HABITAT, 33 (1), 2022, 86-92

DOI: 10.21776/ub.habitat.2022.033.1.9

Comparative Study of SVR, Regression and ANN Water Surface Forecasting for Smart Agriculture

Arief Andy Soebroto[1], Imam Cholissodin[2*], Destyana Ellingga Pratiwi[3], Guruh Prayogi W. P. [4]

[1,2,4]Department of Informatics, Faculty of Computer Science, University of Brawijaya, Jl. Veteran, Fakultas Ilmu Komputer Universitas Brawijaya, Malang 65145, Indonesia
[3]Department of Socio-Economics, Faculty of Agriculture, University of Brawijaya, Jl. Veteran, Fakultas Pertanian Universitas Brawijaya, Malang 65145, Indonesia

Received: 8 January 2022; Revised: 16 March 2022; Accepted: 1 April 2022

ABSTRACT

In the smart agriculture system based on green-based technology of artificial intelligence (AI), flooding can be predicted early by forecasting the water surface and good agricultural irrigation. The process of rising and falling of the water surface in a water basin area can be explained theoretically, but since there are many related variables and the complexity of dependencies between variables, the mathematical model is difficult to construct. Forecasting water surface in the field of irrigation needs too many variable parameters, such as cross-sectional area, depth, volume of rivers and so on. Based on patterns in each period, forecasting can be done using a statistical method and AI. This study uses the support vector regression (SVR) method, regression, multiple linear regression, and algorithm backpropagation, all compared to one another. The results of tests carried out between SVR and multiple linear regression show that SVR is superior. This can be seen from the result of the mean square error (MSE) obtained for each method. SVR 0.03 and for multiple linear regression, 0.05. The result is also supported by the best MSE result in the regression method, which is 0.338, and the best MSE value in artificial neural network (ANN), which is 0.428.

Keywords: forecasting; smart; agriculture; system; water

How to cite:

Soebroto, A. A., Cholissodin, I., Pratiwi, D. E., & P, G. P. W. (2022). Comparative Study of SVR , Regression and ANN Water Surface Forecasting for Smart Agriculture. *HABITAT*, *33*(1), 86–92. https://doi.org/10.21776/ub.habitat.2022.033.1.9

1. Introduction

Agriculture has an important position in the economic sector. For developing countries, like Indonesia, the agricultural sector provides employment to 50% of the country's workforce. The agricultural development will urge the rural development, further leading toward rural and structural transformation (Shah et al., 2019; Mogili and Deepak, 2018). Moreover, The food demand is increasing along with the population growth the demand. The conventional methods which had been utilizing by the farmers, had been no longer sufficient to fulfil those need. As a consequence,new computerized techniques were delivered to satisfied the food requirements and additionally supplied employment opportunities to billion of people.

The automation in agriculture is the main challenge across the globe. Artificial Intelligence (AI) has started to play a primary position in human daily lives, extending our perceptions and ability to alter the environment round us (Gandhi et al., 2020; Kundalia et al., 2020; Ahir et al., 2020).

Agriculture, even though being the least sector that has digitized, has opportunity for the improvement and commercialization of agricultural technologies. Artificial Intelligence (AI) is a rising technology in the area of agriculture. AI-based totally equipment and machines, has taken state-of-the-art agriculture machine to an exclusive stage.

This technology has improved crop produc tion and progressed actual-time monitoring, harvesting, processing and advertising (Yang et

*Correspondence Author.
E-mail: imamcs@ub.ac.id
Phone: +62-856-48067872

al., 2007). These kinds of technology prevent the crop yield fluctuation from various factors like the climate exchange. One of high-risk impact of climate exchange on agriculture is flood.

Flooding can be handled early through prediction by forecasting the water surface in a particular place that is considered a place of flooding, or prevented by good agricultural irrigation. Based on the theory, the high and low water level in a watershed can be explained easily, but it is difficult to build a mathematical model, because there are many variables in the irrigation field, involving time series data patterns, interrelated and complex (Lingemann, 2012; Burgin and Ristschard, 2015), for forecasting in the future, such as parameters of cross-sectional area, depth, volume of the river and so on (Ishak, Ku-Mahmud, and Norwawi, 2011). Unstable river conditions cause forecasting that cannot be done at any time, therefore must take data variables as a reference for forecasting.

