



Original Paper

# A Novel Hybrid Machine Learning Approach and Basin Modeling for Thermal Maturity Estimation of Source Rocks in Mandawa Basin, East Africa

Christopher N. Mkono,<sup>1</sup> Chuanbo Shen,<sup>1,5</sup> Alvin K. Mulashani,<sup>2</sup> Mbega Ramadhani Ngata,<sup>3</sup> and Wakeel Hussain<sup>4</sup>

Received 11 April 2024; accepted 6 June 2024

Basin modeling and thermal maturity estimation are crucial for understanding sedimentary basin evolution and hydrocarbon potential. Assessing thermal maturity in the oil and gas industry is vital during exploration. With artificial intelligence advancements, more accurate evaluation of hydrocarbon source rocks and efficient thermal maturity estimation are possible. This study employed 1D basin modeling using PetroMod and a novel hybrid group method of data handling (GMDH) neural network optimized by a differential evolution (DE) algorithm to estimate thermal maturity (T<sub>max</sub>) and assess kerogen type in Triassic–Jurassic source rocks of the Mandawa Basin, Tanzania. The GMDH–DE addresses the limitations of conventional methods by offering a data-driven approach that reduces computational time, overcomes overfitting, and improves accuracy. The 1D thermal maturity basin modeling suggests that the Mbuo source rocks reached the gas–oil window in late Triassic times and began expulsion in the early Jurassic while located in an immature-to-mature zone. The GMDH–DE model effectively estimated T<sub>max</sub> with high coefficient of determination ( $R^2 = 0.9946$ ), low root mean square error (RMSE = 0.004), and mean absolute error (MAE = 0.006) during training. When tested on unseen data, the GMDH–DE model yielded an  $R^2$  of 0.9703, RMSE of 0.017, and MAE of 0.025. Moreover, GMDH–DE reduced the computational time by 94% during training and 87% during testing. The results demonstrated the model's exceptional reliability compared to the benchmark methods such as artificial neural network–particle swarm optimization and principal component analysis coupled with artificial neural network. The GMDH–DE T<sub>max</sub> model offers a unique and independent approach for rapid real-time determination of T<sub>max</sub> values in organic matter, promoting efficient resource assessment in oil and gas exploration.

<sup>1</sup>Key Laboratory of Tectonics and Petroleum Resources, Ministry of Education, China University of Geosciences, Wuhan 430074, China.

<sup>2</sup>Department of Geoscience and Mining Technology, College of Engineering and Technology, Mbeya University of Science and Technology, Box 131 Mbeya, Tanzania.

<sup>3</sup>Key Laboratory of Theory and Technology of Petroleum Exploration and Development in Hubei Province, China University of Geosciences, Wuhan 430074, China.

<sup>4</sup>Hubei Multiscale Image Key Laboratory, School of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China.

<sup>5</sup>To whom correspondence should be addressed; e-mail: cbshen@cug.edu.cn; alvinmulashani@yahoo.com

**KEY WORDS:** Thermal maturity, 1D basin modeling, Group method of data handling, Source rock, Well logs, Differential evolution.

## INTRODUCTION

East Africa has emerged as a leading oil and gas exploration hub, drawing global attention due to its extensive hydrocarbon reserves and promising geological formations. The region's strategic significance is underscored by substantial discoveries in sedimentary basins such as the Mandawa Basin (Tanzania), Lake Albert Rift Basin (Uganda, Kenya), and the Rovuma Basin (Tanzania, Mozambique) as detailed by Zongying et al. (2013) and Purcell (2014). This surge in exploration activity, fueled by advancements in technology and the attraction of vast reserves, has positioned East Africa as a frontier for innovative exploration techniques and international investment. Major oil companies are engaged actively in assessing the region's hydrocarbon potential, with estimates indicating reserves of up to 2.8 billion barrels<sup>1</sup> of oil and 2.2 billion barrels of natural gas liquids from Tanzania's basin alone (Brownfield, 2016). This influx of exploration activity underscores East Africa's emergence as a vital player in meeting global energy demands, emphasizing the need for innovative exploration techniques and efficient resource assessment strategies.

Recently, many researchers have focused on utilizing new technology to minimize the time and cost of exploration due to the expected rise in the future energy demand. According to the International Energy Agency (IEA), energy demand is projected to increase by 25% by 2040, with fossil fuels remaining the dominant energy source and accounting for 75% of the total energy mix (IEA, 2021). This has led to the emergence of unconventional resources as a substitute. Basin modeling, a numerical simulation technique, plays a vital role in deciphering the complex subsurface geological processes that govern the formation and distribution of these valuable resources (Abdel-Fattah et al., 2017; Ehsan et al., 2023; Feng et al., 2023b). Thermal maturity (T<sub>max</sub>) prediction is a vital component of basin modeling as it assesses the extent to which organic matter has been converted inside a source

rock, hence influencing its capacity to produce oil and gas (Farouk et al., 2023).

However, accurate T<sub>max</sub> estimation is essential in assessing and evaluating any unconventional hydrocarbon resources, and it is usually measured from core samples using geochemical analysis (Huijun et al., 2020; Kibria et al., 2020; Stokes et al., 2023). Maturity indices, such as maturity temperature (T<sub>max</sub>), are used widely for assessing the T<sub>max</sub> of a source rock (Zhang and Li, 2018; Yang and Horsfield, 2020). T<sub>max</sub> is a crucial T<sub>max</sub> index that can be used to estimate the maximum temperature reached by a source rock during the burial history of a basin (Albriki et al., 2022; Wu et al., 2023). T<sub>max</sub> is obtained through the pyrolysis process and corresponds to the S<sub>2</sub> (remain potential of hydrocarbon generation) peak that results from the thermal breakdown of kerogen during temperature-programmed pyrolysis at temperatures between 300 and 600 °C (Synnott et al., 2018; Yang and Horsfield, 2020; Thana'Ani et al., 2022; Osukuku et al., 2023).

When estimating the maturity of drilled wells, geochemical methods like pyrolysis have long been thought to be the most reliable and accurate. However, numerous drawbacks are associated with this method, like time-consuming, operating costs, and inability to cover an extensive range of depth (Wood, 2018; İnan, 2022). Moreover, several studies have claimed that pyrolysis methods may expose samples to air for an extended period; measurements can often be inaccurate because the effect increases the likelihood of free organic matter oxidizing and escaping (Dembicki, 2022). As an alternative to geochemical methods, wireline logs offer an accessible and affordable data source and have become increasingly popular recently (Zhao et al., 2019; Malki et al., 2023).

Numerous researchers have shown that the evaluation of T<sub>max</sub> has been a primary concern in oil and gas exploration, and various conventional techniques have been implemented by different researchers to assess it (Gu et al., 2022; Hackley et al., 2022; Singh et al., 2022; Feng et al., 2023a; Thankan et al., 2023; Wu et al., 2023). These techniques include the bitumen reflectance (Hackley and Lünsdorf, 2018; Jubb et al., 2020; Adeyilola et al., 2022),

<sup>1</sup> 1 billion barrel = 6.118 × 109 liters

## A Novel Hybrid Machine Learning Approach

thermal alteration index (Craddock et al., 2018; Deaf et al., 2022), Rock–Eval pyrolysis (Cheshire et al., 2017; Chen et al., 2019; Pang et al., 2020; Arysanto et al., 2022; Farouk et al., 2023; Sohail et al., 2024), and fluid inclusion analysis (Petersen et al., 2022). While conventional approaches to basin modeling and Tmax prediction have proven valuable, they have significant limitations as they provide discrete data that can be rigorous and lead to poor evaluation of source rock (İnan et al., 2017; Katz and Lin, 2021; Lohr and Hackley, 2021; Sadeghtabaghi et al., 2021; Safaei-Farouji and Kadkhodaie, 2022a). However, irrespective of the conventional solver adopted for Tmax computation, the procedure typically involves significant computational overheads and consumes time.

Recently, due to the advancement of technology, various machine learning techniques have become the focal point of researchers and have been adopted to predict the Tmax of source rock (Abdizadeh et al., 2017; AlSinan et al., 2020; Ehsan and Gu, 2020; Shalaby et al., 2020; Tariq et al., 2020; Amosu et al., 2021; Barham et al., 2021; Aliakbar-doust et al., 2024; Li et al., 2024). Hybrid methods have been reported to be more accurate in predicting different source rock parameters (Ahangari et al., 2022; Safaei-Farouji and Kadkhodaie, 2022b; Saporetti et al., 2022; Mkono et al., 2023). However, a group method of data handling (GMDH) method was suggested by Mulashani et al. (2021) as an alternative method for predicting total organic carbon (TOC) from well logs. The methods include input factors such as neutron porosity, spontaneous potential, gamma-ray, resistivity log, sonic travel time, and bulk density. Compared to ANN and log *R*, the results demonstrated that the methods accurately estimate TOC from log data. However, the study of organic matter Tmax estimation was not presented in detail.

In addition, some studies have been reported to predict Tmax using hybrid methods (Tariq et al., 2020; Barham et al., 2021). In their research, Tariq et al. (2020) used a hybrid technique of artificial neural network–particle swarm optimization (PSO–ANN) to predict Tmax from well log. Another researcher, Barham et al. (2021), applied ANN coupled with principal component analysis (ANN–PCA) to predict Tmax from geophysical well logs. These methods have shown some limitations that lead to inaccurate estimation of Tmax (Table 1). For this, a hybrid of group method of data handling (GMDH) neural network and differential evolution

(DE) algorithm (i.e., GMDH–DE) is proposed in this study to overcome the drawbacks of previously utilized hybrid methods used in predicting Tmax.

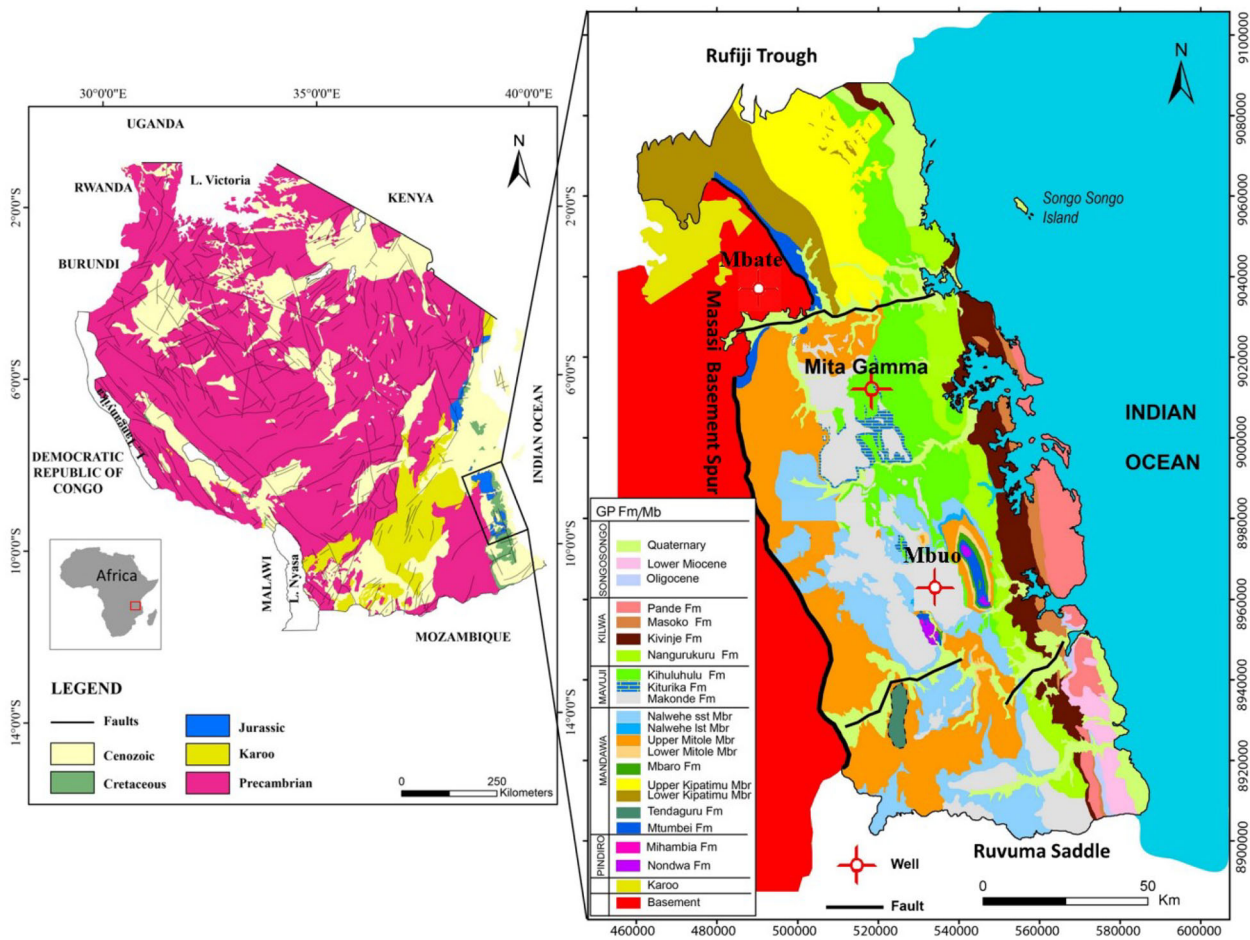
This study presents, for the first time, an integral technique of basin modeling with a novel hybrid GMDH–DE method to enhance the computational process, assess the kerogen type, and simplify the estimation of Tmax in source rocks. The GMDH–DE serves as an improved neural network model for estimating Tmax as a maturity index using geophysical well logs. During the training phase, the GMDH–DE exhibits a remarkable self-organizing characteristic that automatically adjusts model parameters and generates an ideal model structure. Unlike previous hybrid machine learning models for predicting Tmax, the GMDH–DE eliminates the need to manually adjust learning parameters to achieve optimal results. Hence, the performance of the proposed GMDH–DE model in forecasting Tmax is an adequate improvement compared to that of previously employed hybrid machine learning algorithms, namely PCA–ANN and PSO–ANN. Moreover, this study performed a sensitivity analysis to determine how much each input parameter affected the suggested GMDH–DE model in the estimation of Tmax. The results of this study ranked the GMDH–DE model as a reasonably new computational intelligent learning model for the reliable estimation of Tmax. This research contributes to advancing exploration techniques and efficient resource assessment strategies, making it a valuable asset for the oil and gas industry’s ongoing efforts to meet global energy demands.

