Australian Journal of Basic and Applied Sciences
2019 December;13(12): pages 10-18
DOI: 10.22587/ajbas.2019.13.12.3
Original paper
AENSI Publications
Journal home page: www.ajbasweb.com

Water Flow forecasting using Artificial Intelligence techniques and Markov chain model: The Blue Nile Scenario

[1]Atika Hussein,[2]Johnson Agbinya

[1]*Sudan University of Science & Technology, Faculty of Computer Science &Information Technology, Name of Faculty, Box. 407.City. Khartoum Sudan*
[2]*Melbourne Institute of Technology, School of Information Technology and Engineering, Box.3000. Melbourne. Australia*

Correspondence Author: **Atika Malik Hussein:** Sudan University of Science & Technology, Faculty of Computer Science &Information Technology, computer Science Department Box.407.City. Khartoum Sudan,
E-mail: atikahussien@gmail.com
Phone number: 0024991287023.

Received date: 15 September 2019, Accepted date: 12 December 2019, Online date: 31 December 2019

Abstract

Water flow forecasting is expected to be achieved to overcome water disasters in Sudan. Daily Flow Forecasting of the Blue Nile is expected to be achieved using real data sets of river flow and weather parameters from metrological station Sudan.When the Renisunse Dam is filled up for the first time, the margin of safety has to be high to avoid unpredictable disasters. If there is a defect in Renisunse Dam, such as overflow or collapse of the Dam partially or wholly, it is going to endanger Sudan up to the Great Desert.So, it is imperative in hydrology to allow accurate evaluation in water budget, floods erosion, and even for local river navigation. In this paper, three models for forecasting water flow in Blue Nile using Artificial Neural Network, Support Vector Machine, and Markov Chain are built. An ensemble model for forecasting water flow periodically will be built and compared to the results of each model separately. Single models usually give predictions that do not consider all phenomena or events. Ensemble modeling gives better accuracy than single classifiers. In this research, real data was collected from the metrological stations (Soba and Eldeim) and the Ministry of Irrigation for the years 2003 until 2015. It includes the daily data of river flow, level, discharge, relative humidity, Sunshine Duration (SSD), rainfall, temperature maximum, temperature minimum, pressure, wind speed, and wind direction from the ground measurement. This data was used for building an ensemble model predicting the flow of the Blue Nile using three different algorithms. These algorithms, which are (Artificial Neural Network, Support vector regression, and Markov chain), were trained and applied separately for the prediction of flow. The results were compared, showing that the Markov chain gives the best accuracy for predicting river flow in the Blue Nile. Although tested on the Blue Nile, the models should apply to other rivers provided the parameters are also derived from the statistics for those rivers.

Two ensemble techniques, which were voting and bagging, were implemented. The results showed that using ensemble models with bagging and voting improved the accuracy of prediction. Also, the analysis indicated that bagging gives better accuracy than voting. The software used in this research is the R language, and Rapid Miner. It is concluded that it is difficult to determine the best algorithm to be used in a specific application. The only way to solve this problem is by trying many algorithms to find which one is better. This paper focused attention on the importance of selecting the right data or algorithm before using a particular modeling technique. The performance is compared using the correlation coefficient and accuracy.

Keywords: Support Vector Machine, Artificial Neural Network, Markov Chain, Ensemble Modeling.

INTRODUCTION

Knowledge is no longer separable from a defector and universal adequate causality. An experience by now has to be encountered in functional ecosystem functions. To cross the limitation of unusual prevalence into reality and conductivity to the actual happenings, this will have its indigenous effect in human life. The research subject is revealed as per disastrous and catastrophic events, and here we are, giving way to new research domains in real life, which are well interweaving with problem-solving techniques addressing humanitarian, environmental and physical attributes.

River flow modelling for time series is essential in managing river water resources. It allows policymakers to generate effective water utilization techniques to maximize the use of scarce water resource. Time series analysis has been widely used for river flow data modeling and weather forecasting. Using machine learning algorithms such as Support Vector regression and neural network models is getting more popular these days (Kotu *et al*., 2014).

Recently, accurate forecasting of river flow can help in many fields and jobs, such as (1) Perfect design of water storage networks. (2) They help us to avoid and manage the catastrophes coming from floods and droughts; (3) to determine the future size of reservoir capacities; (4) to enhance the effectiveness of power generation and (5) help in avoiding hydrologic hazards such as erosion and sediment movement and environmental pollutants (Toro *et al*.2013). The river flow forecasting techniques are divided into three sections (1) The physical models. (2) Conceptual models and (3) empirical models. The empirical models are linear such as autoregressive moving average. These three models cannot handle nonlinear problems involved in the hydrological process. For this reason, machine learning methods are used in the hydrological process. The most comprehensive methods used are ANN and SVM (Bourdon *et al*., 2012; Valipour *et al*., 2013).

The support vector machine is one of the data mining methods used in the machine learning field. It finds an optimum hyperplane to divide the data by using a chosen set of data. The hyperplane with the maximum margin is known as the ideal one. It is prevailed by fixing the restricted optimization problem. When the data is nonlinearly separable, it is represented in higher dimensional space using nonlinear representation. This representation allows linear regression implementation by SVM in high dimensional space (Pawar *et al*.,2019).

A Markov chain is an unsystematic procedure where the probability distribution of the present state is separate from the previous states, and this is known as the Markov property. Figure 1 indicates the Markov Chains procedure (Tiwari et al., 2010).

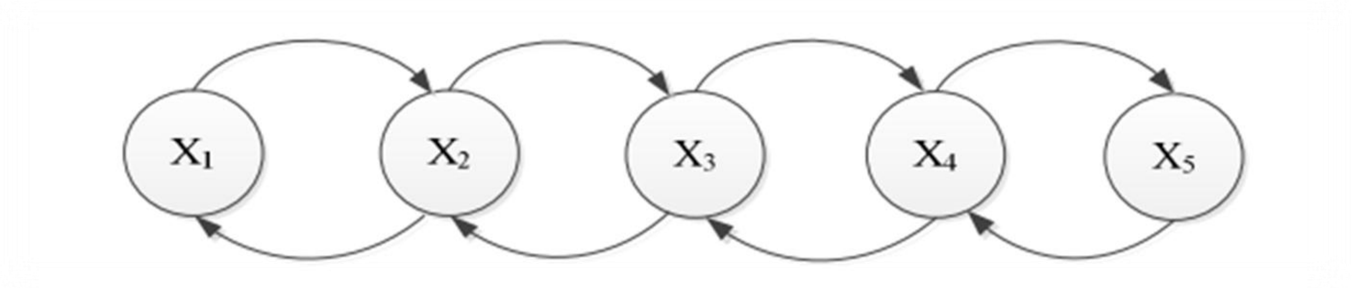

Fig 1: Markov Chain Process

After training and learning an algorithm its performance will improve and become more accurate. But if the outputs of many individual algorithms solving the same problem are joined together the concert will be enhanced strongly and vividly. The method of joining the output of different models is known as the ensemble method. (Sattari *et al*.,2018)
Ensemble models are considered one of the best prediction models used in real life implementations. An ensemble always enhances the results obtained from single predictors it combines (Galicia *et al*., 2019).

