Abstract
This research tackles the feasibility of using Artificial Neural Networks to capture nonlinear interactions between various soil parameters. In this study an attempt was conducted to predict the compaction parameter ($\gamma_{\text{dmax}}$ & O.M.C) using database comprising a total of 177 case records of laboratory measurements.

Eight parameters are considered to have the most significant impact on the magnitude of compaction parameters have been used as the model's inputs; liquid and plastic limits, plasticity index, specific gravity, soil type, gravel, sand, and fines content. The model output is the maximum dry unit weight and optimum moisture content.

A Multi–layer perceptron trainings using the back–propagation algorithm, are used in this work. A number of issues in relation to ANN's construction such as the effect of ANN's geometry and internal parameters on the performance of ANN's models are investigated. A parametric study was conducted for the three models to investigate the effect of the input variables on the output of the model.

Based on statistical criterion, it was found that ANN's have the ability to predict the compaction parameter with a good degree of accuracy.

Keywords: soil, Compaction parameter, Artificial Neural Network (ANN), Back–Propagation Algorithm, Matlab.
1. Introduction

Nowadays, Iraqi investment undergoes a fast building stage concerning the construction of various huge engineering projects. In addition, as earthworks and heavy constructions continue to expand, various geotechnical engineering problems are also demanding attention from engineers and geoscientists. Thus, integrity of such structures as well as saving both cost and time are of much importance which must be taken into engineering and geoscientists considerations. Consequently, geotechnical data may justify our demands to create a large soil database covering most of the Iraqi area. Artificial Neural Networks is suitable to fulfill this approach. If properly applied, it will optimize the exploration program by maximizing ground coverage and minimizing the laboratories testing.

In geotechnical engineering, empirical relationships are often used to estimate certain engineering properties of soil, using data from extensive laboratory or field testing, these correlations are usually derived with the aid of statistical methods, Artificial neural Networks, or by other approach.

The development of computer hardware and software that facilitate both the collection of raw data in digital form and the rapid, accurate and economic digitisation of most old analogue records is transforming both the way in which information is stored and the ability of data management organizations to make the information available to users [1].

In recent years, a new field of soft computing has emerged for modeling and controlling geotechnical problems. The ANN's are becoming more reliable than statistical methods due to their special attributes of identifying complex systems when the input and output are known from either laboratory or field experimentation [2].

Compaction of soil may be defined as the process by which the soil particles are artificially rearranged and packed together (densification) into a state of closer contact by mechanical means in order to decrease its porosity and thereby increases its dry unit weight [3]. This mechanical process will increase strength, reduce shrinkage, subsidence and permeability [4]. This is usually achieved by dynamic means such as tamping, rolling, or vibration [3].

Since Proctor's first paper published in 1933, the compaction method has become one of the most widely used soil improvement techniques around the world. However, most laboratory and field test programs are concerned with the physical properties of the soil near or at the maximum dry unit weight.

ANN's techniques in the prediction of compaction's parameter based on easily measured basic soil properties were adopted by many researchers [5-8], many of soil types were considered in their study, and their data base was varied from (82-126) sets of training and testing pattern.

2. Research significance

Since the compaction parameter (maximum dry unit weight ($\gamma_{d_{max}}$), optimum moisture content O.M.C) is normally determined from laboratory tests that are laborious and time-consuming. However, the group of tests performed to obtain index soil properties are relatively inexpensive and simple. The later tests do not require much time or any sophisticated testing systems. It is also essential to conduct these tests with utmost accuracy and to adopt realistic and suitable procedures to evaluate and interpret the results obtained. Thus, it is desirable to develop prediction models so that these desired soil properties can be evaluated from the classification and plasticity characteristic properties of the soil.
The objectives of this paper are to utilize MATLAB [9] based back propagation neural network (BPNN) by using data obtained from case studies, research work, site investigation reports, theses and dissertations, and published articles to investigate the feasibility of ANN's technique to predict \( \gamma_{\text{dmax}} \) & (O.M.C); study the effect of ANN's geometry and internal parameters on the performance of ANN's models; introduce a series of parametric studying to investigate if the model was able to generalize well, rather than memorize.

3. Artificial Neural Networks

The engineering properties of soil and rock exhibit varied and uncertain behavior due to the complex and imprecise physical processes associated with the formation of these materials. This is in contrast to most other civil engineering materials, such as steel, concrete and timber, which exhibit far greater homogeneity and isotropy.

In order to cope with the complexity of geotechnical behavior, and the spatial variability of these materials, traditional forms of engineering design models are justifiably simplified using artificial neural network.

