The Hybrid of Classification Tree and Extreme Learning Machine

for Permeability Prediction in Oil Reservoir

Thesis by

Chandra Prasetyo Utomo

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

King Abdullah University of Science and Technology

Thuwal, Kingdom of Saudi Arabia

June 2011

The thesis of Chandra Prasetyo Utomo is approved by examination committee.

Committee Chairperson: Mikhail Moshkov

Committee Member: Basem Shihada

Committee Member: Shuyu Sun

ABSTRACT

The Hybrid of Classification Tree and Extreme Learning Machine

for Permeability Prediction in Oil Reservoir

Chandra Prasetyo Utomo

Permeability is an important parameter connected with oil reservoir. Predicting the permeability could save millions of dollars. Unfortunately, petroleum engineers have faced numerous challenges arriving at cost-efficient predictions. Much work has been carried out to solve this problem. The main challenge is to handle the high range of permeability in each reservoir. For about a hundred year, mathematicians and engineers have tried to deliver best prediction models. However, none of them have produced satisfying results. In the last two decades, artificial intelligence models have been used. The current best prediction model in permeability prediction is extreme learning machine (ELM). It produces fairly good results but a clear explanation of the model is hard to come by because it is so complex. The aim of this research is to propose a way out of this complexity through the design of a hybrid intelligent model. In this proposal, the system combines classification and regression models to predict the permeability value. These are based on the well logs data. In order to handle the high range of the permeability value, a classification tree is utilized. A benefit of this innovation is that the tree represents knowledge in a clear and succinct fashion and thereby avoids the complexity of all previous models. Finally, it is important to note that the ELM is used as a final predictor. Results demonstrate that this proposed hybrid model performs better when

compared with support vector machines (SVM) and ELM in term of correlation coefficient. Moreover, the classification tree model potentially leads to better communication among petroleum engineers concerning this important process and has wider implications for oil reservoir management efficiency.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ANN	Artificial Neural Networks
CART	Classification And Regression Tree
ELM	Extreme Learning Machine
SVM	Support Vector Machines

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter I Introduction

This introductory chapter consists of three sections. The first section is problem statement which explains the importance of predicting permeability, the brief history of scientists' efforts in this problem, the state of the art and its weaknesses, the explanation of proposed model, and the contributions. The second section is research objective which states the purposes of this research. The final section is thesis structure which presents the organization of this report.

I.1 Problem Statement

Permeability is the flow capacity of fluid to be transmitted through a rock's pore space. According to the latest study in oil reservoir, millions of dollars can be saved or lost depending on the quality of permeability prediction. The information of permeability values in reservoirs is important because it is needed to find out the quantity of oil or gas exists in reservoirs, the quantity that can be retrieved, its flow rate, the prediction of future production, and the production facilities design. Based on that, correct knowledge of permeability is required for the whole reservoir management and development [1].

Conventional method used to obtain the permeability values is by taking rock samples in some depths then measuring its permeability in the laboratory. This method is very expensive, complex, and time consuming. In addition, laboratorial measurement is limited to the rock samples. So that, the continuous picture of permeability values can't be captured. Based on this reasons, a new method which is quite accurate, less expensive, simpler, faster, and able to deliver permeability distribution along the depth is highly needed.

A huge number of efforts have been carried out to obtain new method to predict permeability values from well log data. From 1927 to 1981, scientists had tried empirical models by delivering mathematical formulas to get permeability values. None of this formula gives satisfying result in general case. Since 1961, multiple variable regressions models had been applied. The distribution of predicted values gained from this model is

still far from actual values. However, empirical and regression models gave hint about factors controlling permeability [2].

In the past two decades, computational intelligent techniques, such as artificial neural networks (ANN), have been utilized in permeability prediction. An ANN is a powerful and flexible tool for many applications including in petroleum area. This model is able to learn from previous data in order to predict values from new data. It gives better performance than previous models in predicting permeability from well logs in new wells [3]. Nevertheless, back propagation neural network suffers some drawbacks. It has some tuning parameters such as number of hidden neurons, learning rate, and momentum so it needs more efforts to find the best model. In addition, the gradient based learning algorithm used by ANN makes the training process becomes time consuming.

Many works have been tried to develop new ANN model to solve its weaknesses. In 2004, Huang [4] proposed new learning algorithm for single-hidden layer feed forward neural networks which is called extreme learning machine (ELM). Both in theory and experimental results, this learning algorithm gives better generalization performances and extremely faster learning speed than traditional popular gradient based learning algorithm [5]. Based on that, ELM has been highly exploited in many applications including in petroleum engineering area. In comparison with support vector machines (SVM) and conventional ANN for predicting permeability from well log data, ELM gives better generalization ability and faster speed [1]. This result stated that ELM is the current best single model in permeability prediction problem.

Although ELM gives fairly good results and faster speed, it still has some limitations. First, ELM can't deal with high data distribution of permeability values. One of the main challenges in predicting permeability is high range of its values in each well [6]. There are a lot of very low permeability values and there are also a lot of very high permeability values. Second, ELM can't give knowledge representation of developed model. Because of its structure which is dense combination of simple computation, trained ELM is hard and complex to be written in mathematical formulas. As a result, it is impossible to

produce understandable knowledge representation which is needed to communicate with expert for future study and research.

In this research, a new hybrid intelligent model which can manage high data distribution and give knowledge representation is proposed. To deal with high range data, a single model is not enough. The data should be classified into low permeability and high permeability then applied different models to predict the value.

This proposed hybrid model is basically combination of classification and regression models. Classification model is responsible to classify the data into low and high permeability. On the other hand, regression models are responsible to give final prediction value of its associated data. Classification tree is utilized as classification model since it can produce knowledge representation which is close to human intuition. ELM is used as regression model since it is currently the best single model in permeability prediction.

The wells log data from Saudi Aramco in Middle East oil reservoir are used to test the performance of this proposed method. Results demonstrate that this proposed hybrid system performs better when compared with pure ELM and SVM in term of correlation coefficient. Moreover, the classification tree model potentially leads to better communication among petroleum engineers concerning this important process and has wider implications for oil reservoir management efficiency.

I.2 Research Objectives

The main objectives of this research are:

1. To address the limitations of ELM, the current single best Artificial Intelligence model, in permeability prediction.
2. To propose new model, the hybrid of classification tree and ELM, to overcome ELM's limitations.
3. To investigate whether the proposed model give better prediction performance than other models such as SVM and ELM.

4. To apply the proposed method in the real world wells log data and then show that the method can provide reasonable prediction accuracy and produce understandable knowledge representation.

I.3 Thesis Structure

The rest of this thesis is organized as the following. Chapter 2 is dedicated as literature review. In this chapter, the previous works in permeability prediction, the overview of classification tree, and ELM are presented. In Chapter 3, design and implementation of the proposed model are explained. In Chapter 4, experiments, results, and analysis are provided. Finally, conclusions and future works are given in Chapter 5.

Chapter II Literature Review

This Literature Review chapter is divided into three sections. The first section describes previous works in permeability prediction. The second section explains about Extreme Leaning Machine. The third section describes about Classification Tree.

II.1 Previous Works in Permeability Prediction

There are huge efforts from scientists and engineers in order to deliver best model to predict permeability values based on well logs data. This section describes previous works in permeability prediction which can be categorized into empirical models, multiple regression variable models, and artificial intelligence models.

II.1.1 Empirical Models

Empirical models are predicting permeability by defining mathematical formulas based on its correlation with some rock properties. Kozeny [2] introduced the first equation of permeability in 1927. He measured permeability as a function of empirical Kozeny constant, porosity, and surface area. The major limitations are the surface area can be computed only by core analysis with special tools and it is only valid for pack of uniformly sized spheres. Archie [7] established the concept of "formation resistivity factor" in 1941. His concept indirectly influenced the computation of permeability since it affected the way to calculate water saturation.

Tixier [8] proposed a formula in 1949 to determine permeability from resistivity gradients by using empirical correlation between resistivity and water saturation, water saturation and capillarity pressure, and capillarity pressure and permeability. In 1950, Wyllie & Rose [9] modified the formula proposed by Tixier. Their model is based on quantitative log interpretation theoretical analysis and some assumptions.

In 1956, Sheffield [10] delivered permeability formula based on Kozeny's equation and formation of a correlation coefficient for some water well-known water-wet sands. However, he recommended his formula is suitable only for clean sands. In 1963, Prison [10] proposed formula which was determined by multiple correlation from relatively few

data. For high gravity crudes (API > 40°) and for depths greater than 6500 ft, the formula must not be utilized.

Tabel 1. History of Proposed Empirical Models of Permeability Prediction

Author(s)	Year
Kozeny	1927
Archie	1941
Tixier	1949
Wyllie & Rose	1950
Sheffield	1956
Prison	1963
Timur	1968
Coates & Dumanoir	1974
Coates	1981

Timur [11] generalized permeability equation based on the work of Kozeny and Willy & Rose. His model is appropriate where condition of residual water saturation exists. In 1974, Coates & Dumanoir [12] proposed an improved empirical permeability formula which is satisfied the condition of zero permeability at zero porosity and when irreducible water saturation is 100%. Coates and Denoo [13] simplified the previous proposed formulas and still satisfied the zero permeability condition. However, the formation must be at irreducible water saturation.

