Implementation of Objects Recognition in Seismic Image via Artificial Neural Network (ANN)

By

Wan Chin Ee

Dissertation submitted in partial fulfillment of
the requirements for the
Bachelor of Technology (Hons)
(Business Information System)

SEPTEMBER 2012

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledges, and that the original work contained herein have not been undertaken or done unspecified sources or persons.

WAN CHIN EE

ABSTRACT

Seismic image processing is necessary in oil and gas exploration to identify the existence of potential reservoir by classifying the seismic image into different sections. These sections, also known as objects made up of different patterns which portraying the structure of subsurface. This project aims to develop a data mining algorithm embedded in a system that has ability to recognize the objects of channel and fault in seismic image. The method chosen is artificial neural network (ANN) which consists of input layer, hidden layer and output layer. Each layer is made up of numbers of neuron nodes to receive input data from preceding layers and output value to next layer until final output is determined from output layer. The ANN is trained and tested via MATLAB Neural Network Pattern Recognition Toolbox (nprtool) and MATLAB Neural Network Toolbox (nntool). 2-dimension (2D) seismic image is converted into gray scale image via MATLAB Image Processing Toolbox (imtool) and Grey-level co-occurrence matrix (GLCM) which serve as input to the ANN is retrieved from the gray scale image. Result is displayed by the system informing user whether the input image is channel, fault or neither both.

ACKNOWLEDGEMENT

First and foremost, I would like to take this opportunity to acknowledge utmost gratitude to my Final Year Project supervisor, Dr. Vivian Yong Suet Peng, who has been given abundance of constructive advices as well as guidance until completion of this project. Her willingness to spend her time so generously has been very much appreciated.

Secondly, I am grateful for the assistance given by Research and Innovation Office of Universiti Teknologi PETRONAS for booking me a desktop to do seismic image processing. Not forgotten to thank Master and PhD students from Faculty of Geosciences and Petroleum Engineering who have been providing me number of seismic images.

Last but not least, the entire corporation given by 27 respondents in testing and using this project is highly appreciated because their feedbacks are really meant to the improvement of this project.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND NOMENCLATURES

IT	Information technology
AI	Artificial intelligence
2D	2-dimensional
3D	3-dimensional
ANN	Artificial neural network
MPNN	Multilayer perceptron neural network
U1	Input layer
US	Feature extracting layer
UC	Intensity layer
UO	Output layer
CWT	Continuous wavelet transform
BPNN	Back propagation neural network
IANN	Important-aided neural network
fri	Feature relative importance
SOM	Self organizing map

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND OF STUDY

Seismic technique is originally developed by oil and gas industry and practiced in reservoir exploration since the 1930's to delineate the structure of the subsurface [1]. Accumulation of reservoir underneath can be visualize and predicted based on the processed data and information, hence decisive strategy can be made whether to further explore the natural resources stored in that particular location or vice versa. Reflection seismology is a widely implemented technique during the exploration phase to construct an accurate profile of the subsurface geology [2]. These techniques involve the measurement of travel time of seismic energy from surficial shots through subsurface to arrays of ground motion sensors or geophone [3]. The reflection and refraction wave of seismic energy will form seismic image at the end of activity.

A seismic image is characterized by a series of wiggle traces with alternating peak and trough amplitudes aligned laterally to form a strata reflection pattern [4]. Both the magnitude and variation of amplitude along or across the wiggle traces define the term seismic texture where it is analyzed by oil and gas experts to predict the presence of hydrocarbon in the subsurface via seismic image segmentation. This segmentation process is aim to partition a seismic image into multiple sections based on seismic texture (hereafter the term object will be used).

At current stage, the experts are responsible to differentiate and recognize all the existing objects in the seismic image because characteristic and possibility of the structure of subsurface in accommodating reservoir can be reviewed from these objects. Examples of the objects are fault, channels, salt domes, and strong reflectors. Seismic image segmentation process has already been carried out in oil and gas exploration decade years ago yet it has not been fully automated because of the heavy amount of knowledge involved in the decision making process. The human brain is well-known as the good pattern recognizer and until today this outstanding capability is not surpassed by any computer [2]. Although human interpretation is carrying the possibility of human error occurrence and argument in defining section that is not obvious to be recognized through naked eyes, yet the tacit knowledge and years of experience embedded in experts' brain are still acting as the most powerful tool in recognizing the existence of the objects in seismic sections.

Investment in oil and gas development involves extremely complex procedures and problems in which precise decision is the key to determine the worth-value of a project to ensure profitability. Thank to the great number of technological alternatives existing in this society and advancement of information technology (IT), people are start looking to develop intelligent systems to assist the experts in completing seismic image interpretation. These systems provide faster speed, greater consistency and concrete specification to support wise decision making before development stage is started.

Hence, more and more researches have been carried out in research field to study how a seismic image can be segmented by extracting the image pixel values via image processing software and develop data mining algorithm to train an artificial intelligence (AI) to learn the behavior and relationship which can be used to distinguish the differences between the values to differentiate the objects. Seismic image is typically displayed as a 2-dimensional (2D) section of stacked traces placed side-by-side [5]. Time travelled of the seismic energy is represented vertically and horizontal direction is linear distance on the surface of the earth.

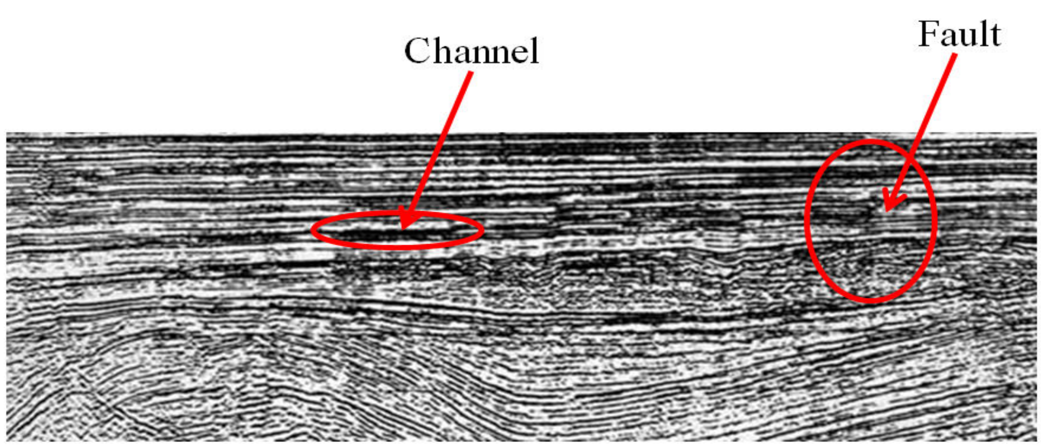


Figure 1.1: Example of 2D Seismic Image with Channel and Fault

For this project, data mining concepts are used to unseal the behavior and relationship among pixel values of channel and fault. ANN is chosen as the algorithm to process input, which is GLCM of these objects and learn to retrieve the image texture pattern for recognition purpose.

1.2 PROBLEM STATEMENT

Seismic image is studied by industry experts and the objects in the image is being interpreted manually based on their years of experience and knowledge in which problem of over-depending on the experts could be raised. Efficiency and result of the manual interpretation is not guaranteed as it is strongly relying on the experience and knowledge of an individual has in mind.

1.3 OBJECTIVE AND SCOPE OF STUDY

The main objectives of this study are:

- To create ANN which is able to categorize channel and fault by analyzing Gray Level Co-Matrix (GLCM) of the objects.
- To build a seismic objects recognition system that is able to facilitate the recognition of channel and fault.

The scope of the study embodied:

- Feasibility of ANN in performing pattern recognition.
- Conversion of seismic image into GLCM and as input to the ANN.
- 3D seismic models.

Experts from petroleum engineering department such as lecturers and postgraduate students are engaged during the development of the research and project to further understand the way seismic image being interpreted manually which their experiences could help in developing the algorithm that performing the same task.

CHAPTER 2
LITERATURE REVIEW

2.1 DATA MINING AND ITS PROCESSES

Rapid advancement in computer hardware and software technology can be seen in 21 century where it allows more and more government and non-government organizations to generate as well as to store large volume of data. The size of the database equipped and volume of data collected are not the key factors that guarantee valuable information as a return. The mounted information available could become liability in the sense that it could divert the focus in strategy planning or turn out costly to install additional infrastructure to accommodate the volume if it is not being well processed and filtered.

Data mining, also known as knowledge discovery and data mining serves its purpose by discovering hidden valuable knowledge or information by analyzing huge amount of data with a degree of certainty. It can be defined as a process of discovering new, interesting knowledge, such as patterns, associations, rules, changes, anomalies and significant structures from large amounts of data stored in data banks and other information repositories [6]. The output of data mining is information patterns [7]. These patterns are made up from non-trivial information that implicitly embedded inside the readily dataset which cannot be easily summarized via simple computations. For example, in this project, the unseen relationship between pixel values of seismic image will be revealed by using mathematical algorithms to differentiate the faults and channel objects.

As mentioned in Chapter 1, data mining tasks can be divided into 2 categories, which are predictive and descriptive that encompass 8 methods in total. The data mining process flows in same way regardless which category and method are use in knowledge discovery. There are 3 major steps to produce an output patterns: preparing input data, mining patterns, and last but not least post-processing patterns [7].

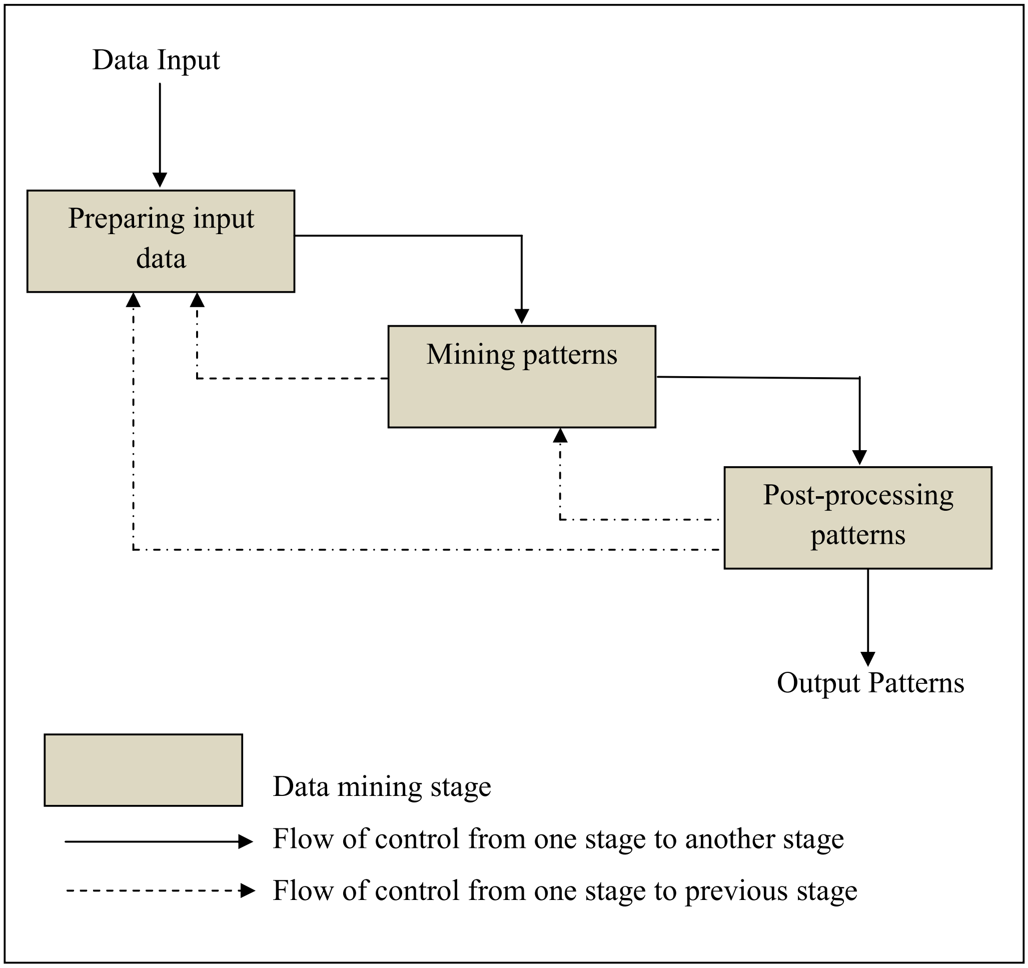


Figure 2.1: Flow of Data Mining Processes

There are many sources which data can be taken from, such as databases, web, text as well as image. These data sources are identified according to the objective of the data mining process and data is taken from the relevant sources. In our case, seismic image's pixel values are the target data and will be acquired via image processing software, MATLAB. Once the target data is identified and gathered, it needs to be pre-processed before proceed to data mining step. Data pre-processing tasks such as data cleaning, reducing data dimensionality and dealing with unknown data values are playing important role in ensuring integrity and quality of data which could affect the accuracy of the output.

The next step in this process is the mining process from input data to patterns. At this stage, a sagacious data mining task must be determined to serve the objective of knowledge discovery which has been set at the very first place. The selected model or solution must be able to cope with the data types. Classification and clustering model are the examples of data mining task that can be used to recognize object faults and channel in seismic image. The ultimate choice of the model is heavily relying on the sensibility of the patterns being captured. Different datasets are used to test those models to evaluate the outcome and the model that produces the highest quality of patterns is picked.

Post-processing patterns take place after the mining stage in which the further processing of the discovered patterns is carried out. This processing includes pattern selection, pattern evaluation, and pattern interpretation. Only relevant patterns are taken for further purposes. A seismic image may encompass several objects and our focus is on object faults and channel. So the patterns of these 2 objects will be selected. In data mining, patterns are the main subject to be studied to aid decision making process. Hence, the selected patterns must be evaluated with its credibility and significance at stake. Lastly, the patterns that have no issue are interpreted and have to be understood by experts to ensure the right patterns are generated and serve the objective.

Data mining processes will not end until the output of the patterns is satisfied or meet the objective. All the steps could be repeated from data collection to analyzing the output and iteration of works may happens at the same stage as well before proceeding to next stage. In the first stage, pre-processing of input data may undergoes several cleaning process to filter unnecessary values and new data may be added to enrich the value of the existing dataset. Repetitive tasks are unavoidable in mining stage as different data mining methods have to be tested with selected dataset in order to find out the best solution that produce the most recognized patterns. .

2.2 DATA MINING APPROACHES

In data mining, there are 2 approaches can be taken to complete the process: hypothesis testing and discovery. Hypothesis approach marked as a traditional way of knowledge discovery process [7]. The purpose of this approach is definitely to test the hypothesis propose aforehand and evaluate the outcome of the result whether that hypothesis is supported by reliable evidence or disprove by the patterns shown. In this approach, a null hypothesis as a default belief is assumed and is used to compare with the proposed hypothesis when patterns are produced. The proposed hypothesis is correct and accepted when support demonstrated from the output is relatively higher than the support that null hypothesis gained.

For instance, hypothesis of students who burn midnight oil score better in final exam than students who have early sleep. Data about students who stay up until late night and students who do not, such as exam result and hours of study are collected. A null hypothesis stated that students sleep late because of revision have higher score than who do not do so. In order to find out the truthfulness of the hypothesis, the exam result is computed and total revision hours for both types of students are measured. If the patterns of the output showing that students who burn midnight oil are able to score better, the propose hypothesis is substantiated and the null hypothesis is disproved.

Meanwhile, discovery approach does not set any specific hypothesis at first but starts with data and induces possible patterns from dataset. The procedures involved are almost the same with those in hypothesis testing, such as collect and prepare data of interest, conduct data analysis, interpret possible patterns and measure the usefulness and credibility of the patterns to the area of investigation. In conducting knowledge discovery, any method that does not presume hypothesis is considered taking the discovery approach where process of data learning takes place.

There are 2 types of data learning [8] [9]:

i. Directed learning
ii. Undirected learning

In directed learning (also known as supervised learning), the classes are predetermined which means the discovery is driven by the predetermined classes or outcomes of an output variable. The computerized system is obligated to search for the patterns relating to the outcome. The data mining methods that reflects this characteristic is classification where target variable is selected, for example list of male participants. The computer is directed to search among a given dataset the name of male participants based on gender and all the names are classified under same group. Decision tree induction and naïve Bayes are the examples of supervised learning as well [8]. Undirected learning (also known as unsupervised learning) is not provided with the criteria of desired outcome like supervised learning but to develop the classification labels by the learner or computer itself. This algorithm is created to find out similarity between data and their relationship to determine whether the data can be categorized together as a group or vice versa. Clustering and association rule utilize undirected learning to perform data classification.

