LOG DATA-DRIVEN PREDICTIVE MODELS AND FEATURE RANKING IN RESERVOIR CHARACTERIZATION

By

©Mohammad Islam Miah, B.Sc. (Engg.), M. Engg.

A Thesis Submitted to the School of Graduate Studies

In Partial Fulfilment of the Requirements for the Degree of

Doctor of Philosophy

(Oil and Gas Engineering)

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

St. John's, NL, Canada.

May 2020

DEDICATIONS

This dissertation is dedicated to my beloved daughter (Udita Islam Amirah) and wife (Ms. Mafruha Akhter Ovi).

ABSTRACT

Log-based reservoir characterization is one of the widely used techniques to estimate the reservoir properties and make decisions for hydrocarbon production. Use of the machine learning tools is becoming a more accessible approach for data-driven model development. The objective of this research is to identify and rank the most contributing log variables considering their relative performance for prediction of water saturation and rock strength using the machine learning tools. The single layer and multi-layer perception (MLP) artificial neural network (ANN) and the kernel function-based least-squares support vector machine (LS-SVM) techniques are employed for model development. The models can capture the non-linear behavior and high-dimensional complex relationships among real field log data variables. The mutual information (MI) is used to investigate the dependency of predictor subset variables in a model and to rank log variables according to their importance. The connectionist models are also examined to find reliable data-driven predictive models to estimate reservoir properties and rock strength. A new correlation is developed to obtain the in-situ rock strength of the siliciclastic rocks using the most important log parameters. The model predictions are compared/validated against the measured values as well as results obtained from existing log-based correlations. The approaches suggested in this study (connectionist and MI strategies) can assist engineers/operators to run a few numbers of logging tools for prediction of reservoir and rock properties to save the exploration costs. Also, it is expected that the introduced robust data-driven predictive models will enable engineers to better manage the wellbore stability and formation analysis in terms of technical, economic, and environmental aspects.

ACKNOWLEDGEMENTS

At first, I am very much grateful to the most powerful, the gracious Almighty Allah for giving me knowledge, energy, and patience for completing the research work successfully.

I want to express my deepest indebtedness and gratitude to my research supervisors Dr. Salim Ahmed (Associate Professor) and Dr. Sohrab Zendehboudi (Associate Professor and Equinor Research Chair), Faculty of Engineering and Applied Science (FEAS), for their continuous guidance, valuable suggestions, constructive comments, mentoring, and endless encouragement throughout the research work.

This dissertation would not have been possible without the financial support of Equinor (formerly Statoil) Canada Ltd., Natural Sciences and Engineering Research Council of Canada (NSERC), InnovateNL, School of Graduate Studies (SGS) and Departmental support through the FEAS, Memorial University.

I would like to express my sincere gratitude towards my former supervisor Dr. M. Enamul Hossain and supervisory committee member of Dr. Stephen Butt for their valuable research advice, guidance, and mentoring for the research ideas during the study period.

It is my pleasure to express my gratitude to Dr. Syed Imtiaz (Head, Dept. of Process Engineering), Dr. Faisal Khan (Associate Dean) and Dr. Greg F. Naterer (Dean), FEAS, MUN, for providing the different logistic supports during my PhD program. Also, I want to recognize and extend my sincere gratitude to research group students, academic officers/staffs of both FEAS and SGS, and friends for their continuous support during my

stay at MUN. In the meantime, the help of Ms. Tina Dwyer and Ms. Colleen Mahoney is so much appreciated.

Last but not least, I would like to express my deepest appreciation to my lovely wife (Ms. Mafruha Akhter Ovi) and parents for their continuous sacrifices, support, and inspiration to accomplish the dissertation.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Research Background

Reservoir characterization plays a critical role in appraising the economic success of reservoir management and development methods. It also helps in decision making for hydrocarbon exploration, perforation for production, and improving the reliability of the models for predicting reservoir properties and rock strength. Furthermore, it focuses on not only the understanding of past reservoir properties, but also predicting the future reservoir conditions. The ultimate target of reservoir characterization is to show the nature of rock-fluid properties and to develop a reservoir model with high accuracy and minimal uncertainties.

The formation evaluation or reservoir characterization is a continuous process that includes integrated tasks (e.g., coring, logging, and well testing). In the oil and gas industry, it can be accomplished by various techniques such as direct measurement through experimental core analysis or indirect measurement techniques. The most common indirect techniques of reservoir characterization are downhole well tests, wireline logging, geostatistical and geophysical interpretation, reservoir geomechanical analysis, geo-modeling (Dubois et al., 2006; Close and Caycedo, 2011; Tiab and Donaldson, 2015; Sylvester et al., 2015; Azevedo and Soares, 2017; Yusuf et al., 2019; Elkatatny et al., 2019), and soft computing approaches. The most popular machine learning (soft computing) techniques are artificial neural network (ANN), least-square support vector machine (LS-SVM), gene expression

programming (GEP), fuzzy logic (FL), and hybrid computational methods (Cuddy, 2000; Masoud, 2004; Wong et al., 2013; Onalo, 2019; Ertekin and Sun, 2019) which have been used in various engineering disciplines including chemical and petroleum engineering.

It is essential for petroleum and reservoir engineers, drillers, and geologists to obtain detailed information about the type and characteristics of rocks, and flow related properties of a reservoir through estimation of a variety of rock properties (e.g., resistivity, porosity, permeability, and fluid saturation) and rock strength parameters. These features are not the same or uniform in hydrocarbon formations throughout the world due to the various heterogeneities in pore geometry, permeability, and fluid saturation. The most extensive technique for reservoir characterization is petrophysical analysis using coring and well logging data. Water saturation is one of the most important reservoir properties that enables to determine perforation depth for hydrocarbon production in both offshore and onshore fields. Meanwhile, the uniaxial compressive strength (UCS) is one of the crucial rock strength parameters that can be employed to assess reservoir sanding potential during formation pressure drawdown, drilling optimization as well as wellbore stability analysis before smart well completion.

The most widespread strategy for reservoir evaluation/characterization is direct measurements such as experimental core analysis and indirect estimation using wireline log-based correlations. Although the laboratory-based direct method for water saturation estimation is more accurate, it is time-consuming and expensive. A limited number of

samples can capture a few depth-intervals of the well (Adeniran et al., 2009). Due to the complex geological behavior and heterogeneity of reservoirs, a large set of samples are required to properly characterize a reservoir's rock and geomechanical properties. Most of the times, high-quality cylindrical core specimens' preparation with regular geometry or a large scale core sample is not possible to collect due to the complex geological behaviour and heterogeneity of the reservoir.

Over the past few decades, numerous studies have been performed to develop empirical models to estimate rock strength and reservoir properties using petrophysical logs such as resistivity, sonic, density, and neutron porosity logs. Several researchers developed models or correlations to estimate water saturation and rock strength. The existinmg log-based models or correlations for water saturation and rock strength of the formation were listed and summarized by different authors (Chang et al. 2006; Shedid and Saad, 2017). All those petrophysical models are limited to the nature of the lithology, and lithological parameters such as tortuosity factor, cementation, and saturation exponent of the rock.

Nowadays, the machine learning (ML) or artificial intelligence (AI) approaches such as artificial neural network (ANN) and least squares support vector machine (LS-SVM) are becoming more popular tools for data-driven model development and prediction of rock properties to save the experimental and operational costs. The connectionist tools do not depend on the geological characteristics and lithological parameters (e.g., lithology type, saturation exponent, and cementation factor); they overcome some drawbacks of the

existing empirical models for reservoir rock properties. ANN and LS-SVM are powerful deterministic tools in the petroleum industry to capture the uncertainty and non-linear behaviour of the input variables.

1.2 Research Motivation

The formation specimen collection and processing of high-quality cores from a reservoir are complex, tedious, expensive, and time-consuming operations. Lack of experimental analysis of reservoir properties should not be a limiting factor in reservoir characterization. The wireline log data is an alternative and viable option to obtain the in-site profile of reservoir rock properties. If the reservoir core samples are not available, reservoir properties and strength of rock can be predicted using models/correlations relating wireline logs (such as gamma-ray, resistivity, porosity, and nuclear-magnetic resonance logs) and drilling data as well as downhole testing data.

A comprehensive study is required to identify the relative contribution and significance of log parameters in the data-driven reservoir predictive models while estimating in-situ reservoir properties using several log data. The current research aims to identify the most significant log variable in the model while predicting in-situ reservoir petrophysical/mechanical rock properties using machine learning approaches. It is believed that the research strategies employed in this study can be less time consuming and cost-effective for efficient reservoir characterization as well as formation evaluation.

1.3 Objectives of the Research

Based on the literature review, a comprehensive investigation is further required to identify the relative contributions of input log variables in the predictive models for reservoir properties and/or geomechanical properties of the rock formation. To the best of our knowledge, feature ranking of logging variables to predict in-situ water saturation as well as rock strength profile by coupling logging data and machine learning tools have not been investigated systematically. The attribute selection and ranking of logging data with very good precision appear to be a serious challenge for not only petroleum engineers and drillers, but also geologists to estimate the water saturation and rock strength using the wireline log data. The current research work is planned to fill in the knowledge gap by finding the most contributing parameters and feature ranking of log variables, according to their relative contribution while predicting water saturation and strength of rock. The objectives of this research are listed as follows:

- To use data-driven models for prediction of water saturation and rock strength using log data
- To rank input parameters contributing to the output variables using mutual information and smart connectionist models
- To perform parametric sensitivity analysis and compare the predictive performance of deterministic tools through statistical analysis

1.4 Organization of the Thesis

A manuscript style format has been followed in preparing this thesis. A graphical representation of the research approach in the study is shown in Figure 1.1. The outline of the thesis and individual chapters are presented below:

- Chapter 1 highlights the research background, motivation, and objectives of the study.

- Chapter 2 presents a brief review of the literature on theories and techniques for reservoir characterization as well as formation evaluation. To identify the knowledge gaps and research scopes in the area of memory-based reservoir characterization as well as reservoir modeling of thermal flooding technique capturing the rock-fluid interactions in formation porous media, two review articles are written and published. A version of one of the published review articles entitled "Application of Memory Concept on Petroleum Reservoir Characterization: A Critical Review" is available in the www.onepetro.org archive of the Society of Petroleum Engineers (SPE-187676-MS). The second review article (Modeling of Temperature Distribution and Oil Displacement during Thermal Recovery in Porous Media: A Critical Review) is given in Appendix A. These studies were done during the research period in collaboration with other researchers. Also, the concepts and fundamentals of machine learning approaches are briefly described in this chapter. The relevant literature review on the relevant machine learning techniques is presented in the subsequent chapters 3, 4 and 5.

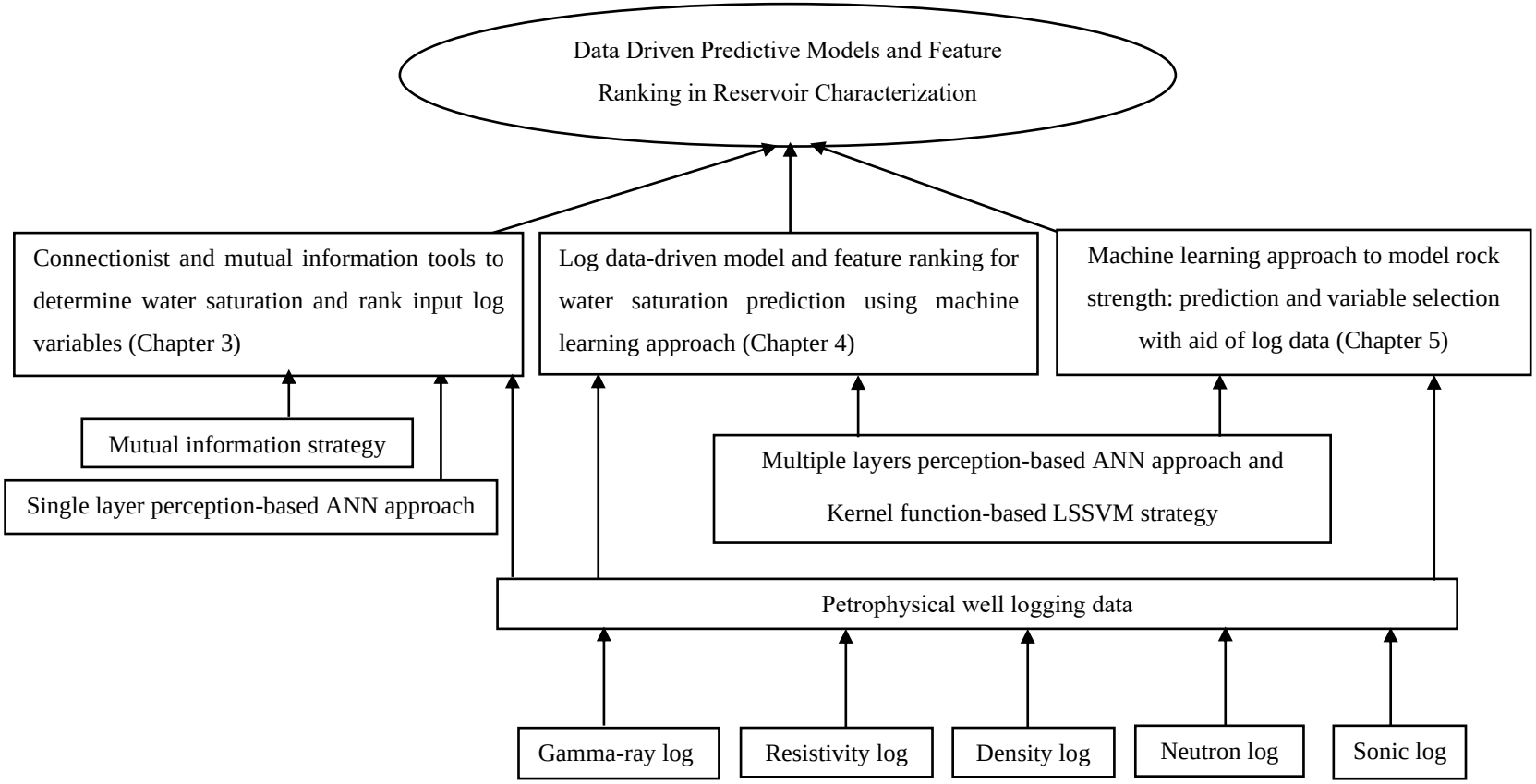

Figure 1.1: Integration of research strategies in the study.

- Chapter 3 presents the theoretical concepts and background regarding the petrophysical models for water saturation. It also covers the methodology for modeling through single layer perception-based ANN and MI techniques. This chapter has been published in the "Journal of Petroleum Science and Engineering".

- Chapter 4 presents the knowledge gap on log data-driven models and feature ranking while predicting water saturation using machine learning approaches. In this chapter, the multi-layer perception-based ANN and CSA-based LSSVM were employed to develop a model for water saturation. This chapter has been published in the "Journal of Petroleum Science and Engineering".

- Chapter 5 presents the investigation on rock strength model and variable selection. This chapter includes detailed information about the log-based rock strength models and knowledge gap on the AI-based connectionist model for variable selection while predicting rock strength. In the study, the smart connectionist AI models (ANN and LS-SVM) are employed to construct the data-driven model and variable selection. This chapter has been reviewed by the supervisory committee and submitted to the "Rock Mechanics and Rock Engineering" Journal in the form of a revised manuscript.

- Chapter 6 presents the overall conclusions and recommendations of the thesis. Thus, the readers also will obtain some new directions for future research in the area of reservoir analysis as well as formation evaluation.

1.5 Research Novelty and Contributions

The novelty and contribution of this research are in the reliable prediction of water saturation and rock strength in reservoir characterization using less number of log data than typically used in the existing methods. The novelties and contributions are highlighted below:

- The mutual information (MI) is implemented for the first time to rank input parameters contributing to the water saturation when log data are available. In this study, development and optimization techniques on the basis of connectionist models for water saturation estimation are illustrated systematically to investigate the dependency of predictor variables using MI and ANN. The MI approach reveals that the important log variables are the true resistivity and bulk density while the gamma-ray and neutron porosity have minor effects on the water saturation. An important finding of this study is that the primary log variables (e.g., true resistivity and bulk density) can be used to accurately obtain the water saturation during the exploration of a reservoir using efficient deterministic strategies such as smart connectionist models. Utilization of such tools requires less time and computational complexities, leading to lower costs in the exploration stage while using fewer log variables. This contribution is presented in Chapter 3.

- The data-driven model is developed by coupling log data and machine learning techniques to identify and rank the most contributing log variables for predicting continuous water saturation profile. It is found that the significance order of influential

(higher to lower) predictor log variables are the true resistivity, bulk density, neutron porosity, photoelectric factor, gamma-ray, and sonic travel time based on their relative contribution to the water saturation. The strategy introduced in this study assists to forecast water saturation with a relatively few number of log variables, and thus, reduces the number of necessary logs to run during exploration, considerably lowering the exploration costs. This contribution is highlighted in chapter 4.

- The connectionist models of MLP-based ANN and LSSVM with radial basis kernel function are employed to predict the continuous in-situ rock strength with the aid of log data. One of the main findings is that the formation acoustic travel time and gamma-ray are the most significant contributing log variables (compared to other variables) while estimating the in-situ rock strength. A new correlation is also developed to predict the in-situ rock strength using influential log parameters for the siliciclastic sedimentary rocks. It is expected that the introduced robust data-driven predictive models/strategies and new correlations will enable rock engineers, drillers, and researchers to better manage the wellbore stability, rate of penetration, and rock formation analysis in terms of technical, economic, and environmental aspects. This contribution is presented in Chapter 5.

References

Adeniran, A., Elshafei, M., & Hamada, G. (2009). Functional Network soft sensor for formation porosity and water saturation in oil wells. In 2009 IEEE Instrumentation and Measurement Technology Conference (pp. 1138-1143). IEEE.

Chang, C., Zoback, M. D., & Khaksar, A. (2006). Empirical relations between rock strength and physical properties in sedimentary rocks. Journal of Petroleum Science Engineering, 51(3-4), 223–237. doi:10.1016/j.petrol.2006.01.003.

Close, D., & Caycedo, F. (2011). Integrated geophysics and geomodelling workflows for reservoir characterization: A case study of waterflood optimization. In SEG Technical Program Expanded Abstracts 2011 (pp. 1840-1844). Society of Exploration Geophysicists.

Dubois, M. K., Byrnes, A. P., Bohling, G. C., & Doveton, J. H. (2006). Multiscale geologic and petrophysical modeling of the giant hugoton gas field (permian), Kansas and Oklahoma, USA.

Elkatatny, S., Tariq, Z., Mahmoud, M., Abdulraheem, A., & Mohamed, I. (2019). An integrated approach for estimating static Young's modulus using artificial intelligence tools. Neural Computing and Applications, 31(8), 4123-4135.

Ertekin, T., & Sun, Q. (2019). Artificial Intelligence Applications in Reservoir Engineering: A Status Check. Energies, 12(15), 2897.

Masoud, N., (2004). Soft Computing-Based Computational Intelligent for Reservoir Characterization, J. Expert Systems with Applications 26 (1), Pp. 19-38.

Onalo, D. O. (2019). Dynamic data driven investigation of petrophysical and geomechanical properties for reservoir formation evaluation, Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL. Canada.

Shedid, S. A., & Saad, M. A. (2017). Analysis and field applications of water saturation models in shaly reservoirs. J. of Petroleum and Gas Engineering, 8(10), 111-122.

Sylvester, O., Bibobra, I., & Ogbon, O. N., (2015). Well Test and PTA for Reservoir Characterization of Key, American Journal of Engineering and Applied Sciences, Volume 8, Issue 4, Pp. 638-647.

Tiab, D., Donalson, E.C., 2015. In: Petrophysics: theory and practice of measuring reservoir rock and fluid transport properties. Gulf professional publishing.

Wong, P., Aminzadeh, F., & Nikravesh, M. (Eds.). (2013). Soft computing for reservoir characterization and modeling (Vol. 80). Physica, Heidelberg. Pp. 587.

Yusuf, B. O. (2019). Elastic properties prediction for drilling and subsurface evaluation for the Grand Banks and Niger Delta (Doctoral dissertation, Memorial University of Newfoundland).

CHAPTER 2: LITERATURE REVIEW

This section presents a comprehensive literature review to find the research scopes for the study. The aim of sub-section 2.1 is to revisit the basic concepts, existing techniques and models for reservoir characterization in porous media with the main focus on water saturation and rock strength estimation. Subsections 2.2 and 2.3 present the basic concepts and theory of mutual information and machine learning (artificial intelligent-based model) approaches with highlighting ANN and LS-SVM.

2.1 Reservoir Characterization

Reservoir characterization encompasses the understanding and procedures to describe the behavior of reservoir rock or fluid properties in a porous medium. Several methods and advanced technologies have been used to properly characterize clastic and carbonate reservoirs. It is a continuous process that can be accomplished using several techniques such as experimental measurements through core analysis, geophysical log interpretation, and down-hole testing (Lucia, 2007).

Also, the in-situ rock and fluid properties behavior can be analyzed through mathematical modeling. The continuous alteration of rock/fluid properties can be characterized using the memory concept (e.g., the effect of past events on the present and future course of developments (Zhang, 2003). It is also significant to consider the rock, and fluid properties as a function of time, and the inclusion of recently introduced memory concept in

petroleum engineering study. A detailed review of the existing techniques and models of rock and fluid properties related to memory concept in reservoir characterization is presented by Miah et al. (2017). In the study, the researchers revisited the literature survey and listed the knowledge gaps in the models developed for flow in porous media. For instance, the detailed information about the present status of memory-based fluid flow modeling, rock and fluid properties models under several assumptions during reservoir characterization is provided.

Moreover, the reservoir rock and fluid properties play an important role for both inductive and conductive heat transfer process, and considerably affect the energy balance during the thermal flooding processes. The effective thermal conductivity of formation is employed to characterize the thermal conductivity of reservoir rocks as thermal properties are determined by mineral constituents, porosity, and water saturation, as well as fluid saturating pore space (Green and Willhite, 1998). For instance, thermal conductivity can be obtained from the core and log data (Gąsior and Przelaskowska, 2014). The influence of temperature on reservoir rock and fluid properties plays a vital role in accurately predicting reservoir temperature distribution, oil displacement, and steam oil ratio. In Appendix A, a systematic review discusses about the impacts of rock properties (i.e., porosity and permeability), water saturation, and fluid properties on recovery performance during the thermal displacement process. This review aticle highlights the assumptions and limitations of the current models for thermal conductivity, and temperature distribution during the thermal flooding process in porous media. The following subsections cover the

literature review on rock properties, and modeling and estimation techniques in the context of reservoir characterization/analysis.

2.1.1 Reservoir petrophysical properties

The reservoir petrophysical properties such as rock resistivity, porosity, permeability, and fluid saturation can be measured through experimental core analysis and/or using wireline log data. The wireline logs are reliable to estimate in-situ porosity, water saturation, and log-based permeability in the absence of core data. The wireline logs (such as gamma-ray , resistivity, density, neutron, and sonic logs) are used to obtain the reservoir rock (petrophysical) properties.

The gamma-ray (GR) log measures the strength of the natural radioactivity (i.e., spontaneous decay of the atoms of certain isotopes into other isotopes) present in the formation. The gamma radioactivity is generally expressed in API unit (American Petroleum Institute). Note that 1 micrograms (μg) is equal to 16.5 API (Serra, 1984). The API unit can be defined as the one two-hundredth of the difference in curve deflection between zones of high and low radiation in the API gamma ray calibration pit in Houston, Texas (Serrra, 1984). The magnitude of GR is a function of concentration (by weight) of radioactive minerals and rock density. The major three radioactive minerals are potassium, thorium, and uranium. Potassium is more abundant in the formation, and it has a significant contribution to the GR response rather than other clay minerals in the formation. The presence of clay minerals (e.g., potassium, thorium, and uranium) in a rock formation may

considerably affect the reservoir petrophysical properties. The presence of clay minerals or shale in a reservoir may either be good or bad in terms of reservoir quality. On the other hand, the presence of a large amount of clay may result in a decrease in the porosity (i.e., void spaces in a rock) and permeability (i.e., the ability of fluid flow in a rock) of the reservoir so that the reservoir becomes non-productive. More commonly, clay minerals are found in sandstone reservoirs; however, not in carbonate rocks. The GR log is extensively used to obtain the shaliness (clay content or shale volume) that appreciably affects the predicted values of effective porosity (i.e. interconnected pore space in rock volume that is capable of transmitting the fluid) and water saturation (i.e. the ratio of water volume to pore volume) in shaly sand reservoirs.

Rock porosity measures the void (pore) space in the reservoir, which is available for accumulation of fluids. There are three types of logs, namely, density, neutron, and sonic logs that are used to estimate the amount of pore space in sedimentary rocks. Porosity in sandstones normally varies with the grain size distribution, grain shape, packing arrangement, cementation, and clay content (Asquith and Krygowski, 2004). The formation density log measures the electron density of a formation, which is used to detect gas-bearing zones, determine hydrocarbon density, and evaluate shaly-sand reservoirs as well as complex lithologies of the formation. The neutron log measures the hydrogen concentration or hydrogen index in the rock. In clean formations (shale free) where porosity is filled with water, oil or gas, the neutron log measures liquid-filled porosity. Whenever pores are filled with gas rather than oil or water, the reported neutron porosity

is less than the actual formation porosity. This occurs because there is a lower concentration of hydrogen in gas than oil or water. This lower concentration is not accounted for by the processing software of the logging tool; this is interpreted as a low porosity. A decrease in the neutron porosity due to the presence of gas is called the gas effect. Also, an increase in neutron porosity because of the presence of clays is called shale effect (Asquith and Krygowski, 2004). Neutron porosity increases for the cases with high amount of clay minerals to the reservoir. Effective porosity is defined as the porosity available to free fluids in the reservoir. The values of neutron and density porosity, corrected for the presence of clays, are used to estimate the effective porosity of the formation of interest for gas or oil reservoirs (Asquith and Krygowski, 2004); relevant equations are included in chapters 3 and 5.

On the other hand, the sonic tool (or borehole compensated device) measures interval transit time (DT) of a compressional sound wave travelling through the formation along the axis of the borehole. The sonic log device consists of one or more ultrasonic transmitters and two or more receivers. The true porosity can be calculated from the compressional sonic travel time (the reciprocal of wave velocity) using the Wyllie time-average equation (Wyllie et al. 1958). The interval transit time of a formation is increased due to the presence of hydrocarbons (e.g., hydrocarbon effect). If the effect of hydrocarbons is not corrected, the sonic derived true porosity will be too high.

Resistivity logs are used for the evaluation of reservoir pore fluids within formations. The resistivity (inverse of conductivity) of a formation can be measured from the laterolog or induction type logging tool; this is a crucial parameter in determining water saturation, which depends on the formation water resistivity (fully saturated water case), pore geometry, and amount of fluids (water and/or hydrocarbon) in the reservoir. The deep induction log measures the true resistivity (R_t, ohm-m) of the virgin formation. The formation water resistivity (R_w) can be estimated using the inverse Archie's formula (Archie's, 1941). More information about the common types of logging tools, their applications, and uncertainties can be found in the literature (Bassiouni, 1994; Asquith and Krygowski, 2004; Masoudi et al., 2017).

2.1.2 Reservoir formation properties and strength of rock

The unconfined rock compressive strength (UCS) has significant importance as a main rock strength parameter with applications in various fields including petroleum and mining engineering. It is a vital parameter for reservoir stimulation design, planning the mud weight window, bit selection, and real-time wellbore stability analysis. The drilling engineer would be able to optimize proper weight on bit and weight of drilling fluid as well as to investigate the efficiency of a drilling operation using the rock strength profile (Moos et al., 2003; Chatterjee et al., 2013). A drilling operation is an interaction between the rock and bit; the rock will fail when the resultant stress is greater than the rock strength (resistance of rock against loading). Different methods can be employed to estimate the rock strength. Estimation of rock strength (UCS) at the laboratory is very costly because

drilled core samples are required to be collected from well sites and then prepared for surface testing. When laboratory measurement of rock strength parameters (uniaxial compressive strength, UCS, and tensile rock strength) is time-consuming and expensive, the UCS is typically estimated from wireline logs such as density and sonic logs.

There are two types of body waves (compressional and shear), which are commonly used to estimate rock strength, physical properties profile, and dynamic elastic properties (e.g., Poisson's ratio, modulus of elasticity, and shear and bulk moduli) of the rock formation. The compressional wave (primary or longitudinal wave) has the particle motion in the direction of wave propagation in rock formations. The compressional wave has the fastest arrival times from an acoustic energy source. In rock porous media that are saturated with fluids, the primary wave travels through both the liquid and solid phases. The shear wave (secondary or transverse wave) has its particle motion perpendicular to its direction of wave propagation. The shear wave is the second-fastest wave, which can travel in the rock-solid; however, it does not travel through fluids (gas, oil, and water) in a saturated rock formation because fluids do not shear. Several empirical equations exist for predicting in-situ UCS profile along the wellbore using rock porosity, acoustic velocity (sonic travel time), dynamic elastic moduli, and other formation properties (Rabbani et al., 2012). Several correlations can be used to predict rock strength for the sedimentary rock types using log data, which can be found in the literature (Chang et al. 2006; Odunlami et al., 2011).

2.2 Information Theoretic Measures Approach

The most common measures of information used in different fields of engineering are entropy (H) and mutual information (MI). MI is one of the widely used concepts in communication engineering, which offers a procedure to measure the strength of the relationship between discrete and/or continuous variables. MI has been used to capture the dependencies among random variables relevant to a model using available data and feature selection methods in different disciplines (Shannon, 1948; Cover and Tomas, 2012; Ghaeinia et al., 2017; Pascoal et al., 2017). The detailed theory and procedure to estimate MI are provided in Chapter 3.

2.3 Theory and Application of Machine Learning Approach

The machine learning (ML) is a scientific computational aspect of artificial intelligent (AI) that deals with the design and development of different algorithms that allow models to learn based on real datasets such as wireline log and/or experimental core data. A major focus of ML research is to automatically learn to recognize complex patterns and make intelligent decisions based on feed data. Moreover, ML is closely related to different disciplines such as probability theory, artificial intelligence, data mining, pattern recognition, and theoretical computer science such as computational learning theory (Anifowose et al. 2011).

The most common ML techniques are supervised, unsupervised, and hybrid. The supervised learning is the ML technique in which the algorithm generates a function that

maps inputs to the desired outputs with the least error. Unsupervised learning is another ML strategy in which a set of inputs are examined without the target output. This is also known as clustering. The most common ML techniques are artificial neural network (ANN), fuzzy logic (FL), adaptive neuro-fuzzy inference system (ANFIS), gene expression programming (GEP), classic support vector machine (SVM), least squares support vector machine (LSSVM), and/or hybrid (two or more combinations) methods. Also, the hybrid machine learning-based model can be developed through a combination of LSSVM and global optimization technique of coupled simulated annealing (CSA) or particle swarm optimization (PSO). More information about the theory and mathematical formulation of ML (or AI) techniques can be found in the literature (Jang, 1993; Ashena and Thonhauser, 2015; Sebtosheikh et al. 2015; Barzegar et al. 2016; Barati-harooni et al. 2017; Gholami and Fakhari, 2017). The following two sub-sections briefly explain the ANN and LSSVM techniques due to the interest of the current study.

2.3.1 Fundamentals of artificial neural network

The ANN is a parallel distributed information processing model. The major components of ANN are input layer, hidden layer (s), training algorithm, transfer (activation) function, and output layer. The ANN has been established as a simplification of the mathematical model of the neural networks by having some assumptions listed below (Ashena and Thonhauser, 2015):

a) Variables (input or output) information processing is performed in many simple individual processors, called neurons.

b) The variable information is transmitted between neurons using connection links in the network.
c) A specified weight factor is assigned to each link to be multiplied by the variable information, which is passing through each link.
d) Each neuron assigns a desired bias term or a threshold value to be added to the sum for yielding a net value.
e) The net value is provided as an input to an activation (transfer) function; the output of the neuron would then be determined.

Ultimately, the function of the whole ANN structure is the calculation of the output of all the neurons existing in the connectionist network-based model. The number of neurons in the input layer relates to the number of input variables. The number of neurons in the output layer corresponds to the number of the output variable (s). The number of hidden layers and also the number of neurons of hidden layers can be defined by following a trial and error procedure (Ham and Kostanic, 2001). The single-layer perception and multi-layer perception (MLP) connectionist network solely depend on the adjustment of the weight factors between the layers. The output or hidden layers are assigned with activation (transfer) function. The most common transfer functions are threshold (unit step), piecewise linear, logistic sigmoid, sigmoid hyperbolic tangent function, and purlin transfer function. Generally, the output layer is associated with a purlin (linear) transfer function.

There are different kinds of ANN learning algorithms used at the summation stage between the layers such as probabilistic neural networks, generalized regression neural networks, and feed-forward back-propagation neural networks. The purpose of the back-propagation algorithm is to find the least minimum error surface, while it computes the local gradient in the error surface and afterwards updates the weights along the direction of the steepest local gradient (Adedigba et at. 2017). Each iteration in the back-propagation process, the forward and backward pass sweep, is performed repeatedly until the output variable is the same as the target value within an allowable predetermined tolerance level (Basheer and Hajmeer, 2000).

Generally, the data stratification of the ANN network is divided into three phases such as training, testing (calibration), and validation (verification). The data points can be divided as per the data points available as well as the experience of the users to attain the expected outcomes. The detailed procedure of feedforward backpropagation, as well as data processing of the ANN model, can be found in the literature (Ashena and Thonhauser, 2015; Adedigba et al. 2017). The generalized mathematical expression for the ANN model is given below:

$$y_m = f_o\left[\sum_{j=1}^{m} K_{jm} f_h\left(\sum_{i=1}^{n} K_{ij} x_i + b_j\right) + b_m\right] \qquad (2.1)$$

In equation (2.1), y_m stands for the output variables; x_i is the vector of target variables (e.g., i = 1, 2, 3, 4, …,n); b_m symbolizes the bias term for output layers; K_{ij} represents the connection weight on the link from i to j node between the input and hidden layers; m refers

to the number of hidden nodes; and *n* represents the number of input variables. f_h and f_o are the transfer function for the hidden layer and output layer, respectively.

A typical single neuron-based and multiple neurons structure is shown in Figure 2.1. According to Figure 2.1, the inputs are multiplied by the corresponding weight factors (W) and a bias term (b) is added as an error correction to obtain the summation in the network. This summation value is called the net value of the connectionist model. The net value from the input variables is passed through a transfer function (f) (Figure 2.1).

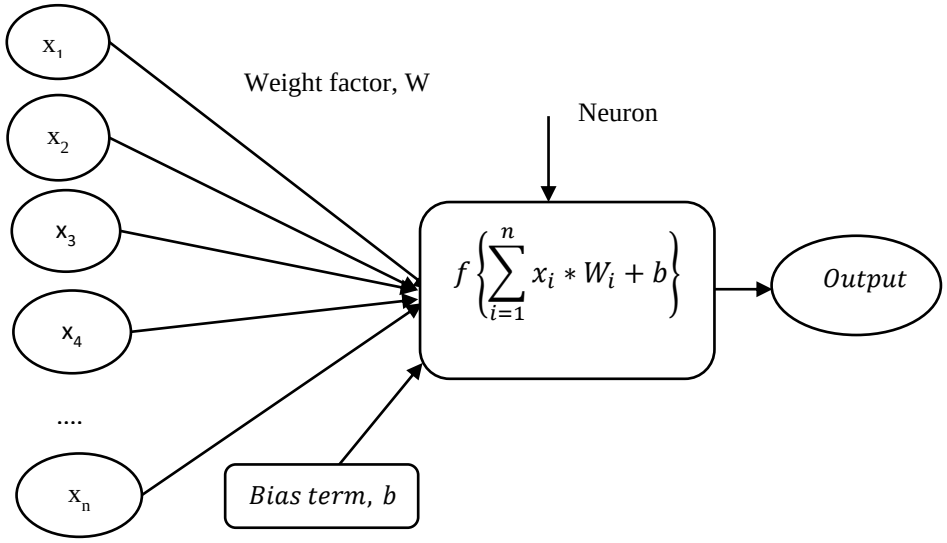

Figure 2.1: A graphical representation of a typical neuron-based network.

A good trained ANN would assign higher weight factors to be multiplied by the more important or stronger input variable values. Back propagation is one of the supervised training algorithms in which initially all weight factors are randomly optimized in the

algorithm, then the neural output variable is compared with the target variable in the training datasets and the error (the difference between target and neuron output) is propagated backward to the neural network. During this back-propagation, the weight factors are changed to decrease the error in the model. This procedure is repeated frequently until the produced output variables are acceptably close to the target value. In the connectionist model, training is performed by adjusting the weight factors until convergence between obtained outputs and desired outputs (target) is acquired (Figure 2.2).

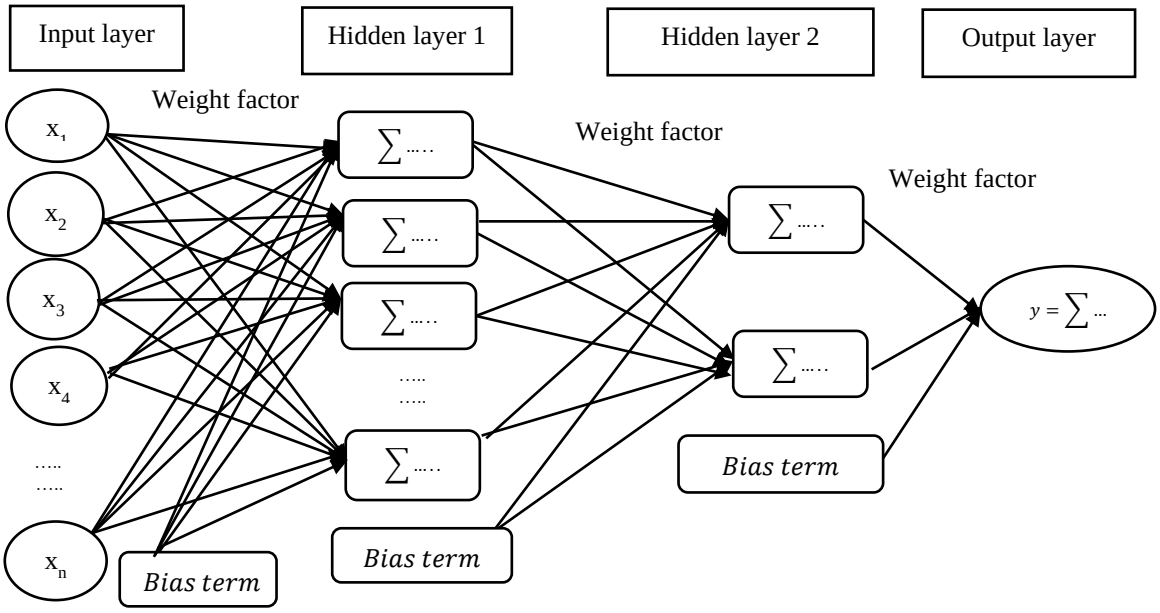

Figure 2.2: A graphical representation of an ANN structure with several neurons and two hidden layers.

The most common learning algorithms of backpropagation are Levenberg Marquardt (LM), Bayesian Regularization (BP), and Scaled Conjugated Gradient (SCG) to train the

MLP ANN network. The LM algorithm is faster with the least epochs (e.g., the number of iterations it takes for the ANN model to converge to a solution) and more reliable second-order nonlinear optimization training algorithm than the other standard back-propagation techniques (Ceryan et al. 2013). It represents a simplified version of the Newton's method (Marquardt, 1963); the steps in the training algorithm can be found in the literature (Hagan and Menhaj, 1994; Ceryan et al. 2013). When training with the LM optimization algorithm, the change of weights (ΔW) can be calculated using the following equation:

$$\Delta W = [J_k^T J_k + \mu_k I]^{-1} J_k^T e_k \quad (2.2)$$

Then, the weight factors can be updated using the following expression:

$$W_{k+1} = \Delta W + W_k \quad (2.3)$$

where μ (mu) is the Marquardt parameter in the training state of the network; $e_k$is the training error for each k step. I is an identity matrix and J is the Jacobian matrix.

2.3.2 Fundamentals of least square support vector machine

The primary concept of support vector machine (SVM) was first introduced by Vapnik (1995). The SVM is a supervised learning technique applicable to non-linear solutions for both classification and regression analysis. The special features of SVM are: a) ability to learn well with only a very small number of datasets/features, b) robustness against the error of models, and c) computational efficiency compared to other AI techniques such as ANN (Gholami and Fakhari, 2017). The major drawback of the SVM is that it needs to solve a large-scale quadratic programming problem (Suykens et al., 2002) which can be overcome by a modification to the classic SVM called the least-squares SVM (LS-SVM).

The LS-SVM can help to solve the complex solution with a more efficient way by setting up a linear set of equations employing SVM, instead of the quadratic programming problems to reduce the complexity of optimization process (Suykens and Vandewalle, 1999; Arabloo and Rafiee-Taghanaki, 2014). More information regarding SVM or LS-SVM principles and different features such as strength and weakness can be found in the literature (Suykens and Vandewalle, 1999; Smola et al., 2004; Ceryan, 2014; Gholami and Fakhari, 2017). The mathematical formulation of the LS-SVM is provided in Chapter 4. The final equation for the LS-SVM function estimation can be expressed as follows:

$$y = \sum_{i}^{n} \alpha_i K(x, x_k) + b \qquad (2.4)$$

where b and α are the solutions to the linear system; and K(x, x_k) represents the kernel function that should satisfy the Mercer's condition (Pelckmans et al., 2002). The weight factor (α) is a vector with a size of n×1.

There are many kernel functions such as linear, polynomial, radial basis, and sigmoidal kernel function employed for LS-SVM. The Gaussian radial basis kernel function has been widely used in learning strategy to attain the best output; it is computationally simpler than the other types of functions and it nonlinearly maps the training data into an infinite-dimensional space (Suykens et al., 2002; Ceryan, 2014). The simulated annealing (SA) is a technique among several other global optimization approaches developed to solve difficult non-convex problems. In order to improve performance of the model during the learning process, coupled simulated annealing (CSA) algorithm is adopted to optimize two tuning parameters, nemely, as the annealing term (gamma) and the kernel width (sigma),

which control the model accuracy and convergence, respectively (Xavier-de-Souza et al., 2009).

2.3.3 Applications of artificial intelligent for reservoir characterization

The AI application has become more widespread in different science and engineering fields including geosciences, and petroleum and mining engineering due to the ability to predict reasonable outputs for different important variables in the context of reservoir analysis. A typical workflow of AI-based connectionist approaches to predict rock properties using wireline log data is shown in Figure 2.3.

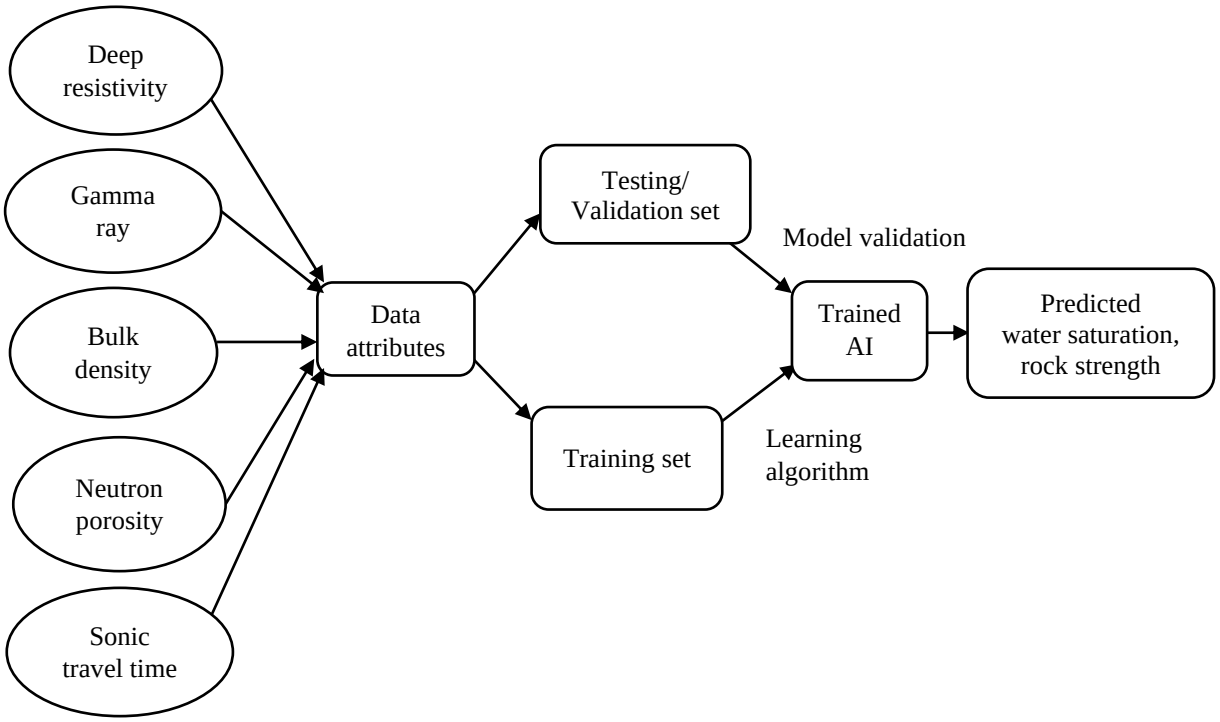

Figure 2.3: Application of AI technique for reservoir characterization

Several scholars have investigated the use of AI techniques, specially ANN and SVM (or LS-SVM) in reservoir engineering, reservoir characterization, and production engineering (Nikravesh and Aminzadeh, 2001; Mohaghegh, 2005; Rolon et al., 2009; Helmy et al., 2010; Anifowose, 2011; Wong et al., 2013; Zendehboudi et al., 2014; Anifowose et al., 2014; Esfahani et al., 2015; Attia et al., 2016; Anifowose et al., 2017; Tariq et al., 2017; Onalo, 2019; Ertekin and Sun, 2019). The most commonly used AI techniques are ANN and SVM (or LS-SVM), which can be employed to determine the reservoir rock porosity and permeability as well as hydraulic flow unit in reservoir characterization using core and wireline log data (Olson (1998), Aminian et al. (2003), Aggoun et al. (2006), Ahmadi et al. (2008), Anifowose et al. (2010), Moghadam et al. (2011), Gholami et al. (2012), Ali et al. (2013), Singh et al. (2016), Rafik and Kamel (2017), Elkatatny et al. (2018), and Eriavbe and Okene (2019)). The above mentioned research studied the performance of AI-based predictive models for various cases.

To date, AI techniques have also been used intensively to predict water (fluid) saturation and strength of rock using log datasets. A critical review on different research investigations, assumptions, and limitations of AI-based models for determination of water (fluid) saturation and strength of rock has been presented in chapters 3, 4, and 5. Also, the detailed model formulation and prediction procedure for ANN and LS-SVM-based connectionist models are illustrated systematically in chapters 3, 4, and 5.

References

Adedigba, S. A., Khan, F., & Yang, M. (2017). Dynamic failure analysis of process systems using neural networks. Process Safety and Environmental Protection, 111, 529-543.

Aggoun, R. C., Tiab, D., & Owayed, J. F. (2006). Characterization of flow units in shaly sand reservoirs—Hassi R'mel Oil Rim, Algeria. Journal of Petroleum Science and Engineering, 50(3-4), 211-226.

Ahmadi, M., Saemi, M., & Asghari, K. (2008). Estimation of the reservoir permeability by petrophysical information using intelligent systems. Petroleum Science and Technology, 26(14), 1656-1667.

Ali, S. S., Nizamuddin, S., Abdulraheem, A., Hassan, M. R., & Hossain, M. E. (2013). Hydraulic unit prediction using support vector machine. Journal of Petroleum Science and Engineering, 110, 243-252.

Aminian, K., Ameri, S., Oyerokun, A., & Thomas, B. (2003, January). Prediction of flow units and permeability using artificial neural networks. In SPE Western Regional/AAPG Pacific Section Joint Meeting. Society of Petroleum Engineers.

Anifowose, F. A. (2011, January). Artificial intelligence application in reservoir characterization and modeling: whitening the black Box. In SPE Saudi Arabia section Young Professionals Technical Symposium. Society of Petroleum Engineers.

Anifowose, F. A., & Abdulraheem, A. (2010, January). Prediction of porosity and permeability of oil and gas reservoirs using hybrid computational intelligence

models. In North Africa Technical Conference and Exhibition. Society of Petroleum Engineers.

Anifowose, F. A., Labadin, J., & Abdulraheem, A. (2014). Non-linear feature selection-based hybrid computational intelligence models for improved natural gas reservoir characterization. Journal of Natural Gas Science and Engineering, 21, 397-410.

Anifowose, F. A., Labadin, J., & Abdulraheem, A. (2017). Hybrid intelligent systems in petroleum reservoir characterization and modeling: the journey so far and the challenges ahead. Journal of Petroleum Exploration and Production Technology, 7(1), 251-263.

Arabloo, M., & Rafiee-Taghanaki, S. (2014). SVM modeling of the constant volume depletion (CVD) behavior of gas condensate reservoirs. Journal of Natural Gas Science and Engineering, 21, 1148-1155.

Archie, G.E., (1941). The electrical resistivity log as an aid in determining some reservoir characteristics. Transactions of AIME 146, 54–62.

Ashena, R., & Thonhauser, G. (2015). Application of Artificial Neural Networks in Geoscience and Petroleum Industry. In Artificial Intelligent Approaches in Petroleum Geosciences (pp. 127-166). Springer, Cham.

Asquith, G., & Krygowski, D., (2004). Basic well log analysis, American Association of Petroleum Geologists, Second Edition, Tulsa, Oklahoma, p. 216.

Attia, M., Mahmoud, M., Abdulraheem, A., & Abido, M. A. (2016, April 25). Prediction of the Gas Compressibility Factor Using Coefficient-Matrices Based on ANN. Society of Petroleum Engineers. doi:10.2118/182780-MS

Barati-Harooni, A., Najafi-Marghmaleki, A., Mohebbi, A., & Mohammadi, A. H. (2017). On the estimation of viscosities of Newtonian nanofluids. Journal of Molecular Liquids, 241, 1079-1090.

Barzegar, R., Sattarpour, M., Nikudel, M. R., & Moghaddam, A. A. (2016). Comparative evaluation of artificial intelligence models for prediction of uniaxial compressive strength of travertine rocks, case study: Azarshahr area, NW Iran. Modeling Earth Systems and Environment, 2(2), 76.

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. J. of microbiological methods, 43(1), 3-31.

Bassiouni, Z., (1994). Theory, Measurement and Interpretation of Well Logs, First Printing, Henry L. Doherty Memorial Fund of AIME, SPE, Richardson, p. 372.

Ceryan, N. (2014). Application of support vector machines and relevance vector machines in predict- ing uniaxial compressive strength of volcanic rocks. Journal of African Earth Sciences, 100, 634–644.

Chang, C., Zoback, M. D., & Khaksar, A. (2006). Empirical relations between rock strength and physical properties in sedimentary rocks. Journal of Petroleum Science Engineering, 51(3-4), 223–237. doi:10.1016/j.petrol.2006.01.003

Chatterjee, R., Paul, S., & Mourya, V. K. (2013). Prediction of Uniaxial Compressive Strength from well log data in Jharia Coalfield. Department of Applied Geophysics, Indian School of Mines, Dhanbad–826, 4.

Cover, T. M., & Thomas, J. A. (2012). Elements of information theory. John Wiley & Sons.

Elkatatny, S., Mahmoud, M., Tariq, Z., & Abdulraheem, A. (2018). New insights into the prediction of heterogeneous carbonate reservoir permeability from well logs using artificial intelligence network. Neural Computing & Applications, 30(9), 2673-2683.

Eriavbe, F. E., & Okene, U. O. (2019, August). Machine Learning Application to Permeability Prediction Using Log & Core Measurements: A Realistic Workflow Application for Reservoir Characterization. In SPE Nigeria Annual International Conference and Exhibition. Society of Petroleum Engineers.

Ertekin, T., & Sun, Q. (2019). Artificial Intelligence Applications in Reservoir Engineering: A Status Check. Energies, 12(15), 2897.

Esfahani, S., Baselizadeh, S., & Hemmati-Sarapardeh, A. (2015). On determination of natural gas density: least square support vector machine modeling approach. Journal of Natural Gas Science and Engineering, 22, 348-358.

Gąsior, I., & Przelaskowska, A. (2014). Estimating thermal conductivity from core and well log data. Acta Geophysica, 62(4), 785-801.

Gholami, R., & Fakhari, N. (2017). Support Vector Machine: Principles, Parameters, and Applications. In Handbook of Neural Computation (pp. 515-535). Academic Press.

Gholami, R., Shahraki, A. R., & Jamali Paghaleh, M. (2012). Prediction of hydrocarbon reservoirs permeability using support vector machine. Mathematical Problems in Engineering, 2012.

Green, D. W., & Willhite, G. P. (1998). Enhanced Oil Recovery, SPE textbook series. Society of Petroleum Engineers, Richardson, Texas.

Hagan, M. T., & Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. IEEE transactions on Neural Networks, 5(6), 989-993.

Ham, F. M., & Kostanic, I. (2000). Principles of neurocomputing for science and engineering. McGraw-Hill Higher Education.

Helmy, T., Fatai, A., & Faisal, K. (2010). Hybrid computational models for the characterization of oil and gas reservoirs. Expert Systems with Applications, 37(7), 5353-5363.

Jang, J. S. (1993). ANFIS: adaptive-network-based fuzzy inference system. IEEE transactions on systems, man, and cybernetics, 23(3), 665-685.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. science, 220(4598), 671-680.

Lucia, F. J. (2007). Carbonate reservoir characterization: An integrated approach. Springer Science & Business Media. Pp. 223.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. Journal of the society for Industrial and Applied Mathematics, 11(2), 431-441.

Masoudi, P., Aïfa, T., Memarian, H., & Tokhmechi, B. (2017). Uncertainty assessment of volumes of investigation to enhance the vertical resolution of well-logs. Journal of Petroleum Science and Engineering 154, pp. 252-276.

Miah, M. I., Deb, P. K., Rahman, M., & Hossain, M. E. (2017). Application of Memory Concept on Petroleum Reservoir Characterization: A Critical Review. In SPE Kuwait Oil & Gas Show and Conference. Society of Petroleum Engineers.

Moghadam, J. N., Salahshoor, K., & Kharrat, R. (2011). Intelligent Prediction of Porosity and Permeability from Well Logs for An Iranian Fractured Carbonate Reservoir. Petroleum Science and Technology, 29(20), 2095-2112.

Mohaghegh, S. D. (2005). Recent developments in application of artificial intelligence in petroleum engineering. Journal of Petroleum Technology, 57(04), 86-91.

Moos, D., Peska, P., Finkbeiner, T., & Zoback, M. (2003). Comprehensive wellbore stability analysis utilizing quantitative risk assessment. Journal of Petroleum Science and Engineering, 38(3-4), 97-109.

Nikravesh, M., & Aminzadeh, F. (2001). Past, present and future intelligent reservoir characterization trends. Journal of Petroleum Sci. and Engineering, 31(2-4), 67-79.

Odunlami, T., Soroush, H., Kalathingal, P., & Somerville, J. (2011). Log-Based Rock Property Evaluation-A New Capability in a Specialized Log Data Management Platform. In SPE/DGS Saudi Arabia Section Technical Symposium and Exhibition.

Olson, T. M. (1998, January 1). Porosity and Permeability Prediction in Low-Permeability Gas Reservoirs From Well Logs Using Neura Networks. Society of Petroleum Engineers. doi:10.2118/39964-MS.

Pascoal, C., Oliveira, M. R., Pacheco, A., & Valadas, R. (2017). Theoretical evaluation of feature selection methods based on MI. Neurocomputing, 226, 168-181.

Pelckmans, K., Suykens, J.A., Van Gestel, T., De Brabanter, J., Lukas, L., Hamers, B., De Moor, B. & Vandewalle, J. (2002). LS-SVMlab: a matlab/c toolbox for least squares support vector machines. Tutorial. KULeuven-ESAT. Belgium, 142.

Rabbani, E., Sharif, F., Salooki, M. K., & Moradzadeh, A. (2012). Application of neural network technique for prediction of uniaxial compressive strength using reservoir formation properties. International Journal of Rock Mechanics and Mining Sciences, (56), 100-111.

Rafik, B., & Kamel, B. (2017). Prediction of permeability and porosity from well log data using the nonparametric regression with multivariate analysis and neural network, Hassi R'Mel Field, Algeria. Egyptian Journal of Petroleum, 26(3), 763-778.

Rolon, L., Mohaghegh, S. D., Ameri, S., Gaskari, R., & McDaniel, B. (2009). Using artificial neural networks to generate synthetic well logs. Journal of Natural Gas Science and Engineering, 1(4-5), 118-133.

Sebtosheikh, M. A., Motafakkerfard, R., Riahi, M. A., Moradi, S., & Sabety, N. (2015). Support vector machine method, a new technique for lithology prediction in an Iranian heterogeneous carbonate reservoir using petrophysical well logs. Carbonates and Evaporites, 30(1), 59-68.

Serrra, O. (1984). Fundamentals of Well-Log Interpretation. 1. The Acquisition of Logging Data. Elsevier, New York.

Shannon, C.E., (1948). The mathematical theory of communication (parts 1 and 2). Bell Syst. Tech J. vol. 27, pp. 379-423 and 623-656, 1948.

Singh, S., Kanli, A. I., & Sevgen, S. (2016). A general approach for porosity estimation using artificial neural network method: a case study from Kansas gas field. Studia Geophysica et Geodaetica, 60(1), 130-140.

Smola, A.J., Sch, B., & lkopf (2004). A tutorial on support vector regression. Stat. Comput. 14, 199–222.

Suykens, J. & Vandewalle, J. (1999). Neural Processing Letters 9, 293. https://doi.org/10.1023/A:1018628609742.

Suykens, J.A.K., Van Gestel, T., Brabanter, J., De Moor, B., & Vandewalle, J. (2002). Least Squares Support Vector Machines. World Scientific, Singapore.

Tariq, Z., Elkatatny, S. M., Mahmoud, M. A., Abdulraheem, A., Abdelwahab, A. Z., & Woldeamanuel, M. (2017). Estimation of Rock Mechanical Parameters Using Artificial Intelligence Tools. American Rock Mechanics Association.

Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer, New York.

Wong, P., Aminzadeh, F., & Nikravesh, M. (Eds.). (2013). Soft computing for reservoir characterization and modeling (Vol. 80). Physica, Heidelberg. Pp. 587.

Wyllie, M. R. J., Gregory, A. R., & Gardner, L. W. (1956). Elastic wave velocities in heterogeneous and porous media. Geophysics, 21(1), 41-70.

Xavier-de-Souza, S., Suykens, J. A., Vandewalle, J., & Bollé, D. (2009). Coupled simulated annealing. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 40(2), 320-335.

Zendehboudi, S., Shafiei, A., Bahadori, A., James, L. A., Elkamel, A., & Lohi, A. (2014). Asphaltene precipitation and deposition in oil reservoirs–Technical aspects, experimental and hybrid neural network predictive tools. Chemical Engineering Research and Design, 92(5), 857-875.

Zhang, H. M. (2003). Driver memory, traffic viscosity and a viscous vehicular traffic flow model. Transportation Research Part B: Methodological, 37(1), 27-41.

CHAPTER 3: CONNECTIONIST AND MUTUAL INFORMATION TOOLS TO DETERMINE WATER SATURATION AND RANK INPUT LOG VARIABLES

Preface

A version of this chapter has been published in the Journal of Petroleum Science and Engineering (106741). I am the primary author of this article and prepared the first draft of the manuscript. The co-author, Dr. Salim Ahmed, provided technical assistance to find the research scopes, as well as expoloring the knowledge gap. Dr. Ahmed supervised for data analysis and review of the first draft of the manuscript. Co-author Dr. Sohrab Zendehboudi also reviewed and provided valuable insights on how to improve the model development and result analysis subsections in the manuscript to a journal-level publication. I have revised the manuscript based on the feedback from co-authors. Finally, all authors contributed to prepare the final version of the manuscript as per the reviewers' and editor's comments.

Abstract

Characterization of petroleum reservoirs plays an important role to effectively manage and forecast the recovery performance. A number of subset log variables such as gamma-ray, resistivity, density, neutron, and sonic porosity logs are generally used to characterize/predict the reservoir properties. The data attributes selection and ranking in reservoir characterization are vital to determine the output variables with the best performance and cost-effective manner during exploration and production operations. The objectives of this research work are to estimate the water saturation in the reservoir with an acceptable accuracy and to rank the log variables according to their importance. To achieve the objectives, the mutual information (MI) and artificial neural network (ANN) techniques are implemented with the non-linear predictors using log variables. The feed-forward ANN model is employed and optimized to predict the water saturation, where the Levenberg-Marquardt algorithm is used for the network training. There is a good match between the real data and predictions so that the regression coefficient and the maximum error is 99.98% and 5.55%, respectively. In addition, both ANN and MI approaches lead to the same ranking levels of log variables, implying high accuracy and reliability of the introduced strategy. It is found that the primary (or most important) log variables are true resistivity and bulk density to obtain the pore fluid saturation. The approach suggested in this study (connectionist and MI strategies) can assist engineers/operators to run a few numbers of logging tools for prediction of water saturation to save the exploration costs through a timely manner. In addition, further understanding is attained to conduct proper data selection for determination of petrophysical properties.

Keywords: Reservoir characterization; Well logs; Prediction tools; Information theory; Variables ranking

3.1 Introduction

Reservoir characterization involves various approaches to describe the reservoir rock and fluid properties/behaviors at different process and thermodynamic conditions. There are several conventional and advanced methods (and technologies) to properly characterize underground formations such as sandstones, shales, and carbonates. The reservoir evaluation/characterization is a continuous process that includes integrated tasks (e.g., logging and well testing). In the petroleum industry, the most common techniques for reservoir characterization are geophysics (Close and Caycedo, 2011), petrophysics and well-logging (Luthi, 2001; Worthington, 1985; Asquith and Krygowski, 2004; Al-Ruwaili and Al-Waheed, 2004; Aggoun et al., 2006; Lucia, 2007; Ellis and Singer, 2007; Tiab and Donaldson, 2011, Yang and Wei, 2017, Miah et al., 2017; Baouche et al., 2017), geostatistics (Azevedo, and Soares, 2017), reservoir modeling (Dubois et al., 2006), well testing (Sylvester et al., 2015), and soft computing tools such as artificial neural networks (ANN), fuzzy logic (FL), and evolutionary computing methods (Cuddy, 2000; Wong et al., 2002; Masoud, 2004; Wang et al., 2013; Ali et al., 2013; Hakiki and Wibowo, 2014). Wireline log is considered as one of the most widely used methods for reservoir characterization in the oil and gas industry (Onalo et al., 2018). It is essential for petroleum/ reservoir engineers and geologists to obtain detailed information about the type, characteristics, and quality of reservoirs through estimation of a variety of rock properties

(e.g., rock resistivity, shale content, porosity, permeability, and pore fluid saturation). These properties are not the same or uniform in hydrocarbon formations throughout the world due to the various heterogeneities in structure, permeability, porosity, and wettability. In addition, the wireline log is useful to detect water and hydrocarbon-bearing zones and to evaluate the shale (clay) content and hydrocarbon volume. The reservoir quality can be also assessed by making various strategies such as the field development, economic analysis, and reservoir management plan.

Several water saturation models were developed based on reservoir behaviors (Shedid and Saad, 2017). The mostly used petrophysical model in the petroleum industry to estimate the water (or hydrocarbon) saturation includes the Archie's equation for clean sandstone formations (Archie, 1941) and shaly sand formations (Poupon et al., 1954; Simandoux, 1963; Schlumberger, 1972; Clavier et al., 1984; Worthington, 1985). The empirical models may have some limitations due to the nature and type of reservoir formations (Waxman and Smits, 1968; Ipek, 2002; Al-Ruwaili and Al-Waheed, 2004; Shedid and Saad, 2017). For instance, the models may lead to either overestimation or underestimation of the volume of water and hydrocarbon in the reservoir.

On the other hand, handing the big data of well logging seems crucial and challenging in different activities attributed to the oil industry such as reservoir characterization (Shedid, 2018; Shedid-Elgaghah et. al., 2001) and selection of proper production scenarios and/or recovery methods (Khamidy et al., 2019). The smart connectionist models such as artificial neural networks (ANN), fuzzy logic (FL), and least square support vector machine (LS-SVM) have shown a great potential to predict the reservoir properties using the data

obtained from different well logs, as seen in research works of several researchers including Mohaghegh et al., 1995-1996; Mohaghegh, 2000a-c; Helle et al., 2001; Eidsvik et al., 2004; Aminian and Ameri, 2005; Al-Bulushi et al., 2007; Olatunji et al., 2011; El-Sebakhy et al., 2012; Mollajan et al., 2013; Aïfa et al., 2014; Mollajan, 2015; Baziar et al., 2018; Zendehboudi et al, 2018; and Alkinani et al., 2019. The connectionist techniques can explore the nonlinear relationships between the wireline log data, generate suitable correlations, and eventually forecast the reservoir properties such as porosity, permeability, and water saturation (Mohaghegh et al., 2000a-d; Banchs and Michelena, 2002; Al-Bulushi et al., 2007; Rolon et al., 2009).

Also, the information theoretic measure approaches such as entropy and mutual information (MI) have been used to capture the dependencies among the random variables of the model using available big databanks and feature selection methods (Cover and Tomas, 2006; Vergara, J.R. & Estévez, 2014; Pascoal et al., 2017). Most of the feature selection methods are applied using conventional statistical methods such as linear discriminant analysis, and univariate and multivariate statistical analysis (Biu et al., 2010; Graf et al., 2011). There are only a limited number of research studies in the open sources that deal with the natural non-linear phenomena and feature selection for improved petroleum recovery and characterization of oil and gas reservoirs (Anifowose et al., 2014; 2015; 2016; 2018). In other words, the applications of MI concept in the oil and gas industry are currently limited. For instance, Abellan and Noetinger (2010) employed the information theory to find the optimum well placement. Le and Reynolds (2012) investigated one-dimensional water flooding case and proposed a model for translating the

magnitude of mutual information to the expected values of probability levels of 90 and 10. After that, Le and Reynolds (2014) extended the same approach to simulate two dimensional and three-dimensional water flooding cases. The most influencing parameters were not identified in their model.

Helle and Bhatt (2002) employed an ANN tool to predict fluid saturation using log variables. It was concluded that the ANN offers a greater precision, compared to the petrophysical models. In their study, the gamma-ray (GR) log to identify/characterize the shaly sand reservoir was not considered. Shokir (2004) implemented ANN modeling to predict the hydrocarbon saturation of low resistivity shaly sand reservoirs using log data, showing high potential of ANN. For instance, the regression coefficient was obtained to be close to one. The feature ranking was not investigated in their research study. Al-Bulushi et al. (2009) utilized an ANN model to obtain the saturation degree. It was found that the true resistivity has the highest contribution (e.g., 40 %) to water saturation. They did not take into account GR as an input log variable to figure out the shale effect on the magnitude of water saturation. Kamalyar et al. (2011) also constructed an ANN model using core data where no field log data was used in the model development. Based on their study, the ANN model offers more reliable and accurate predictions, compared to available theoretical and empirical models. Mardi et al. (2012) proposed an ANN model based on both core and wireline log data to determine the cementation factor and saturation exponent of the water saturation model. In the study, the researchers implemented an ANN model to estimate water saturation in Sarvak limestone formation of an Iranian oilfield, Azadegan. It was concluded that ANN is able to more accurately estimate the target variable, compared to

the dual water model. The relative importance of the input parameters was not reported in their work. Baneshi et al. (2013) conducted a parametric sensitivity analysis on the log variables while predicting reservoir parameters using an ANN model with the Levenberg-Marquardt training function. They developed the ANN model using the available dataset from Iranian onshore hydrocarbon basin to obtain porosity index and water saturation. It was found that ANN is a suitable tool to forecast reservoir porosity, saturation degrees, and petrophysical index using true resistivity, density, and neutron porosity logs.

Additionally, Anifowose et al. (2014) used the information gain concept to determine the relative significance of feature attributes for the decision tree approach. Based on their research findings, the grain density and grain volume are the important input variables for porosity estimation, while porosity log, density log, water saturation, and micro-spherically focused log are good data attributes to determine reservoir permeability. Water saturation was not estimated using input data in their work. Gholanlo et al. (2016) also employed a radial-based kernel function in the ANN model to obtain water saturation using log variables. In their model, the Levenberg-Marquardt (LM) algorithm was used to optimize the weight factor between neurons of the layer. Hamada et al. (2018) developed an ANN approach where the activation function was the tans- and log sigmoid function; and two hidden layers and one output layer with 16 and 5 neurons, respectively, were utilized. They estimated the porosity and water saturation using the core data. The contributions of the input variables were not discussed in their work. Khan et al. (2018) claimed that the adaptive neuro-fuzzy inference system gives better results compared to the ANN approach while predicting water saturation. However, the sensitivity analysis and relative

performance of variables are missing in their research investigation. In addition, Hamada et al. (2019) determined the porosity and water saturation with ANN and conventional petrophysical approaches using wireline log data. It was found that the ANN model has a greater deterministic performance than conventional petrophysical models for the shaly sand reservoir. In their study, the input variables ranking was not conducted.

Multimin is a statistical technique (on the basis of probability concept) for estimating properties of fluids and minerals in petroleum reservoirs using petrophysical data including cores and logs (Chabock et al., 2017; Kumar et al., 2018, Seyyedattar, 2019a, 2019b; Zendehboudi et al., 2018). This approach is probabilistic that possess some inherent randomness. Thus, using the same input data and variables might lead to a different set of outputs at various runs. However, ANN is a deterministic model in which the output is fully obtained by the input data and parameters (Seyyedattar, 2019a, 2019b; Zendehboudi et al., 2018). To accurately predict a target variable, more input data and knowledge about the problem physics are normally needed while using probabilistic methods such as Multimin. In addition, it is more difficult to implement Multimin, compared to ANN. Highlighting other limitations of Multimin, the dependencies between variables are not transparent and well defined in this strategy. Link between statistical realizations and the real world is also unclear; Multimin might cause considerable errors if the input ranges and algorithms are not proper (Chabock et al., 2017; Kumar et al., 2018, Seyyedattar, 2019a, 2019b; Zendehboudi et al., 2018). It is not easy and straightforward to conduct parameter ranking using Multimin. It has been also proven that ANN is more cost effective and simpler, compared to Multimin (Seyyedattar, 2019a, 2019b; Zendehboudi et al., 2018).