## GEOLOGICAL SETTING

The Mandawa Basin is a sedimentary basin in southeastern Tanzania. It is bounded to the north by the Rufiji Trough, to the south by the Ruvuma Basin, to the west by the metamorphic basement, and to the east by offshore basins (Fig. 1). The basin covers an area of approximately 16,000 km<sup>2</sup> (Fossum, 2020; Abay et al., 2021). The basin was formed during the Permo–Triassic rifting event, which resulted in the separation of East Gondwana from West Gondwana (Hudson, 2011; Hudson and Nicholas, 2014; Godfray and Seetharamaiah, 2019). The rifting event began in the Late Permian and continued into the Early Triassic. The sediments deposited during this time are known as the Pindiro Group, which comprises various lithologies, includ-

**Table 1.** Summary of previous hybrid methods' strength and limitation in predicting Tmax

Model	Strength	Limitation
PSO-ANN	Has the ability to identify better solutions than traditional ANN's ability to use fewer neurons and connections than a traditional ANN	Computationally expensive and time-consuming (Li et al., 2014; Devi et al., 2015; Jahed Armaghani et al., 2017; Pooria et al., 2022; Kedia et al., 2023). It is difficult to determine the optimal number of parameters. It can be easily trapped in local minima. High computational complexity and slow convergence
PCA-ANN	It uses fewer training samples and is more efficient than a single ANN model. It allows for improved generalization performance compared to using a single ANN model	It can be computationally intensive and requires much processing power to run. It can be affected by the curse of dimensionality. It may be affected by outliers. It is not suitable for datasets with nonlinear relationships between variables (Gallo & Capozzi, 2019; Mostaar et al., 2019; Bayatvarkeshi et al., 2020)



**Figure 1.** Location of the Mandawa Basin and position of the Mbate, Mbuo, and Mita Gamma Wells (map modified from Hudson (2011) and Barth et al. (2016)).

## A Novel Hybrid Machine Learning Approach

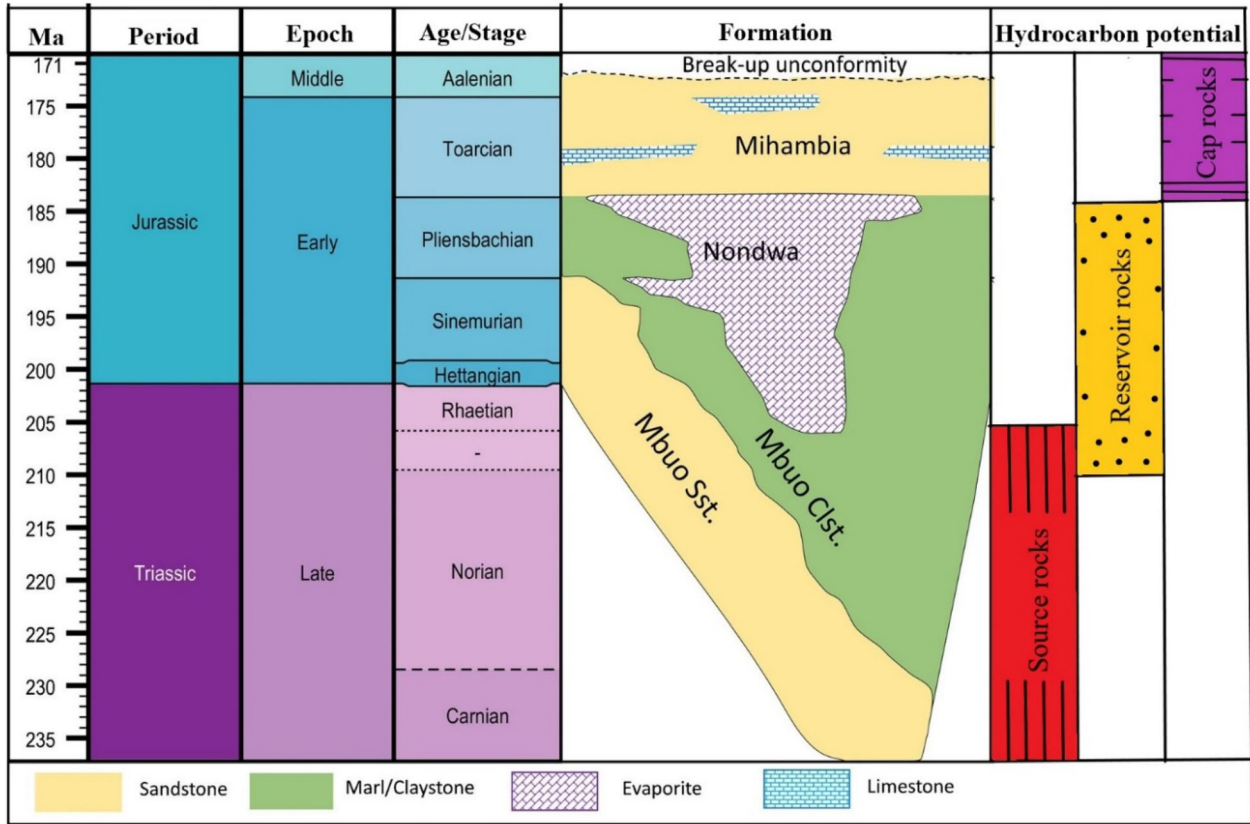


Figure 2. Lithostratigraphic details of the Pindiro Group in the Mandawa Basin (modified from Gama & Schwark, 2022).

ing sandstones, mudstones, and limestones (Gama & Schwark, 2022, 2023). The rifting event ended in late Triassic, and the Mandawa Basin entered a period of relative stability. During this time, the basin was filled with shallow marine sediments known as the Mavuji Group, which is composed of sandstones, mudstones, and limestones. In the Late Cretaceous, the Mandawa Basin was uplifted and eroded (Hou, 2015; McCabe, 2021). This uplift resulted in the formation of a series of cuestas and valleys. The cuestas are long, sloping ridges that are formed by resistant sandstones. The valleys are low-lying areas formed by softer mudstones and limestones (Einvik-Heitmann, 2016).

Moreover, according to the geological setting of the Mandawa Basin, it contained and exposed the Kilwa Group, a succession of Late Cretaceous to Paleogene age. The group comprises four formations: the Nangurukuru, Kivinje, Masoko, and Pande Formations (Nicholas et al., 2006). The Nangurukuru Formation seems to be the oldest. It is composed of variably lithified sandstones, mud-

stones, and shales, while the Kivinje Formation is composed of marine shales and mudstones that contain abundant fossils of planktonic foraminifera (McCabe et al., 2023). The Masoko Formation comprises shallow marine sandstones and mudstones with abundant fossils of benthic foraminifera and ostracods (Fossum et al., 2019). The youngest Pande Formation is composed of fluvial sandstones and mudstones with abundant fossils of land plants (Zhou et al., 2013). The source rock of the Mandawa Basin is the Mbuo Claystone in the late Triassic Pindiro Group and Nondwa evaporites in the early Jurassic Pindiro Group (Maganza, 2014) (Fig. 2).

## METHODOLOGY

### Data Description and Pre-processing

This study was conducted on three wells in the Mandawa Basin. The source rocks in the study area are composed mainly of Jurassic shale and Triassic

**Table 2.** Statistical features of the input datasets

Statistical features	DT ( $\mu\text{s/m}$ )	GR (API)	LLD (Ohm.m)	NPHI (v/v)	RHOB (gm/cc)	SP (mV)
<b>Mbate Well</b>						
Minimum	59.31	33.18	1.52	3.12	2.23	52.48
Maximum	127.73	88.65	22.15	37.79	2.47	80.33
Mean	95.63	62.61	6.29	19.24	2.37	66.71
Standard deviation	15.19	13.84	5.05	8.92	0.08	9.90
<b>Mbuo Well</b>						
Minimum	185.57	5.70	0.88	0.09	2.14	– 31.06
Maximum	439.10	84.08	6.33	0.42	2.58	– 5.61
Mean	285.17	53.43	2.96	0.29	2.35	– 23.16
Standard deviation	76.98	28.42	1.73	0.08	0.14	7.89
<b>Mita Gamma Well</b>						
Minimum	217.33	57.95	0.68	0.14	1.78	– 32.43
Maximum	444.30	125.54	43.08	0.39	2.61	– 2.56
Mean	307.75	85.38	8.35	0.28	2.39	– 19.21
Standard deviation	54.46	16.79	10.94	0.07	0.21	7.39

claystone, both from the Nondwa and Mbuo Formations of the Pindiro Group (Mshiu et al., 2022; Gama & Schwark, 2023). The input variables used to develop the model included the well log suite of deep lateral resistivity log (LLD), neutron porosity (NPHI), sonic travel time (DT), gamma-ray (GR), spontaneous potential log (SP), and bulk density log (RHOB). The holdout validation method was used to split the dataset into two parts: 70% of the data were used to train the model (data from Mbate Well and Mbuo Well), while 30% were used to validate the model's performance (data from Mita Gamma Well).

During data processing, feature selection was conducted to remove outliers that have the potential to compromise the accuracy of an estimating model or diminish its predictive performance. The relative impact of the input parameters was evaluated using the Pearson correlation coefficient ( $R$ ), thus:

$$b = \frac{1}{n} \sum_{i=1}^N b_i \quad (1)$$

where  $R_{a,b}$  is the Pearson correlation coefficient of variables  $a$  and  $b$ ,  $\bar{a} = \frac{1}{n} \sum_{i=1}^N a_i$  is the mean of  $a$ , and

$\bar{b} = \frac{1}{n} \sum_{i=1}^N b_i$  is the mean of  $b$ . The  $R_{a,b}$  is calculated to measure the linear association between two variables. It ranges from  $-1$  to  $1$ , with  $-1$  indicating a strong negative correlation and  $1$  indicating a strong positive correlation. The coefficient is calculated by dividing the covariance of the two variables by the product of their standard deviations. This calcula-

tion can be done for both normal and binary responses and can also be extended to fuzzy numbers (Cohen et al., 2009; Zhou et al., 2016).

The measured Tmax and well log data were normalized to a scale from 0 and 1 to reduce redundancy and improve data integrity. The normalization process was performed as:

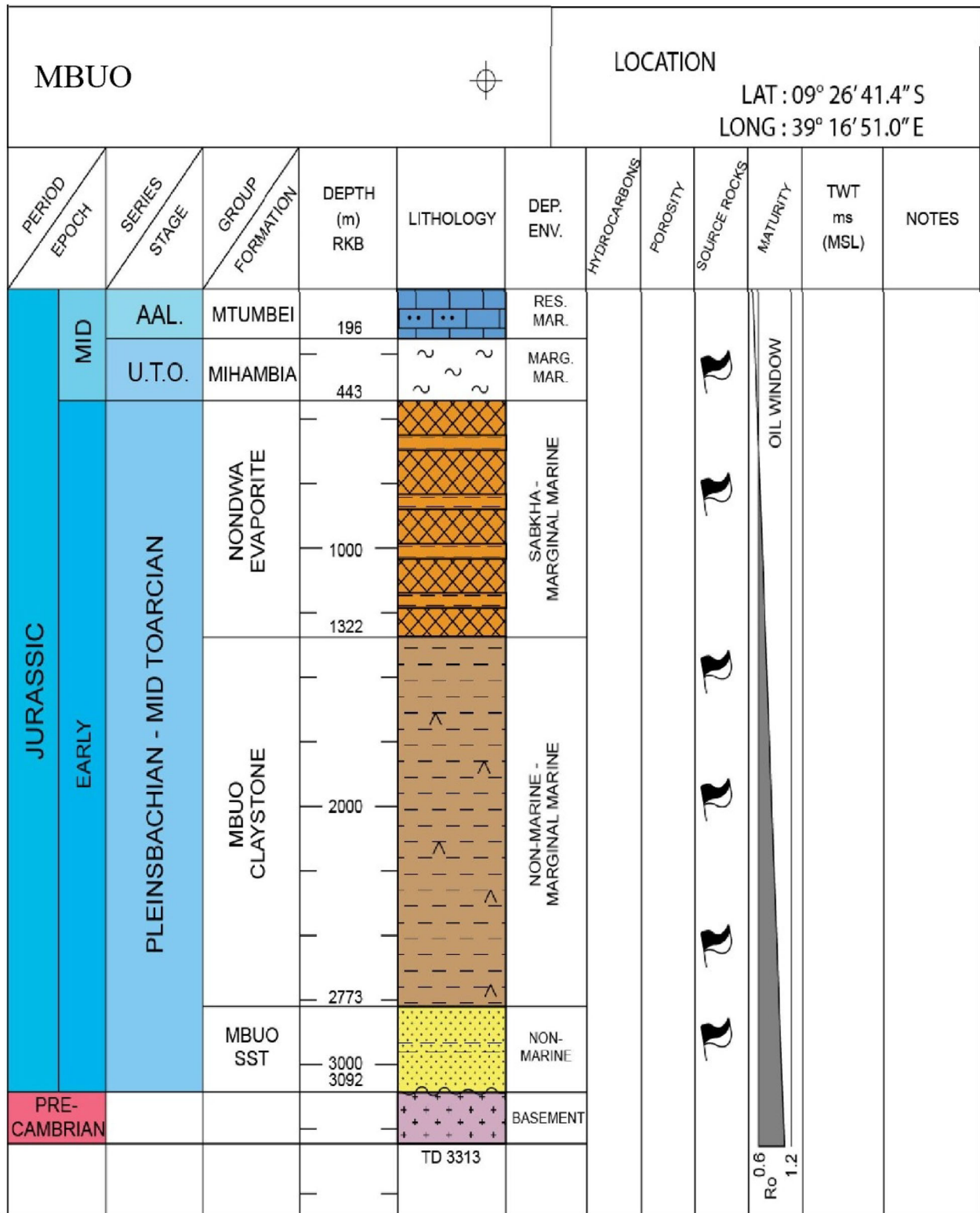
$$X_{\text{NORM}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \quad (2)$$

where  $x$  is the original value,  $X_{\text{NORM}}$  represents the normalized value in a dataset,  $x_{\text{max}}$  is the maximum value, and  $x_{\text{min}}$  is the minimum value. Table 2 presents the statistical features for the three well suites: Mbate, Mbuo, and Mita Gamma.