II.1.2 Multiple Variable Regression Models

Multiple variable regression models are expansions of the regression analysis that include extra independent variables in the equation. The model can be generalized as:

$$Y = C_0 + C_1X_1 + C_2X_2 + \cdots + C_nX_n + e \tag{2.1}$$

where Y is the dependent variable, X_1, X_2,…, X_n are the independent variables, and e is a random error or residual. The regression coefficients C_1, C_2,…,C_n are the parameters to be approximated.

A general procedure of multiple variable regression for permeability prediction was established by Wendt and Sakurai [14] in 1986. The main drawback of using this model

is the predicted permeability values is narrower than the actual values. Kendall and Stuart [15] enlightened above phenomena by stating this model gave the best prediction on the average. Weighting the high and low values are applied to improve the capability of regression model to predict outlier data. However, this may turn the predictor into unstable and statistically biased. Pereira [16] reported that density, derivative of density, gamma ray, and derivative of gamma ray are the best combination to be utilized as independent variables in multiple regression analyses.

II.1.3 Artificial Intelligence Models

Artificial Intelligence (AI) is set of models inspired by nature such as neural networks, fuzzy logic, and genetic algorithm. A lot of neural networks applications can be found in the petroleum industry, from exploration, drilling exploration, to reservoir and production engineering [17]. In predicting permeability, neural networks gave significant improvement [18-21]. This opened the door of others AI models to be applied in the petroleum industry area especially in the permeability prediction problem.

The combination of two or more AI models is called hybrid model. It complements the weaknesses of one model with the advantage of others. Since neural networks is one of the best AI model, most of published hybrid model are neural network based model. There are some proposed hybrid models in permeability prediction. Deni [22] proposed a hybrid of genetic algorithm and fuzzy/neural network inference system. Helmi [23] developed a hybrid of fuzzy logic, support vector machine, and functional network. Karimpouli [6] built up supervise committee machine neural network. Li [24] enhanced decision tree learning approach for neural decision tree model.

Although previous hybrid model gave better results than single model, it has some drawbacks due to the limitation of neural networks model. As a "black box" model, neural networks cannot give clear relationships among variables. Other limitations are it can fall into local minima, need to adjust too many parameters, and time consuming. Huang [5] proposed extreme learning machine which is theoretically proved that it can tackle the neural networks' limitation. In permeability prediction problem, extreme

learning machine give better prediction results than neural networks and support vector machines [1, 25]. Some studies to develop understandable model in petroleum area have been done by utilizing decision tree [26-28].

The proposed hybrid model in this research is hybrid of classification tree and extreme learning machine. The details of classification tree and extreme learning machine are given in next sections. The explanation of design and implementation of proposed hybrid model is presented in the following chapter.

II.2 An Overview of Classification Tree

Classification tree is widely used tool in classification area. This section explains about classification problem and decision tree induction.

II.2.1 Classification Problem

Classification is the process of assigning class label t from given attribute set $\mathbf{x}$ by using a classification model [29]. The implementation of classification model is applied in many applications such as credit card fraud detection, face recognition, spam email detection, early disease diagnosis, and marketing campaign target customer selection.
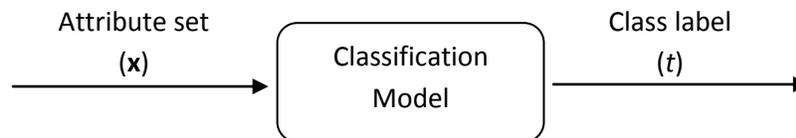
Figure 1. Classification Problem

Classification problem can be seen as the work of building classification model by using learning algorithm with given m rows of training dataset $D_{train} = \{\mathbf{x}^{(k)}, t^{(k)}, k = 1,\ldots,m\}$ such that the built classification model is able to predict the class of the new rows which is called testing dataset as accurate as possible.
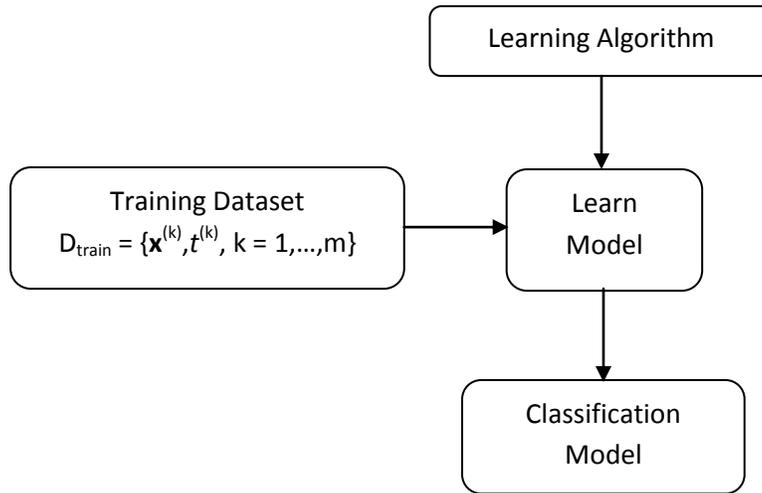
Figure 2. Builidng Classification Model

II.2.2 Decision Tree Induction

There are some classification models such as support vector machines, neural networks, decision tree, Bayesian methods, and logistic regression. Decision tree is one of the widely used models for inductive inference [30]. It produces classification tree as knowledge representation which is close to human intuition.
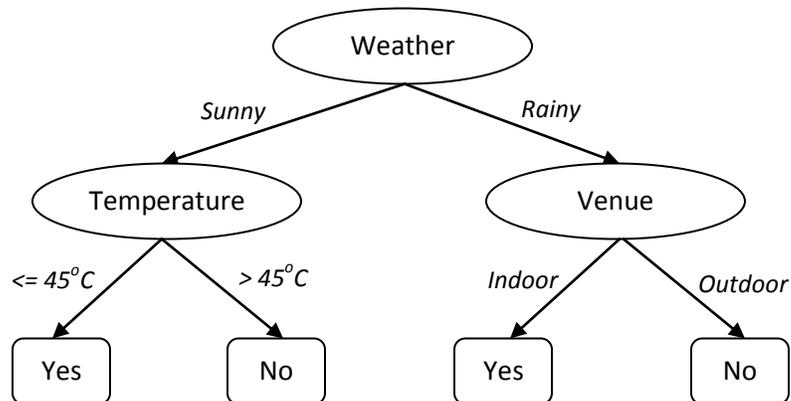
Figure 3. Example of Classification Tree for tennis match decision

The tree representation consists of three parts which are internal nodes, branches, and terminal nodes. Figure 3 presents an example of classification tree for tennis match. The decision whether a tennis match will be held or not can be predicted by some attributes such as weather, temperature, and venue. The internal nodes assessing the attributes are represented by ellipse. The branches associated with attribute's values are represented by arrows. The terminal nodes assigning the class labels or decisions are represented by rectangle.

The learning algorithm to build classification tree as classification model is decision tree induction. The previous proposed tree induction algorithms are Hunt's Algorithm [31], CART [32, 33], ID3 [34, 35], and C4.5 [36]. Classification and Regression Tree (CART) is one of the widely used and successful algorithms. The procedure of CART induction [33] can be defined as the following:

Step (1): Define the splitting criteria and stopping rules

Step (2): Find each attribute's best split

Step (3): Find the node's best spilt among all attribute's best split from step (2)

Step (4): If any of stopping rule is satisfied, then assign the node with class label. Otherwise, split the node using best split found in the step (3). Repeat step (2) – step (4) for all remaining non terminal nodes.

Splitting and stopping criteria are two important parameters for CART induction. The main purpose of node splitting is to reduce the impurity. Based on that, the first thing to be measured is node impurity. The best split is the one that achieves most reduction. Some common splitting criteria are Gini, Entropy, and Twoing.
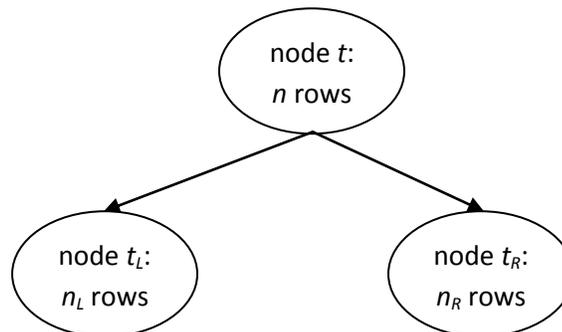
Figure 4. Binary splitting of CART

Let t_L be the left child and t_R be the right child of node t, $p(j|t)$ be the relative frequency of class j in node t, and n_i be number of rows in node t_i. The best split s is chosen to maximize the splitting criterion $\Delta i(s,t)$ which reduces the impurity $i(t)$.