This problem can be solved by applying the statistical method as a forecasting method. Statistical methods can perform learning from past noted data of water surface. One of the statistical methods that can be used to handle forecasting problems is a regression. By using regression, future water surface based on the pattern formed from the available data of water surface can be known. Artificial intelligence is one method that can also be implemented to address issues such as the process of forecasting the rise and fall of water in addition to the regression method (Kuo, 2002; Ventura, 2013).

Regression in the modern sense is a study of the dependence of one variable, that is, variables depend on one or more other variables or so-called explanatory variables to make estimates and/or predict the average population or the mean value of the dependent variable in relation to the value that is already known from explanatory variables (Gujarati, 2003). Artificial intelligence studies patterns that are formed from previously obtained data. Patterns that have been formed will produce a decision function, which can be used to predict water surfaces in the future. Research using Artificial intelligence is influenced by the data that is used as a training (Kohler, 2015; Burgin and Ristschard, 2015; Cholissodin and Riyandani, 2016; Cholissodin, et. al., 2020). The data used as training has to be accurate, to generate predictions with a small error rate value. In the case of the rise and fall of the water surface, there are two variables, time and water surface (TMA). Both of these data will be processed and used as training data in its application.

2. Theoretical Underpinning

Support Vector Regression (SVR) is one method of artificial intelligence that has high accuracy results. SVR is one method of Support Vector Machine (SVM) to overcome the problem of regression. SVR method will form a decision function that will be used to make predictions (Cholissodin and Riyandani, 2016). SVR is the implementation of SVM for the regression case (Ventura, et. al. 2013; Darwiche & Mokhiamar, 2021). SVR is a method that can overcome overfitting. The purpose of overfitting is that the data used in the training result in a much better accuracy than the test data. SVR is the development of SVM method with a statistical method that is Regression (Were, 2015).

In addition to these two methods, Artificial Neural Network is one method that can be used to perform forecasting. Artificial neural network (ANN) or also called Neural Network (NN), is an adaptive system that can change its structure to solve problems based on external or internal information that flows through the network. Simply put, ANN is a processing algorithm whose functioning is inspired through the layout and functioning of a human mind (Shah et al., 2020a, 2020b).

Neural networks have an ability of self-organization, and adaptive gaining knowledge of. It has replaced many conventional strategies in numerous fields like laptop science, mathematics, Physics, Engineering image/signal processing, monetary/ Finance, Philosophy, Linguistics, Neurology. ANN undergoes the procedure of gaining knowledge. mastering is the technique of adapting the alternate in itself as and when there is a alternate in surroundings.

There are two getting to know techniques, are supervised getting to know and unsupervised mastering. The paintings of Jha et al., 2019, encloses the linked members of the family among the numerous embedded structures and the AI era coherent with the agricultural discipline, it gave a quick about the diverse programs of neural networks, ML in this region for precision farming (Yang et al., 2007).

This study was conducted to compare the SVR method with another method; artificial intelligence (Artificial Neural Networks), Regression, and Multiple Linear Regression.

3. Research Methodology

Preparation of data is required for research to proceed according to the draft of the research design. Preparation of data required for the implementation of this study are as follows:

a. From the data on the water surface of the Gadang river area, Malang city, there are 2 columns, the first column is the time and the second column is the water surface (TMA) in the time column there is the date, month, year, hours, minutes, seconds, milliseconds. In the second column, there is only one TMA value in a given unit.

b. Convert unit time values in columns Prepare tabular data from TMA to be used as an additional feature in the testing process. This treatment is added because, when performing the test using the SVR method, it was found that the distance between the TMA value and time value that was too far, resulted in the system unable to predict it for a remote range of the training data. The system can only predict 15 of the data after. Examples of data that will be used in the research are shown in Table 1.

