

**PREDICTION OF WATER SATURATION FROM
WELL LOGS USING ARTIFICIAL
INTELLIGENCE TECHNIQUES**

BY

Mohamed Farouk Mohamed Gad

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

GEOPHYSICS

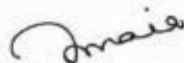
January 2023

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

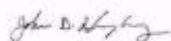
DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **Mohamed Gad** under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN GEOPHYSICS**.



Dr. Umair Bin Waheed
(Advisor)



Dr. John D. Humphrey

Department Chairman



Dr. Pantelis Soupios
(Member)



Dr. Suliman Al-Homidan
Dean of Graduate Studies



Dr. Salaheldin M. Elktatny
(Member)

Date



© Mohamed Gad

2023

Dedication

To Mr. Magid Talhah, my parents, my wife, my kids and my parents in law |

ACKNOWLEDGMENTS

First, This researcher expresses the most profound gratitude to his adviser, Dr. Umair bin Waheed, for his dedicated mentorship throughout his study. His expertise was invaluable in understanding the problem tackled in the study. Special thanks to Prof. Pantelis Soupios for his help and time. His valuable guidance, amiable attitude and thoughtful suggestions are among many things that made his master's journey a much better experience. He was there when the researcher needed him. He always offers his help with the most welcoming attitude. Third, he is grateful to Dr. Salaheldin M. Elktatny for his kind support and cooperation. Finally, he appreciated Ahmed Abdulhamid's support, constructive feedback, and kindness. The discussions we had, helped him develop the research question and methodology.

Most importantly, The researcher would like to thank his business partner, Mr. Majed Talhah, for his endless support and taking over duties during his weekly travel to KFUPM to attend the classes. The researcher wants to thank his parents; without them, He has never achieved anything. He wants to thank his wife for her constant support through his learning journey and her significant sacrifices to whom he will remain indebted for the rest of his life. Special thanks to the researcher's sons, Omar and Anas, for their lovely and emotional attitude toward him and for bearing with his busy life. The author wants to thank his parents-in-law for their tremendous support during his education journey. Also, He would like to thank his brothers and sisters for their emotional support during that time. The researcher wants to thank all the friends He knew here at KFUPM from around the world for our shared experiences. Special thanks to the colleague at TDS company for the good

times and their emotional support, especially during the Covid pandemic. Last but not least, thanks, KFUPM, represented by Geosciences Department, for allowing him to have this incredible journey to study Geophysics. The researcher learned a lot as a graduate student in this vital field of science. |

[TABLE OF CONTENTS]

ACKNOWLEDGMENTS.....	V
TABLE OF CONTENTS	VII
LIST OF TABLES	VIII
LIST OF FIGURES	X
ABSTRACT	XI
ملخص الرسالة	XII
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 MACHINE LEARNING AND ARTIFICIAL NEURAL NETWORKS	7
2.1 Machine Learning.....	7
2.2 Artificial Neural Network (ANN)	8
2.2.1 The Neuron: the building block.....	8
CHAPTER 3: THE STUDY AREA (PRINOS BASIN)	15
3.1 Field Background.....	15
CHAPTER 4: METHODOLOGY	19
4.1 Data preparation.....	19
4.2 ANN model learning	19
4.3 Extracting equation from the optimized ANN model.....	22
CHAPTER 5: CHAPTER 5 RESULTS AND DISCUSSION	25
5.1 Results and Discussion	25

5.1.1 Learning the artificial neural networks model	25
5.1.2 Training the artificial neural networks model	27
5.1.3 Testing the artificial neural networks model	28
5.2 Conclusions	29
REFERENCES	30
VITAE	36

LIST OF TABLES

- Table 1-** **The optimized parameters for water saturation prediction.**
- Table 2-** **The statistical features of the 252 datasets used to train the ANN for water
 saturation estimation.**
- Table 3-** **The extracted weights and biases for water saturation estimation.**

LIST OF FIGURES

Figure 1: Neuron is building block of ANN.....	9
Figure 2: Input vales and Synapse	10
Figure 3: Standardized the independent variables of ANN	11
Figure 4: ANN output values examples.....	12
Figure 5: ANN structure and Output values	12
Figure 6: input and output single observation	13
Figure 7: ANN model wights	14
Figure 8: Activation Function	14
Figure 9: The Prinos basin location.....	15
Figure 10: Prinos basin stratigraphic column. (Proedrou and Sidiropoulos (2004)).	17
Figure 11: Schematic of the optimized ANN model for water saturation prediction.	20
Figure 12: Flowchart of building ANN model.....	21
Figure 13: Training data	26
Figure 14: testing data	26
Figure 15: Prediction of water saturation by the ANN model (training data)	27
Figure 16: Testing data to predict the water saturation by ANN model	28

ABSTRACT

Full Name : Mohammed Farouk Mohamed Gad

Thesis Title : Prediction of water saturation from well logs using artificial intelligence techniques

Major Field : Geophysics

Date of Degree : [Dec. 2022]

Water saturation is a parameter that helps evaluate the volume of hydrocarbon in reservoirs. Improper water saturation calculation leads to remarkable errors in the subsequent decisions related to the development and production from the reservoir. Several methods are available to estimate the water saturation in clean and shaly formations, but the most commonly used methods in the petroleum industry are those based on petrophysical models, such as Archie's equation, Waxman-Smiths, Simandoux, Indonesia and dual-water models. Unfortunately, these models have limitations, and their input parameters are often not readily available. Consequently, it leads to either underestimated or overestimated fluid saturations. Furthermore, the main impediment remains in these approaches due to their reliance on experimental analysis, which is costly and time-consuming.

In this research, machine learning techniques including artificial neural networks (ANN), adaptive neuro-fuzzy interference system (ANFIS), function neural networks (FN), support vector machine (SVM), and random forests (RF) will be applied to predict the water saturation from well-logs of resistivity, gamma ray (GR), bulk formation density, sonic porosity and neutron porosity, in addition to the core data collected from five wells in Greece. These data will be used since it is usually affect/determine the water saturation. A portion of the data will be used for training the ML models, while others will be used for testing and validation of the trained models.

ملخص الرسالة

الاسم الكامل: محمد فاروق محمد جاد

عنوان الرسالة: التنبؤ بالتشبع المائي من خلال تسجيلات سبر غور الآبار باستخدام تقنيات الذكاء

الاصطناعي

التخصص: جيوفيزياء

الدرجة العلمية: ماجستير

إن حساب المحتوى المائي بالخرانات الجوفية البترولية يلعب دوراً حيوياً في تقييم حجم الهيدروكربونات في الخزانات التحت سطحية، يؤدي حساب المحتوى المائي غير السليم إلى أخطاء ملحوظة في القرارات اللاحقة المتعلقة بالتطوير والإنتاج من الخزانات التحت سطحية. هناك العديد من الطرق لتقدير نسبة المحتوى المائي في الخزانات التحت سطحية بأنواعها المختلفة، ولكن الطرق الأكثر استخداماً في مجال النفط هي تلك القائمة على النماذج البتروفيزيائية مثل معادلة Archie's equation, Waxman- (Smith, Simandoux, Indonesia, and dual-water models). ولكن هناك بعض القيود والتحديات عند استخدام هذه المعادلات حيث تحتاج إلى تدقيق في حساب الثوابت الموجودة بها والتي تختلف من منطقة لأخرى وأحياناً لا تكون متاحة بسهولة وبالتالي قد يؤدي ذلك إلى الخطأ في حساب نسبة المحتوى المائي إما بالزيادة أو النقصان عما هي عليه. علاوة على ذلك، لا يزال العائق الرئيسي في هذه النهج بسبب اعتمادها على التجارب العملية والتي عادة تكون مكلفة وتستغرق وقتاً طويلاً. مع وجود طفرة ملحوظة في مجال الذكاء الاصطناعي وتقنيات تعلم الآلة ظهرت طرق وأنماط جديدة قد تلعب دوراً رئيسياً في التغلب على التحديات القائمة الخاصة بالطرق المعتادة لحساب المحتوى المائي بالخرانات التحت سطحية وتظهر هذه الدراسة مدى قدرة تقنية التعلم الآلي على تقدير المحتوى المائي من سجلات الآبار التقليدية بجودة ودقة عالية مقارنة بنظيراتها المعتادة. وتناقش الطرق القائمة على الشبكة العصبية

القائمة على الذكاء الاصطناعي وكيفية التنبؤ بالمحتوى المائي من المعلومات المتاحة من تسجيلات الابار
دون اللجوء للتجارب العملية واخذ عينات صخرية من أعماق الخزانات التحت سطحية.

CHAPTER 1

[INTRODUCTION]

Most of the reservoirs consist of a minimum of two different fluid phases, some reservoirs have all three phases of gas, oil, and water (Dandekar, 2006). Therefore, water saturation (S_w) is essential for determining hydrocarbon saturation. Kamel and Mabrouk (2002) assumed that all void spaces in the reservoir consist of water and hydrocarbon; therefore: $S_h=1-S_w$. Furthermore, water saturation is critical for formation evaluation, reserve estimation, and future field planning. The static and dynamic models of the reservoir use it to estimate the original oil/gas in place (OOIP, OGIP) and form the foundation for the future production forecasts and the determination of the economic viability of the discovered reservoir.

Petrophysicists always pay attention to water saturation determination. Because of that, various methods have been developed through the past decades and are still ongoing. Therefore, two main categories of methods are used: (1) direct and (2) indirect.

The direct methods include laboratory analysis of rock samples, such as the Retort method and Dean–Stark and Soxhlet extraction. These methods consider that the rock sample has representative fluid saturation, which is impossible except for costly uncommon sampling methods such as sponge core barrel or pressure core barrel. In addition, they are exhaustive, time-consuming, and provide discrete data points.

Moreover, Dean-Stark extraction is a destructive method which measures water and oil saturation by distillation extraction when the water in the core is vaporized by boiling

solvent, then condensed and collected in a calibrated trap. Furthermore, considering the importance of the samples, they cannot be used in other experiments. Other laboratory approaches, such as using rock centrifuge, displacing fluid, and mercury-air capillary pressure curve for a water-air system, are expensive, time-consuming, and non-destructive.

Determination of water saturation is complex, it started in 1942 by incorporating some well-logs in clean sandstones. Then, many scientists proposed various equations to validate this procedure in shaly sands and carbonates. Later, many scientists proposed different equations to validate this procedure in shaly sands and carbonates. The main challenge in this process is that dependency on core analysis in labs takes time and is costly and for the accuracy, lab. measurements are not usually available for all existing wells. In addition, it does not provide a continuous recording through the well, but it just provides a discrete measurement at specific depths. To tackle the dependency of water saturation estimation on core analysis, researchers suggested using well-logs log data which are commonly acquired in most of the wells; and performing formation evaluation to calculate the water saturation as continuous values with depths. The relationship between water saturation and formation resistivity, and porosity logs as shown in Archie's equation, Eq. (1):

$$S_w = \left(\frac{aR_w}{\phi^m R_t} \right)^{\frac{1}{n}} \quad (1)$$

where; S_w is the water saturation, R_t is the true formation resistivity, ϕ^m is the formation porosity that can be estimated from NPHI, density, NMR, or acoustic logs, R_w is the formation water resistivity that can be calculated from the self-potential log (SP), a , m , and

n are constants (a: tortuosity constant, m: cementation exponent, n: saturation exponent) that are based on the pore geometry, cementation, and rock type.