Artificial neural networks are a form of artificial intelligence which attempts to mimic the behavior of the human brain and nervous system [10][11].

Over the last few years or so, the use of ANN's have increased in many areas of engineering [10][11]. Neural networks are often used for statistical analysis and data modeling, in which their role is perceived as an alternative to standard nonlinear regression or cluster analysis techniques.

ANN's learn from data examples presented to them in order to capture the subtle functional relationships among the data even if the underlying relationships are unknown or the physical meaning is difficult to explain. This is in contrast to most traditional empirical and statistical methods which need prior knowledge about the nature of the relationships among the data. ANN's are thus well suited to modeling the complex behavior of most geotechnical engineering materials which, by their very nature, exhibit extreme variability [11][12].

3.1 Artificial Neural Networks architecture and operation

A neural network consists of a number of interconnected processing elements, commonly referred to as neurons, as in Figure (1)[11]. The neurons are logically arranged into two or more layers and interact with each other via weighted connection. The scalar weight determines the nature and strength of the influence between the interconnected neurons. Each neurons is connected to all the neurons in the next layer (i.e. fully connected) [13].

There is an input layer, where data are presented to the neural network, and an output layer that holds the response of the network to the input. It is the intermediate layer, also known as hidden layer(s), that enables these network to represent and compute complicated association between inputs and outputs [13][14].

The hidden layers (so called because they do not interact directly with the external environment) provide the network with a high degree of nonlinearity. Furthermore, that layer takes the task to do all the necessary important calculations [5][10].
Figure (1) A simple ANN's architecture with one neuron [13]

The most widely used connection pattern is the three–layer backpropagation neural network (see figure 2), and the ideal number of nodes in the hidden layer has to be found only through trial and error. Using too few hidden neurons could result in large training errors and errors during testing, due to under–fitting and high statistical bias. On the other hand, using too many hidden neurons might give low training errors but could still have high testing errors due to over–fitting and high variance [7].

Figure (2) Three–layer back propagation neural network architecture [15]

Supervised neural networks [11] and multilayer feed forward networks [16] was adopted in the present work. Note that, determining the network architecture is one of the most important and difficult tasks in ANN's model development.

3.2 Optimization

Training or learning can be defined as the process of optimizing the connection weights between the nodes in the hidden layers. In other words, training (learning) is defined as self–adjustment of the network weights as a response to changes in the information environment [17].

The primary goal of training is to minimize an error function by searching for a set of connection strengths (weights) that cause the ANN's to produce outputs that are equal or close to targets [14].

In brief, training consists of three running steps:
1. Calculating outputs from input data.
2. Comparing the measured and calculated outputs.
3. Adjusting the weights for each node to decrease the difference between the measured and calculated values.
The training in this research was applied using Levenberg–Marquardt (LM) algorithm with variable learning rate (see figure 3).

The Levenberg–Marquardt (LM) back propagation algorithm is a powerful optimization technique which was introduced to the neural net research because it provided methods to accelerate the training and convergence of the algorithm. This algorithm is well suited for neural net training where the performance index is the mean square error (MSE) [18][19].

The performance of the steepest descent algorithm can be improved if the learning rate is allowed to change during the training process. An adaptive learning rate will attempt to keep the learning step size as large as possible while keeping learning stable. The performance of the steepest descent algorithm can be improved if the learning rate is allowed to change during the training process [18].

4. Methodology
4.1 Data acquisition and preparation

For the compaction network analysis, a total of (20) data patterns were considered out of the original (177) for testing the performance. The remaining records were considered to be for training. Sometimes, the designer collides with one of the input or output variables that have a text formula (i.e. not numeral), such as soil type in the present work.

Data was collected from published and authentic results from many faculties, centers, universities, institute, etc.

Five types of soils are considered here followed by their designations in the network; (CL=1), (CH=2), (SC=3), (GC=4), and (SM=5). The overall numbers of each soils type were 82,28,45,12,and 10 respectively. Thus, it could be seen that soil type (CL) was dominating. The main statistical characteristics of the data (input–output) variables are summarized in (table 1).