Based on the current research interest, we employ new tools such as mutual information (MI) and artificial neural network (ANN) to obtain the water saturation with the aid of log data. The novelty of this manuscript is utilization of MI and ANN approaches for the first time to forecast the water saturation in shale sand formation and to rank the input parameters.

To the best of our knowledge, ranking the logging variables and reservoir characterization using the extended mutual information and non-linear regression and/or artificial intelligence models have not been studied systematically. The attribute selection and ranking of logging subset data appear to be a serious challenge for oil and gas engineers and geologists to determine the reservoir properties using the logging data with sufficient precision. The logging data is not easy to collect and represent the entire reservoir due to the complex nature of geological structures and reservoir locations. The current research work is planned to fill in the knowledge gap by finding proper relationships between logging parameters for the reservoir characterization where effective predictive strategies such as mutual information and computational intelligence techniques with the aid of real-life log data information are employed. It is believed that the methods introduced in this work can considerably lower the costs and time for efficient reservoir characterization. The objectives of this research are listed as follows:

- To obtain the water saturation using logging data and deterministic tools
- To investigate the dependency of predictor subset variables in developed model, and
- To rank the log variables according to their importance while predicting water saturation

This paper is organized as follows: After the introduction part, Section 3.2 presents the theoretical concepts and background regarding the petrophysical models and development of ANN and MI techniques. Section 3.3 briefly covers the methodology for modeling implementation. The data collection and processing are included in Section 3.4. The results and discussions on important points/trends of this study are described in Section 3.5. The conclusions and recommendations are highlighted in the last section.

3.2 Theory and Background

The sources of log variables and log interpretation steps for estimation of water saturation are illustrated in this section. In addition, a brief theory on the fundamentals and structure of ANN and MI is covered.

3.2.1 Pore fluid prediction using log data

Hydrocarbon saturation is one of the important petrophysical parameters that is needed for reservoir characterization as well as estimation of hydrocarbon volume. Thus, accurate prediction of fluid saturation is vital in terms of reservoir analysis. The common logs such as lithology log, porosity log, and resistivity log are used to identify the rock lithology and to estimate the reservoir thickness and petrophysical properties such as porosity and water saturation. More information about the common types of logging tools and uncertainties of well logging data can be found in the literature (Crain, 1986; Bassiouni, 1994; Asquith and Krygowski, 2004; Masoudi et al., 2017).

The gamma-ray (GR) log measures the strength of the natural radioactivity present in the formation. The presence of clay minerals in a formation may considerably affect the reservoir petrophysical properties. The GR log is extensively used to obtain the shaliness (clay content) that appreciably affects the predicted values of effective porosity (ϕ_e) and water saturation (Sw) in shaly sand reservoirs. In this study, the shale index (I_{GR}) is calculated by the following equation (Schlumberger, 1998):

$$I_{GR} = \frac{GR_{log} - GR_{min}}{GR_{max} - GR_{min}} \quad (3.1)$$

where GR_{log} stands for the gamma-ray value of the zone of interest; and GR_{min} and GR_{max} represent the minimum and maximum values of gamma-ray log over the entire log, respectively.

The shale volume (V_{sh}) for the tertiary rocks of non-linear response is obtained as follows *(*Larionov, 1969)*:*

$$V_{sh} = 0.083(2^{3.7\, I_{GR}} - 1) \quad (3.2)$$

The resistivity (inverse of conductivity) of a formation can be measured from the laterolog or induction type logging tool. It is a crucial parameter in determining water saturation, which depends on the formation water resistivity (fully saturated water case), pore geometry, and amount of fluids (water and/or hydrocarbon) in the reserve. The deep induction log measures the true resistivity (Rt, ohm-m) of the virgin formation. The formation water resistivity (R_w) can be estimated using the inverse Archie's formula (Archie's, 1941), which is shown below:

$$R_w = \frac{\phi_e^m}{a} \times R_t \quad (3.3)$$

where *a* and *m* denote the tortuosity and cementation exponent, respectively. The procedure proposed by Miah and Howlader (2012) is followed to estimate R_{wa} using the inverse Archie's analysis.

Rock porosity measures the void space in the reservoir, which is available for accumulation/ storage of fluids. Porosity changes slightly laterally and vertically within reservoirs. Also, the porosity is affected by geomechanical reservoir conditions such as effective stress (Hakiki and Shidqi, 2018). Santamarina et al. (2019a) investigated the anisotropic behaviors to estimate the hydraulic conductivity (intrinsic permeability) of the porous medium. The variations of porosity (void space of total volume) and the specific surface area also affect the permeability measurements. This property can be determined from the responses of sonic log (estimation of sonic transit time), density log (bulk electron density), and neutron log (a measure of the hydrogen concentration) in a formation. In addition, porosity can be estimated using the advanced technology of nuclear magnetic resonance (NMR) and dielectric logging tools (Santamarina et al., 2019b). Porosity in sandstones normally varies with grain size distribution, grain shape, packing arrangement, cementation, and clay content (Asquith and Krygowski, 2004). The density porosity (ϕ_D) can be calculated using the bulk density response (ρ_b) by the following equation:

$$\phi_D = \frac{\rho_{ma} - \rho_b}{\rho_{ma} - \rho_{fl}} \tag{3.4}$$

In Equation (3.4), ρ_{ma} and ρ_{fl} denote the matrix density of the rock and fluid density of the drilling mud, respectively.

The effective density porosity ($\phi_{D,e}$) and effective neutron porosity ($\phi_{N,e}$) with clay correction are obtained as follows (Miah, 2014):

$$\rho_{b.c} = \rho_b + V_{sh}\,(\rho_{ma}\text{-}\rho_{cl}) \tag{3.5}$$

$$\phi_{D,e} = \frac{\rho_{ma}\text{-}\,\rho_{b.c}}{\rho_{ma}\text{-}\,\rho_{fl}} \tag{3.6}$$

$$\phi_{N,e} = \mathrm{PHIN} - V_{sh} * \phi_{N,sh} \tag{3.7}$$

in which, $\rho_{b.c}$ represents the corrected bulk density (g/cm^3); PHIN introduces the neutron porosity obtained from the neutron log; and ρ_{cl} and $\phi_{N,sh}$ stand for the bulk density (g/cm^3) and neutron porosity (%) of the shale zone.

Combining the effective neutron and density porosities, the effective porosity (ratio of the interconnected pore space to the total bulk volume of the rock, ϕ_e or PHIND) for gas reservoirs is estimated using the following expression:

$$\phi_e = \sqrt{\frac{{\phi_{D,e}}^2 + {\phi_{N,e}}^2}{2}} \tag{3.8}$$

Based on a research study on shaly sand reservoirs, the following model (Indonesian model) can be employed to predict the water saturation (S_w) (Hamada, 1996):

$$S_w = \frac{1}{R_t}\left[\frac{V_{sh}^{(1-0.5V_{sh})}}{\sqrt{R_{sh}}} + \frac{{\phi_e}^{0.5m}}{\sqrt{aR_w}}\right] \tag{3.9}$$

where R_{sh} refers to the shale resistivity of the shale zone of the formation. The generalized methodology of log data interpretation is depicted in Figure 3.1. Figure 3.2 represents the main sources to obtain the reservoir petrophysical parameters and eventually to predict the pore fluid (water) saturation.

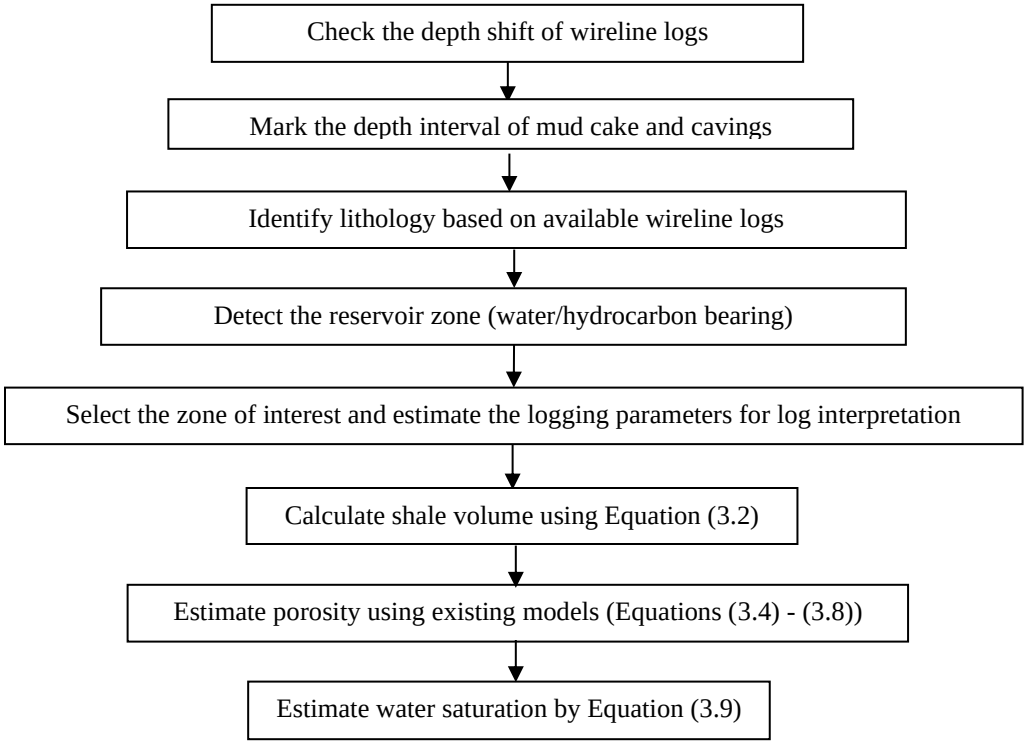

Figure 3.1: A simple procedure to conduct log interpretation for prediction of pore water saturation.

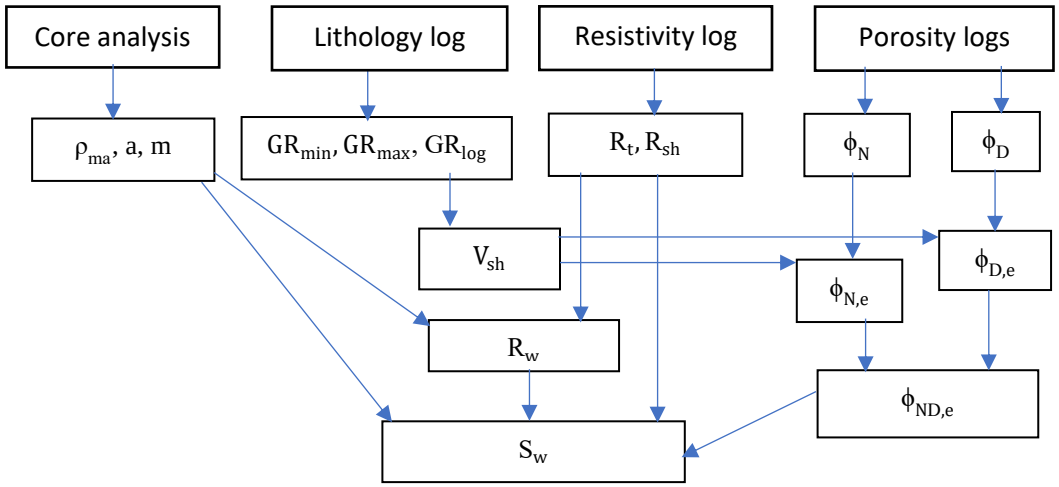

Figure 3.2: Graphical representation of log variables to estimate the pore fluid saturation.

3.2.2 ANN model

The ANN model is an information system based on the adaptive learning and parallel computing scheme, which was originated from the concept/mechanism of the biological neural network (Lippmann, 1987; Mhaskar, 1993; Faucett, 1994; Mohaghegh et al., 1995-1996; Yao and Liu, 1997; Baev, 1998; Poulton, 2002). This strategy has the ability to find highly complex nonlinear relationships between variables. Faucett (1994) listed the main assumptions and/or limitations for ANN mathematical models. For instance, the ANN model does not provide detailed physics and/or knowledge on the process and mechanisms of cases under study.

The main goal in implementation of ANN is to correlate the objective(s) to input parameters where it is targeted to minimize differences between the real output values and predictions obtained from the model (Yao et al., 2005). Different learning algorithms such as back-propagation, genetic algorithm, artificial bee colony (ABC), and imperialist competitive (ICA) can be utilized while applying the ANN approach, even when the input information is less defined and/or noisy (Ali, 1994).

The ANN can be classified by feeding direction of the input data such as feed backward, and feedforward which is generally used in many engineering cases/problems (White et al., 1995; Benardos and Vosniakos, 2007; Razavi and Tolson, 2011; Ahmadi et al., 2013; Talebi et al., 2014; Masoudi et al., 2014; Ghaffarian et al., 2014; Akande et al., 2015; Adedigba et at., 2017). Feedforward ANNs can be utilized as single-layer perceptron, multilayer perceptron (MLP), and radial basis function (RBF) neural networks (Gardner

and Dorling, 1998; Razavi and Tolson, 2011). The data in the neural network are divided into three main subsets such as training, validation, and testing subcategories (Al-Bulushi et al., 2012).

A set of available data is considered as the input data, inserted into the input layer and then propagated through the network to offer the model predictions as the outputs. The mathematical expression of this operation can be shown using the following equation (Jorjani et al., 2008) for each neuron (in either hidden or output layer), which primarily acts as a summing junction for combination/modification of the inputs from the previous layer:

$$Y_i = \sum_{j=1}^{n} X_i W_{ij} + b_j \qquad (3.10)$$

where Y_i refers to the net input to neuron j in the hidden or output layer; X_i stands for the output of previous layer or inputs to neuron j; W_{ij} is the connected weight between the i-th neuron and j-th neuron; and b_j represents the bias associated with neuron j.

The typical activation (transfer) functions are the threshold, piecewise linear, logistic sigmoid, and sigmoid hyperbolic tangent function for the neurons of the model (Ashena and Thonhauser, 2015). In most engineering applications, the sigmoid function can be utilized (Amiri et al., 2015). The neural network workflow and main involved steps were illustrated by several researchers such as Al-Bulushi et al. (2012) and Onalo et al. (2018). The important feature of a data-driven prediction model is to build an appropriate model that is capable of solving the linear and non-linear problems with a high reliability and precision. The vital steps of the ANN workflow are: a) data acquisition and quality

checking b) selection of the input and target variables, c) data stratification, d) selection of a training algorithm, e) selection of an optimum structure (number of hidden layer and neurons), f) weight optimization, and g) model performance evaluation. In general, the training process of an ANN model involves the selection of network structure, numbers of layers and neurons, and type of transfer function.

3.2.3 Information theories measure approach

The information theory was first developed by Shannon (1948) to measure the mutual information in the field of signal and image processing. Mutual information (MI) is one of the useful concepts/approaches in communication engineering subject, which offers a procedure to measure the strength of the relationship between discrete and/or continuous variables. The most common techniques of information theories are entropy (H) and mutual information. Further background/theory on the information theory can be found in several research works such as Vergara and Estévez (2014), Cover and Thomas (2006). This section briefly presents the fundamental theories of entropy and mutual information (see Figure 3.3). The definition of entropy is a measure of uncertainty of a random variable such as X and Y, as shown below:

$$\text{Entropy of } X \text{ variable}, H(X) = -\sum_{x \in X} p(x) log\ p(x) \tag{3.11}$$

$$\text{Entropy of } Y \text{ variable}, H(Y) = -\sum_{y \in Y} p(y) log\ p(y) \tag{3.12}$$

$$\text{Conditional entropy}, H(X \mid Y) = -\sum_{i=1}^{n} \sum_{j=1}^{n} p(x_i, y_j) \log p\left(\frac{x_i}{y_j}\right) \tag{3.13}$$

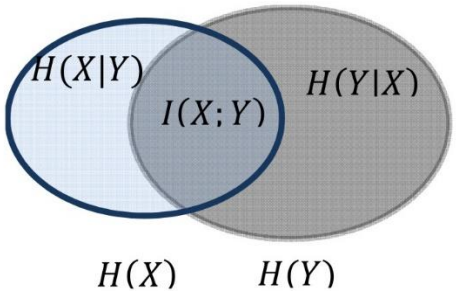

Figure 3.3: Venn diagram to show relationship between H and MI.

The MI is a measure of the amount of dependency information between the random variables. Mathematically, the MI can be expressed for the discrete random variables as follows:

$$MI = I(X;Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\left\{\frac{p(x,y)}{p(x)p(y)}\right\} \quad (3.14)$$

3.3 Research Methodologies

This section describes the prediction procedure using ANN tool as well as ranking of input variables through utilizing both ANN and MI approaches.

3.3.1 Prediction of water saturation using ANN model

The current study considers five well log parameters including gamma ray, true resistivity, density log, neutron porosity, and sonic log as input variables and the water saturation is used as the target variable. The ANN modeling is conducted in the Matlab® environment. A simple schematic of ANN structure used in this research work is illustrated in Figure 3.4. N 1 in Figure 3.4 represents the 1st neuron of hidden layer for the ANN feedforward backpropagation system. The main features listed in Table 3.1 are used to optimize the supervised learning model of ANN.

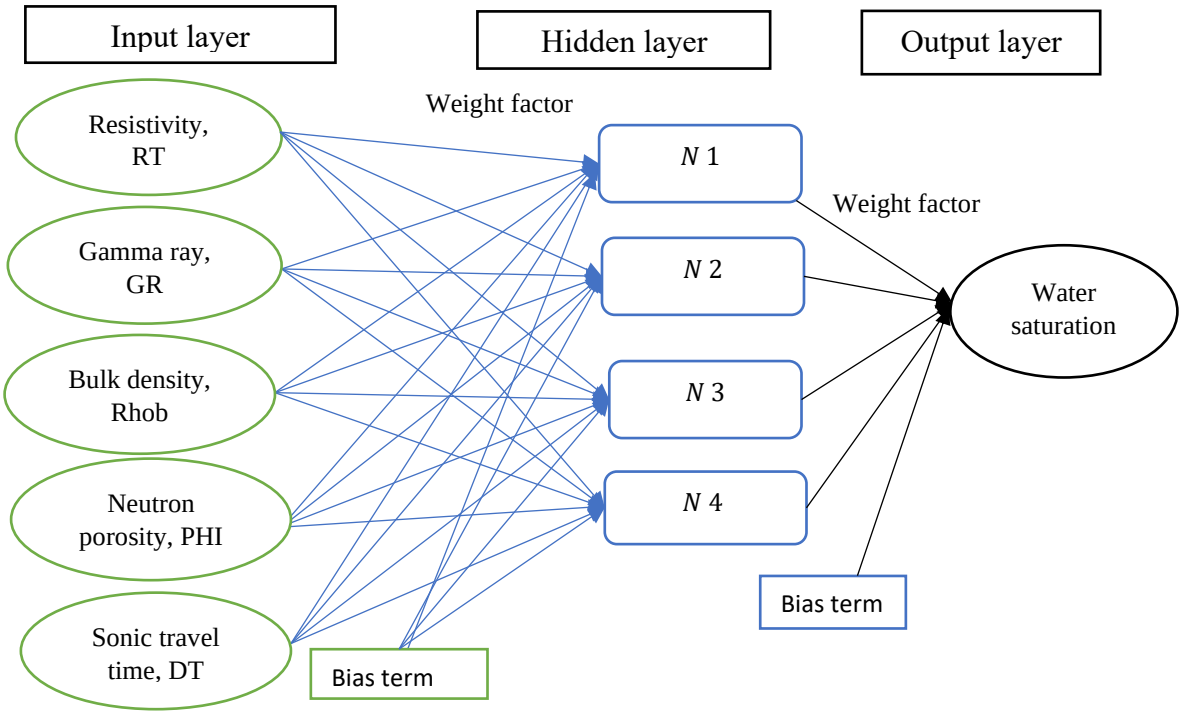

Figure 3.4: Schematic of a single hidden layer-based ANN architecture employed in the current work.

Table 3.1: Vital features of the supervised learning ANN approach.

Model features	Parameters
Input data	RT, RHOB, NPHI, GR, and DELT
Target output (data)	Water saturation (Sw)
Network architecture	Feedforward back propagation
Training algorithm	Levenberg-Marquardt (trainlm)
Activation function	Tansig (hidden)-Purelin (output)
Performance (validation) function	Mean squared error
Model performance dependency parameter	Regression coefficient

Based on the procedure described above, a proper flowchart to implement ANN modeling is described in Figure 3.5.

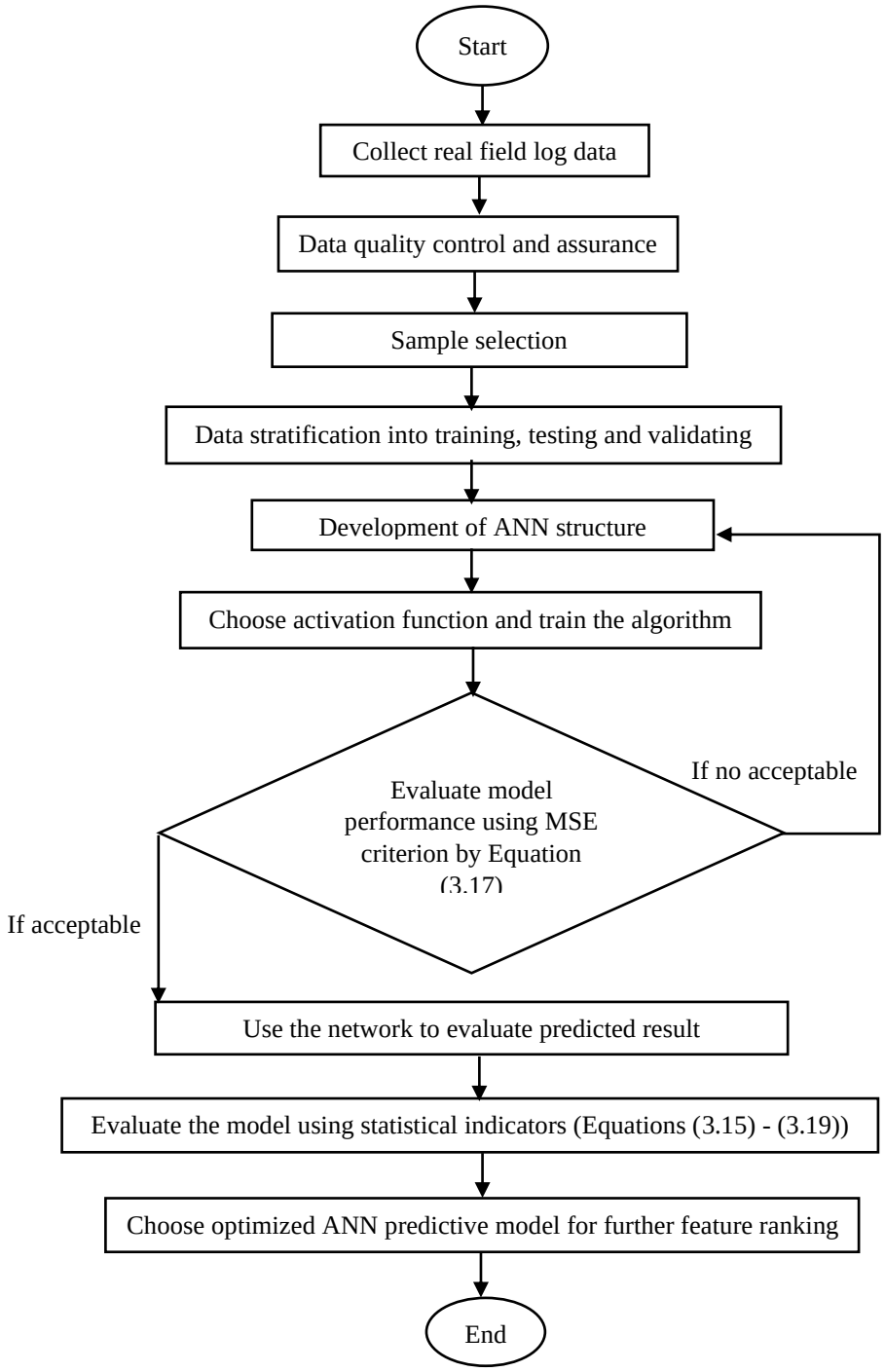

Figure 3.5: A typical flowchart for ANN model to predict the water saturation.

The 182 data points used in this work are from the hydrocarbon-bearing zone in Bengal Basin's gas field. The data points are divided into training (75%), validation (15%), and testing (10%) phases to predict the water saturation using the feed-forward ANN model. The important statistical parameters such as correlation coefficient (*R*), average absolute percent relative error (AAPE), mean squared error (MSE), and maximum error percentage (MAPE) are used to evaluate the performance/accuracy of various ANN models tested in this study (Ashena and Thonhauser, 2015). The AAPE and MSE show the deviation of the predicted value ($\widehat{Y_i}$) from the real data (Y_i). The correlation coefficient (*R*), which falls between 0 and 1, indicates how well a model can forecast the target parameter. When the R-value is high, it implies that the model can offer accurate predictions. The magnitude of R is obtained as follows:

$$R = 1 - \frac{\sum_{i=1}^{n}(Y_i - \widehat{Y_i})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{3.15}$$

in which, Y_i stands for the target real value; $\widehat{Y_i}$ resembles the predicted value of Y_i; and $\bar{Y}$ is the mean value of Y_i.

The mathematical expressions for the AAPE and MSE are given below, respectively (Ashena and Moghadasi, 2011):

$$\text{AAPE} = \frac{1}{n}\sum_{i}^{n}\left(1 - \frac{\widehat{Y_i}}{Y_i}\right) \tag{3.16}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y_i}\right)^2 \tag{3.17}$$

The performance of the models is also assessed based on the values of MAPE, as defined below:

$$MAPE = \text{Maximum}\left|\frac{(Y_i - \widehat{Y}_i)}{Y_i}\right| * 100 \qquad (3.18)$$

3.3.2 Ranking of petrophysical parameters

The ranking/dependency quantification of parameters in the MI and ANN models is conducted by appropriate flowcharts, as shown in Figures 3.6 and 3.7, respectively.

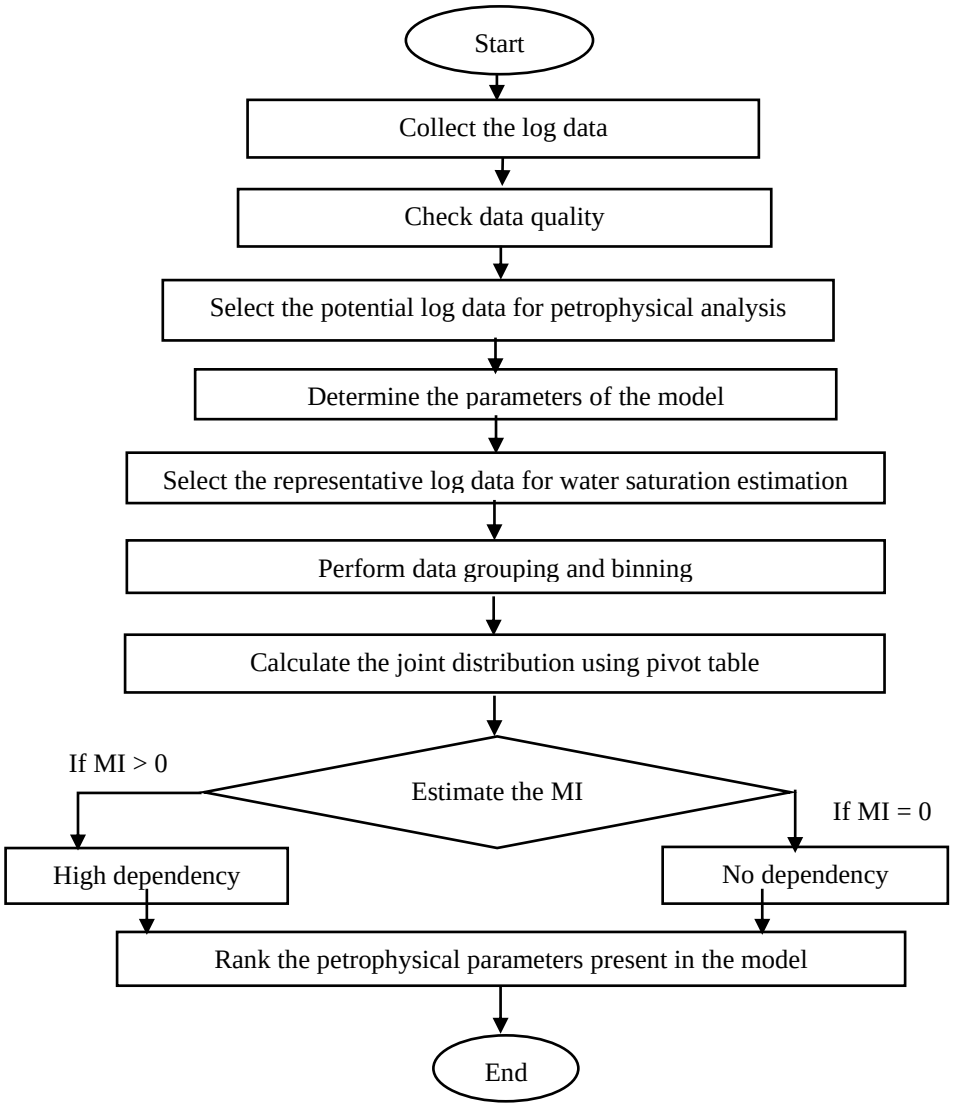

Figure 3.6: Flowchart to illustrate the parameter ranking methodology using the MI approach.

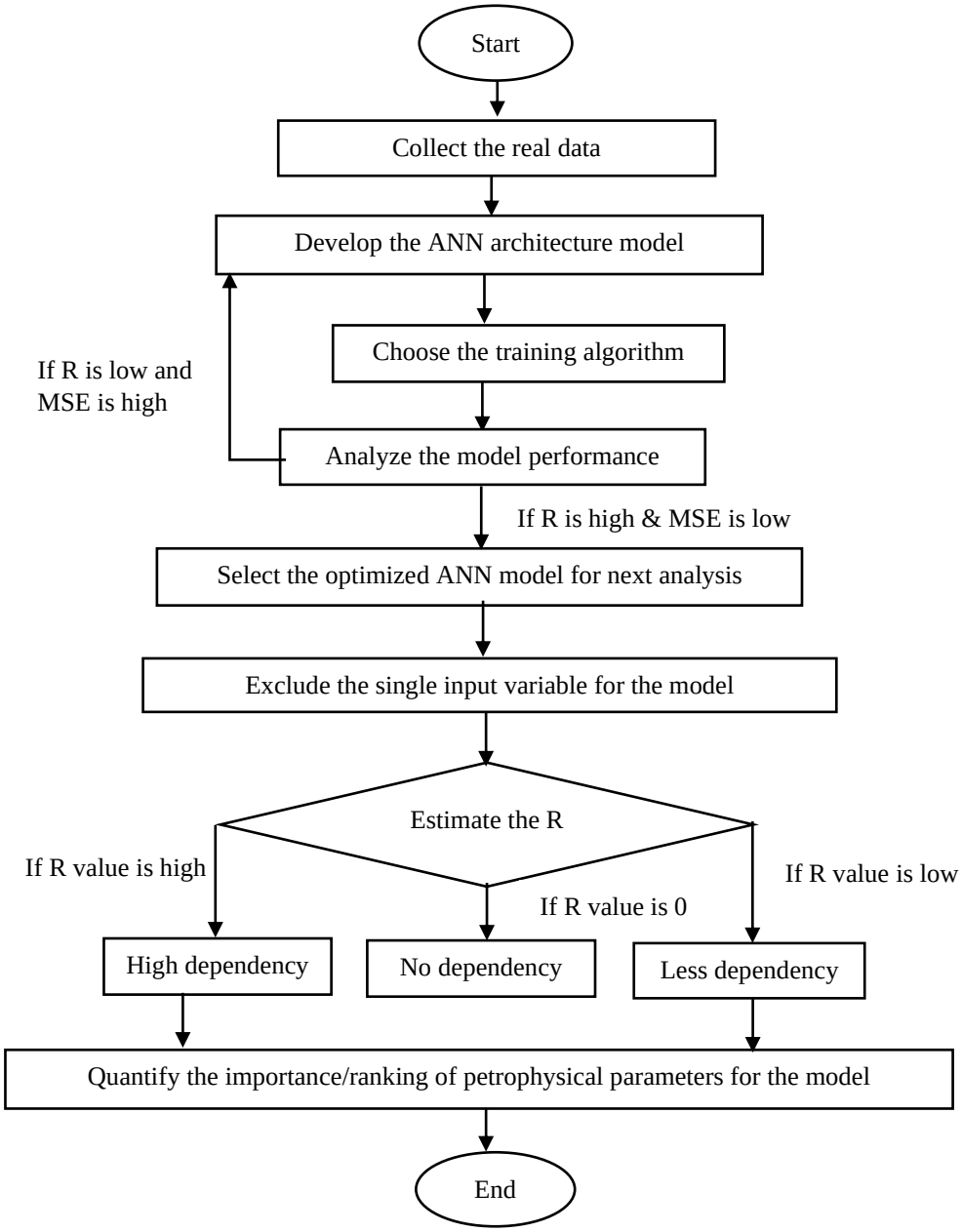

Figure 3.7: Methodology to determine the parameter ranking using the ANN approach.

The higher dependency of two variables in a model can be attributed to the higher positive value of MI. For instance, the MI is zero when variables are statistically independent. The MI value can be determined using the mutual information approach. In addition, the

regression coefficient is an appropriate criterion to quantify the dependency of parameters involved in the optimized ANN model.

3.4 Data Collection and Processing

This study is conducted based on the available log data of Bengal Basin's gas field. In the literature, the geological setting of this basin is well illustrated by several researchers such as Alam (2003), Imam (2013), Miah and Tamim (2015). The rock lithology has been identified with the help of spectral gamma-ray log data. Both water and hydrocarbon-bearing zones have been detected using gamma-ray, resistivity, and porosity logs. There are three major types of logs used in the current study, as listed Table 3.2.

Table 3.2: List of available log data.

Major log type	Name of log type	Parameters
Lithology log	Gamma ray	Gamma Ray (GR), API
Resistivity log	Array compensated resistivity	Deep (true) resistivity (RT), ohm-m
Porosity log	Spectral density, Dual spaced neutron	Bulk density (RHOB), Neutron porosity (NPHI)

3.4.1 Statistical analysis of log data

According to the log data, there is no significant depth shift among the depth of logs such as gamma-ray, resistivity, density, neutron, and sonic logs. Table 3.3 summarizes the results of statistical analysis performed on the reservoir parameters. According to Table 3.3, it is concluded that the reservoir log properties are heterogeneous.

Table 3.3: Statistical information on predictor log variables.

Statistical parameter	GR (API)	RT (ohm-m)	RHOB (g/cm^3)	NPHI (%)	DELT (µs/ft)
Mean	100.19	22.67	2.367	16.979	92.90
Median	98.40	22.47	2.355	16.752	92.83
Standard deviation	13.87	4.96	0.043	1.296	2.42
Variance	192.51	24.58	0.002	1.680	5.89

3.4.2 Data grouping and binning

The log data are grouped into three categories such as Low (L), Medium (M), and High (H) based on the statistical analysis (such as maximum, minimum, mean, and range of log data values) of collected data where the standard data binning process is considered to categorize the cut-off petrophysical parameters. The ranges of scaled parameters and frequency level of data binning are reported in Tables 3.4 and 3.5, respectively.

Table 3.4: Variations of log parameters.

Group	GR (API)	RT (ohm-m)	RHOB (g/cm^3)	NPHI (%)	PHIND (%)	V_{sh} (%)
L	≤80	≤ 15	≤ 2.32	≤ 10	≤ 15	≤ 10
M	80.1-100	15.1-30	2.321-2.400	10.1-20	15.1-20	10.1-20
H	100.1-120	30.1-45	2.401-2.520	≥ 20.1	≥ 20.1	20.1-45

Table 3.5: Frequency level of data binning.

Parameter/level	H	M	L
GR (API)	2	103	65
RT (ohm-m)	12	162	8
RHOB (g/cm^3)	11	140	31
NPHI (%)	0	176	6
PHIND (%)	6	176	0
V_{sh} (%)	7	130	45

3.5 Limitations of Research Approach

One of the limitations of this study is the relatively low number of data points used for the model development. In addition, the logging parameters are considered as discrete random variables to investigate the dependency level of input variables using the MI approach. The single hidden layer perception is applied in the current study, while the multilayer perception algorithms can be employed to construct a more reliable model. Other deterministic tools such as least-square support vector machines and fuzzy logic can be also used to estimate the reservoir characteristics and to rank the input parameters. The comparison between the performances of the models would be then possible to select the best tool for reservoir analysis of various cases with different properties and heterogeneities.

3.6 Results and Discussions

This section presents the main results obtained from this study in the form of tables and figures so that predictions of water saturation and ranking of input parameters through

utilizing the ANN and MI tools are given. The systematic discussions on the findings and trends are then provided.

3.6.1 Prediction of water saturation using correlation

The lithology of the reservoir is the sand and shaly sand based on gamma-ray and resistivity log data analysis. Several water-bearing sandstone reservoirs are detected based on the available test results. Some shale zones are also found at the top reservoir pools, which act as the cap rock (non-permeable). Below the sand zones, there are a number of shaly sand zones, which are porous and permeable. The radioactivity properties, resistivity, bulk density, and other log responses considerably change with depth throughout the reservoir. The maximum and minimum values of GR log of the studied wells are 155 and 77 API, respectively, and the average value is 91-103 API for the main hydrocarbon-bearing zone. The water resistivity (R_w in ohm-m) is estimated using R_{wa} analysis where it gives 0.104 ohm-m at water-bearing zone. The R_w value ranges from 0.102 to 0.092 ohm-m in various zones, considering a reference value of 0.104. The corrected porosity obtained from the neutron-density combination formula is used to estimate the magnitude of water saturation. The water saturation is determined where the tortuosity coefficient (a=1), cementation (m=2), and saturation exponent (n=2) are considered. The ranges of petrophysical parameters are listed in Table 3.6.

Table 3.6: Predicted reservoir parameters.

Category	Shale volume	Corrected porosity	Water saturation
Minimum	0.2622	0.0589	0.2774
Maximum	0.5605	0.1792	0.5426
Average	0.2788	0.1540	0.3950

According to the Indonesian model, an average water saturation of 39.50% is obtained using sample data studied in this study. The estimated water saturation is close to the core water saturation (e.g., 40%). The water saturation predicted from the Indonesian model is also utilized to investigate the dependency of the log variables of the model through implementing MI and ANN approaches.

3.6.2 ANN model to determine water saturation

The feed-forward (FF) backpropagation algorithm is employed to optimize the ANN network where the main target is to obtain the water saturation using five input variables with high reliability and precision. We conduct the ANN model optimization in the training stage. The results obtained from the analysis of the ANN model with one single hidden layer but different numbers of neurons are tabulated in Table 3.7. According to Table 3.7, case no. 3 with 4 neurons exhibits the least error. The optimal ANN model is 5:4:1. It means that the model has one input layer with 5 neurons, one hidden layer with 4 neurons, and one output layer with one neuron; where R is 99.988 and the minimum values

of AAPE and maximum error percentage are obtained for this network. This optimal model is then used in both testing and validation phases.

Table 3.7: Sensitivity analysis on the number of hidden neurons in terms of network performance.

FF-ANN	No. of neurons in the hidden layer	AAPE (%) for training data	MAPE (%) for training data	AAPE (%) entire data points	Regression Coefficient (R)
Case 1	2	0.7464	5.8821	0.7211	99.442
Case 2	3	0.4644	1.9387	0.4111	99.903
Case 3	4	0.165	2.5542	0.1469	99.988
Case 4	5	0.4623	8.3562	0.3874	99.924
Case 5	6	0.1179	3.1173	0.1056	99.995
Case 6	7	0.1437	6.1428	0.1210	99.999
Case 7	8	1.1913	12.7150	1.0655	98.703
Case 8	9	0.5279	7.1232	0.4583	99.969
Case 9	10	0.2590	2.5203	0.2598	99.971

To avoid timely training and over-fitting, the optimal epoch in the validation stage is also identified. The best validation performance is attained at epoch 14 with an MSE of 0.0000173 (see Figure 3.8). As Figure 3.8 demonstrates, the errors of the training and testing phases exhibit a decreasing trend with increasing number of epochs, though the validation error shows small fluctuations at the beginning. It is found that the minimum error occurs at epoch 14 after which no change in error percentage is noticed, as depicted in Figure 3.8. The variations of water saturation with reservoir depth are presented in Figure 3.9 based on both real data and predictions.

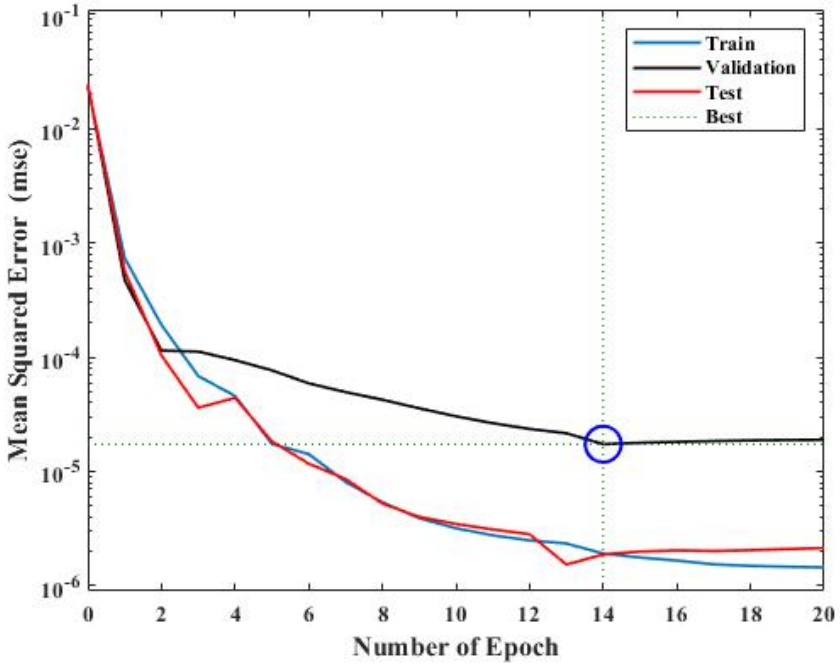

Figure 3.8: Validation performance curve for the optimized ANN approach.

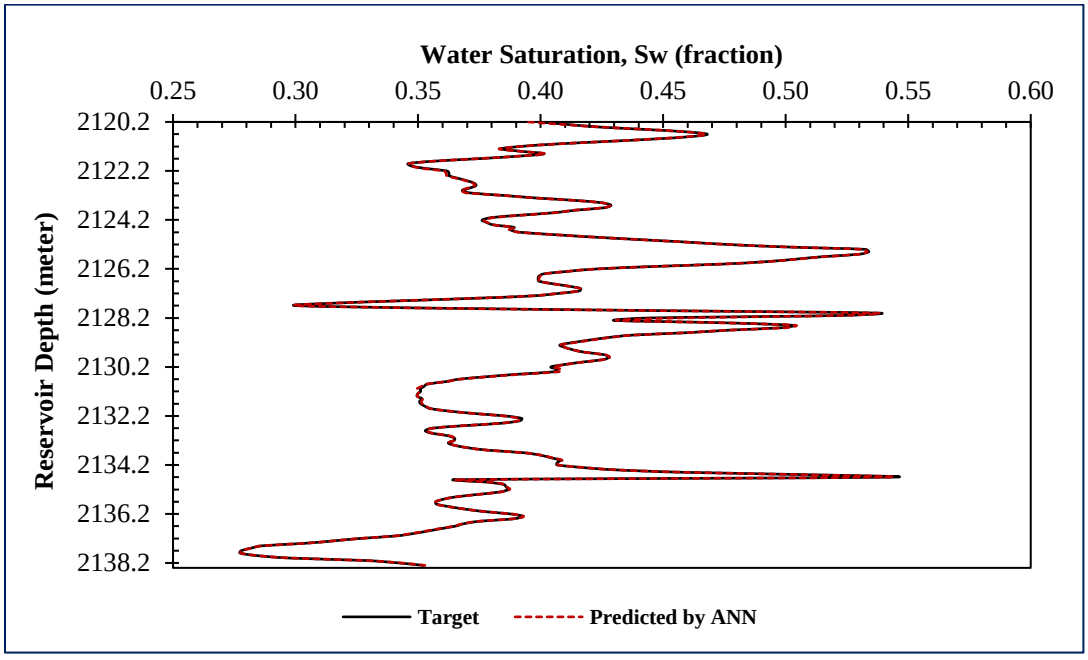

Figure 3.9: Comparison between target and predicted water saturation while using optimized ANN model.

The predicted water saturations versus the real data are plotted in Figure 3.10 for all phases including testing, training, and validation.

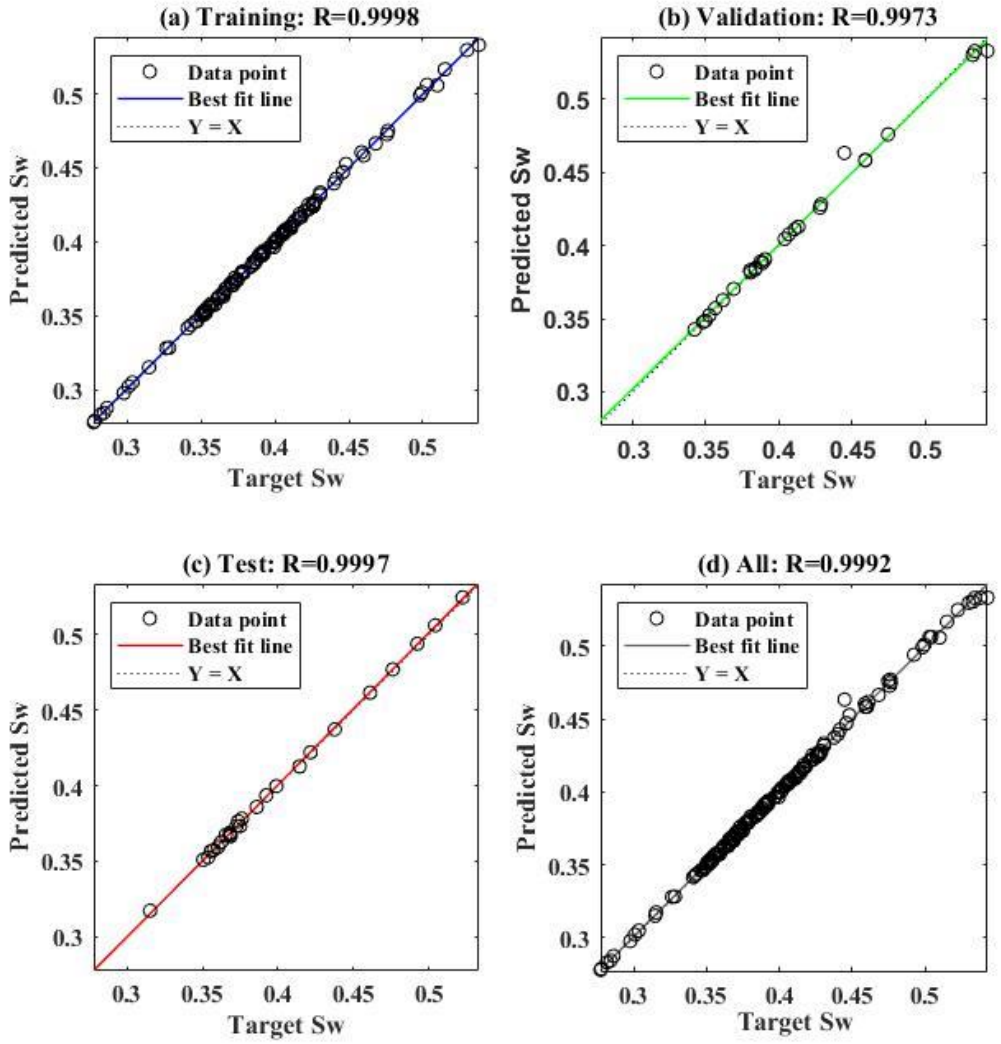

Figure 3.10: Network performance on the basis of targeted and predicted Sw values while using optimized ANN model: (a) Training, (b) Validation, (c) Testing, and (d) All data

As clear from Figure 3.10, all data points are placed on the line Y=X. As the R-value is close to 1, the performance of the connectionist tool is very satisfactory (Coulibaly and Baldwin, 2005). Based on Figures 3.9 and 3.10, there is very good agreement between the predicted and real water saturations, showing high predictive capability of ANN method in reservoir characterization.

3.6.3 Ranking of log variables

The range and frequency distribution of Indonesian model based-water saturation are reported in Table 3.8. The probability distribution is also presented in Table 3.9.

Table 3.8: Range of water saturation data grouping and frequency distribution.

Group	Range	Frequency level of sample
L (Low)	≤ 35	22
M (Medium)	35.1-45	140
H (High)	$\geq$45.1	20

According to Table 3.8, it is found that the M group has a considerably higher frequency level, compared to the other two groups of L and H. Referring to Table 3.9, it is noticed that the joint distribution between RT and Sw in the M-M category (81.32%) is greater than that of other categories. Table 3.10 provides the dependency of petrophysical variables with the water saturation. It is concluded that RHOB and RT show higher importance (dependency) in the Sw model compared to other variables, while GR has the minimum importance.

Table 3.9: Matrix table of the probability distribution between two variables.

Log subset variable with S_w	Group	H	L	M
True resistivity (RT)	H	0	3.85	0.55
	L	4.95	0	1.65
	M	7.14	0.55	81.32
Shale volume (Vsh)	H	5.5	0.55	5.5
	L	2.2	2.75	52.8
	M	4.4	1.1	25.3
Effective porosity (PHI)	H	1.1	4.4	65.4
	L	1.65	0	1.65
	M	9.34	0	16.5
Gamma Ray (GR)	H	9.9	1.1	30.8
	L	0	0	1.10
	M	2.2	3.3	51.65
Bulk density (RHOB)	H	7.7	0	1.65
	L	0	3.3	1.65
	M	4.4	1.1	80.00
Neutron porosity (NPHI)	H	0.55	0.55	2.2
	L	0	0	0
	M	11.5	3.85	81.30

Table 3.10: Importance ranking of log variables to predict water saturation.

Predictor variable	MI	Rank
RHOB	0.271	1
RT	0.237	2
PHIND	0.189	3
NPHI	0.107	4
VSH	0.038	5
GR	0.025	6

Considering the same number of neurons in the hidden layer, the optimized ANN strategy is implemented to explore effect of each variable through excluding it from the model (see Figures 3.11 to 3.14). It is found that different regression coefficients are obtained for different scenarios where one input parameter is removed from the model.

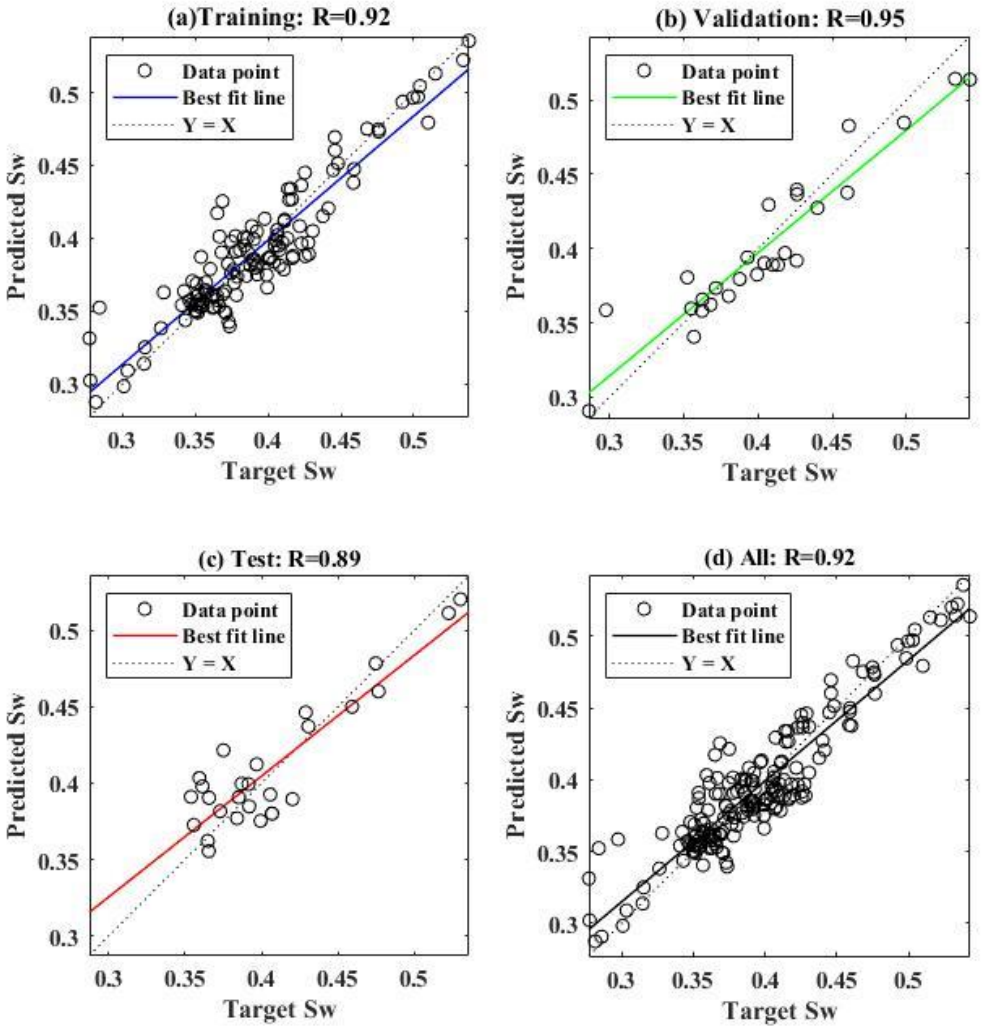

Figure 3.11: Regression coefficient for the corrected ANN model when RT is excluded: (a) Training, (b) Validation, (c) Testing, and (d) All data

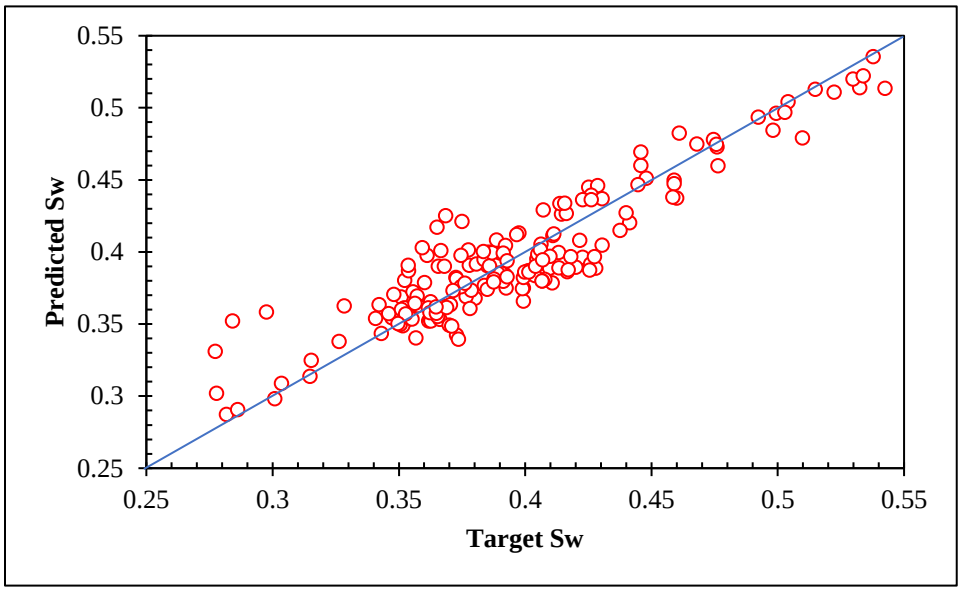

Figure 3.12: Cross plot to compare the actual and predicted water saturations for exploring the impact of RT on the performance of optimized model.

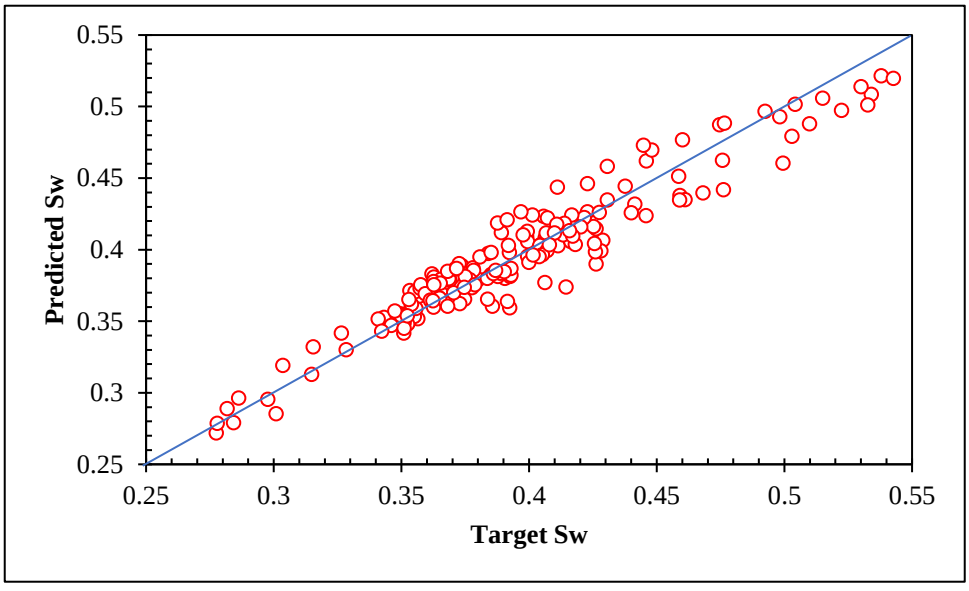

Figure 3.13: Comparing the real data and predictions based on the modified ANN model when RHOB is excluded.

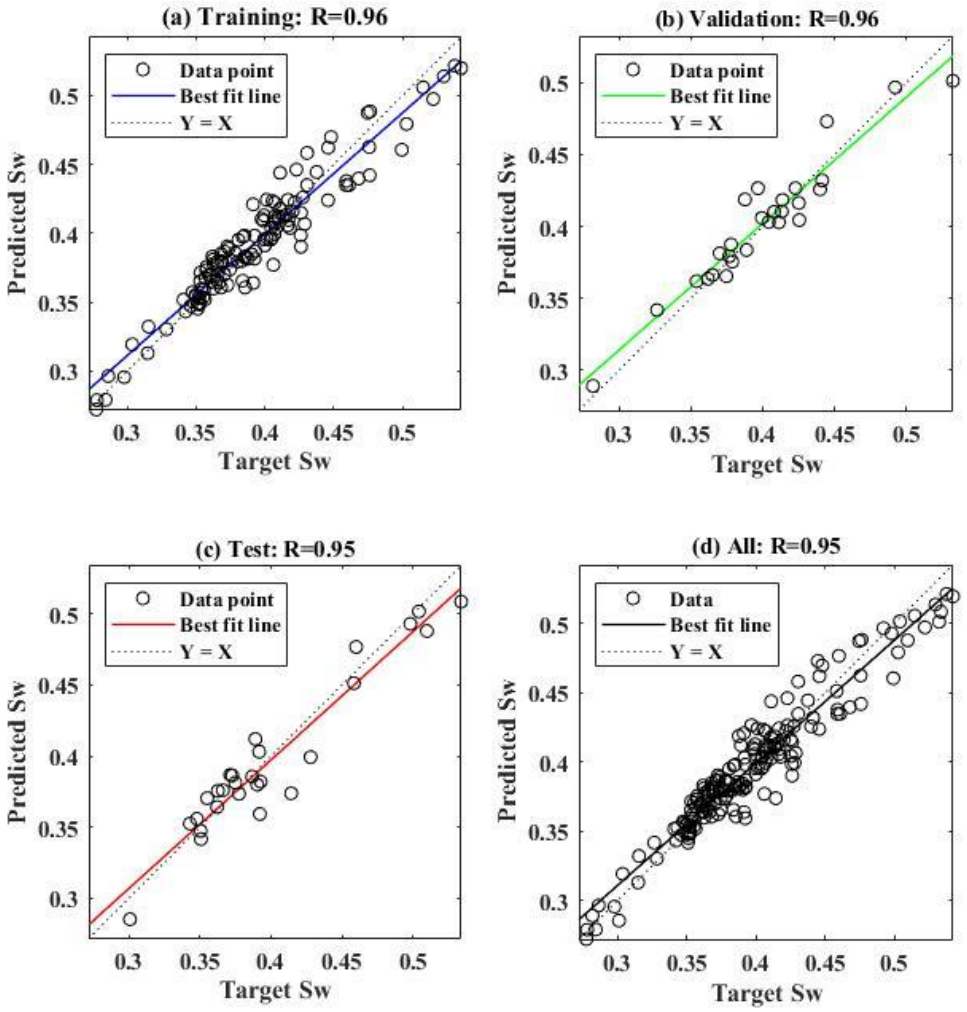

Figure 3.14: Regression coefficient of the reduced ANN approach to highlight the importance of RHOB: (a) Training, (b) Validation, (c) Testing, and (d) All data

According to Figures 3.11 to 3.14, the impacts of RT and RHOB in the model are greater than other parameters. For the scenario of excluding RT as shown in Figure 3.11, the regression coefficient of the training, testing, and validation phases are 0.924, 0.886, and

0.950, respectively. It should be noted that the model gives an R of 0.9234 considering all data points when RT as an input parameter is excluded gamma-ray the water saturation, Sw. Table 3.11 summarizes the statistical parameters including error percentage for all cases (when one input parameter is ignored).

Table 3.11: Error analysis of various versions of reduced ANN model.

	Excluded	AAPE (%)				Maximum	Regression	Ranking
Scenario	parameter	Training	Testing	Validation	All data	error (%)	Coefficient	
a	RT	3.55	3.38	5.97	4.0	25	92.34	1
b	RHOB	2.95	2.99	2.85	2.99	11.45	95.44	2
c	NPHI	0.93	0.98	1.05	0.96	3.55	99.56	3
d	GR	0.62	0.69	0.77	0.66	4.6	99.86	4
e	DELT	0.2	0.08	0.21	0.19	2.52	99.96	5

Based on Table 3.11, high error percentage and low regression coefficient are obtained for the input parameters with high dependency to the output variable, implying the considerable significance of the input parameter. The impact of the parameters excluded from the model is also illustrated in Figure 3.15. According to Figure 3.15, the input parameters of RT and RHOB are highly significant in the optimized ANN approach to estimate the fluid saturation. The comparison between the real and predicted Sw to further highlight the impact of input parameters is illustrated in Figure 3.16, implying a great match between them.

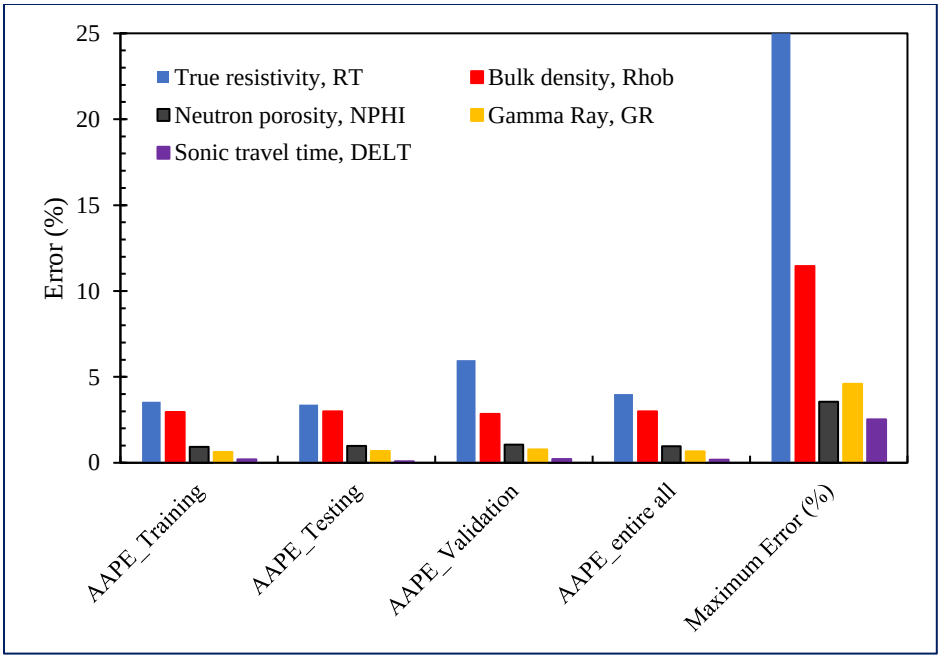

Figure 3.15: Comparison of the performance of the reduced ANN scenarios based on statistical analysis.

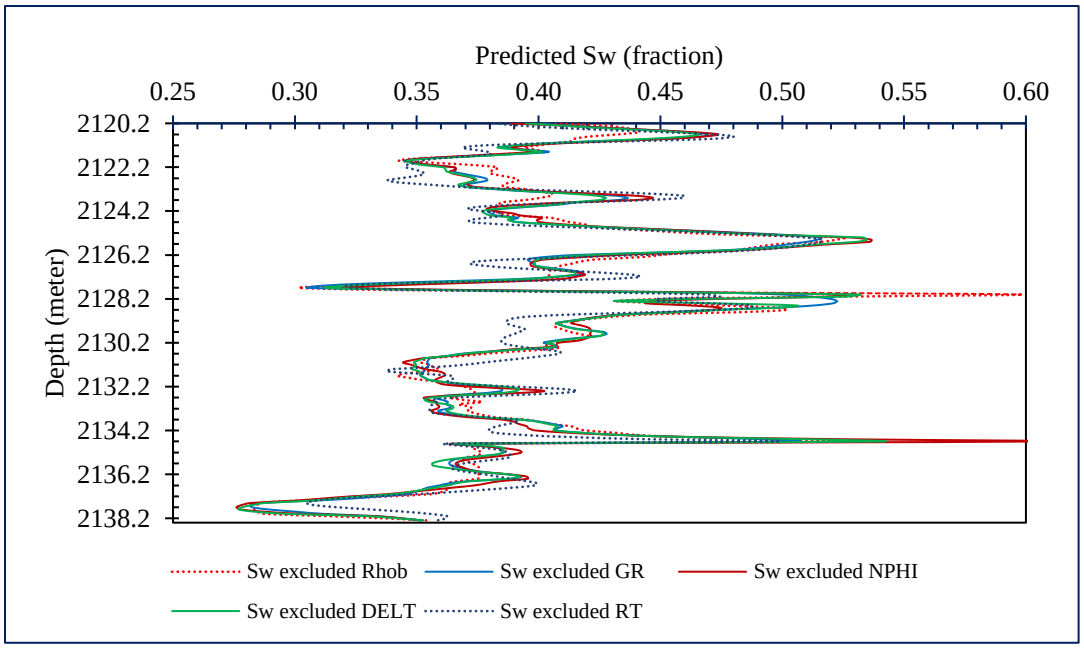

Figure 3.16: Water saturation versus reservoir depth based on the real data and results obtained from the reduced input variable models.

The optimized model is also validated using different data sets of the same reservoir to rank the log variables incorporated in the model. The model introduced for different data banks exhibits almost the same ranking for the input parameters. In addition, the MI approach reveals that the most important log variables are the bulk density and true resistivity while the gamma-ray and neutron porosity have minor effects on the fluid (water) saturation.

Highlighting the importance of this study, the vital log variables (e.g., true resistivity and bulk density) can be employed to accurately obtain the water saturation during exploration phase in the oil and gas industry using efficient deterministic strategies such as smart models. Utilization of such tools requires less time and computational complexities, leading to lower capital and operating costs in the exploration stage.

3.7 Conclusions

This study presents two efficient tools to characterize the reservoir and rank the input parameters affecting the target variable. It is found that the reservoir log properties are heterogeneous based on the nature and distribution of the real-life log data sets of the reservoir. The comparison of the results obtained from the petrophysical model and artificial neural network (ANN) model indicates that the reservoir quality is good in terms of reservoir properties. Based on the study outputs, it is clear that the performance of the optimized model is satisfactory in terms of reliability and accuracy while obtaining the water saturation so that very good agreement is noticed between the real data and estimated water saturations. The strategies used in this research work reveal that the true resistivity

and bulk density have higher impacts on the target variable, compared to other variables. In other words, there is a strong relationship between water saturation and two input variables including true resistivity and bulk density. The ANN model gives an R of 0.9234 and 0.9996 considering entire data points when true resistivity and sonic travel time as the input parameters are excluded, respectively, while predicting the water saturation. It follows that bulk density and true resistivity have a mutual information (MI) value in the range of 0.237-0.271, confirming high significance of these two parameters in the model compared to other variables, while gamma-ray has the minimum importance with an MI of 0.025. It is also found that the ranking orders of the variables with both ANN and MI are almost the same. The ranking implies that the secondary logging variables have the minimum importance to predict the water saturation.

Acknowledgements

The authors would like to thank Equinor (formerly Statoil) Canada Ltd., Natural Sciences and Engineering Research Council of Canada (NSERC), and InnovateNL for providing financial support to accomplish this study at the Memorial University, St. John's, NL, Canada.

