JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 06 Issue 02
2024

Research Article
https://doi.org/10.56979/602/2024

Impact Assessment of Antecedent Hydro-meteorological Parameters Data on the Performance of Support Vector Machines (Regression) Based Stream Flow Prediction Model Integrating Genetic Algorithm

Ateeq-ur-Rauf[1], Naveed Jan[2], Laeeq Ahmed[3], Asif Nawaz[4], Shahid Latif[5], Shahzad Hameed[6], and Sheeraz Ahmed[7*]

[1]Department of Civil Engineering, University of Engineering & Technology, Peshawar, 25000, Pakistan.
[2]University of Technology Nowshera, 24100, Pakistan.
[3]Department of Computer Science and IT, University of Engineering & Technology, Peshawar, 25000, Pakistan.
[4]Faculty of Electronics, Higher Colleges of Technology, Dubai, UAE.
[5]Department of Electrical Engineering, Iqra National University, Peshawar, 25000, Pakistan.
[6]Institute of Computer Science and IT, FMCS, The University of Agriculture, Peshawar, 25000, Pakistan.
[7]Department of Computer Science, Iqra National University, Peshawar, 25000, Pakistan.
*Correspondence Author: Sheeraz Ahmad. Email: sheeraz.ahmad@inu.edu.pk

Received: November 11, 2023 **Accepted:** February 01, 2024 **Published:** March 01, 2024

Abstract: Accurate and reliable river flow predictions are obvious for appropriate planning, development and management of water resources, particularly for a country like Pakistan where cultivation is mostly by canal irrigation system. This is particularly important for sustainable socio-economic growth, proper management of the canal system and flood mitigation under changing climatic conditions. In this study, thirty years' (1985 – 2014) monthly temperature, precipitation and streamflow data from Astore sub-basin of the Upper Indus River Basin, UIRB in Pakistan have been analysed. The streamflow of the Astore River, which is a tributary of the Indus River, is predicted ahead of time, considering the impact of antecedent precipitation, the temperature and streamflow data. During the recent past decades, artificial intelligence-based modeling with several categories of models has been presented as an important technique for the prediction of hydrological phenomenon. In this paper, the performance of four Support Vector Machines Regression (SVR) models have been probed to predict the streamflow of Astore River. The Four SVR model types were compared on the basis of radial basis function, polynomial, linear and sigmoid kernels. Number of input combinations with input variables (temperature, precipitation, and streamflow) with reference to time lag were determined by Genetic Algorithm test. The best input combination for SVR models was identified using a genetic algorithm upon the bases of the smallest values of gamma and Standard Error. The Nash-Sutcliffe efficiency and Mean Bias error were used to evaluate the performance of SVR Models. The SVR model, based on radial basis function kernel forecasted the stream flows with higher accuracy as compared to the other kernels.

Keywords: Water resource management; Genetic Algorithm; Short-term Streamflow Forecast; Support Vector Machine.

1. Introduction

Adequate planning and development of water resources possess crucial importance to a region as it affects many important areas i.e. hydropower, hydraulic structure design, irrigation system, river improvement, the agricultural yield, food security, economy and many other important areas of life. Streamflow simulation has an important role in sustainable water resources planning and management. The non-linear, high dimensional nature of such simulations makes their rendering a complicated process [1]. One way is to predict streamflow using numerous modeling techniques such as black box models, stochastic models, distributed physical models and lumped conceptual models [2].

For a country like Pakistan where the rainfall is uneven, region to region and timely therefore the availability of water for crops in times of water scarcity is of extreme importance for its agriculture needs and hence demands for crop water normally increases at times when the rainfall recedes or when there is no rainfall. This needs to focus more for a better management of the water resources of the country to boost the country's economy and ensure the food security. Also, a cheaper electricity production is a vital need of the country which is only possible to shift all the thermos power generation to hydropower production. All these issues can be dealt with to a large extent through precise water resources management which in turn depends upon correct prediction of stream flows [3]. Runoff prediction is a complex process to be predicted due to its non-linear, multi-dimensional dynamics [4]. A new dimension has been introduced in hydro meteorological prediction using artificial intelligence based data driven modeling techniques for the identification of input models [5]. This new dimension is useful in other predictions such as solar radiation estimation [6] and [7], wind speed modeling [8] and land use classification [9]. Data driven modeling is an effective tool for the prediction of daily streamflow. Over the years, several data driven techniques have been used to predict the streamflow, which includes Support Vector Machines (SVM), Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Genetic Programming (GP), Crisp Distributed Artificial Neural Networks (CDANN), Model Trees (MT) are effective tools for the prediction of streamflow. Previously a lot of work has been done by various researchers using the above techniques e.g. [10], predicted floods using the SVM, similarly [11], applied SVM to predict the rainfall and runoff, and [12] applied Support Vector Regression (SVR) to predict floods in real time. Additionally, [13] applied the SVM model to foresee a-day ahead streamflow and then compared the results with hybrid techniques of ANN (ANN integrated with genetic algorithms ANN-GA) and results showed that SVM based prediction models outperform the rest with relatively high degree of accuracy. [14] used modified SVM for prediction of streamflow of the Shihmen Reservoir, Taiwan. [15] used SVM to predict streamflow for ungauged sites. [16] used SVM to forecast the river flows. [17] used SVM to predict water levels in lakes. [18] used hydro meteorological data sets to predict the inflow in Tarbela reservoir using ANN and regression techniques.

The ANN and SVR models have shown accurate results when solving high dimensional non-linear problems such as streamflow simulation. As such, recently, the ANN and SVR models have been widely and successfully used to perform hydrologic and streamflow forecasting [19], [20] and [21]. [22] and [23] presented streamflow simulation for high altitude catchments in Pakistan. In another study [24], [25], floods were predicted using SVR models. The SVR models are classified into various types comprising the techniques for the selection of inputs, different parametric optimizations and training processes. Finding the optimal SVR model type for a particular problem is a hard exercise. This study facilitates scientists, engineers and researches to select an accurate model to predict streamflow, comparing the performance of a few such techniques.