Previously researchers used to look for best models through the optimization of different algorithms. In today's terms, it is challenging to accept the best models, especially in the field of hydrology. The diversity of the algorithms used in our model is obtained from the application of different algorithms used (Duan *et al*., 2007). The main objective of this paper is to implement and build three time series models for water flow forecasting in Blue Nile using Artificial Neural Network, Support Vector Machine, and Markov Chain algorithms. Then the results of the three models are compared. After that an ensemble model of the three algorithms is implemented using voting and bagging techniques three-time and comparison of the ensemble modeling approach with a single model application is achieved.

1. LITERATU REREVIEW

In this section, some studies concerning flow forecasting using different algorithms and different ensemble models is summarized. Sattari et al. (2013) had investigated two models for predicting water flow forecasting in Sohu. They used seven days ahead of streamflow. After comparing the two models, it was found that SVM and M5 model tree have the same accuracy for flow forecasting within the same day. The M5 model has less computation time than SVM, while SVM needs many parameters. Chen et al. (2007) had presented two techniques for pruning the support vector networks for flood forecasting. The first one had used a cross-correlation method for pruning. The second one studied the relation between support vector machine parameters and pruned the support vector machine based on this relation. After comparing the original and the pruned model, it was found that the pruned model decreases the network complexity but still did not degrade the forecasting ability. It is proved in this research that pruning SVM depends mainly on the data set and the application of the study. Kisi et al. (2011) had developed a study to test the accuracy of an ensemble model using the wavelet and support vector machine in monthly streamflow forecasting. They applied the models to two data sets from different rivers in Turkey. Both algorithms were joined, and their accuracy was compared to a separate support vector machine. It was concluded that the discrete wavelet transform had improved the efficiency of the support vector machine model in stream flow forecasting.
Kisi, Ozgur, et al. (2015) had designed a model using SVM with a firefly algorithm (FA). The firefly algorithm was used for pruning SVM parameters. A comparison was made between this result and genetic programming with artificial neural network models. The SV with flying algorithm resulted in better accuracy and has better generalization ability compared to ANN and genetic programming.
Thinn et al. (2016) had developed a Markov chain model for flood forecasting in the Mandalay weather station. They had analyzed by analyzing the current state of every variable, and then the prediction of future events was made using the present probabilities or rules. Several variables affect river floods, such as snowmelt, water level, tide, and temperature; in this study, only water level was used. Tiwari et al. (2010) had developed an hourly water level forecasting model and uncertainty evaluation for level forecasting was performed. They had checked the assessment through a bootstrap based artificial neural network. They implemented an ensemble modeling technique for level forecasting by averaging the outcome from a over-fitting and under-fitting problems during training using BANN.

2. MATERIALS AND METHODS

The proposed ensemble model for river flow forecasting is shown in figure 2.

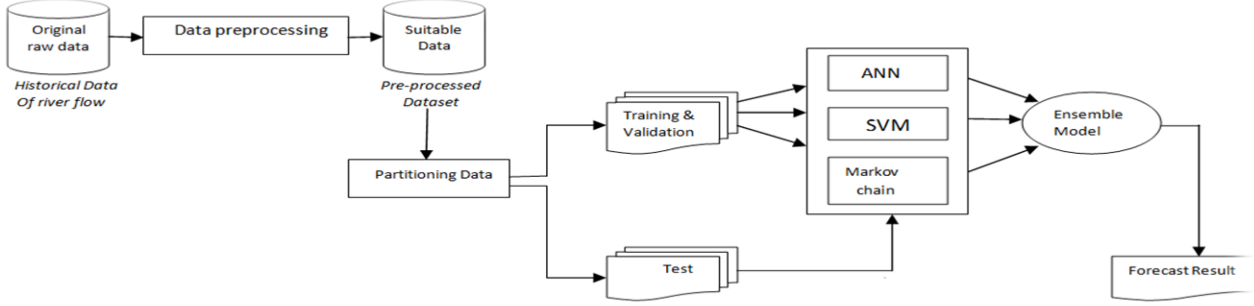

Fig 2: Block diagram of the methodology

2.1 Steps of the design

Five steps are involved in the design of the algorithm and models used in this paper. They are:

1. Data collection and preprocessing.
2. Developing three models and applying the data set on them (SVR, ANN, and Markov chain).
3. Each algorithm is optimized based on specific parameters, then their results are compared.
4. Then, the three algorithms are merged to form an ensemble model. The ensemble methods used are both voting and bagging.
5. The results of the two types of ensemble models are compared.

2.1.1 Study Area and Data Preparation

Water flow renders the total water budget along with the river specific measurement points or stations, which has to be designated to measure the flow (both velocity and quantity) and also monitor the level of river flow. These stations must be selected using divergence, topology, and meandering. In this study, the data set is taken from Soba station and Eldeim station as approved stations. Data from the metrological station and ministry of irrigation is available and represents the daily data of many parameters. These parameters are river flow (m^3/s), level, discharge, relative humidity, SSD, rainfall, temperature maximum, temperature minimum, pressure, wind speed, and wind directions) from ground measurement provided by the ministry of irrigation and metrological stations (Soba and Eldeim stations) from years 2003-2015.In order to make the data set ready to use for our model including training, validation, and testing, first preprocessing and analysis of the raw data had to be achieved. The first step was treating wrong and missing values. The second preprocessing step was applying or finding correlation between different variables. Using the correlation matrix, it was found that water flow renders the total water budget along with the river specific measurement points or stations, which has to be designated to measure the flow both velocity and quantity and also monitor the level of river flow.

From the preprocessing step, the relative humidity was identified as the most dependent variable in the flow variables. Some variables were dropped as input variables because of their low correlation with the river flow variables.

There are two essential operators in time series forecasting task, which are the window or the time lead operators. The window is the number of past days used to generate the forecasting model. Lead is the number of middle days between the previous day for creating the attributes and the one that should be predicted as a target variable.