Nomenclatures

Acronyms

AAPE	Average Absolute Percentage Error
ABC	Artificial bee colony
ANN	Artificial Neural Network
DELT	Sonic travel time (μs/ft)
FL	Fuzzy logic

GR	Gamma-ray (API)
GR_{loa}	Gamma-ray value of the zone of interest
GR_{max}	Maximum value of gamma-ray log over the entire log
GR_{min}	Minimum value of gamma-ray log over the entire log
$H(X)$	Entropy of X variable
$H(Y)$	Entropy of Y variable
$H(X \mid Y)$	Conditional entropy of X at fixed variable of Y
ICA	Imperialist competitive
I_{GR}	Shale Index
LM	Levenberg-Marquardt
MAPE	Maximum absolute error percentage
MSE	Mean Square Error
MI	Mutual Information
N	Number of neurons
NPHI	Neutron Porosity
LSSVM	Least Square Support Vector Machine
PHIN	Neutron porosity (frac.)
PHIND	Porosity from the combination of density and neutron log
R	Regression Coefficient
RHOB	Bulk density (gm/cc)
RT	True (Deep) Resistivity (ohm-m), Rt
R_{sh}	Resistivity of shale zone for virgin zone (ohm-m)
Rw	Formation Water Resistivity (ohm-m)
Sw	Water saturation (frac.)
Vsh	Shale volume (shaliness)
W_{ij}	Connected weight between the *i*-th neuron and *j*-th neuron
X	Random variable
Xi	Output of previous layer or inputs to neuron *j*
Y	Random variable
Yi	Net input to neuron *j* in the hidden or output layer
$\widehat{Y}_i$	The predicted value of Y_i
$\bar{Y}$	Mean value of Y_i

Variables, Parameters and Subscripts

a	tortuosity factor
b	bias
cl	clay
e	effective
m	cementation exponent
n	saturation exponent
$p(x)$	Probability of x variable

$p(y)$	Probability of y variable
$p(x,y)$	Probability between x and y variable
sh	shale
w	Water

Greek Letters and subscript

ϕ	Porosity (frac.)
ϕ_e	Effective porosity (frac.)
$\phi_{D,e}$	Effective density porosity (frac.)
$\phi_{N,e}$	Effective neutron porosity (frac.)
$\phi_{N,sh}$	Neutron porosity of the adjacent shale zone (frac.)
ρ_b	Bulk density (gm/cc)
$\rho_{b.c}$	Clay corrected density porosity (gm/cc)
ρ_{fl}	Fluid density (gm/cc)
ρ_{ma}	Matrix density (gm/cc)

References

Abellan A. & Noetinger, B. (2010). Optimizing Subsurface Field Data Acquisition Using Information Theory. Mathematical Geosciences. 42: 603–630.

Adedigba, S. A., Khan, F., & Yang, M. (2017). Dynamic failure analysis of process systems using neural networks. Process Safety and Environmental Protection, 111, 529-543.

Aggoun, R. C., Tiab, D., & Owayed, J. F. (2006). Characterization of flow units in shaly sand reservoirs—Hassi R'mel Oil Rim, Algeria. Journal of Petroleum Science and Engineering, 50(3-4), 211-226.

Ahmadi, A. M., Zendehboudi, S., Lohi, A., Elkamel, A., & Chatzis, I. (2013). Reservoir permeability prediction by neural networks combined with hybrid genetic algorithm and particle swarm optimization. Geophysical Prospecting, 61(3), 582-598.

Aïfa, T., Baouche, R., & Baddari, K. (2014). Neuro-fuzzy system to predict permeability and porosity from well log data: A case study of Hassi R′ Mel gas field, Algeria. Journal of Petroleum Science and Engineering, 123, 217-229.

Akande, K. O., Owolabi, T. O., & Olatunji, S. O. (2015). Investigating the effect of correlation-based feature selection on the performance of neural network in reservoir characterization. Journal of Natural Gas Science and Engineering, 27, 98-108.

Alam, M, Alam MM, Curray J., Chowdhury A., & Gani M., (2003). An overview of the sedimentary geology of the Bengal Basin in relation to the regional tectonic framework and basin-fill history. Sedimentary Geology. 155:179-208.

Al-Bulushi, N., Araujo, M., Kraaijveld, M., & Jing, X. D., (2007). Predicting Water Saturation Using Artificial Neural Networks (ANNs). In SPWLA Middle East Regional Symposium. Society of Petrophysicists and Well-Log Analysts.

Al-Bulushi, N., King, P. R., Blunt, M. J., & Kraaijveld, M. (2009). Development of artificial neural network models for predicting water saturation and fluid distribution. Journal of Petroleum Science and Engineering, 68(3-4), 197-208.

Al-Bulushi, N.I., King, P.R., Blunt, M.J. & Kraaijveld, M., (2012). Artificial neural networks workflow and its application in the petroleum industry, Computing and Applications, 21(3), 409-421..

Ali J.K., (1994). Neural network: a new tool for petroleum industry. Proceedings of SPE European petroleum computer conference, UK, SPE Paper 27561.

Ali, S. S., Nizamuddin, S., Abdulraheem, A., Hassan, Md. R., & Hossain, M. E., (2013). Hydraulic unit prediction using support vector machine, Journal of Petroleum Science and Engineering, October 2013, Vol.110, 243-252.

Alkinani, H. H., Al-Hameedi, A. T. T., Dunn-Norman, S., Flori, R. E., Alsaba, M. T., & Amer, A. S., (2019). Applications of Artificial Neural Networks in the Petroleum Industry: A Review. Society of Petroleum Engineers. doi:10.2118/195072-MS

Al-Ruwaili, S.B., & Al-Waheed, H.H., (2004). Improved petrophysical methods and techniques for shaly sands evaluation. Proceeding of SPE International Petroleum Conference, Mexico. Paper SPE, 89735.

Aminian, K., & Ameri, S. (2005). Application of artificial neural networks for reservoir characterization with limited data. Journal of Petroleum Science and Engineering, 49(3-4), 212-222.

Amiri, M., Ghiasi-Freez, J., Golkar, B., & Hatampour, A. (2015). Improving water saturation estimation in a tight shaly sandstone reservoir using artificial neural network optimized by imperialist competitive algorithm–A case study. Journal of Petroleum Science and Engineering, 127, 347-358.

Anifowose, F., (2018). Data-driven Approach to Handling High-dimensional Data Input Space using Feature Selection Based Hybrid Machine Learning Methodology. In: Proceedings of the AAPG Middle East Region GTW, Digital Subsurface Transformation, Dubai, UAE, May 7-8.

Anifowose, F., Adeniye, S., Abdulraheem, A., & Al-Shuhail, A. (2016). Integrating seismic and log data for improved petroleum reservoir properties estimation using non-linear feature-selection based hybrid computational intelligence models. Journal of Petroleum Science and Engineering, 145, 230-237.

Anifowose, F., Labadin, J., & Abdulraheem, A., (2014). Non-linear feature selection-based hybrid computational intelligence models for improved natural gas reservoir characterization. J. Nat. Gas Sci. Eng. 21, 397–410.

Anifowose, F., Labadin, J., & Abdulraheem, A., (2015). Ensemble model of non-linear feature selection-based Extreme Learning Machine for improved natural gas reservoir characterization. J. Nat. Gas Sci. Eng. 26, 1561–1572.

Archie, G.E., (1941). The electrical resistivity log as an aid in determining some reservoir characteristics. Transactions of AIME 146, 54–62.

Ashena, R. & Moghadasi. (2011). Bottom Hole Pressure Estimation Using Evolved Neural Networks by Real Coded Ant Colony Optimization and Genetic Algorithm. Journal of Petroleum Science and Engineering 77.3-4 (2011): 375-85.

Ashena, R., & Thonhauser, G. (2015). Application of Artificial Neural Networks in Geoscience and Petroleum Industry. In Artificial Intelligent Approaches in Petroleum Geosciences (pp. 127-166). Springer, Cham.

Asquith, G., & Krygowski, D., (2004). Basic well log analysis, American Association of Petroleum Geologists, Second Edition, Tulsa, Oklahoma, p. 216.

Azevedo, L. & Soares, A. (2017). Geostatistical Methods for Reservoir Geophysics, Advances in Oil and Gas Exploration & Production, DOI 10.1007/978-3-319-53201-1_4.

Baev, K. V., (1998). Biological Neural Networks: Hierarchical Concept of Brain Function. Boston: Birkhäuser.

Banchs, R., & Michelena, R., (2002). From 3D seismic attributes to pseudo-well-log volumes using neural networks: practical considerations. The Leading Edge 21 (10), 996–1001.

Baneshi, M., Behzadijo,M., Schaffie, M. & Nezamabadi-Pour, H., (2013). Predicting Log Data by Using Artificial Neural Networks to Approximate Petrophysical Parameters of Formation, Petroleum Science and Technology, 31:12, 1238-1248, DOI: 10.1080/10916466.2010.540611

Baouche, R., Aïfa, T., & Baddari, K. (2017). Intelligent methods for predicting nuclear magnetic resonance of porosity and permeability by conventional well-logs: a case study of Saharan field. Arabian Journal of Geosciences, 10(24), 545.

Bassiouni, Z., (1994). Theory, Measurement and Interpretation of Well Logs, First Printing, Henry L. Doherty Memorial Fund of AIME, SPE, Richardson, TX, Pp. 372.

Baziar, S., Shahripour, H. B., Tadayoni, M., & Nabi-Bidhendi, M. (2018). Prediction of water saturation in a tight gas sandstone reservoir by using four intelligent methods: a comparative study. Neural Computing and Applications, *30*(4), 1171-1185.

Benardos & Vosniakos. (2007). Optimizing Feedforward Artificial Neural Network Architecture. Engineering Applications of Artificial Intelligence 20.3 (2007): 365-82.

Biu, V.T., Biu, E.O., Buduka, S., Ajienka, J., & Onyekonwu, M., (2010). Well and reservoir parameters estimation (K and skin) using the statistical diagnostic approach part II. In: Proc. the SPE Nigeria Annual International Conference and Exhibition, Tinapa e Calabar, Nigeria.

Chabock, R., Riahi, M.A., & Memariani, M., (2017). Determination of the petrophysical parameters using geostatistical method in one of the hydrocarbon reservoirs in South West of Iran. Journal of Scientific and Engineering Research 4(12), 44-55.

Clavier, C., Coates, G., & Dumanoir, J., (1984). Theoretical and experimental bases for the dual water model for interpretation of shaly sands. Society of Petroleum Engineers Journal 24, 153–168.

Close, D., & Caycedo, F. (2011). Integrated geophysics and geomodelling workflows for reservoir characterization: A case study of waterflood optimization. In SEG Technical Program Expanded Abstracts 2011 (pp. 1840-1844). Society of Exploration Geophysicists.

Coulibaly, P., & Baldwin, C.K. (2005). Nonstationary hydrological time series forecasting using nonlinear dynamic methods. J Hydrol 307:164–174.

Cover, T.M., Thomas, J.A., 2006. In: Elements of information theory. John Wiley & Sons.

Crain E. R. (1986). The Log Analysis Handbook, Volume-1, Quantitative Log Analysis Methods, PennWell Publishing Company, Tulsa, Oklahoma, USA, p. 328.

Cuddy, S. (2000). Litho-facies and permeability prediction from electrical logs using fuzzy logic, SPE Reservoir Evaluation and Engineering, 3(4), 319–324.

Dubois, M.K., A.P. Byrnes, G.C. Bohling, & J.H. Doveton, (2006). Multiscale geologic and petrophysical modeling of the giant Hugoton gas field (Permian), Kansas and Oklahoma: in P. M. Harris and L. J. Weber, eds., Giant reservoirs of the world: From rocks to reservoir characterization and modeling: American Association of Petroleum Geologists Memoir 88, p. 307- 353.

Eidsvik, J., Mukerji, T., & Switzer, P., (2004). Estimation of geological attributes from a well log: an application of hidden Markov chains. Mathematical Geology 36, 379–397.

Ellis, D.V., & Singer, J.M., (2007). Well Logging for Earth Scientists: Springer, London, 692 P.

El-Sebakhy, E., Abdulraheem, A., Ahmed, M., Al-Majed, A., Awal, M., Azzedin, F., & Raharja, P., (2012). Functional networks as a novel approach for prediction of permeability in a carbonate reservoir. Expert Systems with Applications 39 (12), 10359–10375.

Faucett, L., (1994). Fundamentals of Neural Networks. Architectures, Algorithms, and Applications. Prentice Hall, Englewood Cliffs, NJ.

Gardner, & Dorling. (1998). Artificial Neural Networks (the Multilayer Perceptron)—a Review of Applications in the Atmospheric Sciences. Atmospheric Environment 32.14 (1998): 2627-636.

Ghaffarian, N., Eslamloueyan, R., & Vaferi, B. (2014). Model identification for gas condensate reservoirs by using ANN method based on well test data. Journal of Petroleum Science and Engineering, 123, 20-29.

Gholanlo, H. H., Amirpour, M., & Ahmadi, S. (2016). Estimation of water saturation by using radial based function artificial neural network in carbonate reservoir: A case study in Sarvak formation. *Petroleum*, *2*(2), 166-170.

Graf, T., Zangl, G., May, R., Hartlieb, M., Randle, J., & Al-Kinani, A., (2011). Candidate selection using stochastic reasoning driven by surrogate reservoir models. SPE J. Res. Eval. Eng. 14 (4), 413-422.

Hakiki F., & Wibowo A.T. (2014). Formulation of Rock Type Prediction in Cored Well Using Fuzzy Substractive Clustering Algorithm. Proceeding of the 38th Indonesian Petroleum Association Conference and Exhibition, Paper IPA14-SE-118, Jakarta, Indonesia.

Hakiki, F., & Shidqi, M. (2018). Revisiting fracture gradient: Comments on "A new approaching method to estimate fracture gradient by correcting Matthew–Kelly and Eaton's stress ratio". Petroleum, *4*(1), 1-6.

Hamada, G. M., (1996). An integrated Approach to determine shale volume and hydrocarbon potential in shaly sand", 1996 SCA Conference Paper Number 9641. pp. 1-16.

Hamada, M. G., ElKadi., A., Nyein, C.Y., & Elsakka., A., (2019). Artificial neural network (ANN) can better determine petrophysical properties than conventional techniques

in shaly sand reservoirs, EGYPS 2019 Technical Conference, taking place on 11-13 February 2019 in Cairo, Egypt.

Hamada, M. G., Elsakka., A., & Nyein, C.Y., (2018). Artificial Neural Network (ANN) Prediction of Porosity and Water Saturation of Shaly Sandstone Reservoirs, Advances in Applied Science Research, 2018, 9(2):26-31.

Helle, H.B., & Bhatt, A., (2002). Fluid saturation from well logs using committee neural networks. Petroleum Geosciences 8, 109–118.

Helle, H.B., Bhatt, A., & Ursin, B., 2001. Porosity and permeability predication from wireline logs using artificial neural networks: a north sea case study. Geophysical Prospecting 49, 431–444.

Imam, B., (2013). Energy Resources of Bangladesh, Second Edition, University Grants Commission Publication No. 151, ISBN 984-809-020-1, Bangladesh, 2013, p. 324.

Ipek, I., (2002). Log- derived cation exchange capacity of shaly sands: application to hydrocarbon detection and drilling optimization. PhD thesis, Louisiana State University.

Jorjani, E., Chelgani, S.C. & Mesroghli, S., (2008). Application of artificial neural networks to predict chemical desulfurization of Tabas coal. Fuel, 87(12): 2727-2734.

Kamalyar, K., Sheikhi, Y., & Jamialahmadi, M., (2011). Using an Artificial Neural Network for Predicting Water Saturation in an Iranian Oil Reservoir, Petroleum Science and Technology, 30:1, 35-45, DOI: 10.1080/10916461003752561

Khamidy, N. I., Tariq, Z., & Syihab, Z., (2019). Development of ANN-Based Predictive Model for Miscible CO Flooding in Sandstone Reservoir. Society of Petroleum Engineers. doi:10.2118/194726-MS

Khan, M. R., Tariq, Z., & Abdulraheem, A., (2018). Machine Learning Derived Correlation to Determine Water Saturation in Complex Lithologies. Society of Petroleum Engineers. doi:10.2118/192307-MS.

Kumar, M., Dasgupta, R., Singha, D.K., & Singh, N. P., (2018). Petrophysical evaluation of well log data and rock physics modeling for characterization of Eocene reservoir in Chandmari oil field of Assam-Arakan basin, India. Journal of Petroleum Exploration and Production Technology 8(2), 323–340.

Larionov, V.V. (1969). Radiometry of boreholes (in Russian). Nedra, Moscow.

Le, D.H. & Reynolds, A.C. (2012). Optimal Choice of a Surveillance Operation Using Information Theory. In Proceedings of the 13th Euro- peak Conference of the Mathematics of Oil Recovery, ECMOR XIII, Biarritz, France.

Le, D.H. & Reynolds, A.C. (2014). Estimation of Mutual Information and Conditional Entropy for Surveillance Optimization, SPE Journal, 648-663.

Lippmann, R.P. (1987). An introduction to computing with neural networks, ASSP Magazine, April 4–22.

Lucia, F. J. (2007). Carbonate reservoir characterization: An integrated approach. Springer Science & Business Media.

Luthi, S.M., (2001). Geological Well Logs: Their Use in Reservoir Modeling: Springer-Verlag, Berlin, 333 P.

Mardi, M., Nurozi, H., & Edalatkhah, S., (2012). A Water Saturation Prediction Using Artificial Neural Networks and an Investigation on Cementation Factors and Saturation Exponent Variations in an Iranian Oil Well, Petroleum Science and Technology, 30:4, 425-434, DOI: 10.1080/10916460903452033

Masoud, N., (2004). Soft Computing-Based Computational Intelligent for Reservoir Characterization, J. Of Expert Systems with Applications 26 (1), Pp. 19-38.

Masoudi, P., Aïfa, T., Memarian, H., & Tokhmechi, B. (2017). Uncertainty assessment of volumes of investigation to enhance the vertical resolution of well-logs. Journal of Petroleum Science and Engineering 154, pp. 252-276.

Masoudi, P., Arbab, B., & Mohammadrezaei, H. (2014). Net pay determination by artificial neural network: Case study on Iranian offshore oil fields. Journal of Petroleum Science and Engineering, 123, 72-77.

Mhaskar, H. (1993). Approximation Properties of a Multilayered Feedforward Artificial Neural Network. Advances in Computational Mathematics 1.1 (1993): 61-80.

Miah, M. I., & Howlader, M. F. (2012). Prediction of Formation Water Resistivity from Rwa Analysis of Titas Gas Field Using Wireline Log Data. Journal of Petroleum and Gas Exploration Research (JPGER), 2(4), 57-60.

Miah, M. I., (2014). Porosity Assessment of Gas Reservoir Using Wireline Log Data: A Case Study of Bokabil Formation, Bangladesh, Elsevier-Procedia Engineering 90, (2014) Pp. 663 – 668.

Miah, M. I., Deb, P. K., Rahman, M. S., & Hossain, M. E. (2017). Application of Memory Concept on Petroleum Reservoir Characterization: A Critical Review. Society of Petroleum Engineers. doi:10.2118/187676-MS

Miah, M.I., & Tamim, M., (2015). Hydrocarbon Saturation Assessment of Thick Shaly Sand Reservoir Using Wireline Log Data: A Case Study, The 10th International Forum on Strategic Technology 2015 June 3 - June 5, 2015 Universitas Gadjah Mada, Indonesia.

Mohaghegh, Arefi, Bilgesu, Ameri, & Rose. (1995). Design and Development of an Artificial Neural Network for Estimation of Formation Permeability. SPE Computer Applications 7.6 (1995): 151-54.

Mohaghegh, S., Arefi, R., Ameri, S., Aminiand, K., & Nutter, R. (1996). Petroleum reservoir characterization with the aid of artificial neural networks. Journal of Petroleum Science and Engineering, 16(4), 263-274.

Mohaghegh, S., Goddard, C., Popa, A., Ameri, S., & Bhuiyan, M., (2000d). Reservoir characterization through synthetic logs. In: Proceedings Society of Petroleum Engineers (SPE) Eastern Regional Meeting, Morgantown, West Virginia, Paper 65675

Mohaghegh, S.D. (2000b). Virtual Intelligence Applications in Petroleum Engineering: Part 2–Evolutionary Computing, paper SPE 61925, JPT (October 2000) 40.

Mohaghegh, S.D., (2000a). Virtual Intelligence Applications in Petroleum Engineering: Part 1–Artificial Neural Networks, paper SPE 58046, JPT (September 2000) 64.

Mohaghegh, S.D., (2000c). Virtual Intelligence Applications in Petroleum Engineering: Part 3–Fuzzy Logic, paper 62415, JPT (November 2000) 82.

Mollajan, A. (2015). Application of Local Linear Neuro-fuzzy Model in Estimating Reservoir Water Saturation from Well Logs. Arabian Journal of Geosciences8.7 (2015): 4863-872.

Mollajan, A., Memarian, H., & Jalali, MR. (2013). Prediction of reservoir water saturation using support vector regression in an Iranian carbonate reservoir. 47th US Rock Mechanics/Geomechanics Symposium, San Francisco, CA, USA, 3: 1872–1877.

Olatunji, S.O., Selamat, A., & Abdulraheem, A., (2011). Modeling the permeability of carbonate reservoir using type-2 fuzzy logic systems. Comput. Ind. 62 (2), 147–163.

Onalo, D., Adedigba, S., Khan, F., James, L. A., & Butt, S. (2018). Data driven model for sonic well log prediction. J. of Petro. Sci. and Eng., 170, 1022-1037.

Pascoal, C., Oliveira, M. R., Pacheco, A., & Valadas, R. (2017). Theoretical evaluation of feature selection methods based on mutual information. Neurocomputing, 226, 168-181.

Poulton, M., (2002). Neural networks as an intelligence amplification tool: a review of applications. Geophysics 67 (3), 979–993.

Poupon, A., Loy, M.E., & Tixier, M.P., (1954). A Contribution to electrical log interpretation in shaly sands. J. Petrol. Technol. 6 (6), 27–34.

Razavi, S., & Tolson, B.A. (2011). A New Formulation for Feedforward Neural Networks. IEEE Transactions on Neural Networks 22.10 (2011): 1588-598.

Rolon, L., Mohaghegh, S. D., Ameri, S., Gaskari, R., & McDaniel, B. (2009). Using artificial neural networks to generate synthetic well logs. Journal of Natural Gas Science and Engineering, 1(4-5), 118-133.

Santamarina, J. C., Garcia, A., Hakiki, F., Park, J., & Zhao, B. (2019b). Multiphysics low-perturbation methods for sediment characterization and process monitoring. ICEG2019, 21-24 October 2019, Al-Ain, UAE.

Santamarina, J. C., Park, J., Terzariol, M., Cardona, A., Castro, G. M., Cha, W., ... & Shen, Y. (2019a). Soil Properties: Physics Inspired, Data Driven. In Geotechnical Fundamentals for Addressing New World Challenges (pp. 67-91). Springer, Cham.

Schlumberger (1998). Log Interpretation Principles/Applications, 7th printing, Houston, p. 235.

Schlumberger, (1972). Log Interpretation; Vol 1-Principles, New York, Texas.

Seyyedattar, M., Zendehboudi, S., & Butt, S., (2019a). Technical and non-technical challenges of development of offshore petroleum reservoirs: Characterization and production. Natural Resources Research, 1-43 (in press).

Seyyedattar, M., Zendehboudi, S., & Butt, S., (2019b). Molecular dynamics simulations in reservoir analysis of offshore petroleum reserves: A systematic review of theory and applications. Earth-Science Reviews 192, 194-213.

Shannon, C.E., (1948). The mathematical theory of communication (parts 1 and 2). Bell Syst. Tech J. vol. 27, pp. 379-423 and 623-656, 1948.

Shedid, A. S., & Saad, M.A., (2017). Comparison and sensitivity analysis of water saturation models in shaly sandstone reservoirs using well logging data, Journal of Petroleum Science and Engineering, Volume 156, 2017, 536-545.

Shedid, S., (2018). A New Technique for Identification of Flow Units of Shaly Sandstone Reservoirs. Journal of Petroleum Exploration and Production Technology 8.2 (2018): 495-504.

Shedid-Elgaghah, S.A., Tiab, D., & Osisanya, S., (2001). New approach for obtaining Jfunction in clean and shaly reservoirs using in-situ measurements. J. Can. Petrol. Technol. 40, 30–37. July.

Shokir, E. M. (2004). Prediction of the Hydrocarbon Saturation in Low Resistivity Formation via Artificial Neural Network. Society of Petroleum Engineers.

Simandoux, P., (1963). Dielectric measurements on porous media, application to the measurements of water saturation: study of behavior of argillaceous formations. Revue de l'Instiut Francase du Petrol l8, 193–215 (Translated in shaly snads reprint volume, SPWLA, Housten, pp. 97–124).

Sylvester, O., Bibobra, I., & Ogbon, O. N., (2015). Well Test and PTA For Reservoir Characterization of Key, American Journal of Engineering and Applied Sciences, Volume 8, Issue 4, Pp. 638-647.

Talebi, R., Ghiasi, M., Talebi, H., Mohammadyian, M., Zendehboudi, S., Arabloo, M., & Bahadori, A. (2014). Application of Soft Computing Approaches for Modeling Saturation Pressure of Reservoir Oils. Journal of Natural Gas Science and Engineering 20 (2014): 8-15.

Tiab, D., Donalson, E.C., 2015. In: Petrophysics: theory and practice of measuring reservoir rock and fluid transport properties. Gulf professional publishing.

Vergara, J. R., & Estévez, P. A. (2014). A review of feature selection methods based on mutual information. Neural computing and applications, 24(1), 175-186.

Wang, B., Wang, X. & Chen, Z., (2013). A Hybrid Framework for Reservoir Characterization Using Fuzzy Ranking and An Artificial Neural Network, Computers & Geosciences 57 (2013) Pp.1–10.

Waxman, M.H., & Smits, L.J.M., (1968). Electrical conductivities in oil-bearing shaly sands. Society of Petroleum Engineers Journal 8, 107–122.

White, A., Molnar, D., Aminian, K., & Mohaghegh, S., (1995). The Application of artificial neural networks to zone identification in a complex reservoir. SPE 30977, Proceedings, SPE Eastern Regional Conference and Exhibition September 19–21, 1995, Morgantown, West Virginia.

Wong P.M., Aminzadeh F., & Nikravesh M., (2002). Intelligent Reservoir Characterization. In: Wong P., Aminzadeh F., Nikravesh M. (eds) Soft Computing for Reservoir Characterization and Modeling. Studies in Fuzziness and Soft Computing, vol 80. Physica, Heidelberg

Worthington, P.F., (1985). The evolution of shaly sand concepts in reservoir evaluation. The Log Analyst 26, 23–40.

Yang, S. & Wei, J., (2017). Fundamentals of Petrophysics, Spinger Mineralogy, Pp. 501, DOI 10.1007/978-3-662-53529-5.

Yao, Vuthaluru, Tadé, & Djukanovic. (2005). Artificial Neural Network-based Prediction of Hydrogen Content of Coal in Power Station Boilers. Fuel 84.12-13 (2005): 1535-542.

Yao, X. & Liu, Y. (1997). A new evolutionary system for evolving artificial neural networks. IEEE Transactions on Neural Networks 8, 694–713.

Zendehboudi, S., Rezaei, N., & Lohi, A., (2018). Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review. Applied energy 228, 2539-2566.

CHAPTER 4: LOG DATA-DRIVEN MODEL AND FEATURE RANKING FOR WATER SATURATION PREDICTION USING MACHINE LEARNING APPROACH

Preface

A version of this chapter has been published in the Journal of Petroleum Science and Engineering (PETROL 107291). I am the primary author of this article and prepared the first version of the manuscript. The co-author Dr. Salim Ahmed provided technical assistance to find the research scopes, as well as developing the knowledge gap. Dr. Ahmed reviewed the first draft of the manuscript. Co-author Dr. Sohrab Zendehboudi also reviewed and provided valuable insights on how to improve the model development and result analysis subsections in the manuscript to a journal-level publication. I have revised the manuscript based on the feedback from co-authors. Finally, all authors contributed to prepare the final version of the manuscript as per the reviewers' and editor's comments.

Abstract

Log-based reservoir characterization is one of the widely used techniques to estimate the reservoir properties and make decisions for hydrocarbon production. Use of the machine learning tools is becoming a more accessible approach for data-driven model development. The objective of this research is to identify and rank the most contributing log variables for estimation of water saturation using the machine learning tools. The multilayer perception artificial neural network (MLP-ANN) and kernel function-based least-squares support vector machine (LS-SVM) techniques are employed to develop predictive models for water saturation. The model can capture the non-linear behavior and high-dimensional complex relationships among real field log data variables. Based on the prediction performance of the models, the Levenberg-Marquardt algorithm-based MLP-ANN and the radial kernel function-based LS-SVM model optimized with coupled simulated annealing optimization technique were found to perform better compared with other models. The MLP-ANN and radial kernel function-based LS-SVM approaches lead to the same feature ranking of predictor variables. It was found that the significance order of influential (higher to lower) log variables are the true resistivity, bulk density, neutron porosity, photoelectric factor, gamma-ray, and sonic travel time based on their relative contribution to the water saturation. The strategy introduced in this study assists to forecast water saturation with a relatively few number of log variables, and thus, reduces the number of necessary logs to run during exploration, considerably lowering the exploration costs.

Keywords: MLP-ANN; LS-SVM; Model accuracy; Log variables; Feature ranking; Cost-effective.

4.1 Introduction

4.1.1 Background

Reservoir characterization plays an important role in making decisions regarding exploration, reservoir simulation as well as the management of a reservoir. It also helps in decision making for hydrocarbon perforation to production, and improving the reliability of the models for determination of reservoir properties. Thus, it focuses on not only the understanding of past reservoir properties, but also predicting the future reservoir conditions. The ultimate target of reservoir characterization is to show the nature of rock-fluid properties and to develop a reservoir model with a high accuracy and minimal uncertainty. It is a continuous process that can be accomplished using information sources from several disciplines such as geological information, petro-physical knowledge, well testing, geophysics, reservoir engineering, and production history (Aminian and Ameri, 2005; Wong et al., 2005; Sylvester et al., 2015; Miah and Tamim, 2015; Zhang and Xu, 2016; Yang and Wei, 2017; Miah et al., 2017; Movahhed et al., 2019). The most extensive technique for reservoir characterization is petrophysical analysis using coring and well logging data.

Water saturation is one of the most vital reservoir properties that enables researchers and engineers to determine perforation depth for hydrocarbon production in both offshore and onshore fields. Although the laboratory-based direct method for water saturation (Sw) estimation is more accurate, it is time-consuming and expensive. A limited number of samples can capture a few limited depth-intervals of a targeted production/injection well (Adeniran et al., 2009). Due to the complex geological behavior and heterogeneities of

reservoirs, a large set of samples are required to properly characterize an underground formation. Nowadays, log-based petrophysical reservoir characterization approach is one of the widely used methods to obtain the reservoir properties such as hydrocarbon saturation, porosity, permeability, hydraulic flow unit, and reservoir quality index. To estimate water saturation, many researchers have attempted to develop models using a variety of deterministic approaches (Amiri et al., 2015a; Si et al., 2016). Over the past few decades, numerous studies have been performed to introduce empirical models for estimating water saturation using petrophysical logs such as resistivity, sonic, density, and neutron porosity logs. For instance, the common models are the Archie's formula (1942) for clean sand zones; DeWittee model (1950) for dispersed shale; Simandoux's model (1963) for shaly sand formations; Waxman and Smits (1968) for shaly sand reservoirs; Poupon et al. model (1954) for laminated shaly sand formations; Poupon and Leveaux (1971) for Indonesia's laminated shaly formations; and Schlumberger (1972) for shaly sand reserves. Applicability of these models is limited to the nature of the formation (lithology), and lithological parameters such as tortuosity factor, cementation, and saturation exponent. Shedid and Saad (2017) summarized the existing petrophysical water saturation models where their drawbacks were highlighted.

Nowadays, the connectionist tools or machine learning (ML) approaches such as artificial neural network (ANN) and least squares support vector machine (LS-SVM) are becoming more popular for data-driven model analysis and prediction of rock properties to save the experimental and operational expenses. The connectionist tools do not use the geological characteristics and lithological parameters (e.g., lithology type, saturation exponent, and

cementation factor) in the models and overcome the drawbacks (data range and generalization) of the existing empirical models for obtaining reservoir rock properties. ANN and LS-SVM are powerful tools in the petroleum industry to capture the uncertainties and non-linear behaviors of the input variables.

4.1.2 Literature survey on ANN and LSSVM application

The intelligent systems and/or machine learning tools such as ANN, LS-SVM, functional network (FN), generic algorithms (GA), imperialist competitive algorithm (ICA), fuzzy decision tree (FDT), fuzzy logic (FL), particle swarm optimization (PSO), neuro-fuzzy inference system (ANFIS), and recurrent neural network (RNN) models have been extensively employed for industrial applications to capture high dimensional non-linear data, find best patterns, and predict target variables (Al-Bulushi et al., 2012; Anifowose et al., 2013; 2014; 2015; Mollajan, 2015; Amiri et al., 2015b; Dongxiao et al., 2018; Onalo et al., 2018; Zendehboudi et al., 2018). Among those approaches, ANN and SVM are more popular in applications related to the petroleum industry such as reservoir rock characterization, (Wong et al., 2005; Ghaffarian et al., 2014; Anifowose et al., 2014; Esmaeili et al., 2019;), rock mechanical properties and formation evaluation (Salehi et al., 2017; Tariq et al., 2016, 2017; Onalo et al., 2018; Wang and Peng, 2019), drilling optimization (Amer et al., 2017; Ashrafi et al., 2019), screening criteria and performance of enhanced oil recovery techniques (Ahmadi et al., 2015, Shan et al., 2018), compartmentalized reservoir analysis and multi-phase fluid flow modeling in oil reservoirs (Esmaeilzadeh et al., 2019, 2020; Temirchev et al., 2020), reservoir history matching and

optimization (Hutahaean et al., 2015, 2019), and production optimization (Jreou, 2012; Ebrahimi and Khamehchi, 2016).

Due to the unique features such as simple architecture, ease of training, and stability in convergence, LS-SVM and ANN are widely utilized as powerful computational tools to improve predictions of the reservoir properties such as water saturation (Sw), porosity (PHI), and permeability (K). It was found from the literature survey that many scholars including Erofeev et al. (2019), Anifowose et al. (2019), Rafik and Kamel (2017), Anifowose et al. (2014), Ahmadi et al. (2014), Tahmasebi and Hezarkhani (2012), Saffarzadeh and Shadizadeh (2012), Ahrimankosh et al. (2011), Karimpouli et al. (2010), Helmy et al. (2010), Basbug and Karpyn (2007), and Smaoui and Garrouch (1997) implemented different tools including the ANN and LS-SVM algorithms to obtain the porosity and permeability using core and wireline logging data. The machine learning-based models are beneficial for various chemical and energy industries due to their capability for handling of high dimensional non-linear data for modelling purposes. Feature selection and ranking of the input log variables are also a crucial task to maximize the model performance and to avoid the redundancy of input variables for a model development.

However, only a few limited attempts have been made to provide insight into feature selection in reservoir characterization (Helmy et al., 2010, Saffarzadeh and Shadizadeh, 2012; Anifowose et al., 2014, 2016; Miah et al., 2019). Researchers used the available wireline logging data such as self-potential (SP) or gamma-ray (GR), deep induction log (ILD) or true resistivity (RT), bulk density (RB), neutron porosity (NP), photoelectric

factor (PE), and sonic compressional travel time (DT) to estimate reservoir properties such as water saturation. To the best of the authors' knowledge, only a few studies have been conducted to predict water saturation using machine learning techniques. Based on the literature, Table 4.1 lists most of the available research works as well as a brief information on development approach for determination of the water saturation using machine learning (ML) tools. To assess the model performance, the statistical indicators such as the mean square error (MSE), the root mean square error (RMSE), the coefficient of determination (R^2), the average percentage relative error (APRE), the absolute average percentage relative error (AAPRE), and the standard estimation error (SEE) have been used.

Helle and Bhatt (2002) used the ANN tool to predict water (fluid) saturation using log input variables. It was concluded that ANN gives results with higher accuracy than the petrophysical models; however, they did not incorporate the shale effect to diagnose the clastic reservoir. Shokir (2004) used an ANN model to predict the water (hydrocarbon) saturation of a low resistivity shaly sand reservoir using well log data. Based on the results, the correlation coefficient was found to be close to one for the prediction of water saturation; however, the study did not investigate the feature ranking of the variables.

Table 4.1: Implications of machine learning (ML) tools for water saturation (Sw) prediction in reservoir characterization.

Authors, year	Input log variables	Output variables	ML approach	Statistical parameter (s)	Feature ranking	Comments/Remarks
Helle and Bhatt (2002)	RT, NP, DT, and RB	Sw	ANN	RMS	No	• Used a three-layer ANN model to predict fluid saturation. • ANN gives results with greater accuracy than the petrophysical model • Did not include GR as an input log.
Shokir (2004)	GR, SP, RT, NP, and RB	Sw	ANN	R^2	No	• It is a powerful tool to predict hydrocarbon saturation. • Based on the results, the R^2 value is close to 1 for predicting Sw.
Zhao et al. (2006)	Resistivity and density logs	Sw	SVM	No	No	• Shale volume was calculated using the gamma-ray log. • Archie's formula was used for Sw calculation.
Al-Bulushi et al. (2009)	RT, NP, RB, PE, and core data	Sw	ANN	RMS	Yes	• It was found that ANN offers the superior results, compared to the conventional statistical regression methods. • It was concluded that RT is the main factor with a 40 % contribution to the estimation of Sw. • Authors did not include GR as an input log to consider the shale effect.
Adeniran et al. (2009)	GR, RT, NP, RB, PE, and core data	PHI and Sw	FN and ANN	RMS and R^2	No	• The model network was analyzed through a trial and error approach for selecting the number of hidden layers and number of nodes. • They also examined various learning algorithms and transfer functions.
Mardi et al. (2012)	RT, NP, RB, DT, and core porosity	Sw, m, and n	ANN	RMS and R^2	No	• They proposed ANN model by employing both core and log data. • It was found that the ANN-based model is more reliable than the dual water model.
Kamalyar et al. (2011)	Core PHI, K, and height above free water level	Sw	ANN	APRE, AAPRE, RE, and SEE	No	• The researchers proposed a model that offers more accurate results than the available published models. • They did not consider any real field log data.

Kenari and Mashohor (2013)	SP, RT, NP, RB, PE, and effective PHI	Sw	ANN, FL, and ANFIS	MSE, R^2, and MSPE	No	• It was concluded that ANFIS is more accurate intelligence algorithm than the other smart methods. • This study used SP lithology log instead of GR log.
Mollajan et al. (2013)	RT, NP, RB, and DT	Sw	ANN and SVR	RMSE, MAE, and R^2	No	• It was found that the RBF-based SVR model gives better result than ANN and dual water model proposed by Clavier et al. (1984). • They did not include GR to find the shale effect.
Amiri et al. (2015c)	GR, RT, NP, RB, and effective NP-DP	Sw	ICA-ANN	MSE, RME, and R^2	No	• It was found that ICA-based ANN model performs better with a limited number of parameters, compared to the conventional ANN model.
Bageri et al. (2015)	Resistivity logs and core data	Sw	ANN and FL	MAPE, MIPE, RE, and R^2	No	• It was claimed that the artificial intelligence models reduce the long time required for analytical petrophysical tools. • The ANN and FL performed better with a higher precision than other studied models.
Gholanlo et al. (2016)	NP, DT, RB, and core data	Sw	RBF-ANN	MSE and R^2	No	• Log data were collected from carbonate reservoirs. • It was concluded that ANN is more reliable compared to Archie's method.
Baziar et al. (2018)	GR, ILD, NP, DP, and DT	Sw	DT, ANN, and SVR	AAE, RMSE, and R^2	No	• Two hidden layers were used to construct the MLP model using the sigmoid function and conjugate gradient algorithm. • Grid search method was employed to optimize the parameters of SVM.
Hamada et al. (2018)	GR, LLD, RB, PE and NP	PHI and Sw	ANN	MSE	No	• Developed an MLP based ANN model to predict output variables. • The researchers did not discuss about the relative importance of the model's variables.
Khan et al. (2018)	GR, RT, Rxo, NP, RB, and Caliper log	Sw	ANN and ANFIS	R^2	No	• It was concluded that ANFIS has a greater performance than ANN. • This study did not include a part of the most important input parameters for predicting Sw.

Al-Bulushi et al. (2007) developed an ANN model with a resilient back-propagation learning algorithm to predict the water saturation using well logs and Dean-Stark core data. The researchers did not include GR log as an input variable to take into account the shale effect while determining water saturation. Rolon et al. (2009) investigated the application of ANN with three layers using wireline logs to generate the synthetic logs such as resistivity, bulk density, and neutron logs. It was reported that the ANN model results in better predictions to generate resistivity log using density and neutron logs. Kamalyar et al. (2011) developed an ANN model using core data without considering real field log data; the smart model exhibited more accurate results than the available published models (Kamalyar et al., 2011). Mardi et al. (2011) developed an ANN approach by employing both core and well log data to investigate the saturation exponent and cementation factor of a water saturation model. The researchers utilized the data samples from the carbonate rocks of Sarvak formation, Iran. It was found that ANN is more reliable than the dual water model. However, the relative importance of log variables was not discussed in their research. Baneshi et al. (2013) investigated the sensitivity of log variables for predicting reservoir rock parameters using Levenberg-Marquardt training function in an ANN model. They found that the ANN is a satisfactory tool to predict not only the reservoir porosity and water saturation but also the petrophysical index using well logs. Anifowose et al. (2014) employed the concept of information gain to study the relative importance of feature attributes for the decision tree approach. It was concluded that density volume of grain is an appropriate predictor variable for rock porosity estimation. It is worth noting that some

logs (e.g., porosity, density, and micro-spherically focused resistivity logs) offer good data attributes to predict reservoir rock permeability.

As mentioned earlier, a majority of researchers did not take into account the relative importance of predictor variables while forecasting the water saturation. Gholanlo et al. (2016) also used a kernel function-based ANN model to estimate water saturation by employing log predictor variables. In the model, they used the algorithm of Levenberg-Marquardt to optimize the magnitude of weights, representing the strength of connections between the units in the layers. Baziar et al. (2018) investigated a comparative analysis using different machine learning approaches such as decision tree forest, tree boost, multilayer perception network, and kernel-based support vector machine (SVM) to forecast the water saturation of a gas reservoir. The researchers also analyzed the model performance using different scenarios of data set. According to their analysis, it was concluded that radial basis function (RBF) kernel SVM model is more reliable as it gives less error than other models in predicting water saturation. Hamada et al. (2018) introduced an ANN model by adopting the tans and log sigmoid transfer function, and 2 hidden layers with sixteen and five neurons, respectively, and 1 output layer. In their study, the porosity and water saturation were estimated, and then the predictions were compared with the core data. Khan et al. (2018) evaluated the predictive performance of the ANFIS while obtaining water saturation; the suggested approach provided better results than the ANN model. They did not identify the most influential input variables in the model while estimating water saturation. Hamada et al. (2019) calculated the water saturation and porosity with ANN and conventional petrophysical models using logging data. It was concluded that the ANN

approach yields a better performance and/or higher precision than the conventional petrophysical approaches for the clastic reservoirs. Recently, Miah et al. (2019) constructed a single hidden layer perception-based ANN model with the Levenberg Marquardt training algorithm to obtain water saturation (Sw). They concluded that true resistivity and formation bulk density log variables are the significant parameters for estimation of Sw. The impact of various learning algorithms on the performance of data-driven predictive model with linearized predictor variables was not investigated in their study.

The above literature review suggests that more reliable data-driven models are still needed to improve the prediction performance/ accuracy in reservoir characterization. To the best of our knowledge, data-driven predictive model and feature ranking of logging variables to determine the water saturation by coupling logging data and machine learning tools have not been investigated systematically. The ranking of logging parameters appears to be a serious challenge for not only petroleum engineers, but also petrophysicists to forecast the water saturation using the wireline log data with sufficient precision. The well log data is expensive to collect and represent the entire reservoir profile. The current research work aims to fill in the knowledge gap by finding the most contributing predictor variables and feature ranking of log variables, according to their relative significance obtained from ANN and LS-SVM while predicting water saturation. Indeed, this is the first time that these two predictive approaches (ANN and LS-SVM) are used to forecast the water saturation profile in the shale sandy formation and to rank the input parameters. Systematic investigation about dependency of predictor variables to the output parameter is another important

objective of this study. It is believed that the research strategies employed in this study can be less time consuming and cost-effective approaches for efficient reservoir characterization as well as formation evaluation.

The remaining of the paper is structured as follows: Section 2 includes data collection and processing, model performance strategy, brief background and formulation of machine learning models, and feature ranking strategy. The results and discussions are presented in Section 3. In the last section, the conclusions and recommendations are briefly listed.

4.2 Theory and Research Methodology

This section covers the log data collection, normalization process and, model performance analysis procedures. It also describes the model formulation and prediction procedure using ANN and LS-SVM tools as well as feature ranking of log variables with both strategies.

4.2.1 Data collection and processing

The most common wireline logs are resistivity log, porosity log, and lithology log, which are utilized to identify the lithology type and hydrocarbon-bearing zone, and to estimate petrophysical properties. The gamma-ray (GR) log is a complementary log that can be used to identify the lithology and measure the amount of radioactive elements (API) and clay fraction (shale content, Vsh). The density log measures the electron density (bulk density, RB) as well as photoelectric effect (PE) factor while the neutron log counts the hydrogen concentration (NP) of a formation. The sonic log helps to measure the acoustic wave in the form of compressional travel (DT) time or shear slowness. On the other hand, the resistivity

log measures the conductivity (inverse of resistivity, RT) in the rock. The detailed information about the major types of logging principles and their application can be found in the open sources (Bassiouni, 1994; Schlumberger, 1972; Asquith and Krygowski, 2004; Rolon et al. 2009; Miah and Howlader, 2012; Miah, 2014). These log variables are utilized in this study to meet the research goals; field data collected from a hydrocarbon reservoir located in the Bengal basin are the input data for this research investigation. The available field well logs data (GR, RT, RB, NP, PE, and DT) are used as predictor (input) variables while the target variable is the water saturation in the reservoir. The log data quality process is also conducted to ensure about the reliability of each log dataset variables through checking the depth shift and borehole conditions, and comparing hole size with caliper log.

The procedure to determine water saturation obtained using wireline log data is given by Miah et al. (2019) where a series of standard log interpretation steps need to be followed. The literature also reports proper procedures for log interpretation, and estimation of water saturation (Bassiouni, 1994; Asquith and Krygowski, 2004). A total of 182 data samples are collected. The field log data points are divided into training, testing, and validation phases to develop the machine learning system and train the model. All the programming tasks related to this study are carried out using MATLAB programming environment. The statistical information of the data samples is given in Table 4.2.

Table 4.2: Summary of the statistical values of the used data.

Parameters	Max.	Min.	Mean	St. Dev.	Sample var.	Skewness	Kurtosis
DT (µs/ft)	97.40	85.84	92.90	2.43	5.89	-0.2173	0.2263
GR (API)	157.82	76.28	100.19	13.87	192.51	1.58	3.78
RT (ohm-m)	39.70	13.70	22.67	4.96	24.58	0.7432	1.57
RB (g/cc)	2.53	2.30	2.37	0.0425	0.0018	1.559	2.714
PE (barns/e)	4.456	2.876	3.2294	0.2539	0.0645	1.7812	5.0704
NP (v/v)	0.2039	0.1455	0.17	0.013	0.0002	0.819	0.035
Sw (v/v)	0.5426	0.2774	0.3950	0.0515	0.0027	0.6296	0.8851

All available log variables are normalized (scaled) in the range of 0 (zero) to 1 (one) which can help to obtain better results from the machine learning approaches. The following equation is used for normalizing the variables (Wang et al., 2013; Ashena and Thonhauser, 2015):

$$X_{(i,nor)(k)} = \frac{x_{i(k)} - x_{(i,min)}}{x_{i,max} - x_{i,min}} \quad (4.1)$$

where i =1,2,…N with N being the number of variables and k =1,2,…,n with n being the number of data for the variable; X_{nor} stands for the normalized value of x (input variable); and x_{min} and x_{max} refer to the minimum and maximum magnitudes of the log variable, x_i.

4.2.2 Model performance indicators and accuracy

There are several statistical indicators used to analyze the model performance, to find the model accuracy, and to select a proper algorithm (Olatunji et al., 2014; Akande et al. 2015a-

b; 2016; Ashrafi et al., 2019). In the ANN approach, it is important to assess the optimum number of hidden layers and neurons in each hidden layer. The statistical performance indices are used to quantify the difference between the actual and predicted outcomes for finding the optimum number of hidden layers. In this study, five statistical indicators are utilized to evaluate the predictive model performance. The indicators are RMSE, R^2, average absolute percentage error (AAPE), maximum average percentage error (MAPE), and performance index (PI). The mathematical expressions for all performance indices are listed below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_{t,i} - Y_{p,i}\right)^2} \tag{4.2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_{t,i} - Y_{p,i}\right)^2}{\sum_{i=1}^{n}\left(Y_{t,i} - Y_{t,mean}\right)^2} \tag{4.3}$$

$$AAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{(Y_{t,i} - Y_{p,i})}{Y_{t,i}} \tag{4.4}$$

$$MAPE = Max.\left|\frac{(Y_{t,i} - Y_{p,i})}{Y_{t,i}}\right| * 100 \tag{4.5}$$

$$PI = \left(R + \frac{VAF}{100} - RMSE\right) \text{ where } VAF = \left\{1 - \frac{Var(Y_t - Y_p)}{Var(Y_{p)}}\right\} \tag{4.6}$$

In the above equations, *n* indicates the total number of samples; Y_t is the target (actual) output variable; $Y_{t,mean}$ introduces the mean value of Y_t; VAF stands for variance account factor; and Y_p represents the predicted output variable. The accuracies of the data-driven models have been analyzed in the scale of the low or high value of statistical indices.

Lower values of RMSE, AAPE and MAPE, and higher magnitudes of R^2 and PI imply a model with a greater precision and reliability.

4.2.3 Fundamentals and development of ANN model

The artificial neural network (ANN) approach is a computing system based adaptive learning model, which was developed from the concept of the biological neural network (Lippmann, 1987; Yao and Liu, 1997; White et al., 1995; Mohaghegh et al., 1996; Poulton, 2002). The ANN network contains an input layer, one or more hidden layers, transfer functions, and an output layer. The basics and drawbacks/ limitations of ANN models can be found in the available literature (Faucett, 1994; Lawrence et al., 1997; Rolon et al., 2009; Zendehboudi et al., 2018).

A major advantage of ANN is its ability to find highly complex nonlinear relationships among variables. The feedforward neural network can be adopted by a single-layer perception or multi-layer perceptron (MLP), and radial basis function networks (Ali, 1994; Mohaghegh et al., 1994; 1996; Razavi and Tolson, 2011; Akande et al., 2015a; Adedigba et at., 2017). The data used in the ANN are typically divided into three main subsets, namely, training, testing, and validation data series. The MLP is one of the effective and conceptually useful feedforward neural network approaches that can be employed in estimation of reservoir characteristics, leading to attaining reliable results (Rogers et al., 1995; Huang et al., 1996; Fung et al., 1997; Helle et al., 2001; Helle and Bhatt, 2002; Zendehboudi et al., 2018). The deep learning MLP-ANN model consists of at least four layers with one input layer, at least two hidden layers, and one output layer for target

variables. A typical MLP-ANN architecture of deep learning feedforward network is depicted in Figure 4.1. The hidden layer is characterized by several numbers of computation connectionist units called neurons (e.g., N1, N2, N3, and N4).

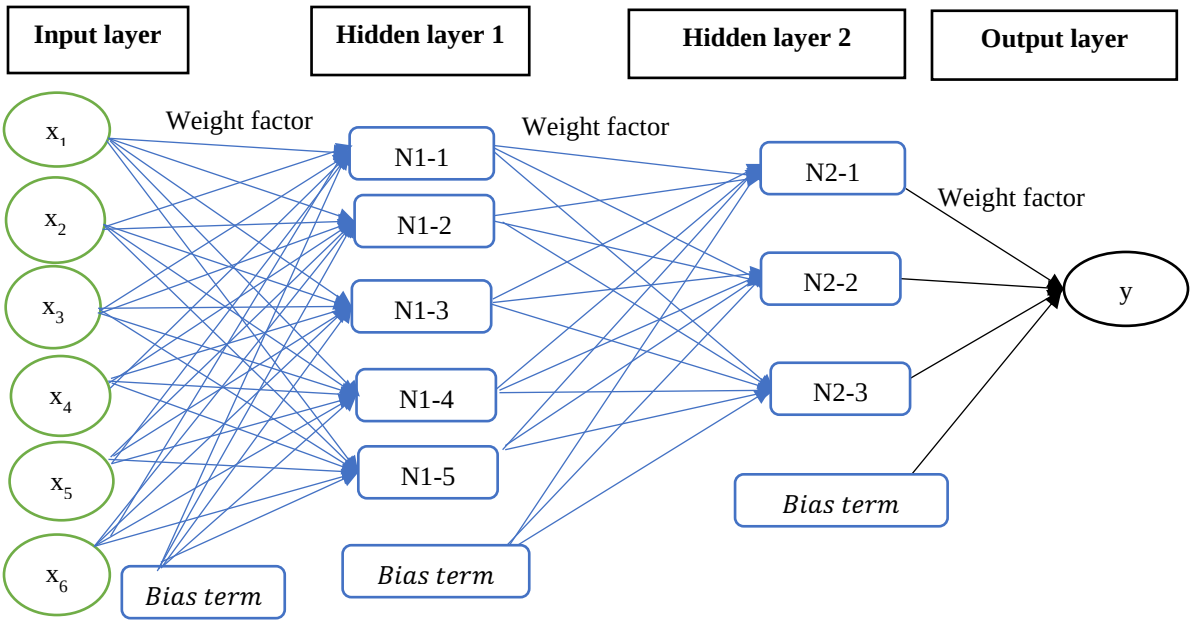

Figure 4.1: Schematic of a typical multilayer perception-based ANN architecture.

The generalized mathematical expression for the deep learning ANN is given below (Ceryan and Can, 2018):

$$y_m = f_o\left[\sum_{j=1}^{m} \omega_{jm} f_h\left(\sum_{i=1}^{n} \omega_{ij} x_i + b_j\right) + b_m\right] \tag{4.7}$$

In equation (4.7), y_m stands for the output variable (s); x_i is the vector of target variables (e.g., $i = 1, 2, 3, 4, 5, 6,\ldots,n$); b_m symbolizes the bias term for output layers; ω_{ij} represents

the connection weight on the link from *i* to *j* node between the input and hidden layers; *m* refers to the number of hidden nodes; and *n* resembles the number of input variables. f_h and f_o are the transfer function for the hidden layer and output layer, respectively.

Selection of the number of connectionist neurons and the hidden layers depend on the nonlinearity of the problem that is being studied (Hamedi et al. 2019). As illustrated in Figure 4.1, the input layer of network is linked to the hidden layer by a defined weight factor while it is constantly adjusted by a training algorithm based on the field data. Both hidden layers and the output layer of the network are all connected together in a forward direction (Esene et al. 2019). The MLP network is solely dependent on the adjustment of the weight factor between the layers. The output layer is associated with a 'purelin' transfer function and the hidden layers are assigned with an activation (transfer) function such as Heaviside (threshold), precise linear, and sigmoidal (Gaussian, logistic and arctan) function. The purpose of the back-propagation algorithm is to find the least error surface, while it calculates the local gradient in the error surface and afterwards updates the weight magnitudes along the direction of the steepest local gradient (Gardner and Dorling,1998; Adedigba et al., 2017). In each iteration of the back-propagation process, the forward and backward pass sweep is performed repeatedly until the output variable becomes the same as the target variable within an allowable predetermined tolerance level in the model (Basheer and Hajmeer, 2000). A flow chart shown in Figure 4.2 presents the ANN model development steps. The detailed procedure of feedforward backpropagation can be found in the literature (Mazur and Marry, 2015; Adedigba et al., 2017; Zendehboudi et al., 2018).

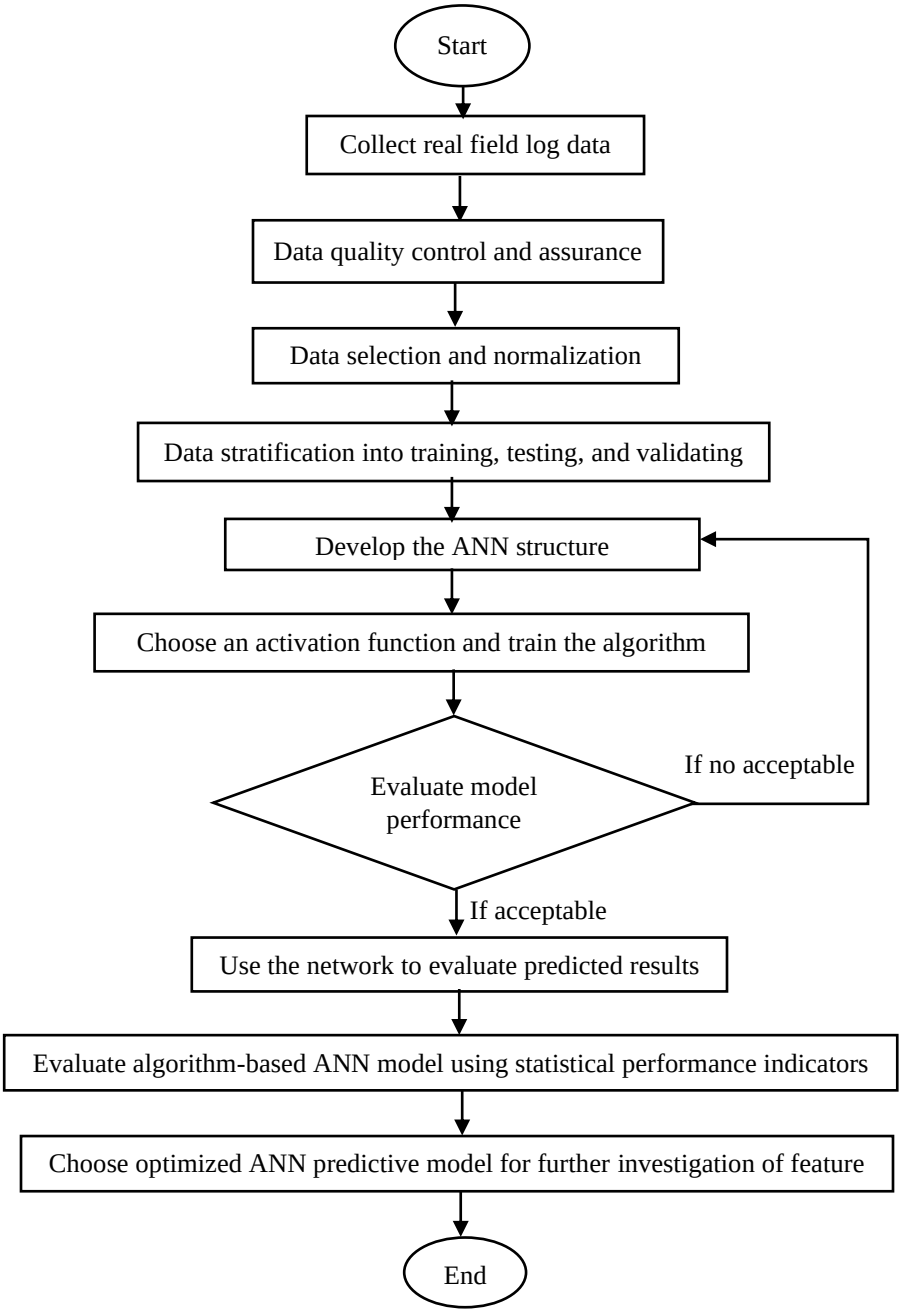

Figure 4.2: A flowchart for ANN model development and estimation of the output variable.

Three learning algorithms of backpropagation, namely, Levenberg Marquardt (LM), Bayesian Regularization (BP) and Scaled Conjugated Gradient (SCG) are implemented to

train the MLP ANN network in this study. The tan-sigmoidal activation function is adopted for the hidden layers as the logistic sigmoidal is well suited to demands of backpropagation, while pure line transfer function is chosen for the output layer. To construct the machine learning algorithm or kernel function-based models, the learning data subsets (70% of total samples) are utilized to adjust the weights of the trained neural network. Furthermore, the testing data subsets (15%) are employed for testing the final solution to assess the performance of the deterministic neural network. The remaining 15% validation data subsets are used to minimize the overfitting of the model. The numbers of hidden layers and neurons are chosen by the trial and error, considering a minimal error.

Providing a brief information on the methodology, the training portion of the data points are used to train the model. The fitted model is then employed to forecast the responses for the observations in a second phase, known as the validation phase. This stage is planned to provide an unbiased assessment of the model suggested from the training phase while tuning the model parameters. Validation datasets can be utilized for regularization by early stopping, meaning the training is stopped when an increase in the error extent is noticed during the validation phase, due to overfitting. Thus, the first two stages (training and validation) offer an optimal model. The constructed model is then tested using the testing datapoints. The selected ANN network is run for training, testing, and validation purposes with the collected data points to ensure model consistency.

4.2.4 Background and mathematical formulation of LS-SVM model

The support vector machine (SVM), a recognized powerful machine learning tool, was first introduced by Vapnik (1995). It is used to locate an optimum hyperplane from which all real experimental data have a minimum distance or maximum margin (Cristianini and Shawe-Taylor, 2000; Suykens and Vandewalle, 1999; Cortes and Vapnik, 1995). More information regarding SVM or SVR algorithm with different features can be found in the literature (Haifeng and Dejin, 2005; Smola et al., 2004). The LS-SVM, proposed by Suykens and Vandewalle (1999), is a modified version of the classic SVM algorithm. This modified version is less complex than the classic SVM algorithm. It can help to reach the solution of a worsening problem more efficiently by setting up a linear set of equations employing SVM instead of the quadratic programming (Suykens and Vandewalle, 1999; Suykens et al., 2002; Pelckmans et al., 2002). Comparing with SVM, the LS-SVM learning method is less time consuming and it can utilize the real field data to construct an appropriate model.

Generally, the LS-SVM optimization of the model problem is formulated through engagement of the following objective function (Suykens and Vandewalle, 1999; Esfahani et al., 2015; Esmaeili et al., 2019); Given a training set $\{x_k, y_k\}_{k=1}^{n}$ with input data $x_k \in R^n$ and the output variable $y_k \in R^n$, the LS-SVM model for estimation of the objective function has the following mathematical expression in the feature space:

$$y(x) = \omega^T \varphi(x) + b \tag{4.8}$$

In Equation (4.8), the nonlinear function $\varphi(.)R^n \rightarrow R^{n_k}$ represents the primal space to a higher dimensional feature space. The dimension n^k of this space is only defined in an

implicit way while *b* introduces a bias term and $\omega \in R^{n_k}$ is the weight factor. In the primal space, the optimization problems can be written as follows:

$$Minimize\ J(\omega, e_k) \cong \frac{1}{2}\omega^T\omega + \gamma \sum_{k=1}^{n} {e_k}^2 \tag{4.9}$$

Subject to

$$y_k = \omega^T\varphi(x_k) + b + e_k\ , k = 1,2,3, \ldots., n \tag{4.10}$$

in which, e_k represents the error term; J(.) resembles the loss function; and γ is the adjustable constant.

Using Equation (4.8), the Lagrangian function ($\mathcal{L}$) can be written as follows:

$$\mathcal{L}(\omega, b, e_k, \alpha_k) = J(\omega, e_k) - \sum_{k}^{n} \alpha_k \{\omega^T\varphi(x_k) + b + e_k - y_k\} \tag{4.11}$$

where α_k refers to the Lagrange multiplier.

The optimality upper function can be expressed as follows:

$$\frac{\partial \mathcal{L}}{\partial \omega} = 0 \rightarrow \omega = \sum_{k}^{n} \alpha_k \varphi(x_k) \tag{4.12}$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \rightarrow \sum_{k}^{n} \alpha_k = 0 \tag{4.13}$$

$$\frac{\partial \mathcal{L}}{\partial e_k} = 0 \rightarrow \alpha_k = \gamma e_k \tag{4.14}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_k} = 0 \rightarrow \omega^T\varphi(x_k) + b + e_k - y_k = 0 \tag{4.15}$$

where k =1,2,3,……,n. After eliminating ω and e_k, Equations 4.12 to 4.15 can be presented as follows:

$$\begin{bmatrix} 0 & 1^T \\ 1 & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{4.16}$$

In Equation (4.16), $y = [y_1, y_2, y_3, y_4, y_5, \ldots\ldots\ldots, y_n]^T$; $1 = [1,1,1, \ldots.1]^T$; $\alpha = [\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \ldots, \alpha_n]$; Kernel matrix, $\Omega = K(x, x_k) = \varphi(x)^T \varphi(x_k)$; and K(x, x_k) represents the kernel function that should satisfy the Mercer's condition (Pelckmans et al., 2002).

The final expression can be formulated for the LS-SVM function estimation as follows:

$$y = \sum_{i}^{n} \alpha_i K(x, x_k) + b \tag{4.17}$$

where b and α are the solutions to the linear system expressed through Equation (4.17). The weight factor, α is a vector with the size of n×1.

There are many kernel functions for the LS-SVM such as linear (Lin), polynomial (Poly), spinal, radial basis, and sigmoidal kernel function. Among these functions, the Gaussian radial basis function (RBF) has been widely used in the LS-SVM learning strategy to attain the best output (Suykens et al., 2002). The methodology for kernel-based LS-SVM model development is summarized in Figure 4.3. The most common types of the kernel functions employed in regression learning approach are listed in Table 4.3 (Zendehboudi et al., 2018). In Table 4.3, *t* and *d* are the intercept and the degree of polynomial kernel function, respectively.

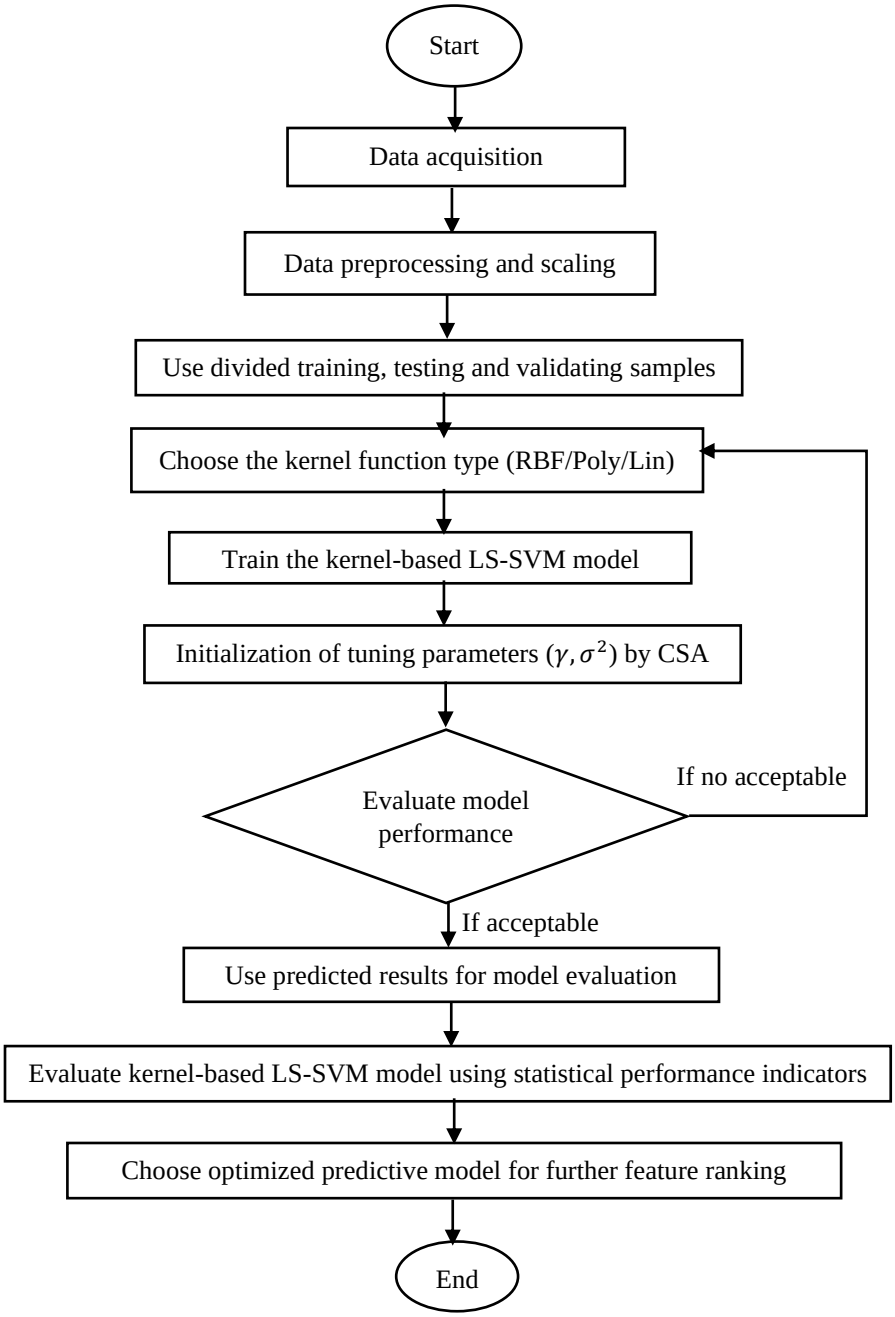

Figure 4.3: A simple flowchart for kernel-based LS-SVM model development.

Table 4.3: Kernel function types and associated mathematical expressions.

Type of kernel function	Mathematical expression	Syntax for simulation
Linear	$K(x, x_k) = x^T x_k$	lin_kernel
Polynomial	$K(x, x_k) = (t + x^T x_k)^d$ with t ≥ 0	poly_kernel
Radial based (Gaussian)	$K(x, x_k) = \exp\left(-\frac{\left(\lvert\lvert x - x_k \rvert\rvert\right)^2}{2\sigma^2}\right)$	RBF_kernel

In the kernel function- based LS-SVM, the regularization and kernel parameters (also named as tuning parameters) are γ and σ^2. These parameters are adjusted through a global optimization technique of coupled simulated annealing (CSA) during the training process. A small value of γ indicates high regularization, which leads to a less nonlinear model. The tuning parameter σ^2 affects the number of neighbors. A larger value of σ^2 means more neighbors in the RBF model and thus a more nonlinear model. In order to improve the performance of two tuning parameters, CSA is used to guarantee the model accuracy and convergence through an iterative random way. The CSA optimization process is has been proven to be more effective than the multi-start gradient descent technique (Suykens et al., 2002; Rostami et al., 2019). Similar to the MLP-ANN model, the database for log data is divided randomly into three sub-datasets to construct the LS-SVM models using different kernel functions. All samples are categorized into three groups such as 70% for training, 15% for testing, and 15% for validation.