Archie's equation is restricted to the clean formation, which is not the case in most hydrocarbon reservoirs since the existence of clay causes some error in resistivity measurements because clay increase conductivity. Moreover, it is not common in oil and gas fields to have SP logs which in turn which affect R_w estimation. In addition, If the rock matrix is unknown, then the low accuracy of porosity estimation will have occurred. Furthermore, m and n cannot be measured. Consequently, the estimation of S_w from Archie's equation needs to be calibrated and corrected.

Other techniques such as the Simandoux, Eq. (3), Waxman–Smits, Indonesia, Eq. (2), and dual-water models were proposed to modify Archie's equation by considering the shale factor into Archie's equation.

$$S_w = \left(\frac{\sqrt{1/R_t}}{\left(\frac{V_{sh}^{1-0.5V_{sh}}}{\sqrt{R_{sh}}} \right) + \sqrt{\frac{\phi^m}{aR_w}}} \right)^{2/n} \quad (2)$$

$$S_w = \frac{aR_w}{2\phi^m} \left(\sqrt{\left(\frac{V_{sh}}{R_{sh}} \right)^2 + \frac{4\phi^m}{aR_w R_t}} - \frac{V_{sh}}{R_{sh}} \right) \quad (3)$$

where V_{sh} is the shale volume, R_{sh} is shale resistivity, ϕ is the porosity, and R_t is the true formation resistivity. Also, the existence of non-conductive kerogen underestimates the water saturation.

Well logging may have some flaws, like it can be expensive. Thus we may skip running specific tools in a well. Because of the limitation of the logging tools' size, the resistivity log is unavailable. Also, Technically, the resistivity log data might be unreliable in view of unfavorable geology.

Due to the limitation of calculation a , m , and n and well logs data., Thus, we cannot count on the data in such cases because it defines only the initial state and cannot use for water saturation forecasts. Nevertheless, numerical model prediction might be a promising tool for optimising expenses and filling the gaps in well-log data.

The oil and gas industry faces several challenges and issues in data processing and calculations. Machine learning techniques can play a vital role in providing a proper solution for such challenges. Various machine learning methods such as artificial neural networks (ANN), support vector machine (SVM), and random forests (RF) can forecast the water saturation without any additional cost.

ANN is an artificial computational system configured to simulate organizational principles of the nervous system functions. Including the ability to learn automatically from the given training patterns, ANN can solve a mapping problem by finding a close approximation of the correlation between input and output data. Such a feature distinguishes it from other typical expert systems. ANN computing system is composed of artificial neurons that play the role of fundamental computing units and resemble the parallel process of a biological brain to find the answer.

SVM is a supervised machine learning algorithm dealing with classification and regression challenges. SVM can deal with complex problems by transforming the input dataset into a

higher-degree dimensional (n-dimensional) feature space to perform the optimal hyper-plane. In addition, SVM provides a proper way for estimating different petrophysical and geomechanical properties of rocks.

RF is a supervised machine learning algorithm used widely in Classification and Regression problems. It uses ensemble learning, a technique that combines many decision trees to solve complex problems. RF has several advantages, such as reducing the datasets' overfitting and increasing precision. In addition, it provides an effective way of handling missing data and can produce a proper prediction without hyper-parameter tuning.

Special Core Analysis (SCAL) accurately estimates water saturation at specific depths, but it requires time and is highly expensive. Unlike models developed based on conventional well logs to indirectly provide a continuous profile of the water saturation, SCAL provides discrete values of the water saturation. However, there is still a challenge in estimating water saturation due to the uncertainty of the constant calculations. This research proposes the use of different ML techniques such as ANN, ANFIS, FN, SVM, and RF to estimate the water saturation from conventional well logs.

Ahmed Farid Ibrahim et al. (2022) applied different machine learning techniques such as FN, SVM, and RF on 727 data points to calculate the water saturation in tight gas sandstone formation and they got accurate prediction comparing to other conventional equations.

Ahmed Farid Ibrahim et al. (2022) comparing between ANN and ANFIS with the conventional equations such as Archie, Simandoux, Indonesia for water saturation

prediction. They got results from AI methods with very low error comparing to normal methods. In addition, ANFIS showed more capability for the accurate prediction.

In chapter 2, I first give a brief background on machine learning with special emphasis on artificial neural networks, which I will use in this study. In chapter 3, I give a detailed information about the study area and reservoir. Afterwards, I explain the methodology used to train and test the network (Chapter 4). Lastly, I show main results of the study and discuss their implications (Chapter 5).

Chapter 2 Machine learning and Artificial Neural Networks

2.1 Machine Learning

Machine learning is fast emerging as a potential disruptive tool to tackle longstanding research problems across the sciences. This is mainly driven by its ability to find complex patterns in large datasets

The learning process is crucial in generalizing the problem by acting on its historical experience. Considering the training datasets as experience aids in achieving accurate results on new and unseen tasks. The learner uses The training datasets to build a general model about an existing problem domain. This process enables the model to generate largely accurate predictions in new cases.

Traditional computer programs are very good at doing massive computations fast and performing specific operations with specific instructions efficiently. However, if faced with a complex problem, they require a large set of specific rules to deal with it. On the other hand, machine learning algorithms learn fewer guidelines that help them learn how to get the desired outputs from the original data.

Machine learning (ML) algorithms have two branches: unsupervised and supervised. Unsupervised ML algorithms aim to discover the implicit relations between data points without needing training examples. Unsupervised algorithms have two main branches: clustering and association—the clustering algorithms sort data into separate categories based on the input data. For example, they are clustering a group of customers based on the amount of their spending and the type of goods they buy. Association algorithms attempt to find association rules between different data inputs. An example of association is cancer patients and the possession of specific genes.

Supervised ML algorithms, on the other hand, attempt to approximate a particular output (Y) from given inputs (X). They require a dataset of known inputs and outputs to train on and predict future outcomes.

Supervised ML algorithms, on the other hand, attempt to approximate a certain output (Y) from given inputs (X). They require a dataset of known inputs and outputs to train on, and then predict future outcomes. The process, in general, has the following steps: The algorithm is introduced to a dataset of inputs (X) and their corresponding outputs (Y), called labels. The algorithm trains on the dataset and

There are many supervised ML algorithms, such as linear regression, logistic regression, support vector machine, and others. However, one of the most potent supervised algorithms is the artificial neural network (ANN). The following section will discuss ANN's fundamental concepts as a prelude to discussing the methodology and results.

2.2 Artificial Neural Network (ANN)

The artificial neural network is a supervised machine learning algorithm that attempts to emulate the behaviour of biological neurons. Its edge is its ability to find complex non-linear relations between input variables and to optimize itself to output the correct values in the light of a labelled training dataset. The following subsections will explain the building block, the neuron, then the ANN, how it introduces non-linearity to the data and how it optimizes itself.

2.2.1 The Neuron: the building block

The building unit of a neural network is the neuron. The neuron, as shown in figure 1, is the simplest neural network. Its name is inspired by the nervous system cell, which is the

human brain's building block. The neuron, also sometimes called the node. The neuron gets some input signals and it has an output signal.

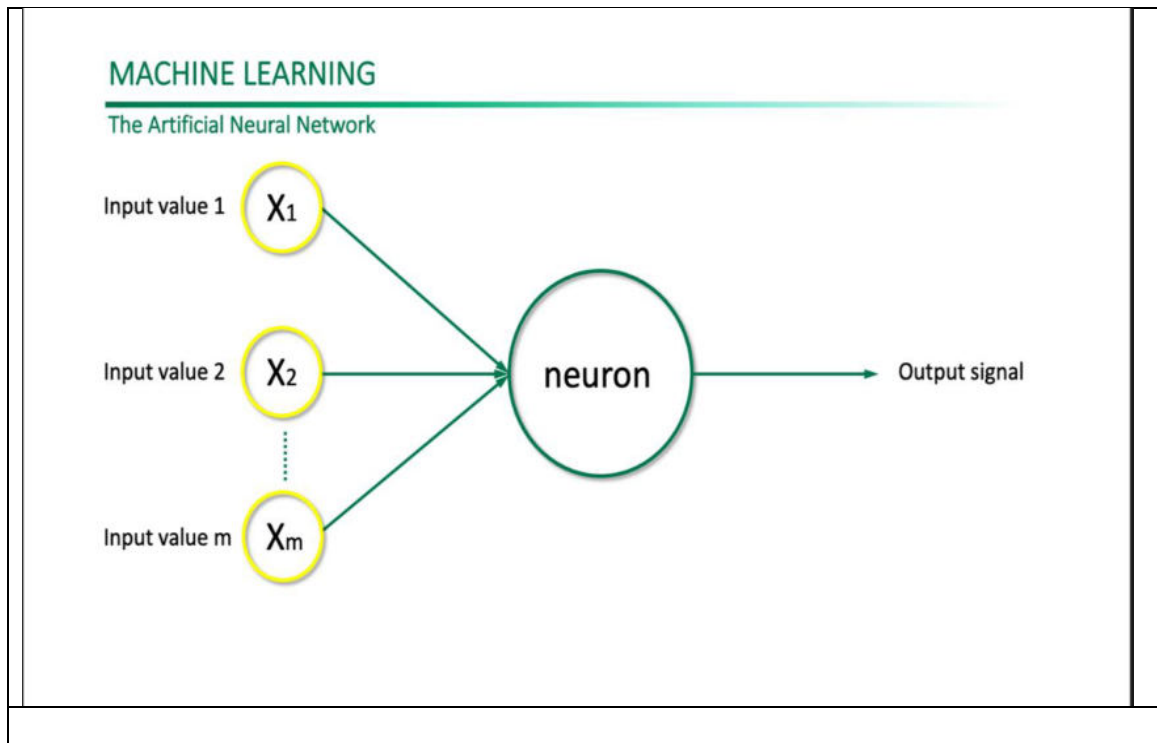


Figure 1: Neuron is building block of ANN

So the input values, as shown in figure 2, the signal is passed through synapses to the neuron, and then the neuron has an output value that passes it further down the chain.