2.2 Markov chain model

A Markov process is defined as one that does not depend on the past; it works with the present data. The sample space of every process is taken to be finite. When considering time as a discrete value, a Markov chain model becomes precisely simple. When taking all the variables as integers, it becomes a matrix of transition probabilities, and the modeling process is based on matrix algebra (Xiao fan Liu et al., 2009). In this research, an ensemble model depends on the spatial dependencies between variables for the water flow of the Blue Nile at Khartoum is built using Markov Chain. The main target of the model is analyzing the current state and movement of a variable to forecast the water flow by using presently known probabilities. The variables which are used in this data are level, discharge, and relative humidity, sunshine duration (SSD, rainfall, temperature maximum, temperature minimum, pressure, wind speed, and wind direction. Among these, the system considered only water flow and relative humidity as input variables.

The main idea of the Markov model is to find Markov states such as Flood, Approximately Flood, and Non-Flood using water flow. Then the transition probability values were found from which the transition probability matrix T is found, as shown in equation (1).

$$T = \begin{array}{c} \\ F \\ AF \\ NF \end{array} \begin{array}{c} \begin{array}{ccc} F & AF & NF \end{array} \\ \begin{bmatrix} 0.75 & 0.2 & 0.05 \\ 0.1 & 0.65 & 0.25 \\ 0.05 & 0.75 & 0.2 \end{bmatrix} \end{array} \quad (1)$$

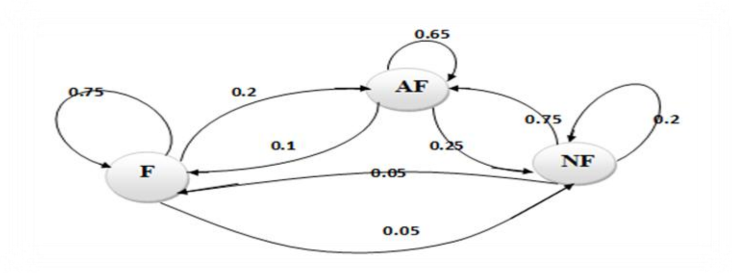

Fig. 3: Markov Model for Soba Station.

The transition matrix T shows the Soba station states; a similar matrix is available for the Eldeim station but is not required for the subsequent processing as what is obtained for the Soba station can be repeated for the Eldeim station.

2.3 Support Vector Regression

The primary task of the SVR is to find the ideal separating hyperplane that maximizes the margin of the training data. From the review, it is found that Support vector machines represent one of the most critical developments in the field of tackling hydrological parameters. It comes after (chronologically) fuzzy and artificial neural networks (Deka and Paresh, 2014). If the classes are separable, the classifier that reduces the generalization error is selected. After that, the model had chosen the hyperplane with the greatest margin between the separable sections. The margin is the aggregation of lengths of the hyperplane from the nearest dot of the separable classes. Vapnik (2013) had suggested support vector regression based on ε insensitive loss function using the principle of margin when using regression problems. The task of the SVR is to bring the error of training data into a value less than ε (Singh *et al*., 2017).

For any training data with k number of samples, denoted by (x1, y1), (xk, yk) a linear decision function can be denoted by

$$f(x, \alpha) = \langle w, x \rangle + b \quad (2)$$

Where (w, x) represents the dot product in space. A smaller value of w indicates the flatness of Eq. (2), which can be achieved by minimizing the Euclidean norm as defined by $||w||^2$ (Zhou et al.,2002) Thus, an optimization problem for regression can be written as:

Citation: Atika Hussein and Johnson Agbinya.,2019. **Water Flow forecasting using Artificial Intelligence techniques and Markov chain model: The Blue Nile Scenario**. Australian Journal of Basic and Applied Sciences, 13(12): 10-18. DOI: 10.22587/ajbas.2019.13.12.3

$$Minmize\frac{1}{2}\|w\| \quad (3)$$

$$Minmize\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{k}(\varepsilon_i - \varepsilon_i) \quad (4)$$

C represents the constant cost; ε represents the radius of the insensitive tube. These parameters are reciprocally depending on each other, and so changing one of the parameters will affect the others. If C is small, this leads to sparse approximation because of the under-fitting of training data. If C is higher, its over-fits the training data, and it will only reduce the empirical risk leading to complex learning. The parameter ε is used for smoothening the complexity of the approximation function and fixes the breadth of the ε-insensitive zone used for fitting the training data. It has an impact on the number of support vectors. When ε becomes small, it leads to a higher number of support vectors (Zhou et al., 2002).
In SVR the word tuning means selecting the suitable values of the kernel parameters which are parameters c, γ, and α. These parameters should have optimal values that are gained by trial and error methods and this is referred to as tuning. The term training means to optimize the coefficients in the decision function by solving the quadratic problem. Every time two of them are fixed and the third one is changed randomly using trial and error method until we get the best correlation

2.4 Artificial Neural Network
Figure 4 shows a three-layer feed-forward artificial neural network. It contains input, output, and hidden layer in the middle. Neuron in a layer is linked to the neurons in the following layer. Neurons found in the same layer are unconnected to each other. When the data passes from a neuron to the other, they are multiplied by the weights. The purpose of the weight is to control the strength of the signal. If the weights are changed, the data also changes and so the whole network will change. The solution of the network is found at the output node.
The neuron has three main functions, to multiply each input by its corresponding weight, sums the product, and keeps the sum in a transfer function to give the result. In this paper, Sigmoid Function is used as a transfer function (Asadi et al., 2013). The product **yj** is bounded between **yj** from the **jth** neuron in a given layer is

$$Y_j = f\left(\sum w_{ji}x_i\right) = \frac{1}{1+e^{(\sum w_{ji}x_i)}} \quad (5)$$

Where *wji*= weight of the connection joining the *j***th** neuron in a layer with the *i***th** neuron in the previous layer; and *xi* = value of the *i***th** neuron in the last layer.

