Journal of Rock Mechanics and Geotechnical Engineering xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Rock Mechanics and Geotechnical Engineering

journal homepage: www.jrmge.cn

Full Length Article

An improved permeability estimation model using integrated approach of hybrid machine learning technique and shapley additive explanation

Christopher N. Mkono [a, b], Chuanbo Shen [a, *], Alvin K. Mulashani [c], Patrice Nyangi [d]

[a] Key Laboratory of Tectonics and Petroleum Resources of Ministry of Education, China University of Geosciences, Wuhan, 430074, China
[b] Department of Petroleum Engineering, School of Earth Resources, China University of Geosciences, Wuhan, 430074, China
[c] Department of Geoscience and Mining Technology, College of Engineering and Technology, Mbeya University of Science and Technology, Box 131, Mbeya, Tanzania
[d] Civil Engineering Department, College of Engineering and Technology, Mbeya University of Science and Technology, Box 131, Mbeya, Tanzania

ARTICLE INFO

Article history:
Received 9 February 2024 Received in revised form
6 July 2024
Accepted 1 September 2024
Available online 21 September 2024

Keywords:
Permeability
Hydrocarbon
Differential evolution
Shapley additive explanation (SHAP)
Group method of data handling
Well logs

ABSTRACT

Accurate reservoir permeability determination is crucial in hydrocarbon exploration and production. Conventional methods relying on empirical correlations and assumptions often result in high costs, time consumption, inaccuracies, and uncertainties. This study introduces a novel hybrid machine learning approach to predict the permeability of the Wangkwar formation in the Gunya oilfield, Northwestern Uganda. The group method of data handling with differential evolution (GMDH-DE) algorithm was used to predict permeability due to its capability to manage complex, nonlinear relationships between variables, reduced computation time, and parameter optimization through evolutionary algorithms. Using 1953 samples from Gunya-1 and Gunya-2 wells for training and 1563 samples from Gunya-3 for testing, the GMDH-DE outperformed the group method of data handling (GMDH) and random forest (RF) in predicting permeability with higher accuracy and lower computation time. The GMDH-DE achieved an R^2 of 0.9985, RMSE of 3.157, MAE of 2.366, and ME of 0.001 during training, and for testing, the ME, MAE, RMSE, and R^2 were 1.3508, 12.503, 21.3898, and 0.9534, respectively. Additionally, the GMDH-DE demonstrated a 41% reduction in processing time compared to GMDH and RF. The model was also used to predict the permeability of the Mita Gamma well in the Mandawa basin, Tanzania, which lacks core data. Shapley additive explanations (SHAP) analysis identified thermal neutron porosity (TNPH), effective porosity (PHIE), and spectral gamma-ray (SGR) as the most critical parameters in permeability prediction. Therefore, the GMDH-DE model offers a novel, efficient, and accurate approach for fast permeability prediction, enhancing hydrocarbon exploration and production.

1. Introduction

Hydrocarbon reservoirs are complex geological systems that contain oil and gas in porous and permeable rock formations. The accurate determination of permeability, which is a measure of a rock's ability to transmit fluids, is a critical factor in hydrocarbon exploration, CO_2 injection, enhanced oil recovery, petroleum production and hydrogen storage (Zhang and Wu, 2023). Permeability is used to estimate the amount of fluid that can be produced from a reservoir and to optimize the design of production wells (Al-Amri, 2017; Lee et al., 2021; Jiang et al., 2022; Liu et al., 2023; Masroor et al., 2023; Sun et al., 2023). This parameter is crucial for assessing reservoirs, making management decisions, and determining the most effective strategy for enhanced oil recovery. The primary factors determining the permeability at certain points are influenced by the porosity, pore geometry, wettability, and water saturation of the reservoir rock formations (Mahdaviara et al., 2020; Zakharov et al., 2022).

Conventional methods for permeability prediction such as empirical correlations, Archie's law, Kozeny-Carman equation, and

* Corresponding author.
E-mail address: cbshen@cug.edu.cn (C. Shen).
Peer review under responsibility of Institute of Rock and Soil Mechanics, Chinese Academy of Sciences.

https://doi.org/10.1016/j.jrmge.2024.09.013

Table 1
The previous ML methods for permeability prediction and their limitations.

Method	Limitation	Source
Support vector machine (SVM)	Difficult in handling large dataset	Anifowose et al. (2019)
Group method of data handling (GMDH)	Model complexity; Overfitting	Mathew Nkurlu et al. (2020)
Genetic algorithm (GA)	May lead to overfitting	Al Khalifah et al. (2020)
Deep residual neural network (DRNN)	Prone to overfitting; Problem of gradient disappearance	Zhang et al. (2021a)
Extreme gradient boosting (XGBoost)	Computationally expensive; Complex interpretability	Mohammadian et al. (2022)
Multilayer perception with social ski-driver (MLP-SSD)	Slow convergence; Computationally expensive	Matinkia et al. (2023)
Stacked extreme learning machine (ELM)	Time consuming; Computationally expensive	Kalule et al. (2023)
Generalized additive model (GAM)	Prone to overfitting; Computationally expensive	Mahdy et al. (2024)

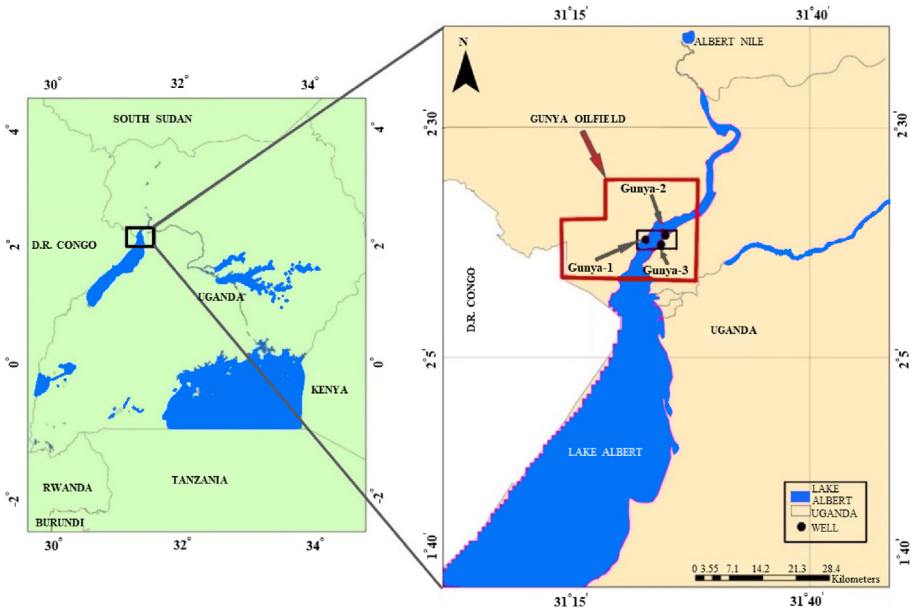

Fig. 1. Location of the Gunya oilfield showing three studied wells.

flow-zone indicator (FZI) analysis (Archie, 1942; Amaefule et al., 1993; Uguru et al., 2005; Costa, 2006; Rashid et al., 2015) are employed in the oil and gas industry on a large scale. However, these methods have some limitations. Empirical correlations rely on statistical relationships between core permeability measurements and petrophysical properties, making them potentially inapplicable in different geological settings (Wang, 2019; Belhouchet et al., 2021; Makarian et al., 2023). Archie's law assumes electrical homogeneity and known fluid saturations (Zhang et al., 2021b), which may not always hold true in permeability determination. The Kozeny-Carman equation is restricted to homogeneous and isotropic formations, failing to account for rock heterogeneity and anisotropy (Wong et al., 2000; Shokir et al., 2006). FZI method is a time-consuming process that requires core analysis, and the results may not be representative of the entire reservoir (Mirzaei-Paiaman et al., 2015, 2020). In contrast, conventional methods relying on empirical correlations and assumptions for permeability prediction from petrophysical logs can introduce inaccuracies, uncertainties, high operational cost, and time consumption (Mohammadlou and Mørk, 2012; Zhong and Carr, 2019; Martyushev et al., 2023b).

To overcome these limitations, there is an increasing interest in applying machine learning for permeability prediction from petrophysical logs (Akande et al., 2017; Elkatatny et al., 2018; Zhang et al., 2018; Zhong et al., 2019; Subasi et al., 2020; Kabwe, 2022; Sun et al., 2022; Tian et al., 2022; Martyushev et al., 2023a; Rashid et al., 2023; Xu et al., 2023). Machine learning algorithms can optimize model parameters to increase accuracy and capture the intricate, nonlinear correlations between petrophysical characteristics and permeability. Sheykhinasab et al. (2023) used a multilayer extreme learning machine coupled with the cuckoo optimization algorithm (MELM-COA) to predict the permeability of highly heterogeneous reservoirs from petrophysical logs. Rezaee and Ekundayo (2022) explored machine learning techniques for permeability prediction using well logs and core data from five Surat Basin boreholes in Australia. The study revealed that the artificial neural network (ANN), utilizing all seven input logs, achieved the highest performance. It demonstrated an R^2 of about 0.93 during training and 0.87 during testing. Notably, the research did not incorporate data standardization or feature engineering, potentially limiting the efficacy of other machine learning models and hindering their accuracy in permeability prediction. Another study by Matinkia et al. (2023) predicted permeability from well logs by combining multilayer perception with social ski-driver (MLP-SSD) although their developed model encountered a slow convergence problem and it was computationally expensive to run the model. Nazari and Hajizadeh (2023) estimated the permeability by using an intelligent method called relevance vector regression with grey wolf optimization (RVR-GWO) in the Azadegan oilfield, Iran. Moosavi et al. (2022) applied the fuzzy support vector regression (FSVR) to predict permeability. Liu and Liu (2022) proposed a particle swarm optimization-optimized XGBoost model to

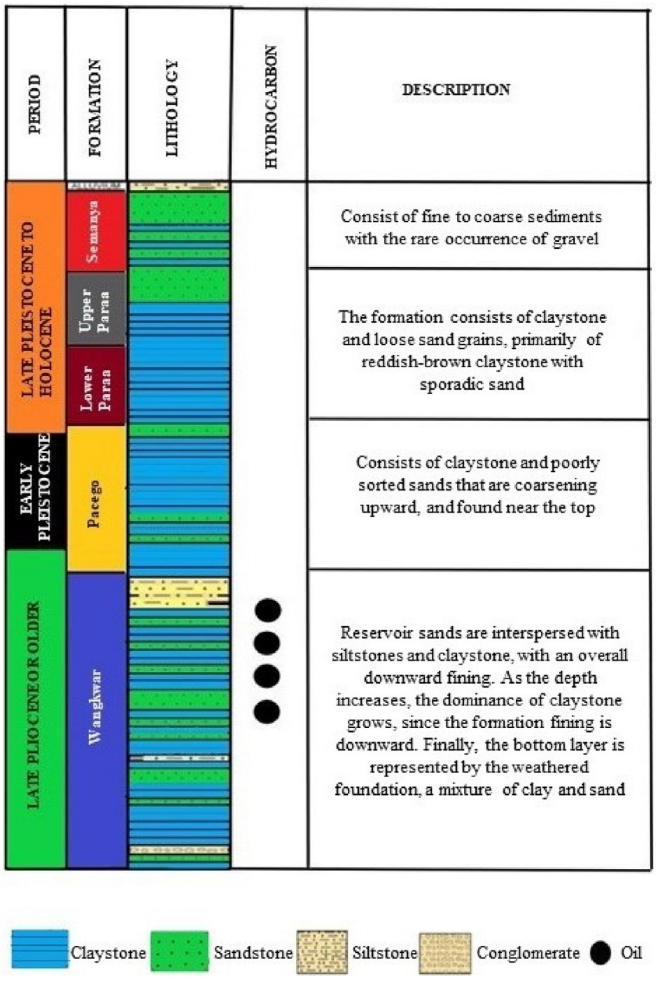

Fig. 2. Stratigraphic column of Gunya Oilfield.

forecast the permeability of confined sandstone reservoirs.

