Tech Science Press

Flood Forecasting of Malaysia Kelantan River using Support Vector Regression Technique

Amrul Faruq[1], Aminaton Marto[2] and Shahrum Shah Abdullah[3,*]

[1]Department of Electrical Engineering, Faculty of Engineering, Universitas Muhammadiyah Malang, Malang, 65144, Indonesia
[2]Centre for Tropical Geoengineering, Universiti Teknologi Malaysia, Johor Bahru, 81310, Malaysia
[3]Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur, 54100, Malaysia
*Corresponding Author: Shahrum Shah Abdullah. Email: shahrum@utm.my

Abstract: The rainstorm is believed to contribute flood disasters in upstream catchments, resulting in further consequences in downstream area due to rise of river water levels. Forecasting for flood water level has been challenging, presenting complex task due to its nonlinearities and dependencies. This study proposes a support vector machine regression model, regarded as a powerful machine learning-based technique to forecast flood water levels in downstream area for different lead times. As a case study, Kelantan River in Malaysia has been selected to validate the proposed model. Four water level stations in river basin upstream were identified as input variables. A river water level in downstream area was selected as output of flood forecasting model. A comparison with several benchmarking models, including radial basis function (RBF) and nonlinear autoregressive with exogenous input (NARX) neural network was performed. The results demonstrated that in terms of RMSE error, NARX model was better for the proposed models. However, support vector regression (SVR) demonstrated a more consistent performance, indicated by the highest coefficient of determination value in twelve-hour period ahead of forecasting time. The findings of this study signified that SVR was more capable of addressing the long-term flood forecasting problems.

Keywords: Flood forecasting; support vector machine; machine learning; artificial intelligence; disaster risk reduction; data mining

1 Introduction

Research on the advancement of flood forecasting has been increasing since it contributes to disaster risk reductions, presenting a difficult, challenging, and complex application to model [1]. According to Sendai Frameworks for disaster risk reduction (SFDRR) of 2015-2030, the DRR is stated in priority number three and four, stipulated as 'investing in disaster risk reduction for resilience' and 'enhancing disaster risk preparedness for effective response' respectively [2]. Hence, in connection with these viewpoints, flood modelling and forecasting are crucial for disaster risk management. In many regions of the world, flood forecasting is among the few feasible options to manage flood disasters.

To date, several flood forecasting models generally focus on data-specification involving simplified various input assumptions [3]. Thus, to mimic the complex mathematical expression of physical processes and river behaviors, models with specific techniques (empirical black-box models, stochastic, and hybrids) were applied [4]. The physically and statistically based models improve the usage of advanced data-driven methods, such as in Machine Learning technique. The most well-known works of flood forecasting modelling include artificial neural networks (ANNs) [5–7], support vector machines (SVMs) [8,9], and adaptive neuro-fuzzy inference systems (ANFIS) [3,10], which have been effectively employed for both short-term and long-term flood forecasting.

ANNs model provides considerable flexibility in solving nonlinear problems, successfully applied in various hydrological areas [11,12]. ANNs has been employed for flood forecasting due to its ability and efficiency in terms of computing time. Although ANNs performed more efficiently in solving time series hydrological data rather than in a physical-based model, SVM has also been incredibly effective in improving flood forecasting techniques due to its high accuracy and capability [13]. The high accuracy of SVM compared with ANN indicated as an appropriate method for rapidly producing flood inundation forecasting and early warning system [14]. Furthermore, Wu [15] presented SVM effectiveness in different lead time of flood forecasting. The result shows that the SVM model provides a strong capability and satisfying regression model performance for one to three hour ahead of forecasting.

For more than a few decades, researchers have used conventional SVM algorithms and supervised learning algorithms such as neural networks successfully utilized for classification problems [16]. These learning approaches have been applied for regression task analysis, including function estimation by fitting a curve to a set of data points. The application of SVMs in addressing general problem of regression analysis is called Support Vector Regression (SVR). SVM has been proven in hydrological modelling and its application owing to the robustness of the system. SVM-Regression has played an important role in numerous time series forecasting applications, including flood forecasting [17]. Khaled Boukharouba [18] employed SVR for flash flood forecasting in the absence of rainfall forecast, based on the hierarchical flood events, and demontrated that SVR performed efficiently for flash flood forecasting.

