RESERVOIR OPERATION USING MACHINE LEARNING ALGORITHM

A PROJECT REPORT

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF DEGREE

OF

MASTER OF TECHNOLOGY
IN
HYDRAULIC AND WATER RESOURCE ENGINEERING

Submitted by

Shivani Singh
(2k20/HFE/06)

Under the supervision of

Prof. Munendra Kumar

DEPARTMENT OF CIVIL ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi 110042

MAY, 2022

<u>CANDIDATE'S DECLARATION</u>

I, **Shivani Singh,** Roll no. **2k20/HFE/06**, student of **M. Tech. (Hydraulic and Water Resource Engineering)**, hereby declare that the project Dissertation titled "**Reservoir operation using machine learning algorithm**" is submitted by me to the Civil Engineering Department, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of degree of **Master of Technology in Hydraulic and Water Resource Engineering**. The content of this thesis is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associates, Fellowship or other similar title or recognition.

Place: Delhi **SHIVANI SINGH**

Date: 31-05-2022 (2k20/HFE/06)

DEPARTMENT OF CIVIL ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi 110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "**Reservoir operation using machine learning algorithm**" which is submitted by **Shivani Singh**, Roll no. **2k20/HFE/06**, Civil Engineering Department, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of **Master of Technology in Hydraulic and Water Resource Engineering**, is a record of the project work carried out by her under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Date:

PROF. MUNENDRA KUMAR

SUPERVISOR

Professor,

Department of Civil Engineering

Delhi Technological University

Bawana Road, Delhi -110042

DEPARTMENT OF CIVIL ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi 110042

ACKNOWLEDGEMENT

I wish to express my sincerest gratitude to Prof. Munendra Kumar for his continuous guidance and mentorship that he provided me during the project. He showed me the path to achieve my targets by explaining all the tasks to be done and explained to me the importance of this project as well as its industrial relevance. He was always ready to help me and clear my doubts regarding any hurdles in this project. Without his constant support and motivation, this project would not have been successful.

Place: Delhi **SHIVANI SINGH**

Date: (2k20/HFE/06)

ABSTRACT

Generating hydropower, supporting water supply, and blocking over lasting droughts are few crucial tasks of water stored in reservoir. During floods, the water delivery from the reservoir must be acceptable, to confirm that the gross volume of water is at a safe level and release from the reservoir will not trigger flooding downstream. This study aims to develop the well-versed assessments for management of reservoir and pre-release water outflow using machine learning, a new and exciting area of artificial intelligence and the most valuable, time, supervised, and cost-effective approach. In this study two data-driven forecasting models, Regression Tree (RT) and Support Vector Machine (SVM) are employed using approximately 30 years of hydrological records to simulate reservoir outflow. Obtaining accurate monthly river flow discharge prediction has always been a challenging task in water resources management for that different models of SVM and RT are applied to the data accurately to predict the fluctuations in the water outflow of a Bhakra reservoir. Different input combinations were used to find the most effective release such as reservoir level (M), monthly reservoir storage (BCM), the previous inflow of reservoir (MCM), the current inflow of reservoir (MCM), evaporation of reservoir (MCM), the previous outflow of the reservoir (MCM) and time (months) and release of the reservoir. . Findings indicate that SVM (medium guassian) combination having seven different parameters gives minimum RSME (720.2), maximum R^2 (0.8), minimum MAPE(14.0197), minimum scatter index(.4239) and minimum MAE (360.69) and therefore, can be considered as the best model for the dataset with these techniques. The ability to accurately estimate changes in reservoir outflow can aid in the planning and management of reservoir water usage in the long run.

CONTENT

LIST OF FIGURES

LIST OF TABLES

ABBREVATIONS

API-RP-2A	=	American Petroleum Institute Recommended Practice -2A
LQ	=	Living Quarters
SACS	=	Structural Analysis Computer System
RSR	=	Reserve strength ratio
RP	=	Recommended practices
SIM	=	Structural Integrity Management
GoM	=	Gulf of Mexico
ISO	=	International Organization for Standardization
UC	=	Unity Check
CD	=	Chart Datum
LAT	=	Lowest Astronomical Tide
AT	=	Astronomical Tide
Cm	=	Coefficient of Mass

CHAPTER 1

INTRODUCTION

1.1 Background

A reservoir is a tangible construct (artificial or natural) used as water storage for water supervision, monitoring, and maintenance of water supply [1]. Reservoirs are the most valuable element among the various components of a water resources system. Besides, due to environmental issues, construction of a new dam is not easy-going, therefore it is important that active reservoir needs to be boosted for maximum effective plan to handle with the future and present water challenges. Across a flow, a reservoir is built by constructing a dam. The major feature of a reservoir is the rule of herbal streamflow with the aid of storing surplus water withinside the moist season and liberating the saved water in a destiny dry season to complement the discount in river flow. The intention is to balance the streamflow and to change the sequential and three-dimensional water availability. The water stored in a reservoir can be distributed later for advantageous uses giving rise to sequential changes or rerouted through waterways or pipelines to outlying locations resulting in three-dimensional changes. Reservoir outflow projection is guided by various potential constraints example water storage, inflow of water, water level, evaporation, infiltration, geomorphology, and others which requires to be considered to understand the ambiguity. There have been plentiful methods used in forecasting hydrological practices over the past years. Traditional tactics used were of linear mathematical relationships based on capability of machinist, simple set of curve fitment, and standards employed to quote reservoir water outflows [3]**.** However, undermining, and poor performances of numerical models due to unavailability and complexity of statistics, missing datapoints and overemphasized constraints. Various machine learning algorithms had been used in previous study in intent to overcome the concern and to estimate reservoir water outflows [4, 5]. Subsequently many Machine Learning (ML) models, such as Artificial Neural Networks (ANNs), Radial Basis Neural Networks (RBNN), Support Vector Machines (SVMs), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), Logistic Regression

(LR), etc have been deployed in water management systems progressively owing to improve the consistency and precision of the estimated models [6-8]. Modelling a machine to work and improvise on its own without each time explicit programming is called ML. In intellectual studies, ML has the capability to solve the complex problems with high level of accuracy and predictions as demanded for future periods [9]. Nowadays, AI models came to be extended successfully to the reservoir operation field. Compared to conventional physical prediction models, ML models with the help of historical dataset can learn numerous hydrological operations independently on correct operating rates. Advantage of this modeling is the capability of the software system to map the input-output models [10-12]. To forecast daily water levels, five different ANN models were tested each with an increasing number of inputs. It has been seen that the accuracy began to decrease after the addition of many inputs. The reason for this is that the network starts to be obsolete and irrelevant as explained in the study [13]. On comparing the performance of SVM and multilayer perceptron (MLP), it is found that due to the optimization algorithm SVM has a great deal of capacity to resolve a linearly constrained quadratic programming function, and the optimum kernel function in this case is the radial basis kernel function. [14]**.** During the process of creating fuzzy membership functions, a study on the ANFIS technique indicated that triangular and trapezoid membership functions are deemed to be more suited than bell-shaped membership functions. [15]. A genetic algorithm (GA) is successfully utilized in optimizing reservoir operation and by using data from a longer period of time, the GA model could be further improved for reservoir water levels [16]. Many more AI methods, such as the adaptive network-based fuzzy inference system (ANFIS), genetic algorithm (GA), and decision tree, have been effectively applied to the reservoir operation field in addition to these AI algorithms (DT). Many reservoirs in California use an improved decision tree (DT) algorithm, classification, and regression tree to estimate storage or release. [17].

1.2 Reservoir operation

Reservoir operation is frequently viewed as a multilevel stochastic control problem, with large gains in benefits even with tiny increases in operational efficiency. Despite the emergence of new techniques, implicit stochastic optimisation (ISO) models solved with linear programming (LP) continue to be one of the most often used strategies for complex reservoir systems. Finding dependencies from data was done using statistical analysis like correlation analysis and multiple regressions in ISO models. However, the use of these strategies is limited since the mathematics form of the link between independent and dependent variables must be defined explicitly. This method may not produce appropriate answers for highly nonlinear issues or those with substantial uncertainty.

Reservoirs are the most significant of the different components of a water management system. A dam is built across a stream to create a reservoir. The principal function of a reservoir is to control natural streamflow by storing extra water during the rainy season and releasing it during the dry season to adjust for the reduction in river flow. In a nutshell, the goal of a reservoir is to balance natural streamflow while also adjusting the water's temporal and geographic availability. Water kept in a reservoir can be transported to distant locations via pipelines or canals, causing spatial changes, or it can be held in the reservoir and eventually released for beneficial purposes, causing temporal changes. Water is either held in the reservoir or provided from the storage, depending on the size of natural inflows and demands at any given time. A reservoir provides a head of water that can be used to generate electricity as a result of storing water. In the case of flood control projects, it creates vacant space for water storage, reducing hydrograph peaks. A reservoir also serves as a pool for navigating rapids, as well as providing habitat for aquatic life and recreational and sporting amenities. It improves visual attractiveness, encourages afforestation, and protects wildlife. In India, reservoirs are built on a regular basis for flood prevention and conservation. The Indian subcontinent does have a tropical monsoon, which means that the most of the water is gathered from June to September. The conservation standards are best met when the reservoir is as full as possible at the completion of the filling period. Flood control, on the other hand, necessitates empty storage space in order to collect and regulate incoming floods to suitable levels. The conflict between the functions is handled in terms of storage capacity limits by effective reservoir functioning. Reservoir operating policy specifies the amount of water to be released from storage at any particular time, based on the state of the reservoir, consumption levels, as well as any information about the reservoir's likely intake. A single-purpose reservoir's operational difficulty is choosing which discharges should be taken from the reservoirs in order to reap the benefits for that purpose. In the case of a multifunctional reservoir, the discharge must be optimally allocated across various functions in addition to the foregoing.

Characteristics and requirement of water of various water uses

The degree to which the multiple intended aims are compatible determines the difficulty of the reservoir management challenge. When the goals are more compatible, it takes less effort to coordinate them. The following are the numerous uses for which a reservoir is utilised, as well as the technical specifications for these applications:

- Irrigation needs vary depending on farming trends in the command area. Irrigation demands are high, and only a small portion of the water delivered is accessible as return flow to the system. The precipitation with in command area has a direct impact on these requirements. Demands will be lowest during the monsoon and highest throughout the summer and winter months. Unless the command area grows or the cropping pattern changes dramatically from year to year, the average annual demands remain rather stable. Drought resistance is determined by the amount of available storage in the reservoir, so it's best to keep as much backup water in stored as possible while meeting current demands.
- Seasonally, and to a smaller extent daily and hourly, hydroelectric power demand varies. The degree of variability is determined by the types of loads provided, which include industrial, municipal, and agricultural loads. Hydroelectric demands, for example, are highest in municipal areas during the prime summer months. Furthermore, two demand peaks are recorded throughout the day, one in the early and the other in the evening. Because water can be used for consumptive purposes downstream after passing through turbines, hydroelectric power demand falls under the non-consumptive use of water. The quantity of power generated is determined by the water volume and effective head.
- Flood prevention reservoirs are built to manage flood flows into them. Flood control is accomplished by holding a portion of inflows in a reservoir and releasing the rest. The amount of flood attenuation or moderation is determined by the amount of empty storage space in the reservoir at the time the flood hits. The

availability of vacant storage capacity in the reservoir is required to accomplish this goal. The discharges from such storage are kept as low as possible in relation to the downstream safe capacity.

- Storage reservoirs are frequently constructed to keep a stretch of river erupting from the reservoir passable by maintaining a suitable flow depth in the river channel used for transportation. Seasonal differences in water needs for navigation can be seen. When enough depth of flow is present in the channel during the monsoon season, there is rarely any demand. During the dry season, when significant releases are needed to maintain the required depth, the demands are at their peak. The kind and volume of travel on navigable waterways determines demand at any given time.
- This component of the reservoir provides benefits when it is used for swimming, boating, fishing, and other water activities and picnics. Recreational benefits are usually unintended consequences of other reservoir activities, and reservoirs are rarely used for recreation. During the recreation season, when the reservoir is nearly full, the recreational activities are best supported. Large and quick changes in water level are detrimental to recreational interests since they can result in marshy plains around the reservoir's rim.

