Effects of significant variables on compressive strength of soil-fly ash geopolymer: Analytical approach based on neural networks and genetic programming

Hsiao Yun Leong[1], Dominic Ek Leong Ong[2*], Jay G. Sanjayan[3], Ali Nazari[4], Sze Miang Kueh[5]

[1] Ph.D scholar, Research Centre for Sustainable Technologies, Faculty of Engineering, Science & Computing, Swinburne University of Technology Sarawak Campus, 93350 Kuching, Sarawak, Malaysia. E-mail: hyleong@swinburne.edu.my

[2] Associate Professor, Research Centre for Sustainable Technologies, Faculty of Engineering, Science & Computing, Swinburne University of Technology Sarawak Campus, 93350 Kuching, Sarawak, Malaysia. E-mail: elong@swinburne.edu.my

[3] Professor, Centre for Sustainable Infrastructure, Faculty of Science, Engineering and Technology, Swinburne University of Technology, PO Box 218, Hawthorn, Victoria 3122, Australia. E-mail: jsanjayan@swin.edu.au

[4] ARC Future Fellowship, Centre for Sustainable Infrastructure, Faculty of Science, Engineering and Technology, Swinburne University of Technology, PO Box 218, Hawthorn, Victoria 3122, Australia. E-mail: alinazari@swin.edu.au

[5] Ph.D graduate, Research Centre for Sustainable Technologies, Faculty of Engineering, Science & Computing, Swinburne University of Technology Sarawak Campus, 93350 Kuching, Sarawak, Malaysia. E-mail: skueh@swinburne.edu.my

*corresponding author

Abstract

The identification of significant input variables to the output provides very useful information for mix design for soil-fly ash geopolymer in order to obtain the optimum compressive strength. The importance of input variables to the output of soil-fly ash geopolymer was quantified by Garson's algorithm and connection weights approach in artificial neural networks (ANN) model, whereas model analysis and fitness method were used in genetic programming (GP) model. The former approaches in the ANN model used the connection weights among the input-hidden-output layers to evaluate the importance of the input variables. The latter methods in the GP model assessed the frequency of variables used in the model and the value of fitness for the evaluation. The assessment results identified the percentages of fly ash, water and soil as important input variables to the output. The percentage of hydroxide, the ratios of silicate/hydroxide and alkali activator/ash were ranked as less important input variables. The positive or negative relationships between these input variables and the output demonstrated a very significant influence on the strength development of soil-fly ash geopolymer, showing a positive or negative effect on the compressive strength.

Keywords

Artificial intelligence approaches; geopolymer; compressive strength; fly ash; importance analysis

Introduction

Geopolymer is a recent popular environmental friendly construction material. It is commonly produced by activating fly ash, an industrial by-product from the combustion of coal-fired power plant, with alkali activators such as sodium silicate and sodium hydroxide in an alkaline environment at the low curing temperature (Davidovits 2008). Being a green material, geopolymer is found to emit carbon footprint 80 % less than the traditional Portland cement (Duxson et al. 2007). It has been used to construct structural elements of buildings such as beams, columns and piles (Power Pile 2013; Sarker 2008; Shrest 2013; Uretek 2014), besides also being viable as a ground treatment method when mixed with weaker clayey soils (Cristelo et al. 2013; Cristelo et al. 2012; Zhang et al. 2013). The strength development of geopolymer is mainly attributed to the formation of geopolymer gel (i.e. three-dimensional polymer chains) in the geopolymer structure through geopolymerization (Davidovits 2008; Komljenovi et al. 2010; Tennakoon et al. 2014). Factors influencing the compressive strength of geopolymer were reported to be the ratio of alkali activator to ash, the ratio of silicate to hydroxide, types of alkali activator, percentage of fly ash etc. (Bakri et al. 2009; Hardjito and Rangan 2005; Kong and Sanjayan 2008; Leong et al. 2016; Xu and Deventer 2000). As the compressive strength of geopolymer could be affected by various factors, a prediction model could effectively predict the strength performance of geopolymer; for instances, the modeling methods in the artificial neural network (ANN) and genetic programming (GP).

Recently, GP is used in many civil engineering applications (Cladera et al. 2014; Garg et al. 2014a; Garg et al. 2014b; Sarıdemir 2014). The novelty of GP is its capability to generate the prediction equation and expression trees from the model, providing a very useful information for strength prediction. For ANN, it is known for being applicable to solve problems in a wide range of different fields such as biological, business, environmental, manufacturing etc. In civil engineering, it can be used to detect structural damage, to model material behavior, to monitor groundwater, to predict mix ratios etc (Kewalramani and Gupta 2006).

It is known that the most common method to acquire the compressive strength of geopolymer is through laboratory tests on samples made from various mix ratios. The compressive strength is attainable from the selection of the appropriate mix ratio, where each parameter has the capability to influence the strength development of the geopolymer. Hence, a vast number of laboratory tests is required to understand the contribution of each parameter to strength development. This process is not only costly and time-consuming, but requires a large quantity of raw materials for sample preparation. Therefore, the use of predictive models such as artificial neural networks (ANN) and genetic programming (GP) could be effective tools in predicting the strength capability of each mix design. The predictive model is developed by employing the experimental dataset through minimizing the difference between the target and experimental output in order to obtain optimal solutions. The modification in the mix design could then aid in identifying the appropriate mix ratio for an idealized compressive strength, besides being more cost effective during production (Chopra et al. 2016; Flood and Kartam 1994; Nikoo et al. 2015).

Both types of modeling methods have been widely utilized to predict compressive strength of concrete and geopolymer by many researchers (Atici 2011; Castelli et al. 2013; Chopra et al. 2016; Hong-Guang and Ji-Zong 2000; Kostic and Vasovic 2015; Leong et al. 2015; Mozumder and Laskar 2015; Nazari 2013; Nazari and Torgal 2013a; Nazari and Torgal 2013b; Sarıdemir 2010; Sobhani et al. 2010; Topcu and Sarıdemir 2008; Yadollahi et al. 2015). The models have generated very close results between experimental data and predictive models. Hence, it indicates that both modeling methods are capable of being used for strength prediction. However, the importance of the input variables to the output of a typical geopolymer model has rarely been investigated.

Factors influencing the compressive strength of soil-fly ash geopolymer were studied and evaluated based on the experimental results in Leong et al. (2017). Subsequently in this research, the experimental results were used as input data to the predictive ANN and GP models. These models were then used to predict the compressive strength of the soil-fly ash geopolymer. The main objective of this paper is to identify the importance of each input variable to the corresponding output through various assessment methods such as Garson's algorithm, connection weights approach, fitness method and model analysis.

Data Collection from the Experimental Works

Materials

Sarawak fly ash obtained from Sejingkat Power Station, Kuching, was used in this research. It was observed to be gray and classified as Class F type in accordance to ASTM-C618 (2005). The Sarawak fly ash was produced from the combustion of sub-bituminous coal.

The residual soil was sourced from the outskirt of Kuching city, Malaysia. It is classified as reddish brown slightly sandy silt. The natural moisture content of the soil is 53 %. It consists of 11 % gravel, 13 % sand, 34 % silt and 42 % clay. The liquid limit, plastic limit and plasticity index of the residual soil are 66 %, 36 % and 30 % respectively. The optimum moisture content (OMC) and the maximum dry density (MDD) of the residual soil are 47 % and 1184 kg/m^3.

The chemical compositions of both fly ash and soil were obtained using WD-X-ray Fluorescence Spectrometer (WD-XRF) as shown in **Table 1**. Both types of materials are rich in SiO_2, Al_2O_3 and Fe_2O_3. **Figs. 1a** and **b** shows the morphologies of fly ash and soil respectively, under scanning electron microscopy (SEM) (brand: ZEISS SUPRA 40 VP SEM). The former was observed to have smooth spheres whereas the latter consisted irregular shapes with porous-like structure. **Fig. 2.** shows the particle size distributions (PSD) of both fly ash and soil. It was conducted using a laser particle size analyzer (brand: CILAS 1190). The PSD plot shows that the fly ash particles are predominantly smaller than soil particles, comprising 90 % of the overall particles.

The combination of 8M KOH (or NaOH) and sodium silicate (17% Na_2O and 35% SiO_2 by weight) was selected as the alkali activators used in this research.

Sample Preparation and Test

The residual soil was placed into the oven for dehydration for 24 hours. The dry soil was initially premixed with fly ash in a mixer prior to the addition of alkali activators and water. The ratios of alkali activator/ash were varied in the range of 0.4 to 0.7 whereas the ratios of Na_2SiO_3/NaOH (or KOH) were 0.5 and 1. Different percentages of additional water (10%, 20%, 30% and 40%) were added to the mixture. The ratio of fly ash/soil was kept constant throughout the study (i.e. 0.8). Although a higher ratio of fly ash/soil could lead to higher compressive strength, strength gain between these two ratios was not significant as observed in Leong et al. (2017). Moreover, increasing the fly ash content could increase the amount of alkali activators required for producing the geopolymer. Hence, fly ash/soil of 0.8 was selected in term of cost effectiveness and potentially good strength development. The mixing process was conducted until the mixture was mixed homogeneously. The mixture was poured into the modified cube mold (50 mm × 50mm × 120 mm). The compression machine was used to press the sample into cubes with dimensions of 50 mm × 50 mm × 50 mm at a press load of 10 kN. The samples were then demolded and cured in an oven at 100 °C for 24 hours. After 24 hours, the samples were tested for compressive strength using compression test machine in accordance to ASTM-C109/C109M (2005). Three samples were tested for each mix ratio. The mix ratios and the corresponding compressive strengths of the samples are tabulated in **Table 2**. A number of 64 datasets were collected from the experiments and these data were used in the predictive modeling, which shall be discussed further in the next section. **Table 3** summarizes the statistical values used in the predictive modeling. **Fig. 3.** illustrates the cumulative percentage and frequency distribution of input variables and target.

Predictive modeling

Artificial Neural Network

Artificial neural network (ANN) is inspired by the biological neural network. The structure of neuron consists of soma (i.e. where the nucleus is to process the inputs), dendrite, synapse and axon. The tree-like fibre namely dendrite is a receptor to accept the signal or input. The long single fibre cell namely axon transfers the signal from a synapse to another receiving end of the synapse of the neutron. The fundamental structure and operation of the biological neural network motivate the development of the artificial neural network with some significant features relating to computing model for pattern recognition tasks (Yegnanarayana 2006).

ANN is a mapping of input into the desired output. It consists of weighted inputs, transfer function and output. Learning is defined as a process where the weights are adjusted to obtain the desired input-to-output mapping. A simple feed-forward neural network is commonly used in ANN. The input at first layer is fed into the interconnecting layer namely hidden layers, follows by the transfer function, which does not affect the feed-forward behavior in neural network and finally the output layer.

In this research, the ANN model was developed in MATLAB R2014b using the neural network toolbox. Seven input parameters, one hidden layer with ten hidden neurons and one output parameter were employed in the ANN model. The input parameters were the percentage of soil, percentage of fly ash, percentage of hydroxide, percentage of silicate, the percentage of water, the ratio of alkali activator/ash and the ratio of silicate/hydroxide. The compressive strength of soil-fly ash geopolymer was assigned as the output parameter. The interpretation diagram of the ANN model is illustrated in **Fig. 4.** Levenberg-Marquardt back-propagation was used as a training algorithm and log sigmoid function was selected. The gradient descent algorithm reduces the error by adjusting the connection weights along the gradients. The performance of the ANN model was examined by mean square error (MSE).

Genetic Programming

Genetic Programming (GP) is an evolutionary algorithm-based methodology inspired by biological evolution to solve the task of relating independent input parameter to an output parameter through linear or nonlinear equations. It is an evolving program, which is an extension of Genetic Algorithm (GA) (Affenzeller et al. 2009; Koza 1994). It shows analogy to GA, where a population of solution candidates are given a problem and it works based on the Darwinian principle (i.e. the survival of fitness). The basic genetic operators in GP are similar to GA such as crossover, mutation and reproduction, with the exception of tree representation (Sivanandam and Deepa 2008). Mutation and crossover are the main operators used. The individual computer programs in an initial population are randomly generated. The selection of individuals is based on its fitness for crossover. GP is expressed by syntax tree (expression tree or explicit tree) instead of traditional lines of code. It can be expressed in linear formulae. Terminals are defined as the variables (i.e. x, y etc.) and constants in the syntax tree whereas functions are defined as the internal nodes such as arithmetic operations (i.e. +, -, *, /), mathematical operations (i.e. sin, cos, exp. etc.), conditional operations (i.e. if, then, else) etc. The combination of both terminals and functions is termed as primitive set in GP.

In this research, GeneXProTools 5.0 was implemented to construct the GP model. The input parameters and the output parameter for the GP model used were identical to the parameters used for the ANN model. For instances, percentage of soil (d0 denoted as S), percentage of fly ash (d1 denoted as FA), percentage of water (d2 denoted as W), percentage of silicate (d3 denoted as Si), percentage of hydroxide (d4 denoted as H), ratio of silicate/hydroxide (d5 denoted as Na) and ratio of alkali activator/ash (d6 denoted as A) were chosen as the input parameters whereas the compressive strength (d7 denoted as fs) of soil-fly ash geopolymer was employed as the output parameter. The dataset (a total number of 64) was divided into 70% for training, 15% for validating and 15% for testing in the GP model. Ten functions used in this GP model including $+$, $-$, $\times$, $\div$, $\sqrt{}$, $\sqrt[3]{}$, *ln* and *exp*. Multiplication was selected as the linking function whereas Root Relative Squared Error (RRSE) was used. A number of 30 chromosomes, 4 genes and 10 head size were used in the GP model.

It is worth mentioning that the predictive ANN and GP models discussed in this research were trained in a supervised manner based on reliable experimental dataset. Hence, models capable of providing high accuracy and good performance were generated through the modeling process. Although these models were generated based on the input dataset, it could also be applicable to other types of materials. In this research, the input parameters used in the models were the (i) percentage of fly ash, (ii) percentage of water, (iii) percentage of silicate, (iv) percentage of soil, (v) percentage of hydroxide, (vi) ratio of alkali activator to ash and (vii) ratio of silicate to hydroxide, whereas the output parameter was the compressive strength of the geopolymer. These input parameters were classified as experimental mix design. Neither soil properties nor types of mineralogies were used as input parameters. The results obtained from these models were quantified from the aspect of experimental mix design instead of soil mineralogies. Hence, the results presented in this research could be applicable to other types of soils or materials with different mineralogies. It is important to note that materials with different mineralogies could have significant effect on the strength development of geopolymer. The amorphous phase of SiO_2 and the Al_2O_3 enhances the strength capability of geopolymer. However, SiO_2 and Al_2O_3 in crystalline phase indicates that the formation of geopolymer gel due to the chemical reaction amongst the studied materials and alkali activators could be unlikely.

Results and discussion

ANN model

Fig. 5. depicts the regression plot of the experimental data and the predicted data obtained in the ANN model. The coefficient of determination (R^2=0.9535) indicates that the model has fitted the data very well. The difference between the predicted value and the experimental value is minimal. Details of the experimental data and the predicted results obtained by the ANN model are tabulated in **Appendix A**.

For the ANN model, there are several methods to quantify the importance of the input variables. For examples, connection weights approach, Garson's algorithm, partial derivatives, forward stepwise addition, backward stepwise elimination etc (Gevrey et al. 2003; Olden et al. 2004). In this research, Garson's algorithm and connection weights approach were chosen due to their simplicity and were commonly used by other researchers (Mozumder and Laskar 2015; Olanrewaju et al. 2012; Olden and Jackson 2002).

Garson's algorithm

In Garson's algorithm, the connection weights are used to calculate the contribution of each variable (Garson 1991). **Table 4** presents the connection weights between input-hidden-output layer. The connection weight of input-hidden neurons was multiplied by the connection weight of hidden-output. The multiplication product was designated as the contribution of input-hidden-output. The relative importance of the input variables was calculated by evaluating the absolute values of the contribution. As the absolute values were used, the relationships between the input variables and the output variable were directionless since non-negative values were present. Sample calculations according to Garson's algorithm can be referred to Olden and Jackson (2002). **Fig. 6.** shows the relative importance of the input variables according to Garson's algorithm. The importance of the input variables was in the order as follows: percentage of water> ratio of alkali activator/ash> percentage of hydroxide> percentage of fly ash> ratio of Na_2SiO_3/NaOH (or KOH)> percentage of soil> percentage of silicate. The difference of each contribution is negligible, indicating all the input variables have strong relationships with the output, with the exception of the percentage of silicate.

Connection weights approach

Apart from Garson's algorithm, connection weights approach was also used to assess the importance of each input variable to the output. Similar to Garson's algorithm, connection weights approach justified the importance of each input variable to the output via multiplying the connection weights between input-hidden neurons and hidden-output neurons. However, this method did not evaluate the absolute value of each contribution. The importance of each input variable was ranked by the summation of each contribution of input-hidden-output. As presented in **Fig. 7a.**, the percentage of fly ash, percentage of hydroxide and the ratio of silicate to hydroxide were ranked as the most important input variables influencing the output whereas the ratio of alkali activators to ash and the percentage of soil showed the weakest relationships to the output.

As the result was ranked according to its summation product, which means the positive or negative value determined the assessment. It is worth to mention that the negative value implies negative effect to the output, thus indicating the importance of the inverse relationship of these input variables to the output. Hence, **Fig. 7b.** illustrates another ranking method according to the magnitude of the summation product regardless of its positive or negative sign. Nevertheless, both positive and negative relationships were shown in this figure. A value close to zero indicates the weakest relationship of input variable to the output. For instance, the percentage of silicate. The importance of the input variables according to this ranking method was in the order as follows: percentage of fly ash > percentage of soil > percentage of hydroxide > ratio of alkali activator/ash > ratio of silicate/hydroxide > the percentage of water > the percentage of silicate.

Genetic Programming

Apart from the ANN model, a GP model was also developed to study the performance of soil-fly ash geopolymer. More importantly, it aims to study the importance of each input variable to the output by another modeling method. The results of the GP model were presented in term of expression tree as depicted in **Fig. 8**. Four sub-expression trees with multiplication as linking function were obtained. The expression tree can be interpreted through the **Equation 1** as follows:

$$f_s = \frac{\mathrm{FA}-\mathrm{S}}{\left((0.84\mathrm{S}-\mathrm{FA})+\frac{\sqrt[3]{W}}{\mathrm{Na}}\right)} \cdot \Bigg((\exp \sqrt[3]{0.89FA} + \left(\sqrt[3]{0.89-H}\,.\,(A.W - 0.45)\right)\Bigg) \cdot \sqrt[3]{(H.W.\sqrt[3]{Si}) + \left((FA+Na).\frac{FA}{-5.48}\right)} \cdot \frac{A}{2Si + \frac{1.51-Na}{4.51-W}}$$

-- **Equation 1**

As shown in the prediction equation, all the input parameters were used, indicating the importance of all input parameters to the output. It may play a major or minor role in strength development. The values of R^2 in training and validation phases were 0.9197 and 0.9301 respectively. **Table 5** summarizes the values of error such as RMSE, RAE and RRSE. R^2 greater than 0.9 and low values of error show that the GP model has been trained well. It can predict the compressive strength of geopolymer with high accuracy and reliability. **Fig. 9.** illustrates the plot of experimental data compared to the predicted results in the GP model. The results show that the GP model can predict the compressive strength close to the experimental results. It demonstrates that this GP model can be an efficient and reliable model for the prediction of the compressive strength of geopolymer with high accuracy. **Appendix A** presents the details of experimental data and the predicted results obtained in the GP model.

In this research, two types of assessment method namely model analysis and fitness method were used to evaluate the importance of the input variables in the GP model. Detailed discussions are presented subsequently.

Model analysis

Fig. 10. presents the importance of the input variables to the output obtained from GP. This result is evaluated by quantifying the recurrence of each independent variable in the GP models from the best model structures (Sreekanth and Datta 2012). The results show that the percentage of fly ash and the percentage of water were the most important input variables to the output, indicating a high frequency of these input variables used in the GP model. The percentage of silicate, the percentage of hydroxide and the ratio of silicate to hydroxide show moderate significance to the output, whereas the percentage of soil and the ratio of alkali activator to ash show the least significance to the output.

