7th International Conference on Debris-Flow Hazards Mitigation

On the regression of the velocity distribution of debris flows using machine learning techniques

Li-Jeng Huang[a,*], Dar-Horng Hsiao[a]

[a]*Department of Civil Engineering, National Kaohsiung University of Science and Technology, 807 Kaohsiung, Taiwan, R.O.C.*

Abstract

Five machine learning techniques— classical nonlinear regression (NLR), multi-layer perceptrons (MLP), support vector machines (SVM) with radial-basis function (RBF) kernel, *k* nearest neighbour (*k*NN) and decision tree (DT) schemes— were applied for regression of velocity distribution along the depth of debris flows by using experimental data of steady uniform open-channel flows. Programs coded in Python and package *scikit-learn* were developed for machine learning analyses. Experimental results of two cases conducted and published by Matsumura and Mizuyama (1990) were adopted for training and prediction curves of the velocity distributions using the five different machine learning techniques. Three theoretical formulas were employed for comparison and investigation, the power-law derived by Takahashi (1978) based on Bagnold dilatant flow, theory modified by Matsumura and Mizuyama (1990), and the two-region formula derived by Su et al. (1993). *R*-squared scores for each case were calculated to check the fitness of the machine learning results to the experimental data and then to verify the fitness of the theoretical formulas to the machine learning predictions. The quantified results revealed that machine learning schemes provide powerful approaches for building prediction models for velocity distribution of debris flows.

Keywords: data analysis; debris flows; machine learning; nonlinear regression; velocity distribution

1. Introduction

Disasters caused by debris flows often occur in Japan, Taiwan and elsewhere in the world (Takahashi, 1977; Jan, 2000). Development of the disaster prevention techniques is based on the understanding and analysis of the mechanical characteristics of debris flow. Debris flows are inherently non-Newtonian flows from the viewpoints of fluid mechanics, in which the rheological behavior is highly nonlinear and complicated.

Many flow models have been proposed for analysis of the mechanical characteristics of debris flows. Among them, the following models are useful and significant: the dilatant fluid model initiated by Bagnold (1954) and extended by Takahashi (1977, 1978); the Bingham fluid model, and the pseudo- or generalized visco-plastic fluid models proposed by Chen (1986), O'Brien and Julien (1988), Chen et al. (1991), and Julien and Lan (1991); the Prandtl mixing-length model employed by Matsumura and Mizuyama (1990); the modified turbulent flow model proposed by Yu and Chen (1990); the mixed-layer model proposed by Su et al. (1993); and the two-layer model, proposed by Ho (1997) in which an inertia sub-region and a viscous sub-region exist.

Conversely, machine learning and artificial intelligence technologies have been developed widely during the past decades (Muller and Guido, 2017; Bonaccorso, 2017). Among these schemes the supervised learning algorithms employed for regression can be applied for the prediction of flow velocity profiles.

In this study we applied five machine learning schemes, i.e., classical nonlinear regression (NLR), multi-layer perceptrons (MLP), support vector machines (SVM) with radial-basis function (RBF) kernel, k nearest neighbour (kNN) and decision tree (DT) schemes to predict the velocity profiles of debris flows using the experimental data from the study of Matsumura and Mizuyama (1990). Two objectives were emphasized in this investigation: (1) to check the fitness of the five machine learning techniques to the experimental data; and (2) to compare the fitness of three theoretical formulas to the machine learning predictions.

* Corresponding author e-mail address*: ljhuang@nkust.edu.tw*

2. Some Supervised Machine Learning Techniques for Regression

The prediction of velocity profile for a debris flow is in general a nonlinear regression problem due to the inherent non-Newtonian characteristics of the debris flow. Regression problems pertain to supervised learning because of the existence of targets of value type for training data. The relationship between velocity (the target), *u*, and depth (the data), *y*, can be expressed as follows:

$$u = f(y) \tag{1}$$

where *f* is in general a non-linear function. Some researchers have attempted to derive the relationship based on mechanics of debris flows, including the equations of continuity, momentum, energy and the kinematics of non-Newtonian flows in which some special term such as Bagnold stress term was introduced (Takahashi, 1978; Matsumura and Mizuyama, 1990; Su et al., 1993). However, there are some parameters that make theoretical analyses difficult to be applied to practical cases, for example, the constant *a* in three theoretical formulas and mixing lengths present in the turbulence flow models.