Data sensitive machine learning models can be applied to particular sites and are also generally standardized for a particular set of data. Therefore, there is a lot of research gap and the models need to investigate further, while working with different record lengths. Similarly the performance of models with a variety of data sets in terms of complexity and sizes need to be investigated. The work will surely be useful to researchers and engineers in this area of research as such methods have been sparsely applied for the predictions of stream flow in the past, especially in the selected study area of the Indus Basin. For stream flow predictions for various time lags (t in months), the monthly observed data for precipitation ((P(t), P(t-1), P(t-2), P(t-3), P(t-4), P(t-5)), temperature (T(t), T(t-1), T(t-2), T(t-3), T(t-4), T(t-5)) and discharge (Q(t), Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5)), used being the inputs whereas the output variable was the streamflow (Q(t+1)). The most essential modeling objective was to find the ideal combination of inputs. (Bray and Han 2004) urged that the vast amount of inputs makes the process of developing the model, a complex activity and therefore, a more robust method is needed to find the best possible input combination.

Here, the genetic algorithm (GA) is utilized for choosing the best possible combination of inputs for the Support Vector Regression models.

Although regarding the time lag, a certain correlation always exists between the input and output, a number of varying input combinations can also be there including the prime input variables (precipitation, temperature, evaporation, streamflow, river stage etc.). Previous studies indicates that three of these variables i.e. temperature T, precipitation P and streamflow Q with respect to time, t had been rarely used simultaneously for the development of input combinations. Mostly the precipitation and streamflow have

been used together and, in some studies streamflow, has been used as a sole input parameter. [26], [27], [28], [23], [19] and [20]. As an example, [26] used runoff depth and rainfall as input variables. [27] used precipitation, temperature and streamflow as input variables. [19], [23] and [28] used precipitation and streamflow as input variables whereas water level and precipitation were used as input variables by [29]. [30], [31], [32], [20] and [33] used a unary variable streamflow as input.

Here in this work, the streamflow simulation results are compared, computed using three main input variables, precipitation, temperature and streamflow, simultaneously in a single input combination. Monthly data of the past three decades (1985 to 2014) have been used for this study.

2. Materials and Methods

2.1 Study Area

One of the world's biggest basins lying in Pakistan is the Indus River Basin (IRB) surrounding about 970,000km^2 of area. The Upper Indus River Basin (UIRB) is a sub-catchment of Indus River Basin (IRB) and it spreads from the origin of Indus River towards the first water reservoir at Tarbela with a total coverage area of approximately 175,000 km2 [34]. The River Indus is the biggest river in Pakistan and annually provides 75% of water to irrigation canals and 80% to the hydro-power generation projects [34]. Three largest mountain ranges, the Karakorum, the Himalaya and the Hindukush also exist in the UIRB. The three mountain ranges meet at the juncture located about 40 kilometers from the Gilgit city. Among all the world's glaciers, UIRB glaciers comes at the second spot after polar region as the largest glaciers. Glacier melt and precipitation over UIRB are the major waterflow sources of the Indus River and its tributaries. The major tributaries of UIRB includes Gilgit, Skardu, Astore, Gupiz and Drosh rivers. All the plain areas of the Lower Indus region throughout towards the Arabian Sea depends upon the flows of River Indus for their groundwater resources.

In order to study the Upper Indus Basin Cryosphere, Water and Power Development Authority (WAPDA) of Pakistan has undertaken various initiatives. During 1960s, WAPDA established Hydro-meteorological networks to monitor the Cryosphere through its Surface Water Hydrology Project, [refer Figure 1]. In this study, precipitation (P), streamflow (Q) and temperature (T) data for 30 years (1985 to 2014) was considered. The data was collected from Astore climate and streamflow gauging station (as illustrated in Figure 1)

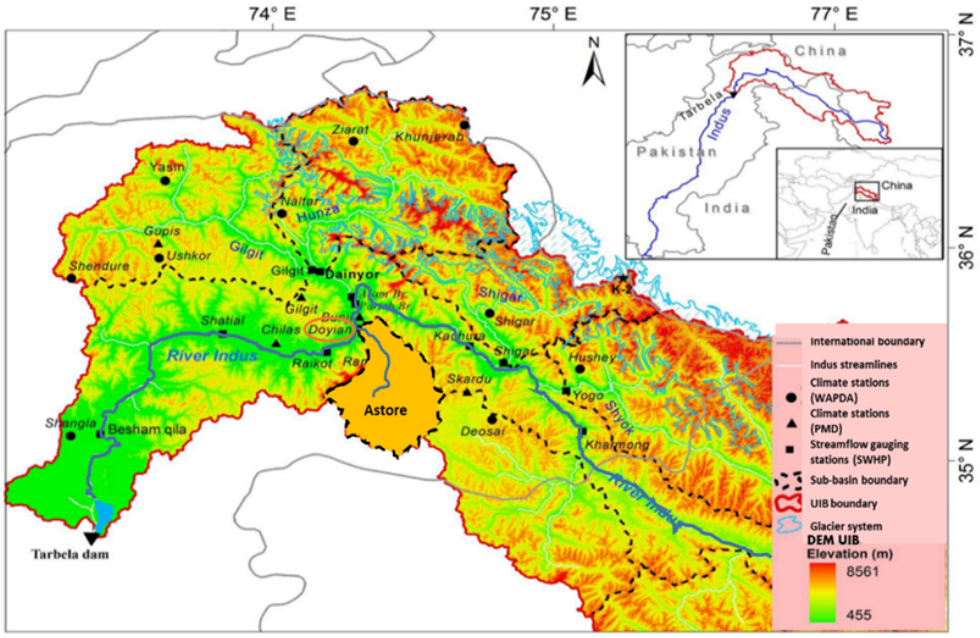

Figure 1. The figure illustrates the study area of this work, mentioning climatic and river gauging stations of the Astore sub basin of the Upper Indus Basin, Pakistan.