The importance of each input variable to the output was also quantified by assessing the correlation coefficient of each input variable in training and validation phases as tabulated in **Table 6**. The percentage of soil and the percentage of fly ash exhibited the strongest positive relationship between the input variables and the output, whereas the percentage of water shows the strongest negative relationship to the output. This result is consistent with the result as reported in the ANN model. For the remaining input variables, correlation coefficients yield close to zero, thus denoting the weak relationships between the input variables and the output.

Fitness method

In this section, the frequency of variables used in the GP models was further evaluated by analyzing its contribution to the fitness in training and validation phases. As mentioned earlier on, GP works based on the Darwinian principle, which means survival of the fittest. Evaluation of the input variables according to the fitness could effectively provide significant insight into the importance of each input variable to the output. Therefore, this method is proposed and described in details as follows:

50 GP models with different frequency of input variables used were selected. The selection was determined based on the values of R^2, which is greater than 0.9 to ensure the reliability and accuracy of the model. These 50 GP models might not fall in the range of the best models in term of R^2 because the best 50 models with the highest R^2 were observed to have a consistent frequency of input variables used. In this case, the contributions of different frequency of input variables used to the value of fitness might not be significant. Consequently, the assessment of the importance of the input variables to the output according to the value of fitness might not reliable. The selected 50 GP models are represented in **Appendix B**.

The relative frequency of variable used for each input variable was evaluated by dividing the individual frequency of variable used from the overall frequency of variables used (see **Appendix C**: Step A). Hence, it shows the proportion of each input variable in respective GP model. The individual relative frequency of variable used obtained in **Appendix C**: Step A was then multiplied by the fitness as shown in **Appendix B**. This step evaluated the contributions of fitness according to the relative frequency of variable used in training phase (see **Appendix C**: Step B) and validation phase (see **Appendix C**: Step C), respectively. The results showed that the difference between the contributions of fitness in training and validation phase is minimal. The summation of the contribution of fitness for each input variable was calculated in order to quantify the relative importance of the input variable to the output (see **Appendix C**: Step D).

The assessment method was ranked in descending order as illustrated in **Fig. 11**. Both training (see **Fig. 11a**) and validation phases (see **Fig. 11b**) show similar results, representing consistency of the importance of the input variables to the output regardless of the computational phases. The significant input variables were ranked as the percentage of water and the percentage of fly ash, followed by the ratio of silicate to hydroxide, the percentage of silicate and the percentage of hydroxide. The ratio of alkali activator to ash and the percentage of soil were identified as the least significant input variable to the output.

Significance of input variables to the output

The assessment results of the importance of the input variables to the output in both ANN and GP models is summarized in **Table 7**. Different ranks of the importance of the input variables were observed. Hence. the reliability of each evaluation approach is assessed hereinafter.

Garson's algorithm and connection weights approach in the ANN model

Dissimilarities between the ranks of the significance of input variables were found on both Garson's algorithm and connection weights approach. It can be attributed to the respective assessment method as the former used absolute values to calculate the variable contributions as mentioned earlier on. Hence, the positive and negative influences to the output were negligible. For the latter assessment method, it used original values instead of absolute values. Both positive and negative relationships were shown. From the point of geopolymer study, it is important to show the true direction of relationships between the input variable and the output. For instances, the positive relationship between the percentage of fly ash and compressive strength indicates that the increase of fly ash content could lead to higher compressive strength. As a negative relationship was present between the percentage of water and compressive strength, the addition of water to the mixture could reduce, in order to prevent the decline of strength capability. It showed that connection weights approach demonstrated better assessment method to interpret the importance of the input variables than Garson's algorithm due to its accuracy and precision. Moreover, it showed the individual interacting relationship between the input variable and the output, which forms an important study and contribution to the strength development of geopolymer.

Model analysis and fitness method in the GP model

Comparing the assessment methods according to the model analysis and fitness method, it is found that the former could not compute the actual rank for all the input variables. It is because model analysis method evaluated the importance of the input variables by quantifying the recurrence of each input variable from the best model structures. These structures were identical to one another with a slight variation of R^2 but similar functions and constant used in each gene. Similar frequencies of the variable used among these models could not significantly contribute to the assessment. Therefore, this method could not rank all the input variables accordingly.

Unlike model analysis, fitness method used the contributions of fitness evaluated from the frequency of variable used to assess the importance of each input variable. It showed a very prominent result as the input variables which could not be ranked using model analysis, were ranked accordingly using fitness method. Moreover, the rank of these input variables was in the range as ranked using model analysis. It demonstrates that the fitness method could accurately assess the importance of the input variables to the output. Furthermore, fitness method quantified the GP models with different frequency of variable used instead of selecting the best models with similar structures. It considered a wider range of data than the model analysis. The superior assessment results have shown that fitness method is suitable and more reliable to evaluate the importance of the input variables to the output in the GP model.
Therefore, connection weights approach and fitness method were selected as the most suitable assessment methods for the ANN and GP models. However, both assessment methods show dissimilar ranks for the importance of the input variables to the output. It could be attributed to their respective evaluation method, i.e. the former used the connecting weights among the input-hidden-output layers for the assessment, whereas the latter evaluated the contributions of fitness according to the frequency of input variable used. Hence, statistical analysis evaluation method might help to analyze the results obtained from both assessment methods, as well as being used to interpret the significance of each input variable to the output.

Statistical analysis

Table 8 shows the statistical analysis of the importance of the input variables to the output. The correlation coefficient represents the magnitude and direction of the relationship between each input variable and the output. The analysis shows that the percentage of soil and the percentage of fly ash have the strongest positive relationships with the output. Contrariwise, the percentage of water shows a negative relationship to the output. The percentage of silicate demonstrated the weakest relationship as the correlation coefficient was close to zero. p-value indicates the significance of input variables to the output by the significance level of 0.05.

In this research, the null hypothesis was defined as the insignificant relationship between the input variable and the output. The null hypothesis is true when the p-value is greater than 0.05; however, it is rejected when the p-value is less than 0.05, indicating the input variable is important to the output. The results show that the input variables such as the percentage of soil, the percentage of fly ash and the percentage of water were statistically significant to the output. However, the remaining input variables such as the percentage of silicate, the percentage of hydroxide, the ratio of silicate to hydroxide and the ratio of alkali activators to ash showed insignificant relationships to the output. It implies that the significant input variables were relatively more important than those insignificant input variables. Effects of each input variable on the compressive strength of geopolymer are discussed subsequently.

As shown in **Fig. 12a**, increasing the percentage of fly ash has a positive effect on the compressive strength of geopolymer. It is known that fly ash is the main source of aluminosilicate that produces geopolymer, whereas the alkali activators play a role in activating the fly ash particles. Increasing the fly ash content thus enhances the geopolymerization process, which is crucial in strength development. A rigid geopolymer structure is formed when fly ash completely reacts with the alkali activators. Hence, the formation of geopolymer gel binds the soil particles together, resulting in higher strength capabilities (Heah et al. 2012; Liew et al. 2011; Sukmak et al. 2013a; Sukmak et al. 2013b). Any further increase in the percentage of fly ash but with insufficient increase in alkali activators in the geopolymer system, however, could reduce the strength development of geopolymer. This phenomenon could be attributed to the excessive unreacted fly ash particles present in the geopolymer sample.

As depicted in **Fig. 12b**, the percentage of soil used shows similar development of geopolymer compressive strength as observed in the percentage of fly ash used (see **Fig. 12a**). The compressive strength of geopolymer increases when the percentage of soil increases. As reported in Leong et al. (2017), some of the cations from the alkali activators such as Na^+ and K^+ could be attracted to the negatively charged surfaces of the soil particles. Increasing the percentage of soil could result in high concentration of absorbed cations near the surfaces of the soil particles. Hence, anions such as OH^- and SiO_3^{2-} from alkali activators are available surrounding the fly ash particles. These are the essential ions required to dissolve the fly ash particles for the geopolymerization process to occur. As the fly ash particles are smaller than the soil particles, both types of particles can be closely packed together via the "filler effect". Thus, the formation of geopolymer gel binds the soil particles together, improving the strength capabilities. However, the presence of a high percentage of soils may lead to a negative effect on the geopolymer compressive strength. The negatively charged surfaces of the soil particles may interact with water (Mitchell and Soga 2005), creating an interface layers detrimental to forming the necessary bonds amongst particles. Furthermore, the presence of organic matters in the soil may absorb some of the alkali activators, which play an essential role in the geopolymerization process, thus reducing the strength development.

The effect of water on compressive strength of geopolymer is illustrated in **Fig. 12c**. It demonstrates that water has a negative correlation with the strength development of geopolymer. It is known that alkali activators play an important role in strength development of geopolymer, however, the addition of water could intensively reduce its alkalinity and concentration (Aliabdo et al. ; Leong et al. 2016; Patankar et al. 2013; Zuhua et al. 2009).

Fig. 12d and **Fig. 12e** show the effects of both hydroxide and silicate on the development of the compressive strength of geopolymer. The data points are scattered with no particular trend, indicating very low correlation with the strength development of geopolymer, thus presenting very contrasting effects from **Fig. 12a-c** which evidently show obvious correlations (either positive or negative direction) with strength development. These results show good agreement with the assessment results reported in the previous section of this paper. Both silicate and hydroxide show relative insignificant importance to the output in comparison to the more significant input parameters such as fly ash, soil and water. Theoretically, increasing the percentage of silicate and the percentage of hydroxide improves the geopolymerization process and strength capabilities of geopolymer. However, the addition of water reduces the alkalinity and concentration of the alkali activators. Besides, some of the alkali activators may interact with soil particles or absorbed by the organic matters present. Alkali activators enhance the activation of fly ash and stimulates the geopolymerization process. However, with an increase in alkali activators in the mixture but having reduced fly ash content could result in strength retardation.

The effects of both ratios of silicate to hydroxide and alkali activators to ash on compressive strength of geopolymer are illustrated in **Fig. 12f and Fig. 12g,** respectively. It shows that both lines of the plot in **Fig. 12f** are similar, indicating that reducing the amount of silicate but increasing the amount of hydroxide in the geopolymer system does not significantly improve the compressive strength. It implies that an adequate amount of hydroxide and silicate should be used. Increasing the ratio of alkali activator/ash increases the compressive strength of geopolymer as shown in **Fig. 12g**. It could be due to the greater dissolution of fly ash and better geopolymerization process, forming a rigid and coherent geopolymer structure. However, the geopolymer structure tends to become weaker and porous as the ratio of alkali activator/ash increases (Leong et al. 2016).

Conclusions

The prediction models in both ANN and GP have been successfully developed. The values of R^2 greater than 0.9 reveals that the results obtained from the generated models fitted the experimental data very well, indicating the reliability of the prediction models. The prediction equation obtained from the GP model provides useful information for the prediction of compressive strength derived from the various mix design, besides shortening the time required to estimate the strength value effectively. Other than this, the importance of the input variables to the output in both ANN and GP models have been evaluated through various assessment methods. As the input variables could influence the output significantly, the relationships between the input variables and the output results could be important for the mix design of soil-fly ash geopolymer. Variables with high significant value contribute to the positive effect of the output and vice versa. Insignificant variables might reduce the accuracy and performance of the model. Therefore, the evaluation of the importance of the input variables to the output results from this research could provide useful insights into the development of compressive strength of soil-fly ash geopolymer. The conclusions derived from this research are as follows:

1. Garson's algorithm and connection weights approach were selected to quantify the importance of the input variables to the output in the ANN model. Connection weights approach demonstrated better assessment capability then Garson's algorithm as it showed the true relationships (i.e. positive or negative effect) between the input variables and the output. Garson's algorithm used absolute values to evaluate the assessment, hence showing a directionless result.

2. For the GP model, fitness method and model analysis were used to assess the importance of the input variables. The former exhibited superior assessment ability to evaluate the importance of the input variables than the latter. It is because fitness method used the contribution of fitness according to the frequency of input variable used in the model to assess the significance of each input variable. However, for model analysis, it quantified the recurrence of each input variable from the best model structures. These models presented identical structures to one another, hence, it could not rank all the input variables accordingly.

3. The percentage of soil, the percentage of fly ash and the percentage of water were identified as the significant input variables to the output. The positive or negative relationship provides a very important insight into the strength development of soil-fly ash geopolymer. These variables could influence the formation of geopolymer gel in the structure and alkalinity of the alkali activators that could subsequently lead to strength differences.

4. Factors affecting the compressive strength of geopolymer have been evaluated in details. Increasing the amounts of fly ash and soil improve the geopolymer strength. It could be due to better geopolymerization within the sample, forming a rigid and coherent structure. However, water has a negative effect on the development of the compressive strength of geopolymer. The amounts of silicate and hydroxide used show low correlations to the strength development. High ratios of silicate to hydroxide and alkali activator to ash do not necessarily lead to higher compressive strength. An adequate ratio should be used to obtain the optimum compressive strength.

Acknowledgements

The first author would like to show her gratitude to Swinburne Sarawak Research Centre for Sustainable Technologies for the financial support throughout her study.

References

Affenzeller, M., Winkler, S., Wagner, S., and Beham, A. (2009). *Genetic Algorithms and Genetic Programming* Taylor & Francis Group, FL.

Aliabdo, A. A., Elmoaty, A. E. M. A., and Salem, H. A. "Effect of water addition, plasticizer and alkaline solution constitution on fly ash based geopolymer concrete performance." *Construction and Building Materials*, 121, 694-703.

ASTM-C109/C109M (2005). "Standard Test Method for Compressive Strength of Hydraulic Cement Mortars (Using 2-in. or [50-mm] Cube Specimens)." ASTM International.

ASTM-C618 (2005). "Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in Concrete."

Atici, U. (2011). "Prediction of the strength of mineral admixture concrete using multivariable regression analysis and an artificial neural network." *Expert Systems with Applications*, 38, 9609–9618.

Bakri, A. M. M. A., Kareem, O. A. K. A. A., and Myint, S. (2009). "Study on the Effect of Alkaline Activators Ratio in Preparation of Fly Ash-based Geopolymer." University Malaysia Perlis (UniMAP), Malaysia.

Castelli, M., Vanneschi, L., and Silva, S. (2013). "Prediction of high performance concrete strength using Genetic Programming with geometric semantic genetic operators." *Expert Systems with Applications* 40, 6856–6862.

Chopra, P., Sharma, R. K., and Kumar, M. (2016). "Prediction of Compressive Strength of Concrete Using Artificial Neural Network and Genetic Programming." *Advances in Materials Science and Engineering*

Cladera, A., P´erez-Ord´o˜nez, J. L., and Mart´ınez-Abella, F. (2014). "Shear strength of RC beams. Precision, accuracy, safety and simplicity using genetic programming." *Computers and Concrete*, 14(4), 479–501.

Cristelo, N., Glendinning, S., Fernandes, L., and Pinto, A. n. T. (2013). "Effects of alkaline-activated fly ash and Portland cement on soft soil stabilisation." *Acta Geotechnica (2013) 8:395–405*.

Cristelo, N., Glendinning, S., Miranda, T., Oliveira, D., and Silva, R. (2012). "Soil stabilisation using alkaline activation of fly ash for self compacting rammed earth construction." *Construction and Building Materials 36 (2012) 727–735*.

Davidovits, J. (2008). *Geopolymer Chemistry and Application* Institute Geopolymer France.

Duxson, P., Provis, J. L., Lukey, G. C., and Deventer, J. S. J. v. (2007). "The role of inorganic polymer technology in the development of 'green concrete'." *Cem. Con. Res.*, 37, 1590-1597.

Flood, A., and Kartam, N. (1994). "Neural networks in civil engineering I: principles and understanding." *Journal of Computing in Civil Engineering*
8(2), 131-148.

Garg, A., Garg, A., and Tai, K. (2014a). "A multi-gene genetic programming model for estimating stress-dependent soil water retention curves." *Computational Geosciences*, 18(1), 45–56.

Garg, A., Garg, A., Tai, K., and Sreedeep, S. (2014b). "An integrated SRM multi-gene genetic programming approach for prediction of factor of safety of 3-D soil nailed slopes." *Engineering Applications of Artificial Intelligence*, 30, 30-40.

Garson, G. D. (1991). "Interpreting neural-network coneection weights." *Artif. Intell. Expert* 6, 47-51.

Gevrey, M., Dimopoulos, I., and Lek, S. (2003). "Review and comparison of methods to study the contribution of variables in artificial neural network models." *Ecological Modelling* 160, 249-264.

Hardjito, D., and Rangan, B. V. (2005). "Development and Properties of Low-Calcium Fly Ash-based Geopolymer Concrete." Curtin University of Technology, Perth, Australia 48.

Heah, C. Y., Kamarudin, H., Bakri, A. M. M. A., Bnhussain, M., Luqman, M., Nizar, I. K., Ruzaidi, C. M., and Liew, Y. M. (2012). "Study on solids-to-liquid and alkaline activator ratios on kaolin-based geopolymers." *Constr. Build. Mater.*, 35, 912–922.

Hong-Guang, N., and Ji-Zong, W. (2000). "Prediction of compressive strength of concrete by neural networks." *Cem. Con. Res.*, 30, 1245-1250.

Kewalramani, M. A., and Gupta, R. (2006). "Concrete compressive strength prediction using ultrasonic pulse velocity through artificial neural networks." *Automation in Construction*, 15, 374 – 379.

Komljenovi, M., Bascarevi, Z., and Bradic, V. (2010). "Mechanical and microstructural properties of alkali-activated fly ash geopolymers." *J. Hazard. Mater.*, 181, 35–42.

Kong, D. L. Y., and Sanjayan, J. G. (2008). "Damage behavior of geopolymer composites exposed to elevated temperatures." *Cem. Con. Compos.*, 30, 986–991.

Kostic, S., and Vasovic, D. (2015). "Prediction model for compressive strength of basic concrete mixture using artificial neural networks." *Neural Comput & Applic* 26, 1005–1024.

Koza, J. R. (1994). "Genetic programming as a means for programming computers by natural selection." *Stat Comput* 4(2), 87-112.

Leong, H. Y., Ong, D. E. L., Sanjayan, J. G., and Nazari, A. (2015). "A genetic programming predictive model for parametric study of factors affecting strength of geopolymers." *Royal Society of Chemistry Advances*, 5, 85630–85639.

Leong, H. Y., Ong, D. E. L., Sanjayan, J. G., and Nazari, A. (2016). "The effect of different Na2O and K2O ratios of alkali activator on compressive strength of fly ash based-geopolymer." *Constr. Build. Mater.*, 106, 500–511.

Leong, H. Y., Ong, D. E. L., Sanjayan, J. G., and Nazari, A. (2017, under review). "Soil-fly 617 ash geopolymer: Strength and microstructure assessment of soil, fly ash, alkali 618 activators and water " *ASCE, Journal of Materials in Civil Engineering*.

Liew, Y. M., Kamarudin, H., Bakri, A. M. M. A., Binhussain, M., Luqman, M., Nizar, I. K., Ruzaidi, C. M., and Heah, C. Y. (2011). "Influence of Solids-to-liquid and Activator Ratios on Calcined Kaolin Cement Powder." *Physics Procedia* 22, 312 – 317.

Mitchell, J. K., and Soga, K. (2005). "Fundamentals of Soil Behavior." John Wiley & Sons, Inc, Canada.

Mozumder, R. A., and Laskar, A. I. (2015). "Prediction of unconfined compressive strength of geopolymer stabilized clayey soil using Artificial Neural Network." *Computers and Geotechnics* 69, 291–300.

Nazari, A. (2013). "Compressive strength of geopolymers produced by ordinary Portland cement: Application of genetic programming for design." *Mater. Design*, 43, 356–366.

Nazari, A., and Torgal, F. P. (2013a). "Modeling the compressive strength of geopolymeric binders by gene expression programming-GEP." *Expert Systems with Applications* 40, 5427–5438.