4.2.5 Feature ranking of the predictive model

The following systematic strategy, demonstrated in Figure 4.4, is employed to find the relative importance/ performance of the input log variables on the machine learning (ML) based predictive water saturation model.

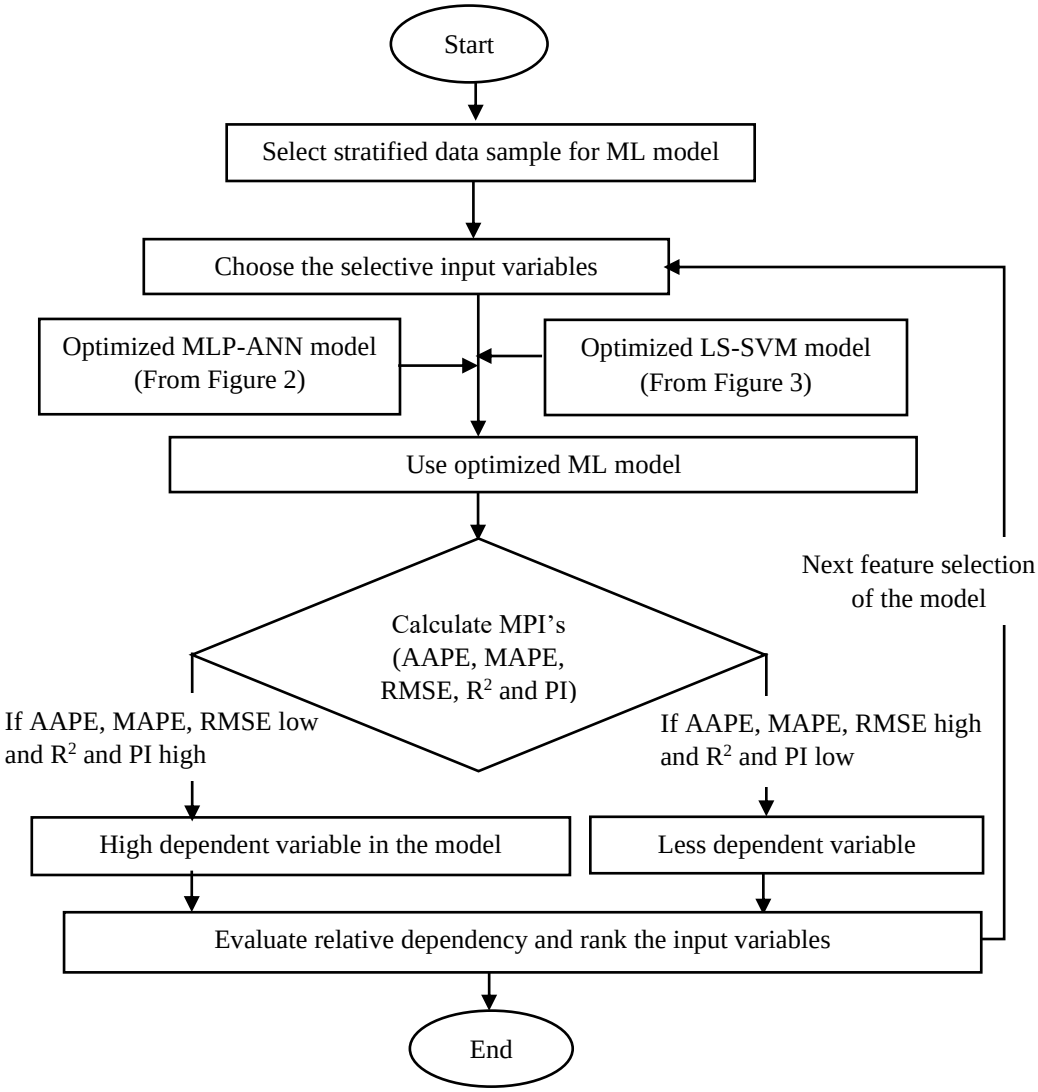

Figure 4.4: A flowchart to conduct feature ranking using the ML approach in the study.

Furthermore, the optimized ML model is employed to predict the output variable; the model performance is examined using statistical performance indicators for each case. Based on the contribution of an input variable to the predictive models, the input parameter is ranked using two approaches. The first approach is a preliminary screening scenario for feature ranking by adopting a single-input and single-output strategy. The input variable having a high impact on the predictive model has low AAPE, MAPE and RMSE, and high R^2 and PI values. The second approach is called 'single variable elimination strategy' (selective multiple input variables to determine the output variable using the optimized ML structure). In the second method, a total number of input variables, N_t (as feature selection), is chosen based on the relative model performance and contribution using the first approach. The low contributing input variable is omitted to avoid the overfitting and run model simulation smoothly. After that, the (N_t -1) number of input variables is selected to find their contribution in the model consequently. In the second strategy, if the model performance exhibits high extent of AAPE, MAPE, and RMSE, and low values of R^2 and PI, it means that the eliminated input variable has more impact on that model. Finally, the input log variables are ranked based on their contribution to the predictive model.

4.3 Results and Discussions

We study the application of ANN and LS-SVM approaches for obtaining the water saturation in the reservoir. According to log data analysis and nature of log variables, there are no significant outlier data points in the studied reservoir depth. To eliminate the scale

effect, the input variables are normalized, as shown in Figure 4.5. All the log variables are employed to find the best ANN and kernel-based LS-SVM models.

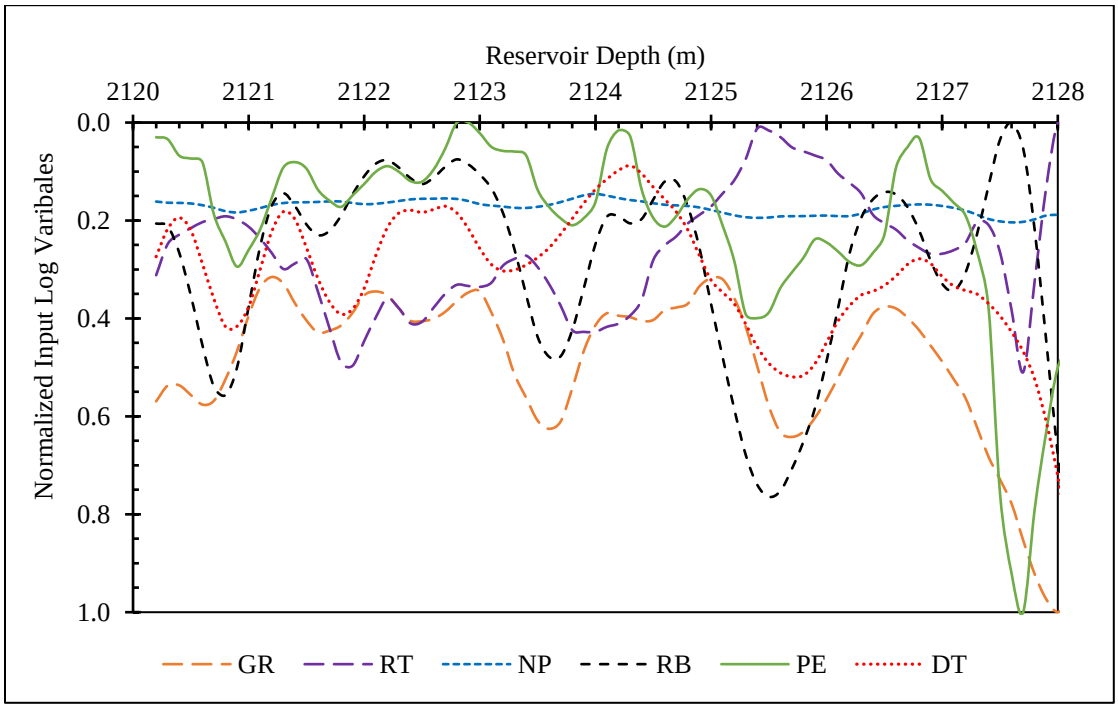

Figure 4.5: Normalized input log variables in the study.

4.3.1 MLP-ANN model performance

To obtain a reliable topology of MLP network using stratified sub-set data, the optimal structure of MLP-ANN (6-5-3-1) is trained with three algorithms, namely, Levenberg Marquardt (LM), Bayesian regularization (BR), and scaled conjugated gradient (SCG). Based on the optimization phase, the model has one input layer with 6 neurons, the first hidden layer with 5 neurons, the second hidden layer with 3 neurons, and one output layer with one neuron. The performance of the constructed models, for each training-based network, is illustrated in Figures 4.6 through 4.11.

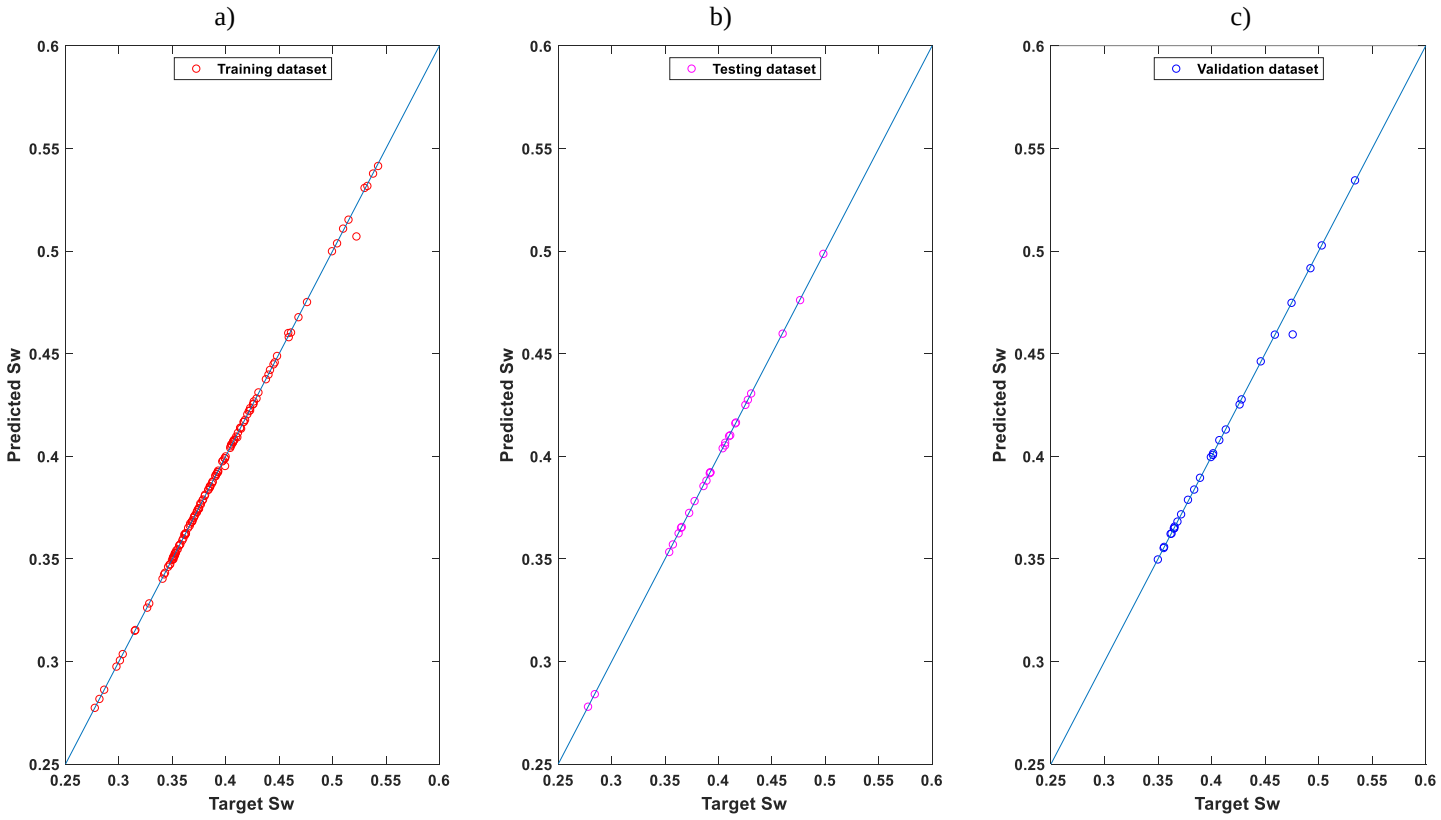

Figure 4.6: Scatter plot of MLP-ANN model for a) training, b) testing, and c) validation datasets with Levenberg Marquardt training algorithm.

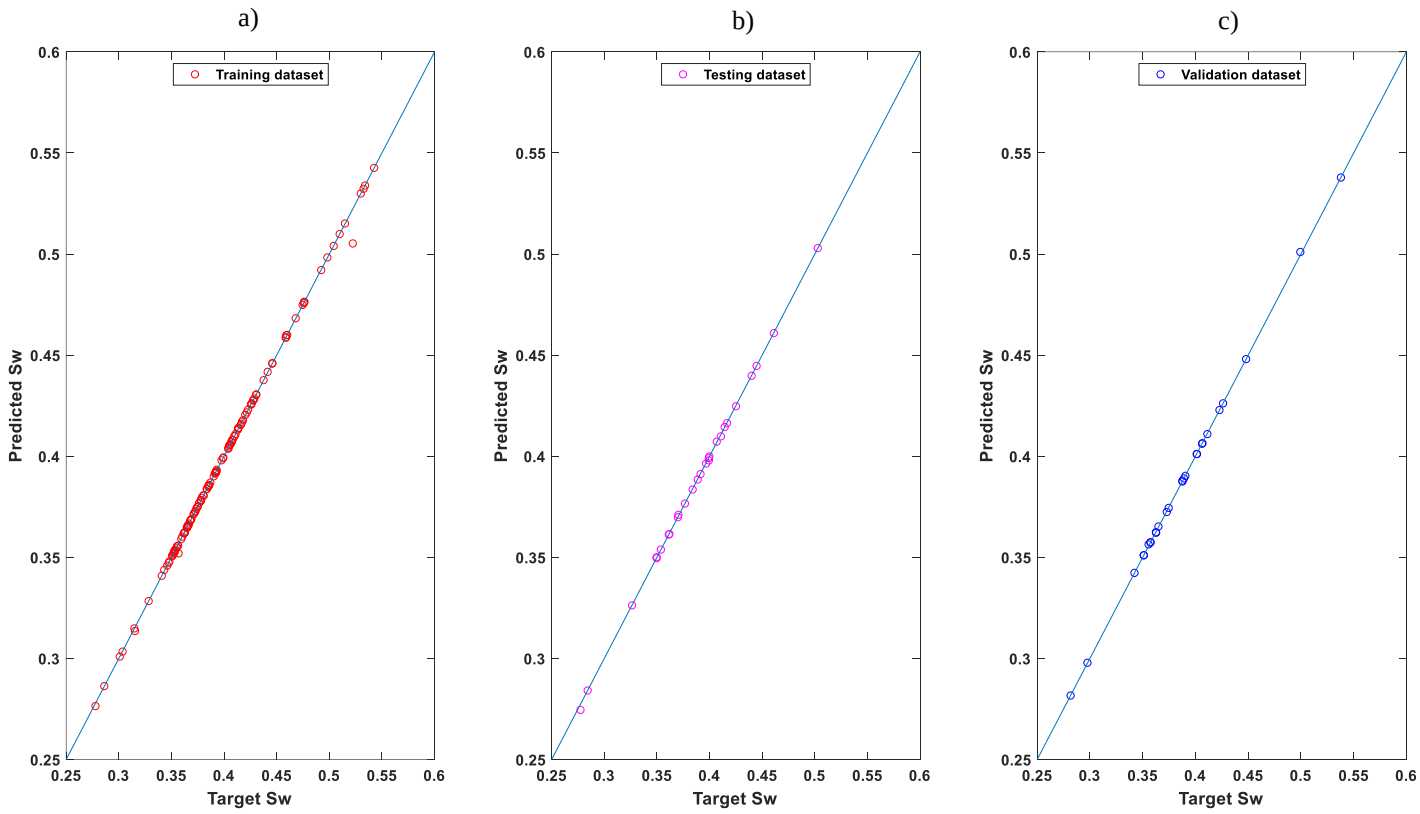

Figure 4.7: Scatter plot of Bayesian regularization training algorithm-based MLP-ANN model for a) training, b) testing, and c) validation datasets.

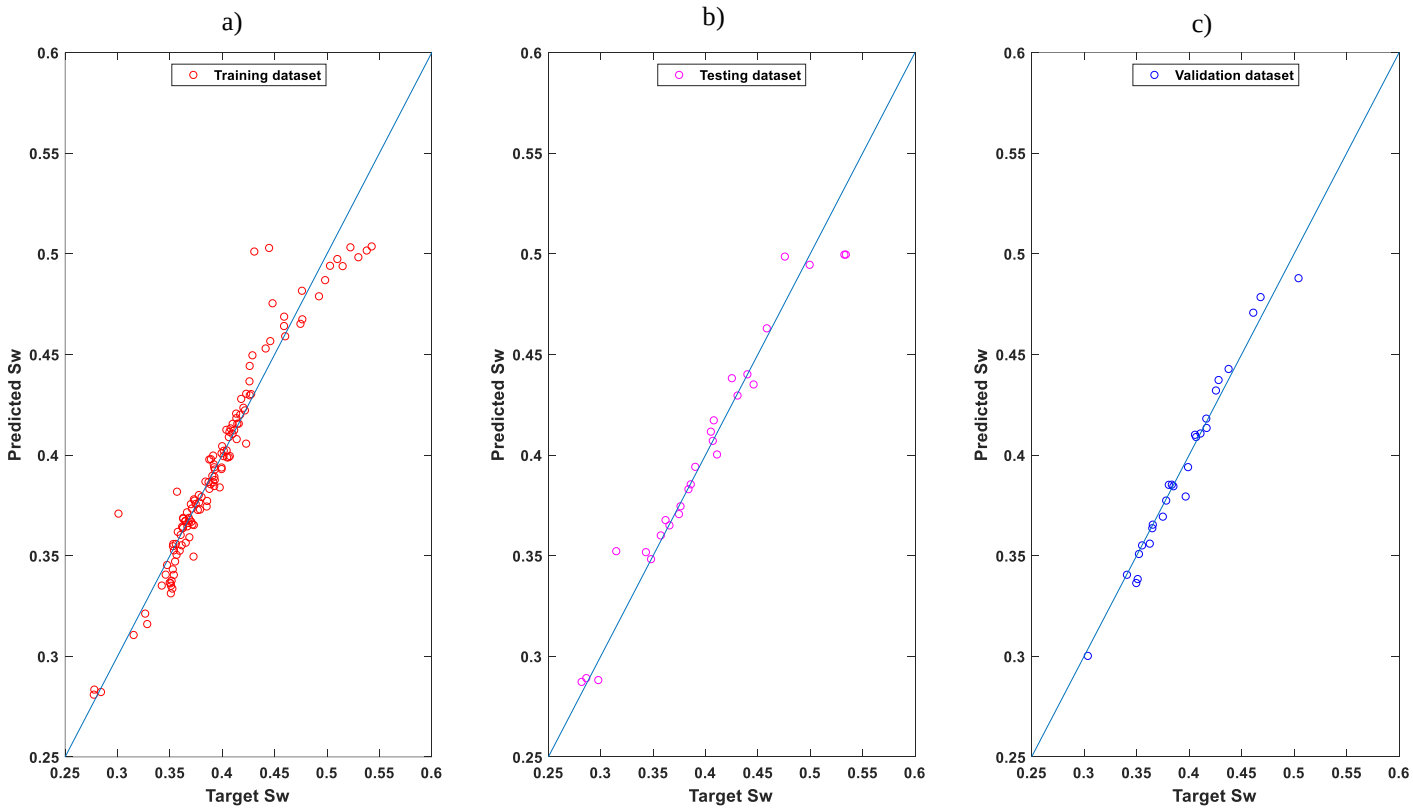

Figure 4.8: Scatter plot of scaled conjugated gradient training algorithm-based MLP-ANN model for a) training, b) testing, and c) validation datasets.

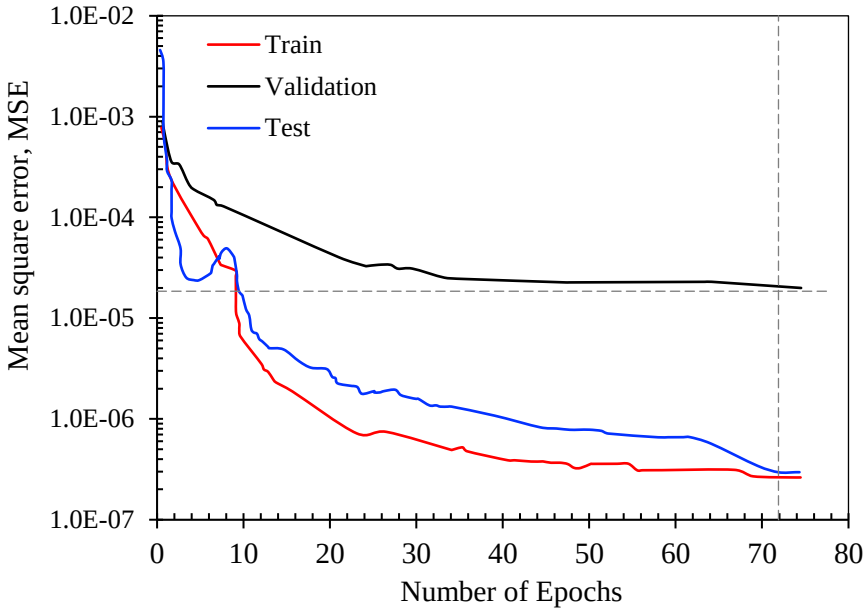

Figure 4.9: Validation performance plot of ANN model with Levenberg Marquardt training algorithm.

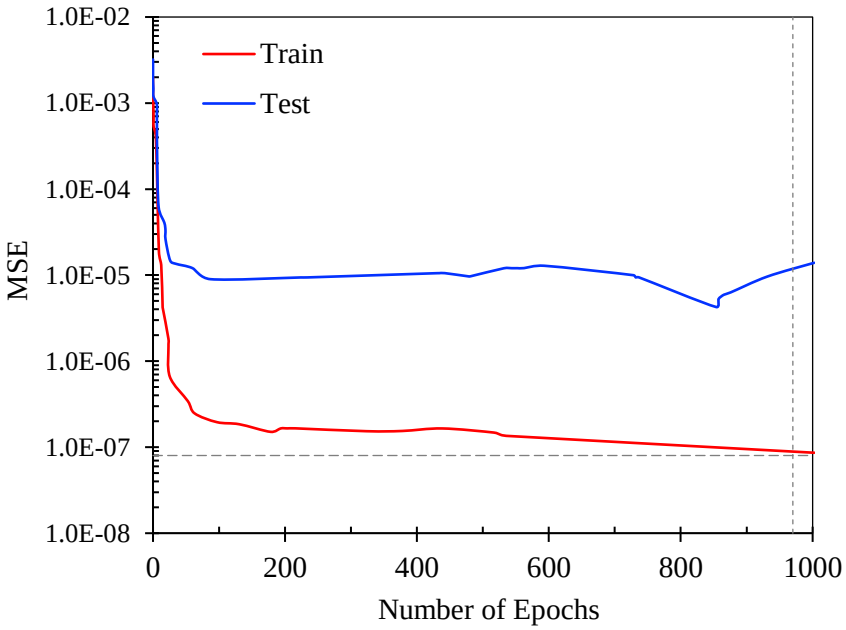

Figure 4.10: Validation performance plot of ANN model with Bayesian regularization training algorithm.

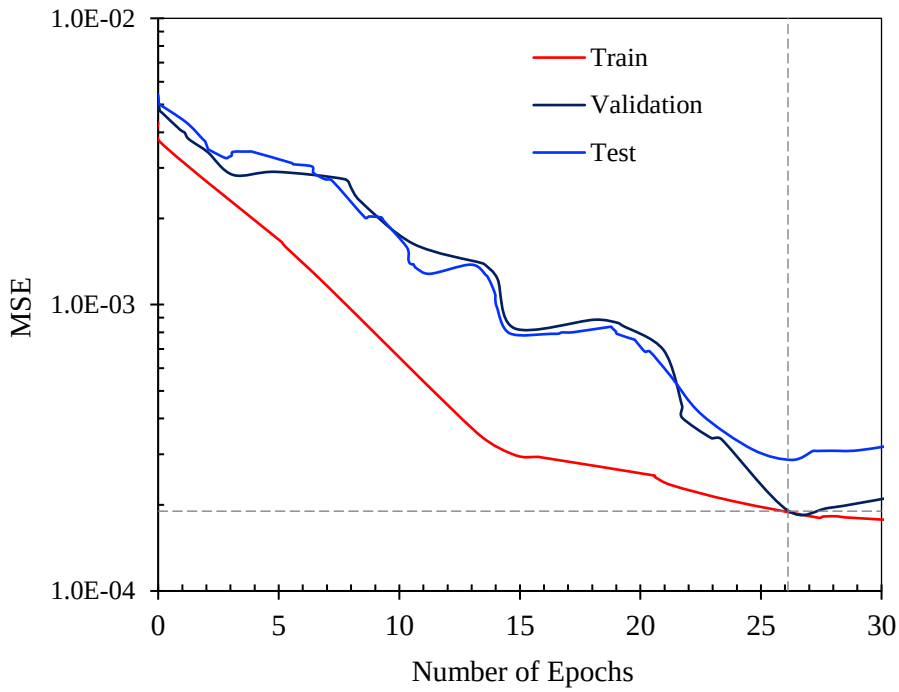

Figure 4.11: Validation performance plot of ANN model with scaled conjugated gradient training algorithm.

According to Figure 4.6, the LM based training algorithm offers the best match between the target and predicted values of water saturation, compared to the others algorithms for MLP-ANN network. Based on Figure 4.10, the BR-based ANN network needs more time to attain the best training performance with more iterations, compared to other algorithms. The model performance of different algorithms for prediction of water saturation using MLP-ANN is listed in Table 4.4 and Figure 4.12 for both training (Trn) and testing (Tst) datasets.

Table 4.4: Evaluation of algorithm based MLP ANN model based on statistical criteria.

Trained algorithm	No. of epoch (iterations)	RMSE Trn (Tst)	AAPE Trn (Tst)	MAPE Trn (Tst)	R^2 Trn (Tst)	PI Trn (Tst)
LM-ANN	75	0.0015 (3.62e-04)	0.111 (0.071)	2.912 (0.207)	1.000 (0.999)	1.008 (1.010)
BR-ANN	975	0.0016 (6.79e-07)	0.092 (0.127)	3.270 (1.012)	0.999 (0.999)	1.008 (1.009)
SCG-ANN	116	0.0144 (0.0137)	2.183 (2.135)	23.257 (11.893)	0.917 (0.957)	0.952 (0.974)

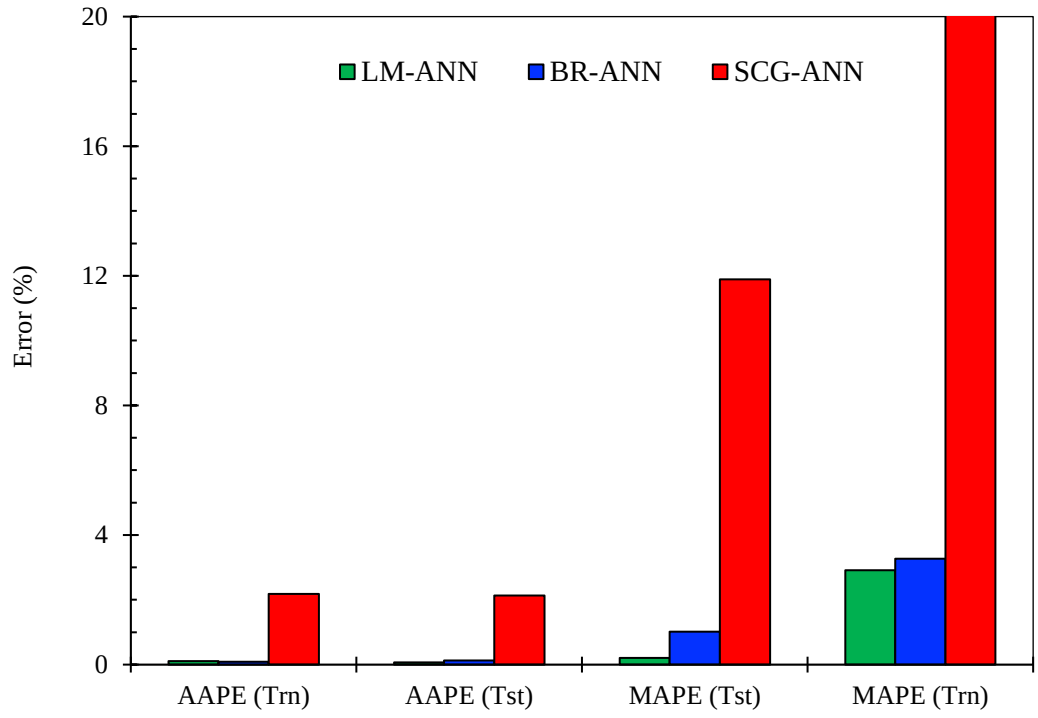

Figure 4.12: Comparison of error measures for different algorithms.

As demonstrated in Table 4.4, the LM and BR-based algorithms exhibit a better performance than SCG-ANN on the basis of RMSE, AAPE, R^2, and PI. Although the BR-ANN model and LM-ANN model have almost the same values for statistical parameters, the former has a higher computational time. Comparing various algorithms, the LM-ANN results in a low MAPE (%) of 2.912 and 0.207 for the training and testing schemes, respectively. It can be concluded that the LM-ANN performance is better than other algorithm-based models according to Table 4.4 and Figure 4.12. Thus, the optimized LM-ANN model with a topology 6-5-3-1 (total number of layers: 4; input layer with five predictor variables; the first and second hidden layers containing five and three neurons, respectively with Tansig transfer function; and output layer with a single variable) is used for further analysis and to find the influencing input log variables while predicting water saturation.

4.3.2 Kernel function-based LS-SVM model performance

In this paper, the CSA optimization technique is used in the LS-SVM model as an iterative random search strategy such that the optimization procedure is repeated several times to achieve the optimum global point. Figure 4.13 displays the scatter plot that compares the target and predicted (output) results. According to Figure 4.13, the RBF-based model leads to a better match compared to the other two kernel function-based models for training, testing, and validation phases. The initial values of two tuning parameters of γ and σ^2 are 16134.9131 and 21.7973, respectively.

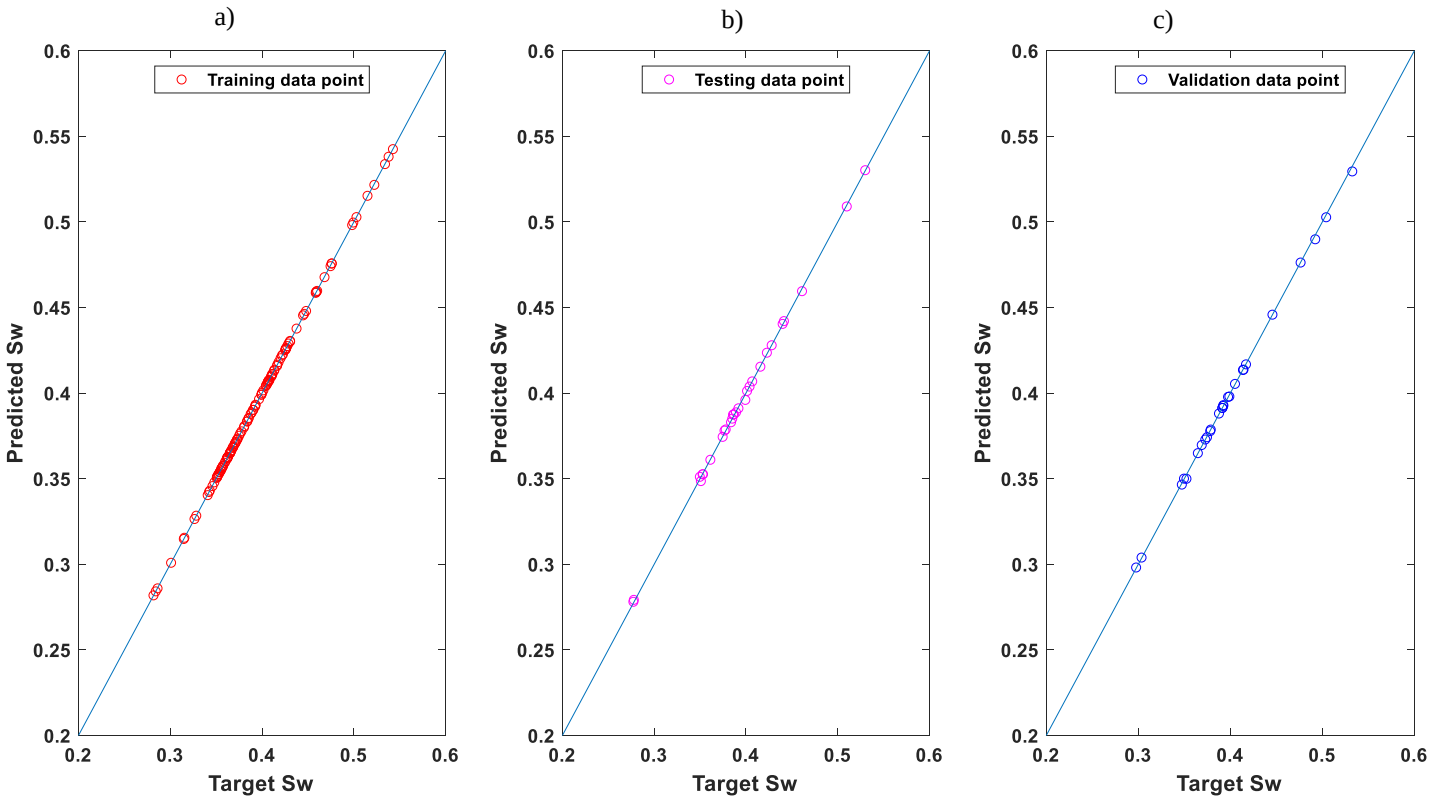

Figure 4.13: Scatter plot of the target and predicted results of RBF based LS-SVM for a) training, b) testing, and c) validation phases.

The final tuning parameters of γ and σ^2 for the RBF (Gaussian) based LSSVM model for water saturation prediction are 1.24e+02 and 4.69e+07 through employing 15 iterations. The optimization results and performance of the three different kernel-based LS-SVM predictive models are tabulated in Table 4.5. The polynomial-based LS-SVM predictive model gives the magnitudes of the hyper-parameters equal to 2.62 (t) and 3 (degree of polynomial) with 13 iterations.

Table 4.5: Statistical parameters corresponded to kernel function-based LS-SVM while obtaining tuning and hyper-parameters.

Kernel function	**No. of iterations**	**Obtained parameters**		**RMSE** Trn (Tst)	**AAPE** Trn (Tst)	**MAPE** Trn (Tst)	**R^2** Trn (Tst)	**PI** Trn (Tst)
		γ	b					
RBF	15	4.69e+07	-6.88	0.0002 (0.0011)	0.039 (0.203)	0.169 (0.819)	1.000 (0.999)	1.010 (1.009)
Polynomial	13	0.124	-0.049	0.0013 (0.0022)	0.237 (0.344)	1.072 (1.767)	0.999 (0.997)	1.008 (1.006)
Linear	7	0.351	1.78e-15	0.0133 (0.012)	1.891 (2.416)	18.067 (7.652)	0.929 (0.952)	0.956 (0.975)

Although, the RBF model results in the minimum values for MAPE (0.169 for the training and 0.819 for the testing), the polynomial function-based model and RBF LS-SVM model show almost the same performance based on the magnitudes of RMSE, AAPE, R^2, and PI (see Table 4.5). Comparing the three-kernel function based predictive models, the RBF

based LS-SVM predictive model has the highest accuracy (or best performance) as it exhibits the lowest RMSE, AAPE, and MAPE, but the highest R^2 and PI.

4.3.3 Feature ranking of the predictive model

Impact of a single predictor variable: In this section, the topography of LM-ANN (1-5-3-1: input layer with 1 neuron; the first hidden layer with 5 neurons; the second hidden layer with 3 neurons; and one output layer with one neuron) is employed to find the relative importance of the input variables in the optimized model while determining water saturation. The single log variable is used as the input variable for predicting the water saturation by employing the same optimized hidden layers to investigate the contribution of that variable in the predictive model. According to the results of conducted simulations, Table 4.6 summarizes the statistical analysis and/or performance assessment of different model cases. Figures 4.14 through 4.16 demonstrate the relative performance and visual contrast amongst the models for training and testing data. According to the different single variable-based data connectionist models, the true resistivity (RT) is more significant than the other input log variables. In the current study, the LM-based ANN model with RT input variable results in the best fit with the highest correlation coefficient (average 0.85) as well as high performance index (0.91). In addition, the RT input log-based LM-ANN model gives the least RMSE (0.020-0.019), AAPE (3.602-3.765), and MAPE (20.252-19.635), compared to other five predictive models.

Table 4.6: Performance evaluation of different cases for the optimized LM-ANN model based on statistical analysis.

Case No. of model (1-5-3-1)	Input variable	No. of iterations (validation epochs)	RMSE Trn (Tst)	AAPE Trn (Tst)	MAPE Trn (Tst)	R^2 Trn (Tst)	PI Trn (Tst)
1	RT	15	0.0202	3.602	20.252	0.861	0.916
		(9)	(0.0194)	(3.765)	(19.635)	(0.845)	(0.911)
2	RB	15	0.0263	5.686	22.40	0.7581	0.8512
		(9)	(0.0257)	(5.553)	(12.764)	(0.6056)	(0.7569)
3	NP	10	0.039	7.579	32.93	0.476	0.651
		(4)	(0.0361)	(7.085)	(28.83)	(0.272)	(0.481)
4	PE	15	0.0483	9.038	29.005	0.1689	0.3295
		(9)	(0.0349)	(8.038)	(28.297)	(0.4369)	(0.6219)
5	GR	10	0.0453	8.917	51.525	0.2231	0.4059
		(4)	(0.0406)	(8.720)	(30.842)	(0.0795)	(0.2192)
6	DT	11	0.0464	8.781	39.00	0.1977	0.3591
		(5)	(0.0514)	(10.974)	(40.76)	(0.2752)	(0.4341)

Based on the predictive performance of the various models, the rank of the influencing input log variables from the highest to lowest is as follows: RT, RB, NP, PE, GR, and DT. To make sure about the generalization of this finding, the RBF-based LS-SVM is implemented to find the most significant parameters for predicting water saturation. Table 4.7 reports the results of different RBF LS-SVM models to assess the importance of single input parameters.

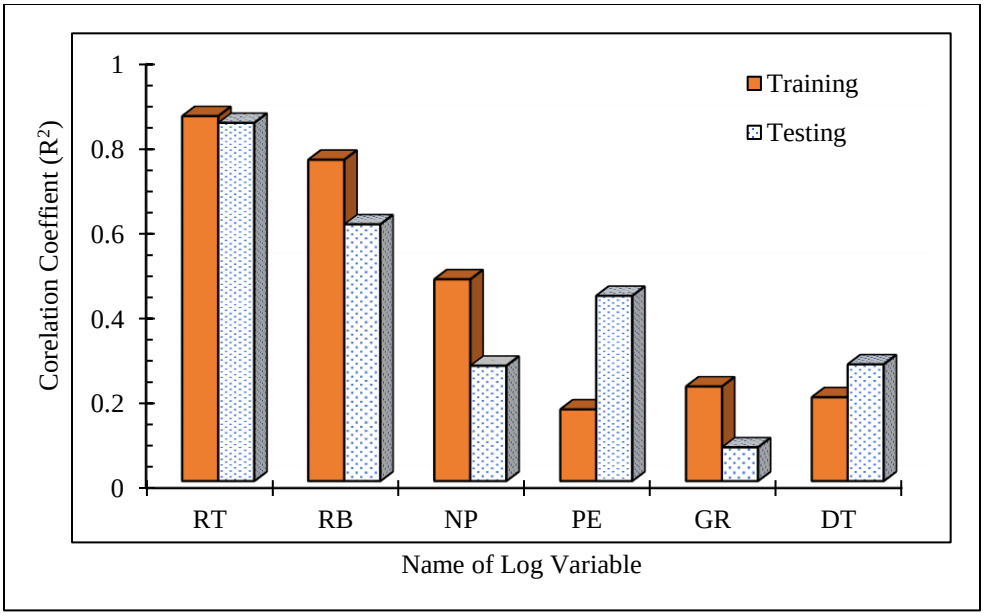

Figure 4.14: Correlation coefficient of single input variables while predicting Sw.

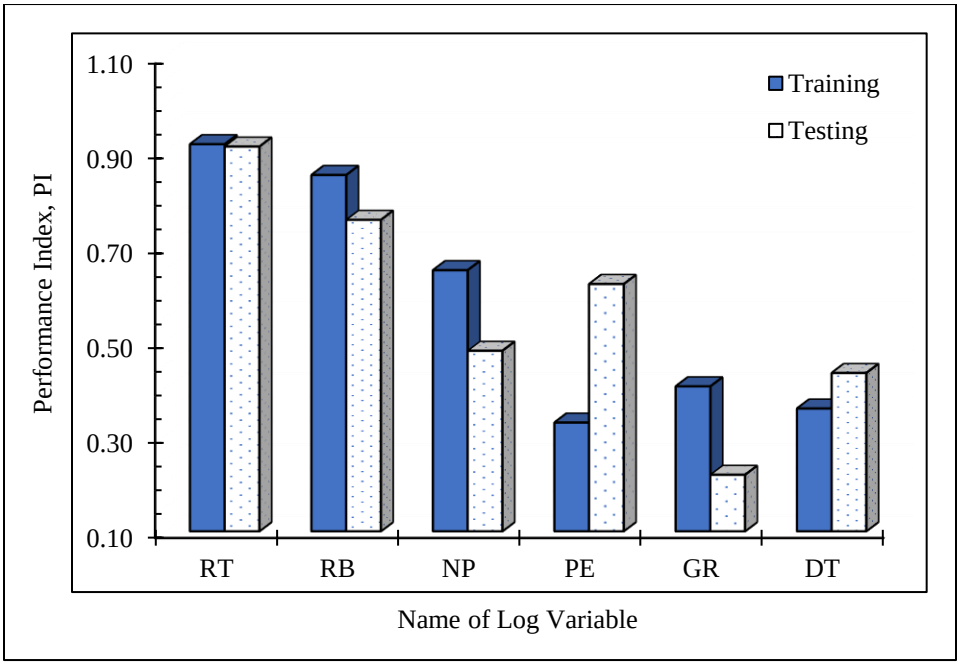

Figure 4.15: Comparison of performance index of input variables when obtaining water saturation.

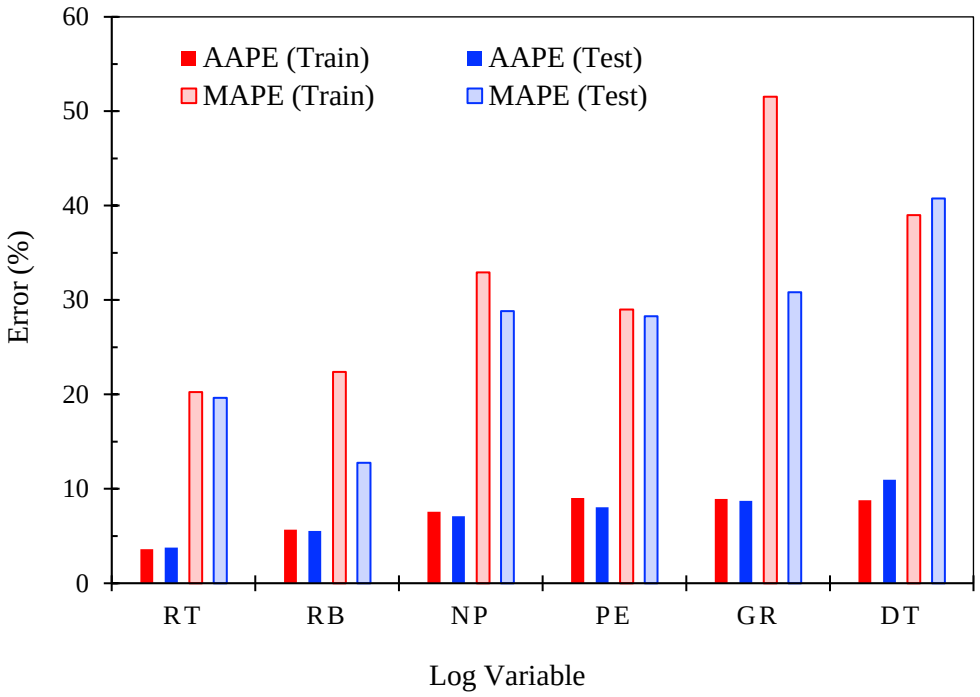

Figure 4.16: Statistical error indicators to show importance of single input variables for prediction of Sw.

Table 4.7: Performance of RBF model with respect to single input variables based on statistical analysis.

Input variable	Tuning parameters: γ, σ^2 and bias (no. of iteration)	RMSE Trn (Tst)	AAPE Trn (Tst)	MAPE Trn (Tst)	R^2 Trn (Tst)	PI Trn (Tst)
RT	352, 6.925; 2.792 (9)	0.0181 (0.0168)	3.350 (3.329)	19.736 (19.296)	0.880 (0.877)	0.929 (0.928)
RB	53.123, 3.188; -1.262 (12)	0.026 (0.020)	5.671 (4.170)	21.453 (13.649)	0.703 (0.884)	0.818 (0.928)
NP	4.912, 0.0283;	0.031	6.137	28.999	0.670	0.792

	0.143 (12)	(0.032)	(5.987)	(18.267)	(0.517)	(0.692)
PE	5.412, 1.948;	0.046	8.875	33.530	0.256	0.438
	-0.294 (14)	(0.048)	(10.601)	(34.100)	(0.090)	(0.225)
GR	102871.765, 6.335;	0.044	8.322	49.935	0.231	0.412
	69.452 (13)	(0.051)	(9.147)	(21.694)	(0.060)	(0.130)
DT	1.642, 19.764;	0.047	9.146	40.734	0.088	0.121
	0.642 (12)	(0.056)	(11.110)	(38.079)	(0.170)	(0.151)

Based on Figures 4.17 to 4.19, the model performance is evaluated using different statistical indicators to demonstrate the importance of various input variables in predicting water saturation.

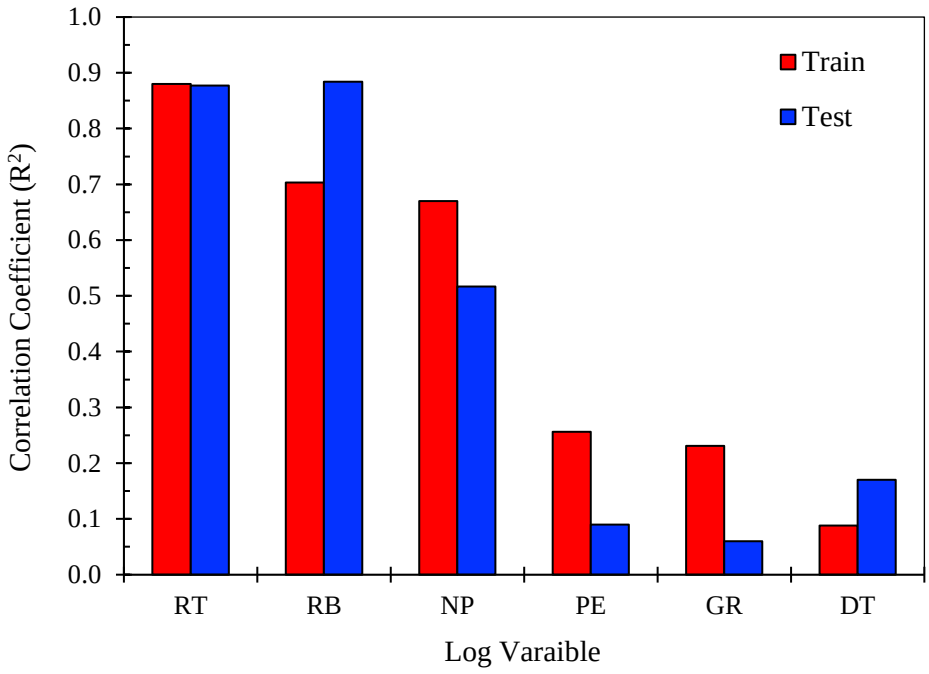

Figure 4.17: Comparison of R^2 for different RBF LS-SVM models based on once single input variable.

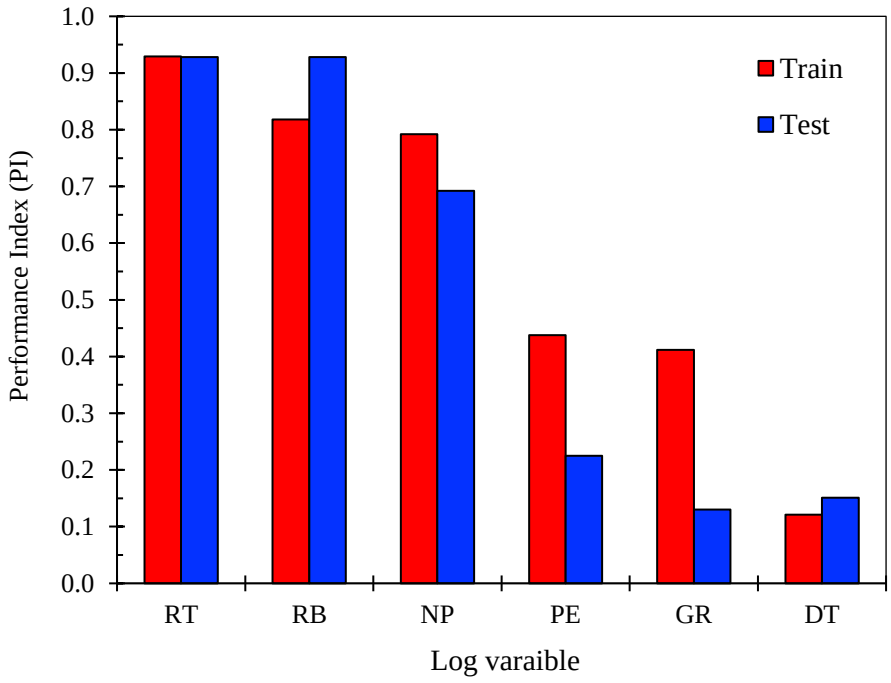

Figure 4.18: Comparison of PI for different RBF LS-SVM models with respect to one single input variable.

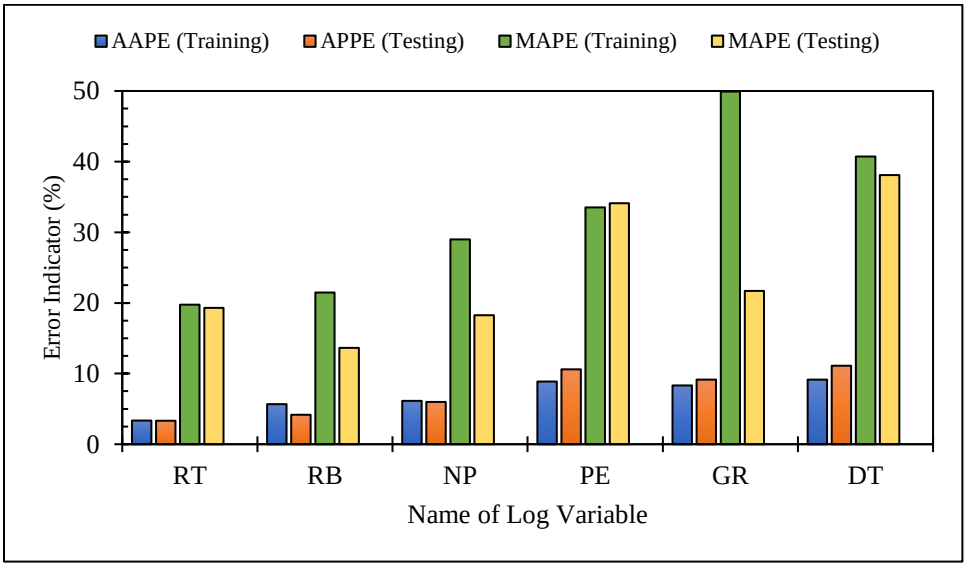

Figure 4.19: Comparison of error indicators for different RBF LS-SVM models based on single input variables.

According to the performance criteria, the DT log variable-based model is found to have the lowest performance with an R^2 of 0.09 and 0.17; PI of 0.12 and 0.15; and the highest errors, AAPE (%) of 9.15 and 11.11; and MAPE (%) of 40. 73 and 38.08 for the training and testing phases, respectively. Referring to the statistical parameter of RMSE, the performance of the optimized LM-ANN and RBF LS-SVM with respect to input log variables is depicted in Figure 4.20. The DT variable has an RMSE of about 0.05 and a total relative error of about 25% for both LM-ANN and RBF LS-SVM approaches, implying it has the minimum significance among the input variables in water saturation prediction. Based on the results, the RT log leads to a relative error of about 9% for LM-ANN and 8% for RBF LS-SVM models, while the relative error for RB case is 12% and 9% in the training and testing phases, respectively, for estimation of water saturation.

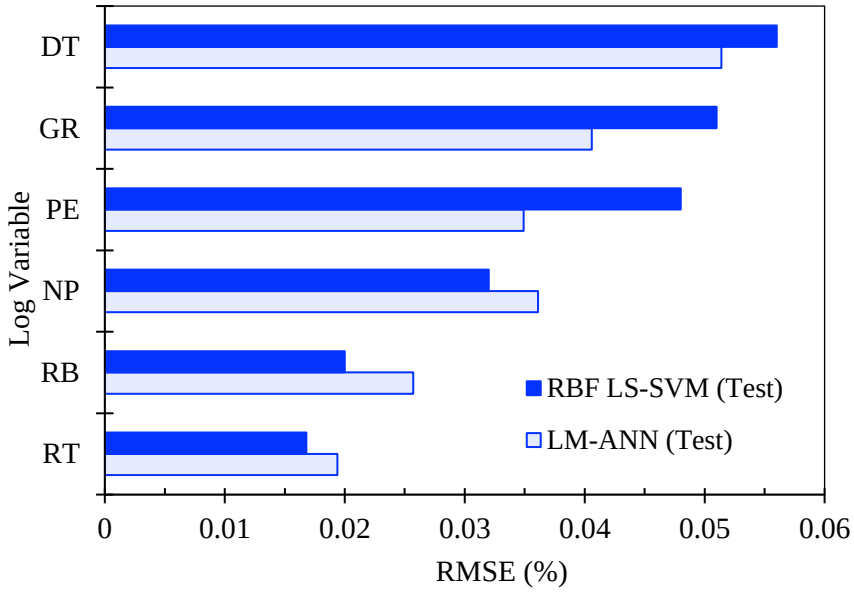

Figure 4.20: Performance comparison of different predictive models for predicting Sw using one single input variable based on RMSE.

It follows that RT and RB exhibit the highest importance among the variables. The study also confirms that NP and PE logs have almost the same relative contribution in both LM-ANN and RBF-based LS-SVM techniques. The GR log variable has also a low impact in the model for estimation of water saturation. Comparing the models performance using the statistical indicators, it is evident that the influential log parameters (higher to lower) are RT, RB, NP, PE, GR, and DT in the predictive model.

Effect of single variable elimination on the optimized predictive model: For further analyzing the effect of individual variables, the DT log is first excluded. From the remaining variables, another variable is then excluded each time and the remaining variables are used to develop the prediction model. Table 4.8 reports the model performance in the absence of each single variable. Figures 4.21 and 4.22 demonstrate the correlation coefficient and statistical errors to evaluate the effectiveness of the models used in the current study.

Table 4.8: Performance of four input variables based-LM ANN models in the absence of one single variable.

Model scheme	**Input variables**	**Excluded variable**	**RMSE Trn (Tst)**	**AAPE Trn (Tst)**	**MAPE Trn (Tst)**	**R^2 Trn (Tst)**	**PI Trn (Tst)**
1	GR, RB, NP, PE	RT	0.0176 (0.0217)	3.492 (3.954)	23.056 (26.225)	0.8810 (0.7926)	0.9296 (0.8766)

2	GR, RT, NP, PE	RB	0.0125	2.271	12.549	0.9347	0.9637
			(0.0148)	(2.701)	(8.758)	(0.9097)	(0.9481)
3	GR, RT, RB, PE	NP	0.0074	1.398	6.960	0.9772	0.9905
			(0.0073)	(1.394)	(5.647)	(0.9858)	(0.9954)
4	GR, RT, RB, NP	PE	0.0013	0.1967	3.023	0.9994	1.0084
			(0.0033)	(0.2373)	(3.199)	(0.9965)	(1.0050)
5	RT, RB, NP, PE	GR	0.0013	0.1567	2.259	0.9994	1.0085
			(0.0007)	(0.1443)	(0.3439)	(0.9998)	(1.0092)

Table 4.9: Performance comparison of four input variable-based RBF LS-SVM model while excluding one input parameter.

Model scheme	**Input variables**	**Excluded variable**	**RMSE Trn (Tst)**	**AAPE Trn (Tst)**	**MAPE Trn (Tst)**	**R^2 Trn (Tst)**	**PI Trn (Tst)**
1	GR, RB, NP, PE	RT	0.0153	2.9006	23.7608	0.9113	0.9483
			(0.0443)	(4.7283)	(13.0177)	(0.7472)	(0.8466)
2	GR, RT, NP, PE	RB	0.0094	1.7935	7.2478	0.9651	0.9826
			(0.0230)	(3.7944)	(28.4404)	(0.8465)	(0.9050)
3	GR, RT, RB, PE	NP	0.0043	0.8072	3.633	0.9926	1.0019
			(0.0049)	(0.9912)	(2.4825)	(0.9936)	(1.0001)
4	GR, RT, RB, NP	PE	0.0002	0.0397	2.2162	1.000	1.0098
			(0.0025)	(0.1928)	(2.5041)	(0.9953)	(1.0051)
5	RT, RB, NP, PE	GR	0.0006	0.1065	0.4121	0.9999	1.0094
			(0.0007)	(0.1637)	(0.3621)	(0.9998)	(1.0091)

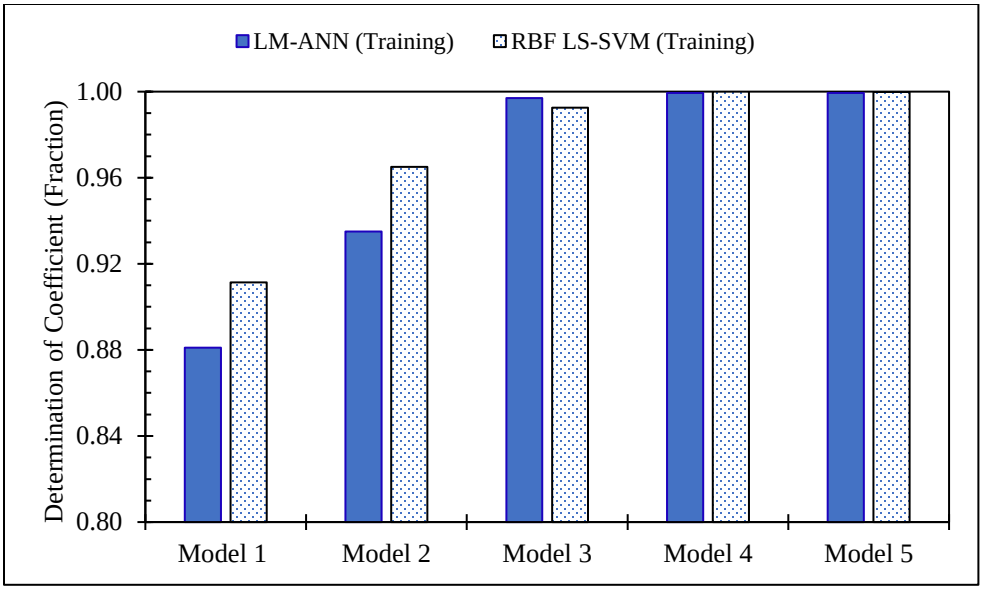

Figure 4.21: Comparison of R^2 for different model schemes in the absence of once single variable.

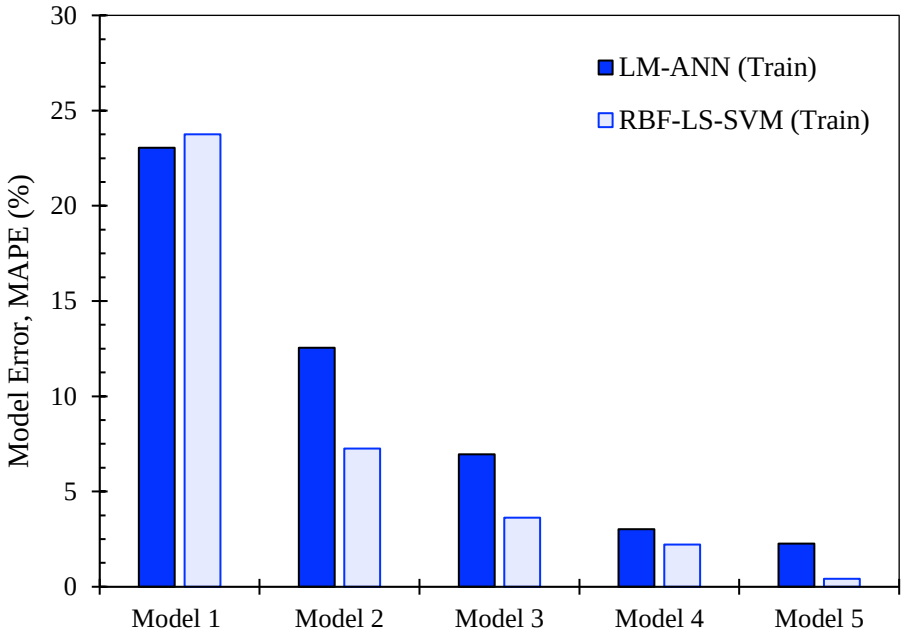

Figure 4.22: Comparison of MAPE for various model schemes in the absence of once single variable.

To evaluate the reliability and precision of the five predictive models, various statistical parameters including correlation coefficient and relative errors are used. The predictive model 1 (see Tables 4.8 and 4.9) in the absence of RT log variable exhibits a high error percentage and low correlation coefficient because the RT has a considerable influence on the water saturation model. Similarly, model 5 without considering GR log variable leads to a high performance (e.g., a low statistical error and high correlation coefficient), since the GR has a lower impact than other variables in the data-driven model schemes for water saturation prediction. According to the statistical analysis conducted for the LM-ANN approach, the decreasing order of important input variables for predicting water saturation is RT, RB, NP, PE, and GR. The RBF based LS-SVM technique also leads to almost the same finding. In the literature, it is claimed that the resistivity (RT) and porosity logs (RB or NP) are essential to determine water saturation using petrophysical models. According to the current study, the log variables are ranked in terms of significance as follows: RT, RB, NP, PE, GR, and DT for estimation of water saturation. The engineers and/or operators in the oil and gas industry can utilize the developed deterministic approaches using the data from the most contributing logs for prediction of water saturation to save the exploration expenses and time through an effective manner.

4.4 Conclusions

In this study, using real field well log data such as RT, RB, NP, PE, GR and DT, the MLP-ANN and the LS-SVM techniques are implemented to develop data-driven predictive models for obtaining water saturation. The statistical indicators, namely, RMSE, AAPE,

MAPE, R^2, and PI are utilized to examine the model performance while using MLP based ANN algorithms and different kernel functions-based LS-SVM models. Based on the research outcomes, the concluding remarks are listed below:

- The water saturation extent depends on the log variables including RT, RB, PE, GR, and DT. The RBF kernel function-based LS-SVM model and ANN approach use the above variables for objective function prediction. Both strategies exhibit an acceptable accuracy.

- The MLP-ANN model including the Levenberg Marquardt training algorithm shows a superior performance with a lower error percentage, MAPE (2.91%), compared to the other algorithms such as the Bayesian regularization (with an MAPE of 3.27%) and the scaled conjugated gradient (with an MAPE of 23.26%).

- It reveals that the RBF based LS-SVM model has a greater reliability and deterministic capability, compared to the polynomial and linear kernel function-based models for predicting water saturation so that the lowest statistical errors and high magnitudes of R^2 and PI for this model are attained.

- Based on the predictive performance of the MLP-ANN and the LS-SVM models, bulk density and true resistivity are found to be the most important predictor variables, while gamma-ray and sonic travel time have the least contribution to the water saturation.

- Good agreement is noticed between the predictions and field data, implying the effectiveness and precision of the proposed predictive tools.

It is recommended that RT and RB with high influence degrees are included as dataset attributes in the data-driven predictive models for estimating water saturation. The feature ranking strategy with machine learning technique might be proper to predict other reservoir properties (e.g., relative permeability) using log variables and to find the minimum number of logging parameters for saving the geophysical exploration costs, because a lower number of logging tools is employed. It appears that the research strategies and feature ranking approaches in the data-driven models of reservoir rock mechanical properties can be useful to field specialists, practitioners, geologists, and engineers in the oil and gas industry.

The LS-SVM is a more desirable tool for investigation of reservoir performance as well as enhanced oil recovery (EOR) analysis where deterministic models with high accuracy are required. Finally, such a research approach can be implemented using more data points for various purposes such as predictive model development, optimization problems, and feature ranking that can have various applications in engineering, science, and health disciplines.

Acknowledgements

The authors would like to thank Equinor (formerly Statoil) Canada Ltd., Natural Sciences and Engineering Research Council of Canada (NSERC) and InnovateNL for providing financial support to accomplish this study at Memorial University, St. John's, NL, Canada.

Nomenclature

Acronyms	
AAPE	Absolute Average Relative Percentage Error

ANFIS	Adaptive Neuro-fuzzy Inference System
ANN	Artificial Neural Network
BP	Bayesian Regularization
CSA	Coupled Simulated Annealing
DT	Sonic Transit (Compressional) Time (μs/ft)
FDT	Fuzzy Decision Tree
FN	Functional Network
FL	Fuzzy Logic
GA	Genetic Algorithm
GR	Gamma Ray (API)
ICA	Imperialist Competitive Algorithm
ILD	Deep Induction Log (ohm-m)
NP	Neutron Porosity (percentage)
LM	Levenberg Marquardt
LS-SVM	Least Square Support Vector Machine
MSE	Mean Square Error
MAPE	Maximum Absolute Percentage Error (percentage)
ML	Machine Learning
MLP	Multi-Layered Perceptron
PE	Photo Electric Index (barns/electron))
PSO	Particle Swarm Optimization
PI	Performance Index
RNN	Recurrent Neural Networks
RBF	Radial Basis Kernel Function
RB	Formation Bulk Density (gram/cm^3)
RT	True (Deep) Resistivity (ohm-m)
R^2	Regression Coefficient (fraction)
SCG	Scaled Conjugated Gradient
SP	Self-Potential Log (millivolt)
SVR	Support Vector Regression
SVM	Support Vector Machine
Sw	Water Saturation (fraction)
VAF	Variance Account Factor

Variables and Parameters

b	Bias term
f_h	Transfer function for hidden layer,
f_o	Transfer function for output layer
N_t	Total number of input variables

N1-1	1st neuron in the 1st hidden layer
N2-1	1st neuron in the 2nd hidden layer
n	Number of data available for the variable
$_{nor}$	Normalized value of variable
max	Maximum magnitude of variable
min	Minimum magnitude of variable
y	Output variable
y_p	Predicted value
y_t	Target (actual) variable
x	Input log variable
x_k	Support vector

Greek Letters and subscript

α	Weight factor for LS-SVM
α_k	Lagrange multiplier
γ	Tuning (Annealing) parameter
ω_{ij}	Connectionist weight of ANN
σ	Kernel width of LSSVM
Ω	Kernel matrix
t	Target
p	Predicted
o	Output

References

Adedigba, S. A., Khan, F., & Yang, M. (2017). Dynamic failure analysis of process systems using neural networks. Process Safety and Environmental Protection, 111, 529-543.

Adeniran, A., Elshafei, M., & Hamada, G. (2009, May). Functional Network soft sensor for formation porosity and water saturation in oil wells. In 2009 IEEE Instrumentation and Measurement Technology Conference (pp. 1138-1143). IEEE.

Ahmadi, M. A., Soleimani, R., Lee, M., Kashiwao, T., & Bahadori, A. (2015). Determination of oil well production performance using artificial neural network (ANN) linked to the particle swarm optimization (PSO) tool. Petroleum, 1(2), 118-132.

Ahmadi, Mohammad-Ali, Ahmadi, M. R., Hosseini, S. M., & Ebadi, M. (2014). Connectionist Model Predicts the Porosity and Permeability of Petroleum Reservoirs by Means of Petro-physical Logs: Application of Artificial Intelligence. Journal of Petroleum Science and Engineering 123, 183-200.

Ahrimankosh, M., Kasiri, N. & Mousavi, SM. (2011). Improved Permeability Prediction of a Heterogeneous Carbonate Reservoir Using Artificial Neural Networks Based on the Flow Zone Index Approach, Petroleum Science and Technology, 29:23, 2494-2506, DOI: 10.1080/10916461003735103.

Akande, K. O., Owolabi, T.O., & Olatunji, S.O. (2015a). Investigating the Effect of Correlation-based Feature Selection on the Performance of Neural Network in Reservoir Characterization. Journal of Natural Gas Science and Engineering 27, 98-108.

Akande, K. O., Owolabi, T.O., & Olatunji, S.O., (2015b). Investigating the Effect of Correlation-Based Feature Selection on the Performance of Support Vector

Machines in Reservoir Characterization. Journal of Natural Gas Science and Engineering 22:515–22.

Akande, K. O., Owolabi, T.O., Olatunji, S.O., & Abdulraheem, A.-A. (2016). A Novel Homogenous Hybridization Scheme for Performance Improvement of Support Vector Machines Regression in Reservoir Characterization. Applied Computational Intelligence and Soft Computing 2016:1–10.