1.3 Machine learning

A subset of artificial intelligence is machine learning (AI). Rather than being explicitly programmed to do so, it focuses on educating computers to learn through data and improve over time. AI is described as a programme with cognitive abilities comparable to those of a person. One of the major concepts of artificial intelligence is to have computers think like people and solve problems in the same manner we do. Artificial intelligence (AI) is a broad phrase that refers to all computer programmes that can think like humans. AI is defined as any computer programme that exhibits features such as self-improvement, inference learning, or even basic human tasks like image recognition and language processing. Algorithms are taught in machine learning to detect patterns and connections in massive data sets and then make good decisions and estimates based on that analysis. Machine learning algorithms grow over time and be more accurate as more data is available. Machine learning is used in our homes, shopping carts, entertainment media, and healthcare. ML theory is linked to pattern classification and statistical inference, and it states that a model can learn to enhance its task performance based on its own prior experience [18]. Artificial neural networks (ANNs), support vector machines (SVMs), and relevance vector machines are examples of machine learning models (RVMs).

Deep learning and neural networks are machine learning subcategories that are concentric. AI analyses data to make decisions and forecasts. Machine learning approaches allow AI not only to analyse data, but to understand from it and grow without requiring additional programming. Artificial intelligence is the ancestor of all machine learning subsets. The first subgroup includes machine learning, deep learning, and neural networks.

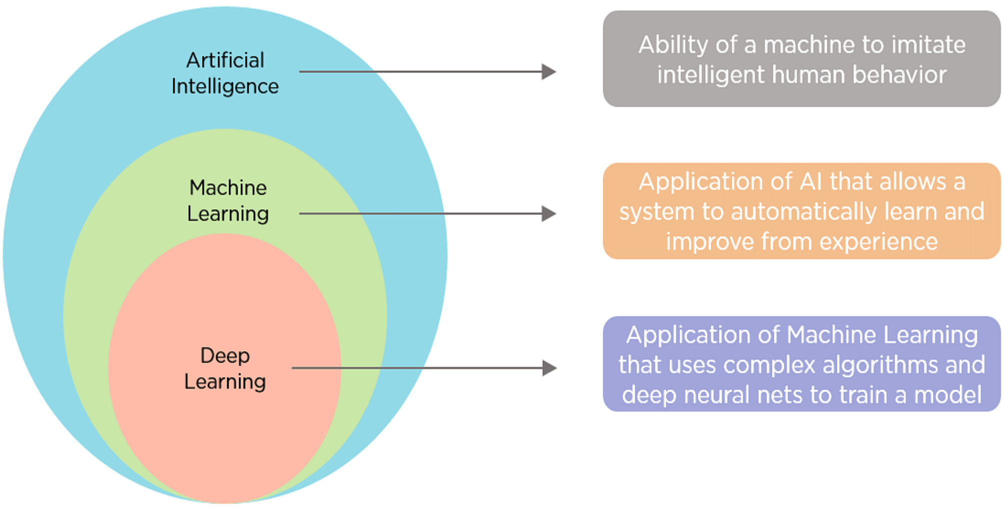

Fig 1.1 Diagram of the relationship between AI and machine learning

Because it uses multiple layers of neural networks and large amounts of complex and heterogeneous data, this type of machine learning is referred to as "deep." The system interacts with numerous layers of the network to achieve deep learning, collecting increasingly higher-level outputs. By human standards, big data is time-consuming and difficult to handle, yet good data is the finest feed for training a machine learning algorithm. In a large dataset, the more clean, useable, and machine-readable information there is, the more effective the machine learning algorithm's training will be. Machine learning algorithms, as previously said, can enhance themselves through training. Today, three common strategies are used to train machine learning algorithms. Machine learning can be divided into three categories: supervised learning, unsupervised learning, and reinforcement learning.

1.4 Working of Machine learning

Various types of machine learning models employ distinct algorithmic strategies in machine learning. Depending on the nature of the information and the desired outcome, one of four learning models can be used: unsupervised, supervised, semi-supervised, or reinforced. One or even more algorithmic strategies may be used under both of those models, based on data sets being used and outputs expected. Machine learning algorithms are used to categorise objects, detect patterns, anticipate outcomes, and make informed judgments. Algorithms can be used alone or in combination to achieve the highest level of accuracy when dealing with complex and unexpected data.

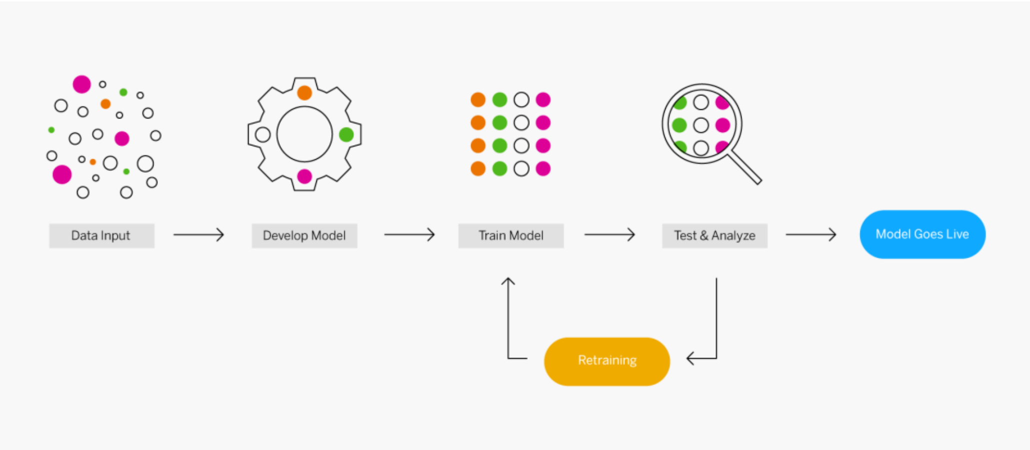

Fig 1.2 Working of Machine learning process

i) Supervised learning

Amongst the most basic types of machine learning is supervised learning. In this example, the machine learning algorithm is trained using labelled data. Regardless of the fact that data should be appropriately labelled for it to work, supervised learning is very successful when utilised in the right situations.

In supervised learning, the machine learning algorithm is given a small training dataset. This training dataset is a subset of the broader dataset that provides the algorithm with a foundational understanding of the problem, solution, and data points to be dealt with. The training dataset has many of the same qualities as the final dataset and provides the labelled parameters that the algorithm requires to solve the problem. At the completion of the training, the algorithm learns how the data operates and the relationship between input and output.

After that, the solution is deployed for use with the testing dataset, from which it learns in the very same way it did with the training dataset. As it learns from new data, supervised machine learning algorithms will enhance even after they have been deployed, recognising new patterns and relationships.

ii) Unsupervised learning

Unsupervised machine learning has the advantage of being able to deal with unlabeled data. This eliminates the need for human labour in the creation of the dataset device, allowing the computer to handle considerably larger datasets. In supervised learning, the labels allow the algorithm to find the correct nature of any connection between the two data points. Unsupervised learning, on the other side, lacks labels to work with, causing hidden structures to emerge. The programme perceives relationship between the data points in an abstract manner, with little or no human input required. Because they can generate these hidden structures, unsupervised learning techniques are versatile. Instead of a planned and fixed problem statement, unsupervised learning algorithms can respond to the data by dynamically altering hidden structures. This enables for further post-deployment development than supervised learning techniques.

iii) Reinforcement learning

Reinforcement learning is based on how people learn from data. It includes a trial-and-error algorithm that builds onto itself learns from different scenarios. Positive outcomes are rewarded or ‘reinforced,' while negative outcomes are discouraged or 'punished’. Reinforcement learning is based on the psychology concept of conditioning, and it works by putting the algorithm in a workplace environment with an interpreter and a reward system. The interpreter receives the final output of each step of the algorithm and determines whether the outcome is useful or not. The interpreter reinforces the algorithm by rewarding it if the programme discovers the correct solution. If the outcome is unfavorable, the algorithm is forced to repeat until a good results is discovered. The reward system is usually closely tied to the productivity of the result.

ML strives to automate the knowledge engineering process by replacing much of the time-consuming human effort with automated algorithms that improve the results or efficiency by detecting and utilizing regularities in training data. [**19**]. Given this, machine learning has long been utilised in mathematics, statistics, and engineering with models or algorithms such as linear, polynomial, and time series regression. Some of the tasks learning machines have been used for are include in Table 1.1:

Table 1.1 Example of Data-driven models in Water Resources

Task	Example of use
Anomaly detection	Identification of anomalous data records in meteorological or hydrologic time series variables (outliers, pattern shift, data deviation)
Association rule learning	Finding links (dependence) between variables from several sources to describe a specific phenomenon, such as identifying crucial meteorological factors, vegetation cover, and urban growth data to explain the change in lake water levels over time.
Clustering	Detection of similar groupings and structures in data without using existing data structures or relationships. The Western United States, for example, has been identified as having similar weather-hydrological trends.
Classification	Structures in data are discovered in order to identify patterns. Identification of vegetation covertures in aerial or satellite images, for example.
Regression	Finding a mathematical phrase or equation that accurately models the data. Prediction of river flow based on the weather parameters and specific geographical variables, for example.

1.5 Importance of machine learning in water resource

To better understand the behaviour of hydrological and water resources systems, many modelling tools based on physical principles have been created. The input-output relationship is determined in physically based modelling by developing and solving fluid mechanics and thermodynamics equations with adequate and detailed boundary conditions to explain the dynamics of water across the hydrologic system in issue.

Because the physiographic and geomorphic properties of most hydrological systems are intricate, and the boundary conditions are typically unclear, solutions for physically based models frequently require simplifying assumptions [**20**].

Furthermore, the lack of essential data and the cost of data collecting can limit the practical implementation of physically based models. Researchers have employed data-driven models based on machine learning as an alternative to physically based models to overcome these constraints [**21,22**].A model is constructed in the ML technique to link the macro-description of a system's behaviour (output) to the behaviour of its constituents (inputs) [**23**].

Many, if not all, time series characteristics of hydrology and water resource systems are nonstationary nowadays. As a result, to optimise water systems, approaches that can represent nonstationary behaviours of environmental variables are needed [**24**].This means that traditional statistics (such as ARMA models) that presume the temporal series are stationary are ineffective. ARIMA models could be employed because they account for non-stationary time series behaviour. They do, however, take a linear parametric approach, that can result in poor performance when the model is evaluated with unknown data. ARIMA models are also unsuitable for long-term forecasting (e.g., streamflow forecasting up to a year ahead), because the long-term forecast asymptotically approaches the mean value of the time series data [**25**].Machine learning techniques have been shown to be more effective than ARIMA models in understanding the nonlinear dynamics and nonstationary behaviour of water resources systems with the goal of providing accurate predictions for previously unknown variables [**26,27**].

1.6 Objective of the study

1. Application of Support vector machine (SVM) and regression tree (RT) algorithms for Prediction of outflow
2. To compare the forecasting efficiency of SVM and RT model
3. To summarize the influence of parameters settings on model performance.
4. Comparison of above mentioned technique on basis of calculations of RMSE MAPE, Scatter index , R^2 and Mean Absolute Error.

1.7 Organization of thesis:

Chapter 1 presents a brief introduction covering an overview of Reservoir operation, Machine Learning, its type and how it works. Importance of Data Driven Model in water resource engineering. Need of machine learning in water resource response study and its impact, also objectives behind this study have been described.

Chapter 2 gives the details about the literature review done for this dissertation work and which have been used throughout for investigation work.

Chapter 3 deals with the methodology i.e. support vector machine(SVM), Regression tree(RT) involved for the whole and sole process of analysing and predicting the release from the reservoir. and gives description about study area, Bhakra Nangal Dam in Uttarakhand, about the data used in the study.