## Geochemical Analysis

Eighty-three core samples from the Mandawa Basin were collected for geochemical analysis. Fifty-nine core samples from Mbuo and Mbate Wells were used as training data, and the remaining 24 core samples from Mita Gamma Well were used as testing data. The depth intervals of the core samples were 1058–2135 m in Mbate Well, 1661–3145 m in Mbuo Well, and 1630–2150 m in Mita Gamma Well (Fig. 3). The samples were collected and brought to a laboratory for examination. A sample weighed 67.5 g and was subjected to crushing, sieving, and subsequent extraction and analysis using Rock-Eval pyrolysis, which employed a 25 °C min  $-1$  temperature schedule, with pyrolysis oven temperatures exceeding 750 °C and oxidation oven temperatures

**A Novel Hybrid Machine Learning Approach**



**Figure 3.** Lithostratigraphic column of the Mbuo Well.

**Table 3.** Rock pyrolysis data of Tmax and HI of the Mandawa Basin

Mbate Well			Mbuo Well			Mita Gamma Well		
Depth (m)	Tmax (°C)	HI (mg/g TOC)	Depth (m)	Tmax (°C)	HI (mg/g TOC)	Depth (m)	Tmax (°C)	HI (mg/g TOC)
1058	426	36	1661	425	64	1630	447	825
1128	434	25	1691.5	429	72	1661	446	689
1241	432	26	1710	419	610	1691.5	446	527
1298	429	27	1722	418	825	1710	446	389
1320	431	43	1813.5	419	689	1722	450	646
1348	431	51	1820	433	527	1764	435	272
1417	433	30	1844	433	389	1820	447	760
1445	425	92	1859	434	646	1844	440	997
1472	433	92	1874.5	436	272	1859	440	1017
1475	417	36	1905	432	64	1874.5	447	119
1606	425	34	1935	436	760	1905	448	78
1622	429	53	1942	430	997	1935	433	91
1661	419	153	1981	433	1000	1942	436	74
1715	418	153	1996	457	76	1981	436	44
1759	419	87	2003	446	318	1996	436	61
1832	433	171	2012	458	108	2003	441	45
1844	433	76	2012	441	99	2012	443	63
1859	434	216	2017	422	182	2017	433	188
1881	436	43	2026.5	445	96	2026.5	446	257
1887	433	153	2057	448	52	2057	437	139
1905	437	43	2057.5	473	575	2070.5	446	13
1935	443	35	2070.5	447	69	2089	444	16
1942	445	285	2089	435	58	2118	443	69
1981	441	120	2118	432	88	2150	445	197
1996	446	66	2148.5	442	69			
2017	432	65	2889	437	390			
2057	430	175	3068	441	66			
2076	426	228	3090	446	54			
2103	435	65	3145	432	215			
2135	433	79						

above 800 °C. The Rock–Eval pyrolysis results for Tmax and hydrogen index (HI) are presented in Table 3.

### Basin Modeling

Basin modeling is a tool used to predict maturation, hydrocarbon generation, expulsion, and migration in exploration geology (Abdel-Fattah et al., 2017). It involves building 1D models to analyze burial and temperature histories, as well as maturity and hydrocarbon generation and expulsion. This study utilized 1D basin modeling through PetroMod software (version 2012) to analyze the Tmax of the Pindiro Group. Essential parameters, including formation names, depths, thicknesses, and

deposition ages, were used as inputs (Allen & Allen, 2013; Ahmed et al., 2019). Tmax and vitrinite reflectance were measured for model calibration. A constant heat flow of 64 mW/m<sup>2</sup> was employed as described by Wygrala (1989), while the Burnham and Sweeney (1989) kinetic model was utilized due to the oil/gas-prone nature of the source rock. Relative petroleum system elements were assigned to each formation, along with TOC and HI parameters as input.

### Back-Propagation Neural Network (BPNN)

Wang et al. (2019) reported the back-propagation approach as a supervised learning algorithm commonly used in neural networks. This approach



## A Novel Hybrid Machine Learning Approach

adjusts the network weights to minimize the error between estimated and actual output (Titus et al., 2022). It is based on the gradient descent method that calculates the error gradient in response to the weight using the chain rule (Wu & Tong, 2022). It has been shown that the bias concept often works as a set of weights where the signals are sent in opposite directions during the back-propagation learning phase (Sun et al., 2021; Dai, 2023). BPNN was built as a way to solve the multilayer perceptron training problem. However, the two main improvements of the BPNN were the addition of a differentiable function at each node and the internal network weight change due to back-propagation error after each training epoch (Che Nordin et al., 2021).

### Group Method Data Handling (GMDH)

GMDH is the association of a multilayer algorithm that generates a network of layers and nodes by utilizing several inputs from the analyzed data stream (MolaAbasi et al., 2021). 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 (Lal and Datta, 2021). The self-organizing inductive propagation algorithm is a technique that can solve complex problems (Roshani et al., 2020; Lv et al., 2023). 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. Volterra's series function, the discrete analog of the Kolmogorov–Gabor polynomial, can describe a generic relation of output–input (Nelles, 2020).

$$u = a + \sum_{i=1}^n b_i x_i + \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n d_{ijk} x_i x_j x_k + \dots \quad (3)$$

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

### GMDH Optimized by Differential Evolution (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, the DE is an innovative parallel direct search optimization technique introduced by Price and Storn (1995). 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 & Price, 1995; 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 GMDH–DE can effectively handle complex nonlinear relationships and improve predictive performance. The basic equation for the method is:

$$x^* = \operatorname{argmin} F(x) \quad (4)$$

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

In the GMDH–DE, the process starts with the creation of an initial population of candidate models using the GMDH algorithm. These models are then evaluated based on their fitness, typically using a performance metric like mean squared error or correlation coefficient, to determine their effectiveness in predicting Tmax. The DE algorithm is then applied to evolve and optimize the parameters of the candidate models, such as the coefficients of the polynomial regression, to further improve accuracy. This iterative process continues until a satisfactory model with optimized parameters is obtained, providing a reliable prediction of Tmax in geological formations. Table 4 summarizes the workflow steps followed by the GMDH–DE to predict Tmax.

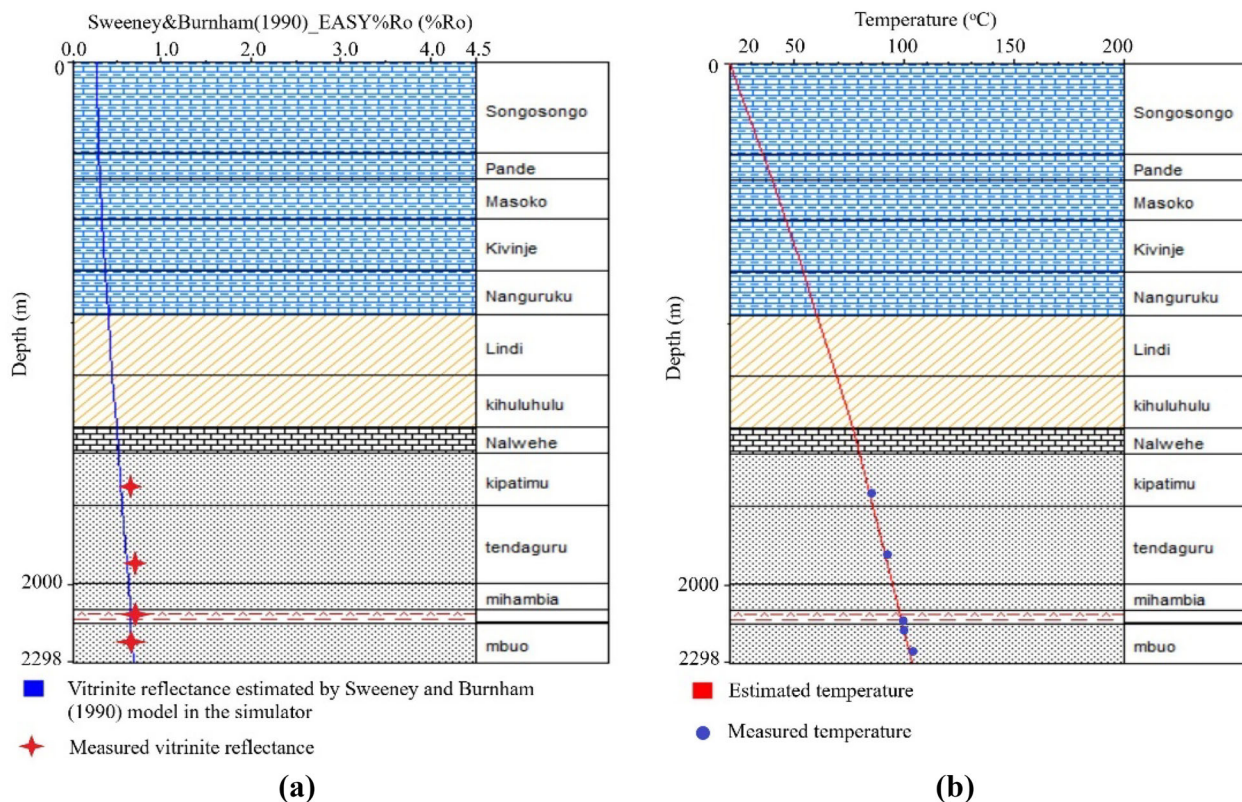
**Table 4.** Workflow of the GMDH–DE algorithm

No.	Working
1.	Start
2.	Initialize the inputs
3.	Setting GMDH's parameters
4.	Adjust the hidden layers
5.	Modify the hidden layer's neuron count
6.	Selection of GMDH's learning algorithm
7.	Selection of learning rate
8.	Training the GMDH model, followed by testing
9.	Determine the minimum convergence value by evaluating the objective function
10.	Determine the trained model's weights and biases
11.	Initialize the DE algorithm's parameters
12.	Limit sample search space for extracted weights and biases
13.	Use the retrieved weights and biases as the starting population in a DE method
14.	Continue iteratively until the stopping criterion* is met
15.	Select the optimal global solution
16.	Set optimum weights and biases from the globally best model in the network for the estimation of Tmax parameters
17.	Ends

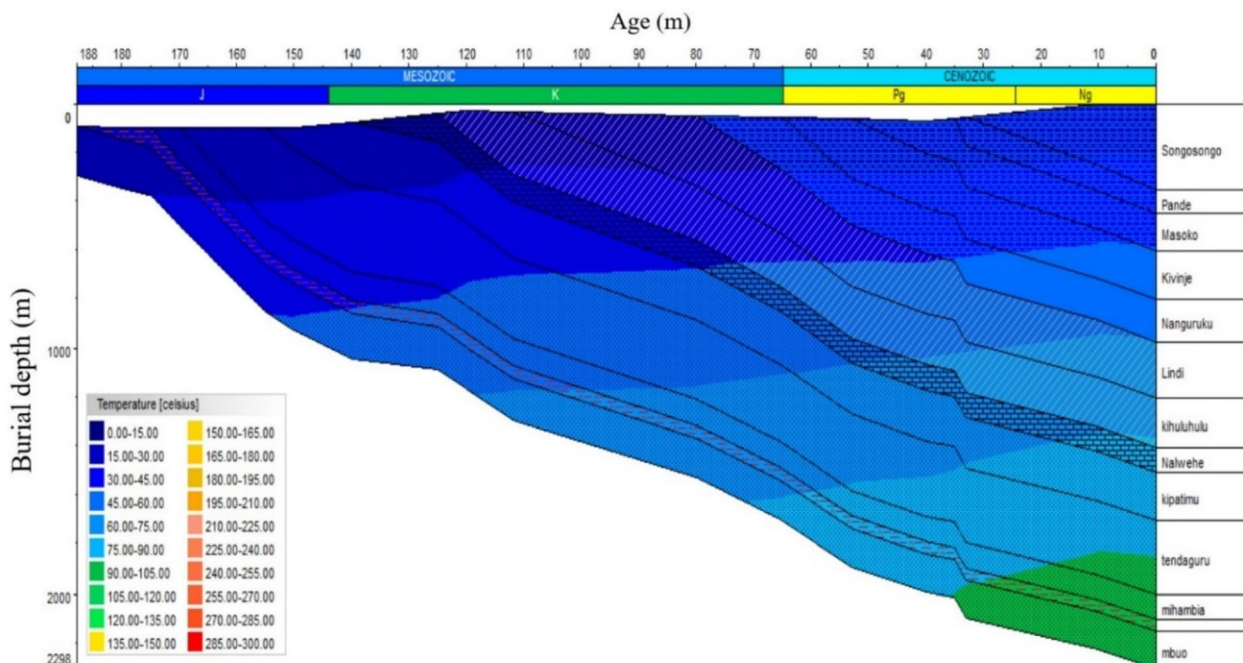
## RESULTS AND DISCUSSION

### 1D Basin Modeling Analysis

A change in heat flow ranging from 50 to 70 mW/m<sup>2</sup> leads to a depth difference of approximately 1 km for the 100 °C isotherm, often considered as the lower limit for the oil generation window. This variation in heat flow across a basin can significantly influence the maturity stages of potential source rocks, resulting in considerable differences in their thermal evolution. It is essential to calibrate the thermal and maturity history in basin modeling by utilizing borehole temperature data and the vitrinite reflectance (Ro) measurements of source rocks (Hantschel & Kauerauf, 2009). According to 1D burial profiles, the source rock from the early Jurassic period shows a maximum burial depth that is comparable to the current day (2001–2287 m). As shown in Figure 4a, modeling and calibration data agreed well. Based on the Sweeney and Burnham (1990) classification, the analysis revealed that the Mbuo Formation's source

**Figure 4.** Tmax history showing (a) the fitness of calibrated and measured Ro; (b) temperature vs. depth.

## A Novel Hybrid Machine Learning Approach



**Figure 5.** Burial depth history with temperature variation curve for the Triassic source rock in the Mandawa Basin.

rock displayed vitrinite reflectance levels of 0.50–0.68% Ro (Fig. 4a), indicative of temperatures ranging from 90 to 103.3 °C (Fig. 4b). This suggests that the source rock spans the immature to mature stages of hydrocarbon generation, specifically within the gas–oil window.