Al-Bulushi, N. I., King, P.R., Blunt, M.J., & Kraaijveld, M. (2012). Artificial Neural Networks Workflow and Its Application in the Petroleum Industry. Neural Computing and Applications 21(3):409–21.

Al-Bulushi, N., Araujo, M., Kraaijveld, M., & Jing, X. D. (2007). Predicting Water Saturation Using Artificial Neural Networks (ANNs). Society of Petrophysicists and Well-Log Analysts.

Al-Bulushi, N., King, P.R., Blunt, M.J., & Kraaijveld, M. (2009). Development of artificial neural network models for predicting water saturation and fluid distribution. Journal of Petroleum Science and Engineering 68, 197–208.

Ali J.K. (1994). Neural network: a new tool for petroleum industry. Proceedings of SPE European petroleum computer conference, UK, paper 27561.

Amer, M. M., Dahab, A. S., & El-Sayed, A. A. H. (2017, June). An ROP predictive model in Nile Delta area using artificial neural networks. In SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition. Society of Petroleum Engineers.

Aminian, K. & Ameri, S. (2005). Application of Artificial Neural Networks for Reservoir Characterization with Limited Data. Journal of Petroleum Science and Engineering 49(3–4):212–22.

Amiri, M., Ghiasi-Freez, J., Golkar, B., & Hatampour, A. (2015c). Improving water saturation estimation in a tight shaly sandstone reservoir using artificial neural network optimized by imperialist competitive algorithm–A case study. Journal of Petroleum Science and Engineering, 127, 347-358.

Amiri, M., Zahedi, G., & Yunan, M. H. (2015a). Reducing predictive uncertainty in log-derived water saturation models in a giant tight shaly sandstones–A case study from Mesaverde tight gas reservoir. Journal of Natural Gas Science and Engineering, 23, 380-386.

Amiri, M., Zahedi, G., & Yunan, M. H. (2015b). Water saturation estimation in tight shaly gas sandstones by application of Progressive Quasi-Static (PQS) algorithm–A case study. Journal of Natural Gas Science and Engineering, 22, 468-477.

Anifowose, F. A., Labadin, J., & Abdulraheem, A. (2015). Ensemble model of non-linear feature selection-based extreme learning machine for improved natural gas reservoir characterization. Journal of Natural Gas Science and Engineering, 26, 1561-1572.

Anifowose, F., Adeniye, S., Abdulraheem, A., & Al-Shuhail, A. (2016). Integrating seismic and log data for improved petroleum reservoir properties estimation using non-linear feature-selection based hybrid computational intelligence models. Journal of Petroleum Science and Engineering, 145, 230-237.

Anifowose, F., Ayadiuno, C., & Rashedan, F., (2019). Comparative Analysis of Machine Learning Based Feature Selection Approach for Carbonate Reservoir Cementation Factor Prediction. International Petroleum Technology Conference. doi:10.2523/IPTC-19330-MS

Anifowose, F., Jane Labadin, and Abdulraheem, A.-A. (2013). Prediction of Petroleum Reservoir Properties Using Different Versions of Adaptive Neuro-Fuzzy Inference System Hybrid Models. International Journal of Computer Information Systems and Industrial Management Applications 5:413–26.

Anifowose, F., Labadin, J., & Abdulraheem, A., (2014). Non-linear feature selection based hybrid computational intelligence models for improved natural gas reservoir characterization. Journal of Natural Gas Science and Engineering 21:397–410.

Archie, G.E., (1941). The electrical resistivity log as an aid in determining some reservoir characteristics. Transactions of AIME 146, 54–62.

Ashena, R., & Thonhauser, G. (2015). Application of Artificial Neural Networks in Geoscience and Petroleum Industry. In Artificial Intelligent Approaches in Petroleum Geosciences (pp. 127-166). Springer, Cham.

Ashrafi, S. B., Anemangely, M., Sabah, M., & Ameri, M. J. (2019). Application of hybrid artificial neural networks for predicting rate of penetration (ROP): A case study from Marun oil field. Journal of Petroleum Science and Engineering, 175, 604-623.

Asquith, G., & Krygowski, D. (2004). Basic well log analysis, American Association of Petroleum Geologists, Second Edition, Tulsa, Oklahoma, p. 216.

Bageri, B. S., Anifowose, F. A., & Abdulraheem, A. (2015). Artificial intelligence based estimation of water saturation using electrical measurements data in a carbonate reservoir. In SPE Middle East Oil & Gas Show and Conference. Society of Petroleum Engineers.

Baneshi, M., Behzadijo, M., Schaffie, M., & Nezamabadi-Pour, H. (2013). Predicting log data by using artificial neural networks to approximate petrophysical parameters of formation. Petroleum Science and Technology, 31(12), 1238-1248.

Basbug, B., & Karpyn, Z. (2007). Estimation of permeability from porosity, specific surface area and irreducible water saturation using an artificial neural network. SPE 107909, Latin American and Caribbean Petroleum Engineering Conference, Buenos Aires, Argentina, April 15–18.

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. Journal of microbiological methods, 43(1), 3-31.

Bassiouni, Z. (1994). Theory, Measurement and Interpretation of Well Logs, First Printing, Henry L. Doherty Memorial Fund of AIME, SPE, Richardson, TX, p. 372.

Baziar, S., Shahripour, H. B., Tadayoni, M., & Nabi-Bidhendi, M. (2018). Prediction of water saturation in a tight gas sandstone reservoir by using four intelligent methods: a comparative study. Neural Computing and Applications, 30(4), 1171-1185.

Ceryan, N., & Can, N. K. (2018). Prediction of The Uniaxial Compressive Strength of Rocks Materials. In Handbook of Research on Trends and Digital Advances in Engineering Geology (pp. 31-96). IGI Global.

Clavier, C., Coates, G., & Dumanoir, J. (1984). Theoretical and experimental bases for the dual-water model for interpretation of shaly sands. Society of Petroleum Engineers Journal, 24(02), 153-168.

Cortes, C., & Vapnik, V. (1995). Support vector networks. Mach. Learn. 20(3), 273–297.

Cristianini, N. & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.

DeWitte, L. (1950). Relations between resistivities and fluid contents of porous rocks. Oil Gas J. 120–134. August 24.

Dongxiao, Z., Chen, Y., & Meng, J. (2018). Synthetic well logs generation via Recurrent Neural Networks. Petroleum Exploration and Development, 45(4), 598-607.

Ebrahimi, A., & Khamehchi, E. (2016). Developing a novel workflow for natural gas lift optimization using advanced support vector machine. Journal of Natural Gas Science and Engineering, 28, 626-638.

Erofeev, A., Orlov, D., Ryzhov, A., & Koroteev, D. (2019). Prediction of porosity and permeability alteration based on machine learning algorithms. Transport in Porous Media, 128(2), 677-700.

Esene, C., Zendehboudi, S., Shiri, H., & Aborig, A. (2019). Deterministic tools to predict recovery performance of carbonated water injection. Journal of Molecular Liquids, 111911.

Esfahani, S., Baselizadeh, S., & Hemmati-Sarapardeh, A. (2015). On determination of natural gas density: least square support vector machine modeling approach. Journal of Natural Gas Science and Engineering, 22, 348-358.

Esmaeili, S., Sarma, H., Harding, T, & Maini, B. (2019). A data-driven model for predicting the effect of temperature on oil-water relative permeability, Fuel, 236, 264-277.

Esmaeilzadeh, S., Salehi, A., Hetz, G., Olalotiti-lawal, F., Darabi, H., & Castineira, D. (2020). Multiscale modeling of compartmentalized reservoirs using a hybrid clustering-based non-local approach. Journal of Petroleum Science and Engineering, 184, 106485.

Esmaeilzadeh, S., Salehi, A., Hetz, G., Olalotiti-lawal, F., Darabi, H., & Castineira, D. (2019). A General Spatio-Temporal Clustering-Based Non-local Formulation for Multiscale Modeling of Compartmentalized Reservoirs. arXiv preprint arXiv:1904.13236. https://doi.org/10.2118/195329-MS

Faucett, L. (1994). Fundamentals of Neural Networks. Architectures, Algorithms, and Applications. Prentice Hall, Englewood Cliffs, NJ.

Fung, C., Wong, K., & Eren, H. (1997). Modular artificial neural network for prediction of petrophysical properties from well log data. IEEE Transactions on Instrumentation and Measurement 46 (6), 1295–1299.

Gardner, M.W., & Dorling, S.R. (1998). Artificial Neural Networks (the Multilayer Perceptron)—a Review of Applications in the Atmospheric Sciences. Atmospheric Environment 32.14: 2627-636.

Ghaffarian, N., Eslamloueyan, R., & Vaferi, B. (2014). Model identification for gas condensate reservoirs by using ANN method based on well test data. Journal of Petroleum Science and Engineering, 123, 20-29.

Gholanlo, H. H., Amirpour, M., & Ahmadi, S. (2016). Estimation of water saturation by using radial based function artificial neural network in carbonate reservoir: A case study in Sarvak formation. Petroleum, 2(2), 166-170.

Haifeng, W., Dejin, H. (2005). Comparison of SVM and LS-SVM for regression, neural networks and brain, 2005. In: ICNN&B '05. International Conference on, pp. 279–283.

Hamada, M. G., ElKadi., A., Nyein, C.Y., & Elsakka., A. (2019). Artificial neural network (ANN) can better determine petrophysical properties than conventional techniques in shaly sand reservoirs, EGYPS 2019 Technical Conference, taking place on 11-13 February 2019 in Cairo, Egypt.

Hamada, M. G., Elsakka., A., & Nyein, C.Y. (2018). Artificial Neural Network (ANN) Prediction of Porosity and Water Saturation of Shaly Sandstone Reservoirs, Advances in Applied Science Research, 9(2):26-31.

Hamedi, H., Ehteshami, M., Mirbagheri, S. A., & Zendehboudi, S. (2019). New deterministic tools to systematically investigate fouling occurrence in membrane bioreactors. Chemical Engineering Research and Design, *144*, 334-353.

Helle, H. B., Bhatt, A., & Ursin, B. (2001). Porosity and permeability prediction from wireline logs using artificial neural networks: a North Sea case study. Geophysical Prospecting, 49(4), 431-444.

Helle, H.B., & Bhatt, A. (2002). Fluid saturation from well logs using committee neural networks. Petroleum Geosciences 8 (2), 109–118.

Helmy, T., Anifowose F, & Faisal K. (2010). Hybrid computational models for the characterization of oil and gas reservoirs. Elsevier Int J Expert Syst Appl 37:5353–5363.

Huang, Z., Shimeld, J., Williamson, M., & Katsube, J., (1996). Permeability prediction with artificial neural network modeling in the Venture gas field, offshore eastern Canada. GEOPHYSICS 61, 422–436.

Hutahaean, J. J., Demyanow, V., & Christie, M. A. (2015). Impact of model parameterisation and objective choices on assisted history matching and reservoir forecasting. In SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition. Society of Petroleum Engineers.

Hutahaean, J., Demyanov, V., & Christie, M. (2019). Reservoir development optimization under uncertainty for infill well placement in brownfield redevelopment. Journal of Petroleum Science and Engineering, 175, 444-464.

Jreou, N. S. G. (2012). Application of neural network to optimize oil field production. Asian Tran. Eng, 2(3), 10-23.

Kamalyar, K., Sheikhi, Y., & Jamialahmadi, M. (2011). Using an Artificial Neural Network for Predicting Water Saturation in an Iranian Oil Reservoir, Petroleum Science and Technology, 30:1, 35-45, DOI: 10.1080/10916461003752561.

Karimpouli, S., Fathianpour, N., & Roohi, J., (2010). A new approach to improve neural networks' algorithm in permeability prediction of petroleum reservoirs using supervised committee machine neural network (SCMNN). Journal of Petroleum Science and Engineering 73, 227e232.

Kenari, S. A. J., & Mashohor, S. (2013). Robust Committee Machine for Water Saturation Prediction. Journal of Petroleum Science and Engineering 104:1–10.

Khan, M. R., Tariq, Z., & Abdulraheem, A., (2018). Machine Learning Derived Correlation to Determine Water Saturation in Complex Lithologies. Society of Petroleum Engineers. doi:10.2118/192307-MS

Lawrence, S., Giles, C.L., & Tsoi, A. (1997). Lessons in neural network training: training may be harder than expected. In: Proceedings of the Fourteenth National Conference on Artificial Intelligence, AAAI-97. AAAI Press, Menlo Park, California, pp. 540–545.

Lippmann, R.P. (1987). An introduction to computing with neural networks, ASSP Magazine, April 4–22.

Mardi, M., Nurozi, H., & Edalatkhah, S. (2012). A water saturation prediction using artificial neural networks and an investigation on cementation factors and saturation exponent variations in an Iranian oil well. Petroleum Science and Technology, 30(4), 425-434.

Mazur, M., and Marry, L. (2015). A step by step backpropagation example, Available at: https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/Accessed on 9th November 2019.

Miah, M. I. (2014). Porosity Assessment of Gas Reservoir Using Wireline Log Data: A Case Study of Bokabil Formation, Bangladesh, Elsevier-Procedia Engineering 90, 663 – 668.

Miah, M. I., & Howlader, M. F. (2012). Prediction of Formation Water Resistivity from Rwa Analysis of Titas Gas Field Using Wireline Log Data. Journal of Petroleum and Gas Exploration Research (JPGER), 2(4), 57-60.

Miah, M. I., Deb, P. K., Rahman, M. S., & Hossain, M. E. (2017). Application of Memory Concept on Petroleum Reservoir Characterization: A Critical Review. Society of Petroleum Engineers. doi:10.2118/187676-MS

Miah, M.I., & Tamim, M. (2015). Hydrocarbon Saturation Assessment of Thick Shaly Sand Reservoir Using Wireline Log Data: A Case Study, The 10th International Forum on Strategic Technology 2015 June 3 - June 5, 2015 Universitas Gadjah Mada, Indonesia.

Miah, M.I., Ahmed, S. & Zendehboudi, S. (2019). Connectionist and mutual information tools to determine water saturation and rank input log variables, Journal of Petroleum Science and Engineering, PETROL 106741.

Mohaghegh, S., Arefi, R., Bilgesu, I., Ameri, S., & Rose, D. (1994). Design and development of an artificial neural network for estimation of formation permeability. Proceeding of the SPE Petroleum Computer Conference, Dallas. Paper SPE, 28237.

Mohaghegh, S.D., Arefi, R., Ameri, S., Aminiand, K., & Nutter, R. (1996). Petroleum reservoir characterization with the aid of artificial neural networks. Journal of Petroleum Science and Engineering 16, 263–274.

Mollajan, A. (2015). Application of Local Linear Neuro-Fuzzy Model in Estimating Reservoir Water Saturation from Well Logs. Arabian Journal of Geosciences 8(7):4863–72.

Mollajan, A., Memarian, H., & Jalali, M. R. (2013). Prediction of reservoir water saturation using support vector regression in an Iranian carbonate reservoir. In 47th US Rock Mechanics/Geomechanics Symposium. American Rock Mechanics Association.

Movahhed, A., Bidhendi, M.N., Masihi, M. & Emamzadeh, A. (2019). Introducing a method for calculating water saturation in a carbonate gas reservoir. Journal of Natural Gas Science and Engineering 70, 102942.

Olatunji, S. O., Selamat, A., & Raheem, A. A. A. (2014). Improved sensitivity based linear learning method for permeability prediction of carbonate reservoir using interval type-2 fuzzy logic system. Applied Soft Computing, 14, 144-155.

Onalo, D., Oloruntobi, O., Adedigba, S., Khan, F., James, L., & Butt, S. (2018). Static Young's modulus model prediction for formation evaluation. Journal of Petroleum Science and Engineering, 171, 394-402.

Pelckmans, K., Suykens, J.A., Van Gestel, T., De Brabanter, J., Lukas, L., Hamers, B., De Moor, B. & Vandewalle, J. (2002). LS-SVMlab: a matlab/c toolbox for least squares support vector machines. Tutorial. KULeuven-ESAT. Belgium, 142.

Poulton, M. (2002). Neural networks as an intelligence amplification tool: a review of applications. Geophysics 67 (3), 979–993.

Poupon, A., & Leveaux, J. (1971). Evaluation of water saturation in shaly formation. Proceeding of SPWLA 12th Annual Logging Symposium.

Poupon, A., Loy, M.E., & Tixier, M.P. (1954). A Contribution to electrical log interpretation in shaly sands. J. Petrol. Technol. 6 (6), 27–34.

Rafik, B., & Kamel, B. (2017). Prediction of permeability and porosity from well log data using the nonparametric regression with multivariate analysis and neural network, Hassi R'Mel Field, Algeria. Egyptian journal of petroleum, 26(3), 763-778.

Razavi, S., & Tolson, B. A. (2011). A new formulation for feedforward neural networks. IEEE Transactions on neural networks, 22(10), 1588-1598.

Rogers, S.J., Chen, H.C., Kopaska-Merkel, D.C., & Fang, J.H. (1995). Predicting permeability from porosity using artificial neural networks. AAPG Bulletin 79, 1786–1797.

Rolon, L., Mohaghegh, S. D., Ameri, S., Gaskari, R., & McDaniel, B. (2009). Using artificial neural networks to generate synthetic well logs. Journal of Natural Gas Science and Engineering, 1(4-5), 118-133.

Rostami, S., Rashidi, F., & Safari, H. (2019). Prediction of oil-water relative permeability in sandstone and carbonate reservoir rocks using the CSA-LSSVM algorithm. Journal of Petroleum Science and Engineering, 173, 170-186.

Saffarzadeh, S., & Shadizadeh, S. R. (2012). Reservoir rock permeability prediction using support vector regression in an Iranian oil field. Journal of Geophysics and Engineering, 9(3), 336-344.

Salehi, M. M., Rahmati, M., Karimnezhad, M., & Omidvar, P. (2017). Estimation of the non records logs from existing logs using artificial neural networks. Egyptian Journal of Petroleum, 26(4), 957-968.

Schlumberger (1972). Log Interpretation; Vol 1-Principles, New York, Texas.

Shan, L., Cao, L., & Guo, B. (2018). Identification of flow units using the joint of WT and LSSVM based on FZI in a heterogeneous carbonate reservoir. Journal of Petroleum Science and Engineering, 161, 219-230.

Shedid, S. A., & Saad, M. A. (2017). Comparison and sensitivity analysis of water saturation models in shaly sandstone reservoirs using well logging data. Journal of Petroleum Science and Engineering, 156, 536-545.

Shokir, E. E. M. (2004, January). Prediction of the hydrocarbon saturation in low resistivity formation via artificial neural network. In SPE Asia Pacific Conference on Integrated Modelling for Asset Management. SPE

Si, W., Di, B., Wei, J., & Li, Q. (2016). Experimental study of water saturation effect on acoustic velocity of sandstones. Journal of Nat. Gas Sci. and Eng., 33, 37-43.

Simandoux, P. (1963). Dielectric measurements on porous media, application to the measurements of water saturation: study of behavior of argillaceous formations. Revue de l'Instiut Francase du Petrol l8, 193–215.

Smaoui, N., & Garrouch, A. A. (1997). A new approach combining Karhunen-Loeve decomposition and artificial neural network for estimating tight gas sand permeability. Journal of Petroleum Science and Engineering, 18(1-2), 101-112.

Smola, A.J., Sch, B., & lkopf (2004). A tutorial on support vector regression. Stat. Comput. 14, 199–222.

Suykens, J. & Vandewalle, J. (1999). Neural Processing Letters 9, 293.

Suykens, J.A.K., Van Gestel, T., Brabanter, J., De Moor, B., & Vandewalle, J. (2002). Least Squares Support Vector Machines. World Scientific, Singapore.

Sylvester, O., Bibobra, I., & Ogbon, O. N. (2015). Well Test and PTA for Reservoir Characterization of Key Properties. American J. of Eng. and App. Sci., 8(4), 638.

Tahmasebi, P., & Hezarkhani, A. (2012). A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation. Computers & geosciences, 42, 18-27.

Tariq, Z., Elkatatny, S., Mahmoud, M., & Abdulraheem, A. (2016). A holistic approach to develop new rigorous empirical correlation for static Young's modulus. In Abu Dhabi International Petroleum Exhibition & Conference. SPE.

Tariq, Z., Elkatatny, S., Mahmoud, M., Ali, A. Z., & Abdulraheem, A. (2017, June). A new approach to predict failure parameters of carbonate rocks using artificial intelligence tools. In SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition. Society of Petroleum Engineers.

Temirchev, P., Simonov, M., Kostoev, R., Burnaev, E., Oseledets, I., Akhmetov, A., ... & Koroteev, D. (2020). Deep neural networks predicting oil movement in a development unit. Journal of Petroleum Science and Engineering, 184, 106513.

Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer, New York.

Wang, B., Wang, X. & Chen, Z. (2013). A Hybrid Framework for Reservoir Characterization Using Fuzzy Ranking and An Artificial Neural Network, Computers & Geosciences 57:1–10.

Wang, P., & Peng, S. (2019). On a new method of estimating shear wave velocity from conventional well logs. Journal of Petroleum Sci. and Eng., 180, 105-123.

Waxman, M.H., & Smits, L.J.M. (1968). Electrical conductivities in oil-bearing shaly sands. Society of Petroleum Engineers Journal 8, 107–122.

White, A., Molnar, D., Aminian, K., & Mohaghegh, S. (1995). The Application of artificial neural networks to zone identification in a complex reservoir. SPE 30977, Proceedings, SPE Eastern Regional Conference and Exhibition September 19–21, 1995, Morgantown, West Virginia.

Wong, K. W., Fung, C. C., Ong, Y. S., & Gedeon, T. D. (2005). Reservoir characterization using support vector machines. In International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IA, WTIC'06) (Vol. 2, pp. 354-359). IEEE.

Yang, S., & Wei, J. (2017). Fundamentals of petrophysics. Berlin, Heidelberg: Springer Berlin Heidelberg. Doi: 10.1007/978-3-662-55029-8

Yao, X. & Liu, Y. (1997). A new evolutionary system for evolving artificial neural networks. IEEE Transactions on Neural Networks 8, 694–713.

Zendehboudi, S., Rezaei, N., & Lohi, A. (2018). Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review. Applied Energy 228, 2539-2566.

Zhang, B., & Xu, J. (2016). Methods for the evaluation of water saturation considering TOC in shale reservoirs. Journal of Nat. Gas Sci. and Eng., 36, 800-810.

Zhao, B., Zhou, H., Li, X., & Han, D. (2006). Water saturation estimation using support vector machine. In SEG Technical Program Expanded Abstracts 2006 (pp. 1693-1697). Society of Exploration Geophysicists.

CHAPTER 5: MACHINE LEARNING APPROACH TO MODEL ROCK STRENGTH: PREDICTION AND VARIABLE SELECTION WITH AID OF LOG DATA

Preface

A version of this chapter has been submitted to the "Rock Mechanics and Rock Engineering" Journal. I am the primary author of this article and prepared the first draft of the manuscript. All authors had contribution to the conceptualization. The co-author Dr. Salim Ahmed provided technical assistance in finding the research scopes and the knowledge gaps. Dr. Ahmed supervised for data analysis and review of the first draft of the manuscript. Co-author, Dr. Sohrab Zendehboudi also reviewed and provided valuable insights on how to improve the model development and results analysis subsections in the manuscript to a journal-level publication. Co-author, Dr. Stephen Butt also provided technical assistance, presentation of the well log data, expert analysis and review. I have revised the manuscript based on the feedback from co-authors.

Abstract

Comprehensive knowledge and analysis of in-situ rock strength and geo-mechanical characteristics of rocks are crucial in hydrocarbon and minerals exploration stage to maximize wellbore performance, maintain wellbore stability, and optimize hydraulic fracturing process. Due to the high cost of laboratory-based measurements of rock mechanics properties, the log-based prediction is a viable option. Nowadays, the machine learning tools are being used for estimation of the in-situ rock properties using wireline log data. This article proposes a machine learning approach for rock strength (uniaxial compressive strength) prediction. The main objectives are to investigate the performance of data-driven predictive model in determining this vital parameter and to select features of predictor log variables in the model. The backpropagation multilayer perception (MLP) artificial neural network (ANN) with Levenberg-Marquardt training algorithm as well as the least squares support vector machine (LSSVM) with coupled simulated annealing (CSA) optimization technique are employed to develop the dynamic data-driven models. Capturing nonlinear, high dimensional, and complex nature of real field log data, the rock strength models' performances are evaluated using statistical criteria to ensure concerning the model reliability and accuracy. The model predictions are compared and validated against the measured values as well as the results obtained from existing log-based correlations. Both the MLP-ANN and the CSA-based LSSVM connectionist strategies are able to predict the rock strength so that there is a very good match between the model results and corresponding measured values. The input log parameters are ranked based on their contributions in prediction performance. The acoustic travel time and gamma-ray are

found to have the highest relative significance in estimating in rock strength. A new correlation is also developed to obtain the in-situ rock strength of the siliciclastic sedimentary rocks using the most important log parameters. It is expected that the proposed models and tools will enable reservoir engineers to better predict rock strength and thus enhance wellbore performance.

Keywords: Log variables; Connectionist models; Uniaxial compressive strength; Variable selection; Rock stability.

5.1 Introduction

5.1.1 Background

Reservoir rock mechanical properties (RMPs) play an important role in making decisions regarding exploration development, wellbore stability analysis, sanding potentiality, and safe drilling operations. The RMPs are closely related to the rock physical properties such as rock density, porosity, and pore fluid saturation. The uniaxial compressive strength (UCS) is one of the key rock strength parameters that can be used in wellbore stability assessment, in-situ stress analysis, drilling optimization as well as reservoir sanding likelihood prediction (Nouri et al. 2006, Crawford et al. 2011). It is also utilized for mining and geotechnical engineering applications including the rock mass characterization, underground rock excavation, and geotechnical project design (ISRM, 1981). The reservoir rock strength provides the limit of axial stress in-situ condition before rock failing (Fjær et al., 2008).

The most common techniques for determining rock strength are in-situ measurements using geophysical log-based models and laboratory measurements using core samples. The direct and indirect measurements of rock strength include UCS test, triaxial test, Brazilian test, and basic mechanical tests such as point load index, non-destructive ultrasonic test, Schmidt hammer rebound test, impact strength test, scratch test, indentation test, and Equotip hardness tester (Miller, 1965; Broch and Frankline, 1972; Hoek and Brown, 1997; Fener et al., 2005; Yilmaz, 2013; Mousavi et al., 2018). The rock mechanical laboratory (RML) testing on real core samples is the most reliable method to determine the rock strength using the procedures recommended by the American Society for Testing Materials (ASTM, 1986) and/or the International Society of Rock Mechanics (ISRM, 1981). The RML testing cannot provide continuous rock strength profile along with the in-situ wellbore condition. Also, the core sample preparation is tedious, time-consuming, expensive, and very sensitive to stress unloading (Raeen et al., 1996). Most of the times, high-quality cylindrical core specimens preparation with regular geometry is difficult due to the nature of fractured thinly bedded and clay-bearing sedimentary rocks. In addition, a large scale core sample preparation is often challenging due to time limitation, complex variations in rock composition, and geological behavior and heterogeneity of the formation. When the formation rock sample is not attainable from the deeper depth of oil and gas and/or mining fields, the well logging data can be used to estimate the rock mechanical and physical properties.

Over the past few decades, numerous studies have been performed to develop empirical models for estimation of rock strength using petrophysical logs such as the sonic and

density porosity logs. The log-based existing models or correlations for rock strength were listed/ summarized by different authors such as Chang et al. (2006) and Odunlami et al. (2011). A continuous in-situ UCS profile is essential to conduct rock failure analysis during a drilling operation.

Nowadays, the machine learning (ML) or artificial intelligence (AI) approaches such as artificial neural network (ANN), functional network (FN), adaptive neuro-fuzzy inference system (ANFIS), decision tree (DT), classic support vector machine (SVM) and/or least square support vector machine (LS-SVM) are extensively used to develop the data-driven models for rock mechanics related to petroleum and mining (Asoodeh and Bagheripour, 2012; Oscak and Seker, 2012; Khandelwal and Monjezi, 2013; Ashena and Thonhauser, 2005; Dehghan et al., 2010; Koolivand-Salooki; 2017; Ceryan and Can, 2018; Anemangely et al., 2019; Miah et al., 2019). Recently, ANN and LS-SVM are becoming a more popular strategy for data-driven model development, prediction of formation elastic properties, and failure analysis to save the operational expenses and time.

5.1.2 Literature Survey

Several empirical correlations exist for calculation of UCS along the wellbore using rock porosity (ϕ), sonic travel time (Δt), elastic moduli, and other formation properties such as shale volume and resistivity (Fjaer et al., 1992; Edimann et al., 1998; Sonmez et al., 2004; Cheng et al., 2006; Odunlami et al., 2011; Singh et al., 2012; Rabbani et al., 2012). The rock strength can be estimated from the empirical correlations considering the lithology type (e.g., carbonate or clastic rocks) and availability of log variables. In addition, there are

so-called “worldwide” models which can be used irrespective of the rock type (Chang et al. 2006). Well-known correlations for estimating rock strength (UCS) using log data in the oil and gas industry are listed in Table 5.1.

Onyia (1988) investigated the in-situ rock strength and developed a correlation to estimate ultimate rock strength using drilling and logging data such as gamma-ray (GR) log, resistivity log, density log, and borehole compensated sonic logs. Fjaer et al. (1992) developed a correlation that is most commonly used to obtain the rock strength for shaly sand reservoir using three rock parameters, namely, Poisson’s ratio (ν), sonic velocity (Vp), and shale volume (Vsh). The shale volume measurement is costly and time-consuming approach through experimental core analysis, while it can be found from the wireline log data including GR log. For instance, the shale volume can be obtained using over the entire GR log of formation to capture the maximum and minimum magnitudes of GR if the full-length log is available. Moos et al. (2001) developed the UCS correlation for the clastic sandstone rocks using sonic travel time or acoustic velocity, and formation bulk density. Also, Chang et al. (2006) took into account only sonic travel time to obtain the UCS correlation. Both correlations of Moos et al. (2001) and Chang et al. (2006) did not consider the shale content effect (i.e., GR parameter) to estimate the in-situ UCS estimation. A number of studies have investigated the effect of various rock properties on rock strength; however, the ranking of variables was not discussed (Rajabzadeh et al., 2012; Haftani et al., 2015; Jamshidi et al., 2016; Kong and Shang, 2018; Jamshidi et al., 2018). In-situ UCS models have not been developed yet to capture the lithology indicator (GR), number of electron density, and acoustic travel time of the porous shay sand formation.

Table 5.1: Selected empirical correlations to estimate rock strength (UCS, MPa) for sandstone reservoir rocks.

Authors, year	Input variables	Correlation for UCS	Region where developed	Comments/Remarks
Freyburg, 1972	Vp	$0.035V_p - 31.5$	Germany	• It was developed for the relatively strong rocks.
McNally, 1987	Δt	1200exp(-0.036Δt)	Bowen Basin, Australia	• It is applicable to sandstones (fine-grained, both consolidated and unconsolidated).
Onyia, 1988	R_t, GR, ρ_b, Δt_c	$0.98(P + QLog_{10}(R_t) + R \times GR + S \times \rho_b + T \times \Delta t_c)$	Rogers Country, Oklahoma	• The correlation was developed based on multivariate regression analysis (MVRA) incorporating log properties.
Fjaer *et al.*, 1992	ν, Vsh, Vp	$3.3 \times 10^{-20}\rho^2 V_p^4(1-2\nu)\{(1+\nu)^2/((1-\nu)^2\}(1+0.78\text{Vsh})$	Gulf coast	• This correlation is applicable for sandstones if UCS$> 30\ MPa$. • Shale effect was considered; it is applicable for shaly sand reservoirs.
Vernik *et al.*, 1993	ϕ	$254(1-2.7\phi)^2$	Global	• It can be used for very clean, well-consolidated sandstones with $0.3 > \phi$.
Sarda *et al.*, 1993	ϕ	a) $(357 \exp(-10.8\phi)$ b) $(258 \exp(-9\phi)$ c) $115 \exp(-11.6\phi)$	-	• It is developed using Germingy-sous-Coulombs reservoir. • It is applicable for the range of $0.07 > \phi > 0$ (Eq. a)., $0.3 > \phi$ (Eq. b), and $\phi > 0.3$ (Eq. c).
Farquhar *et al.*, 1994	ϕ	$(208.08 \exp(-0.074\ \phi)$	North Sea	• It was developed for sandstones.
Weingarten and Perkins, 1995	E, K, Vsh	$(114 + 97 \times Vsh)K \times E$	USA	• It is applicable for unconsolidated sandstone reservoir.
Raaen *et al.*, 1996	Δt	$1.40 - 2.15\Delta t + 0.0083 \times \Delta t^2$	Norway, North Sea	• It was developed for relatively clean sandstones. • It is applicable for $140 > \Delta t > 90\ \mu s/ft$.
Raaen *et al.*, 1996	ϕ	$(43 - 140\phi + 63\phi^2)$	Norway, North Sea	• It is applicable for the range of $0.35 > \phi > 0.2$ • It is valid for relatively clean sandstone reservoirs.
Edlmann *et al.*, 1998	ϕ	$(129.54 - 3.225\phi)$	North Sea	• It was developed for sandstones.
Bradford *et al.*, 1998	E	$2.28 + 4.1089E$	Worldwide	• It is a more representative model for the sandstone reservoir of the Chicontepec basin.
Moos *et al.*, 2001	Vp, ρ_b	$1.745 \times 10^{-9} \times \rho_b \times {V_p}^{-3} - 21$	Cook Inlet, Alaska	• It was developed for coarse-grained sandstones and conglomerates.
Chang *et al.*, 2006	Δt	$1.4138 \times 10^7 \Delta t^{-3}$	Gulf Coast	• It is applicable for the weak and unconsolidated sandstone rocks.

Thus, it seems necessary to develop new correlation(s) for estimation of realistic continuous in-situ rock strength profile using dynamic wireline logging data in the absence of drilling or laboratory-based data.

The connectionist models, for example ANN and LS- SVM, are being increasingly used for predicting rock mechanical properties for wellbore failure analysis, reservoir geomechanics, underground excavation as well as rock mass characterization (Yang and Zhnag, 1997; Yılmaz and Yuksek, 2008; Ocak and Seker, 2012; Asoodeh and Bagheripour, 2012; Khandelwal and Monjezi, 2013; Barzegar et al. 2016; Negara et al. 2017; Matin et al. 2018; Abdi et al. 2018; Onalo et al. 2018; Ceryan and Can, 2018; Onalo, 2019). Although several research works have attempted to determine the elastic or geomechanical characteristics of rock using experimental core and log data, only a few studies have forecasted the rock strength (UCS) using wireline logs with AI or machine learning tools. The existing literature reveals that several input variables are adopted in soft computing smart models to obtain the UCS or Young's modulus (E) using rock physical properties or features such as porosity, unit weight (uw), dry density (dd), compressional wave velocity (Vp), water absorption (wa), water content (wc), Schmidt hammer rebound number or hardness number (SRn), Brazilian tensile strength (BST), point load index ($Is_{(50)}$), and slake durability index (I_d). The main studies conducted for rock strength prediction are listed in Table 5.2 with a focus on the corresponding research methods and shortcomings. Only a few studies investigated ranking of the input variables while determining the rock strength (UCS) (Majidi and Rezaci, 2013; Torabi-Kaveh et al., 2015; Matin et al., 2018).

Table 5.2: Applications of AI-based connectionist tools for rock strength and mechanical properties prediction.

Authors, year	Input variables	AI approach	Variable ranking	Comments/Remarks
Yilmaz and Yuksek, 2009	SRn, $I_{S(50)}$, wc, sonic velocity	ANN, ANFIS	No	• Authors used tansig transfer function for the feed-forward neural network that consists of 4 neurons for an input layer, 9 neurons for the hidden layer, and one output layer. • The predictive model performance was acceptable with a root means square error (RMSE) of 5.98% and an correlation coefficient (R^2) of 0.887.
Sharma *et al.*, 2010	RB, Vsh, Vp, Vs	ANN	No	• It was concluded that the neural network is better at capturing the complex relations between geophysical properties and the strength of rock.
Rabbani *et al.*, 2012	Total ϕ, RB, Sw	ANN	No	• To train the ANN network, the optimal number of neurons was used for the first layer (20 perceptions) and the second hidden layer (25) using the trial and error approach. • The correlation coefficient about 0.903 was found for the UCS deterministic model.
Yagiz *et al.*, 2012	SRn, ϕ, I_d, Vp, uw	ANN	No	• It was revealed that ANN gives relatively more precise results than a regression approach.
Majdi and Rezaei, 2013	Rock type, SRn, RB, ϕ	ANN	Yes	• Authors constructed a multilayer neural network model using different rock samples. • It offered better results, close to laboratory results, with an R^2 of 0.97, compared to the MVRA. • It was claimed that the most effective variables are density and SRn.
Nabaei and Shahbazi, 2012	Well trajectory	ANN	No	• The feedforward backpropagation network applied to predict the output using tansig transfer function for the first two layers.
Ceryan *et al.*, 2013	Vp, Vm, total ϕ	ANN	No	• Authors constructed the Levenberg–Marquardt algorithm-based network using input variables including solid part (Vm) of the carbonate rock samples.
Ceryan, 2014	P-wave durability, ϕ	ANN, SVM	No	• Ceryan (2014) concluded that the vector machine approach performs better than the ANN model. • The model's performance was investigated by adopting porosity and microstructural properties.
Momeni *et al.*, 2015	dd, $I_{s(50)}$, SRn, Vp	ANN	No	• It was concluded that the particle swarm optimization-based ANN predictive model exhibits greater reliability than the conventional ANN model for limestone and granite. • It was revealed that the dd and SRn have more importance than other variables ($I_{s(50)}$, Vp) to predict UCS.
Mohamad *et al.*, 2015	Density, BTS, Vp, $I_{s(50)}$	ANN	No	• The PSO-based ANN model was employed to find the relationship between input variables and UCS of soft rocks. • The predictive model showed a good match with real date where R^2 was equal to 0.971.

Torabi-Kaveh et al., 2015	RB, Vp, ϕ	ANN	Yes	• The researchers evaluated the performance of the ANN predictive model, and linear and non-linear regression equations using predictor variables of limestone rocks. • It was revealed that the Vp has a higher contribution than other input variables.
Barzegar et al., 2016	RRn, Vp, ϕ	ANN, ANFIS, SVM	No	• Different AI models were utilized to analyze the model performance. • It was claimed that SVM performs better than other models for travertine rocks.
Behnia et al., 2017	Quartz content, dd ϕ	GEP	No	• Gene expression programming (GEP) tool was employed to predict the intact rock strength parameters; the statistical criteria were used to assess the models.
Tariq et al., 2017	Ed, RB, Vs, Vp	ANN ANFIS SVM	No	• The neural network was developed with tansig transfer function between input layer (4 neurons) and hidden layer (20 neurons). • It was concluded that Ed has a higher relative importance than wave velocities and RB in carbonate rocks.
Negara et al., 2017	Grain density, ϕ, elemental spectroscopy	SVR	No	• They applied the SVR technique using the core data to build the predictive model. • It was demonstrated that elemental spectroscopy has a significant impact to estimate the UCS.
Rastegarnia et al., 2018	$Is_{(50)}$, Vp, wc, ϕ	ANN	No	• The sigmoidal and linear transfer functions were used between the input and hidden layers, and hidden and output layers to train the network. • Levenberg-Marquardt feedforward backpropagation algorithm was adopted to train the data.
Matin et al., 2018	SRn, $Is_{(50)}$, Vp, ϕ	SVR, DT	Yes	• It was found that a rain forest is a powerful tool for variable selection, and Vp is the most effective variable to predict UCS.
Onalo et al., 2018	GR, RB, Vsh	ANN	No	• They utilized the ANN model performance to predict compressional, shear travel time, and elastic moduli. • The sanding potentiality was also estimated and the results of the ANN-based predictive model and real data are compared. • The study did not examine variable ranking while estimating UCS with the predictive model.
Abdi et al., 2018	dd, wa, Vp, ϕ	ANN	No	• It was found that the neural network model is more powerful than the MLR technique to determine the UCS and/or E of sedimentary rocks. • They proposed the empirical correlations to predict UCS or E using the input variables.

Also, the use of AI approaches to estimate elastic properties and rock strength parameter was found to be limited (Sharma et al., 2010; Tariq et al., 2017; Onalo et al., 2018).

Sharma et al. (2010) introduced a correlation between UCS and physical properties of weakly cemented sandstone rock such as RB, synthetic Vp and Vs, and Vsh (volume of shale). An extended version of the correlation was built through employing multiple regression analysis; the model is valid for the range of 0 to 4 MPa of UCS using synthetic velocities and multiple parameters from wireline log data. The performance of the regression-based model was also compared with that of an ANN model. Torabi-Kaveh et al. (2015) utilized the ANN model and regressive equation for predicting the rock strength and elasticity modulus of carbonate and limestone rocks. They claimed that higher accuracy can be obtained from the ANN predictive model, compared with the multiple regression analysis (MVRA). Tariq et al. (2017) investigated the applicability of machine learning approaches such as ANN, ANFIS, and SVM to predict UCS. It was concluded that ANN outperforms the other techniques; an empirical correlation to predict the UCS was also proposed that gives better results, compared to a number of existing correlations. It was claimed that four geomechanical parameters such as RB, BP, VS, and dynamic Young's modulus (E) are sufficient to forecast the rock strength; however, the relative contribution of the variables was not investigated.

Furthermore, Onalo et al. (2018) constructed an MLP-based ANN model using input log variables such as GR, Vsh, and RB to predict sonic acoustic velocities. In their study, they obtained the elastic moduli using correlations; the results obtained from the correlation

were also compared with ANN predictions. The researchers also investigated sanding possibility as a case study of shaly sand reservoir, Niger delta basin. In their study, ranking of input variables for the predictive ANN model was not discussed. Abdi et al. (2018) applied the ANN and the MLR approaches to obtain the sedimentary rock strength parameters using experimental data sets. It was found that ANN results in higher accuracy than MLR. They introduced a correlation to estimate UCS using data sets for different rock samples of limestone, sandstone, conglomerate, and marl. In their research study, the correlation did not include the lithology type and they did not rank the input variables. Based on the literature review, a comprehensive investigation is further required to identify relative contributions of the input variables in prediction of rock strength as well as mechanical properties of shaly sand formations.

5.1.3 Research Objectives

To the best of our knowledge, data-driven predictive model and feature ranking of logging variables to study the rock strength with the aid of logging data have not been investigated systematically. The variable selection and ranking of logging parameters appear to be an important challenge for not only petroleum and mining engineers and drillers but also geologists to estimate ultimate rock strength profile using the wireline log data with acceptable accuracy. The logging data is expensive to assemble and depict the entire formation profile. The current study is planned to fill in the knowledge gap by finding the most contributing predictor log variables, corresponding to their relative performance while predicting the ultimate rock strength. It is believed that the research strategies

employed in this study can be less time consuming and cost-effective for efficient rock formation evaluation. The objectives of this research work are listed as follows:

- To investigate the performance of rock strength models and compare the predictive performance of deterministic tools through statistical analysis.
- To perform comprehensive parametric sensitivity analysis and find the most contributing variables for predicting UCS.
- Develop a new correlation to estimate in-situ formation UCS, capturing most effective log parameters.

5.2 Theory and Model Development

5.2.1 Log-based model to predict rock strength

The strength of rock is a function of rock mineralogy and lithology, acoustic properties, density, porosity, geometric factors, depositional environments, and compaction level. The most common logging types are the gamma-ray log, resistivity log, and porosity log. The comprehensive information on well logging principles, advantages, limitations and applications can be found in the available literature (Asquith and Krygowski, 2004; Mondol, 2015). The presence of clay minerals in a formation may strongly affect the rock materials properties. The GR log is extensively used to identify the lithology as well as to estimate the shale volume (clay content); this corrects their presence in shaly sand rock, yielding more accurate effective porosity (ϕ_e). Rock porosity (void space in the rock volume) can be obtained from the responses of neutron log (counts the hydrogen

concentration), density log (measures the bulk electron density), and sonic log (estimates acoustic travel time) in a formation. Also, the rock porosity varies primarily with particle size distribution, grain shape, packing arrangement, compaction, cementation, and clay content (Asquith and Krygowski, 2004).

The conductivity (inverse of resistivity) of a rock can be obtained from the deep induction resistivity log, which converts to true resistivity (RT, ohm-m) of the virgin zone in a formation. In the present study, several well log responses are incorporated to estimate the in-situ shale volume, acoustic primary velocity, and effective porosity; these rock properties are then used to estimate the rock strength.

The shale index (I_{GR}) is obtained in the current study by the following equation (Schlumberger, 1998):

$$I_{GR} = \frac{GR_{log} - GR_{min}}{GR_{max} - GR_{min}} \quad (5.1)$$

where GR_{log} is the gamma-ray value of the zone of interest; and GR_{max} and GR_{min} represent the maximum and minimum magnitudes of gamma-ray log over the entire formation, respectively. The clay content or shale volume (V_{sh}) is estimated from the following equation to simulate the nonlinear response of the tertiary rock formation (Larionov, 1969):

$$V_{sh} = 0.083(2^{3.7*I_{GR}} - 1) \quad (5.2)$$

The density porosity (ϕ_D) is determined using the formation bulk density response (ρ_b) by the following equation:

$$\phi_D = \frac{\rho_{ma} - \rho_b}{\rho_{ma} - \rho_{fl}} \quad (5.3)$$

In Equation (5.3), ρ_{fl} and ρ_{ma} refer to the density of the drilling mud, and the rock matrix density, respectively.

The clay corrected density porosity ($\phi_{D,e}$) and neutron porosity ($\phi_{N,e}$) are calculated as follows (Miah, 2014):

$$\rho_{b.c} = \rho_b + V_{sh} (\rho_{ma}-\rho_{cl}) \quad (5.4)$$

$$\phi_{D,e}=\frac{\rho_{ma}- \rho_{b.c}}{\rho_{ma}- \rho_{fl}} \quad (5.5)$$

$$\phi_{N,e} = PHIN - V_{sh} * \phi_{N,sh} \quad (5.6)$$

in which, $\rho_{b.c}$ refers to the clay corrected formation bulk density (g/cm^3); ϕ_N represents the neutron porosity (fraction) from the neutron log; and $\phi_{N,sh}$ and ρ_{cl} denote the neutron porosity (fraction) and formation bulk density (g/cm^3) of the shale zone, respectively.

The clay corrected effective porosity (ϕ_e) is determined through the combination of clay corrected neutron and density porosities in the porous rock formation using the following equation:

$$\phi_e = \sqrt{\frac{\phi_{D,e}{}^2+ \phi_{N,e}{}^2}{2}} \quad (5.7)$$

Although, the non-destructive testing for a rock sample is a more accurate way to predict the sonic primary velocity; however, it is a time-consuming, expensive, and tedious method. The sonic log is employed to obtain the profile of sonic acoustic compressional wave velocity (Vp). The compressional sonic acoustic time (DT, $\frac{\mu s}{ft}$) is inversely

proportional to the sonic primary wave velocity (Vp, km/sec) of the formation, as expressed below (Anemangely et al., 2017):

$$V_p = \frac{304.8}{DT} \tag{5.8}$$

When the shear sonic slowness is not available from the wireline log data, the synthetic shear wave velocity (Vs) can be estimated from the available correlations for various rock formations (Pickett, 1963; Caroll, 1969; Krishna et al., 1989; Greenberg and Castagna, 1992; Vernik et al., 2003; Brocher, 2005; Hossain et al., 2012; Bailey and Dutton, 2012; Lee, 2013; Ojha and Sain, 2014; Anemangely et al., 2017). A limited number of studies have attempted to obtain the shear wave velocity for the siliciclastic or shaly sand rock formations (Castagna et al., 1985; Han et al., 1986; Williams, 1990; Miller and Stewart, 1990; Ramcharitar and Hosein, 2016). Recently, Oloruntobi and Butt (2019) have proposed an empirical correlation to estimate V_s using V_p and ρ_b for any type of sedimentary rocks such as shaly sand reservoirs, as given below:

$$V_s = A\left[\frac{V_p}{\sqrt{\rho_b}}\right]^4 + B\left[\frac{V_p}{\sqrt{\rho_b}}\right]^3 + C\left[\frac{V_p}{\sqrt{\rho_b}}\right]^2 + D\left[\frac{V_p}{\sqrt{\rho_b}}\right] + E \tag{5.9}$$

In Equation (5.9), A (0.094), B (-0.881), C (2.605), D (-1.415), and E (-0.435) are the coefficients of the polynomial expression.

In the literature, there are several empirical correlations to calculate the rock strength as well as elastic properties using sonic acoustic travel time and formation density. Also, elastic properties such as Young's modulus (E), shear modulus (G), bulk modulus (K), and

Poisson's ratio (υ) can be estimated using density and sonic logs. The following equations can be utilized to obtain the four main isotropic elastic properties at dynamic conditions using bulk density (ρ_b), sonic acoustic compressional (Vp), and shear wave velocity or V_s (Li and Fjaer, 2012):

$$\nu = 0.5({V_p}^2 - 2{V_s}^2)/({V_p}^2 - {V_s}^2) \quad (5.10)$$

$$E = \rho_b {V_s}^2(3{V_p}^2 - 4{V_s}^2)/({V_p}^2 - {V_s}^2) \quad (5.11)$$

$$G = \rho_b({V_s}^2) \quad (5.12)$$

$$K = \rho_b({V_p}^2 - \frac{4}{3}{V_s}^2) = \frac{E}{3(1-2\nu)} \quad (5.13)$$

Any of the elastic properties (such as E) can be calculated using two remaining properties (e.g., K and ν) (Fjær et al., 2008). The dynamic moduli is always larger than the static ones, as the dynamic strains are always smaller than static strains of rocks (Demirdag et al., 2010; Najibi et al., 2015). The conversion factor from static to dynamic moduli is dependent on the formation porosity, confining pressure, degree of loading, and other lithological factors (Rasouli et al., 2011). In the current study, the UCS parameter is estimated using the following model for the shaly-sand formation (Fjær et al., 2008; Rabbani et al., 2012):

$$UCS = 3.3 * 10^{-20}{\rho_b}^2 * {V_p}^4(1-2\nu)\frac{(1+\nu)^2}{(1-\nu)^2}(1+0.78V_{sh}) \quad (5.14)$$

5.2.2 Log data collection and quality assurance

A combination of several wireline logs such as gamma-ray, deep resistivity, porosity, and caliper logs is used to identify the lithology, and to estimate the formation thickness, and

rock physical and mechanical properties. The gamma-ray (GR) log is a complementary log that is able to identify the formation rock type, and to measure the amount of radioactive elements (API) and clay content. The density log measures the electron density (bulk density, RB) to estimate the formation true porosity, while the neutron log counts the hydrogen concentration (NP) as well as porosity of fluid-filled formation. The sonic log helps to measure the acoustic wave in the form of compressional travel time (DT) or shear slowness. On the other hand, the resistivity log measures the conductivity (inverse of resistivity, RT) in the rock formation. All of these log variables are used in this study to develop the data-driven model; field log data are collected from a shaly sand reservoir located in the Bengal basin.

The available field well logs data (such as GR, RT, RB, and DT) and neutron-density porosity are utilized to predict the in-situ rock strength profile in the shaly sand rock formation. A total of 182 data samples are used for each log variable in the study. The log data quality is also confirmed to ensure the reliability of each log dataset variable by checking the depth shift and borehole conditions. The output and input variables are classified into training, testing, and validation phases with 75%, 15%, and 10%, respectively. All the programming tasks related to this study are carried out using Matlab programming environment.

5.2.3 Development of connectionist predictive models

ANN model: The ANN is composed of different components such as artificial neurons (connected in each layer), weight factors, bias term, and transfer function in a connectionist process unit system. The ANN can be adopted with a single-layer perception, multilayer perception (MLP), and/or radial basis function networks (Ali, 1994; Mohaghegh et al., 1996; Razavi and Tolson, 2011). The MLP is one of the conceptually attractive feedforward neural network approach employed in estimation of rock formation properties, rock mechanics, stability of underground openings, wellbore failure analysis, and rock engineering (Yang and Zhang, 1997; Meulenkamp and Grima, 1999; Helle and Bhatt, 2002; Yılmaz and Yuksek, 2009; Ocak and Seke, 2012; Khandelwal and Monjezi, 2013). The deep learning MLP model consists of at least four layers with one input layer, at least two hidden layers, and an output layer for target variables. The training algorithm and activation (transfer) function are the most important components of an ANN structure.

In this study, the ANN model is processed with the Levenberg–Marquardt (LM) algorithm for training function (Trainlm) to adjust the weight factors through simulation in the connectionist model. The LM algorithm is usually faster and more reliable in the back-propagation system for ANN model than other standard back-propagation methods (Ceryan et al., 2013). The Tansig type activation function is used between input and hidden layers while Purelin (linear) transfer function is employed between the last hidden layer and output layer in the model. Taheri-Garavand et al. (2015) claimed that a hidden layer with a less number of neurons is desired in an ANN model because of shrinking the neural network structure as well as increasing the learning potential.

In the current study, the number of hidden layers is chosen by a trial and error strategy to optimize the ANN structure such that the maximum regression coefficient and minimum mean squared error (mse) as the selection criteria are attained to find the best configuration in the smart approach. The number of neurons in the hidden layer is optimized to enhance the network performance as well as to save the computational training time. An ANN structure is shown in Figure 5.1. A flow chart is also depicted in Figure 5.2 to present the ANN model development steps and the optimization strategy in the study. In Figure 1, N1 and N2 refer to the number of neurons in the hidden layers 1 and 2, respectively.

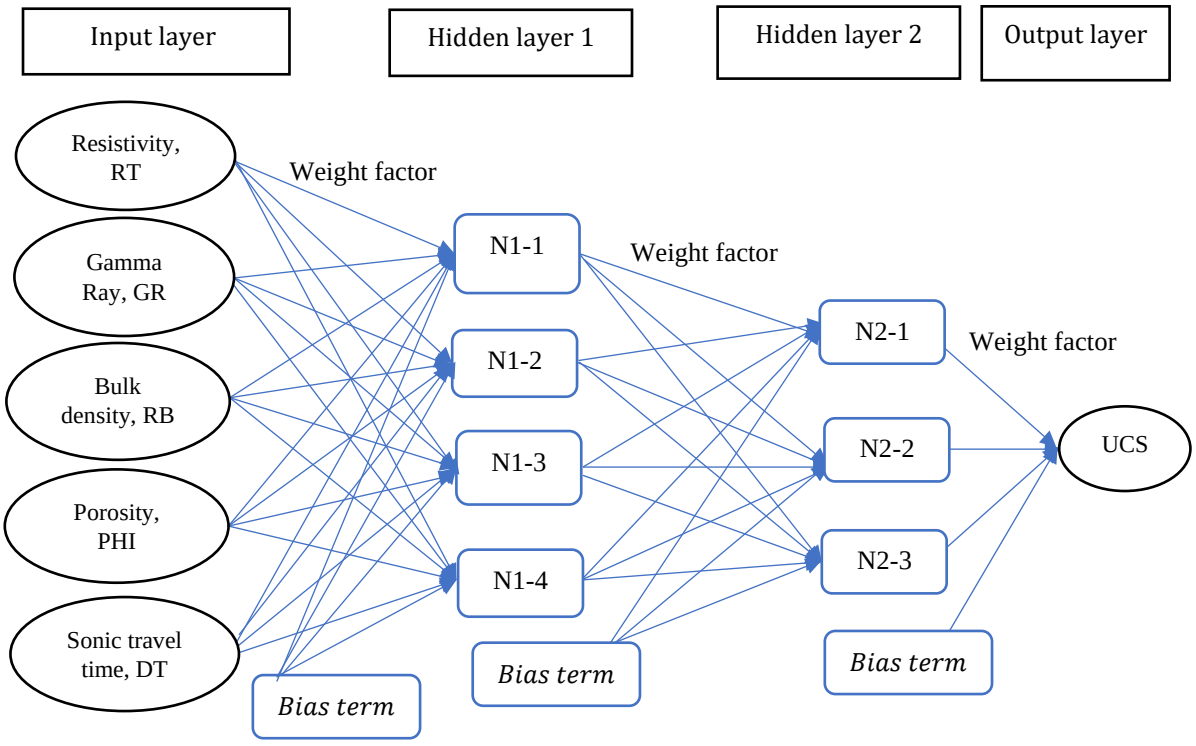

Figure 5.1: Schematic of ANN architecture employed in the current work.

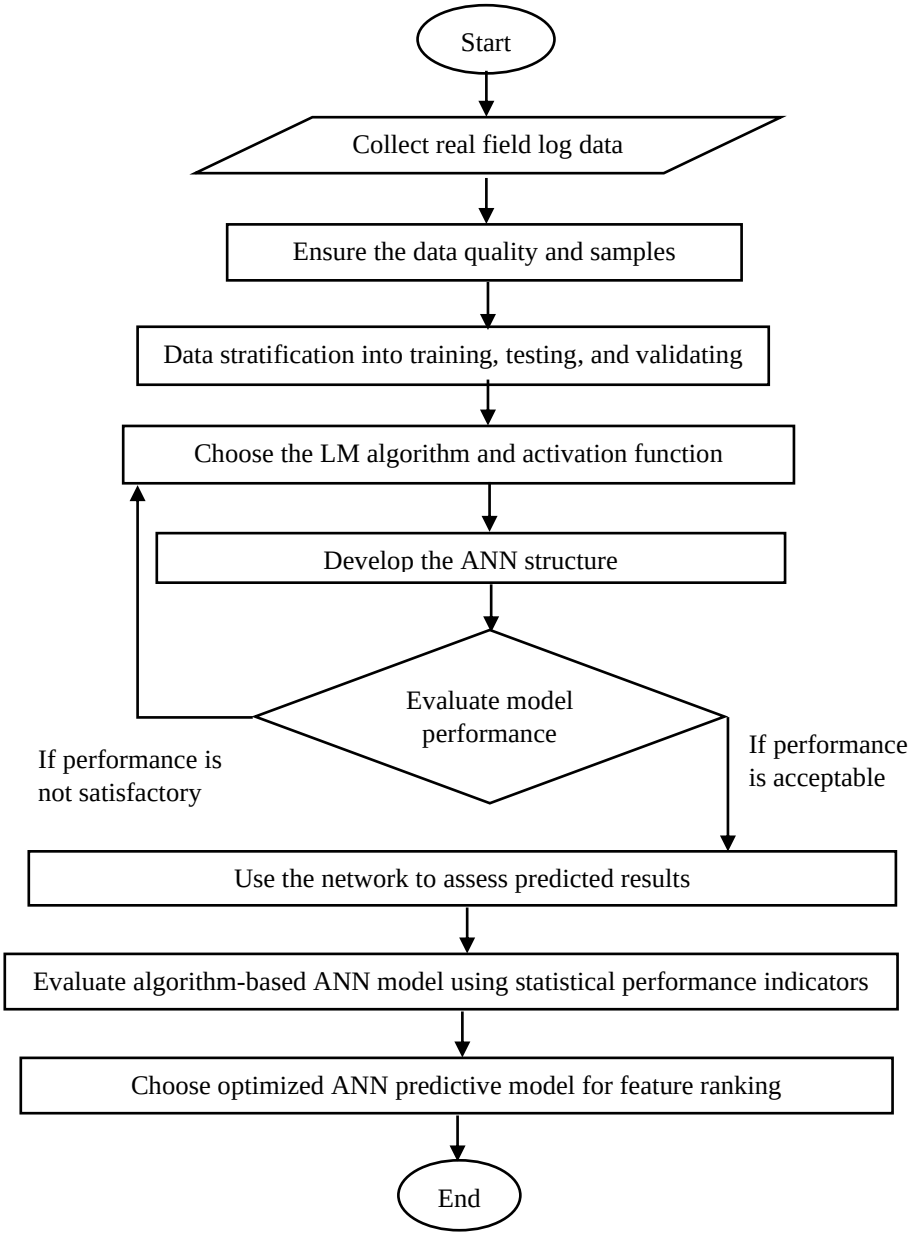

Figure 5.2: A flowchart for ANN model development and to predict the output variable

When the transfer function of purlin is used for output neuron, the generalized mathematical expression for MLP-based neural network with Tansig transfer function is given as follows:

$$y_m = \left[\sum_{j=1}^{m} \omega_{jm}\, \text{Tansig}\left(\sum_{i=1}^{n} \omega_{ij} x_i + b_j\right) + b_m\right] \quad (5.15)$$

In Equation (5.15), y_m stands for the output variables; x_i introduces the vector of target variables (i = 1, 2, 3, 4, 5…); b_m resembles the bias term for output layers; ω_{ij} denotes the connection weight on the link from i to j node between input and hidden layers; m is the number of hidden nodes; and n refers to the number of input variables.

LS-SVM model: The SVM is one of the powerful artificial intelligence approaches used in data classification and regression analysis. The SVM uses a subset of training points in the support vectors (decision function), which was first introduced by Vapnik (1995). The LS-SVM is a modified version of the classic SVM algorithm proposed by Suykens and Vandewalle (1999). This modified version is less complex than the classic SVM algorithm. It helps to reach the solution of a worsening problem with less data points more efficiently by setting up a linear set of equations through employing SVM instead of the quadratic programming (Suykens and Vandewalle, 1999). Compared to SVM, the LS-SVM learning method is less time consuming. More information regarding the theory and algorithm with different features of SVM or LS-SVM can be found in the literature (Smola et al., 2004; Sebtosheikh et al., 2015; Esfahani et al., 2015). The following equation is used in the LS-SVM:

$$y(x_i) = \omega^T \varphi(x_i) + b \quad where\ x_i \in R^n \text{ and } y_i \in R \quad (5.16)$$

where the nonlinear function $\varphi(.)R^n \rightarrow R^{n_i}$ represents the primal space to a feature space with higher dimensions. The dimension n_i of this space is only defined in an implicit way; b is a bias term; and $\omega \in R^{n_i}$ introduces the weight factor. The optimization problem can be written for function estimation in the LS-SVM (Suykens et al. 2002, Esfahani et al., 2015), as shown below:

$$\text{Minimize } J(\omega, e_i) \cong \frac{1}{2}\omega^T\omega + \gamma \sum_{i=1}^{n} {e_i}^2 \tag{5.17}$$

Subject to the following constraint,

$$y_i = \omega^T\varphi(x_i) + b + e_i\,, i = 1,2,3, \ldots, n \tag{5.18}$$

In Equation (5.18), e_i represents the error variable, and γ resembles the regularization (annealing) parameter to prevent overfitting.

The kernel function is $K(x, x_i) = \varphi(x_i)^T\varphi(x_i)$, which needs to be satisfied with Mercer's condition. After eliminating ω and e_i, the final expression can be formulated for the LS-SVM function estimation as follows:

$$y(x_i) = \sum_{i}^{n} \alpha_i K(x, x_i) + b \tag{5.19}$$

where b and α are the solutions to the linear system expressed through Equation (5.19). The α (weight factor) is a vector with the size of $n \times 1$; x is the training sample; and x_i refers to the support vector.

The kernel function-based LS-SVM structure used in this study is illustrated in Figure 5.3. A summarized methodology for development of kernel-based LS-SVM model is shown in Figure 5.4.

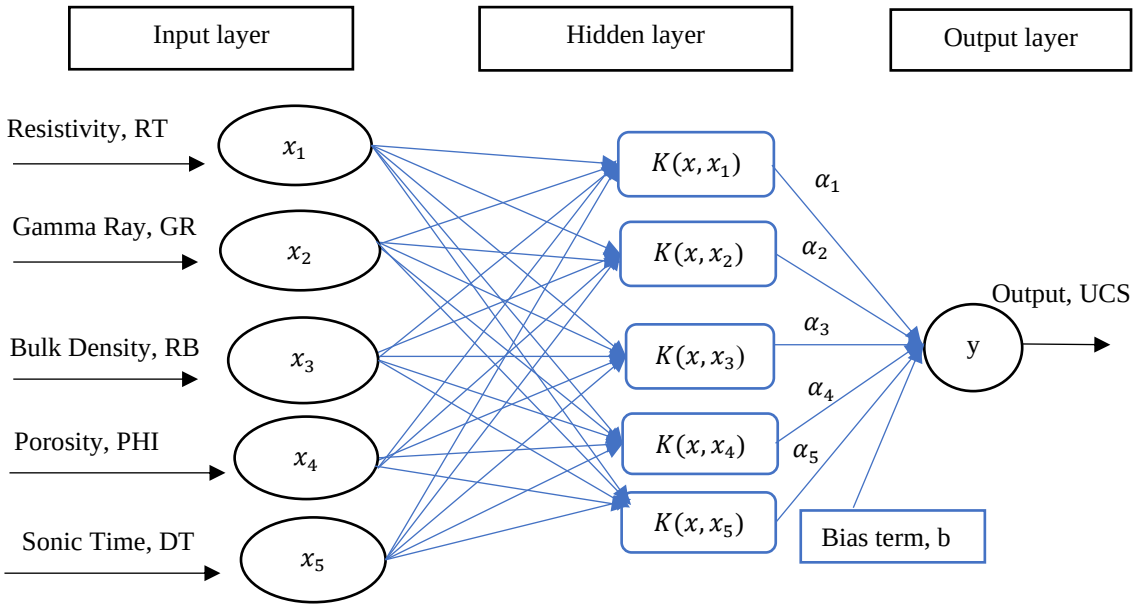

Figure 5.3: A generalized structure of LS-SVM proposed in the study.

There are many kernel functions that can be used in LS-SVM such as linear, polynomial, spinal, radial basis, and sigmoidal. Among all kernel functions, the Gaussian radial basis kernel function (RBF) is mostly used in the LS-SVM learning strategy to find the best output (Suykens et al., 2002; Samui, 2008) due to its computational simplicity and other features (e.g., capable of solving nonlinear problems). The Gaussian RBF can be defined mathematically as follows (Samui, 2008):

$$K(x, x_i) = \exp\left(-\frac{\left(||x_i - x||\right)^2}{2\sigma^2}\right) \tag{5.20}$$

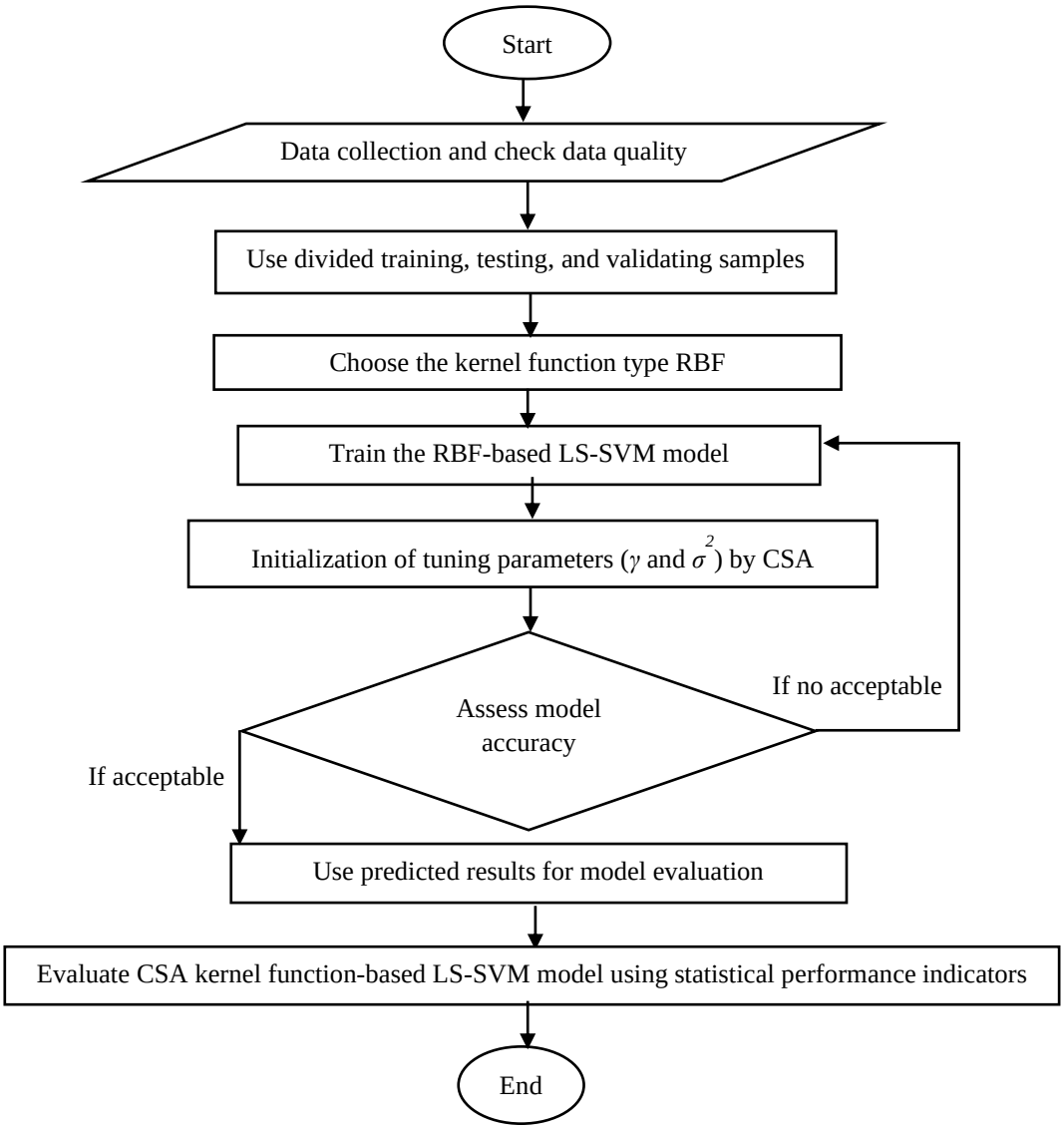

Figure 5.4: A flowchart for kernel-based LS-SVM model development

In the kernel function of RBF-based LS-SVM, the regularization and kernel parameters (also named as tuning parameters) γ and σ^2 are adjusted through a global optimization technique, namely, the coupled simulated annealing (CSA) (Xavier-de-Souza et al., 2009). The CSA optimization process is proven to be more effective than multi-start gradient descent technique (Suykens et al., 2002; Rostami et al., 2019). Similar to the MLP-ANN

model, the databank for the log data is divided randomly into three sub-datasets to construct the SS-SVM models using different kernel functions. The total samples are categorized into three groups including 75% for training, 15% for testing, and 10% for validation in the LS-SVM connectionist model with the CSA optimization approach.