Although some attempts have been devoted to address time-series issues by using SVM approach, published research works implementing SVM as a machine learning approach in the hydrological engineering area have been limited especially for flood forecasting. This study intends to evaluate these SVM models' performance against other models such as ANNs and linear regression models in predicting river water levels to address flood forecasting problems. In addition, this study aims to expand the results of a previous study [19]. This study proposes the multi-time ahead data-driven models that simulate and predict river water levels from historical-observed data by implementing SVM technique. In this study, the two machine learning algorithms, namely radial basis function and nonlinear autoregressive exogenous neural networks, have been successfully examined. The comparison among the three mentioned methods was investigated.

2 Methodology and Study Area

The proposed method has been evaluated by examining a case study in Malaysia, specifically in Kelantan River, as a representative of flood forecasting point (FFP). The area was selected due to its proximity to reservoir frequently causing seasonal-flood disasters in Malaysia. The state of Kelantan was situated in the eastern region and in the northeast of peninsular Malaysia, with Kota Bharu as the capital city of Kelantan. Kelantan state fronts China South Sea boundary in northeast, Terengganu state in east, Pahang and Perak in south and west respectively, and Thailand boundary in north. Kelantan state has a total area of about 15,101 km^2 with the population of approximately 1.76 million in 2015 [20].

Kelantan river basin covers about 13,000 km^2 with tributaries including Lebir river, Galas river, Pergau river and Nenggiri river [21]. Kelantan river is to approximately 105 km in length, including Lebir river and Galas river in Kuala Krai city, as the central part of Kelantan river, comprising approximately 2,430 km^2 and 7,770 km^2 respectively [21]. Fig. 1 illustrates the river network of Kelantan watershed, major cities, and water level stations. The total length of Kelantan main river comprises approximately 388 km from the head of its longest tributary, draining an area of about 13,000 km^2 and occupying more than 85% of Kelantan State [22].
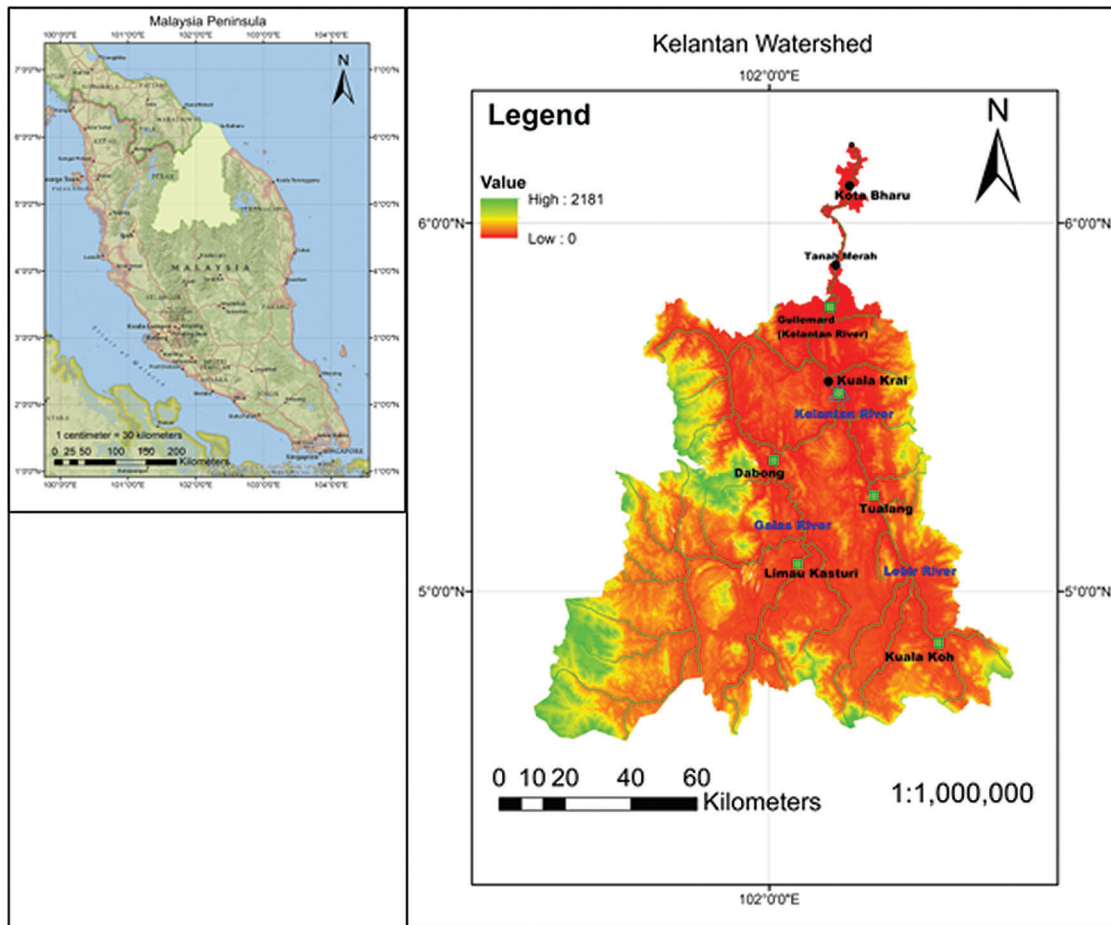
Figure 1: Location of the study area, Kelantan state, Malaysia