Gini Criterion

The Gini impurity $i_{Gini}(t)$ is defined as

$$i_{Gini}(t) = 1 - \sum_j [p(j|t)]^2 \qquad (2.2)$$

The Gini splitting criterion $\Delta i_{Gini}(s,t)$ is defined as

$$\Delta i_{Gini}(s,t) = i_{Gini}(t) - \frac{n_L}{n} i_{Gini}(t_L) - \frac{n_R}{n} i_{Gini}(t_R) \qquad (2.3)$$

Entropy Criterion

The Entropy impurity $i_{Ent}(t)$ is defined as

$$i_{Ent}(t) = - \sum_j p(j|t) \log p(j|t) \qquad (2.4)$$

The Entropy splitting criterion $\Delta i_{Ent}(s,t)$ is defined as

$$\Delta i_{Ent}(s,t) = i_{Ent}(t) - \frac{n_L}{n} i_{Ent}(t_L) - \frac{n_R}{n} i_{Ent}(t_R) \qquad (2.5)$$

Twoing Criterion

The Twoing splitting criterion $\Delta i_{Two}(s,t)$ is defined as

$$\Delta i_{Two}(s,t) = \frac{n_L}{n} \frac{n_R}{n} \left[\sum_j |p(j|t_L) - p(j|t_R)| \right]^2 \qquad (2.6)$$

Another important parameter for CART induction is stopping criteria. This parameter determines whether the tree growing process must be continued or not. The stopping criteria could be maximum value of tree depth, minimum number of rows in a node, minimum number of row in a child node, and minimum value of splitting criterion for the best split.

II.3 An Overview of Extreme Learning Machine

Extreme learning machine is special kind of learning algorithm for artificial neural network. This section describes neural network structure, the popular conventional back propagation learning algorithm and its limitations. The Extreme learning machine algorithm and how it can handle the drawbacks of previous algorithm are explained in details.

II.3.1 Artificial Neural Networks Structure

Artificial neural networks (ANN) is widely used computational intelligence model in many applications including regression, classification, and clustering problems. It is also become popular option among petroleum prediction models [37]. ANN can learn from training dataset to approximate complex function which is often impossible to be generated by other methods. In training process, ANN is updating its internal structure such that at the end of this process, it will produce trained ANN which is able to predict new dataset. The high level abstraction of training process in ANN can be seen in the Figure 5.

Suppose we have training dataset with m vectors or samples. Each vector consists of n inputs and p target outputs. Formally, the training dataset can be defined as

$$\mathbf{D} = \{(\mathbf{x}^{(k)}, \mathbf{t}^{(k)}) \mid \mathbf{x}^{(k)} \in \mathbf{R}^n, \mathbf{t}^{(k)} \in \mathbf{R}^p, k = 1,..,m\}. \qquad (2.7)$$
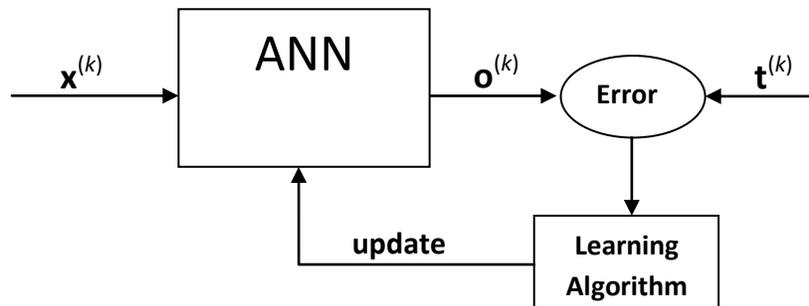
Figure 5. High level abstraction of training process in ANN

ANN consists of several layers. The first layer is input layer which takes n inputs from dataset. The last layer is output layer which produces p outputs. Layers between input layer and output layer are hidden layers.
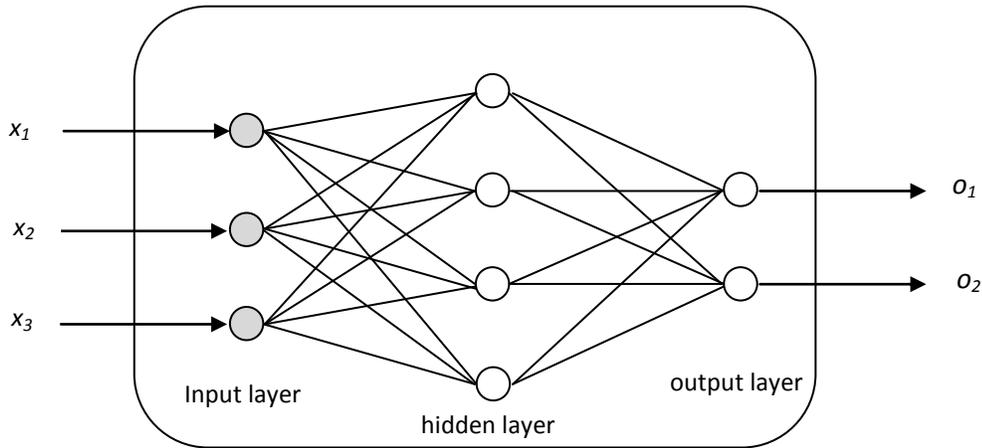
Figure 6. The example of single hidden layer feedforwad neural networks with four hidden neurons. This neural network receives three inputs and produces two outputs

In each layer, there are neurons. The number of neurons in the input layer is associated with the number of input which is n. The number of neurons in the output layer is associated with the number of output which is p. The number of neurons in the hidden layers can be set in any value. There are weights which are connecting any two neurons in the adjacency layers.
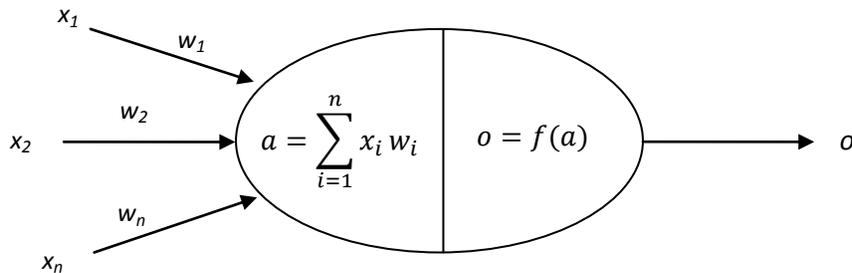
Figure 7. The neuron in the hidden and output layers of ANN

Each neuron in the hidden and output layers consists of two parts. The first part is summation and the second part is activation. In the first part, neuron receives input from all neurons in the previous layer. Each input is multiplied by its associated connecting

weight. After that, it sums the multiplications. In the second part, neuron uses activation function to produce neuron's output. The input for activation function is summation result from the first part.

II.3.2 Conventional Back Propagation Learning Algorithm

The ANN which produces output straight forward from input layer to hidden layer is called feedforward neural networks. The widely used learning algorithm for this kind of ANN is back propagation with gradient descent based learning algorithm.

Suppose we have m vectors in training dataset $\mathbf{D} = \{(\mathbf{x}^{(k)}, \mathbf{t}^{(k)}), k = 1,...,m \}$ where $\mathbf{x}^{(k)} \in \mathbf{R}^n$ is the input and $\mathbf{t}^{(k)} \in \mathbf{R}^p$ is the targets output. The training process of feedforward neural networks with back propagation gradient descent based learning algorithm is the following:

Step (1): Initialization

Initialize randomly all the weights and define some parameters such as learning rate, momentum, and termination criteria.

Step (2): Learning Process

(a) Choose one vector ($\mathbf{x}^{(k)}$, $\mathbf{t}^{(k)}$) from training dataset $\mathbf{D}$

(b) Feed forward the vector into the neural networks and compute the output $\mathbf{o}^{(k)}$

(c) Compute the error which is the difference between target $\mathbf{t}^{(k)}$ and output $\mathbf{o}^{(k)}$

(d) Update the weights by back propagation gradient descent based learning algorithm

(e) Repeat the Step (2a) until Step (2e) for the rest vectors from training dataset

Step (3): Evaluation

Check if one of the determined termination criteria is achieved. Termination criterion could be threshold of cumulative error, number of iterations, or changing of weights norm. If one of them is achieved, then the training process is done. Otherwise, repeat from the Step (2).

This kind of widely used neural networks has some drawbacks. First, the training process is time consuming because all the weights should be updated in every iteration. Second,

the gradient descent based learning algorithm does not guarantee to reach the optimal solution. There are often terminated in the local minima in many highly nonlinear problems. In addition, this algorithm only works for differentiable activation function. Third, several parameters should be determined intuitively in the beginning such as learning rate, momentum, and termination criteria so that there are extra effort to tune all of these parameters.

II.3.3 Extreme Learning Machine Learning Algorithm

A lot of works has been tried to resolve the drawbacks of ANN. Huang and Babri [38] proved that single hidden layer feedforward neural networks (SLFN) with at most m hidden nodes is able to approximate function for m distinct vectors in training dataset.