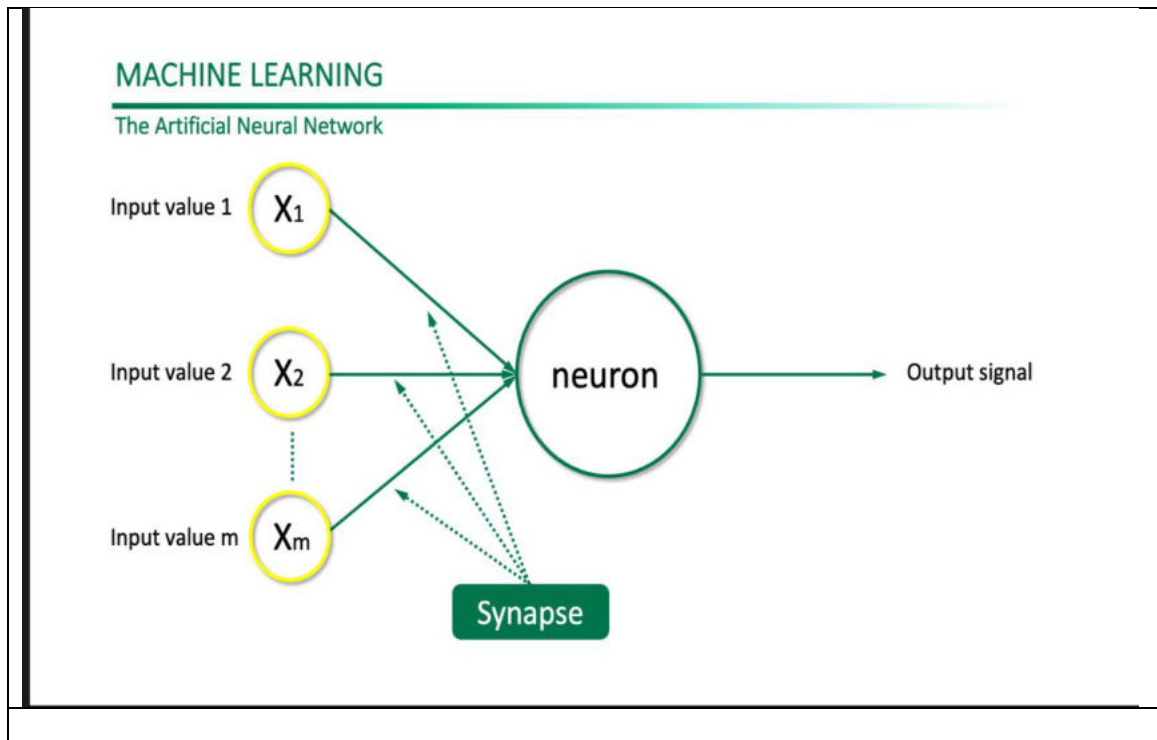


Figure 2: Input vales and Synapse

Well, we have these inputs, which are, in fact, independent variables. These independent variables are all for one single observation. It is just like just row in the database, one observation. So we take all of the independent variables, and the other thing we need to know about them is that we need to standardize them. So we need to either standardize them or ensure they have a mean of zero and a variance one. So sometimes, it needs not to be standardized; it is required to normalize them.

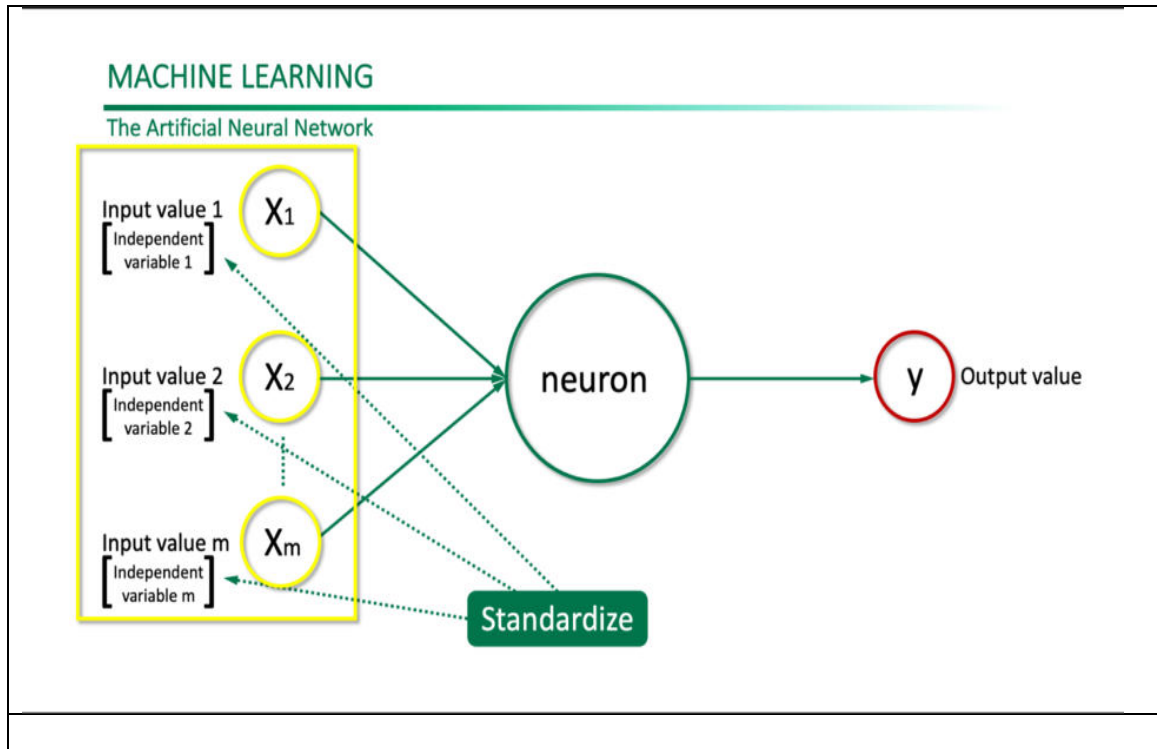


Figure 3: Standardized the independent variables of ANN

Output value can be continuous. It can be binary (a person will exit or stay) or a categorical variable. Furthermore, if it is a categorical variable, the important thing to remember here is that, in that case, the output value will not be just one. There will be several output values because these will be dummy variables representing the categories. But let us go back to a simple case of one output value.

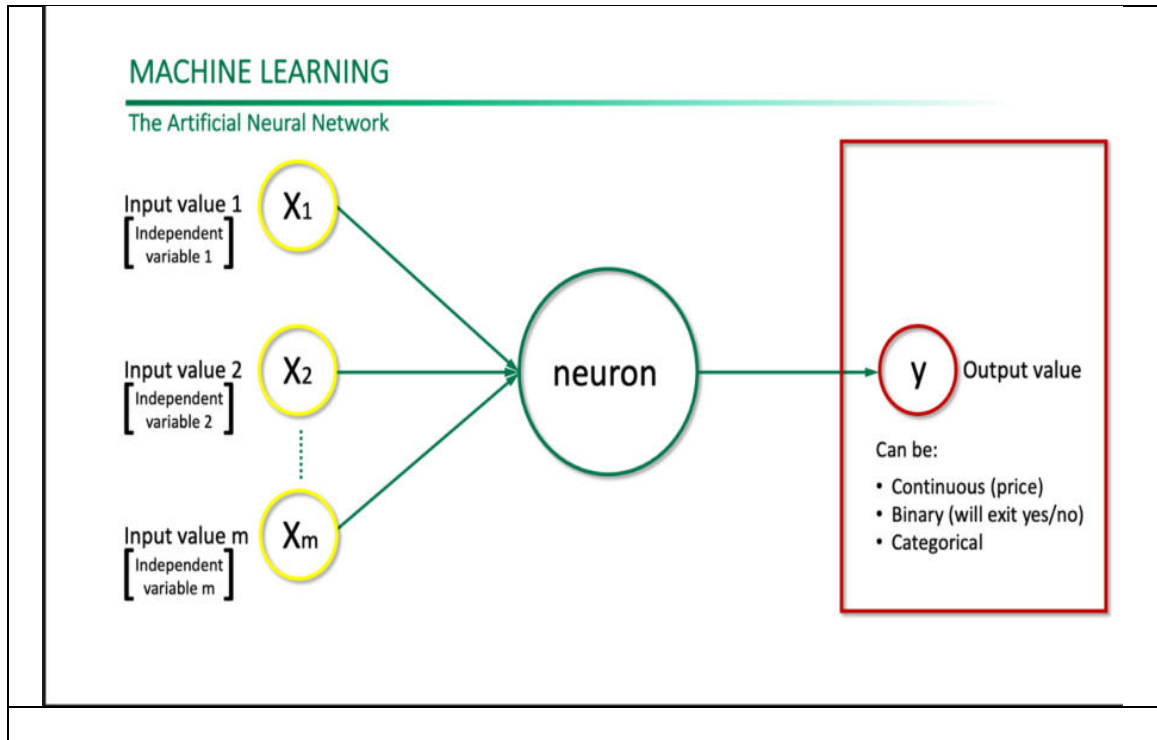


Figure 4: ANN output values examples

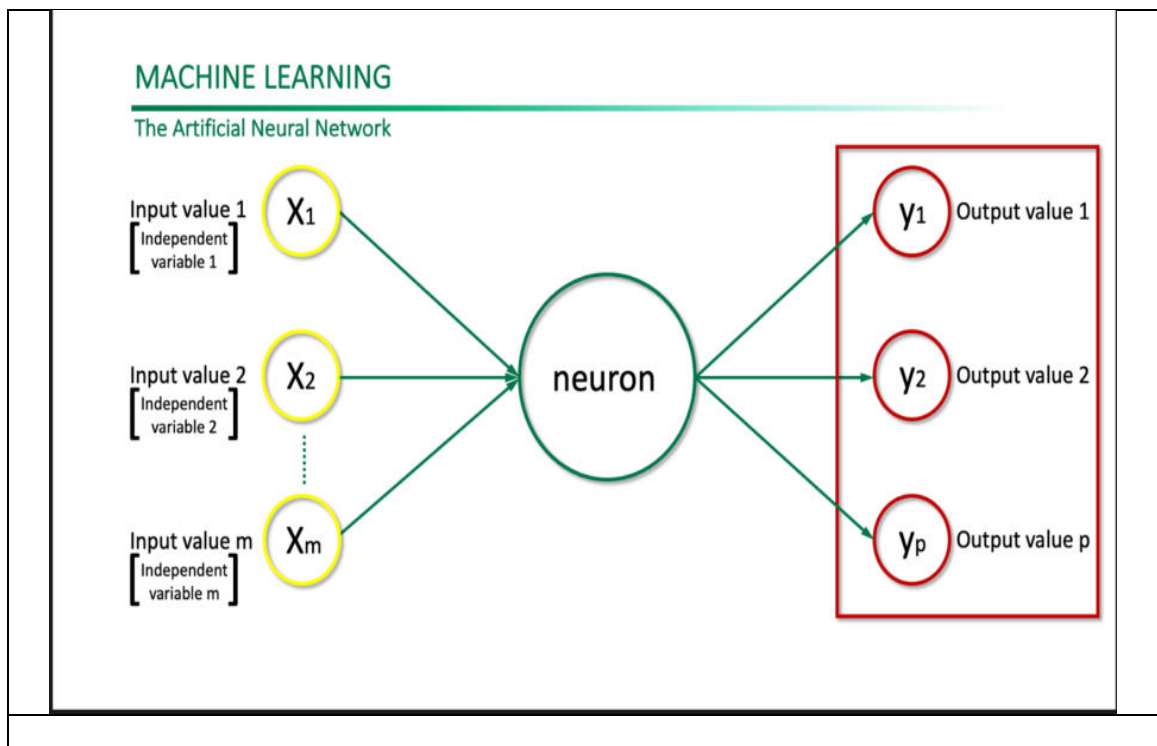


Figure 5: ANN structure and Output values

To reiterate this point, in figure 6, On the left, there is a single observation. Hence, there is also a single observation on the right as well. And that is the same observation.

