

Atterberg Limits Prediction Comparing SVM with ANFIS Model

Mohammad Murtaza Sherzoy^{1,*}

¹Academy of Sciences of Afghanistan, Sher Ali Khan Watt, Shari-e-naw, Kabul, POBox 894, Afghanistan

Corresponding author: murtaza_sherzoy2000@yahoo.com

Phone: +93780151633

Received: Jan 15, 2017. Revised : 18 Jan 2017, Accepted: Feb 20, 2017, Published: 1 March 2017

DOI:10.24273/jgeet.2017.2.1.16

Abstract

Support Vector Machine (SVM) and Adaptive Neuro-Fuzzy inference Systems (ANFIS) both analytical methods are used to predict the values of Atterberg limits, such as the liquid limit, plastic limit and plasticity index. The main objective of this study is to make a comparison between both forecasts (SVM & ANFIS) methods. All data of 54 soil samples are used and taken from the area of Peninsular Malaysian and tested for different parameters containing liquid limit, plastic limit, plasticity index and grain size distribution and were. The input parameter used in for this case are the fraction of grain size distribution which are the percentage of silt, clay and sand. The actual and predicted values of Atterberg limit which obtained from the SVM and ANFIS models are compared by using the correlation coefficient R^2 and root mean squared error (RMSE) value. The outcome of the study show that the ANFIS model shows higher accuracy than SVM model for the liquid limit ($R^2 = 0.987$), plastic limit ($R^2 = 0.949$) and plastic index ($R^2 = 0.966$). RMSE value that obtained for both methods have shown that the ANFIS model has represent the best performance than SVM model to predict the Atterberg Limits as a whole.

Keywords: Atterberg limit, Support Vector Machine (SVM), Adaptive Neuro-Fuzzy inference System (ANFIS), sand, clay, silt.

1. Introduction

The Atterberg limits can be used to distinguish between sand, silt and clay, and it can distinguish between different types of sand, silt and clays. These limits were created by Albert Atterberg, a Swedish chemist. They are then refined by Arthur Casagrande. Knowledge of the grain size distribution is very important for the behavior of soil under load and soil that come in contact with water can be identified. Water is also a part of the soil component, and its presence reduces the strength of the soil (Ali, 2011). If a particular soil grain size distribution is known, an accurate prediction of how the soil when acting as a basis for or a component of the structural works such as buildings, dams, and roads and other can be made. Once you know how to soil tend to behave, engineers can design and estimate the best foundation to support an initiatory safer and more durable. Previously, the study of the grain size distribution and geological characteristics of the other soil has been done, for example, (Berbenni 2007) conducted a study on the impact of the size distribution of soil to the yield stress. Reproduction of his results showed a yield stress decreased with increasing grain size distribution.

However, in this study, the grain size distribution of soil fractions and percentages will

be used to predict the Atterberg limits using analytical methods *Support Vector Machine* (SVM) and *Adaptive Neuro- Fuzzy inference System* (ANFIS). Considering the main objective and aim of this work the prediction of the Atterberg Limits, it is convenient to review fundamental principles related to the comparing a Support Vector Machine (SVM) model with Adaptive Neuro -Fuzzy Inference System (ANFIS). The Atterberg limits are a convenient means to describe the plastic type properties of a soil. They are defined by limits on different types of behavior, and are expressed as a water content for a detailed description.

SVM is generally utilized in classification and regression problems (Chen et al. 2010). SVMs have the ability to enable a learning machine to generalize well to unseen data with their strong statistical learning theory grasp and very promising in empirical performance (Lin & Yeh 2009). There are a wide number of applications that can be utilized by using SVMs such as regression, pattern recognition, Bioinformatics and artificial intelligence (Tripathi et al. 2006). Support vector machine is a machine learning method that is widely used for data analyzing and pattern recognizing. The algorithm was invented by Vapnik and the current standard incarnation

was proposed by (Cortes and Vapnik 1995). This application note is to help understand the concept of support vector machine and how to build a simple support vector machine using Matlab.

The ANFIS has the ability to learn from data, such as that owned by an artificial neural network. ANFIS models can also quickly achieve optimal results even if the target is not given. Additionally, there is no ambiguity in the ANFIS, unlike in a neural network. Because ANFIS combines both neural networks and fuzzy logic, it can handle complex problems and non-linear problems.

2. Material And Methods

A. Data Distribution

The distribution of the sample can be divided into two areas, area 1 (Fig 1) and area 2 as shown in (Fig 2). The first sample was taken around the state of Pahang, while in the second, the distribution of the sample is in the state of Johor. In this study, all sample data for the grain size distribution were prepared by IKRAM and tests of

soil classification and testing the limits Atterberg has been obtained from the results of laboratory tests. All distributions of soil samples taken as casual as the distance between the distributions of samples is almost 400 km. A total of 54 soil samples taken in the neighbourhood of the Peninsular Malaysian and its distribution is shown in Table 1.

B. Revision of Area

The Atterberg limits value and Grain size distribution were obtained through laboratory test carried out by (IKRAM) the Malaysian Institute of Public Works. The ANFIS and SVM models were then examined by applying 54 data records collected from these tests, the actual data value compared with the predicted Atterberg limit values. For use as a training data set the ANFIS and SVM models need a set of input and output data. The grain size distribution was employed for the purpose of this study, as input parameters in the development of the ANFIS and SVM models for the prediction of Atterberg limit values.