The river water level data is retrieved from Department of Irrigation (DID) Malaysia on fifteen minutes basis. DID supervisory control and data acquisition systems collected about three month period of data in October - December 2011. Only the specific season on the whole one-month recorded in November is used as a dataset in this study. It is about 2880 records of dataset were used, employed for training and validation test. As shown in Tab. 1, four variables indicate the river water level as input data required for SVR network, with one observed water level as an output target.

2.1 Support Vector Regression

A software package, known as LIBSVM developed by Chi-Chung Chang and Chi-Jen [23] is used in this study. At the same time, the Matlab® data normalization function was applied to normalize inputs

and targets. LIBSVM serves as a library for support vector machines (SVMs) in solving SVM optimization problem in different types, including classification SVM, support vector classification (SVC), one-class SVM for distribution estimation, support vector regression (SVR), and SV regression (SVR). In this study, SVR is employed to investigate river water level for flood forecasting model. In prior study, SVR has been successfully employed for flood forecasting in China river basin by Bafitlhile and Zhijia Li [24]. This method has been compared to ANNs models in simulating and forecasting the stream flow. Results indicated that SVR generally performs better than ANNs in stream flow forecasting of catchments.

Table 1: Input and output variables used for flood forecasting model

River WL-1	River WL-2	River WL-3	River WL-4	Observed WL
Lebir river at Tualang	Galas river at Dabong	Lebir river at Kuala Koh	Galas river at Limau Kasturi	Kelantan river at Kuala Krai

In examining the proposed model's effectiveness, it is significant to compare the previous studies. Therefore, a case study applied in [19] was examined to verify the models' performance. Two approaches, which were radial basis function neural network (RBFNN) and nonlinear autoregressive exogenous neural network (NARX), had been successfully implemented for twelve-hour period ahead of flood forecasting, with the formulation as described in [25]. The observed event-based water level data was divided into training and testing sets, where 80% of the available data was allocated for training data and the remaining 20% was allocated for testing data.

According to Vapnik's theory [26,27], SVM equations are illustrated in Eq. (1–4), respectively. Further, a set of N data points by $\{x_i, d_i\}_i^n$ is depicted in SVM-Regression function as in Eq. (1–2):

$$f(x) = w\varphi(x) + b \tag{1}$$

$$R_{SVMs}(C) = \frac{1}{2}\| w \|^2 + C\frac{1}{n}\sum_{i=1}^{n}\{L(x_i), d_i\} \tag{2}$$

In which: x_i serves as input space vector, and d_i serves as target value. Meanwhile, $\varphi(x)$ represents high-dimension feature space for mapping the input x; b is a scalar; w is a normal vector; and $C\frac{1}{n}\sum_{i=1}^{n}L(x_i, d_i)$ represents empirical error. SVR problem is formulated in the following optimization problem:

$$minR = (w, b, \xi, \xi') = \frac{1}{2} + C\sum_{i=1}^{n}(\xi_i + \xi_i') \tag{3}$$

Subject to:

$$d_i - w\varphi(x_i) + b_i \leq \varepsilon + \xi_i \tag{4}$$

$$w\varphi(x_i) + b_i - d_i \leq \varepsilon + \xi_i'$$

$$\xi_i\xi_i' \geq 0, \ i = 1, \ldots l$$

In which: regularization term is $\frac{1}{2}\| w \|^2$, ξ is loss function related to approximation accuracy of training data point, C represents error penalty factor, and l represents size of training data set. By solving Eq. (1) and (2), a generic function is obtained through Eq. (5):

$$f\left(x,\, a_i, a_i^{'}\right) = \sum_{i=1}^{n} \left(a_i -\, a_i^{'}\right) K(x, x_i) + b \tag{5}$$

In which: n is the number of support vectors, x_i is the support vector, and $K(x, x_i) =\, \varphi(x_i)\varphi(x_j)$ is a kernel function to map SVR input vector into a higher-dimensional feature space. In this study, RBF kernel is employed due to the efficiency of this kernel proven in previous studies [28]. Based on the literature, RBF kernel has worthy interpolation capabilities, mathematically expressed in Eq. (6):

$$K\left(x_i, x_j\right) = \exp\left(-\gamma \parallel x_i, x_j \parallel^2\right) \tag{6}$$

In which: variable x_i and x_j are input space vectors (vectors computed from the training or testing data set). The choice of three parameters (γ, ε, and C) determines RBF kernel function predictive accuracy. It is demonstrated that RBF outperformed than other kernel functions in SVM model [29]. Thus, in this study, RBF would be implemented as an optimization of kernel function.