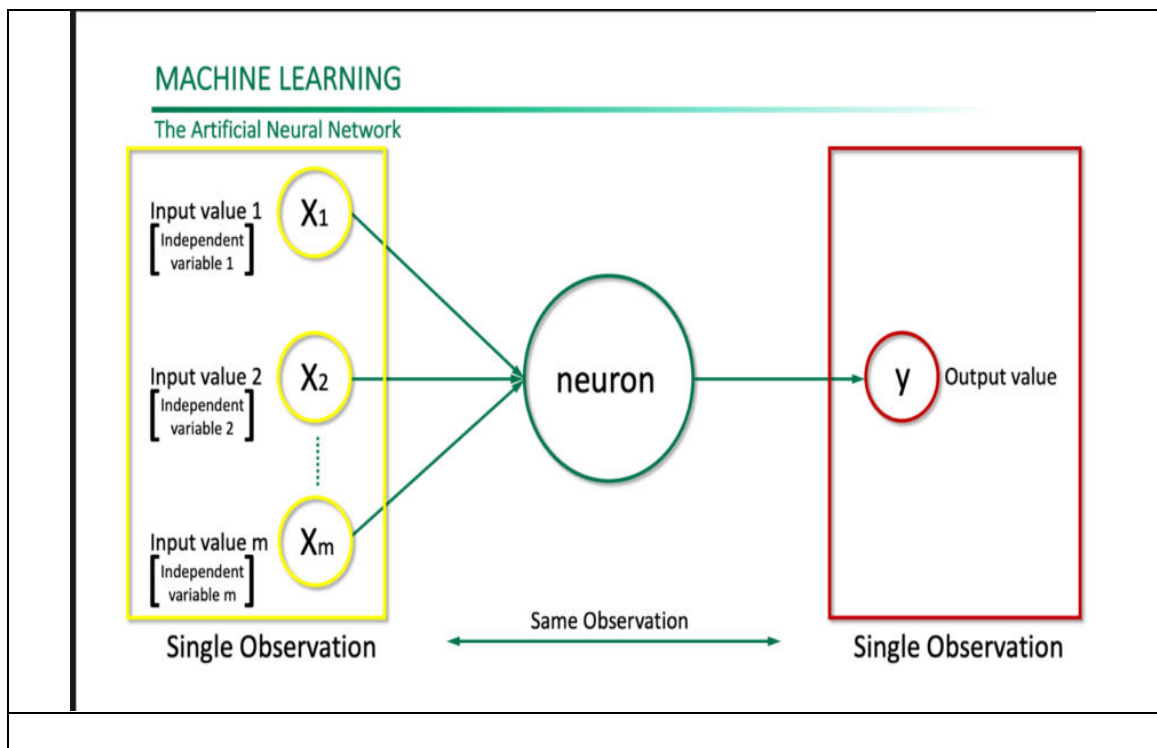


Figure 6: input and output single observation

So next thing we want to talk about here is the Synopsise and they all actually get assigned weights. weights are crucial to artificial neural network networks functioning because weights are how neural networks learn.

By adjusting the weights, the neural network decides in every single case what single is important, what signal is not important to certain neuron, what signal gets passed along and what signal doesn't get passed along.

Inside the neuron, So signals go into the neuron and what happens in the neuron. And the first step we multiply each input value by its weight then sum all of these values

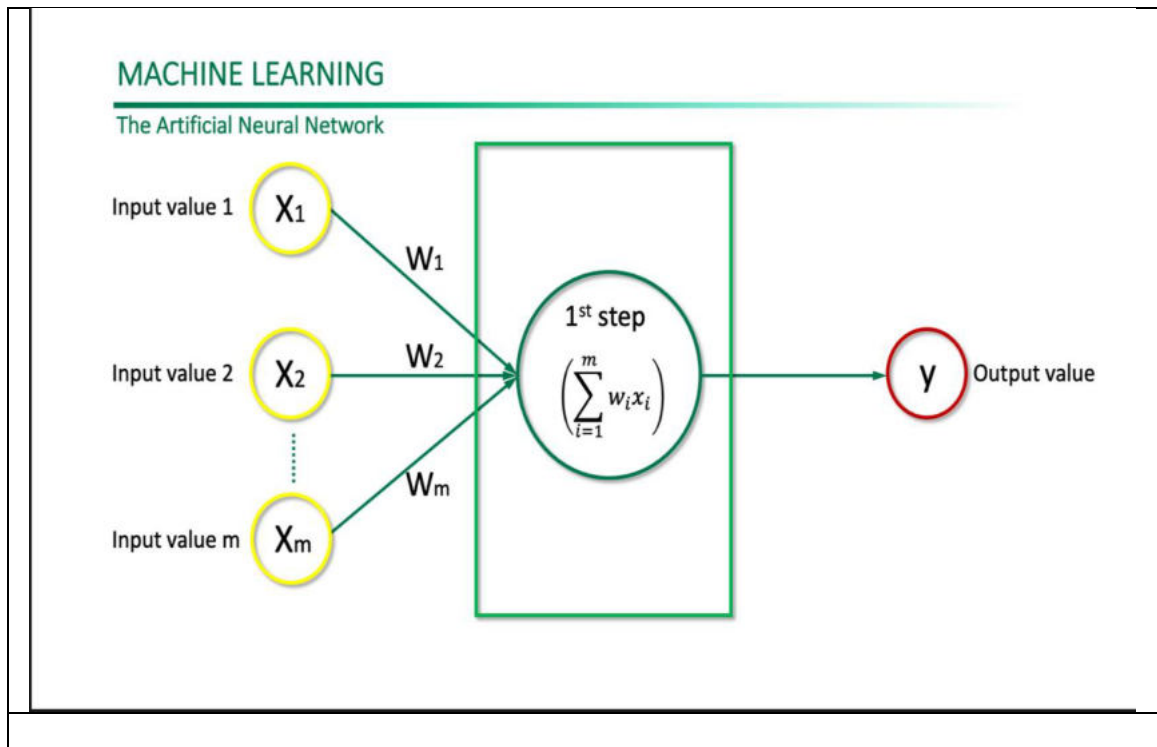


Figure 7: ANN model wights

As shown in figure 8, The second step is to multiply it by activation function.

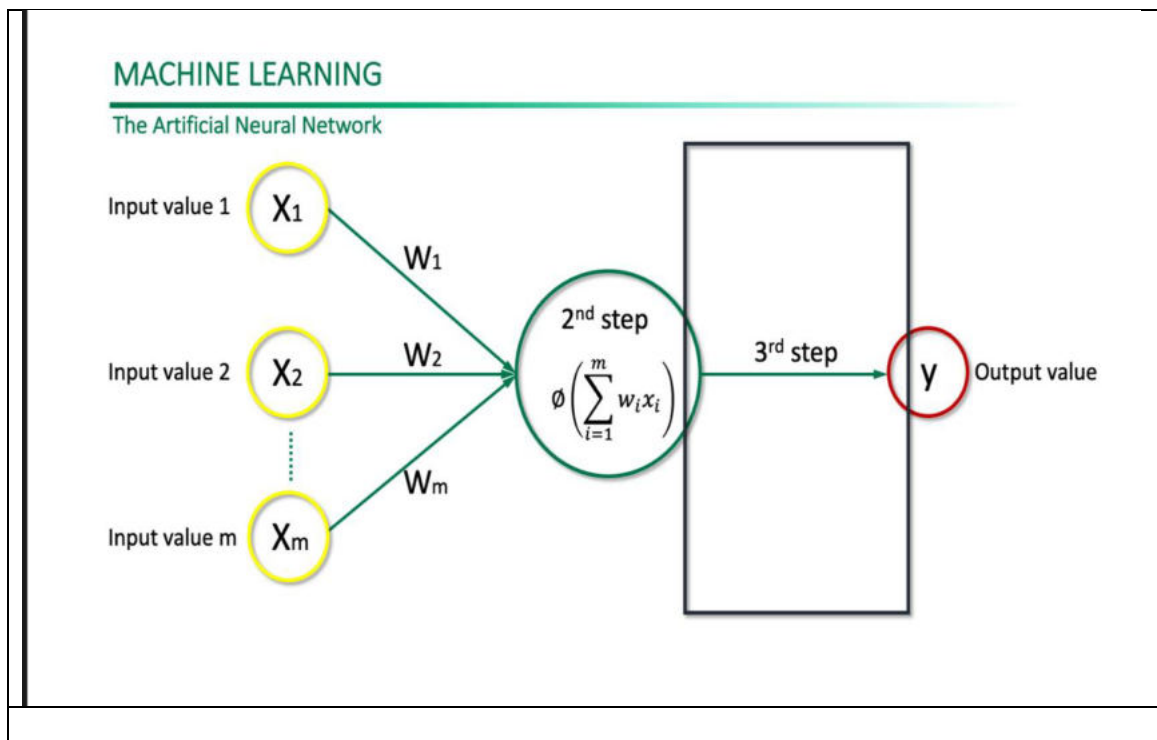


Figure 8: Activation Function

Chapter 3: The Study Area (Prinos Basin)

3.1 Field Background

Greece has only The Prinos basin as geological area which produce hydrocarbon. The Epsilon oil field exists between the Greek mainland and the island of Thassos, about 11 km South-SE of the city of Kavalaat, as shown in figure 9. This field is the latest one brought to production.

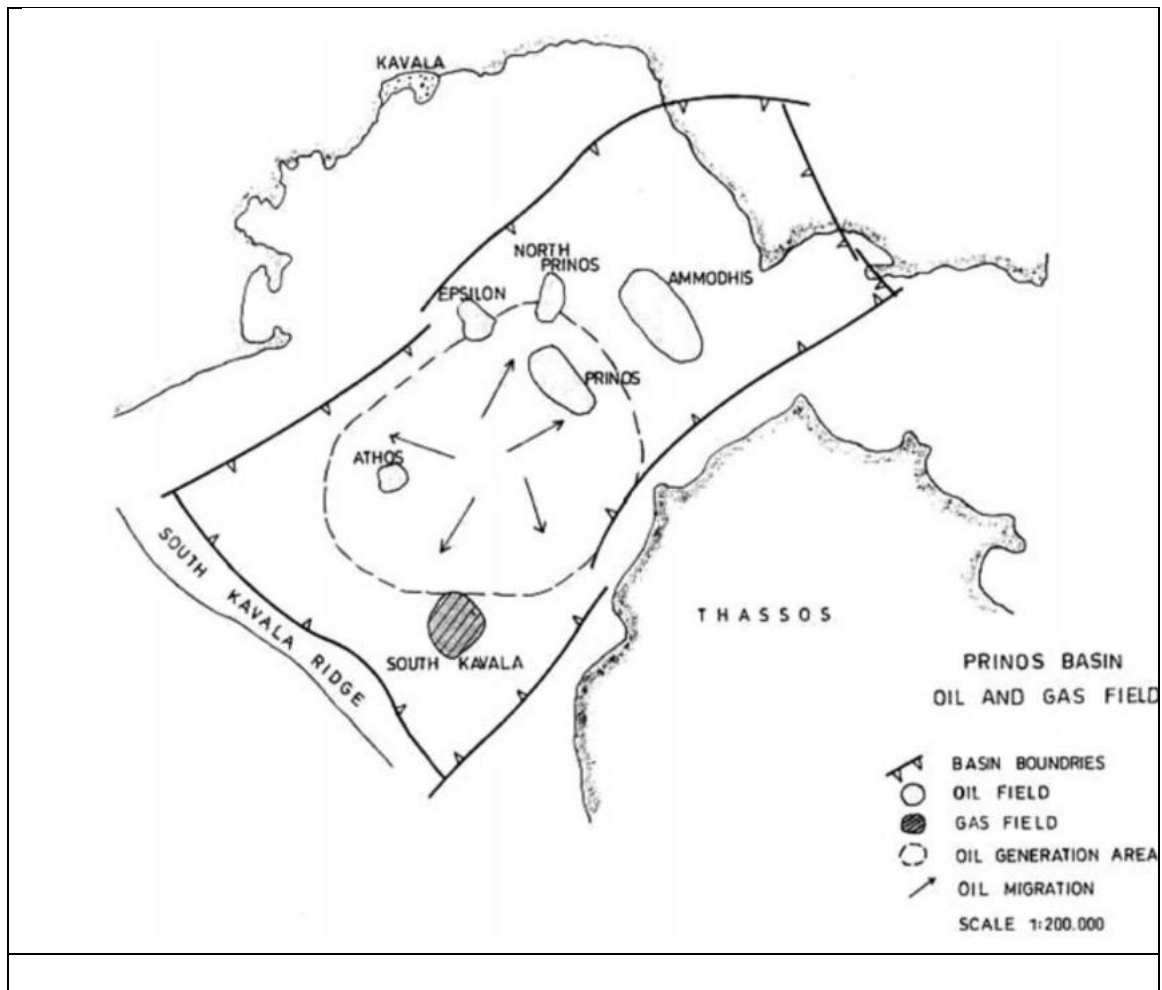


Figure 9: The Prinos basin location.