2.2 Genetic Algorithm (GA)

A biologically inspired algorithm known as Genetic Algorithm, conforms to genetic norms by generating numerous input combinations. The best input combination, with least complex SVR models and least output error rate, is selected. GA is basically an optimization algorithm where it tries to find

improved solutions by performing a global search. It has a fixed sized population of individuals (inputs and weights), known as generation. Through an iterative process, in each iteration, the algorithm selects new individuals or improves the individuals to form a new generation, eventually reaching to improved solutions. Following this methodology, the GA has the ability to choose the appropriate feature subsets and the learning rate. Further details regarding this method can be found at [35], [36]. In order to run the GA algorithm, the Win Gamma software was used [37].

2.2.1 Model Input Selection Using Genetic Algorithm Test (GA)

The most and the least influential parameters on the model predictions have been identified in this work. As mentioned before, the most important steps for data driven machine learning model development is the selection of the right set of input variables. A right selection of input variables leads to a model with increased efficiency whereas a poor selection of input variables can cause overfitting. An overfitted model overreacts to small variations in the training data and as such depicts inferior predictive performance.

According to the theory of Genetic Algorithms, the input combination with the least values of Standard Error (SE) and Gamma is considered to be the best combination. A low value of SE and gamma show that the model is more likely to have better results determined from the developed models using the data with the given input combination. Generally, input combinations with low SE and gamma are quite rare, hence special care is required when selection of the best input combinations is performed. For modeling purposes, the monthly data points for discharge, temperature and precipitation were used for the year 1985 to 2014 for the sub-basin Astore. The units of almost all the input variables were different, hence the data was made smooth and uniformed by normalizing it between 0 and 1 using Microsoft Excel. Once normalization was complete, the data was saved as a CSV (Comma Separated variable) file, so it can be imported to win-Gamma application [37]. In the win-Gamma application, the Genetic Algorithm test was performed to obtain the superior input combinations. The following variables can be modified using GA, however, default values provided by the software were applied in this study.

- Population size: Number of masks used in the current generation (default value =100)
- Mutation Rate: Probability of single bit mutation in each generation. (default values = 0.05)
- Crossover Rate: Probability of crossover in each generation (default values = 0.5)
- Gradient Fitness: The fitness function weight utilized for the masks in the Gamma Test having a lower gradient. (default value = 0.1). By increasing this weight further simplicity can also be achieved.
- Intercept Fitness: The fitness function weight utilized for the masks having a low Gamma statistic absolute value. (default value = 1.0). More Accuracy is obtained by increasing this weight.
- Length Fitness: The fitness function weight utilized for masks having a specific number of '1's. (default value = 0.1). Masks having small numbers of '1's have a higher selection probability and by increasing this weight simpler models can be obtained.
- Run Time: For a particular chosen GA, a longer run time enables a bigger population which eventually leads to a better fitness of the most suitable mask. (default value = 5 min). Several hours of run time is required for GA with longer masks (having large numbers of inputs) and bigger data sets.

Using the Genetic Algorithmic simulations, 100 possible input combinations were found and then the 10 best input combinations were chosen based on the smallest Standard Error (SE) and Gamma (Γ) values. Eventually, among these ten input combinations the one having the smallest Gamma value was chosen for the analysis as the best input combination. (Figure 2(b)).

Gamma value (Γ): The Gamma Γ is an estimate of that part of the output variance that cannot be attributed for by the presence of a smooth data model. Infact, the Gamma is the y-intercept of the regression line. [38]

2.3 Support Vector Machines

Support Vector Machines (SVM) (Bray and Han, 2004) is comparatively a new tool in artificial intelligence based on supervised learning that analyzes data and recognizes patterns. It can be applied successfully to classification tasks such as Pattern Recognition (PR) and Optical Character Recognition (OCR) but the most recent success was when applied to regression and time series analysis (Xie, 2009). SVM is a two layered network i.e., in first layer the weights are non-linear while for the second layer, the weights are linear. [39]. Figure 2 (a, b) shows the flow chart and a general structure of SVR model.

In statistical learning process, the basic mathematical function is represented by equation "(1)".

$$y = f(x) = \left[\sum_{i=1}^{N} \alpha_i K(x_i, x) \right] - b \tag{1}$$

For the case of SVM, nonlinear conversion in usually carried out by φ(x) and linearly weighted sum of M is the output. decision function is given as

$$y = f(x) = \left[\sum_{i=1}^{N} \alpha_i K(x_i, x)\right] - b \quad (2)$$

In "(2)" N represents the number of training data points, xi are vectors used in training process while x is an independent vector. Where αi and b are parameters that have been derived by maximizing the objective function. An important variable in "(2)" is the kernel function K, which simplifies the learning process to a higher dimensional feature space from the input space. Four standard types of kernels namely linear, polynomial, sigmoid and radial basis are commonly applied, and these are given as

Linear: $u' \times v$

Polynomial: $\left(\gamma \times u'v + coef\right)^{\deg ree}$

Sigmoid: $\tanh\left(\gamma \times u'v + coef\right)$

Radial basis: $e^{-\gamma|u-v|^2}$

2.3.1 Support Vector Machine (SVM) Modelling

The Epsilon-SVR model for predicting the stream flow was used for models development. The Epsilon-SVR model kernels i.e., Linear, Polynomial, Sigmoid and Radial Basis Function (RBF) were considered for the analysis. A number of SVM implementations are available, LIBSVM [40] was used in this study, supported by the National Science Council of Taiwan. SVM modelling was performed using MATLAB version R2013a. For predicting stream flows (Qt+1), the "tolerance of termination criterion (e)" and "the parameter C (c)" were set to 0.1 and 2 respectively for the Linear, Polynomial, Sigmoid and RBF kernels modelling. The coding for both the training and testing stages was done in MATLAB, for the best selected input combination.