Table 1. Sample data with 1 feature

No	Data Number (DN)	*F*1	DN	*F*2
1	1	387.11	2	387.13
2	2	387.13	3	387.17
3	3	387.17	4	387.24
4	4	387.24	5	387.33
5	5	387.33	6	387.3
6	6	387.3	7	387.22
7	7	387.22	8	387.16
8	8	387.16	9	387.04
9	9	387.04	10	386.95
10	10	386.95	11	386.88

Sample data in Table 1 is an example of data used during training and testing using one feature training data. Data in column F1 will be used as the features, and the data in column F 2 will be used as the target of training data. Then, for example, data used for training and testing using two or more features, are shown in Table 2.

Table 2 shows that the data used is the same as the existing data in Table 1. However, there is one additional column; F3. An example of the data shown in Table 2 is an example of data that will be used for training and testing in cases of 2 features or more. Data in column F1 and F2 are used as training data for feature 1 and feature 2, while data from the F3 column is used as a target. If tests were to be done using 3 features of training data, the data in column F1, F2, and F3, which will be made a feature, should add the data tabulation new (F4) which starts from the data 4 to the *n*-th as target and applies also to training using training data above 3 features.

Table 2. Sample data with 2 features

No	DN	F1	DN	F2	DN	F3
1	1	387.11	2	387.13	3	387.17
2	2	387.13	3	387.17	4	387.24
3	3	387.17	4	387.24	5	387.33
4	4	387.24	5	387.33	6	387.3
5	5	387.33	6	387.3	7	387.22
6	6	387.3	7	387.22	8	387.16
..						
10	10	386.95	11	386.88	12	386.86