Table (1) Statistical characteristics of compaction's network data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Sample variance</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>L.L%</td>
<td>15</td>
<td>89</td>
<td>39.62</td>
<td>11.87</td>
<td>140.79</td>
<td>29</td>
</tr>
<tr>
<td>P.L%</td>
<td>8.35</td>
<td>52.4</td>
<td>21.09</td>
<td>5.32</td>
<td>28.31</td>
<td>21</td>
</tr>
<tr>
<td>P.I%</td>
<td>1</td>
<td>50</td>
<td>18.53</td>
<td>8.45</td>
<td>71.48</td>
<td>8</td>
</tr>
<tr>
<td>Gs</td>
<td>2.33</td>
<td>5.82</td>
<td>2.71</td>
<td>0.10</td>
<td>0.01</td>
<td>2.76</td>
</tr>
<tr>
<td>Gr%</td>
<td>0</td>
<td>67.1</td>
<td>8.92</td>
<td>11.79</td>
<td>139.08</td>
<td>0</td>
</tr>
<tr>
<td>S%</td>
<td>0</td>
<td>85</td>
<td>31.43</td>
<td>19.17</td>
<td>367.66</td>
<td>8</td>
</tr>
<tr>
<td>F%</td>
<td>13</td>
<td>100</td>
<td>59.64</td>
<td>23.19</td>
<td>537.87</td>
<td>92</td>
</tr>
<tr>
<td>γdmax kN/m</td>
<td>12.7</td>
<td>20.57</td>
<td>17.50</td>
<td>1.41</td>
<td>1.98</td>
<td>18.45</td>
</tr>
<tr>
<td>O.M.C %</td>
<td>7.6</td>
<td>37</td>
<td>16.37</td>
<td>4.35</td>
<td>18.89</td>
<td>14</td>
</tr>
</tbody>
</table>

Figure (3) Idealize learning curve of Levenberg–Marquardt algorithm [17]
An important issue that has to be addressed is how to arrange the data. It is a common practice to divide the available data into two subsets: a training set, to construct the neural network model, and an independent validation set to estimate model performance in the deployed environment.

However, recent studies have found that the way the data are divided (see table 2) can have a significant impact on the obtained results [11][12][20].

<table>
<thead>
<tr>
<th>overall data</th>
<th>training data</th>
<th>testing data</th>
<th>percentage of training data</th>
</tr>
</thead>
<tbody>
<tr>
<td>177</td>
<td>157</td>
<td>20</td>
<td>11%</td>
</tr>
</tbody>
</table>

4.2 Model development

Choosing a successful network geometry is highly problem dependent, and obtaining an optimal combination of these parameters is a difficult task as well.

The selection of the model input variables that have the most significant impact on the model performance is an important step in developing ANN's models [20]. (See table 3)

<table>
<thead>
<tr>
<th>inputs variable</th>
<th>output(s) variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>liquid limit (LL), plastic limit (PL), plasticity index (PI), specific gravity (Gs), soil classification (c), Gravel (Gr), Sand (S), and Fines content (F).</td>
<td>(γ_{dmax}) &amp; (O.M.C)</td>
</tr>
</tbody>
</table>

It is important to pre-process the data in a suitable form before it's applied to the ANN's. Data pre-processing (normalization, scaling, transformation) is necessary to ensure all variables receive equal attention during the training process. Moreover, pre-processing usually speeds up the learning process, and obtains better convergence.

4.2.1 Pre-processing of data

Data pre-processing (normalization, scaling, transformation) is necessary to ensure all variables receive equal attention during the training process. Moreover, pre-processing usually speeds up the learning process, and obtains better convergence. Pre-processing can be in the form of data scaling, normalization and transformation. Scaling the output data is essential, as they have to be commensurate with the limits of the transfer functions used in the output layer [11][12].

4.2.2 Architecture and activation function

Determining the network architecture is one of the most important and difficult tasks in ANN's model development.

In general, most researchers recommend that two-layer model with differentiable transfer functions, such as log-sigmoid, and sufficient number of neurons in the hidden layer can approximate any nonlinear function [6].
4.2.3 Training operation

In the present work, the training operation is performed mainly by combining Levenberg–Marquardt back propagation algorithm beside delta rule supervised by variable learning rate.

4.2.4 Model validation

Once the training phase of the model has been successfully accomplished, the performance of the trained model should be validated. The purpose of the model validation phase is to ensure that the model has the ability to generalize in a robust fashion.

Usually, the progress of the training is checked by plotting the training MSE versus the performed number of epochs (as seen in figure 4). Note that the programmer who deals with training ANN's will have an experience in how the shape of this plot Vs time should be.

This performance can be assessed by several criteria (see table 4). These criteria include coefficient of determination, root mean squared error, mean absolute error, minimal absolute error, maximum absolute error and variance account. A well–trained model should result in a coefficient of determination value close to (1) and small values of error terms.