In the following sections we summarize the five machine learning techniques used in this study for regression of Eq. (1) that is obtained from experimental data, especially those developed and provided in *scikit-learn* package (Scikit-learn.org, 2018):

2.1. Nonlinear Regression (NLR)

In this scheme a power-law form of the nonlinear relation can be expressed as

$$u(y) = c\ y^n \tag{2}$$

This equation can be transformed into a linear one by taking the natural logarithm on both sides. Then we obtain

$$\ln u = \ln c + n \ln y = A + BY \tag{3}$$

After the linear regression analysis we can obtain the two parameters: $c = e^A$, $n = B$. The approach is direct and simple and the obtained value *n* in the power-law can be compared with theoretical results. In the *scikit-learn* package, *LinearRegression* class can be imported to solve Eq. (3).

2.2. Neural Network Using MLPs

Multi-layer perceptron (MLP), such as the *MLPRegressor* class in *scikit-learn* package, can be employed for conducting regression of a nonlinear function by training from input data to target values by constructing a specific neural network topology along with, input, output and hidden layers using different activation (transfer) functions. The errors are resolved using the back-propagation scheme. Some activation functions usually employed are as follows: (1) sigmoid (logistic); (2) tanh; and (3) relu. Parameters such as the learning rate, momentum factor, and iteration number can be adjusted.

2.3. SVM with RBF Kernel

An SVM is a powerful tool that is employed for classification and regression problems (linear or nonlinear). The concept is to search for the separation boundary for classification problem and the fitting curve for regression problem based on the so-called supporting vectors. Various kernel functions can be used, among which the Gaussian (RBF) kernel function is often employed. In the *scikit-learn* package, the *SVR* class can be employed for regression analysis.

2.4. kNNs

The *kNN*, such as the *KNeighborsRegressor* class in the *scikit-learn* package, is a non-parametric technique in which the predicted value of a point is obtained by taking the simple average or weighted average using inverse of distance of values of the *k* nearest neighbors. This algorithm is very simple.

2.5. DT

A DT is also a famous non-parametric supervised learning scheme that is used for classification and regression. In this method, the dataset is continuously partitioned into smaller subsets as the size of a tree is increased, and the final result is a tree with decision nodes and leaf nodes. The partitioning process is repeated until the criterion is satisfied. The aim is to create a model that predicts the value of target by learning simple decision rules inferred from the data features. In the *scikit-learn* package, the *DecisionTreeRegressor* class can be imported for analysis.

2.6. Pros and Cons

The pros and cons of the above five machine learning techniques are summarized and compared in Table 1 when they are applied for nonlinear regression. Moreover, the coefficient of determination, R^2, used for measurement index for all machine learning algorithms, is defined as follows:

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}} \qquad SS_{total} = \sum_{i=1}^{N} (u_i - \bar{u})^2, \; SS_{res} = \sum_{i=1}^{N} (u_i - \hat{u}_i)^2 \tag{4}$$

Here R^2 is a statistical measure that represents the portion of the variance for a dependent variable that is explained by an independent variable. The value of R^2 approaches 1.0 implies good fitness.