The proposed models' effectiveness, can be evaluated by comparing their root mean square error ($RMSE$) and their coefficient of determination (R^2) value [30]. These formulations are illustrated in Eq. (7 and 8) respectively, in which n represents number of data points, Q_f is forecasted value, Q_0 is actual value and $\overline{Q_0}$ is average value of actual or observed records.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left|Q_0 - Q_f\right|^2} \tag{7}$$

$$R^2 = \frac{\sum_{i=1}^{n} \left(Q_f - \overline{Q_0}\right)^2}{\sum_{i=1}^{n} \left(Q_0 - \overline{Q_0}\right)^2} \tag{8}$$

3 Result and Discussions

As a result, the four input variables represent river water level in upstream and downstream area. One output variable represents the observed river water level in downstream area as flood forecasting point. The illustration of single line time series from the input and output is presented in Fig. 2, indicating that four water level inputs from upstream stations significantly impact the flood water level as observed in downstream station. Each river with its levels contributes to river water level in output area due to heavy rain at the observed time. Thus, flood disasters are inevitable due to overflow of river water level.

This study constructed multi-step models to forecast river water level at different leading time steps. The trained SVR model is utilized to hourly forecast flood water level hydrograph in one until twelve hours ahead of forecasting time. The result of actual flood data and simulated floodwater level is summarized in Fig. 3, indicating that the predicted peak levels match the recorded peak levels for all flood events. SVR model from one-hour step size is closer to the measured water level, while other step sizes are considered one step behind. However, RMSE and R^2 indicate different performances among the simulated models. Both RMSE and R^2 are calculated to evaluate model performance as illustrated in Fig. 4. For one-hour to twelve-hour period ahead of forecasting time, it is obvious that the change in RMSE and R^2 is not very significant. However, results indicate that the proposed method performs with sufficient reliability when examined in four-hour period ahead of forecasting time, depicted by the highest R^2 value and the least RMSE value obtained in this study. This finding emerges since the $t - 4$ means of four hour period before the time t has the most significant correlations for the forecasted water level. The twelve-hour period ahead of forecasting time is considered fit than other models, indicating that longer time-step of forecasting time could not reflect the expected predictions [15].
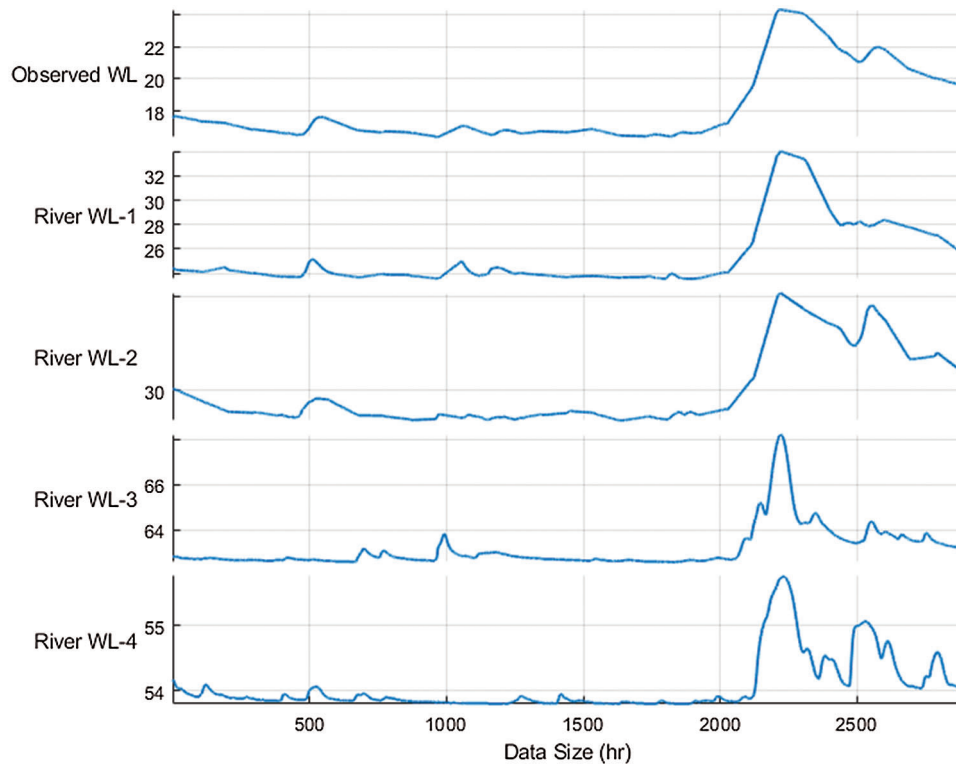
Figure 2: Single line time series of river water level input variables and output (observed) water level

