

Research Article

Application on shallow neuron network (SNN) in flood forecasting, case study in Vu Gia Thu Bon river basin

Van Anh Truong ^{*1}, Nguyet Minh Hoang Thi¹, Ngoc Minh Vu Thi¹

¹ Hanoi University of Natural Resources and Environment, Vietnam. Add: 41A Phu Dien Street, Bac Tu Liem District, Ha Noi, Viet Nam; tvanh@hunre.edu.vn ; htnminh.tnn@hunre.edu.vn;

* Correspondence: tvanh@hunre.edu.vn; Tel: (+84) 24 3837 0598 (1509)

Received: 08 October 2020; Accepted: 22 December 2020; Published: 25 December 2020

Abstract: Flood forecasting is one of the most important thing for flood prevention. Upto date, there are many techniques that can be used for this work, from simple ones like linear regression model (AR, ARX, ARMA, etc.) to very complex models like hydrological and hydrodynamic models. Recently, Artificial Intelligent (AI) become an cleve approach for many field including hydrological forecasting. Shallow neuron network is one of a simplest algorithm of AI but it can help to get a great result of forecasting problem due to its non-linear and automata technique. This paper present the test on applying Shallow neuron network for flood forecasting in Vu Gia Thu Bon river basin. The result show comparetable with the complex hydrological and hydraudynamic model.

Keywords: Artificial Intelligent; Flood forrecasting; Shallow neuron network; Vu Gia–Thu Bon; Machine learning.

1. Introduction

Flood is one of the most frequent and dangerous natural disaster in Vietnam [1]. They can affect an area as small as a local neighborhood or community in the mountain watersheds, to as large as an entire river basin in the Central part of Vietnam. In the past, the first option to reduce the flood damage is structural measures [2] such as dikes, reservoirs, division dam, etc. However, due to the limit of structural scale, budget and their truly effective function, they are not always the first or only option in flood management [3]. Nowadays, early warning system is usually designed and operated instead of/or parallel to structural measures to give flood forecasting services, civil protection authorities and the public adequate preparation time to eliminate the lost [4]. The key part of early warning system is flood forecasting. Flood forecasting provides the advance flow's information (magnitude and timing) at key locations of a river which helps to accelarate response system to prevent flood impact on the community exposed to flood event [5]. Unlike several other disasters, approaching flood can be forecast ahead of its occurrence with advance collection of hydro–meteorological data, and its transformation into flood water level or flood hydrograph. Therefore, there are many techniques have been developed to implement the flood forecasting, ranging from the simple ones like correlation/coaxial diagrams between two variables and mathematical equations developed using regression/ multiple linear regression to the more complecated ones like hydrological models or hydrodynamic models [6]. These methods usually contain many kind of uncertainties in their results. The linear regression models has the assumption of linearity between the dependent variable and the independent variables which can not exist in the real work therefore the error of this one usually larger than other methods [7]. In the other hand,

hydrological models and hydrodynamic models which describe the process of transforming of rain water to flow rate in the river can have more accuracy results. However, they errors still come from many sources such as the uncertainty of meteorological forecast as their inputs [8]; or models' initial conditions which are assumed in the networks of hydrodynamic models [9] or initial soil moisture, overland flows, intermediated flow, baseflow in the case of hydrological modes [10]; or the uncertainty of model parameters due to the ways of their estimation such as try and error method [11]. In the 4.0 era, data driven approach becomes more reasonable ones among flood forecast techniques in which machine learning (ML) algorithm are the popular one. ML are known as a computer can learn to do some tasks by itself without giving them the instruction of how to do these tasks [12]. Therefore they can overcome these above uncertainties. Infact, they describe the nonlinear relation of inputs and outputs instead of linear ones in traditional regression model. In addition, unlike physical based model like hydrological/ hydrodynamic models, they can use only the historical data in forecasting without requiring the initial conditions and automatically estimating their parameters by iteratively correcting their values until the criteria's termination matched [13]. This paper test a simplest algorithm of machine learning: Shallow Neuron Network (SNN) in the task of flood forecasting in Vu Gia Thu Bon river basin in Vietnam.