There were five drilling attempts 1976, 1994, 1996, 2004 and 2009, kept the field in production till today. The in first productive well was drilled in 1971, 20 km east of Thassos island but discovered very low gravity oil. In 1972-1973 the next two wells were drilled

west of the island and led to the discovery of South Kavala gas field in 52 m of water depth.

The fourth drilling attempt, was at the end of 1973, at the centre of Prinos basin and led to the discovery of Prinos oil field in a water depth of approximately 30 m.

The stratigraphy model of the taphrogenetic basin has been presented by Pollak (1979) and Proedrou (1979) (figure 10). According to this, the generalized stratigraphic column is divided into three series:

- The Pre-Evaporitic Series.
- The Evaporitic Series.
- The Post-Evaporitic Series.

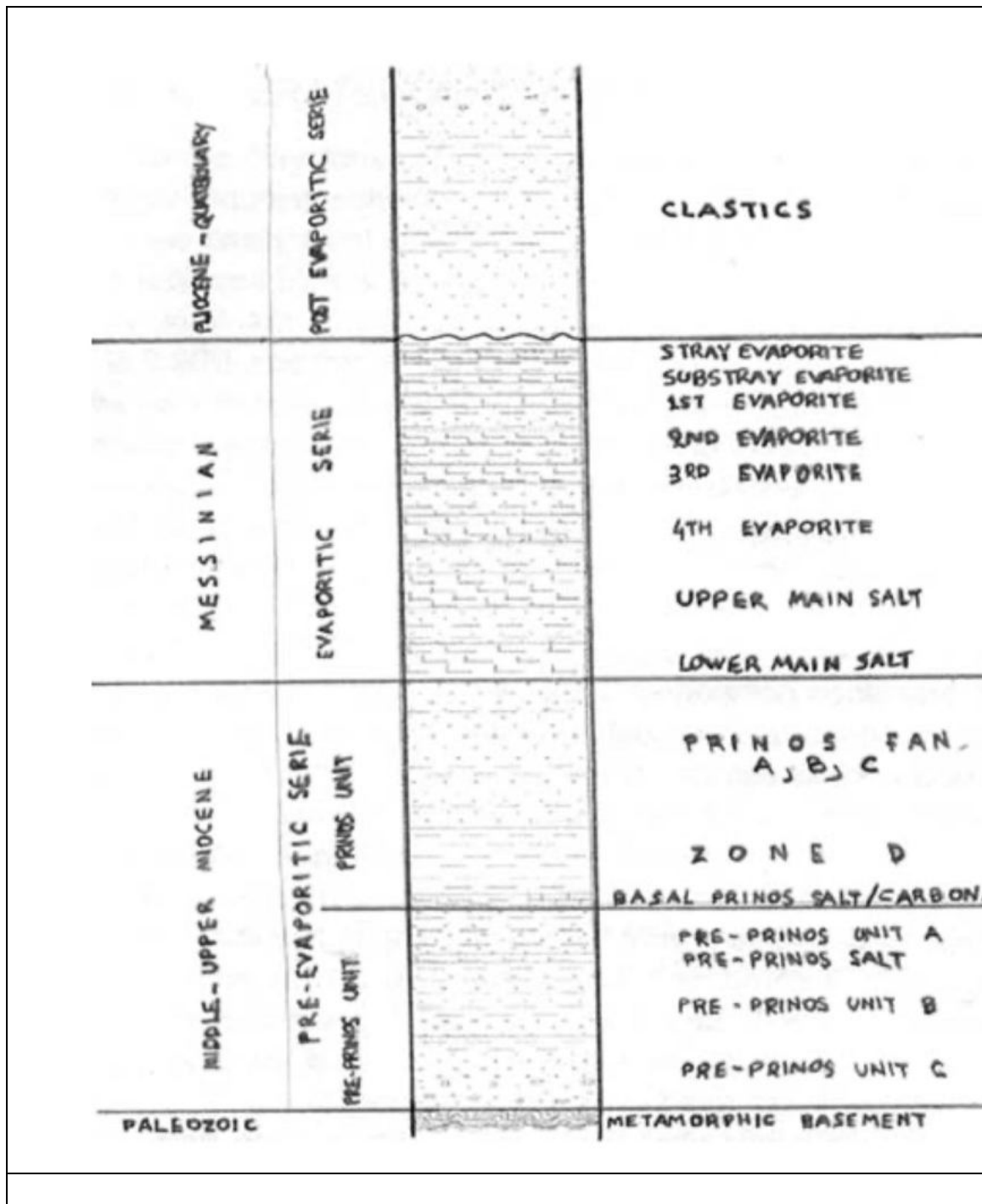


Figure 10: Prinos basin stratigraphic column. (Proedrou and Sidiropoulos (2004)).

The general structure of the taphrogenetic basin, it is a dome-like anticline formed by NW-striking and SW-dipping syngenetic faults also related to the trapping mechanism (Proedrou & Papaconstantinou, 2004).

Reservoirs are mainly sandstones and, secondly, siltstones, which resulted in the uppermost Miocene depositions from deltaic, marine and turbiditic environments. The evaporites cover the whole basin, keeping hydrocarbons below them, except for an upward movement of hydrocarbons in South Kavala and Ammodhis fields, probable due to fault activation. Generally, as depth increases, porosity and permeability decrease because of the weight of overburden layers, clay content and dolomitization (Proedrou & Papaconstantinou, 2004).

Chapter 4: Methodology

Herein, we describe the methodology for training the ML models, testing and validating the resulting models.

4.1 Data preparation

For water saturation prediction, 364 data points of shale volume (V_{shale}), formation porosity, permeability, and deep resistivity measurement, in addition to their corresponding water saturation, were collected from Well-1. In the first stage, the data were preprocessed to remove all non-real values; for example, all inputs with water saturation values less than zero or greater than 1.0 were considered non-real inputs and removed from the input parameters. Then the outliers were removed; for this purpose, the input parameters were studied mathematically. Finally, all values outside a ± 3.0 standard deviation range were treated as outliers and removed from the input dataset.

4.2 ANN model learning

After the data preprocessing stage, 334 datasets were left as valid data for model learning. Then the artificial neural network model was learned to predict the water saturation; during this learning step, the percentage of training data points in addition to the design parameters of the ANN model, including the learning function, the number of learning layers, the number of learning neurons per each layer, and the transferring function were optimized. The percentage of the training data was optimized between (40% to 90%). The performance of several training functions, including the Levenberg-Marquardt backpropagation function, resilient backpropagation function, and Bayesian regularization backpropagation function. The ability of various transferring functions to predict water saturation was also performed, including the tangential sigmoidal function, logarithmic sigmoidal function,

and linear function. The use of single, two, or three layers for learning the model with 4 to 30 neurons per layer was also studied.

The learning stage results concluded that 70% of the data for training using a single learning layer associated with five neurons, the Bayesian regularization backpropagation function for training, and the Logarithmic sigmoidal transferring function was the optimum for water saturation. **Error! Reference source not found.** shows a schematic of the optimized ANN model proposed for water saturation, other properties of this optimized model are summarized in Table . The statistical features and ranges of the training datasets are listed in

Table .

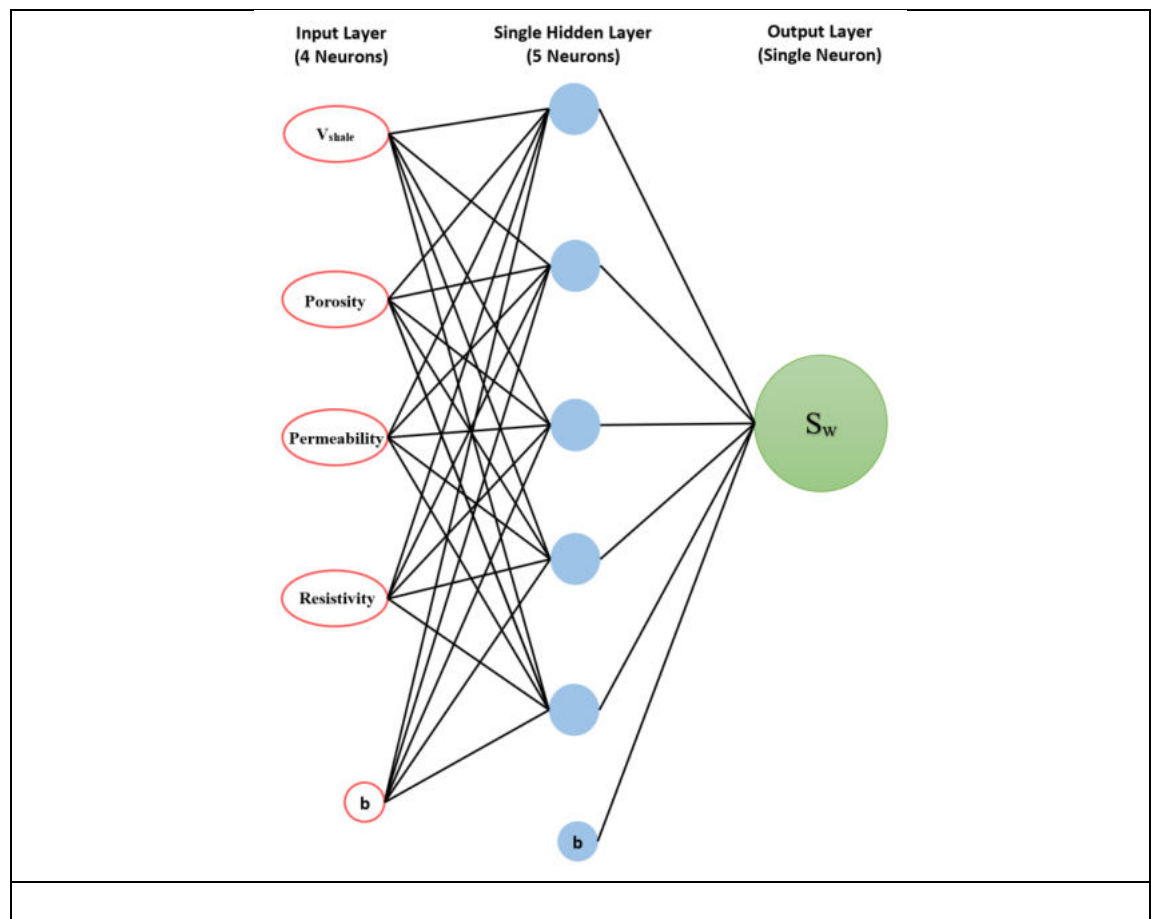


Figure 11: Schematic of the optimized ANN model for water saturation prediction.