Nazari, A., and Torgal, F. P. (2013b). "Predicting compressive strength of different geopolymers by artificial neural networks." *Ceram. Int.*, 39, 2247–2257.

Nikoo, M., Moghadam, F. T., and Sadowski, A. (2015). "Prediction of Concrete Compressive Strength by Evolutionary Artificial Neural Networks." *Advances in Materials Science and Engineering*

Olanrewaju, O. A., Jimoh, A. A., and Kholopane, P. A. (2012). "Evaluating GHG Components using Artificial Intelligence: Connection Weight Approach." *Proceedings of the 2012 IEEE*, 1214-1217.

Olden, J. D., and Jackson, D. A. (2002). "Illuminating the ‘‘black box’’: a randomization approach for understanding variable contributions in artificial neural networks." *Ecological Modelling* 154, 135–150.

Olden, J. D., Joy, M. K., and Death, R. G. (2004). "An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data." *Ecological Modelling* 178, 389–397.

Patankar, S. V., Jamkar, S. S., and Ghugal, Y. M. (2013). "Effect of water-to-geopolymer binder ratio on the production of fly ash based geopolymer concrete." *International Journal of Advanced Technology in Civil Engineering*, 2(1), 79-83.

Pile, P. (2013). "What is Expanding PowerPile Geopolymer Pillar?" *PowerPile Pillars*, <http://powerpile.com/what-is-powerpile-expanding-polymer-pillar>.

Sarıdemir, M. (2010). "Genetic programming approach for prediction of compressive strength of concretes containing rice husk ash." *Constr. Build. Mater.*, 24, 1911–1919.

Sarıdemir, M. (2014). "Effect of specimen size and shape on compressive strength of concrete containing fly ash: Application of genetic programming for design." *Mater. Des.*, 56, 297–304.

Sarker, P. K. (2008). "Analysis of geopolymer concrete columns." *Materials and Structures (2009) 42:715–724*.

Shrest, P. (2013). "Development of Geopolymer Concrete for Precast Structures." The University of Texas, Arlington.

Sivanandam, S. N., and Deepa, S. N. (2008). *Introduction to Genetic Algorithms*, Springer, Berlin.

Sobhani, J., Najimi, M., Pourkhorshidi, A. R., and Parhizkar, T. (2010). "Prediction of the compressive strength of no-slump concrete: A comparative study of regression, neural network and ANFIS models." *Constr. Build. Mater.*, 24, 709–718.

Sreekanth, J., and Datta, B. (2012). "Genetic Programming: Efficient Modeling Tool in Hydrology and Groundwater Management " *Genetic Programming - New Approaches and Successful*, S. V. Soto, ed., InTech, 225-238.

Sukmak, P., Horpibulsuk, S., and Shen, S.-L. (2013a). "Strength development in clay-fly ash geopolymer." *Constr. Build. Mater.*, 40, 566–574.

Sukmak, P., Horpibulsuk, S., Shen, S.-L., Chindaprasirt, P., and Suksiripattanapong, C. (2013b). "Factors influencing strength development in clay–fly ash geopolymer." *Constr. Build. Mater.*, 47, 1125–1136.

Tennakoon, C., Nazari, A., Sanjayan, J. G., and Sagoe-Crentsil, K. (2014). "Distribution of oxides in fly ash controls strength evolution of geopolymers." *Constr. Build. Mater.*, 71, 72–82.

Topcu, I. B., and Sarıdemir, M. (2008). "Prediction of compressive strength of concrete containing fly ash using artificial neural networks and fuzzy logic." *Computational Materials Science* 41, 305–311.

Uretek (2014). "Deep Injection." *Methods*, <http://www.uretekworldwide.com/solutions/methods/geopolymer-injection>.

Xu, H., and Deventer, J. S. J. V. (2000). "The geopolymerisation of alumino-silicate minerals." *Int. J. Miner. Process.*, 59, 247–266.

Yadollahi, M. M., Benli, A., and Demirboğa, R. (2015). "Prediction of compressive strength of geopolymer composites using an artificial neural network." *Materials Research Innovations*, 19(6), 453-458.

Yegnanarayana, B. (2006). *Artificial Neural Networks*, Prentice-Hall of India, New Delhi.

Zhang, M., Guo, H., El-Korchi, T., Zhang, G., and Tao, M. (2013). "Experimental feasibility study of geopolymer as the next-generation soil stabilizer." *Construction and Building Materials 47 (2013) 1468–1478*.

Zuhua, Z., Xiao, Y., Huajun, Z., and Yue, C. (2009). "Role of water in the synthesis of calcined kaolin-based geopolymer." *Applied Clay Science*, 43(2), 218-223.

Appendices

Appendix A Details of experimental data and predicted results in the ANN and GP models

No	Soil	Fly ash	Water	Silicate	Hydroxide	Silicate/ hydroxide	Alkali activator/ ash	Compressive strength-experimental data	Compressive strength-prediction in the ANN model	Compressive strength-prediction in the GP model
1	45.0	36.0	4.5	4.8	9.6	0.5	0.4	4.35	3.78	1.99
2	43.1	34.5	8.6	4.6	9.2	0.5	0.4	13.9	14.1	17.6
3	41.3	33.1	12.4	4.4	8.8	0.5	0.4	12.0	12.3	11.1
4	39.7	31.7	15.9	4.2	8.5	0.5	0.4	4.24	4.41	3.67
5	43.5	34.8	4.3	5.8	11.6	0.5	0.5	14.0	15.1	19.7
6	41.7	33.3	8.3	5.6	11.1	0.5	0.5	17.3	17.5	16.4
7	40.0	32.0	12.0	5.3	10.7	0.5	0.5	6.72	6.42	5.58
8	38.5	30.8	15.4	5.1	10.3	0.5	0.5	2.42	2.14	2.61
9	42.0	33.6	4.2	6.7	13.4	0.5	0.6	24.5	28.5	22.9
10	40.3	32.3	8.1	6.5	12.9	0.5	0.6	8.64	11.1	11.4
11	38.8	31.0	11.6	6.2	12.4	0.5	0.6	2.88	3.58	4.36
12	37.3	29.9	14.9	6.0	11.9	0.5	0.6	1.60	1.83	1.45
13	40.7	32.5	4.1	11.4	11.4	0.5	0.7	15.1	16.8	14.1
14	39.1	31.3	7.8	10.9	10.9	0.5	0.7	2.96	4.78	5.29
15	37.6	30.1	11.3	10.5	10.5	0.5	0.7	1.20	1.40	1.72
16	36.2	29.0	14.5	10.1	10.1	0.5	0.7	0.880	1.04	0.290
17	45.0	36.0	4.5	7.2	7.2	1	0.4	5.42	5.60	4.33
18	43.1	34.5	8.6	6.9	6.9	1	0.4	11.7	12.1	15.6
19	41.3	33.1	12.4	6.6	6.6	1	0.4	10.4	11.9	10.7
20	39.7	31.7	15.9	6.3	6.3	1	0.4	5.24	5.11	6.26
21	43.5	34.8	4.3	8.7	8.7	1	0.5	22.7	21.9	20.4
22	41.7	33.3	8.3	8.3	8.3	1	0.5	14.5	13.9	14.7
23	40.0	32.0	12.0	8.0	8.0	1	0.5	9.70	7.98	7.53
24	38.5	30.8	15.4	7.7	7.7	1	0.5	1.79	2.96	2.20
25	42.0	33.6	4.2	10.1	10.1	1	0.6	23.1	24.5	21.8
26	40.3	32.3	8.1	9.7	9.7	1	0.6	12.9	16.7	13.4
27	38.8	31.0	11.6	9.3	9.3	1	0.6	2.44	2.88	3.72
28	37.3	29.9	14.9	9.0	9.0	1	0.6	2.06	1.59	1.43
29	40.7	32.5	4.1	11.4	11.4	1	0.7	19.5	22.2	22.1
30	39.1	31.3	7.8	10.9	10.9	1	0.7	5.90	6.94	8.50
31	37.6	30.1	11.3	10.5	10.5	1	0.7	2.21	2.94	2.77
32	36.2	29.0	14.5	10.1	10.1	1	0.7	1.63	1.29	0.480
33	45.0	36.0	4.5	4.8	9.6	0.5	0.4	2.09	3.78	1.99
34	43.1	34.5	8.6	4.6	9.2	0.5	0.4	15.6	14.1	17.6
35	41.3	33.1	12.4	4.4	8.8	0.5	0.4	12.1	12.3	11.1
36	39.7	31.7	15.9	4.2	8.5	0.5	0.4	3.17	4.41	3.67
37	43.5	34.8	4.3	5.8	11.6	0.5	0.5	22.7	15.1	19.7
38	41.7	33.3	8.3	5.6	11.1	0.5	0.5	19.7	17.5	16.4

39	40.0	32.0	12.0	5.3	10.7	0.5	0.5	5.09	6.42	5.58
40	38.5	30.8	15.4	5.1	10.3	0.5	0.5	1.87	2.14	1.45
41	42.0	33.6	4.2	6.7	13.4	0.5	0.6	32.1	28.5	22.9
42	40.3	32.3	8.1	6.5	12.9	0.5	0.6	13.2	11.1	11.4
43	38.8	31.0	11.6	6.2	12.4	0.5	0.6	3.36	3.58	4.36
44	37.3	29.9	14.9	6.0	11.9	0.5	0.6	2.00	1.83	2.20
45	40.7	32.5	4.1	11.4	11.4	0.5	0.7	14.0	16.8	14.1
46	39.1	31.3	7.8	10.9	10.9	0.5	0.7	5.33	4.78	5.29
47	37.6	30.1	11.3	10.5	10.5	0.5	0.7	1.59	1.40	1.72
48	36.2	29.0	14.5	10.1	10.1	0.5	0.7	1.73	1.04	0.290
49	45.0	36.0	4.5	7.2	7.2	1	0.4	4.88	5.60	4.33
50	43.1	34.5	8.6	6.9	6.9	1	0.4	12.6	12.1	15.6
51	41.3	33.1	12.4	6.6	6.6	1	0.4	15.5	11.9	10.7
52	39.7	31.7	15.9	6.3	6.3	1	0.4	4.62	5.11	6.26
53	43.5	34.8	4.3	8.7	8.7	1	0.5	22.6	21.9	20.4
54	41.7	33.3	8.3	8.3	8.3	1	0.5	12.9	13.9	14.7
55	40.0	32.0	12.0	8.0	8.0	1	0.5	7.46	7.98	7.53
56	38.5	30.8	15.4	7.7	7.7	1	0.5	1.98	2.96	2.20
57	42.0	33.6	4.2	10.1	10.1	1	0.6	26.3	24.5	21.8
58	40.3	32.3	8.1	9.7	9.7	1	0.6	17.4	16.7	13.4
59	38.8	31.0	11.6	9.3	9.3	1	0.6	3.29	2.88	3.72
60	37.3	29.9	14.9	9.0	9.0	1	0.6	1.31	1.59	1.43
61	40.7	32.5	4.1	11.4	11.4	1	0.7	24.7	22.2	22.1
62	39.1	31.3	7.8	10.9	10.9	1	0.7	7.29	6.94	8.50
63	37.6	30.1	11.3	10.5	10.5	1	0.7	3.75	2.94	2.77
64	36.2	29.0	14.5	10.1	10.1	1	0.7	0.980	1.29	0.480

Appendix B Details of the GP model: Frequency of variables used and fitness in training and validation phases

No	Frequency of variables used in the model, fr							Fitness-Training (Ft)	Fitness-validation (Fv)	R^2-training	R^2-validation
	Soil (S)	Fly ash (FA)	Water (W)	Silicate (Si)	Hydroxide (H)	Silicate/ Hydroxide (Na)	Alkali activator/ Ash (A)				
1	2	5	4	3	3	3	2	778.9	776.3	0.919	0.935
2	2	5	4	3	3	3	3	778.5	771.3	0.919	0.931
3	2	5	4	3	3	4	2	777.9	777.6	0.919	0.931
4	2	5	4	3	3	3	2	777.7	777.9	0.919	0.931
5	2	4	4	3	3	3	2	776.8	771.7	0.918	0.931
6	2	6	4	3	3	3	2	776.8	771.7	0.918	0.931
7	2	4	6	3	3	4	2	776.8	771.7	0.918	0.931
8	2	4	4	3	3	4	2	776.8	771.3	0.918	0.931
9	2	4	4	3	3	6	2	776.8	770.7	0.918	0.930
10	2	4	4	3	3	4	2	776.8	770.7	0.918	0.930
11	2	4	6	3	3	4	2	776.8	770.4	0.918	0.930
12	2	4	4	3	3	4	2	776.8	770.3	0.918	0.930
13	2	4	4	3	3	6	2	776.8	770.6	0.918	0.930
14	2	4	4	3	3	4	2	776.8	770.2	0.918	0.930
15	3	4	4	3	3	4	2	776.6	770.3	0.917	0.929
16	1	4	4	3	3	4	2	776.6	770.3	0.917	0.929
17	1	4	4	3	3	6	2	776.6	771.4	0.917	0.930
18	1	4	4	5	3	4	2	776.6	771.4	0.917	0.930
19	1	4	4	3	3	4	2	776.6	771.4	0.917	0.930
20	3	4	4	3	3	4	2	776.6	771.4	0.917	0.930
21	1	4	4	3	3	4	2	776.6	771.4	0.917	0.930
22	1	4	4	5	3	4	2	776.5	771.4	0.917	0.930
23	1	4	4	3	3	4	2	776.5	771.4	0.917	0.930
24	1	4	4	3	3	4	4	776.5	771.3	0.917	0.930
25	1	4	4	3	3	4	2	776.5	771.2	0.917	0.930
26	3	4	4	3	3	4	2	776.5	771.4	0.917	0.930
27	1	4	4	3	3	4	2	776.5	771.3	0.917	0.930
28	1	4	4	3	3	3	2	776.5	771.8	0.917	0.930
29	1	4	6	3	3	3	2	776.5	771.2	0.917	0.930
30	1	4	4	3	3	3	2	776.5	771.2	0.917	0.930
31	1	4	6	3	3	3	2	776.4	771.1	0.917	0.930
32	1	4	4	3	3	3	2	776.4	771.6	0.917	0.930
33	1	4	6	3	3	3	2	776.4	771.4	0.917	0.930
34	1	4	4	3	3	3	2	776.4	771.5	0.917	0.930
35	1	4	4	3	3	3	4	776.4	771.9	0.917	0.930
36	1	4	4	3	3	3	2	776.4	771.9	0.917	0.930
37	1	4	6	3	3	3	2	776.4	771.5	0.917	0.930
38	1	4	4	3	3	3	2	776.4	771.3	0.917	0.930
39	1	4	4	3	5	3	2	776.4	771.3	0.917	0.930
40	1	4	4	3	3	3	2	776.4	771.1	0.917	0.930

41	1	4	4	3	3	5	2	776.3	772.1	0.917	0.931
42	1	4	4	3	3	3	2	776.3	771.9	0.917	0.931
43	1	4	6	3	3	3	2	776.1	770.5	0.917	0.930
44	1	4	6	3	3	3	2	776.1	770.2	0.917	0.930
45	1	4	6	3	3	3	2	775.9	771.0	0.917	0.930
46	1	4	6	3	3	5	2	775.9	768.9	0.917	0.929
47	1	4	6	3	3	6	2	775.7	769.4	0.917	0.929
48	1	4	4	3	3	4	4	775.5	763.3	0.917	0.928
49	1	4	5	3	3	4	4	775.4	767.0	0.917	0.928
50	1	4	4	3	3	6	2	774.6	759.2	0.917	0.928

Appendix C Sample calculations for evaluating the importance of the input variable to the output according to the contribution of fitness

A. Relative frequency of variables used, r

No	Soil (S)	Fly ash (FA)	Water (W)	Silicate (Si)	Hydroxide (H)	Silicate/hydroxide (Na)	Alkali activator/ash (A)
1	0.0909	0.227	0.182	0.136	0.136	0.136	0.0909
2	0.0870	0.217	0.174	0.130	0.130	0.130	0.130
3	0.0870	0.217	0.174	0.130	0.130	0.174	0.0870

*The individual frequency of variable used is divided by the total of frequency of variables used

e.g. $r_S = fr_S/(fr_S + fr_{FA} + fr_W + fr_{Si} + fr_H + fr_{Na} + fr_A) = 2/(2+5+4+3+3+3+2) = 0.0909$

B. Contribution of fitness according to the relative frequency of variables used in training phase, Ct

No	Soil (S)	Fly ash (FA)	Water (W)	Silicate (Si)	Hydroxide (H)	Silicate/hydroxide (Na)	Alkali activator/ash (A)
1	70.8	177.0	141.6	106.2	106.2	106.2	70.8
2	67.7	169.2	135.4	101.5	101.5	101.5	101.5
3	67.6	169.1	135.3	101.5	101.5	135.3	67.6
Sum (∑)	2447.1	7264.3	7840.8	5428.8	5364.3	6648.9	3833.2

*The individual relative frequency of variable used is multiplied by the fitness in training phase

e.g. $Ct_S = r_S \times Ft = 0.0909 \times 778.9 = 70.8$

C. Contribution of fitness according to the relative frequency of variables used in the GP model in validation phase, Cv

No	Soil (S)	Fly ash (FA)	Water (W)	Silicate (Si)	Hydroxide (H)	Silicate/hydroxide (Na)	Alkali activator/ash (A)
1	70.6	176.4	141.2	105.9	105.9	105.9	70.6
2	67.1	167.7	134.1	100.6	100.6	100.6	100.6
3	67.6	169.0	135.2	101.4	101.4	135.2	67.6
Sum (∑)	2430.1	7213.4	7784.9	5390.4	5326.3	6599.9	3805.2

*The individual relative frequency of variable used is multiplied by the fitness in validation phase

e.g. $Cv_S = r_S \times Ft = 0.0909 \times 776.3 = 70.6$

D. Relative importance of the input variable to the output, RI

Phase	Soil (S)	Fly ash (FA)	Water (W)	Silicate (Si)	Hydroxide (H)	Silicate/hydroxide (Na)	Alkali activator/ash (A)
Training	6.30	18.7	20.2	13.9	13.8	17.1	9.87
Rank	7	2	1	4	5	3	6
Validation	6.30	18.7	20.2	13.9	13.8	17.1	9.87
Rank	7	2	1	4	5	3	6

*Summation of contribution of fitness for individual input variable is divided by the summation of contribution of fitness for all input variable in respective training or validation phase, then multiplies by 100. The relative importance is ranked in descending order.

e.g. $RI_S = [\sum Ct_S/\sum(Ct_S + Ct_{FA} + \ldots + Ct_A)] \times 100 = [2447.1/(2447.1 + 7264.3 + \ldots + 3833.2)] \times 100 = 6.3$

Table 1 Chemical composition of fly ash and residual soil

Elements (%)	Sarawak fly ash	Residual soil
SiO_2	43.8	32.7
Al_2O_3	18.1	25.3
Fe_2O_3	7.7	21.3
CaO	3.9	0.04
MgO	0.5	0.22
MnO	22.8	-
K_2O	2.0	0.03
Na_2O	0.3	0.07
SO_3	0.1	0.01
TiO_2	0.6	-
P_2O_5	0.1	-
LOI	0.5	17.2

Table 2 Mixture ratios and compressive strength of each dataset collected from the experiments

No.	Alkali activator/ ash	NaOH (or KOH)	Na_2SiO_3/ NaOH (or KOH)	Water (%)	Compressive Strength (MPa)
1	0.4	KOH	0.5	10	4.35
2	0.4	KOH	0.5	20	13.9
3	0.4	KOH	0.5	30	12
4	0.4	KOH	0.5	40	4.24
5	0.5	KOH	0.5	10	14
6	0.5	KOH	0.5	20	17.3
7	0.5	KOH	0.5	30	6.72
8	0.5	KOH	0.5	40	2.42
9	0.6	KOH	0.5	10	24.5
10	0.6	KOH	0.5	20	8.64
11	0.6	KOH	0.5	30	2.88
12	0.6	KOH	0.5	40	1.6
13	0.7	KOH	0.5	10	15.1
14	0.7	KOH	0.5	20	2.96
15	0.7	KOH	0.5	30	1.2
16	0.7	KOH	0.5	40	0.88
17	0.4	KOH	1	10	5.42
18	0.4	KOH	1	20	11.7
19	0.4	KOH	1	30	10.4
20	0.4	KOH	1	40	5.24
21	0.5	KOH	1	10	22.7
22	0.5	KOH	1	20	14.5
23	0.5	KOH	1	30	9.7
24	0.5	KOH	1	40	1.79
25	0.6	KOH	1	10	23.1
26	0.6	KOH	1	20	12.9
27	0.6	KOH	1	30	2.44
28	0.6	KOH	1	40	2.06