<table>
<thead>
<tr>
<th>No.</th>
<th>Statistical criteria</th>
<th>equation</th>
<th>Equ. No.</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Coefficient of correlation (R)</td>
<td>$R = \frac{\sum (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum (O_i - \bar{O})^2 \sum (P_i - \bar{P})^2}}$</td>
<td>(1)</td>
<td>[18]</td>
</tr>
<tr>
<td>2</td>
<td>Mean square error (MSE)</td>
<td>$MSE = \frac{1}{n} \sum (O_i - P_i)^2$</td>
<td>(2)</td>
<td>[14][18]</td>
</tr>
<tr>
<td>3</td>
<td>Error equation (E)</td>
<td>$E = \frac{O_i - P_i}{O_i} \times 100$</td>
<td>(3)</td>
<td>[14][18]</td>
</tr>
<tr>
<td>4</td>
<td>Coefficient of efficiency (CE)</td>
<td>$CE = 1 - \frac{\sum (O_i - P_i)^2}{\sum (O_i - \bar{O})^2}$</td>
<td>(4)</td>
<td>[19]</td>
</tr>
<tr>
<td>5</td>
<td>Variance account for (VAF)</td>
<td>$VAF = \left[ 1 - \frac{\text{var}(O_i - P_i)}{\text{var}(O_i)} \right]$</td>
<td>(5)</td>
<td>[21]</td>
</tr>
<tr>
<td>6</td>
<td>Over–fitting (over–training) ratio (OFR)</td>
<td>$OFR = \frac{(MSE)<em>{\text{training}}}{(MSE)</em>{\text{testing}}}$</td>
<td>6</td>
<td>[22]</td>
</tr>
</tbody>
</table>

Where:-

$O_i$ = observed (target) value for ith data,
$\bar{O}$ = mean of observed value, $n$ = number of observations.

$P_i$ = predicted value for ith data,
$\bar{P}$ = mean of predicted value,
Once the training and testing phase of the model have been successfully accomplished, (i.e. strong coefficient of correlation is obtained ((R) value approach to 0.8), the performance of the trained model should be validated by conducting a parametric study.

5. Results and discussion

One of the most thorny problems in the model's development is the model's architecture ((see table (5). Choosing the right architecture normally takes many iterations and may be considered time consuming.

Table (5) Compaction network properties

<table>
<thead>
<tr>
<th>Index</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architect</td>
<td>8–20–20–2</td>
</tr>
<tr>
<td>activation functions</td>
<td>Logsig–Logsig–Logsig</td>
</tr>
<tr>
<td>Epochs</td>
<td>62</td>
</tr>
<tr>
<td>MSE</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Table (6) examine some proposed formulas by other researchers to conduct the neuron's number of the hidden layers.

Table (6) Equations for the prediction of the hidden neuron's number

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$H = \sqrt{I \times O}$</td>
<td>Yeh (1997) [15]</td>
<td>4</td>
<td>(7)</td>
</tr>
<tr>
<td>2</td>
<td>$H = 2 \times I + 1$</td>
<td>Hecht–Nielsen (1987) [21]</td>
<td>9</td>
<td>(8)</td>
</tr>
<tr>
<td>3</td>
<td>$H = \frac{I + O}{2}$</td>
<td>Yeh (1997) [23]</td>
<td>5</td>
<td>(9)</td>
</tr>
<tr>
<td>4</td>
<td>$H = 2 \times I$</td>
<td>Kanellopoulos and Wilkinson (1997) [24]</td>
<td>8</td>
<td>(10)</td>
</tr>
<tr>
<td>5</td>
<td>$H = \frac{T - O}{I + O + 1}$</td>
<td>Najjar (1999) [25]</td>
<td>14</td>
<td>(11)</td>
</tr>
</tbody>
</table>

Where,
$H$= No. of hidden neurons,
$I$= No. of input neurons,
$O$= No. of output neurons,
and $T$ = training sample.

For current work, Najjar's equation seems to be the closest to the obtained result (i.e. 20 neurons). It was also observed that a better performance could be obtained using two hidden layers. This could be attributed to the complexity and nonlinearity involved in mapping the input to the output.
The optimal ANN's model’s structure that resulted in minimum error and maximum efficiency during both training and testing was selected for validation. The validation criterion is materialized in table (7).

Table (7) shows that MSE is considered as an acceptable value to reach during the training and test since it is a three digits after comma (i.e. 0.001). Also, the coefficient of correlation shows a good agreement between observed and predicted values.

For CE and VAF values, it can be deemed fairly acceptable, and it is plausible to notice that is no sign for overtraining since the value of OFR is close to unity. The training progress is depicted in figure (4). It clearly shows a smoothly gradient with no sign of over–training (see figure 3).