Evidence from the Mbuo Well indicates that the Mbuo Formation’s base has been heated to 103.3 °C throughout 0.17 Ma, having descended to a depth of 2287.69 m (Fig. 5). The generation started during the late Triassic to the early Jurassic in both Mbuo Claystone and Mbuo Sandstone and continues up to recent. Other overlying formations such as the Nondwa evaporites (intercalated with shales) and minor claystone in the Mihambia Formation are immature based on the modeling result. The measured data for Ro and temperature are provided in the Appendix along with different inputs used for basin modeling. The beginning of the immature stage (0.50% Ro) of the Mbuo Formation was noted in the Mbuo Well at a depth of 1728 m during the Paleogene period (62 Ma). At a depth below 1965 m, during the middle Paleogene (42 Ma), the source rock of interest had its early oil window

(0.56% Ro). In the Neogene period (0.69 Ma), at a depth of 2287.7 m, the primary oil window began (Fig. 6)

### GMDH–DE Model Development

The GMDH–DE model comprised six input neurons and two hidden layers, namely,  $h_1$ ,  $h_2$ ,  $h_3$ , and  $h_4$  in the first layer and  $v_1$  and  $v_2$  in the second layer. The output of the model was represented as  $y$ .

Figure 7 presents a neural network structure of the proposed model in predicting Tmax. The equations for the layer of the neural network model needed to provide the Tmax estimation are presented in Table 5.

### Performance Indicators

The GMDH–DE, GMDH, and BPNN models were coded and implemented in MATLAB R2022a on an AMD Ryzen 5 5600U with Radeon Graphics 2.30 GHz running Windows 10 operating system.

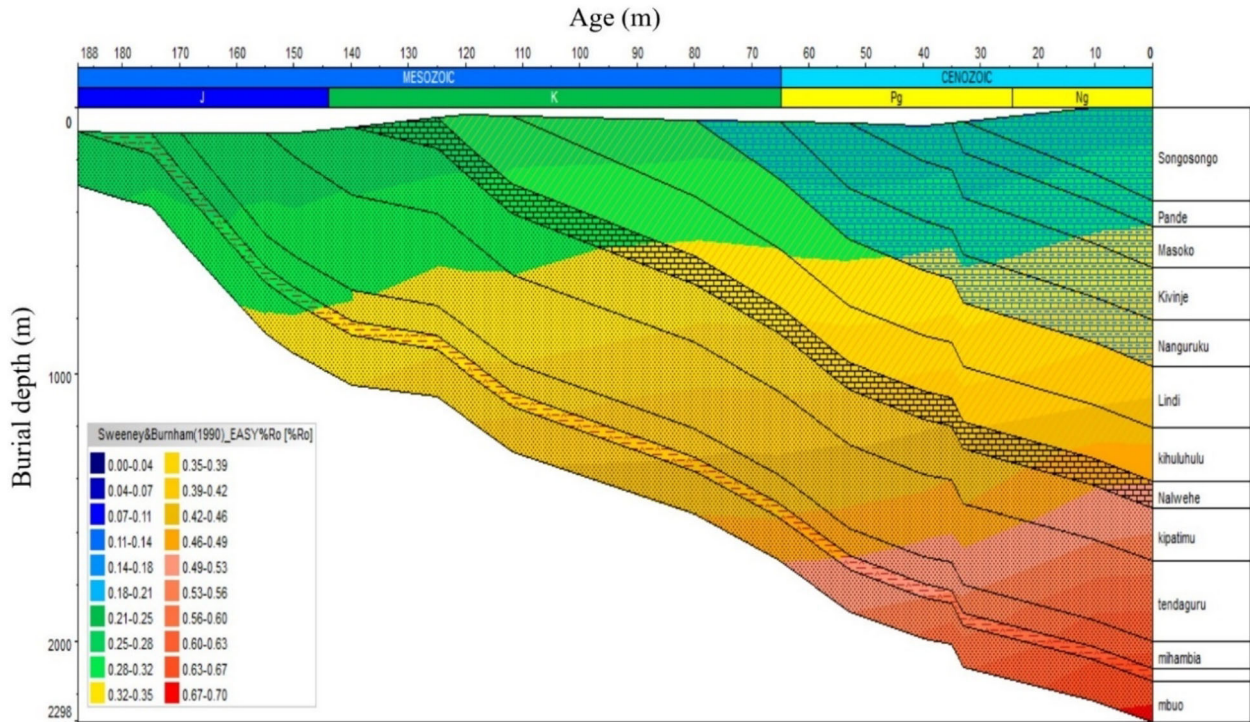


Figure 6. Burial depth history with calibrated vitrinite reflectance for the Triassic source rocks in the Mandawa Basin.

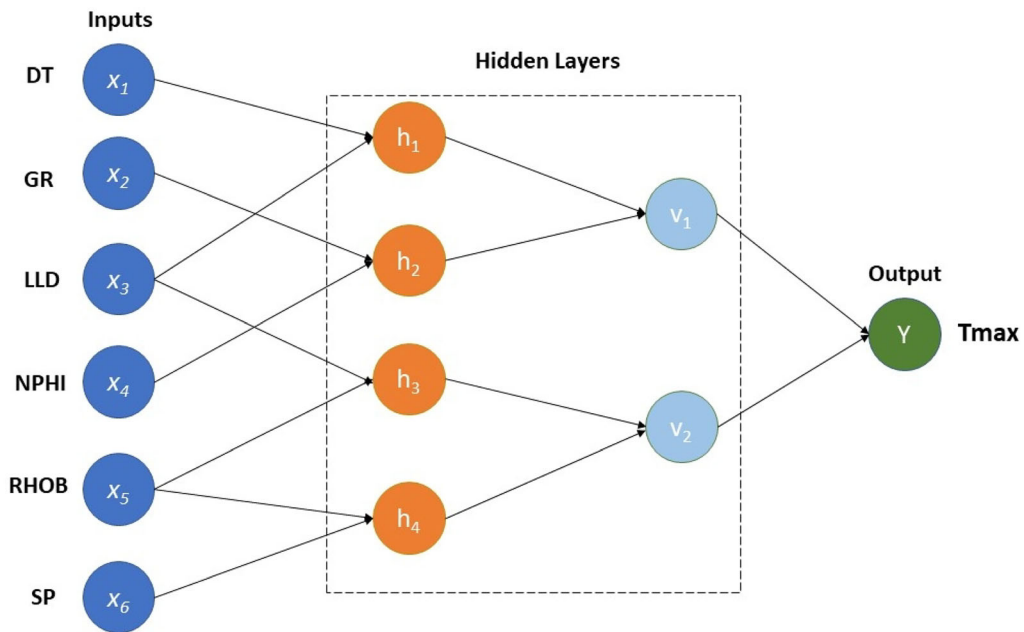


Figure 7. The GMDH-DE network structure.

## A Novel Hybrid Machine Learning Approach

**Table 5.** Proposed equations for Tmax estimation

Layers	Neurons	Equations
1	$h_1$	$h_1 = -631.8285 + 38.8594x_1 - 752.9471x_3 - 0.0179x_1^2 + 456.0425x_3^2 - 14.4068x_1x_3$
	$h_2$	$h_2 = -1334.1104 - 13.7616x_2 + 57.5324x_4 + 0.0905x_2^2 - 0.4525x_4^2 + 0.1442x_2x_4$
	$h_3$	$h_3 = 45065.4234 - 172.3706x_3 - 35171.0970x_6 + 0.1667x_3^2 + 6930.8268x_6^2 + 67.7218x_3x_6$
	$h_4$	$h_4 = 338.0869 + 17.8998x_5 + 2.4854x_6 - 0.0019x_5^2 - 0.0149x_6^2 - 0.2579x_5x_6$
2	$v_1$	$v_1 = 0.2037 + 65.2220h_1 - 65.7311h_2 - 0.1553h_1^2 - 0.0063h_2^2 + 0.1650h_1h_2$
	$v_2$	$v_2 = 0.0059 + 3.2159h_3 - 2.5669h_4 + 0.0069h_3 + 0.0138h_4 - 0.0198h_3h_4$
Output (y)		$y = -0.0119 + 15.4842v_1 - 14.8671v_2 + 0.0249v_1^2 + 0.0591v_2^2 - 0.0832v_1v_2$

where  $x_1$  is sonic travel time (DT),  $x_2$  is gamma-ray (GR),  $x_3$  is deep lateral resistivity log (LLD),  $x_4$  is neutron porosity (NPHI),  $x_5$  is bulk density log (RHOB), and  $x_6$  is spontaneous potential log (SP). At the same time,  $h_1, h_2, h_3, h_4, v_1,$  and  $v_2$  are the neurons for the hidden layers.

The coefficient of determination ( $R^2$ ), root mean square error (RMSE), and mean absolute error (MAE) were the evaluation metrics used to assess the performance of the estimation models. The values of  $R^2$  vary between 0 and 1; a model is more effective when its  $R^2$  value is higher, and 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; Mula-shani et al., 2022). At the same time, RMSE is a measure of the differences between predicted values and observed values. Excellent model accuracy is defined by RMSE of < 10%, good model accuracy by RMSE of 10–20%, fair model accuracy by RMSE of 20–30%, and poor model accuracy by RMSE of > 30% (Yao et al., 2021; Hussain et al., 2023). Moreover, MAE is a metric used to measure the average magnitude of errors in a regression model; a lower MAE indicates better performance, as it represents a smaller average error between estimated and actual values (Ali et al., 2023). The  $R^2$ , RMSE, and MAE mathematical expressions can be presented, respectively, as follows (Chong et al., 2022; Ramos et al., 2023):

$$R^2 = \left( \frac{\sum_{i=1}^n (a_i - \bar{a})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (a_i - \bar{a})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}} \right)^2 \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (a_i - P_i)^2}{n}} \quad (6)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{a_i - P_i}{a_i} \right| \quad (7)$$

where  $P_i$  is the estimated  $T_{\max}$  value from each model,  $a_i$  represents the actual value of  $T_{\max}$  measured from core samples,  $\bar{a}$  and  $\bar{P}$  are the true mean and estimated mean values for  $T_{\max}$ , respectively, and  $n$  represents the number of samples.

### Hyperparameters Tuning

Hyperparameter tuning refers to adjusting the parameters of a machine learning model to optimize its performance. These parameters, known as hyperparameters, are set before training and not learned during training (Pravin et al., 2022). Hyperparameter tuning is essential because the performance of a machine learning model can be significantly affected by the values of its hyperparameters (Yang & Shami, 2020). It involves experimenting with different combinations of hyperparameters to find optimal values that produce the best performance. In this study, hyperparameters optimization was done using the DE algorithm, a parallel direct search technique for determining the structure of polynomial neurons in the GMDH. The DE generates improved values for the hyperparameters in each model loop and then inputs them to the GMDH to assess the model's performance on the testing data. The performance of the GMDH was re-evaluated in the following phase. If it was satisfactory, the optimization was terminated; otherwise, the process continued until the stopping criteria were met, and the optimal hyperparameters were obtained. The hyperparameter configuration yielding optimal results comprised a population size of 70 individuals, a mutation rate of 0.7, and a cross-

over rate of 13. This configuration fostered a balanced exploration of the solution space, facilitating the discovery of robust models. The choice of two layers with 17 neurons each effectively captured complex relationships within the data, enhancing the model's capacity to discern patterns indicative of  $T_{max}$ . By setting a stopping criterion of 300 iterations, the model achieved convergence while mitigating the risk of overfitting. Additionally, employing a moderate learning rate of 0.1 ensured stable and efficient training, allowing the model to effectively learn from the data without diverging or stagnating. Overall, the setting optimally balanced exploration and exploitation, thereby facilitating accurate estimation of the  $T_{max}$  index in the GMDH-DE model. The optimal hyperparameter setting for the GMDH-DE estimation model is shown in Table 6.