Additionally, this study employed a *LIBSVM* package which is ε-SVR. SVR was trained by RBF kernel function to transform a nonlinear problem into linear function by mapping input data into a high dimensional feature space. The performance of SVR model is exceedingly sensitive based on the hyperparameter values, including cost constant C, radius of insensitive tube, ε value, and kernel parameter γ of RBF function. After several configurations, scale of C is set as 2^{-5}, 2^{-4}, …, 2^{10}, and scale of γ is set as 2^{-5}, 2^{-4}, …, 2^{5}. Further, ε-SVR has been tuned according to [31] to get the best C and best γ. Following some explorations, the best values were set as 1, 6.9644, and 0.01 respectively for C, γ, and ε.

In evaluating SVR model's effectiveness, it is necessary to compare previous studies [19] in which FFP and data observed are the same. The evaluated SVR model in twelve-hour period ahead of forecasting time was compared with the presented models. Twelve-hour period ahead of forecasting time was selected to measure the time sufficiently, preventing flood disasters. It was reported that NARX neural network outperformed RBF neural network model in forecasting a twelve-hour period ahead of forecasting time to observe flood from river water level. Fig. 5 illustrates that the studied models could perform with the actual flood value, indicating that all the proposed models are proficient in following and fitting the observed flood data. To investigate the model performance, RMSE and R^2 are calculated to get insight into the detailed performance.
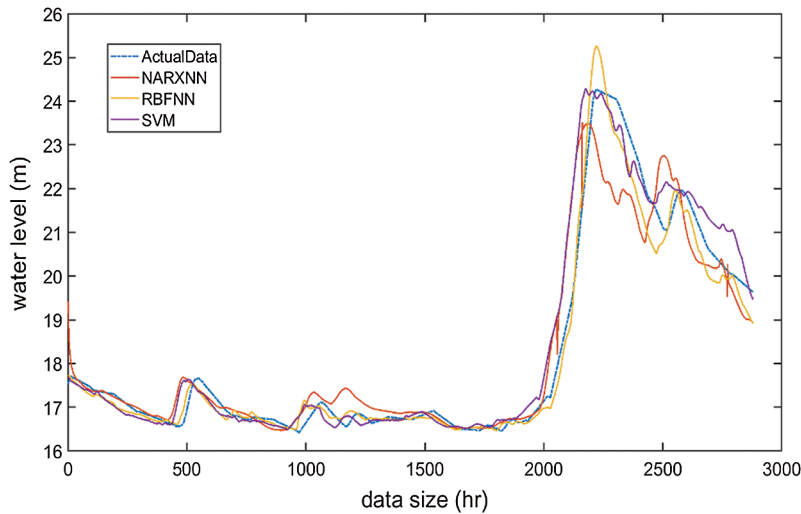
The overall comparison of model performances are calculated and summarized in Tab. 2. It can be seen that, in term of RMSE performance, the NARXNN still outperformed over the other two models. However, SVR model indicates better result as seen from the highest R^2 value; therefore, the proposed SVR model is have a great potential in long-term time ahead of flood forecasting time [32].
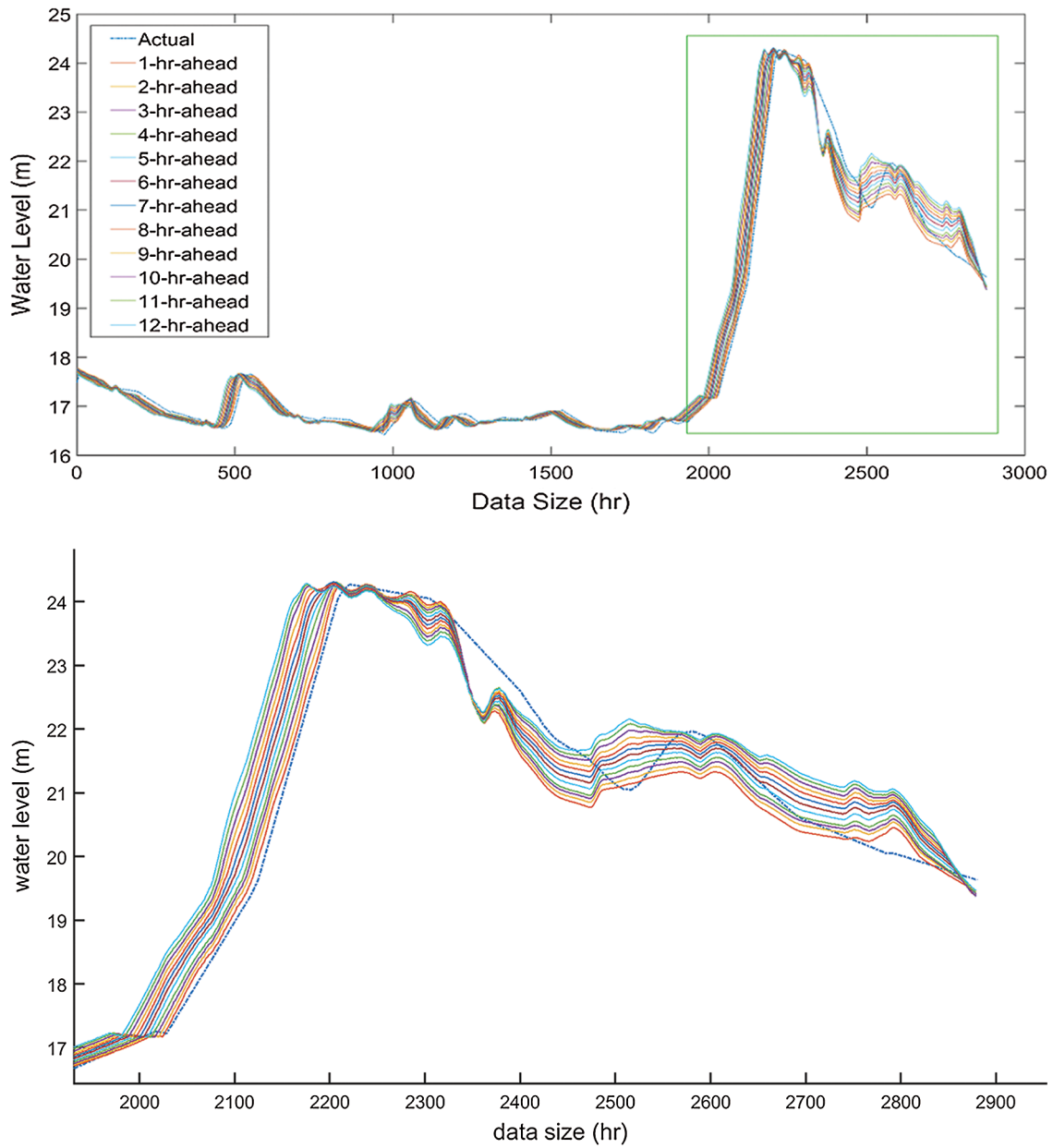
Figure 3: T-hour ahead of flood water level forecasting results

(a) R^2 (b)   RMSE
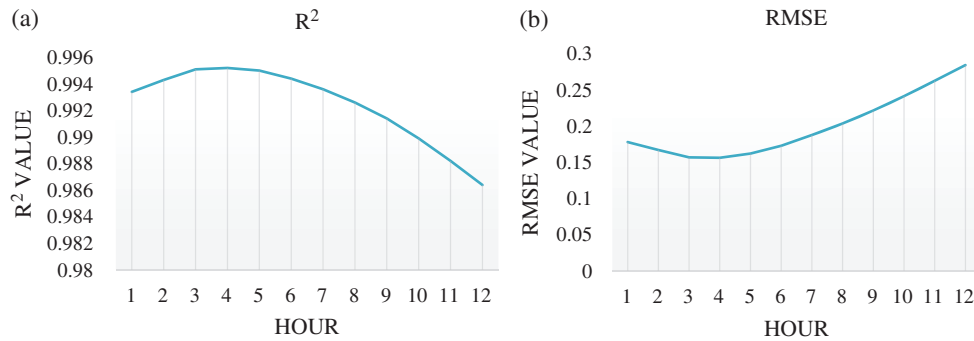
Figure 4: SVR hourly performance result ahead of settings. (a) coefficient of determination, (b) error value by RMSE