Recently, GMDH has shown promising results in the field of petroleum, especially in the estimation of permeability. For instance, Mathew Nkurlu et al. (2020) estimated the permeability of the hydrocarbon reservoir by using the GMDH model with 6 inputs and 366 datasets. Moreover, Mulashani et al. (2022) enhanced the GMDH in estimating permeability by using five well-log inputs and 1356 datasets from the Mpyo oilfield. Zanganeh Kamali et al. (2022) applied the GMDH model and 212-point dataset from heterogeneous carbonate gas condensate reservoirs to predict permeability. Despite its usefulness, the GMDH method still faces challenges such as setting the optimal dataset partition and removing effective input parameters which can produce outliers (Buryan, 2006). It can lead to problems with data convergence and overfitting due to the use of generic architecture (Park et al., 2004). It requires the use of assumptions to establish initial parameters during model training, which can consume time and lead to low accuracy (Oh and Pedrycz, 2006). Table 1 summarizes the previous ML methods used to predict permeability with their encountered limitations.

In this study, we propose for the first time a novel approach for permeability estimation by integrating a hybrid machine learning technique and well logs. The hybrid machine learning model proposed in this study is the group method of data handling (GMDH) optimized by differential evolution (DE) (GMDH-DE). The GMDH-DE is a powerful and versatile algorithm that can overcome the challenge of data overfitting, uses minimal computational time, and handles complex, nonlinear relationships between input variables and output variables. GMDH-DE is a variant of the GMDH algorithm, which is an adaptive network-based machine learning technique that uses a self-organizing algorithm to select the best subset of input variables for predicting the output variable. GMDH-DE extends the GMDH algorithm by incorporating DE to optimize the model parameters. The computational efficiency is increased with the hybridization of GMDH and DE. As presented in Table 1, the proposed GMDH-DE technique has not previously been utilized for permeability prediction. Furthermore, the performance of the GMDH-DE was compared to GMDH and RF, and it was observed that the GMDH-DE outperforms the others in terms of

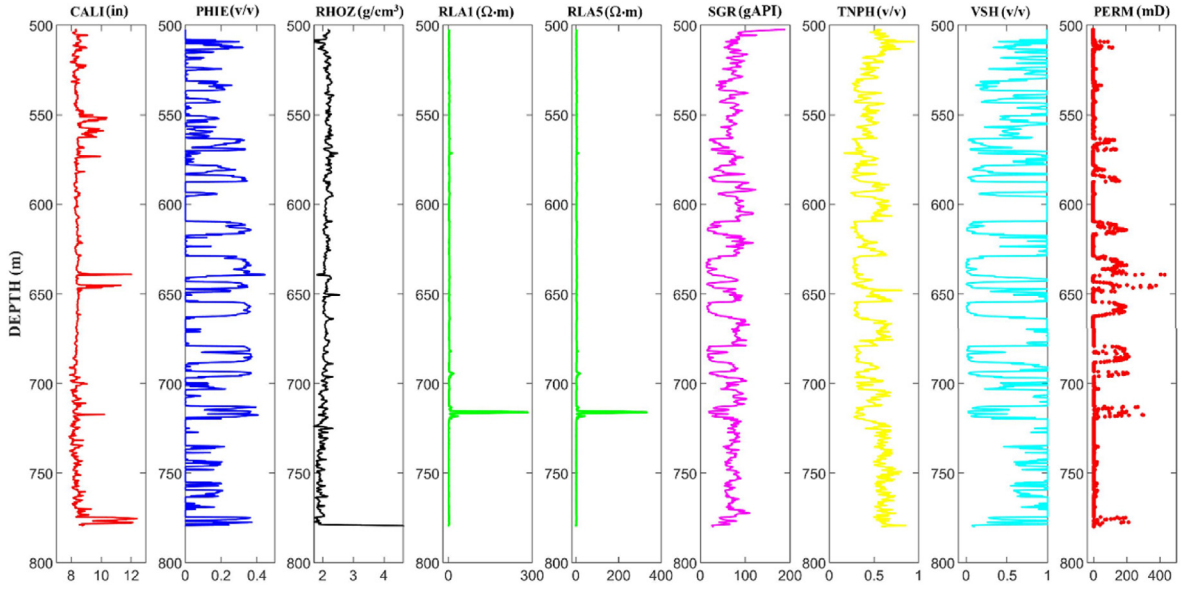

Fig. 3. Sample input well logs and permeability used for model development.

Table 2
Statistical parameters of inputs well logs.

Well	Variable	CALI (in)	PHIE (v/v)	RHOZ (g/cm^3)	RLA1 (Ω m)	RLA5 (Ω m)	SGR (gAPI)	TNPH (v/v)	VSH (v/v)
Gunya-1	Count	1053	1053	1053	1053	1053	1053	1053	1053
	Mean	5.58	0.321	2.292	2051.337	2051.446	46.855	0.803	0.283
	Minimum	3.917	0.009	1.766	0.0033	0.006	17.386	0.273	0.008
	Maximum	10.525	0.491	3.942	100,000	100,000	78.632	2.113	0.944
	Standard deviation	2.298	0.136	0.422	14175	14175	12.567	0.222	0.236
Gunya-2	Count	900	900	900	900	900	900	900	900
	Mean	8.364	0.228	2.153	7.722	8.359	46.493	0.354	0.302
	Minimum	7.884	0.011	1.955	1.93	1.951	15.403	0.186	0.001
	Maximum	10.257	0.383	2.45	176.081	191.373	78.897	0.737	0.95
	Standard deviation	0.159	0.103	0.092	18.545	19.847	19.841	0.094	0.287
Gunya-3	Count	1563	1563	1563	1563	1563	1563	1563	1563
	Mean	8.829	0.189	2.121	6.146	6.547	48.993	0.425	0.392
	Minimum	7.899	0.002	1.478	0.125	0.087	13.241	0.165	0.0005
	Maximum	15.928	0.445	4.149	491.265	499.749	78.913	1.493	0.95
	Standard deviation	0.672	0.118	0.238	22.408	23.962	17.731	0.16	0.293

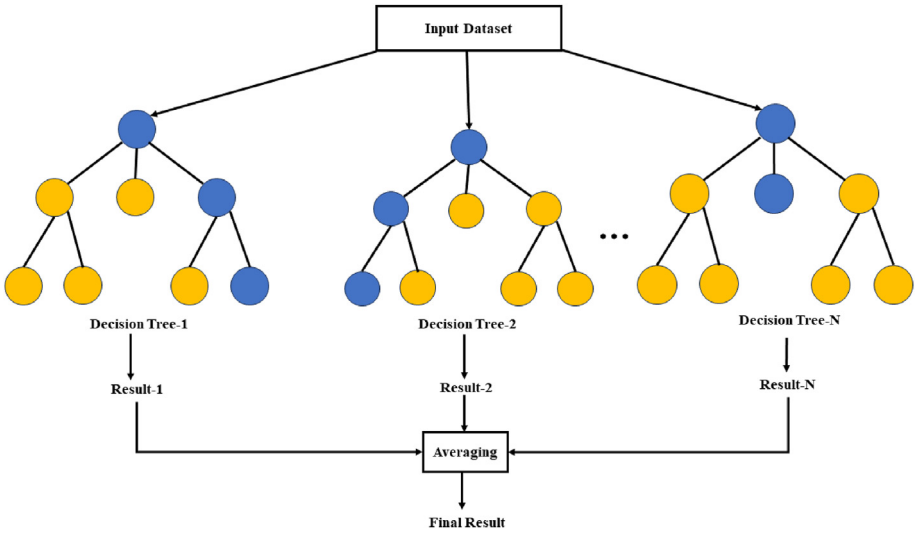

Fig. 4. Schematic of the RF method.

computational time and accuracy. Moreover, parameter importance was evaluated by using Shapley additive explanation (SHAP) analysis to determine which input parameter has the most effect on GMDH-DE model building. Finally, the developed GMDH-DE model was applied to predict permeability in an uncored well of the Mita gamma in the Mandawa basin, Southern Tanzania.

2. Geological setting and lithostratigraphy

Gunya oilfield is situated in the northwestern region of Uganda, on the onshore of Lake Albert (Fig. 1). The field extends to the northeast of the Victoria Nile region, which is situated near the Albertine Graben (Tonny et al., 2015; Simon et al., 2017; Guma et al., 2021). The Albertine Graben, which lies across the border of Uganda and the Democratic Republic of Congo, is a 500 km-long rift basin of Mesozoic-Cenozoic origin. It sits on top of the African Craton's Precambrian orogenic belts and is surrounded by steep normal faults (Dou et al., 2004a; Mutebi et al., 2021). The uplifted flanks of these faults are made up of Precambrian basement rocks, including gneisses, quartzites, and matie intrusions (Dou et al., 2004b). The region is encircled by main rift faults, with the strata gradually sinking towards the southwest. The predominant sedimentary rocks in the area are clastic sediments with tertiary to quaternary ages, and it is believed that the major lithologies are influenced by the climate. Fine to coarse sediments are found in the Semanya formation, with the rare occurrence of gravel. The Upper Paraa Formation consists of claystone and loose sand grains, whereas the Paraa Formation is made up primarily of brown to reddish-brown claystone with sporadic sand (Fig. 2). Sand in the Pacego Formation is badly sorted, coarsening upward, and found near the top (Lukaye et al., 2015). In Wangkwar, reservoir sands are interspersed with siltstones and claystone, with an overall downward fining. As the depth increases, the dominance of claystone grows, since the formation fining is downward. Finally, the bottom layer is represented by the weathered foundation, a mixture of clay and sand.

3. Methodology

3.1. Data description and pre-processing

This study focused on three wells, Gunya-1, Gunya-2 and Gunya-3 located in the Gunya oilfield in Uganda. These wells are located in the eastern portion of Ondyek and the southern portion of Job East discovery. A total of 3516 samples were collected from three wells (1053 for Gunya-1, 900 for Gunya-2, and 1563 for Gunya-3). Collected data comprise of core permeability and nine well logs of spectral gamma-ray (SGR), apparent resistivity focusing mode 1

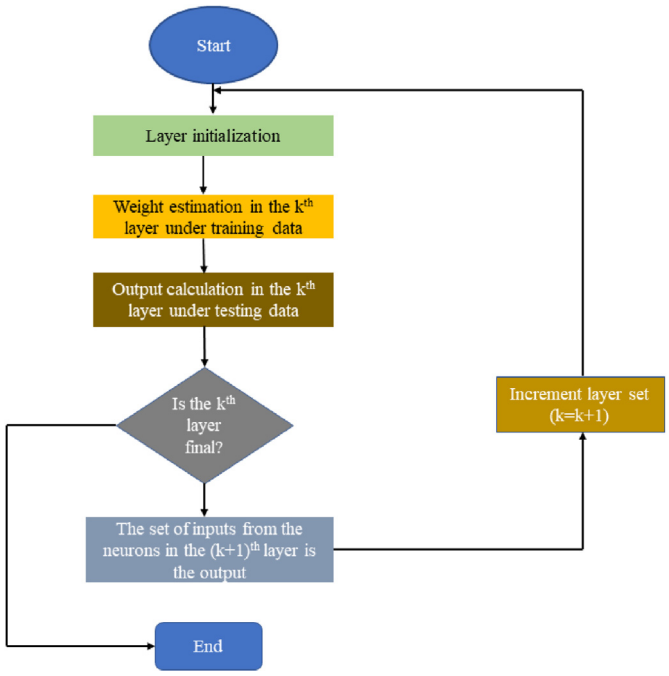

Fig. 5. Flowchart of the GMDH algorithm.