2.3 DATA MINING ALGORITHM - ARTIFICIAL NEURAL NETWORK (ANN)

There are many algorithms developed by experts to provide better version of problem solving solution in the field of medical, military, oil and gas exploration and many more. In this seismic image object recognition project, our focus will be on ANN and MPNN will be used to recognize the fault and channel objects. ANN is an artificial intelligence inspired by human brain neural network that consists of nodes and links located on different neuron layers. It is not a new concept used in data mining field as it has been introduced by Warren S. McCulloch (neuroscientist) and Walter Pitts (logician) in a paper entitled "A Logical Calculus of the Ideas Immanent in Nervous Activity" [10].

Each node (neuron) in ANN receives input data from nodes on the previous layer and summarizes the data via sum function and output the result in a single value with the assistance of transfer function. Generally weighted sum is used as the sum function and sigmoid function as transfer function. Then the output is channeled to other neurons as input and same process is repeated until final node is reached.

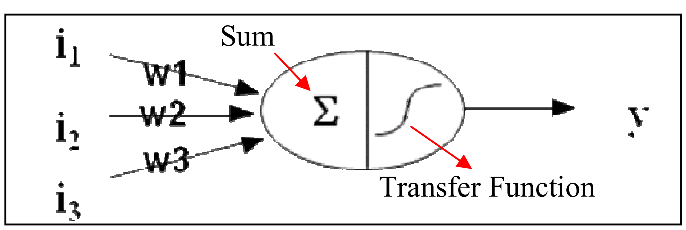


Figure 2.2: Single Neuron Node Structure with Sum Function, $x = w_1*i_1 + w_2*i_2 + w_3*i_3$ and Transfer Function, $Sigmoid(x) = 1/(1+e^{-x})$

McCulloch and Pitts's effort to form a neural network with mathematical model has been further developed and improvised until today where an ANN can be categorized into 2 groups:

i. **Feed-forward network**

This type of network does not provide feedback from the output layer to the input layer, hence the output data will not be automatically incurred to the input neuron and return the whole process. There is no connection between neurons in the same layer and the different layers as well. The network structure as portrayed in Figure 2.3.

ii. **Recurrent network**

Recurrent network contains feedback from the output produced to the input layer which means the result of the output is manipulated by the current input as well as next output (feedback). Once a new input is taken, the output is calculated and feedback is sent to modify the input to narrow the gap difference of each output produced.

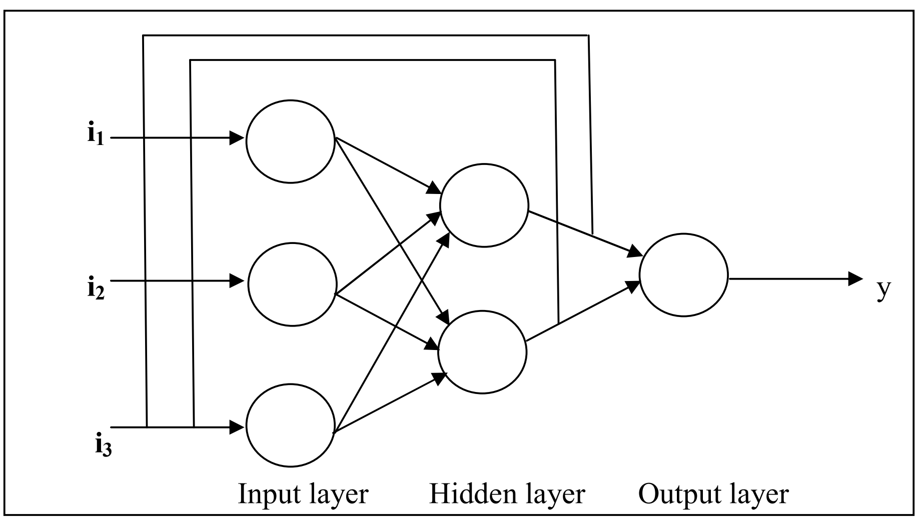


Figure 2.3: Recurrent ANN

In a paper completed by Harrigan et al, the structure of MPNN is well described [11]. A complete MPNN should consists of layers of artificial neuron nodes, named input layer, hidden layer and output layer. Nodes located in input layer receive weighted attribute value as input from previous nodes and output layer responsible to produce ultimate output values from the network. Hidden layer is the layer that formed in between input layer and output layer where it receives input from other nodes and sent out value to the next layer. The flow of input and output data are executed by number of links that connect all the layers together to form a neural network.

The transfer function of node j can be defined as:

$$o_{pj} = f_i(net_{pj}) \tag{1}$$

If this node receives total input of o_{pj} that coming along weighted connections from the nodes in the previous layer for training pattern p is set as:

$$net_{pj} = \sum_{i=0}^{N-1} w_{ij} o_{pi} + \theta_j \tag{2}$$

Where θ_j = value of the threshold.

According Harrigan and partners, back propagation algorithm is a popular training method for MPNN. Error that calculated at the output layer is sent back reversely to previous layers of nodes and the weight is modified to reduce the error rate. So generally this training process involves 2 main steps:

i. Forward propagation of input to calculate value o_{pj} for every node and

ii. Error value encounter at the output later is passing backward to each node in the previous layers and weight is adjusted.

Following is the function used to adjust the weight value according to training pattern p :

$$\Delta_p w_{ij} = \eta \delta_{pj}(t) i_{pj} + a[\delta_{pj}(t) - \delta_{pj}(t-1)] \tag{3}$$

Where η = learning rate

$\delta(t)$ = function of error at the output layer

a = momentum function

For output nodes, δ_{pj} is given by:

$$\delta_{pj} = (t_{pj} - o_{pj}) f_j^{'}(net_{pj}) \quad (4)$$

Where $f_j^{'}(net_{pj})$ = derivative of nonlinear function $f_j\ (net_{pj})$

t_{pj} = target signal at node j for pattern p

For hidden nodes, δ_{pj} is given by:

$$\delta_{pj} = f_j^{'}(net_{pj}) \sum_k \delta_{pk} w_{kj} \quad (5)$$

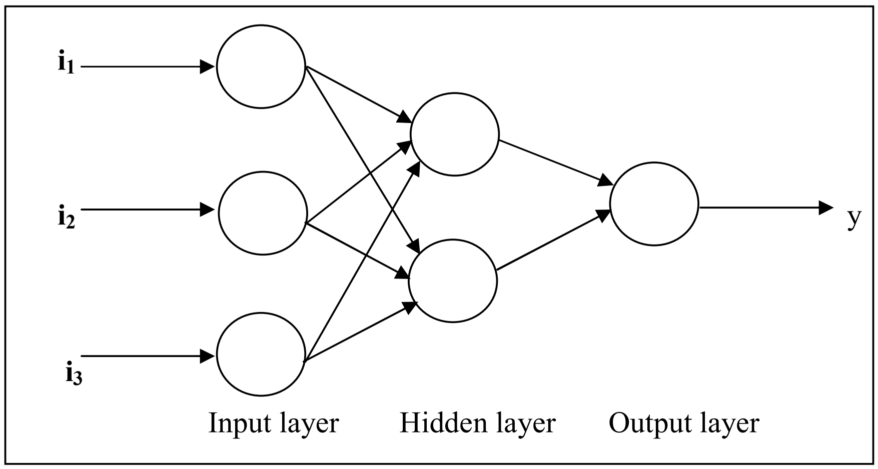


Figure 2.4: Multilayer Perceptron Neural Network (MPNN)

Huang (2001) presents a MPNN which has been trained in several stages by training set which includes noise-free, low-noise, and misclassified seismic patterns for robust recognition of seismic image [12]. 3 types of seismic patterns are chosen to analyze, which are bright spot, pinch-out and horizontal reflection patterns. Seven moments such as invariant to translation that introduced by Hu in 1962 are also applied in this study for characterization of each seismic pattern [13]. Given a two-dimensional digital image function $f(x, y)$, the formula that defines the moment of order $(p+q)$ is:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (6)$$

For $p\ ,q$ = 0, 1, 2, 3…

The central moment can be expressed as

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (7)$$

Where $\bar{x} = \frac{m_{10}}{m_{00}}, \bar{y} = \frac{m_{01}}{m_{00}}$

η_{pq} represents the normalized central moment and is defined by the following function:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \tag{8}$$

Where $\gamma = \frac{p+q}{2} + 1$

For $p+q$ = 2, 3, 4…

The MPNN created by Huang contains 1 input layer with 7 neurons, 1 hidden layer with 50 neurons and 1 output layer with 3 neurons. In the input layer, the purpose of set one of the neuron to 1 instead of $x_1, x_2 \ldots$ is to create a constant term in a complete linear combination with weighting coefficient. Formula shown in hidden layer is:

$$\begin{aligned} s_j &= \sum_{i=1}^{I+1} w_{ji} xi, x_{I+1} = 1, I = 7, j = 1,...,50 \\ o_j &= f(s_j) \end{aligned} \tag{9}$$

Where the number 50 represents the total number of hidden nodes chosen which are near to the order of the number of training patterns.

For the activation function, the following sigmoid function is used:

$$\begin{aligned} f(s) &= \frac{1}{1+e^{-s}} \\ o_j &= f(s) \end{aligned} \tag{10}$$

And as for output layer:

$$s_k = \sum_{j=1}^{J+1} w_{kj} o_j, o_{j+1} = 1, J = 50, k = 1,2,3... \tag{11}$$

$$o_k = f(s_j) = \frac{1}{1+e^{-s_k}} \tag{12}$$

Back propagation learning rule is suggested by Huang to train the weighting of this network. Properly trained back propagation networks tend to give reasonable answers when presented with inputs that they have never seen [14]. The objective of training is to reduce the sum of the squared-error between desired output and actual output from the network. Hence, weighting vector within neurons is adjusted repetitively to get a lower sum of the squared-error via standard back propagation, Gradient Descent method. The following functions define the weight adjustment within input layers, hidden layers and output layers:

Error at output layer, E is defined as $\frac{1}{2}\sum_{k=1}^{k=K} (d_k - o_k)^2$ (13)

Between input and hidden layers,
$$\Delta w_{ji} = w_{ji}(t+1) - w_{ji}(t) = -\eta \frac{\partial E}{\partial W_{ji}} = \eta(\sum_k (d_k - o_k) f_k(s_k) w_{kj}) f_j(s_j) o_i \tag{14}$$

Between hidden and output layers,
$$\Delta w_{kj} = w_{kj}(t+1) - w_{kj}(t) = -\eta \frac{\partial E}{\partial W_{kj}} = \eta(d_k - o_k) f_k(s_k) o_i \tag{15}$$

As mentioned in earlier part, noise-free bright spot, right pinch-out and horizontal reflection patterns are used as training patterns to fulfill the requirement of training and recognition of this project. Binary numbers 1 and 0 are used as the indicator to show whether the training patterns are belonging or k-th class or not. During the recognition stage, 10%, 15%, 20%, 25%, 30% and 35% of uniform random noise are added into the initial noise-free training seismic patterns.

Other than back propagation, Robust Training concept is introduced to improve the performance of the network. This training method is also adapted by Song et al (1999) to ensure the training process does not involve the convergence of noise [15]. The process of neural network conventional training carries the possibility of being interfered by the presence of noise [16]. Misclassified noisy patterns are added into the previous training dataset and the MPNN is retrained so that it has greater ability in noise classification. The result in Huang's study showed Robust Training does contribute to the decreasing rate of classification error in MPNN when using noisier error patterns in retraining.

Other than conventional ANN method, researchers have put in effort to develop non-conventional seismic pattern recognition algorithm which is able to differentiate each seismic pattern accurately via advance learning and training solutions. A hybrid neural network presented by Huang and Yang (1992) breakthrough the traditional design of neural network because both unsupervised and supervised learning are practiced in this neural network [17]. 4-layer network that contains an input layer (U1), a feature extracting layer (US), an intensity layer (UC) and an output layer (UO) are created.

Generally, the input layer that consist 30*60 i cells is playing the same roles with other common neural networks, which is receiving input data and passes the output to US. In this model, US have 30 s planes where each plane contains an array of 20*50 s cells. The output value of a S cell in the p -th plane is represented by $us(p,n)$. This output value is computed according to Euclidean distance between the cell's input stimulus and its weights:

$$us(p,n) = f(\sqrt{\sum_{v \in A}(ui(n+v) - ws(p,v))^2}) \tag{16}$$

Meanwhile, in UC layer there are 30 planes. Each of the planes receives input via the connection that connected to the previous US plane. An array of 7*13 c cells is structured and the inward connections are set to 1.0. $uc(p,n)$ is used to represent the value output value of a c cell where n is its 2-dimensional coordinates in the plane. All the cells in intensity layer are responsible to sum up the intensity of a particular region on previous layer's plane and channel the output to UO layer which is the output layer. There are 5 neurons assigned in this layer in which each of the neuron is corresponding to 1 type of seismic pattern, which are: pinch-out seismic pattern, flat-spot seismic pattern, sea level fall seismic pattern, sea level rise seismic pattern and lastly bright-spot seismic pattern. The weights between UC and UO layer are subject to change as they are fully connected and trained by delta rule, which is the supervised learning.

The learning process in this study is divided into 2 phases. The first phase of learning takes place from UI layer to US layer to extract features and the second phase is clustering learning from UC layer to UO layer. In US layer learning, training pattern is formed by value computed from all the cells in each US planes via:

$$us(p,n) = f\left(\sqrt{\sum_{v \in A}(ui(n+v) - ws(p,v))^2}\right) \tag{17}$$

After that, specific steps are carried out for several times until there is no strongest cell, also known as seed can be selected. When a seed's cell is found, that particular weights are reinforced according to the following functions:

$$\begin{aligned} ws(p,v) &= ws(p,v) + alpha(p) \times (ui(n+v) - ws(p,v)) \\ alpha(p) &= alpha(p) \times 0.8 \end{aligned} \tag{18}$$

Once the weighting is adjusted, new Euclidean distance is calculated and the new value is saved as $D(p)$. The process is back to the finding of strongest cell and the whole learning process is completed after all strongest cells are selected.

Moving on to UO layer learning, indexed cell and non-indexed cell are identified and their weights are computed in the following way:

i. For indexed cell,

$$wo(p,v) = wo(p,v) + uc(p,n+v) \tag{19}$$

is used if its value is smaller than a specified threshold value.

ii. For non-indexed cell,

$$wo(p,v) = wo(p,v) - uc(p,n+v) \tag{20}$$

is used if its value is larger than a specified threshold value.

5 sets of training patterns are prepared with each set consists of 2 different seismic patterns that are in same class and the US layer is trained for 5 times. Meanwhile, specified requirement is set to UO layer that only specified output has response. From the earlier study that has been discussed, same approach taken which is insertion of 10%, 15%, 20%, 25%, 30% random noise is seen in this study to test the ability of the layer to recognize patterns with certain percentage of noise. Result shows that this hybrid neural network is able to recognize all the 5 group of seismic patterns even with 20% of noise existing and in clustering part, 89% of seismic patterns are successfully being classified. More importantly, it is able to learn to

extract critical features in training patterns and cluster them according to defined requirement.

In the study of seismic image recognition via ANN, different researchers will prefer different data pre-processing process which is the preparation of the input data that feed the input layer. Other than pixel values of the image, as portrayed in Diersen et al research in 2011, seismic wavelet is used for this algorithm as well. The ANN algorithm developed in this study is based on Continuous Wavelet Transform (CWT) that allows the trained neural network to analyze waveforms in the time-frequency domain [18]. The difference between the preparation of seismic wave and GLCM is that the seismic pattern is converted into waveform based on time and amplitude of the image instead of converted into pixel values based on the grayscale intensity.

Standard feed-forward back propagation neural network (BPNN) is chosen and the activation function is Sigmoid function, due to the range of output, which are 1 and 0. The initial weights for all links are set to number between -1 and 1. Another type of neural network named importance-aided neural network (IANN) is introduced by the authors to further discover the performance of ANN in seismic image classification. IANN is a bit different from the conventional ANN in the sense that Feature Relative Importance (fri) that extracted from experts' knowledge is embedded into the ANN. According to Iqbal, fri is a real valued approximation of a feature's importance given by the experts [19]. Another difference between these 2 neural networks is the change to the back-propagation algorithm where the links between input layer and the first hidden layer are affected. fri value is multiplied with learning rate to highlight the attributes that are more importance. Thus, the function of the link in this layer is:

$$\Delta w_{j,i}(n) = afri_i \delta_j x_{j,i}(n-1) \tag{21}$$

Significant result and performance can be seen by doing some simple modifying on standard ANN [19]. This statement is proved by Diersen and his partners as their experiment result clearly shows that IANN manage to get 99.60% of testing set classified correctly while conventional ANN achieved 99.21%. With the embedded of experts knowledge into ANN, the accuracy of result that produced will be further improved, hence pushing forward the ability of ANN in running seismic patterns recognition.