Table 1. Distribution Area Sample Data Collection Peninsular Malaysia (Ikram, 2011)

No	Area	Location	Total sample
1	1	Genting Sempah, Pahang	9
2	1	Gua Tempurung, Perak	3
3	1	Lentang, Pahang	4
4	1	Simpang Pulai, Perak	10
5	1	Kuala Kubu Baru, Selangor	7
6	1	Fraser Hill, Pahang	10
7	1	Logging, Pahang	1
8	2	Gunung Pulai, Johor	10
Total			54

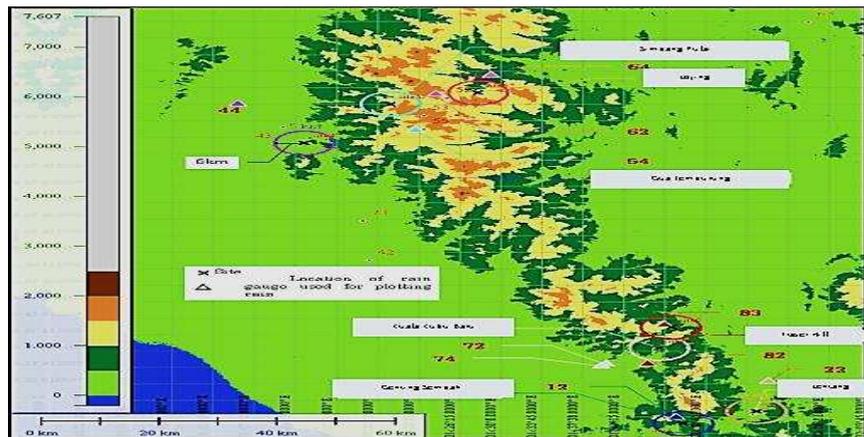


Fig 1. Distribution of sample data (Area 1) Peninsular Malaysia (Ikram, 2011)

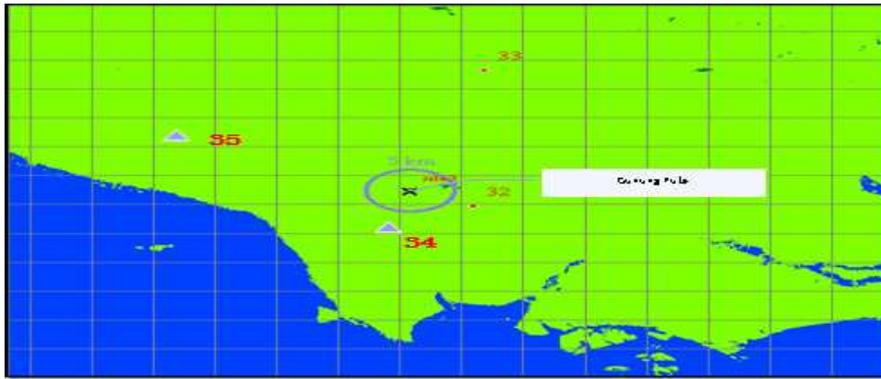


Fig 2. Distribution of sample data (Area 2) Peninsular Malaysia (Ikram, 2011).

The soil sample data were taken based on the occurrence of debris flow event across Peninsular Malaysia, as recorded in Table 1. Fig 1 presents the location of the grain size distribution sample used in the study. The sampling area can effectively be divided into two areas, including the state of Perak and Pahang (Area 1) and Johor (Area 2), respectively. All the 54 soil samples were collected and for different parameters tested, including grain size distribution, liquid limit (LL), plastic limit (PL), plasticity index and grain size distribution.

Methods of data collection for this study is to gather existing data for analysis SVM and ANFIS method. Both input and output parameters such as soil grain size distribution, liquid limit (LL), plastic limit (PL) and plasticity index (PI) will be identified and studied. The Methodology was established for comparing the output parameters will be analyzed based on the two methods mentioned SVM and ANFIS.

C. Support Vector Machine (SVM) Model

Support vector machine (SVM) is a technique valuable for data classification, regression and prediction. SVMs are a set of learning methods that analyses data and recognize patterns, the first introduced in computer science. SVM algorithm is the current standard proposed by (Cortes and Vapnik 1995). SVM has originated from statistical learning theory pioneered by (Boser et al. 1992). Since SVM is a relatively new technique, a brief explanation of how it works is given below. More detail can be found in many publications. The second learning technique uses the support vector machine (SVM) that is firmly based on the theory of statistical learning theory, uses regression method. The SVM developed to predict the Plastic Limit (PL), Liquid Limit (LL) and Plastic index (PI). Further, an attempt has been made to simplify the models, requiring only three parameters plastic limit, liquid limit and plastic index as input for prediction.

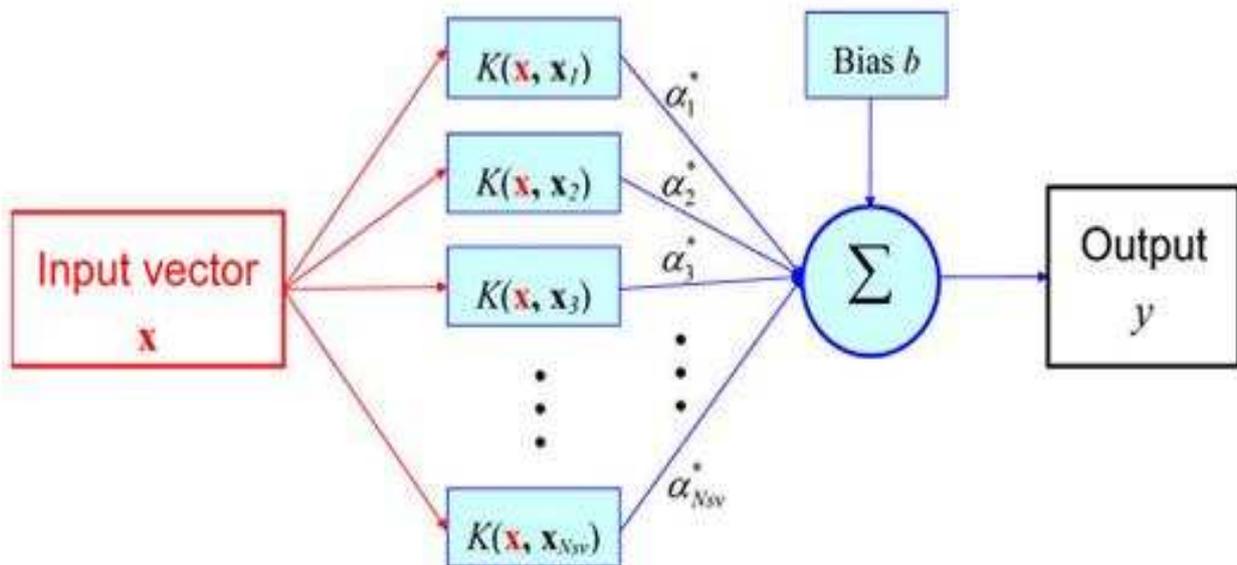


Fig 3. Architectural graph of Support Vector Machine (Lin et al, 2009).