2. Methodology and Materials

2.1. Methodology

In this research, Shallow Neuron Network (SNN) is exploited and applied to forecast the flood in Vu Gia–Thu Bon river basin. SNN is the simplest supervised learning algorithm of the modern machine learning technique. However, in many cases including forecasting problem, it gives a very good result [14]. SNN, as its name, is composed by a neuron network with a simple feed–forward structure. They contain only one input layer, one hidden layer and one output layer (Figure 1).

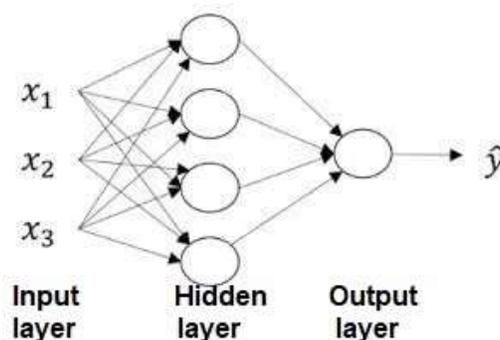


Figure 1. An example structure of SNN.

In SNN structure, each hidden neuron will receive the information from all inputs and transmit them to the outputs. In other word, each hidden neuron can be considered as the combination of 2 parts (Figure 2):

- The first part estimates its intermediated output z using the input x , the weight w and the bias b .
- The second part implements an action on z to give the final output a of the hidden neuron.

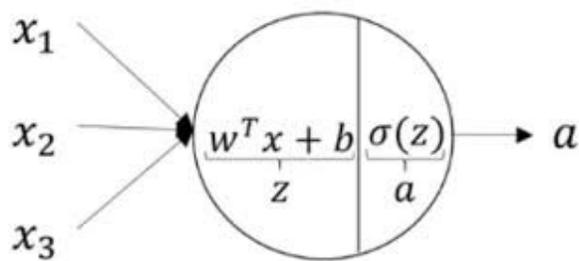


Figure 2. The structure of one hidden neuron.

Therefore, in the mathematic term, the hidden layer can be vectorized and written down as in Eq.1:

$$Z^{[1]} = W^{[1]T} X + b^{[1]} \tag{1}$$

$$A^{[1]} = \sigma(Z^{[1]})$$

where $Z^{[1]}$ is the intermediated output vector of hidden layer; $W^{[1]}$ is the weighted vector of hidden layer; $b^{[1]}$ is bias vector of hidden layer; $A^{[1]}$ is the final output of hidden layer as the active function σ of $Z^{[1]}$.

If we call $Z^{[2]}$ is the intermediated output of output layer, the final result of output layer \hat{y} can be estimated as Eq. 2:

$$Z^{[2]} = W^{[2]T} A^{[1]} + b^{[2]} \tag{2}$$

$$\hat{y} = A^{[2]} = \sigma(Z^{[2]})$$

At the beginning, the random values of parameter set (weighted matrix and bias matrix) are automatically generated. Through training process, they are corrected at each iterative loop. The technique used for parameter correction is backpropagation. The principle of the backpropagation approach is modifying internal weightings of input signals to produce an expected output signal. The system is trained using a supervised learning method, where the error between the system’s output and a known expected output is presented to the system and used to modify its internal state [13].

To optimize the parameter sets, Levenberg–Marquardt algorithm was used because they typically require less calculated time [14]. Training automatically stops when the generalization stops improving, as indicated by an increase in the mean square error of the validation samples (Figure 3).

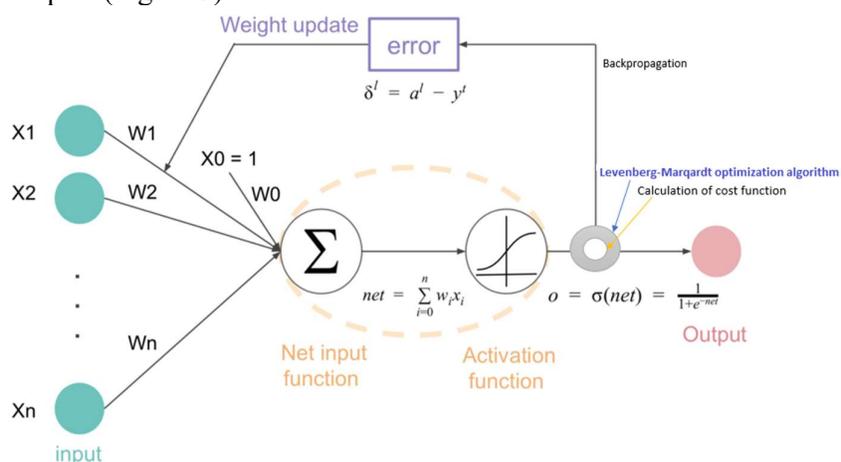


Figure 3. The training SNN’s procedure (adapted figure from [15]).