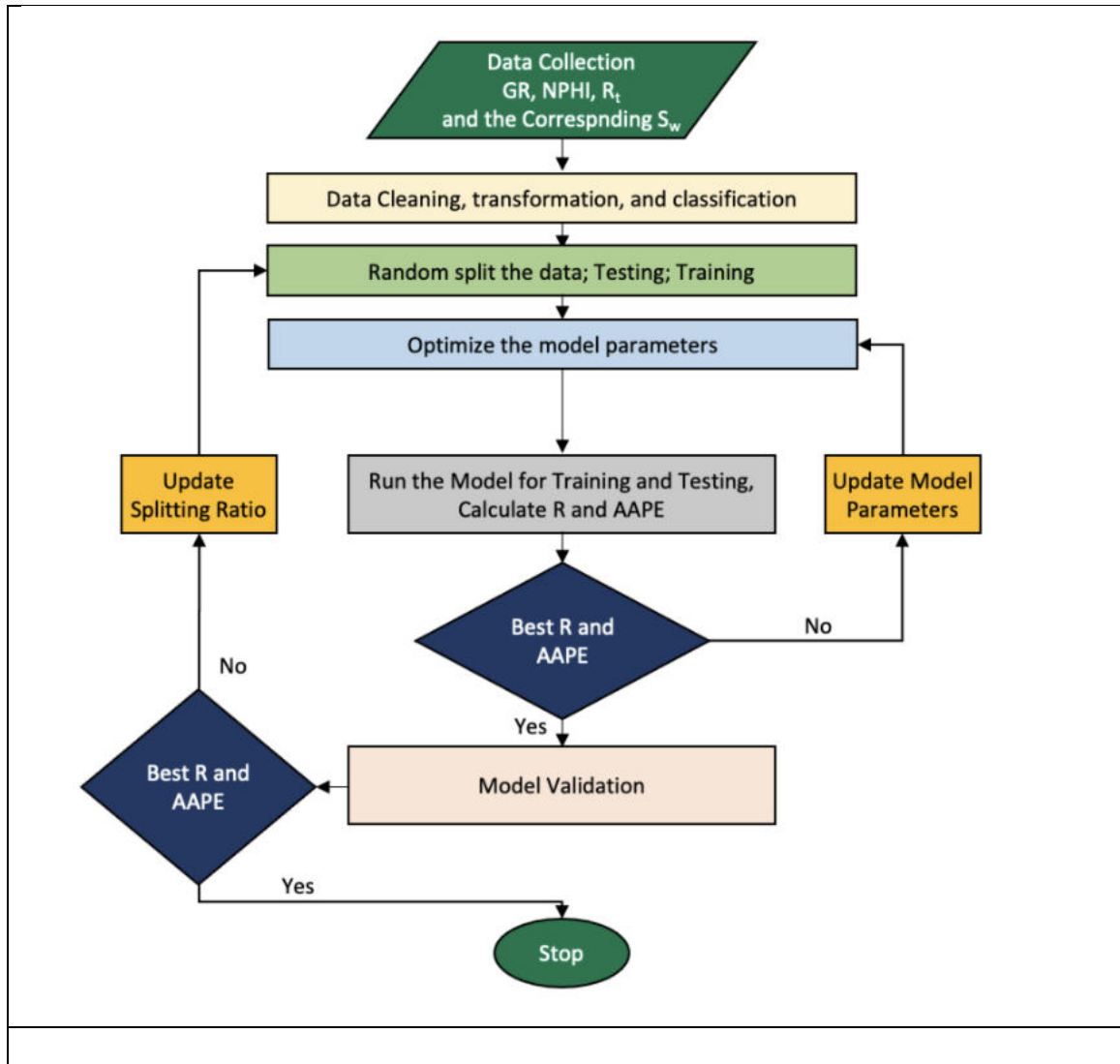


Figure 12: Flowchart of building ANN model

Table 1. The optimized parameters for water saturation prediction.

Parameter	Optimum Value
Training layers	Single
Number of neurons	Five
Training function	Bayesian regularization backpropagation
Transferring function	Logarithmic sigmoidal function

Table 2. The statistical features of the 252 datasets used to train the ANN for water saturation estimation.

	V _{shale} (frac.)	Porosity (frac.)	Permeability (mD)	Resistivity (Ohm.m)	S _w (frac.)
Minimum	0.012	0.005	0.0003	2.537	0.174
Maximum	0.852	0.173	13.86	19.63	1.000
Average	0.291	0.096	3.494	6.624	0.489
Mean	0.227	0.087	2.443	5.790	0.448
Median	0.231	0.099	2.952	5.535	0.427
Sample variance	0.037	0.001	5.989	14.80	0.041
Standard deviation	0.193	0.036	2.447	3.847	0.203

4.3 Extracting equation from the model

The ANN model with the properties in Table was then considered for developing a new equation for water saturation estimation. This model could be generally expressed as shown in Eq. (ξ).

$$\begin{aligned}
 S_{w,n} &= \sum_{i=1}^I w_{oi} \logsig \left(\sum_{j=1}^J w_{t_{ij}} Y_j + b_{t_i} \right) + b_o \\
 &= \sum_{i=1}^I w_{oi} \frac{1}{1 + e^{-\left(\sum_{j=1}^J w_{t_{ij}} Y_j + b_{t_i} \right)}} + b_o
 \end{aligned} \tag{ξ}$$

where $S_{w,n}$ is the normalized water saturation, i and J denote the total number of neurons and input parameters which are five and four, respectively, in this case. w and b represent

the weights and biases, respectively. t denotes the training layer, while o represents the output layer. It is important to mention that the parameters introduced into the ANN model are usually automatically normalized by the two-point slop explained by Eqs. (5) and (6) to have a value between 1 and -1.

$$\frac{Y_n - Y_{n,min}}{Y_{n,max} - Y_{n,min}} = \frac{Y - Y_{min}}{Y_{max} - Y_{min}} \quad (5)$$

$$Y_n = \left(\frac{Y - Y_{min}}{Y_{max} - Y_{min}} \right) (Y_{n,max} - Y_{n,min}) + Y_{n,min} \quad (6)$$

where Y denotes the input parameter, which could be, in this case, the V_{shale} , porosity, permeability, resistivity, or water saturation, the subscript n indicates the normalized parameter, and min and max represent the minimum and maximum, respectively. For example, since the input parameters are normalized between -1 and 1, $Y_{n,min}$ will be -1 and $Y_{n,max}$ will be 1. Therefore Eq. (6) could be written as in Eq. (7)

$$Y_n = 2 \left(\frac{Y - Y_{min}}{Y_{max} - Y_{min}} \right) + 1 \quad (7)$$

General expressions for normalizing the four inputs considered in this study could be derived by substituting for the minimum and maximum values of each parameter, as listed in

Table , to get the expressions in Eqs. (8) to (11)

$$V_{shale,n} = 2.383 (V_{shale} - 0.0121) + 1 \quad (8)$$

$$\phi_n = 11.89 (\phi - 0.0052) + 1 \quad (9)$$

$$K_n = 0.144 (K - 0.0003) + 1 \quad (10)$$

$$Res_n = 0.117 (Res - 2.537) + 1 \quad (11)$$

where $V_{shale,n}$, Φ_n , K_n , and Res_n are the normalized shale volume, porosity, permeability, and resistivity, respectively. Now Eq. (1) could be expanded and written as a function of these normalized parameters as in Eq. (12).

$$S_{w,n} = \sum_{i=1}^I w_{oi} \frac{1}{1 + e^{-(w_{t_{i,1}} V_{shale,n} + w_{t_{i,2}} \Phi_n + w_{t_{i,3}} K_n + w_{t_{i,4}} Res_n + b_{t_i})}} + b_o \quad (12)$$

The normalized parameters in Eq. (12)) could be calculated using Eqs. (8) to (11), while the weights and biases needed in this equation were extracted from the optimized ANN model and listed in Table .

Table 3. The extracted weights and biases for water saturation estimation.

		Input Layer					Output Layer	
		Weights (w1)				Biases	Weights	Bias
		j = 1	j = 2	j = 3	j = 4	(b1)	(w2)	(b2)
Number of neurons	i = 1	1.765	-0.061	-0.113	-1.094	-2.308	-3.828	3.302
	i = 2	0.569	7.871	0.855	-1.174	5.078	-0.579	
	i = 3	-0.269	-1.088	-1.479	-0.515	-2.930	3.360	
	i = 4	-2.487	-2.741	1.120	0.899	2.414	-3.865	
	i = 5	-0.346	-2.111	0.574	-3.528	-4.541	2.913	

The water saturation, as calculated from Eq. (12), is in the normalized form, requiring denormalization to obtain the actual water saturation values. The denormalized water

saturation could be calculated using Eq. (13) which is derived by rearranging Eq. (7) and substituting it for water saturation instead of Y.

$$S_w = \frac{(S_{w,n} + 1)}{2} (S_{w,max} - S_{w,min}) + S_{w,min} \quad (13)$$

The developed correlation of Eq. (13) was then tested using the remaining datasets collected from Well-1, 82 data points. Finally, the accuracy of Eq. (13) was evaluated by considering the AAPE and correlation coefficient between the actual and predicted water saturation.

Chapter 5: Chapter 5 Results and discussion

5.1 Results and Discussion

5.1.1 Learning the artificial neural networks model

ANN was implemented on 334 data points to train and test the model. The optimum training testing ratio was found to be 70/30. The best model quality control determined and selected the optimized hyperparameters. The optimum ANN model was developed with a single hidden layer and 5 neurons. The training function was “Bayesian regularization backpropagation” with the “Logarithmic sigmoidal function” data transfer function. Figures 13 and 14 present the cross plots for the actual vs the estimated S_w from the ANN model in both training and testing data sets.