Chapter 4 describes and discuss the results obtained from the two different methodologies used in the present study.

Chapter 5 gives the conclusion of the present study resulting from two different methodologies and gives the need for future.

CHAPTER 2

LITERATURE REVIEW

The necessary literature review and studies were carried out through national and international journals, periodicals, conferences, books, codes of practice of different countries and the data available on the internet and sources.

Tomasa and Milosa et al. (2020)

He performed single research to looks into the possibilities of applying adaptive control on the Dyje River reservoir Vranov Reservoir. The control technique employs a fuzzy model to approximate the I/O relationships found in the target reservoir's behaviour matrix, which was built using the differential evolution optimization method. A fuzzy model is used to create recurring projections of water inflows into to the reservoir, which would be built on the concept of consistency of the path of a true series of average monthly flows throughout the course of the year. The entire controls is tested for the span 2004–2018 after the control and prediction models have been calibrated. The outcomes of the described models are compared to the outcomes of graph dispatching. The adaptive control findings suggest that the system is well suited to driving during long periods of low water. When there is plenty of water, the results are almost same [28]

Yaseen et al.(2015)

As AI has made great strides in forecasting and modelling nonlinear hydrological applications, as well as in capturing the dataset's noise complexity. His research looks on a state-of-the-art use of artificial intelligence in stream-flow forecasting, with a focus on defining data-driven AI, the advantages of complementary models, and the literature and their possible future application in stream-flow modelling and forecasting. A new method for modelling inflow, a unique manner of pre-processing time series frequency utilising Fast Orthogonal Search (FOS) approaches, and Swarm Intelligence (SI) as an optimization strategy are all discussed in the study. [29].

Gavahi et al.(2018)

The performance of a proposed adaptive real-time optimum reservoir operation model is studied utilising two data-driven inflow prediction approaches, according to his research. The model consists of three modules: a forecasting module that forecasts monthly future inflows, a reservoir operation optimization module that estimates monthly optimum reservoir releases through the end of the year, and an updated module that updates the current state of the system and feeds the other two modules the most recent seen information on future inflows. The K-nearest neighbour (KNN) and adaptive neuro fuzzy inference system (ANFIS) approaches are used to anticipate monthly inflows to the reservoir. The results demonstrate that ANFIS surpasses the KNN approach by 25, 23, and 25 percent in terms of RMSE, PWRMSE, NSCE, and correlation coefficient indices. [30].

Bai et al.(2015)

He suggested that inflow forecasting uses data to help with reservoir operations and management. This research proposes a multiscale deep feature learning (MDFL) strategy using hybrid models for daily reservoir inflow predictions. Multiscale (trend, period, and random) characteristics are extracted via ensemble empirical mode decomposition & Fourier spectrum, which are subsequently represented by 3 deep belief networks (DBNs). Following that, the weights of the each DBN are used to initialise a neural network (D-NN). The predicting results are eventually reconstructed utilising the outputs of the 3 D-NNs using a sum-up technique. The suggested MDFL using hybrid models investigates a historical everyday inflow dataset (from 1/1/2000 - 31/12/2012) of a Three Gorges reservoir in China. For the same goal, four peer models are being used for comparison. In terms of mean absolute percentage error (MAPE = 11.2896%), normalised root-mean-square error (NRMSE = 0.2292), determination coefficient criteria (R2 = 0.8905), and peak percent threshold statistics (PPTS(5) = 10.0229%), the current model outperforms all peer models. The suggested method merges a deep framework using multiscale and hybrid observations to investigate complicated natures in reservoir inflow predictions[31].

Zainab et al. (2019)

He claimed that his research demonstrates how to predict streamflow patterns in semi-arid environments using a novel hybrid model known as Multivariate Adaptive Regression Spline coupled with Differential Evolution (MARS-DE). This is accomplished by inspecting month time series river discharge data at the Baghdad station, which is coordinated at the Tigris River in Iraq. The model is validated using Least Square Support Vector Regression (LSSVR) and standalone MARS models. To demonstrate the analysis of the undertaken models, various statistical indicators are developed to verify the modelling accuracies. The MARS-DE model demonstrated outstanding hybrid predictive modelling capabilities over monthly time scale river discharge in the semi-arid region based on the obtained results. For the same goal, 4 peer models are used for comparison. In respect of mean absolute percentage error (MAPE = 11.2896%), normalised root-mean-square error (NRMSE = 0.2292), determining coefficient criteria (R2 = 0.8905), & peak percent threshold statistics (PPTS(5) = 10.0229%), the current model outperforms other peer models. The suggested method incorporates a deep framework with multiscale and hybrid observations to investigate complicated natures in reservoir inflow predictions [32].

Yilmaz et al. (2018)

He estimated suspended sediment load (SSL) at two gauging stations on the Oruh River in Turkey using artificial bee colonies (ABC), teaching-learning-based optimization algorithm (TLBO), and multivariate adaptive regression splines (MARS). These models were compared to one another as well as traditional regression analysis (CRA). Model inputs were river discharge values and previously acquired SSL data, with anticipated SSL data as an output. To improve model accuracy, 2 different testing and training dataset combinations were used. The root mean square error value for the MARS technique was found to vary between 35 and 39 percent for both the test two gauging stations, which would have been lower than other models' errors. Using a different dataset, error values were considerably lower (7 to 15 percent). Our findings show that simultaneous streamflow and SSL measurements are the most effective factor for developing reliable predictive models, and also that MARS is the best accurate model for forecasting SSL [33].

Jieqiong et al. (2014)

According to him, reservoir storage forecasting is critical for reservoir operation and management. In this paper, a genetic algorithm (GA)-based support vector machine (SVM) model was formulated to predict monthly reservoir storage at Miyun Reservoir (Beijing's only surface drinking` water source) from 1995 to 2011. Simultaneously, two more SVM-based models were utilised for comparison, each combining linear search and particle swarm approaches for parameter optimization. The findings revealed that the created GA-SVM model seemed to have the best calibration and prediction performance. The GA-SVM model will be widely used to predict reservoir storage in various places due to its excellent accuracy [34].

Hipni et al.(2013)

The research done by him proposes a new expert system approach that employs SVM to predict the daily dam level of water of the Klang gate. The input situation, the kind of SVM regression, the amount of V-fold cross-validation, and the time lag are all factors in determining the optimal model. The both rainfall R(t-i) as well as the dam water level L are used in the optimum input scenario (t-i). The best regression kind is Type 2 SVM regression, and most accurate results are obtained using 5-fold cross-validation. The findings are compared to those produced with ANFIS: all of the RMSE, MAE, and MAPE values show that SVM outperforms ANFIS. Finally, all of the findings are added together to establish the optimal time lag, yielding R(t-2) L(t-2) as the best system with only 1.64 percent inaccuracy[35].

Fathiana et al. (2019)

He proposed that two time series analysis approaches, namely self-exciting threshold autoregressive (SETAR) and generalised autoregressive conditional heteroscedasticity (GARCH) models, be investigated first, followed by three artificial intelligence approaches, namely artificial neural networks (ANN), multivariate adaptive regression splines (MARS), and random forests (RF) models, to predict monthly river flow. Monthly river flow data from the Brantford and Galt stations on the Grand River in Canada were used for this purpose, and its performances were evaluated using multiple assessment criteria during October 1948 to September 2017. In terms of predicting flow of the river at the study stations, the SETAR model outperformed the GARCH model. Furthermore, the standalone MARS and RF models outperformed the ANN by a small margin. Following that, hybrid models were created by combining the previously employed ANN, MARS, and RF models using non-linear time series models SETAR and GARCH[36].

Shaokun et al. (2021)

In this study, the popular Génie Rural à 9 Paramètres Journalier (GR4J-9) hydrological model is utilised to predict streamflow, with the simulated series and remote sensing data being used to train the long short-term memory (LSTM) technique. The improved GR4J9–LSTM model chain effectively improves the performance of the river discharge simulation by integrating more remote sensing data connected with the hydrological response variables, according to the findings. We were able to demonstrate the potential use of our suggested methodology by integrating the LSTM-based simulation findings into a reservoir operation model. [37].

Zhanga et al. (2018)

The goal of this research is to summarise the impact of parameters on model performance and to investigate the LSTM model's applicability to reservoir operation modeling. The following are the findings: (1) The impacts of the amount of maximum iterations of model performance should be prioritised for the BP neural network and LSTM models; for the SVR model, simulation performance is closely linked to the kernel function selection, and sigmoid and RBF kernel functions should indeed be prioritised; (2) the BP neural network and SVR are suited for the model to learn reservoir operating rules from a limited quantity of data. (3) The LSTM model can effectively minimise the time and memory storage requirements of existing AI models, as well as display good capabilities in modelling low-flow conditions as well as the outflow curve during peak operation periods[38].

Ahmadi et al.(2029)

This research presents 2 reservoir operation optimization models regarding water allocation to various customers. Both models' goal functions are dependent on the

Nash Bargaining Theory, which may take into account the utility functions of water users and stakeholders, as well as their relative water allocation authority. The initial optimization model is known as GA–KNN (Genetic Algorithm–K Nearest Neighbourhood). In this model, a KNN strategy for predicting initial solutions is employed to speed up the GA convergence process. In addition, each month, KNN is used to generate operational rules based on the optimization findings. The Bayesian Stochastic GA (BSGA) optimization model is the second model. This model takes into account the conditional probability of inflow and the reservoir's forecast. The inherent and predicted uncertainty of reservoir input are therefore taken into account. The proposed models are tested by applying them to a Satarkhan reservoir system in Iran's north western region[39].

CHAPTER 3

METHODOLOGY

3.1 Study Area

Bhakra Dam on the Sutlej River(Bilaspur, Himachal Pradesh) is a concrete gravity dam in the northern part of India. The dam has geographic coordinates of 31°24′39″N latitude and 76°26′0″ E longitude. The dam is considered the highest gravity dam in the world. The Sutlej River, a major tributary of the Indus River, originates in Tibet and flows into the Indo-Gangetic plains nearby Bhakra. The overall catchment area upriver of the Bhakra River is 56,980 km^2. The precipitation in the catchment changes according to an annual average of about 875 mm. The dam is situated in a canyon near the (now submerged) upstream Bhakra community in the Himachal Pradesh district of Bilaspur, at a height of 226 metres. The dam is 518.25 metres long and 9.1 metres wide. Its "Gobind Sagar" reservoir can hold up to 9.34 billion cubic metres of water. The Bhakra dam generated a 90-kilometer-long reservoir that covers 168.35 square kilometres. It is India's third-largest reservoir in terms of water storage capacity. The Bhakra Beas Management Board is in charge of the dam's operation and maintenance (BBMB). Bhakra Dam is a straight gravity cum concrete dam with four radial spillway gates and an 8212 cumec designed overflow capacity. The location of the study area on the map is shown in Figure 3.1. The Nangal reservoir is built by a 28-95 m high dam which is situated about 11 km downstream of the Bhakra dam. It controls irrigation releases by acting as a head regulator. During the monsoon, the dam retains extra water and releases it gradually throughout the year. It also protects against flood damage caused by monsoon rains. This dam feeds the Bhakra canal, which irrigates 10 million acres (40,000 km2) of land in, Haryana Punjab, and Rajasthan. Table 3.1 shows the characteristics of the Bhakra Nangal dam and reservoir.