2.2. Materials

Vu Gia–Thu Bon (VGTB) River basin is one of the most prominent river basins in the Central region of Vietnam. The total length of the river is 205 km while the total surface of the river basin is 10,350 km² (Figure 4). The river runs through 3 provinces Quang Ngai, Kon Tum, Quang Nam and Da Nang city, starting in Truong Son mountain in the West and flow toward the sea into Da Nang bay in Da Nang city and at Cua Dai in Quang Nam province. The system consists of two main tributaries: Vu Gia river and Thu Bon river. Finally, Quang Hue river connects the two rivers throughout the year. The Vu Gia river consists of significant tributaries like Cai River, Bung River, A Vuong River, and Con River. The river basin is one of the most strategic and productive areas of Vietnam with an average growth rate of the GDP in the last 5 years of 11.8% but with an average poverty rate of 66.8%.

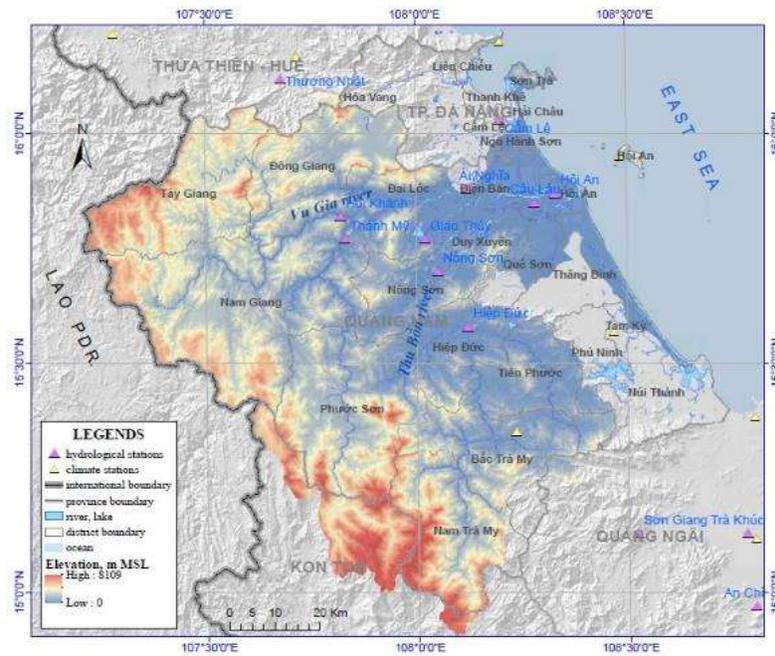


Figure 4. Vu Gia Thu Bon River Basin.

The main damages and disasters in the river basin are caused by tropical storms, flooding, drought, saline intrusion and landslide, of these, the most dangerous natural phenomena are storms and floods that causing the most significant damages in terms of human lives and property. Storms occur from May to July and October to November and typically associated with heavy rain leading to flooding. According to the provincial reports, from 1997 to 2009, the disasters due to these natural events caused 765 deaths, 63 missing persons and 2,403 injuries, with total property damage of over 18,000 BVND in Quang Nam and Da Nang city.

This research present the evaluation of applied SNN to flood forecasting at Cam Le and Hoi An stations. In data driven model, data set is the most importain one decided the success of the model. Therefore the related data have been collected. After analysing the correlation between Cam Le and Hoi An’s flow with the surrounding location, the most affect ones are Nong Son, Thanh My, Ai Nghia, Giao Thuy and Son Tra. Therefore, the following time series are collected with their longest availaible data set of flood events. This costly data can guarantee the sufficient condition for the data driven model with more than 100 past flood events:

- Discharge time series at Thanh My and Nong Son from 1978 up to 2018.
- Water level time series at Ai Nghia, Giao Thuy, Cam Le and Hoi An from 1978 up to 2018.
- Water level time series at Son Tra station from 1990 up to 2018.