D. Kernel function

Once applying the SVM to linearly separable data we have started by generating a matrix H from the dot product of our input variables:

$$H_{ij} = y_i y_j k(x_i, x_j) = x_i \cdot x_j = x_i^T x_j$$

The $k(x_i; x_j)$ is an example of a family of functions in the above equation, called Kernel Functions ($k(x_i, x_j) = x_i^T x_j$ being known as a Linear

Kernel). The set of kernel functions is composed of variants of (2) in that they are all based on calculating inner products of two vectors. This means that if the functions can be recast into a higher dimensionality space by some potentially non-linear feature mapping function $x \rightarrow \phi(x)$.

Only inner products of the mapped inputs in the feature space need be determined without us needing to explicitly calculate ϕ .

One of the reason that this Kernel Trick is valuable is that there are many regression and classification problems that are not linearly regress able and separable in the space of the inputs x , which might be in an advanced dimensionality feature space given a suitable mapping.. $x \rightarrow \phi(x)$. The kernel function

can be defined as in equation (2) if we define our kernel to be:

$$k(x_i, x_j) = e^{-\left(\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)}$$

(2)

As show in the left side of the Fig 5 the data set that is not linearly separable in the two dimensional dataspace x could be separable in the nonlinear feature space, which is on the right hand of Fig 5. Because the data set defined implicitly by this non-linear kernel function is known as a Radial Basis Kernel

(1)

E. Adaptive Neuro Fuzzy Interference System (ANFIS) Model

The proposed neuro-fuzzy model of ANFIS is a multilayer neural network-based fuzzy system. Its topology is shown in Fig. 5, and the total of the system are five layers. In this connected structure, the input and output nodes represent the training values and the predicted values, respectively, and in the hidden layers, there are nodes functioning as rules and membership functions (MFs). This architecture has the benefit that it omit the disadvantage of a normal feed forward multilayer network, where it is difficult for an observer to modify or understand the network. For simplicity, we assume that the examined fuzzy inference system has two inputs x and y and one output. For a first-order Sugeno fuzzy model, a common rule set with two fuzzy if-then rules is defined as:

Rule 1 : If x is A_1 and y is B_1 : Then $f_1 = p_1x + q_1y + r_1$

Rule 2 : If x is A_2 and y is B_2 : Then $f_2 =$

$$p_2 x + q_2 y +$$

$$r_2 \tag{4}$$

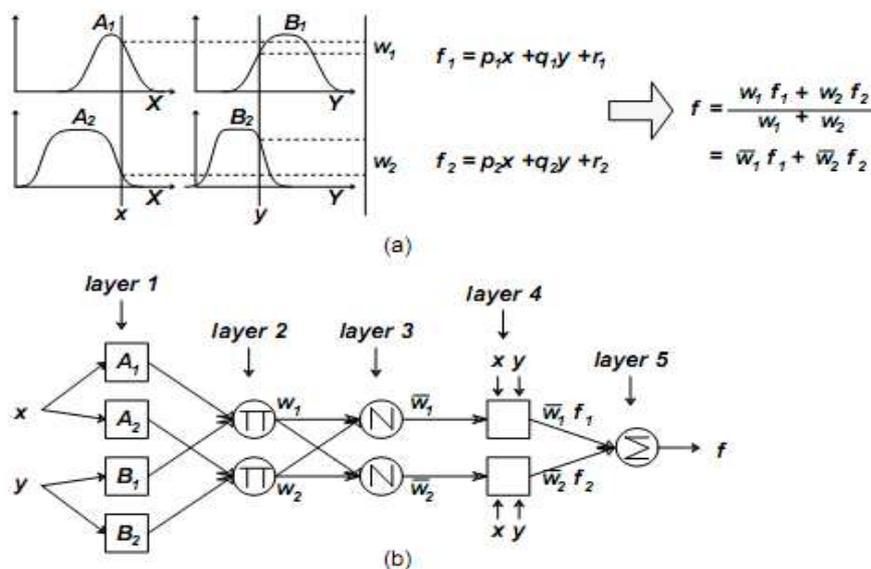


Fig 5. (a) First- order Sugeno fuzzy model; (b) Equivalent ANFIS architecture, (Jang,1993)

Fig 5 (a) graphically illustrated mechanism fuzzy reasoning to get a f output from a given input vector $[x, y]$. That w_1 and w_2 shoot strength usually obtained as a result of grade of membership in the premises, and output f is the weighted average of each rule's output. To fascinate learning (or adaptation) Surgeon fuzzy model, it is easy to put into the framework of fuzzy model adaptive network that can compute the gradient vector in a systematic manner. Resultant network architecture, called ANFIS (Adaptive Neuro-Fuzzy inference system), and shown from Fig. 1b, different layers of ANFIS have different nodes. Each node in a layer is either fixed or adaptive (Jang, 1993). Different layers with their associated nodes are described below :

F. Performance Avaluation

This part is important to have a fair comparison of the predicting result obtained from ANFIS and SVM. Addition, there are a lot of criteria included in the models which will prove difficult to perform simply by using conventional mathematic formula. Data obtained from both SVM and ANFIS parameters compared to see the difference. This is to see the effect of changes to the output and error when various renovations

G. Root Mean Square Error (RMSE)

The correlation coefficient (R), root mean squared error (RMSE) was used to evaluate the performance of the proposed models. By this formula determines the residual value between the actual and predicted Atterberg limits. The effect on coefficient is more obvious by larger error in predicted values than the smaller ones. The best fit can be seen when the value of RMSE is zero. The formula for RMSE can be calculated using Equation (5).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (h_i - t_i)^2} \quad (5)$$

Where n is amount of data, h_i is observed value, t_i is the predicted value.