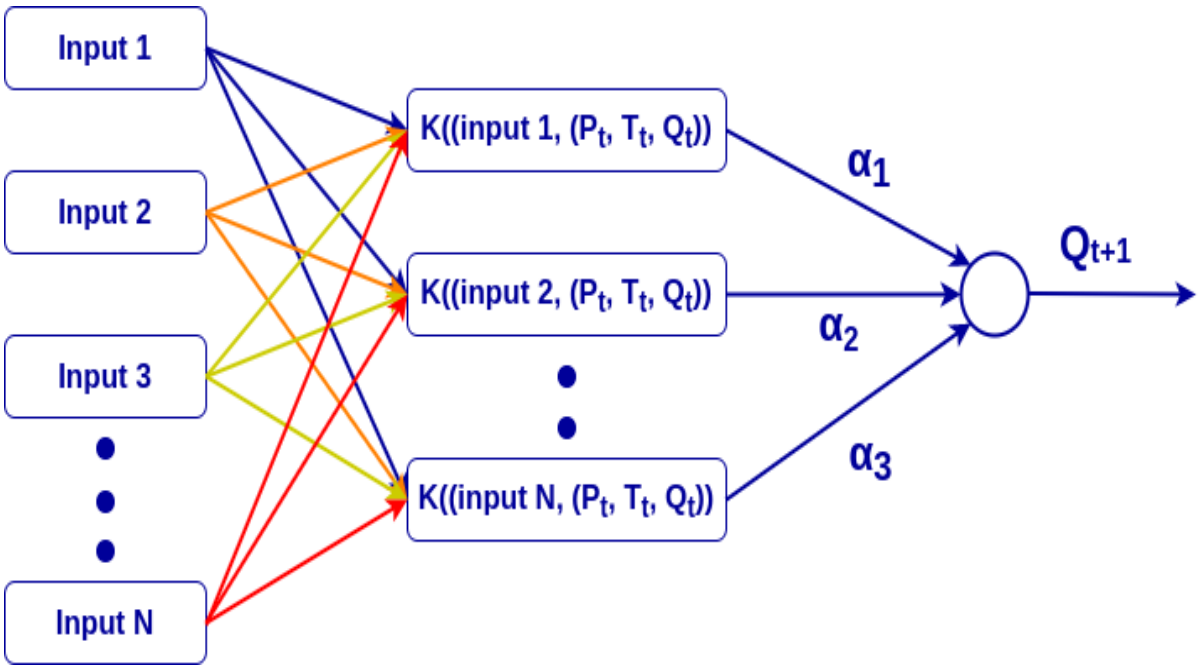

Figure 2. (a) Schematic diagram of support vector regression structure illustrating input and output parameters.

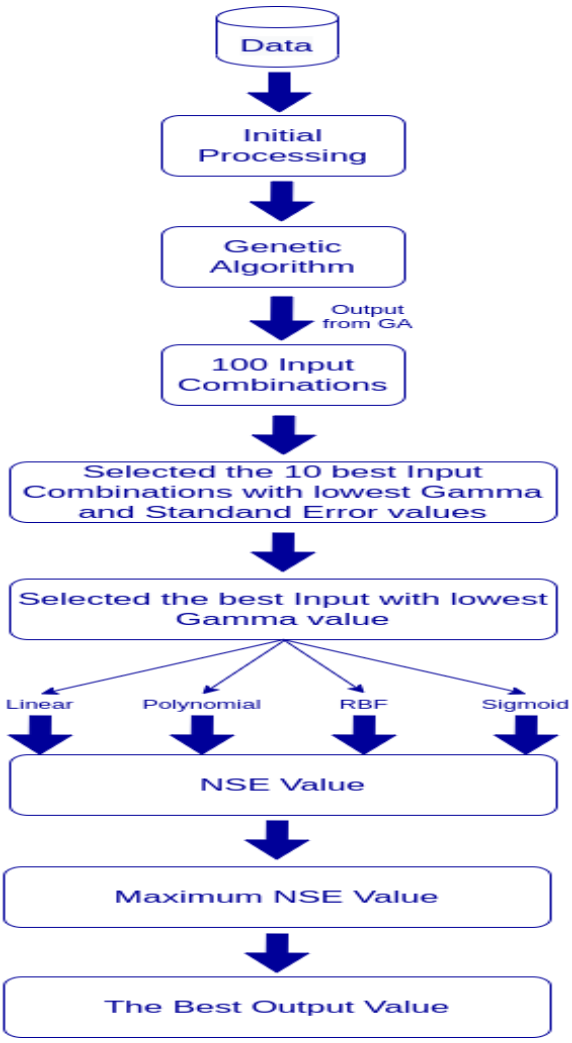

Figure 2. (b)The methodology framework of the GA-SVR modeling for selecting the best input combination and output prediction based on performance indices.

2.4. Evaluating Model Performance

Table 1. Provides a summary for the adopted model's performances found using statistical parameters.

Indices	Value	Classification of performance	Reference
Nash–Sutcliffe model efficiency coefficient (NSE) $NSE = 1 - \frac{\sum_{i=1}^{n}(Q_i^o - Q_i^p)^2}{\sum_{i=1}^{n}(Q_i^o - Q_{avg})^2}$	0.75 - 1.00 0.65 - 0.75 0.50 - 0.65 0.4 - 0.50 0.4 ≥ NSE	Very Good Good Satisfactory Acceptable Unsatisfactory	Boskidis, et al., (2012) Moriasi, et al., (2007)
Mean Bias Error (MBE). $MBE = \sum_{i=1}^{n}\frac{(Q_i^p - Q_i^o)}{n}$	MBE > 0(Positive) MBE< 0(negative)	Over Estimated Predictions Under Estimated Predictions	Ines & Hansen (2006).

In the given equations, Q_i^p , Q_i^o and Q_{avg} Represents the predicted, observed and the average observed stream flows respectively whereas the total number of input samples are represented by n.

