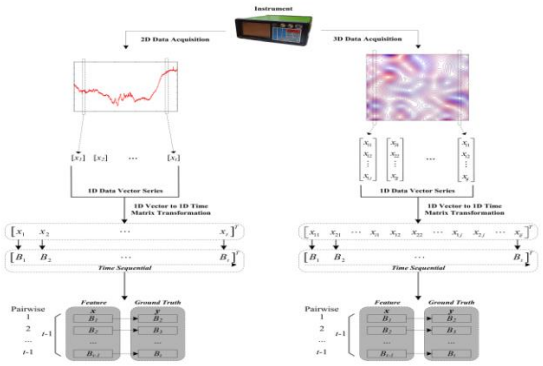


Fig. 5. Procedure of preprocess geomagnetic time-series data

that have been trained in the previous stage, thus allowing us to obtain the missing geomagnetic data

EXPERIMENTAL RESULTS AND DISCUSSION

First, a data set is built; next, the regression model is trained; then the missing geomagnetic traces are recovered, and finally, the model is tested using fresh data sets to verify the findings. R-squared [39] and RMSE [39] were both utilised to assess the accuracy of various modelling approaches in each experiment.

Equations like these [13] are used in their work.

$$\begin{cases} \mathrm{RMSE} = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \\ R_{-\mathrm{squared}} = 1 - \dfrac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \end{cases}$$

There is a difference between R-squared and RMSE. The smaller the RMSE, the more accurate the model is. R-squared has a range of zero to one, and the closer to one it is, the better the model is at predicting. Using Python scikit-learn and Keras on an Intel Core i7 with a 3.2GHz clock speed and 16GB of random access memory, the experiments and numerical analysis were carried out (RAM).

Geomagnetic Data Preprocessing

For using machine learning models in geomagnetic data reconstruction, it is necessary to include a rule to transform the geomagnetic data into the form of data pairs (feature, ground truth), which should be suitable for the input of each model. The overall data preprocessing flowchart of our proposed method is shown in Fig. 5. Typically, for a set of 2-D data, first, the input variables x = {x1, x2,..., xt} whose t subscript indicates the order of the data collected. The values in it are the magnetic field strength values varying with respect to time. Then, x can be transposed into a 1-D column vector [x1, x2,..., xt] T directly

TABLE I

MACHINE LEARNING PARAMETERS USED FOR EACH OF THE DIFFERENT MODELS

Model	Main parameter definitions
SVM	cache_size=200, degree=3, gamma='auto', kernel='rbf', decision_function_shape='ovo'
Gradient Boosting	learning_rate=80, criterion='mse', n_estimators=100, max_depth=3, min_samples_leaf=10
Random Forests	random_state=80, criterion='entropy', n_estimators=100, max_depth=3, min_samples_leaf=20
LSTM	visible_layer=1, hidden_layer=10, look_back=1, epochs=100

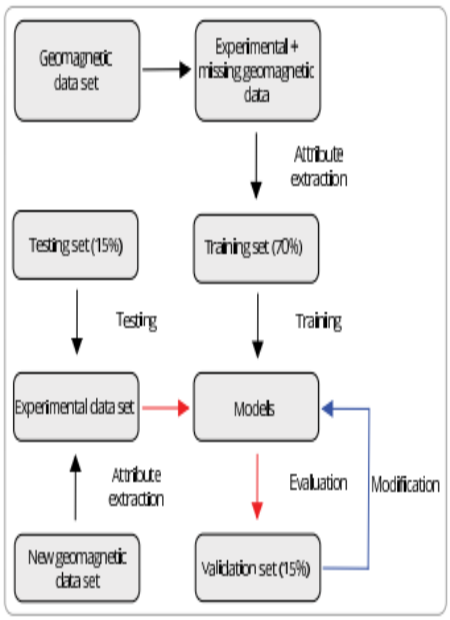


Fig. 6. Flowchart of training a regression model

There are two spatial dimensions I and j) in 3-D, which necessitates a different approach to creating the database than in 2-D. Each magnetic field strength at each place is represented by a sequential pixel along the time axis, whereas I and j are each the number of pixels along the row and column, respectively. The original data may be depicted as a series of photographs. For example, the input variables in this situation may be summarised as follows: the first two variables are x11 (the first one), and the second one (the second one), and the third one (the third one). It begins by reshaping 3-D sequential information in multiple 1-D vectors. A 1-D column vector ([x11, x12,...]) T is then produced by taking the 1-D space (x11–xij) and sorting it. Reshaped data is thus an ensemble of pixels from the original dataset. The width I times the height j of the picture is used to calculate the number of samples. A two- or three-dimensional geomagnetic data set is then transformed into an array, with each element representing one location's measured magnetic field strength (measured in Bt units), and each element representing the location's measured magnetic field strength (measured in Bt+1 units). An array is a collection of related data values. Additional emphasis should be placed on one important point: the geomagnetic data have their own database B (training separately) based on their own time sequence during measurement.