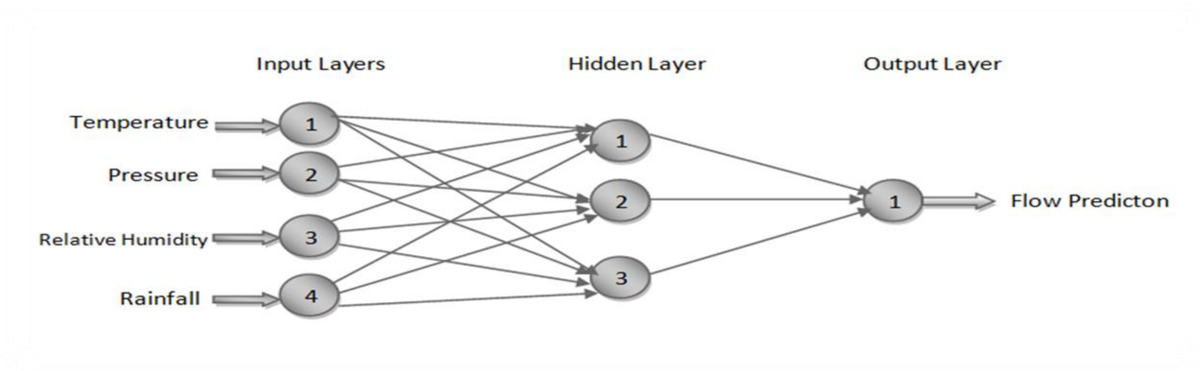

Fig4: Three-layer artificial neural network architecture used for flow prediction

2.4.1 Back Propagation (BP):
The training phase has a set of examples each example has input and equivalent target output. These example sets are presented to the neural network one by one until the network is adapted and reaches to the desired behavior. This sequential process continues until the required mapping is achieved. Every iteration of this process is called an epoch (Moreira et al., 1995). To decrease the error function gradient descent technique is used. To correct the weights for every iteration the partial derivative of the error function is calculated for each weight which leads to steepest descent. The delta rule gives the value of weight update with step size and shown in equation

$$\Delta w_{ij}(n) = -\eta\frac{\delta E(n)}{\delta w_{ij}(n)} \quad (6)$$

wij represents the weight from unit i in layer I to unit j in layer L+1.The parameter η shows the step size and is known asthe learning rate.

Each iteration of the algorithm consists of three steps:
i) The forward pass which presents an example as an input to the network and then pass through all the layers till it reaches the output. The activation value aj of unit j in layer l (2 < l < L) is calculated using a sigmoid activation function f

$$\Delta w_{ij}(n) = -\eta\frac{\delta E(n)}{\delta w_{ij}(n)} \quad a_j = f(net_j) \equiv \frac{1}{i+e^{-net_j}} \quad net_j = \sum_{i=1}^{N_{t-1}} w_{ij}a_i + \theta_j \quad (7)$$

Every $i$shows one of the units of layer L-1 connected to unit j and **θ**j represents the bias or woj.
ii) The generalized delta rule is to determine the values of the local gradients. Each weight update is shown as

$$\Delta w_{ij}(n) = \eta \delta_j a_i \quad (8)$$

And the equations of this rule are used for calculating the values are

$$\delta_j = a_j(1-a_j)(t_j - a_j) \text{ For output neurons} \quad (9)$$

$$\delta_j = a_j(1-a_j)\sum_{k=1}^{N_{t-1}} \delta k_w k_j \text{ For hidden neurons} \quad (10)$$

The $\boldsymbol{\delta j}$ for the output units is found using already available data becausethe error measure depends on the difference between the desired$\boldsymbol{tj}$ and actual$\boldsymbol{aj}$ values. For the hidden neuron, this measure is not applied. That is why the values of **j** should be back propagated layer by layer through the network(Moreira *et al*., 1995)
iii) Finally, the weights are updated.
The term momentum determines the effect of the preceding iteration on the current one. It calculates the value affecting the present iteration based on the previous one. The momentum has a basic impact on the search process to avoid oscillations in the error surface by average out gradient elements having different signs and speed up the convergence time in flat areas. Furthermore, many values of the learning rate are able to have the same convergence time.

2.5 Ensemble Models
In this study, an ensemble model is designed for water flow forecasting. It should consider the diversity of the predictors to be efficient. If similar models are combined, they give no gain. A capable ensemble should contain not only accurate classifiers but also dissimilar. If many identical models are combined, they result in no gain. So, in this study, different base learners with diversity are used, which provides different generalization, since they produce varying global and local minima, which leads to different predictions (Partalas *et al*.,2007). Three different algorithms are used as ensemble models which are support vector machine, Artificial Neural Network (back propagation algorithm).

2.5.1 Ensemble model using Voting
Ensemble models consist of many models that receive the same inputs and produce the outcome separately. All the outputs coming from the individual models are aggregated, giving an ensemble output.
This combination is called Voting, and it produces one separate strong model from many individual weak models. Figure 5 shows the design of ensemble models. (Kotu *et al*., 2014).

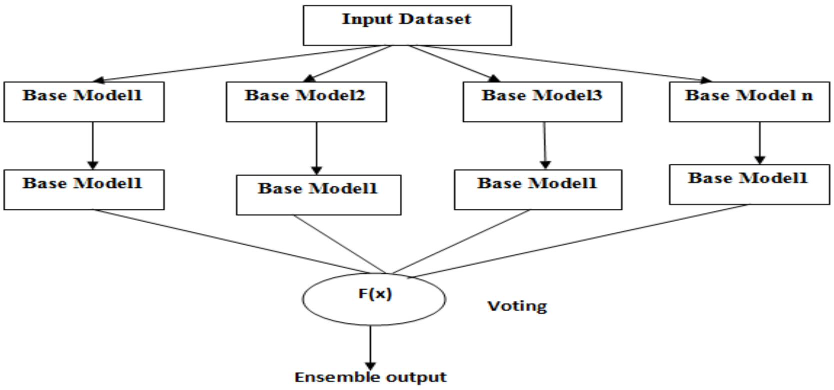

Fig 5: Frame work of Ensemble Model (Duan *et al.,* 2019)

The main drawback of voting is that all the base models use the identical data set, and the models can handle different data types.
2.5.2 Ensemble model using Bagging
Bootstrap aggregation (bagging), was proposed by Breiman (1996) for classification and regression. Bagging actually depends on bootstrap statistical resampling techniques (Efron and Tibshirani, 1993), which create different training sets for training the predictors forming an ensemble. Bagging is proved to be successful with unstable learning algorithms (Toro and et al., 2013).
In bagging, the predictors are formed by varying the training set for each predictor. For every training set T of n record, m training sets are created each has n records, by sampling with replacement. Every training set T1, T2, T3, Tm will have the same number of records as the basic training set T. Since they are sampled with replacement, they can have repeated record, which is known as bootstrapping. Every training set can be used for constructing the model after sampling. In bootstrapping, all the predictors and their results are combined to form an ensemble model. This integration of bootstrapping and aggregating is known as bagging (Kotu et al., 2014). In this study, the bagging technique is used with three algorithms artificial neural network, support vector machine, and Markov chain model.