3. Results and Discussion

3.1 Combinations of Inputs using Genetic Algorithms

For such an input type where all the three parameters are taken, the GA identified a number of different input combinations which are presented in table 2 whereas SE and Ƭ variations are presented in Figure 3. It was found that the 111100011100111000 (the ten input values as Pt, Pt-1, Pt-2, Pt-3, Tt-1, Tt-2, Tt-3, Qt, Qt-1, Qt-2) for a single output (Qt+1) was the best choice of the available input variables. The selection of this combination was based on the least Gamma and SE values for analysis. For the selected combination, the precipitation during time delay of the running month (t) and temperature with three previous months (t-1, t-2, t-3) have shown an effect on streamflow Qt+1, however in case of streamflow, only two previous values and and one running month value i.e. (Qt , Qt-1, Qt-2) have shown their influence on Qt+1. It had been suggested in a study that for high stream flows, variation in precipitation is the key parameter. In the weather and basins having significant snowmelt, temperature is the prime factor for the predictions of streamflow. [41] The flows in UIRB stream comes mostly from glacier melts, groundwater and from rain flows. Therefore, best input combination which is developed by the GA test is logical.

Table 2. Ten selected input combinations on basis of lowest Gamma (Ƭ) and Standard Error (SE) values developed by GA simulations.

Input combinations	Mask	(Ƭ)	SE
*$P_{(t)}$, $P_{(t-1)}$, $P_{(t-2)}$, $P_{(t-3)}$, $T_{(t-1)}$, $T_{(t-2)}$, $T_{(t-3)}$, $Q_{(t)}$, $Q_{(t-1)}$, $Q_{(t-2)}$	111100011100111000	0.001231	0.00145
$P_{(t)}$, $P_{(t-1)}$, $P_{(t-2)}$, $P_{(t-3)}$,$T_{(t)}$, $T_{(t-1)}$, $T_{(t-2)}$, $T_{(t-3)}$, $T_{(t-4)}$, $Q_{(t)}$, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	111100111110111100	0.001313	0.00122
$P_{(t)}$, $P_{(t-2)}$, $P_{(t-4)}$, $P_{(t-5)}$, $T_{(t)}$, $T_{(t-1)}$, $T_{(t-2)}$, $T_{(t-5)}$, $Q_{(t)}$, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	101011111001011100	0.001338	0.00134
$P_{(t)}$, $P_{(t-2)}$, $P_{(t-4)}$, $P_{(t-5)}$, $T_{(t)}$, $T_{(t-1)}$, $T_{(t-2)}$, $Q_{(t)}$, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	101011111000111100	0.001432	0.00153
$P_{(t)}$, $P_{(t-2)}$, $P_{(t-4)}$, $P_{(t-5)}$, $T_{(t-1)}$, $T_{(t-2)}$, $T_{(t-3)}$, $T_{(t-5)}$, $Q_{(t)}$, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	101011011101111100	0.001446	0.00118
$P_{(t)}$, $P_{(t-2)}$, $P_{(t-4)}$, $P_{(t-5)}$, $T_{(t-1)}$, $T_{(t-2)}$, $T_{(t-3)}$, $T_{(t-4)}$, $T_{(t-5)}$, $Q_{(t)}$, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	101011011111111100	0.001489	0.00085
$P_{(t)}$, $P_{(t-2)}$, $P_{(t-4)}$, $P_{(t-5)}$, $T_{(t)}$, $T_{(t-2)}$, $T_{(t-3)}$, $T_{(t-4)}$, $T_{(t-5)}$, $Q_{(t)}$, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	101011101111111100	0.001495	0.00058
$P_{(t)}$, $P_{(t-2)}$, $P_{(t-4)}$, $P_{(t-5)}$, $T_{(t)}$, $T_{(t-1)}$, $T_{(t-2)}$, $T_{(t-5)}$, $Q_{(t)}$, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	101011111001111100	0.001497	0.00071
$P_{(t)}$, $P_{(t-2)}$, $P_{(t-4)}$, $P_{(t-5)}$, $T_{(t-1)}$, $T_{(t-2)}$, $T_{(t-3)}$, $T_{(t-5)}$, $Q_{(t)}$, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	101011011101111100	0.001498	0.00069
$P_{(t)}$, $P_{(t-2)}$, $P_{(t-4)}$, $P_{(t-5)}$, $T_{(t-1)}$, $T_{(t-2)}$, $T_{(t-3)}$, $T_{(t-4)}$, $T_{(t-5)}$, $Q_{(t)}$, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	101011011111111100	0.001510	0.00132

*** The best combination is in bold.**

Figure 3. Gamma and Standard Errors found in the data in response to the combinations of inputs developed by performing the Genetic Algorithm test.

3.2 Results of SVR Models

The Figures 4 – 6 represents the comparison of the results, obtained by the analysis of the four selected SVR kernel models for which the input combinations were selected by Genetic Algorithm tests. Table 3 shows the various indicators (i.e., NSE, and MBE) applied in the analysis to measure the performance ratings of the various SVR Kernel Models. SVR with RBF kernel having maximum value of NSE (minimum

error) and minimum MBE showed best results when for a single output a combination of ten inputs were applied. SVR -RBF Kernel have good performance during the training and testing phases. Results suggested that the SVR-Linear kernel proved better performance as compared to Polynomial and Sigmoid kernels. The SVR-Polynomial kernel model stood inferior of all the four models. The SVR-RBF kernel model also had the best correlation coefficient (R2) shown in figure 6. The result obtained by the Epsilon-SVR models for validation phase are presented graphically (hydrographs) for the predicted stream-flows (Qt+1) and observed stream-flows (Q) in figure 5.