H. Correlation Coefficient (R)

Generally, this formula is the root of ratio between the explained variations where it range from 0 to 1. It represent the relationship's value

between the actual value and the predicted value. This formula is best shown by equation (6).

$$R = \frac{\sum_{i=1}^n (h_i - \bar{h}_i)(t_i - \bar{t}_i)}{\sqrt{\sum_{i=1}^n (h_i - \bar{h}_i)^2 \sum_{i=1}^n (t_i - \bar{t}_i)^2}}$$

(6)

Where n is amount of data, h_i is observed value, t_i is the predicted value, \bar{h}_i and \bar{t}_i are the average of the observed and predicted values respectively.

Correlation coefficient R^2 indicates the strength of the linear relationship and the relationship of those variables. R^2 value closer to 1 indicates the efficiency of a model.

3. Result And Discussion

Comparison of both SVM and ANFIS methods of analysis necessary to determine the best methods of both, and to calculate the uncertainty for both these models. Determination of the best and efficient analysis is important that the accurate method can be used for a reference primarily associated with Atterberg limits or engineering properties of soil in the future. For SVM analysis method, two criteria are discussed modification of renovation and modification of the input training data set. As for the method of analysis ANFIS, modification total input will be carried out for comparison purposes. All data obtained were analysed and a comparison is made through tables and graphs.

Fig 6 shows a comparison of the predictive values of the liquid limit for SVM and ANFIS models. From the Fig, it was found that the ANFIS model is represented by the red line is closer to the actual value compared with the SVM predictions that indicate by green line.

Fig 7 also clearly shows the red line representing the results of the ANFIS model predictions are seen getting closer to the actual value of the plastic limit is represented by the green line (SVM model). Fig 8 ANFIS prediction is seen closer to the actual value than the SVM for analysis of Plasticityc index. In terms of observations on all of these Figs, it is seen that the results of ANFIS prediction closer to the experimental data for the analytical testing laboratory liquid limit, plastic limit and plasticity index analysis where revenue forecasts ANFIS model is closer to the actual value.

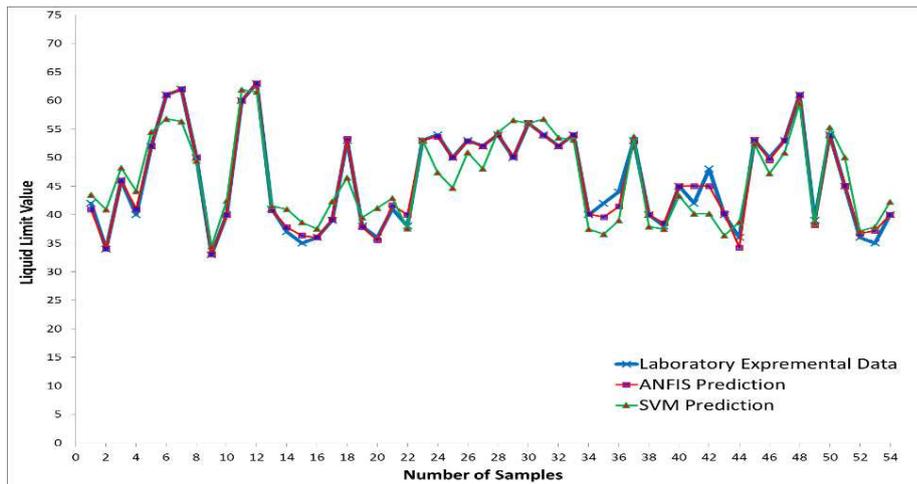


Fig 6. Predicted and actual liquid limit values using SVM and ANFIS models with 3 input

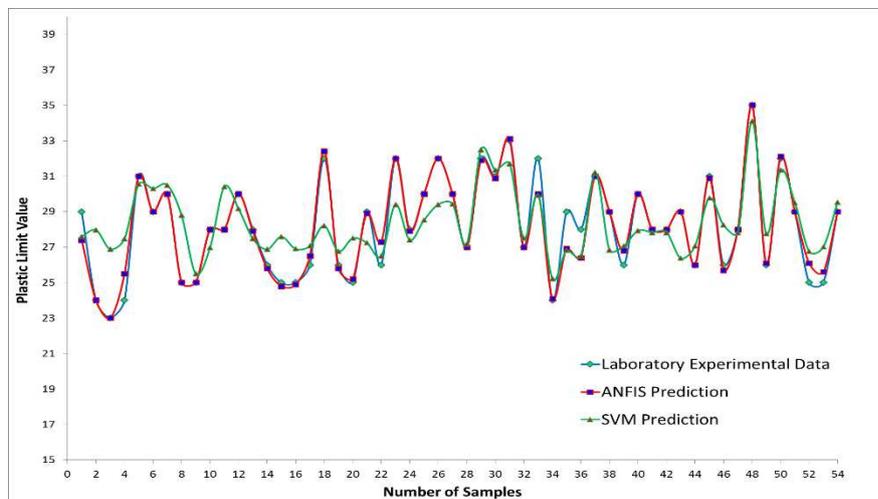


Fig 7. Predicted and actual plastic limit values using SVM and ANFIS models with 3 input

4. Comparison of SVM and ANFIS best models RMSE and R of 3 Input

In this study, the performance of both ANFIS and SVM model can be assessed by looking at the difference between the values predicted by the correlation coefficient, R^2 and root mean squared error RMSE. The R^2 value closer to 1 indicates the efficiency of such a model. The smaller RMSE values indicate smaller errors produced by the model. Comparison of R^2 values for the two models are briefly described by Table 2 Referring to Table 2 the value of R^2 obtained results ANFIS is better than SVM model for the liquid limit, plastic limit and plasticity index. However, the results indicate that ANFIS is more accurate the SVM model.