Another popular ANN algorithm is called self organizing map (SOM) [20]. This method is adapted by Moraes et al (2006) in investigating cluster analysis of 3D seismic data for oil and gas exploration [21]. The SOM clustering method is a network model that practice neighborhood concept where the network learns to recognize neighbor sections in the training and also the topology of the learning set [22]. In this study, it is used to perform crisp clustering in which a dataset with over 223000 records about seismic attributes from a Brazilian oil field are used to test the algorithm. Fuzzy c-means algorithm is tested together with SOM but only SOM part is taken to be reviewed as it fall under ANN category.

Neurons in SOM's network layer are arranged in a fixed position according to a topology function and it can be in rectangular or hexagonal shapes. Hence this method can be named according to their arrangements, which are SOM rectangular or SOM hexagonal. Competitive learning is applied in this network and the neuron that stays nearest to the input data is the output unit which located in output layer. Kohonen's rule is used update all the neurons from a certain area of that particular winner neuron and the value of radius determines the neighborhood space. Winner neuron is selected via the following function:

$$\| X - G_W = \min_j\{\| X - G_j \|\} \text{ or } w = \arg\min_j\{\| X - G_j \|\} \tag{22}$$

Where X = input data vector

G_j = synapses vector and

w = index of the winning neuron.

Linear combination's mean between the preceding and current weight is used to adjust the synapses of the winner neuron. The formula is as below:

$$G_j(t+1) = G_j(t) + h_{wj}(t)[X(t) - G_j(t)] \tag{23}$$

Where t = discrete-time coordinate and

h_{wj} = neighborhood kernel.

Usually the value of the kernel is calculated from the function:

$$h_{wj}(t) = h(\| r_w - r_j \|, t) \tag{24}$$

Where r_w and r_j = radius of the neuron w and j.

In this study, bubble neighborhood concept that act as a constant function in the defined neighborhood of the winner neuron [14] is used by the authors. It is defined by

$$h_{wj} = \{ \begin{matrix} a(t) \\ 0 \end{matrix} \tag{25}$$

If neuron j belongs to neuron w neighborhood, the bubble neighborhood will be $a(t)$ which is a monotonically decreasing function of time (0< $a(t)$ <1) or 0 if this is not the case.

In order to allow comparison of the results can be made between fuzzy c-means and SOM, only 1 neuron is assigned to the first layer and second layer contains the number of groups it is applied to. There is no definite rule to define the number of neurons have to be existed in a neural network to produce good clustering result. Hence, the decision of number of neurons in SOM layer can be identified by cluster validation index. The standard used in this study is PBM index:

$$PBM(K) = (\frac{1}{k} \times \frac{E_1}{E_K} \times D_K)^2 \tag{26}$$

Where K = number of clusters

E_K = sum of intra-cluster distances

E_1 = sum of distances of all points to the data center and

D_K = maximum of within-cluster distances. These distance values are calculated from the function:

$$E_K = \sum_{j=1}^{K} \sum_{i=1}^{n} u_{ij} d(x_1, w_j) \tag{27}$$

$$D_K = \max_{i,j=1}^{k} d(w_1, w_j) \tag{28}$$

Degree of membership of object X_i to group G_j with center W_j is represented by u_{ij} and d is the distance function.

Once the cluster validation index is set, the SOM algorithm is run from 2 to 10 partitions to determine the best number of neuron groups in the network. This process is repeated for 10 times in SOM network with rectangular topology and SOM network with hexagonal topology to get a higher credibility of means results. Result from all the three algorithms including fuzzy c-means unanimously showing that 3 groups are the best number of group to partition the seismic data. Values for all the algorithms created by PBM also showing that there are not much difference can be seen, indicating the groups formed are credible and dependable.

The advantage of using SOM compared to other ANN is that it is able to perform faster and simpler because it has only 2 layers of neuron. However, it required the image data in a seismic image to behave similarly or else the quality of the output will be affected. The brief view of the SOM algorithm used listed as below.

Table 2.1: Finalized SOM details

Neurons on first layer	1
Neurons on second layer	From 2 to 10 (3 has the highest result quality)
Epochs	10^6
Learning rate	0.1
Radius	0
Neighborhood function	bubble
Topology	rectangular/hexagonal

CHAPTER 3
METHODOLOGY

3.1 SYSTEM DEVELOPMENT CYCLE

In this project, the “Waterfall Model” is chosen. This approach is the first “Process Model” to be introduced and has been widely implemented in project completion [23]. There are 5 phases included in this model: requirements definition, system and software design, implementation and system testing, integration and system testing and lastly operation and maintenance. In order to complete this project, the model has been customized based on the project requirements. The third phase is named as implementation and unit training because ANN training is a major step need to be done to complete the project. Each phase consists of specific set of activities that need to be completed before the following phase can proceed. This model maintains a disciplinary approach in developing a solution as all the activities throughout the project are clearly defined in each phase.

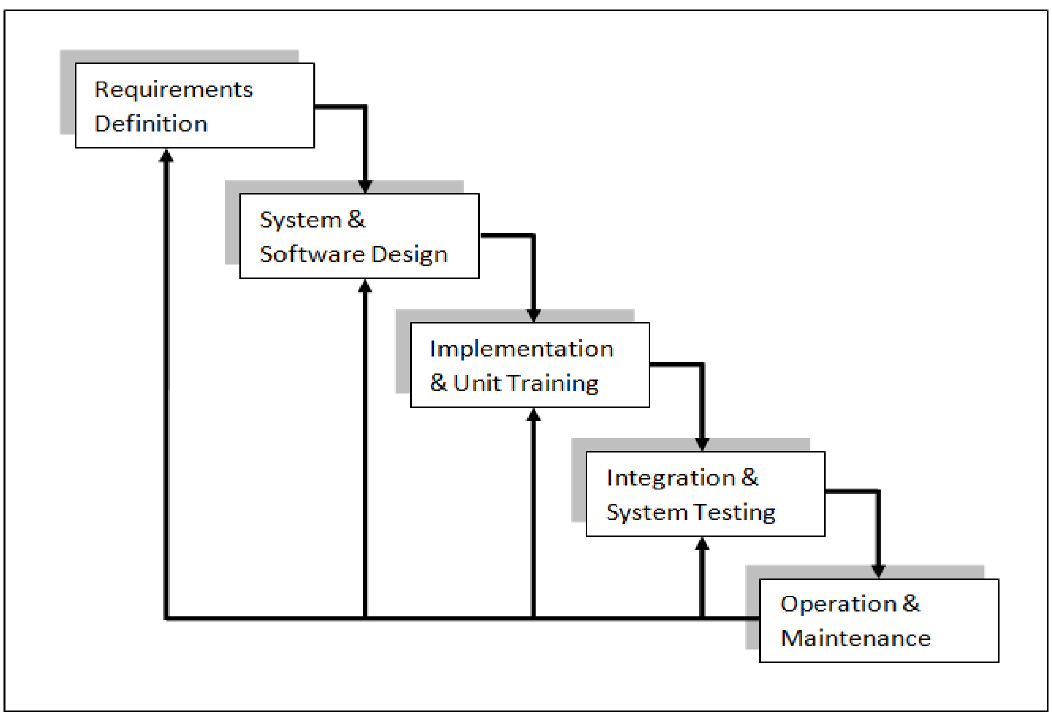


Figure 3.1: Waterfall Model

3.1.1 Requirement Definition

The existence of this project is clearly defined by the problem statement which mainly involves manual interpretation issue. Hence, the requirements of this project are analyzed via validation of background of study as well as definition of study scope and objective. In addition, literature review is carried out to further understand the data mining algorithms that have been developed to serve the purpose of seismic image segmentation and classification in oil and gas exploration.

The goal of this project is to develop a ANN that is able to recognize fault and channel objects. In order to meet the objective, resource and dataset of seismic image is determined because this information is crucial in providing input data. Training and testing of the algorithm required as many dataset as possible to increase the accuracy as well as validity of the output patterns. Last but not least, the software used to fulfillment all these requirements is MATLAB 7.9.0 (R2009b) where the ANN tool box, functions as well as image processing tool box are prepared in a platform without sophisticated programming skill is needed. Defining of such necessary requirements could ease the task in later stage, such as system design and implementation.

3.1.2 System and Software Design

3.1.2.1 System Architecture

During the design phase, system architecture is established to identify and describe the fundamental structure abstractions as well as process flow of the whole system. The system architecture consisted of 3 core focuses of the whole system, which are image processing, classification and result. Selected seismic image with the dimension of 130 x 70 is processed and converted into GLCM in the first part and the GLCM is input to desire ANN after that. Supervised learning is adopted in the ANN learning process as training dataset and target dataset are provided to the ANN in order to learn to recognize the pattern of training dataset and categorized it based on the target dataset. Result produced is in binary form and it is converted into understandable phrases by the system to enhance system's user friendliness.

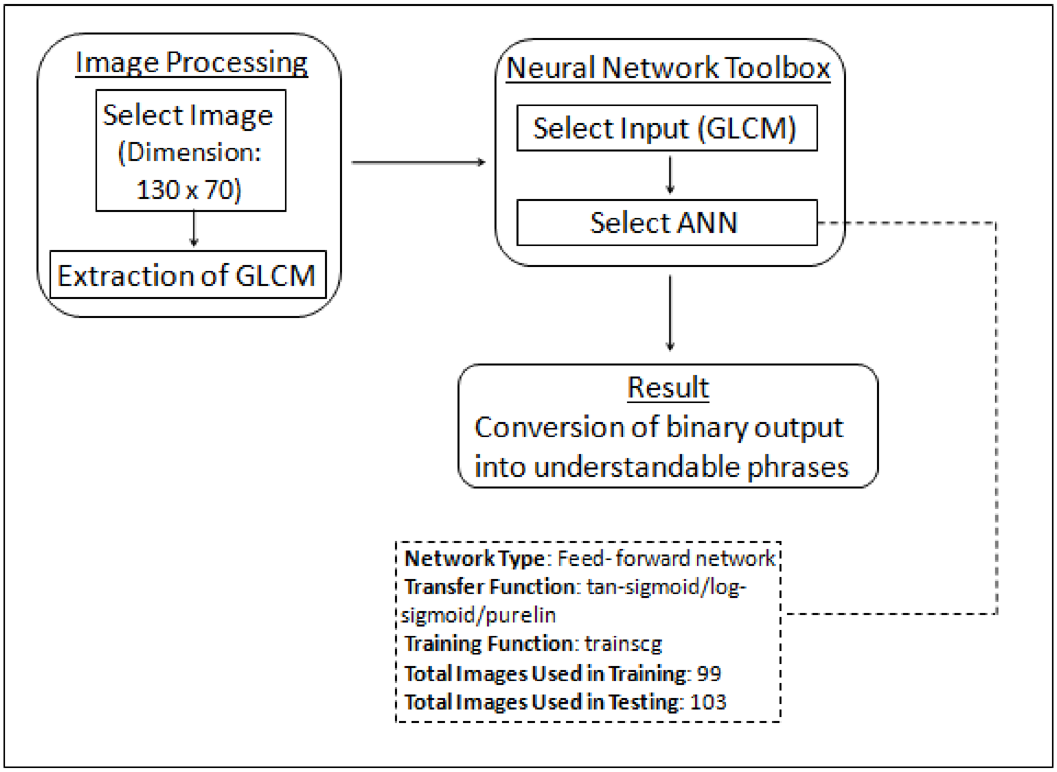


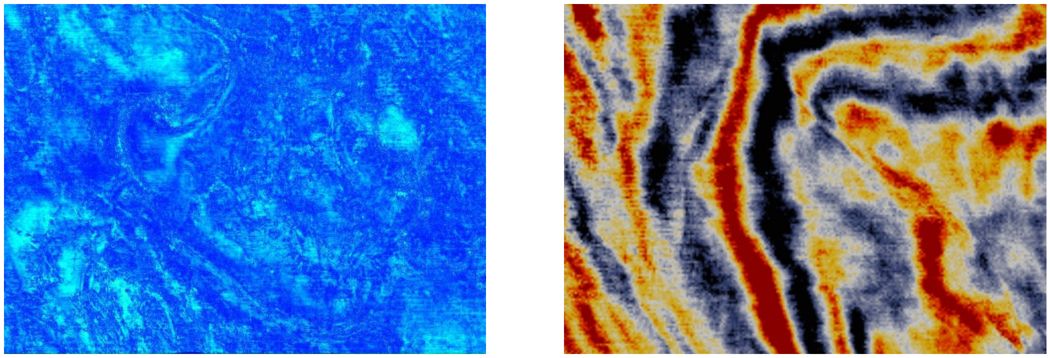
Figure 3.2: System Architecture

As shown above, the type of ANN used is feed-forward network with the transfer function of *tan-sigmoid* which is the default ANN used in MATLAB Neural Network Pattern Recognition Tool. In order to come out with better system, comparison is made between 3 type of transfer functions available- *tan-sigmoid*, *log-sigmoid* and *purelin*. The function of transfer function in ANN is to calculate a layer's output from its net input. Hence, different transfer function may have different quality and accuracy of output. The ANN with one of these transfer functions which has highest accuracy is selected in the last phase of the development.

3.1.2.2 Datasets Preparation

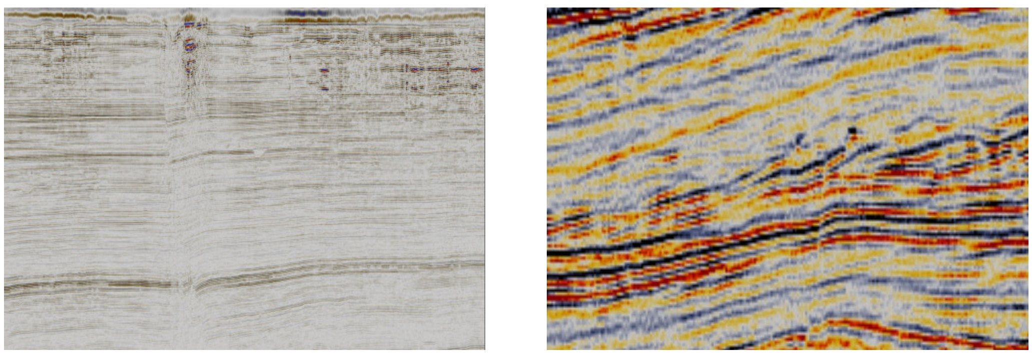
Datasets made up of GLCM are the main element needed in training and testing a ANN. It defines the whole structure of the neural network as pattern learning is based on the datasets given as input. Before hand, 2D seismic images that contain channel and fault are collected. These images are sliced from 3D seismic models provided by OpendTect as well as Universiti Teknologi PETRONAS (UTP)

Petroleum Geosciences Department's PhD students who are studying on seismic interpretation. Following are the examples of 2D seismic images extracted from 2 different seismic models:



(*a*) (*b*)

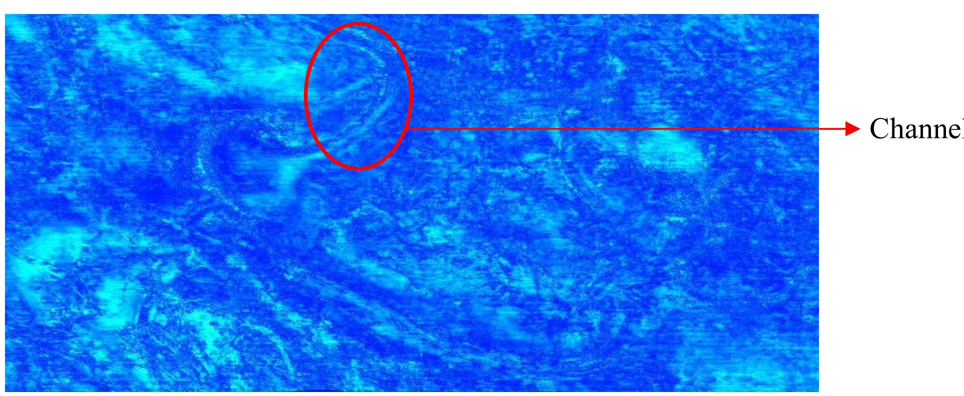
Figure 3.3: Top View of Different 3D Seismic Models



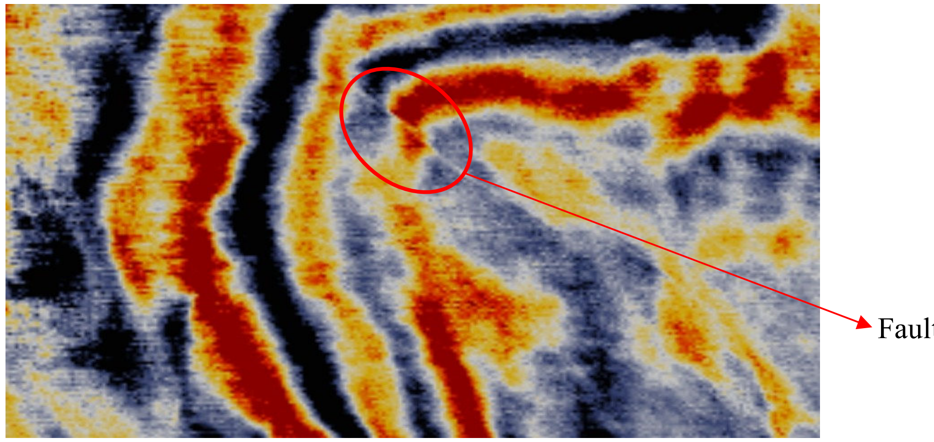
(a) *(b)*

Figure 3.4: Side View of Different 3D Seismic Models

After seismic images are collected, 2 PhD students from Petroleum Geosciences Department are invited to recognize fault and channel from the seismic images. Below are some of the channels and faults that have been identified by the students:



(a)



(b)

(c)

Figure 3.5: Examples of Channel and Fault Recognized from Top and Side View

Once the channel and fault have been indentified from the seismic images, these objects are cropped out as shown in Figure 9 and Figure 10 and loaded into MATLAB to start image processing work, which are image conversion, dimensions trimming as well as generate and concatenate Gray Level Co-Matrix (GLCM). GLCM is the final output from this process which will serve as input data to ANN.