Let given m vectors in training dataset $\mathbf{D} = \{(\mathbf{x}^{(k)}, \mathbf{t}^{(k)}) \mid \mathbf{x}^{(k)} \in \mathbf{R}^n, \mathbf{t}^{(k)} \in \mathbf{R}^p, k = 1,..,m\}$ where $\mathbf{x}^{(k)} = [x_1^{(k)}, x_2^{(k)}, ..., x_n^{(k)}]^T$ dan $\mathbf{t}^{(k)} = [t_1^{(k)}, t_2^{(k)}, ..., t_p^{(k)}]^T$. A SLFN with M hidden nodes, activation function $g(x)$ in hidden nodes, and linear activation function in output nodes is mathematically modeled as:

$$\sum_{i=1}^{M} \beta_i g_i(\mathbf{x}^{(k)}) = \sum_{i=1}^{M} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}^{(k)} + b_i) = \mathbf{o}^{(k)},$$
$$k = 1, ..., m \tag{2.8}$$

where

$\mathbf{w}_i \in \mathbf{R}^n$ is the weights attached to the edge connecting input nodes and the i-th hidden node

$$\mathbf{w}_i = [w_{i1}, w_{i1}, \ldots, w_{in}]^T, \tag{2.9}$$

$\boldsymbol{\beta}_i \in \mathbf{R}^p$ is the weights attached to the edge connecting the i-th hidden node and the output nodes

$$\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{ip}]^T, \tag{2.10}$$

$\mathbf{w}_i \cdot \mathbf{x}^{(k)}$ is the inner product of $\mathbf{w}_i$ and $\mathbf{x}^{(k)}$,

b_i is the bias of the i-th hidden node,

$\mathbf{o}^{(k)} \in \mathbf{R}^p$ is the output of neural network for k-th vector.

The meaning of SLFN can approximate m vectors is there are exist $\mathbf{w}_i$, $\boldsymbol{\beta}_i$, and b_i, such that:

$$|| \mathbf{o}^{(k)} - \mathbf{t}^{(k)} || = 0 \tag{2.11}$$

$$\sum_{i=1}^{M} \beta_i g\left(\mathbf{w}_i \cdot \mathbf{x}^{(k)} + b_i \right) = \mathbf{t}^{(k)},$$

$$k = 1, \ldots, m \tag{2.12}$$

Those m equations can be written as:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T}, \tag{2.13}$$

where

$\mathbf{H} \in \mathbf{R}^{m \times M}$ is the hidden layer output matrix of the neural networks

$$\mathbf{H} = \begin{bmatrix} g(w_1 \bullet x^{(1)} + b_1) & \cdots & g(w_M \bullet x^{(1)} + b_M) \\ \vdots & \ddots & \vdots \\ g(w_1 \bullet x^{(m)} + b_1) & \cdots & g(w_M \bullet x^{(m)} + b_M) \end{bmatrix}, \tag{2.14}$$

$\boldsymbol{\beta} \in \mathbf{R}^{M \times p}$ is the weights connecting hidden layer and output layers

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_M^T \end{bmatrix}, \tag{2.15}$$

$\mathbf{T} \in \mathbf{R}^{m \times p}$ is the target values of m vectors in training dataset

$$\mathbf{T} = \begin{bmatrix} t^{(1)^T} \\ \vdots \\ t^{(m)^T} \end{bmatrix}, \tag{2.16}$$

In the conventional gradient descent based learning algorithm, weights $\mathbf{w}_i$ which is connecting the input layer and hidden layer and biases b_i in the hidden nodes are needed to be initialized and tuned in every iteration. This is the main factor which often makes training process of neural networks become time consuming and the trained model may not reach global minima.

Huang [5] proposed minimum norm least-squares solution of SLFN which doesn't need to tune those parameters. Training SLFN with fixed input weights $\mathbf{w}_i$ and the hidden layer biases b_i is similar to find a least square solution $\widehat{\boldsymbol{\beta}}$ of the linear system $\mathbf{H}\boldsymbol{\beta} = \mathbf{T}$:

$$\|\mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_M, \boldsymbol{b}_1, \dots, \boldsymbol{b}_M,)\widehat{\boldsymbol{\beta}} - \mathbf{T}\| = \min_{\boldsymbol{\beta}} \|\mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_M, \boldsymbol{b}_1, \dots, \boldsymbol{b}_M,)\boldsymbol{\beta} - \mathbf{T}\|. \quad (2.17)$$

The smallest norm least squares solution of the above linear system is

$$\widehat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{T} \qquad (2.18)$$

where $\mathbf{H}^{\dagger}$ is the *Moore-Penrose generalized inverse* of matrix H. This solution has three important properties which are minimum training error, smallest norm of weights, and unique solution which is $\widehat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{T}$.

The above minimum norm least-square solution for SLFN is called extreme learning machine (ELM). Let given m vectors in training dataset $\mathbf{D} = \{(\mathbf{x}^{(k)}, \mathbf{t}^{(k)}) \mid \mathbf{x}^{(k)} \in \mathbf{R}^n, \mathbf{t}^{(k)} \in \mathbf{R}^p, k = 1,..,m\}$, activation function $g(x)$, and number of hidden node M. The training process of ELM is the the following:

Step (1): Randomly set input-hidden layer weights $\mathbf{w}_i$ and bias b_i, $i = 1,\dots,M$.

Step (2): Compute the matrix of hidden layer output $\mathbf{H}$

Step (3): Compute the hidden-output layer weights $\widehat{\boldsymbol{\beta}}$ for $\widehat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{T}$
where $\mathbf{T} = [\mathbf{t}^{(1)},\dots, \mathbf{t}^{(m)}]$.

The comparison between conventional widely used neural networks and ELM is summarized in the Table 2.

Table 2. The comparison between Back Propagation ANN and ELM

No.	Points of Comparison	Back Propagation ANN	ELM
1.	Learning Algorithm	Gradient based learning	Minimum least-squares
2.	Training Parameters	Need to tuning: • Number of hidden nodes, • Learning rate, • Momentum, and • Termination criteria	Simple tuning-free algorithm. The only one to be defined is number of hidden nodes.

3.	Activation Function	Works only for differentiable functions	Not only works for differentiable functions but also many non differentiable functions
4.	Speed	Very slow especially in the large dataset. All of weights are updated in every iteration.	Extremely faster than BP ANN. Only three steps without any iteration.
5.	Result	Get trained model which has minimum training error. There is possibility to finish in the local minima.	Get trained model which has minimum training error and smallest norm of weight. Better generalization model and reach global minima.

Chapter III Design and Implementation Model

This chapter explains about the proposed model which is basically hybrid of classification tree and extreme learning machine. It is divided into two sections. The first section describes about the design model. The second section presents the implementation of the model.

III.1 Design Model

The main challenge in permeability prediction is high range of permeability. A single model is not enough to deal with that. The data should be classified into low permeability and high permeability then applied different model to predict the value.
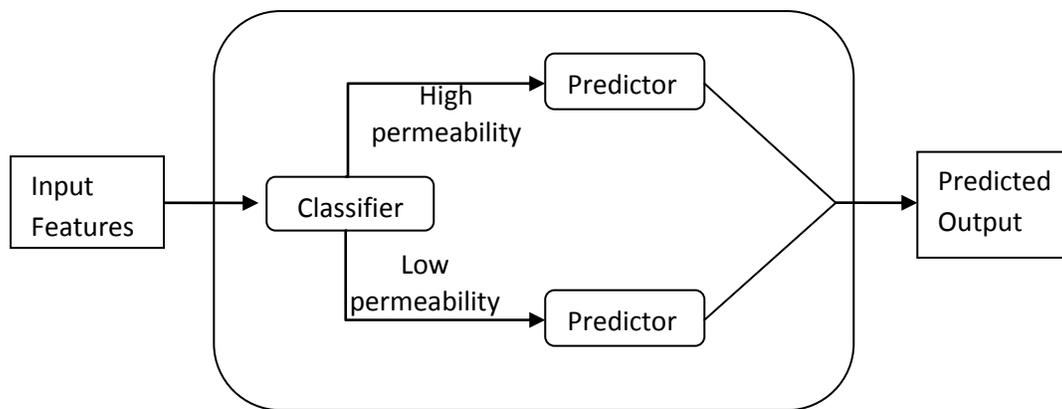
Figure 8. Design of the proposed hybrid model

Hybrid model which is basically combination of classification and regression models is proposed. Classification model is responsible to classify the data into low and high permeability based on a threshold value. On the other hand, regression models are responsible to give final prediction value of its associated data. Design of this model can be seen in the Figure 9.

III.2 Implementation Model

One of the objectives in this research is to propose new model which gives understandable knowledge representation. The best representation model which is close

to human reasoning is classification tree. For this reason, Classification Tree model is used in the classification part. Since Classification and Regression Tree (CART) from Salford System [39] is one of the best tools for classification tree design, it is implemented in this proposed model.