(RLA1), shale volume of rock (VSH), effective porosity (PHIE), photoelectric factor (PERFZ), apparent resistivity focusing mode 5 (RLA5), standard resolution formation density (RHOZ), caliper log (CALI) and thermal neutron porosity (TNPH). The collected dataset was processed to eliminate outliers and missing data. In this study, the dataset was normalized and missing data points and outliers were eliminated using the *z*-score method (Singh and Singh, 2020). Data from the entire set were separated into training and testing sets. A total of 1953 samples, or 70% of the whole data, were utilized as training data, while 1563 samples, or 30% of the full data, were used as testing data. The well log inputs of SGR, RLA1, RHOZ, PHIE, VSH, RLA5, CALI and TNPH were selected for model network development as visualized in Fig. 3. Table 2 summarizes the statistical evaluation of the input data used in this study.

3.2. Random forest (RF)

RF is the method of ensemble learning that is mostly used for regression, classification, and other tasks. During training, it is generally focused on developing multiple decision trees and giving out the classes or predicting each tree (Zhu et al., 2021). RF combines two methods of bagging and features randomness which help us to obtain highly accurate results, avoid overfitting problems, and be able to handle larger input datasets and thus make it suitable for the prediction purpose (Mahmoodzadeh et al., 2023; Wu et al., 2023). From the set of training data, the Bagging technique is often used to train each individual tree (Bhattacharya and Mishra, 2018). To obtain a split at each node, this approach just looks at a random subset of variables. Each tree in an RF can only be selected from a random subset of features (Feature randomness). The increased diversification and lower correlation are the results of significant trees variation in the model. As a result, in an RF, we finish up with trees that are not only trained on different sets of data but also make decisions based on the use of different features

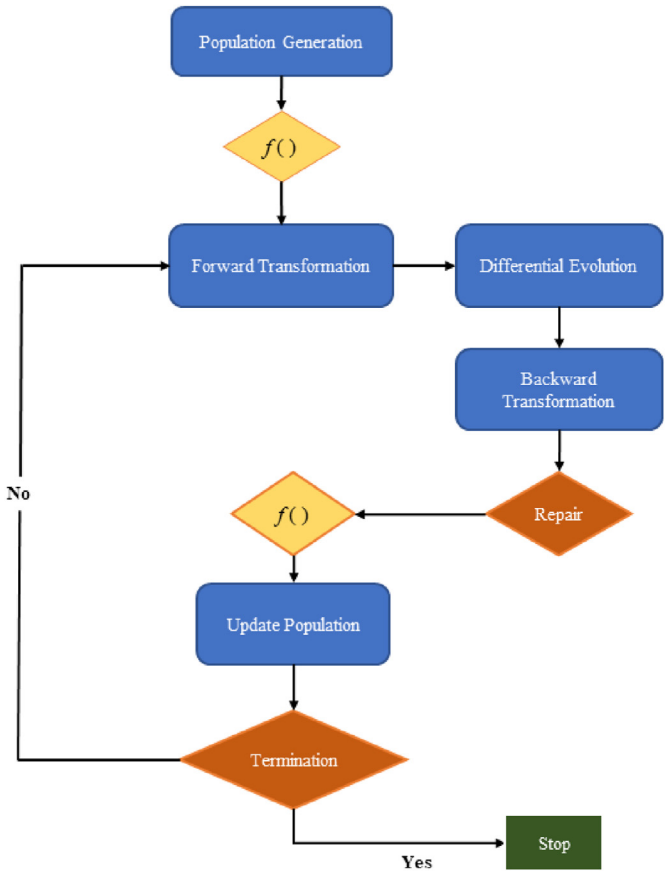

Fig. 6. Flowchart of the GMDH-DE.

(Meng et al., 2020). The schematic diagram of RF method is illustrated in Fig. 4.

3.3. Group method data handling (GMDH)

GMDH is the association of a multi-layer algorithm that generates a network of layers and nodes by utilizing several inputs from the analyzed data stream. It includes probabilistic, analogs complexing parametric, rebinarization, and clusterization techniques. Modeling of complex processes, function approximation, nonlinear regression, and pattern recognition are the core applications of GMDH. The technique refers to a self-organizing inductive propagation algorithm that can solve complex problems (Roshani et al., 2020). In addition, it is possible to derive a mathematical model from data samples, which can then be used for pattern recognition and identification.

Most GMDH algorithms employ polynomial reference functions. The Volterra's series function, which is the discrete analogs of the Kolmogorov-Gabor polynomial, can be used to describe a generic relation of output-input (Nelles, 2020).

$$u = a + \sum_{i=1}^{n} b_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} d_{ij_k} x_i x_j x_k + \ldots \quad (1)$$

where $\{x_1, x_2, x_3 \ldots\}$ represents the inputs, $\{a, b, c, d \ldots\}$ are the coefficients of the polynomials, and u is the output node.

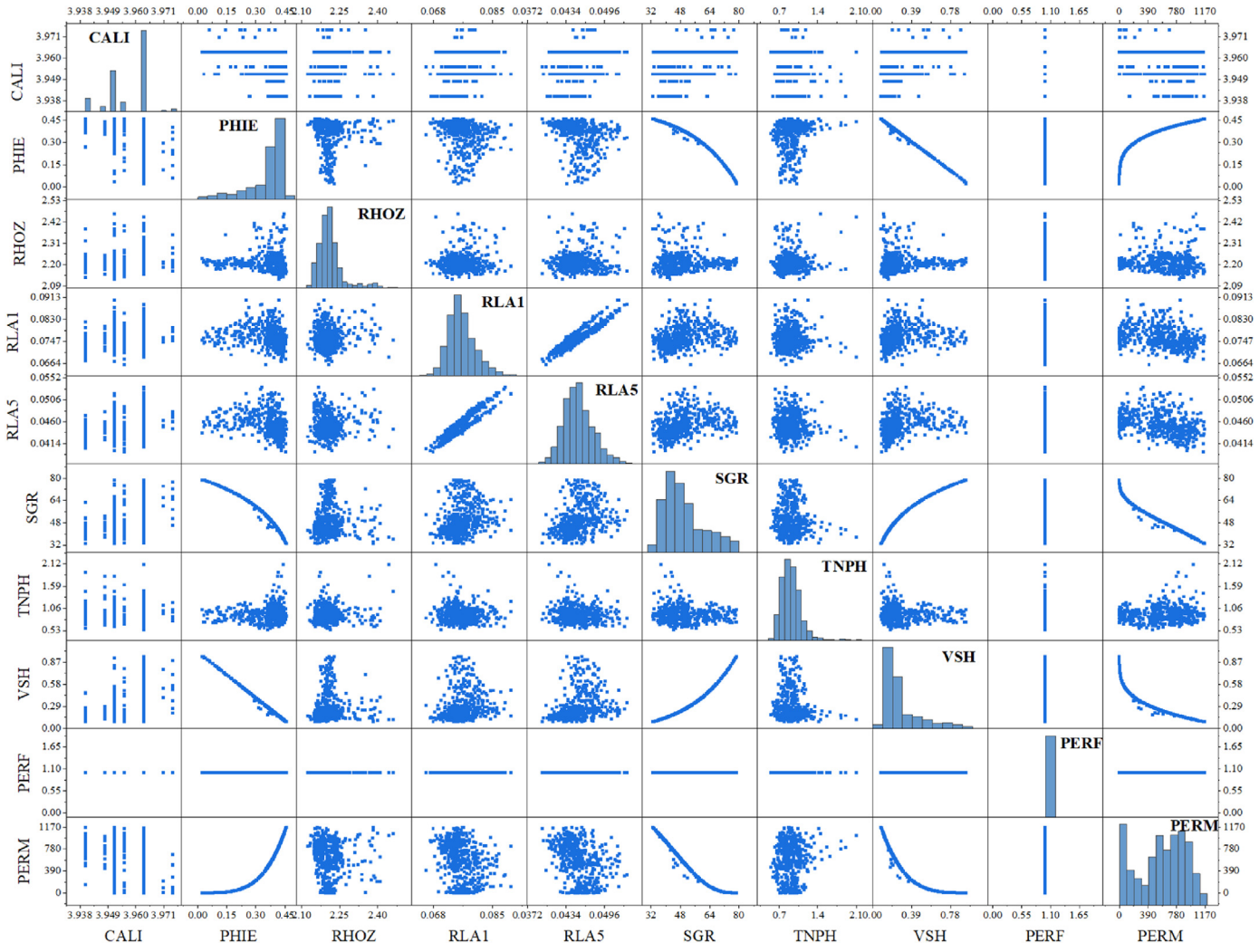

Fig. 7. Correlation coefficient matrix between input and output parameters.

The above discrete polynomial equation (Eq. (1)) can be expressed by the following system of partial quadratic polynomials, which contains only two neurons.

$$\hat{u} = a_0 + a_1 x_i + a_2 x_j + a_3 {x_i}^2 + a_4 {x_j}^2 + a_5 x_i x_j \quad (2)$$

The flowchart for GMDH model development is presented in Fig. 5.

3.4. GMDH optimized by DE

The GMDH-DE method is a novel approach for solving optimization problems, particularly those involving nonlinear systems. This method combines the strengths of the GMDH algorithm and DE algorithm to produce efficient and reliable solutions (Onwubolu, 2008).

The GMDH algorithm is a self-organizing method that generates a hierarchical structure of models, starting with simple linear models and gradually building up to more complex nonlinear models (Aljarrah et al., 2022). On the other hand, DE is an innovative parallel direct search technique introduced by Storn and Price (Storn and Price, 1997; Fathi et al., 2020). It uses a population for each generation made up of NP parameter vectors. The DE was reworked to address permutative issues even though it was initially intended for continuous domain space formulation (Storn and Price, 1997; Pourghasemi et al., 2020). The DE configuration is usually expressed in *DE/x/y/z* form, given that *x* is the perturbation solution, *y* is the difference vector's number used to modify *x*, and *z* represents the recombination operator used, such as exp for exponential and bin for binomial. The basic equation for the method is as follows:

$$X^* = \text{argmin} F(x) \quad (3)$$

Where X^* is the optimal solution, $F(x)$ stands for the objective function, while the population of solutions is represented by *x*.