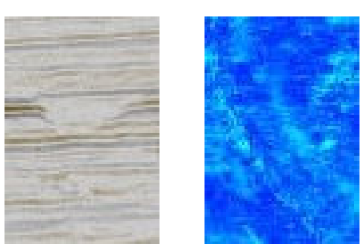


Figure 3.6: Channel

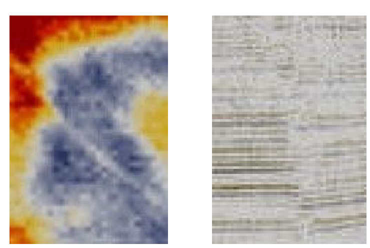


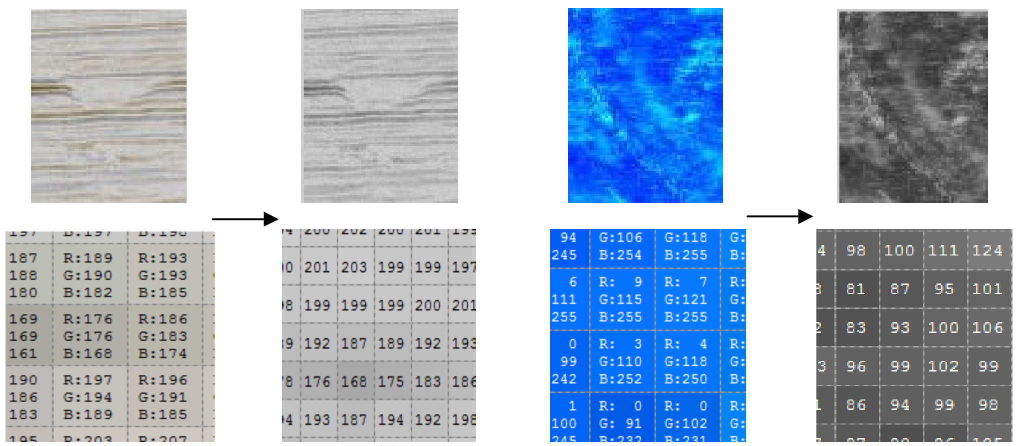
Figure 3.7: Fault

i. Image Conversion

Before pixel values are extracted, the images are converted into gray scale from RGB scale because gray scale contains only one value in each pixel and speed up the extraction process without having to look into 3 values in each pixel as display in RGB scale. The command used as below:

a= imread(image_name);

b = rgb2gray(a);



(a) (b) (a) (b)

Figure 3.8: RGB Image (*a*) to Gray Scale Image (*b*)

ii. Dimensions Trimming

The purpose of trimming the dimensions is to standardize the images size as different dimensions of images are produced after cropped out from original images. The dimension of all the images is set to 130 x 70.

This step will improve the consistency of GLCM which will be generated based on the occurrence of total number of pixel values. By using same number of dimensions in each image, the quality of the GLCM will be better as well. However, it has to be done precisely to ensure the reducing or increasing of image dimensions will not negatively affect the section of channel and fault in the image, which could probably lead to poorer input data.

iii. Generation and Concatenation of GLCM

After pixel values are extracted into vector form and dimensions are trimmed, GLCM is generated via the command:

[glcm, SI] = graycomatrix(image_name);

The generated GLCM is in 8 x 8 dimensions as shown below:

	1	2	3	4	5	6	7	8
1	0	0	0	0	0	0	0	0
2	0	3	20	0	0	0	0	0
3	0	20	1945	382	5	0	0	0
4	0	0	380	1368	163	2	0	0
5	0	0	15	152	218	11	0	0
6	0	0	0	1	13	6	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0
9								

Figure 3.9: Example of GLCM Generated from a Channel

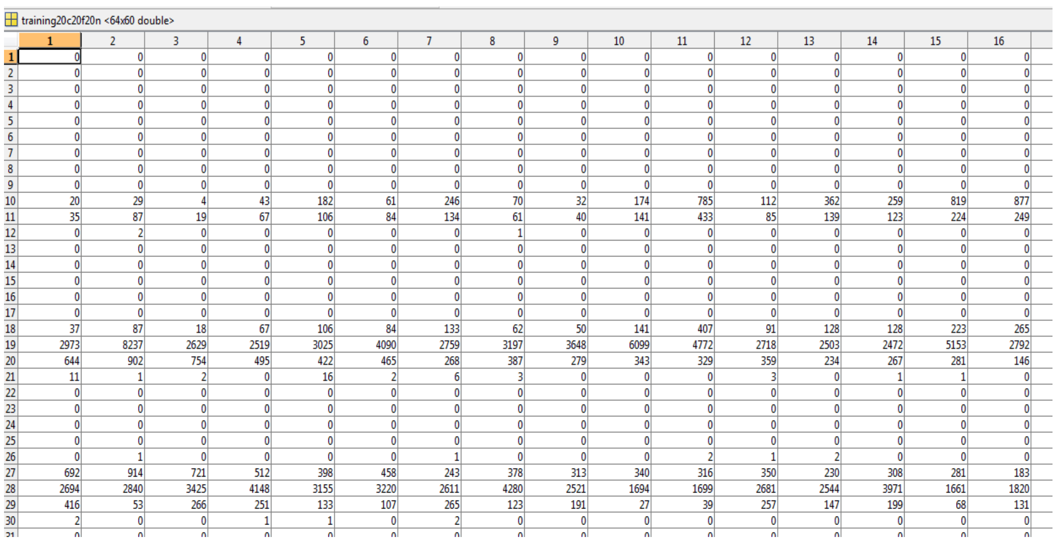
GLCM of a single image has to be arranged into either a row or a column. In this project, the alternative chosen is to use *glcm(:)* command to arrange the GLCM into column form. As *glcm(:)* command is triggered, it automatically brings down the values from column 2 to column 8 and connects all these columns under column 1. This will be the input data where ANN used to learn and recognize its pattern. Same image processing steps are repeated for all seismic images to prepare GLCM datasets.

	1	2	3	4	5	6	7
1	0						
2	0						
3	0						
4	0						
5	0						
6	0						
7	0						
8	0						
9	0						
10	3						
11	20						
12	0						
13	0						
14	0						
15	0						
16	0						
17	0						
18	20						
19	1945						
20	380						
21	15						
22	0						
23	0						
24	0						
25	0						
26	0						
27	382						
28	1368						
29	152						

Figure 3.10: GLCM Arranged in Column Form

3.1.3 Implementation and Unit Training

Different seismic objects contained different image texture and GLCM is used to recognize the differences of the objects. The main purpose of ANN training is to teach the neural network to learn and capture input's pattern. This is a supervised learning as target data is provided during the training process and the ANN will try to learn to achieve the target by analyzing the GLCM. The training dataset used to train ANN consisted of 99 sets of GLCM - 25 sets from channel, 23 sets from fault and 51 sets are neither. Both the training and testing of ANN is done via MATLAB Neural Network Pattern Recognition Tool (nprtool) and MATLAB Neural Network Tool (nntool) where a simpler platform is prepared to build a ANN.



training20c20f20n <64x60 double>

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	20	29	4	43	182	61	246	70	32	174	785	112	362	259	819	877
11	35	87	19	67	106	84	134	61	40	141	433	85	139	123	224	249
12	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	37	87	18	67	106	84	133	62	50	141	407	91	128	128	223	265
19	2973	8237	2629	2519	3025	4090	2759	3197	3648	6099	4772	2718	2503	2472	5153	2792
20	644	902	754	495	422	465	268	387	279	343	329	359	234	267	281	146
21	11	1	2	0	16	2	6	3	0	0	0	3	0	1	1	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	1	0	0	0	0	1	0	0	0	2	1	2	0	0	0
27	692	914	721	512	398	458	243	378	313	340	316	350	230	308	281	183
28	2694	2840	3425	4148	3155	3220	2611	4280	2521	1694	1699	2681	2544	3971	1661	1820
29	416	53	266	251	133	107	265	123	191	27	39	257	147	199	68	131
30	2	0	0	1	1	0	2	0	0	0	0	0	0	0	0	0

Figure 3.11: Training Dataset with 99 Sets of GLCM Arranged Vertically with 64 Rows Each

i. Training Process

Step 1:

To start the training, training inputs and training targets are imported from MATLAB workspace with the summary of the datasets shown under "Summary" section.

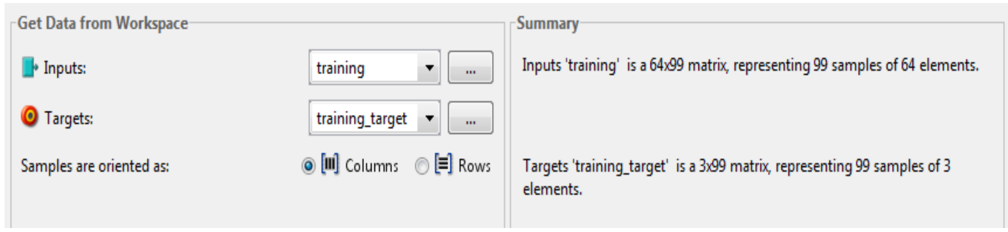


Figure 3.12: Selection of Training Dataset

Step 2:

In this training, 70% of the datasets which are 69 samples in total are used to train the network whereas 30% are used to validate the training stopping time as well as to test the network internally.

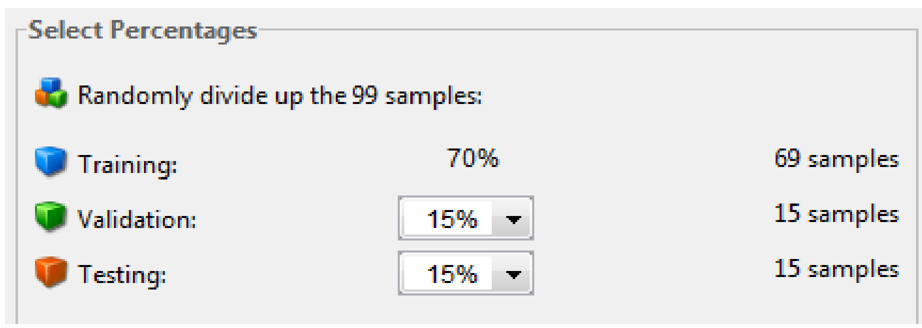


Figure 3.13: Division of Training Dataset into 3 Parts – Training, Validation and Testing

Step 3:

20 hidden neurons are used in the hidden layer to process input data. It can be adjusted according to different requirements but the number is fixed throughout the training process as it does not yield much difference to the result produced.

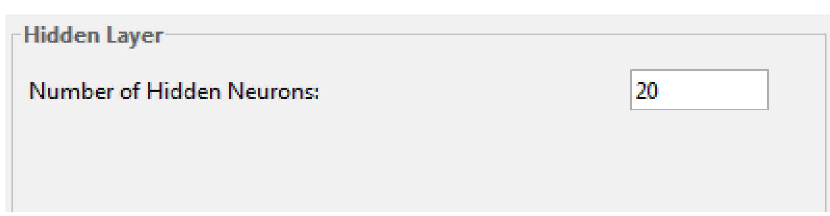


Figure 3.14: Number of Hidden Neurons in ANN

Step 4:

Training is started once the previous steps are done. Under the "Plot Confusion" section, 4 Confusion Matrix which are Training Confusion Matrix, Validation Confusion Matrix, Testing Confusion Matrix and All Confusion Matrix are constructed to show percentage of correctly classified of samples. The quality of the results can be judged based on the mean square error (MSE), percentage of classification error (%E) and All Confusion Matrix (percentage of correctly and incorrectly classified of data). However, these results are not taken into account because good training result does not guarantee the ANN is able to achieve high preciseness in the testing stage.

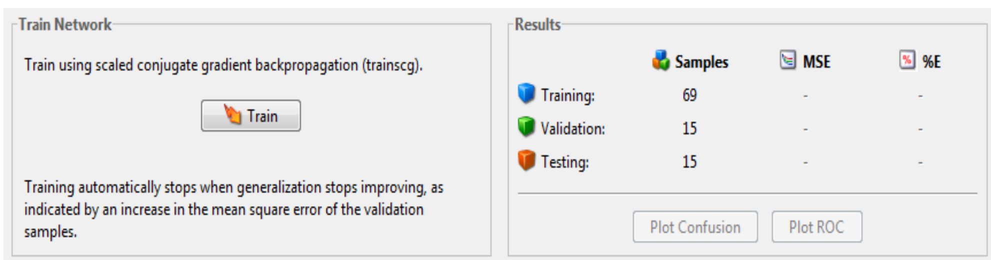


Figure 3.15: ANN Training Function with Analysis of Result

3.1.4 Integration and System Testing

After the ANN is trained, it is tested by integrating testing dataset to verify the quality of the trained neural network. The results are measured by MSE, %E and also All Confusion Matrix which plotted under "Plot Confusion". However, only All Confusion Matrix is chosen as benchmark to evaluate the preciseness of the ANN because classification results are clearly shown instead of showing figure only.

i. Testing Process

Step 1:

To begin the testing process, testing inputs and testing targets with total number of 103 sets of GLCM are selected. MSE and %E are shown as well to show the quality of the result.

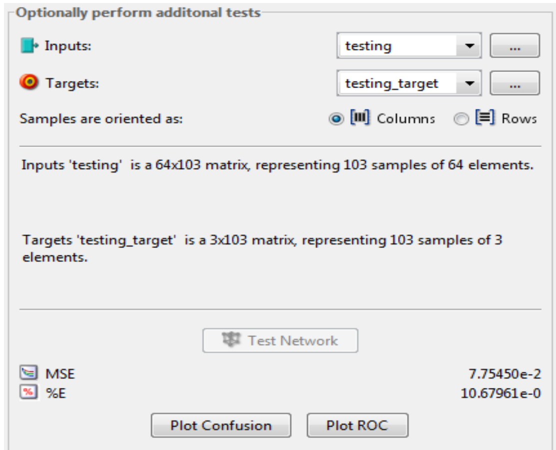


Figure 3.16: Selection of Testing and Target Datasets to Test the Trained ANN

Step 2:

All Confusion Matrix is plotted under the "Plot Confusion" function and from this matrix the number of correctly and incorrectly classified GLCM are shown. Output Class 1 representing channel, fault represented by Output Class 2 and GLCM that is not belongs to channel and fault is categorized in Output Class 3.