Figure 5: Overall comparison of the models performance of 12-hr period ahead of forecasting time

Table 2: Performance comparison over the proposed models (12-hr period ahead of forecasting time)

Models	RMSE	R^2
RBFNN	0.1439	0.9752
NARX-NN	0.1154	0.9836
SVM-R	0.2843	0.9864

4 Conclusions

This study is set out to assess the support vector machine algorithm's feasibility for the time-series forecasting problem. SVM-Regression is used as a technique to establish river water level in flood forecasting model. The experiment was conducted by applying river water level data, measured in Kelantan River, Malaysia. A comparison of the three methods, including SVR, RBF, and NARX neural networks, is described in this study. This study investigated that SVR could easily forecast river water level in one to twelve-hour period ahead of forecasting time. Although SVR is presented outperforms in

coefficient of determination result over the two published models, NARX neural networks still leading through error RMSE output.

This study examined three essentials machine learning methods to achieve river water level forecasting for flood disasters. These findings make several contributions to current in intelligent frameworks to build a committee machine with an intelligent system (CMIS), currently in development by the present authors. These individual learning machines could improve the proposed models to obtain the generalization and robustness of flood forecasting technique. For future research work, CMIS could also help as a promising optimization tool in the hydrological time-series forecasting topics in the context of advanced computational methods. Besides, correlation analysis between the time inputs variable and time forecasted data could be explored more in further studies.

Conflicts of Interest: Authors declare no conflict of interest in relation to this work.

References

[1] S. K. Jain, P. Mani, K. J. Sanjay, P. Prakash, V. P. Singh et al., "A brief review of flood forecasting techniques and their applications," *Int. Journal of River Basin Management*, vol. 16, no. 3, pp. 329–344, 2018.

[2] UNISDR, "Sendai framework for disaster risk reduction 2015-2030," 2015.

[3] A. K. Lohani, N. K. Goel and K. K. S. Bhatia, "Improving real time flood forecasting using fuzzy inference system," *Journal of Hydrology*, vol. 509, no. 5, pp. 25–41, 2014.

[4] T. Zhao, B. Minsker, F. Salas, D. Maidment, V. Diev et al., "Statistical and hybrid methods implemented in a web application for predicting reservoir inflows during flood events," *JAWRA Journal of the American Water Resources Association*, vol. 54, no. 1, pp. 69–89, 2018.

[5] G. Napolitano, L. See, B. Calvo, F. Savi and A. Heppenstall, "A conceptual and neural network model for real-time flood forecasting of the Tiber river in Rome," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 35, no. 3-5, pp. 187–194, 2010.

[6] S. H. Elsafi, "Artificial neural networks (ANNs) for flood forecasting at Dongola Station in the River Nile," *Alexandria Engineering Journal*, vol. 53, no. 3, pp. 655–662, 2014.

[7] Z. M. Yaseen, M. Fu, C. Wang, W. H. M. W. Mohtar, R. C. Deo et al., "Application of the hybrid artificial neural network coupled with rolling mechanism and grey model algorithms for streamflow forecasting over multiple time horizons," *Water Resources Management*, vol. 32, no. 5, pp. 1883–1899, 2018.

[8] W. C. Hong, "Rainfall forecasting by technological machine learning models," *Applied Mathematics and Computation*, vol. 200, no. 1, pp. 41–57, 2008.

[9] S. Zhu, J. Zhou, L. Ye and C. Meng, "Streamflow estimation by support vector machine coupled with different methods of time series decomposition in the upper reaches of Yangtze River," *Environmental Earth Sciences*, vol. 75, no. 1, pp. 1201–1212, 2016.

[10] Ashrafi M., Chua L. H. C., Quek C. and Qin X., "A fully-online neuro-fuzzy model for flow forecasting in basins with limited data," *Journal of Hydrology*, vol. 545, no. 1, pp. 424–435, 2017.

[11] A. Jabbari and D. H. Bae, "Application of artificial neural networks for accuracy enhancements of real-time flood forecasting in the Imjin basin," *Water*, vol. 10, no. 11, pp. 1626, 2018.

[12] A. A. Alexander, S. G. Thampi and N. R. Chithra, "Development of hybrid wavelet-ANN model for hourly flood stage forecasting," *ISH Journal of Hydraulic. Engineering*, vol. 24, no. 2, pp. 266–274, 2018.