2.6 Statistical Analysis
Data analysis in this paper is performed and implemented using Excel 2016. Correlation coefficient and accuracy are the performance measures used to measure the results in this research.

3. RESULTS
This section is divided into four parts. The first part compares the results between SVM, ANN and Markov chain. The second part is on partitioning and pruning of ANN. SVM pruning is given as the third part. Ensemble modeling is the fourth part of the results.
3.1 The comparison between the three algorithms (SVM, ANN, and Makrov chain model)
The datasets were implemented using the three models separately for river flow forecasting in the Blue Nile. Table 1 and Fig 6 show the results. The Markov chain has withan accuracy has an accuracy of 93 %; it is followed by the SVM and lastly ANN, which has an accuracy of 82 %.

Table 1: The results of the three algorithms

Citation: Atika Hussein and Johnson Agbinya.,2019. **Water Flow forecasting using Artificial Intelligence techniques and Markov chain model: The Blue Nile Scenario**. Australian Journal of Basic and Applied Sciences, 13(12): 10-18. DOI: 10.22587/ajbas.2019.13.12.3

	Markov chain	*ANN*	*SVM*
Accuracy	93%	82%	92%

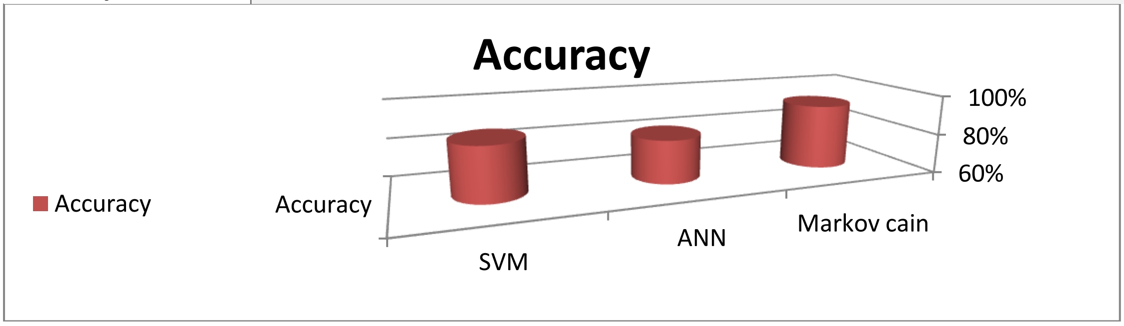

Fig 6. Comparison between the three algorithms

3.2 *Artificial Neural Network (Partitioning and Pruning)*
For Artificial Neural Network the data is divided into training and testing partitions. Many trials for the partitions have been done to improve the accuracy of the ANN and the results are shown in Table 2 and Fig 7.

Table 2: The results of Split Data Ratio (ANN)

Cases	Partition 1	Partition 2	Accuracy
1	0.3	0.7	80.7 %
2	0.4	0.60	82%
3	0.35	0.65	80 %
4	0.1	0.9	81 %
5	0.2	0.8	80%

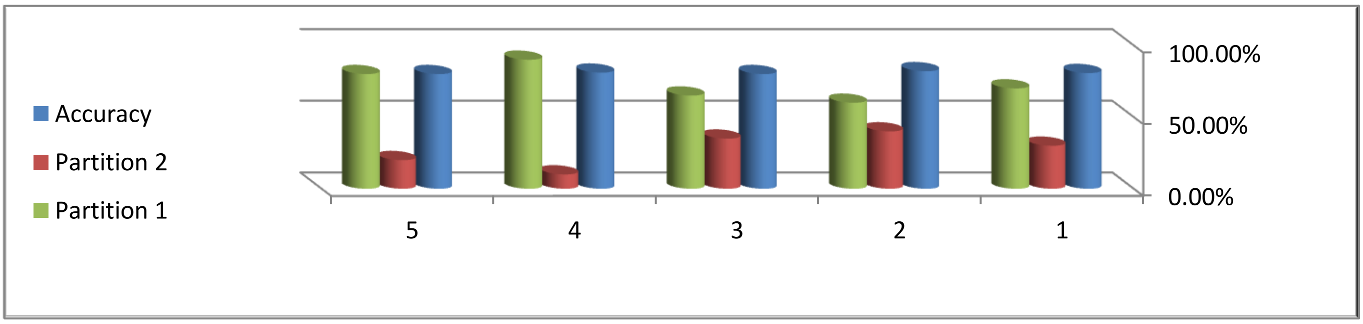

Fig 7: Accuracy obtained with different partitions of data set (ANN).
The ANN parameters are pruned using trial and error method to find the best accuracy, by changing the momentum, learning rate and the number of epochs. The results are shown in Table 3 and Fig 8.

Table 3: Changing parameters of ANN

No of epoch	Learning rate	Momentum	Time/seconds	Accuracy
500	0.4	0.3	25.5	75
500	0.3	0.3	22.45	82
500	0.5	0.3	23.64	79
500	0.1	0.3	25.56	81

Citation: Atika Hussein and Johnson Agbinya.,2019. **Water Flow forecasting using Artificial Intelligence techniques and Markov chain model: The Blue Nile Scenario**. Australian Journal of Basic and Applied Sciences, 13(12): 10-18. DOI: 10.22587/ajbas.2019.13.12.3