29	0.7	KOH	1	10	19.5
30	0.7	KOH	1	20	5.9
31	0.7	KOH	1	30	2.21
32	0.7	KOH	1	40	1.63
33	0.4	NaOH	0.5	10	2.09
34	0.4	NaOH	0.5	20	15.6
35	0.4	NaOH	0.5	30	12.1
36	0.4	NaOH	0.5	40	3.17
37	0.5	NaOH	0.5	10	22.7
38	0.5	NaOH	0.5	20	19.7
39	0.5	NaOH	0.5	30	5.09
40	0.5	NaOH	0.5	40	1.87
41	0.6	NaOH	0.5	10	32.1
42	0.6	NaOH	0.5	20	13.2
43	0.6	NaOH	0.5	30	3.36
44	0.6	NaOH	0.5	40	2
45	0.7	NaOH	0.5	10	14
46	0.7	NaOH	0.5	20	5.33
47	0.7	NaOH	0.5	30	1.59
48	0.7	NaOH	0.5	40	1.73
49	0.4	NaOH	1	10	4.88
50	0.4	NaOH	1	20	12.6
51	0.4	NaOH	1	30	15.5
52	0.4	NaOH	1	40	4.62
53	0.5	NaOH	1	10	22.6
54	0.5	NaOH	1	20	12.9
55	0.5	NaOH	1	30	7.46
56	0.5	NaOH	1	40	1.98
57	0.6	NaOH	1	10	26.3
58	0.6	NaOH	1	20	17.4
59	0.6	NaOH	1	30	3.29
60	0.6	NaOH	1	40	1.31

61	0.7	NaOH	1	10	24.7
62	0.7	NaOH	1	20	7.29
63	0.7	NaOH	1	30	3.75
64	0.7	NaOH	1	40	0.98

Table 3 Statistical values of input variables and output used in the predictive modelling

Statistical parameter	Soil (wt%)	Fly ash (wt%)	Water (wt%)	Silicate (wt%)	Hydroxide (wt%)	Silicate/ hydroxide (weight ratio)	Alkali activator/ash (weight ratio)	Compressive strength (MPa)
Minimum	36.2	29.0	4.07	4.23	6.35	0.5	0.4	0.876
Maximum	45.0	36.0	15.9	11.4	13.4	1.0	0.7	32.0
Range	8.81	7.05	11.8	7.15	7.10	0.5	0.3	31.2
Average	40.3	32.2	9.87	7.78	9.82	0.8	0.6	9.48
Standard deviation	2.29	1.83	4.08	2.27	1.81	0.3	0.1	8.00
Sample variance	5.19	3.32	16.9	5.05	3.32	0.1	0.0	64.7
Median	40.00	32.0	11.3	7.69	10.1	1.0	0.6	5.90
Skewness	0.251	0.251	-0.0718	0.0498	-0.185	-0.1	-0.1	0.844
Kurtosis	-0.479	-0.479	-1.32	-1.35	-0.615	-2.0	-1.3	-0.283

Table 4 Connection weights amongst input-hidden-output layer

Hidden layer	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
Input layer										
Soil	4.266	-0.0883	-0.0764	0.808	1.51	-2.41	-0.908	3.09	1.75	2.31
Fly ash	0.0634	0.170	4.38	-1.06	2.02	-2.35	-0.277	5.26	1.06	2.21
Water	-0.884	-1.829	-1.04	6.39	0.611	1.01	-0.0219	-2.69	3.64	-2.11
Silicate	0.460	-0.00160	-0.655	-1.93	-1.94	3.07	3.29	-0.499	0.903	-1.54
Hydroxide	-1.27	1.50	1.42	-3.60	-2.41	-0.334	-1.70	-2.486	-1.82	0.934
Silicate/hydroxide	-0.954	0.0807	0.818	2.47	-2.11	3.79	5.02	2.36	1.32	0.688
Alkali activator/ash	-0.237	2.17	-0.387	-2.14	0.649	3.35	0.962	-5.07	-4.17	-0.930
Output layer										
Compressive strength	-2.69	2.56	2.47	-0.285	-1.05	-0.815	-0.435	1.09	1.17	-1.62

Table 5 The values of R^2 and errors in training and validation phases of the GP model

Phase	R^2	***MAE***	***RMSE***	***RAE***	***RRSE***
Training	0.920	1.513	2.01	0.243	0.284
Validating	0.935	1.754	2.73	0.218	0.288

Table 6 Correlation coefficient of the GP model in training and validating phases

Input variable	**Correlation coefficient, r**	
	Training phase	**Validation phase**
Soil	0.717	0.464
Fly ash	0.717	0.461
Water	-0.713	-0.688
Silicate	-7.78×10^{-2}	0.159
Hydroxide	0.180	0.219
Silicate/hydroxide	-2.08×10^{-2}	0.201
Alkali activator/ash	-0.149	3.21

Table 7 Summary of importance of the input variable to the output in the ANN and GP models

Model		**ANN**	**GP**	
R^2 obtained in the model		0.9535	0.9301	
Input variable	**Garson's algorithm**	**Connection weights approach**	**Model analysis**	**Fitness method**
Soil	6	2 (negative relationship)	6-7	7
Fly ash	4	1 (positive relationship)	1	2
Water	1	6 (negative relationship)	2	1
Silicate	7	7 (negative relationship)	3-5	4
Hydroxide	3	3 (positive relationship)	3-5	5
Silicate/hydroxide	5	5 (positive relationship)	3-5	3
Alkali activator/ash	2	4 (negative relationship)	6-7	6

Table 8 Statistical values of the input variables

Input variable	Covariance	Correlation Coefficient, r	Standard error of t, Sr	Test statistic, t	p-value	Significance of input variable to the output
Soil	10.9	0.591	0.102	5.77	2.68×10^{-7}	Significant (positive relationship)
Fly ash	8.71	0.591	0.102	5.77	2.68×10^{-7}	Significant (positive relationship)
Water	-22.6	-0.701	0.0906	-7.74	1.13×10^{-7}	Significant (negative relationship)
Silicate	0.295	0.0165	0.127	0.130	0.897	Not Significant
Hydroxide	2.67	0.190	0.125	1.52	0.133	Not Significant
Silicate/ hydroxide	0.119	0.0604	0.127	0.477	0.635	Not Significant
Alkali activator/ash	-0.0730	-0.0827	0.127	-0.654	0.516	Not Significant

Notes: The input variable shows the significant effect on the output when p<0.05; it is insignificant when p>0.05 (null hypothesis).

Fig 1a

Click here to download Figure FIG. 1a.tif

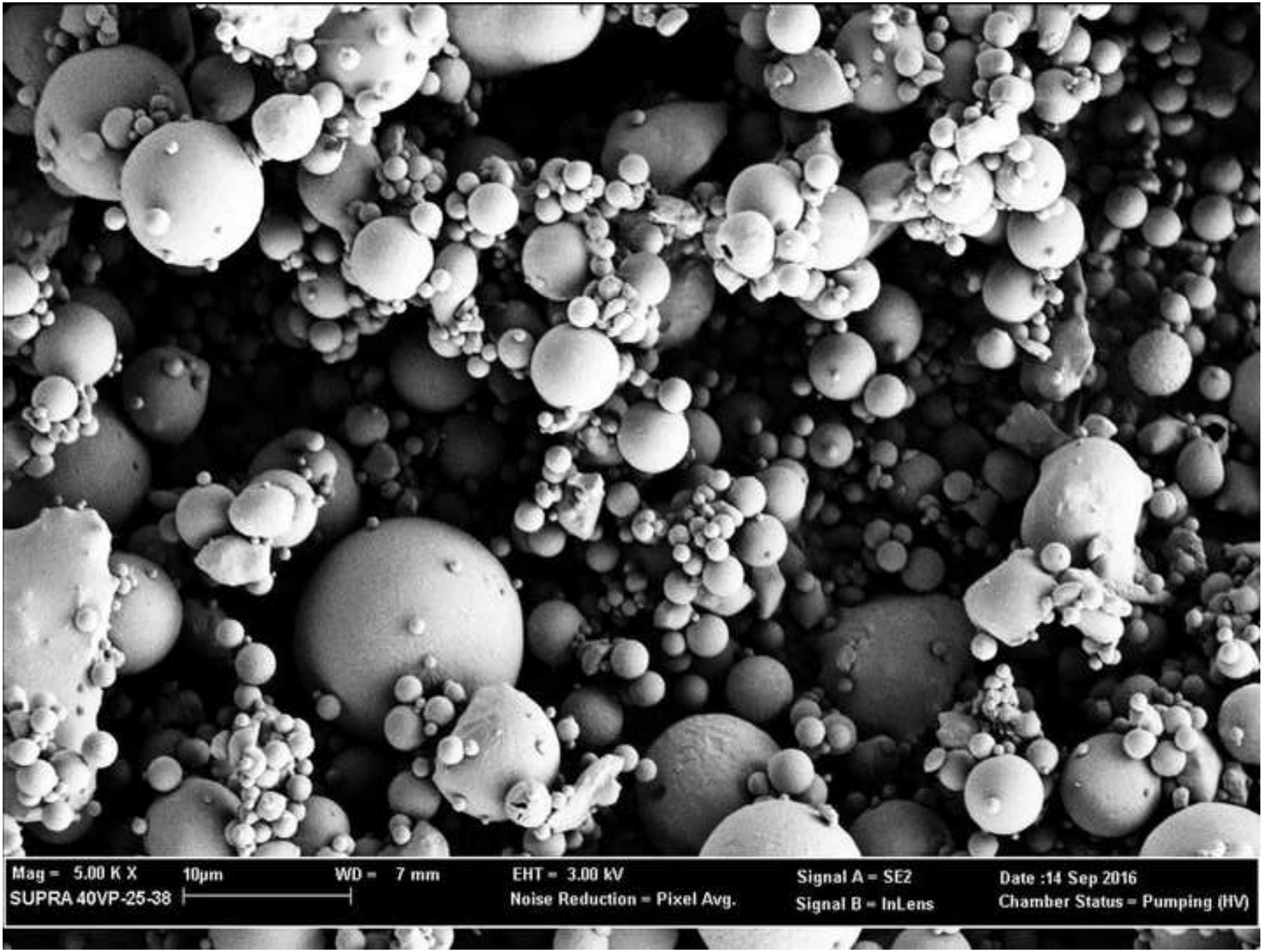


Fig 1b

Click here to download Figure FIG. 1b.tif

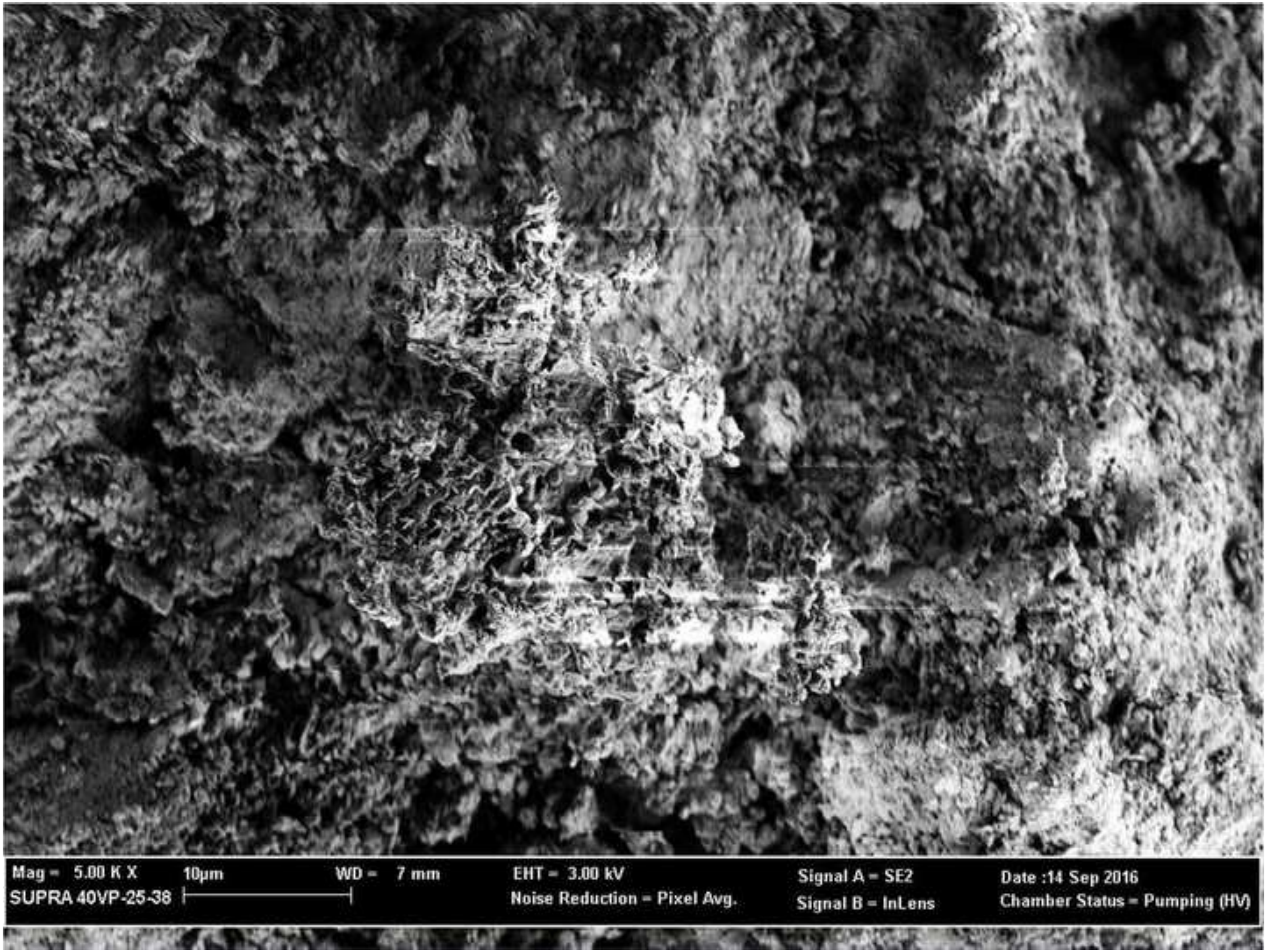


Fig 2 Click here to download Figure FIG. 2.tif

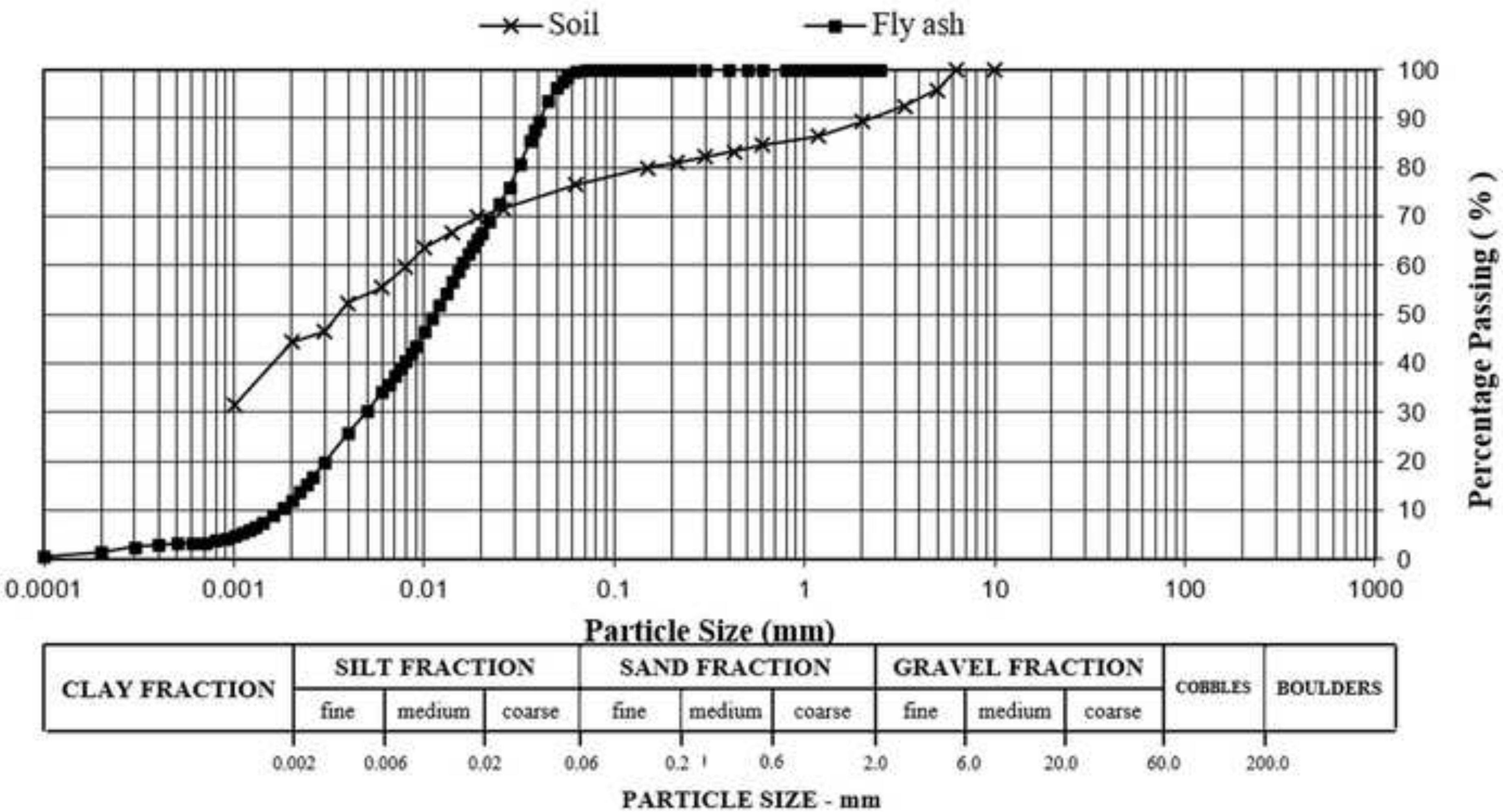
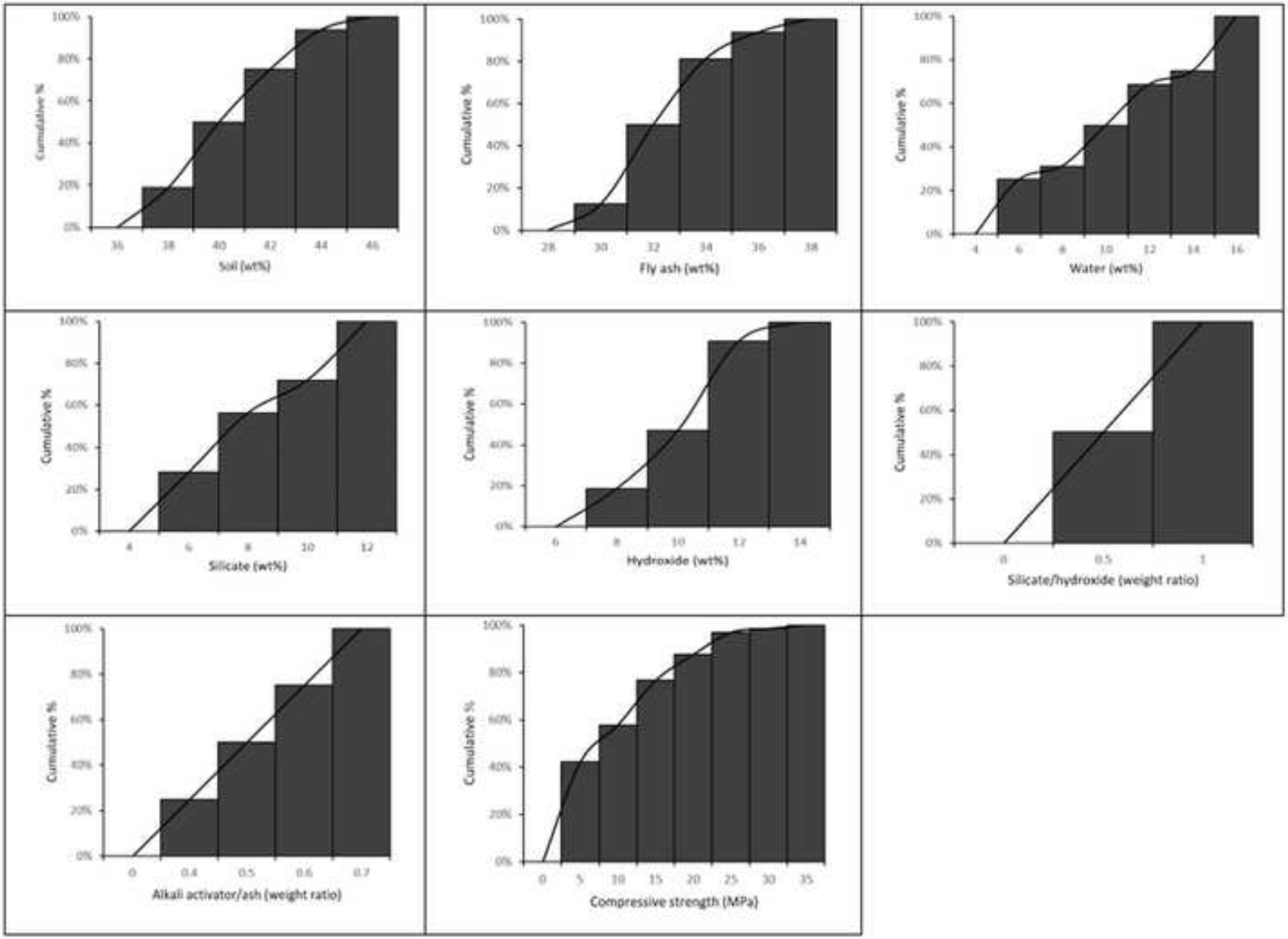