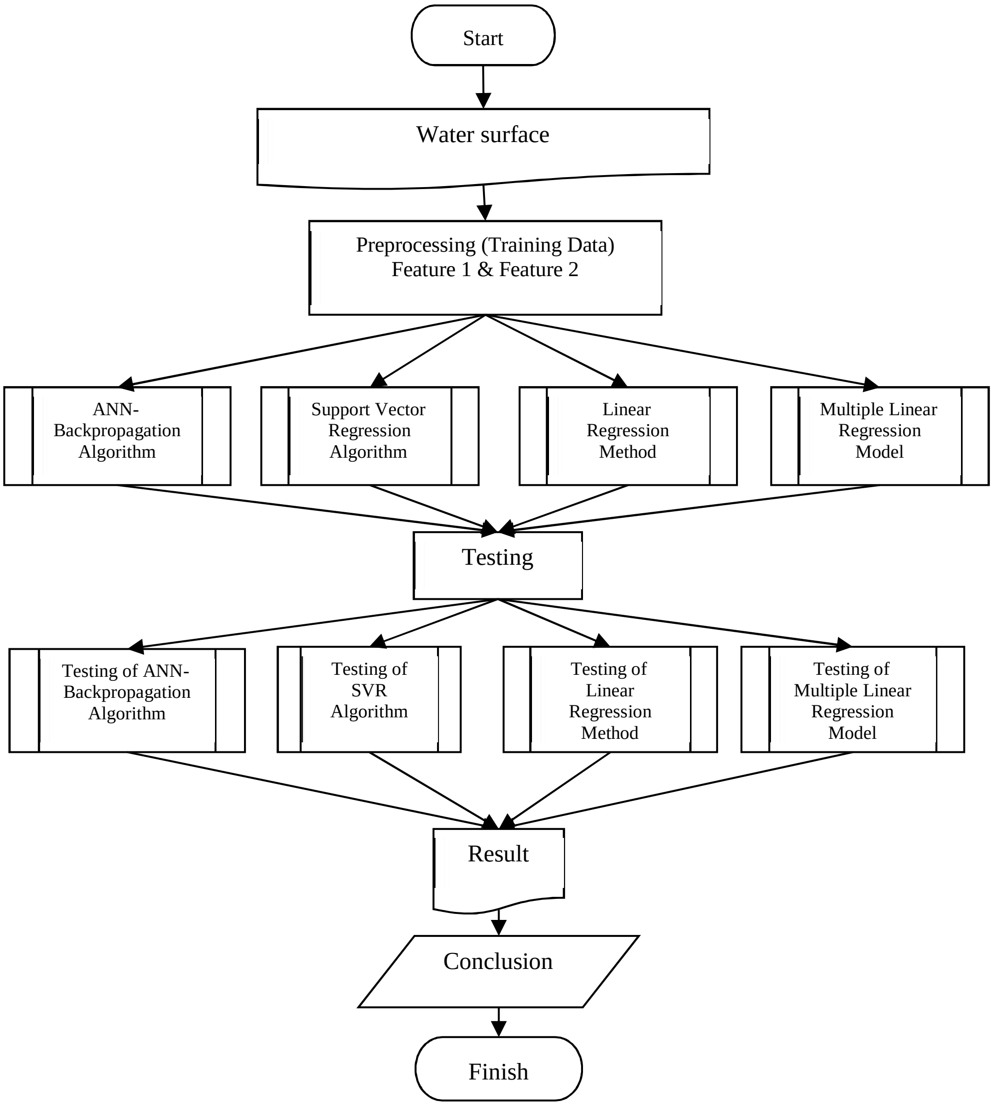

Figure 1. Flow Chart of Research Process

This stage will explain the process by comparing all algorithms. The first stage should be done to perform data preparation and preprocessing. The second stage is the initialization of parameter values to be used in the training process. The process is done continuously as much as the amount of data used in the training. The flow of the ANN, SVR, Linear Regression, and Multiple Linear Regression is shown in Fig. 1. All processes used as much as 100 preliminary data in January 2006.

4. Results and Discussion

Testing this system is done to ensure that applications that have been built to work by the specifications of the predetermined requirements. This experiment tests the Mean Square Error (MSE) of the results of prediction data with actual data using methods Support Vector Regression (SVR), Regression, or ANN. The test uses 100 preliminary data. Data is calculated from January 1st, 2006 at 00.00. The test is intended to compare the test results of regression methods (linear and multiple), and ANN method with SVR method. The purpose of the testing/comparison is to determine the extent the SVR method can forecast compared with statistical methods (regression) and artificial intelligence methods (ANN/ Backpropagation).

Comparison method is a comparison between SVR method, regression, and ANN based on testing that is done. Comparisons were performed by the best MSE results for each of the previously tested methods. This test was conducted to determine the SVR methods test results for forecasting water surfaces compared