In this study comparison of the Root mean square error or RMSE will be conducted. RMSE is a mathematical method for measuring the magnitude of the average error. The lower the RMSE value of a data means more accurate predictions. Table 3 shows the RMSE values obtained for the three analyzes the Atterberg limits.

The results show that the low RMSE values obtained by ANFIS model for all liquid limit, plastic limit and plasticity index analysis.

Meanwhile, finally the ANFIS model shows the RMSE is lower than SVM. In conclusion, the three Atterberg limits tests conducted, three tests that test the liquid limit plastic limit and plasticity index, ANFIS models give a more accurate prediction of the actual value compared with the SVM model.

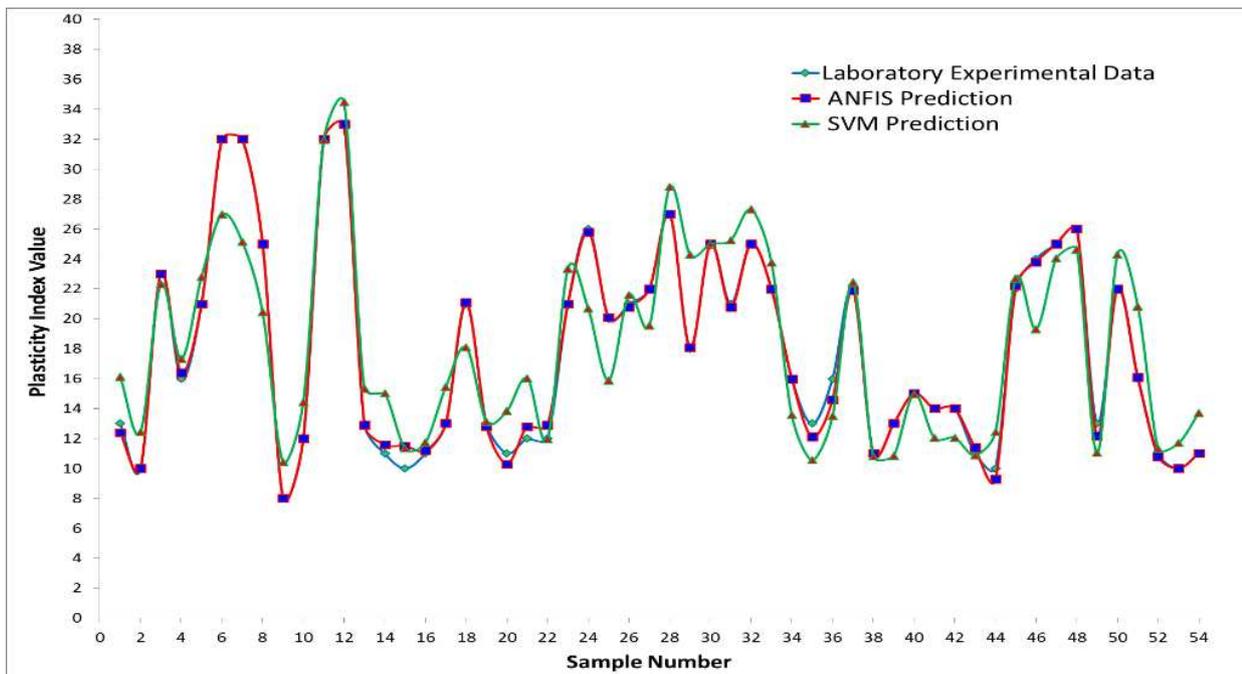


Fig 8. Predicted and actual plasticity index values using SVM and ANFIS models with 3 input

Table 2. Comparison of correlation coefficient values, R^2 for SVM and ANFIS models

No.	Parameter	SVM	ANFIS
1	Liquid limit	0.835	0.987
2	Plastic Limit	0.578	0.949
3	Plasticity Index	0.831	0.996

Table 3. Comparison of RMSE values for SVM and ANFIS models

No.	Parameter	SVM	ANFIS
1	Liquid Limit	3.378	0.957
2	Plastic Limit	1.798	0.615
3	Plasticity Index	2.776	0.421

5. Modification Of Svm Model

To find out how the number of total input can change the outcome of the prediction by the SVM model, the model is analyzed by carrying out modifications for the amount of inputs used. The amount of inputs used for both models are modified from two inputs to the three inputs by using the percentage of silt and clay fraction was then added to the three inputs of the percentage of sand, silt and clay. These modifications are briefly described in Table 5 below.

A. Total Input SVM

To find out how the number of total input can change the outcome of the prediction by the SVM

model, the model is analyzed by carrying out modifications for the amount of inputs used. The amount of inputs used for both models are modified from two inputs to the three inputs by using the percentage of silt and clay fraction was then added to the three inputs of the percentage of sand, silt and clay.

Fig 9, 10 and 11 show the results of the SVM model predictions for the three tests Atterberg limits on the amount of inputs used. As shown in Fig 4.16, the SVM model predictions for the liquid limit test that uses three input be represented by the red line is closer to the actual data (green lines) than the two input be represented by yellow line. Large errors also occur in most of the samples as an example, the samples 2, 4, 6, 7, 15,

16, 17, 25, 26, 27, 30, 36, 43, 44, 53, 54 for the two-input SVM model predictions away from the true value.