### Estimation of $T_{max}$ during Training Performance

The training results summarized in Table 7 showcase the superior performance of the GMDH-DE model in estimating  $T_{max}$  compared to traditional models like GMDH and BPNN. This is evidenced by its notably higher  $R^2$  of 0.995 (Fig. 8), indicating strong correlation between estimated and observed values, coupled with lower error metrics including RMSE (0.004) and MAE (0.006) (Fig. 9) and a substantially shorter computational time of 0.2 seconds. The success of the GMDH-DE model can be attributed to its adeptness in mitigating

**Table 6.** Hyperparameter settings for the GMDH-DE model

Hyperparameters	Selected optimum values
Population size	70
Mutation rate	0.7
Cross-over rate	13
Number of layers	2
Number of neurons in each layer	17
Stopping criterion	300 iterations
Learning rate	0.1

overfitting, a common challenge in model training. The choice of a moderate learning rate (0.1) and a stopping criterion of 300 iterations effectively regulated the model's complexity and prevented it from overfitting noise in the data, as discussed above in the context of the hyperparameter settings. Moreover, in the geological setting of the Mandawa Basin, where variations in  $T_{max}$  are influenced by complex interactions of organic matter and geological processes, the GMDH-DE model's capacity to capture intricate patterns and nonlinear relationships between input variables and  $T_{max}$  is particularly advantageous. Its ability to adaptively select features and construct hierarchical models aligns well with the complex nature of geological systems, thereby facilitating more accurate and robust predictions compared to the simpler architectures of GMDH and BPNN. Consequently, the GMDH-DE model emerges as the preferred choice for  $T_{max}$  estimation in such a geological environment, offering superior predictive performance and computational efficiency.

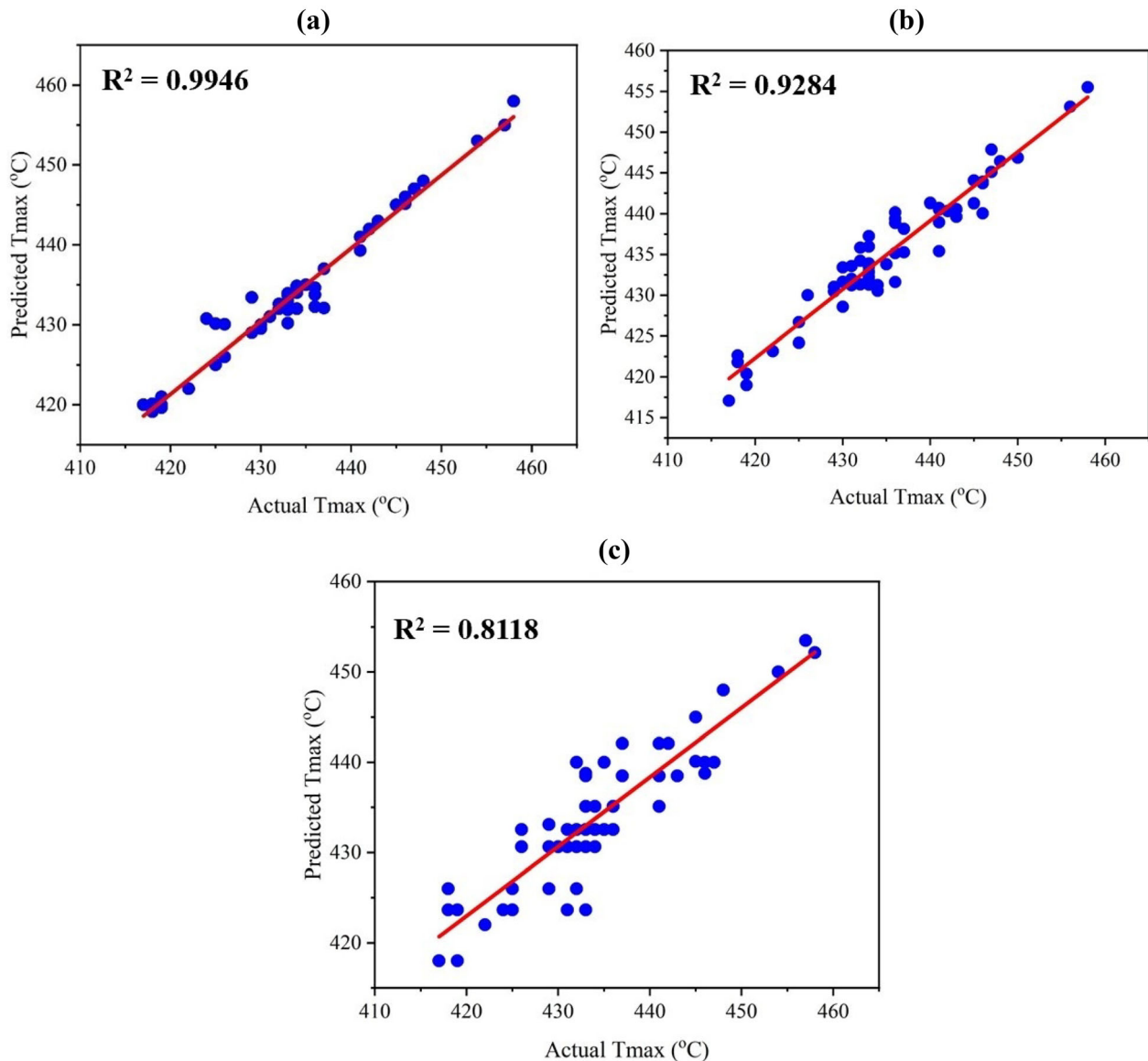
### Estimation of $T_{max}$ During Testing Performance

The results of the  $T_{max}$  estimation during model testing (Table 8) reveal notable performance differences among the models considered. The GMDH-DE model exhibited superior performance, achieving  $R^2$  of 0.970 (Fig. 10), low RMSE of 0.017, and minimal MAE of 0.025 (Fig. 11), all within a remarkably short computational time of 0.5 seconds. Contrastingly, the traditional GMDH and BPNN models demonstrated inferior performance, with lower  $R^2$  values accompanied by higher RMSE and MAE, and longer computational times. The efficacy of the GMDH-DE model can be attributed to its hyperparameter settings, notably the learning rate and stopping criterion, which contributed to the model's generalizability. By employing a learning rate of 0.1 and a stopping criterion of 300 iterations, the GMDH-DE model effectively balanced the trade-off between model complexity and overfitting,

**Table 7.** Value of  $R^2$  for the fitted equations per model during the training phase

Model	Equation	$R^2$	RMSE	MAE	Computational time (s)
GMDH-DE	$y = 37.731 + 0.9313x$	0.995	0.004	0.006	0.2
GMDH	$y = 68.510 + 0.8424x$	0.928	0.013	0.018	3.5
BPNN	$y = 100.096 + 0.7688x$	0.812	0.039	0.043	5.2

## A Novel Hybrid Machine Learning Approach



**Figure 8.** Cross-plots of actual Tmax vs. Tmax predicted by (a) GMDH-DE, (b) GMDH, and (c) BPNN during training.

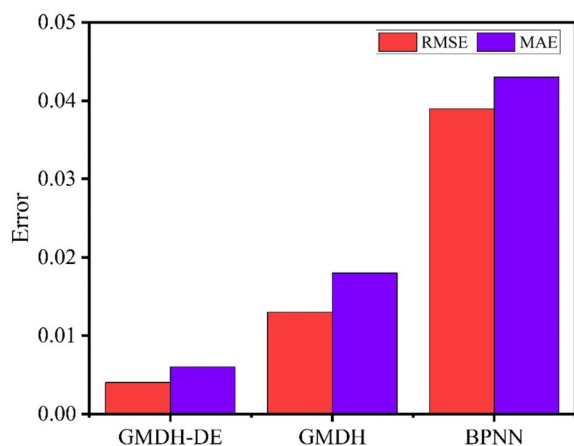
allowing it to generalize well to unseen data. This is particularly crucial in geological settings such as the Mandawa Basin, characterized by diverse and intricate geological processes influencing Tmax. Additionally, the GMDH-DE model's adaptive and self-organizing nature enabled it to capture complex nonlinear relationships inherent in geological data,

thereby outperforming traditional models like GMDH and BPNN. Moreover, the success of the GMDH-DE model in Tmax estimation is linked to the capacity to reveal the hierarchical connections within the data through the multilayer structure and evolutionary optimization process. This enables the model to effectively leverage the geological features

specific to the Mandawa Basin, thus enhancing its predictive accuracy. Overall, the GMDH-DE model's robust hyperparameter settings, coupled with its adaptive nature and ability to capture complex geological processes, culminated in its superior performance compared to traditional models in estimating Tmax in the Mandawa Basin.

### Comparison with Previous Studies

Generally, the results of the proposed GMDH-DE were further compared with the previously developed hybrid models of PSO-ANN and PCA-ANN, which were used in the estimation of Tmax (Table 9). The GMDH-DE performed better than both models suggested by Tariq et al. (2020) and Barham et al. (2021) during training by obtaining higher  $R^2$  of 0.995, while the PSO-ANN and PCA-ANN had  $R^2$  of 0.917 and 0.88, respectively. Likewise, during testing, the GMDH-DE performed better by obtaining higher  $R^2$  of 0.9703, followed by PSO-ANN with  $R^2$  of 0.918 and PCA-ANN with  $R^2$  of 0.8518 (Fig. 12).



**Figure 9.** RMSE and MAE of GMDH, GMDH-DE, and BPNN models for Tmax estimation during training.

### SHAP (SHapley Additive exPlanations)

In this study, the GMDH-DE model estimated Tmax and provided valuable insights into feature relevance through SHapley Additive exPlanations (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 importance in Fig. 13 highlighted the substantial impact of the SP parameter on the GMDH-DE model's Tmax estimation, with mean SHAP value of 4.29. Additionally, the DT, GR, RHOB, and LLD parameters had a moderate impact on Tmax estimation, as indicated by their mean SHAP values of 1.76, 1.26, 1.04, and 0.91, respectively, reflecting their role in conveying information about clay content in the Wangkwar Formation. NPHI contributed the least to Tmax estimation, with mean SHAP value of 0.35. Moreover, Figure 14 illustrates that an increase in SP, DT, and GR led to an increase in Tmax, while higher values of RHOB, LLD, and NPHI resulted in a decrease in Tmax.

### Assessment of Kerogen Type and Maturity Stage

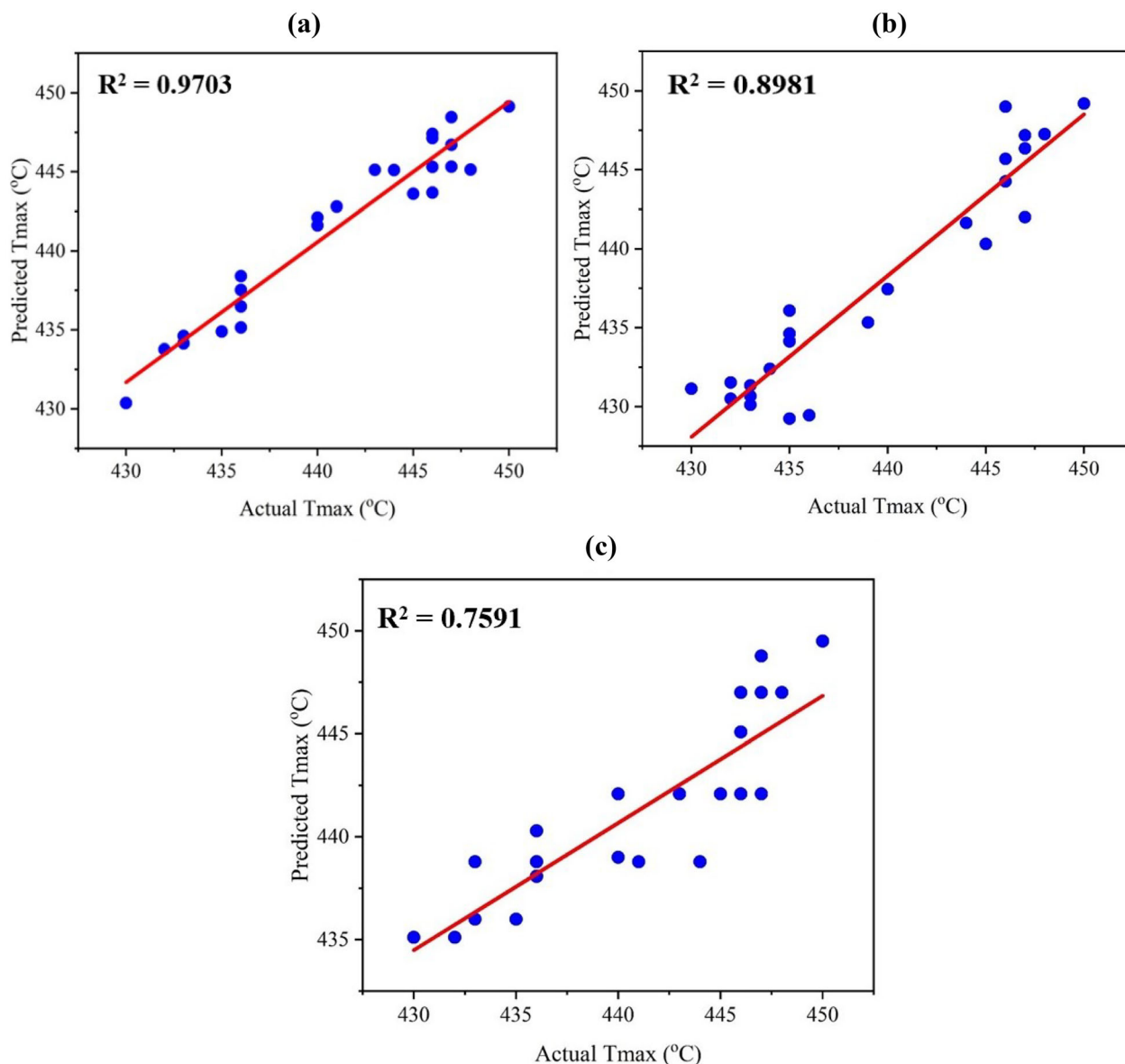
Based on the results estimated by GMDH-DE, kerogen classification diagrams were constructed using the HI vs. Tmax plot utilized by the earliest researcher to determine the maturity stage and kerogen type (Fig. 15). Overall, the findings indicated that most of the analyzed samples from the Triassic-Jurassic source rocks in the Mbate and Mbuo Wells are typically plotted in the immature zone of Types I to III kerogens with Tmax of < 435 °C belonging to the gas-oil generation window and signifying the incapability of the rocks to generate hydrocarbons (Tissot & Welte, 2013; Al-Areeq et al., 2018). These findings correspond to the one attained from basin modeling analyzed from the PetroMod. Their HI values justify this in the 52-

**Table 8.** Value of  $R^2$  for the fitted equations per model during the testing phase

Model	Equation	$R^2$	RMSE	MAE	Computational time (s)
GMDH-DE	$y = 49.877 + 0.9079x$	0.970	0.017	0.025	0.5
GMDH	$y = 94.135 + 0.7816x$	0.898	0.031	0.037	3.8
BPNN	$y = 168.820 + 0.6178x$	0.759	0.056	0.064	5.6



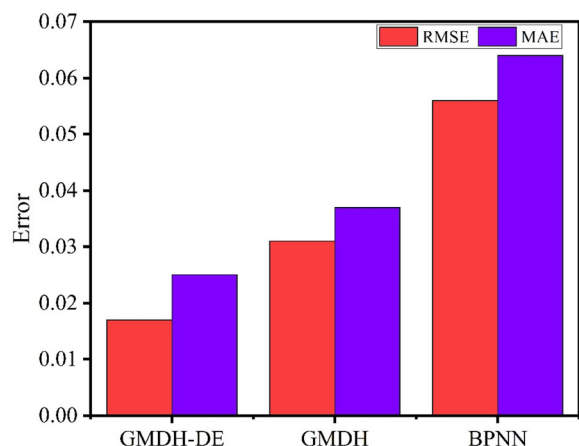
## A Novel Hybrid Machine Learning Approach



**Figure 10.** Cross-plots of actual Tmax vs. Tmax predicted by (a) GMDH-DE, (b) GMDH, and (c) BPNN models during the testing.