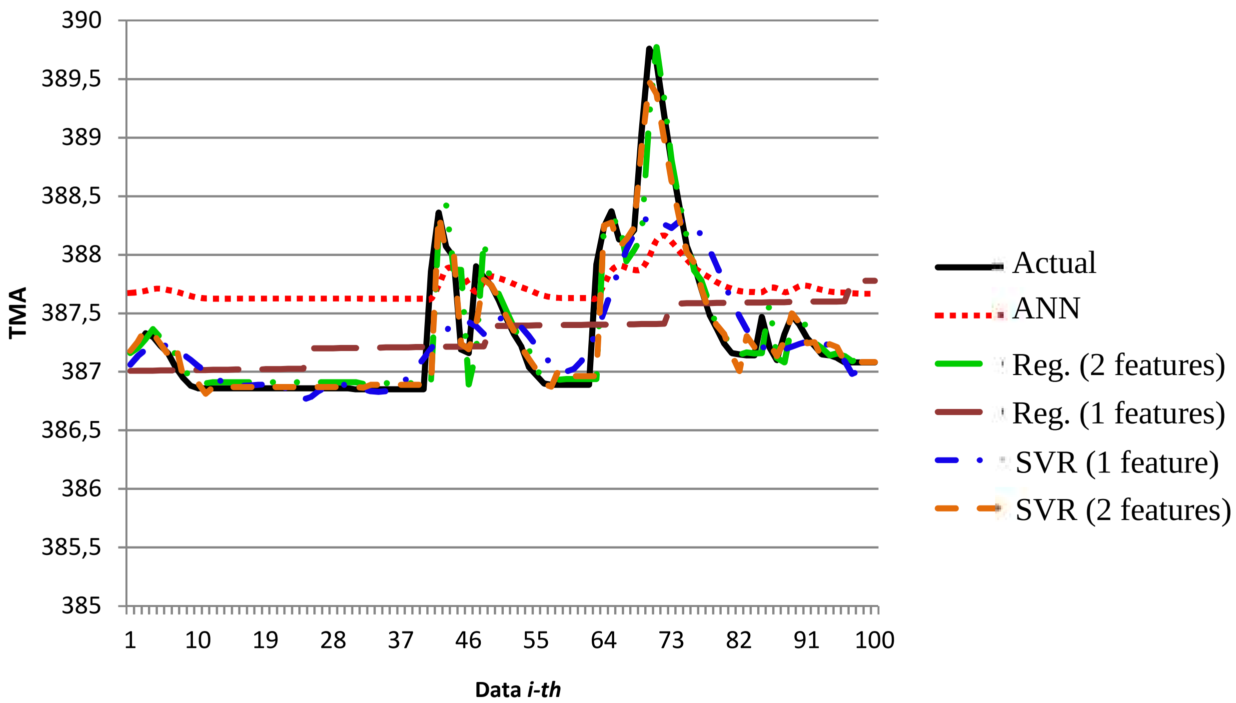

Figure 2. Results of Methods Comparison

with statistical methods and other methods of artificial intelligence (ANN / Backpropagation). The results can be seen in the comparative method in Fig. 2. Fig. 2 shows the comparison method for each testing done. The test results for SVR using 2 features is superior when compared to other methods, it is supported by the results obtained by the MSE value of 0.03. The method almost equalling the results of the accuracy of the SVR method is multiple linear regression with MSE value of 0.05. Poor results are shown from by the test results using the algorithm backpropagation, 0.42. These results can occur because the algorithm backpropagation is not as good in training. The training, which requires the generation of random weight values can be a problem that appears when performing data training. Then the number of feature data (data input) in the backpropagation algorithm is also restricted to 2 only to compare the results obtained by using SVR.

Based on test results and comparative methods done, it can be concluded that the SVR method is superior when compared with the regression method and ANN in addressing water surface forecasting. SVR can provide predictive results with a fault tolerance of 0.03, using only two features. This suggests that SVR methods are effective for forecasting data with a limited number of features.

5. Conclusion

Based on the test results, it can be concluded that:

a. Support Vector Machine Regression (SVM-R) Method can be applied to forecast water surface. Based on the comparison results of SVR with regression, SVR is more superior than the regression method. It can be seen from the results of MSE (Mean Square Error) obtained for each method, for SVR is 0.10 and Regression is 0.33.
b. SVR methods using two features as training will give better results. It can be seen from the MSE obtained when using two features, that is equal to 0.03.
c. Based on the results of the comparison between the SVR and multiple linear regression, SVR was also more superior. It can be seen from the results of MSE (Mean Square Error) obtained for each method, for SVR 0.03 and 0.05 for Multiple Linear Regression.
d. Based on the results of the comparison between SVR with artificial neural networks (backpropagation), SVR was still superior to the ANN method. It can be seen from the results of MSE (Mean Square Error) obtained for each method, for SVR 0.03 and for 0.42 ANN.

The test results of the methods done show that the SVR method is more superior than the other methods, but the parameter use of SVR method is still generated randomly. The best results obtained in the SVR method were performed with 10 tries. The method in comparison method is the same, primarily the ANN method (backpropagation). To cope with the results of uncertain value on SVR and weight parameters on the backpropagation, future research is recommended to use evolution algorithms like Parallel Swarm Optimization on SVR to perform parameter optimization and Nguyen Widrow on backpropagation for the optimization of weight at the start of data training. The addition of evolutionary algorithms in the training process can maximize the value of the parameter optimization process on SVR and weights on ANN-Backpropagation.

Acknowledgment

We would like to thank the Faculty of Computer Science (a.k.a Filkom), University of Brawijaya (UB), that financed the continuation of this article journal.

References

Ahir, K., Govani, K., Gajera, R., Shah, M., 2020. Application on virtual reality for enhanced education learning, military training and sports. Augmented Human Research (2020) 5:7.