3. Results and discussion

The forecasted locations are Cam Le in Vu Gia branch and Hoi An in Thu Bon branch. The lag time of flood propagation along the Vu Gia river from Thanh My and Ai Nghia to Cam Le are around 24 and 10–12 hours, respectively. Therefore lead time of 12 hours and 24 hours were chosen as ones of requested lead times by MONRE [16].

The first model (Model 1) predicts the water level at Cam Le station based on the information of last 12 hours discharge at Thanh My station, last 12 hours water level at Ai Nghia station and last 2 hours water level at Son Tra station. Because the water level at Son Tra can be forecast nearly 1 year in advance except the case of wave raising in the storm. In that case, Son Tra’s water level still can forecast 24 hours in advance with certain accuracy. Therefore the last 2 hours water level at Son Tra can be known in advance of 24 hours. The model has 1 input layer with three neurons, 1 hidden layer with 10 neurons and one output layer with one neuron (Equation 4).

$$H_{t+12}^{CL12h} = f(Q_t^{TM12h}, H_t^{AN12h}, H_{t+10}^{STr2h}) \tag{4}$$

The result of model are so good. The mean square error go down expotentially through 10 epoches and can not improve significantly after 24 epoches (Figure 5).

The correlation coefficient R also very good (larger than 0.95) through training, testing and validating the model. Scater plot between the forecasted data set with the recorded ones locating along the fitted line in all cases (Figure 6).

The profile of forecasted water level at Cam Le show the same good result with the magnitude and time at peak matching the recorded one through three big events (Figure 7). The error is smaller than 28 cm which is an accepted value by MONRE. Other criteria also show the good values as in Table 1.

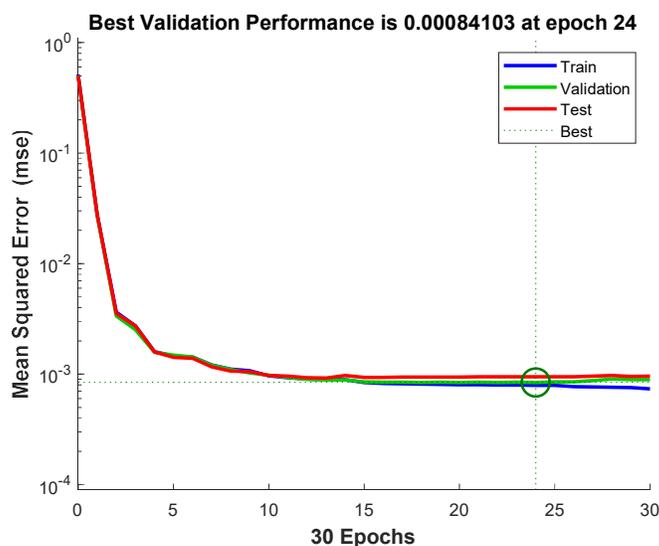


Figure 4. Mean squared error improving through epoches.