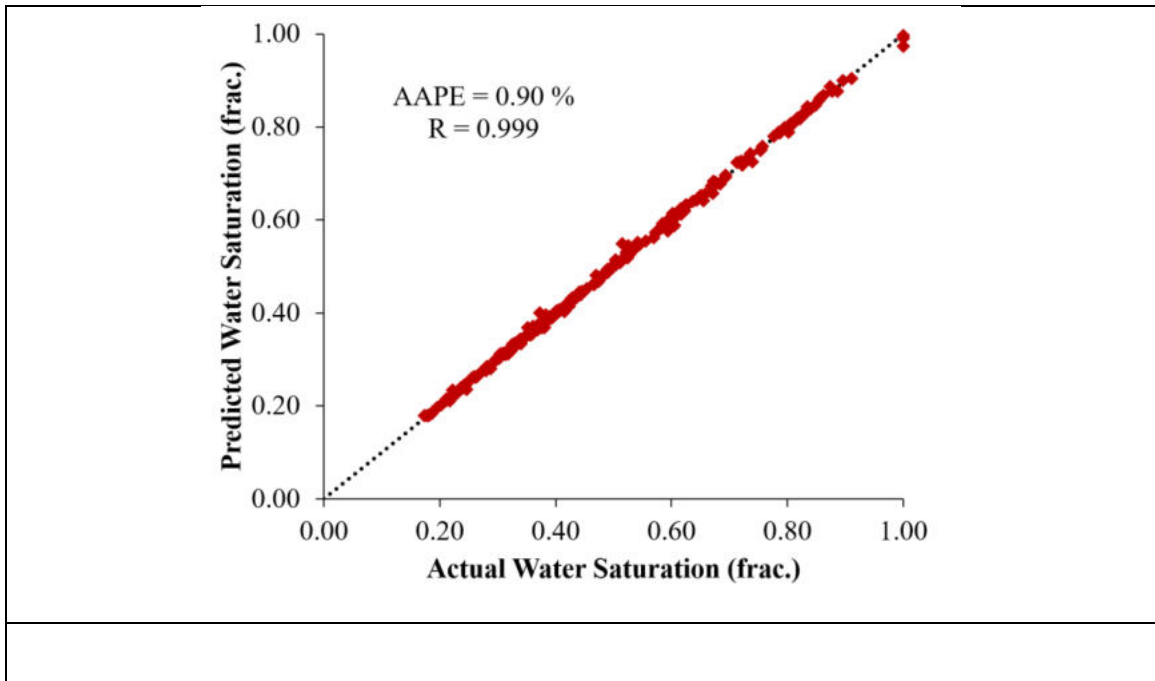


Figure 13: Training data

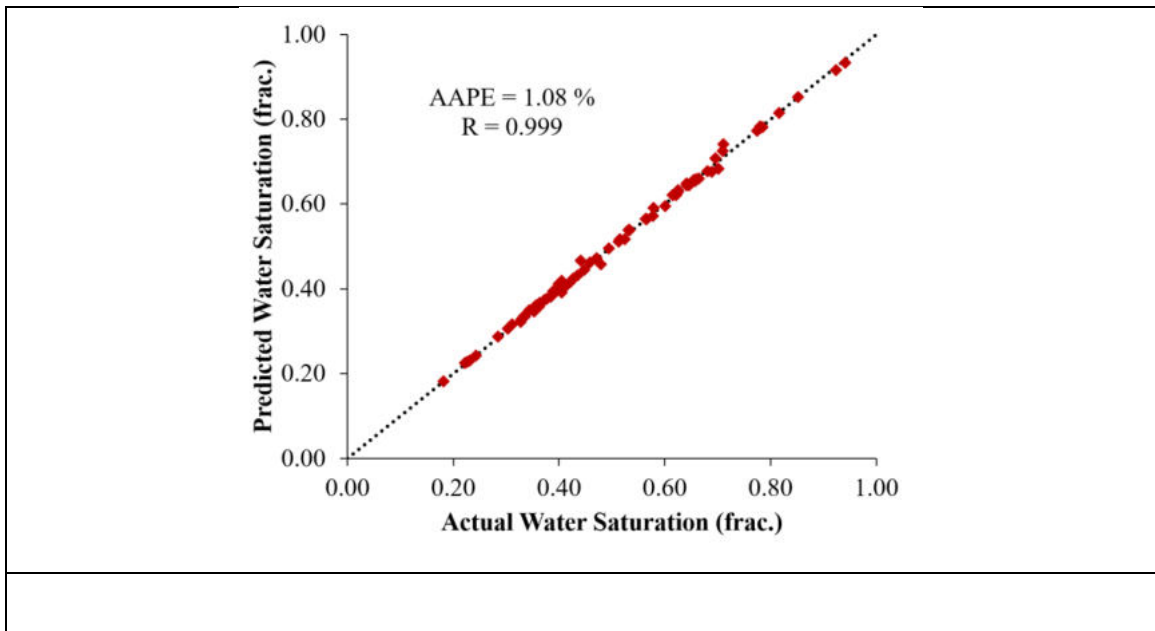


Figure 14: testing data

5.1.2 Training the artificial neural networks model

Figures 15 and 16 show that the Neural Network model is capable to predict the water saturation (S_w) from the well logs data with a good alignment. The training and testing dataset have an average absolute percent error (AAPE) as 0.90 and 1.08% with correlation coefficient (R) values as 0.999 and 0.9999. The Well no. 2 dataset was used to validate the ANN model.

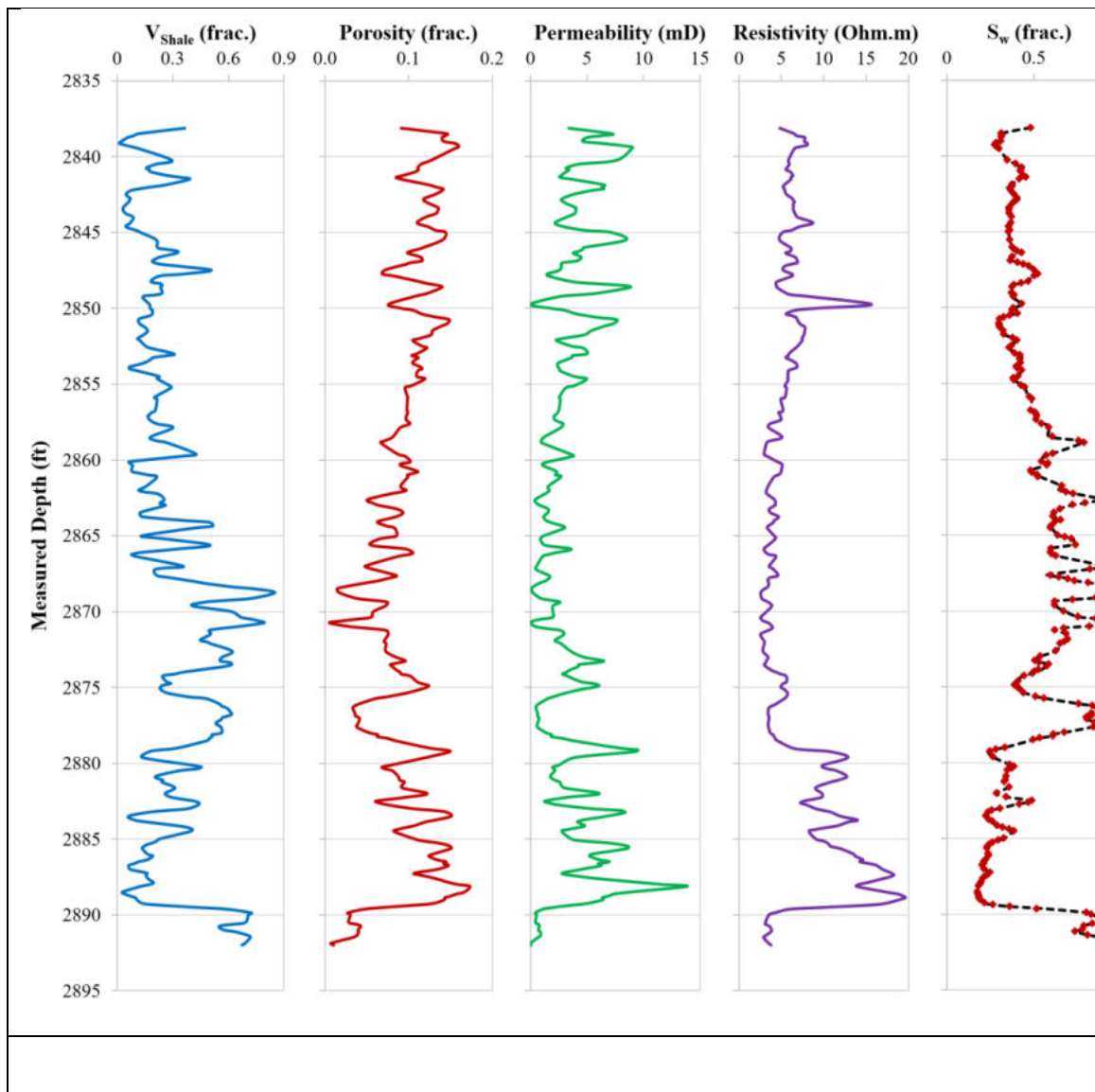


Figure 15: Prediction of water saturation by the ANN model (training data)

5.1.3 Testing the artificial neural networks model

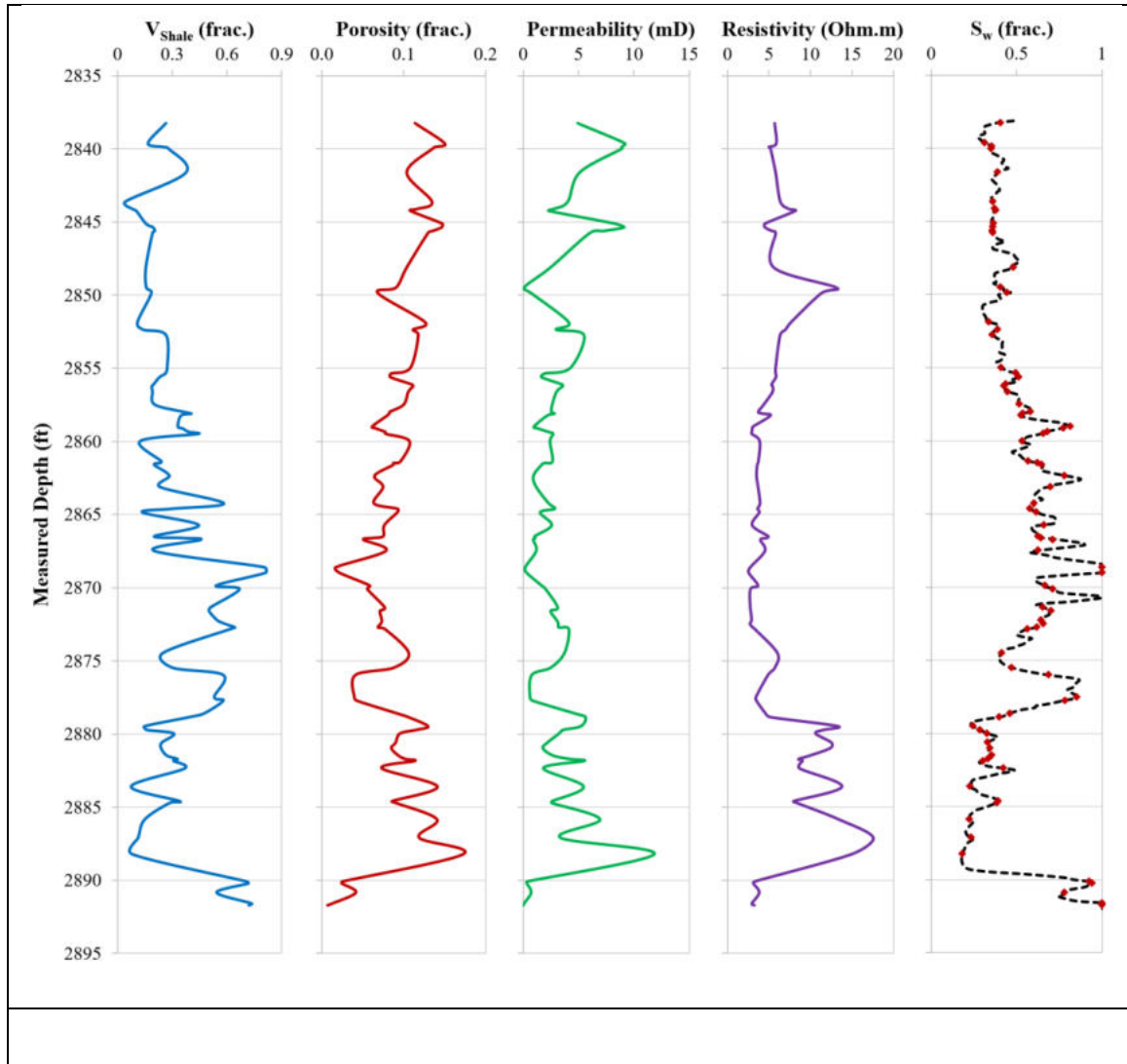


Figure 16: Testing data to predict the water saturation by ANN model

5.2 Conclusions

This research introduces the AI technique's capability (ANN) to estimate S_w using conventional logging data, including NPHI, Resistivity, and GR. The following are the main findings:

- The R values between the actual and estimated water saturation values based on the Artificial Neural Network model were 0.999 and 0.999 from the training and testing processes.
- The AAPE was less than 2% in the Artificial Neural Network model.
- The ANN Model produces an empirical correlation validated with unseen data and results in the correlation coefficient between the actual and the estimated water saturation values was 0.91 with an AAPE of 6%.