Burgin, R., Ristschard, G. 2015. *Tree-based varying coefficient regression for longitudinal ordinal responses*. Computational Statistics and Data Analysis 86, 65–80

Cholissodin, I., Riyandani, E. 2016. *Analisis Big Data*. Fakultas Ilmu Komputer (FILKOM), Universitas Brawijaya (UB), Malang.

Cholissodin, I., Riyandani, E. 2016. *Swarm Intelligence* Fakultas Ilmu Komputer (FILKOM), Universitas Brawijaya (UB), Malang.

Cholissodin, I., Sutrisno, S., Soebroto, A.A., Hasanah, U., Febiola, Y.I. 2020. *AI, Machine Learning & Deep Learning*, Fakultas Ilmu Komputer, Universitas Brawijaya, Malang.

Darwiche, M., Mokhiamar, O. 2021. *SVR approach for predicting vehicle velocity for comfortable ride while crossing speed humps*. Alexandria Engineering Journal, Alexandria University.

Gandhi, M., Kamdar, J., Shah, M., 2020. Preprocessing of non-symmetrical images forEdge detection. Augment Hum Res 5, 10. https://doi.org/10.1007/s41133-019-0030-5.

Gujarati, D. N. 2003. *Basic Econometrics*. 4th ed. New York: McGraw-Hill Companies, Inc.

Ishak, W.H., Ku-Mahmud, K.R., Norwawi, N. M. 2011. *Neural Network Application in Reservoir Water Level Forecasting and Release Decision*. International Journal of New Computer Architectures and their Applications (IJNCAA) 1(2): 265-274

Jha, K., Doshi, A., Patel, P., Shah, M., 2019. A comprehensive review on automation in agriculture using artificial intelligence. Artificial Intelligence in Agriculture. 2, 1–12.

Kohler, M., et. al. 2015. *Estimation of a regression function corresponding to latent variables*. Journal of Statistical Planning and Inference 162:88–109

Kundalia, K., Patel, Y., Shah, M., 2020. Multi-label movie genre detection from aMovie poster using knowledge transfer learning. Augment Hum Res 5 (2020), 11. https://doi.org/10.1007/s41133-019-0029-y.
Kurkute, S.R., Deore, B.D., Kasar, Payal, Bham.

Kuo. R.J., et. al. 2002. *An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy wight elimination*. Neural Networks 15: 909-925.

Lingemann, I. et. al. 2021. *Deliverable 2.1.3 Improved Procedures For Medium And Seasonal Water Level Forecast*. European Commission Directorate-General for Energy and Transport 1049 Brussels.

Mogili, U.M.R., Deepak, B.B.V.L., 2018. Review on application of drone systems in precision agriculture. International Conference on Robotics and Smart Manufacturing. Procedia Computer Science 133, pp. 502–509.

Ventura, C., et. al. 2013. *Comparison of Multiple Linear Regressions and Neural Networks base QSAR models for the design of new* *antitubercular compounds*. European Journal of Medicinal Chemistry 70, 831e845

Were, K., et. al. 2015. A *Comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane Landscape*. Ecological Indicators 52:394–403

Shah, D., Dixit, R., Shah, A., Shah, P., Shah, M., 2020a. A comprehensive analysis regarding several breakthroughs based on computer intelligence targeting various syndromes. Augment Hum Res 5, 14 (2020).

Shah, K., Patel, H., Sanghvi, D., Shah, M., 2020b. A comparative analysis of logistic regression, random Forest and KNN models for the text classification. Augment Hum Res 5, 12 (2020).

Shah, G., Shah, A., Shah, M., 2019. Panacea of challenges in real-world application of big data analytics in healthcare sector. Data Inf. Manag., 1–10 https://doi.org/10.1007/s42488-019-00010-1. 2007. 4th IEEE. pp. 197–201 Las Vegas, Nevada.\https://doi.org/10.1007/s41133-020-00033-z.

Yang, H., Liusheng, W., Junmin, X. Hongli, Jan. 2007. Wireless Sensor Networks for Intensive Irrigated Agriculture, Consumer Communications and Networking Conference, 2007. CCNC