Table 1. Summary of the pros and cons of the five machine learning techniques employed in this study

	NLR	*MLP*	*SVM*	*kNN*	*DT*
Pros	• non-linear models • can obtain analytical curve	• Capability to learn non-linear models	• Memory efficient • Versatile in usage of kernel functions	• non-parametric method • simple	• Data scaling not required. • Simple • Robustic
Cons	• requires to assume the form of function	• Local minimum problem • Including Many parameters • Scaling sensitive	• SVMs do not directly provide probability estimates	• Prediction curve is not smooth	• May be unstable • Prediction curve is not smooth

3. Typical Theoretical Formulas for the Velocity Profiles of Debris Flows

(1) *Takahashi (1978)*: Based on the dispersive stress concept proposed by Bagnold, the velocity can be expressed as

$$u(y) = \frac{2}{3d_s}\frac{1}{\lambda}\sqrt{\frac{g\sin\theta}{a\sin\phi}[C_d + (1-C_d)\frac{\rho_f}{\rho_s}]} \bullet [h^{3/2} - (h-y)^{3/2}] \tag{5}$$

(2) *Matsumura and Mizuyama (1990)*: By adding the Reynolds turbulent stress to the theory provided by Takahashi, the following can be obtained:

$$u(y)=\frac{2}{3d_s}\sqrt{\frac{\rho_d g\sin\theta}{a\sin\phi\rho_S\lambda^2+\rho_f(\frac{1-C_d}{C_d})^{2/3}}}\bullet[h^{3/2}-(h-y)^{3/2}] \tag{6}$$

(3) *Mixed-layer theory (Su et al., 1993)*: Considering two layers, visco-layer and inertia layer, exist in the profile, the velocity distributions in each layer and interface can be derived as follows:

(a) Within the visco-layer:

$$u_{VL}(y)=\sqrt{\frac{\rho_d g\sin\theta}{a\sin\phi\rho_s h_V d_s^2\lambda^2}}\bullet\left[\frac{\pi h^2}{16}-\frac{h-2y}{4}\sqrt{y(h-y)}-\frac{h^2}{8}\sin^{-1}\frac{h-2y}{h}\right]\quad 0\le y\le h_{BL} \tag{7a}$$

(b) On the interface:

$$u_{Inter}=u_{VL}(y=h_{VL}) \tag{7b}$$

(c) Within the inertia layer:

$$u_{IL}(y)=u_{Initer}+\frac{2}{3}\sqrt{\frac{\rho_d g\sin\theta}{a\sin\phi\rho_s d_s^2\lambda^2+\rho\zeta^2 d_s^2\lambda^{-2}}}\bullet\left[(h-h_{BL})^{3/2}-(h-y)^{3/2}\right]\quad h_{BL}\le y\le h \tag{7c}$$

In the above equations, the height of the visco-layer can be deduced as follows:

$$\frac{h_{VL}}{h}=\frac{3-\sqrt{9-12[\frac{\rho\tan\theta-(\rho_s-\rho)C_d(\tan\phi-\tan\theta)}{\rho_d\tan\theta}]}}{2} \tag{7d}$$

Some important parameters are defined in Table 2.

4. Application of Five Machine Learning Techniques to Regression of the Velocity Profiles for Debris Flows

4.1. Collection of Experiment Data

We used the experimental results of the study conducted by Matsumura and Mizuyama (1990) and summarized by Su et al. (1993) as listed in Table 2. Matsumura and Mizuyama (1990) employed natural sand ($\rho_S=2.65\,g/cm^3$) and conducted the flow experiments on a channel with dimensions $7\,cm\times30\,cm\times500\,cm\,(W\times H\times L)$.

Table 2. Data for debris-flow experiments conducted by Matsumura and Mizuyama (1990)

Case	*Material Characteristics*			Experimental Conditions				Mixing Layer Parameters			
	Density g/cm^3	*Internal Friction Angle* ϕ *(deg)*	d_{50} *(cm)*	Bed Slope θ (deg)	Debris-Flow Height *(cm)*	Averaged Concentration	Shear Velocity U^* *(cm/s)*	λ	*a (VL)*	*a (IL)*	h_{VL}
1-1	2.65	38	0.30	15	3.5	0.348	21.4	4.32	0.15	0.15	0.55
1-2	2.65	38	0.30	15	3.4	0.348	21.3	4.21	0.08	0.07	0.57