Steps used for GMDH-DE are described as follows:

(1) Step 1: Find the system's input variables as x_i ($i = 1, 2, 3, \ldots, n$) and variable *y* as output.
(2) Step 2: Set the testing and training data; the input-output data set (x_i, y_i) = (x_{1i}, x_{2i},.., x_n, y_n) ($i = 1, 2, 3, \ldots, n$) (n is the whole set of data) is divided into testing and training dataset. The GMDH-DE model is built by training data. The quality of the model is then assessed using the testing data.
(3) Step 3: Identify the primary data needed to build the GMDH-DE structure: We ascertain the following preliminary data concerning the GMDH-DE design: (i) The termination

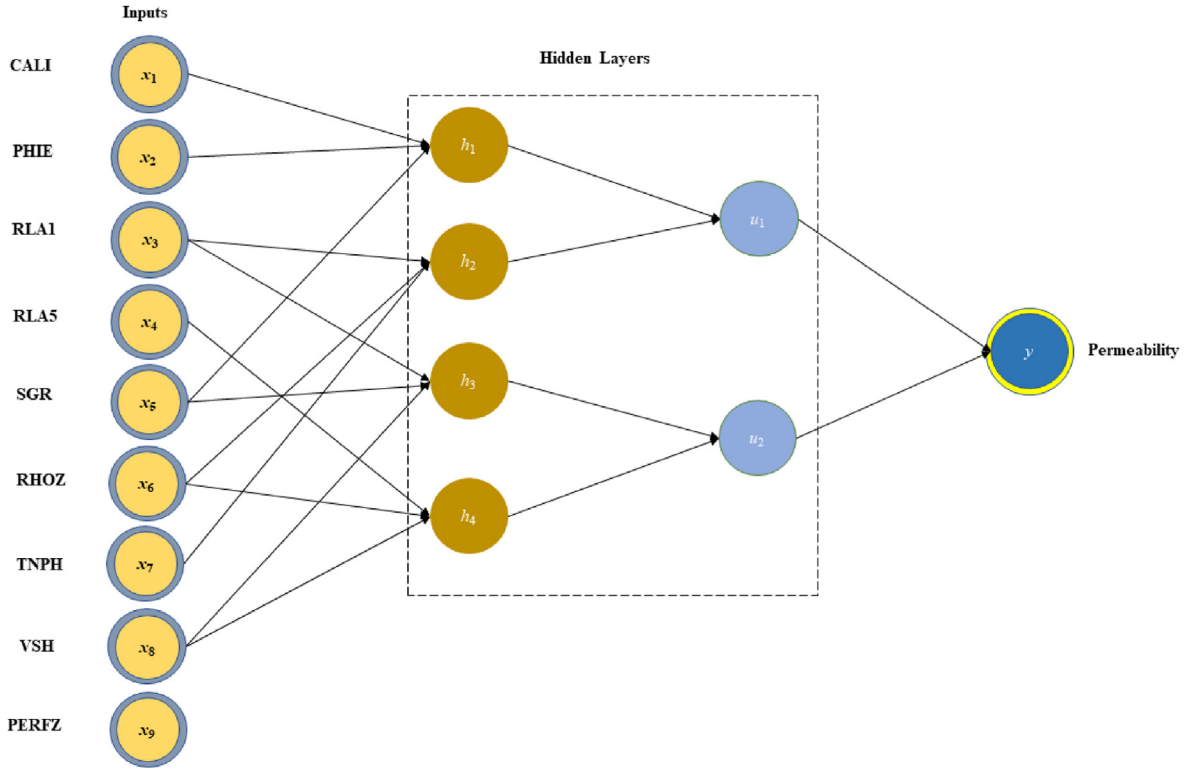

Fig. 8. GMDH-DE network structure.

Table 3
The hyperparameters setting for the GMDH-DE model.

Hyperparameter	Optimum value selected
Population size	30
Mutation rate	0.7
Crossover rate	0.7
Number of hidden layers	2
Number of neurons in each layer	10
Stopping criterion	50 iterations
Learning rate	0.1

method. (ii) most input variables can be used at a given node of the correlating layer. (iii) The aggregate objective function's weighting factor value.

(4) Step 4: Using DE design to determine the structure of polynomial neurons (PN). The input variables and polynomial order are all important factors for the polynomial neuro's identification.

(5) Step 5: Estimating the polynomial coefficients related to the designated node (PN). The standard mean squared error is used to determine the PDs' vector coefficients as stipulated in Eq. (4):

$$E_r = \frac{1}{n_{\text{training}}} \sum_{i=1}^{n_{\text{training}}} (y_i - w_{ki})^2 \ (k = 1, 2, \ldots, r) \tag{4}$$

where w_{ki} indicates the kth node's output of the ith data, r represents the parameter of the second system, n_{training} represents the number of training datasets. Nevertheless, the conventional method of least squares computes the PN node coefficients in each layer. Repeat the process up to the output layer from the input layer for all nodes and GMDH-DE layers.

(6) Step 6: Build the layer corresponding to the PN's nodes with the most robust predictive capability. The DE optimization process is used to build each layer's corresponding node in the GMDH-DE architecture:

(i) Sub-step 1: To generate the GMDH-DE architecture, and determine the initial DE information.

(ii) Sub-step 2: DE chooses the population of input variables, the number of input variables, and the order of the polynomials. The conventional least squares method is used to create the polynomial parameters.

(iii) Sub-step 3: Assess the efficiency of the PNs (nodes) in each population.

(iv) Sub-step 4: Using the DE initial information and the fitness values obtained from sub-step iii, we perform a mutation, crossover, and selection processes to create the next generation. The population's overall fitness generally increases following these DE operations. The best fitness values lead us to select some PNs. For the GMDH-DE algorithm's next iteration to run as efficiently as possible, we choose the node with the highest fitness value in this case. The outputs of the maintained nodes (PNs) are used as inputs in the network's next layer. In the GMDH-DE model, an iterative procedure creates the best nodes for a layer.

Table 4
Statistical metrics of the developed permeability models during training.

Model	RMSE	MAE	ME	R^2	Computational time (s)
GMDH-DE	3.157	2.366	0.001	0.9985	1.21
GMDH	135.63	109.458	0.0538	0.9168	2.04
RF	153.71	112.434	2.9789	0.8988	6.23

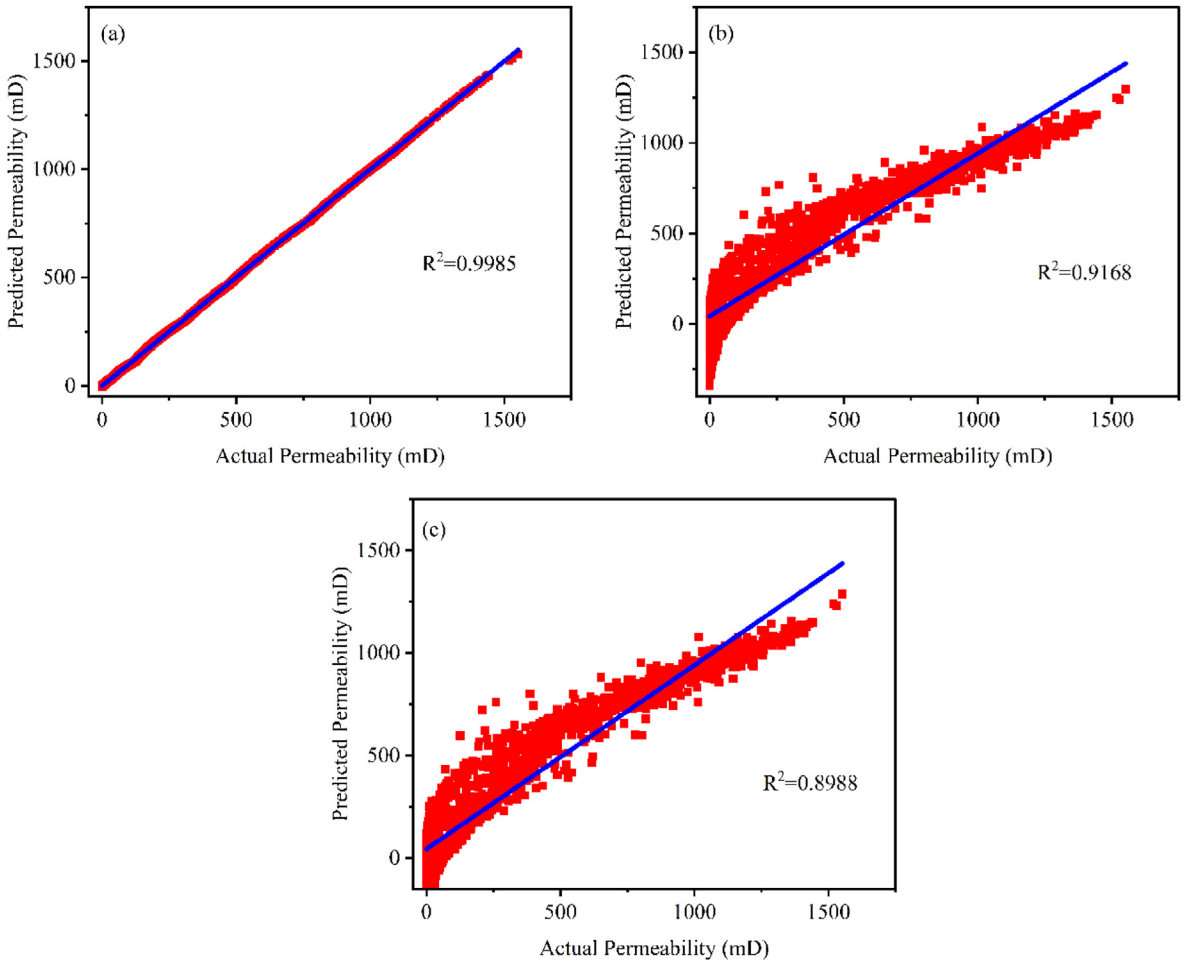

Fig. 9. The cross plot of actual permeability versus estimated permeability during training for (a) GMDH-DE, (b) GMDH, and (c) RF.

Table 5
Statistical metrics of the developed permeability models during testing.

Model	RMSE	MAE	ME	R^2	Computational time (s)
GMDH-DE	21.3898	12.503	1.3508	0.9534	1.52
GMDH	36.9449	24.161	2.1778	0.9017	2.75
RF	57.8274	45.813	10.3032	0.7289	7.36

(7) Step 7: Criteria for termination: To balance model complexity and accuracy, the termination approach used the maximum number of generations. The GMDH-DE algorithm sequentially repeats steps 4–6. The final population generation comprises solution vectors that are highly well-fitted and offer the best solutions. The generalized flowchart of GMDH-DE is shown in Fig. 6.

4. Results and discussion

4.1. Correlation analysis

A correlation coefficient matrix was generated between input and output parameters, using Pearson's product-moment correlation coefficient approach as illustrated in Fig. 7. Among nine input variables, only eight were selected to build the model due to their good correlation with the permeability. Furthermore, the analysis shows that PHIE and TNPH were positively correlated with permeability which signifies that their value increases with the rise of permeability while CALI, RHOZ, RLA1, RLA5, SGR and VSH were negatively correlated with permeability means that their value decreases as the permeability increase. PERF had a relatively negligible effect on permeability; hence it was discarded in model building.

4.2. GMDH-DE model development

The developed GMDH-DE model has eight input neurons and two hidden layers which are presented by h_1, h_2, h_3, and h_4, from the first layer and v_1 and v_2 from the second layer while y serves as a representation of the model's output shown in Fig. 8. The model was coded and executed in MATLAB R2021a. The hyperparameter values of the GMDH-DE model were determined through trial and error, the results of this method indicated that the model performed well under the hyperparameter values presented in Table 3. Keep in mind that the best hyperparameters chosen are solely dependent on the dataset and may vary for different datasets.