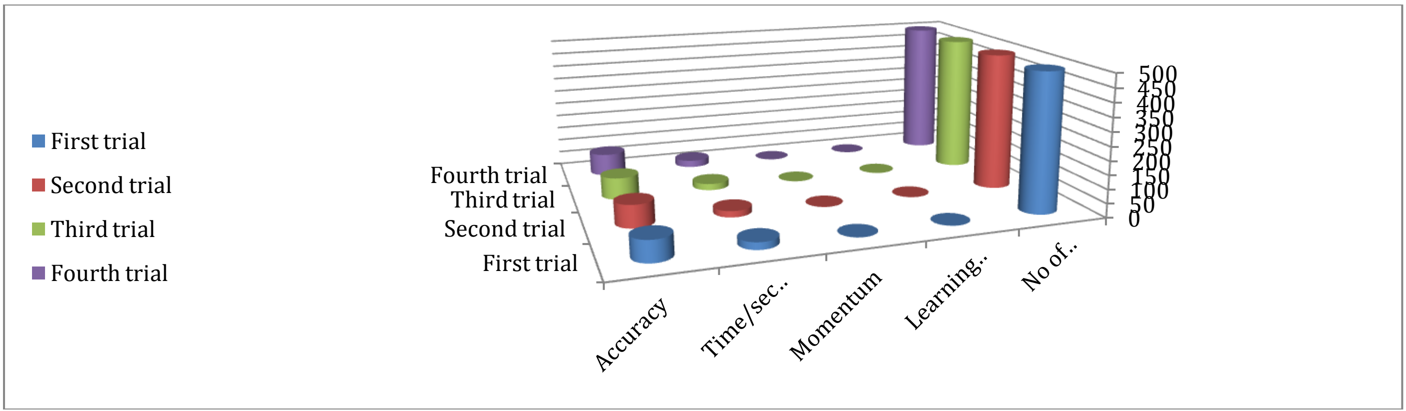

Fig 8: Changing parameters of ANN

3.3 Support Vector Regression Pruning

In this paper, all three parameters are tuned. Table 4 and Fig9 shows the c parameter is changed while γ and α are fixed until the best correlation is obtained. It also shows the processing time in seconds. Table 5and Fig 10 shows the parameterα is changed while the c and γ are fixed. Finally, Table 6 and Fig 11 show the γ parameter is changed while c and αare fixed.

Table 4: Changing the parameter c

c	γ	α	Accuracy	Time/sec
0.05	1	0.001	90	25.5
0.01	1	0.001	87	22.45
0.02	1	0.001	88	23.64
0.001	1	0.001	85	25.56
0.005	1	0.001	77	22
0.8	1	0.001	78	24
0.1	1	0.001	82	25
0.5	1	0.001	78	22
0.6	1	0.001	81	20

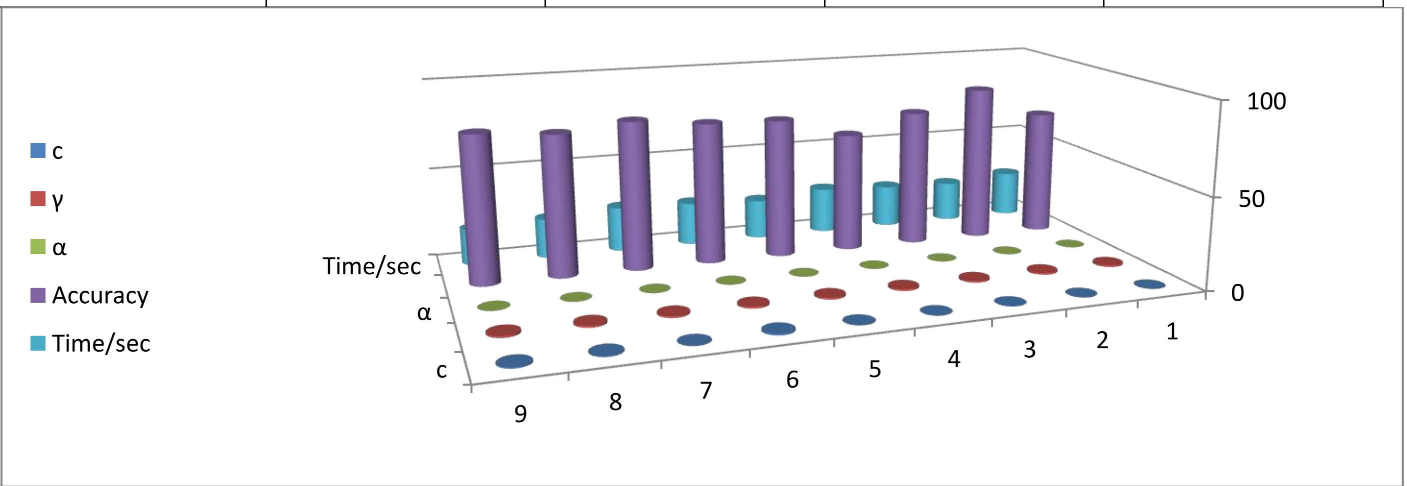

Fig 9: Changing the parameter c

Table 5: changing the parameter α

c	γ	α	Accuracy	Time/sec
0.05	1	0.002	79	23.35
0.05	1	0.003	91	22.01
0.05	1	0.01	85	20
0.05	1	0.02	89	25.12
0.05	1	0.1	88	23.19
0.05	1	0.2	91	21.36
0.05	1	0.3	90	22.01
0.05	1	0.03	91	22.1

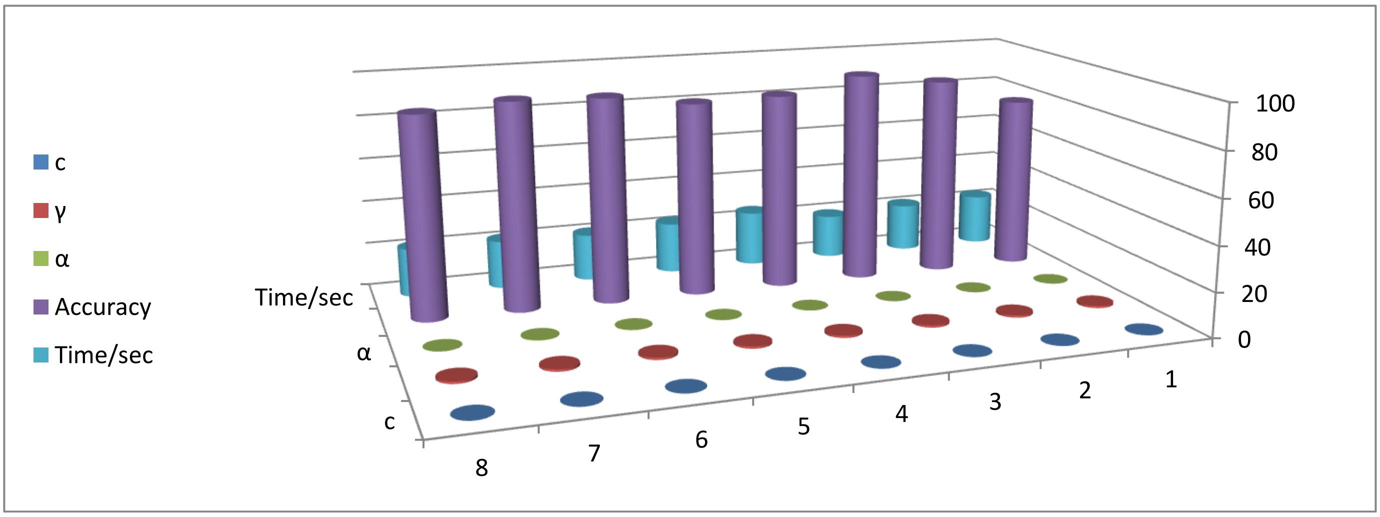

Fig 10: Changing the parameter α

Table 6: Changing the parameter γ

c	γ	α	Accuracy	Time/sec
0.05	6	0.01	88	23.35
0.05	2	0.01	83	22.01
0.05	3	0.01	80	20
0.05	1.5	0.01	78	25.12
0.05	2.5	0.01	77	23.19
0.05	1	0.01	92	21.36
0.05	5	0.01	90	22.01
0.05	7	0.01	91	22.1