Table 3. Comparisons of results among four SVR models

SVM Models	Calibration		Validation	
	NSE	MBE	NSE	MBE
Linear Kernel	0.89	-9.54	0.88	-10.68
Polynomial Kernel	0.76	-28.57	0.72	-34.38
RBF Kernel	**0.93**	**-4.23**	**0.90**	**-5.83**
Sigmoid Kernel	0.86	10.68	0.85	-12.76

*__Best result is bolded__

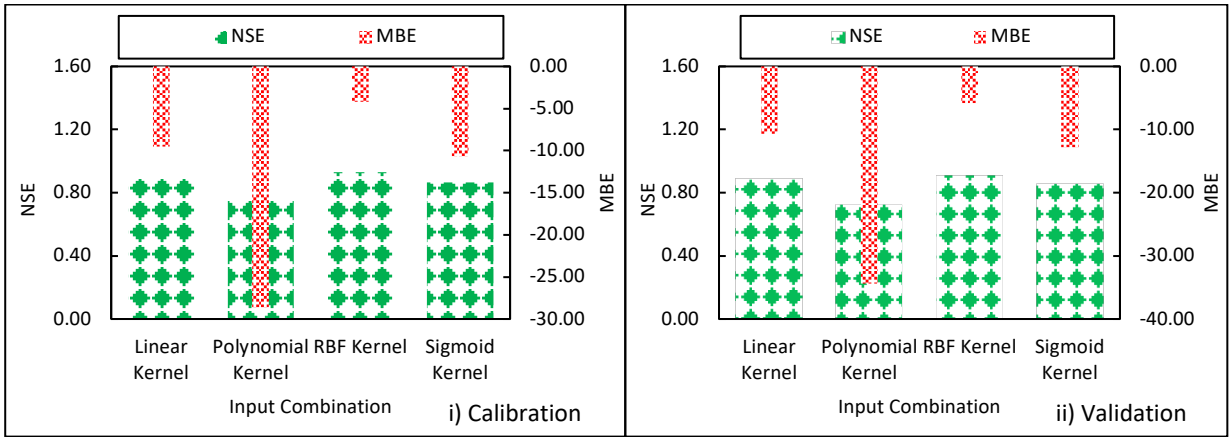

Figure 4.. Comparison of performance indices i.e. Mean Bias Error (MBE) and Nash–Sutcliffe model efficiency coefficient (NSE) for the SVR Kernel models both for i) Calibration and ii) Validation phases.

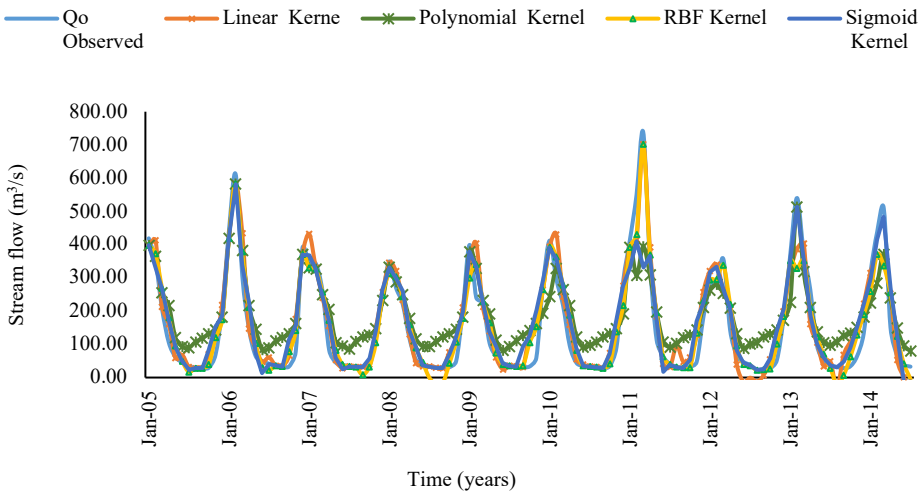

Figure 5. Observed stream flow comparison with the stream flow predicted by Support Vector Regression models using Genetic Algorithm Test for a validating phase of 10 years from 2005 to 2014.