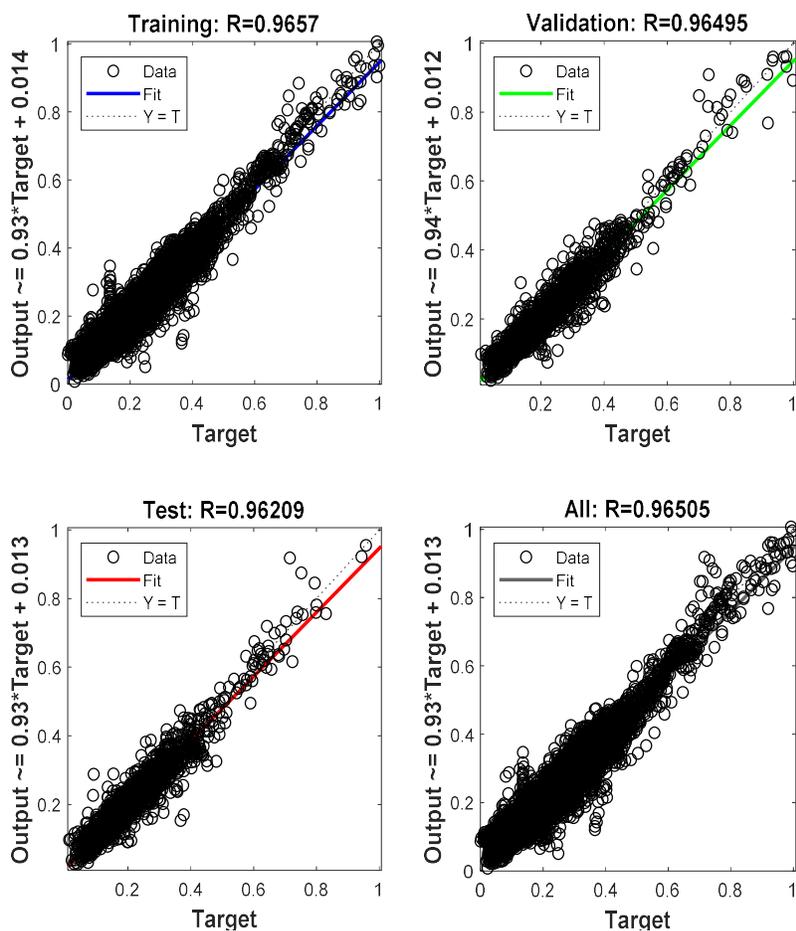


Figure 5. The correlation coefficient through learning process.

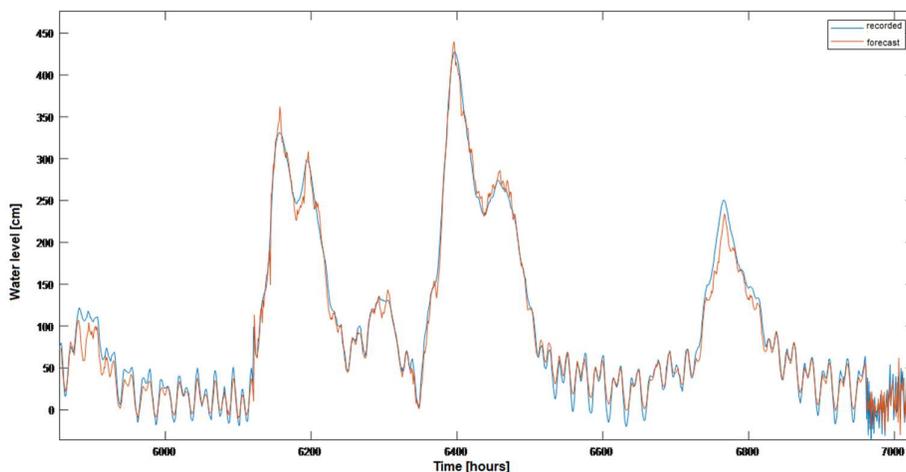


Figure 6. Forecasted water level (red) by 12 hours forecasted model and recorded ones (blue) at Cam Le.

The model of 24 hours lead time at Cam Le show similar results. It has a form as in Eq.5

$$H_{t+24}^{CL24h} = f(Q_t^{TM24h}, H_t^{AN24h}, H_{t+22}^{STr2h}) \tag{5}$$

One can see that, 2 inputs are past events which is known in advance, the only one need the forecasted value is the water level at Son Tra. However, as above analysis, Son Tra level can be predicted with a very high accuracy based on the tide lookup table created one year in advance. Therefore, we can eliminate the uncertainty of forecasted input data. The model find the best parameter set after 20 epoches. R values are still high in all learning phases (larrger than 0.94). Other criteria are good as shown in Table 1.

Table 1. forecasted criteria for flood forecast model at Cam Le station with the lead time of 12 hours and 24 hours.

Model	Summary evaluation	Acceptable Error	$\frac{S}{\sigma}$	η	P(%)
12 hours lead time at Cam Le	good	35.75	$0.36 < 0.4$	$0.93 > 0.9$	89.31%
24 hours lead time at Cam Le	good	35.75	$0.36 < 0.4$	$0.916 > 0.9$	88.94%

Hoi An station locates in downstream of Thu Bon branch. The lag time of flood propagation along the Thu Bon river from Nong Son and Giao Thuy to Hoi An are around 20 – 22 hours and 7–9 hours, respectively. Therefore lead time of 12 hours and 24 hours were chosen as ones of requested lead times by MONRE [16].