4.2. Nonlinear Regression Using Machine Learning Techniques

In this study, we used Python 3.6 and the associated package *scikit-learn* for conducting regression of experimental data of the debris flows. The training data are presented in Table 3. The imported and called functions employed in each machine learning schemes are presented as follows:

```
LinearRegression(),\
MLPRegressor(hidden_layer_sizes= (20,), activation='logistic',solver='adam',alpha=0.001,batch_size='auto',\
  learning_rate='constant',learning_rate_init=0.01,max_iter=5000,random_state=0,tol=0.0001, momentum=0.),\
SVR(kernel="rbf",gamma = 1.0, C=1),\
KNeighborsRegressor(5),\
DecisionTreeRegressor(max_depth=3)]
```

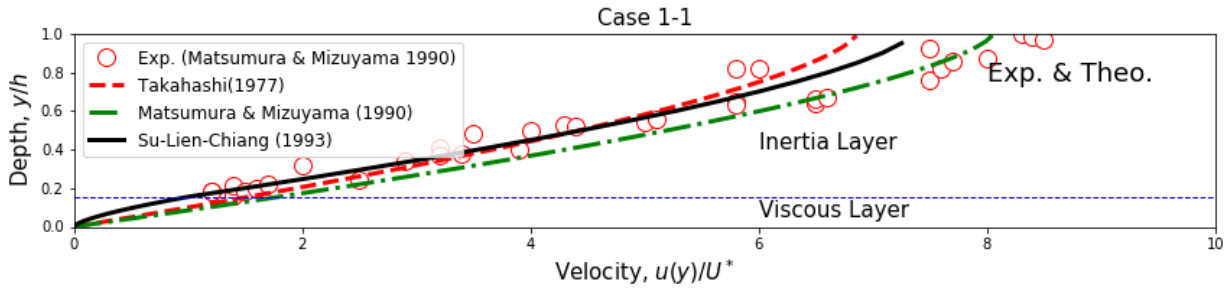
Table 3. Training data using machine learning techniques for debris-flow experiments conducted by Matsumura and Mizuyama (1990)

		Case 1-1 (N = 34)
Training Data (*X*)	Depth y/h	0.18; 0.16; 0.21; 0.18; 0.20; 0.22; 0.32; 0.24; 0.34; 0.41; 0.37; 0.38; 0.48; 0.40; 0.50; 0.53; 0.52; 0.54; 0.56; 0.65; 0.63; 0.82; 0.82; 0.64; 0.66; 0.67; 0.92; 0.76; 0.82; 0.86; 0.87; 1.00; 0.98; 0.97
Target (*y*)	Velocity u/U^*	1.20; 1.40; 1.40; 1.50; 1.60; 1.70; 2.00; 2.50; 2.90; 3.20; 3.20; 3.40; 3.50; 3.90; 4.00; 4.30; 4.40; 5.00; 5.10; 5.80; 5.80; 5.80; 6.00; 6.50; 6.50; 6.60; 7.50; 7.50; 7.60; 7.70; 8.00; 8.30; 8.40; 8.50
		Case 1-2 (N = 38)
Training Data (*X*)	Depth y/h	0.16; 0.18; 0.20; 0.22; 0.26; 0.23; 0.20; 0.26; 0.27; 0.29; 0.29; 0.40; 0.38; 0.37; 0.48; 0.44; 0.52; 0.46; 0.53; 0.48; 0.54; 0.60; 0.70; 0.71; 0.76; 0.77; 0.70; 0.76; 0.82; 0.83; 0.80; 0.79; 0.84; 0.98; 0.94; 0.84; 0.99; 0.88
Target (*y*)	Velocity u/U^*	1.00; 1.20; 1.60; 1.60; 1.60; 1.80; 2.00; 2.00; 2.10; 2.00; 2.80; 3.00; 3.20; 3.30; 3.60; 3.70; 4.00; 4.40; 4.40; 4.60; 4.90; 5.80; 5.80; 6.00; 6.10; 6.30; 6.40; 6.80; 7.00; 7.30; 7.40; 7.50; 7.70; 8.00; 8.20; 8.40; 8.80; 8.90