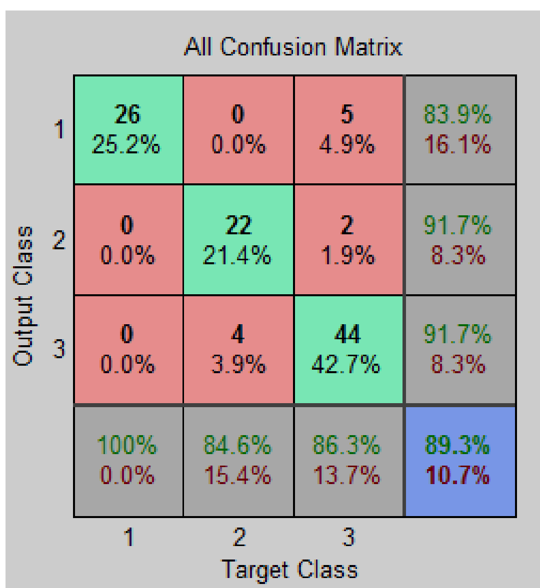


Figure 3.17: All Confusion Matrix

3.1.5 Operation and Maintenance

The quality and sophistication of the seismic objects recognition system are depending on the number of seismic images used in training process as well as expert guidance in recognizing channel and fault before training is started. It needs to be maintained by increasing the number of seismic numbers to training datasets so that the ANN able to learn more of these objects' patterns. More experts' guidance is needed as well to incorporate their tacit knowledge in order to improve the system sophistication and functionality.

3.2 HARDWARE AND SOFTWARE REQUIRED

i. **Required software:** Image Processing
   - OpendTect
   - MATLAB

   Construction of ANN
   - MATLAB

   Graphical User Interface
   - MATLAB

ii. **Required hardware:** Laptop or desktop with the minimum requirement of Window Vista, 2 GB RAM and 32-bit OS.

CHAPTER 4
RESULT AND DISCUSSION

4.1 SYSTEM INTERFACE AND OPERATION

Following are the screenshots of the channel and fault recognition system. Further details of system usage and user manual are shown in Appendix 1.

4.1.1 Main Interface

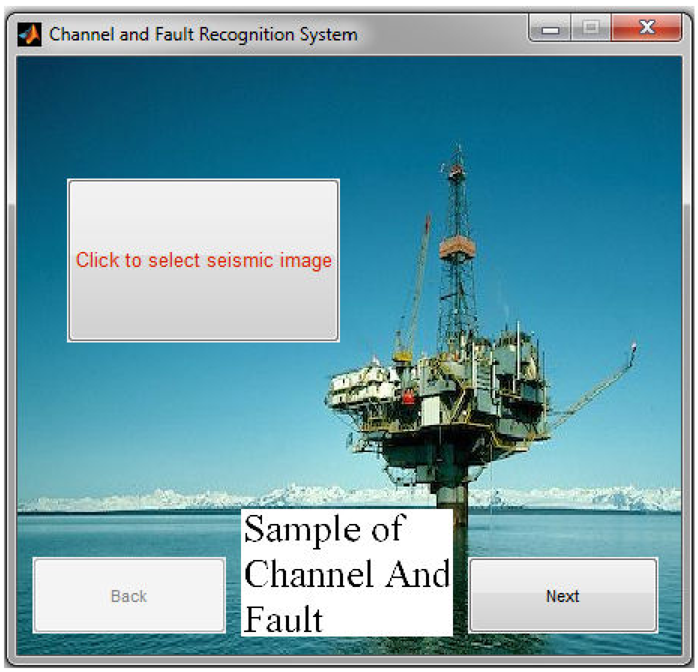


Figure 4.1: System Main Interface

4.2 ACCURACY TESTING

Testing results are gathered in confusion matrix form after the training and testing stage are completed. 3 type of ANN with different transfer function each are built and the testing results from these ANN are compared to determine which of the neural network has the highest accuracy of classification. Training of ANN with *log-sigmoid* and *purelin* function stopped at 4^{th} attempt while *tan-sigmoid* ANN training's result varied in every round of training. Hence, to make the analysis more comparable, 4 best testing results from *tan-sigmoid* ANN are taken and compare with the 4 testing results from the other 2 ANN. Other than All Confusion Matrix, F-score of each ANN is computed as well to measure test's accuracy.

4.2.1 All Confusion Matrix

ANN with *tan-sigmoid*

Table 4.1: Test 1 (ANN with *tan-sigmoid*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	26	0	5	83.9%
Fault	0	20	1	95.2%
Neither	0	6	45	88.2%
Percentage	100%	76.9%	88.2%	**88.3%**

Table 4.2: Test 2 (ANN with *tan-sigmoid*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	26	0	4	86.7%
Fault	0	22	2	91.7%
Neither	0	4	45	91.8%
Percentage	100%	84.6%	88.2%	**90.3%**

Table 4.3: Test 3 (ANN with *tan-sigmoid*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	26	0	4	86.7%
Fault	0	20	1	95.2%
Neither	0	6	46	88.5%
Percentage	100%	76.9%	90.2%	**89.3%**

Table 4.4: Test 4 (ANN with *tan-sigmoid*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	26	0	5	83.9%
Fault	0	22	2	91.7%
Neither	0	4	44	91.7%
Percentage	100%	84.6%	89.3%	**89.3%**

Table 4.5: Overall Average of Result's Accuracy (ANN with *tan-sigmoid*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	26	0	4.5	85.2%
Fault	0	21	1.5	93.3%
Neither	0	5	45	90.0%
Percentage	100%	80.8%	88.2%	**89.3%**

ANN with *log-sigmoid*

Table 4.6: Test 1 (ANN with *log-sigmoid*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	25	8	8	61.0%
Fault	0	12	3	80.0%
Neither	1	6	40	85.1%
Percentage	96.2%	46.1%	78.4%	**74.8%**

Table 4.7: Test 2 (ANN with *log-sigmoid*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	25	8	7	62.5%
Fault	0	13	0	100.0%
Neither	1	5	44	88.0%
Percentage	96.2%	50.0%	86.3%	**79.6%**

Table 4.8: Test 3 (ANN with *log-sigmoid*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	25	8	5	65.8%
Fault	0	14	1	93.3%
Neither	1	4	45	90.0%
Percentage	96.2%	53.8%	88.2%	**81.6%**

Table 4.9: Test 4 (ANN with *log-sigmoid*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	25	8	5	65.8%
Fault	0	14	1	93.3%
Neither	1	4	45	90.0%
Percentage	96.2%	53.8%	88.2%	**81.6%**

Table 4.10: Overall Average of Result's Accuracy (ANN with *log-sigmoid*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	25	8	6.3	63.6%
Fault	0	13.3	1.2	91.7%
Neither	1	4.7	43.5	88.4%
Percentage	96.2%	51.2%	85.3%	**79.4%**

ANN with *purelin*

Table 4.11: Test 1 (ANN with *purelin*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	26	6	10	61.9%
Fault	0	20	8	71.4%
Neither	0	0	33	100.0%
Percentage	100.0%	76.9%	64.7%	**76.7%**

Table 4.12: Test 2 (ANN with *purelin*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	21	0	6	77.8%
Fault	1	18	4	78.3%
Neither	4	8	41	77.4%
Percentage	80.8%	69.2%	80.4%	**77.7%**

Table 4.13: Test 3 (ANN with *purelin*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	26	6	10	61.0%
Fault	0	20	4	83.3%
Neither	0	0	37	100.0%
Percentage	100.0%	76.9%	72.5%	**80.6%**

Table 4.14: Test 4 (ANN with *purelin*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	26	6	10	61.0%
Fault	0	20	4	83.3%
Neither	0	0	37	100.0%
Percentage	100.0%	76.9%	72.5%	**80.6%**

Table 4.15: Overall Average of Result's Accuracy (ANN with *purelin*)

Seismic Objects	Channel	Fault	Neither	Percentage
Channel	24.75	4.5	9	64.8%
Fault	0.25	19.5	5	78.8%
Neither	1	2	37	92.5%
Percentage	95.2%	75.0%	72.5%	**78.9%**

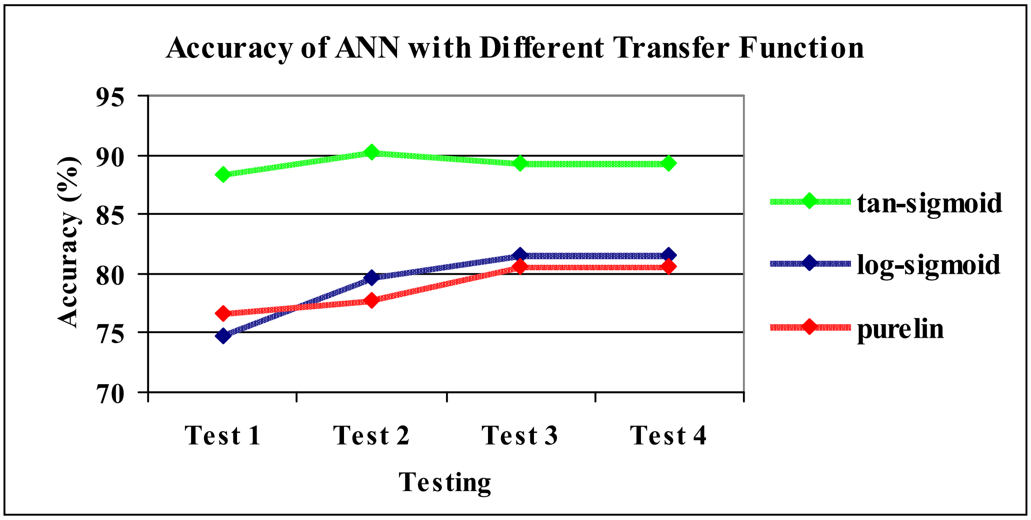


Figure 4.2: Overall Performance of 3 ANN with Different Transfer Functions

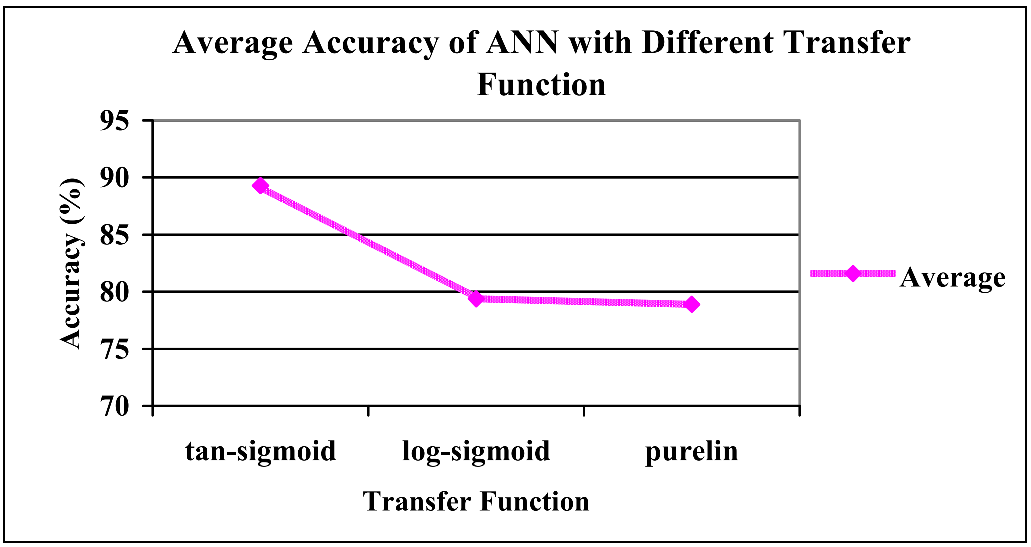


Figure 4.3: Average Percentage of Accuracy Achieved by Each ANN

Based on Figure 4.6, generally ANN with transfer function *tan-sigmoid* managed to achieve higher average percentage of correctly classified GLCM, which 89.3% as compared to the other 2 (*log-sigmoid,* 79.4% and *purelin,* 78.9%). The highest percentage recorded by this ANN is 90.3% or 93 out of 103 GLCM are correctly classified. Meanwhile, 81.6% and 80.6% are the highest record achieved by ANN with *log-sigmoid* and *purelin*. The results shown by third ANN are relatively poorer as the average percentage is 10.4% and 0.5% lower than the first and second ANN. The average of the ANN that applied *tan-sigmoid* transfer function is 89.3%.

All the ANN do not have difficulty in classifying channel but not fault and neither. Overall, ANN with tan-sigmoid showed that it is unable to recognize 5 fault's GLCM and 6 which do not belong to channel and fault. Misclassified rate of these 2 categories is relatively higher in the other ANN as well.

4.2.2 F-score

As mentioned earlier, other than All Confusion Matrix, another statistic measurement called F-score is used to measure the testing accuracy as well. It considers both the precision *p* and the recall *r* of the test to compute the final score. The formulas used are as below:

$$F_1 = 2 \bullet \frac{precision \bullet recall}{precision + recall} \tag{29}$$

$$\Pr ecision = \frac{tp}{tp + fn} \tag{30}$$

$$\operatorname{Re} call = \frac{tp}{tp + fn} \tag{31}$$

Where tp (true positive) = when both system and target output match with the current condition

fp (false positive) = when both system and target output do not match and the target output match with the condition

fn (false negative) = when system output, target output and condition are different

The terms *positive* and *negative* refer to the classifier's prediction (target output), and the terms *true* and *false* refer to whether that prediction corresponds to the external judgment (condition). Meanwhile, precision is the fraction of retrieved results that are relevant, while recall is the fraction of relevant results that are retrieved. In another word, high recall means that an algorithm returned most of the relevant results and high precision shows more relevant results are returned than irrelevant. Testing result from each ANN is tested by 3 conditions, which are the seismic object objects itself - channel, fault and neither. Following matrixes demonstrated how the true positive, false negative and false negative are determined according to the conditions.

Table 4.16: Condition = Channel

System output

Target Output

Seismic Objects	Channel	Fault	Neither
Channel	TP	FP	FP
Fault	FN	TN	FN
Neither	FN	FN	TN

Table 4.17: Condition = Fault

System output

Target Output

Seismic Objects	Channel	Fault	Neither
Channel	TN	FN	FN
Fault	FP	TP	FP
Neither	FN	FN	TN

Table 4.18: Condition = Neither

System output

Target Output

Seismic Objects	Channel	Fault	Neither
Channel	TN	FN	FN
Fault	FN	TN	FN
Neither	FP	FP	TP

The overall average of All Confusion Matrix of each ANN calculated earlier is used to measure the recall and precision in order to get the F-score. Following table showed the summary of each ANN's recall, precision as well as F-score.

Table 4.19: Summary of 3 ANNs' F-score

Type of ANN / Condition	ANN with transfer function *tan-sigmoid*		ANN with transfer function *log-sigmoid*		ANN with transfer function *purelin*	
	Recall	Precision	Recall	Precision	Recall	Precision
Channel	0.85	0.8	0.64	0.78	0.65	0.75
Fault	0.93	0.69	0.92	0.40	0.79	0.60
Neither	0.90	0.88	0.88	0.74	0.89	0.57
Average F-score	**0.83**		**0.69**		**0.69**	

As shown in the average of All Confusion Matrix, ANN with *tan-sigmoid* has better performance in term of higher percentage of correctly classified GLCM. Moving on to the F-score measurement, the average recall and precision of each condition of ANN is calculated to get the final F-score. Statistic showed that ANN with *tan-sigmoid* has the highest F-score compared to the rest, which is 0.83. The main contribution to this figure is from the high recall and precision average- 0.89 and 0.79 which are relatively higher than the other 2.

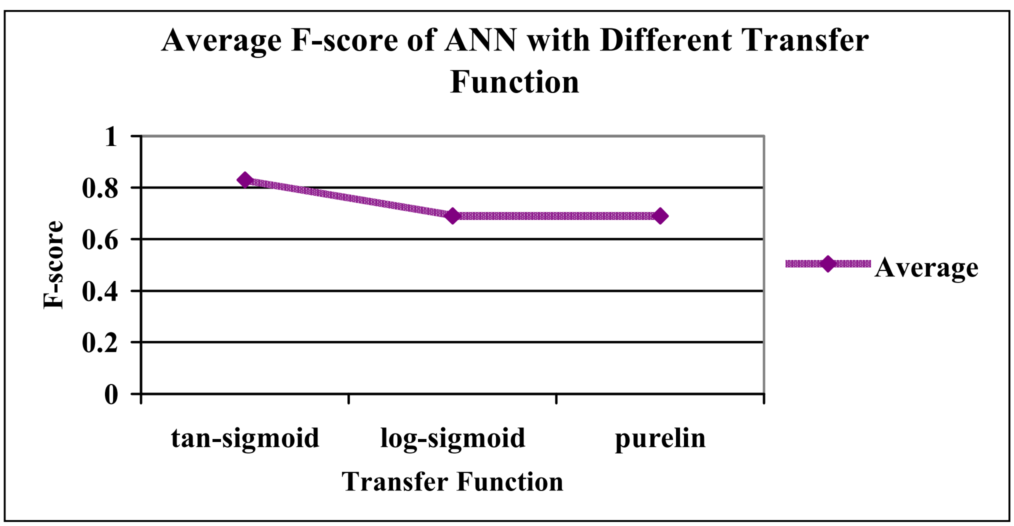


Figure 4.4: Average F-score Recorded by Each ANN with Different Transfer Function

Both of the ANN with *log-sigmoid* and *purelin* shared the same F-score, which is 0.69. Even though both of these ANN are having the F-score yet interpretation on recall and precision may differ. According to the average recall, ANN with *log-sigmoid* transfer function is able to return more relevant results as compared to ANN with *purelin*. By looking at this factor, since the precision of both are the same, ANN with *log-sigmoid* can be categorized as better classifier because its recall is higher.