Figure 3.1. Bhakra Dam's location

Table 3.1 Characteristics of Bhakra Nangal Dam and Reservoir

Item	**Description**
Catchment area	56980 square kilometers
Normal reservoir level	EL. 512.06 meters (EL.1680 feet)
Dead storage level	EL.445.62 meters
New area irrigated	60 lakh acres
Area of the reservoir.	162.48 square kilometers (62.78 sq. miles)
Length of the reservoir.	96.56 kilometers
Live storage capacity at EL.1680 ft.	6911 million cum (5.60 MAF)
Gross storage capacity at EL.1680 ft.	9340 million cum (7.57 MAF)
Dead storage capacity	2430 million cum (1.97 MAF)

3.2 Data Collection

In this study, 20 years data were collected from Jan 1989 to Dec 2009 and then extrapolated for the next 10 years from Jan 2010 to Dec 2019 by decade wise simple moving average method. A total of 2976 (30 years) historical data points used are reservoir level (M), monthly reservoir storage (BCM), the previous inflow of reservoir (MCM), the current inflow of reservoir (MCM), evaporation of reservoir (MCM), the previous outflow of the reservoir (MCM) and time (months) and release of the reservoir. All the data were acquired from the websites of "UK Centre for Ecology and Hydrology", "Bhakra Beas Management Board" and "India Meteorological Department". The range for reservoir water level was determined by the hydraulic features of the Bhakra dam; the maximum water level was 512.06 metres and the minimum operating level was 450.45 metres. Table 3.2 shows the essential statistical properties of the inputs, such as minimum, maximum, and total count.

During modeling nonlinear hydrological processes, one of the tasks is selecting the most significant variables from the entire set of input variables [40]. when there is a vast dataset and a large number of variables mainly during the identification process, selection of input is critical for learning systems[41].

Table 3.2 Data acquired with descriptive statistic

Input	Unit	Minimum	Maximum	Average	Count
Inflow	MCM	352.04	10267.8	1764.88	372
Reservoir level	Meter	450.45	512.06	489.01	372
Monthly reservoir storage	BCM	0.2	6.23	3.13	372
Evaporation	MCM	-28.10	41.20	-6.80	372
Previous inflow	MCM	352.04	10267.8	1764.62	372
Previous outflow	MCM	402.3	8942.3	1819.95	372

The major goal of data collecting and study is to choose appropriate input variables depending on the data available. The selection of the best subset of inputs in the model, also known as feature selection, is a method of selecting the best subset of inputs based on defined governing rules[42]. To increase model accuracy and efficiency through such selection of input for change of models. During the modelling phase in this study, various input variable combinations are used. For this study, initially five scenarios are defined shown in Table 3.3 at a different number of folds to find the most effective output and then, for each scenario, the prediction accuracy was evaluated.

Table 3.3 The selected scenarios for input combination

Number	Different Input Combinations	Output
1	Inflow, Evaporation	Outflow
2	Inflow, Evaporation, Reservoir level	Outflow
3	Inflow, Evaporation, Reservoir level, Monthly reservoir storage	Outflow
4	Inflow, Evaporation, Reservoir level, Monthly reservoir storage, Previous inflow	Outflow
5	Inflow, Evaporation, Reservoir level, Monthly reservoir storage, Previous inflow, Previous outflow	Outflow

3.3 Matlab

Data Preparation

Data for this study were secondary data, data points of 20 years were collected from Jan1989 to Dec 2009 and then extrapolated for next 10 years from Jan 2010 to Dec 2019 by decade wise simple moving average method. A total of 2976 (30 years) historical data points used are reservoir level (M), monthly reservoir storage(BCM), previous inflow of reservoir(MCM), current inflow of reservoir (MCM), evaporation of reservoir (MCM), previous outflow of reservoir (MCM) and time(Months) and release of reservoir from Jan1989 to Dec 2019 were acquired from the the website of "UK Centre for Ecology and Hydrology" and from the website of "Bhakra Beas Management Board and from the websites of India Meteorological Department (IMD).

SVM regression model, RT model has been trained by using hyperparameter optimization in the regression learner app in the MATLABR2021b. Figure 3.2 shows the flow chart using machine learning algorithms to forecast values.

Data set variable has been selected from workspace for the new session. Default validation option is used to protect against overfitting. Then all the model are trained and at each iteration, the app tries a different combination of hyperparameter values. Then list out the optimized hyperparameters in both the Optimization Results section to the right of the plot and the Model Hyperparameters section of the model Summary tab. The model is exported to the MATLAB workspace.

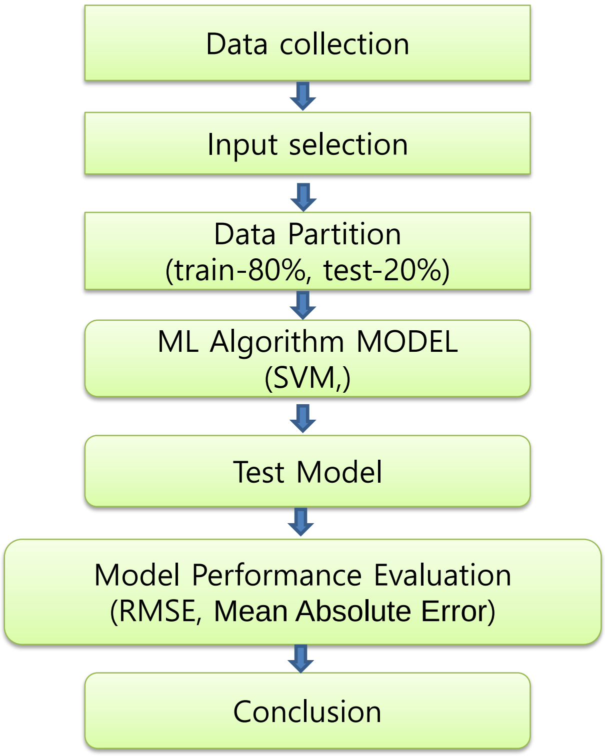

Fig-3.2 Methodology flow chart using machine learning algorithms to forecast values.

3.4 Support Vector Machine (SVM)

Support Vector Machine has gained popularity as a novel statistical learning method in the recent two decades, and for both classification and regression, it has been demonstrated to be an efficient and reliable approach. [43-45]. when compared to the use of traditional chaotic methods, The SVM method is based on the idea of mapping input data into a high-dimensional feature space to aid classification and simulate an unknown relationship between both the set of input variables and the set of output variables. The sights to see the mechanism's simplicity, the two advantages of this method are that it is sufficiently known by scientists and that it dominates prediction. It had a level of precision that set it apart from several approaches; for example, nearest neighbours and ANNs. The SVM is a strategy that uses a kernel trick to learn everything there is to know about an issue while simultaneously lowering the complexity of the model and prediction error. SVM Classification is the first step in making a decision limitation for the feature space, which stands out by generating an ideal separation hyperplane between the two classes to maximize the margin by minimizing the generalization error. In theory, SVM classification has the potential to predict outcomes that can be comprehended using three essential concepts: (1) function of the kernel (2) the soft-margin (3) separation hyperplane[46,47]. Polynomial, radial basis and sigmoid, functions are exemplary kernel functions. Algorithm like SVM is mostly used to forecast classification problems, and the support vector regression (SVR) is the expansion of the support vector machine (SVM) that adds an insensitive loss function to allow it to be used in regression analysis. [46, 48]. In other words, in a classification problem, the SVM is utilised for partition of data into "+1" and "-1" classes, the SVR, on the other hand, is a generalized SVM approach for predicting random real values. [49, 50]. To improve reservoir inflow forecasting, a modified SVM-based prediction system was created. [51]. Climatic data from the previous time was used, as well as highly connected climate precursors. To understand the non-linear pattern underlying climatic systems more flexibly, the SVM parameters were determined using a genetic algorithm-based parameter determination approach. The median of forecasts from the created models was then used to reduce the variation in the prediction by using bagging to construct several SVM models. In terms of predictive ability, the suggested modified SVM-based model outperformed a bagged multiple linear regression (MLR), a simple SVM, and a simple MLR model. In terms of prediction ability, the suggested modified SVM-based model outperforms the previous models.

A regression with an alternate loss function is an example of an SVM. Loss functions are frequently used in the estimate, model selection, and prediction, and they are critical in determining any disparity between the null and nonparametric models' fitted values [52]. In terms of hydrology, researchers must consider loss function while making predictions. In this study, the usage of the hydrologic loss function is linked to two primary variables: rainfall and runoff. A distance measure must be supplied, which necessitates a change in the loss function [53]. SVR's main notion is to nonlinearly translate the initial data into a higher-dimensional feature space and solve the linear regression issue there. (Figure. 3.3). As a result, as demonstrated in Eq. (4), SVR is usually required to construct a suitable function f(x) to reflect the non-linear relationship between feature xi and target value yi.

$$\mathrm{f}(x_i)=\mathrm{w}.\varphi(x_i)+\mathrm{b}\|w\|^2 \tag{1}$$

where w is the coefficient vector, $\varphi\ (x_i)$ is the transformation function, and w and b denotes the weight and bias. W and b are calculated by minimising the so-called regularised risk function, as shown in Eq. (5).

$$\mathrm{R(w)}=\frac{1}{2}\|w\|^2+c\sum_{i=1}^{n} L_\varepsilon(y_i\,,\mathrm{f}(x_i\,)) \tag{2}$$

where $\frac{1}{2}\|w\|^2$is the regularization term; c is the penalty coefficient; $L\varepsilon(y_i, \mathrm{f}(x_i\,))$the ε-insensitive loss function, which is calculated according to Eq. (6).

$$L\varepsilon\big(y_i, \mathrm{f}(x_i\,)\big)=\max\{0,|y_i-\mathrm{f}(x_i\,)|\text{-}\varepsilon\} \tag{3}$$

Where ε signifies the allowed error threshold, and is ignored if the projected value is inside the threshold; otherwise, the loss equals a number greater than ε

To solve the optimization boundary, two slack factors $\xi^+$and $\xi^-$are introduced:

$$\mathrm{minf}(\mathrm{w},\xi^+,\xi^-)=\frac{1}{2}\|w\|^2+c\sum_{i=1}^{n}(\,\xi^+,\xi^-) \tag{4}$$

Subject to

$$y_i-[\mathrm{w}.\varphi\ (x_i\,)]-\mathrm{b}\ \leq\ \varepsilon+\xi^-,\xi^-\ \geq 0$$

$[w.\varphi\ (x_i\ )] + b - y_i \leq\ \varepsilon + \xi^+, \xi^+\ \geq 0$ (5)

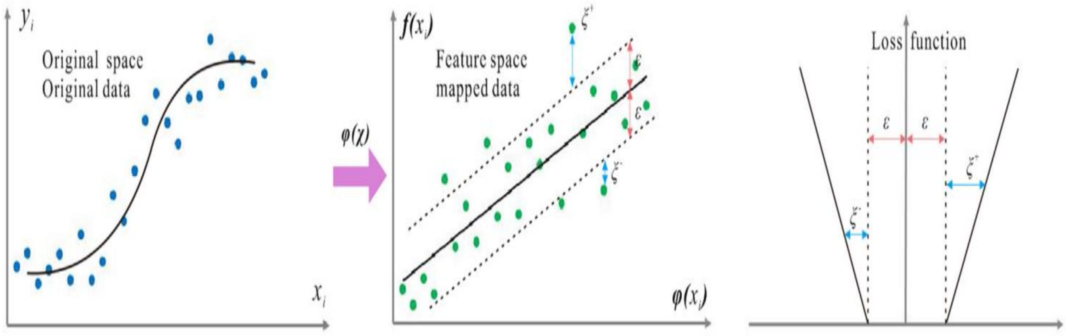

Figure 3.3. SVR schematic diagram [54]

process normally used for assessment of existing offshore structures involves two different type of analysis:

The minimization of a Lagrange function, which is formed from the objective function and the problem constraints, yields the dual version of this optimization problem:

$$\max\alpha,\ \alpha^* = \frac{1}{2}\sum_{i,j=1}^{N}\ (\alpha_i\ - \alpha^*{}_i), (\alpha_j\ - \alpha^*{}_j\ )K(x_i, x_j) - \varepsilon\sum_{i=1}^{N}(\ \alpha_i + \alpha^*{}_i)$$

$$+\ \varepsilon\sum_{i=1}^{N} y_i(\ \alpha_i - \alpha^*{}_i),$$

s.t $\sum_{i=1}^{l}\ (\alpha_i\ - \alpha^*{}_i) = 0,$ i=1to N, (7)

$\alpha_i\ , \alpha^*{}_i\ \geq 0$ i=1to N,

$-\alpha_i\ , -\alpha^*{}_i\ \geq -C$ i=1to N.