4.3. Evaluation metrics

GMDH-DE, GMDH and BPNN models were coded and implemented in MATLAB R2021a on Intel(R) Core (TM) i7 1065G7 CPU running Windows 10 at 2.5 GHz and an Iris (R) Plus Graphics. The root mean square error (RMSE) and coefficient of determination (R^2) were the evaluation metrics used to assess the performance of

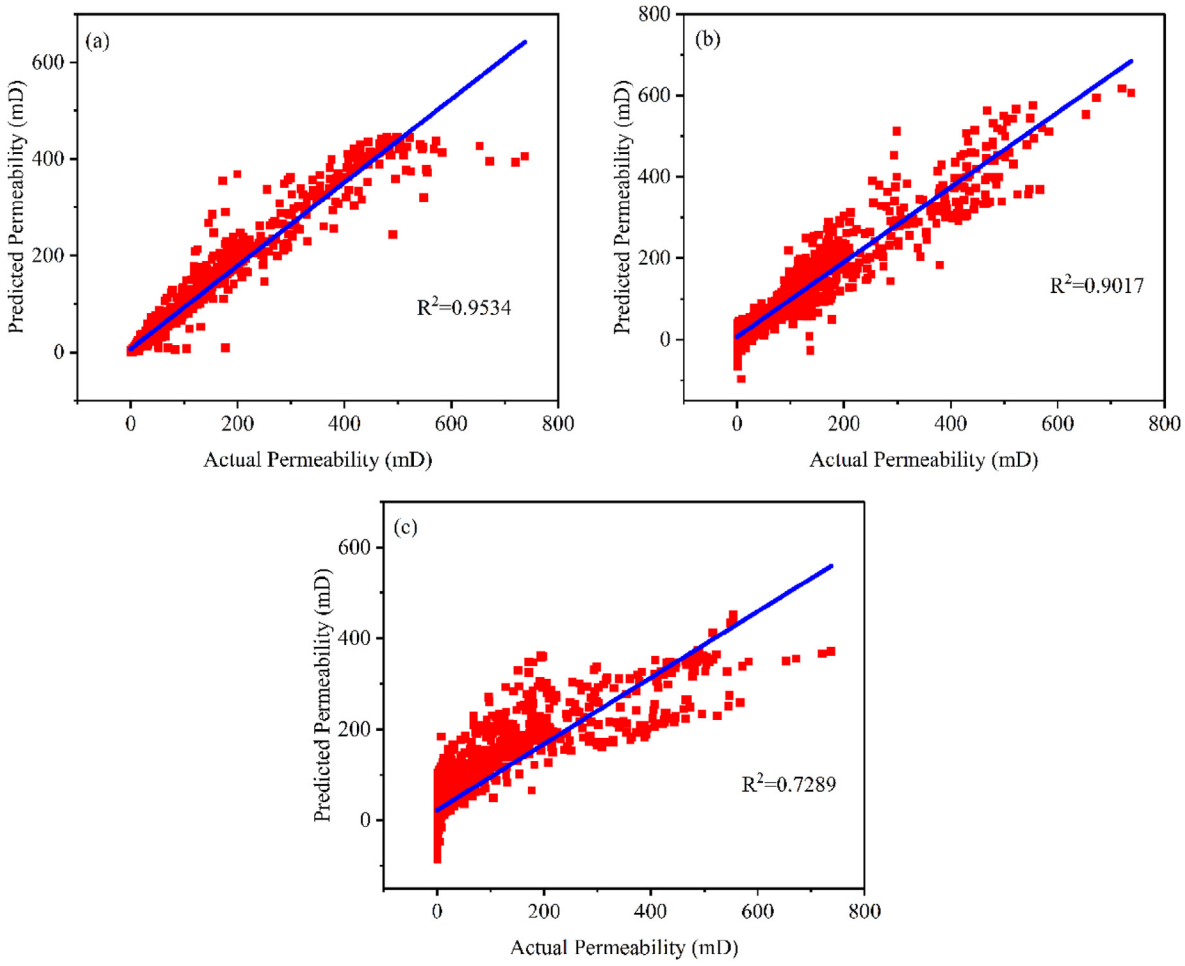

Fig. 10. The cross plot between actual permeability and estimated permeability models during testing for (a) GMDH-DE (b) GMDH and (c) RF.

prediction models. The values of R^2 can be between 0 and 1. A model is more effective when its R^2 value is higher. When a model's R^2 score is higher than 0.8 and close to 1, it is regarded as effective (Chicco et al., 2021). At the same time, RMSE expresses the stability or quality of the models by measuring the relative average square of the errors. Excellent model accuracy is defined as RMSE less than 10%, good model accuracy as RMSE between 10% and 20%, fair model accuracy as RMSE between 20% and 30%, and poor model accuracy as RMSE greater than 30% (Yao et al., 2021). ME is a statistical measure that quantifies the average difference between actual and predicted permeability values while MAE is used to measure the average magnitude of errors between actual and predicted permeability values. The R^2, RMSE, MAE and ME mathematical expressions can be presented in Eqs. (5)–(8), respectively.

$$R^2 = \left[\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}}\right]^2 \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i-y_i)^2}{n}} \tag{6}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i-y_i| \tag{7}$$

$$ME = \frac{\sum_{i=1}^{n}(x_i-y_i)}{n} \tag{8}$$

where x_i is the measured permeability value, n represents the total number of data points, $\overline{x}$ is the average value of measured permeability, y_i is the value of predicted permeability, and ⌈ is the average value of the predicted permeability.

4.4. Model training results

Table 4 presents the results of three models of GMDH-DE, GMDH, and RF in terms of their R^2, RMSE, MAE and ME during the model training as well as the computational time utilized by each model in the prediction of the permeability.

The error result shows that the GMDH-DE model has the lowest RMSE of 3.157 which indicates the highest prediction accuracy. The GMDH model has an RMSE of 135.63, which is higher than GMDH-DE, suggesting low prediction accuracy. The RF model has the highest RMSE of 153.71, indicating the lowest accuracy in prediction. Moreover, the GMDH-DE model stands out with an impressively low ME of 0.001 and a corresponding MAE of 2.366,

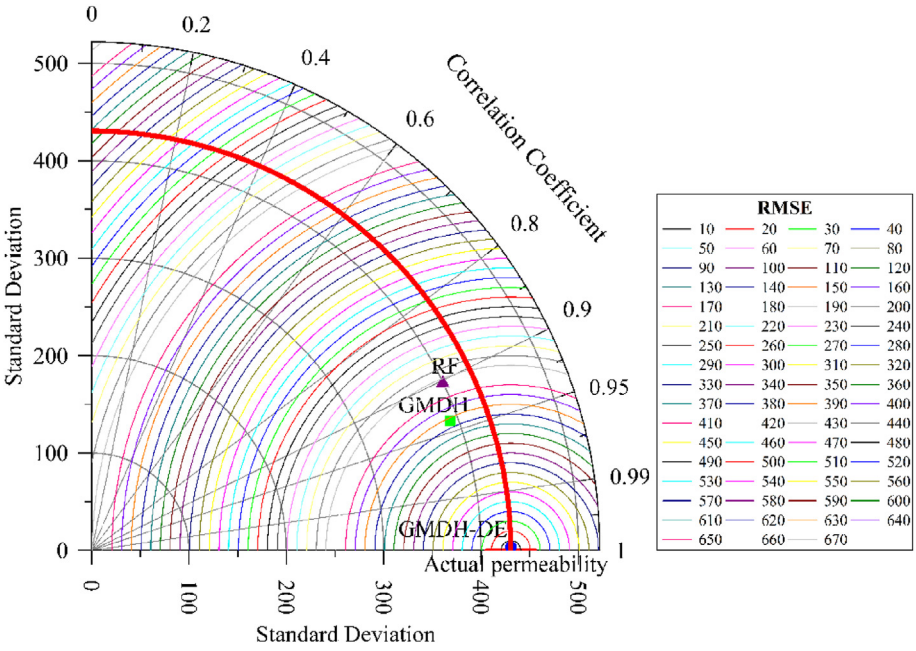

Fig. 11. Taylor diagram for estimated permeability models during training.

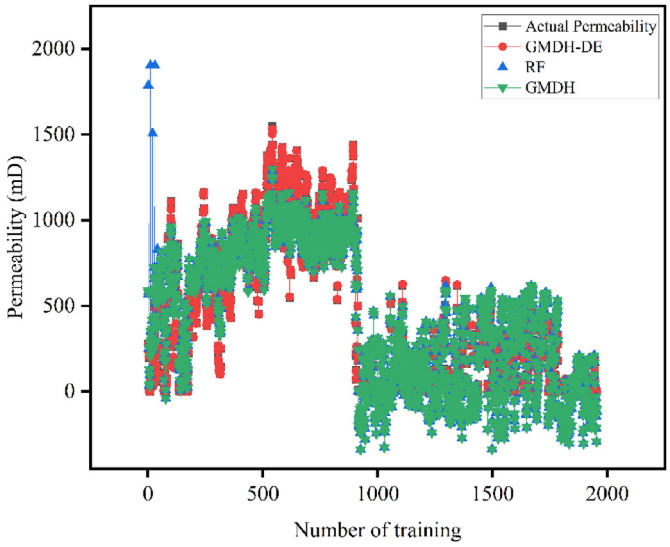

Fig. 12. Scatter plot for comparison of actual permeability and estimated permeability models during training.

indicating its superior ability to capture the underlying patterns in the training data. On the other hand, the standard GMDH model exhibits significantly higher errors with an ME of 0.0538 and an MAE of 109.458, suggesting that it struggles to generalize well to unseen data. The RF model also shows relatively high errors compared to GMDH-DE, further highlighting the latter's excellence in predicting permeability during the training phase. Fig. 9 shows that the GMDH-DE model exhibits the highest R^2 of 0.9985, followed by the GMDH model with 0.9168, and the RF model with 0.8988. These results indicate that the GMDH-DE model provides the best fit to the data, suggesting that it can make precise predictions of permeability values in hydrocarbon reservoirs. Additionally, the GMDH model also displays a relatively high R^2, implying that it is capable of making accurate predictions as well. Conversely, the RF model exhibits a lower R^2, indicating that it may not be as effective in predicting permeability values. Moreover, the GMDH-DE showed a less computational time of 1.21 s compared to the GMDH and RF which had 2.04 s and 6.23 s, respectively, hence improving the permeability model estimation by almost 41% as compared to the one suggested by Mulashani et al. (2022) and 40% from the results obtained by Matinkia et al. (2023).

Overall, these findings have significant implications for the petroleum industry, as accurate predictions of permeability values are crucial in identifying and developing hydrocarbon reservoirs.

These results imply that the GMDH-DE model can be a reliable tool for predicting the permeability of hydrocarbon reservoirs. The model can help geologists and petroleum engineers better understand the reservoir's characteristics and optimize their drilling and production strategies. On the other hand, the RF model may not be as reliable for permeability prediction, and its results may need to be used with caution.

4.5. Model testing results

Table 5 presents the results of three models of GMDH-DE, GMDH, and RF in terms of their R^2, RMSE, MAE and ME during the model testing as well as the computational time utilized by each model in the prediction of the permeability.