As presented in previous chapters, ELM is the current best single model in permeability prediction. ELM developed by Huang [40] is implemented in this proposed model as final predictor.
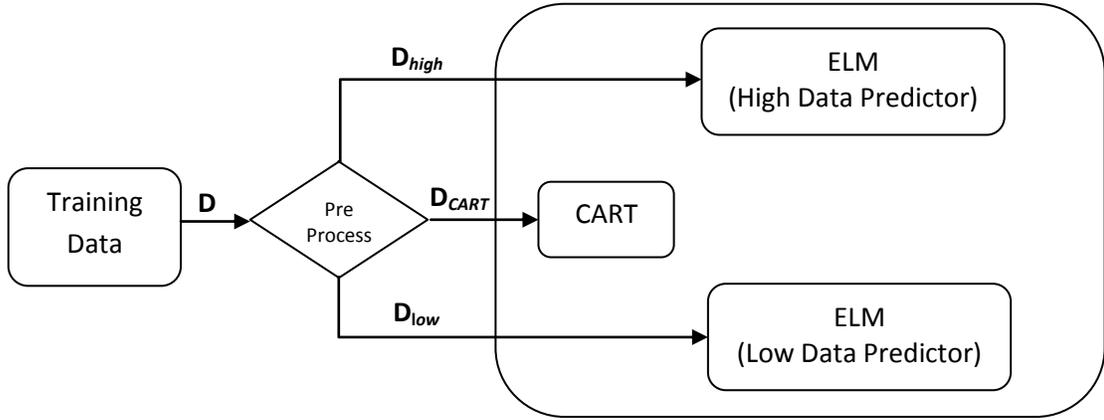

Figure 9. Training procedure of proposed hybrid model

Let we have *m* vectors in training dataset $\mathbf{D} = \{(\mathbf{x}^{(k)}, t^{(k)}) \mid \mathbf{x}^{(k)} \in \mathbf{R}^n, k = 1,..,m.\}$. The training algorithm of this hybrid model is designed as the following:

Step (1):   Add Discretized Target

Dicretize the target output $t^{(k)}$ into two classes *"low"* and *"high"* based on selected threshold value. The new training dataset is $\mathbf{D}_I = \{(\mathbf{x}^{(k)}, t^{(k)}, t_d^{(k)}) \mid \mathbf{x}^{(k)} \in \mathbf{R}^n, k = 1,..,m.\}$ with $t_d^{(k)}$ is "low" if $t^{(k)} \leq$ threshold, otherwise $t_d^{(k)}$ is "high".

Step (2):   Produce The Associated Training Data

In this step, three training dataset $\mathbf{D}_{CART}$, $\mathbf{D}_{low}$, $\mathbf{D}_{high}$ are produced. The training dataset for CART $\mathbf{D}_{CART}$ is $\mathbf{D}_I$ without original target value $t^{(k)}$. The vector ($\mathbf{x}^{(k)}$, $t^{(k)}$, $t_d^{(k)}$) in $\mathbf{D}_I$ is putted into $\mathbf{D}_{low}$ if $t_d^{(k)} =$ "low", otherwise it is putted into $\mathbf{D}_{high}$. The $t_d^{(k)}$ element in the $\mathbf{D}_{low}$ and $\mathbf{D}_{high}$ are removed at the end of this step.

Step (3):   Train the CART

Train the CART by training dataset $\mathbf{D}_{CART} = \{(\mathbf{x}^{(k)}, t_d^{(k)}) \mid \mathbf{x}^{(k)} \in \mathbf{R}^n, k = 1,..,m.\}$

Step (4):   Train the ELMs

Train the low ELM by training dataset $\mathbf{D}_{low}$ $\{(\mathbf{x}^{(l)}, t^{(l)}) \mid \mathbf{x}^{(l)} \in \mathbf{R}^n, k = 1,..,y.\}$

Train the high ELM by training dataset $\mathbf{D}_{high}$ $\{(\mathbf{x}^{(h)}, t^{(h)}) \mid \mathbf{x}^{(h)} \in \mathbf{R}^n, k = 1,..,z.\}$

After finish four steps above, the trained hybrid model is produced and ready to predict permeability from new dataset.
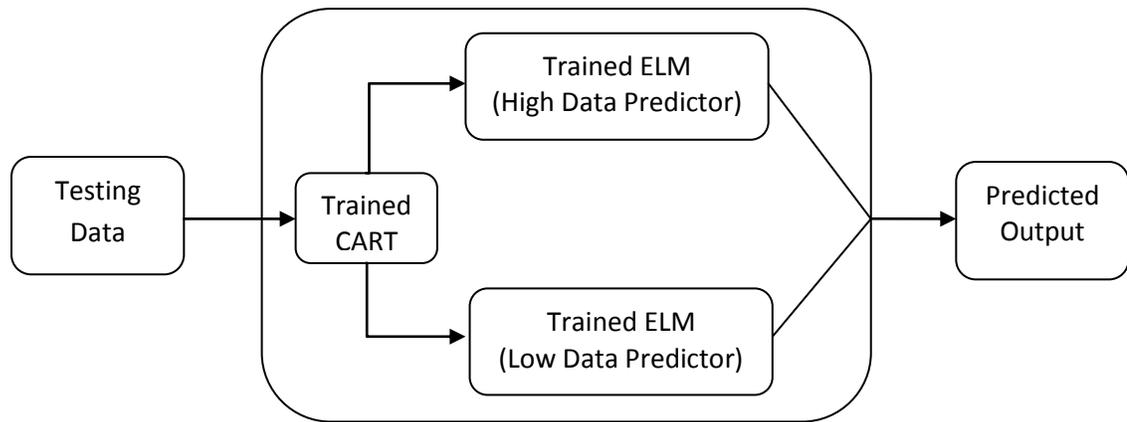

Figure 10. Testing procedure of trained model

Chapter IV Experiments, Results, and Analysis

This chapter presents experiments, results, and analysis. It consists of two sections. The first section describes design of experiments. The second section provides results and analysis.

IV.1 Design of Experiments

The data used in this experiment are 5 well logs data from Saudi Aramco given by instructor of Machine Learning course in KAUST. Data for Well 1 has 145 rows (vectors), for Well 2 has 141 rows, for Well 3 has 193 rows, for Well 4 has 147 rows, and for Well 5 has 141 rows. There are 5 input variables which are DT (sonic travel time), GR (Gamma Ray), PHIE (Effective Porosity), RHOB (Density), and SWT (Water Saturation). The target output to be predicted is PERM (Permeability).

Two kinds of experiments are conducted in this research. In the first experiment, one well is chosen as tested well and the rest wells are used to train the model. Because there are 5 wells, this experiment is repeated up to 5 times with different combination of training and testing wells. In the second experiment, all data are combined then divided randomly into training and testing data with ratio 80:20. The training data is used to train the model. Then, the trained model is tested by testing data to predict the permeability values.

The input features are normalized into [-1,1] and the output target is kept in the original value. The threshold used in this experiment is 1. This means, if the permeability value is less or equal than 1, then it is considered as low permeability. Otherwise, it is high permeability. A number of experiments had been tried to get the best parameters combination of CART such as in splitting criteria, stopping conditions, and thresholds.

Both classification and final prediction performance will be measured. The performance measurements for classification are Accuracy (ACC), True Positive Rate (TPR), and False Positive Rate (FPR).

Tabel 3. Confusion Matrix

Predicted Label		Actual Label	
		+1 (high)	-1 (low)
	+1 (high)	TP	FP
	-1 (low)	FN	TN

True Positive (TP)   : high permeability predicted as high permeability

False Positive (FP)   : low permeability predicted as high permeability

False Negative (FN)   : high permeability predicted as low permeability

True Negative (TN)   : low permeability predicted as low permeability

$$Accuracy\ (ACC) = \frac{TP+TN}{TP+FP+FN+TN} \tag{4.1}$$

$$True\ Positive\ Rate\ (TPR) = \frac{TP}{TP+FN} \tag{4.2}$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP+TN} \tag{4.3}$$

In order to measure the performance of the whole model, Root Mean Square Error (RMSE) and Correlation Coefficient (R) are used as performance criteria. The proposed model will be compared with SVM [41] and ELM based on this performance criteria. The RMSE is computed by taking the square root of the averages of the square differences between each actual permeability x_i and associated predicted permeability value y_i. The lower the RMSE, the better the model. The formula to compute RMSE is defined as follow:

$$RMSE = \sqrt{\frac{(x_1-y_1)^2+(x_2-y_2)^2+...+(x_n-y_n)^2}{n}} \tag{4.4}$$

The parameter R measures statistical correlation between actual and predicted values. Its values range from -1 to 1 inclusive. A value of "1" means perfectly correlated, a value "0" means no correlation, and a value "-1" means perfectly uncorrelated. The higher the R, the better the model. The formula to compute R is defined as follow:

$$R = \frac{\Sigma(x-x')(y-y')}{\sqrt{\Sigma(x-x')^2\ \Sigma(y-y')^2}} \tag{4.5}$$

where *x* and *y* are the actual and predicted values, and *x'* and *y'* are mean of the actual and predicted values.

IV.2 Results and Analysis

ELM assigns randomly input weights and biases in the first step of execution. To reduce the influence of random generator, 10 sequences of executions are applied in each model and the average results are obtained.