The inner product $\{\varphi\ (x_i\ ), \varphi\ \left(x_j\ \right)\}$in the feature space is denoted by the function K(x i,x j) in the dual formulation of the issue.

Any function $K(x_i, x_j)$ can become a kernel function if it satisfies the inner product criteria. Hence, the regression function is as follows:

$$f(x)=\sum_{i=1}^{N} (\alpha_i - \alpha^*{}_i)K(x_i, x_j) + b \tag{8}$$

3.5 Regression tree (RT)

Regression Trees are machine-learning algorithm for building prediction models off of datasets. Regression Trees employs a clustering tree with post-pruning processing. Various papers have referred to the clustering tree algorithm as the forecasting clustering tree and the monothetic clustering tree [55, 56]. Regression Trees are used to model-dependent variables having a finite number of values which is not arranged in order, with prediction error commonly assessed as the squared difference between predicted and observed values [57]. The clustering tree algorithm is based on the top-down induction technique of the decision tree. [58]. Regression Trees algorithm takes a collection of data for training and creates a new internal node that is as good as possible. Based on their decreased variance, the system chooses the top test scores. The lower the variance, the more homogeneous the cluster is and the more accurate the forecast. The programme generates a leaf and marks it as data representative if none of the tests significantly reduce variance [55,56]. By recursively splitting the data space and fitting a prediction model within each partition, a hierarchical tree -like division of the input space can be created.[59]. A sequence of recursive splits divide the input space into local regions, which are designated by a series of recursive splits. Internal decision nodes and terminal leaves make up the tree. Starting at the root node, a sequence of tests and decision nodes will determine the path through the tree till it approaches a terminal node, provided a test data point. A prediction is made at the terminal node based on the model linked with that node locally

3.6 Water Balance

The theory of mass conservation wherein variations in total water volume, intake (precipitation), and outflow (surface and subsurface runoff, evaporation, transpiration) on a particular area are balanced is known as water balance. The learning of the water balance combined with prior knowledge of the meteorological and physical parameters of the basin gives information on current and future water volumes as well as additional insight into the complicated process of basin runoff.

The determination of water balance is utilised for a variety of research and applied challenges, including estimating a regional water balance and assessing the impact of human activities and climate fluctuations on basin discharge. Understanding water balance in connection to climatic and morphological basin factors provides insight into complicated processes that take place over a variety of geographical and temporal scales. Droughts will be more frequent and last longer, and floods will be more intense, indicating the need for a more complete understanding of current and future watershed conditions. The computation of water balance is necessary for this understanding, and it may also provide trustworthy information for developing climate change mitigation plans for the watershed.

Water is almost always in motion in the natural world, and it can change state from liquid to solid or vapour given the right conditions. Water inflows must equal water outflows inside a specified area during a specific length of time, plus or minus any change in storage within the area of interest, according to the conservation of mass principle. Simply put, water that enters a region must either depart or be kept within the space. The most basic form of the water balance equation 9 is shown below:

$$P = Q + E \pm \Delta S \qquad (9)$$

Therefore, $Q = P - (E \pm \Delta S)$

Where, P is precipitation, Q is release , E is evaporation and ΔS is the storage in the reservoirs

An analysis of water balance can be used to:

i) Examine the present state and trends in the availability of water resources in a given area throughout time.
ii) Assess and improve the validity of visions, scenarios, and strategies to improve water management decision-making.

Estimates of water balance are frequently reported as precise. In fact, due to insufficient data capture networks, sampling error, and the complex geographical and temporal heterogeneity that defines hydrological processes, there is always uncertainty. It is often reasonable to eliminate components that do not effect changes when data sources are imprecise. For example, if year-to-year storage variations (such as reservoirs) are minor, storage might be omitted from a yearly water balance. The prediction capability of water balance equation, for predicting monthly outflow of Bhakra reservoir is carried out and the result shows RMSE(1166.331), R2 (0.52), MAE(659.338), MAPE(30.16). Figure 3.4 show variation of observed outflow with the outflow predicted from conventional method.

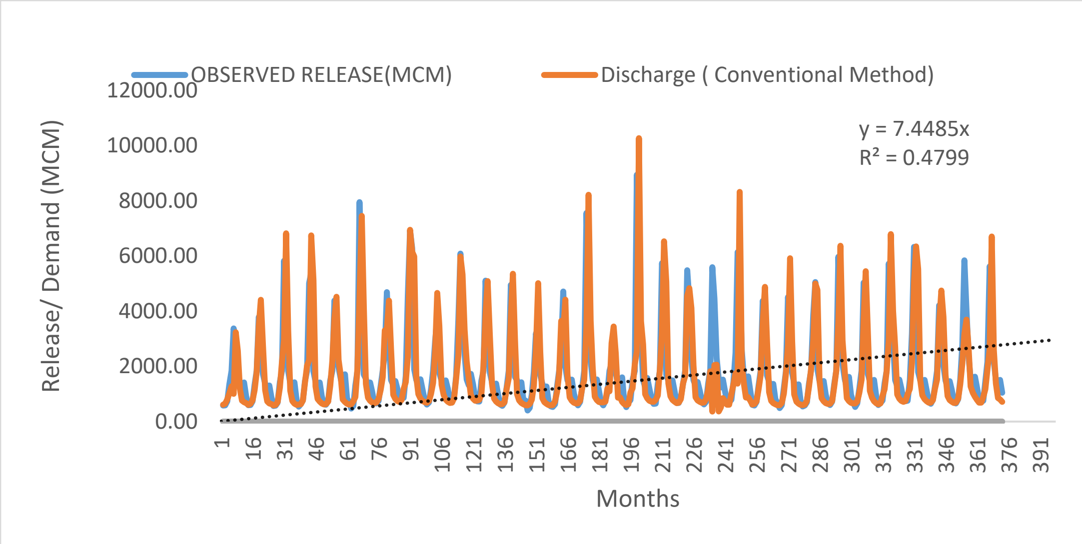

Figure 3.4 Comparison of predicted outflow using conventional method and observed monthly outflow

3.7 K- fold Cross Validation

The holdout approach is employed when the amount of data available for training and testing is restricted. A subset of the data is saved for validation, while the rest is used for training. It is customary in engineering practice to keep one-third of the data for validation and utilize the other two-thirds for training and testing [60]. Divide the obtained data into a specified number of equally sized observations or folds to improve on the holdout approach (k). The dataset used for testing is chosen from among these (k) folds, whereas the rest (k-1) are employed in the training process. This procedure is repeated k times, with each time a different fold being tested and the remaining folds (k-1) serving as the training dataset. As a result, the approach generates k different degrees of accuracy. The variance of the resulting estimate diminishes as, k is increased. Consider the 5-fold cross-validation scenario (K=5). Figure 3.5 shows how the dataset is divided into five folds. The first fold is used to test the model, while the others are used to train it in the first iteration. The second iteration uses the second fold as the testing set and the rest as the training set. This procedure is repeated until each of the five folds has served as a testing set.

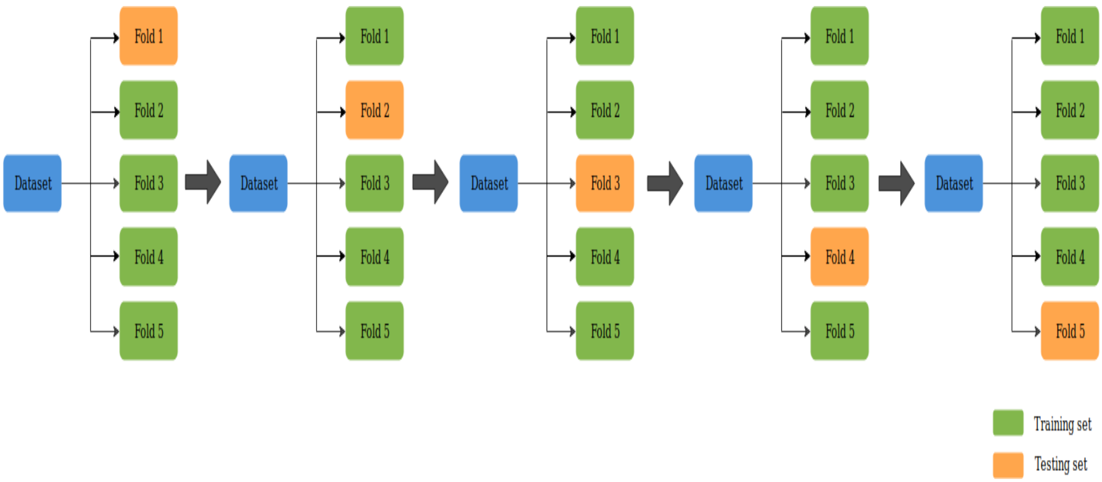

Figure 3.5 Different Fold 5 Cross-Validation

CHAPTER 4

RESULT AND DISCUSSION

In this section, the observed Bhakra Nangal reservoir outflows are compared with simulated results based on the two different AI models, i.e. SVM, and RT, combined with various model parameters. Best value of each model results in most precise estimating model, and later the unsurpassed grouping combination is chosen. The data was divided into two categories: Regression tree and SVM analysis. Following that, two models are employed to make data projections based on five diverse scenarios. The most appropriate and exact prediction scenario is determined by each model's best estimation. To decide whether model was best, estimating power of both models was examined. To evaluate the suggested model's execution in varied in preparing, checking, and testing information, five of measurable assessments are used i.e., RMSE, MAE , MAPE, Scatter Index and R^2. Figure 4.1- 4.5 represents the comparison of observed values with predicted values of outflow on monthly basis using SVM for 5-fold cross validation of different scenario. It is clear from the figure 4.5 that predicted values are much closer with observed ones for scenario 5 a medium gaussian as a kernel function. As for the scenario 5, the points are more narrowed around the line, which shows a better correlation of its outputs with respect to the observed inflows as presented in figure 4.5. Also, the R-square value is maximum amid all the situations i.e.(R^2=.8). In all other scenario points are more scattered, which is indicating lower prediction performance, especially for discharge between 4000 to 1000 MCM.

Data observed from the table 4.1 clearly shows that in SVM, scenario 5 relates with the lowest RMSE(720.1), MAPE(14.0197),SI(0.423962929) and MAE(360.69) values and a maximum R^2 (.8) value amid all the situations Moreover, Data forecasting in scenario 5 offers the most accurate results, while Scenario 4 provides the second-lowest. Besides, scenario 1 has erratic forecasts compared to all other scenarios as also shown in figure 4.1. Lowest value for authentication oversights RMSE is also held by Scenario 5 i.e.(RMSE = 720.1). Continuous development can be seen in results from scenario 3-4 in SVM Regression. A big inaccuracy from scenarios 2 to 3 of the SVM was recorded in the values of MAPE, MAE, SI increased while the coefficient of correlation decreased from 0.68 to 0.67. All the result of statistical calculation of Support Vector Machine for all the five scenarios are summarized in table 4.1.