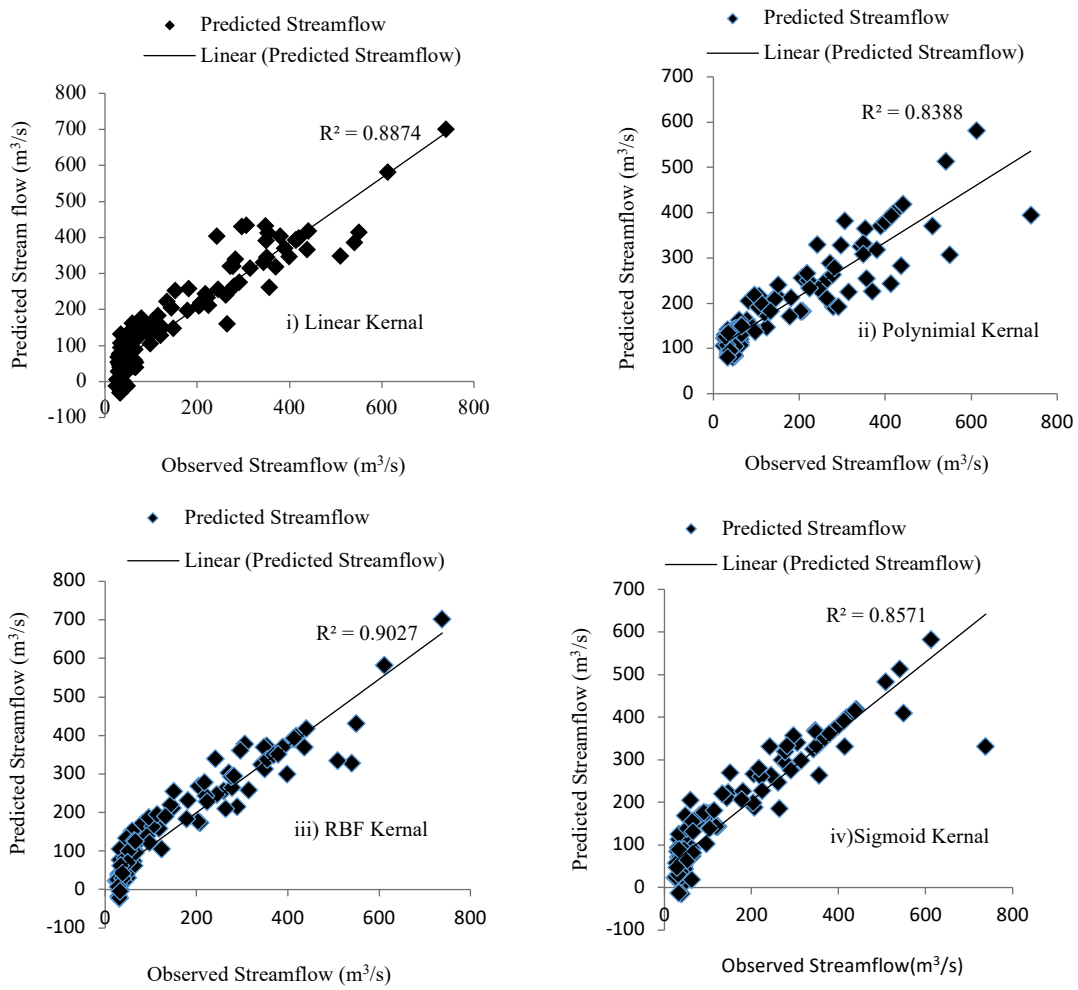

Figure 6. Scatter plot for Observed discharge values to the predicted discharge values using the best Epsilon-SVR models demonstrating the values of R^2 (coefficient of determination) elaborating the effective results of the model developed for validation phase

4. Conclusion

In this work, four different SVR kernels (i)Linear, (ii)Radial Base Forecast, (iii)Polynomial and (iv)Sigmoid were utilized. The RBF kernel outperformed the other three kernels during each phase i.e., training and testing phase. The outperforming results of statistical parameters show that SVM has a good predictive ability in the field of hydrology which can be utilized in watershed and river management. Antecedent rainfall and antecedent discharges can be used as input for the prediction of streamflow. Also, the results indicate that it is not necessary that a greater number of data input combinations will yield good results as it has been seen that the best results have been produced by the above stated combination consisting only ten inputs as compared to the total selected seventeen inputs.

Author Contributions: Research and Analysis were done by Dr. Ateeq-ur-Rauf, the Graphs and figures were plotted by Dr. Laeeq Ahmad while Dr. Sheeraz Ahmad helped in computer programming. The original article was written by Dr. Ateeq-ur Rauf while Dr. Shahid Latif and Mr. Naveed Jan did the proof reading and formatting of the entire article. All authors contributed equally through out the manuscript.

References

1. Ghumman, A. R., Al-Salamah, I. S., AlSaleem, S. S., & Haider, H. (2017). Evaluating the impact of lower resolutions of digital elevation model on rainfall-runoff modeling for ungauged catchments. Environmental monitoring and assessment, 189(2), 54.
2. Jonsdottir, H. (2006). Stochastic modelling of hydrologic systems. Technical University of Denmark.
3. Brooks, K.N., Ffolliott, P.F., Gregersen, H.M., DeBani, L.F., 2003. Hydrologyand the Management of Watersheds. Iowa State Press, Ames, IA.
4. Remesan, R., Shamim, M. A., Han, D., & Mathew, J. (2009). Runoff prediction using an integrated hybrid modelling scheme. Journal of Hydrology, 372(1-4), 48-60.
5. Remesan, R., Bray, M., & Mathew, J. (2018). Application of PCA and Clustering Methods in Input Selection of Hybrid Runoff Models. Journal of Environmental Informatics, 31(2).
6. Remesan, R., Shamim, M. A., & Han, D. (2008). Model data selection using gamma test for daily solar radiation estimation. Hydrological processes, 22(21), 4301-4309.
7. Shamim, M. A., Bray, M., Remesan, R., & Han, D. (2015). A hybrid modelling approach for assessing solar radiation. Theoretical and Applied Climatology, 122, 403-420.
8. Ishak AM, Remesan R, Srivastava PK, Islam T, Han D (2013) Error correction modeling of wind speed through hydro-meteorological parameters and mesoscale model: a hybrid approach. Water Resource Manag 27:1–23
9. Srivastava, P. K., Han, D., Rico-Ramirez, M. A., Bray, M., & Islam, T. (2012). Selection of classification techniques for land use/land cover change investigation. Advances in Space Research, 50(9), 1250-1265.
10. Liong SY, Sivapragasam C. (2002). Flood stage forecasting with support vector machines. Journal of the American Water Resources Association 38(1): 173–186.
11. Sivapragasam, C., Liong, S. Y., & Pasha, M. F. K. (2001). Rainfall and runoff forecasting with SSA–SVM approach. Journal of Hydroinformatics, 3(3), 141-152.
12. Yu, P. S., Chen, S. T., & Chang, I. F. (2006). Support vector regression for real-time flood stage forecasting. Journal of hydrology, 328(3-4), 704-716.
13. Behzad, M. K. Asghari, M. Eazi, M. Palhang. 2009. Generalization performance of support vector machines and neural networks runoff modeling. Expert System with Applications, 36: 7624-7629.
14. Li, P. H., Kwon, H. H., Sun, L., Lall, U., & Kao, J. J. (2010). A modified support vector machine-based prediction model on streamflow at the Shihmen Reservoir, Taiwan. International Journal of Climatology, 30(8), 1256-1268.
15. Zakaria, Z. A., & Shabri, A. (2012). Streamflow forecasting at ungaged sites using support vector machines. Applied Mathematical Sciences, 6(60), 3003-3014.
16. Shahbazi, A. N., & Pilpayeh, A. R. (2012). River flow forecasting using support vector machines. In Proceedings of 14th International Conference on Computing in Civil and Building Engineering.
17. Khan MS, Coulibaly P. 2006. Application of Support Vector Machine in Lake Water Level Prediction. Journal of Hydrologic Engineering 11 (3): 199–205.
18. Hassan, M., Shamim, M. A., Hashmi, H. N., Ashiq, S. Z., Ahmed, I., Pasha, G. A., ... & Han, D. (2015). Predicting streamflows to a multipurpose reservoir using artificial neural networks and regression techniques. Earth Science Informatics, 8, 337-352.
19. Wang, D., Luo, H., Grunder, O., & Lin, Y. (2017). Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction. Renewable Energy, 113, 1345-1358.
20. Adnan, M., Nabi, G., Poomee, M. S., & Ashraf, A. (2017). Snowmelt runoff prediction under changing climate in the Himalayan cryosphere: A case of Gilgit River Basin. Geoscience Frontiers, 8(5), 941-949.
21. Goyal M. K., B. Bharti, J. Quilty, J. Adamowski, and A. Pandey, (2014). Modelling of daily pan evaporation in subtropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. Expert Systems with Applications, vol. 41, no. 11, pp. 5267–5276
22. Shamim, M. A., Hassan, M., Ahmad, S., & Zeeshan, M. (2016). A comparison of artificial neural networks (ANN) and local linear regression (LLR) techniques for predicting monthly reservoir levels. KSCE Journal of Civil Engineering, 20, 971-977.
23. Rauf, A., Ahmed, S., Ghumman, A. R., Ahmad, I., Khan, K. A., & Ahsan, M. (2016, December). Data driven modelling for real-time flood forecasting. In Proceedings of the 2nd International Multi-Disciplinary Conference, Gujrat, Pakistan (pp. 19-20).
24. Kisi, O. (2015). Streamflow forecasting and estimation using least square support vector regression and adaptive neuro-fuzzy embedded fuzzy c-means clustering. Water resources management, 29(14), 5109-5127.
25. Ghorbani, M. A., Zadeh, H. A., Isazadeh, M., & Terzi, O. (2016). A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. Environmental Earth Sciences, 75(6), 476.
26. Dhamge, N. R., Atmapoojya, S. L., &Kadu, M. S. (2012). Genetic algorithm driven ANN model for runoff estimation. Procedia Technology, 6, 501-508.
27. Jajarmizadeh, M., Lafdani, E. K., Harun, S., & Ahmadi, A. (2015). Application of SVM and SWAT models for monthly streamflow prediction, a case study in South of Iran. KSCE Journal of Civil Engineering, 19(1), 345-357.
28. Aichouri, I., Hani, A., Bougherira, N., Djabri, L., Chaffai, H., &Lallahem, S. (2015). River flow model using artificial neural networks. Energy Procedia, 74, 1007-1014.
29. Seyam, M., & Mogheir, Y. (2011). Application of artificial neural networks model as analytical tool for groundwater salinity. Journal of Environmental Protection, 2(01), 56.

