INTELLIGENT SYSTEM FOR ESTIMATING THE POROSITY IN SEDIMENTS FROM THE ANALYSIS OF SIGNALS GPR

Eduardo Henrique Silveira de Araújo
Departamento de Estatística - DEST
Universidade Federal do Rio Grande do Norte- UFRN
Natal/ Rio Grande do Norte
ehsa@ccet.ufrn.br

Adrião Duarte Dória Neto
Programa de Pós-Graduação em Ciência e Engenharia de Petróleo – PPGCEP/CCET – DCA/CT - UFRN
Universidade Federal do Rio Grande do Norte- UFRN
Natal/ Rio Grande do Norte
adriao@dca.ufrn.br

Francisco Pinheiro Lima Filho
Programa de Pós-Graduação em Geodinâmica e Geofísica – PPGG/CCET – Departamento de Geologia – DG
Universidade Federal do Rio Grande do Norte- UFRN
Natal/ Rio Grande do Norte
pinheiro@geologia.ufrn.br

Abstract - This paper presents the elaboration of a methodological propose for the development of an intelligent system, able to automatically achieve the effective porosity, in sedimentary layers, from a database bank built with information from the Ground Penetrating Radar – GPR. The intelligent system was built to model the relation between the porosity (response variable) and the electromagnetic attribute from the GPR (explicative variables). Using it, the porosity was estimated using the artificial neural network (Multilayer Perceptron – MLP) and the multiple linear regression. The data from the response variable and from the explicative variables were acquired in laboratory and in GPR surveys outlined in controlled sites, on site and in laboratory. The proposed intelligent system has the capacity of estimating the porosity from any database bank available, which has the same variables used in this paper. The architecture of the neural network used can be modified according to the existing need, adapting to the data bank available. The use of the multiple linear regression model allowed the identification and quantification of the influence (level of effect) of each explicative variable in the estimation of the porosity. The proposed methodology an innovative approach the use of the GPR, not only for the imaging of the sedimentary geometry and faces, but mainly for the automatically achievement of the porosity – one of the most important parameters for the characterization of reservoir rocks (from petroleum or water).

Keywords: Porosity, Artificial neural networks (ANN), GPR, Intelligent system

1. INTRODUCTION

The oil industry has made a big effort to have a better comprehension of the heterogeneities (depositional and deformational) and geometries of the oil reservoirs. These informations is the base for success of the exploration and exploitation of hydrocarbonets. Therefore, it is necessary the knowledge of the depositional architecture of the depositional systems involved, and also the petrophysical characteristics of the reservoir rocks that might function as ducts or flow barriers. On the other hand, the knowledge of the petrophysical properties from the reservoir rocks makes it possible to elaborate models that explain in a more realistic and coherent way the fluid migration. Among these properties, the porosity and permeability are the most important parameters for the engineer and/or reservoir geologist.

As the available information about the reservoirs is, most of the time, restricted to boreholes (polls testimonials, through samples and electric/radioactive profiles), the oil industry has been recently using, the so called "reservoirs analogous outcrops". These outcrops are chosen due to possessing some characteristics that hold a strict relation with the reservoir of interest, such as: the sedimentary packages geometry, stratigraphic relations, fracturing patterns, its age and associated tectonic. One of the most used geophysical methods on the characterization of reservoirs analogous outcrops is the GPR, referred in the international literature as Ground Penetrating Radar, or simply GPR, and is used by several authors, including [1].