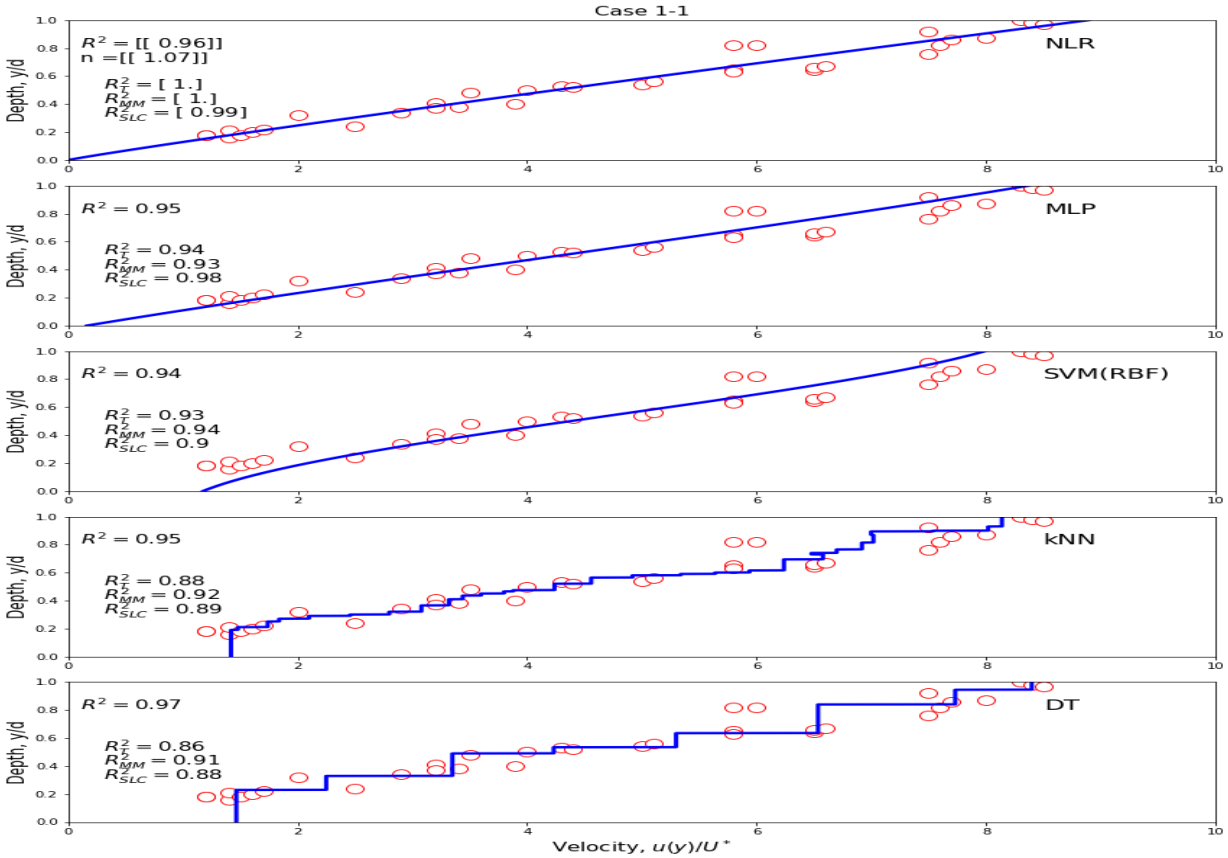
The regression results obtained using the five schemes are plotted in Fig. 1 and Fig. 2 for case 1-1 and case 1-2, respectively. In these plots the R^2 scores are presented (in the upper left corner) to depict the measure of fitness of the machine learning results to the experimental data. In the *NLR* scheme the value of the power law *n* was obtained. The value of *n* was *1.07* and *1.13* for case 1-1 and case 1-2, respectively, and the theoretical value was $n = 3/2 = 1.5$. Moreover, all the results predicted by the five machine learning schemes fit well with the original data because their R^2 scores are all near one. Note that although we can adjust some parameters in each scheme to obtain higher scores over-fitting should be avoided. The first row of Tables 4 and 5 present the averaged R^2 scores.