30. Cheng, C. T., Niu, W. J., Feng, Z. K., Shen, J. J., & Chau, K. W. (2015). Daily reservoir runoff forecasting method using artificial neural network based on quantum-behaved particle swarm optimization. Water, 7(8), 4232-4246.
31. Tayyab, M., Zhou, J., Adnan, R., & Zeng, X. (2017). Application of artificial intelligence method coupled with discrete wavelet transform method. Procedia Computer Science, 107, 212-217.
32. Yaseen, Z. M., Ebtehaj, I., Bonakdari, H., Deo, R. C., Mehr, A. D., Mohtar, W. H. M. W., ... & Singh, V. P. (2017). Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model. Journal of Hydrology, 554, 263-276.
33. Mehr, A. D., & Kahya, E. (2017). A Pareto-optimal moving average multigene genetic programming model for daily streamflow prediction. Journal of hydrology, 549, 603-615.
34. NESPAK (1997). Flood Early Warning System Manual for Indus Basin (p. 425). Lahore: National Engineering Services Pakistan (NESPAK).
35. Blanco, A., Delgado, M., &Pegalajar, M. C. (2000). A genetic algorithm to obtain the optimal recurrent neural network. International Journal of Approximate Reasoning, 23(1), 67-83.
36. Arena, P., Caponetto, R., Fortuna, L., &Xibilia, M. G. (1992, August). Genetic algorithms to select optimal neural network topology. In Circuits and Systems, 1992., Proceedings of the 35th Midwest Symposium on (pp. 1381-1383). IEEE.
37. Durrant, P. J. (2001). winGamma: A non-linear data analysis and modelling tool with applications to flood prediction. Unpublished PhD thesis, Department of Computer Science, Cardiff University, Wales, UK.
38. Evans, D., & Jones, A. J. (2002). A proof of the Gamma test. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 458(2027), 2759–2799 The Royal Society.
39. Bray, M., & Han, D. (2004). Identification of support vector machines for runoff modelling. Journal of Hydroinformatics, 6(4), 265-280.
40. Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. ACM transactions on intelligent systems and technology (TIST), 2(3), 27..
41. Slater, L., & Villarini, G. (2017). Evaluating the drivers of seasonal streamflow in the US Midwest. Water, 9(9), 695.
42. Boskidis, I., Gikas, G. D., Sylaios, G. K., &Tsihrintzis, V. A. (2012). Hydrologic and water quality modeling of lower Nestos River basin. Water resources management, 26(10), 3023-3051.
43. Hassan M, Shamim MA, Hashmi H N, Ashiq S Z, Ahmed I, Pasha A P, Naeem U A, Ghumman A R, Han D (2014). Predicting streamflows to a multi- purpose reservoir using artificial neural networks and regression techniques. Earth Science Informatics (in press). ISSN: 1865-0481
44. Ines, A. V., & Hansen, J. W. (2006). Bias correction of daily GCM rainfall for crop simulation studies. Agricultural and forest meteorology, 138(1-4), 44-53.
45. J. Xie, Optical character recognition based on least square supportvector machine, in: 2009 Third International Symposium onIntelligent Information Technology Application, vol. 1, IEEE, 2009,pp. 626–629.