Table 4. The performances of CART as classifier

Tested Well	TP	TN	FP	FN	TPR	FPR	ACC
1	30	63	46	6	0.8333	0.4220	0.6414
2	35	94	5	7	0.8333	0.0693	0.9148
3	48	79	31	35	0.5783	0.2818	0.6500
4	15	89	3	40	0.2727	0.0326	0.7075
5	34	63	5	39	0.4658	0.0735	0.6879

Table 5. The performances comparison of models

Tested Well	RMSE			R		
	SVM	ELM	CART+ELM	SVM	ELM	CART+ELM
1	7.8782	9.7741	12.245	0.5564	0.4432	0.4427
2	19.4780	14.489	13.987	0.6793	0.7756	0.7330
3	16.8240	15.528	15.294	0.3803	0.3954	0.4219
4	9.3820	8.5143	9.5150	0.4038	0.4479	0.3590
5	10.4002	8.4206	9.6068	0.3802	0.4702	0.4793

The performances of CART as classifier to classify the high and low permeability data are shown in Table 4. These performances are obtained after tree pruning. When there is no pruning mechanism in classification tree induction, the classifier testing performances are bad and the final predictions of hybrid model are not reliable. Table 5 shows that the performances of proposed model are similar with current single best prediction model in permeability prediction.

The comparison of models based on RMSE can be clearly seen in Figure 11. Except the models for tested Well 1, SVM models give the highest errors. The proposed models are better than ELMs in tested Wells 2 and Well 3.
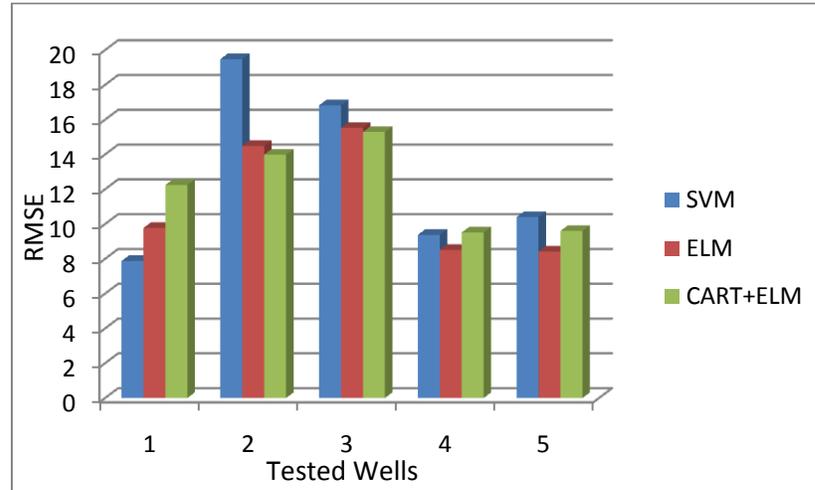

Figure 11. The performances comparison of models in each Well based on RMSE

Figure 12 shows the comparison of models based on Correlation Coefficient R. SVMs give the worst performances in tested Wells 2, 3, and 5. The proposed models are better than ELMs in tested Well 3 and 5, worse in tested Wells 2 and 5, and almost equal in tested Well 1.
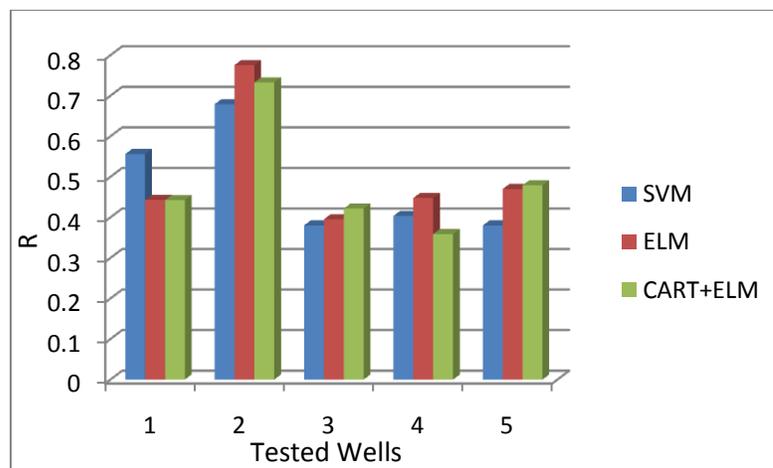

Figure 12. The performances comparison of models in each Well based on R

The performances results of the second experiment which is randomly divided data into training and testing data can be seen in table 6. In term of RMSE, the proposed model is worse than SVM and ELM. In term of R, the proposed model is better than SVM and ELM.

Table 6. The performances comparison of models in general Wells

Model	RMSE	R
SVM	12.88770	0.20670
ELM	12.49098	0.23453
CART + ELM	13.07807	0.26898

Another way to see differences of prediction is by looking the plot of actual and predicted values. Figure 13 gives the permeability data plot of actual value and predicted value by ELM and proposed model. This figure shows that the proposed model can handle high distribution data and predict accurately the low permeability values. However, it is still not good enough to predict the high permeability values.
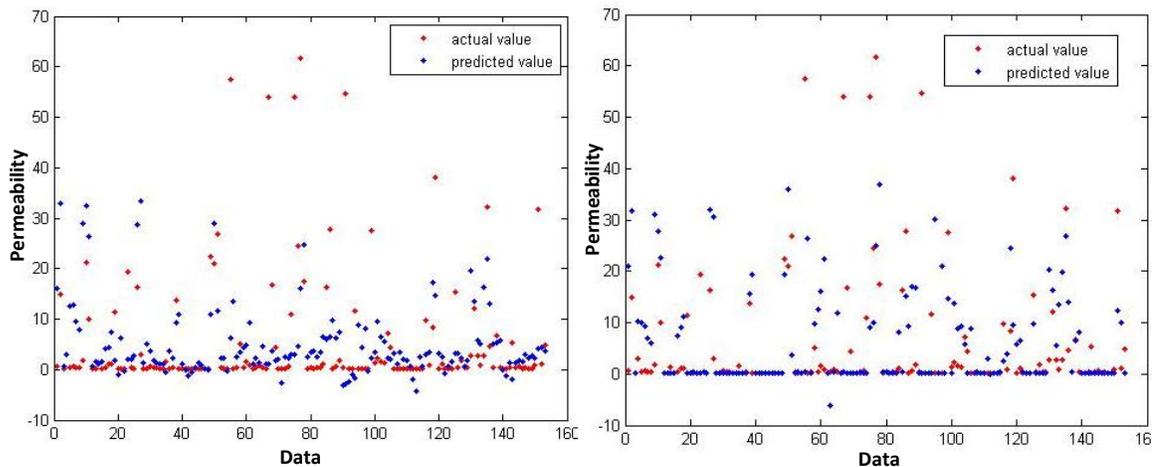

Figure13. Plotting permeability data of actual values and predicted values by ELM (left) and CART+ELM (right)

One of the most important objectives in this research is deliver knowledge representation. The classification tree is produced in the classification part. The classification tree produced in the second experiment can be seen in the Figure 14 (other trees can be found in the appendices). This tree is simple and understandable. Some rules connected with relationship between permeability and the predictors can be drawn. It can be used to communicate with experts and researchers in domain problem.
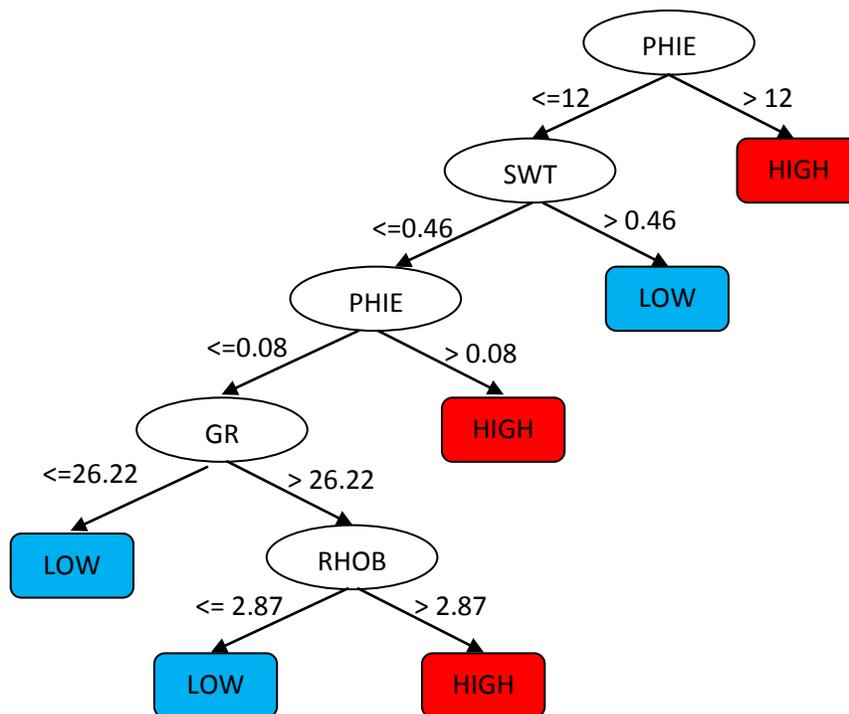

Figure 14. Classification tree generated by CART in the classification part

Chapter V Conclusions

V.1 Conclusions

Based on the results and analysis of the experiments, some conclusions can be drawn.

1. The proposed hybrid model, which is combination of Classification Tree as classifier and ELM as predictor, gives better performance than SVM and ELM in term of correlation coefficient in general Wells. The prediction in low permeability data is excellent but still not good enough in high permeability data.