Training Model

For reconstruction of missing geomagnetic traces, we utilized three classic machine learning models and one typ

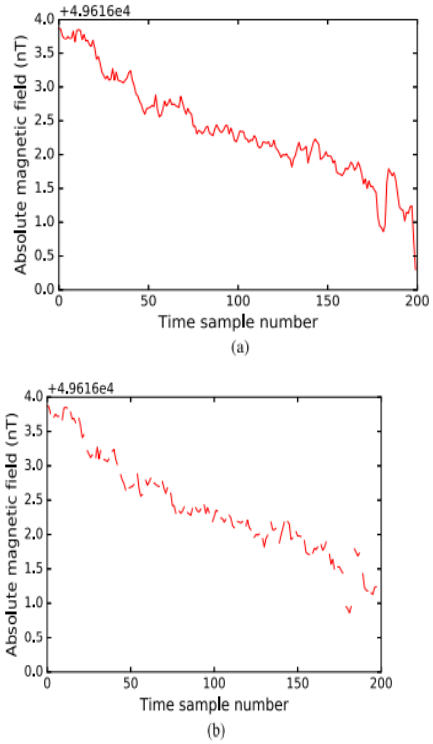
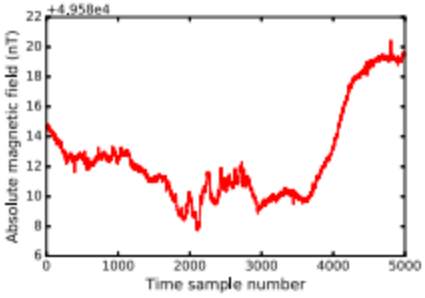
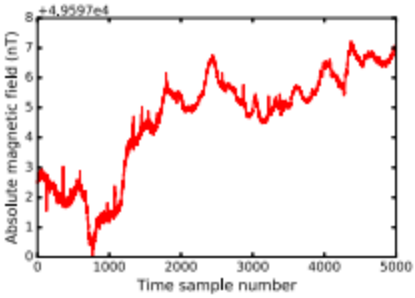


Fig. 7 shows the 2-D field geomagnetic data used for the experiment. Geomagnetic field data from the original 2-D model. data with 30 percent normal missing traces in decimated form

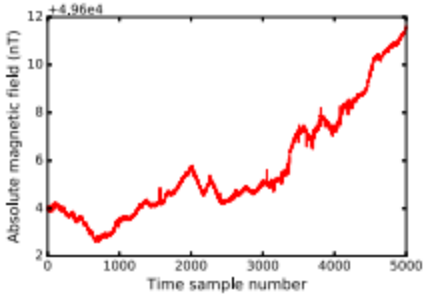
data instances with multiple labels may be learned from using a deep learning model. Machine learning models assigned the occurrences of a certain event the label of input variables. The four regression models stated before were applied and tested in order to obtain the best relative optimist model. These trained models were also subjected to a slew of comparative studies including various geomagnetic data sets. Table I summarises all parameter definitions for each Python-based machine learning model. 1) Preparation: When it comes to machine learning, training is the most important step. For the purpose of selecting the intrinsic parameters of the regression model, 70% of the geomagnetic data set was utilised to train the model's input variables. Cross-validation was allocated for 15% of the geomagnetic data, and performance testing was reserved for another 15%.



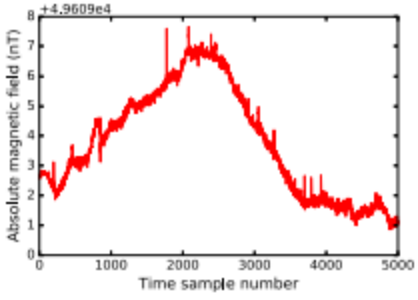
(a)



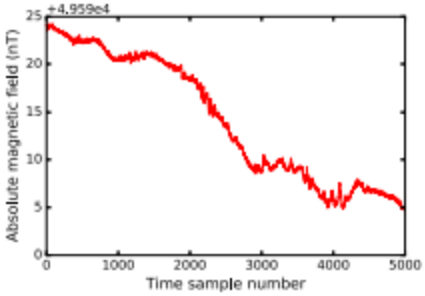
(b)