Fig 3

Click here to download Figure FIG. 3.tif



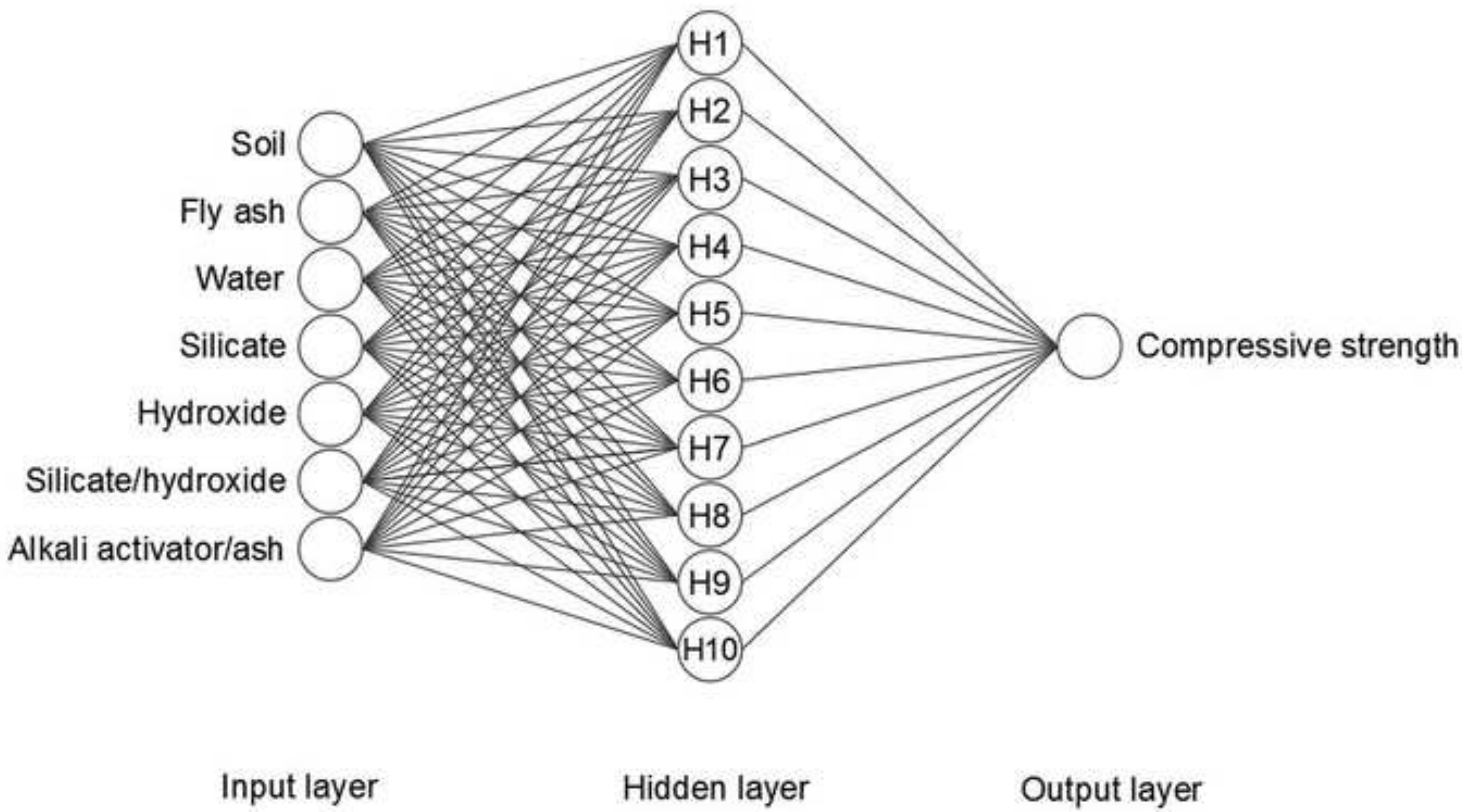
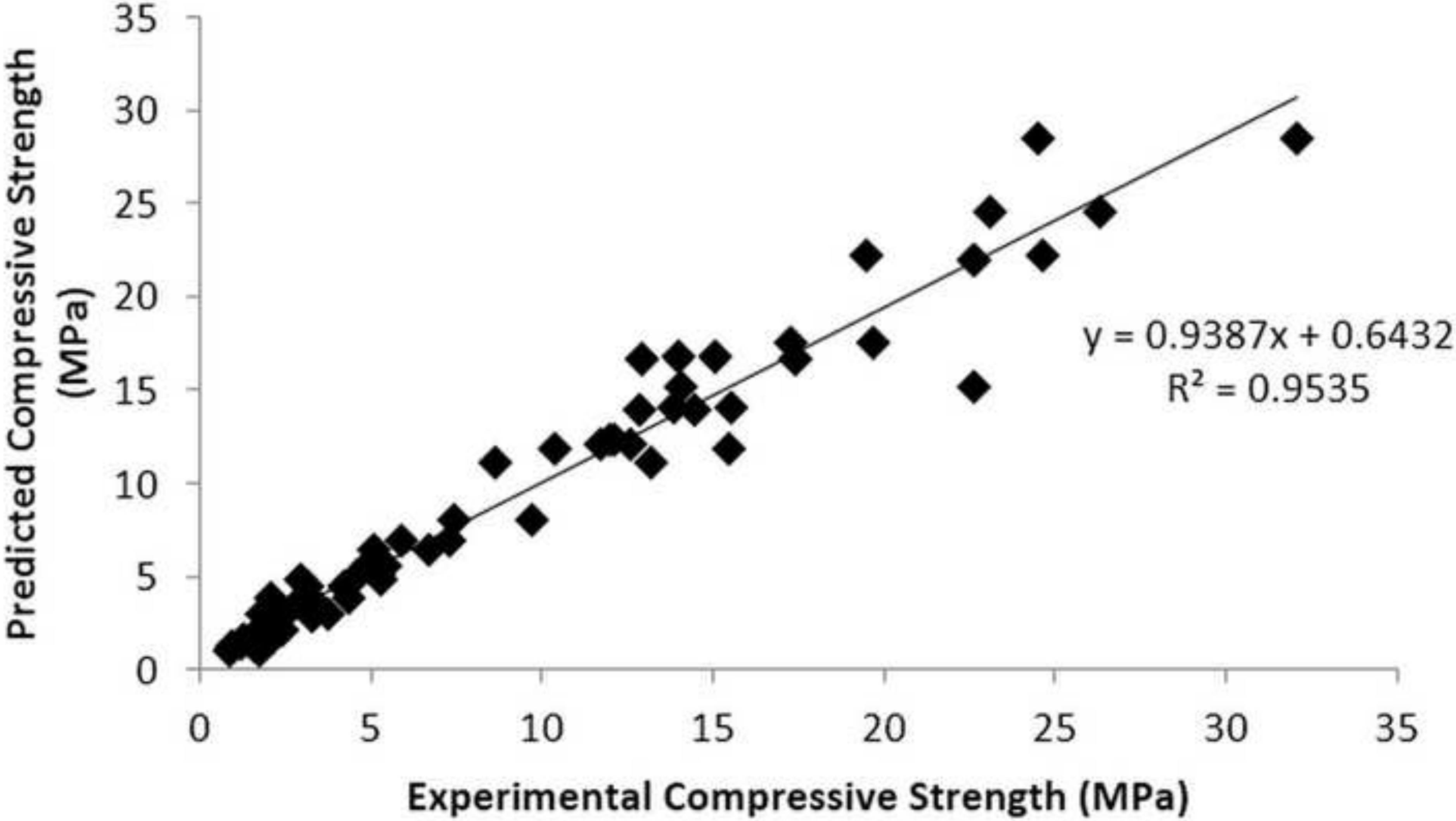
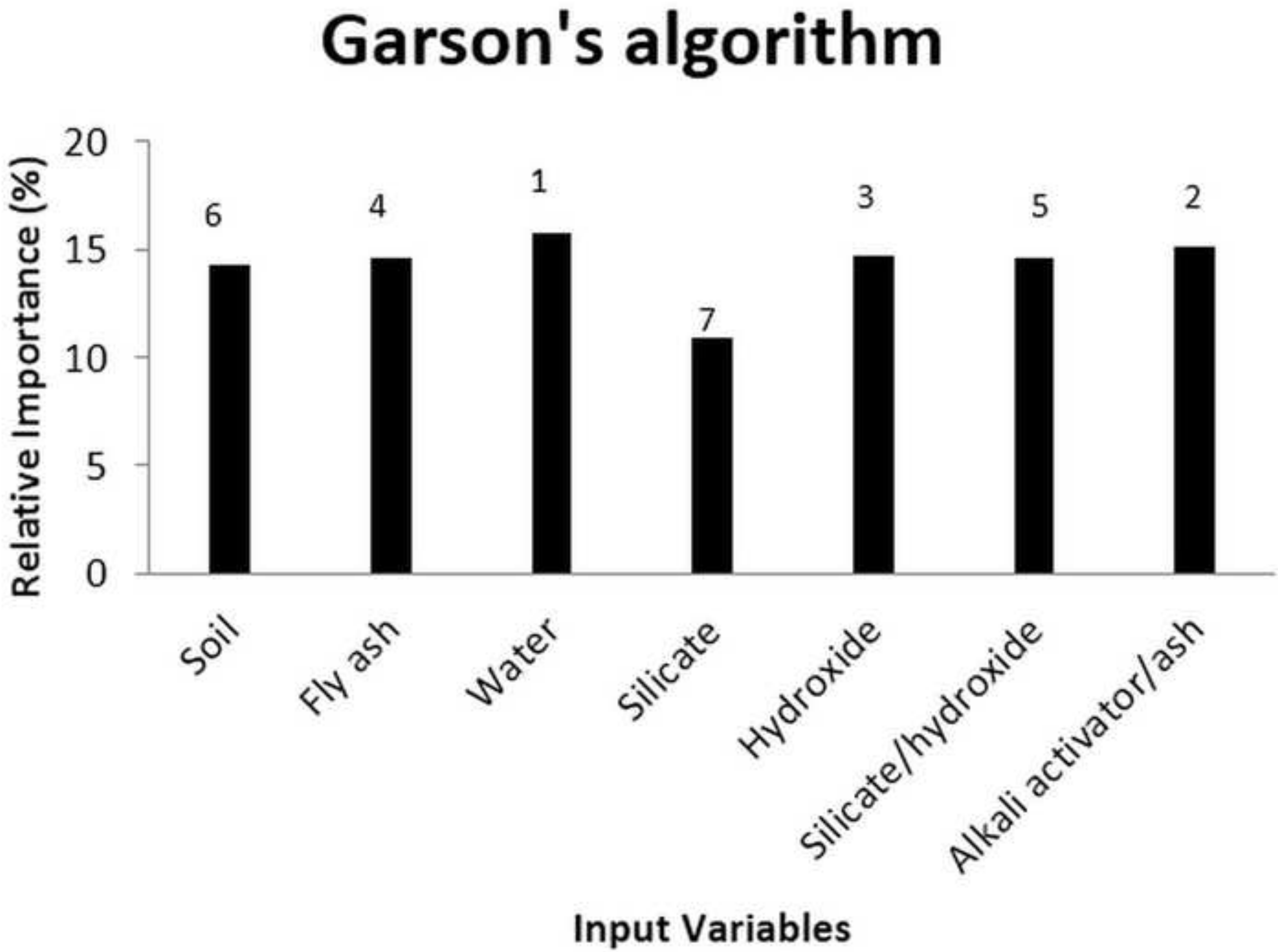

Soil
Fly ash
Water
Silicate
Hydroxide
Silicate/hydroxide
Alkali activator/ash
H1
H2
H3
H4
H5
H6
H7
H8
H9
H10
Compressive strength
Input layer
Hidden layer
Output layer

Fig 5

Click here to download Figure FIG. 5.tif




Garson's algorithm
Relative Importance (%)
20
15
10
5
0
6
4
1
7
3
5
2
Soil
Fly ash
Water
Silicate
Hydroxide
Silicate/hydroxide
Alkali activator/ash
Input Variables

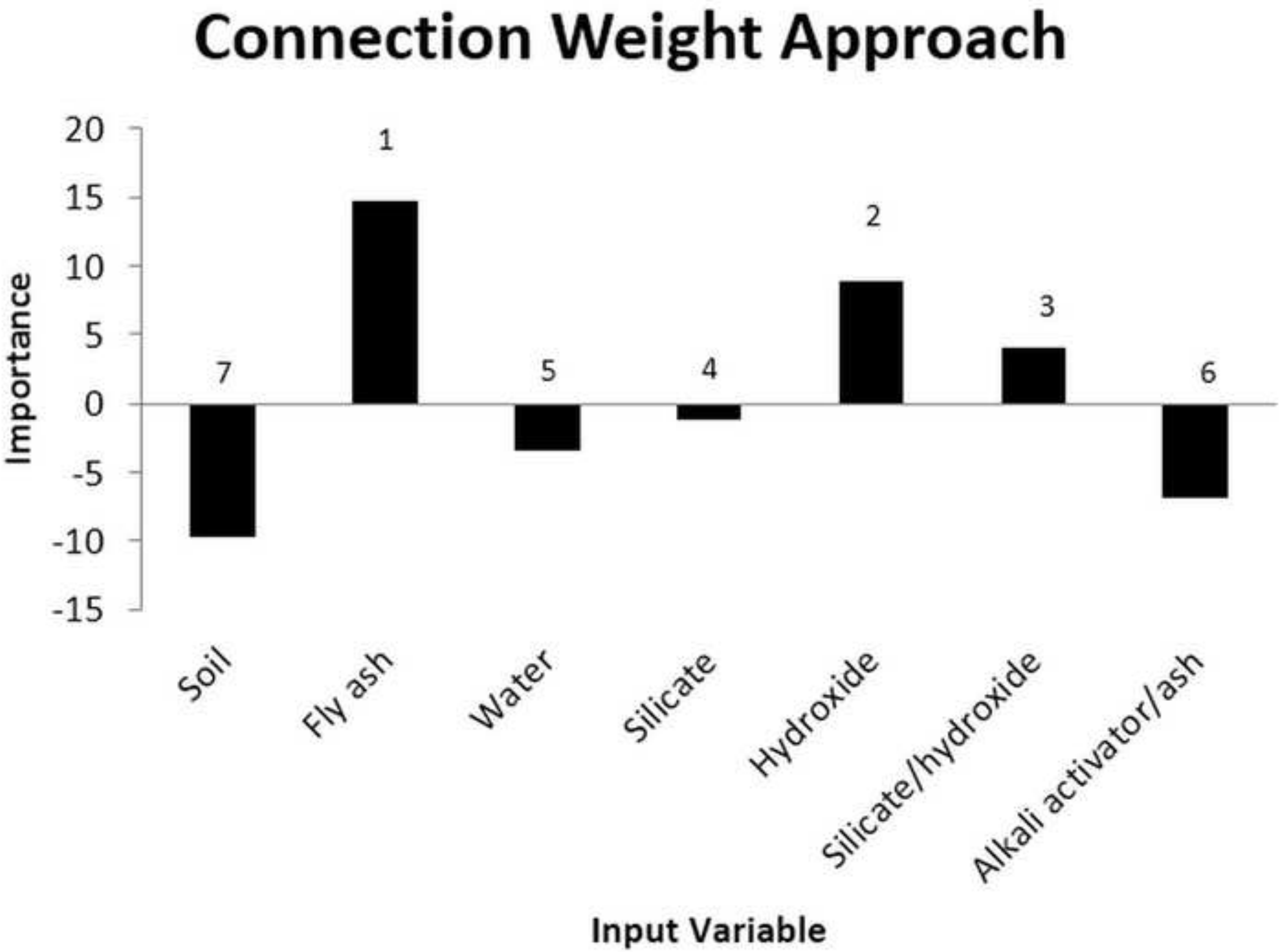
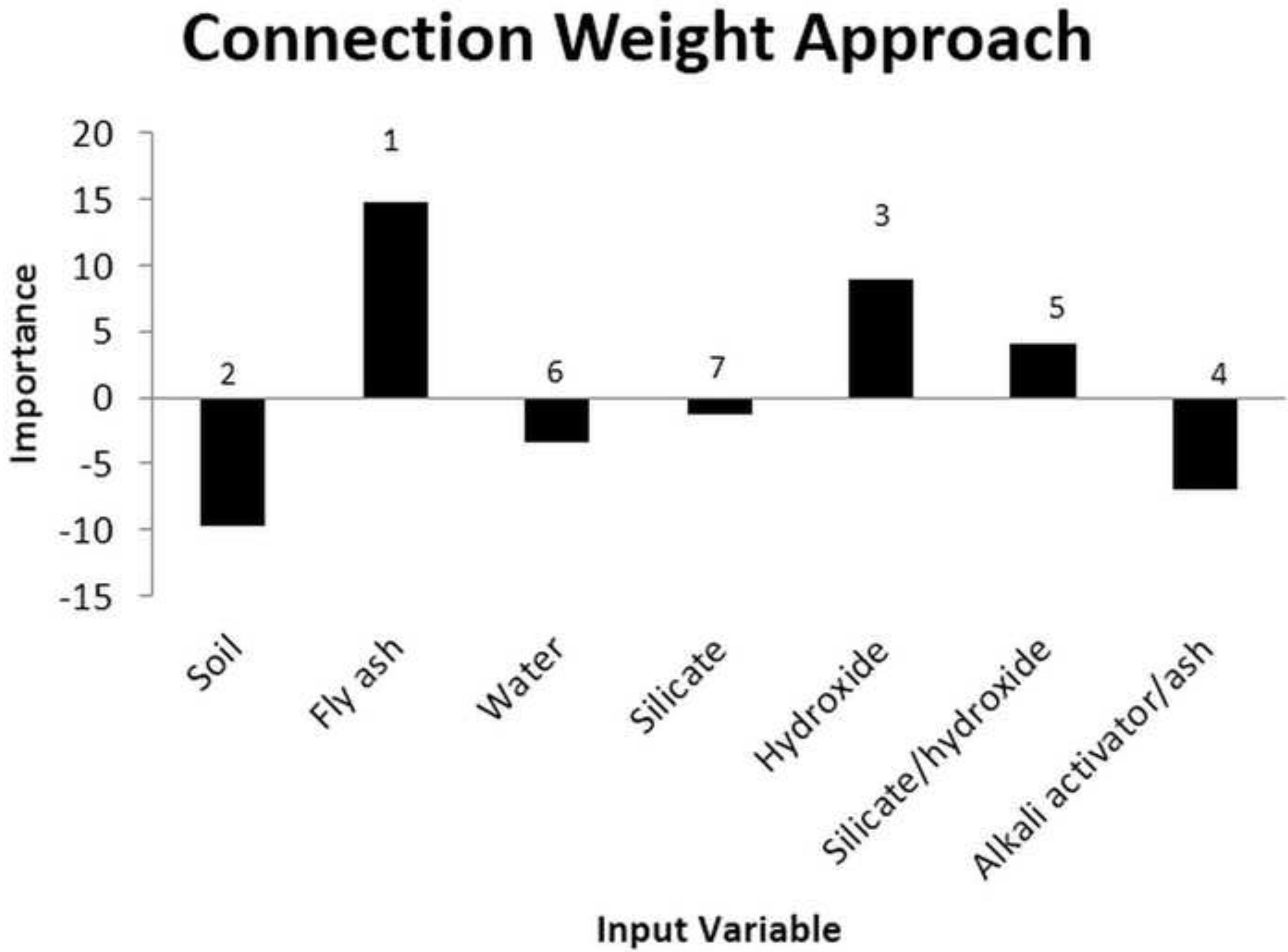
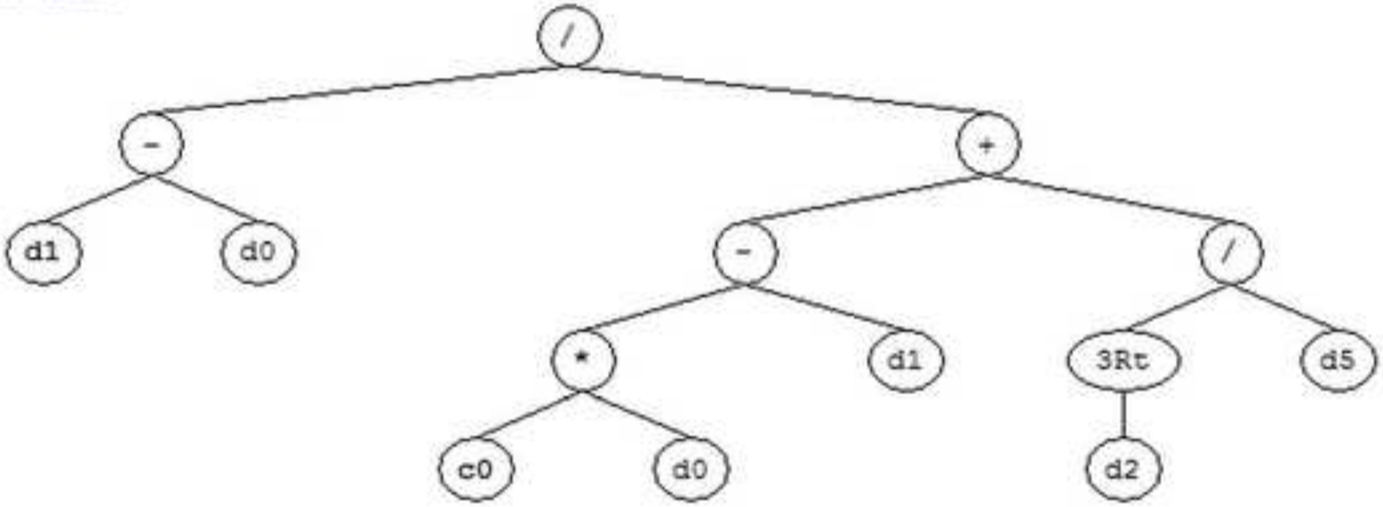