5.2.4 Model performance assessment

Four statistical indicators are used in this study to analyze the predictive model performance. The indicators are root mean square error (RMSE), correlation coefficient (R), average absolute percentage relative error (AAPE), and maximum average absolute percentage error (MAPE). The mathematical expressions for all performance indices are listed below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_{m,i} - Y_{p,i}\right)^2} \tag{5.21}$$

$$R = 1 - \frac{\sum_{i=1}^{n}\left(Y_{m,i} - Y_{p,i}\right)^2}{\sum_{i=1}^{n}\left(Y_{m,i} - Y_{m,mean}\right)^2} \tag{5.22}$$

$$AAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(Y_{m,i} - Y_{p,i})}{Y_{m,i}}\right) * 100 \tag{5.23}$$

$$MAPE = Max.\left|\frac{(Y_m - Y_p)}{Y_m}\right| * 100 \tag{5.24}$$

In the preceding equations, *n* indicates the total number of samples; Y_m resembles the measured variable; $Y_{m,mean}$ is the mean value of Y_m; and Y_p represents the predicted output variable. The accuracies of the data-driven models are analyzed on the basis of the low or

high value of statistical indices. In the study, the model is best for the high value of R (close to 1) and low values of RMSE, AAPE, and MAPE.

5.2.5 Sensitivity analysis and variable selection

In this study, a systematic strategy is employed to perform the parametric sensitivity analysis as well as to find the relative importance of the input log variables in the AI-based predictive rock strength models (see Figure 5.5).

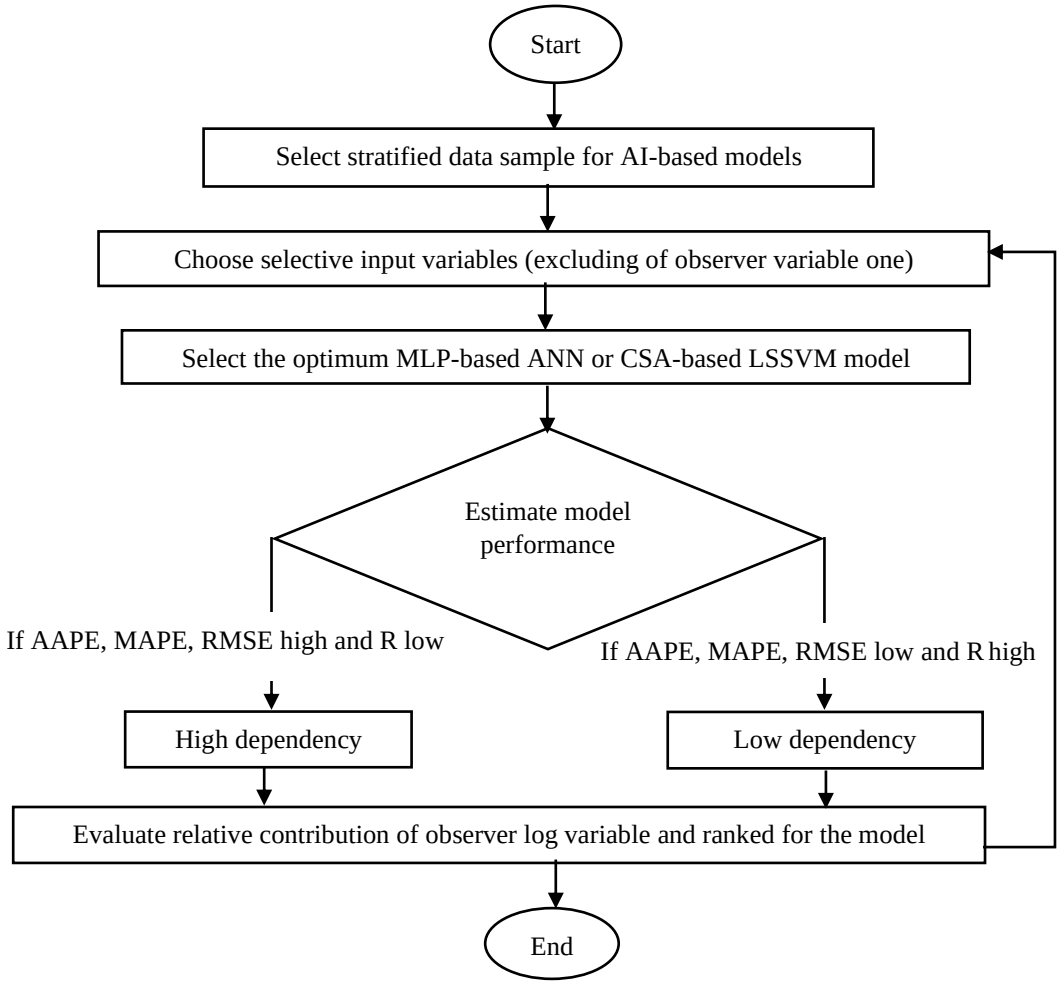

Figure 5.5: A summarized flow chat of parameter sensitivity analysis and ranking using AI approaches.

Furthermore, the optimized AI models predict the UCS; the model performance is assessed using the statistical criteria. Based on the contribution of an input variable to the predictive models, the input parameters are ranked. The variable ranking approach is called 'single variable elimination' of the data-driven model, while it has selective multiple input variables to predict the output variable using the optimized AI structure. In the log parameter ranking through the data-driven model, if the model results in high AAPE, MAPE, and RMSE and low R, implying that the eliminated variable has high impact on the model. It is worth noting that only most influential parameters are used to develop the new correlation for obtaining continuous in-situ UCS profile for the clastic sedimentary reservoir rocks through multivariable regression analysis with real field application.

5.3 Results and Discussions

5.3.1 Data analysis

The radioactivity properties including gamma-ray, deep resistivity, formation bulk density, and sonic log responses considerably change with depth throughout a shaly sand formation. For the data set under study, the minimum and maximum values of gamma-ray over the entire lithology log of the formation are 77 and 155 API, respectively, which are used to calculate the shale volume. The average shale volume extent is 11.84%, while the minimum value is 0.26% in the studied depth interval. The statistical information on the log data samples and the estimated values of rock characteristics are presented in Tables 5.3 and 5.4 over the entire depth of interest reservoir rock zone, respectively. Note that the

formation properties in Table 5.4 were derived from the statistical values in Table 5.3 following the methodology outlined in Section 5.2.1.

Table 5.3: Summary of the statistical values of the used log data.

Log parameters	Max.	Min.	Mean	St. Dev.	Sample var.
GR (API)	157.82	76.28	100.19	13.87	192.51
RT (ohm-m)	39.70	13.70	22.67	4.96	24.58
RB (g/cc)	2.53	2.30	2.37	0.0425	0.0018
NPHI (v/v)	0.2039	0.1455	0.17	0.013	0.0002
DT (µs/ft)	97.40	85.84	92.90	2.43	5.89

Table 5.4: Summary of log-based formation properties magnitudes in the study

Formation properties	Maximum	Minimum	Mean
Compressional wave, Vp (Km/s)	3.59	3.13	3.28
Shear wave, Vs (Km/s)	1.59	1.51	1.54
Poison's ratio, (fraction)	0.31	0.25	0.29
Porosity, PHI (v/v)	0.18	0.06	0.15
Rock strength, UCS (MPa)	78.82	24.40	33.19

The clay corrected effective porosity (PHI) is estimated using neutron and density porosity; it is then used to obtain the in-situ UCS of shaly sand formation using an empirical correlation presented in Fjær et al. (2008). The clay corrected porosity of a siliciclastic rock

changes with the heterogeneity of pore diameter, size and shape of pores, and formation density along the direction of vertical depth. Also, the primary acoustic velocity (inverse of sonic compressional travel time) and in-situ Poisson's ratio varies with respect to the vertical depth (6955.95-7015.33 ft) due to the rock formation compaction, cementation, and heterogeneity, as shown in Figure 5.6.

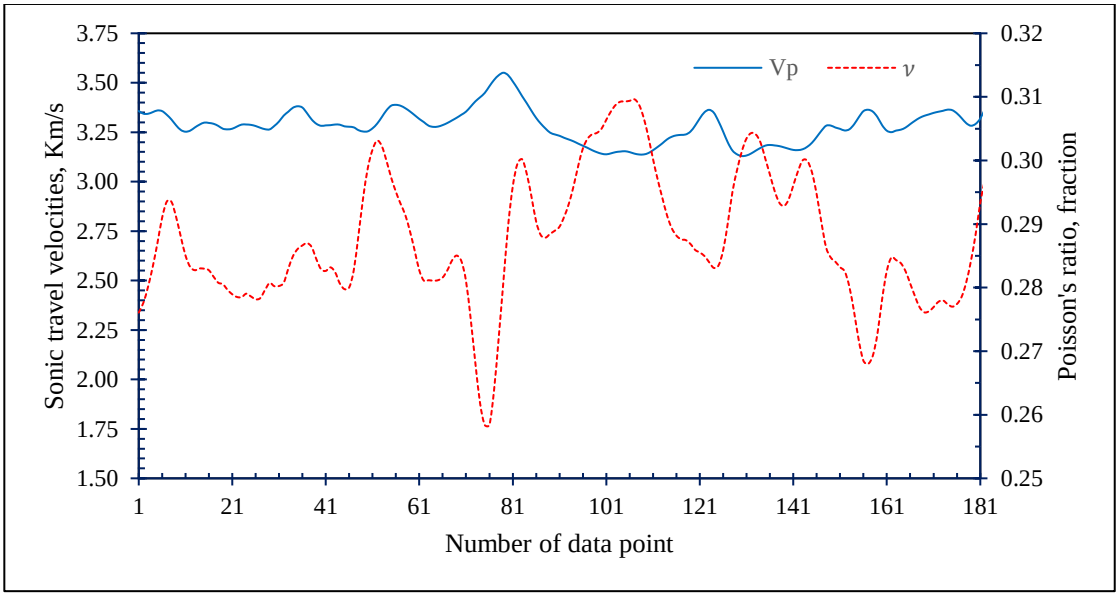

Figure 5.6: Profile of compressional wave and Poison's ratio in the rock formation.

The in-situ rock strength, UCS, varies vertically due to the overall effect of rock heterogeneity, formation radioactive mineral depositions, compaction, pore structure, grain size, packing, and density of the rock, as depicted in Figure 5.7. The rock strength of a shaly sand formation increases with an increase in gamma-ray and bulk density, while it is inversely proportional to the acoustic travel time as well as rock resistivity and porosity. The correlation matrix between UCS and other formation properties (e.g., porosity and true resistivity) is shown in Table 5.5.

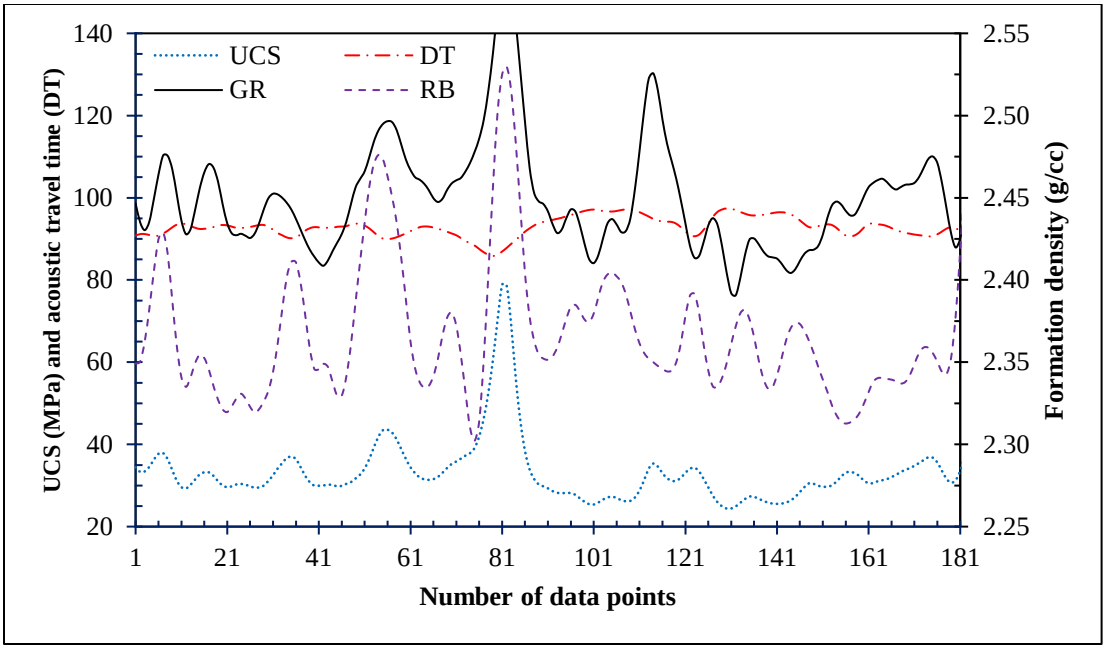

Figure 5.7: Variation of rock strength with formation gamma-ray, sonic travel time, and bulk density, depending on the number of data points.

Table 5.5: Correlation matrix between rock strength and formation characteristics

	UCS	RT	PHI	RB	DT	GR
UCS	1					
RT	-0.44106	1				
PHI	-0.84662	0.473699	1			
RB	0.664292	-0.59155	-0.84257	1		
DT	-0.77644	0.311042	0.464583	-0.29861	1	
GR	0.852674	-0.37541	-0.8225	0.523704	-0.58251	1

5.3.2 Data-driven model performance

A feed-forward artificial neural network is designed using stratified data samples. The optimum structure is obtained with the algorithm of LM; the optimal model has one input layer with 5 neurons, the first hidden layer with 4 neurons, the second hidden layer with 3 neurons, and one output layer with one neuron. The validation performance versus the number of epochs is illustrated in Figure 5.8a for the constructed MLP-based ANN model. Figure 5.8b also shows the performance of training phase in terms of number of epochs. In this study, the validation phase is conducted to tune the model parameters and terminate the neural network training stage before overfitting; μ (mu) is the Marquardt parameter in the training step of the network. The best validation performance is found at the epoch number of 10 with a mean square error of 0.02243. The magnitudes of the gradient and mu are 0.02034 and 0.01 at epoch 16, respectively, in the training stage. The correlation coefficient is close to one for both training and testing stages in the optimized ANN model (see Figure 5.9). It follows that the predicted results are in good agreement with the actual results for all training, testing, and validation data sets. The MLP based ANN gives less error (percentage) with an MAPE of 2.3692 and 1.1956% for the training and testing phases, respectively.

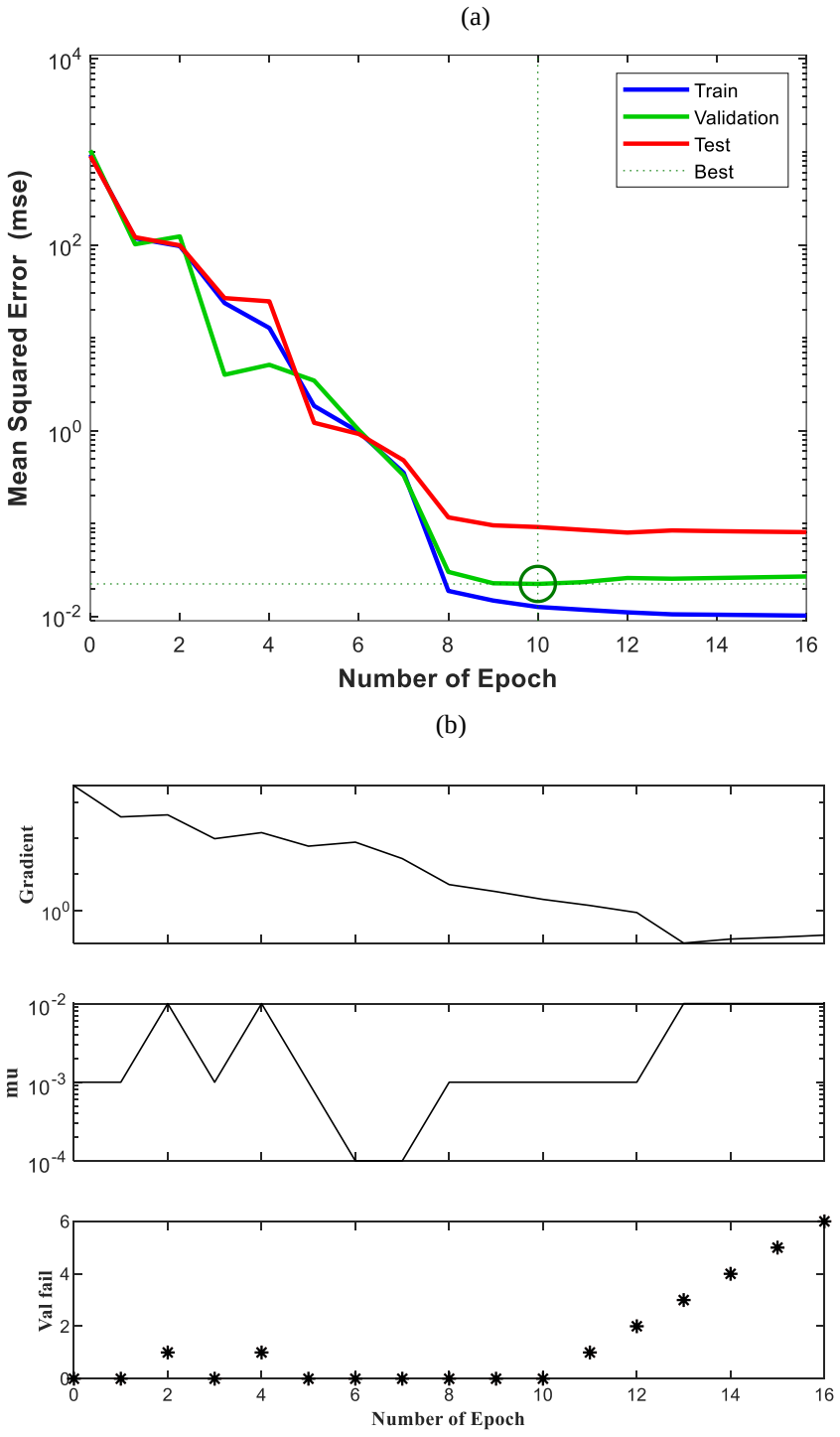

Figure 5.8: Graphically representation of a)Validation performance and b) training phase for optimized ANN model.

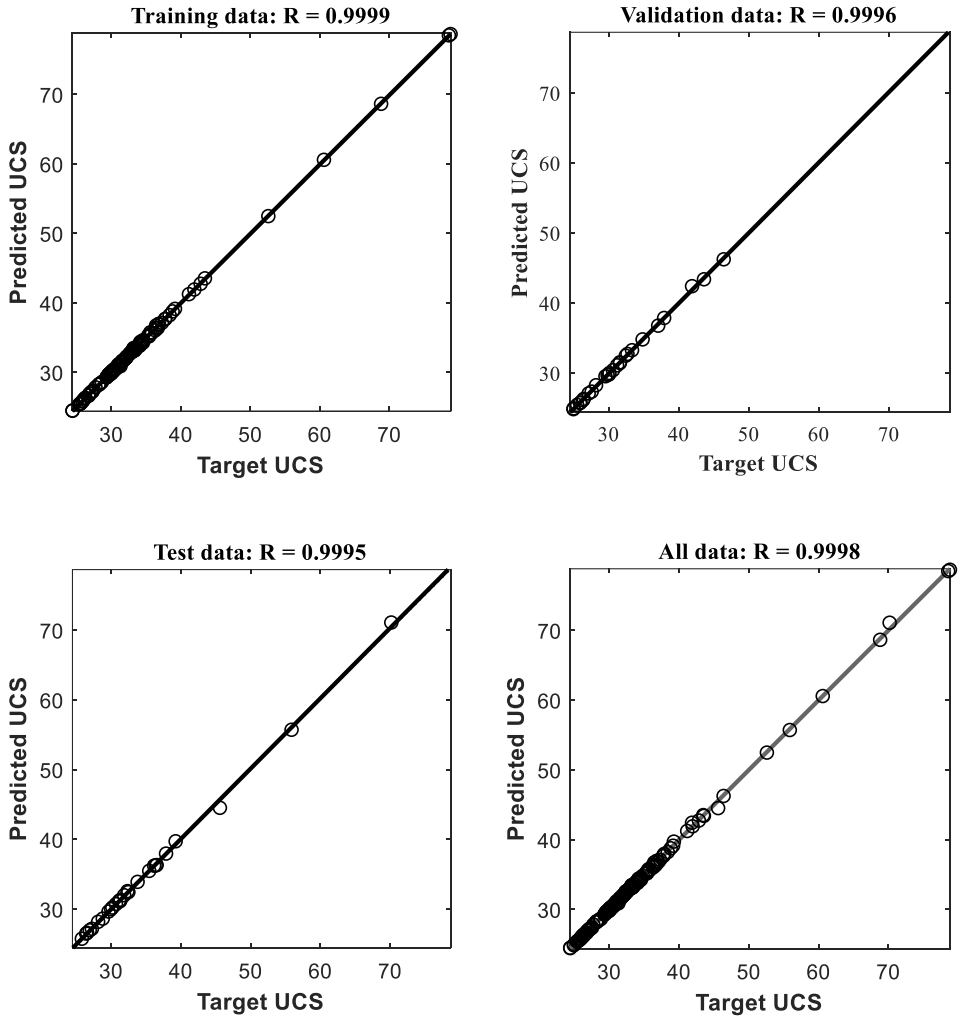

Figure 5.9: Predictive performance of the optimized ANN model.

In addition, the CSA optimization technique is used in the LS-SVM model as an iterative random search strategy. The optimization procedure is repeated several times to reach the optimum global point. Figure 5.10 displays the scatter plots (target values versus predictions) of the training, testing, and validation steps for the optimized LS-SVM model.

The initial values of the annealing and tuning parameters (γ and σ^2) are 505962.56 and 104.38, respectively. The final tuning parameters γ and σ^2 of the radial kernel function (Gaussian) based log data-driven model for UCS are 3.62e+11 and 1.74e+03, after 14 iterations. The RBF-based model results in a value of MAPE (%) equal to 0.0476 for the training and 0.2656 for the testing, respectively; the correlation coefficient is close to 1. The RBF based LS-SVM predictive model has a greater performance in terms of accuracy and reliability; it leads to the lowest RMSE, AAPE, and MAPE and high R. The statistical information for both MLP-based ANN and kernel function-based LS-SVM models is summarized in Table 5.6.

Table 5.6: Comparison of data-driven predictive models performance in the study

AI approach	Prediction phase	AAPE	MAPE	RMSE
MLP-ANN	Training	0.2803	2.3692	0.1735
	Testing	0.2779	1.1956	0.1314
LS-SVM	Training	0.0118	0.0476	0.0048
	Testing	0.0306	0.2656	0.0199

5.3.3 Parametric sensibility analysis and variable Selection

The optimized ANN structure in this study has two hidden layers with the topography of (4-4-3-1). The selected ANN model is used to perform parametric sensitivity analysis. Based on the statistical analysis, the log parameters are ranked based on their relative

importance in the predictive UCS model. To further understand the effect of individual variables, one variable is excluded at each time and the remaining four variables are used in both ANN and LS-SVM models. In the model schemes A through E, the UCS predictions are obtained in the absence of one input variable. Due to the less significance of RT and PHI (input parameters excluded from the predictive models), models D and E show better performance (such as lower error and greater correlation coefficient), compared with the other models. Model A leads to a weaker performance with a low value of the correlation coefficient and high magnitudes of statistical errors due to the absence of an important log variable, DT, in the model (Figure 5.10). Tables 5.7 and 5.8 present the performance metrics for the developed models. The statistical information of both connectionist techniques is presented in Figure 5.11.

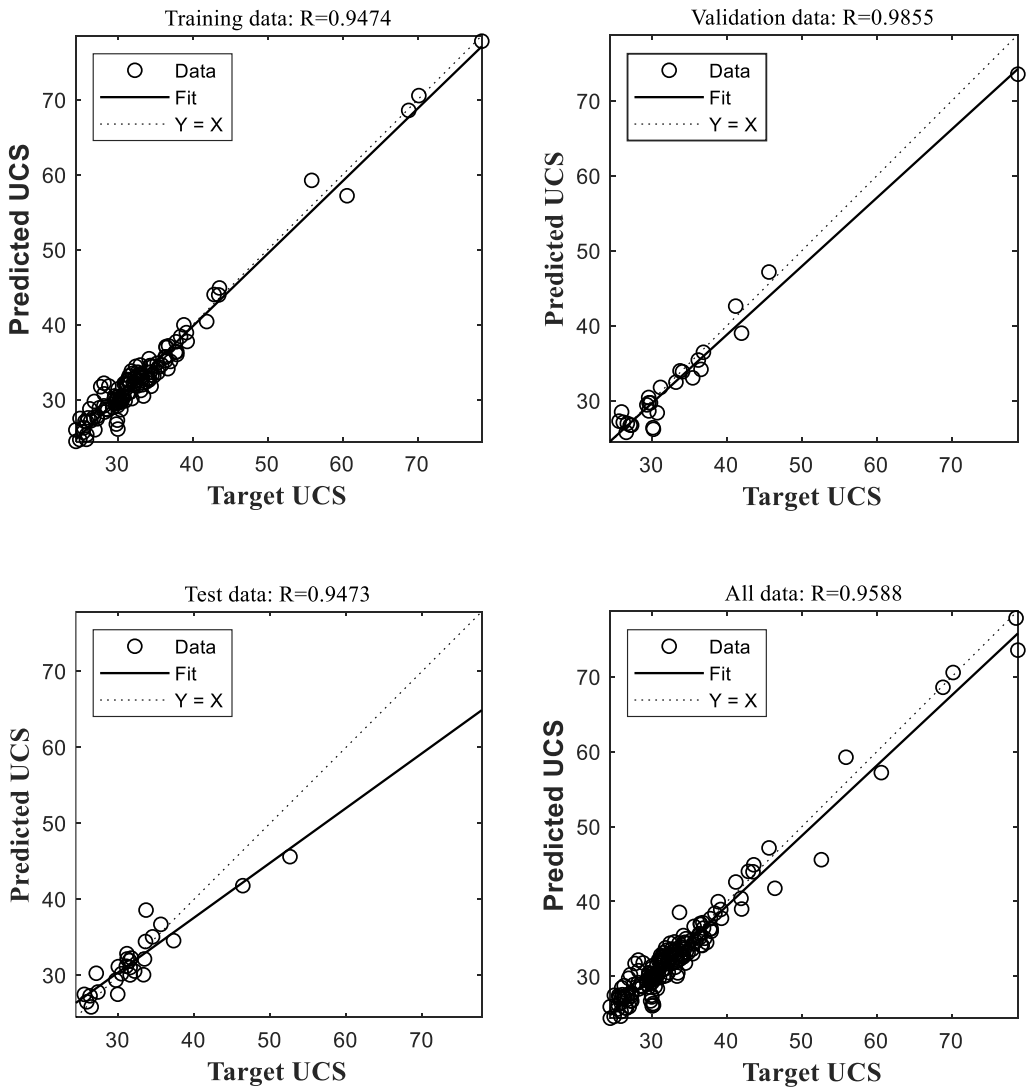

Figure 5.10: Performance of model A at different stages in the absence of DT (observer) variable.

Table 5.7: Investigating Investigating the impact of excluded variable on the performance of the MLP-based ANN model with four layers.

Model scheme	Predictor variables	Excluded variable	Phase	AAPE	RMSE	R
A	RT, GR, RB, PHI	DT	Training	3.60	1.60	0.947
			Testing	4.84	2.26	0.947
B	RT, RB, PHI, DT	GR	Training	1.85	0.96	0.983
			Testing	1.90	0.84	0.991
C	RT, GR, PHI, DT	RB	Training	0.63	0.29	0.999
			Testing	0.53	0.23	0.999
D	GR, RB, PHI, DT	RT	Training	0.18	0.10	0.999
			Testing	0.16	0.07	0.999
E	RT, GR, RB, DT	PHI	Training	0.005	0.003	1
			Testing	0.004	0.004	1

Table 5.8: Importance of input variables in the RBF-based LS-SVM model.

Model scheme	Predictor variables	Observer variable	Phase	AAPE	RMSE	R
A	RT, GR, RB, PHI	DT	Training	2.85	1.20	0.973
			Testing	4.41	2.78	0.938
B	RT, RB, PHI, DT	GR	Training	1.08	0.61	0.994
			Testing	1.42	0.69	0.952
C	RT, GR, PHI, DT	RB	Training	0.44	0.19	0.999
			Testing	1.71	2.38	0.956
D	GR, RB, PHI, DT	RT	Training	0.01	0.002	0.999
			Testing	0.02	0.018	1
E	RT, GR, RB, DT	PHI	Training	0.005	0.003	1
			Testing	0.004	0.004	1

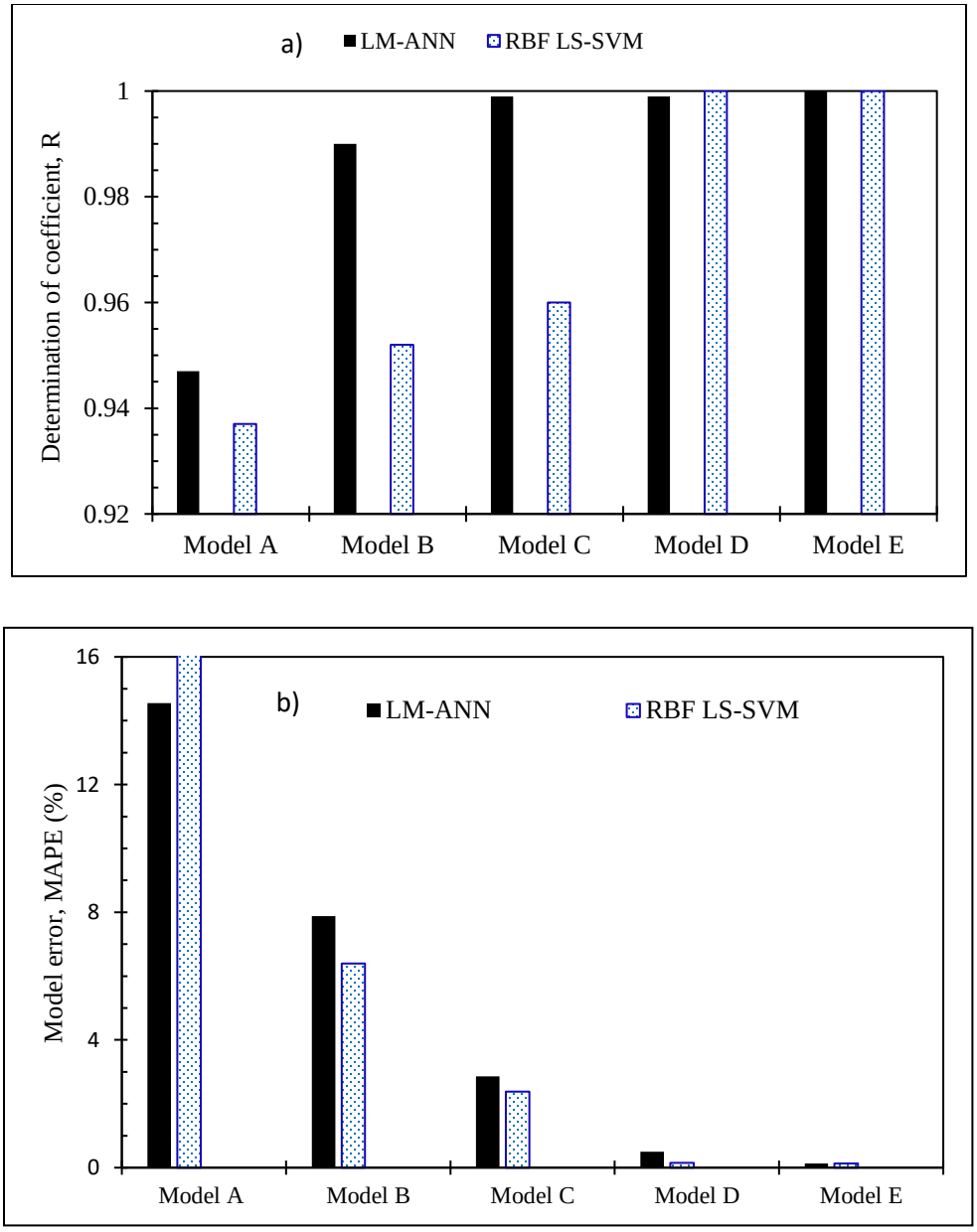

Figure 5.11: Comparison of a) R (top) and b) statistical error (bottom) performance for various model schemes in the absence of observer variable.

According to the statistical analysis conducted on the LM-based ANN approach, the significant input variables are the sonic travel time, gamma-ray, formation bulk density,

porosity, and true resistivity (high to low ranked order) for prediction of UCS. Also, the same order of variable importance is attained for the RBF-based LSSVM model. Based on the literature, the sonic porosity log (e.g., sonic travel time) is needed to estimate the rock strength using lithology-based correlations. The current study demonstrates that acoustic compressional travel time (DT) parameter has the highest contribution to the UCS value in a shaly sand formation. The rock resistivity and porosity have minor significance while predicting in-situ UCS profile. According to the different testing, validation, and generalization approaches used in the current study, the most significant predictor variables are DT, GR, and RB to estimate the rock strength of siliciclastic rock formations. Also, these variables are vital in capturing dense minerals effect, clay content, acoustic velocity, and bulk density of rocks.

5.3.4 Development of new correlation for rock strength estimation

Several correlations have been developed to estimate rock strength (UCS) using either core or wireline log properties data. Thus, it seems that a new log-based UCS correlation is required to obtain in-situ rock strength profile that can capture the mineralogical or clay effect, dynamic formation acoustic travel time, and bulk density. A set of 175 in-situ data samples of a shaly sand reservoir in the Bengal basin is used in this study to figure out the influences of formation properties on the rock strength as well as to develop a new correlation. The relationships between the rock strength and formation bulk density, gamma-ray, and acoustic travel time are illustrated in Figures 5.12 through 5.14. The formation bulk density is highly heterogeneous with respect to depth in the studied

formation. The in-situ rock strength increases with an increase in the formation bulk density due to the higher number density of electrons. The rock strength also increases with increasing gamma-ray due to the radioactive minerals and clay content in the shaly sand rock. The in-situ rock strength decreases with increasing acoustic travel time or decreasing compressional wave in the clastic rock formation.

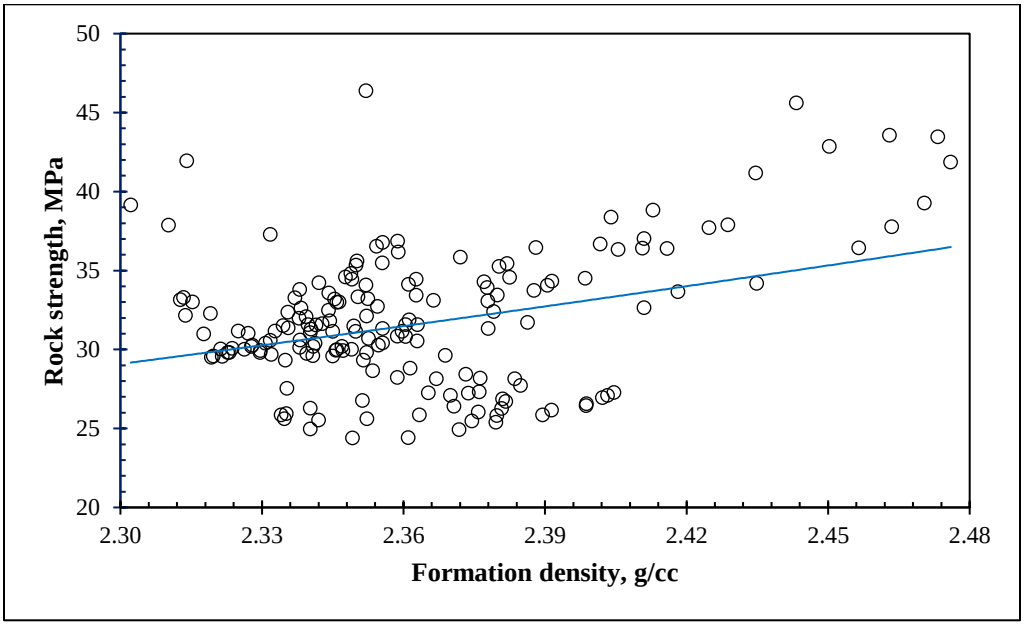

Figure 5.12: Relationship between in-situ rock strength and bulk density.

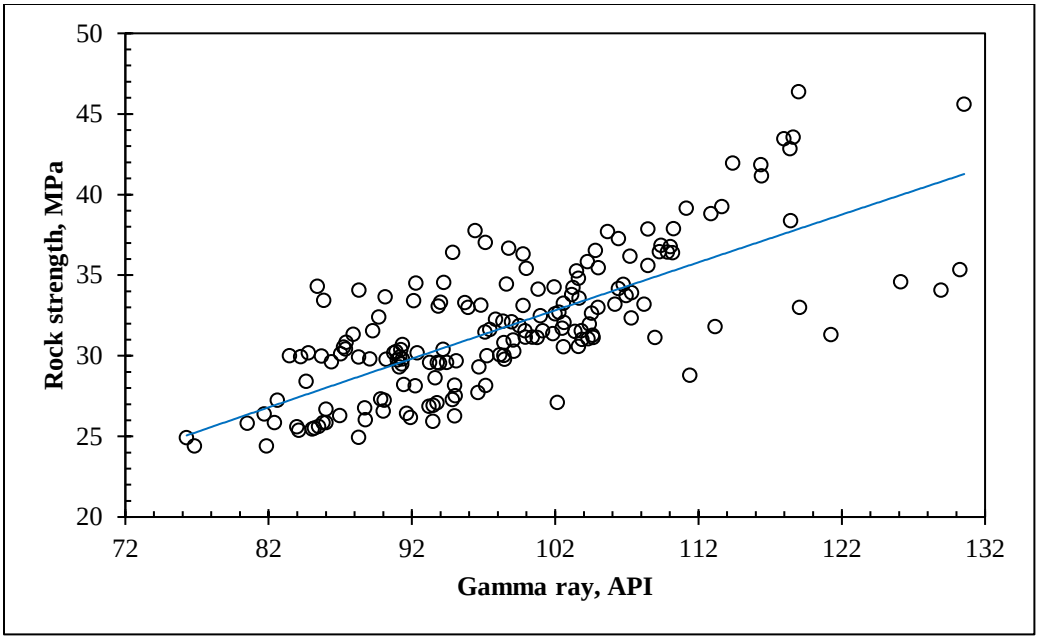

Figure 5.13: Relationship between in-situ rock strength and gamma ray.

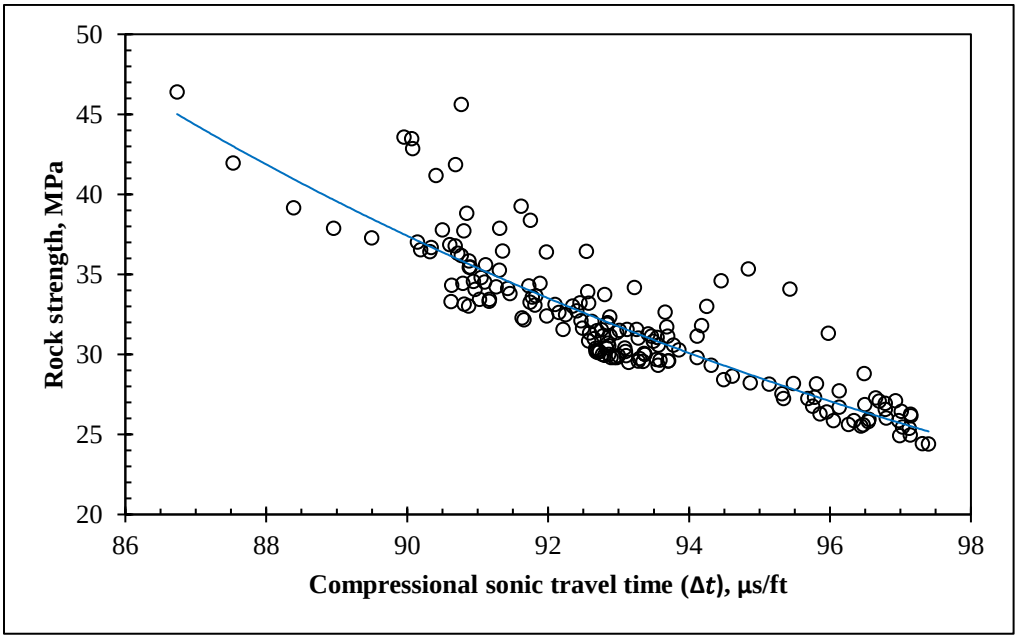

Figure 5.14: In-situ rock strength versus acoustic travel time.

To capture the lithology indicators, degree of formation density and touchstone of compaction, the formation gamma-ray (GR), bulk density (ρ_b), and sonic transit time (Δt) are considered to develop the new correlations for predicting the UCS of clastic rocks such as clean or shaly sand rock formation. The following two UCS models are proposed through linear regression analysis:

Model 1: $$UCS = 68.158 + 31.347\rho_b + 0.156GR - 1.349\Delta t \quad (5.25)$$

Model 2: $$UCS = 0.149\ GR^{0.93} + 6.67 * 10^{10}\Delta t^{-5} + 0.75{\rho_b}^{3.07} \quad (5.26)$$

The performance of proposed models is compared with that of other models in Figure 5.15.

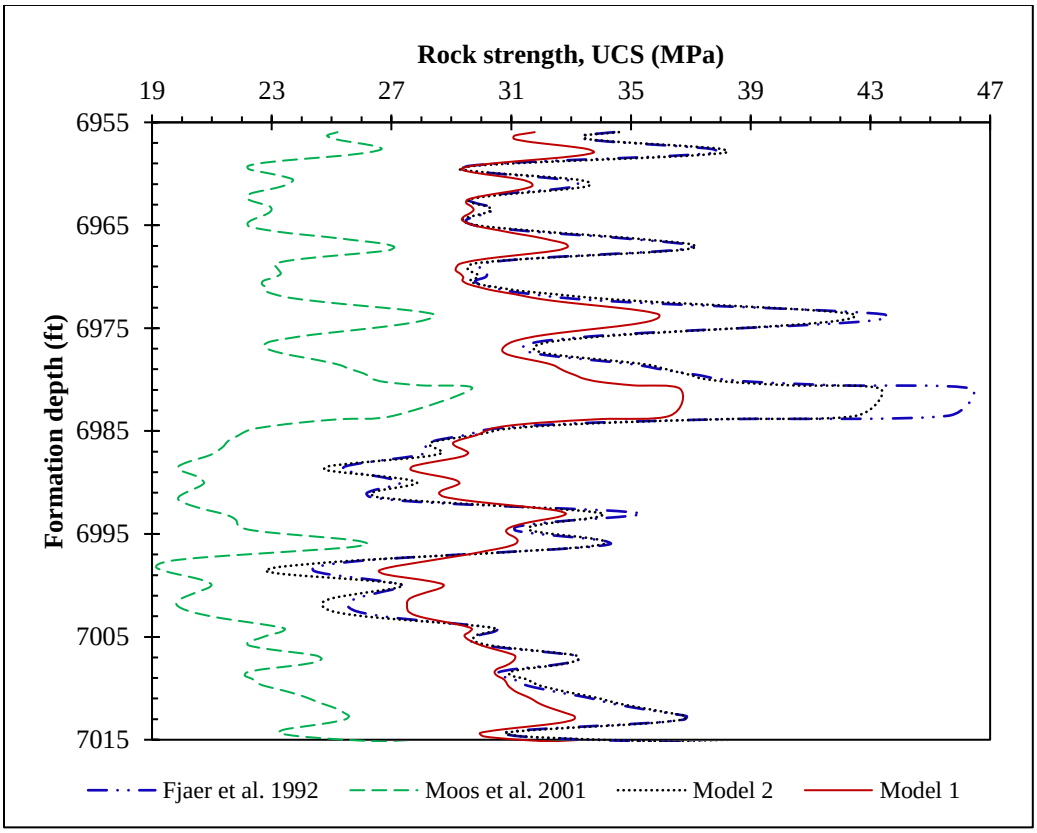

Figure 5.15: A comparison of UCS profile with different models.

To develop the above models, 175 field log data samples are collected from the shaly sand clastic sedimentary reservoir rock while it has an average UCS of 31.85 MPa, GR (API) of 98.39, ρ_b of 2.36 g/cm^3, and Δt of 93.12 μs/ft in the depth interval of 6955.5-7015.34 ft. The proposed model has a lower prediction error (a RMSE of 2.593 and an MAPE of 21%), compared to Moos et al. (2001) model (a RMSE of 8.92 and MAPE of 40.63%), while determining the UCS.

To further validate the robustness of the new correlation for UCS, real data taken from Volve field in the Norwegian North Sea is used. The detailed geological information about the filed can be found in the literature (Brekke et al., 2001; Faleide et al., 2010). Due to the availability of well log data in the open sources, the Well no. 15/9-F-1-A is chosen to examine the validity of the proposed correlations. The most common lithologies of this well are carbonate rocks (limestone, 9091 -10538 ft), marlstones with the trace of limestone and sandstone (10538-11017 ft), claystone (11017–11250 ft), and predominantly sandstones with some claystone in the depth interval of 11250-11948 ft. The field log data of GR, RB, and DT is employed to estimate continuous UCS profile using proposed models and other correlations, namely Chang et al. (2006) and Moos et al. (2001) where the formation depth is in the range 11250 to 11850 ft of Volve field (Equinor, 2018). A total of 1823 log samples are used from the selected depth of formation with average values of 75.50 μs/ft, 130.52 API, and 2.48 g/cm^3 for DT, GR, and RB, respectively. The predicted profile of UCS at 200 ft is depicted in Figure 5.16.

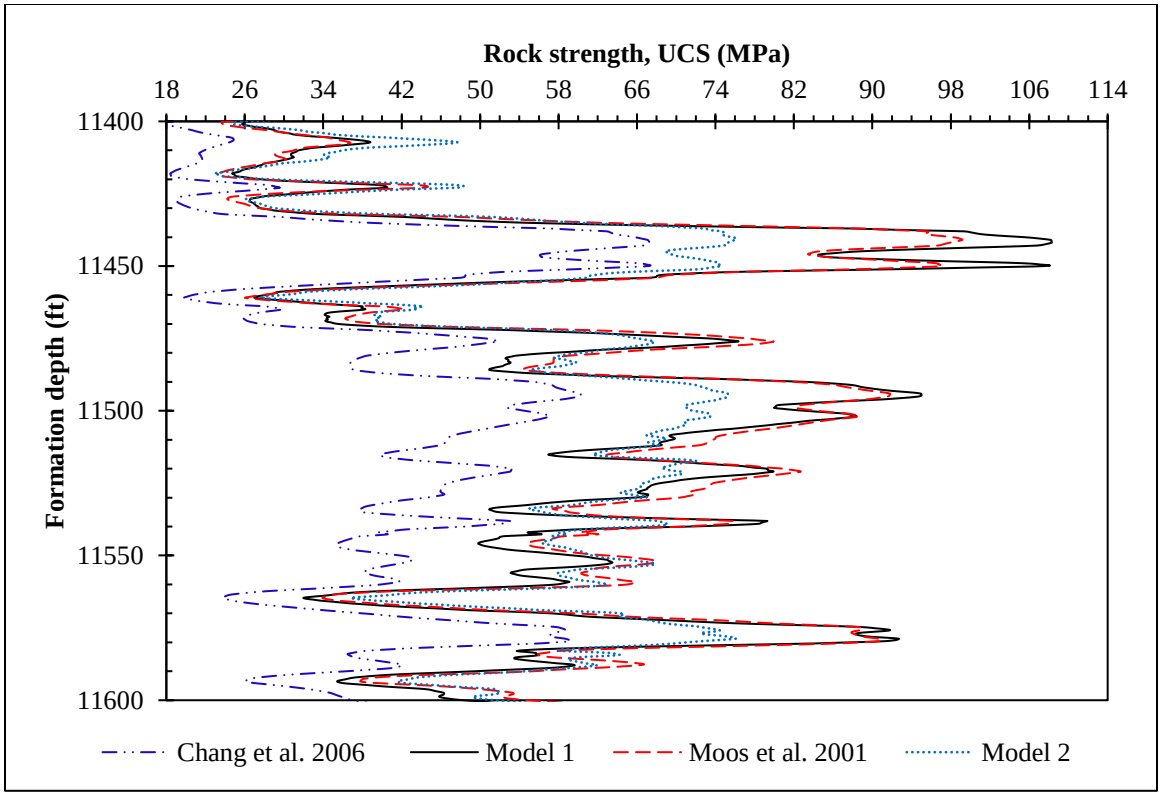

Figure 5.16: A comparison of UCS profile obtained from different models of in the Volve field.

Based on the UCS profile shown in Figure 5.16, model 2 predictions are closer to the results of Moos et al. (2001)'s model; while model 1 gives slightly lower values than the outputs based on Moos et al. (2001)'s model. Chang et al. (2006)'s model underestimates the in-situ rock strength. Using the North sea fields' log data, Chang et al. (2006) model exhibits higher error (an AAPE of 32.52%), compared to Moos et al (2001) model, and models 1 and 2 introduced in this study. Chang et al. (2006)'s model does not include the lithology feature (e.g., gamma-ray log) as well as formation density to capture the clay content and formation electron density effect in the model. The drilling engineers and/or geomechanists

can employ model 1 to obtain more reliable in-situ UCS profile for the shaly sand rock formations, compared to the model proposed by Moos et al. (2001). The new models take into account the most contributing log parameters to figure out the lithology effect, number of electron density, and acoustic travel time of the porous formation. Furthermore, the obtained UCS profile can be used to investigate the wellbore stability or rock failure criterion, and drilling performance analysis, reducing the exploration costs during the field development phase.

5.4 Conclusions

In this study, ANN and LS-SVM techniques are employed to predict the continuous profile of in-situ rock strength (UCS) of clastic sedimentary rocks using several field log data such as true resistivity, gamma-ray, bulk density, porosity, and compressional sonic travel time. The statistical indicators such as AAPE, MAPE, RMSE, and coefficient of determination (R) are used to evaluate the AI-based connectionist model performance. The key findings of this study are listed as follows:

- The connectionist models based on ANN with Levenberg-Marquardt training algorithm and the LS-SVM with CSA optimization strategy are capable of accurately estimating the reservoir rock strength (UCS) using log data.
- The compressional sonic travel time is the most influential parameter for determination of continuous in-situ rock strength profile of siliciclastic rocks.
- The formation acoustic travel time, gamma-ray, and the bulk density are essential to attain an accurate continuous formation UCS profile to capture lithology indicator,

dense minerals, number of electron density, and acoustic travel time of the underground formation.

- A correlation (model) is developed to estimate UCS through multivariate regression analysis by incorporating three influential log parameters. Similar to existing correlations, the introduced model exhibits a good performance.
- It is proven that newly developed log-based correlation can be used to predict actual in-situ unconfined rock strength for clastic sedimentary rocks including shaly sand rock formations.
- The deterministic tools, log variable ranking approach, and developed correlations can be useful for field specialists, researchers, and rock engineers while dealing with rock failure analysis, geomechanics, drilling optimization as well as formation evaluation.

Acknowledgements

The authors would like to thank Equinor (formerly Statoil) Canada Ltd., Natural Sciences and Engineering Research Council of Canada (NSERC) and InnovateNL for providing financial support to accomplish this study at the Memorial University, St. John's, NL, Canada.

Nomenclatures

Acronyms	
AAPE	Average Absolute Percentage Error
AI	Artificial Intelligence
ANN	Artificial Neural Network

BTS	Brazilian Tensile Strength
DT	Sonic Travel Time (µs/ft**)**
FL	Fuzzy Logic
GEP	Gene Expression Programming
GR	Gamma-ray (API)
LS-SVM	Least Square Support Vector Machine
LM	Levenberg-Marquardt
MAPE	Maximum Absolute Error Percentage
MSE	Mean Square Error
ML	Machine Learning
MLP	Multilayer Perception
MVRE	Multivariate Regression Analysis
RBF	Radial Basis Kernel Function
RMSE	Root Mean Square Error
UCS	Unconfined Compressive Strength (MPa), Rock

Variables/Parameters

b	Bias
dd	Dry density
E	Young's modulus (MPa)
G	Shear modulus (MPa)
GR_{log}	Gamma-ray value of the zone of interest
GR_{max}	Maximum value of gamma-ray log over the entire
GR_{min}	Minimum value of gamma-ray log over the entire
I_d	Slake durability index
$Is_{(50)}$	Point load index test
I_{GR}	Shale Index (Clay index)
K	Bulk modulus (MPa)
N	Number of neurons
NPHI	Neutron Porosity
PHI	Effective Porosity
PHIN	Neutron Porosity (frac.)
PHIND	Porosity from the combination of density and
R	Correlation coefficient
RB	Formation Bulk Density
RT	True (Deep) Resistivity (ohm-m), Rt
SRn	Schmidt hammer rebound number harness number
Vp	Compressional wave velocity (km/s)
Vs	Shear wave velocity (km/s)
Vsh	Shale volume (shaliness)
wc	Water content

x_i	Input variables
y_p	Predicted value (y)
y_m	Target (actual) output variable (y)

Greek Letters

ϕ	True porosity (frac.)
ϕ_e	Effective porosity (frac.)
$\phi_{D,e}$	Effective density porosity (frac.)
$\phi_{N,e}$	Effective neutron porosity (frac.)
$\phi_{N,sh}$	Neutron porosity of the adjacent shale zone (frac.)
ν	Poison's ratio
ρ_b	Bulk density (g/cm^3)
$\rho_{b.c}$	Clay corrected density porosity (g/cm^3)
ρ_{fl}	Fluid density (g/cm^3)
ρ_{ma}	Matrix density (g/cm^3)
ω_{ij}	Connected weight between the *i*-th neuron and *j*-th

References

Abdi, Y., Garavand, A. T., & Sahamieh, R. Z. (2018). Prediction of strength parameters of sedimentary rocks using artificial neural networks and regression analysis. Arabian Journal of Geosciences, 11(19), 587.

Al-Bulushi, N., King, P. R., Blunt, M. J., & Kraaijveld, M. (2009). Development of artificial neural network models for predicting water saturation and fluid distribution. Journal of Petroleum Science and Engineering, 68(3-4), 197-208.

Ali J.K., (1994). Neural network: a new tool for petroleum industry. Proceedings of SPE European petroleum computer conference, UK, SPE Paper 27561.

Anemangely, M., Ramezanzadeh, A., & Tokhmechi, B. (2017). Shear wave travel time estimation from petrophysical logs using ANFIS-PSO algorithm: A case study from Ab-Teymour Oilfield. Journal of Natural Gas Science and Engineering, 38, 373-387.

Anemangely, M., Ramezanzadeh, A., Amiri, H., & Hoseinpour, S. A. (2019). Machine learning technique for the prediction of shear wave velocity using petrophysical logs. Journal of Petroleum Science and Engineering, 174, 306-327.

Ashena, R., & Thonhauser, G. (2015). Application of Artificial Neural Networks in Geoscience and Petroleum Industry. In Artificial Intelligent Approaches in Petroleum Geosciences (pp. 127-166). Springer, Cham.

Asoodeh, M., & Bagheripour, P. (2012). Prediction of compressional, shear, and stoneley wave velocities from conventional well log data using a committee machine with intelligent systems. Rock mechanics and rock engineering, 45(1), 45-63.

Asquith, G., & Krygowski, D., (2004). Basic well log analysis, American Association of Petroleum Geologists, Second Edition, Tulsa, Oklahoma, p. 216.

ASTM, A. (1986). Standard test method of unconfined compressive strength of intact rock core specimens. ASTM Publication.

Bailey, T., & Dutton, D. (2012). An Empirical Vp/Vs Shale Trend for the Kimmeridge Clay of the Central North Sea. In 74th EAGE Conference and Exhibition incorporating EUROPEC 2012.

Barzegar, R., Sattarpour, M., Nikudel, M. R., & Moghaddam, A. A. (2016). Comparative evaluation of artificial intelligence models for prediction of uniaxial compressive

strength of travertine rocks, case study: Azarshahr area, NW Iran. Modeling Earth Systems and Environment, 2(2), 76.

Behnia, D., Behnia, M., Shahriar, K., & Goshtasbi, K. (2017). A new predictive model for rock strength parameters utilizing GEP method. Procedia Engineering, 191, 591-599.

Bradford, I. D. R., Fuller, J., Thompson, P. J., & Walsgrove, T. R. (1998). Benefits of assessing the solids production risk in a North Sea reservoir using elastoplastic modelling. In SPE/ISRM Rock Mechanics in Petroleum Engineering. Society of Petroleum Engineers.

Brekke, H., Sjulstad, H. I., Magnus, C., & Williams, R. W. (2001). Sedimentary environments offshore Norway—an overview. In Norwegian Petroleum Society Special Publications (Vol. 10, pp. 7-37). Elsevier.

Broch, E., & Franklin, J. A. (1972). The point-load strength test. In International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts (Vol. 9, No. 6, pp. 669-676). Pergamon.

Brocher, T. M. (2005). Empirical relations between elastic wave speeds and density in the Earth's crust. Bulletin of the seismological Society of America, 95(6), 2081-2092.

Castagna, J. P., Batzle, M. L., & Eastwood, R. L. (1985). Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks. Geophysics, 50(4), 571-581.

Ceryan, N. (2014). Application of support vector machines and relevance vector machines in predicting uniaxial compressive strength of volcanic rocks. Journal of African Earth Sciences, 100, 634–644.

Ceryan, N., & Can, N. K. (2018). Prediction of The Uniaxial Compressive Strength of Rocks Materials. In Handbook of Research on Trends and Digital Advances in Engineering Geology (pp. 31-96). IGI Global.

Ceryan, N., Okkan, U., & Kesimal, A. (2013). Prediction of unconfined compressive strength of carbonate rocks using artificial neural networks. Environmental Earth Sciences, 68(3), 807–819. doi:10.1007/ s12665-012-1783-z

Chang, C., Zoback, M. D., & Khaksar, A. (2006). Empirical relations between rock strength and physical properties in sedimentary rocks. Journal of Petroleum Science Engineering, 51(3-4), 223–237. doi:10.1016/j.petrol.2006.01.003

Crawford, B., Alramahi, B., Gaillot, P., Sanz, P., & DeDontney, N. (2011). Mechanical rock properties prediction: Deriving rock strength and compressibility from petrophysical properties 12th ISRM Congress.

Dehghan, S., Sattari, G. H., Chelgani, S. C., & Aliabadi, M. A. (2010). Prediction of uniaxial compressive strength and modulus of elasticity for Travertine samples using regression and artificial neural networks. Mining Science and Technology (China), 20(1), 41-46.

Demirdag, S., Tufekci, K., Kayacan, R., Yavuz, H., & Altindag, R. (2010). Dynamic mechanical behavior of some carbonate rocks. International Journal of Rock Mechanics and Mining Sciences, 47(2), 307-312.

Edimann, K., Somerville, J. M., Smart, B. G. D., Hamilton, S. A., & Crawford, B. R. (1998). Predicting rock mechanical properties from wireline porosities. In SPE/ISRM Rock Mechanics in Petroleum Engineering. Society of Petroleum Engineers.

Equinor (2018). Volve Data Village, URL https://data.equinor.com/dataset/Volve

Esfahani, S., Baselizadeh, S., & Hemmati-Sarapardeh, A. (2015). On determination of natural gas density: least square support vector machine modeling approach. Journal of Natural Gas Science and Engineering, 22, 348-358.

Faleide, J. I., Bjørlykke, K., & Gabrielsen, R. H. (2010). Geology of the Norwegian continental shelf. In Petroleum Geoscience (pp. 467-499). Springer, Berlin, Heidelberg.

Farquhar, R. A., Somerville, J. M., & Smart, B. G. D. (1994). Porosity as a geomechanical indicator: an application of core and log data and rock mechanics. In European Petroleum Conference. Society of Petroleum Engineers.

Fener, M., Kahraman, S., Bilgil, A., & Gunaydin, O. (2005). A comparative evaluation of indirect methods to estimate the compressive strength of rocks. Rock Mechanics and Rock Engineering, 38(4), 329-343.

Fjær, E., Holt, R. M., Raaen, A. M., Risnes, R., & Horsrud, P. (2008). Petroleum related rock mechanics (Vol. 53). Elsevier.

Fjær, E., Holt, R.M., Horsrud, P., Raaen, A.M., and Risnes, R. (1992). Petroleum Related Rock Mechanics, 1st ed., Elsevier, Amsterdam, 346p.

Freyburg, E. (1972). Der Untere und mittlere Buntsandstein SW-Thuringen in seinen gesteinstechnicschen Eigenschaften. Deustche Gesellschaft Geologische Wissenschaften. A; Berlin, 176, 911-919.

Gaviglio, P. (1989). Longitudinal waves propagation in a limestone: the relationship between velocity and density. Rock Mechanics and Rock Engineering, 22(4), 299-306.

Greenberg, M. L., & Castagna, J. P. (1992). Shear-Wave Velocity Estimation In Porous Rocks: Theoretical Formulation, Preliminary Verification And Applications 1. Geophysical prospecting, 40(2), 195-209.

Haftani, M., Bohloli, B., Nouri, A., Javan, M. R. M., Moosavi, M., & Moradi, M. (2015). Influence of penetration rate and indenter diameter in strength measurement by indentation testing on small rock specimens. Rock Mechanics and Rock Engineering, 48(2), 527-534.

Han, D. H., Nur, A., & Morgan, D. (1986). Effect of porosity and clay content on wave velocity in sandstones. *Geophysics*, *51*(11), 2093-2107.

Hoek, E., & Brown, E. T. (1997). Practical estimates of rock mass strength. International journal of rock mechanics and mining sciences, 34(8), 1165-1186.

Hossain, Z., Mukerji, T., & Fabricius, I. L. (2012). Vp-Vs relationship and amplitude variation with offset modeling of glauconitic greensand. Geophysical Prospecting, 60(1), 117-137.

ISRM (1981). Rock characterisation testing and monitoring. In: Brown, E. T. (ed.) Pergamon Press, Oxford.

Jamshidi, A., Nikudel, M. R., Khamehchiyan, M., & Sahamieh, R. Z. (2016). The effect of specimen diameter size on uniaxial compressive strength, P-wave velocity and the correlation between them. Geomechanics and Geoengineering, 11(1), 13-19.

Jamshidi, A., Zamanian, H., & Sahamieh, R. Z. (2018). The effect of density and porosity on the correlation between uniaxial compressive strength and P-wave velocity. Rock Mechanics and Rock Engineering, 51(4), 1279-1286.

Khandelwal, M., & Monjezi, M. (2013). Prediction of backbreak in open-pit blasting operations using the machine learning method. Rock mechanics and rock engineering, 46(2), 389-396.

Kong, F., & Shang, J. (2018). A validation study for the estimation of uniaxial compressive strength based on index tests. Rock Mechanics and Rock Engineering, 51(7), 2289-2297.

Koolivand-Salooki, M., Esfandyari, M., Rabbani, E., Koulivand, M., & Azarmehr, A. (2017). Application of Genetic Programing technique for predicting Uniaxial Compressive Strength using reservoir formation properties. Journal of Petroleum Science and Engineering, 159, 35-48.

Krishna, K. S., Rao, D. G., Murty, G. P. S., & Ramana, Y. V. (1989). Sound velocity, density, and related properties along a transect across the Bay of Bengal. Geo-marine letters, 9(2), 95-102.

Larionov, V.V. (1969). Radiometry of boreholes (in Russian). Nedra, Moscow.

Lee, M. W. (2013). Comparison of Methods for Predicting Shear-Wave Velocities of Unconsolidated Shallow Sediments in the Gulf of Mexico. US Department of the Interior, US Geological Survey.

Majdi, A., & Rezaei, M. (2013). Prediction of unconfined compressive strength of rock surrounding a roadway using artificial neural network. Neural Computing and Applications, 23(2), 381-389.

Matin, S. S., Farahzadi, L., Makaremi, S., Chelgani, S. C., & Sattari, G. (2018). Variable selection and prediction of uniaxial compressive strength and modulus of elasticity by random forest. Applied Soft Computing, 70, 980-987.

McNally, G. H. (1987). Estimation of coal measures rock strength using sonic and neutron logs. Geoexploration, 24(4-5), 381-395.

Meulenkamp, F., & Grima, M. A. (1999). Application of neural networks for the prediction of the unconfined compressive strength (UCS) from Equotip hardness. International Journal of rock mechanics and mining sciences, 36(1), 29-39.

Miah, M. I. (2014). Porosity assessment of gas reservoir using wireline log data: a case study of bokabil formation, Bangladesh. Procedia Engineering, 90, 663-668.

Miah, M.I., Ahmed, S. & Zendehboudi, S. (2019). Connectionist and mutual information tools to determine water saturation and rank input log variables, Journal of Petroleum Science and Engineering, PETROL 106741.

Miller, R. P. (1965). Engineering classification and index properties for intact rock. PhD Thesis, University of Illinois.

Miller, S. L., & Stewart, R. R. (1990). Effects of lithology, porosity and shaliness on P-and S-wave velocities from sonic logs. Canadian Journal of Exploration Geophysics, 26(1-2), 94-103.

Mohaghegh, S., Arefi, R., Ameri, S., Aminiand, K., & Nutter, R. (1996). Petroleum reservoir characterization with the aid of artificial neural networks. Journal of Petroleum Science and Engineering, 16(4), 263-274.

Mohamad, E. T., Armaghani, D. J., Momeni, E., & Abad, S. V. A. N. K. (2015). Prediction of the unconfined compressive strength of soft rocks: a PSO-based ANN approach. Bulletin of Engineering Geology and the Environment, 74(3), 745-757.

Momeni, E., Jahed Armaghani, D., Hajihassani, M., & Amin, M. F. M. (2015). Prediction of uniaxial compressive strength of rock samples using hybrid particle swarm optimization-based artificial neural networks. Measurement, 60, 50–63. doi:10.1016/j.measurement.2014.09.075

Mondol, N. H. (2015). Well logging: Principles, applications and uncertainties. In Petroleum Geoscience (pp. 385-425). Springer, Berlin, Heidelberg.

Moos, D., Zoback, M. D., & Bailey, L. (2001). Feasibility study of the stability of openhole multilaterals, Cook Inlet, Alaska. SPE Drilling & Completion, 16(03), 140-145.

Mousavi, E., Cheshomi, A., & Ashtari, M. (2018). Estimating elasticity modulus and uniaxial compressive strength of sandstone using indentation test. Journal of Petroleum Science and Engineering, 169, 157-166.

Nabaei, M., & Shahbazi, K. (2012). A new approach for predrilling the unconfined rock compressive strength prediction. Petroleum science and technology, 30(4), 350-359.

Najibi, A. R., Ghafoori, M., Lashkaripour, G. R., & Asef, M. R. (2015). Empirical relations between strength and static and dynamic elastic properties of Asmari and Sarvak limestones, two main oil reservoirs in Iran. Journal of Petroleum Science and Engineering, 126, 78-82.

Negara, A., Ali, S., AlDhamen, A., Kesserwan, H., & Jin, G. (2017). Unconfined Compressive Strength Prediction from Petrophysical Properties and Elemental Spectroscopy Using Support-Vector Regression. In SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition. Society of Petroleum Engineers.

Nouri, A., Vaziri, H., Kuru, E., & Islam, R. (2006). A comparison of two sanding criteria in physical and numerical modeling of sand production. Journal of petroleum science and engineering, 50(1), 55-70.

Ocak, I., & Seker, S. E. (2012). Estimation of elastic modulus of intact rocks by artificial neural network. Rock Mechanics and Rock Engineering, 45(6), 1047-1054.

Odunlami, T., Soroush, H., Kalathingal, P., & Somerville, J. (2011). Log-Based Rock Property Evaluation-A New Capability in a Specialized Log Data Management Platform. In SPE/DGS Saudi Arabia Section Technical Symposium and Exhibition. Society of Petroleum Engineers.

Ojha, M., & Sain, K. (2014). Velocity-porosity and velocity-density relationship for shallow sediments in the Kerala-Konkan basin of western Indian margin. Journal of the Geological Society of India, 84(2), 187-191.

Oloruntobi, O., & Butt, S. (2019). The Shear-wave Velocity Prediction for Sedimentary Rocks. Journal of Natural Gas Science and Engineering, 103084.

Onalo, D. O. (2019). Dynamic data driven investigation of petrophysical and geomechanical properties for reservoir formation evaluation (Doctoral dissertation, Memorial University of Newfoundland).

Onalo, D., Adedigba, S., Khan, F., James, L. A., & Butt, S. (2018). Data driven model for sonic well log prediction. Journal of Petroleum Science and Engineering, 170, 1022-1037.

Onyia, E. C. (1988). Relationships between formation strength, drilling strength, and electric log properties. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Pickett, G. R. (1963). Acoustic character logs and their applications in formation evaluation. Journal of Petroleum Technology, 15(06), 659-667.

Raaen, A. M., Hovem, K. A., Joranson, H., & Fjaer, E. (1996). FORMEL: A step forward in strength logging. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Rabbani, E., Sharif, F., Salooki, M. K., & Moradzadeh, A. (2012). Application of neural network technique for prediction of uniaxial compressive strength using reservoir formation properties. International Journal of Rock Mechanics and Mining Sciences, (56), 100-111.

Rajabzadeh, M. A., Moosavinasab, Z., & Rakhshandehroo, G. (2012). Effects of rock classes and porosity on the relation between uniaxial compressive strength and some rock properties for carbonate rocks. Rock mechanics and rock engineering, 45(1), 113-122.

Ramcharitar, K., & Hosein, R. (2016). Rock Mechanical Properties of Shallow Unconsolidated Sandstone Formations. In SPE Trinidad and Tobago Section Energy Resources Conference. Society of Petroleum Engineers.

Rasouli, V., Pallikathekathil, Z. J., & Mawuli, E. (2011). The influence of perturbed stresses near faults on drilling strategy: a case study in Blacktip field, North Australia. Journal of Petroleum Science and Engineering, 76(1-2), 37-50.