The error result revealed that the GMDH-DE model had the lowest RMSE of 21.3898, while the GMDH model had an RMSE of 36.9449, and the RF model had an RMSE of 57.8274. The implication here is that the GMDH-DE model had the smallest average deviation between the predicted and actual permeability values, while the RF model had the largest deviation. The GMDH-DE model continues to demonstrate its robustness with a low ME of 1.3508 and an MAE of 12.503. These metrics indicate that GMDH-DE maintains its accuracy and generalization capabilities even when applied to new, unseen data. In contrast, both the GMDH and RF models show higher errors with GMDH having an ME of 2.1778 and an MAE of 24.161, and RF exhibiting an ME of 10.3032 and an MAE of 45.813. These results reaffirm the superiority of the GMDH-DE model, as it consistently outperforms the other models in terms

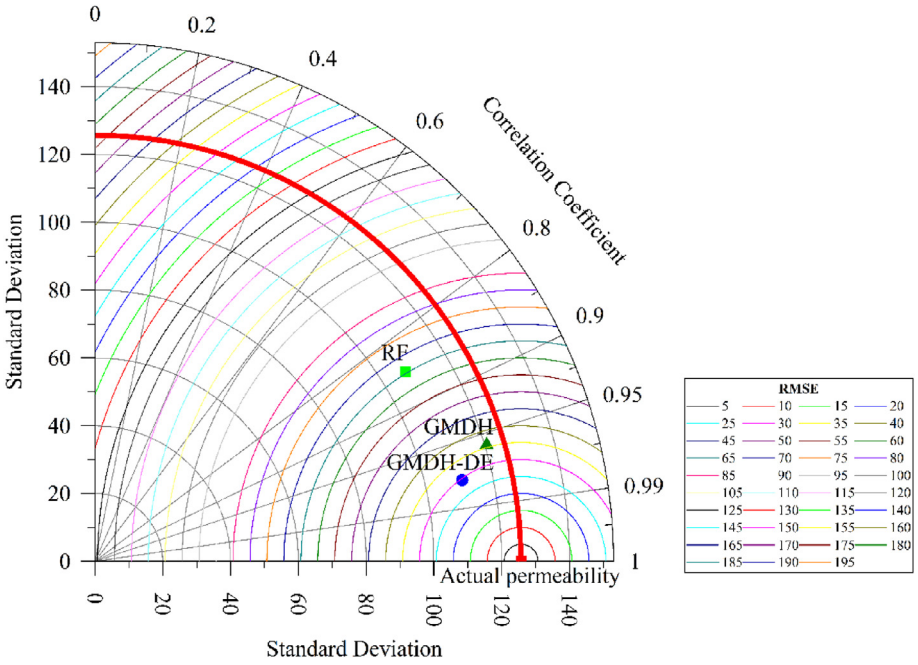

Fig. 13. Taylor diagram for estimated permeability models during testing.

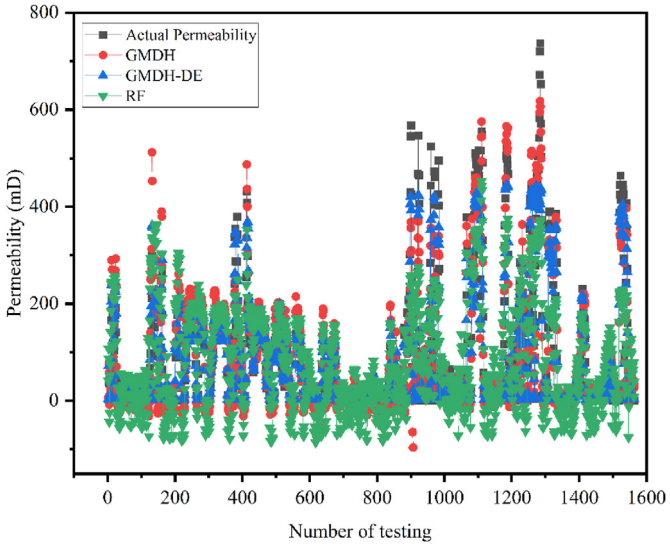

Fig. 14. Scatter plot for comparison of actual permeability and estimated permeability models during testing.

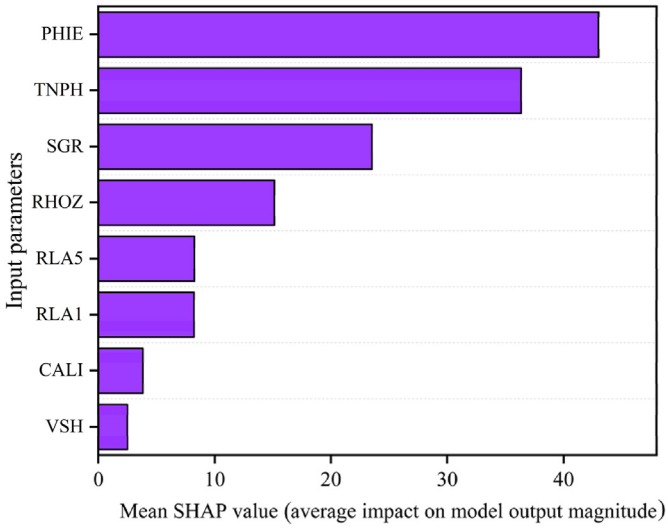

Fig. 15. SHAP parameter importance of inputs to permeability prediction.

of prediction accuracy and robustness across both training and testing scenarios. Meanwhile, the GMDH-DE shows less computational time of 1.52 s compared to the GMDH and RF which had 2.75 s and 7.36 s, respectively. Hence, this result proved the generalization ability of GMDH-DE when used in unseen data to predict the permeability of rocks.

Based on the results presented in Fig. 10, it is evident that the GMDH-DE model yields the highest R^2 value of 0.9534. This signifies that the model can predict 95.34% of the variance in permeability based on the independent variables utilized in the study. Therefore, it can be concluded that the GMDH-DE model is the most precise and dependable in predicting permeability, in comparison to others. The GMDH model, on the other hand, exhibits an R^2 value of 0.9017, indicating that it can predict 90.17% of the variance in permeability based on the independent variables used in the study. Even though the GMDH model has a lower R^2 value than the GMDH-DE model, it still demonstrates a good level of accuracy in predicting permeability. Conversely, the RF model obtains the lowest R^2 value of 0.7289, signifying that the model can only predict 72.89% of the variance in permeability based on the independent variables used in the study. This suggests that the RF model is the least accurate and reliable model for predicting permeability, compared to the other models assessed in the study. Since precise permeability prediction is essential for effective reservoir characterization and management, these findings have a substantial influence on the petroleum sector. The GMDH-DE excels in capturing nonlinear relationships between input variables and permeability. Through iterative self-organizing processes, it constructs a series of polynomial models that adaptively capture

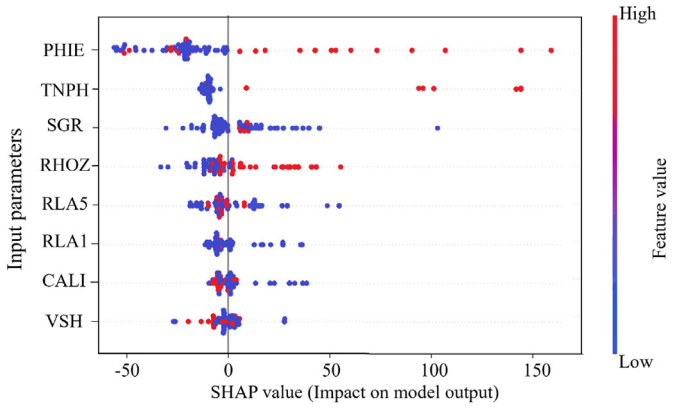

Fig. 16. SHAP parameter influence of inputs to permeability prediction.

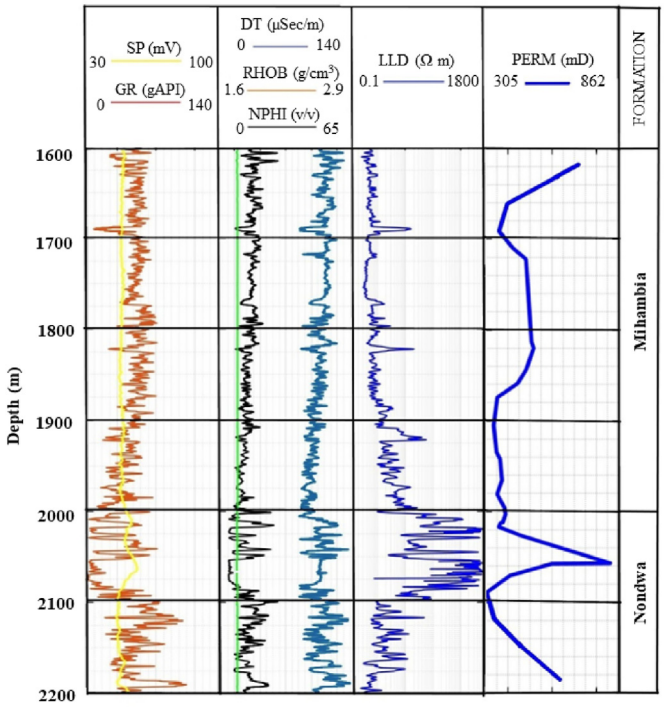

Fig. 17. Well logs and GMDH-DE predicted permeability values from Mita Gamma well.

complex patterns in the data, leading to improved predictive performance. The self-organizing nature of GMDH-DE enables it to dynamically adjust model complexity based on the data characteristics. This adaptability ensures that the model is neither underfitting nor overfitting, thus improving generalization and prediction accuracy.

Furthermore, DE plays a crucial role in optimizing the GMDH model parameters to achieve global optima, enhancing the search for optimal model configurations and leading to improved predictive performance compared to the conventional optimization methods. Moreover, DE's efficient search capabilities contribute to faster convergence towards optimal solutions, enabling GMDH-DE to handle large datasets and complex model structures efficiently, thus making it suitable for permeability prediction tasks with varying degrees of complexity. Additionally, DE's robustness to noise and its ability to handle multimodal and non-convex optimization landscapes make it well-suited for addressing challenges in permeability prediction, especially in scenarios where data exhibit nonlinearities, outliers, and complex interactions among variables.

4.6. Model comparison

The comparison of the models used was done to verify the results obtained during the model training and testing. Taylor diagrams show the correlation and variability of how well various models reproduce the observed data. Four main parts make up Taylor's diagram: points indicating the model, a circular line indicating the standard deviation (SD), actual data (represented by reference point), and a straight line representing the correlation coefficient (R). This variability and association can be better captured by a model that is closer to the reference point. It is also possible to measure the model-observation correlation by calculating the angle between the reference point, model point and the origin. Lower angles signify stronger relationships. Another way to see the model's diversity standard deviation is by measuring the distance along the circular contours from the reference point. Accurate models will have a standard deviation that is comparable to the reference data. From Taylor's diagram in Fig. 11 it revealed that the GMDH-DE outperformed both GMDH and RF in the permeability prediction during training with the support of a higher correlation coefficient. Furthermore, the scatter plot in Fig. 12 proved the GMDH-DE to be more precise and accurate in permeability prediction as it produces many values which correlate with the actual permeability.

Moreover, during the testing, the results of the model comparison showed that the GMDH-DE performed excellently than the GMDH and RF as shown in the Taylor diagram in Fig. 13. As we can see the correlation coefficient is higher in GMDH-DE compared to other models, the same result can be observed in Fig. 14 which shows that the predicted GMDH-DE permeability values are closer to the actual permeability values.

4.7. SHAP

In this study, the GMDH-DE model predicted permeability and provided valuable insights into feature relevance through SHAP values. The SHAP values calculate each feature's average marginal contribution to the model's prediction for every combination of features that may be present (Kannangara et al., 2022; Zhao et al., 2022). The SHAP parameter's importance in Fig. 15 highlighted the substantial impact of the PHIE and TNPH features on the GMDH-DE model's permeability predictions, with mean SHAP values of 43.02 and 36.35, respectively. The model effectively captured the influence of porosity on permeability, considering variations in pore system interconnectivity. Additionally, the SGR feature significantly affected permeability predictions, as indicated by its mean SHAP value of 23.54, reflecting its role in conveying information about clay content in the Wangkwar formation. RHOZ also demonstrated an influence on permeability prediction (mean SHAP value: 15.14) by indirectly determining permeability through porosity measurements. The resistivity logs of RLA5 and RLA1 moderately impacted permeability prediction (mean SHAP values: 8.26 and 8.22, respectively). CALI and VSH contributed the least to permeability prediction, with mean SHAP values of 3.83 and 2.49, respectively. Furthermore, Fig. 16 illustrates that the increases in NPHI, TNPH, and RHOZ led to increased permeability, while higher values of SGR, RLA1, RLA5, CALI, and VSH resulted in decreased permeability.