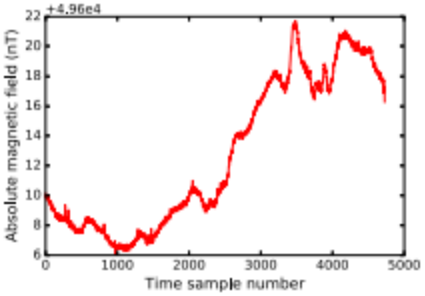
(c)



(d)

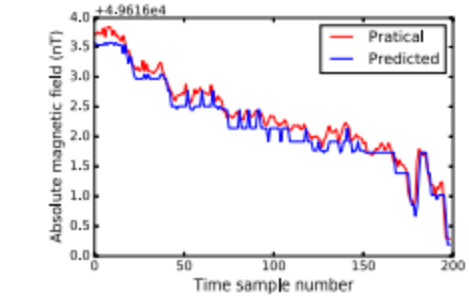


(e)

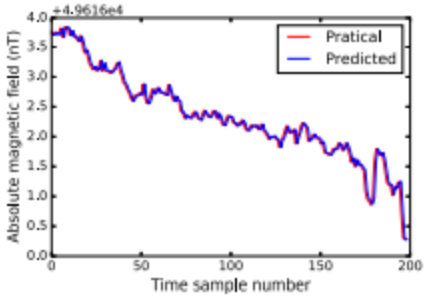


(f)

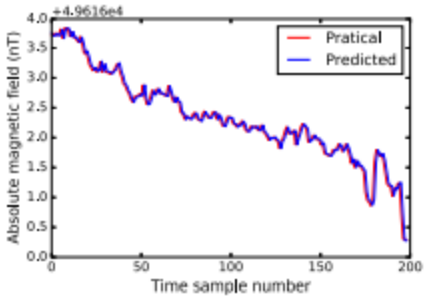
Fig. 8. Six 2-D field geomagnetic data used to feed into the training model. (a) 1st dataset. (b) 2nd dataset. (c) 3rd dataset. (d) 4th dataset. (e) 5th dataset. (f) 6th dataset.



(a)

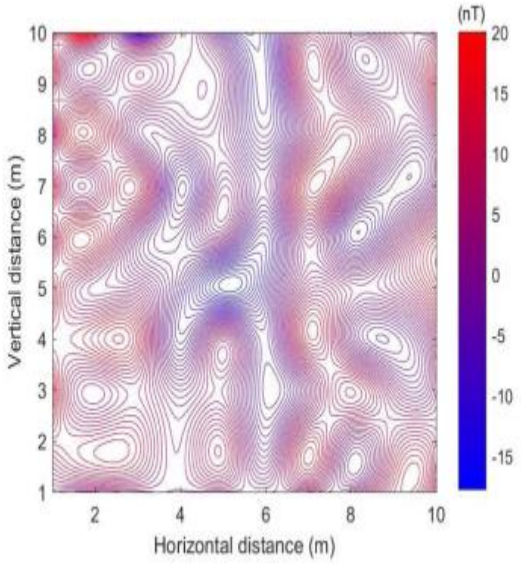


(b)

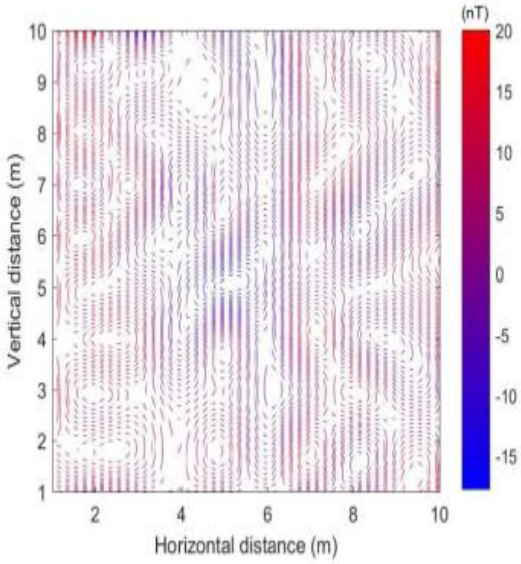


(b)