Rastegarnia, A., Teshnizi, E. S., Hosseini, S., Shamsi, H., & Etemadifar, M. (2018). Estimation of punch strength index and static properties of sedimentary rocks using neural networks in south west of Iran. Measurement, 128, 464-478.

Razavi, S., & Tolson, B.A. (2011). A New Formulation for Feedforward Neural Networks. IEEE Transactions on Neural Networks 22.10 (2011): 1588-598.

Rostami, S., Rashidi, F., & Safari, H. (2019). Prediction of oil-water relative permeability in sandstone and carbonate reservoir rocks using the CSA-LSSVM algorithm. Journal of Petroleum Science and Engineering, 173, 170-186.

Sarda, J. P., Kessler, N., Wicquart, E., Hannaford, K., & Deflandre, J. P. (1993). Use of porosity as a strength indicator for sand production evaluation. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Schlumberger (1998). Log Interpretation Principles/Applications, 7th printing, Houston, p. 235.

Sebtosheikh, M. A., Motafakkerfard, R., Riahi, M. A., Moradi, S., & Sabety, N. (2015). Support vector machine method, a new technique for lithology prediction in an

Iranian heterogeneous carbonate reservoir using petrophysical well logs. Carbonates and Evaporites, 30(1), 59-68.

Sharma, M. R., O'Regan, M., Baxter, C. D. P., Moran, K., Vaziri, H., & Narayanasamy, R. (2010). Empirical relationship between strength and geophysical properties for weakly cemented formations. Journal of Petroleum Science and Engineering, 72(1-2), 134-142.

Singh, T. N., Kainthola, A., & Venkatesh, A. (2012). Correlation between point load index and uniaxial compressive strength for different rock types. Rock Mechanics and Rock Engineering, 45(2), 259-264.

Smola, A.J., Sch, B., & lkopf (2004). A tutorial on support vector regression. Stat. Comput. 14, 199–222.

Sonmez, H., Tuncay, E., & Gokceoglu, C. (2004). Models to predict the uniaxial compressive strength and the modulus of elasticity for Ankara Agglomerate. International Journal of Rock Mechanics and Mining Sciences, 41(5), 717-729.

Suykens, J. & Vandewalle, J. (1999). Neural Processing Letters 9, 293. https://doi.org/10.1023/A:1018628609742.

Suykens, J.A.K., Van Gestel, T., Brabanter, J., De Moor, B., & Vandewalle, J. (2002). Least Squares Support Vector Machines. World Scientific, Singapore.

Taheri-Garavand, A., Ahmadi, H., Omid, M., Mohtasebi, S. S., Mollazade, K., Smith, A. J. R., & Carlomagno, G. M. (2015). An intelligent approach for cooling radiator fault

diagnosis based on infrared thermal image processing technique. Applied Thermal Engineering, 87, 434-443.

Tariq, Z., Elkatatny, S., Mahmoud, M., Ali, A. Z., & Abdulraheem, A. (2017). A new technique to develop rock strength correlation using artificial intelligence tools. In SPE Reservoir Characterisation and Simulation Conference and Exhibition. Society of Petroleum Engineers.

Torabi-Kaveh, M., Naseri, F., Saneie, S., & Sarshari, B. (2015). Application of artificial neural networks and multivariate statistics to predict UCS and E using physical properties of Asmari limestones. Arabian Journal of Geosciences, 8(5), 2889–2897. doi:10.1007/s12517-014-1331-0

Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer, New York.

Vernik, L., Bruno, M., & Bovberg, C. (1993). Empirical relations between compressive strength and porosity of siliciclastic rocks. In International journal of rock mechanics and mining sciences & geomechanics abstracts (Vol. 30, No. 7, pp. 677-680). Pergamon.

Weingarten, J. S., & Perkins, T. K. (1995). Prediction of sand production in gas wells: methods and Gulf of Mexico case studies. Journal of Petroleum Technology, 47(07), 596-600.

Williams, D. M. (1990). The acoustic log hydrocarbon indicator. In SPWLA 31st Annual Logging Symposium. Society of Petrophysicists and Well-Log Analysts.

Xavier-de-Souza, S., Suykens, J. A., Vandewalle, J., & Bollé, D. (2009). Coupled simulated annealing. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 40(2), 320-335.

Yagiz, S., Sezer, E. A., & Gokceoglu, C. (2012). Artificial neural networks and nonlinear regression techniques to assess the influence of slake durability cycles on the prediction of uniaxial compressive strength and modulus of elasticity for carbonate rocks. International Journal for Numerical and Analytical Methods in Geomechanics, 36(14), 1636-1650.

Yang, Y., & Zhang, Q. (1997). A hierarchical analysis for rock engineering using artificial neural networks. Rock Mechanics and Rock Engineering, 30(4), 207-222.

Yılmaz, I., & Yuksek, A. G. (2008). An example of artificial neural network (ANN) application for indirect estimation of rock parameters. Rock Mechanics and Rock Engineering, 41(5), 781-795.

Yilmaz, I., & Yuksek, G. (2009). Prediction of the strength and elastic modulus of gypsum using multiple regression, ANN, and ANFIS models. International Journal of Rock Mechanics and Mining Sciences, 46(4), 803–810.

Yılmaz, N. G. (2013). The influence of testing procedures on uniaxial compressive strength prediction of carbonate rocks from Equotip hardness tester (EHT) and proposal of a new testing methodology: Hybrid dynamic hardness (HDH). Rock mechanics and rock engineering, 46(1), 95-106.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

This section includes the main conclusions drawn from this research study as well as the recommendations for further investigations.

6.1 Conclusions

This research study investigates the application of two widely used machine learning tools, namely, the artificial neural networks (ANNs) and the support vector machines (SVM) for reservoir characterization using log data. The study further explores the application of the concept of mutual information to identify and rank log variables for the purpose of developing predictive models for reservoir fluid and rock properties. The dataset includes real field well logs from a clastic rock formation. The research strategies are established to address the knowledge gap and incorporate the in-situ reservoir time-series sequential log data in the deterministic tools. This thesis implements conventional and hybridized data-driven models to predict water saturation and in-situ rock strength in reservoir characterization. The introduced approaches provide less time consuming, cost-effective, and continuous predictions of the formation properties along the wellbore of the reservoir rock formations. These predictive models serve as efficient tools to facilitate reservoir characterization for strategic exploration decisions. Based on the research outcomes, the following conclusions are made:

- The data-driven model development techniques are illustrated systematically to investigate the dependency of water saturation on the predictor variables using mutual

information (MI). The MI approach reveals that the most contributing log variables are the true resistivity and the bulk density, while the gamma-ray and neutron porosity have minor effects on water saturation. The single hidden-layer perception-based ANN technique also shows that the true resistivity and the formation bulk density logs are the primary and major predictor variables for the successful development of the fluid saturation model.

- It is also found that the ranking orders of the variables with both multilayer perception (MLP)-based ANN and LS-SVM techniques are the same.
- The Levenberg Marquardt training algorithm-based MLP-ANN water saturation model shows superior performance with a lower statistical error and higher correlation coefficient (R), compared to the other algorithms such as the Bayesian regularization and the scaled conjugated gradient.
- It is concluded that the radial basis kernel (Gaussian) function-based LS-SVM connectionist model has a greater reliability and prediction capability, compared to the other kernel function-based models for predicting water saturation.
- Based on the parametric sensitivity analysis with the MLP-ANN and the LS-SVM predictive models, true resistivity and the bulk density are found to be the most important predictor variables, while gamma-ray and sonic travel time have the least contribution to the water saturation.
- The acoustic travel time, the gamma-ray, and the formation bulk density are the most important predictor variables for continuous in-situ rock strength (UCS) profile of shaly sand reservoirs.

- The compressional sonic travel time is the most influential parameter to estimate UCS, compared to other log variables in the connectionist MLP-based ANN approach; there is also a good agreement between LS-SVM and MLP-ANN methods in terms of prediction performance.
- Two new correlations for prediction of UCS have been presented which are able to provide an accurate continuous formation UCS profile with capturing lithology indicator, dense minerals, number of electron density, and acoustic travel time of the porous formation.
- The smart connectionist models such as coupled simulated annealing-based LS-SVM and ANN are more reliable and robust (compared to other tools), and are capable of estimating continuous profile of reservoir properties with high accuracy using well logs.

6.2 Recommendations for Future Studies

This section outlines some recommendations that can address in future works while dealing with the data-driven connectionist approaches in reservoir characterization.

- The hybrid computational models such as PSO-based ANN, convolution neural network (CNN), recurrent neural network (RNN), neuro-fuzzy inference system (ANFIS) and gene expression programming (GEP) can be used to predict water saturation and rock strength as well as to find the most influential parameters in the model.

- The log variables of true resistivity and formation bulk density can be considered to develop suitable correlations for predicting water saturation using wireline log data.
- The machine learning (ML) tools requires less time and computational complexities, leading to lower capital and operating costs in the exploration stage; these approaches can be used to generate the synthetic wireline logs such as resistivity and sonic logs using available other petrophysical well logs.
- The information theoretic measures approach (i.e. entropy and mutual information) can be used to investigate the dependency between the predictor and output variables for other reservoir properties (i.e. permeability and porosity), and geo-mechanical properties (i.e. elastic constant, and rock strength parameters) of rock formations.
- The proposed ML research strategies can be implemented to determine formation properties (such as porosity, permeability, sonic acoustic velocities, and formation strength index); the feature ranking of predictor variables in the data-driven models can be also studied.
- Geophysical log data can be used for different aspects of reservoir analysis. The soft computing approaches can be adopted to generate the synthetic well logs such as resistivity, density, and sonic logs using different combinations of input and output log data for costs reduction and time-saving for further field development and formation evaluation.
- The ML research strategies and log variable ranking approaches in the data-driven model can be also used by field specialists, researchers, and petroleum engineers where

reservoir fluid properties prediction, rock failure analysis, geomechanics and drilling optimization as well as formation evaluation are targeted.

APPENDIX

Appendix A: Modeling of Temperature Distribution and Oil Displacement during Thermal Recovery in Porous Media: A Critical Review

Preface

A version of this chapter has been published in the "Fuel" Journal, Vol. 226 (2018). I am the primary author of this article and was the originator of the study concept. The co-author Dr. Enamul Hossain provided technical assistance to find the research scopes, as well as identifying the knowledge gap. The first draft of the manuscript was prepared by myself. The co-author Murtada Elhaj helped to present the figures and references in the manuscript. Co-author Dr. Salim Ahmed also reviewed and provided valuable insights on how to improve the manuscript. I have revised the manuscript based on the feedback from the co-authors. Finally, all authors contributed to prepare the final version of the manuscript as per the reviewers' and editor's comments.

Abstract

Thermal flooding is one of the most successful and widely used processes for heavy oil recovery. The memory-based fluid flow model is effective in characterizing reservoir heat transport mechanism, temperature profile, and predicting the performance of thermal recovery. Temperature has a substantial effect on the thermodynamic properties such as thermal conductivity of the formation. In addition, the influence of temperature on reservoir rock and fluid properties plays an important role in accurately predicting reservoir temperature distribution, oil displacement, and steam oil ratio. This paper presents a critical review and analyses to provide inclusive information on the state-of-the-art memory-based fluid flow modeling during the thermal displacement process. The review highlights the assumptions and limitations of the current models in the areas of thermal conductivity, temperature distribution, oil displacement, and steam oil ratio during the thermal flooding process in porous media. This paper also serves to provide an insight into future research opportunities to fill the knowledge gaps in the subject area by applying the memory concept and further improvement of the current and classical models for heavy oil recovery.

Keywords: Porous media; Heat transport; Thermal conductivity; Heavy oil; Thermal flooding; Memory concept.

A1 Introduction

A1.1 Background and motivation

The current worldwide petroleum industry is facing a great challenge to produce hydrocarbon from mature reservoirs. The produced reserves from new discoveries have provided steadily declining replacement rate over the last few decades. For meeting the global energy demands by producing more oil from mature hydrocarbon reservoirs, enhanced oil recovery (EOR) is the most popular technique. EOR is accomplished by increasing the oil recovery after the primary recovery such as natural drive mechanisms, and secondary recovery such as water flooding techniques. This can be accomplished by many techniques such as thermal recovery, chemical or gas injection, ultrasonic stimulation and microbial injection. These EOR methods allow recovery of oil that was not economically recoverable earlier using conventional techniques. EOR methods involve the injection of substances and/or energy into the oil reservoir to unlock trapped oil, improve sweep and enhance production rates. Among the tertiary oil recovery methods, thermal techniques (i.e. cyclic steam stimulation, steam flooding and in-situ combustion) aim to reduce oil viscosity to increase its mobility through the application of heat into the reservoir formation. Among the EOR processes, thermal injection is the most successful and extensively used process that is applicable to a variety of heavy oil as well as bitumen reservoirs (Green and Willhite, 1998). However, the achievement of a thermal flood is entirely dependent on understanding the mechanism of heat transfer within the reservoir and the complex interactions between the reservoir rock and fluid matrices, temperature-

sensitive rock and fluid parameters, which govern the evolution of the temperature profile during the thermal EOR processes (Hossain and Abu-khamsin, 2012).

A1.2 Fluid flow modeling

It is important to predict reservoir properties and rheology through the porous media during thermal flooding. The classical constitutive equation describing fluid flow in reservoir porous media is Darcy's law (Darcy, 1856). This equation is based on several assumptions such as single phase, isothermal laminar flow, no chemical reaction between rock and fluid, as well as constant permeability and viscosity (Darcy, 1856). When heat energy is introduced, thermal alterations of reservoir rock and fluid properties induce non-Darcy flow effects, which cannot be captured accurately by the Darcy equation. This type of modeling is a complicated research task in oil and gas engineering due to the complex nature of reservoir rock and fluid properties, and geological subsurface behavior. Rheology is the study of fluid flow and its deformation, which are related to the field of physics (Larson, 1999). The study of rheology focuses on materials that possess both viscous and elastic properties. Most underground reservoir fluids do not follow the Newtonian behavior (Hossain et al., 2008a). Heavy crude oil is a non-Newtonian fluid at low temperatures (Dong et al., 2013). Fluid rheology is an important issue for the prediction of rock-fluid interactions within a complex oil reservoir or any formation management. The subsurface structure and reservoir rock properties of a formation are not only dependent on the deposition, but also on the sedimentation process within the earth's interior with respect to geological age (Hossain et al., 2008b). This subsurface geological information is

significant to model the performance of the fluid movement within the pore network of complex reservoir formations. Consequently, research on fluid rheology and geological structure is essential for the development of real flow models in reservoir formations. In addition, determination of time-dependent rock and fluid properties is the most challenging task in the dynamic condition. The continuous alteration of rock and fluid properties can be characterized by the memory concept. This concept is defined as "the properties of rock and fluid that help to account for changes in rock properties (such as permeability and porosity) and fluid properties (such as pressure dependent fluid properties and viscosity) with time and space" (Hossain, 2016). In addition, a simple definition of memory concept is also proposed by Hossain (2016) as "the system can remember its previous state" (Hossain, 2016). The concept of memory has been incorporated into the various model equations to represent the alteration of rock and fluid properties within a reservoir porous medium. Literature shows that the memory-based fluid flow models can be utilized successfully to develop a more rigorous and realistic model for thermal displacement processes by using momentum and energy transport concepts (Obembe et al., 2016a).

A1.3 Thermal flooding

Thermal flooding is one of the most effective techniques for extracting heavy oil from conventional reservoirs. However, it faces challenges in the arctic and offshore regions. Proper modeling of thermal flooding is challenging in these regions, if thermal hysteresis is considered. This phenomenon was recognized experimentally in several disciplines, other than petroleum reservoir engineering (Krober et al., 1993). Mathematically,

numerous formulations have been proposed to analyze the difficulties associated with thermal flooding, each concentrating on one or more aspects of the process, including different assumptions (Krober et al., 1993). For this process, success with modeling is entirely dependent on understanding the heat transfer mechanism within the complex reservoir and interactions among all temperature-sensitive rock/fluid parameters that govern the evolution of the temperature profile (Hossain et al., 2008c; Hossain et al., 2009a). In a thermal flooding process, a more realistic rheological fluid flow model through porous media and accurate prediction of the temperature profile using the memory concept are key factors in the process design, production projects, and reservoir management (Hossain et al., 2015).

A1.4 The aim of this review and its novelty

The aim of this study is to revisit the development of fluid flow models with rock-fluid interactions during thermal EOR processes. This review highlights the present status of memory-based research on fluid flow modeling in complex reservoirs and the thermal flooding process. It also serves as a future research guideline towards the development of complex heat transfer models, temperature distribution profile prediction, and reservoir performance model during thermal oil recovery by application of the memory concept.

A2 Literature Review and Discussions

Reservoir rock (i.e. clastic or non-clastic sedimentary rock) is the medium of fluid transport in hydrocarbon reservoirs. Rock and fluid properties vary during any pressure disturbances

or thermal actions in the reservoir formation. Most of the reservoir rock and fluid properties are functions of pressure and temperature. In many conventional simulators, these properties are considered to vary only as a function of space. However, it is also necessary to include variations in rock properties (e.g. porosity, permeability etc.), and fluid properties (e.g. fluid saturation, viscosity, and PVT properties) by incorporating time function for applications including geothermal activities, chemical reactions and other geological activities in the sub-surface of the reservoir structure. In addition, the memory concept is important for not only proper modeling of fluid flow through reservoir porous media but also temperature effect on thermal flooding processes (Hossain et al., 2011).

A2.1 Fluid flow model in porous media

The equation of fluid motion by Darcy's law (i.e. volumetric fluid velocity is proportional to the negative of flow potential gradient) is not applicable for capturing changes in rock and fluid properties (Darcy, 1856). This model has its own restrictions arising from the assumption of a homogeneous medium, and constant reservoir rock and fluid properties. When porosity (ϕ), permeability (k), total compressibility (c_t) and viscosity (μ) are considered constants over the range of reservoir pressures (p), then the linearized Darcy-based diffusivity equation for fluid flow through porous media at the time (t) can be written (Dake, 1998)as:

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dp}{dr}\right) = \frac{\phi\mu c_t}{k}\frac{dp}{dt} \qquad \text{(A.1)}$$

B2.1.1 Classical models in porous media

Several studies by different researchers were done related to the diffusion of water in porous media. It is possible for some fluids to react chemically with a porous medium leading to changes in the pores. Similarly, solid particles embedded within the reservoir fluids under suitable conditions can be deposited or attached along the pore throats or walls. In addition, pore size has been reported to be affected by temperature variations within the porous media, and it is also reported in the literature that alterations of rock and fluid properties occur during a thermal flooding process (Krober et al., 1993). Several researchers have proposed different extensions of the classic Darcy's law by accounting for slip, inertia and so on (Brinkman, 1949; Bear, 1975; Sposito, 1980; Whitaker, 1986). Some fluid flow models are reported as alternative models; for example, the non-local flow theory is derived using the principles of statistical mechanics (Moroni and Cushman, 2001). The concept of fractional derivatives incorporated into constitutive equations is not a new theme and there are plenty of examples to be found in the literature to model rheological properties of solids, frequency-independent quality factor Fennoscandian uplift, heat diffusion, and other fields of research (Bagley and Torvik, 1986; Hilfer, 2000; Metzler and Klafter, 2004). Similarly, fractional diffusion models have been proposed to model sub or super-diffusion transport in the absence and presence of an external field (Metzler et al., 1999; Barkai et al., 2000); studies related to the dynamics of interfaces between nano-particles and substrate have also been reported (Chow, 2005). Balhoff et al. (2012) investigated the behavior of non-Newtonian flow in porous media for capturing the numerical accuracy in pore-scale network systems. Authors were able to investigate the

accuracy and efficiency of other popular equations used for modeling non-Newtonian fluids considering shear-thinning. The effect of fluid yield stress and zero fluid yield stress were also considered. In addition, for yield-stress fluids, the threshold pressure gradient required to initiate flow, and the qualitative pathway in which the fluid would travel at this threshold were determined. Different approaches were proposed for determining the threshold gradient (Sochi and Blunt, 2008; Sochi, 2010). Hristov (2013) proposed a new technique based on the assumption of final depth of penetration to obtain an approximate solution for heat conduction equation considering the fading memory, which is expressed by Jeffrey's kernel. Hristov observed that infinite speed of propagation of the flux is implicitly assumed for the conventional diffusion conservation equations. For the heat conduction problem, he introduced a Volterra-type integral, which has the damping function as was conceived by Cattaneo (1958). The proposed method consists of the classic Fourier's law, the heat flux and its time derivative related to its history. On the other hand, temperature and capillary hysteresis also play a significant role during thermal flooding process in porous media. Dynamic theories of multi-phase porous media developed for hydrocarbon reservoirs through modified wave propagation problem requirements and drag forces have drawn more attention from researchers than capillary hysteresis (Santos et al., 1990). Ashrafi et al. (2012) investigated that the decrease in residual oil saturation and increase in initial water saturation are expected by increasing temperature. The effects of temperature on capillary properties as well as relaxation time of liquid flux by capillary effects have also been studied by several authors (Molenda et al. (1993). Simonson et al. (1993) investigated an experimental and numerical study on thermal hysteresis effect for a

fiber glass insulation slab. Authors also concluded that it can lead to errors for an initially moist and dry fiber glass slab up to 23% and 51%, respectively, for a transport process at dynamic conditions while hysteresis effects are negligible. Therefore, modeling the dynamic process in the oil-water displacement process with thermal hysteresis and capturing the phenomena through experiments is the most significant research task. Wang and Sheng (2017) presented a low-velocity non-Darcy flow model to investigate the oil recovery performance of vertical and multi-fractured horizontal wells for shale and tight gas reservoirs. Authors concluded that oil production rate for Darcy flow is higher than that of non-Darcy flow for a vertical well. Besides, the ultimate oil recovery for Darcy flow is about 80% for the multi-fractured horizontal reservoir and 48% higher than that of non-Darcy flow vertical well.

A2.1.2 Memory-based fluid flow model

Several researchers have included the dependence and effects of the pathway on fluid history in their definitions of memory (Shin et al., 2003; Zhang, 2003; Zavala-Sanchez et al., 2009). Zhang (2003) stated that forward time events depend on previous time events while the memory is a function of time and space. Memory is the effect of past events on the present and future course of developments (Hossain and Abu-Khamsin, 2012). Caputo (1999) proposed a modified form of Darcy's law by introducing the time fractional derivative to account for the local permeability changes in porous media. In addition, the pressure distribution within a fluid in a half space under varying boundary conditions was derived. Finally, a method to determine the two parameters defining memory diffusion

model was suggested and memory effects on the volumetric flux along with its spectral properties were investigated (Caputo, 1999; 2000; Caputo and Plastino, 2003). Caputo and Plastino (2003) proposed a modified constitutive relation in order to better depict the diffusion process of fluids in porous media. The memory effect was introduced through space fractional derivative in this model. The purpose was to capture the effect on the medium previously affected by the fluid. This proposed constitutive equation implied a volumetric flow proportional to the spatial fractional derivative, plus another term as opposed to the classic Darcy's equation. Furthermore, Authors noticed that the Green function acted as a low-pass filter in the frequency domain. Consequently, while the time-memory is suitable for accounting for local phenomena, the space memory captures the variations in space. Caputo and Plastino (2003) are also associated with modifications to the classic Darcy equation and one or more constitutive equation by further studies. Caputo and Cametti (2008) proposed a space-dependent diffusion constant to Fick's diffusion equation to accurately describe the transport process. According to Caputo and Cametti (2008), it was possible for solid particles to obstruct some of the pores, leading to a non-constant permeability. Finally, Authors concluded that their approach was a generalization of Fick's equation to describe the diffusion process in more complex systems. Sprouse (Sprouse) proposed a numerical solution based on the short memory approach to solving a class of fractional diffusional heat equation through an explicit finite difference scheme. Carillo et al. (2014) showed analytical solutions for the integrodifferential equations describing a rigid heat conductor with memory. According to the authors, the introduction of memory effects provides an alternative way to account for nonlinearities in problems

where linear models cannot be applied. To understand the effect of memory on some physical property of the material, two different models are considered. Carillo et al. (2014) concluded that heat flux is related to the temperature gradient history for heat diffusion problems. Similarly, the stress tensor is related to the strain history of isothermal visco-elasticity problems. In addition, several studies have been devoted to properties of free energy functional for materials with memory (Coleman and Dill, 1973; Gurtin, 1974). Laffaldano et al. (2005) carried out experimental investigations to understand the permeability reduction observed during the diffusion of water in porous sand layers because of grain rearrangement and compaction. Elias and Hajash (1992) carried out an experimental investigation of this phenomenon. Further, the authors have proposed a modified diffusion model applicable to porous media by incorporating fluid memory formalism where it is able to give a good fit with the volumetric flux observed during experiments.

The variation of rock and fluid properties can be demonstrated using the concept of memory in the field of reservoir porous media. Hossain and Islam (2006) presented an extensive review of fluid memory based on the available literature and models, and applications pertaining to fluid flow in porous media problems. The authors showed how different researchers correlated the fluid memory to various fluid properties (e.g. the stress, density, and free energies). Hossain et al. (2008a) proposed a stress-strain relationship applicable to non-Newtonian flow in porous media by taking into account temperature and pressure variations, surface/interfacial tension, and rock and fluid memory. The authors

also investigated the effects of memory on the stress-strain curve assuming a homogeneous and isotropic porous media. It was concluded that the stress-strain behavior was a strong function of time, distance, and the memory parameter. Besides, the stress-strain models are also investigated by several researchers such as Mifflin and Schowalter (1986), Hossain and Islam (2009). Hossain and Islam (2009) introduced a modified material balance equation by including a stress-strain formulation for rock and fluid. Hossain et al. (2009b) developed a model that considers permeability variation over time using the memory concept and without linearization. Hossain et al. (2008d) proposed an extension of the classic fluid flow diffusivity equation by incorporating rock and fluid memory. It is derived by introducing the Caputo fractional derivative to the classic Darcy's law. Authors argued that this introduction is necessary to account for the variation of fluid and formation properties with time. In addition, authors also solved the resulting nonlinear integrodifferential equation using the explicit finite difference scheme. The general form of memory-based non-linear diffusivity equation can be written for any axial flow of a single-phase fluid in a porous media (Hossain et al., 2008a) as:

$$\frac{1}{\eta}\frac{d\eta}{dx}\left[\frac{\int_0^t (t-\zeta)^{-\alpha}\left(\frac{d^2p}{d\zeta dx}\right)d\zeta}{\tau(1-\alpha)}\right]+c_f\frac{dp}{dx}\left[\frac{\int_0^t (t-\zeta)^{-\alpha}\left(\frac{d^2p}{d\zeta dx}\right)d\zeta}{\tau(1-\alpha)}\right]+\frac{d}{dx}\left[\frac{\int_0^t (t-\zeta)^{-\alpha}\left(\frac{d^2p}{d\zeta dx}\right)d\zeta}{\tau(1-\alpha)}\right]=\frac{\phi c_t}{\eta}\frac{dp}{dt} \quad \text{(A.2)}$$

Where η is the composite variable, and α is the fractional order of the derivative related to time and space with $0 < \alpha < 1$. The other symbols are defined in the nomenclature section.

Hassan et al. (2015) presented a comprehensive study on composite pseudo-permeability, fluid velocity and viscosity of memory based diffusivity equation. Hassan et al. (2015) also

investigated the effects of these parameters on the pressure response of a reservoir. Table A.1 shows a summary of some selected equations/models related to fluid flow in porous media. The table explains the contributions of various researchers over recent decades in the areas of porous media flow modeling. Hossain (2016) modified the memory-based diffusivity equation incorporating integro-differential equation of fluid flow. The non-linear memory-based equation was solved using an implicit-explicit finite difference method. The author concluded that the memory-based equation results in less pressure drops compared to Darcy's model for a specified distance and time.

Obembe et al. (2017a) developed a modified memory-based model to represent the flow of fluids in porous media. Authors investigated this model considering the effect of different values of anomalous diffusion exponent and pseudo permeability with different simulation time and compared the results between the numerical scheme and the analytical solution. Authors studied the effect of anomalous diffusion coefficient on the diffusion of fluid within the reservoir. In their study a higher diffusion exponent corresponds to a higher effect of memory. It was concluded that the diffusion of the fluid decreases with the increase in the effect of memory.

Table A.1: Models describing fluid flow in porous media.

Investigator	Model/Equation	Remarks
Darcy (1856)	$\vec{u} = \frac{k}{\mu}[G - \nabla P] = \frac{k}{\mu}[\nabla \emptyset]$	a) Single phase flow b) Appropriate for laminar flow c) No chemical reaction between rock and fluid d) No-slip boundary condition is not considered e) There is no electro-kinetic effect.
Lauriat and Prasad (1989)	$\frac{\rho}{\phi}[\frac{d\vec{u}}{dt} + \frac{(\vec{u}.\nabla)\vec{u}}{\phi}] = \rho g - \Delta P + \mu\nabla^2\vec{u} - \frac{\mu}{k}\vec{u} - \rho\frac{c_F}{\sqrt{k}}\vec{u}\lvert\vec{u}\rvert$	a) The system is considered as isotropic b) Assumed homogenous and fluid-saturated porous medium c) The non-darcy effect is included.
Caputo (1999)	$A\frac{\partial\gamma}{\partial t^\gamma}\nabla^2 p = \frac{\partial p}{\partial t}$ Where $A = \eta\left[\frac{2G(1-\nu_u)}{(1-2\nu)}\right][\frac{B^2(1+\nu_u)^2(1-2\nu)}{9(1-\nu_u)(\nu_u-\nu))}]$	a) Memory formalism to simulate permeability reduction with time in geothermal areas b) Linear, isotropic, and homogeneous porous media c) Incompressible and viscous fluid d) The porosity of the media is not considered.
Caputo and Plastino (2003)	$-\frac{1}{k_1}\frac{\partial p}{\partial t} = \alpha''\frac{\partial^{2+n}}{\partial x^{2+n}}p + \beta\frac{\partial^{2+n}}{\partial x^2}p$ where α''= Modified coefficient of the Darcy's law β= Coefficient of the classical Darcy's law and k_1= Ratio of pressure to fluid density in an undisturbed condition	a) Space memory for diffusion in very thick layers b) Linear, isotropic, and homogeneous porous media c) Incompressible and viscous fluid d) The porosity of the media is not considered.
Iaffaldano et al. (2005)	$\gamma q(x,t) = -\left[c + d\frac{\partial^\gamma}{\partial t^\gamma}\nabla p(x,t)\right];$	a) Original Darcy's law is modified to account for memory b) Linear, isotropic, and homogeneous porous media

	$ap(x,t) = \alpha\rho(x,t);\ \nabla.q(x,t) + \frac{\partial\rho(x,t)}{\partial t} = 0;$ where γ, d and c are real numbers and α/a is the bulk modulus of fluid.	c) Incompressible and viscous fluid d) The porosity of the media is not considered.
Hossain et al. (2008a)	$\frac{1}{\eta}\frac{d\eta}{dx}Z + c_f\frac{dp}{dx}Z + \frac{d}{dx}Z = \frac{\phi c_t}{\eta}\frac{dp}{dt}$ where $Z = \frac{\int_0^t (t-\zeta)^{-\alpha}\left(\frac{d^2p}{d\zeta dx}\right)d\zeta}{\tau(1-\alpha)}$	a) Incompressible and viscous fluid b) The porosity of the medium is not considered
Hossain (2016)	$\frac{1}{\eta}\frac{d\eta}{dx}\sum_{k=1}^{n}(t-\xi_k)^{-\alpha}\left(\frac{d^2p}{d\zeta dx}\right)W_k + c_f\frac{dp}{dx}\sum_{k=1}^{n}(t-\xi_k)^{-\alpha}\left(\frac{d^2p}{d\zeta dx}\right)W_k + \sum_{k=1}^{n}(t-\xi_k)^{-\alpha}\left(\frac{d^3p}{d\zeta dx}\right)W_k = \frac{\phi c_t}{\eta}\Gamma(1-\alpha)\frac{dp}{dt}$	a) This model is written based on approximating the convolution integrals with summations b) It can be solved by explicit finite difference scheme
Obembe et al. (2017a)	$\frac{\partial}{\partial x}\left(\frac{\beta_c k_\alpha}{\mu_o}\frac{A_x}{B_o}\frac{\partial}{\partial x}D_t^{\alpha(t)}p\right)\Delta x + q_{sc} = \left(\frac{V_b\phi c_t}{\alpha_c B_o}\right)\frac{\partial p}{\partial t}$	a) One dimensional fractional diffusion equation b) Considered as single-phase slightly compressible fluid c) Constant fluid and rock properties d) Source term (q_{sc}) included in this model e) It is non-linear time-fractional partial differential equation.
Obembe et al. (2017b)	$\frac{\partial}{\partial x}\left(\frac{\beta_c k_\gamma}{\mu_o}\frac{A_x}{B_o}\frac{\partial p}{\partial x}\right)\Delta x = \left(\frac{V_b\phi c_t}{\alpha_{c\prime} B_o}\right)\frac{\partial^\gamma p}{\partial t^\gamma}$	a) One dimensional fractional diffusive equation b) Considered as a homogeneous porous medium c) Constant fluid and rock properties d) No producer or well present within the reservoir e) Specified flux at the left boundary and no flow at the right boundary f) Constant order fractional derivative.

Rammay et al. (2016) conducted a study on the variations of PVT properties using the Darcy's as well as the memory-based diffusivity equations. Hassan and Hossain (2016) developed a memory-based equation considering both the motion equation and the continuity equation to estimate temperature distributions during a thermal EOR process. It also concluded that coupling of pressure and temperature model can lead to more reasonable temperature distribution which is affected by fluid velocity and pressure variations. Furthermore, Obembe et al. (2018) studied anomalous effects during thermal displacement in porous media employing the order of fractional differentiation. Authors investigated the profiles of both pressure and temperature for different injection rates.

A2.2 Production mechanism and screening criteria for thermal flooding process

Various types of EOR techniques are employed to recover residual oil left in the reservoir when the primary and secondary recovery methods have been performed to their respective economic limits. The final goal of EOR methods is to enhance the overall oil displacement by considering the microscopic and macroscopic displacement efficiency. Thermal recovery processes such as cyclic steam stimulation (CSS), steam flooding (SF) and In-situ-combustion as well as steam-assisted gravity drainage (SAGD) have been the most widely used tertiary recovery methods for extracting heavy crude oil and bitumen. During thermal flooding, hot water or steam is injected into several wells while the oil is produced from adjacent wells. A typical mechanism of an oil recovery system is shown in Figure A.1 (Hossain et al., 2007). The solar energy can be used for a green environment for generating hot water or steam where a direct solar heating system has been established to

be effective by several authors (Krober et al., 1993; Kaye et al., 1998; Chu, 1983; Khan et al., 2005).

The physical and chemical mechanisms of different types of thermal flooding are well described in the literature (Burger et al., 1985; Butler, 1991; Prats, 2005). The list of screening criteria for thermal flooding methods is well described by several authors (Muskat, 1949; Taber and Martin, 1983; Taber et al., 1997; Green and Willhite, 1998; Hama et al., 2014; Kang et al., 2014; Elbaloula et al., 2016)). Green and Willhite (1998) mentioned that the reservoir depth, pressure and average permeability are the major screening criteria in any thermal flooding process evaluation. The performance of heavy oil recovery from thermal flooding depends not only on the reservoir geometry and characteristics but also on selected flooding type, and pattern size. This process classically proceeds through four phases of development. Several studies were done for EOR field project applicability based on lithology type, such as sandstone and carbonate reservoir rocks (Alvarado and Manrique, 2010). Figure A.2 shows the most EOR applications where the authors have extensively implemented the sandstone formations based on an available database of EOR developments covering 1,507 projects combined during the last decade (Alvarado and Manrique, 2010). As shown in Figure A.2, thermal methods are among the most extensively applied techniques in sandstone reservoirs compared to other EOR techniques. Thermal flooding projects have not been popular in carbonate reservoirs due to the chemical alterations in reservoir conditions.

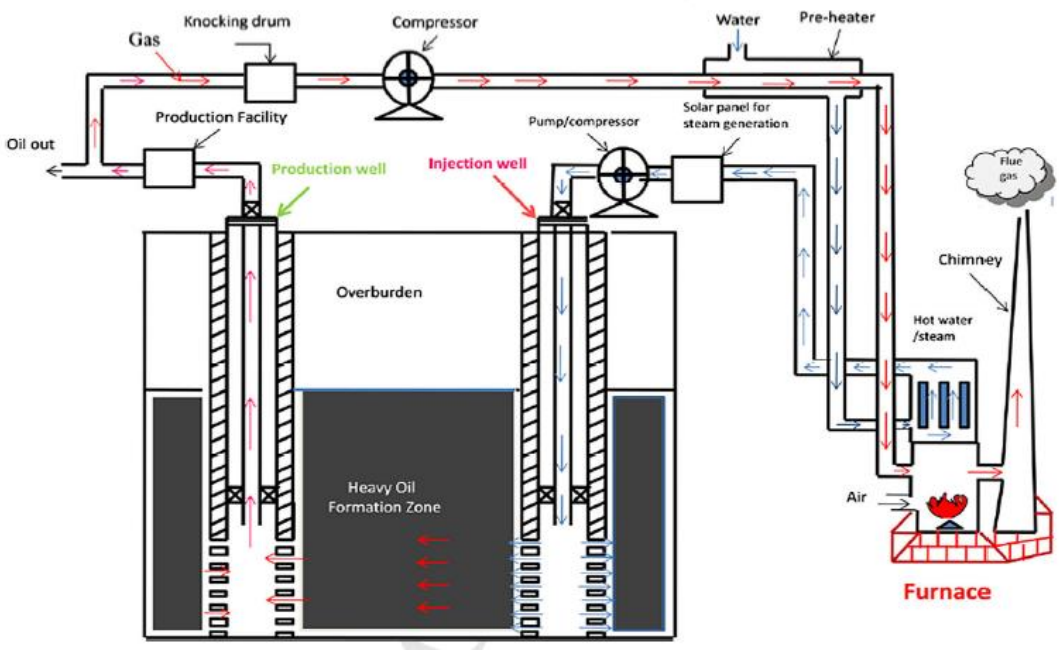

Figure A.1: Mechanism of the thermally oil recovery scheme (Hossain et al., 2007).

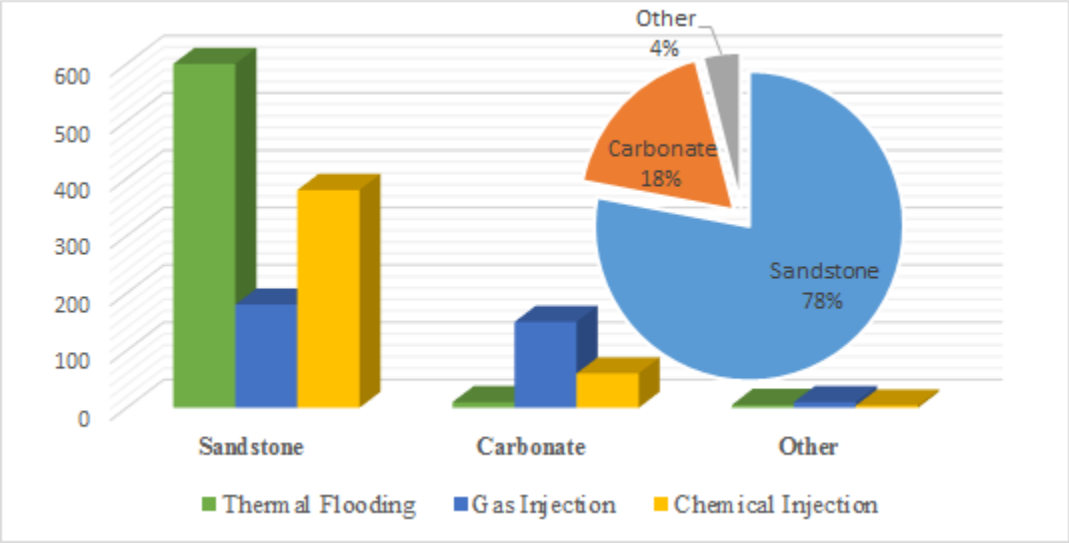

Figure A.2: Thermal and non-thermal EOR methods based on reservoir rocks (Redrawn from Alvarado and Manrique, 2010).

Garland Fields in Wyoming and Yates Field in Texas, USA are two of the few steam-driven projects in carbonate formations reported in the literature (Dehghani and Ehrlich, 1998; Manrique et al., 2007). In general, sandstone formations show the most potential for implementing thermal recovery projects because most of the technologies have been verified at the pilot and commercial scale in this type of formations (Alvarado and Manrique, 2010). Thermal flooding projects have been used mostly in Canada, the Former Soviet Union (FSU), the United States, Brazil, Venezuela and China, and to a lesser extent where steam injection began roughly five decades ago (Alvarado and Manrique, 2010). Globally CSS projects have been used extensively to produce heavy oil in the oil field such as Cold Lake in Alberta (Buckles, 1979), Midway Sunset field in California (Burns, 1969), Liaohe Shuguang field, Liaohe Huanxiling field, Gudao Field, Karamay field, and Gaosheng Field in China (Sheng, 2013).

The Tia Juana and Mene Grande fields in Venezuela (De Haan et al., 1969; Ernandez, 2009), the Kern River and Yorba Linda fields in California (Hanzlik, 2003) are examples of successful steam flooding projects that have been ongoing for over four decades. Some examples of recent steam flooding projects listed in the literature for sandstone formations such as Crude E Field (Ramlal, 2004), the Schoonebeek oil field and Alto do Rodrigues in Brazil (Jelgersma, 2007; Lacerda et al., 2008). Several coupling techniques were proposed to improve steam flooding process by using solvents (Rivero and Mamora, 2007), gases (Bagci and Gumrah, 2004), chemical additives (Ovalles et al., 2001) and foams (Mendez et al., 1992).

Field tests on in-situ combustion (ISC) have been extensively done in a variety of oil reservoirs. An extensive survey of ISC is done by Chu (1982a; 1982b). Some good examples of both dry and wet ISC projects are reported in Getty's Bellevue filed, United States (Bleakley, 1971), Golden Lake, LIoydminster in Canada (Fairfield, 1982), Santhal and Balol fields of Cambay Basin, India (Chattopadhyay, 2002; Chattopadhyay et al., 2003; 2004). An ISC pilot project was conducted by the Peace River in-situ project in the Peace River oil sands area of Canada. The pressure cycle steam drive process was tested as the pilot project in the same oil sands area (Lentz, 1971). Later Peace River Expansion Project (PREV) was commercialized in 1986 (Thimm, 1993). SAGD and CSS projects were also initiated in the same region of Canada (Hamm and Ong, 1995; Brissenden, 2005). CSS and SAGD have been applied on a commercial basis in Alberta for many years by oil producing companies in the different formations such as Cleanwater, Bluesky-Gething, and Lower Grand Rapids Formations (Nasr and Ayodele, 2005; Chang, 2013). Several commercial SAGD projects were being initiated for oil sand extractions in Canada (Suggett et al., 2000; Huc, 2011).

A2.3 Role of thermodynamic and reservoir properties during thermal flooding process

The thermodynamic properties of steam, and physical properties of the fluid and solid matrices are needed to predict the real performance of thermal recovery processes. In addition, the design of hot fluid injection projects requires a clear understanding of the thermal properties of steam, reservoir matrices and fluids. These properties such as

saturated water and steam heat content, saturation temperature and specific heat capacity are important in performing the calculations of heat losses and oil recovery displacement by both steam and hot water injection. The supporting correlations and data for crude oil and rock properties, and thermodynamic properties of water during thermal recovery calculations are well organized in the literature for thermal recovery calculations (Keenan and Keyes, 1936; Farouq-Ali, 1970; Taber and Martin, 1983; Boberg, 1988; Kartoatmodjo and Schmidt, 1994; Green and Willhite, 1998; Hossain et al., 2005). The porosity and permeability of reservoir rocks play an important role for both inductive and conductive heat transfer process considering energy balance equation during thermal flooding. In addition, porosity and temperature also affect the reservoir permeability during thermal EOR processes. Hossain (2017) showed that porosity increases the storage capacity of fluid during heat conduction process in the reservoir. It was concluded that permeability of reservoir increases with the increase in temperature at the outer boundary as well as toward the production well. In his study, two dimensionless numbers were considered to describe the rheological behavior of the rock and fluid interaction system. Temperature changes that take place in a reservoir during steam injection have a significant impact on the oil viscosity of reservoir. An appropriate temperature profile is useful for investigating the profile of temperature distribution and heat exchange between rock and fluid in the reservoir. The distribution of temperature profile is very significant in heavy oil recovery. When the thermal expansion is constrained due to a significant increase in temperature by hot fluid injection, thermal stress is developed (Collins, 2011). Hossain and Abu-khamsin (2012) confirmed that the continuous alterations of rheological properties of rock and fluid

influence heat conduction and convection significantly. The effective thermal conductivity of formation is used to characterize the thermal conductivity of reservoir rocks because thermal properties are determined by mineral constituents, porosity, and fluid saturating pore space (Green and Willhite, 1998). Some researchers studied the thermal conductivity of the rocks such as Somerton; Somerton and Boozer; Tikomirov; Anand et al.; Somerton et al.; Messmer; Setu and Bharatha; Arthur et al.; Peyman et al.; Irani and Cokar (Somerton, 1958; Somerton and Boozer, 1960; Tikomirov, 1968; Anand et al., 1973; Somerton et al., 1974; Messmer, 1984; Arthur et al., 2015; Mohammadmoradi et al., 2016; Irani and Cokar, 2016). Some selected models of thermal conductivity are mentioned in Table A.2.

Somerton (1958) investigated on heat capacities of fluid-saturated rocks and thermal conductivities of rock samples under considerations of fluid saturation. In addition, thermal diffusivity using those data was calculated. It was concluded that vapor or gas has a minor effect on heat capacities. However, heat capacities greatly changed due to variations in pressure, temperature, porosity and water saturation of the rock. In addition, the temperature has a substantial effect on thermal conductivity as well as thermal diffusivity. The influence of elevated pressure and temperature on thermal conductivity and diffusivity of fluid-saturated rocks was not investigated. Later Somerton et al. (1974) developed an empirical correlation for predicting thermal conductivity incorporating the effects of temperature, porosity and fluid saturation of unconsolidated oil sands. In this model, the impact of pressure was ignored.

Table A.2: Thermal conductivity models for unconsolidated and consolidated sands

Investigator	Model/Correlation	Considered Variables				Remarks
		T	P	ϕ	S_i	
Somerton (1958)	The effective thermal conductivity of oil sand, $\lambda_{hR} = 0.735 - 1.30\phi + \sqrt{S_w}$ Where $S_w = water\ saturation\ and$ $\phi = porosity\ of\ rock$	No	No	Yes	Yes	a) Sands (unconsolidated) contained by a mixture of fluids at the temperature of 325 Kelvin. b) Porosity range of 28-37%.
Tiknomirov (1968)	$\lambda_h(T) = 8.787[\frac{e^{0.6(\rho_b+S_w)}}{(T+459.6)^{0.55}}]$	Yes	No	No	Yes	a) Laboratory sample is sandstone saturated with water and gas
Anand et al. (1973)	The effective thermal conductivity of dry to liquid-saturated sandstone, $\frac{\lambda_{hR,l}}{\lambda_{hR,d}} = 1.00 - 0.3\left[\frac{\lambda_{hl}}{\lambda_{ha}} - 1.0\right]^{\frac{1}{3}}$ $+4.57(\frac{\phi}{1-\phi}\frac{\lambda_{hl}}{\lambda_{hR,d}})^{0.48m}(\frac{\rho_{R,l}}{\rho_{R,d}})^{-4.30}$	No	No	Yes	No	a) The thermal conductivity of the liquid is known at 68^0F b) Cementation factor (m) is 2.15 for consolidated sandstone.
Somerton et al. (1974)	λ =a + b$\phi + c\sqrt{S_w} + d\sqrt{S_o} + eT$ Where a, b, c, d and e are constant coefficients	Yes	No	Yes	Yes	a) It allows linear temperature dependence of thermal conductivity b) Considered multi-phase fluids

Prats (2005)	$\lambda_{hR} = \lambda_{hR,w} \times (\frac{\lambda_{ho}}{\lambda_{hR,w}})^{\wedge(\phi S_w)} \times (\frac{\lambda_{hg}}{\lambda_{hR,w}})^{\wedge(\phi S_g)}$	No	Yes	Yes	Yes	a) This correlation is for three phases (oil, water and gas)
Seto and Bharata (1991)	$\lambda = a + b\phi^m + cS_w^n + cS_o^p + cT^q$ Where m, n, p and q are constant exponents	Yes	No	Yes	Yes	a) It is used for clear water formation b) Temperature condition is 55^0F
Arthur et al. (2015)	$\lambda = 1.98S_W^{0.2}$	No	No	No	Yes	a) It is an experimental study for oil sands b) This model did not consider the impact of porosity, pressure and temperature.
Mohammadmoradi et al. (2016)	$\lambda_{eff} = \frac{-q}{\nabla T}$	No	No	Yes	Yes	a) Lithology type: sandstone b) Heat transfer at steady state conditions c) Finite element software (COSMOL) is used
Irani and Cokar (2016)	$\lambda = a + b\phi + c\sqrt{S_w} + d\sqrt{S_o} - eT = A - BT$ Where A=thermal conductivity at T=0, and B= thermal conductivity reduction as a function of T	Yes	No	Yes	Yes	a) Lithology type is unconsolidated oil sand

Setu and Bharatha (1991) modified the model proposed by Somerton et al. (1974) for estimating in-situ thermal conductivity using temperature logs and laboratory measurements core data from the Athabasca oil sands field, Clearwater formation at Cold Lake. The anisotropic condition for estimation of thermal conductivity was ignored in this study. Arthur et al. (2015) investigated changes in thermal conductivity through experimentation, considering the steady state measurement techniques for both radial and axial thermal systems. In addition, Arthur et al. (2015) also studied the performance of this model numerically using the COMSOL software to check its accuracy level. Authors concluded that thermal conductivity is significantly changed with water saturation. Later, Mohammadmoradi et al. (2016) investigated prediction of effective thermal and electric conductivity in porous media on pore-scale using the COMSOL software. It also concluded that rock solids are the crucial carrier and weak function of fluid saturation in the system of steam-water for thermal properties estimation at the pore scale.

A2.4 Modeling of reservoir heat transport and temperature distribution

The heat transport system is crucial to represent the real scenario of heat transfer through a porous medium into an oil reservoir during a thermal flooding process. Several mathematical heat transport models have been established to describe the heat transfer mechanism to predict the heated area, heating efficiency and temperature distribution in heavy oil and bitumen reservoirs. The available models are presented and critically analyzed in the following sections.

A2.4.1 Analytical investigation on heat transport and thermal bleeding

Thermal conduction is accountable for heat losses to the overburden and under-burden rocks of reservoir formation. Heat conduction is caused by temperature difference between adjacent particles without flowing fluid (Kaviany, 2012). Heat transfer can also occur through the movement of heated fluid. Convection heat transfer takes place as the hot fluid flows into the reservoir, and heat is transferred by the movement of particles within the fluid (Yu and Zhao, 2016). The relationship between convective heat flux and Darcy's velocity can be shown as in the following equation (Baston, 2008; Irani and Ghannadi, 2013):

$$q_v = \rho_f C_f (T_{inj} - T_r) \times -\frac{k}{\mu} \nabla P \tag{A.3}$$

Please refer to the nomenclature section for variables introduced in Equation A.3.

Several authors assumed that conduction is the main heat transfer mechanism during a thermal flooding process. Some researchers considered only conduction heat at the edge of the steam chamber from steam to cold oil reservoir in the classical SAGD models (Butler and Stephens, 1981; Butler et al., 1981; Reis, 1992; Akin, 2005; Liang, 2005; Nukhaev et al., 2006; Azad and Chalaturnyk, 2010). However, this assumption is questioned by several other authors (Farouq-Ali, 197; Edmunds, 199; Ito, 1998; Ito and Suzuki, 1999; Sharma and Gates, 2011; Irani and Ghannadi, 2013). Sharma and Gates (2011), Irani and Ghannadi (2013) presented convective heat flux from condensate flow at the edge of the steam chamber. Authors also considered both convection and conduction heat transfer mechanisms in the SAGD process in bitumen reservoirs (Irani

and Ghannadi, 2013). Some analytical models of reservoir heat transport have been critically reviewed in the preceding paragraphs.

Lauwerier (1955) developed an analytical solution for convective heat transport in porous media. Lauwerier assumed that heat transfers in the aquifer just by convection (no conduction) and into the adjacent layers by vertical conduction, and heat transport in a reservoir is uniform over the height. However, the heat loss from both overburden and under-burden cap rocks was not considered. Marx and Langenheim (1959) developed a conceptual heating model for both conduction and convection heat transfer in an idealized reservoir for calculating cumulative heated area, thermal efficiency and theoretical economic limit considering the following assumptions: (i) steam is introduced at a constant rate into a uniform reservoir, (ii) heat losses are negligible to the overburden and underburden rocks, and (iii) temperature is constant within the steam zone. Lawal and Vesovic (2009) investigated a one dimensional mathematical model for conduction heating of a heavy oil reservoir using three major assumptions: (i) homogenous reservoir, (ii) heating plane is maintained at a constant steam temperature, and (iii) local thermal equilibrium in the horizontal direction. It is concluded that induced convection is substantial in the vicinity of the heating plane. Barends (2010) modified Lauwerier's concept by including the effect of thermal bleeding during the conduction process. An analytical solution for plane-symmetric and radial convective heat transport including thermal bleeding effect was presented. These cases are limited to not only incompressible porous media with steady fluid flow patterns but also for Lauwerier's concept. Using the COSMOL 2D heat flow simulator,

it was also shown that thermal bleeding affects the entire front position with respect to the Peclet number.

Farouq-Ali (1997) first criticized the assumption that there is only conductive heat transfer in the SAGD process. It was pointed out that with so much condensate flowing, convective heat transfer would be expected to be dominant, which can be reasonable at high temperature. Yu and Zhao (2015) developed three models with different fluid injection scenarios to describe the transient heat transfer coupled with steady flow conditions during a SAGD process. These models are solved analytically using the Laplace transformation under consideration of continuous fluid injection at a constant temperature, with exponentially decreasing temperature, and periodic fluid injection at high and low temperature with respect to time (Yu, 2014). It was concluded that conductive and convective heat transfers simultaneously occur during a SAGD process. Finally, fluid flow is influenced by convection, and greatly increases the rate of heat transfer. Irani and Ghannadi (2013) analyzed the relative roles of convective and conductive heat transfers at the edge of a SAGD system's steam chambers for bitumen reservoirs and compared with several models (Ito and Suzuki, 1999; Edmunds, 1999; Sharma and Gates, 2011). Authors concluded that convection can transfer relatively large amount of heat at the edges of steam chambers. Later Irani and Gates (2013) presented that the convective heat flux linked with parallel flow to the edges of a chamber is minor compared to perpendicular to the edges. Irani and Cokar (2014) modified the Butler's model by incorporating the impact of temperature on thermal conductivity within the oil reservoir to a SAGD analytical model.

A2.4.2 Hot fluid injection rate and heating efficiency

Marx and Langenheim (1959) introduced a mathematical model incorporating heat transfer model for forecasting the growth of the steam zone during steam injection into a single well. It is assumed that no heat is transferred ahead of the front. However, authors did not consider the variable heat injection rate as a function of time. Later Ramey (1959) updated this model by incorporating the variable heat injection rates. Rubinshtein (1959) developed a heat transport model for calculating the heat efficiency while heat injection rate is constant. Boberg and Lantz (1966) developed a heating model for calculating the production rate and to predict the field performance during cyclic steam stimulation process. It was established for multiple oil sands which separated by shales. It is assumed that steam flow is radial and the heated zone is a cylindrical centred on the well.

Willman et al. (1961) presented an equation to estimate the steam injection rate neglecting both the sensible heat and the conduction process. Authors also suggested a calculation procedure to estimate heat requirements and oil recovery for hot fluid injection and steam drive process. By laboratory tests, it also concluded that both steam and hot water injection recover more oil than an ordinary water flood. Mandl and Volek (1967) developed a linear steam drive model considering a single layer reservoir to predict the total heat flow into the liquid zone and the steam zone growth. In this model, authors made some assumptions: (i) constant thermal properties as well as all saturation, temperature and fluid properties, (ii) heated layer is uniform thickness, (iii) gravitational effects and pressure gradient is negligible, (iv) variable injection rate for both sensible and latent heat as function of time, and so on. Mandl and Volek (1967)

first introduced the critical time across the condensation front which depends on reservoir temperature, thickness and steam quality. Mandl and Volek also showed that the theoretical steam-zone volume to be higher rather than the experimental result at near or after critical time.

Myhill and Stegemeier (1978) developed a mathematical model for prediction of oil and steam ratio. Authors also calculated the steam-zone growth using a slightly modified version of Mandl and Volek model (1967). Authors used the following basic assumptions for the development of models: (i) uniform rock and fluid properties, (ii) thermal properties are assumed constant throughout the zone, (iii) vertical temperature gradients in the reservoir are zero, (iv) heat losses from the steam zone are by conduction only, and occur normal to the reservoir into the base and cap rocks, (v) heat transfer in the reservoir by convection only, and (vi) heat passes only after critical time through the condensation front. Prats (1969) modified the model of Mandl and Volek (1967) considering two nearby layers separated by impermeable center layer during steam injection. In this model, it was considered that the injection rate is constant. It also concluded that steam zone volume is lower while heat transfer between the layers is negligible. Some selected models of reservoir heated area and heating efficiency is shown in Table A.3. Figure A.3 shows the typical profile of reservoir heating efficiency during the thermal flooding process. Butler (1991) presented the heating model for constant displacement rate which gives higher thermal efficiency compared to Rubenshtein (1959) and Marx and Langenheim (1959) models due to sufficient supply of heat to mitigate the heat losses.

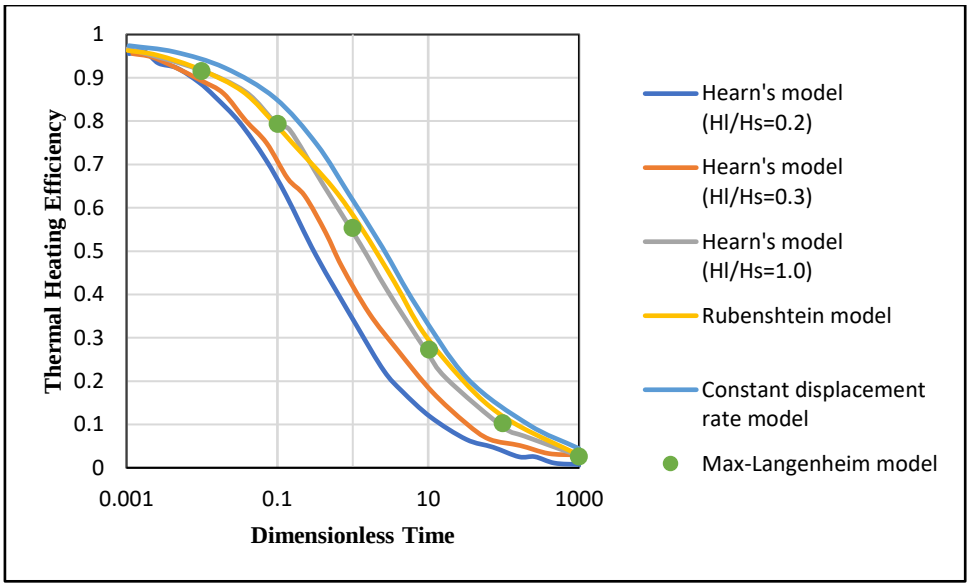

Figure A.3: A typical nature of thermal efficiency for different models (Redrawn from Nian and Cheng, 2017).

Hearn's model (1969) gives similar heating efficiency if we compare Marx and Langenheim model (1959) while there is no sensible heat transfer beyond the condensation front (CF). If latent to sensible heat ratio (H_l/H_s) is lower than 1, Hearn's model gives lower values of thermal heating than other models. Van Lookeren (1983) presented a calculation method to predict the steam-zone shape and oil recovery from radial and linear steam-drive process using the principles of segregated flow and steam override effects. The calculated results of a radial steam zone shape were also verified by experimental study. It is concluded that the shape of the steam zone is controlled by not only steam-injection rate and pressure but also by effective formation permeability to steam. Jensen et al. (1991) proposed a model for steam-drive projects, which showed a reasonably better accuracy over the existing models such as Myhill and Stegemeier (1978), Van Lookeren (1983), and also compared with 15 field scale steam-drive projects.

Table A.3: Mathematical models for reservoir heated area and heating efficiency.

Investigator	Model/Equation	Remarks
Marx and Langenheim (1959)	The cumulative heated area at time t, $A_s(t) = \left(\frac{H_0 MhD}{4\lambda_R{}^2(T_s - T_r)}\right)[e^{t_D} erfc\sqrt{t_D}) + 2\sqrt{\frac{t_D}{\pi}} - 1]$ where $t_D = \left(\frac{4\lambda_R{}^2 MhD}{M^2 h^2 D}\right) t$ and $M = \phi(S_o\rho_o C_o + S_w\rho_w C_w) + (1-\phi)\rho_r C_r$ $Heating\ efficiency, E_H = \left(\frac{1}{t_D}\right)[\left(e^{t_D} erfc(\sqrt{t_D}) + 2\sqrt{\frac{t_D}{\pi}}\right) - 1]$	a) Injection heat content (injection rate) is constant b) The areal shape of the steam zone is not specified c) Heat losses are negligible to the cap and base rocks. d) Heat is not transferred beyond the heat front
Ramey (1959)	$A_s(t) = \left(\frac{H_0(t)}{2\rho Cb(T_{inj} - T)}\right) * (e^{t_D{}^2} erfc\ t_D)$	a) The injection rate is a function of time b) This model is modified based on Marx and Langenheim (1959) c) Constant effective volumetric specific heat throughout the formation
Rubinshtein (1959)	$E_H = 1 - (1-\beta) \times \left\{ \begin{array}{l} \frac{\sqrt{\gamma\tau}}{\pi}\left[1 - (1-\beta)\sum_{n=1}^{m} \beta^{n-1}\left(1 + \frac{n^2}{\gamma\tau}\right) e^{\left(1+\frac{n^2}{\gamma\tau}\right)}\right] \\ +(1-\beta)\sum_{n-1}^{\infty} n\beta^{n-1}\left(1 + \frac{2n^2}{3\gamma\tau}\right) erfc\frac{n}{\sqrt{\gamma\tau}} \end{array} \right\}$ where $\beta = \frac{\gamma a - 1}{\gamma a + 1}$, $\gamma = \frac{k_1}{k_2}$ and $a^2 = \frac{k_2\rho_1 c_1}{k_1\rho_2 c_2}$	a) Thermal properties are same for both cap and base rocks

		b) Heating efficiency in the oil layer in between the overburden and the under-burden formation
Myhill and Stegemeier (1978)	$E_{h,s} = \frac{1}{t_D}\left[\left\{2\sqrt{t_D} - 2(1-f_{h,v}) \times \sqrt{t_D - t_{cD}} - \int_0^{t_{cD}} \frac{e^u erfc\sqrt{u}}{\sqrt{t_D - u}} du - \sqrt{\pi} G(t_D)\right\} + \frac{(1-f_{h,v})U(t_D - t_{cD})}{\sqrt{\pi}} G(t_D)\right]$ $Where\ U(t_D - t_{cD}) = 1\ for\ t_D > t_{cD}\ and\ 0\ for\ t_D \leq t_{cD}$	a) The weighting factor (1-$f_{h,v}$) is empirical b) The correlation should not be used for low-quality steam-drives
Prats (1969)	$H(t) = \int_0^t \dot{Q}(t') K\left(\theta_n \sqrt{t - t'}\right) dt - F \int_0^t H_o(t') dt\, K\left(\theta_n \sqrt{t - t'}\right) dt'$ $where\ K(z) = e^{z^2} erfc(z),\ \theta_n = \frac{\lambda_{h2z}}{h\sqrt{\alpha_2 (\rho c)_1}}$ and $F = \frac{(\rho c)_f - (\rho c)_n}{(\rho c)_n}$	a) It is assumed that effective volumetric heat capacity of the pay zone is independent of the liquid saturation b) This model is considered for hot water injection that there is no distinction between near and far regions, and coefficient, F =0.

Cheng et al. (2016) investigated wellbore heat loss during hot water injection. Authors concluded that injection time and pressure, and air temperature have minimal effect on heat loss while injection rate and temperature, the proportion of hot water as well as insulated tube have a significant effect on heat loss from offshore wells. Nian and Cheng (2017) showed number of heat transport models in pipe wellbore, formation near the wellbore, and in reservoirs to evaluate heat loss from hot fluid to surrounding and for estimating wellbore fluid temperature.

A2.4.3 Models related to temperature distribution

Thomas (1967) proposed an approximate method for determining the temperature profile when injecting fluid is hot. This model approximation is applicable to relatively thin reservoir beds in which the injected fluid flow is high. He considered that convective heat flux in the radial direction is the leading heat transport mechanism. Ziagos and Blackwell (1986) developed a two-dimensional mathematical model in fluid flow in geothermal structures using two major assumptions: (i) the thickness of the aquifer is negligible compared to the layer thickness, and (ii) conduction process is negligible in the horizontal direction. Authors proposed a method to forecast the extent of the zone of influence and its scale for any combination of hydrological and thermal parameters. Numerous models and theory are developed for temperature distribution considering various scenarios in geothermal and naturally fractured reservoirs (Satman, 1988; Kocabas and Horne, 1990), thermal injection modeling by well test analysis (Aeschliman et al., 1983; Jahanbani Ghahfarokhi et al., 2012), and development of analogies to tracer transport (Kocabas, 2001; 2006).

For a CSS process, Boberg and Lantz (1966) developed a model for determining average temperature which was solved by an analytical technique using the principle of superposition. In this model, two adjustments were used: (i) there is no initial temperature spreading in the shale above or below the heated zone and added a hypothetical thickness for each sand thickness; and (ii) used a dimensionless parameter (sigma) to account for energy removed by produced fluids. Kocabas (2004) developed a two-dimensional analytical model and solved it to predict the dimensionless temperature distribution across half of the injection plane. In this conceptual model, several assumptions used such as (i) linear flow unit confined by two layers, (ii) incompressible fluid flow with constant linear flow velocities in the directions of x and z, and (iii) heat conduction coefficients in the boundary layers is constant and equal. A major benefit of this model is that it gives a correct estimation of thermal efficiency than the one-dimensional models considering infinite transverse conductivity. In addition, it serves to detect the roles of boundary conditions and fluid mechanics through dispersion parameters.

Li et al. (2010) developed a mathematical model and solved it by Laplace domain solution approach to simulate the temperature distributions of the aquifer and rocks in an aquifer thermal energy storage system. Authors made some assumptions for this model such as: (i) the underlying and overlying, and confined aquifer is homogeneous, anisotropic, and uniform thickness, (ii) buoyancy flow is negligible, (iii) the initial temperature is constant over the whole aquifer system, and (iv) the temperatures in both rocks are uniformly disseminated over vertical depth before the hot water injection. In addition, the steady state solution obtained by the final value theorem while thermal energy loss is zero at any radial distance from the aquifer into rocks. It observed that

the value of aquifer temperature exhibits higher while the effect of thermal conductivity is negligible. Besides, it is concluded that aquifer temperature will be underestimated when wellbore radius is considered as negligible. Irani and Cokar (2016) developed an analytical SAGD model to calculate the oil production rate and oil production and steam oil ratio (SOR). Authors presented a modification of Butler's model (1985a; 1985b) in which thermal conductivity varies as a function of temperature within the oil reservoir. Authors also investigated on SOR variation under consideration of temperature-dependent thermal conductivity effects. Irani and Cokar (2016) concluded that SOR is independent of the thermal conductivity for both laterally and angularly expanding reservoirs. Barends (2010) developed a model to forecast the temperature profile in porous rocks for both radial and linear flow conditions. The model was solved analytically using the methods of Boltzmann and the Laplace transformation. The model is also validated by the COMSOL software. It was concluded that thermal bleeding (i.e. heat leakage) affects the temperature distribution in both linear and radial flow directions. Li and Cheng (2015) constructed the condensate velocity versus temperature ahead of a steam chamber, and temperature versus distance normal to the interface ahead of the steam chamber using several models such as Birrell (2001), and Sharma and Gates (2011). Li and Cheng (2015) showed that water condensate velocity drops more rapidly rather than Sharma and Gates (2011) model with the decrease of temperature under typical Athabasca oil sands conditions. Besides, the results presented that temperature from Li and Cheng model is approximately 10 ^{0}C lower than that of Sharma and Gates model, and 13 ^{0}C higher than that from Butler's model.

Yu and Zhao (2016) developed a one-dimensional heat transfer model under consideration of steady flow, and pressure gradient between crude oil and injection fluid

zone is constant. Authors solved this mathematical model analytically to depict the temperature behavior as a function of time and location. It was stated that conductive and convective heat transfer occur simultaneously during a SAGD process. Yu and Zhao (2016) also concluded that the heating area and the temperature gradient are smaller for a reservoir and fluid system with larger thermal diffusivity. In addition, the increasing rate of reservoir temperature is smaller when thermal diffusivity is lower due to poor ability to store thermal energy. Lawal and Vesovic (2009) developed a one-dimensional heat transfer model in linear systems to predict the viscosity profile and temperature distribution at different conditions. Authors used the correlation of the temperature-dependent oil density and viscosity of an Athabasca bitumen reservoir. In addition, Lawal and Vesovic (2009) also used the magnitude of Nusselt number for the relative importance of natural convection. Authors concluded that the significance of free convection depends on the rock and fluid properties including duration of heating. Miura and Wang (2012) developed an analytical model to predict the cumulative SOR as a function of both average reservoir properties, and time-dependent operation variables. Wei et al. (2014) proposed a new analytical model to determine the oil production rate, steam chamber development process, water cut, and steam oil ratio including production performance of SAGD under a constant steam injection rate. Authors also compared the results between STARS and proposed a new model. It was concluded that the shape of the steam chamber is affected by injection rate. Finally, Lawal (2016) modified the model of Lawal and Vesovic (2009) for steam flooding process. The author concluded that a maximum of four zones (i.e. conduction, quiescent, convection and condensation) can be recognized at any instant of the thermal flood.