Similarly in Fig 10 below shows the results of the predictive value of the plastic limit of the SVM model that uses three input a little bit accurate than using two input model. The difference between the SVM prediction model that uses two input too much away from the actual value.

In conclusion, based on Fig 9, 10 and 11, the results of SVM model predictions indicate that the modifier amount of inputs used by the model is related to the value of output produced. This is evidenced also by the R^2 obtained as a result of the analysis. Table 5 below shows the value of the coefficient R^2 obtained after doing an analysis of both models. Comparison of the coefficient R^2 obtained from SVM model are shown in Table 5 below.

Table 5. Modification Total Input

Total Input	Percentage (%)	Output		
2 input	Clay and Slt	Liquid Limit	Plastic Limit	Plasticity Index
3 input	Sand, Clay and Slt			

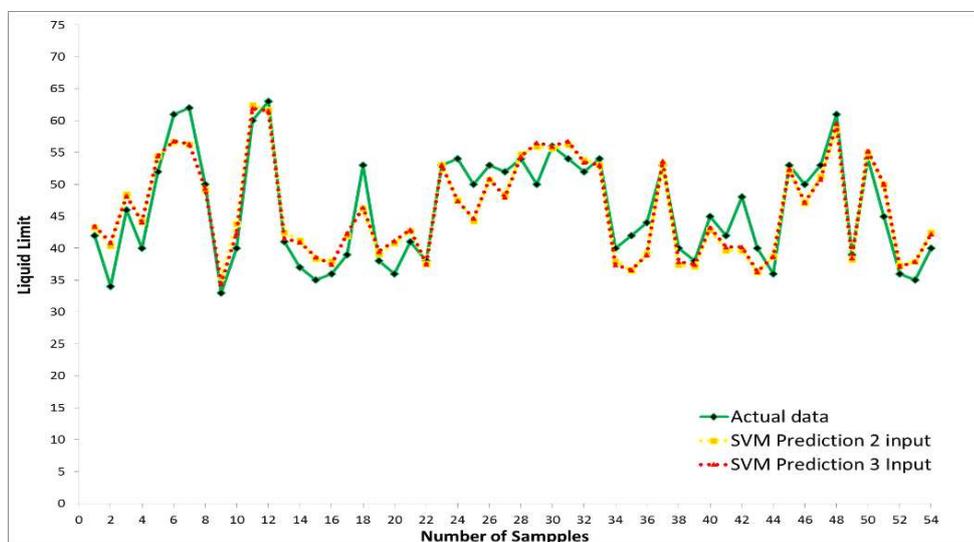


Fig 9. Comparison of results for Liquid Limit Prediction Model Based on SVM Total Input

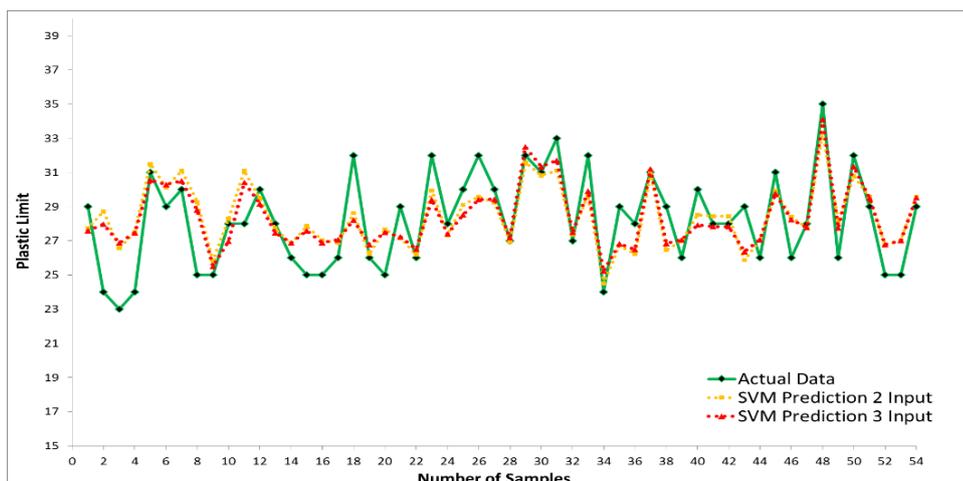


Fig 10. Comparison of Result for Plastic Limit Prediction Model Based on SVM Total Input

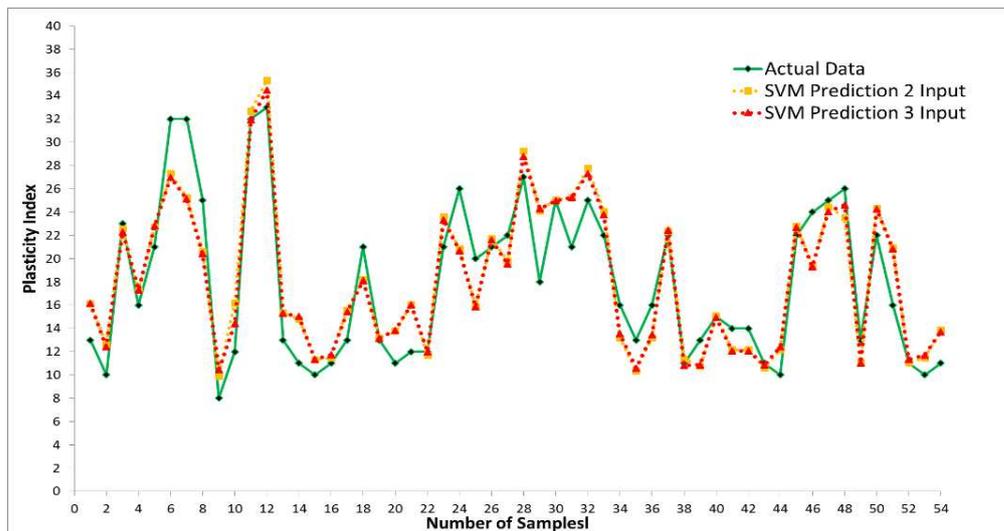


Fig 11. Comparison of results for Plasticity Index Forecast Based on SVM Model Total Input