The first model predicts the water level at Hoi An station based on the information of last 12 hours discharge at Nong Son station, last 12 hours water level at Giao Thuy station and last 2 hours water level at Son Tra station. The model has 1 input layer with three neurons, 1 hidden layer with 10 neurons and one output layer with one neuron (Equation 6).

$$H_{t+12}^{HA12h} = f(Q_t^{NS12h}, H_t^{GT12h}, H_{t+10}^{STr2h}) \tag{6}$$

The result of model are so good. The mean square error go down expotentially through 30 epoches and can not improve significantly after 18 epoches (Figure 8).

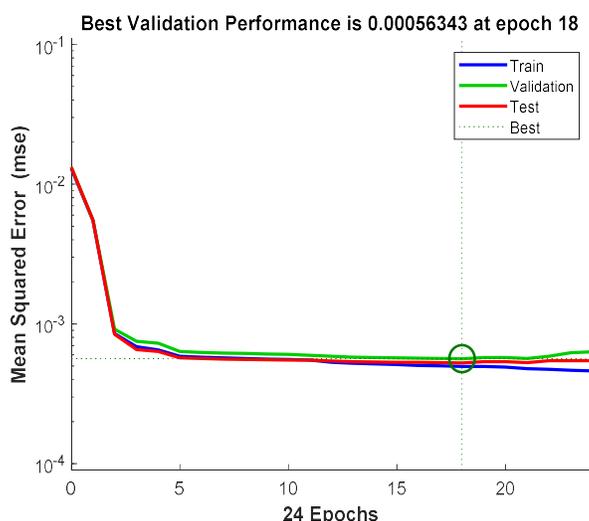


Figure 7. Mean squared error improving through epoches.

The correlation coefficient R also very good (larger than 0.96) through training, testing and validating the model. Scater plot between the forecasted data set with the recorded ones locating along the fitted line in all cases (Figure 9).

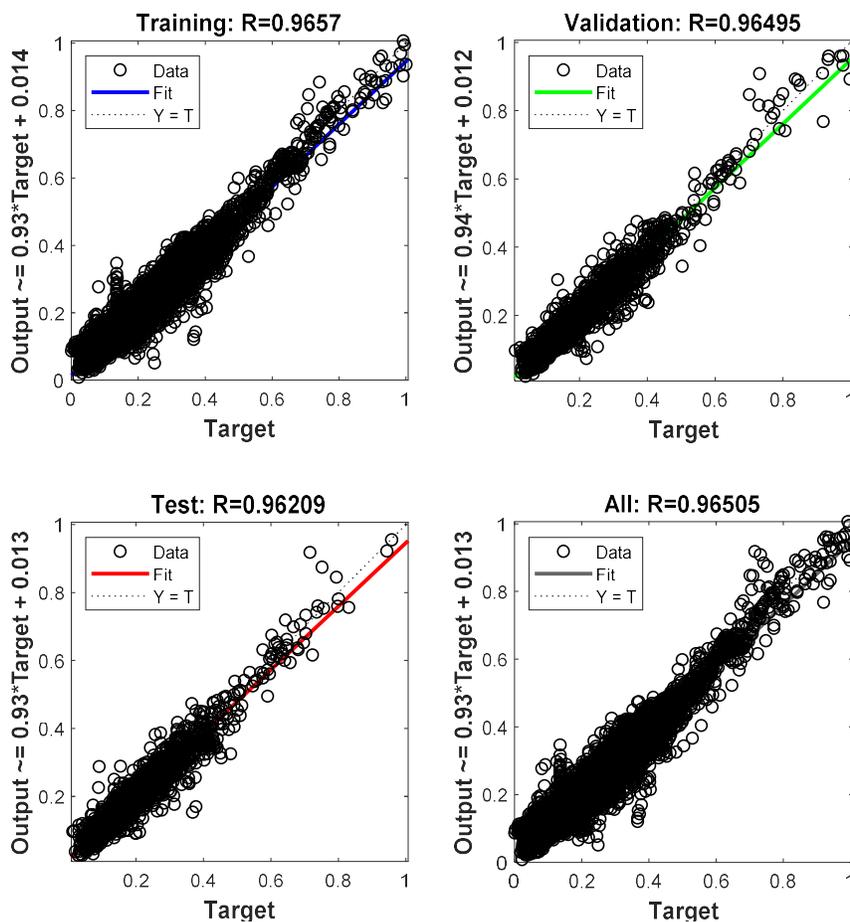


Figure 8. The correlation coefficient through learning process.