4.2 USER TESTING

Upon the completion of the system functionality and accuracy testing, 7 Master and PhD students as well as 20 undergraduate students from Faculty of Geosciences and Petroleum Engineering Department are invited to do user testing which is important to know whether this system meet one of the objective set earlier or vice versa – to facilitate channel and fault recognition. There are 3 criteria used by the respondents to evaluate the system performance, which are functionality, usability and ability to facilitate channel and fault recognition. Their feedbacks are demonstrated in the graphs below.

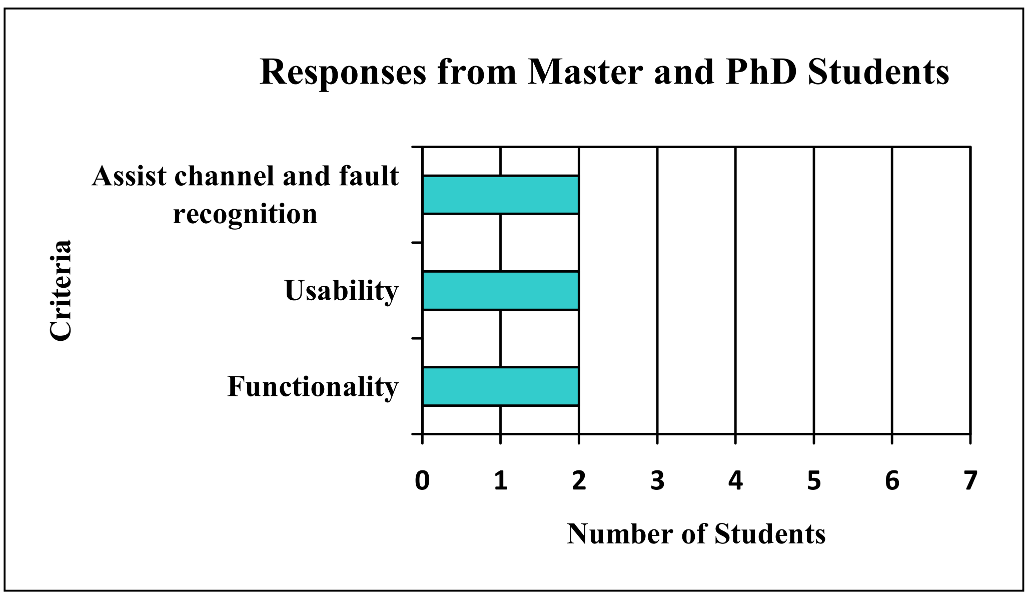


Figure 4.5: Responses Given by Master and PhD Students

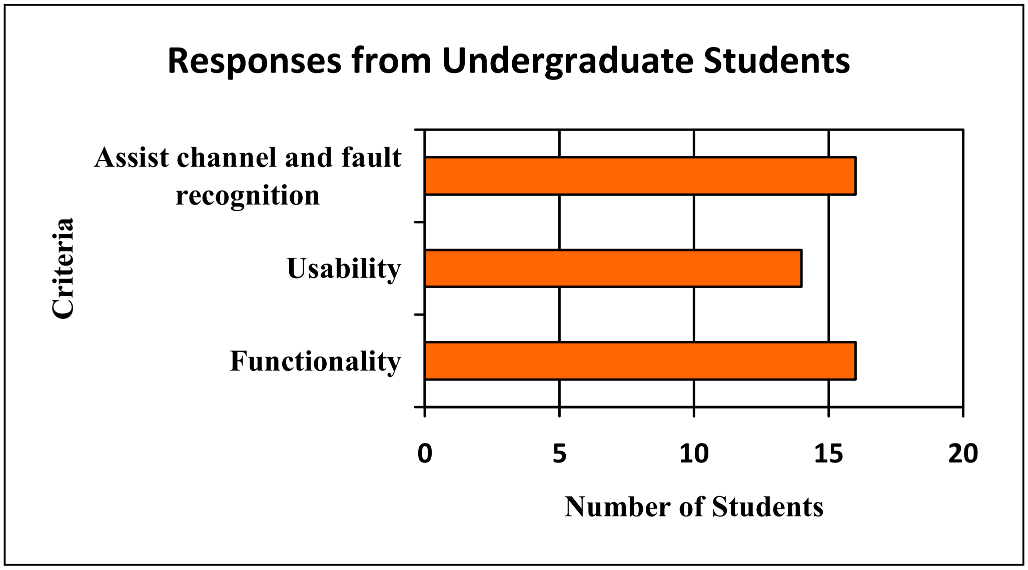


Figure 4.6: Responses Given by Undergraduate Students

Based on the statistic in Figure 4.5 and Figure 4.6, an extreme feedback can be seen where most of the Master and PhD students do not think that the functionality of the system is appropriate enough to involve in seismic objects recognition because they still believing that the recognition task cannot be automated as there are more aspects need to be considered before concluding the existence of channel and fault, such as the size of the objects and volume of reservoir stored beneath. Hence, this group of respondents does not agree that this system is able to facilitate channel and fault recognition.

However, from undergraduate students' point of view, 16 out of 20 of them are positive to this system's ability in guiding them to recognize channel and fault as foundation knowledge in this field. It is useful to them because with this system's help, they can learn independently without assistance from lecturer and tutor.

In short, Master and PhD students do not think that this system is able to help them in recognizing channel and fault because more sophisticated and advanced software has been using by them in their research. Yet this system is able to help undergraduate students in determining the basic pattern of channel and fault without guidance from experts.

CHAPTER 5
CONCLUSION AND RECOMMENDATION

5.1 CONCLUSION

Oil and gas exploration is a very crucial process indeed where it involves a lot of investment and decision making to determine a wealthy reservoir field. With the advancement of technology, the study and interpretation of subsurface are assisted by modern software which makes a seismic model more readable and understandable. In this 21^{st} century, human brain is still playing a vital role in performing seismic image interpretation to identify the existence of reservoir.

Data mining has emerged to be a very important research area that helps organizations to retrieve and make good use of the data on hand. Researchers have been conducting years of researches on the use of ANN to classify image, and this project has successfully demonstrated that ANN is able to classify objects in seismic image via analyzing GLCM which has not been done by other researchers. The ANN testing also proved that the combination of feed-forward network and transfer function *tan-sigmoid* has best performance in classifying the objects as compared to the usage of transfer function *log-sigmoid* and *purelin*. Unlike existing seismic image analyzing tools which required sophisticated skill to utilize, this seismic object recognition system required only one step to know check the existence of channel or fault.

In a nutshell, this project has achieved the target objectives set, which are to create a ANN which is able to recognize channel and fault and secondly to build a seismic objects recognition system which is able to facilitate channel and fault recognition. After undergo several training and testing process, a ANN with the accuracy of 90.3% is built and applied in the system. Seismic image usually contains more information that requires abundance experience and knowledge to interpret. Hence, this system is not aim to replace expert's role in determining the existence of reservoir but hopefully by improving its sophistication and functionality it can helps in facilitating seismic image interpretation in oil and gas industry.

5.2 RECOMMENDATION

Due to limited access to more seismic models, only 2 seismic models are used where seismic images are cropped and processed. The quality of the output is heavily depending on the training and learning process. The more seismic images used to train the network, the better quality of the learning process as the ANN is able to learn more about possible channel and fault's pattern. Hence, more seismic images are needed to expand the "knowledge" of ANN which is crucial in recognizing seismic objects.

Secondly, the intelligence applied this in project is just a tip of the iceberg. Channel and fault are just a small part analyzed by experts in oil and gas exploration. In order to further expand the functionalities of this system, continuous research need to be carried out to learn wider scope of criteria in seismic image classification that could provide more useful information which has higher reliability other than channel and fault, such as involving salt dome and anticline recognition.

Another issue concerned throughout the project is the sophistication of the system. One of the purposes running this project is to incorporate human knowledge and experience into an automated process, which is the seismic objects recognition system. By involving industry expert in the dataset preparation stage, the ANN will be able to learn industry expert's mind which is far more advance if compare to any other individual.

Last but not least, combination of different algorithms such as combination of supervised and unsupervised learning to form a hybrid neural network could be a good approach in the sense that such combination is possible to form a stronger solution which is able to provide greater efficiency and effectiveness in seismic objects recognition.

REFERENCES

[1] (n.d.). Seismic Survey. Retrieved June 11, 2012, from http://www.rriseismic.com/Frame%20Pages/Tech%20Pages/Seismic/seismic.html

[2] Pitas, I., and Kotropoulou, C. (1989). Texture Analysis and Segmentation of Seismic Images.Acoustics, Speech, and Signal Processing. ICASSP-89., International Conference, Vol. 3, pp. 1437-1440, 1989.

[3] (n.d.). Seismic Refraction Versus Reflection. Retrieved June, 11, 2012 from http://www.enviroscan.com/html/seismic_refraction_versus_refl.html

[4] Gao, D. (n.d.). Seismic textures aid exploration. Retrieved June, 11, 2012 from http://www.offshore-mag.com/articles/print/volume-62/issue-9/news/seismic-textures-aid-exploration.html

[5] Simaan, M. A. (2000). Artificial Neural Network Classification of Texture Orientations in Seismic Images. Geoscience and Remote Sensing Symposium, 2000. Proceeding. IGARSS 2000. IEEE 2000 International. Vol. 2, pp. 693-695.

[6] Velickov, S., Solomatine, D. (2000). Predictive Data Mining: Practical Examples. Artificial Intelligence in Civil Engineering. Proc, 2nd Joint Workshop, March 2000, Cottbus, Germany. ISBN: 3-934934-00-5.

[7] Du, H. (2010). Data Mining Techniques and Applications: An Introduction. Introduction and Principle of Data Mining (pp. 1-27). United Kingdom. Brendan George.

[8] Khanal, N., Kehoe, A., Kumar, A., MacDonald, A., Mueller, Martin., Plaisant, C., Ruecker, S., Sinclair, S., and Unsworth, J. (n.d.). Analytics: Naive Bayes. Retrieved 13 Jun, 2012 from http://gautam.lis.illinois.edu/monkmiddleware/public/tutorial/index.html

[9] Orbanz, P. (n.d.). Supervised vs. Unsupervised. Retrieved 16 June, 2012 from http://mlg.eng.cam.ac.uk/porbanz/teaching/slides_ml__supervised_vs_unsupervised.pdf

[10] Marsalli, M. (n.d.). McCulloch-Pitts Neurons. Retrieved 16 June, 2012, from http://www.mind.ilstu.edu/curriculum/modOverview.php?modGUI=212

[11] Harrigan, E., Kroh, R., Sandham, W. A., and Durrani, T. S. (1991). Seismic Wavelet Extraction Using Artificial Neural Networks. Artificial Neural Networks, 1991, Second International Conference, 18-20 November, pp. 95-99, 1991.

[12] Huang, K. Y. (2001) Neural Network for Robust Recognition of Seismic Patterns. Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on, 2001. Vol. 4, pp. 2930-2935.

[13] HU, M. K. (1962) Visual Pattern Recognition by Moment Invariants. Information Theory, IRE Transactions on, Vol. 8, No. 2, pp. 179-187, February 1962.

[14] Backpropagation. (n.d.). Retrieved 17 June, 2012 from http://radio.feld.cvut.cz/matlab/toolbox/nnet/backpr5.html

[15] Song, Q., Xiao, J., and Soh, Y. C. (1999). Robust Backpropagation Training Algorithm for Multilayered Neural Tracking Controller. Neural Networks, IEEE Transactions, Vol. 10, No. 5, pp. 1133-1141, September 1999.

[16] Heiss, M. (1996). Error-Minimizing Dead Zone for Basis Function Networks. Neural Networks, IEEE Transactions, Vol. 7, No. 6, pp. 1503-1506, November 1996.

[17] Huang, K. Y., and Yang, H. Z. (1992). A Hybrid Neural Network for Seismic Pattern Recognition. Neural Networks, IJCNN. International Joint Conference, Vol. 3, pp. 736-741, 7-11 June 1992.

[18] Diersen, S., Lee, E. J., Spears, D., Chen, P., and Wang, L. (2011). Classification of Seismic Windows Using Artificial Neural Networks. Procedia Computer Science. Proceeding of the International Conference on Computational Science, ICCS, Vol. 4, pp. 1572-1581, 2011.

[19] Iqbal, R. A. (2011) Empirical Learning Aided by Weak Domain Knowledge in the Form of Feature Importance. Multimedia and Signal Processing (CMSP), International Conference, Vol. 1, pp. 126-130, May 2011.

[20] Hecht-Nielsen, R. (1989) Theory of the Backpropagation Neural Network. Neural Network. IJCNN. International Joint Conference, Vol. 1, pp. 593-605, 1989.

[21] Moraes, D. R. S., Espindola, R. P., Evsukoff, A. G., and Ebecken, N. F. F. (2006). Cluster Analysis of 3D Seismic Data for Oil and Gas Exploration. WIT Transactions on Information and Communication Technology, Vol. 37, 2006.

[22] Kohonen, T. (1990). The Self-organizing Map. Proceedings of The IEEE, Vol. 78, No. 9, September 1990.

[23] The Waterfall Model Explained. (n.d.). Retrieved 21 June, 2012 from http://www.buzzle.com/editorials/1-5-2005-63768.asp

APPENDICES

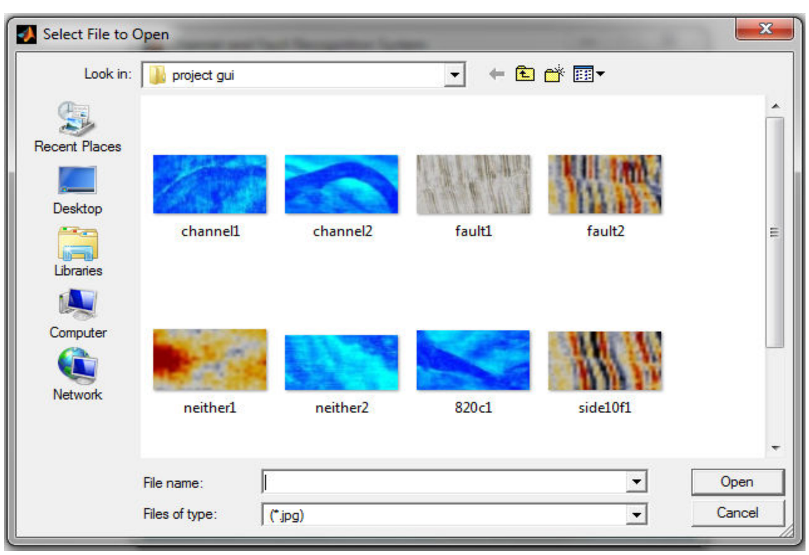
APPENDIX 1 USER INTERFACE MANUAL

Step 1:

Run the GUI.

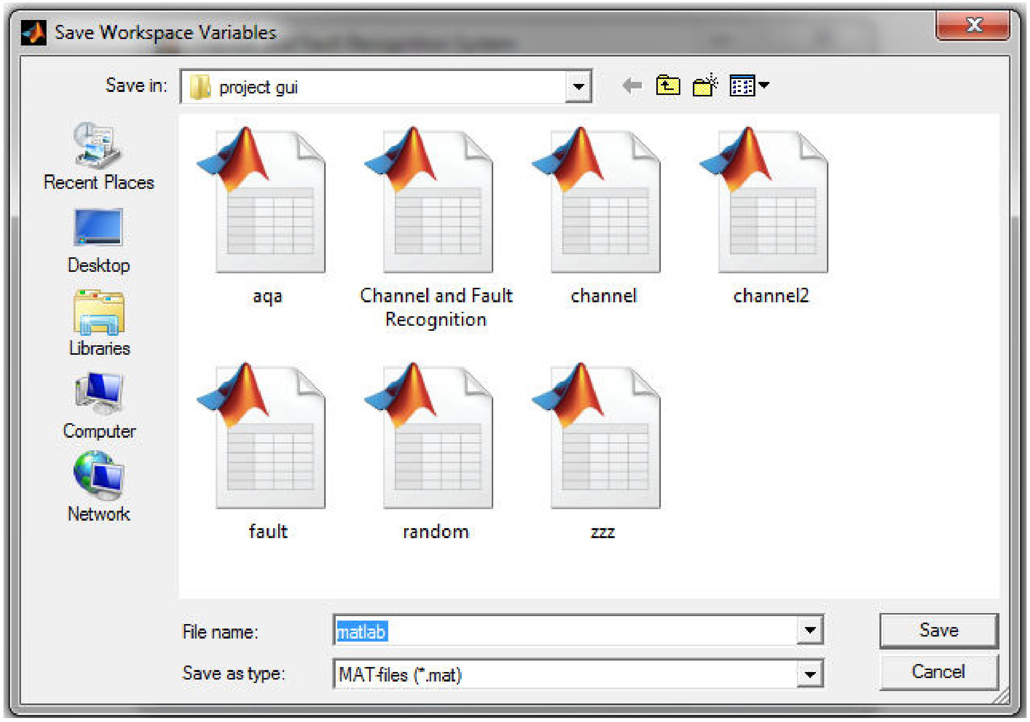
Step 2:

Select seismic image that need to be classified.