[13] Z. Z. Latt and H. Wittenberg, "Improving flood forecasting in a developing country: a comparative study of stepwise multiple linear regression and artificial neural network," *Water Resources Management*, vol. 28, pp. 2109–2128, 2014.

[14] M. J. Chang, H. K. Chang, Y. C. Chen, G. F. Lin, P. A. Chen *et al.,* "A aupport vector machine forecasting model for typhoon flood inundation mapping and early flood warning systems," *Water*, vol. 10, no. 2, pp. 1–19, 2018.

[15] J. Wu, H. Liu, G. Wei, T. Song, C. Zhang *et al.,* "Flash flood forecasting using support vector regression model in a small mountainous catchment," *Water*, vol. 11, no. 1327, pp. 1–16, 2019.

[16] R. G. Brereton and G. R. Lloyd, "Support vector machines for classification and regression," *Analyst*, vol. 135, no. 2, pp. 230–267, 2010.

[17] P. S. Yu, S. T. Chen and I. F. Chang, "Support vector regression for real-time flood stage forecasting," *Journal of Hydrology*, vol. 328, no. 3-4, pp. 704–716, 2006.

[18] K. Boukharouba, P. Roussel, G. Dreyfus and A. Johannet, "Flash flood forecasting using support vector regression: an event clustering based approach," in *2013 IEEE Int. Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, 2013.

[19] A. Faruq, S. S. Abdullah, A. Marto, M. A. Abu Bakar, S. F. Mohd Hussein *et al.,* "The use of radial basis function and non-linear autoregressive exogenous neural networks to forecast multi-step ahead of time flood water level," *Int. Journal of Advances in Intelligent Informatics*, vol. 5, no. 1, pp. 1–10, 2019.

[20] D. of S. M. DoS Department of Statistics Malaysia, "Adjusted population and housing census of Malaysia," *Department of Statistics Malaysia*, 2015. [Online]. Available at: https://www.dosm.gov.my/v1/index.php?r=column/cone&menu_id=RU84WGQxYkVPeVpodUZtTkpPdnBmZz09.

[21] N. H. M. Ghazali and S. Osman, "Flood hazard mapping in Malaysia: case study Sg. Kelantan river basin," *Catalogue of Hydrologic Analysis Flood Hazard Mapping*, vol. 1, pp. 1–30, 2019.

[22] E. D. P. Perera and L. Lahat, "Fuzzy logic based flood forecasting model for the Kelantan River basin, Malaysia," *Journal of Hydro-environment Research*, vol. 9, no. 4, pp. 542–553, 2015.

[23] C. C. Chang and C. J. Lin, "LIBSVM: A Library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.

[24] T. M. Bafitlhile and Z. Li, "Applicability of ε support vector machine and artificial neural network for flood forecasting in humid, semi-humid and semi-arid basins in China," *Water*, vol. 11, no. 85, pp. 85–24, 2019.

[25] M. B. A. Anuar, F. A. A. Aziz, S. F. M. Hussein, S. S. Abdullah and F. Ahamd, "Flood water level modeling and prediction using radial basis function neural network: case study Kedah, in 17th Asia Simulation Conf," *AsiaSim*, vol. 751, pp. 225–234, 2017.

[26] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[27] V. Vapnik, S. E. Golowich and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," in *Neural Information Processing Systems (NIPS)*, pp. 281–287, 1996.

[28] C. Roy, S. Motamedi, R. Hashim, S. Shamshirband and D. Petković, "A comparative study for estimation of wave height using traditional and hybrid soft-computing methods," *Environmental Earth Science*, vol. 75, no. 7, pp. 244–220, 2016.

[29] X. L. Li, H. Lü, R. Horton, T. An and Z. Yu, "Real-time flood forecast using the coupling support vector machine and data assimilation method," *Journal of Hydroinformatics*, vol. 16, no. 5, pp. 973–988, 2014.

[30] K. S. Cheng, Y. T. Lien, Y. C. Wu and Y. F. Su, "On the criteria of model performance evaluation for real-time flood forecasting," *Stochastic Environmental Research Risk Assessment*, vol. 31, no. 5, pp. 1123–1146, 2017.

[31] D. Han, L. Chan and N. Zhu, "Flood forecasting using support vector machines," *Journal of Hydroinformatics*, vol. 09, no. 4, pp. 267–276, 2007.

[32] A. Mosavi, Y. Bathla and A. Varkonyi-Koczy, "Predicting the future using web knowledge: state of the art survey," in *Recent Advances in Technology Research and Education. INTER-ACADEMIA 2017. Advances in Intelligent Systems and Computing*. Vol. 660, pp. 341–349, 2018.