The profile of forecasted water level at Hoi An show the same good result (Figure 10) with the magnitude and time at peak matching the recorded one in three big event. The error is smaller than 28 cm – the accepted error. Other criteria also show the good values as in Table 2.

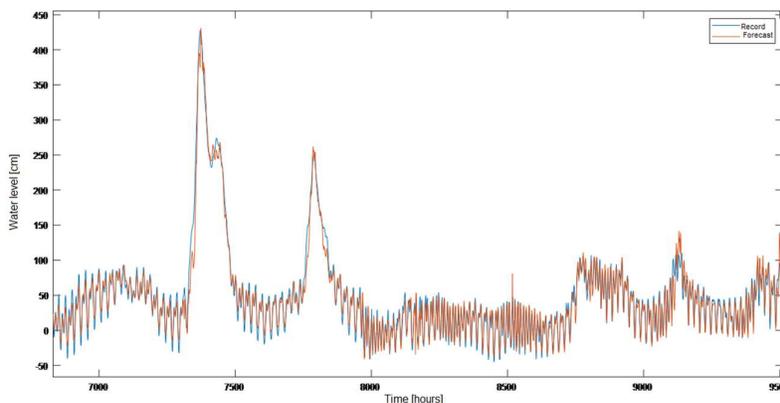


Figure 9. Forecasted water level (red) by 12 hours forecasted model and recorded ones (blue) at Hoi An.

The model of 24 hours lead time at Hoi AN show similar results. It has a form of

$$H_{t+24}^{CL24h} = f(Q_t^{TM24h}, H_t^{AN24h}, H_{t+22}^{STr2h}) \tag{7}$$

One can see in this kind of model, only water level at Son Tay is forecasted in 10 hour advance. However, as above analysis, Son Tay level can be predicted with a very high accuracy based on the tide lookup table created in one year in advance. Therefore, we can eliminate the uncertainty of forecasted input data. The model find the best parameter set after 17 epoches. R values are still high, all learning phases have value R of 0.94. Other criteria are good as shown in Table 2.

Table 2. forecasted criteria for flood forecast model at Hoi An station with the lead time of 12 hours and 24 hours.

Model	Summary evaluation	Acceptable Error	$\frac{S}{\sigma}$	η	P(%)
12 hours lead time at Hoi An	good	28.60	$0.26 < 0.4$	$0.965 > 0.9$	97.52%
24 hours lead time at Hoi An	good	28.60	$0.34 < 0.4$	$0.94 > 0.9$	95.85%

4. Conclusion

It can be seen that, neural network is an advanced approach in hydrological forecasting. Through the application process, some conclusions are drawn as follows:

- SNN is very flexible in using data. One can use discharge series and water level series in cm to predict water level in a given location without any physical process description. It just needs to normalize the data into dimensionless form to have the same range of their values.
- Model learning time is quite fast compared to physical based models. It takes only about 5–10 minutes to train the network with 1 hidden layer of 10 neurons.
- Neuron network with only 10 neurons but it generates a quite close correlation between upstream flow and tidal fluctuation to control station flow. This is an advantage of the nonlinear data-based model compared to traditional ones such as AR, ARMA, ARIMA or even physical based model in some case.

Acknowledgments: This article is built using the results of the grassroots project of Hanoi University of Natural Resources and Environment, namely “Application on machine learning algorithms in flow forecasting in Vu Gia–Thu Bon River basin” grant number: CS.2020.05.14” under the leading of Truong Van Anh, the first author of this article.

Author Contributions: Conceptualization, N.M.V.T: Data sets, V.A.T: Methodology, N.M.H.T: Verification of results, V.A.T: Writing–original draft preparation, N.M.H.T: Writing–review and editing.