On this paper we present a methodological proposition to obtain the porosity of sedimentary layers, indirectly, which might represent a technological leap on the use of the GPR method for characterization of petroliferal reservoirs analogous outcrops. The work had as an objective to develop a modeling methodology through an intelligent system to estimate porosity values, based on results acquired from GPR and data from electromagnetic variables from siliciclastic materials, disposed on strata, which simulates the sedimentary environment. The GPR is a geophysical imaging method for the shallow surface, composed of a set of antennas for transmission and reception of high frequency electromagnetic waves (10 and 2600 MHz) and a control unit for registration of the signals which are reflected and processed producing images of the features found underground, called Radargram. The differences in the electric properties of the materials crossed by the electromagnetic wave produce higher or lower reflection of the GRP signals. The intensity or attenuation of the signal reflection is directly associated to the dielectric constant, the magnetic permissivity, the electric conductivity [2]. The disposition or arrangement of the mineral grains and the mean diameter of the sedimentary particles also control the development of the porosity on the sedimentary deposit, which on the other hand will influence the electric resistivity of the analyzed layer. The higher the porosity, more empty spots there will be and, consequently, it will influence the final dielectric constant obtained. With the value of the reflection intensity, measured by the estimative of the reflection energy of the environment, which can be found according to the signal amplitude, we can, using statistic modeling and an artificial neural network, relate porosity values (response variable) with variations of the dielectric constant, with the variations of pulse frequency of the signal and with variations of the reflection energy (explanatory variables).

Aiming to identify the effects of electromagnetic variables and model the variability of the porosity on a sedimentary deposit, from data acquired from GPR, we made experiments on 2 controlled sites. The database for the variables on this research was obtained from experiments, from synthetic built radargrams and electrical constant measurements of the sedimentary material used on the experiments. The data were treated by the statistic technique of Multivariate Data Analysis. Furthermore, it was also used the Artificial Neural Network technique to model the relation between the "porosity" response and the "electromagnetic explanatory" variables.

Accordingly, for the realization of this study an Intelligent System was built based on computational intelligence techniques and statistic modeling to estimate the petrophysic propriety sedimentary layer porosity of reservoirs analogous outcrops (system output), based on data from GPR acquisitions of the electromagnetic variables (system input).

2. MATERIALS AND METHODS

In order to build the database of the porosity, dielectric constant, GPR antenna frequency and reflection energy variables to be implemented on the intelligent system, a set of procedures were done.

Firstly experiments were idealized and built on controlled sites on field and laboratory. On the experiments in was used samples of sedimentary material of homogeneous, thick, medium and thin sand and also sand with granulometry of 0.125mm and 0.250mm. The second procedure was to obtain porosity values and dielectric constant from the sand samples on the Sedimentology and UFRN's Telecommunication Engineering sand. The third procedure was to obtain data of the dielectric constant variable, antenna frequency and reflection energy from the acquisitions with GPR performed on the programmed experiments and the acquisitions database from the UFRN's Stratigraphic Analysis Laboratory (SAL), of department of geology UFRN. The fourth procedure, also for obtaining data from the variables, was to build, based on the realized experiments, synthetic scenarios of sedimentary layers with the same kind of materials from the experiments and acquisitions from LAE. To build the synthetic scenarios, the Reflex system was used. The fifth procedure was to idealize and implement the intelligent system built with the Artificial Neural Networks models of the MLP kind to estimate the porosity and the use of the multivariate model of data analysis Multiple Linear Regression to identify the effect of the explainable variables.

A. Analysis variables present on the reservoir rocks

Porosity

The reservoir rocks have, in their characteristics, mechanical and physical properties important regarding the storage and migration of fluids. In the oil industry, to present a methodology that allows, with some degree of reliability, to calculate or estimate from geophysical surveys with GPR or seismic, petrophysical property values porosity in reservoir rocks is highly desirable and the size of its values is directly related to the storage capacity of the rock [3] and [4]. The total porosity ϕ_t is defined as the ratio of the void volume of a rock (pores, channels, fissures, vugs), whether or not interconnected, and the total volume thereof. While the effective porosity ϕ_e represents the space occupied by fluids, which can be moved through the porous medium may be determined by the relationship between the volume of voids interconnected in a rock and its total volume.

Dieletric Constant

A material is considered "dielectrical" if it has the ability to store energy when an external electric field is applied to it. The dielectric constant is the material's dielectrical part which forms the rock and is measured on laboratory by an equipment known as Probe, a dimensionless quantity.