2. The classification part plays important role in determining the prediction. The better accuracy of classifier, the better result in final prediction.

3. The classification tree produced by this hybrid model is simple and understandable. This means, it will be promising tool to be widely used to communicate with domain expert.

V.2 Future Work

Although the proposed model just gave small improvement, it concludes that the use of hybrid model in this way is in the right direction. The future work will be improvement in both classification and regression parts of this hybrid model. It is interesting to see how performance of classification tree with others induction tree algorithms such as Hunt's algorithm, ID3, and C4.5. It is also necessary to investigate different possible hybrid models which combine classification tree with other regression models such as support vector regressions and fuzzy systems.

REFERENCES

[1]   S. O. Olatunji, *et al.*, "Modeling Permeability Prediction Using Extreme Learning Machines," in *Mathematical/Analytical Modelling and Computer Simulation (AMS), 2010 Fourth Asia International Conference on*, 2010, pp. 29-33.

[2]   B. Balan, *et al.*, "State-Of-The-Art in Permeability Determination From Well Log Data: Part 1- A Comparative Study, Model Development," in *SPE Eastern Regional Conference & Exhibition*, Morgantown, West Virginia, USA, 1995.

[3]   S. Mohaghegh, *et al.*, "State-Of-The-Art in Permeability Determination From Well Log Data: Part2- Verifiable, Accurate Permeability Prediction, The Touch-Stone of All Models," in *SPE Eastern Regional Conference & Exhibition*, Morgatown, West Virginia, USA, 1995.

[4]   H. Guang-Bin, *et al.*, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, 2004, pp. 985-990 vol.2.

[5]   G.-B. Huang, *et al.*, "Extreme learning machine: Theory and applications," *Neurocomputing,* vol. 70, pp. 489-501, 2006.

[6]   S. Karimpouli, *et al.*, "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,* vol. 73, pp. 227-232, 2010.

[7]   G. E. Archie, "The Electrical Resistivity Log as an Aid in Determining Some Reservoir Charachteristics," *Trans., AIME,* vol. 146, pp. 54-62, 1942.

[8]   M. P. Tixier, "Evaluation of Permeability From Electric-Log Resistivity Gradients," *Oil and Gas Journal,* p. 113, June 1949.

[9]   M. R. J. Wyllie and W. D. Rose, "Some Theoretical Considerations Related to the Quantitative Evaluation of the Physical Characteristics of Reservoir Rock from Electric Log Data," *Trans., AIME,* vol. 189, p. 105, 1950.

[10]   S. J. Pirson, *Handbook of Well Log Analysis*: Englewood Cliffs, N.J., Prentice-Hall, Inc., 1963.

[11]   A. Timur, "An Investigation of Permeability, Porosity, and Residual Water Saturation Relationship for Sandstone Reservoirs," *The Log Analyst,* vol. 9, p. 8, July-August 1968.

[12]   G. R. Coats and J. L. Dumanoir, "A New Approach to Improved Log-Derived Permeability," *The Log Analyst,* January-February 1974.

[13]   SchlumbergerLtd., *Log Interpretation Charts*. Houston, Texas, 1987.

[14] W. A. Wendt, *et al.*, Eds., *Permeability Prediction From Well Logs Using Multiple Regression* (Reservoir Characterization. New York: Academic Press, 1986.

[15] N. R. Draper and H. Smith, *Applied Regression Analysis*: Wiley, 1981.

[16] J. L. L. Pereira, "Permeability prediction from well log data using multiple regression analysis," M.S.PNGE. 1424328, West Virginia University, United States -- West Virginia, 2004.

[17] M. P. McCormak, "Neural Networks in the Petroleum Industry," in *Society of Exploration Geophysicists 1991 Technical Program*, 1991, pp. 285-289.

[18] S. Mohaghegh, *et al.*, "A Methodological Approach for Reservoir Heterogeneity Characterization Using Artificial Neural Networks," presented at the SPE Annual Technical Conference, New Orleans, LA, September 1994.

[19] S. Mohaghegh, *et al.*, "Design and Development of An Artificial Neural Network for Estimation of Formation Permeability," presented at the SPE Petroleum Computer Conference, Dallas, Texas, August 1994.

[20] D. A. Osborne, "Permeability Estimation Using a Neural Network: A Case Study from The Roberts Unit, Wasson Field, Yoakum County, Texas," *AAPG South West Section Transactions,* pp. 125-132, 1992.

[21] J. M. Wiener, "Predicting Carbonate Permeabilities from Wireline Logs Using a Back-Propagation Neural Network," in *Society of Exploration Geophysicists 1991 Technical Program*, 1991.

[22] X. Deny, *et al.*, "Permeability Estimation Using Hybrid Genetic Programming and Fuzzy/Neural Inference Approach," presented at the Society of Petroleum Engineers Annual Technical Conference and Exhibition, Dallas, Texas, USA, 9-12 October 2005.

[23] T. Helmy, *et al.*, "Hybrid computational models for the characterization of oil and gas reservoirs," *Expert Systems with Applications,* vol. 37, pp. 5353-5363, 2010.

[24] X. Li and C. W. Chan, "Application of an enhanced decision tree learning approach for prediction of petroleum production," *Engineering Applications of Artificial Intelligence,* vol. 23, pp. 102-109, 2010.

[25] C. Guo-Jian, *et al.*, "Comparison of Extreme Learning Machine with Support Vector Regression for Reservoir Permeability Prediction," in *Computational Intelligence and Security, 2009. CIS '09. International Conference on*, 2009, pp. 173-176.

[26] T. Jensen, "Estimation of production forecast uncertainty for a mature production license," 1998.

[27]   I. Agbon*, et al.*, "Fuzzy Ranking of Gas Exploitation Opportunities in Mature Oil Fields in Eastern Venezuela," 2003.

[28]   H. H. Perez*, et al.*, "The Role of Electrofacies, Lithofacies, and Hydraulic Flow Unit in Permeability Prediction From Well Logs: A Comparative Analysis Using Classification Trees," *SPE Reservoir Evaluation and Engineering,* vol. 8, pp. 143-155, 2005.

[29]   P.-N. Tan*, et al.*, *Introduction to Data Mining*: Pearson Addison Wesley, 2006.

[30]   T. M. Mitchell, *Machine Learning*: The MIT Press, 1997.

[31]   E. B. Hunt*, et al.*, *Experiments in Induction*. New York: Academic Press, 1966.

[32]   J. H. Friedman, "A recursive partitioning decision rule for non-parametric classification," *IEEE Transaction on Computers,* pp. 404-408, 1977.

[33]   L. Breiman*, et al.*, *Classification and regression tree*. Belmont, CA: Wadsworth International Group, 1984.

[34]   J. R. Quinlan, "Learning efficient classification procedures and their application to chess end games," *Machine learning: An artificial intelligence approach,* vol. 1, pp. 463-482, 1983.

[35]   J. R. Quinlan, "Discovering rules by induction from large collections of examples," *Expert Systems,* 1979.

[36]   J. R. Quinlan, *C4. 5: programs for machine learning*: Morgan Kaufmann, 1993.

[37]   H. H. Nguyen and C. W. Chan, "Applications of data analysis techniques for oil production prediction," *Engineering Applications of Artificial Intelligence,* vol. 18, pp. 549-558, 2005.

[38]   H. Guang-Bin and H. A. Babri, "Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions," *Neural Networks, IEEE Transactions on,* vol. 9, pp. 224-229, 1998.

[39]   S. System, "Salford Predictive Modeler Builder v6.6," ed. San Diego, California: Salford System, 2011.

[40]   G.-B. Huang, "ELM Source Codes: http://www.ntu.edu.sg/home/egbhuang/," ed, 2011.

[41]   C.-C. Chang and C.-J. Lin, " LIBSVM : a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology,* vol. 2, pp. 27:1--27:27, 2011.