4.3. Comparative Study on the Prediction of Velocity Profiles Using the Theoretical Formulas

We attempted to employ the five machine learning schemes to compare the fitness of velocity predictions by using the three theoretical formulas, Eq. (5), (6) and (7a-d). Here the reference bases are the machine learning results because they have been verified to have good fitness to the experimental data and can be considered to be valid velocity profiles for case 1-1 and case 1-2. The R^2 scores R_T^2, R_{MM}^2, R_{SLC}^2 shown in the bottom left corners depict the measure of fitness of theoretical formulas obtained from Takahashi (1978), Matsumura and Mizuyama (1990), and Su et al. (1993), respectively. In the second, third and fourth row of Table 4 and Table 5, the values for each machine learning scheme and averaged R^2 scores are also summarized. We can see that all these values depict good fitness. However, among them, formulas proposed by Matsumura and Mizuyama (1990), Eq. (6), and by Su et al. (1993), Eq. (7a-d), present relatively higher fitness than that by Takahashi (1978), Eq. (5). However, we should emphasize that these analysis results are obtained based on the experimental data set we used.



(a)

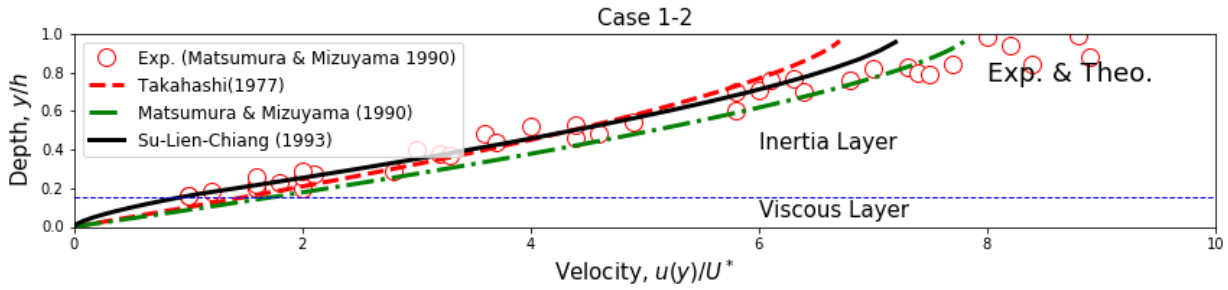


(b)