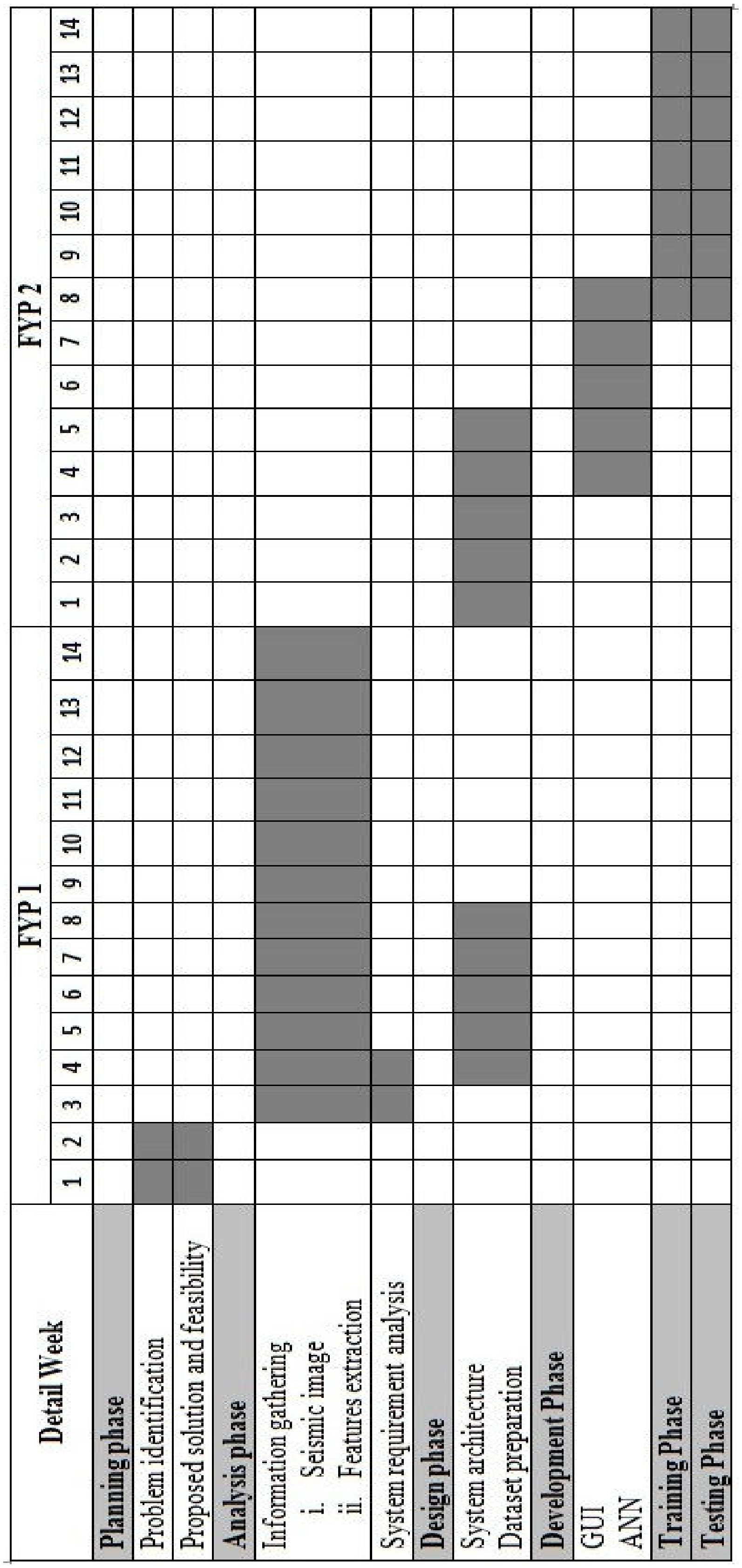


Step 3:

Save the GLCM generated in matfile format and result will be shown.



APPENDIX II PROJECT GANTT CHART



APPENDIX III TECHNICAL REPORT

Implementation of Objects Recognition in Seismic Image via Artificial Neural Network (ANN)

Dr. Vivian Yong Suet Peng [1,a] and Wan Chin Ee [2,b]
[1,2]Department of Computer and Information Sciences, Universiti Teknologi PETRONAS
Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia
[a]yongsuetpeng@petronas.com.my, [b]wance90@gmail.com

ABSTRACT

Seismic image processing is necessary in oil and gas exploration to identify the existence of potential reservoir by classifying the seismic image into different sections. These sections, also known as objects made up of different patterns which portraying the structure of subsurface. This project aims to develop a data mining algorithm embedded in a system that has ability to recognize the objects of channel and fault in seismic image. The method chosen is artificial neural network (ANN) which consists of input layer, hidden layer and output layer. Each layer is made up of numbers of neuron nodes to receive input data from preceding layers and output value to next layer until final output is determined from output layer. The ANN is trained and tested via MATLAB Neural Network Pattern Recognition Toolbox (nprtool) and MATLAB Neural Network Toolbox (nntool). 2-dimension (2D) seismic image is converted into gray scale image via MATLAB Image Processing Toolbox (imtool) and Grey-level co-occurrence matrix (GLCM) which serve as input to the ANN is retrieved from the gray scale image. Result is displayed by the system informing user whether the input image is channel, fault or neither both.

Keywords: *seismic image processing, artificial neural network, objects recognition, MATLAB*

I. INTRODUCTION

Seismic technique is originally developed by oil and gas industry and practiced in reservoir exploration since the 1930's to delineate the structure of the subsurface [1]. Accumulation of reservoir underneath can be visualize and predicted based on the processed data and information, hence decisive strategy can be made whether to further explore the natural resources stored in that particular location or vice versa. Reflection seismology is a widely implemented technique during the exploration phase to construct an accurate profile of the subsurface geology [2]. These techniques involve the measurement of travel time of seismic energy from surficial shots through subsurface to arrays of ground motion sensors or geophone [3]. The reflection and refraction wave of seismic energy will form seismic image at the end of activity.

A seismic image is characterized by a series of wiggle traces with alternating peak and trough amplitudes aligned laterally to form a strata reflection pattern [4]. Both the magnitude and variation of amplitude along or across the wiggle traces define the term seismic texture where it is analyzed by oil and gas experts to predict the presence of hydrocarbon in the subsurface via seismic image segmentation. This segmentation process is aim to partition a seismic image into multiple sections based on seismic texture (hereafter the term object will be used).

At current stage, the experts are responsible to differentiate and recognize all the existing objects in the seismic image because characteristic and structure of subsurface in accommodating reservoir can be reviewed from these objects. Examples of the objects are fault, channels, salt domes, and strong reflectors. Seismic objects recognition has already been carried out in oil and gas exploration decade years ago yet it has not been fully automated because of the heavy amount of knowledge involved in the decision making. Today, more and more researches have been carried out to study how seismic objects can be recognized by using artificial intelligence (AI) build up by data mining algorithm, such as artificial neural network.

A. *Problem Statement*

Although seismic image is studied by industry experts with the assistance of seismic analyzing tools, yet years of experience and tacit knowledge are still needed to interpret the result in which problem of over-depending on the experts could be raised. Efficiency and result of the manual interpretation is not guaranteed as it is strongly relying on the experience and knowledge of an individual has in mind. Furthermore, it would be difficult for students to recognize seismic objects independently without expert's help or utilize complicated seismic analyzing tools because insufficient of knowledge and skill in this field.

B. *Objective*

The objectives of this project are as outlined below.

1. To create ANN which is able to categorize channel and fault by analyzing Gray Level Co-Matrix (GLCM) of these objects.
2. To build a seismic objects recognition system that is able to facilitate the recognition of channel and fault.

C. *Scope of Study*

The scope of the study embodied:

1. Feasibility of ANN in performing pattern recognition.
2. Conversion of seismic image into GLCM and as input to the ANN.
3. 3D seismic models.

II. LITERATURE REVIEW

A. Data Mining and Processes

Data mining, also known as knowledge discovery and data mining serves its purpose by discovering hidden valuable knowledge or information by analyzing huge amount of data with a degree of certainty. It can be defined as a process of discovering new, interesting knowledge, such as patterns, associations, rules, changes, anomalies and significant structures from large amounts of data stored in data banks and other information repositories [5]. The output of data mining is information patterns [6]. These patterns are made up from non-trivial information that implicitly embedded inside the readily dataset which cannot be easily summarized via simple computations. For example, in this project, the unseen relationship between pixel values of seismic image will be revealed by using mathematical algorithms to differentiate the faults and channel objects.

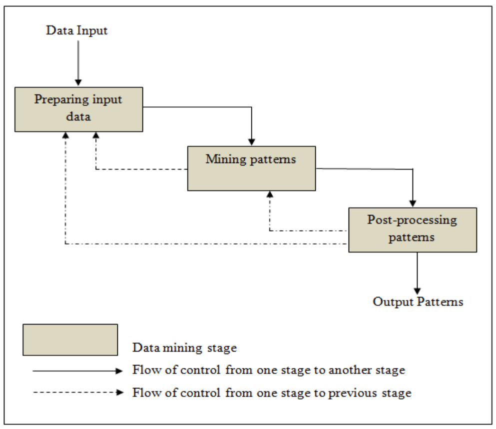


FIGURE 1: Data Mining Processes

B. Data Mining Algorithm – Artificial Neural Network (ANN)

There are many algorithms developed by experts to provide better version of problem solving solution in the field of medical, military, oil and gas exploration and many more. In this seismic image objects recognition project, our focus will be on ANN to recognize channel and fault. ANN is an artificial intelligence inspired by human brain neural network that consists of nodes and links located on different neuron layers. It is not a new concept used in data mining field as it has been introduced by Warren S. McCulloch (neuroscientist) and Walter Pitts (logician) in a paper entitled "A Logical Calculus of the Ideas Immanent in Nervous Activity" [7].

Huang (2001) presents a ANN which has been trained in several stages by training set which includes noise-free, low-noise, and misclassified seismic patterns for robust recognition of seismic image [8]. The ANN created is trained by supervised learning method and contains 1 input layer with 7 neurons, 1 hidden layer with 50 neurons and 1 output layer with 3 neurons. 3 types of seismic patterns are chosen to analyze, which are bright spot, pinch-out and horizontal reflection patterns. Seven moments such as invariant to translation that introduced by Hu in 1962 are also applied in this study for characterization of each seismic pattern [9].

Other than conventional ANN method, researchers have put in effort to develop non-conventional seismic pattern recognition algorithm which is able to differentiate each seismic pattern accurately via advance learning and training solutions. A hybrid neural network presented by Huang and Yang (1992) breakthrough the traditional design of neural network because both unsupervised and supervised learning are practiced in this neural network [10]. 4-layer network that contains an input layer (U1), a feature extracting layer (US), an intensity layer (UC) and an output layer (UO) are created. Result shows that this hybrid neural network is able to recognize all the 5 group of seismic patterns even with 20% of noise existing and in clustering part, 89% of seismic patterns are successfully being classified. More importantly, it is able to learn to extract critical features in training patterns and cluster them according to defined requirement.

In 2011, Diersen et al developed a feed-forward back propagation neural network algorithm is created based on Continuous Wavelet Transform (CWT) that allow the trained neural network to analyze waveforms in the time-frequency domain [11]. The difference between the preparation of seismic wave and GLCM is that the seismic pattern is converted into waveform based on time and amplitude of the image instead of converted into pixel values based on the grayscale intensity. In the same research, importance-aided neural network (IANN) is introduced by the authors to further discover the performance of ANN in seismic image classification. IANN is a bit different from the conventional ANN in the sense that Feature Relative Importance (*fri*) that extracted from experts' knowledge is embedded into the ANN. According to Iqbal, *fri* is a real valued approximation of a feature's importance given by the experts [12]. Significant result and performance are shown as classification accuracy of these 2 types of neural network is at above 99%.

Another popular ANN algorithm is called self organizing map (SOM) [13]. This method is adapted by Moraes et al (2006) in investigating cluster analysis of 3D seismic data for oil and gas exploration [14]. The SOM clustering method is a network model that practice neighborhood concept where the network learns to recognize neighbor sections in the training and also the topology of the learning set [15]. In this study, it is used to perform crisp clustering in which a dataset with over 223000 records about seismic attributes from a Brazilian oil field are used to test the algorithm. The advantage of using SOM compared to other ANN is that it is able to perform faster and simpler because it has only 2 layers of neuron. However, it required the image data in a seismic image to behave similarly or else the quality of the output will be affected.

III. METHODOLOGY

A. System Development Cycle

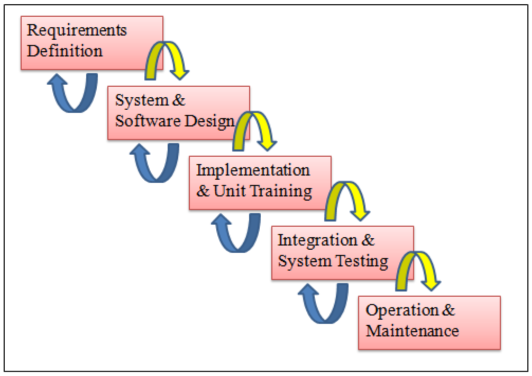


FIGURE 4: Waterfall Model

In this project, the "Waterfall Model" is chosen. This approach is the first "Process Model" to be introduced and has been widely implemented in project completion [16]. There are 5 phases included in this model: requirements definition, system and software design, implementation and system testing, integration and system testing and lastly operation and maintenance. In order to complete this project, the model has been customized based on the project requirements. The third phase is named as implementation and unit training because ANN training is a major step need to be done to complete the project. Each phase consists of specific set of activities that need to be completed before the following phase can proceed. This model maintains a disciplinary approach in developing a solution as all the activities are clearly defined in each phase.

B. System Architecture

FIGURE 5 shows the complete system architecture of the project prototype. The system prototype is working based on 3 main functions – image processing, ANN and result. In image processing section, seismic image selected by user is converted into gray-scale image automatically by the system and GLCM is extracted. Next the GLCM is input into a well-trained ANN which is embedded in the system. The ANN will read the GLCM pattern and group accordingly based on from what it learned in training stage. Eventually, result is shown in a message box informing user whether the selected image is channel, fault or neither both. 3 ANN with different transfer function each (*tan-sigmoid*, *log-sigmoid* and *purelin*) are created and the neural network with highest accuracy is chosen to be implemented in the system.

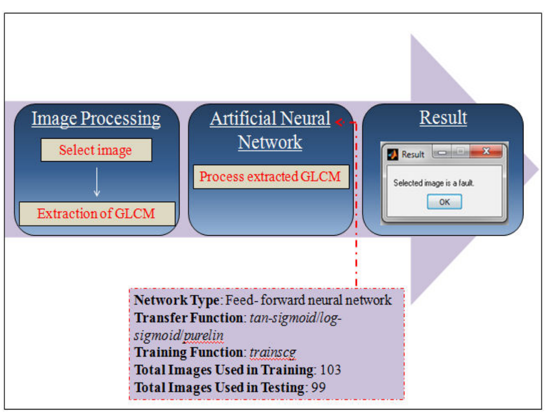


FIGURE 5: System Architecture

C. Development Tools Required

- MATLAB R2009b
- OpendTect
- Platform Windows Vista 32Bit

IV. RESULT AND DISCUSSION

A. Dataset

The training and testing dataset used contained 202 2D seismic images' GLCM which are taken from 2 3D seismic models. First model named Bundi PM-311 with size of 744 square kilometers is provided by a postgraduate student from Department of Geosciences. 40 2D slices from different top and side layers which contain channel and fault are taken from this model. Second model entitled Netherlands Offshore F3 Block is taken from OpendTect Open Seismic Repository web site. This model covered 384 square kilometers of Offshore, Northsea that located at Netherlands. 32 2D slices from top and side layers are captured and sections that contained channel and fault are taken as well to build the dataset.