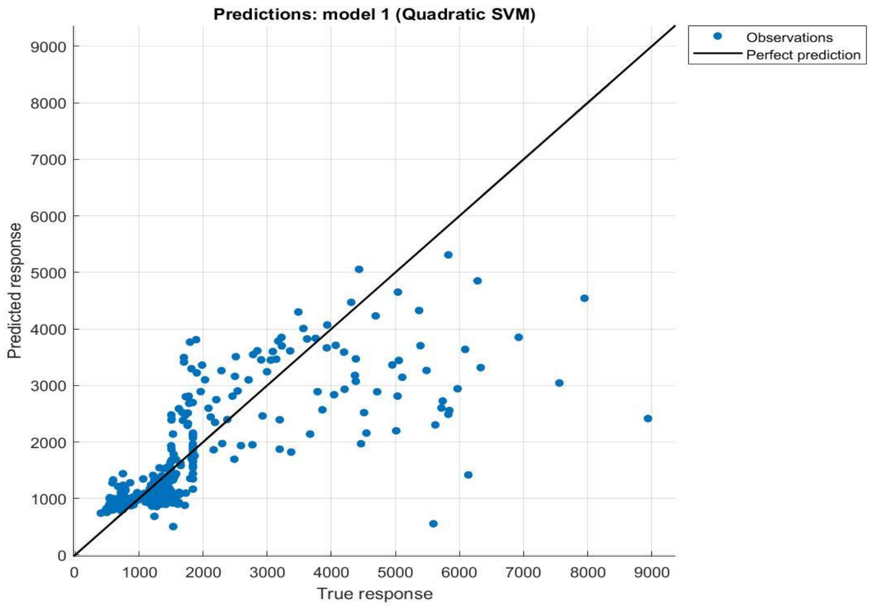

Figure 4.1- Relationship between observed outflow and predicted outflow using SVM (model-1).

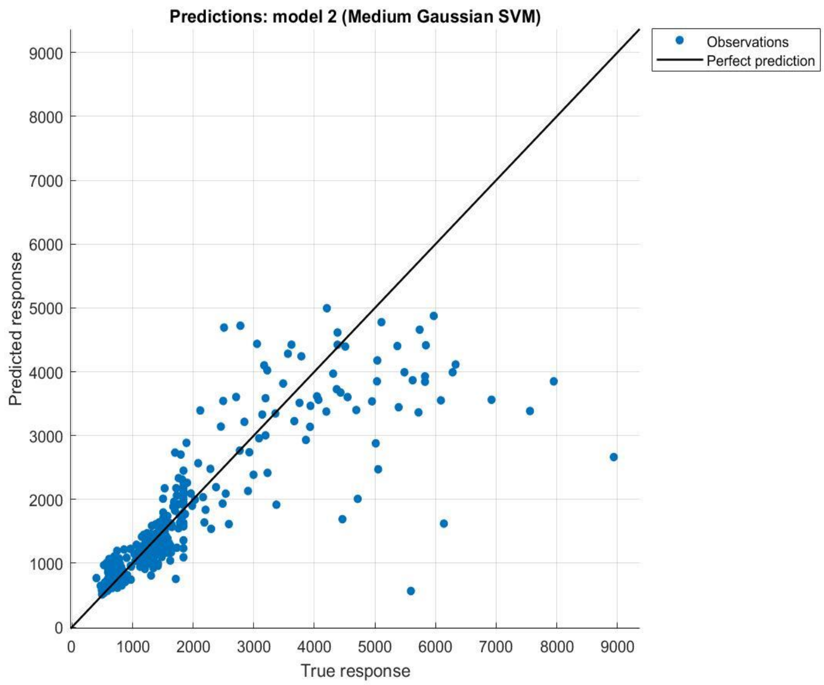

Figure 4.2- Relationship between observed outflow and predicted outflow using SVM (model-2).

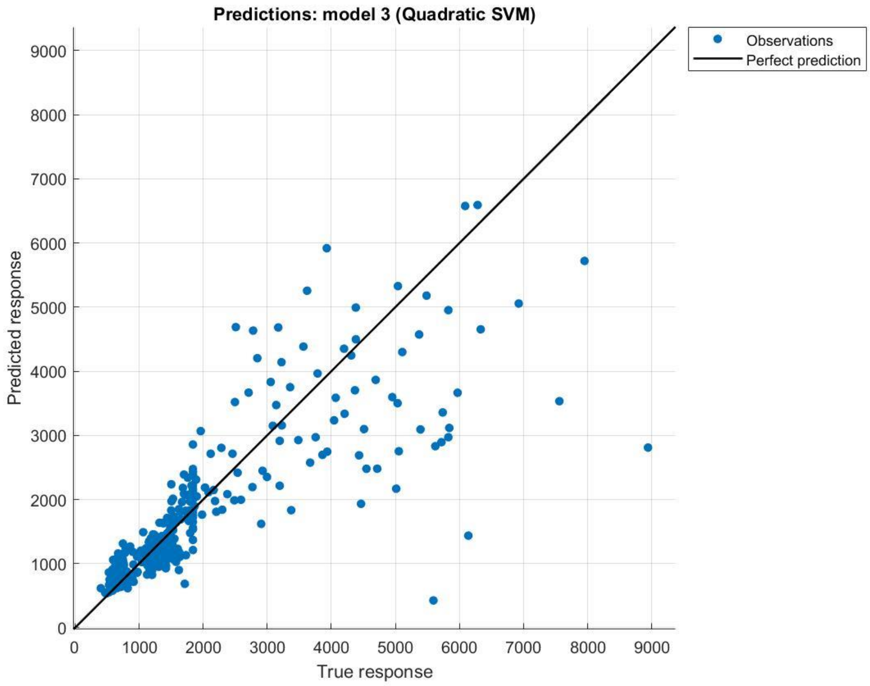

Figure 4.3- Relationship between observed outflow and predicted outflow using SVM (model-3).

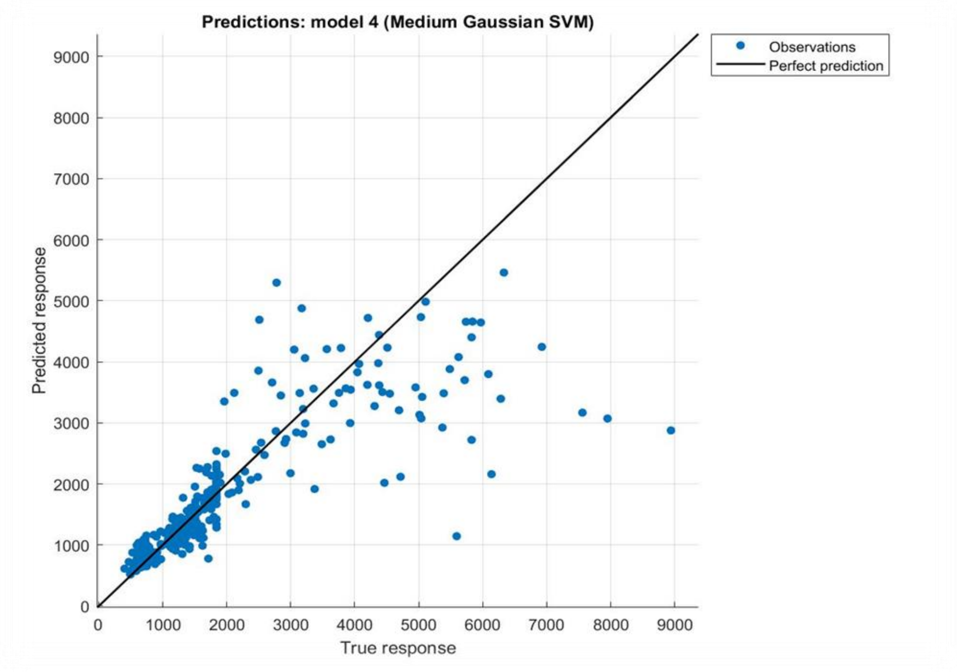

Figure 4.4 - Relationship between observed outflow and predicted outflow using SVM (model-4).

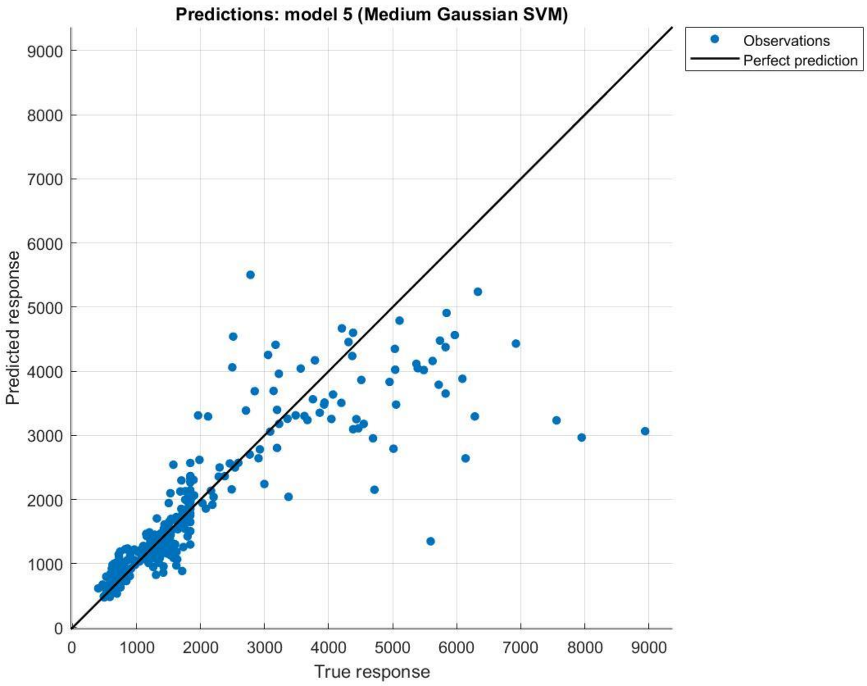

Figure 4.5- Relationship between observed outflow and predicted outflow using SVM (model-5).

Table 4.1 Statistical calculation of Support Vector Machine for 5 scenario.

SVM	BEST SVM	RSME	R-SQUARE	MAE	Scatter Index	MAPE
Scenario-1	QUADRATIC	961.33	0.56	542.71	0.528218721	27.88
Scenario-2	MEDIUM GUASSIAN	827	0.68	402.99	0.454408873	16.114
Scenario-3	CUBIC SVM	838.84	0.67	421.51	0.460914558	18.3536
Scenario-4	MEDIUM GUASSIAN	815.44	0.68	380.25	0.44805704	14.707
Scenario-5	MEDIUM GUASSIAN	720.1	0.8	360.69	0.423962929	14.0197

Figure 4.6 show the scatter plot of observed data and the predicted outflow by SVM algorithm and conventional method on monthly basis. It can be observed from Figure 4.6 that SVM scatter plots clearly demonstrate that the discharged predicted from SVM have less scattered estimates compared to discharge predicted from conventional

method and SVM regression provides a better fit to the data. As for the SVM method, the points are more narrowed around the y=x line, which shows a better correlation of its outputs with respect to the conventional method. In case of conventional method, points are more scattered showing lower prediction performance, especially for outflows greater than 5000 MCM.

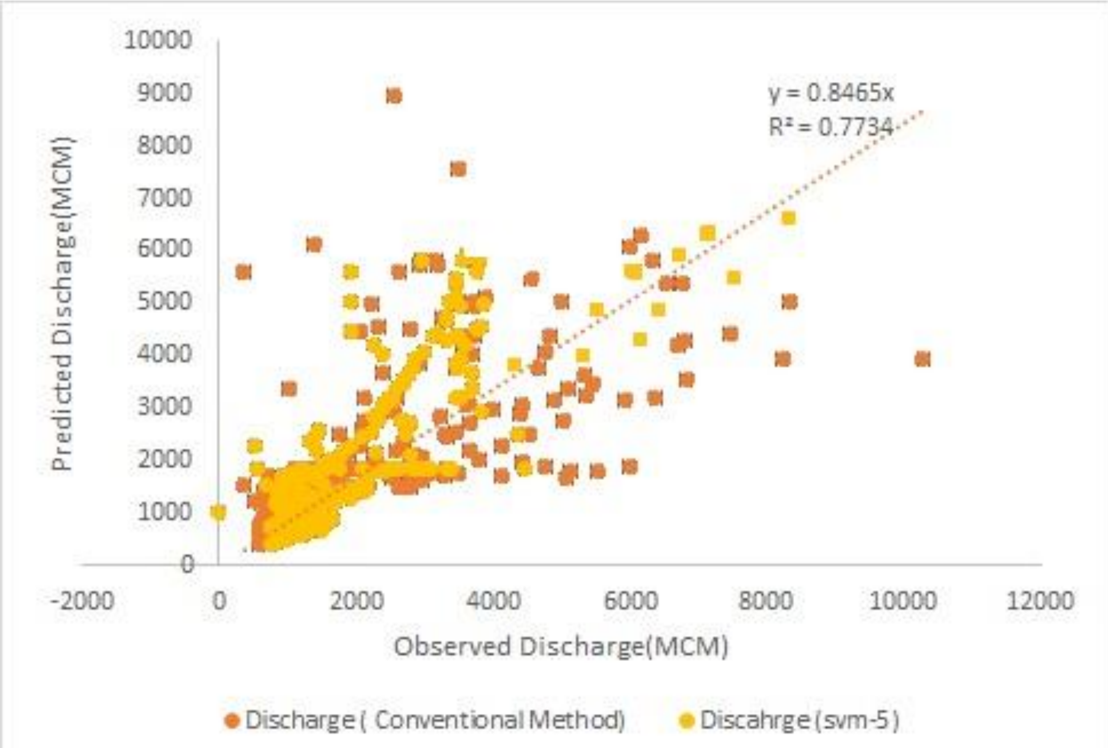

Figure 4.6- scatter plot of observed outflow and predicted discharge by SVM algorithm and conventional method.