Connection Weight Approach
Importance
20
15
10
5
0
-5
-10
-15
7
1
5
4
2
3
6
Soil
Fly ash
Water
Silicate
Hydroxide
Silicate/hydroxide
Alkali activator/ash
Input Variable

Fig 7b

Click here to download Figure FIG. 7b.tif



Sub-ET 1



Sub-ET 2

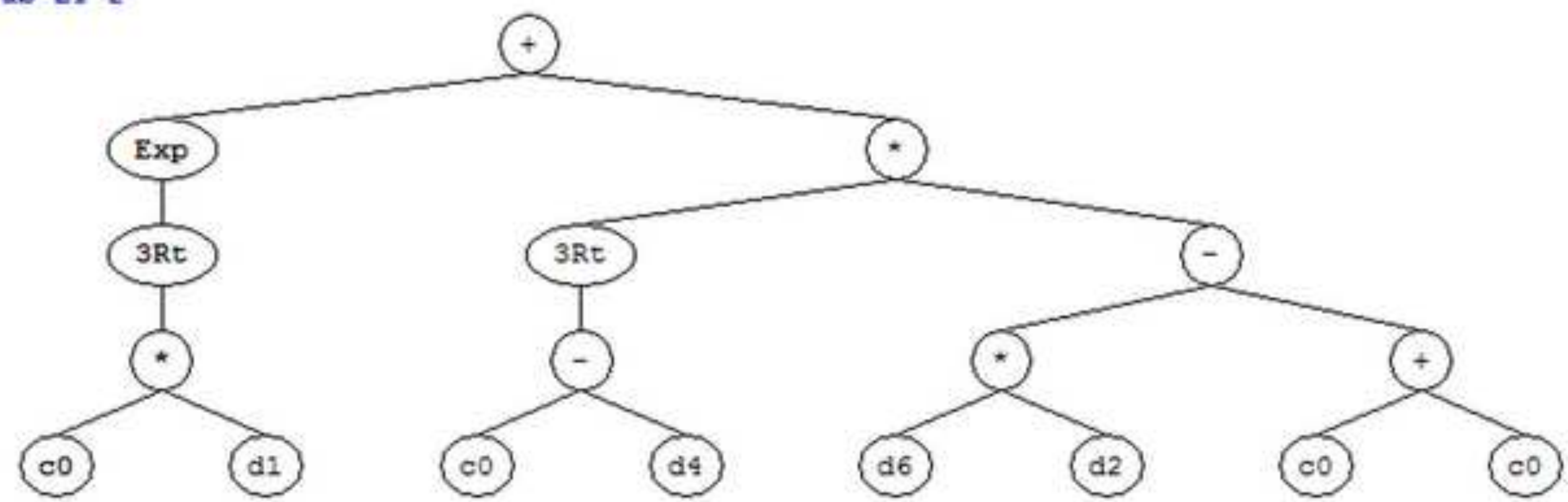
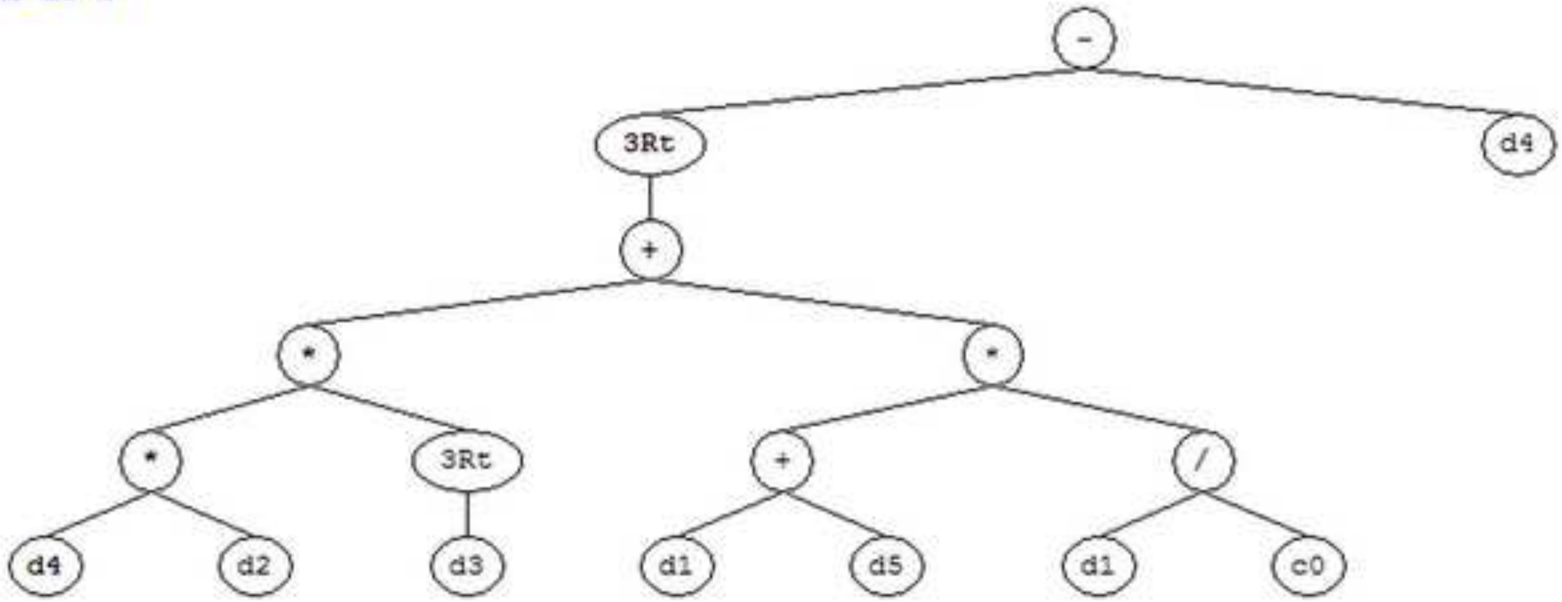