1017 mg HC/g TOC range. Moreover, the results revealed that very few samples from Mbate and Mbuo Wells are plotted in mature zones of Types II to III kerogens as indicated by their Tmax of 435–460 °C and HI of 13–285 mg HC/g TOC. In Figure 15, most samples from the Mita Gamma Well

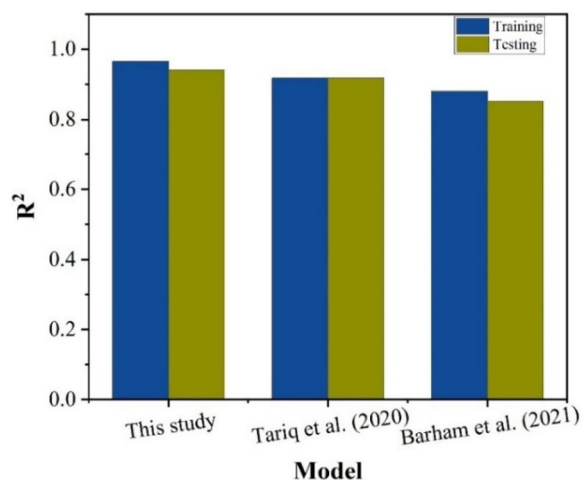
are plotted in the mature zone of the Types II to III kerogens field as indicated by higher Tmax of 440–460 °C, which are in line with the result from the classification suggested by Peters (1986) and Peters and Cassa (1994). In contrast, only a few samples are plotted in an immature zone of Types I to III



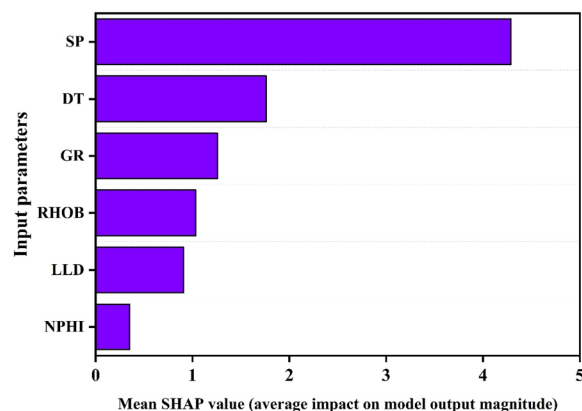
**Figure 11.** RMSE and MAE of GMDH, GMDH-DE, and BPNN models for Tmax estimation during testing.

**Table 9.** Comparison of the proposed model with the previously developed hybrid models for predicting Tmax

Model	Training		Testing		References
	$R^2$	RMSE	$R^2$	RMSE	
GMDH-DE	0.995	1.840	0.970	1.540	This study
PSO-ANN	0.917	0.258	0.918	0.428	Tariq et al. (2020)
PCA-ANN	0.880	-	0.852	-	Barham et al. (2021)



**Figure 12.** Comparison of models from the present study, Tariq et al. (2020) and Barham et al. (2021), for predicting Tmax in terms of  $R^2$ .



**Figure 13.** Input parameter importance for Tmax estimation.

kerogens as indicated by the Tmax of  $< 435$  °C (Fig. 15).

## CONCLUSIONS

The study has shown the proposed GMDH-DE Tmax model, as an independent novel approach, can be adopted for rapid real-time assessment of Tmax values of the organic matter of source rock. Therefore, based on this study, the following conclusions can be drawn:

1. One-dimensional Tmax modeling suggests that the lower Jurassic Mbuo source rocks entered the gas-oil window in late Triassic times and reached the expulsion onset during the early Jurassic. The Tmax thermal maturation and basin modeling vitrinite reflectance showed that Mbuo source rocks are immature to mature.
2. The Tmax estimation using the GMDH-DE model was compared to that using GMDH and BPNN. The GMDH and BPNN models underestimated the Tmax values significantly. In contrast, the GMDH-DE outperformed the other models with estimates very close to the measured values. Therefore, the model accurately and precisely estimated the organic matter of the source rock and can be used in different case studies for positive results.
3. The results of the proposed GMDH-DE model were compared with those of the previously developed hybrid models of PCA-ANN and PSO-ANN; the former model performed better than the latter models. The sensitivity analysis also showed that well log parameters of LLD,

## A Novel Hybrid Machine Learning Approach

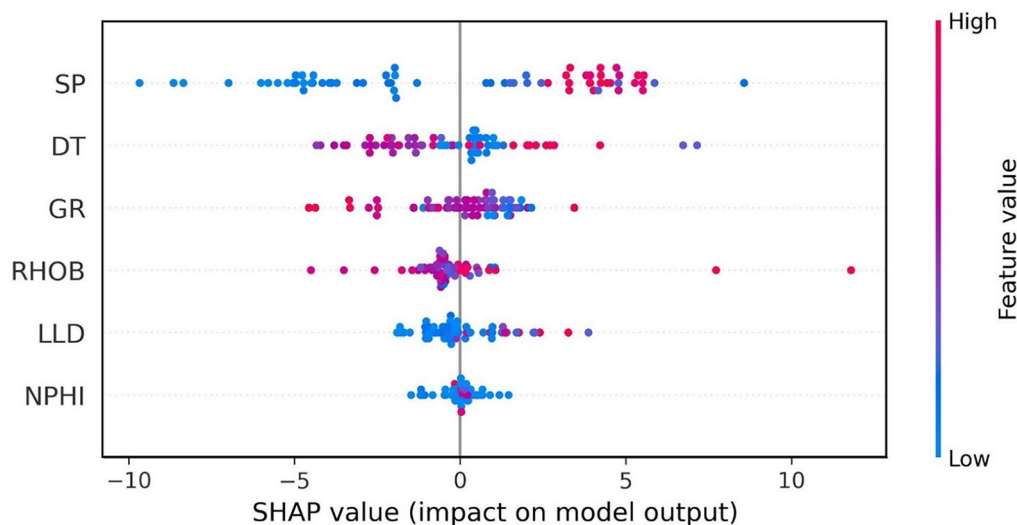


Figure 14. Input parameter influence for Tmax estimation.

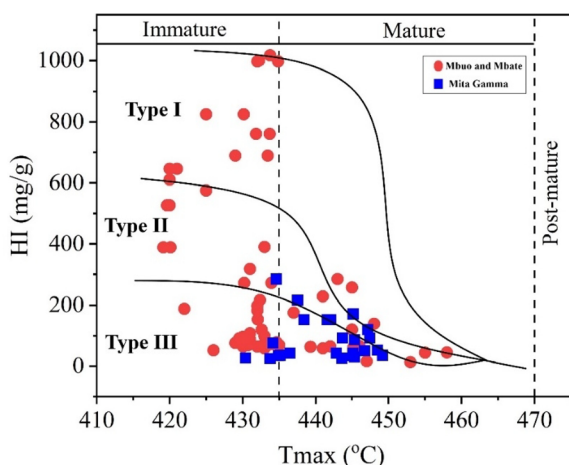


Figure 15. Cross-plots of HI vs. Tmax estimated by GMDH–DE model to show the basin’s maturity stage and kerogen type.

RHOB, and GR had the most significant contribution to the performance of the GMDH–DE model in Tmax estimation.

- The GMDH–DE model outperformed the GMDH, BPNN, PSO–ANN, and PCA–ANN models in predicting Tmax values from well logs. Therefore, exploration and development of oil and gas resources might be significantly facilitated using the proposed hybrid GMDH–DE technique.
- Source rocks analysis showed that the Mbate, Mbuo, and Mita Gamma Wells have fair-to-good

generation potential, as indicated by HI and  $T_{max}$  values. The HI values characterize kerogen Types II and III. The wells lie in the immature-to-mature window zone indicated by the HI vs.  $T_{max}$  plot. Therefore, it could be expected that the wells may have generated oil and gas.

## FUNDING

This work was supported by the Major National Science and Technology Programs in the “Thirteenth Five-Year” Plan period (No. 2017ZX05032-002-004) and The Innovation Team Funding of Natural Science Foundation of Hubei Province (No. 2021CFA031).

## DATA AVAILABILITY

Data used in this study are available from the corresponding author upon reasonable request.

## DECLARATIONS

**Conflict of 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.

## APPENDIX See (Table 10).

Table 10. Data used for basin modeling

Wells	Formation	Top	Base	Age		Temperature (°C)	Vitrinite reflectance (%Ro)
				From	To		
	Songosongo	0	350	33	0		
	Pande	350	450	35	33		
	Masoko	450	600	53	35		
	Kivinje	600	800	65	53		
	Nangurukuru	800	970	80	65		
	Lindi	970	1200	112	80		
	Kihuluhulu	1200	1400	125	112		
	Nalwehe	1400	1500	140	125		
	Kipatimu	1500	1700	155	140	85	0.5
	Tendaguru	1700	2000	170	155	93	0.6
Mita Gamma	Mihambia	2000	2100	175	170	98	0.6
Mbuo and Mbate	Nondwa	2100	2150	188	175	100	0.6
Mbuo	Mbuo	2150	2300	200	188	102	0.5

## REFERENCES

- Abay, T. B., Fossum, K., Karlsen, D. A., Dypvik, H., Narvhus, L. J. J., Haid, M. H. M., & Hudson, W. (2021). Petroleum geochemical aspects of the Mandawa Basin, coastal Tanzania: the origin of migrated oil occurring today as partly biodegraded bitumen. *Petroleum Geoscience*, 27, 2019–2050.
- Abdel-Fattah, M. I., Pigott, J. D., & Abd-Allah, Z. M. (2017). Integrative 1D–2D basin modeling of the cretaceous Beni Suef basin, Western Desert. *Egypt Journal of Petroleum Science and Engineering*, 153, 297–313.
- Abdizadeh, H., Ahmadi, A., Kadkhodaie, A., Heidarifard, M., & Shayeste, M. (2017). Estimation of thermal maturity from well logs and seismic data in the Mansuri oilfield, SW Iran. *Journal of Petroleum Science and Engineering*, 159, 461–473.
- Adeyilola, A., Zakharova, N., Liu, K., Gentzis, T., Carvajal-Ortiz, H., Ocubalidet, S., & Harrison, W. B. (2022). Hydrocarbon potential and Organofacies of the Devonian Antrim Shale, Michigan Basin. *International Journal of Coal Geology*, 249, 103905.
- Ahangari, D., Daneshfar, R., Zakeri, M., Ashoori, S., & Soulgani, B. S. (2022). On the prediction of geochemical parameters (TOC, S1 and S2) by considering well log parameters using ANFIS and LSSVM strategies. *Petroleum*, 8, 174–184.
- Ahmed, M. A., Hegab, O. A., Awadalla, A. S., Farag, A. E., & Hassan, S. (2019). Hydrocarbon generation, in-source conversion of oil to gas and expulsion: Petroleum system modeling of the Duwi Formation, Gulf of Suez. *Egypt Natural Resources Research*, 28, 1547–1573.
- Al-Areeq, N. M., Al-Badani, M. A., Salman, A. H., & Albaroot, M. A. (2018). Petroleum source rocks characterization and hydrocarbon generation of the Upper Jurassic succession in Jabal Ayban field, Sabatayn Basin, Yemen. *Egyptian journal of petroleum*, 27, 835–851.
- Albriki, K., Wang, F., Li, M., El Zaroug, R., & Ali, Z. (2022). Assessment of the thermal maturation, organofacies, and petroleum generation history of sirte shale formation in Sirt basin. *Libya. Journal of African Earth Sciences*, 196, 104710.
- Ali, M., Zhu, P., Huolin, M., Pan, H., Abbas, K., Ashraf, U., Ullah, J., Jiang, R., & Zhang, H. (2023). A novel machine learning approach for detecting outliers, rebuilding well logs, and enhancing reservoir characterization. *Natural Resources Research*, 32, 1047–1066.
- Aliakbardoust, E., Adabi, M. H., Kadkhodaie, A., Harris, N. B., & Chehraz, A. (2024). Integration of well logs and seismic attributes for prediction of thermal maturity and TOC content in the Kazhdumi Formation (central Persian Gulf basin). *Journal of Applied Geophysics*, 222, 105319. <https://doi.org/10.1016/j.jappgeo.2024.105319>.
- Aljarrah, O., Li, J., Heryudono, A., Huang, W., & Bi, J. (2022). Predicting part distortion field in additive manufacturing: a data-driven framework. *Journal of Intelligent Manufacturing*. <https://doi.org/10.1007/s10845-021-01902-z>.
- Allen, P. A., & Allen, J. R. (2013). *Basin analysis: Principles and application to petroleum play assessment*. John Wiley & Sons.
- AlSinan, S., Nivlet, P., Altowairqi, Y., & Poveda, I. L. (2020). Prediction of source rock maturity using semi supervised machine learning algorithms, EAGE 2020 annual conference & exhibition online. EAGE Publications BV.
- Amosu, A., Imsalem, M., & Sun, Y. (2021). Effective machine learning identification of TOC-rich zones in the eagle ford shale. *Journal of Applied Geophysics*, 188, 104311.
- Aryanto, A., Littke, R., Dörner, M., Erdmann, M., & Grohmann, S. (2022). Maturation and migration processes in intact source rock micro plugs induced by chemical and thermal treatment: A new approach combining Rock-Eval pyrolysis and organic petrography. *International Journal of Coal Geology*, 251, 103938.
- Barham, A., Ismail, M. S., Hermana, M., Padmanabhan, E., Baashar, Y., & Sabir, O. (2021). Predicting the maturity and organic richness using artificial neural networks (ANNs): A