APPENDICES
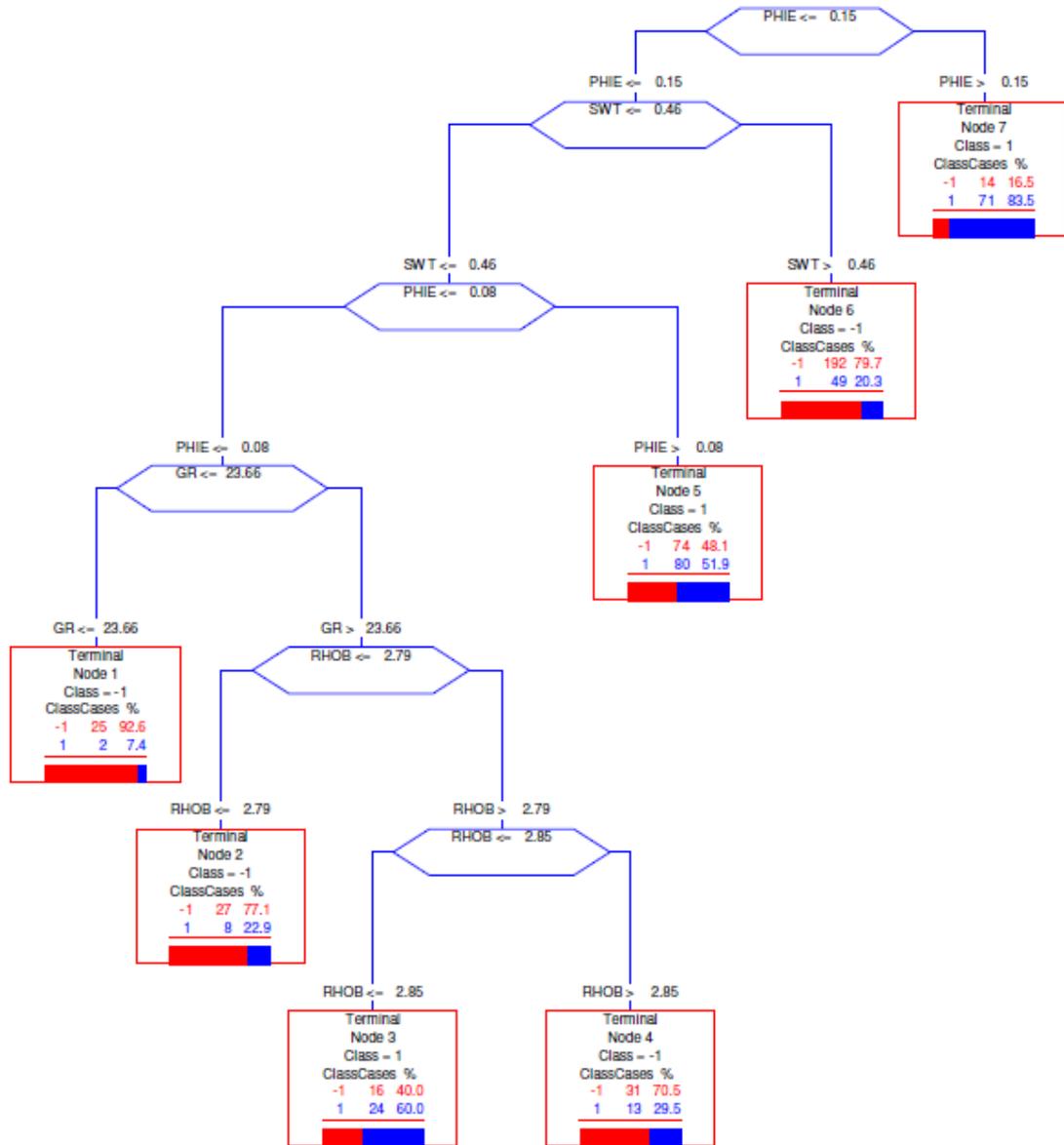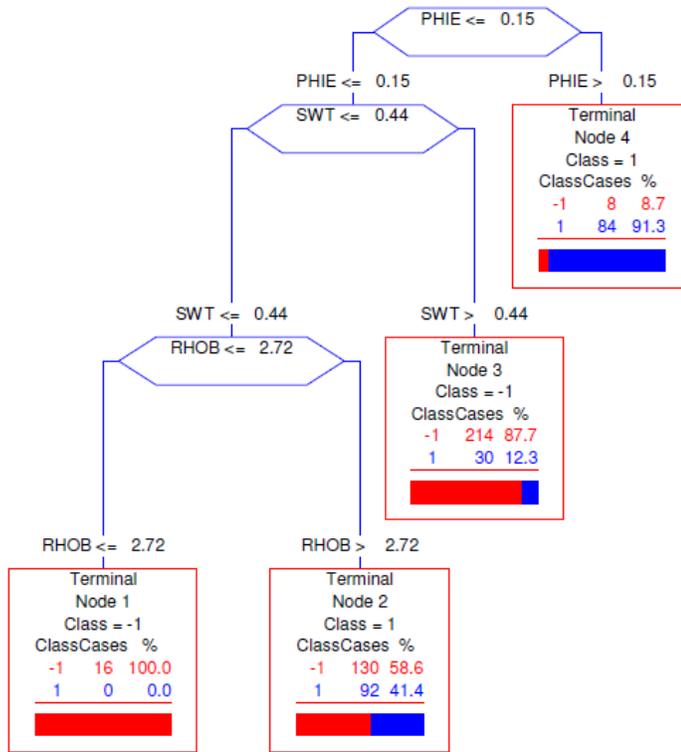
Appendix 1. Classification Tree for Tested Well 1
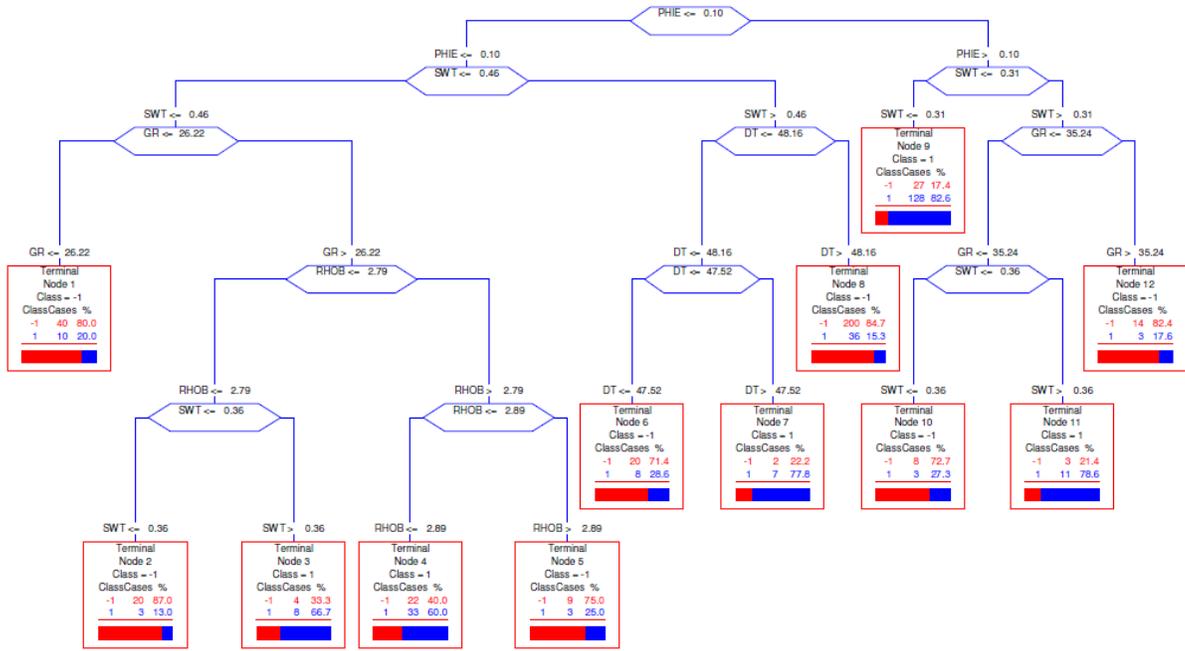
Notes:

 : low permeability class (-1)

 : high permeability class (+1)
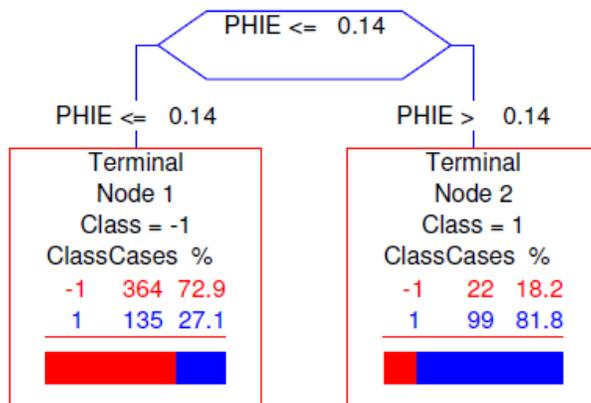
Appendix 2. Classification Tree for Tested Well 2

PHIE <= 0.15

PHIE <= 0.15
SWT <= 0.46

PHIE > 0.15
Terminal
Node 7
Class = 1
ClassCases %
-1   14   16.5
 1   71   83.5

SWT <= 0.46
PHIE <= 0.08

SWT > 0.46
Terminal
Node 6
Class = -1
ClassCases %
-1   192   79.7
 1   49   20.3

PHIE <= 0.08
GR <= 23.66

PHIE > 0.08
Terminal
Node 5
Class = 1
ClassCases %
-1   74   48.1
 1   80   51.9

GR <= 23.66
Terminal
Node 1
Class = -1
ClassCases %
-1   25   92.6
 1   2   7.4

GR > 23.66
RHOB <= 2.79

RHOB <= 2.79
Terminal
Node 2
Class = -1
ClassCases %
-1   27   77.1
 1   8   22.9

RHOB > 2.79
RHOB <= 2.85

RHOB <= 2.85
Terminal
Node 3
Class = 1
ClassCases %
-1   16   40.0
 1   24   60.0

RHOB > 2.85
Terminal
Node 4
Class = -1
ClassCases %
-1   31   70.5
 1   13   29.5

Appendix 3. Classification Tree for Tested Well 3

Appendix 4. Classification Tree for Tested Well 4

Appendix 5. Classification Tree for Tested Well 5