4.8. Model verification in Mandawa basin

The developed GMDH-DE model was further used to estimate

the permeability for the Mita Gamma well in the Mandawa basin, Southern Tanzania which lacks core permeability data. The results show that the estimated permeability in Mita Gamma wells ranges from 305 mD to 862 mD with an average value of 432 mD (Fig. 17). Mita Gamma well intersected the Nondwa and Mihambia formations which are characterized by high permeable lithologies such as sandstone and gravel (Mkono et al., 2023). The predicted permeability results aligned well with the characteristics of the lithologies found in the Nondwa and Mihambia formations. From depth 2020–2060 m, we observed a sharp increase in permeability which is clear evidence of the degree of sorting and compaction of the rocks present in the formation, usually poorly sorted and coarse-sized rock. Hence, this makes the GMDH-DE model more suitable for the prediction of the permeability of rock in a case when there are no available core data.

5. Conclusions

This study proposed a hybrid approach of GMDH-DE as a novel method for predicting the permeability of Wangkwar formation from well logging of SGR, RLA1, VSH, PHIE, RHOZ, RLA5, CALI and TNPH. Based on the findings, we can conclude that:

(1) The results have shown that the GMDH-DE model had a higher generalization strength and better performance in predicting permeability when compared to GMDH and RF models with an accuracy of 0.9985, RMSE of 3.157, MAE of 2.366 and ME of 0.001 during training while for testing the accuracy was 0.9534 with RMSE of 21.3898, MAE of 12.503 and ME of 1.3508. The processing time of GMDH-DE was also reduced by 41%, making it faster than GMDH and RF, hence reducing the computational time in permeability prediction.

(2) The successful GMDH-DE model was further used to predict the permeability of the Mita Gamma well in the Mandawa basin which lacks core data. The results show that the predicted permeability in Mita Gamma wells ranges from 305 mD to 862 mD with an average value of 432 mD. This can be attributed to the intersection of Mita Gamma well to the Nondwa and Mihambia formations which are characterized by high permeable lithologies such as sandstone and gravel. Hence, this makes the GMDH-DE model more suitable for the prediction of the permeability of rock in a case when there are no available core data.

(3) The evaluation of parameter importance performed by SHAP analysis was used to determine how important each well log is to the model's performance. The results revealed that well logging parameters of TNPH, PHIE and SGR had greater contributions to the performance of the GMDH-DE model in permeability prediction. This makes GMDH-DE the most reliable computing intelligence technique for attaining accurate permeability estimates. Therefore, GMDH-DE can be considered as an alternative machine learning for estimating permeability.

CRediT authorship contribution statement

Christopher N. Mkono: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Chuanbo Shen:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization. **Alvin K. Mulashani:** Writing – review & editing, Visualization, Software, Methodology. **Patrice Nyangi:** Writing – review & editing, Visualization, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Major National Science and Technology Programs in the “Thirteenth Five-Year” Plan period (Grant No. 2017ZX05032-002-004), and the Innovation Team Funding of Natural Science Foundation of Hubei Province, China (Grant No. 2021CFA031). The first author thanks the Chinese Scholarship Council (CSC) and Silk Road Institute for their support in terms of stipend.

List of symbols

PHIE	Effective porosity
TNPH	Total porosity
CALI	Caliper log
RHOZ	Standard resolution formation density
RLA1, RLA5	Resistivity logs at 1 ft and 5 ft, respectively
SGR	Sonic gamma ray
VSH	Volume of shale
PERF	Photoelectric factor
RMSE	Root mean square error
R^2	Coefficient of determination
MAE	Mean absolute error
ME	Mean error
GMDH	Group method of data handling
DE	Differential evolution
RF	Random forest
SHAP	Shapley additive explanation
x_i	Measured permeability value
n	Total number of data points
$\overline{x}$	Average value of measured permeability
y_i	Predicted permeability value
$\overline{y}$	Average value of predicted permeability
gAPI	Gamma-ray unit according to the America Petroleum Institute (API)
v	Volume

References

Akande, K.O., Owolabi, T.O., Olatunji, S.O., AbdulRaheem, A., 2017. A hybrid particle swarm optimization and support vector regression model for modelling permeability prediction of hydrocarbon reservoir. J. Pet. Sci. Eng. 150, 43–53.

Al-Amri, M.A., 2017. Integrated petrophysical and reservoir characterization workflow to en hance permeability and water saturation prediction. J. Afr. Earth Sci. 131, 105–116.

Al Khalifah, H., Glover, P.W.J., Lorinczi, P., 2020. Permeability prediction and diagenesis in tight carbonates using machine learning techniques. Mar. Petrol. Geol. 112, 104096.

Aljarrah, O., Li, J., Heryudono, A., Huang, W., Bi, J., 2022. Predicting part distortion field in additive manufacturing: a data-driven framework. J. Intell. Manuf. 34, 1975–1993.

Amaefule, J.O., Altunbay, M., Tiab, D., Kersey, D.G., Keelan, D.K., 1993. Enhanced reservoir description: using core and log data to identify hydraulic (flow) units and predict permeability in uncored intervals/wells. In: OnePetro SPE Annual Technical Conference and Exhibition.

Anifowose, F., Abdulraheem, A., Al-Shuhail, A., 2019. A parametric study of machine learning techniques in petroleum reservoir permeability prediction by integrating seismic attributes and wireline data. J. Pet. Sci. Eng. 176, 762–774.

Archie, G.E., 1942. The electrical resistivity log as an aid in determining some reservoir characteristics. Trans. Am. Inst. Min Eng. 146 (1), 54–62.

Belhouchet, H.E., Benzagouta, M.S., Dobbi, A., Alquraishi, A., Duplay, J., 2021. A new empirical model for enhancing well log permeability prediction, using nonlinear regression method: case study from Hassi-Berkine oil field reservoir – Algeria. J. King Saud Univ. Eng. Sci. 33 (2), 136–145.

Bhattacharya, S., Mishra, S., 2018. Applications of machine learning for facies and fracture prediction using Bayesian Network Theory and Random Forest: case studies from the Appalachian basin, USA. J. Pet. Sci. Eng. 170, 1005–1017.

Buryan, P., 2006. Time Series Analysis by Means of Enhanced GMDH Algorithm. Czech Technical University, Prague, Czech. PhD Thesis.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. 7, e623.

Costa, A., 2006. Permeability-porosity relationship: a reexamination of the Kozeny-Carman equation based on a fractal pore-space geometry assumption. Geophys. Res. Lett. 33 (2). https://doi.org/10.1029/2005GL025134.

Dou, L., Cheng, D., Wang, J., et al., 2004a. Geochemical significance of seepage oils and bituminous sandstones in the Albertine graben. Uganda. J. Pet. Geol. 27 (3), 299–312.

Dou, L., Wang, J., Cheng, D., et al., 2004b. Geological conditions and petroleum exploration potential of the Albertine Graben of Uganda. Acta Geol. Sin. 78 (4), 1002–1010.

Elkatatny, S., Mahmoud, M., Tariq, Z., Abdulraheem, A., 2018. New insights into the prediction of heterogeneous carbonate reservoir permeability from well logs using artificial intelligence network. Neural Comput. Appl. 30, 2673–2683.

Fathi, E., Rezaee, M.J., Tavakkoli-Moghaddam, R., Alizadeh, A., Montazer, A., 2020. Design of an integrated model for diagnosis and classification of pediatric acute leukemia using machine learning. Proc. Inst. Mech. Eng., Part H: J. Eng. Med. 234 (10), 1051–1069.

Guma, B.E., Muwanga, A., Owor, M., 2021. Hydrogeochemical evolution and contamination of groundwater in the Albertine Graben, Uganda. Environ. Earth Sci. 80, 1–17.

Jiang, X., Deng, S., Li, H., Zuo, H., 2022. Characterization of 3D pore nanostructure and stress-dependent permeability of organic-rich shales in northern Guizhou Depression, China. J. Rock Mech. Geotech. Eng. 14 (2), 407–422.

Kabwe, P., 2022. Permeability prediction with integration of log and core data of a south Tanzania gas field using artificial intelligence techniques. In: OnePetro SPWLA 63rd Annual Logging Symposium.

Kalule, R., Abderrahmane, H.A., Alameri, W., Sassi, M., 2023. Stacked ensemble machine learning for porosity and absolute permeability prediction of carbonate rock plugs. Sci. Rep. 13 (1), 9855.

Kannangara, K.K.P.M., Zhou, W., Ding, Z., Hong, Z., 2022. Investigation of feature contribution to shield tunneling-induced settlement using Shapley additive explanations method. J. Rock Mech. Geotech. Eng. 14 (4), 1052–1063.

Lee, H., Lee, J.-W., Oh, T.-M., 2021. Permeability evaluation for artificial single rock fracture according to geometric aperture variation using electrical resistivity. J. Rock Mech. Geotech. Eng. 13 (4), 787–797.

Liu, J.-J., Liu, J.-C., 2022. Permeability predictions for tight sandstone reservoir using explainable machine learning and particle swarm optimization. Geofluids 2022, 1–15.

Liu, Q., Li, Z., Wang, E., Feng, X., Kong, X., Wang, D., 2023. Synchronous inversion of coal seam gas pressure and permeability based on a dual porosity/dual permeability model and surrogate optimization algorithm. Nat. Resour. Res. 32 (5), 2115–2136.

Lukaye, J., Sserubiri, T., Tumushabe, W.M., et al., 2015. Developing a coherent stratigraphic scheme of the albertine graben, east Africa. In: Petroleum Systems in "Rift" Basins. SEPM Society for Sedimentary Geology.

Mahdaviara, M., Rostami, A., Shahbazi, K., 2020. State-of-the-art modeling permeability of the heterogeneous carbonate oil reservoirs using robust computational approaches. Fuel 268, 117389.

Mahdy, A., Zakaria, W., Helmi, A., Helaly, A.S., Mahmoud, A.M.E., 2024. Machine learning approach for core permeability prediction from well logs in Sandstone Reservoir, Mediterranean Sea, Egypt. J. Appl. Geophys. 220, 105249.

Mahmoodzadeh, A., Alanazi, A., Hussein Mohammed, A., Hashim Ibrahim, H., Alqahtani, A., Alsubai, S., Babeker Elhag, A., 2023. Comprehensive analysis of multiple machine learning techniques for rock slope failure prediction. J. Rock Mech. Geotech. Eng. https://doi.org/10.1016/j.jrmge.2023.08.023.

Makarian, E., Abad, A.B.M.N., Manaman, N.S., et al., 2023. An efficient and comprehensive poroelastic analysis of hydrocarbon systems using multiple data sets through laboratory tests and geophysical logs: a case study in an iranian hydrocarbon reservoir. Carbonates Evaporites 38 (2), 37.

Martyushev, D.A., Ponomareva, I.N., Chukhlov, A.S., et al., 2023a. Study of void space structure and its influence on carbonate reservoir properties: X-ray microtomography, electron microscopy, and well testing. Mar. Petrol. Geol. 151, 106192.

Martyushev, D.A., Ponomareva, I.N., Shen, W., 2023b. Adaptation of transient well test results. J. Min. Inst. 264, 919–925.

Masroor, M., Emami Niri, M., Sharifinasab, M.H., 2023. A multiple-input deep residual convolutional neural network for reservoir permeability prediction. Geoenergy Sci. Eng. 222, 211420.

Mathew Nkurlu, B., Shen, C., Asante-Okyere, S., Mulashani, A.K., Chungu, J., Wang, L., 2020. Prediction of permeability using group method of data handling (GMDH) neural network from well log data. Energies 13 (3), 551.