Conflicts of Interest: The authors declare no conflict of interest.

Nhu, O., Thuy, N., Wilderspin, I., & Coulier, M. (2011). A preliminary analysis of flood and storm disaster data in Viet Nam. *Ha Noi, March*, 1–10.
http://www.desinventar.net/doc/Viet_Nam_2011.pdf

References

1. Nhu, O., Thuy, N., Wilderspin, I., & Coulier, M. (2011). A preliminary analysis of flood and storm disaster data in Viet Nam. Ha Noi, March, 1–10. http://www.desinventar.net/doc/Viet_Nam_2011.pdf
2. Jain, S.K.; Mani, P.; Jain, S.K.; Prakash, P.; Singh, V.P.; Tullos, D.; Kumar, S.; Agarwal, S.P.; Dimri, A.P. A Brief review of flood forecasting techniques and their applications. *Int. J. River Basin Manage.* **2018**, *16*, 329–344. <https://doi.org/10.1080/15715124.2017.1411920>
3. Snieder, E.; Shakir, R.; Khan, U.T. A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models. *J. Hydrol.* **2019**, 124299. <https://doi.org/10.1016/j.jhydrol.2019.124299>.
4. Cloke, H.L.; Pappenberger, F. Ensemble flood forecasting: A review. *J. Hydrol.* **2009**, *375*, 613–626. <https://doi.org/10.1016/j.jhydrol.2009.06.005>
5. World Meteorological Organization (WMO). Manual on flood forecasting and warning. In World Meteorological Organization (Zenbakia 1072), 2011.
6. Ranit, A.B.; Durge, P.V. Different Techniques of Flood Forecasting and Their Applications. 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), **2018**, 1–3. <https://doi.org/10.1109/RICE.2018.8509058>.
7. MAXWELL, A. Limitations of the use of the multiple linear regression model. *Br. J. Math. Stat. Psychol.* **2011**, *28*, 51–62. <https://doi.org/10.1111/j.2044-8317.1975.tb00547.x>.
8. Morss, R.E.; Wilhelmi, O.V.; Downton, M.W.; Grunfest, E. Flood risk, uncertainty, and scientific information for decision making: Lessons from an interdisciplinary project. *Bull. Am. Meteorol. Soc.* **2005**, *86*, 1593–1601. <https://doi.org/10.1175/BAMS-86-11-1593>.
9. Madsen, H.; Skotner, C. Adaptive state updating in real-time river flow forecasting—a combined filtering and error forecasting procedure. *J. Hydrol.* **2005**, *308*, 302–312. <https://doi.org/https://doi.org/10.1016/j.jhydrol.2004.10.030>.
10. Götzinger, J.; Bárdossy, A. Generic error model for calibration and uncertainty estimation of hydrological models. *Water Resour. Res.* **2008**, *44*, 1–18. <https://doi.org/10.1029/2007wr006691>.
11. Vrugt, J.A.; Gupta, H.V.; Bouten, W., Sorooshian, S. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res.* **2003**, *39*, 1201. <https://doi.org/10.1029/2002WR001642>.
12. Vu, T.H. Machine Learning Cơ bản. In Journal of Chemical Information and Modeling (Libk. 53, Zenbakia 9), 2013. <https://doi.org/10.1017/CBO9781107415324.004>.
13. Brownlee, J. Master Machine Learning Algorithms, 2020.
14. Erichson, N.B.; Mathelin, L.; Yao, Z.; Brunton, S.L.; Mahoney, M.W.; Kutz, J.N. Shallow neural networks for fluid flow reconstruction with limited sensors. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2020, 476, 2238. <https://doi.org/10.1098/rspa.2020.0097rspa20200097>.
15. Zeeshan. (d.g.). Cost, Activation, Loss Function Neural Network Deep Learning. What are these? <https://medium.com/@zeeshanmulla/cost-activation-loss-function-neural-network-deep-learning-what-are-these-91167825a4de>
16. Tổng cục Khí tượng Thủy văn. QĐ 722/QĐ-TCKTTV về quy định sai số cho phép tại các vị trí dự báo thủy văn trên các sông thuộc phạm vi cả nước, 2018.