Figure 4.7-4.11 shows scatter plots of the monthly observed and predicted outflow by the Regression tree models. It is evident from the error variations that the predictions of scenario 4 is closer to the corresponding observed values than the other scenario having lowest RMSE (748.03), and highest value of R2 (0.73). It is obviously seen from Figure 4.10 that the model 4 of regression tree have less scattered outflow estimate. Table 2 displays the results of 5-fold cross-validation of regression tree models us various assessment criteria. We found that according to our statistical assessment standards. Scenarios had achieved best, due to having the highest R^2 value (0.73) as medium tree, followed by the scenario 5, scenario 3, scenario 2, and scenario 1. In terms of RMSE, scenario 4 had the best predictive power RMSE (748.03), followed by scenarios 2, 3, 2, and 1. In Figure 4.1 points are more scattered, which is indicating lower prediction performance, especially for outflow greater than 4000 MCM. All the result of statistical calculation of Regression tree for all the five scenarios are summarized in table 4.2.

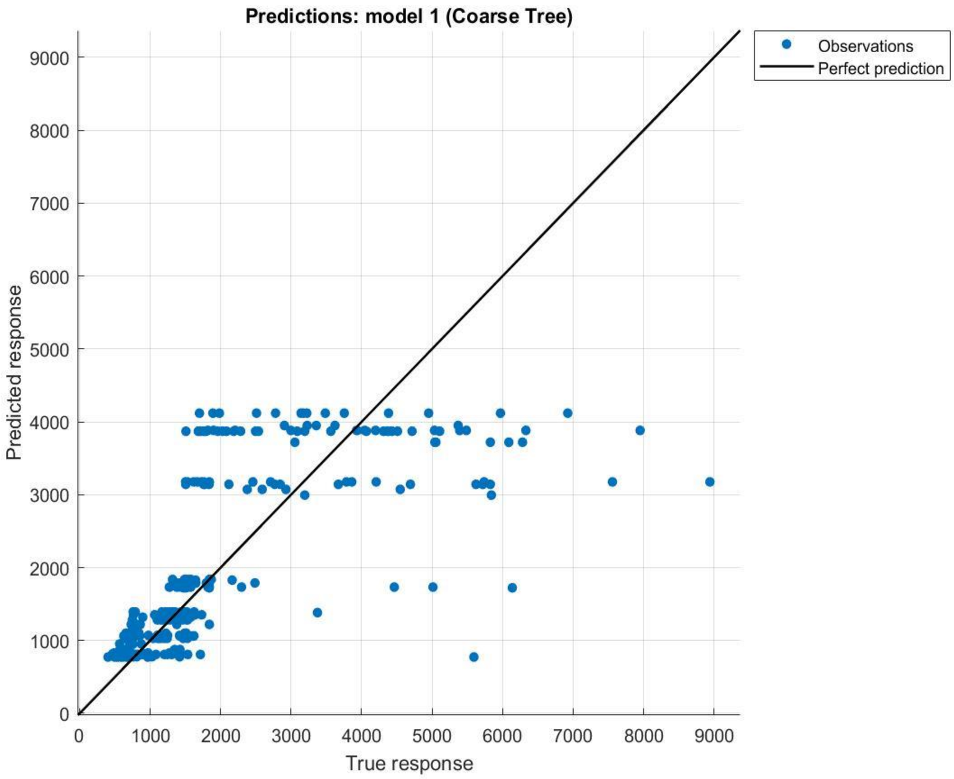

Figure 4.7 - Relationship between observed outflow and predicted outflow using RT (model-1).

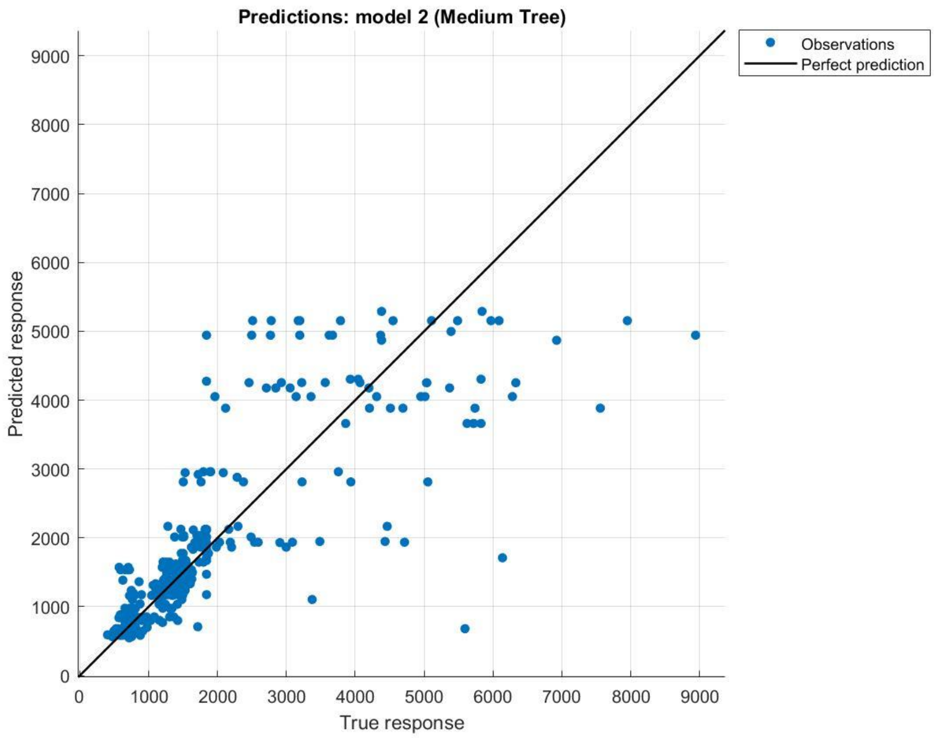

Figure 4.8 - Relationship between observed outflow and predicted outflow using RT (model-2).

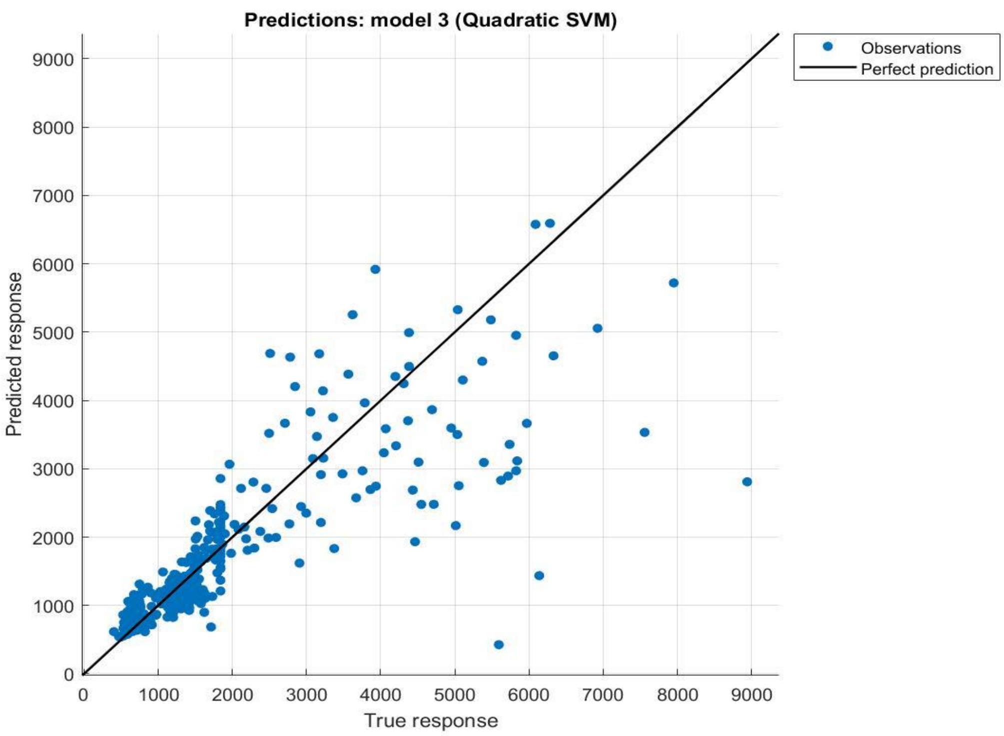

Figure 4.9 - Relationship between observed outflow and predicted outflow using RT (model-3).

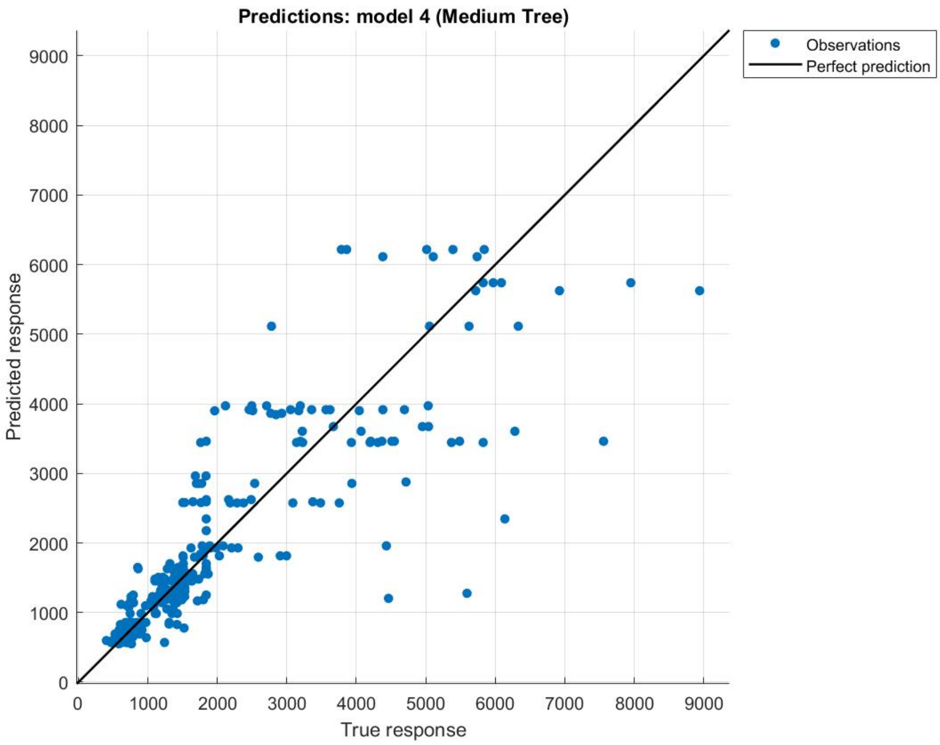

Figure 4.10 - Relationship between observed outflow and predicted outflow using RT (model-4).