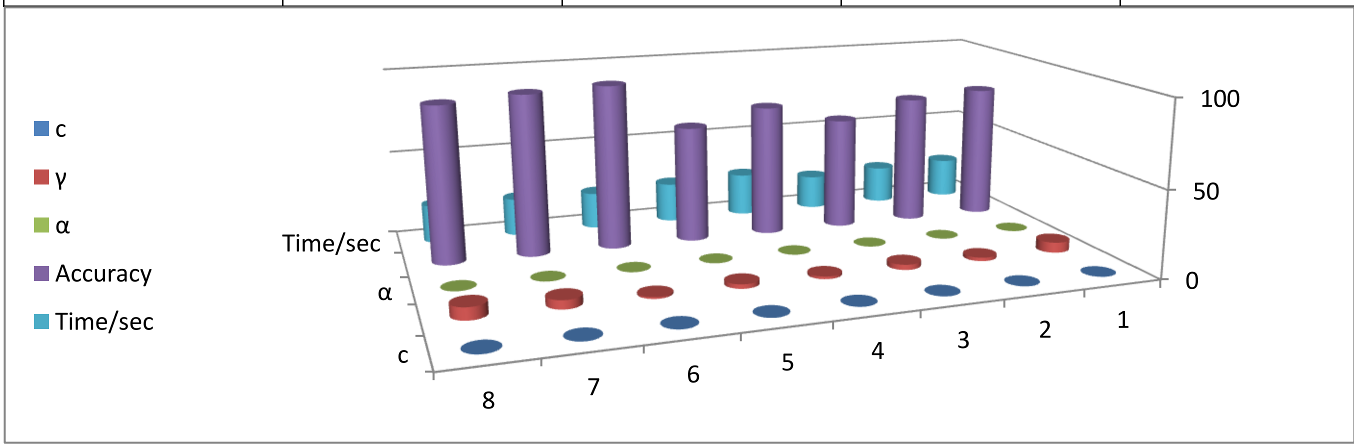

Fig 11: Changing the parameter γ

The best result for the support vector regression is obtained when fixing α to 0.01 and the value of c is 0.05 and γ is. The accuracy given is 92 %.

3.4. Ensemble models techniques (Bagging and Voting)

Bagging and Voting were both implemented on the data set using ANN and SVR. The result of the Bagging after its optimization is found to be as shown in Table 7 and Fig 11. In Bagging, the best result was obtained at SVM with the number of iterations 3 and a sample ratio of 0.6. This means that SVM using Bagging gives the highest accuracy of 95.7, which means ensemble modeling gives higher accuracy than single predictors to water flow forecasting.

Table7: Bagging Optimization

Sample Ratio	Iteration number	SVM	ANN
0.7	3	92.2	90.6
0.7	8	91.1	91.2
0.7	5	93.2	92.8
0.3	3	93.1.	93.2

0.5	3	92.9	92.3
0.6	3	95.7	94

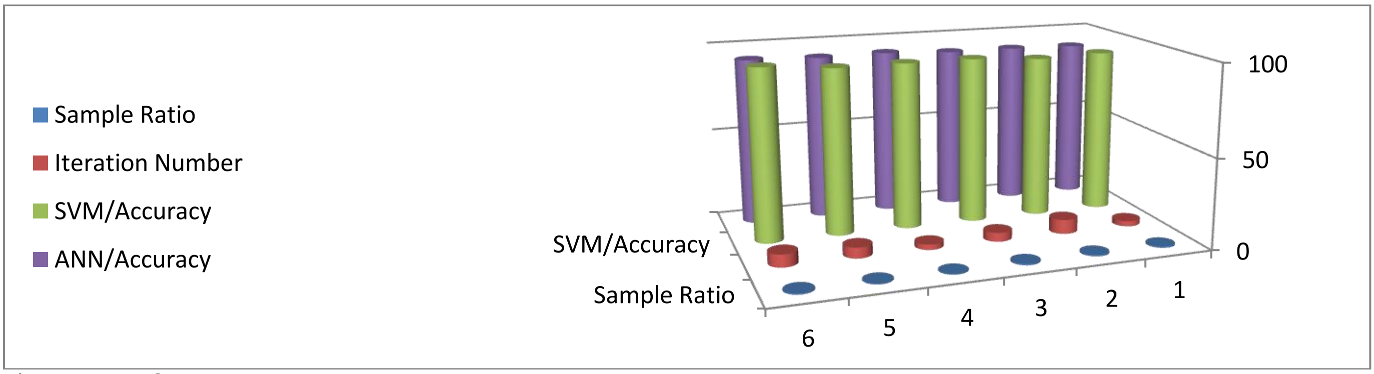

Fig 11: Bagging Optimization

Finally comparing voting and bagging using SVM, the results showed that Bagging gives higher accuracy than voting as shown in Fig 12 and Table 7.