This study showed the capabilities of machine learning techniques to estimate the water saturation from the conventional well logs with a goodness of fit correlation coefficient (R) of 0.999–0.999 and an accuracy error of about $\pm 6\%$.

REFERENCES

- 1] Alkinani, H. H., et al. "Applications of artificial neural networks in the petroleum industry: A review." SPE Middle East Oil and Gas Show and Conference. OnePetro, 2019.
- 2] Kalam, S., et al. "An ANN model to predict oil recovery from a 5-spot waterflood of a heterogeneous reservoir." Journal of Petroleum Science and Engineering (2021): 110012.
- 3] Mahmoud, A. A., et al. "Estimation of oil recovery factor for water drive sandy reservoirs through applications of artificial intelligence." Energies 12.19 (2019): 3671.
- 4] Al-AbdulJabbar, A., et al. "Prediction of the rate of penetration while drilling horizontal carbonate reservoirs using the self-adaptive artificial neural networks technique." Sustainability 12.4 (2020): 1376.
- 5] Tariq, Z., et al. "A Data-Driven Machine Learning Approach to Predict the Natural Gas Density of Pure and Mixed Hydrocarbons." Journal of Energy Resources Technology 143.9 (2021): 092801.
- 6] Alsamadony, K. L., et al. "Deep Learning Enabled Deblurring of Computed Tomography Images of Porous Media." SPE Symposium: Artificial Intelligence-Towards a Resilient and Efficient Energy Industry. OnePetro, 2021.
- 7] Ahmed S, A., et al. "Prediction of pore and fracture pressures using support vector machine." International Petroleum Technology Conference. OnePetro, 2019.

- 8] Elkatatny, S., et al. "New robust model to estimate formation tops in real time using artificial neural networks (ANN)." *Petrophysics-The SPWLA Journal of Formation Evaluation and Reservoir Description* 60.06 (2019): 825-837.
- 9] Mahmoud, A. A., et al. "Application of machine learning models for real-time prediction of the formation lithology and tops from the drilling parameters." *Journal of Petroleum Science and Engineering* 203 (2021): 108574.
- 10] Gussow, W.C., Migration of reservoir fluids. *Journal of Petroleum Technology*, 1968. 20(04): p. 353-365.
- 11] Dandekar, A., *Petroleum reservoir rocks and fluid properties*: Boca Raton. Florida, Taylor amp Francis Group, 2006.
- 12] Kamel, M.H. and W.M. Mabrouk, Estimation of shale volume using a combination of the three porosity logs. *Journal of Petroleum Science and Engineering*, 2003. 40(3-4): p. 145-157.
- 13] Fernandez, G.J., M. Grinberg, and M. Valenzuela. Different Techniques to Calculate Horizontal Permeability and Initial Water Saturation: Case Study. in *SPE Latin American and Caribbean Petroleum Engineering Conference*. 2003.
- 14] Ellis, D.V. and J.M. Singer, Saturation and Permeability Estimation, in *Well Logging for Earth Scientists*, D.V. Ellis and J.M. Singer, Editors. 2007, Springer Netherlands: Dordrecht. p. 653-685.
- 15] Xiao, L., et al., Estimation of Saturation Exponent from Nuclear Magnetic Resonance (NMR) Logs in Low Permeability Reservoirs. *Applied Magnetic Resonance*, 2013. 44(3): p. 333-347.

- 16] Liu, K., et al., A new method for calculation of water saturation in shale gas reservoirs using VP-to-VS ratio and porosity. *Journal of Geophysics and Engineering*, 2018. 15(1): p. 224-233.
- 17] Sam-Marcus, J., et al., A proposed solution to the determination of water saturation: using a modelled equation. *Journal of Petroleum Exploration and Production Technology*, 2018. 8(4): p. 1009-1015.
- 18] Stiles, J. and J. Hutfilz, The use of routine and special core analysis in characterizing Brent Group reservoirs, UK North Sea. *Journal of Petroleum Technology*, 1992. 44(06): p. 704-713.
- 19] Kokkedee, J., et al. Improved Special Core Analysis: Scope for a reduced residual oil saturation. In *Society of core analysis conference paper*. 1996.
- 20] McPhee, C., J. Reed, and I. Zubizarreta, *Core analysis: a best practice guide*. 2015: Elsevier.
- 21] Hearst, J.R. and P.H. Nelson, *Well logging for physical properties*. 1985, United States: Mc Graw-Hill Book Co.
- 22] Archie, G.E., The Electrical Resistivity Log as an Aid in Determining Some Reservoir Characteristics. *Transactions of the AIME*, 1942. 146(01): p. 54-62.
- 23] Pickett, G.R., *Acoustic Character Logs and Their Applications in Formation Evaluation*. *Journal of Petroleum Technology*, 1963. 15(06): p. 659-667.
- 24] Poupon, A. and J. Leveaux. *Evaluation Of Water Saturation In Shaly Formations*. Presented at the SPWLA 12th Annual Logging Symposium, Dallas, Texas, May 1971. 1971.

- 25] Kadjhodaie, A. and R. Rezaee, A new correlation for water saturation calculation in gas shale reservoirs based on compensation of kerogen-clay conductivity. *Journal of Petroleum Science and Engineering*, 2016. 146: p. 932-939.
- 26] Jahanbakhshi, R., et al., Intelligent prediction of differential pipe sticking by support vector machine compared with conventional artificial neural networks: an example of Iranian offshore oil fields. *SPE Drilling & Completion*, 2012. 27(04): p. 586-595.
- 27] Tariq, Z., M. Mahmoud, and A. Abdulraheem, Machine Learning-Based Improved Pressure–Volume–Temperature Correlations for Black Oil Reservoirs. *Journal of Energy Resources Technology*, 2021. 143(11): p. 113003.
- 28] Alsaihati, A., et al., Use of machine learning and data analytics to detect downhole abnormalities while drilling horizontal wells, with real case study. *Journal of Energy Resources Technology*, 2021. 143(4): p. 043201.
- 29] Yang, J., et al., A Novel Method for Fracture Pressure Prediction in Shallow Formation During Deep-Water Drilling. *Journal of Energy Resources Technology*, 2021. 144(3).
- 30] Castillo, E., et al., Some Applications of Functional Networks in Statistics and Engineering. *Technometrics*, 2001. 43(1): p. 10-24. 20
- 31] Hagan, M.T. and M.B. Menhaj, Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, 1994. 5(6): p. 989-993.
- 32] Jang, J.R. and S. Chuen-Tsai, Neuro-fuzzy modeling and control. *Proceedings of the IEEE*, 1995. 83(3): p. 378-406.

- 33] Ahmed, A.A., et al. Application of Artificial Intelligence Techniques in Estimating Oil Recovery Factor for Water Derive Sandy Reservoirs. in SPE Kuwait Oil & Gas Show and Conference. 2017. Kuwait City, Kuwait, October 2017.
- 34] Acar, M.C. and B. Kaya, Models to estimate the elastic modulus of weak rocks based on least square support vector machine. *Arabian Journal of Geosciences*, 2020. 13(14): p. 590.
- 35] Gowida, A., S. Elkatatny, and H. Gamal, Unconfined compressive strength (UCS) prediction in real-time while drilling using artificial intelligence tools. *Neural Computing and Applications*, 2021.
- 36] Zhao, H.-b. and S. Yin, Geomechanical parameters identification by particle swarm optimization and support vector machine. *Applied Mathematical Modelling*, 2009. 33(10): p. 3997-4012.
- 37] Al Dhaif, R., A.F. Ibrahim, and S. Elkatatny, Prediction of Surface Oil Rates for Volatile Oil and Gas Condensate Reservoirs Using Artificial Intelligence Techniques. *Journal of Energy Resources Technology*, 2021: p. 1-14.
- 38] Hegde, C., S. Wallace, and K. Gray. Using Trees, Bagging, and Random Forests to Predict Rate of Penetration During Drilling. in Paper presented at the SPE Middle East Intelligent Oil and Gas Conference and Exhibition, Abu Dhabi, UAE, September 2015. 2015.
- 39] Castillo, E., et al., Functional Networks: A New Network-Based Methodology. *Computer-Aided Civil and Infrastructure Engineering*, 2000. 15(2): p. 90-106.

- 40] Anifowose, F., J. Labadin, and A. Abdulraheem. Prediction of Petroleum Reservoir Properties using Different Versions of Adaptive Neuro-Fuzzy Inference System Hybrid Models. in CISIM 2013. 2013.
- 41] Ahmed, A. and M. Khalid, An intelligent framework for short-term multi-step wind speed forecasting based on Functional Networks. *Applied Energy*, 2018. 225: p. 902-911.
- 42] Durgesh, K.S. and B. Lekha, Data classification using support vector machine. *Journal of theoretical and applied information technology*, 2010. 12(1): p. 1-7.
- 43] Gholami, R. and N. Fakhari, Chapter 27 - Support Vector Machine: Principles, Parameters, and Applications, in *Handbook of Neural Computation*, P. Samui, S. Sekhar, and V.E. Balas, Editors. 2017, Academic Press. p. 515-535.
- 44] Pisner, D.A. and D.M. Schnyer, Support vector machine, in *Machine Learning*. 2020, Elsevier. p. 101-121.
- 45] Suthaharan, S., *Machine learning models and algorithms for big data classification*. *Integr. Ser. Inf. Syst*, 2016. 36: p. 1-12. |

Vitae

Name :Mohamed Farouk Mohamed Gad
Nationality :Egyptian
Date of Birth :02/11/1988
Email mohammed.farouk.gad@gmail.com
Address :3353 Al Tal St. Al Rayan Dist. ,
Riyadh 14213,
Saudi Arabia

Academic Background: King Fahd University of Petroleum and Mineral (KFUPM),

M.Sc. in Geophysics,

December 2022

Cairo University

Diploma in Petroleum Geology,

June 2012

Ain Shams University

B.S. in Geophysics,

June 2010

Publication: **Gad, M.**, Mahmoud, A.A., Panagopoulos, G., Kirmizakis, P., Elkatatny S., Waheed U.B, Souplos, P., 2023, A New Empirical Relation for Water Saturation Prediction Based on the Artificial Neural Networks with Application to an Oilfield in Greece. Submitted to *Journal of Petroleum Science and Engineering*.