A summary of some mathematical models for the temperature distribution in heavy oil reservoir is shown in Table A.4. In addition, a brief review of some selected analytical solutions for describing some aspects of thermal flooding process is presented in Table A.5. A typical reservoir temperature distribution profile is depicted in Figure A.4.

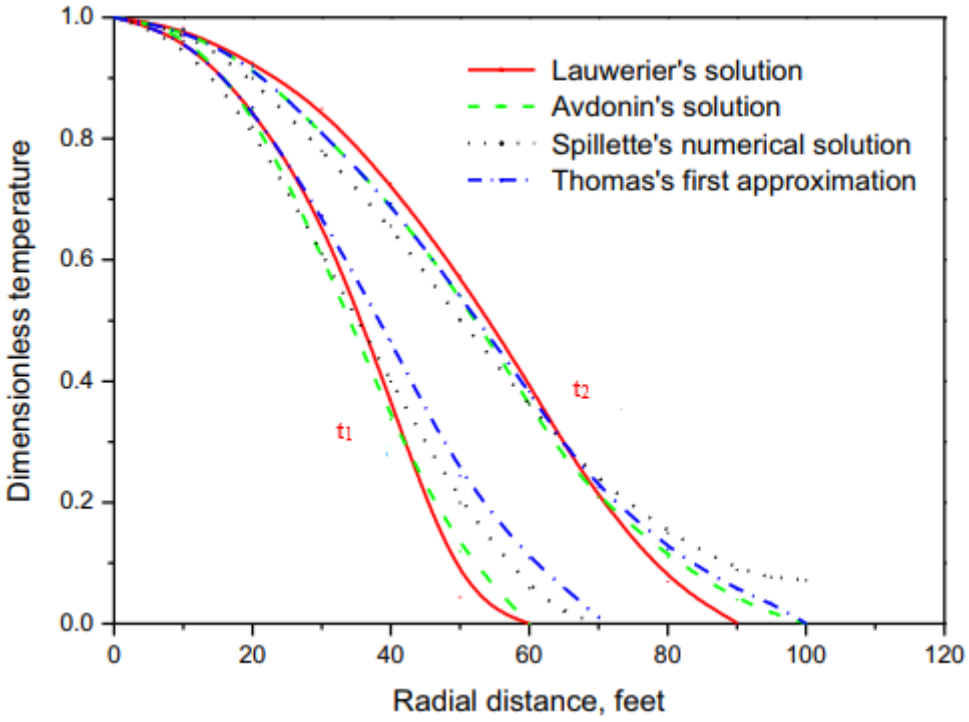

Figure A.4: A typical plot of temperature distribution where $t_2 > t_1$ (Redrawn from Nian and Cheng, 2017).

For radial distances of a different reservoir, the temperature distribution gives a higher value for both analytical and numerical solutions at the longer time (t_2) than less time (t_1) duration. The analytical and numerical solutions also give almost close results for temperature distribution at time t_2. On the other hand, at a short time (t_1), the solutions show slightly different profiles compared to a longer time at the same distance as can be seen from in Figure A.4.

Table A.4: Analytical models of temperature distribution for thermal flooding.

Investigator	Model/Equation	Remarks
Butler (1985a)	$\frac{T-T_r}{T_{st}-T_r} = e^{-\frac{v_s \xi\prime}{\alpha}}$ Where v_s = linear velocity of steam front, and ξ' =distance measured ahead of the front into the coder zone	a) This model is considered as quasi-steady-state condition, and 1 D heat transfer mechanism b) Heat losses for both cap and base rocks are negligible c) Homogeneous porous medium d) Thermal conductivity is constant
Kocabas (2001)	Temperature distribution, $T_D = 1 - \sum_{n=0}^{\infty}\sum_{m=0}^{n} \frac{(-1)^m}{m!}\binom{n}{m}\frac{2^{m+1}}{\sqrt{m}}\int_0^{\infty} F\omega^m e^{-\omega^2} H_m(\omega) d\omega$ Where $F = erfc\left\{\frac{2\sqrt{\frac{\tau}{\theta}}\omega+\alpha_1}{2\sqrt{\eta}}\right\} + erfc\left\{\frac{2\sqrt{\frac{\tau}{\theta}}\omega+\alpha_2}{2\sqrt{\eta'}}\right\}$	a) Assumed that flow rate is constant b) This model considered for injection of low temperature into a hot reservoir
Hossain et al. (2007)	$\frac{\partial T^*}{\partial t^*} + \frac{\rho_s c_{ps}}{M}(u^*)\frac{\partial T^*}{\partial t^*} - \frac{(k_s+k_f)}{MLu_i}\frac{\partial^2 T_s}{\partial x^{*2}} = 0$ Conditions: Initial and boundary conditions	a) One-dimensional numerical model in a one-dimensional reservoir. b) The temperature of rock and fluids is not similar c) Assumed that thermal conductivity of rock and fluid is constant in the system

	$T^*(x,0)=1, T^*(0,t)=\frac{T_{st}}{T_i}\ and\ T^*(L,t)=1$	d) Governing equation solved using an explicit finite difference scheme.
Barends (2010)	$T-T_i=\frac{(T_1-T_i)}{2}\left\{erfc\ (\frac{x-vt}{2\sqrt{\alpha t}})+e^{\frac{xv}{\alpha}}\ erfc\ (\frac{x+vt}{2\sqrt{\alpha t}})\right\}$ Where v= heat convective velocity, and x= special coordinate horizontal	a) Considered the local thermal equilibrium between fluid and rock grains. b) Diffusivity and reservoir height are constant
Irani and Ghannadi (2013)	Dimensionless temperature distribution, $T_D^*=\frac{T-T_r}{T_{st}-T_r}=\frac{\sum_{n=0}^{\infty}\frac{(P^*)^{\frac{U_x}{\lambda\alpha}+n}}{n!\times\frac{U_x}{\lambda\alpha}+n}(-\eta)^n}{\left\{\frac{\lambda\alpha}{U_x}+\sum_{n=1}^{\infty}\frac{1}{n!\times\frac{U_x}{\lambda\alpha}+n}(-\eta)^n\right\}}$ Where $P^*=\frac{P-P_r}{P_{st}-PT_r}\ and\ \eta=\frac{U_x(P_{st}-P_r)kk_{rw}}{\alpha_c\mu_w}$	a) The system is considered as a quasi-steady-state condition b) Assumed a stable bitumen/steam front. c) Condensate velocity is included during convection heat transfer mode d) There is a lack of effect on transverse heat transfer into the oil sand beyond the chamber
Irani and Gates (2013)	$T_D^*=\left[1-\frac{\lambda_r}{n(n+1)U_x^2\rho_r c_{pr}}\frac{kk_r g\ sin\theta}{\mu_{st}}\right]\times\left[e^{\left(-\frac{U_x\rho_r c_{pr}}{\lambda_r}\xi\right)}\right]+\left[\frac{\lambda_r}{n(n+1)U_x^2\rho_r c_{pr}}\frac{kk_r g\ sin\theta}{\mu_{st}}\right]\times e^{\left[-(n+1)\frac{U_x\rho_r c_{pr}}{\lambda_r}\xi\right]}$	a) The system is considered as a quasi-steady-state condition b) Outflow convection is a result of temperature variation

	where n= power constant for viscosity variation versus temperature	
Li and Chen (2015)	$$\frac{\partial^2 T^*}{\partial^2 \xi^2} + \frac{U_x \rho_{os} c_p - \rho_c c_{pc} V_c}{\lambda_r} \frac{dT^*}{d\xi} = 0$$ Temperature gradient, $T^* = c_1 + c_2 e^{\frac{\rho_c c_{pc} V_c - U_x \rho_{os} c_p}{\lambda}} \times \xi$ where V_c = condensate velocity normal to the steam chamber edge	a) Temperature gradient at the boundary condition is zero b) Condensate velocity is constant c) The system is considered as a quasi-steady-state condition
Yu and Zhao (2016)	$T(x,t) =$ $$\begin{cases} T_i + \frac{T_0 - T_i}{2}\left[erfc\left(\frac{x-ut}{2\sqrt{Dt}}\right) + e^{\frac{\alpha\gamma}{D}} erfc\left(\frac{x+ut}{2\sqrt{Dt}}\right)\right], t \le t_0 \\ T_i + \frac{T_0 - T_i}{2}\left[erfc\left(\frac{x-ut}{2\sqrt{Dt}}\right) + e^{\frac{\alpha\gamma}{D}} erfc\left(\frac{x+ut}{2\sqrt{Dt}}\right)\right] + \\ \frac{T_i - T_o}{2}\left[erfc\left(\frac{x-u(t-t_0)}{2\sqrt{D(t-t_0)}}\right) + e^{\frac{\alpha\gamma}{D}} erfc\left(\frac{x+u(t-t_0)}{2\sqrt{D(t-t_0)}}\right)\right], t > t_0 \end{cases}$$	a) Steady-state and 1 D heat transfer coupled model b) Thermal diffusivities of the reservoir and the fluid systems are constant c) Temperature as a function of both time and space

Table A.5: Selected some heat transport models related to temperature distribution and thermal efficiency into the oil reservoir

Investigator	Assumptions							
	Injecetd Fluid	Flow type	Thermal conductivity				Solution	
			Impermeable strata		Productive sand			
			Horizontal	Vertical	Horizontal	Vertical	Temperature distribution	Thermal efficiency
Lauwerier (1955)	Hot water	Linear	Zero	Finite	Zero	Infinite	Yes	Yes
Marx and Langenheim (1959)	Steam	Radial	Zero	Finite	Zero	Infinite	No	Yes
Ramey (1959)	Steam	Radial	Zero	Finite	Zero	Infinite	No	Yes
Rubinshtein (1959)	Hot liquid	Radial	Finite	Finite	Finite	Finite	No	Yes
Willman et al. (1961)	Steam	Radial	Zero	Finite	Zero	Finite	No	Yes
Malofeev and Scheinman (1963)	Hot water	Radial	Zero	Finite	Zero	Infinite	Yes	Yes

Avdonin (1964)	Hot water	Linear, Radial	Finite	Infinite	Zero	Finite	Yes	No
Kocabas (2004)	Cold water	Linear	Zero	Finite	Infinite	Finite	Yes	No
Barends (2010)	Hot water	Linear, Radial	Zero	Finite	Zero	Finite	Yes	No
Lawal and Vesovic (2009)	Hot fluid	Linear	Zero	Infinite	Zero	Finite	Yes	No
Li et al. (2010)	Hot fluid	Radial	Zero	Finite	Finite	Zero	Yes	No
Miura and Wang (2012)	Steam	Linear	Zero	Finite	Zero	Infinite	No	No
Wei et al. (2014)	Steam	Linear	Zero	Infinite	Finite	Finite	No	No
Lawal (2016)	Steam	Linear	Zero	Infinite	Zero	Finite	Yes	No

A2.5 Models related to oil production rate and recovery performance

A variety of empirical and analytical models have been developed to determine the oil flow rate, oil displacement efficiency, steam oil ratio and recovery performance through the injection of hot water and steam for thermal flooding. Several models have been critically analyzed in the following subsections.

A2.5.1 Oil production rate models

The first gravity drainage based model of oil production rate is predicted by Butler et al. (1981). Authors used some assumptions such as (i) constant reservoir height and porosity (ii) uniform thermal diffusivity and effective permeability of oil, (iii) steam pressure is constant in the steam chamber, (iii) heat transfer ahead of the steam chamber to cold oil is only by thermal conduction. This model reportedly overestimated oil drainage rate and predicted an un-realistic steam zone shape. Later, this model was modified by Butler and Stephens (1981) to obtain a better model for the steam zone shape. Reis (1992) modified the model of Butler et al. (1981) using the value of dimensionless temperature coefficient. This model is developed by experimental investigation using an inverted cone geometry of steam-chamber. The above models did not consider the effects of asphaltene deposition and steam distillation. Some selected models of oil production rate have been shown in Table A.6. The model equations are critically analyzed and reported in the remarks of the table.

Table A.6: Oil production rate models during SAGD process.

Investigator	**Model/Equation**	Remarks
Butler et al. (1981)	Oil drainage rate, $q = L\sqrt{\frac{2\phi\Delta S_o k g \alpha h}{a m v_s}}$	a) Constant steam pressure, and only steam flows in the steam chamber b) Considered only conduction heat transfer system c) Assumed temperature coefficient, a=1
Butler and Stephens (1981)	The rate of oil flow, $q = L\sqrt{\frac{1.5\phi\Delta S_o k g \alpha h}{m v_s}}$	a) This model altered the theory of Butler by pinning the base of the chamber to the production-well location b) It is known as "Tandrain" model
Reis (1992)	Oil flow rate, $q = 2L\sqrt{\frac{\phi\Delta S_o k g \alpha h}{2a m v_s}}$	a) This model did not include effects of asphaltene deposition and steam distillation, and assumed temperature coefficient, a= 0.4 b) Assumed that steam-chamber geometry is an inverted cone
Akin (2005)	Cumulative oil production $Q_o = \sqrt{\frac{\phi\Delta S_o k_o g w_s h}{2 m v_s} t}$	a) This is a modification of Butler et al. (1981) and Reis (1992) models b) This model includes the effects of steam distillation and asphaltene deposition

Sharma and Gates (2010)	$q = 2L\sqrt{\frac{2\phi\Delta S_o g h \alpha k k_{rocw} \Gamma(m)\Gamma(a+1)}{v_s \Gamma(m+a+1)}}$	a) This model included relative permeability effects b) It does not consider the effect of Geomechanics
Cokar et al. (2013)	The oil-phase velocity, $u_{oil} = \frac{e^{E+F\phi_o\left[1+\beta(T_s-T_r)e^{-\frac{U\xi}{\alpha}}\right]}}{v_s} \times (1-e^{-\frac{U\xi}{\alpha}})^a (e^{-\frac{U\xi}{\alpha}})^m$	a) This model overcomes the limitations of Sharma and Gates (2010) model b) It includes the dilation caused by thermal expansion
Irani and Cokar (2014)	$q = (A-BT)\frac{U_x \rho_r c_{pr}}{A} \times \frac{\sqrt{1-\frac{2B}{A}T\left(1-\frac{B}{2A}T\right)}}{\left[(\frac{B}{A})^2 T - \frac{B}{A}\right]} \times \frac{1}{\left\{1+\sqrt{1-\frac{2B}{A}T_r}(1-\frac{B}{2A}T_r)\left[\sqrt{X}-\sqrt{Y}\right]\right\}}$ $where\ X = \frac{2B}{A}T\left(1-\frac{B}{2A}T\right) and$ $Y = \frac{2B}{A}T_r\left(1-\frac{B}{2A}T_r\right)$	a) This is modified form of Butler's SAGD model b) Convective heat transfer was not considered c) λ as a function of temperature

Akin (2005) experimentally investigated the prediction of cumulative oil production rate during a SAGD process by incorporating the effects of asphaltene deposition and steam distillation. It was concluded that asphaltene deposition and steam distillation effects dominate over steam zone size and lateral heat transfer effects during late stages of a SAGD operation. Sharma and Gates (2010) modified the theory of Butler and developed a model to predict the oil production rate including the effect of relative permeability. This model is solved by the gamma function solution approach. Sharma and Gates concluded that most mobile oil is not at the edge of steam chambers but at some distances into the oil sands. However, authors did not include the effect of geomechanics as well as that of thermal expansion on the reservoir properties.

Later, Cokar et al. (2013) modified the Sharma and Gates model (2010) to predict the oil flow rate at the edge of steam chamber including the effect of thermal expansion. Cokar et al. (2013) developed an analytical model for SAGD oil rate considering the effect of thermogeo-mechanics at the edge of the steam chamber. This model also considered the effect of relative permeability and the effects of changes in permeability and porosity. It was concluded that geo-mechanical effects should be included in the analysis of oil phase flow at the edge of steam chambers. Besides, authors mentioned that oil drainage and production rates are substantial at the edge of the steam chamber due to the impact of thermal expansion. In addition, Cokar et al. (2013) reported that the peak oil-phase velocity is upto 45% which is higher than that from the Sharma and Gates model (2010) due to the consideration of geo-mechanical effects.

A2.5.2. Investigation on oil displacement and recovery performance

Oil displacement rate is predicted by using the frontal advance models or gravity-override models. Myhill and Stegemeier (1978) modified the frontal advance model to predict the ultimate steam oil ratio (SOR) for a steam drive process. Jones (1981) proposed an analytical model which is an extension of the model proposed by Myhill and Stegemeier (1978) through the introduction of the capture efficiency (three dimensionless factors). The model converts the oil displacement rate obtained from the results of Myhill and Stegemeier (1978) steam flood to the corresponding actual oil production rate. The correlation was developed based on the results of 14 different steam-flood projects. Jones (1981) assumed that any steam flood consists of three production stages. The first stage is controlled by initial oil viscosity, the second stage is controlled by hot oil mobility and reservoir permeability, and the final stage is dominated by the remaining mobile fraction of original oil in place. Chandra and Mamura (20025) improved the analytical model of John's (1981) to predict the capture efficiency as well as steam flood production performance analysis. Neuman (1985) developed a model to predict the post-steam breakthrough performance and steam injection rate. However, the pre-steam breakthrough period was not addressed. Van der Knaap (1993) showed that Neuman model can be derived from Mandl-Volek model (1967), and Authors are completely compatible. Miller and Leung (1985) developed an approximation method to estimate oil displacement rates when gravity drainage is the dominant mechanism. Chu (1985) developed a series of empirical correlation to estimate the

SOR with the various reservoir and crude properties by regression analysis based on 28 field projects.

Edmunds and Peterson (2007) developed an analytical model to predict the cumulative steam-to-oil ratio (CSOR) for steam-based recovery of bitumen from high permeability reservoirs. Authors interpreted as unsteady-state SAGD recovery for cyclic steam which can be operated at effective temperatures about half of that for SAGD. This model captured the fundamentals of the basic physics considering material and energy balances. This model has the limited dependency on time through heat loss term and neglects the practical aspects such as the operating conditions and other time-dependent factors such as rising steam chamber, the chamber, and the temperature difference between injection steam and production fluid. Miura and Wang (2012) developed an analytical model to predict the CSOR. It is a modified form of Edmunds and Peterson model using both the energy balance equation and the gravity drainage theory. Authors extended this model by incorporating operating variables such as the temperature of production and injection and accounting for the effect of time-dependent variables such as oil saturation and height of the rising chamber to mimic the applied SAGD process. Table A.7 and Table A.8 shows the oil displacement capture efficiency model and SOR, respectively.

Davies and Silberberg (1968) presented a technique for predicting the performance of five-spot steam floods under considerations of both Buckley and Leverett (1942), and Marx and Langenheim model (1959). Gomaa (1980) developed a set of

correlation charts for the prediction of oil-to-steam ratio and oil recovery as a function of reservoir characteristics and operating conditions. It was concluded that CSOR depends strongly on initial mobile oil saturation, thickness, and net to gross ratio of a reservoir. In addition, a series of correlations based on numerical simulation using a particular set of fluid and rock properties were developed. Rhee and Doscher (1980) developed a model for predicting oil recovery including the effects of gravity override and distillation during steam flooding process. Vogel (1984) presented a simple, practical and conservative approach to calculate the steam requirements for a steam flood. It was assumed that the steam chamber spreads immediately across the top of the whole reservoir pattern. In addition, Vogel suggested that it is required to inject steam at a higher rate initially in a steam flood, and then to reduce the rate to compensate for the reduced vertical heat flux. Several studies also have been done to characterize the bitumen and crude oil in porous media for applying of thermal flooding process (Phillips et al., 1985; Ayasse et al., 1997; Yoshiki and Phillips, 1985; Lal and Mather, 1999; Cai and Chung, 2001; Sadrameli, 2015; Li et al., 2016; Rodriguez-DeVecchis et al., 2017). Jabbari et al. (2017a; 2017b) investigated the oil recovery performance coupling by thermally-induced wettability alteration from hot-water imbibition in naturally fractured reservoirs. Jabbari et al. concluded that the rate of oil recovery of counter-current imbibition is less than that in co-current imbibition in a hot-water injection process.

Table A.7: Models of capture efficiency for oil displacement during thermal flooding.

Investigator	Capture efficiency = $A_{cD} \times V_{oD} \times V_{pD}$			Remarks
	A_{cD}	V_{oD}	V_{pD}	
Jones (1981)	$A_{cD} = \left[\frac{A_s}{A\left\{0.11ln\left(\frac{\mu_{oi}}{100}\right)\right\}^{0.5}}\right]^2$ $With\ limit: 0 \leq A_{cD} \leq 1.0\ and$ $A_{cD} = 1.0\ at\ \mu_o \leq 100\ cp.$	$V_{oD} = \left[1 - \frac{N_d}{N} \times \frac{S_{oi}}{\Delta S_o}\right]^{\frac{1}{2}}$ $With\ limit: 0 \leq V_{oD} \leq 1.0.$	$V_{pD} = \left[\frac{V_{s,inj} \times 5.615}{43560Ah_n\varphi S_g}\right]^2$ $With\ limit: 0 \leq V_{pD} \leq 1.0$ $and\ V_{pD} = 1.0\ at\ S_g.$	a) The heat capacity of the reservoir rock is 1.2 times less than the base and caprocks b) This model considered that production life cycle is three stages of steam flooding process c) Used a correlation to estimate the critical time as a dimensionless form (t_{cD})
Chandra and Mamora (2005)	$A_{cD} = \frac{\left(\frac{A_s}{A}\right)^4}{A\left\{\alpha \times ln\left(\frac{\bar{\mu}_o}{100}\right)\right\}^{0.5}}$ $With\ limit: 1.0 \leq A_{cD} \leq A_{cD_{max}};$ $A_{cD} = 1.0\ at\ \bar{\mu}_o \leq 100\ cp$ $and\ \alpha = 0.00015i_s + 0.05;$	$V_{oD} = A_{cD_{max}} \times e^{\frac{\beta S_{oi}(N_{d_{max}} - N_d)}{N\Delta S_o}}$ $limit: A_{cD_{max}} > V_{oD} \geq 1.0;$ $\beta = 17.93N_c + 1.3401$ and $N_c = \left[\frac{7758Ah(1-S_{or}-S_{wc})}{365Ai_s t_c}\right];$	$V_{pD} = \left[\frac{V_{s,inj} \times 5.615}{43560Ah_n\varphi S_g}\right]^2$ $With\ limit: 0 \leq V_{pD} \leq 1.0$ $and\ V_{pD} = 1.0\ at\ S_g.$	a) This model has overcome the limitations of Jones model (1981) b) Used linear correlation between alpha (α) and steam injection rate (i_s)

Table A.8: Models of steam oil ratio during thermal flooding.

Investigator	**Model/Equation**	Remarks
Chu (1985)	SOR = 18.744 + 0.001453D – 0.05088h – 0.0008864k – 0.0005915μ - 14.79S_o - 0.0002938 $\frac{kh}{\mu}$	a) This correlation was developed base on 28 different steam flood field projects. b) All reservoir properties are assumed constant c) The memory concept is not considered
Edmund and Peterson (2007)	$CSOR = \frac{\Delta T\left\{C_{vr} + \frac{\sqrt{\lambda_t C_{vo} t}}{h\eta_s}\right\}}{H_{lv}\phi\Delta S_{oi}}$	a) Residual oil saturation is assumed constant b) Neglected the vertical development period of the steam chamber c) Horizontal chamber shape and effective sweep efficiency constant as 50% are assumed
Miura and Wang (2012)	$CSOR = \frac{\Delta T(t)\left\{C_{vr} + \frac{\sqrt{k_t C_{vo} t}}{\beta h_s(t)}\right\}}{\Delta H(t)\phi\left[S_{oi} - \frac{b-1}{b}\left\{\frac{v_s(t)\phi h_s(t)}{bkgt}\right\}\right]^{\frac{1}{b-1}}}$	a) This model is overcoming the limitations of that by Edmund and Peterson (2007) b) It is assumed that the effect on the reservoir-chamber height is very small

Stewart and Udell (1987) developed a quasi-steady model for one-dimensional steam displacement by considering the effects of capillary pressure gradients as well as gravity on steam displacement. Authors presented the distributions for both pressure and saturation to show the mechanisms for semi-analytical calculations. Authors also concluded that Buckley-Leverett displacement velocity is higher than the velocity of steam condensation front from the practical interest point of view. Scott (2002) compared between SAGD and CSS projects in the highlighting points of energy efficiency as well as bitumen recovery performance using available data from Clearwater formation at Cold Lake. This author proposed a conversion factor from the external gas requirement to SOR based on the quality of steam. Li and Chalaturnyk (2006) studied the variations of absolute permeability in response to isotropic unloading and shearing within unconsolidated oil sands during the SAGD operations. Authors concluded that the reservoir permeability is varied due to the changes of pore geometry under considerations of both isotropic and deviatoric system. Besides, authors suggested that geomechanical effects can be incorporated to predict the absolute and effective permeability variations into a SAGD process. Wang et al. (2008) showed that electrothermal oil recovery method is a promising and economical technology to develop low permeability bitumen resources compared to the SAGD process.

A3 Future Research Guideline

To minimize the knowledge gaps in the literature regarding the issues identified in this paper, a comprehensive study is needed to incorporate (i) the memory effects for complex fluid flow modeling in porous media, (ii) alterations of rock and fluid properties, (iii) variations in hot fluid injection rate, (iv) oil displacement rate and steam oil ratio, and (iv)

temperature distribution during thermal flooding process. Some specific research opportunities are included here as a guideline for future research:

- Formation compressibility, porosity and permeability are directly related to the reservoir depth and pressure. Reservoir porosity and pressure decrease over time. Permeability changes over distance because it is directly related to the pressure of the complex reservoir system. Reservoir fluid properties also change with position and time. Besides, the rheological properties also change with space and time as well as the temperature of the complex reservoir in porous media. A comprehensive study in the area will give more understanding for predicting thermal fluid activities in porous media as well as the thermal exchange between rock and fluid and can capture the whole spectrum of the field life during thermal recovery processes.
- The porosity, permeability and fluid saturation change not only with temperature and pressure but also with space and time. In addition, formation temperature and pressure also change with space and time during hot fluid injection of thermal recovery. Therefore, coupling those parameters, several models for thermal conductivity such as Tiknomirov (1968), Somerton et al. (1974), Seto and Bharato (1991), Irani and Cokar (2016) and so on can be modified. Thus, memory-based thermal conductivity model and volumetric heat capacity can help to develop modified heat transport model for prediction of temperature distribution profile during thermal flooding processes.
- Coupling of influential factors into rock and fluid properties, oil flow rate models (e.g. Butler and Stephens (1981), Akin (2005), Sharma and Gates (2010) and so on) can be modified to capture the memory effect during hot fluid injection.

- On the other hand, residual oil situation is also changed by hot fluid injection into oil and bitumen reservoirs. The steam-oil ratio (SOR) models (such as Chu (1985), Edmund and Peterson (2007), Miura and Wang (2010), and capture efficiency of oil displacement models such as Jones (1981), Chandra and Mamora (2005) model can be modified to capture the memory effect during steam (hot fluids) flooding process.
- As stated earlier, thermal conductivity and volumetric heat capacity are not constant with respect to time through heat injection into reservoir formation of thermal recovery. In addition, the thermal diffusivity of reservoir and fluid system is not constant due to non-homogenous reservoir heating systems. So, the reservoir heating models such as Lauwerier (1955), Marx and Langenheim (1959), Myhill and Stegemeier (1978) can be improved by considering both conduction and convection heating process. Memory concept can be implemented into the reservoir heating system. The analytical models of temperature distribution such as Barends (2010), Irani and Ghannadi (2013), Irani and Gates (2013), Li and Chen (2015), and Yu and Zhao (2016) can be modified to include influential parameters as well as using the memory concept in complex reservoir system for thermal recovery.

A4 Concluding Remarks

The reservoir properties, as well as thermal properties are assumed to remain constant with time in most of the current thermal EOR emulators. The rock properties vary with temperature changes during a thermal flooding process. Besides, crude oil viscosity is a strong function of temperature. The critical review rebuilds that controlling parameters of

effective thermal conductivity are formation porosity, permeability, fluid saturation, pressure, temperature and lithology quality. The thermal properties of consolidated rocks differ from unconsolidated oil sands due to changes in cementation factor and deposition of sediments. The reservoir rock and fluid properties, and thermal properties such as thermal conductivity and volumetric heat capacity play a significant role in both inductive and conductive heat transfer process considering energy balance equation during thermal flooding process. It is also identified that thermal properties are not the same for both overburden and underburden rocks. The steam injection models are more complicated than the hot water injection models due to the changes in fluid phase in the reservoir formation and moving thermal boundaries. During thermal recovery methods, the amount of oil recovered is a function of several variables, which are: i) flooding pattern size, ii) heat loss (thermal bleeding) via the flood time, iii) the reservoir pressure and the amount of steam injected, iv) the net sand thickness of the producing interval, v) steam properties and vi) the state of the primary depletion. It is crucial to predicting the real behavior of reservoir rock and fluid properties in porous media and hysteresis effect on thermal flooding in the complex reservoir system. The memory-based oil flow rate, CSOR, heating efficiency, heat transport and temperature distribution models are required to predict the real recovery performance of thermal flooding.

Acknowledgements

The authors would like to thank Natural Sciences and Engineering Research Council of Canada (NSERC); Research & Development Corporation of Newfoundland and Labrador

(RDC), funding no. 210992; and Statoil Canada Ltd., funding no. 211162 for providing financial support to accomplish this research under Statoil Chair in Reservoir Engineering at the Memorial University of Newfoundland, St. John's, NL, Canada.

Nomenclatures

List of symbols

A_{cD}	Dimensionless steam zone size [fraction]
A (t)	Cumulative heated area function of time [m^2]
a	Corey coefficient of the oil relative permeability curve
B_o	Oil formation volume factor [rm^3/sm^3]
C_f	Specific heat capacity of fluid [$JKg^{-1}K^{-1}$]
CSS	Cyclic steam stimulation
C_o	Specific heat capacity of oil [$JKg^{-1}K^{-1}$]
CSOR	Cumulative steam oil ratio (fraction)
C_w	Specific heat capacity of water [$JKg^{-1}K^{-1}$]
C_r	Specific heat capacity of rock [$JKg^{-1}K^{-1}$]
c_F	Non-dimensional form-drag constant
c_t	Total compressibility in porous medium [1/Pa]
D	Thermal diffusivity [m^2/s]
dt	Time step [s]
E_H	Heating efficiency [percentage]
$E_{h,s}$	Overall reservoir thermal efficiency [percentage]
EOR	Enhanced oil recovery
G	The body force term due to gravity [N]
g	Acceleration due to gravitation force [N]
H_o	Heat injection rate [J/s]
H_m	Hermite polynomial of order m

h	Reservoir thickness [m]
k	Reservoir permeability [m^2]
M	Volumetric (overall) heat capacity [$JK^{-1}m^{-3}$]
m	Temperature/viscosity parameter
P	Pressure of condensate [Pa]
P_r	Reservoir pressure [Pa]
P_s	Pressure of the system [Pa]
P_{st}	Steam temperature [Pa]
Q_o	Cumulative oil production rate [m^3/s]
q	Oil flow (drainage) rate [m^3/s]
q_v	Convective heat flux [W/m^2]
r	Radial distance of reservoir in equation (1) [m]
SAGD	Steam-assisted gravity drainage
SF	Steam flooding
S_o	Oil saturation [fraction]
S_g	Gas saturation [fraction]
S_w	Water saturation [fraction]
T	Reservoir temperature [K]
T_{inj}	Injection temperature [K]
T_{st}	Steam temperature [K]
T*	Temperature gradient [K/m]
T_D^*	Dimensionless temperature distribution [dimensionless]
dt/dx	Temperature gradient along direction of heat transfer [K/m]
t	Time [s]
t_D	Dimensionless time [dimensionless]
t_{cD}	Dimensionless critical time [dimensionless]
U_x	Velocity of the advancing front of steam chamber (m/s)
$\vec{u}$	Velocity vector [m/s]
V_{oD}	Volume of displaced oil produce [fraction]

V_{pD}	Initial pore void filled with steam as water [fraction]
v_s	Linear velocity of steam front [m/s]
W_k	A weighting function for the numerical integration
w_s	Steam zone half width [m]

Greek letters

∇P	Pressure gradient [Pa/m]
ΔS_o	Change in oil saturation before/after steam front passage [fraction]
Δx	Size of grid block in x direction
$\nabla \emptyset$	Fluid potendial gradient [N]
ξ	A dummy variable for time i.e., real part in the plane of the integral [s]
ξ'	Distance measured ahead of the front into the coder zone [m]
$d\xi$	Dummy time step [s]
ϕ	Porosity of fluid media [fraction]
μ	Fluid dynamic viscosity at any temperature [Pa-s]
μ_o	Oil (dynamic) viscosity [Pa-s]
μ_w	Water (dynamic) viscosity [Pa-s]
ρ_c	Condensate density [kg/m^3]
ρ_f	Fluid density [kg/m^3]
ρ_o	Oil density [kg/m^3]
ρ_r	Dry rock density [kg/m^3]
ρ_w	Dry rock density [kg/m^3]
λ	Thermal conductivity [Wm^{-1}K^{-1}]
$\lambda_h(T)$	Effective thermal conductivity function of temperature [Wm^{-1}K^{-1}]
λ_{hf}	Thermal conductivity of fluid or formation [Wm^{-1}K^{-1}]
λ_{hR}	Effective thermal conductivity of oil sand [Wm^{-1}K^{-1}]
λ_r	Reservoir thermal conductivity [Wm^{-1}K^{-1}]
$\lambda_{h,s}$	Thermal conductivity of steam [Wm^{-1}K^{-1}]

η	Ratio of the pseudo-permeability of the medium with memory to fluid viscosity [$m^3 s^{1+\alpha}/kg$]
η'	Scaled dimensionless space variable
α	Fractional order of differentiation (related to the time and space), dimensionless
α_1, α_2	Derived variable for dimensionless thickness
α_c	Simplified condensate (water) or convective diffusivity
$\alpha_{c\prime}$	Volumetric conversion factor
β	Coefficient of the classical Darcy's law
β_c	transmissibility conversion factor
γ	Fractional order derivative
Γ	Euler gamma function
ξ	Normal distance to the advancing front of the steam chamber [m]
θ	Inclination of the draining surface from the horizontal plane [angle]
ω	Dummy integral variable

Subscripts

e	Effective
erfc	Complementary error function
f	Fluid
g	Gas
m	Confining layer
o	Oil
r	Rock (matrix)
v	Convection
w	Water

References

Aeschliman, D. P., Meldau, R. F., & Noble, N. J. (1983). Thermal efficiency of a steam injection test well with insulated tubing. In SPE California Regional Meeting. Society of Petroleum Engineers.

Akin, S. (2005). Mathematical modeling of steam assisted gravity drainage. SPE Reservoir Evaluation & Engineering, 8(05), 372-376.

Anand, J., Somerton, W. H., & Gomaa, E. (1973). Predicting thermal conductivities of formations from other known properties. Society of Petroleum Engineers Journal, 13(05), 267-273.

Arthur, J. K., Akinbobola, O., Kryuchkov, S., & Kantzas, A. (2015). Thermal Conductivity Measurements of Bitumen Bearing Reservoir Rocks. In SPE Canada Heavy Oil Technical Conference. Society of Petroleum Engineers.

Ashrafi, M., Souraki, Y., & Torsaeter, O. (2012). Effect of temperature on Athabasca type heavy oil–water relative permeability curves in glass bead packs. Energy and Environment Research, 2(2), 113-126.

Avdonin, N. A. (1964). Some formulas for calculating the temperature field of a stratum subject to thermal injection. Neft'i Gaz, 3, 37-41.

Ayasse, A. R., Nagaishi, H., Chan, E. W., & Gray, M. R. (1997). Lumped kinetics of hydrocracking of bitumen. Fuel, 76(11), 1025-1033.

Azad, A., & Chalaturnyk, R. J. (2010). A mathematical improvement to SAGD using geomechanical modelling. Journal of Canadian Petroleum Technology, 49(10), 53-64.

Bagci, A. S., & Gumrah, F. (2004). Effects of CO2 and CH4 addition to steam on recovery of West Kozluca heavy oil. In SPE International Thermal Operations and Heavy Oil Symposium and Western Regional Meeting. Society of Petroleum Engineers.

Bagley, R. L., & Torvik, P. J. (1986). On the fractional calculus model of viscoelastic behavior. Journal of Rheology, 30(1), 133-155.

Balhoff, M., Sanchez-Rivera, D., Kwok, A., Mehmani, Y., & Prodanović, M. (2012). Numerical algorithms for network modeling of yield stress and other non-Newtonian fluids in porous media. Transport in porous media, 93(3), 363-379.

Barends, F. (2010). Complete solution for transient heat transport in porous media, following Lauwerier. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Barkai, E., Metzler, R., & Klafter, J. (2000). From continuous time random walks to the fractional Fokker-Planck equation. Physical Review E, 61(1), 132.

Baston, D. P. (2008). Analytical and Numerical Modelling of Thermal Conductive Heating in Fractured Rock (Doctoral dissertation, Queen's University).

Bear, J. (1975). Dynamics of Fluids in Porous Media. Soil Science, 120(2), 162-163.

Birrell, G. E. (2001). Heat transfer ahead of a SAGD steam chamber, a study of thermocouple data from phase B of the underground test facility (dover project). In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Bleakley, B.W. (1971). Evolution of the fry in-situ combustion project. Oil & Gas. J., 69 (18).

Boberg, T. C. (1988). Thermal Methods of Oil Recovery, John Wiley and Sons, New York City.

Boberg, T. C., & Lantz, R. B. (1966). Calculation of the production rate of a thermally stimulated well. Journal of Petroleum Technology, 18(12), 1-613.

Brinkman, H. C. (1949). A calculation of the viscous force exerted by a flowing fluid on a dense swarm of particles. Flow, Turbulence and Combustion, 1(1), 27.

Brissenden, S. J. (2005). Steaming uphill: Using J-wells for CSS at peace river. In Canadian International Petroleum Conference. Petroleum Society of Canada.

Buckles, R. S. (1979). Steam stimulation heavy oil recovery at Cold Lake, Alberta. In SPE California Regional Meeting. Society of Petroleum Engineers.

Buckley, S. E., & Leverett, M. (1942). Mechanism of fluid displacement in sands. Transactions of the AIME, 146(01), 107-116.

Burger, J., Sourieau, P., & Combarnous, M. (1985). Thermal Methods of Oil Recovery, Editions Technip.

Butler, R. M. (1985a). A new approach to the modelling of steam-assisted gravity drainage. Journal of Canadian Petroleum Technology, 24(03), 42-51.

Butler, R. M. (1985b). New interpretation of the meaning of the exponent “m” in the gravity drainage theory for continuously steamed wells. AOSTRA Journal of research, 2(1), 67-71.

Butler, R. M. (1991). Thermal Recovery of Oil and Bitumen. Prentice-Hall, Englewood Cliffs.

Butler, R. M., & Stephens, D. J. (1981). The gravity drainage of steam-heated heavy oil to parallel horizontal wells. Journal of Canadian Petroleum Technology, 20(02).

Butler, R. M., McNab, G. S., & Lo, H. Y. (1981). Theoretical studies on the gravity drainage of heavy oil during in-situ steam heating. The Canadian journal of chemical engineering, 59(4), 455-460.

Cai, H. Y., Shaw, J. M., & Chung, K. H. (2001). Hydrogen solubility measurements in heavy oil and bitumen cuts. Fuel, 80(8), 1055-1063.

Caputo, M. (1999). Diffusion of fluids in porous media with memory. Geothermics, 28(1), 113-130.

Caputo, M. (2000). Models of flux in porous media with memory. Water Resources Research, 36(3), 693-705.

Caputo, M., & Cametti, C. (2008). Diffusion with memory in two cases of biological interest. Journal of theoretical biology, 254(3), 697-703.

Caputo, M., & Plastino, W. (2003). Diffusion with space memory. In Geodesy-the Challenge of the 3rd Millennium (pp. 429-435). Springer, Berlin, Heidelberg.

Carillo, S., Valente, V., & Caffarelli, G.V. (2014). Heat conduction with memory: a singular kernel problem, Evolution Equations & Control Theory, 3(3).

Cattaneo, C. (1958). A form of heat-conduction equations which eliminates the paradox of instantaneous propagation. Comptes Rendus, 247, 431.

Chandra, S., & Mamora, D.D. (2005). Improved Steamflood analytical model. In: SPE International Thermal Operations and Heavy Oil Symposium. Society of Petroleum Engineers.

Chang, J. (2013). Understanding HW-CSS for Thin Heavy Oil Reservoir. In SPE Heavy Oil Conference-Canada. Society of Petroleum Engineers.

Chattopadhyay, S. K., Binay, R., Bhattacharya, R. N., & Das, T. K. (2004). Enhanced oil recovery by in-situ combustion process in Santhal field of Cambay basin, Mehsana, Gujarat, India-A case study. In SPE/DOE Symposium on Improved Oil Recovery. Society of Petroleum Engineers.

Chattopadhyay, S.K. (2002). Enhanced oil recovery by in situ combustion process in Santhal and Balol Fields of Cambay Basin, Mehsana, Gujarat, India-A Case Study. Presentation to Shell Holland.

Chattopadhyay, S.K.., Ram, B., Maviya,C., Das,B.K., Mittal,V.K., Meena,H.L.,et al. (2003). Enhanced oilrecoveryby in situ combustion processing Balol Field of Cambay Basin, India—A Case Study, Indian, Oil and Gas Review Symposium IORS-2003, 8-9 September, Mumbai, India.

Cheng, W. L., Han, B. B., Nian, Y. L., & Wang, C. L. (2016). Study on wellbore heat loss during hot water with multiple fluids injection in offshore well. Applied Thermal Engineering, 95, 247-263.

Chow, T. S. (2005). Fractional dynamics of interfaces between soft-nanoparticles and rough substrates. Physics Letters A, 342(1-2), 148-155.

Chu, C. (1982a). Discussion of State-Of-The-Art Fireflood Field Projects-Reply. Journal 0f Petroleum Technology, 34(4), 861-862.

Chu, C. (1982b). State-of-the-Art Review of Fireflood Field Projects, Journal 0f Petroleum Technology, 19-36.

Chu, C. (1983). Current in-situ combustion technology. Journal of Petroleum Technology, 35(08), 1-412.

Chu, C. (1985). State-of-the-art review of steamflood field projects. Journal of petroleum technology, 37(10), 1-887.

Coleman, B. D., & Dill, E. H. (1973). On thermodynamics and the stability of motions of materials with memory. Archive for Rational Mechanics and Analysis, 51(1), 1-53.

Collins, P. M. (2011, January). Geomechanical Screening Criteria for Steam Injection Processes in Heavy Oil and Bitumen Reservoirs. In SPE Heavy Oil Conference and Exhibition. Society of Petroleum Engineers.

Dake, L. P. (1998). Fundamentals of reservoir engineering. Developments in Petroleum Science 8, pp. 438.

Darcy, H. (1856). Les fontaines publiques de la ville de Dijon, Victor Dalmont, Paris. The Flow of Homogeneous Fluids Through Porous Media.

Davies, L.G., & Silberberg, I.H. (1968). A Method of Predicting Oil Recovery in a Five-Spot Steamflood. Society of Petroleum Engineers. doi:10.2118/1730-PA.

De Haan, H. J., & Van Lookeren, J. (1969). Early results of the first large-scale steam soak project in the Tia Juana Field, Western Venezuela. Journal of Petroleum Technology, 21(01), 101-110.

Dehghani, K., & Ehrlich, R. (1998). Evaluation of steam injection process in light oil reservoirs. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Dong, X., Liu, H., Wang, Q., Pang, Z., & Wang, C. (2013). Non-Newtonian flow characterization of heavy crude oil in porous media. Journal of Petroleum Exploration and Production Technology, 3(1), 43-53.

Edmunds, N. (1999). On the difficult birth of SAGD. Journal of Canadian Petroleum Technology, 38(01).

Edmunds, N., & Peterson, J. (2007). A unified model for prediction of CSOR in steam-based bitumen recovery. In Canadian International Petroleum Conference. Petroleum Society of Canada.

Elbaloula, H., Pengxiang, H., Elammas, T., Alwad, F., Rdwan, M., & Musa, T. (2016). Designing and Implementation of the First Steam Flooding Pilot Test in Sudanese Oil Field and Africa. In SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition. Society of Petroleum Engineers.

Elias, B. P., & Hajash Jr, A. (1992). Changes in quartz solubility and porosity due to effective stress: An experimental investigation of pressure solution. Geology, 20(5), 451-454.

Ernandez, J. (2009). EOR Projects in Venezuela: Past and Future. ACI optimising EOR strategy, 11-12.

Fairfield, W. H., & White, P. D. (1982). Lloydminster fireflood performance, modifications promise good recoveries. Oil Gas J.;(United States), 80(6).

Farouq-Ali, S. M. (1997). Is there life after SAGD. Journal of Canadian Petroleum Technology, 36(6), 20-23.

Farouq-Ali, S.M. (1970). Oil Recovery by Steam Injection, Producers Publi. Co. Inc., Bradford, PA., 4-6.

Gomaa, E. E. (1980). Correlations for predicting oil recovery by steamflood. Journal of Petroleum Technology, 32(02), 325-332.

Green, D. W., & Willhite, G. P. (1998). Enhanced Oil Recovery, SPE textbook series. Society of Petroleum Engineers, Richardson, Texas.

Gurtin, M. E. (1974). Modern continuum thermodynamics(first and second laws of thermodynamics). Mechanics today., 1, 168-213.

Hama, M. Q., Wei, M., Saleh, L. D., & Bai, B. (2014). Updated Screening Criteria for Steam Flooding Based on Oil Field Projects Data. Society of Petroleum Engineers (2014, June 10).) doi:10.2118/170031-MS

Hamm, R. A., & Ong, T. S. (1995). Enhanced steam-assisted gravity drainage: a new horizontal well recovery process for peace river, Canada. Journal of Canadian Petroleum Technology, 34(04).

Hanzlik, E. J., & Mims, D. S. (2003). Forty years of steam injection in California-The evolution of heat management. In SPE International Improved Oil Recovery Conference in Asia Pacific. Society of Petroleum Engineers.

Hassan, A. M., & Hossain, M. E. (2016). Coupling Memory-Based Diffusivity Model with Energy Balance Equation to Estimate Temperature Distributions During Thermal EOR Process. In SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition. Society of Petroleum Engineers.

Hassan, A. M., Rammay, M. H., & Hossain, M. E. (2015). Memory-based diffusivity equation: a comprehensive study on variable rock and fluid properties. Journal of Nature Science and Sustainable Technology, 9(4), 699.

Hearn, C. L. (1969). Effect of latent heat content of injected steam in a steam drive. J. Pet. Technol.;(United States), 21.

Hilfer, R. (Ed.). (2000). Applications of fractional calculus in physics (Vol. 35, No. 12, pp. 87-130). Singapore: World scientific.

Hossain, M. E. (2016). Numerical investigation of memory-based diffusivity equation: the integro-differential equation. Arabian Journal for Science and Engineering, 41(7), 2715-2729.

Hossain, M. E. (2017). Role of porosity on energy transport with equal rock-fluid temperatures during thermal EOR process. Arabian Journal for Science and Engineering, 42(4), 1621-1631.

Hossain, M. E., & Abu-Khamsin, S. A. (2012). Utilization of memory concept to develop heat transfer dimensionless numbers for porous media undergoing thermal flooding with equal rock and fluid temperatures. Journal of Porous Media, 15(10).

Hossain, M. E., & Islam, M. R. (2006). Fluid properties with memory–A critical Review and some additions. In Proc. 36 th International Conference on Computers and Industrial Engineering, CIE–00778, Taipei, Taiwan (pp. 20-23).

Hossain, M. E., & Islam, M. R. (2009). A comprehensive material balance equation with the inclusion of memory during rock-fluid deformation. Adv. Sustainable Pet. Eng. Sci, 1(2), 141-162.

Hossain, M. E., Abu-Khamsin, S. A., & Al-Helali, A. A. (2011). Use of the memory concept to investigate the temperature profile during a thermal EOR process. In SPE/DGS Saudi Arabia Section Technical Symposium and Exhibition. Society of Petroleum Engineers.

Hossain, M. E., Abu-Khamsin, S. A., & Al-Helali, A. A. (2015). A mathematical model for thermal flooding with equal rock and fluid temperatures. Journal of Porous Media, 18(7).

Hossain, M. E., Mousavizadegan, S. H., & Islam, M. R. (2007). Rock and fluid temperature changes during thermal operations in EOR processes. Journal of nature science and sustainable technology, 2(3), 347-378.

Hossain, M. E., Mousavizadegan, S. H., & Islam, M. R. (2008b). A Novel Mathematical Tool for Characterizing Petroleum Fluid Rheology within Porous Media. In Mechanics of Time Dependent Materials Conference (MTDM2008). Monterey, California, USA.

Hossain, M. E., Mousavizadegan, S. H., & Islam, M. R. (2008c). The effects of thermal alterations on formation permeability and porosity. Petroleum science and technology, 26(10-11), 1282-1302.

Hossain, M. E., Mousavizadegan, S. H., & Islam, M. R. (2008d). A new porous media diffusivity equation with the inclusion of rock and fluid memories, SPE-114287-MS, Society of Petroleum Engineers.

Hossain, M. E., Mousavizadegan, S. H., & Islam, M. R. (2009a). Variation of rock and fluid temperature during thermal operation in porous media. Petroleum science and technology, 27(6), 597-611.

Hossain, M. E., Mousavizadegan, S. H., & Islam, M. R. (2009b). Modified engineering approach with the variation of permeability over time using the memory concept. In Proceedings of the Third International Conference on Modeling, Simulation and Applied Optimization Sharjah, UAE January (pp. 20-22).

Hossain, M. E., Mousavizadegan, S. H., Ketata, C., & Islam, M. R. (2008a). A novel memory based stress strain model for reservoir characterization. Nat Sci Sustainable Technol Res Prog, 1, 1-29.

Hossain, M. S., Sarica, C., Zhang, H.-Q., Rhyne, L., & Greenhill, K. L. (2005). Assessment and Development of Heavy Oil Viscosity Correlations. Society of Petroleum Engineers.

Hristov, J. (2013). A note on the integral approach to non-linear heat conduction with Jeffrey's fading memory. Thermal Science, 17(3), 733-737.

Huc, A. (2011). Heavy Crude Oils from Geology to Upgrading: An Overview, Editions Technip, Paris.

Iaffaldano, G., Caputo, M., & Martino, S. (2005). Experimental and theoretical memory diffusion of water in sand.

Irani, M., & Cokar, M. (2014, June). Understanding the Impact of Temperature-Dependent Thermal Conductivity on the Steam-Assisted Gravity-Drainage (SAGD) Process. Part

1: Temperature Front Prediction. In SPE Heavy Oil Conference-Canada. Society of Petroleum Engineers.

Irani, M., & Cokar, M. (2016). Discussion on the effects of temperature on thermal properties in the steam-assisted-gravity-drainage (SAGD) process. Part 1: thermal conductivity (includes associated Errata and Addendum). SPE Journal, 21(02), 334-352.

Irani, M., & Gates, I. D. (2013). Understanding the convection heat-transfer mechanism in steam-assisted-gravity-drainage process. SPE Journal, 18(06), 1-202.

Irani, M., & Ghannadi, S. (2013). Understanding the heat-transfer mechanism in the steam-assisted gravity-drainage (SAGD) process and comparing the conduction and convection flux in bitumen reservoirs. SPE Journal, 18(01), 134-145.

Ito, Y. (1998). Effect of reservoir parameter on oil rates and steam oil ratios in sagd projects. In Proc. 7th UNITAR International Conference on Heavy Crude and Tar Sands.

Ito, Y., & Suzuki, S. (1999). Numerical simulation of the SAGD process in the Hangingstone oil sands reservoir. Journal of Canadian Petroleum Technology, 38(09).

Jabbari, H., Afsari, K., Rabiei, M., & Monk, A. (2017a). Thermally-induced wettability alteration from hot-water imbibition in naturally fractured reservoirs—Part 1: Numerical model development & 1D models. Fuel, 208, 682-691.

Jabbari, H., Afsari, K., Rabiei, M., Monk, A., & Ostadhassan, M. (2017b). Thermally-induced wettability alteration from hot-water imbibition in naturally fractured reservoirs—Part 2: 2D models, sensitivity study & heavy oil. Fuel, 208, 692-700.

Jahanbani Ghahfarokhi, A., Jelmert, T. A., Kleppe, J., Ashrafi, M., Souraki, Y., & Torsaeter, O. (2012, January). Investigation of the applicability of thermal well test analysis in steam injection wells for Athabasca heavy oil. In SPE Europec/EAGE Annual Conference. Society of Petroleum Engineers.

Jelgersma, F. (2007, January). Redevelopment of the abandoned Dutch onshore Schoonebeek oilfield with gravity assisted steam flooding. In International Petroleum Technology Conference. International Petroleum Technology Conference.

Jensen, T. B., Sharma, M. P., & Harris, H. G. (1991). An improved evaluation model for steam-drive projects. Journal of Petroleum Science and Engineering, 5(4), 309-322.

Jones, J. (1981). Steam drive model for hand-held programmable calculators. Journal of Petroleum Technology, 33(09), 1-583.

Kang, P. S., Lim, J. S., & Huh, C. (2014). Screening criteria for application of EOR processes in offshore fields. In The Twenty-fourth International Ocean and Polar Engineering Conference. International Society of Offshore and Polar Engineers.

Kartoatmodjo, T. & Schmidt, Z. (1994). Large Data Bank Improves Crude Physical Property Correlations, Oil and Gas J., 51-55.

Kaviany, M. (2012). Principles of heat transfer in porous media. Springer Science & Business Media.

Kaye, S. E., Ting, V. C., & Fair, J. C. (1982). Development of a system to utilize flue gas from enhanced oil recovery combustion projects. Journal of Petroleum Technology, 34(01), 181-188.

Keenan, J. H., & Keyes, F. G. (1936). Thermodynamic Properties of Steam, First Edition, John Wiley and Sons Inc., New York City, 28-39.

Khan, M. M., Prior, D., & Islam, M. R. (2005). Direct-usage solar refrigeration: From irreversible thermodynamics to sustainable engineering. In Jordan International Chemical Engineering Conference V. September (pp. 12-14).

Kocabas, I. (2001). Modeling of Heat/Tracer Transport in Oil Reservoirs. Externally Funded Project Report IV submitted to Exploration and Production Division ADNOC, UAE Univ., Al Ain, UAE.

Kocabas, I. (2004). Thermal transients during nonisothermal fluid injection into oil reservoirs. Journal of Petroleum Science and Engineering, 42(2-4), 133-144.

Kocabas, I. (2006). An analytical model of temperature and stress fields during cold-water injection into an oil reservoir. SPE Production & Operations, 21(02), 282-292.

Kocabas, I., & Horne, R. N. (1990). A new method of forecasting the thermal breakthrough time during reinjection in geothermal reservoirs. In Proceedings, 15th Workshop on Geothermal Reservoir Engineering.

Krober, J., Codjovi, E., Kahn, O., Groliere, F., & Jay, C. (1993). A spin transition system with a thermal hysteresis at room temperature. Journal of the American Chemical Society, 115(21), 9810-9811.

Lacerda, G. D. M., Patriota, J. H., Pereira, J. I., de Lima, L. A. S., & Torres, J. (2008). Alto do Rodrigues GeDIg pilot-Case study for continuous steam injection recovery combined with Real Time Operation. In Intelligent Energy Conference and Exhibition. Society of Petroleum Engineers.

Lal, D., Otto, F. D., & Mather, A. E. (1999). Solubility of hydrogen in Athabasca bitumen. Fuel, 78(12), 1437-1441.

Larson, R. G. (1999). The structure and rheology of complex fluids (Vol. 150). New York: Oxford university press.

Lauriat, G., & Prasad, V. (1989). Non-Darcian effects on natural convection in a vertical porous enclosure. International Journal of Heat and Mass Transfer, 32(11), 2135-2148.

Lauwerier, H. A. (1955). The transport of heat in an oil layer caused by the injection of hot fluid. Applied Scientific Research, Section A, 5(2-3), 145-150.

Lawal, K. A. (2016). Zoning Steam-Heated Reservoirs by Heat-Transfer Mechanisms. In SPE Nigeria Annual International Conference and Exhibition. Society of Petroleum Engineers.

Lawal, K. A., & Vesovic, V. (2009). Analytic investigation of convection during conduction heating of a heavy-oil reservoir. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Lentz, W. (1991). Well Test Improvements at Shell Canada's Peace River Thermal Project. In Annual Technical Meeting. Petroleum Society of Canada.

Li, K. Y., Yang, S. Y., & Yeh, H. D. (2010). An analytical solution for describing the transient temperature distribution in an aquifer thermal energy storage system. Hydrological processes, 24(25), 3676-3688.

Li, P., & Chalaturnyk, R. J. (2006). Permeability variations associated with shearing and isotropic unloading during the SAGD process. Journal of Canadian Petroleum Technology, 45(01).

Li, Q., & Chen, Z. (2015). A new analysis on the convective heat transfer at the edge of the SAGD chamber. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Li, R., Wang, C., Wang, P., & Pei, J. (2016). Preparation of a novel flow improver and its viscosity-reducing effect on bitumen. Fuel, 181, 935-941.

Liang, L. (2005). An analytical model for cyclic steaming of horizontal wells. MS Report Stanford University. online: http://geothermal. stanford. edu/pereports/search. htm.

Malofeev, G. E., & Scheinman, A. B. (1963). The calculation of oil recovery from a stratum upon injecting hot water into it. Neft. Khoz, 41, 31-35.

Mandl, G., & Volek, C. W. (1967). Heat and mass transport in steam-drive processes. In Fall Meeting of the Society of Petroleum Engineers of AIME. Society of Petroleum Engineers.

Manrique, E. J., Muci, V. E., & Gurfinkel, M. E. (2007). EOR field experiences in carbonate reservoirs in the United States. SPE Reservoir Evaluation & Engineering, 10(06), 667-686.

Manrique, E., & Alvarado, V. (2010). Enhanced oil recovery: An update review. Energies, 3, 1529-1575.

Marx, J. W., & Langenheim, R. H. (1959). Reservoir heating by hot fluid injection. Pet. Trans. AIME. **216**, 1959, 312–315.

Messmer, J. H. (1984). The effective thermal conductivity of quartz sands and sandstones. SPE Reprint Series, Richardson, TX.

Metzler, R., & Klafter, J. (2004). The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics. Journal of Physics A: Mathematical and General, 37(31), R161.

Metzler, R., Barkai, E., & Klafter, J. (1999). Anomalous diffusion and relaxation close to thermal equilibrium: A fractional Fokker-Planck equation approach. Physical review letters, 82(18), 3563.

Mifflin, R. T., & Schowalter, W. R. (1986). A numerical technique for three-dimensional steady flows of fluids of the memory-integral type. Journal of Non-Newtonian Fluid Mechanics, 20, 323-337.

Miller, M.A., & Leung, W.K. (1985). A Simple Gravity Drainage Model of Steamdrive, paper SPE 14241 presented at 1985 SPE Annual Technical Conference and Exhibition, Las Vegas, Sept. 22-25.

Miura, K., & Wang, J. (2012). An analytical model to predict cumulative steam/oil ratio (CSOR) in thermal-recovery SAGD process. Journal of Canadian Petroleum Technology, 51(04), 268-275.

Mohammadmoradi, P., Behrang, A., & Kantzas, A. (2016). Effective thermal and electrical conductivity of two-phase saturated porous media. In SPE Canada Heavy Oil Technical Conference. Society of Petroleum Engineers.

Molenda, C. H. A., Crausse, P., & Lemarchand, D. (1993). Heat and humidity transfer in non saturated porous media: capillary hysteresis effects under cyclic thermal conditions. International journal of heat and mass transfer, 36(12), 3077-3088.

Moroni, M., & Cushman, J. H. (2001). Statistical mechanics with three-dimensional particle tracking velocimetry experiments in the study of anomalous dispersion. II. Experiments. Physics of Fluids, 13(1), 81-91.

Myhill, N. A., & Stegemeier, G. L. (1978). Steam-drive correlation and prediction. Journal of Petroleum Technology, 30(02), 173-182.

Nasr, T. N., & Ayodele, O. R. (2005). Thermal techniques for the recovery of heavy oil and bitumen. In SPE international improved oil recovery conference in Asia Pacific. Society of Petroleum Engineers.

Neuman, C. H. (1985). A gravity override model of steamdrive. Journal of petroleum technology, 37(01), 163-169.

Nian, Y. L., & Cheng, W. L. (2017). Insights into heat transport for thermal oil recovery. Journal of Petroleum Science and Engineering, 151, 507-521.

Nukhaev, M. T., Pimenov, V. P., Shandrygin, A., & Tertychnyi, V. V. (2006). A new analytical model for the SAGD production phase. In SPE annual technical conference and exhibition. Society of Petroleum Engineers.

Obembe, A. D., Abu-Khamsin, S. A., & Hossain, M. E. (2016a). A review of modeling thermal displacement processes in porous media. Arabian Journal for Science and Engineering, 41(12), 4719-4741.

Obembe, A. D., Abu-Khamsin, S. A., & Hossain, M. E. (2018). Anomalous effects during thermal displacement in porous media under non-local thermal equilibrium. Journal of Porous Media, 21(2).

Obembe, A. D., Hossain, M. E., & Abu-Khamsin, S. A. (2017b). Variable-order derivative time fractional diffusion model for heterogeneous porous media. Journal of Petroleum Science and Engineering, 152, 391-405.

Obembe, A. D., Hossain, M. E., Mustapha, K., & Abu-Khamsin, S. A. (2017a). A modified memory-based mathematical model describing fluid flow in porous media. Computers & Mathematics with Applications, 73(6), 1385-1402.

Ovalles, C., Vallejos, C., Vasquez, T., Martinis, J., Perez-Perez, A., Cotte, E., ... & Rodriguez, H. (2001, January). Extra-heavy crude oil downhole upgrading process using hydrogen donors under steam injection conditions. In SPE international thermal operations and heavy oil symposium. Society of Petroleum Engineers.

Phillips, C. R., Haidar, N. I., & Poon, Y. C. (1985). Kinetic models for the thermal cracking of athabasca bitumen: the effect of the sand matrix. Fuel, 64(5), 678-691.

Prats, M. (1969). The heat efficiency of thermal recovery processes. Journal of Petroleum Technology, 21(03), 323-332.

Prats, M. (2005). Thermal Recovery, SPE Monograph Series, Richardson, TX., V. 7.

Ramey Jr, H. J. (1959). Discussion of reservoir heating by hot fluid injection. Trans., AIME, 216, 364.

Ramlal, V. (2004). Enhanced oil recovery by steam flooding in a recent steam flood project, Cruse "E" Field, Trinidad (SPE-89411). In Proceedings of 14th SPE/DOE IOR Symposium, Tulsa, OK, USA, 17–21.

Rammay, M. H., Hossain, E. M., & Hassan, A. M. (2016). A comparative study of PVT fluid properties variations using memory-based diffusivity equation. J Nat Sci Sustainable Technol, 10(1).

Reis, J. C. (1992). A steam-assisted gravity drainage model for tar sands: linear geometry. Journal of Canadian Petroleum Technology, 31(10).

Rhee, S. W., & Doscher, T. M. (1980). A method for predicting oil recovery by steamflooding including the effects of distillation and gravity override. Society of Petroleum Engineers Journal, 20(04), 249-266.

Rivero, J. A., & Mamora, D. D. (2007). Oil Production Gains for Mature Steamflooded Oil Fields Using Propane as a Steam Additive and a Novel Smart Horizontal Producer. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Rubinshtein, L. I. (1959). The total heat losses in injection of a hot liquid into a stratum. Neft'i Gaz, 2(9), 41.

Sadrameli, S. M. (2015). Thermal/catalytic cracking of hydrocarbons for the production of olefins: A state-of-the-art review I: Thermal cracking review. Fuel, 140, 102-115.

Santos, J. E., Douglas Jr, J., Corberó, J., & Lovera, O. M. (1990). A model for wave propagation in a porous medium saturated by a two-phase fluid. The Journal of the Acoustical Society of America, 87(4), 1439-1448.

Satman, A. (1988). Solutions of heat, and fluid-flow problems in naturally fractured reservoirs: part 1-heat-flow problems. SPE production engineering, 3(04), 463-466.

Scott, G. R. (2002). Comparison of CSS and SAGD performance in the Clearwater formation at Cold Lake. In SPE International Thermal Operations and Heavy Oil Symposium and International Horizontal Well Technology Conference. Society of Petroleum Engineers.

Seto, A. C., & Bharatha, S. (1991). Thermal conductivity estimation from temperature logs. In SPE International Thermal Operations Symposium. Society of Petroleum Engineers.

Sharma, J., & Gates, I. D. (2011). Convection at the edge of a steam-assisted-gravity-drainage steam chamber. SPE Journal, 16(03), 503-512.

Sheng, J. J. (Ed.). (2013). Enhanced oil recovery field case studies, Chapter 16: Cyclic Steam Stimulation. Gulf Professional Publishing.

Simonson, C. J., Tao, Y. X., & Besant, R. W. (1993). Thermal hysteresis in fibrous insulation. International journal of heat and mass transfer, 36(18), 4433-4441.

Sochi, T. (2010). Modelling the flow of yield-stress fluids in porous media. Transport in Porous Media, 85(2), 489-503.

Sochi, T., & Blunt, M. J. (2008). Pore-scale network modeling of Ellis and Herschel–Bulkley fluids. Journal of Petroleum Science and Engineering, 60(2), 105-124.

Somerton, W. H. (1958). Some thermal characteristics of porous rocks. Society of Petroleum Engineers.

Somerton, W. H., & Boozer, G. D. (1960). Thermal characteristics of porous rocks at elevated temperatures. Journal of Petroleum Technology, 12(06), 77-81.

Somerton, W. H., Keese, J. A., & Chu, S. L. (1974). Thermal behavior of unconsolidated oil sands. Society of Petroleum Engineers Journal, 14(05), 513-521.

Sposito, G. (1980). General criteria for the validity of the Buckingham-Darcy flow law. Soil Science Society of America Journal, 44(6), 1159-1168.

Sprouse, B. P. (2010). Computational efficiency of fractional diffusion using adaptive time step memory and the potential application to neural glial networks (Doctoral dissertation, UC San Diego).

Stewart, L. D., & Udell, K. S. (1987). Mechanisms of residual oil displacement by steam injection. In SPE California Regional Meeting. Society of Petroleum Engineers.

Suggett, J., Gittins, S., & Youn, S. (2000). Christina Lake thermal project. In SPE/CIM International Conference on Horizontal Well Technology. Society of Petroleum Engineers.

Taber, J. J. & Martin, F. D. (1983). Technical screening guides for the enhanced recovery of oil. In SPE annual technical conference and exhibition. Society of Petroleum Engineers.

Taber, J. J., Martin, F. D., & Seright, R. S. (1997). EOR screening criteria revisited-Part 1: Introduction to screening criteria and enhanced recovery field projects. SPE reservoir engineering, 12(03), 189-198.

Thimm, H. F., Capeling, R. R., Stevens, J. P., Kisman, K. E., Ehlers, P. F., & Nosal, M. (1993). A statistical analysis of the early Peace River thermal project performance. Journal of Canadian Petroleum Technology, 32(01).

Thomas, G. W. (1967). Approximate methods for calculating the temperature distribution during hot fluid injection. Journal of Canadian Petroleum Technology, 6(04), 123-129.

Tikhomirov, V. M. (1968). Thermal Conductivity of Rock Samples and Its Relation to Liquid Saturation, Density, and Temperature. Neftyanoe Khozaistvo, 46(4).

Van der Knaap, W. (1993). Physical Aspects of Some Steam Injection Theories An Application. Society of Petroleum Engineers.

Van Lookeren, J. (1983). Calculation methods for linear and radial steam flow in oil reservoirs. Society of Petroleum Engineers Journal, 23(03), 427-439.

Vogel, J. V. (1984). Simplified heat calculations for steamfloods. Journal of petroleum technology, 36(07), 1-127.

Wang, J., Bryan, J. L., & Kantzas, A. (2008). Comparative Investigation of Thermal Processes for Marginal Bitumen Resources. In International Petroleum Technology Conference. International Petroleum Technology Conference.

Wang, X., & Sheng, J. J. (2017). Effect of low-velocity non-Darcy flow on well production performance in shale and tight oil reservoirs. Fuel, 190, 41-46.

Wei, S., Lin-Song, C., Huang, S., & Huang, W. (2014). Steam chamber development and production performance prediction of steam assisted gravity drainage. In SPE Heavy Oil Conference-Canada. Society of Petroleum Engineers.

Whitaker, S. (1986). Flow in porous media I: A theoretical derivation of Darcy's law. Transport in porous media, 1(1), 3-25.

Willman, B. T., Valleroy, V. V., Runberg, G. W., Cornelius, A. J., & Powers, L. W. (1961). Laboratory studies of oil recovery by steam injection. Journal of Petroleum Technology, 13(07), 681-690.

Yoshiki, K. S., & Phillips, C. R. (1985). Kinetics of the thermo-oxidative and thermal cracking reactions of Athabasca bitumen. Fuel, 64(11), 1591-1598.

Yu, K. (2014). Analytical Modeling of Transient Heat Transfer Coupled with Fluid Flow in Heavy Oil Reservoirs During Thermal Recovery Processes (Doctoral dissertation, Faculty of Graduate Studies and Research, University of Regina).

Yu, K., & Zhao, G. (2015). Integration of Conduction and Convection Heat Transfer for Temperature Transient Analysis during Thermal Recovery Processes. In Paper WHOC15-283, World Heavy Oil Congress (pp. 24-26).

Yu, K., & Zhao, G. (2016). Modeling of heat transfer coupled with fluid flow for temperature transient analysis during SAGD process. In SPE Latin America and Caribbean Heavy and Extra Heavy Oil Conference. Society of Petroleum Engineers.

Ziagos, J. P., & Blackwell, D. D. (1986). A model for the transient temperature effects of horizontal fluid flow in geothermal systems. Journal of Volcanology and Geothermal Research, 27(3-4), 371-397.