B. The GPR Method

The GPR or Georadar, as avoid contractions also known, is an electromagnetic geophysical method that generates high resolution images of shallow structures and features found on subsurface, based on register of the double travel time of electromagnetic waves [5]. Avoid contractions a set of non-invasive geophysical techniques which detects electrical discontinuities on shallow sub surfaces depth of up to 80m, from generation, transmission, propagation, reflection and reception of discrete electromagnetic pulses of low and high frequencies [6] (Neal, 2001). The GPR method is based on the emission, done by a transmission antenna, of a short electromagnetic energy pulse with duration of a nanosecond (10-9 s) and a frequency range from 10 to 2600MHz, which is irradiated or transmitted on the ground. On its displacement, the energy alters its speed when it hits materials with different dielectrical properties. On top of that, these pulses emitted by the transmission antenna suffer reflections, refractions and diffractions on discontinuities present on the subsurface (different physical properties). The wave is now reflected and captured when returning to surface by a receptive antenna. These signals are treated by a control unit which records the double travel time, the pulse and, later on, are amplified and registered.

The System Reflex

The RELFEXW 6.1 system [7], produced by the Sandmeier Geophysical Software has the function of processing seismic, electromagnetic or acoustic reflection data and transmission data and signal refraction. The system is composed of processing modules. On this paper it was used the modules: 2D data analysis to process and analyses the GPR acquisition data in 2D and the modeling which allows simulating synthetic scenarios and analyzing reflection time data and seismic refraction.

C. *Artificial Neural Networks: Multilayer Perceptron Networks (MLP's)*

According to [8], an ANN is a parallel distributed processor and constituted by a great number of simple processing units (nodes). The main function of the ANN is to store experimental knowledge and make it available for use, resembling the human brain on 2 aspects: the knowledge is acquired by the network based on its environment and a specific learning process and neuron's connection strength, known as synaptic weights, are used to store the acquired knowledge.

The ANN have been applied to various areas with enought success. They have a number of important features, such as generalization, parallelism, nonlinearity, adaptability, strength, among others [8]. The solution of complex problems through ANN is very attractive, since the way these problems are represented internally by the network and the natural parallelism inherent architecture of ANN creates the possibility of a better performance than the conventional models, which are most often used by researchers. In ANN, the most common procedure in troubleshooting passes initially through a learning phase, in which a set of examples (data) is presented to the network. This type of network will extract the basic characteristics to create a representation of the information provided. These characteristics are used later to generate the answers to the problem.

The ANN widened with the emergence of networks Multilayer Perceptron (MLP's) with units that can be connected to the units of the subsequent layer, generating a greater strength and computational performance with Networks with one or more intermediate layers or "hidden" are an extension of single-layer perceptrons, and can be trained to perform complex nature mappings [8]. An ANN MLP comprises a set of layers, In which each layer has a specific function. The output layer receives the stimuli of the intermediate layer and builds the pattern that will be the answer. The intermediate layers act as extractor characteristics, their weights are an enconding of features presented in the patterns of entry and allow the network to create its own representation, richer and more complex, of the problem. The MLP neural network training is performed with a set of known data (training set) which is extracted from random samples of the input X_p and output Y_p.

The network computes an output vector O_p based on the result obtained in the previous layer. The output vector is compared to the desired response vector y_p . The criterion used to evaluate the performance of the network is the sum of squared error (SSE).

The error of the output and intermediate layers are back-propagated through the network, making adjustments to the weights of their respective layers. The adjustment of the weights is calculated according to (Rumelhart et. Al, 1986):

$$\Delta w_{ij}(n+1) = \eta \delta_j o_i + \alpha \Delta w_{ij}(n) \quad \textbf{(1)}$$

in which, Δw_{ij} is the change in weight between the node k in the hidden layer and neuron i in the input layer; $\eta > 0$, is the learning rate; δ_j ,is the error of the observed value in the hidden layer neuron j; $\alpha \in [0,1]$, is a constant called momentum term. For training of the MLP network it can be used algorithms and Levenberg-Marquardt Backpropagation. In this study we used the Levenberg-Marquardt [9].