Matinkia, M., Hashami, R., Mehrad, M., Hajsaeedi, M.R., Velayati, A., 2023. Prediction of permeability from well logs using a new hybrid machine learning algorithm. Petroleum 9 (1), 108–123.

Meng, M., Zhong, R., Wei, Z., 2020. Prediction of methane adsorption in shale: classical models and machine learning based models. Fuel 278, 118358.

Mirzaei-Paiaman, A., Asadolahpour, S.R., Saboorian-Jooybari, H., Chen, Z., Ostadhassan, M., 2020. A new framework for selection of representative samples for special core analysis. Pet. Res. 5 (3), 210–226.

Mirzaei-Paiaman, A., Saboorian-Jooybari, H., Pourafshary, P., 2015. Improved method to identify hydraulic flow units for reservoir characterization. Energy Technol. 3 (7), 726–733.

Mkono, C.N., Chuanbo, S., Mulashani, A.K., Mwakipunda, G.C., 2023. Deep learning integrated approach for hydrocarbon source rock evaluation and geochemical indicators prediction in the Jurassic-Paleogene of the Mandawa basin, SE Tanzania. Energy 284, 129232.

Mohammadian, E., Kheirollahi, M., Liu, B., Ostadhassan, M., Sabet, M., 2022. A case study of petrophysical rock typing and permeability prediction using machine learning in a heterogenous carbonate reservoir in Iran. Sci. Rep. 12 (1), 4505.

Mohammadlou, M.H.H., Mørk, M.B.B., 2012. Integrated permeability analysis in tight and brecciated carbonate reservoir. SPE Reservoir Eval. Eng. 15 (6), 624–635.

Moosavi, N., Bagheri, M., Nabi-Bidhendi, M., Heidari, R., 2022. Fuzzy support vector regression for permeability estimation of petroleum reservoir using well logs. Acta Geophys. 1–12.

Mulashani, A.K., Shen, C., Nkurlu, B.M., Mkono, C.N., Kawamala, M., 2022. Enhanced group method of data handling (GMDH) for permeability prediction based on the modified Levenberg Marquardt technique from well log data. Energy 239, 121915.

Mutebi, S., Sen, S., Sserubiri, T., Rudra, A., Ganguli, S.S., Radwan, A.E., 2021. Geological characterization of the miocene–pliocene succession in the semliki basin, Uganda: implications for hydrocarbon exploration and drilling in the east African rift system. Nat. Resour. Res. 30, 4329–4354.

Nazari, H., Hajizadeh, F., 2023. Estimation of permeability from a hydrocarbon reservoir located in southwestern Iran using well-logging data and a new intelligent combined method. Carbonates Evaporites 38 (1), 20.

Nelles, O., 2020. Non-linear System Identification: from Classical Approaches to Neural Networks, Fuzzy Systems, and Gaussian Processes. Springer Nature, Switzeland.

Oh, S.-K., Pedrycz, W., 2006. Multi-layer self-organizing polynomial neural networks and their development with the use of genetic algorithms. J. Franklin Inst. 343 (2), 125–136.

Onwubolu, G.C., 2008. Design of hybrid differential evolution and group method of data handling networks for modeling and prediction. Inf. Sci. 178 (18), 3616–3634.

Park, H.-S., Park, B.-J., Kim, H.-K., Oh, S.-K., 2004. Self-organizing polynomial neural networks based on genetically optimized multi-layer perceptron architecture. Int. J. Control Autom. Syst. 2 (4), 423–434.

Pourghasemi, H.R., Razavi-Termeh, S.V., Kariminejad, N., Hong, H., Chen, W., 2020. An assessment of metaheuristic approaches for flood assessment. J. Hydrol. 582, 124536.

Rashid, F., Glover, P.W.J., Lorinczi, P., Hussein, D., Collier, R., Lawrence, J., 2015. Permeability prediction in tight carbonate rocks using capillary pressure measurements. Mar. Petrol. Geol. 68, 536–550.

Rashid, M., Luo, M., Ashraf, U., et al., 2023. Reservoir quality prediction of gas-bearing carbonate sediments in the Qadirpur Field: insights from advanced machine learning approaches of SOM and cluster analysis. Minerals-Basel 13 (1), 29.

Rezaee, R., Ekundayo, J., 2022. Permeability prediction using machine learning methods for the CO_2 injectivity of the precipice sandstone in Surat Basin, Australia. Energies 15 (6), 2053.

Roshani, M., Sattari, M.A., Ali, P.J.M., et al., 2020. Application of GMDH neural network technique to improve measuring precision of a simplified photon attenuation based two-phase flowmeter. Flow Meas. Instrum. 75, 101804.

Sheykhinasab, A., Mohseni, A.A., Barahooie Bahari, A., et al., 2023. Prediction of

permeability of highly heterogeneous hydrocarbon reservoir from conventional petrophysical logs using optimized data-driven algorithms. J. Pet. Explor. Prod. Technol. 13 (2), 661–689.
Shokir, E.M.E.-M., Alsughayer, A.A., Al-Ateeq, A., 2006. Permeability estimation from well log responses. J. Can. Pet. Technol. 45 (11). PETSOC-06-11-05.
Simon, B., Guillocheau, F., Robin, C., et al., 2017. Deformation and sedimentary evolution of the Lake Albert rift (Uganda, east African rift system). Mar. Petrol. Geol. 86, 17–37.
Singh, D., Singh, B., 2020. Investigating the impact of data normalization on classification performance. Appl. Soft Comput. 97, 105524.
Storn, R., Price, K., 1997. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. J. Global Optim. 11 (4), 341–359.
Subasi, A., El-Amin, M.F., Darwich, T., Dossary, M., 2020. Permeability prediction of petroleum reservoirs using stochastic gradient boosting regression. J. Ambient Intell. Hum. Comput. 13, 3555–3564.
Sun, K., Liu, H., Leung, J.Y., Wang, J., Feng, Y., Liu, R., Kang, Z., Zhang, Y., 2023. Impact of effective stress on permeability for carbonate fractured-vuggy rocks. J. Rock Mech. Geotech. Eng. 16 (3), 942–960.
Sun, L., Hao, X., Dou, H., Adenutsi, C.D., Liu, W., 2022. A novel model for predicting tight sandstone reservoir permeability. Int. J. Oil Gas Coal Technol. 29 (1), 75–90.
Tian, J.W., Qi, C., Peng, K., Sun, Y., Yaseen, Z.M., 2022. Improved permeability prediction of porous media by feature selection and machine learning methods comparison. J. Comput. Civ. Eng. 36 (2), 04021040.
Tonny, S., Scholz, A.C., Post, P.J., et al., 2015. Using geochemical data from well samples to reconstruct paleoenvironments of the Central Lake Albert Basin, Uganda. In: Petroleum Systems in "Rift" Basins. SEPM Society for Sedimentary Geology.
Uguru, C.I., Onyeagoro, U.O., Lin, J., Okkerman, J., Sikiru, I.O., 2005. Permeability prediction using genetic unit averages of flow zone indicators (FZIs) and neural networks. SPE Nigeria Annual International Conference and Exhibition. Abuja. SPE-98828-MS.
Wang, X., 2019. Improved permeability prediction based on the feature engineering of p etrophysics and fuzzy logic analysis in low porosity–permeability reservoir. J. Pet. Explor. Prod. Technol. 9 (2), 869–887.
Wong, P.M., Jang, M., Cho, S., Gedeon, T.D., 2000. Multiple permeability predictions using an observational learning algorithm. Comput. Geosci. 26 (8), 907–913.
Wu, K., Meng, Q., Li, R., Luo, L., Ke, Q., Wang, C., Ma, C., 2023. A machine learning-based strategy for predicting the mechanical strength of coral reef limestone using X-ray computed tomography. J. Rock Mech. Geotech. Eng. 16 (7), 2790–2800.
Xu, P., Zhou, H., Liu, X., Chen, L., Xiong, C., Lyu, F., Zhou, J., Liu, J., 2023. Permeability prediction using logging data in a heterogeneous carbonate reservoir: a new self-adaptive predictor. Geoenergy Sci Eng 224, 211635.
Yao, B., He, H., Xu, H., Zhu, T., Liu, T., Ke, J., You, C., Zhu, D., Wu, L., 2021. Determining nitrogen status and quantifying nitrogen fertilizer requirement using a critical nitrogen dilution curve for hybrid indica rice under mechanical pot-seedling transplanting pattern. J. Integr. Agric. 20 (6), 1474–1486.
Zakharov, L.A., Martyushev, D.A., Ponomareva, I.N., 2022. Predicting dynamic formation pressure using artificial intelligence methods. J. Min. Inst. 253, 23–32.
Zanganeh Kamali, M., Davoodi, S., Ghorbani, H., Wood, D.A., Mohamadian, N., Lajmorak, S., Rukavishnikov, V.S., Taherizade, F., Band, S.S., 2022. Permeability prediction of heterogeneous carbonate gas condensate reservoirs applying group method of data handling. Mar. Petrol. Geol. 139, 105597.
Zhang, G., Wang, Z., Li, H., Sun, Y., Zhang, Q., Chen, W., 2018. Permeability prediction of isolated channel sands using machine learning. J. Appl. Geophys. 159, 605–615.
Zhang, G., Wang, Z., Mohaghegh, S., Lin, C., Sun, Y., Pei, S., 2021a. Pattern visualization and understanding of machine learning models for permeability prediction in tight sandstone reservoirs. J. Pet. Sci. Eng. 200, 108142.
Zhang, H., Wu, W., 2023. DBN with IQPSO algorithm for permeability prediction: a case study of the Lizhai Geothermal field, Zhangye Basin (Northern China). Nat. Resour. Res. 32 (5), 1941–1957.
Zhang, Z., Liu, L., Li, C., Cai, J., Ning, F., Meng, Q., Liu, C., 2021b. Fractal analyses on saturation exponent in Archie's law for electrical properties of hydrate-bearing porous media. J. Pet. Sci. Eng. 196, 107642.
Zhao, X., Chen, X., Huang, Q., Lan, Z., Wang, X., Yao, G., 2022. Logging-data-driven permeability prediction in low-permeable sandstones based on machine learning with pattern visualization: a case study in Wenchang A Sag, Pearl River Mouth Basin. J. Pet. Sci. Eng. 214, 110517.
Zhong, Z., Carr, T.R., 2019. Application of a new hybrid particle swarm optimization-mixed kernels function-based support vector machine model for reservoir porosity prediction: a case study in Jacksonburg-Stringtown oil field, West Virginia, USA. Interpr 7 (1), T97–T112.
Zhong, Z., Carr, T.R., Wu, X., Wang, G., 2019. Application of a convolutional neural network in permeability prediction: a case study in the Jacksonburg-Stringtown oil field, West Virginia, USAPermeability prediction via a CNN. Geophysics 84 (6), B363–B373.
Zhu, L., Zhou, X., Zhang, C., 2021. Rapid identification of high-quality marine shale gas reservoirs based on the oversampling method and random forest algorithm. Artif. Intell. Geos 2, 76–81.

Christopher Mkono is a researcher at the China University of Geosciences, specializing in Oil and Natural Gas Engineering under the supervision of Professor Chuanbo Shen. He earned his master's degree in the same field from the China University of Geosciences in 2022. Mkono's research focuses on machine learning applications in source rock evaluation, reservoir characterization, and hydrocarbon resource potential. His work aims to advance the understanding and prediction of oil and natural gas reserves, leveraging cutting-edge technologies to improve the efficiency and accuracy of geological assessments. He is also an active member of the Society of Petroleum Engineers (SPE).