each one of them. Machine learning methods may be used to understand the relationship between the missing and finished traces in this training data set. The values of the missing traces might then be predicted using a trained model. 2) Validation: A cross validation was used to remove the overfitting issue once the model training was completed using the training data set. According to the previous paragraph, 15% of the geomagnetic data set was utilised as the validation dataset. By cross-validating each trained model with this dataset, the effectiveness of the model can be observed. Parameters should be tweaked to enhance the models' performance on the validation dataset if the performance of the trained models is not satisfactory. Following this step, each of the trained models was retrained, as well, and the revalidations were implemented on the same data sets as the retrained models, respectively As soon as the machine learning models attained a satisfactory level of performance, this validation procedure would be terminated. 3) Testing: All of the regression models previously described could be acquired when training and validation were complete. As a result, the testing data set was put to use to assess each model's overall performance. Testing was allowed on 15% of the geomagnetic data. This is an honest and efficient evaluation since the testing results were never utilised.



(a)



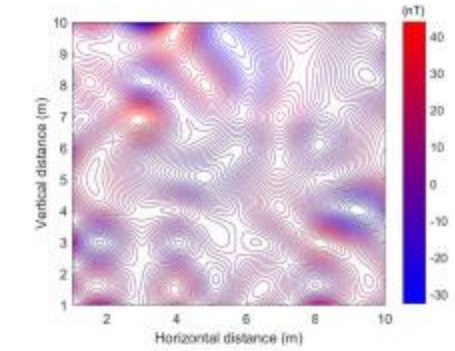
(b)

Fig. 10. 3-D field geomagnetic data for test. (a) Original 3-D field geomagnetic data. (b) Decimated data with 50% regular missing traces.

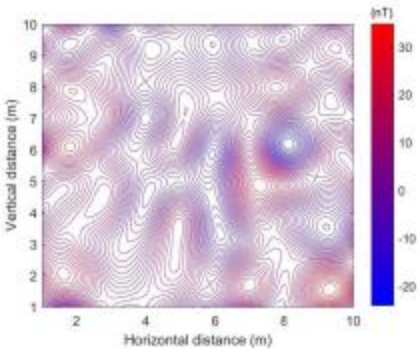
should be fed into each of the trained machine learning models for reconstruction and subsequent interpolation.C. 2-D Example A 2-D completed field geomagnetic data are shown in Fig. 7(a), which is a subset of the data set that was

TABLE II
COMPARISON RESULTS USING DIFFERENT METHODS FOR 2-D DATA

Model	dataset	R-squared	RMSE
Linear	Testing	**0.788**	**0.279**
SVM	Training	0.899	0.188
	Validation	0.889	0.198
	Testing	0.878	0.208
Random Forests	Training	0.896	0.179
	Validation	0.885	0.200
	Testing	0.873	0.220
Gradient Boosting	Training	0.893	0.185
	Validation	0.883	0.204
	Testing	0.872	0.223
LSTM	Training	**0.988**	**0.094**
	Validation	**0.983**	**0.108**
	Testing	**0.978**	**0.122**



(a)



(b)

CONCLUSION

Both machine learning and deep learning were used in this research to reconstruct geomagnetic data. There is a hidden connection (continuous hyperplane) that may be discovered given sufficient example training sets. To enhance the performance of classical machine learning, we introduce the RNN approach (SVM, gradient boosting, and random forests). Another benefit is its universal applicability to a variety of data sets, which we were able to avoid with the past shortcomings of the current reconstruction approaches. It is also possible to save the trained regression model in order to utilise it in the future to recreate the geomagnetic data with a similar geomorphological pattern. Overall, the testing findings revealed that the suggested approach was able to obtain a reconstruction accuracy of more than 90%, which was an improvement of nearly 20% over the conventional method. While deep learning approaches exhibited a high degree of accuracy, For geomagnetic data reconstruction, we used both machine learning and deep learning techniques. Using sufficient exemplary training sets, a hidden connection (continuous hyperplane) may be discovered, and the missing data can be extrapolated from this. The RNN approach is presented here to enhance the performance of the classical machine learning method (SVM, gradient boosting, and random forests). It is worth noting that the LSTM-based methodology was able to bypass past shortcomings in current reconstruction approaches, and it can be used to a wide variety of data sets. It is also possible to save the trained regression model in order to reassemble geomagnetic data with a comparable geomorphological structure in the future. It was shown that the suggested approach can achieve a reconstruction accuracy of more than 90%, which is a 20% improvement over the previous method in terms of the accuracy. The accuracy of deep learning algorithms was shown, although it still has room for improvement.