## A Novel Hybrid Machine Learning Approach

- case study of Montney Formation, NE British Columbia, Canada. *Alexandria Engineering Journal*, 60, 3253–3264.
- Barth, A., Boniface, N., Kagya, M., Knobloch, A., Legler, C., Many, S., Mruma, A., Ngole, T., Stanek, K., & Stephan, T. (2016). The new minerogenic map of Tanzania—An integral part of the geological and mineral information system of the geological survey of Tanzania. Retrieved May 16, 2024, from [https://www.researchgate.net/publication/309152115\\_Download\\_of\\_The\\_Minerogenic\\_Map\\_of\\_Tanzania\\_and\\_Exploratory\\_Notes](https://www.researchgate.net/publication/309152115_Download_of_The_Minerogenic_Map_of_Tanzania_and_Exploratory_Notes).
- Bayatvarkeshi, M., Mohammadi, K., Kisi, O., & Fasihi, R. (2020). A new wavelet conjunction approach for estimation of relative humidity: Wavelet principal component analysis combined with ANN. *Neural Computing and Applications*, 32, 4989–5000.
- Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In I. Cohen, Y. Huang, J. Chen, & J. Benesty (Eds.), *Noise reduction in speech processing* (pp. 1–4). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-00296-0\\_5](https://doi.org/10.1007/978-3-642-00296-0_5).
- Brownfield, M. (2016). *Geologic assessment of undiscovered hydrocarbon resources of Sub-Saharan Africa*. US Geological Survey Digital Data Series.
- Burnham, A. K., & Sweeney, J. J. (1989). A chemical kinetic model of vitrinite maturation and reflectance. *Geochimica et Cosmochimica Acta*, 53, 2649–2657.
- Che Nordin, N. F., Mohd, N. S., Koting, S., Ismail, Z., Sherif, M., & El-Shafie, A. (2021). Groundwater quality forecasting modelling using artificial intelligence: A review. *Groundwater for Sustainable Development*, 14, 100643.
- Chen, Z., Dewing, K., Synnott, D. P., & Liu, X. (2019). Correcting Tmax suppression: A numerical model for removing adsorbed heavy oil and bitumen from upper ordovician source Rocks, Arctic Canada. *Energy & Fuels*, 33, 6234–6246.
- Cheshire, S., Craddock, P. R., Xu, G., Sauerer, B., Pomerantz, A. E., McCormick, D., & Abdallah, W. (2017). Assessing thermal maturity beyond the reaches of vitrinite reflectance and Rock-Eval pyrolysis: A case study from the Silurian Qusaiba formation. *International Journal of Coal Geology*, 180, 29–45.
- 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 Computer Science*, 7, e623.
- Chong, L., Singh, H., Creason, C. G., Seol, Y., & Myshakin, E. M. (2022). Application of machine learning to characterize gas hydrate reservoirs in Mackenzie delta (Canada) and on the Alaska north slope (USA). *Computational Geosciences*, 26, 1151–1165.
- Craddock, P. R., Bake, K. D., & Pomerantz, A. E. (2018). Chemical, molecular, and microstructural evolution of kerogen during thermal maturation: Case study from the woodford shale of Oklahoma. *Energy & Fuels*, 32, 4859–4872.
- Dai, C. (2023). A method of forecasting trade export volume based on back-propagation neural network. *Neural Computing and Applications*, 35, 8775–8784.
- Deaf, A. S., Omran, A. A., El-Arab, E. S. Z., & Maky, A. B. F. (2022). Integrated organic geochemical/petrographic and well logging analyses to evaluate the hydrocarbon source rock potential of the Middle Jurassic upper Khatatba Formation in Matruh Basin, northwestern Egypt. *Marine and Petroleum Geology*, 140, 105622.
- Dembicki, H. (2022). *Practical petroleum geochemistry for exploration and production*. Elsevier.
- Devi, S., Jagadev Alok, K., & Patnaik, S. (2015). Learning an artificial neural network using dynamic particle swarm optimization-backpropagation: Empirical evaluation and comparison. *Journal of Information and Communication Convergence Engineering*, 13, 123–131.
- Ehsan, M., & Gu, H. (2020). An integrated approach for the identification of lithofacies and clay mineralogy through Neuro-Fuzzy, cross plot, and statistical analyses, from well log data. *Journal of Earth System Science*, 129, 1–13.
- Ehsan, M., Latif, M. A. U., Ali, A., Radwan, A. E., Amer, M. A., & Abdelrahman, K. (2023). Geocellular Modeling of the Cambrian to Eocene Multi-Reservoirs, Upper Indus Basin, Pakistan. *Natural Resources Research*, 32, 2583–2607.
- Einvik-Heitmann, V. (2016). Sedimentology, stratigraphy, petrology and diagenesis of an Early Cretaceous drill core. Mandawa Basin, Coastal Tanzania. Oslo: University of Oslo. Master thesis, Geosciences. <https://www.duo.uio.no/handle/10852/52249>.
- Farouk, S., Lofty, N. M., Qteishat, A., Ahmad, F., Shehata, A. M., Al-Kahtany, K., & Hsu, C. S. (2023). Source and thermal maturity assessment of the Paleozoic-Mesozoic organic matter in the Risha gas field. *Jordan. Fuel*, 335, 126998.
- Feng, C., Feng, Z., Mao, R., Li, G., Zhong, Y., & Ling, K. (2023a). Prediction of vitrinite reflectance of shale oil reservoirs using nuclear magnetic resonance and conventional log data. *Fuel*, 339, 127422.
- Feng, D., Liu, C., Tian, J., Ran, Y., Awan, R. S., Zeng, X., Zhang, J., & Zang, Q. (2023b). Natural gas genesis, source and accumulation processes in northwestern Qaidam Basin, China, revealed by integrated 3D basin modeling and geochemical research. *Natural Resources Research*, 32, 391–412.
- Fossum, K. (2020). Jurassic-Cretaceous stratigraphic development of the Mandawa Basin, Tanzania: an integrated sedimentological and heavy mineral study of the early post-rift succession. Oslo: University of Oslo. Doctoral thesis, Geosciences. <https://www.duo.uio.no/handle/10852/75524>.
- Fossum, K., Morton, A. C., Dypvik, H., & Hudson, W. E. (2019). Integrated heavy mineral study of Jurassic to Paleogene sandstones in the Mandawa Basin, Tanzania: Sediment provenance and source-to-sink relations. *Journal of African Earth Sciences*, 150, 546–565.
- Gallo, C., & Capozzi, V. (2019). Feature selection with non linear PCA: A neural network approach. *Journal of Applied Mathematics and Physics*. <https://doi.org/10.4236/jamp.2019>.
- Gama, J., & Schwark, L. (2022). Lithofacies of early Jurassic successions derived from spectral gamma ray logging in the Mandawa Basin. *SE Tanzania. Arabian Journal of Geosciences*, 15, 1373.
- Gama, J., & Schwark, L. (2023). Total organic carbon variability of lower Jurassic successions in the Mandawa Basin. *SE Tanzania. Geoenery Science and Engineering*, 221, 111276.
- Godfray, G., & Seetharamaiah, J. (2019). Geochemical and well logs evaluation of the Triassic source rocks of the Mandawa basin, SE Tanzania: Implication on richness and hydrocarbon generation potential. *Journal of African Earth Sciences*, 153, 9–16.
- Gu, Y., Chen, C., Yang, Y., Song, Z., Chen, X., Jia, W., Lai, X., Li, H., Yin, L., & Huang, X. (2022). Geology, fluid inclusion, bitumen and isotope geochemistry of the organic-matter-rich Nanmushu lead-zinc deposit, Mayuan, the northern margin of the Yangtze platform. *China Arabian Journal of Geosciences*, 15, 221.
- Hackley, P. C., Jubb, A. M., Smith, P. L., McAleer, R. J., Valentine, B. J., Hatcherian, J. J., Botterell, P. J., & Birdwell, J. E. (2022). Evaluating aromatization of solid bitumen generated in the presence and absence of water: Implications for solid bitumen reflectance as a thermal proxy. *International Journal of Coal Geology*, 258, 104016.
- Hackley, P. C., & Lünsdorf, N. K. (2018). Application of raman spectroscopy as thermal maturity probe in shale petroleum systems: insights from natural and artificial maturation series. *Energy & Fuels*, 32, 11190–11202.

- Hantschel, T., & Kauerauf, A. I. (2009). *Fundamentals of basin and petroleum systems modeling*. Springer Science & Business Media.
- Hou, G. (2015). Late Cretaceous Sedimentation (Mavuji Group) in Mandawa Basin, Tanzania. Oslo: University of Oslo. Master thesis, Geosciences. <https://www.duo.uio.no/handle/10852/45504>.
- Hudson, W., (2011). The geological evolution of the petroleum prospective Mandawa Basin southern coastal Tanzania. Trinity College (Dublin, Ireland). Department of Geology.
- Hudson, W., & Nicholas, C. (2014). The pindiro group (Triassic to Early Jurassic Mandawa Basin, southern coastal Tanzania): Definition, palaeoenvironment, and stratigraphy. *Journal of African Earth Sciences*, 92, 55–67.
- Huijun, W., Guiping, Z., Liang, L., Wei, Z., Rong, Q., & Jun, L. (2020). TOC prediction model for muddy source rocks based on convolutional neural network (CNN): a case study of the Hangjinqi area of the Ordos Basin. *Journal of University of Chinese Academy of Sciences*, 37, 103.
- Hussain, W., Luo, M., Ali, M., Hussain, S. M., Ali, S., Hussain, S., Naz, A. F., & Hussain, S. (2023). Machine learning-a novel approach to predict the porosity curve using geophysical logs data: An example from the Lower Goru sand reservoir in the Southern Indus basin. *Pakistan Journal of Applied Geophysics*, 214, 105067.
- IEA, (2021). Global energy review. IEA: Paris; <https://www.iea.org/reports/global-energy-review-2021>, (accessed 04.03.23).
- İnan, S. (2023). Maturity determination of contaminated source rocks by pyrolysis and thermal oxidation methods: A review. In H. El Atfy & B. I. Ghassal (Eds.), *Advances in petroleum source rock characterizations: Integrated methods and case studies: A multidisciplinary source rock approach* (pp. 47–57). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-16396-8\\_3](https://doi.org/10.1007/978-3-031-16396-8_3).
- İnan, S., Henderson, S., & Qathami, S. (2017). Oxidation Tmax: A new thermal maturity indicator for hydrocarbon source rocks. *Organic Geochemistry*, 113, 254–261.
- Jahed Armaghani, D., Shoib, R. S. N. S. B. R., Faizi, K., & Rashid, A. S. A. (2017). Developing a hybrid PSO–ANN model for estimating the ultimate bearing capacity of rock-socketed piles. *Neural Computing and Applications*, 28, 391–405.
- Jubb, A. M., Birdwell, J. E., Hackley, P. C., Hatcherian, J. J., & Qu, J. (2020). Nanoscale molecular composition of solid bitumen from the eagle ford group across a natural thermal maturity gradient. *Energy & Fuels*, 34, 8167–8177.
- 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. *Journal of Rock Mechanics and Geotechnical Engineering*, 14, 1052–1063.
- Katz, B. J., & Lin, F. (2021). Consideration of the limitations of thermal maturity with respect to vitrinite reflectance, Tmax, and other proxies. *AAPG Bulletin*, 105, 695–720.
- Kedia, N. K., Kumar, A., & Singh, Y. (2023). Prediction of underground metro train-induced ground vibration using hybrid PSO-ANN approach. *Neural Computing and Applications*, 35, 8171–8195.
- Kibria, M. G., Das, S., Hu, Q.-H., Basu, A. R., Hu, W.-X., & Mandal, S. (2020). Thermal maturity evaluation using Raman spectroscopy for oil shale samples of USA: Comparisons with vitrinite reflectance and pyrolysis methods. *Petroleum Science*, 17, 567–581.
- Lal, A., & Datta, B. (2021). Application of the group method of data handling and variable importance analysis for prediction and modelling of saltwater intrusion processes in coastal aquifers. *Neural Computing and Applications*, 33, 4179–4190.
- Li, C., Liu, Z., Chen, C., Wang, Y., Liu, F., Xu, M., Yang, Y., Wang, B., & Chen, S. (2024). Predicting the thermal maturity of source rock from well logs and seismic data in basins with low-degree exploration. *Journal of Applied Geophysics*, 221, 105300.
- Li, M., Du, W., & Nian, F. (2014). An adaptive particle swarm optimization algorithm based on directed weighted complex network. *Mathematical Problems in Engineering*, 2014, 434972. <https://doi.org/10.1155/2014/434972>.
- Lohr, C. D., & Hackley, P. C. (2021). Relating Tmax and hydrogen index to vitrinite and solid bitumen reflectance in hydrous pyrolysis residues: Comparisons to natural thermal indices. *International Journal of Coal Geology*, 242, 103768.
- Lv, Q., Zhou, T., Zheng, R., Nakhaei-Kohani, R., Riazi, M., Hemmati-Sarapardeh, A., Li, J., & Wang, W. (2023). Application of group method of data handling and gene expression programming for predicting solubility of CO<sub>2</sub>-N<sub>2</sub> gas mixture in brine. *Fuel*, 332, 126025.
- Maganza, N.E., (2014). Petroleum system modelling of onshore Mandawa Basin-Southern, Tanzania. Institutt for geologi og bergteknikk.
- Malki, M.L., Rasouli, V., Mehana, M., Mellal, I., Saberi, M.R., Sennaoui, B., Chellal, H.A., (2023). The impact of thermal maturity on the organic-rich shales properties: A case study in Bakken, SPE/AAPG/SEG Unconventional Resources Technology Conference. URTEC, p. D031S054R003.
- Mazaheri, P., Rahnamayan, S., & Bidgoli, A. A. (2022). *Designing artificial neural network using particle swarm optimization: A survey*. IntechOpen: In Swarm Intelligence-Recent Advances and Current Applications.
- McCabe, R., (2021). Geochemistry & stratigraphy of the Mesozoic & Cenozoic sedimentary rocks encountered in the Mandawa Basin, South Eastern Tanzania. Trinity College Dublin. School of Natural Sciences. Discipline of Geology.
- McCabe, R., Nicholas, C. J., Fitches, B., Wray, D., & Pearce, T. (2023). Chemostratigraphic and mineralogical examination of the Kilwa Group claystones, coastal Tanzania: An alternative approach to refine the lithostratigraphy. *Journal of African Earth Sciences*, 197, 104746.
- 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*, 129232.
- MolaAbasi, H., Khajeh, A., & Jamshidi Chenari, R. (2021). Use of GMDH-type neural network to model the mechanical behavior of a cement-treated sand. *Neural Computing and Applications*, 33, 15305–15318.
- Mostaar, A., Sattari, M. R., Hosseini, S., & Deevband, M. R. (2019). Use of artificial neural networks and PCA to predict Results of infertility treatment in the ICSI method. *Journal of Biomedical Physics & Engineering*, 9, 679–686.
- Mshiu, E. E., Kiswaka, E. B., & Mohamed, B. (2022). Extensive salt deposition and remobilization influencing petroleum prospectivity of the Mandawa Basin: Remote sensing manifestation confirmed by seismic results. *Journal of Sedimentary Environments*, 7, 147–162.
- Mulashani, A. K., Shen, C., Asante-Okyere, S., Kerttu, P. N., & Abelly, E. N. (2021). Group method of data handling (GMDH) neural network for estimating total organic carbon (TOC) and hydrocarbon potential distribution (S1, S2) using well logs. *Natural Resources Research*, 30, 3605–3622.
- 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.
- Nelles, O. (2020). *Nonlinear system identification: From classical approaches to neural networks, fuzzy models, and gaussian processes*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-47439-3>.