In this work, also, the number of entries is the number porosity vector values, dielectric constant, frequency, antenna and power of reflection. The choice of the appropriate number of hidden layers and their numbers of neurons were found empirically by performing tests with various network architectures and picked up the one with the lowest error for the training set and validation.

D. *The Multiple Linear Regression Model*

Regression analysis is a modeling technique used to help analyzing the relationship between a dependent variable, *Y*, and one or more independent variables X_1, X_2, X_3,..., X_n. The objective of this technique is to identify (estimate) a function that describes the closest possible relationship between these variables, so we can predict more accurately the value that the dependent variable *Y* will take for certain values of the independent variables X_i. The independent variables are also known as covariates [10]. The regression model is written generically as:

$$Y = f(X_1, X_2, X_3, ..., X_n) + \in \quad (2)$$

The term $\in$ representing a random disturbance in the function, or the error of approximation with stochastic distribution.

In some situations, more than one independent variable ($X_1, X_2, ..., X_n$) may be required to predict the value of the dependent variable (Y). The mathematical model for this case is the Multiple Linear Regression Model which is described in the formula [8].

$$Y_i = b_0 + b_1 X_i + b_2 X_{2i} + \ldots + b_k X_{ki} + \varepsilon_i \quad (3)$$

The coefficients b_0, b_1, ...,b_k, are the parameters estimated, $\in_i$ is the random error with the assumption that on average the errors tend to cancel, i.e., E ($\in_i$) = 0. As the errors are random, it is assumed that they possess normal probability distribution or Gaussian. The least squares estimator for the parameter vector *b* is given by the system:

$$\hat{b} = \left(X^{'} X\right)^{-1} \left(X^{'} Y\right) \quad (4)$$

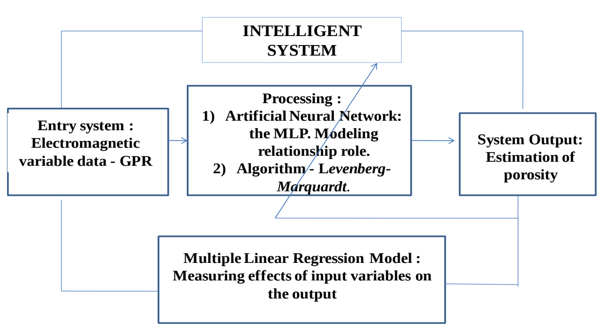
In this paper, the term X is the data matrix of explanatory variables and Y is the data vector of the response variable porosity. The efficiency of the multiple linear regression model is measured based on the results of multivariate analysis of variance (MANOVA - calculations data), performed with variable data.

3. THE SMART SYSTEM

The system structure is composed of input parameters of the computational processing of data and output to the estimate of porosity.

The system was designed to reflect the application of the technique of Artificial Neural Networks to estimate values of porosity mechanical sedimentary deposits using electromagnetic parameters of the data obtained in experiments planned and carried out. Was used testing the neural networks of the type Multilayer Perceptron (MLP), combined with multivariate statistical models.

Fig. 1. Schematic representation of the Intelligent System



Source: Prepared by author

4. RESULTS

E. Structuring of the database

The database was built using the data of the measurements in experiments with acquisitions GPR antenna with frequency range 50-2600 MHz. Was structured it atrix with 62 records containing the vectors with the measured and observed values porosity, dielectric constant, frequency, antenna and power of reflection.