REFERENCES

[1] S. Khomutov, V. Sapunov, A. Denisov, D. Savelyev, and I. Babakhanov, “Overhauser vector magnetometer POS-4: Results of continuous measurements during 2015–2016 at geophysical observatory ‘Paratunka’ of IKIR FEB RAS, Kamchatka, Russia,” in Proc. E3S Web Conf., vol. 11. Les Ulis, France: EDP Sciences, 2016, Art. no. 00007.

[2] H. Dong, H. Liu, J. Ge, Z. Yuan, and Z. Zhao, “A high-precision frequency measurement algorithm for

FID signal of proton magnetometer," IEEE Trans. Instrum. Meas., vol. 65, no. 4, pp. 898–904, Apr. 2016.

[3] R. M. Winslow et al., "Observations of Mercury's northern cusp region with MESSENGER's Magnetometer," Geophys. Res. Lett., vol. 39, no. 8, p. L08112, 2012. [4] F. Lhuillier, J. Aubert, and G. Hulot, "Earth's dynamo limit of predictability controlled by magnetic dissipation," Geophys. J. Int., vol. 186, no. 2, pp. 492–508, 2011.

[5] G. Hulot, F. Lhuillier, and J. Aubert, "Earth's dynamo limit of predictability," Geophys. Res. Lett., vol. 37, no. 6, p. L06305, 2010.

[6] M. Dumberry and C. C. Finlay, "Eastward and westward drift of the Earth's magnetic field for the last three millennia," Earth Planet. Sci. Lett., vol. 254, nos. 1–2, pp. 146–157, 2007.

[7] M. Korte, F. Donadini, and C. Constable, "Geomagnetic field for 0–3 ka: 2. A new series of time-varying global models," Geochem., Geophys., Geosyst., vol. 10, no. 6, p. Q06008, 2009.

[8] I. Wardinski and M. Korte, "The evolution of the core-surface flow over the last seven thousands years," J. Geophys. Res., Solid Earth, vol. 113, no. B5, p. B05101, 2008. [9] M. Korte and R. Holme, "On the persistence of geomagnetic flux lobes in global Holocene field models," Phys. Earth Planet. Interiors, vol. 182, nos. 3–4, pp. 179–186, 2010.

[10] M. Korte, C. Constable, F. Donadini, and R. Holme, "Reconstructing the Holocene geomagnetic field," Earth Planet. Sci. Lett., vol. 312, no. 3, pp. 497–505, 2011.

[11] A. Nilsson, R. Holme, M. Korte, N. Suttie, and M. Hill, "Reconstructing Holocene geomagnetic field variation: New methods, models and implications," Geophys. J. Int., vol. 198, no. 1, pp. 229–248, 2014. [12] L. Kapper, F. Donadini, V. Serneels, E. Tema, A. Goguitchaichvili, and J. J. Morales, "Reconstructing the geomagnetic field in West Africa: First absolute intensity results from Burkina Faso," Sci. Rep., vol. 7, Mar. 2017, Art. no. 45225.

[13] H. Liu, S. Liu, Z. Liu, N. Mrad, and H. Dong, "Prognostics of damage growth in composite materials using machine learning techniques," in Proc. IEEE Int. Conf. Ind. Technol. (ICIT), Mar. 2017, pp. 1042–1047.

[14] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: Machine learning and application to spontaneous behavior," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 2, Jun. 2005, pp. 568–573.

[15] A. M. Prasad, L. R. Iverson, and A. Liaw, "Newer classification and regression tree techniques: Bagging and random forests for ecological prediction," Ecosystems, vol. 9, no. 2, pp. 181–199, 2006.

[16] J. Neter, M. H. Kutner, C. J. Nachtsheim, and W. Wasserman, Applied Linear Statistical Models, vol. 4. Chicago, IL, USA: Irwin Chicago, 1996.

[17] S. K. Murthy, "Automatic construction of decision trees from data: A multi-disciplinary survey," Data Mining Knowl. Discovery, vol. 2, no. 4, pp. 345–389, Dec. 1998.

[18] O. Chapelle and V. Vapnik, "Model selection for support vector machines," in Proc. NIPS, 1999, pp. 230–236.

[19] S. Haykin and N. Network, "A comprehensive foundation," Neural Netw., vol. 2, p. 41, Feb. 2004.

[20] D. M. Dutton and G. V. Conroy, "A review of machine learning," Knowl. Eng. Rev., vol. 12, no. 4, pp. 341–367, 1997.

[21] Y. Jia and J. Ma, "What can machine learning do for seismic data processing? An interpolation application," Geophysics, vol. 82, no. 3, pp. V163–V177, 2017