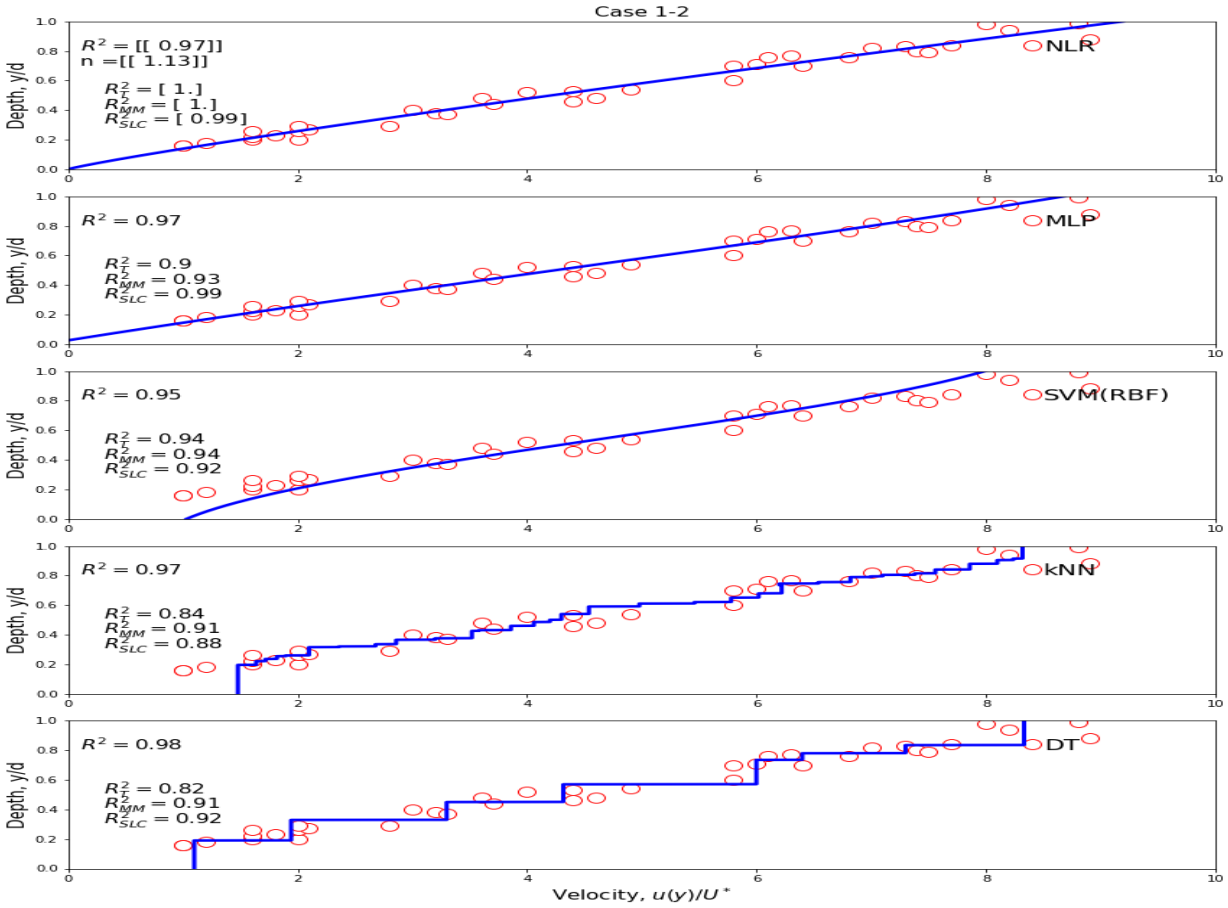
Fig. 1. Velocity profiles of debris flow in case 1-1 (a) experimental data and theoretical predictions; (b) experimental data and machine learning predictions

Table 4. R^2 score of the experimental data and the three theoretical results predicted by the five machine learning algorithms

Algorithms	*NLR*	*MLP*	*SVM (RBF)*	*kNN (k=5)*	*DT*	Ave.
Experiment	0.96	0.95	0.94	0.95	0.97	0.954
Takahashi (1978)	1.0	0.94	0.93	0.88	0.86	**0.922**
Matsumura & Mizuyama (1990)	1.0	0.93	0.94	0.92	0.91	**0.94**
Su et al. (1993)	0.99	0.98	0.90	0.89	0.88	**0.928**



(a)



(b)

Fig. 2. Velocity profiles of debris flow in case 1-2 (a) experimental data and theoretical predictions; (b) experimental data and machine learning predictions

Table 5. R^2 score of the experimental data and the three theoretical results predicted by the five machine learning algorithms

Algorithms	*NLR*	*MLP*	*SVM (RBF)*	*kNN (k=5)*	*DT*	Ave.
Experiment	0.97	0.97	0.95	0.97	0.98	0.968
Takahashi (1978)	1.0	0.9	0.94	0.84	0.82	**0.90**
Matsumura & Mizuyama (1990)	1.0	0.93	0.94	0.91	0.91	**0.938**
Su et al. (1993)	0.99	0.99	0.92	0.88	0.92	**0.94**

5. Concluding Remarks

Data analysis was conducted on using five machine learning technologies, namely, NLR, MLPs, SVM with RBF kernel, kNN and DT schemes, to predict debris-flow velocity profiles of the experimental data presented by Matsumura and Mizuyama (1990). The results are:

(1) Machine learning schemes offer systematic and convenient ways to predict the velocity profile of debris flow by using experimental data. The schemes can achieve good fitness without requiring any physical characteristics and assumptions that are usually employed in the derivation of a theoretical formula. This is a process of description and prediction of data from data.