## A Novel Hybrid Machine Learning Approach

- Nicholas, C. J., Pearson, P. N., Bown, P. R., Jones, T. D., Huber, B. T., Karega, A., Lees, J. A., McMillan, I. K., O'Halloran, A., Singano, J. M., & Wade, B. S. (2006). Stratigraphy and sedimentology of the Upper Cretaceous to Paleogene Kilwa Group, southern coastal Tanzania. *Journal of African Earth Sciences*, 45, 431–466.
- Onwubolu, G. C. (2008). Design of hybrid differential evolution and group method of data handling networks for modeling and prediction. *Information Sciences*, 178, 3616–3634.
- Osukuku, G. A., Osinowo, O. O., Sonibare, W. A., Makhanu, E. W., Rono, S., & Omar, A. (2023). Assessment of hydrocarbon generation potential and thermal maturity of the deep offshore Lamu Basin. *Kenya Energy Geoscience*, 4, 100133.
- Pang, Y., Guo, X., Shi, B., Zhang, X., Cai, L., Han, Z., Chang, X., & Xiao, G. (2020). Hydrocarbon generation evaluation, burial history, and thermal maturity of the lower triassic-silurian organic-rich sedimentary rocks in the central uplift of the South Yellow Sea basin, East Asia. *Energy & Fuels*, 34, 4565–4578.
- Peters, K.E., Cassa, M.R., (1994). Applied source rock geochemistry: Chapter 5: Part II. Essential elements.
- Peters, K. E. (1986). Guidelines for evaluating petroleum source rock using programmed pyrolysis. *AAPG Bulletin*, 70, 318–329.
- Petersen, H. I., Holland, B., & Olivarius, M. (2022). Source rock evaluation and fluid inclusion reconnaissance study of Carboniferous and Zechstein rocks in the northern margin of the Southern Permian basin, onshore Denmark. *International Journal of Coal Geology*, 255, 103985.
- Pourghasemi, H. R., Razavi-Termeh, S. V., Kariminejad, N., Hong, H., & Chen, W. (2020). An assessment of meta-heuristic approaches for flood assessment. *Journal of Hydrology*, 582, 124536.
- Pravin, P. S., Tan, J. Z. M., Yap, K. S., & Wu, Z. (2022). Hyperparameter optimization strategies for machine learning-based stochastic energy efficient scheduling in cyber-physical production systems. *Digital Chemical Engineering*, 4, 100047.
- Price, K., Storn, R., (1995). Differential Evolution—a simple and efficient adaptive scheme for global optimization over continuous space. Technical Report, International Computer Science Institute.
- Purcell, P. (2014). Oil and gas exploration in East Africa: A brief history. *AAPG Search and Discovery Article*, 30388, 14–17.
- Ramos, E. M., Borges, M. R., Giraldi, G. A., Schulze, B., & Bernardo, F. (2023). Prediction of permeability of porous media using optimized convolutional neural networks. *Computational Geosciences*, 27, 1–34.
- Roshani, M., Sattari, M. A., Muhammad Ali, P. J., Roshani, G. H., Nazemi, B., Corniani, E., & Nazemi, E. (2020). Application of GMDH neural network technique to improve measuring precision of a simplified photon attenuation based two-phase flowmeter. *Flow Measurement and Instrumentation*, 75, 101804.
- Sadeghtabaghi, Z., Talebkeikhah, M., & Rabbani, A. R. (2021). Prediction of vitrinite reflectance values using machine learning techniques: a new approach. *Journal of Petroleum Exploration and Production*, 11, 651–671.
- Safaei-Farouji, M., & Kadkhodaie, A. (2022a). Application of ensemble machine learning methods for kerogen type estimation from petrophysical well logs. *Journal of Petroleum Science and Engineering*, 208, 109455.
- Safaei-Farouji, M., & Kadkhodaie, A. (2022b). A comparative study of individual and hybrid machine learning methods for estimation of vitrinite reflectance (Ro) from petrophysical well logs. *Modeling Earth Systems and Environment*, 8, 4867–4881.
- Saporetti, C., Fonseca, D., Oliveira, L., Pereira, E., & Goliatt, L. (2022). Hybrid machine learning models for estimating total organic carbon from mineral constituents in core samples of shale gas fields. *Marine and Petroleum Geology*, 143, 105783.
- Shalaby, M. R., Malik, O. A., Lai, D., Jumat, N., & Islam, M. A. (2020). Thermal maturity and TOC prediction using machine learning techniques: case study from the cretaceous-paleocene source rock, Taranaki Basin, New Zealand. *Journal of Petroleum Exploration and Production Technology*, 10, 2175–2193.
- Singh, D. P., Wood, D. A., Singh, V., Hazra, B., & Singh, P. K. (2022). Impact of particle crush-size and weight on Rock-Eval S2, S4, and kinetics of shales. *Journal of Earth Science*, 33, 513–524.
- Sohail, J., Mehmood, S., Jahandad, S., Ehsan, M., Abdelrahman, K., Ali, A., Qadri, S. T., & Fnais, M. S. (2024). Geochemical Evaluation of Paleocene Source Rocks in the Kohat Sub-Basin, Pakistan. *ACS Omega*, 9, 14123–14141.
- Stokes, M. R., Jubb, A. M., Hackley, P. C., Birdwell, J. E., Barnhart, E. P., Scott, C. T., Shelton, J. L., Sanders, M. M., & Hatcherian, J. J. (2023). Evaluation of portable Raman spectroscopic analysis for source-rock thermal maturity assessments on bulk crushed rock. *International Journal of Coal Geology*, 279, 104374.
- Storn, R., & Price, K. (1995). *Differential evolution—a simple and efficient adaptive scheme for global optimization over continuous spaces: Technical report TR-95-012*. Berkeley, California: International Computer Science.
- Sun, Z., Xu, J., Espinoza, D. N., & Balhoff, M. T. (2021). Optimization of subsurface CO<sub>2</sub> injection based on neural network surrogate modeling. *Computational Geosciences*, 25, 1887–1898.
- Sweeney, J. J., & Burnham, A. K. (1990). Evaluation of a simple model of vitrinite reflectance based on chemical kinetics. *AAPG Bulletin*, 74, 1559–1570.
- Synnott, D. P., Dewing, K., Ardakani, O. H., & Obermajer, M. (2018). Correlation of zooclast reflectance with rock-eval tmax values within upper ordovician cape phillips formation, a potential petroleum source rock from the Canadian Arctic islands. *Fuel*, 227, 165–176.
- Tariq, Z., Mahmoud, M., Abouelresh, M., & Abdurhaem, A. (2020). Data-driven approaches to predict thermal maturity indices of organic matter using artificial neural networks. *ACS Omega*, 5, 26169–26181.
- ThanaAni, N. A. A., Mustapha, K. A., & Idris, M. (2022). Source rock pyrolysis and bulk kinetic modelling of Miocene sedimentary sequences in southeastern Sabah, Malaysia: The variability of thermal maturity to oil-gas producing kerogen. *Journal of Petroleum Science and Engineering*, 208, 109513.
- Thankan, S., Nandakumar, V., & Shivapriya, S. (2023). Hydrocarbon fluid inclusions and source rock parameters: A comparison from two dry wells in the western offshore. *India. Geoscience Frontiers*, 14, 101464.
- Tissot, B. P., & Welte, D. H. (2013). *Petroleum formation and occurrence*. Springer Science & Business Media.
- Titus, Z., Heaney, C., Jacquemyn, C., Salinas, P., Jackson, M. D., & Pain, C. (2022). Conditioning surface-based geological models to well data using artificial neural networks. *Computational Geosciences*, 26, 779–802.
- Wang, H., Wu, W., Chen, T., Dong, X., & Wang, G. (2019). An improved neural network for TOC, S1 and S2 estimation based on conventional well logs. *Journal of Petroleum Science and Engineering*, 176, 664–678.
- Wood, D. A. (2018). Kerogen conversion and thermal maturity modelling of petroleum generation: Integrated analysis applying relevant kerogen kinetics. *Marine and Petroleum Geology*, 89, 313–329.
- Wu, J., Luo, Q., Zhang, Y., Zhong, N., Goodarzi, F., Suchý, V., Li, M., Li, D., Wang, W., Tian, X., & Song, Z. (2023). The organic petrology of vitrinite-like maceral in the Lower Paleozoic shales: Implications for the thermal maturity

- evaluation. *International Journal of Coal Geology*, 274, 104282.
- Wu, Y., & Tong, G. (2022). The evaluation of agricultural enterprise's innovative borrowing capacity based on deep learning and BP neural network. *International Journal of System Assurance Engineering and Management*, 13, 1111–1123.
- Wygrala, B.P., (1989). Integrated study of an oil field in the southern Po basin, northern Italy.
- Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415, 295–316.
- Yang, S., & Horsfield, B. (2020). Critical review of the uncertainty of Tmax in revealing the thermal maturity of organic matter in sedimentary rocks. *International Journal of Coal Geology*, 225, 103500.
- 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. *Journal of Integrative Agriculture*, 20, 1474–1486.
- Zhang, M., & Li, Z. (2018). Thermal maturity of the Permian Lucaogou Formation organic-rich shale at the northern foot of Bogda Mountains, Junggar Basin (NW China): Effective assessments from organic geochemistry. *Fuel*, 211, 278–290. <https://doi.org/10.1016/j.fuel.2017.09.069>.
- Zhao, P., Ostadhassan, M., Shen, B., Liu, W., Abarghani, A., Liu, K., Luo, M., & Cai, J. (2019). Estimating thermal maturity of organic-rich shale from well logs: Case studies of two shale plays. *Fuel*, 235, 1195–1206.
- 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. *Journal of Petroleum Science and Engineering*, 214, 110517.
- Zhou, H., Deng, Z., Xia, Y., & Fu, M. (2016). A new sampling method in particle filter based on Pearson correlation coefficient. *Neurocomputing*, 216, 208–215.
- Zhou, Z., Tao, Y., Li, S., & Ding, W. (2013). Hydrocarbon potential in the key basins in the East Coast of Africa. *Petroleum Exploration and Development*, 40, 582–591.
- Zongying, Z., Ye, T., Shujun, L., & Wenlong, D. (2013). Hydrocarbon potential in the key basins in the East Coast of Africa. *Petroleum exploration and development*, 40, 582–591.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.