F. Application of Neural Network MPL

To estimate the porosity (the network output) as a function of the explanatory variables (network input) it was implemented an architecture of MLP Neural Network in Matlab. For the training and validation, we used the Levenberg-Marquardt algorithm. To presented training the network with an input a matrix with 52 records vectors values of the three explanatory variables and to output a vector with 52 values of porosity. To better define the number of layers and neurons per layer, various simulations were performed with the result that the best architecture is formed respectively of an input layer with three neurons (input variables) and three hidden layers (intermediate) with 10, 15 and 20 neurons and an output layer neuron with a (desired response), i.e., a MLP (3,10,15,20,1). To check the performance of the neural network defined, i.e., to assess their ability to learn and generalize the results, validation has been done to the network was presented an array of size 10x3 formed by 10 rows with values of the input variables that were not presented to the neural network (NR) during training. In training we used the linear activation function.

After training, it was found that the neural network was able to efficiently estimate the associated values of porosity Ø, as shown in the graph in Figure 2. It is noted that in this graph the performance of lines with values of actual and estimated porosity is similar to all other figures in the training of the network, that is, it is hardly noticeable the difference between the two lines.

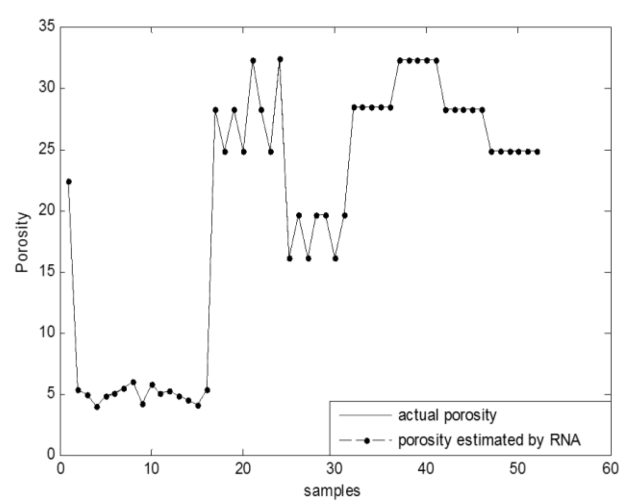
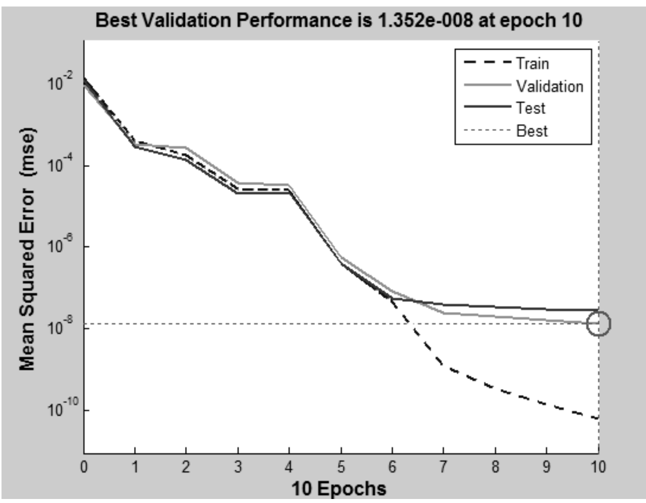


Fig. 2. Graphic lines of data from actual and estimated porosity during the network training

In Fig. 3. there is shown the network performance based on the mean square error, i. e., the error variances for the different epochs (simulations) of training and line training, validation and testing. It is noticed with this, that the mean square error converges to $1,35e^{-8}$ during training. This result is similar during the validation of the values of the training set. Furthermore, the training and validation lines are very close to 10 times of processing. This is an important aspect of network performance, as it presents an adaptive behavior and learning capabilities with much more accuracy.

Fig. 3. Graph with the values of RMSE during times of training, validation and testing.



In Fig. 4. It is the behavior of the lines of the values of actual and estimated porosity for the validation set consists of five rows with the porosity values measured in the sedimentology laboratory at UFRN and in a random sample of five records removed from the training set. In analyzing the behavior of the two rows of actual and estimated porosity, it was verified that they are statistically similar to the values in the ten validations, that is, it is noticed that the network estimated very well values of porosity. The estimates presented by the network are very similar for Low, medium and high values of porosity as well.