Inspection of figures (5) and (6), outlines that O.M.C error is more than $\gamma_{dmax}$ although $R$ for O.M.C is more than $\gamma_{dmax}$.

After the evaluation of the network using many statistical analysis equations, the model will be introduced to a series of parametric studying. Since the parameter of soils under compaction may not be clear directly from the experimental data especially if there is a large database. So, first of all, data were divided and rearranged again according to soil type.
Obtaining the maximum, minimum, and the average for each soil type before starting the parametric study.

Note that, it is rare to conduct the parametric study due to its complicatedness. A few researchers [3][5][12] could carry out it. Even though, they carried out the parametric study, their results did not specify the type of soil and it was not intensive.

The relation between fine material and LL is depicted in figures (7) and (8). Figure (7) illustrates that, generally an increment of fine material in the studied soil leads to a reduction in of $\gamma_{d,\text{max}}$ value. The effect of LL is obvious, since lower LL will give greater values of $\gamma_{d,\text{max}}$. While, the behavior of Figure (8) has an exactly reversed trend of figure (7).

![Figure (7) Effect of fine material and LL on $\gamma_{d,\text{max}}$ (Soil type CL)](image)

![Figure (8) Effect of fine material and LL on O.M.C (Soil type CL)](image)

The relation between LL and fine material was studied and presented in figures (9) and (10) for soil type SC.

![Figure (9) Effect of LL and fine material on $\gamma_{d,\text{max}}$ (soil type SC)](image)

![Figure (10) Effect of LL and fine material on O.M.C (soil type SC)](image)

Note that the above results show that higher dry unit weight can be noticed for soil with lesser amount of fines, that results were agreed by [26].

The relationship between fine material and PI is illustrated in figures (11) and (12) respectively.
It can be seen that a higher dry unit weight can be noticed for soils with lesser amounts of PI. While, soils having a lower fine material gave lower O.M.C. This could be attributed to the electrical and size nature of clay particles [4].

The effect of gravel and LL on $\gamma_{\text{dmax}}$ and O.M.C is depicted and discussed in figures (13) and (14).

It could be seen that if gravel percent increases, this will lead to an increase in $\gamma_{\text{dmax}}$, and the maximum value can be obtained when LL equals to 55%, while that increment will decrease the O.M.C value. This is foreseeable, since adding gravel or plus 4.75 mm material to the soil will increase the maximum unit weight [27].

6. Conclusions
1. ANN’s have the ability to predict the compaction parameters with a good degree of accuracy within the range of data used for developing ANN’s models. And it should be kept in mind that using (ANN’s) does not always guarantee good result.
2. The obtained model can be used for preliminary assessment of preliminary design phases and feasibility studies for obtaining reliable information for site investigation projects and giving an impression about the test results before they are conducted. On the other hand, this models is not intended to nullify the need for executing laboratory tests, but as a tool to obtain reliable preliminary information about the soil behavior.
3. It is essential to understand that the success of ANN's models depends mainly on the quality and validation of data obtained. They also depend on the ANN's parameters (architecture, Activation function, training algorithm, etc).

4. Results show that Logsig activation function was the most appropriate activation function that could be used in the training. Also, Najjar (1999) equation seems to give a number of hidden neurons in the hidden layers close to that in the present work.

5. ANN's model could be translated into a simple and practical formula, if a simple model was adopted.

6. Finally, the researcher would rather join the work with a new equation (equation 12).

$$D \cdot B = 1.8 I \times (I + 1) + 4.6 \frac{N}{H}$$

Where:

- $D \cdot B$ = data base set.
- $I$ = number of input variables.
- $H$ = number of hidden layers
- $N$ = number of neurons in the hidden layers

Since that kind of equations has never been arrived at, this simple equation can be helpful for the designer to start the work. It may also reduce the time to look for the optimal.

7. Recommendations

1. Most of the sites in Mosul city are still suffering from insufficient geotechnical study (especially for depth more than one meter) and do not have a bore hole log and stratum drawing details. The researcher recommends taking a site by site geotechnical/geological study individually and making a complete study to cover Mosul area.

2. Always use high quality laboratory test records in training the neural network. Added to that, data gathered must be examined to assign the confirmation that they are homogenous (i. e. less variance).

3. Using highly speed performance computer with the updated MATHLAB software version are recommended since duration of training phase depends on the size of the training file and computer processor speed.

4. The Coefficient of consolidation $C_v$ needs further attention in future works since it is only can be calculated based on different graphical methods.

8. References


The work was carried out at the college of Engineering, University of Mosul

27