Fig 8ii

Sub-ET 3



Sub-ET 4

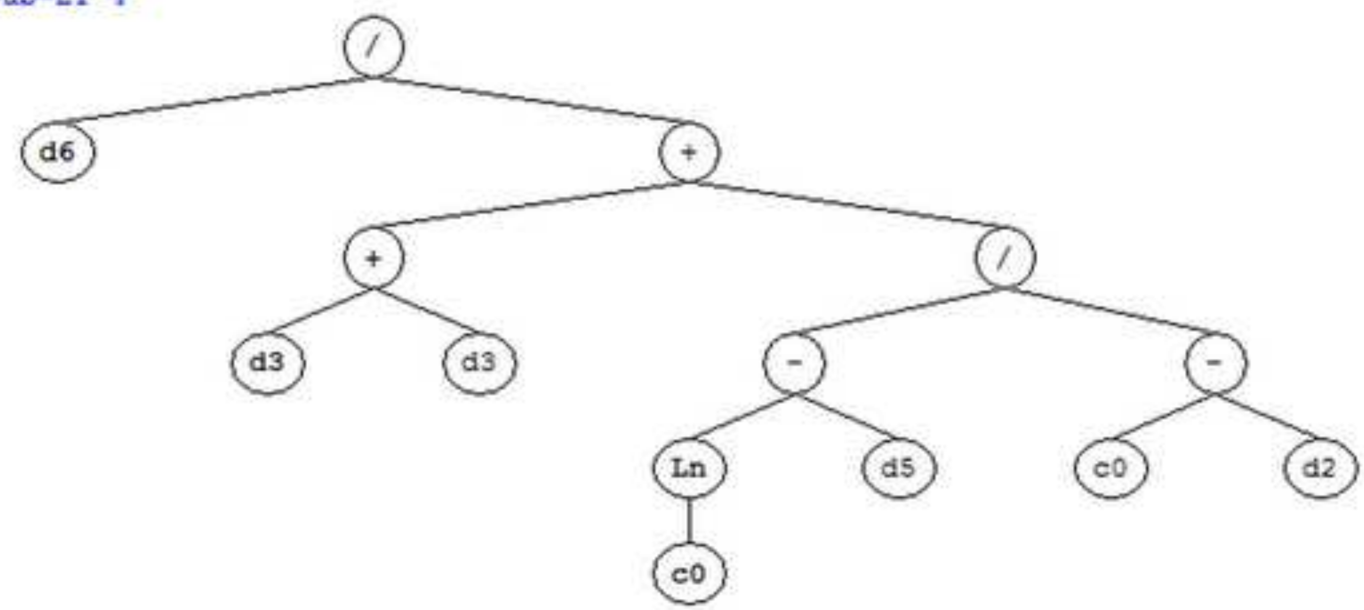


Fig 9a

Click here to download Figure FIG. 9a.tif

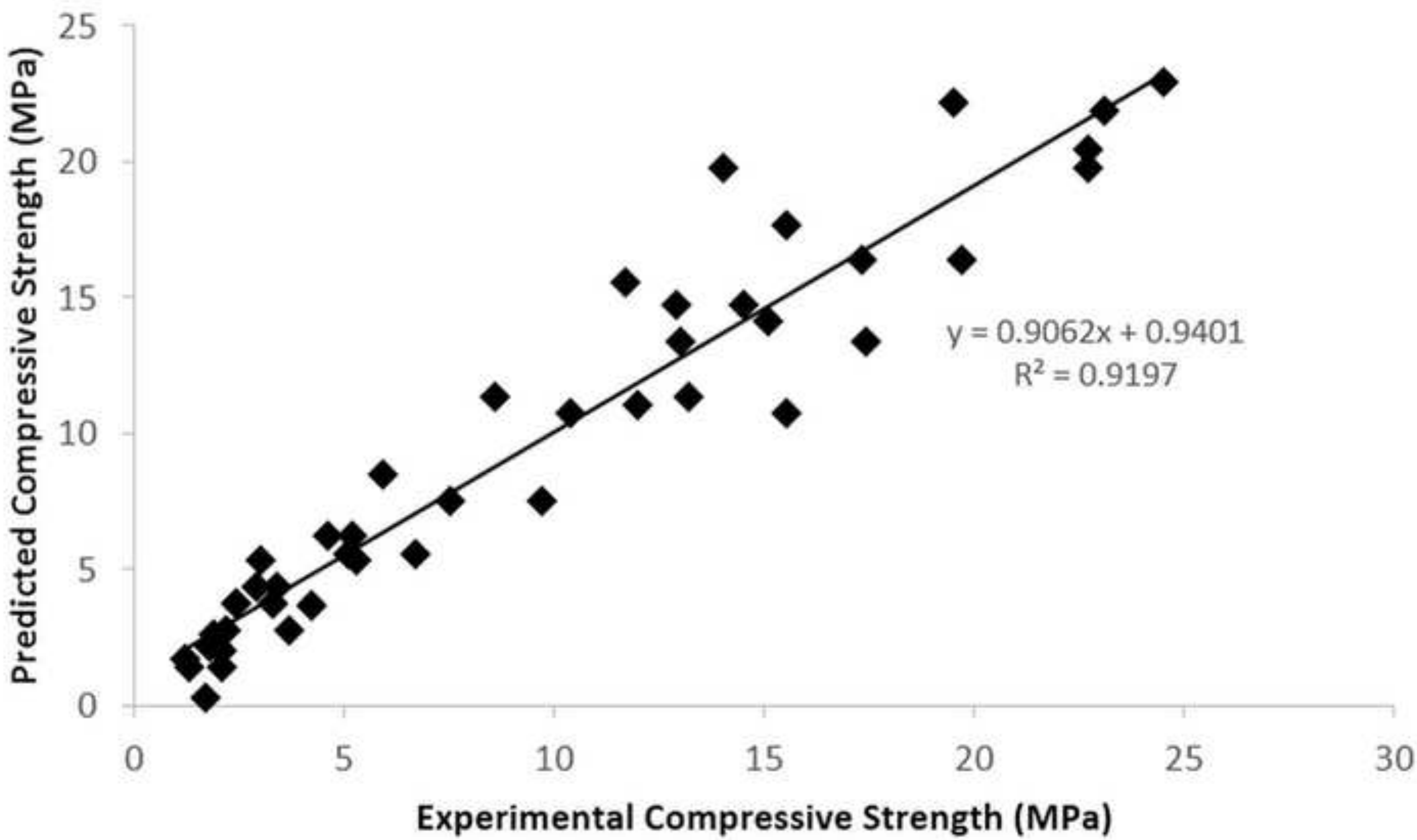
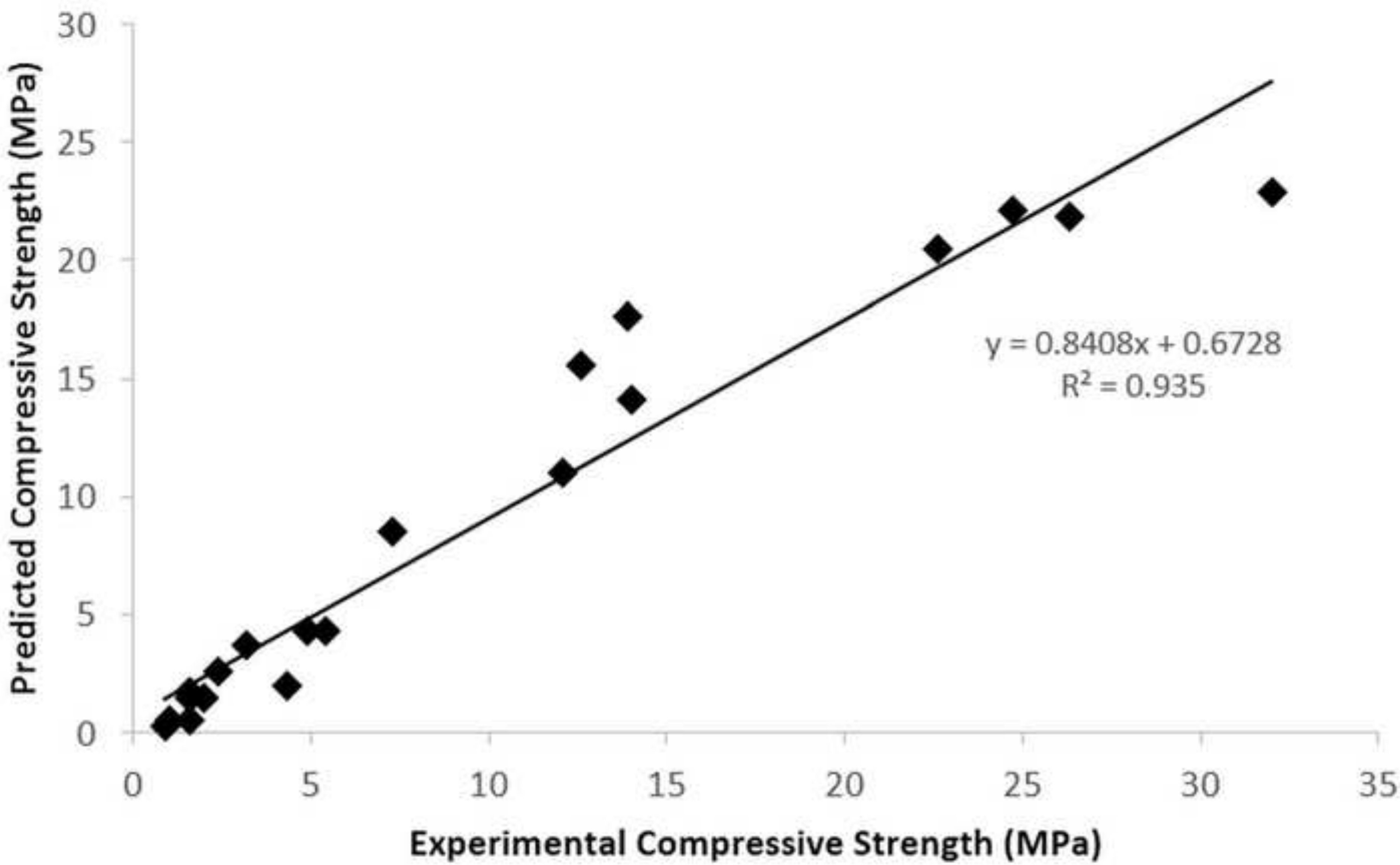
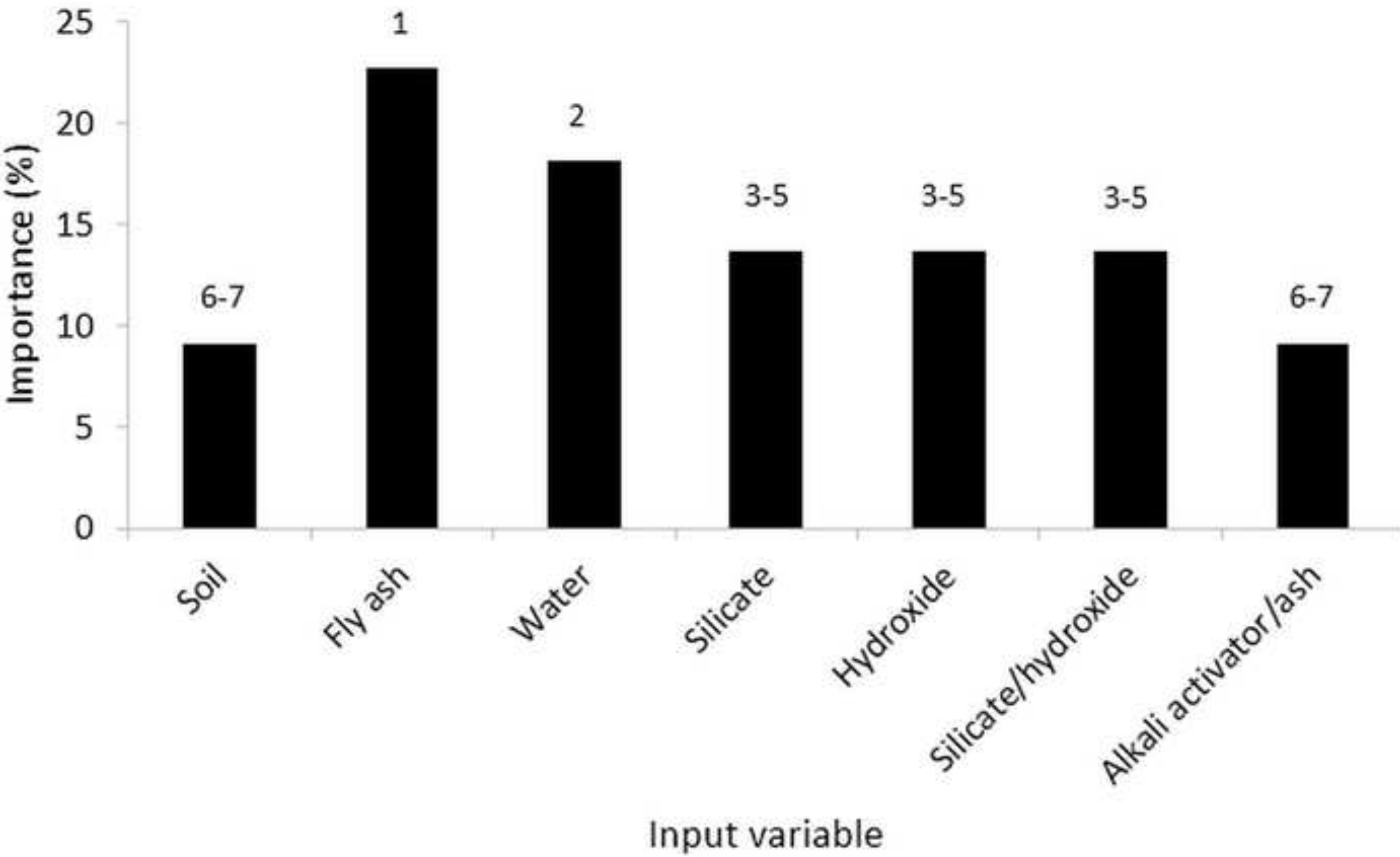
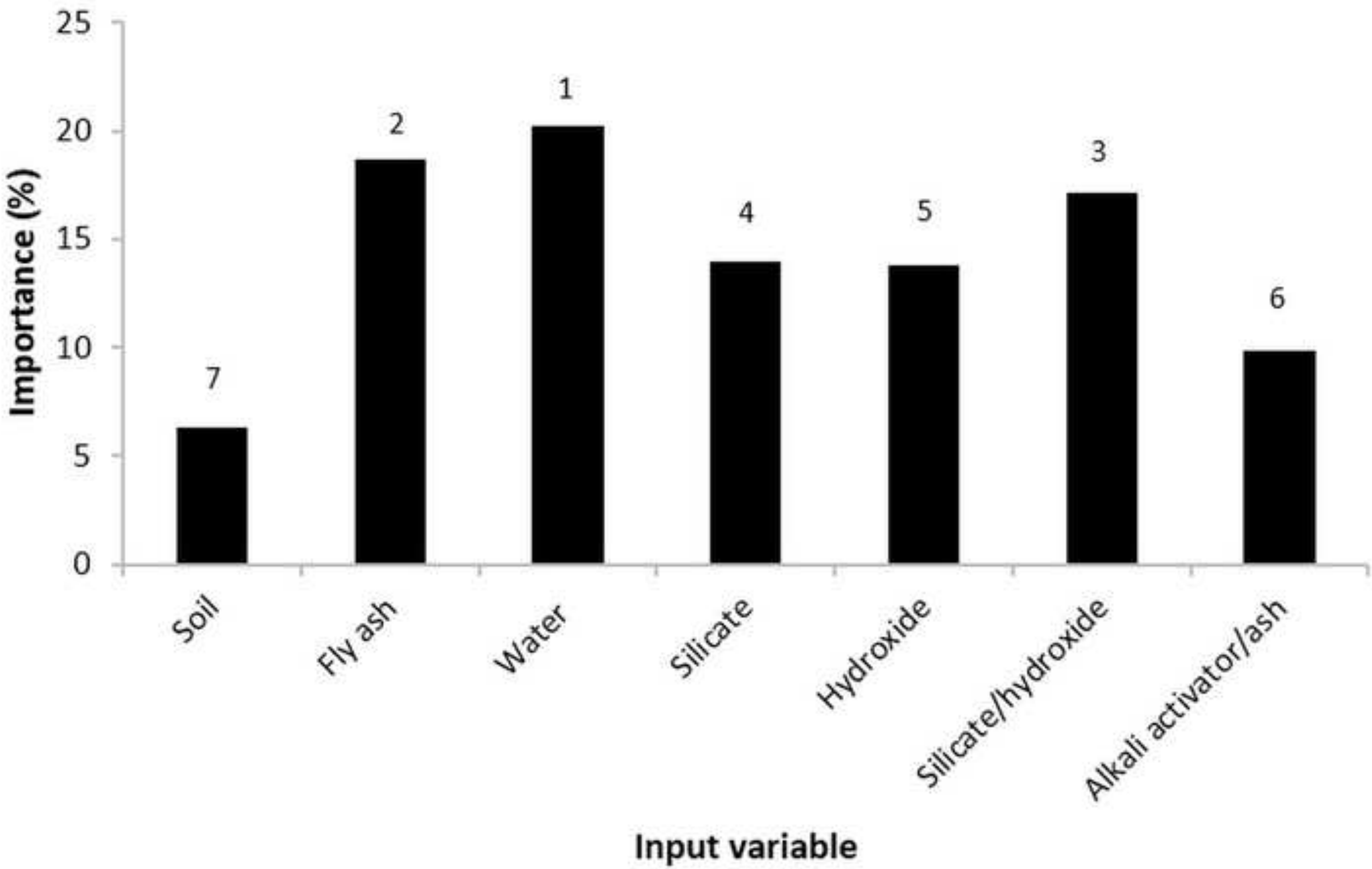


Fig 9b

Click here to download Figure FIG. 9b.tif




25
20
15
10
5
0
Importance (%)
1
2
3-5
3-5
3-5
6-7
6-7
Soil
Fly ash
Water
Silicate
Hydroxide
Silicate/hydroxide
Alkali activator/ash
Input variable


Importance (%)
25
20
15
10
5
0
7
2
1
4
5
3
6
Soil
Fly ash
Water
Silicate
Hydroxide
Silicate/hydroxide
Alkali activator/ash
Input variable

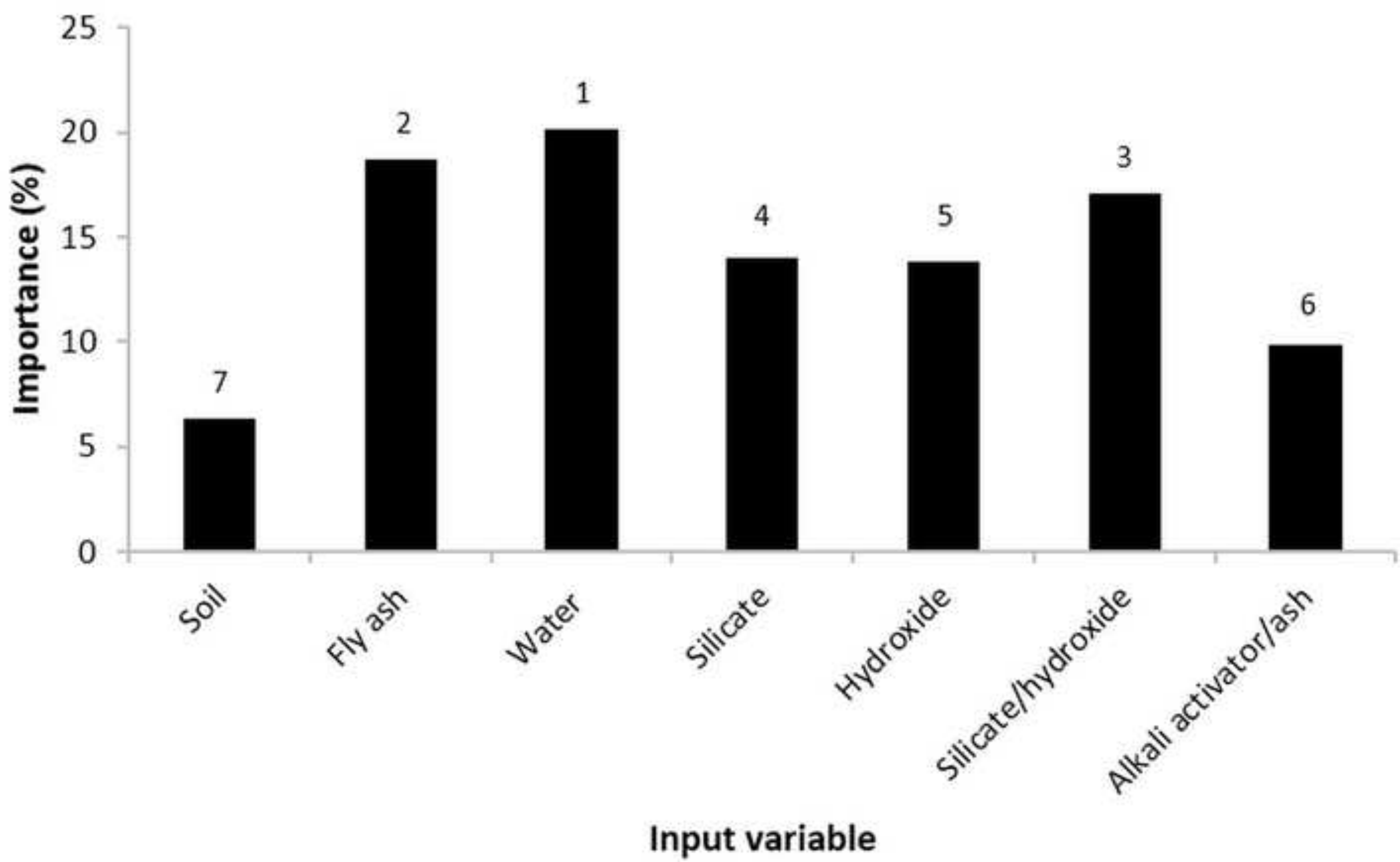
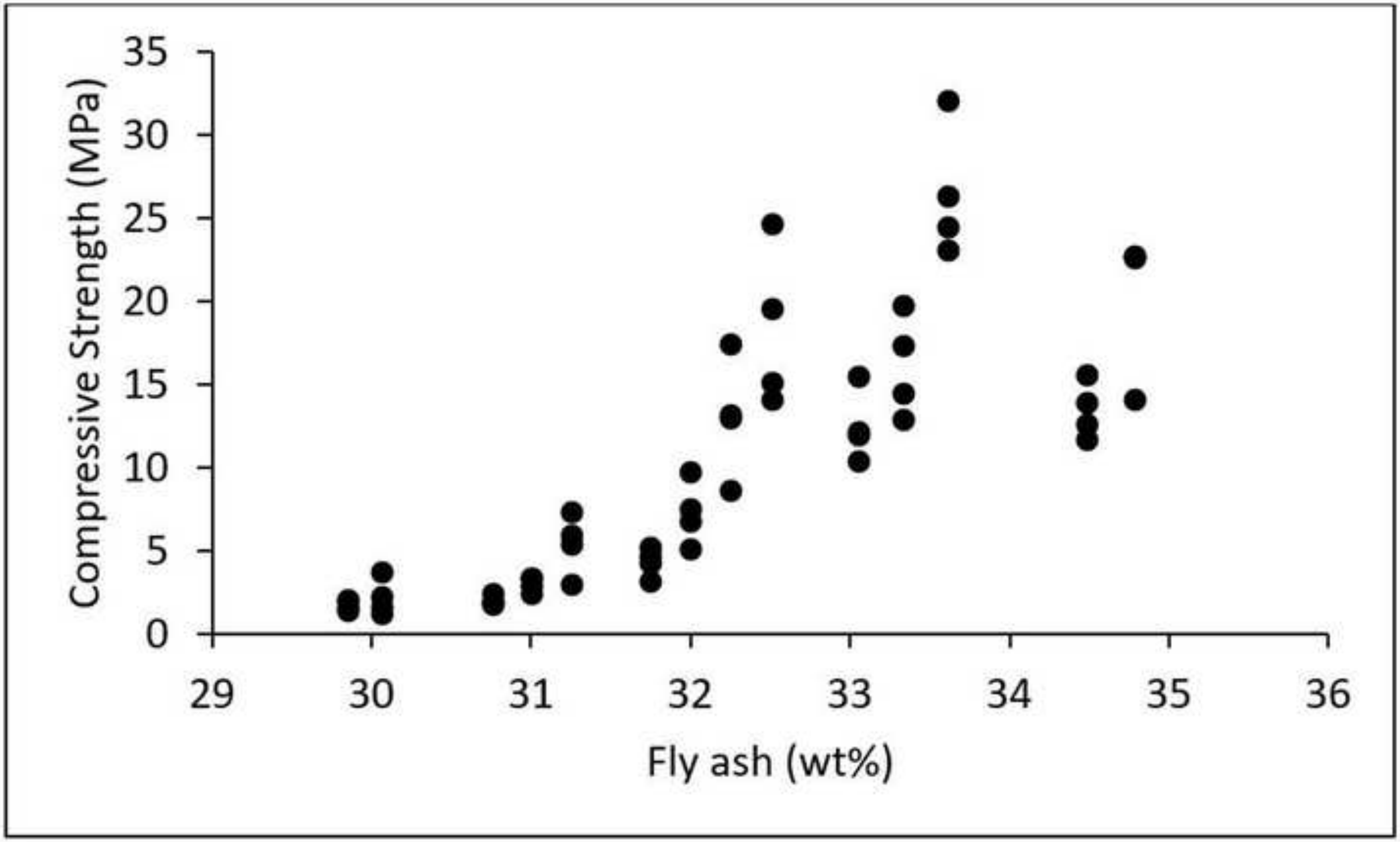
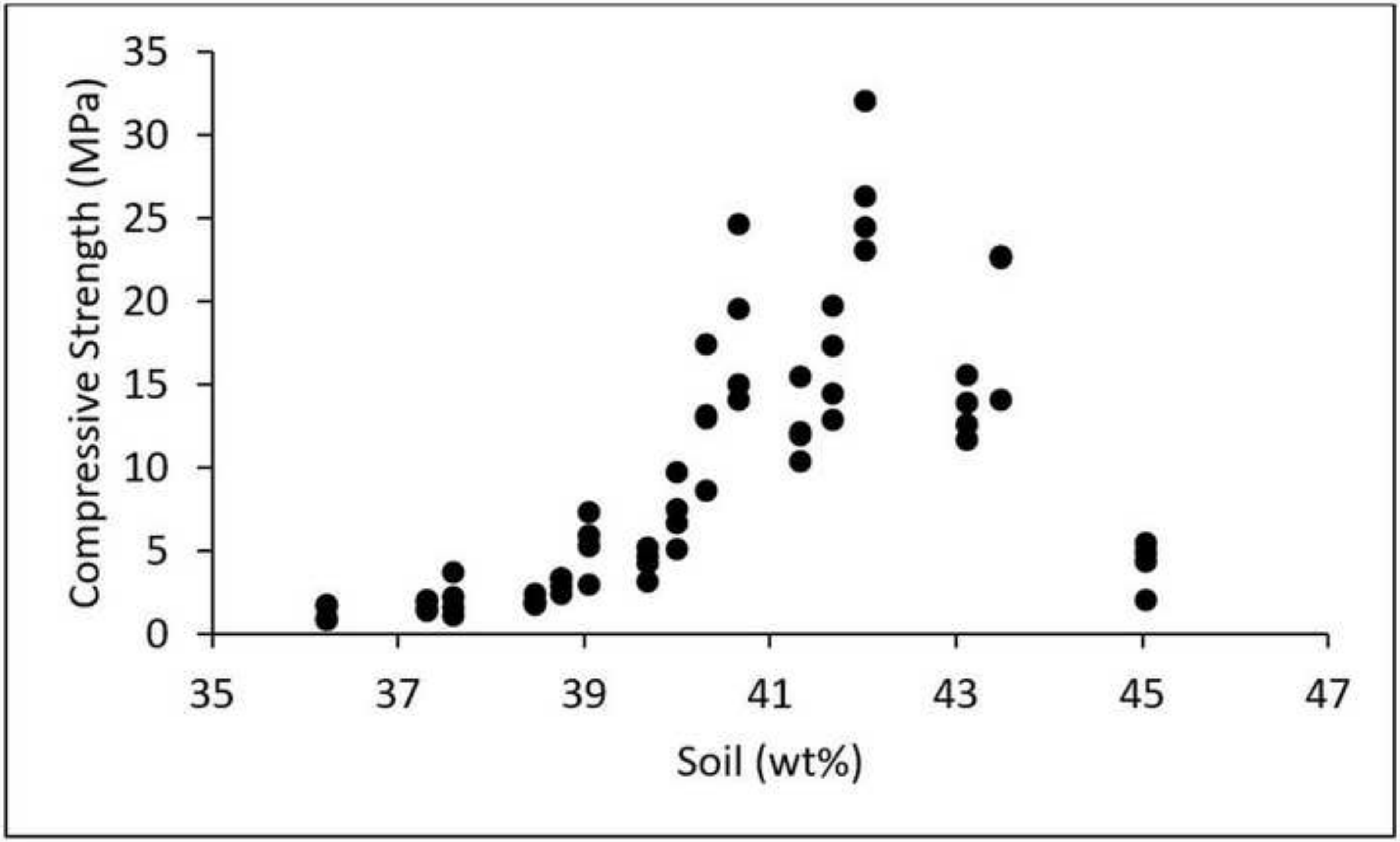

Importance (%)
25
20
15
10
5
0
7
2
1
4
5
3
6
Soil
Fly ash
Water
Silicate
Hydroxide
Silicate/hydroxide
Alkali activator/ash
Input variable