(2) In the NLR analysis we obtained the power-law values n to be *1.07* and *1.13* for case 1-1 and case 1-2, respectively. Both these are smaller than those used in theoretical formulas ($n = 3/2 = 1.5$). The NLR model can be revised using more experimental cases, and some assumptions in theoretical formulas can be re-examined.

(3) The three theoretical predictions depicted good fitness. Among them, the results by Matsumura and Mizuyama (1990) and Su et al. (1993) presented relatively higher fitness than that by Takahashi (1978) in the analysis of the data sets used in this study.

References

Bagnold, R. A., 1954, Experiments on a gravity-free dispersion of large solid spheres in a Newtonian fluid under shear, Proceeding of Royal Society of London, Ser. A. 225, p. 49-63.

Bonaccorso, G., 2017, Machine learning algorithms, Packt Publishing.

Chen, C. L., 1986, Generalized Visco-Plastic Modeling of Debris Flow," J. of Hydraulic Engineering, 114(3), p. 237-258.

Chen, C.L., Lin, C. H. and Jan, C. D., 1991, Rheological model for ring-shear type debris flows. Proceeding. of the 5th International Sediment Conference., (5), p. 1-8.

Ho, M. L., 1997, Study on initiation mechanism and blocking structures of debris flows. [Ph.D Thesis]: Institute of Civil Engineering, National Taiwan University (in Chinese).

Huang, L. J., 2001, Introduction to theory and practice of debris- flow hazards mitigation, Chuan-Hwa Publishing Ltd., Taiwan, R.O.C. (in Chinese).

Jan, C. D. , 2000, Introduction to debris flows, Science and Technology Books Company, Taiwan, R. O. C. (in Chinese).

Julien, P. Y. and Lan, Y., 1991, Rheology of hyperconcentrations. J. of Hydraulic Engineering., 117, p. 346-353.

Matsumura, K. and Mizuyama, T., 1990, Experimental study on mechanism of debris flow using light materials. Shin-Sabo, 43(1), p. 16-22. (In Japanese).

Muller, A. C. and Guido, S., 2017, Introduction to machine learning with Python, O'Reiley, Media, Inc.

O'Brien, J. S. and Julien, P. Y., 1988, Laboratory analysis of mud flow properties. J. of Hydraulic Engineering, 114(8), p. 877-887.

Scikit-learn.org, 2018, User's guide of *scikit-learn*: https://scikit-learn.org/stable/user_guide.html (accessed October 2018).

Su, C. G., Lien, H. P. and Chiang, Y. C., 1993, Study on the velocity distribution of debris flow. J. Chinese Soil & Water Conservation, 24(1), p. 75-82. (in Chinese)

Takahashi, T., 1977, A mechanism of occurrence of mud-debris flow and their characteristics in motion, Annuals, Disaster Prevention Reserach Institute., Tokyo University, 20B(2), p. 405-435 (in Japanese).

Takahashi, T., 1978, Mechanical characteristics of debris flow. J. of Hydraulic Division, ASCE, 104(HY8), p. 1153-1169.

Yu, F. C. and Chen, C. G., 1990, Basic study on the debris flow: (II) preliminary study on the flow velocity of debris flow. J. Soil & Water Conservation, 21-22(2), p. 115-142 .(in Chinese)

Appendix A. Nomenclature

u : Velocity of debris flow

$\bar{u}$: Averaged value of velocity of all samples

u_i : Velocity of the i-th sample point

$\hat{u}_i$: Predicted velocity of the i-th sample point

X : Training data in machine leaning schemes

y : Depth of debris flow; target value in machine learning schemes