B. System Interface

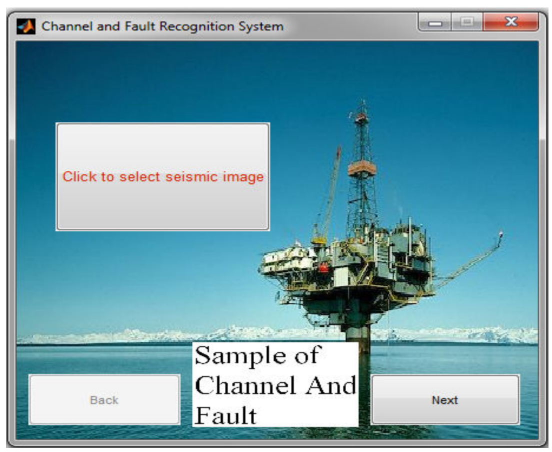


FIGURE 6: System Main Interface

C. Accuracy Testing

Testing results are gathered in matrix form after the training and testing stage are completed. 3 type of ANN with different transfer function each are built and the testing results of these ANN are compared to determine which of the neural network has the highest accuracy of classification. Training of ANN with *log-sigmoid* and *purelin* function stopped at 4^{th} attempt while *tan-sigmoid* ANN training's result varied in every round of training. Hence, to make the analysis more comparable, 4 best testing results from *tan-sigmoid* ANN are taken and compare with the 4 testing results from the other 2 ANN. 2 test's accuracy measure - All Confusion Matrix and F-score are computed to measure accuracy of each ANN. In addition, user testing is conducted as well to collect the feedbacks from target users, which are postgraduate students and undergraduate students.

i. All Confusion Matrix

TABLE 1: Average Accuracy (ANN with *tan-sigmoid*)

Objects	Channel	Fault	Neither	Percentage
Channel	26	0	4.5	85.2%
Fault	0	21	1.5	93.3%
Neither	0	5	45	90.0%
Percentage	100%	80.8%	88.2%	**89.3%**

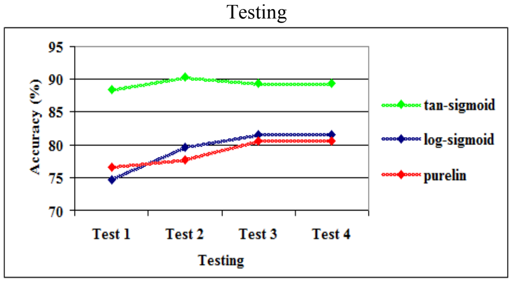
TABLE 2: Average Accuracy (ANN with *log-sigmoid*)

Objects	Channel	Fault	Neither	Percentage
Channel	25	8	6.3	63.6%
Fault	0	13.3	1.2	91.7%
Neither	1	4.7	43.5	88.4%
Percentage	96.2%	51.2%	85.3%	**79.4%**

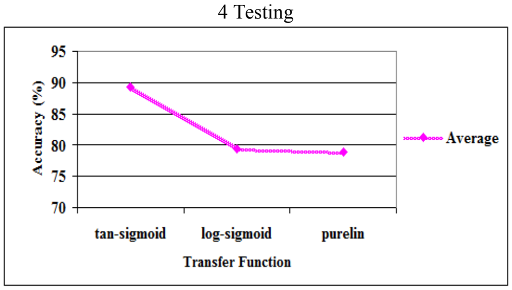
TABLE 3: Average Accuracy (ANN with *purelin*)

Objects	Channel	Fault	Neither	Percentage
Channel	24.75	4.5	9	64.8%
Fault	0.25	19.5	5	78.8%
Neither	1	2	37	92.5%
Percentage	95.2%	75.0%	72.5%	**78.9%**

GRAPH 1: Accuracy of ANN with Different Transfer Function in 4 Testing



GRAPH 2: Average Accuracy of ANN with Different Transfer Function in 4 Testing



Generally, ANN with transfer function *tan-sigmoid* managed to achieve higher average percentage of correctly classified GLCM, which 89.3% as compared to the other 2 (*log-sigmoid,* 79.4% and *purelin,* 78.9%). The highest percentage recorded by this ANN is 90.3% or 93 out of 103 GLCM are correctly classified. Meanwhile, 81.6% and 80.6% are the highest record achieved by ANN with *log-sigmoid* and *purelin*. The results shown by third ANN are relatively poorer as the average percentage is 10.4% and 0.5% lower than the first and second ANN. The average of the ANN that applied *tan-sigmoid* transfer function is 89.3%.

ii. F-score

As mentioned earlier, other than All Confusion Matrix, another statistic measurement called F-score is used to measure the testing accuracy as well. It considers both the precision p and the recall r of the test to compute the final score. The formulas used are as below:

$$F_1 = 2 \bullet \frac{precision \bullet recall}{precision + recall}$$

$$\Pr ecision = \frac{tp}{tp + fn}$$

$$\operatorname{Re} call = \frac{tp}{tp + fn}$$

Where

tp (true positive) = when both system and target output match with the current condition.

fp (false positive) = when both system and target output do not match and the target output match with the condition.

fn (false negative) = when system output, target output and condition are different.

The terms *positive* and *negative* refer to the classifier's prediction (target output), and the terms *true* and *false* refer to whether that prediction corresponds to the external judgment (condition). Meanwhile, precision is the fraction of retrieved results that are relevant, while recall is the fraction of relevant results that are retrieved. In another word, high recall means that an algorithm returned most of the relevant results and high precision shows more relevant results are returned than irrelevant. Testing result from each ANN is tested by 3 conditions, which are the seismic object objects itself - channel, fault and neither.

TABLE 4: Condition = Channel

Seismic Objects		System Output		
		Channel	Fault	Neither
Target Output	Channel	TP	FP	FP
	Fault	FN	TN	FN
	Neither	FN	FN	TN

TABLE 5: Condition = Fault

Seismic Objects		System Output		
		Channel	Fault	Neither
Target Output	Channel	TN	FN	FN
	Fault	FP	TP	FP
	Neither	FN	FN	TN

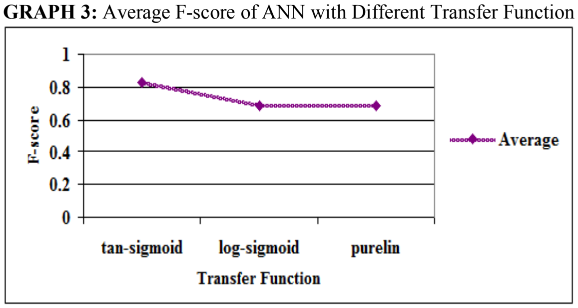
TABLE 6: Condition = Either

	Seismic Objects	System Output		
		Channel	Fault	Neither
Target Output	Channel	TN	FN	FN
	Fault	FN	TN	FN
	Neither	FP	FP	TP

TABLE 7: Summary of 3 ANN F-score

Type of ANN / Condition	ANN with transfer function *tan-sigmoid*		ANN with transfer function *log-sigmoid*		ANN with transfer function *purelin*	
	Recall	Precision	Recall	Precision	Recall	Precision
Channel	0.85	0.8	0.64	0.78	0.65	0.75
Fault	0.93	0.69	0.92	0.40	0.79	0.60
Neither	0.90	0.88	0.88	0.74	0.89	0.57
Average F-score	**0.83**		**0.69**		**0.69**	

GRAPH 3: Average F-score of ANN with Different Transfer Function



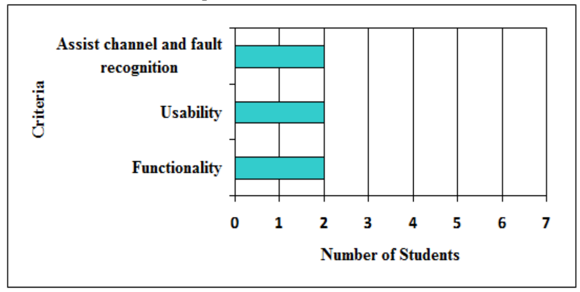
The average recall and precision of each condition of ANN is calculated to get the final F-score. Statistic showed that ANN with *tan-sigmoid* has the highest F-score compared to the rest, which is 0.83. The main contribution to this FIGURE is from the high recall and precision average- 0.89 and 0.79 which are relatively higher than the other 2.

Both of the ANN with *log-sigmoid* and *purelin* shared the same F-score, which is 0.69. Even though both of these ANN are having the F-score yet interpretation on recall and precision may differ. According to the average recall, ANN with *log-sigmoid* transfer function is able to return more relevant results as compared to ANN with *purelin*. By looking at this factor, since the precision of both are the same, ANN with *log-sigmoid* can be categorized as better classifier because its recall is higher.

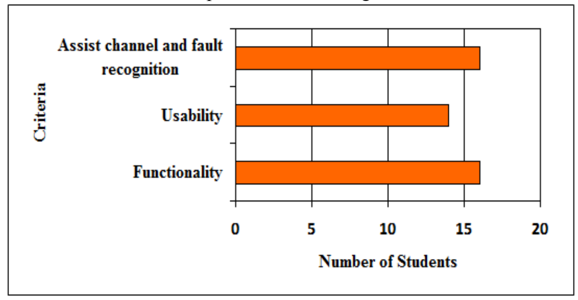
D. User Testing

7 Master and PhD students as well as 20 undergraduate students from Faculty of Geosciences and Petroleum Engineering Department are invited to do user testing which is important to know whether this system meet one of the objective set earlier or vice versa – to facilitate channel and fault recognition. There are 3 criteria used by the respondents to evaluate the system performance, which are functionality, usability and ability to facilitate channel and fault recognition.

GRAPH 4: Responses from Master and PhD Students



GRAPH 5: Responses from Undergraduate Students



Based on the statistic in **GRAPH 4** and **GRAPH 5**, an extreme feedback can be seen where most of the Master and PhD students do not think that the functionality of the system is appropriate enough to involve in seismic objects recognition because they still believing that the recognition task cannot be automated as there are more aspects need to be considered before concluding the existence of channel and fault, such as the size of the objects and volume of reservoir stored beneath. Hence, this group of respondents does not agree that this system is able to facilitate channel and fault recognition.

However, from undergraduate students' point of view, 16 out of 20 of them are positive to this system's ability in guiding them to recognize channel and fault as foundation knowledge in this field. It is useful to them because with this system's help, they can learn independently without assistance from lecturer and tutor.

In short, Master and PhD students do not think that this system is able to help them in recognizing channel and fault because more sophisticated and advanced software has been using by them in their research. Yet this system is able to help undergraduate students in determining the basic pattern of channel and fault without guidance from experts.

V. CONCLUSION

Oil and gas exploration is a very crucial process where it involves a lot of investment and decision making to determine a wealthy reservoir field. Today, human brain is still playing a vital role in performing seismic objects interpretation in this industry.

Data mining has emerged to be a very important research area that helps organizations to retrieve and make good use of the data on hand. Researchers have been conducting years of researches on the use of ANN to classify image, and this project has successfully demonstrated that ANN is able to classify objects in seismic image via analyzing GLCM which has not been done by other researchers. The ANN testing also proved that the combination of feed-forward network and transfer function *tan-sigmoid* has best performance in classifying the objects as compared to the usage of transfer function *log-sigmoid* and *purelin*. Unlike existing seismic image analyzing tools which required sophisticated skill to utilize, this seismic object recognition system required only one step to know check the existence of channel or fault.

In a nutshell, this project has achieved the target objectives set, which are to create a ANN which is able to recognize channel and fault and secondly to build a seismic objects recognition system which is able to facilitate channel and fault recognition. After undergo several training and testing process, a ANN with the accuracy of 90.3% is built and applied in the system.

VI. RECOMMENDATION

Due to limited access to more seismic models, only 2 seismic models are used where seismic images are cropped and processed. The quality of the output is heavily depending on the training and learning process. The more seismic images used to train the network, the better quality of the learning process as the ANN is able to learn more about possible channel and fault's pattern. Hence, more seismic images are needed to expand the "knowledge" of ANN which is crucial in recognizing seismic objects.

The intelligence applied this in project is just a tip of the iceberg. Channel and fault are just a small part analyzed by experts in oil and gas exploration. In order to further expand the functionalities of this system, continuous research need to be carried out to learn wider scope of criteria in seismic image classification that could provide more useful information which has higher reliability other than channel and fault, such as involving salt dome and anticline recognition.

Another issue concerned throughout the project is the sophistication of the system. One of the purposes running this project is to incorporate human knowledge and experience into an automated process, which is the seismic objects recognition system. By involving industry expert in the dataset preparation stage, the ANN will be able to learn industry expert's mind which is far more advance if compare to any other individual.

VII. ACKNOWLEDGEMENT

First and foremost, I would like to take this opportunity to acknowledge utmost gratitude to my Final Year Project supervisor, Dr. Vivian Yong Suet Peng, who has been given abundance of constructive advices as well as guidance until completion of this project. Her willingness to spend her time so generously has been very much appreciated. Secondly, I am grateful for the assistance given by Research and Innovation Office of Universiti Teknologi PETRONAS for booking me a desktop to do seismic image processing. Not forgotten to thank Master and PhD students from Faculty of Geosciences and Petroleum Engineering who have been providing number of seismic images to this project.

REFERENCES

[1] (n.d.). Seismic Survey. Retrieved June 11, 2012, from http://www.rriseismic.com/Frame%20Pages/Tech%20Pages/Seismic/seismic.htm

[2] Pitas, I., and Kotropoulou, C. (1989). Texture Analysis and Segmentation of Seismic Images.Acoustics, Speech, and Signal Processing. ICASSP-89., International Conference, Vol. 3, pp. 1437-1440, 1989.

[3] (n.d.). Seismic Refraction Versus Reflection. Retrieved June, 11, 2012 from http://www.enviroscan.com/html/seismic_refraction_versus_refl.html

[4] Gao, D. (n.d.). Seismic textures aid exploration. Retrieved June, 11, 2012 from http://www.offshore-mag.com/articles/print/volume-62/issue-9/news/seismic-textures-aid-exploration.html

[5] Velickov, S., Solomatine, D. (2000). Predictive Data Mining: Practical Examples. Artificial Intelligence in Civil Engineering. Proc, 2nd Joint Workshop, March 2000, Cottbus, Germany. ISBN: 3-934934-00-5.

[6] Du, H. (2010). Data Mining Techniques and Applications: An Introduction. Introduction and Principle of Data Mining (pp. 1-27). United Kingdom. Brendan George.

[7] Marsalli, M. (n.d.). McCulloch-Pitts Neurons. Retrieved 16 June, 2012, from http://www.mind.ilstu.edu/curriculum/modOverview.php?modGUI=212

[8] Huang, K. Y. (2001) Neural Network for Robust Recognition of Seismic Patterns. Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on, 2001. Vol. 4, pp. 2930-2935.

[9] HU, M. K. (1962) Visual Pattern Recognition by Moment Invariants. Information Theory, IRE Transactions on, Vol. 8, No. 2, pp. 179-187, February 1962.

[10] Huang, K. Y., and Yang, H. Z. (1992). A Hybrid Neural Network for Seismic Pattern Recognition. Neural Networks, IJCNN. International Joint Conference, Vol. 3, pp. 736-741, 7-11 June 1992.

[11] Diersen, S., Lee, E. J., Spears, D., Chen, P., and Wang, L. (2011). Classification of Seismic Windows Using Artificial Neural Networks. Procedia Computer Science. Proceeding of the International Conference on Computational Science, ICCS, Vol. 4, pp. 1572-1581, 2011.

[12] Iqbal, R. A. (2011) Empirical Learning Aided by Weak Domain Knowledge in the Form of Feature Importance. Multimedia and Signal Processing (CMSP), International Conference, Vol. 1, pp. 126-130, May 2011.

[13] Hecht-Nielsen, R. (1989) Theory of the Backpropagation Neural Network. Neural Network. IJCNN. International Joint Conference, Vol. 1, pp. 593-605, 1989.

[14] Moraes, D. R. S., Espindola, R. P., Evsukoff, A. G., and Ebecken, N. F. F. (2006). Cluster Analysis of 3D Seismic Data for Oil and Gas Exploration. WIT Transactions on Information and Communication Technology, Vol. 37, 2006.

[15] Kohonen, T. (1990). The Self-organizing Map. Proceedings of The IEEE, Vol. 78, No. 9, September 1990.

[16] The Waterfall Model Explained. (n.d.). Retrieved 21 June, 2012 from http://www.buzzle.com/editorials/1-5-2005-63768.asp