Table 7: The results of ensemble models

	Vote	Bagging
Correlation coefficient	0.94	0.957
Accuracy	94 %	95.7 %

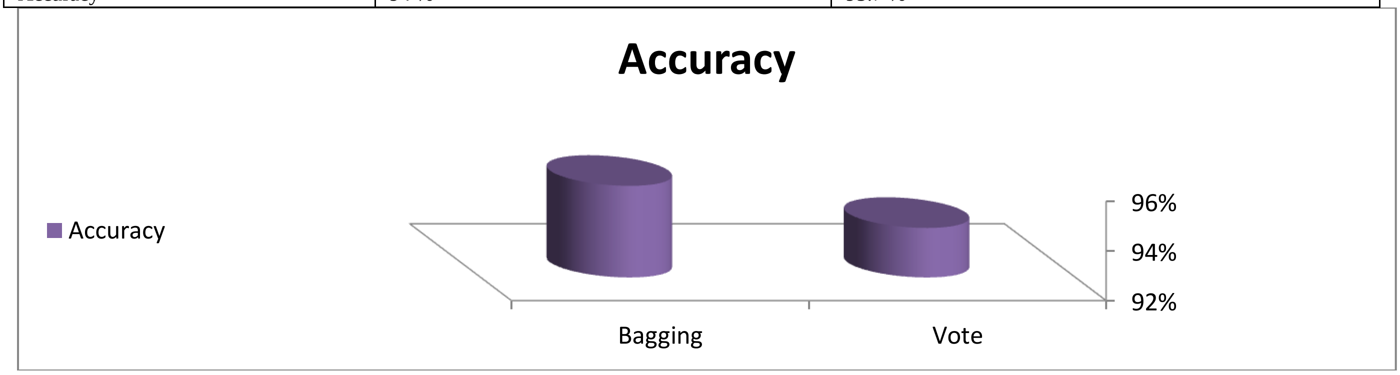

Fig 12: Comparison between ensemble models

4. DISCUSSION AND CONCLUSION

After reviewing many research publications, it is found that it is not possible to determine the best algorithm to be used in any task, or even not possible to prove which one is better than the other. The only way to handle this problem is by trying many algorithms to find which one is better. Developing an ensemble model from already tuned algorithms leads to better accuracy and results, which as shown in this paper.

As bagging and voting ensemble methods are used in this research, the Adaboost algorithm is suggested to be implemented and added as an ensemble technique in the future. It is also recommended to add more predictors to the ensemble model to increase the accuracy.

ACKNOWLEDGENT

We acknowledge the scientific, technical writing help rendered to us by our professors during the preparations, findings, and analysis of the data. We are greatly indebted to thank them for the marvelous contribution which they rendered.

REFERENCES

Asadi, S., Shahrabi, J., Abbaszadeh, P., & Tabanmehr, S. (2013). A new hybrid artificial neural networks for rainfall–runoff process modeling. *Neurocomputing*, *121*, 470-480..

Bourdin, Dominique R., Sean W. Fleming, and Roland B. Stull. "Stream flow modeling: a primer on applications approaches and challenges." Atmosphere-Ocean 50.4 (2012): 507-536.

Chen, Shien-Tsung, and Pao-Shan Yu. "Pruning of support vector networks on flood forecasting." Journal of hydrology 347.1 (2007): 67-78.

Chen, Xiao-yun, Kwok-wing Chau, and Wen-chuan Wang. "A novel hybrid neural network based on continuity equation and fuzzy pattern-recognition for downstream Daily River discharge forecasting." Journal of hydro informatics 17.5 (2015): 733-744.

Deka, Paresh Chandra. "Support vector machine applications in the field of hydrology: a review." Applied soft computing 19 (2014): 372-386.Crossref

Duan, Qingyun, et al. "Multi-model ensemble hydrologic prediction using Bayesian model averaging." Advances in Water Resources 30.5 (2007): 1371-1386.

Duan, Qingyun, et al., eds. Handbook of Hydro meteorological Ensemble Forecasting. Springer, 2019.

Galicia, A., Talavera-Llames, R., Troncoso, A., Koprinska, I., & Martínez-Álvarez, F. (2019). Multi-step forecasting for big data time series based on ensemble learning. *Knowledge-Based Systems*, *163*, 830-841..

Gong Li, Jing Shi, Junyi Zhou. "Short Term Wind Speed Forecasting Based on Bayesian Model Averaging Method", Volume 6: Emerging Technologies: Alternative Energy Systems; Energy Systems: Analysis, Thermodynamics and Sustainability, 2009.

Kisi, Ozgur, and Mesut Cimen. "A wavelet-support vector machine conjunction model for monthly streamflow forecasting." Journal of Hydrology 399.1 (2011): 132-140.

Kisi, O., Shiri, J., Karimi, S., Shamshirband, S., Motamedi, S., Petković, D., & Hashim, R. (2015). A survey of water level fluctuation predicting in Urmia Lake using support vector machine with firefly algorithm. *Applied Mathematics and Computation*, *270*, 731-743.

Kotu, Vijay, and Bala Deshpande. Predictive analytics and data mining: concepts and practice with rapidminer. Morgan Kaufmann, 2014.

Moreira, Miguel, and Emile Fiesler. Neural networks with adaptive learning rate and momentum terms. No. REP_WORK. Idiap, 1995.

Partalas, I., Hatzikos, E., Tsoumakas, G., & Vlahavas, I. (2007). Ensemble selection for water quality prediction. In *Proeedings of 10th International Conference on Engineering Applications of Neural Networks* (pp. 428-435).

Pawar, Archana, V. S. Jape, and Seema Mathew. "Wind Power Forecasting Using Support Vector Machine Model in RStudio." Cognitive Informatics and Soft Computing. Springer, Singapore, 2019. 289-298.

San, Thinn HtetHtet, and Mie MieKhin. "River Flood Prediction Using Markov Model." Genetic and Evolutionary Computing. Springer, Cham, 2016. 435-443.

Sattari, M. T., Pal, M., Mirabbasi, R., & Abraham, J. (2018). Ensemble of M5 model tree based modelling of sodium adsorption ratio. *Journal of AI and Data Mining*, *6*(1), 69-78.

Sattari, M. Taghi, et al. "M5 model tree application in daily river flow forecasting in Sohu Stream, Turkey." Water Resources 40.3 (2013): 233-242.

Singh, Gyanendra, Som Nath Sachdeva, and Mahesh Pal. "Support vector machine model for prediction of accidents on non-urban sections of highways." Proceedings of the Institution of Civil Engineers-Transport. Vol. 171. No. 5. Thomas Telford Ltd, 2017.

Tiwari, Mukesh Kumar, and Chandranath Chatterjee. "Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs)." Journal of Hydrology 382.1-4 (2010): 20-33.

Toro, Carlos H. Fajardo, et al. "A hybrid artificial intelligence model for river flow forecasting." *Applied Soft Computing* 13.8 (2013): 3449-3458.Valipour, Mohammad, Mohammad Ebrahim Banihabib, and Seyyed Mahmood Reza Behbahani. "Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir." Journal of hydrology 476 (2013): 433-441.

Vapnik, Vladimir. The nature of statistical learning theory. Springer science & business media, 2013.

Xiaofan Liu, Liliang Ren, Fei Yuan, Bang Yang. "Meteorological Drought Forecasting Using Markov Chain Model", 2009 International Conference on Environmental Science and Information Application Technology, 2009.

Zhou, Zhi-Hua, Jianxin Wu, and Wei Tang. "Ensembling neural networks: many could be better than all." Artificial intelligence 137.1-2 (2002): 239-263.