Fig. 4. Graph lines of data from the actual and estimated porosity for all ten validation records

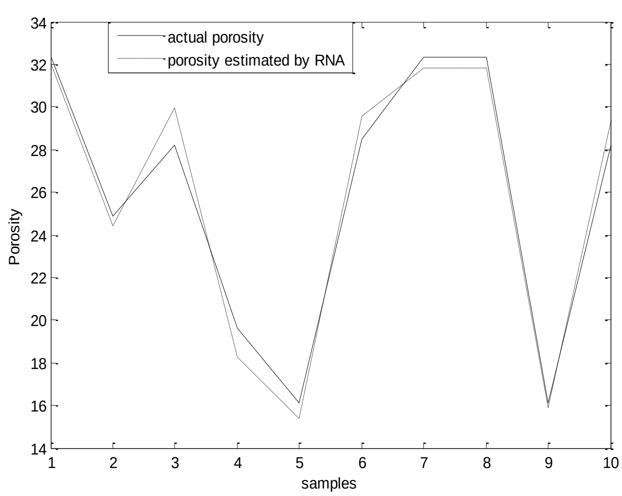
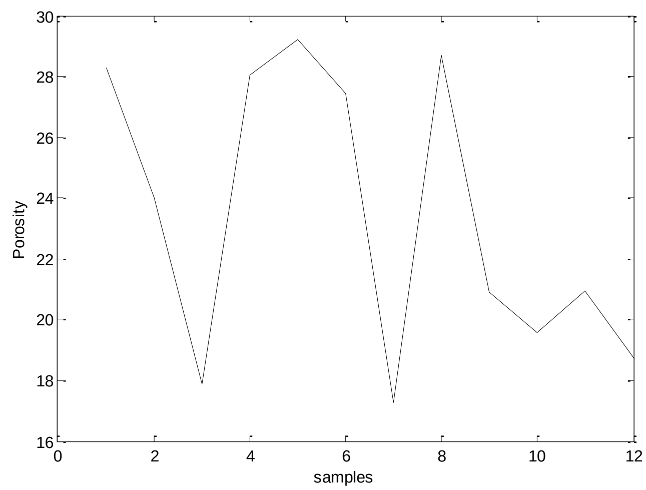


Figure 5 represents the estimated values of porosity for the twelve samples of random values of the explanatory variables presented by the network. It was found that the estimated values are within the minimum and maximum range of the actual values presented in the database. This proves The network estimates really well the porosity values for any group of numbers of the explanatory variables as long as they are between the minimum and maximum values given in the training and validation for these variables.

Fig. 5. Graph the line of the porosity estimated by MLP network for the whole twelve records random values of the explanatory variables



G. Application of Multiple Linear Regression Model

For the definition of the model and the calculations of the overall variability, we performed a multivariate analysis of variance - MANOVA. We estimated the model parameters. Statistica was used for data processing. Before MANOVA was used, an exploratory analysis was performed on the values of each variable, with the intention of determining the total amplitude, the maximum and minimum values and calculate the statistics of central tendency and variability of the variables. The adjusted model with the parameters estimated for the sample of 62 vectors of values of the variables is:

$$Y_i = 60,37 - 11,98\, K_i - 0,0019\, F_i - 0,1e^{-12}\, E_i, \quad i = 1, 2,., 62. \quad (5)$$

According to the parameters of the multiple regression model estimated it is seen by the magnitude of the parameter (11.98) that the dielectric constant is a variable that produces more effect on the variability of porosity. Although the estimated parameter value of the variable power of reflection is to small ($0,1e^{-12}$), it is significant and produces the second largest effect on the variability of porosity. The frequency antenna produces a lesser effect than the other variables on variability in porosity. The important result of the multivariate analysis of variance is that the model fits perfectly with the values of the variable, i.e., the model captures all variations of values of explanatory and response variables. Analyzing the results of the hypothesis test using the values from Table 1, it is verified that the multiple linear regression model is significant and that the explanatory variables dielectric constant, antena frequency and energy reflection explain 53.2% of the total variation of the values of the response variable **porosity**.