Table 5. Comparison of R^2 values for SVM Model Based on Total Input

No.	Parameter	SVM Model	
		2 Input (Clay and Silt)	3 Input (Sand, Clay and Silt)
1	Liquid Limit	0.830	0.835
2	Plastic Limit	0.538	0.578
3	Plasticity Index	0.827	0.831

The results show that the higher the number the more accurate the inputs used for the prediction model. This is evidenced by the difference in the coefficient R^2 obtained for the SVM model with the input of more than the number of inputs. The three tests of the liquid limit, plastic limit and plasticity index indicate that by using more number of inputs, the higher the performance of the SVM model.

The results of the comparative value of RMSE of the amount of inputs used are shown in Table 6 below. Referring to Table 6, the SVM model performed better when using more inputs for the three tests Atterberg limits are. Lower RMSE values obtained when using three input than two inputs.

6. Modification Of Anfis Model

ANFIS model has also been modified in this study for comparison and does not respond to the modification of the model studied. The

modification is done in terms of modification of the input.

A. Total Input ANFIS

The amount of inputs used for both models are modified from two inputs to the three inputs by using the percentage of silt and clay fraction was then added to the three inputs of the percentage of sand, silt and clay. The results and the prediction of ANFIS model for the three values of Atterberg limits are shown in Figs 12, 13 and 14 ANFIS prediction that uses three input is represented by the blue line, while the ANFIS predictions for the two input lines are represented in pink.

For liquid limit test, it was found that using the ANFIS model predictions of three input is closer to the true value compared to the analysis using two inputs Similarly, the analysis of plastic limit testing and plasticity index indicate that the ANFIS prediction for the three inputs closer to the true value than two inputs.

Table 6. Comparison RMSE values for SVM Model Based on Total Input.

No.	Parameter	SVM Model	
		2 Input (Clay and Silt)	3 Input (Sand, Clay and Silt)
1	Liquid Limit	3.425	3.378
2	Plastic Limit	1.876	1.798
3	Plasticity Index	2.824	2.776

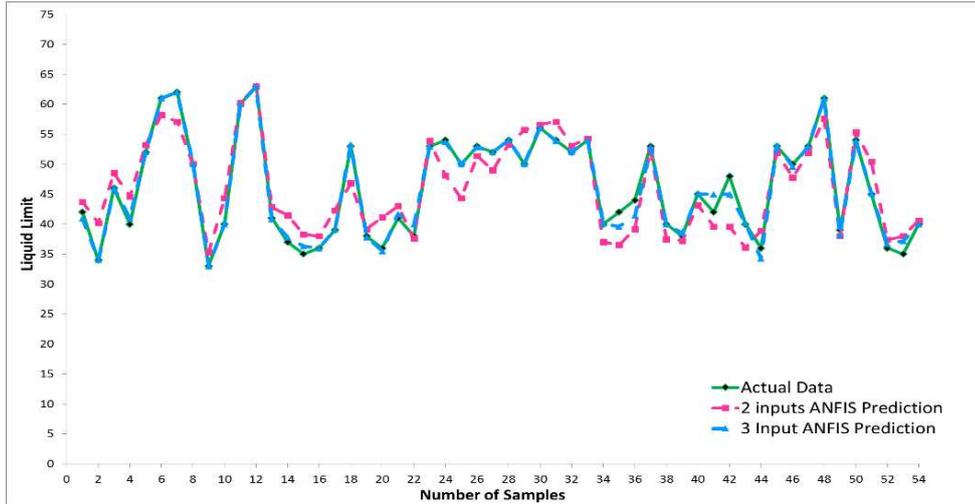


Fig 12 Comparison of results for Liquid Limit Prediction Based on ANFISModel Total Input

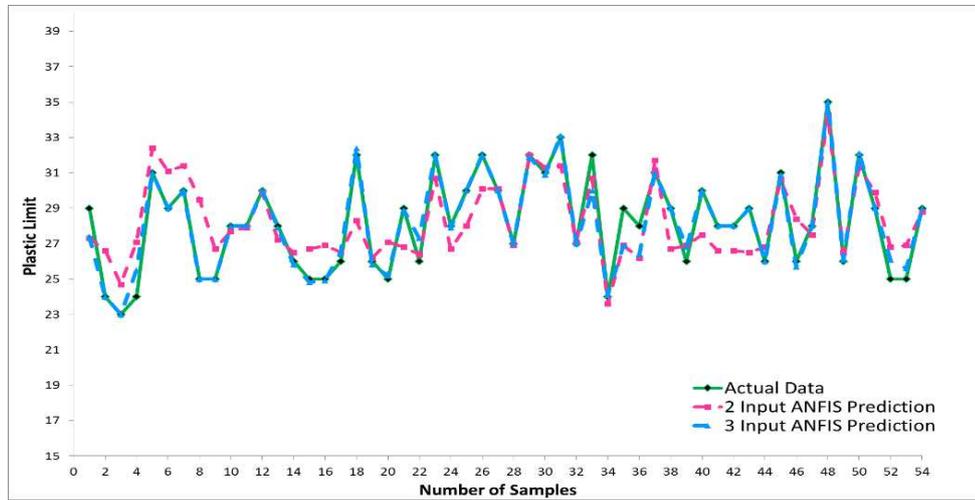


Fig 13. Comparison of Result for Plastic Limit Prediction Based on ANFISModel Total Input