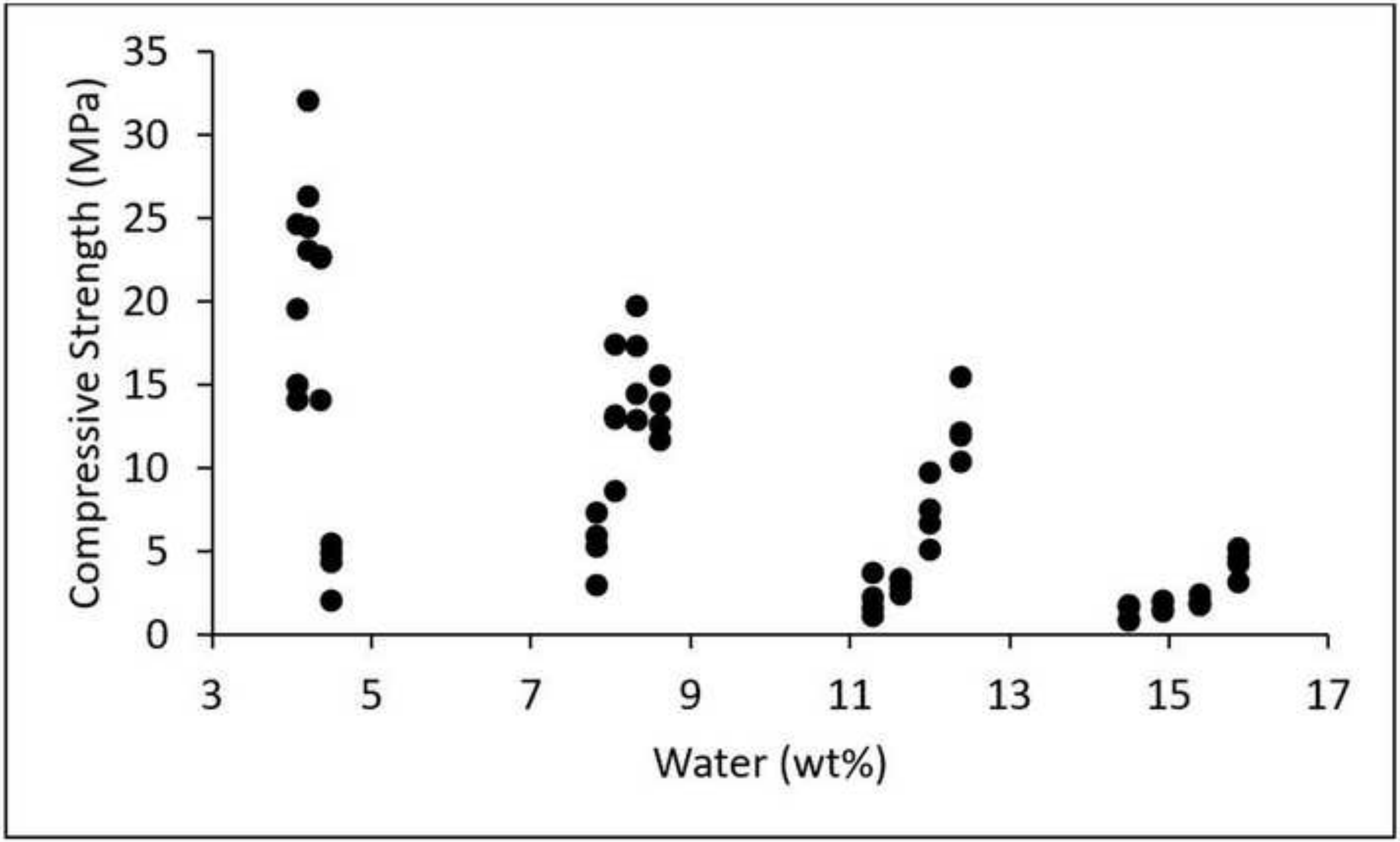
Fig 12a

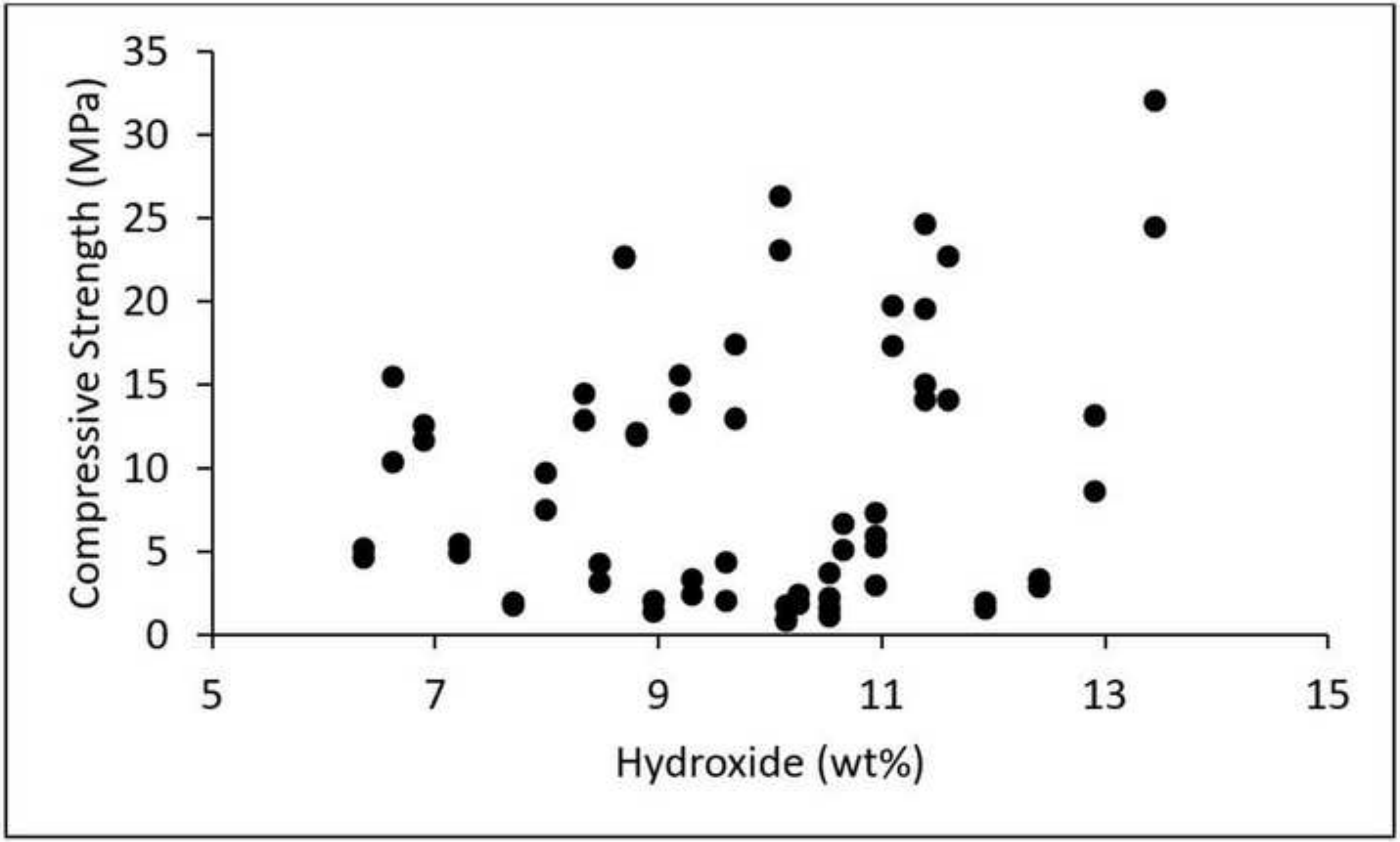


Compressive Strength (MPa)
35
30
25
20
15
10
5
0
29
30
31
32
33
34
35
36
Fly ash (wt%)



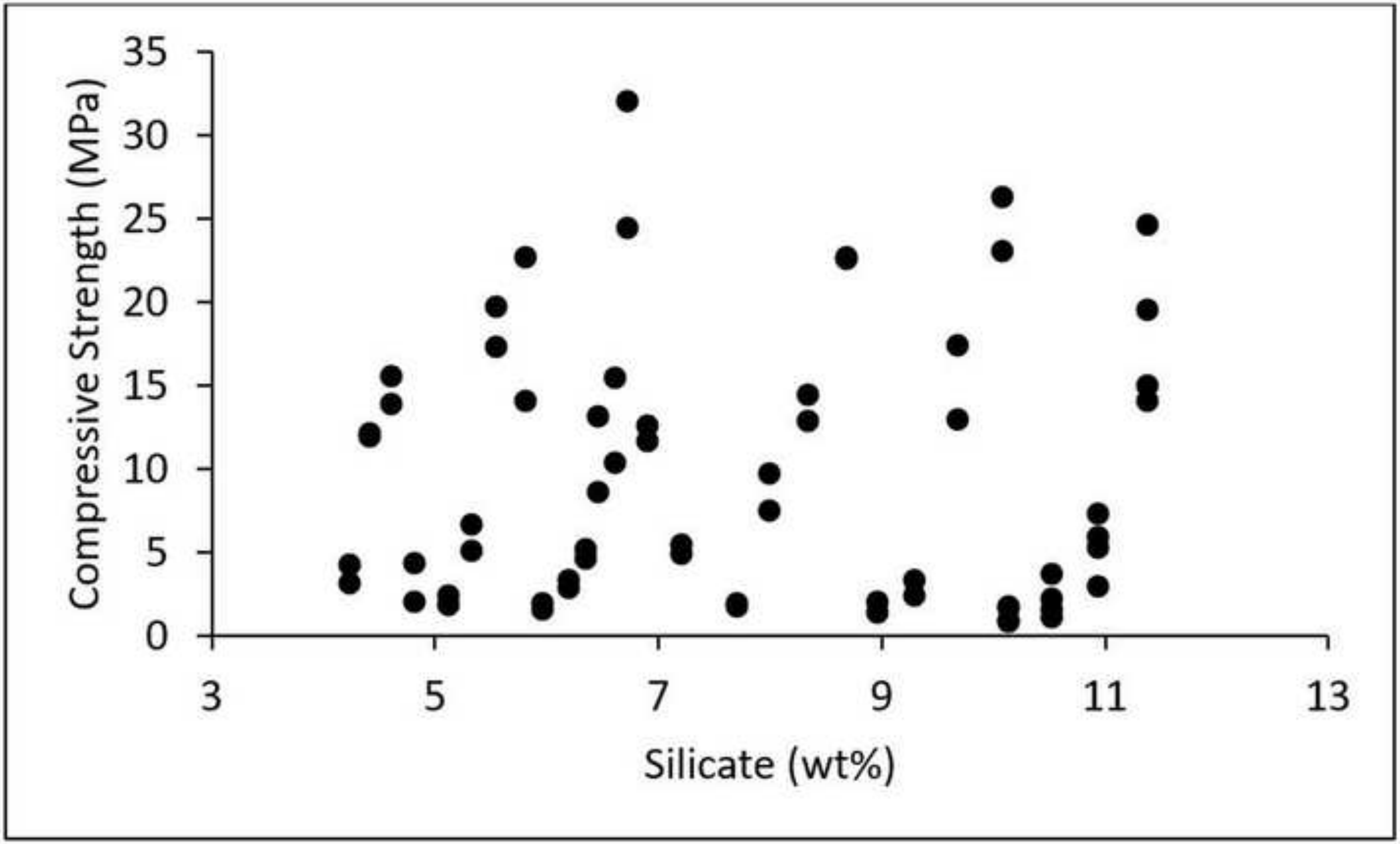
Compressive Strength (MPa)
35
30
25
20
15
10
5
0
35
37
39
41
43
45
47
Soil (wt%)


Compressive Strength (MPa)
35
30
25
20
15
10
5
0
3
5
7
9
11
13
15
17
Water (wt%)



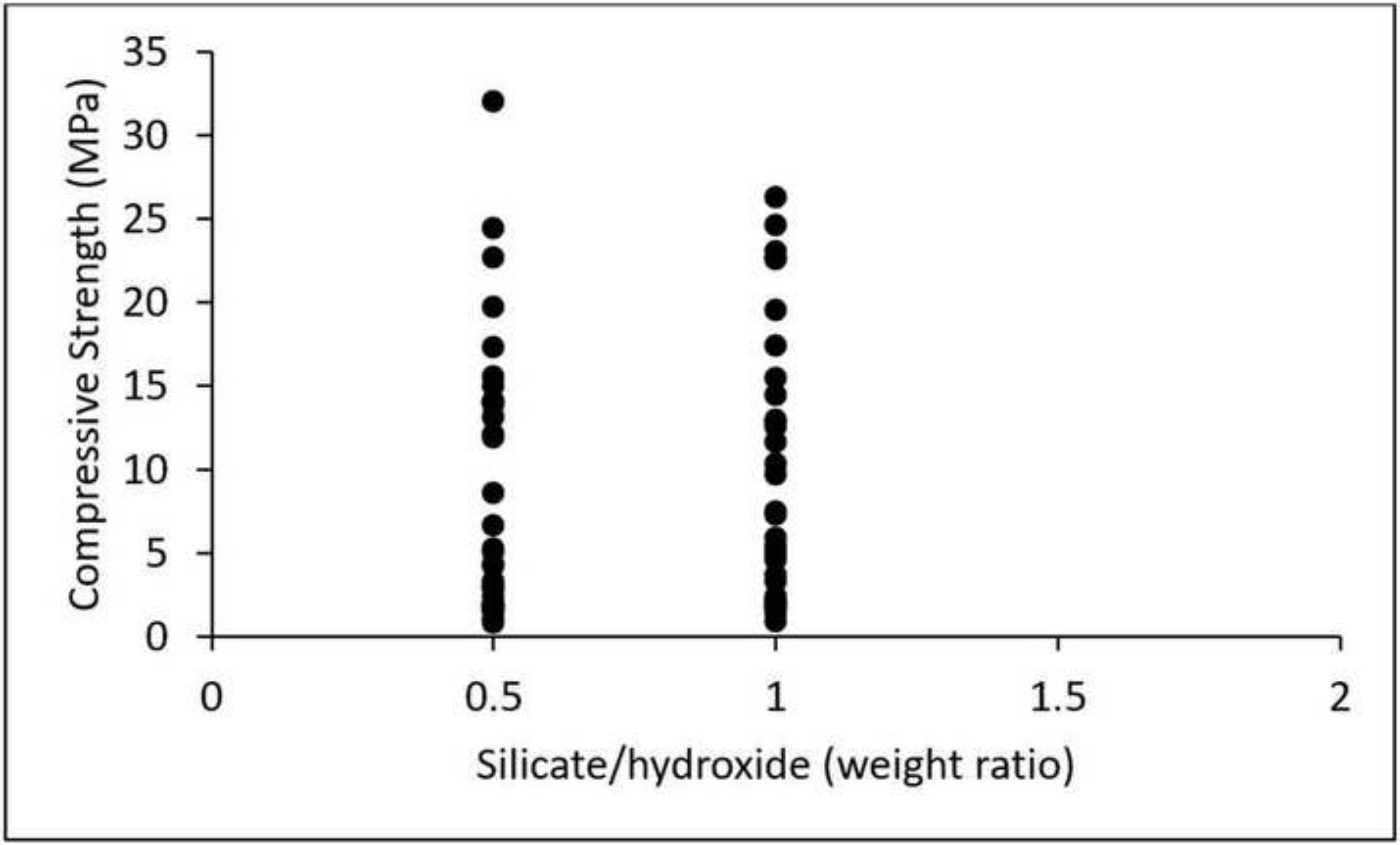
Compressive Strength (MPa)
0
5
10
15
20
25
30
35
5
7
9
11
13
15
Hydroxide (wt%)

Fig 12e

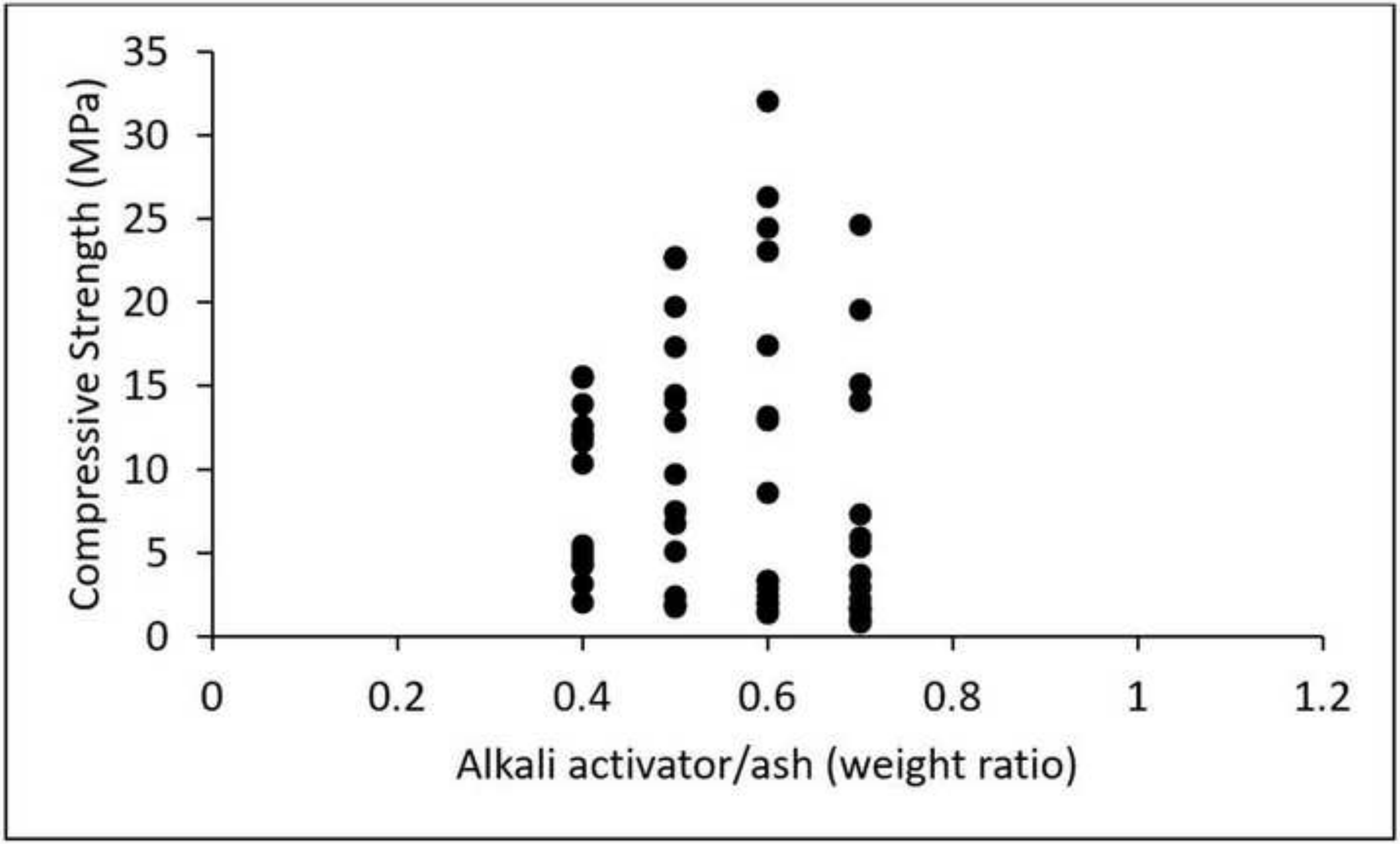


Compressive Strength (MPa)
35
30
25
20
15
10
5
0
3
5
7
9
11
13
Silicate (wt%)

Fig 12f



35
30
25
20
15
10
5
0
Compressive Strength (MPa)
0
0.5
1
1.5
2
Silicate/hydroxide (weight ratio)


35
30
25
20
15
10
5
0
Compressive Strength (MPa)
0
0.2
0.4
0.6
0.8
1
1.2
Alkali activator/ash (weight ratio)

List of Figures