Table 1. Analysis values for multivariate variance

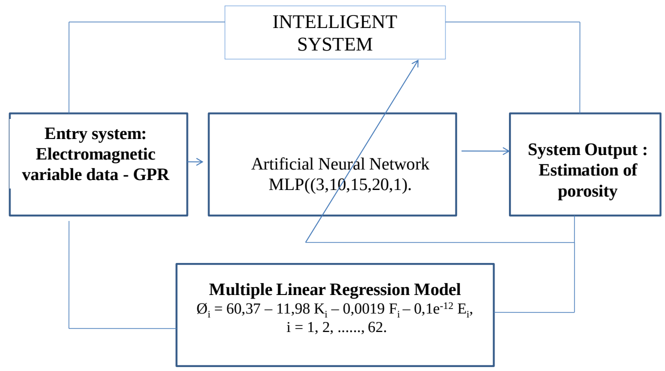
Source of variation	Degrees of freedom	Sum of Squares	Mean Square	Statistic F and α
Regression	3	3338,80	1112,93	F= 21,98
Residue	58	2936,42	50,62	α < 0,00001
Total	61	6275,23		

The magnitude of the parameter values and the result of the significance test of the model parameters indicate the degree of effect of each explanatory variable. The variable dielectric constant shows greater degree of effect on the porosity, power of reflection, although the value of the parameter b3 estimated to be small in magnitude ($0,1e^{-12}$), had a high and significant moderate effect the frequency of the antenna has the lowest degree of effect (0,0019), so before these results can be stated that the model of multiple linear regression is a important statistical tool to complement the proposed intelligent system.

5. FINAL CONSIDERATIONS

The intelligent system proposed on Fig. 6. can be adapted to other geological situations similar to the ones observed in this study. A system with this simple and efficient configuration enables the estimation of the porosity, having a new group of numbers for the independent electromagnetic variables. This intelligent system is able to estimate porosity from any database available, with the same variables used. The architecture of the neural network (MLP's) can be modified in accordance the necessity need, adapting to available databases.

Fig. 6. Illustration of Proposed Intelligent System



6. REFERENCES

[1] ANNAN, A. P. *Ground Penetrating Radar Workshop Notes*. Canada: Sensors & Software, 2001.

[1] ANNAN, A. P. GPR: History, Trends, and Future Developments. *Subsurface Tecnologiesand Aplications*, Canada: Sensors & Software Inc., 3, 4, 2002.

[2] PORSANI, J. L. *Metod GPR*: Applications in Geology, Geotechnical, Environment and Urban Planning. IV Summer School of Geophysics, 2002.

[3] RABELO, S. K. L. Study on the feasibility of calculating the porosity based on a reduced set of profiles and its application in Almada basin / BA. Doctoral Thesis at the State University of North Fluminense – UENF, 2004.

[4] BONET, E.J. & GABRIELLI, M.L. Properties of the rocks – PETROBRAS, Salvador, 1991.

[5] SOUZA, A.M. Use of GPR in the generation of static 3D models for mine planning of heavy minerals deposited in eolian dune. Dissertation, Federal University of Rio Grande do Norte / Program Graduate Geodynamics and Geophysics, Natal / RN, 110p, 2008.

[6] NEAL, A., ROBERTS C. L. Internal structure of a trough blowout, determined from migrated ground penetrating radar profiles. Sedimentology, 48:791-810, 2001.

[7] SOFTWARE REFLEXW, versão 6.1, Sandmeier Geophysical Software

[8] HAYKIN, S. “Neural Networks: Principles and Practice.” 2. ed. Porto Alegre: Bookman, 2001. 900 p. Translation Paulo Martins Engel.

[9]RANGANATHAN, ANANTH. “The Levenberg-Marquardt Algorithm”. Tutoral on LM Algorithm, 2004.

[10] BUSSAB, W. O. Analysis of Variance and Regression. Atual Editora, 2006.