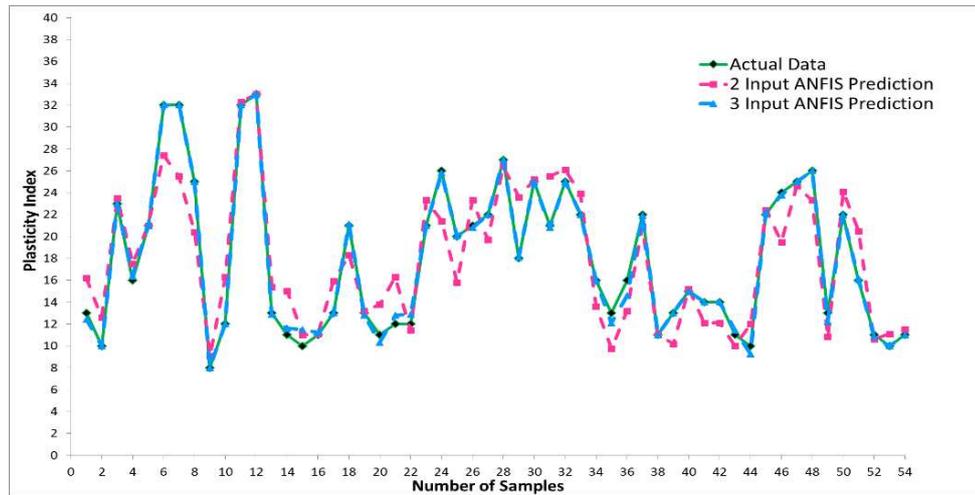


Fig 14. Comparison of Result for Plasticity Index Prediction Based on ANFISModel Total Input

Table 1. Comparison of the R^2 value for ANFIS model by Total Input

No.	Parameter	ANFIS Model	
		2 Input (Clay and Silt)	3 Input (Sand, Clay and Silt)
1	Liquid Limit	0.838	0.987
2	Plastic Limit	0.636	0.949
3	Plasticity Limit	0.835	0.996

Table 2. Comparison of RMSE values for ANFIS Model based on Total Input

No.	Parameter	ANFIS Model	
		2 Input (Clay and Silt)	3 Input (Sand, Clay and Silt)
1	Liquid Limit	3.345	0.957
2	Plastic Limit	1.647	0.615
3	Plasticity Index	2.739	0.421

Comparison of the total input ANFIS model is also reflected in the value of R^2 obtained as shown in Table 7 R^2 values obtained for ANFIS model that uses two inputs for limit liquid testing is 0.838 increasing to 0.987 for the model using three inputs. Similar results were also obtained for analysis of plastic limit testing and plasticity index of the value of R^2 is also increased when the input is increased from two to three input.

Referring to Table 8 the results for the low RMSE also obtained by ANFIS model for the analysis of the three liquid limit, plastic limit and plasticity index when the three inputs used RMSE values for liquid limit decreased from 3.345 to 0.957 Similarly, the plastic limit testing RMSE values decreased from 1.647 to 0.615 The index test plastic, the RMSE values obtained decreased from 2.739 to 0.421 Thus, we can conclude that, the RMSE obtained was dependent on the modification of the number of inputs used in the ANFIS model.

Conclusion

From the results obtained, it can be concluded that the prediction by ANFIS method shows higher accuracy than the SVM method for the liquid limit plastic limit and plasticity index. R^2 coefficient and RMSE values obtained for both methods also showed ANFIS model performed better than the SVM model in predicting the Atterberg limits as a whole. The outcome of the study show that the ANFIS model shows higher accuracy than SVM model for the liquid limit ($R^2 = 0.987$), plastic limit ($R^2 = 0.949$) and plastic index ($R^2 = 0.966$). RMSE value that obtained for both methods have shown that the ANFIS model has represent the best performance than SVM model to predict the Atterberg Limits as a whole. Modifications of SVM and ANFIS models have been done in order to evaluate the response of the output to the modification and the efficiency of the model.

References

- Ali, T.Y., 2011. Effect of fine particles on shear strength parameter of sand (thesis). Faculty of Civil Engineering.
- Berbenni, S., Favier, V., Berveiller, M., 2017. Impact of the grain size distribution on the yield stress of heterogeneous material. Impact of the grain size distribution on the yield stress of heterogeneous material 23, 114–142.
- Boser, B.E., Guyon, I.M., Vapnik, V.N. 1992 A training algorithm for optimal margin classifiers. In: In D. Haussler, editor, 5th Annual ACM Workshop on COLT, pages 144–152, Pittsburgh, PA, ACM
- Chen, S.T., Yu, P.S. & Tang, Y.H. 2010. Statistical downscaling of daily precipitation using support vector machine and multivariate analysis. Journal of hydrology. 385: 13–22
- Cortes, C. and Vapnik, V. 1995. Support-vector networks. Machine Learning. Volume 20, Number 3. pp 273–297
- Fletcher, T. 2009. Support Vector Machine Explained (online) <http://www.tristanfletcher.co.uk/SVM%20Explained.pdf> (22December2011)
- IKRAM 2011. Report of Study on Debris Flow Controlling Factor and Triggering System in Peninsular Malaysia. Institute of Public Work. Malaysia
- Jang, J.S.R. 1993. ANFIS: Adaptive –Network-based Fuzzy Interference System. IEEE Transaction on System, Man, and Cybernetic, 23(03):665–685
- Lin, H.J. & Yeh, J.P. 2009. Optimal reduction of solution for support vector machines. Applied Mathematics and Computation. 214: 17–29
- Tripathi, S., Srinivas, V.V. & Nanjundiah, R.S. 2006. Downscaling of precipitation climate change scenarios: A support vector machine approach. Journal of Hydrology. 330(3–4): 621–640
- Vapnik, V.N. 1995. The nature of statistical learning theory. New York: Springer Verlag.