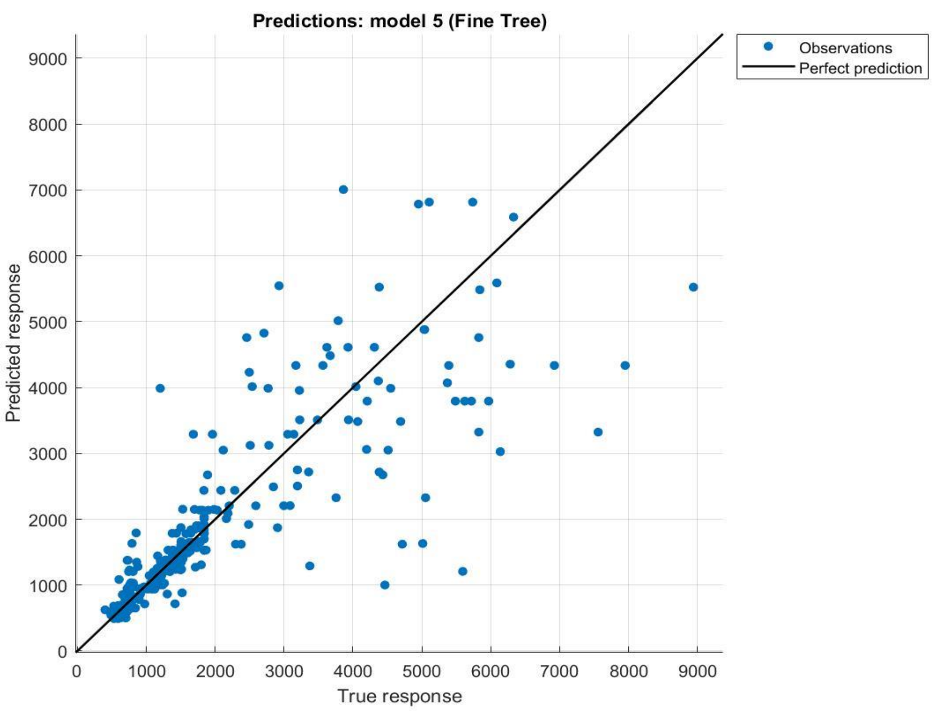

Figure 4.11 - Relationship between observed outflow and predicted outflow using RT (model-5).

Table 4.2 Statistical calculation of Regression tree for 5 scenario.

	BEST RT	RMSE	R-SQUARE	MAE	SCATTER Index	MAPE
Scenario-1	COARSE	989.36	0.53	572.36	0.54362027	27.23
Scenario-2	MEDIUM	921.53	0.6	470.05	0.506349951	18.448
Scenario-3	MEDIUM	871	0.64	470.01	0.478585404	18.518
Scenario-4	MEDIUM	748.03	0.73	408.41	0.411017496	15.711
Scenario-5	FINE	809.47	0.69	381.68	0.444776724	9.0982

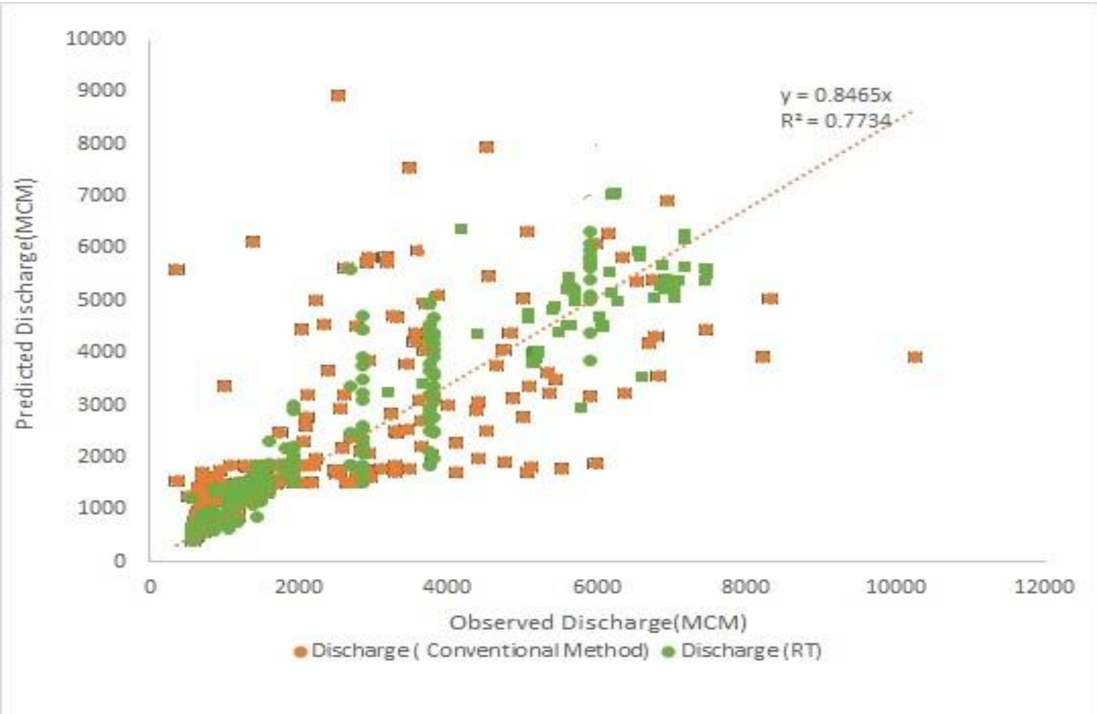

Fig 4.12 scatter plot of observed outflow and predicted outflow by regression tree and conventional method.

It is clear from the above figure 4.12 that points are more scattered in case conventional method, which is indicating lower prediction performance when compare to regression tree. As for the Regression tree method, the points are more narrowed around the line, which shows a better correlation of its outputs with respect to the observed inflows.

In Table 4.3, comparison of predicted outcomes based on the two distinct AI models, i.e., Regression Tree and Support Vector Machine united with various parameters of the model are presented. The findings reveal that SVM model (scenario-5) having (RMSE (452.17), R2 (0.9)) performs the best when compared to other SVM and RT models. It is also evident from the figure 4.13 that the predicted discharge of reservoir by SVM is in good fit with the original streamflow in comparison of that RT model and also discharge calculated by conventional method.

Table 4.3 Statistical calculation of Regression Tree Model and Support Vector Machine Model.

BEST MODEL	RMSE	R-SQUARE	MAE	SCATTER INDEX	MAPE
SVM (MEDIUM GUASSIAN)	720.1	0.8	360.69	0.423962929	14.0197
RT (FINE)	809.47	0.69	381.68	0.444776724	9.0982

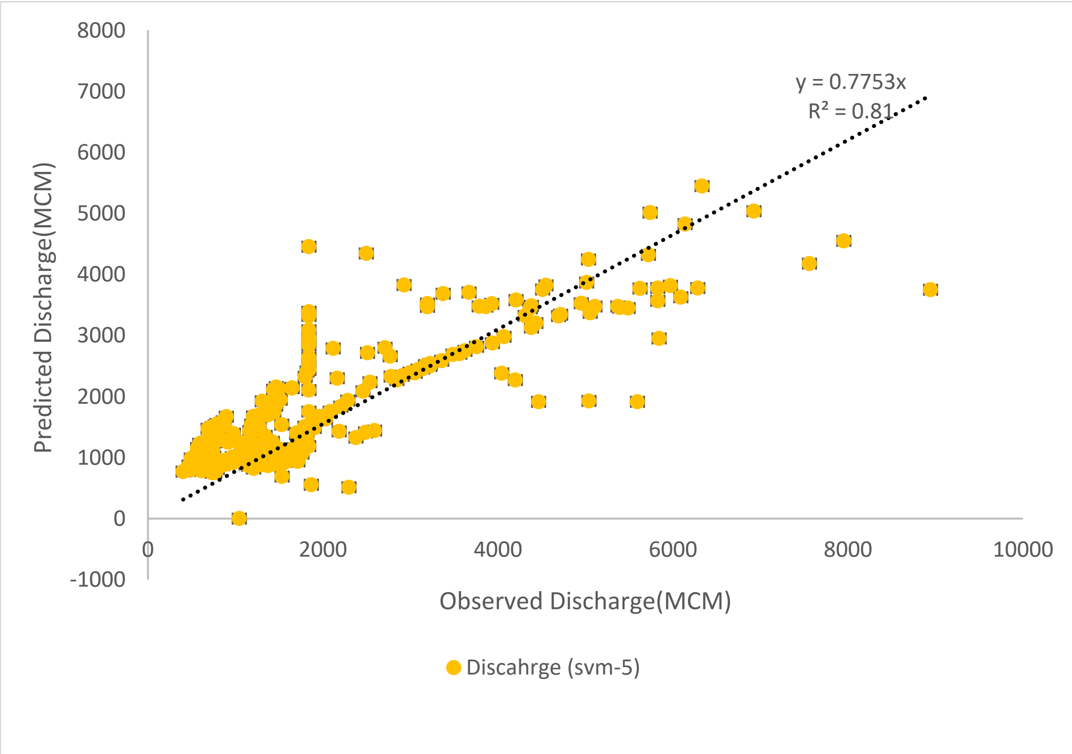

Figure 4.13: Scatter plots of forecasted outflow using SVM for scenario 5 and observed outflow.

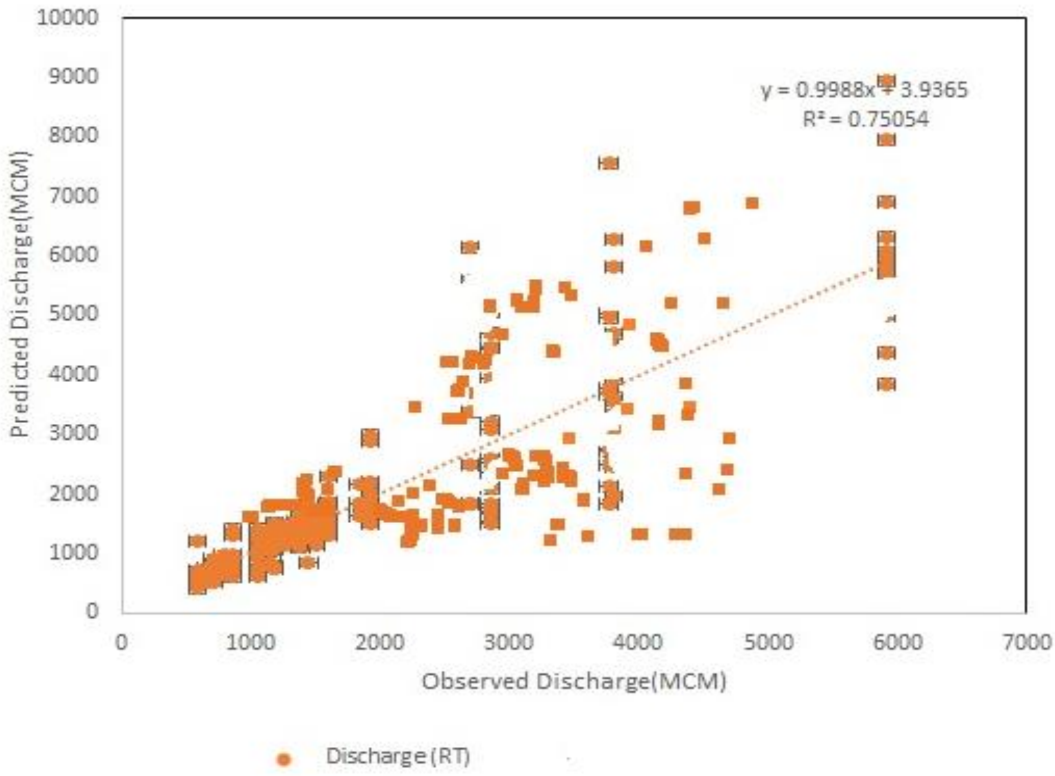

Figure 4.14: : Scatter plots of forecasted outflow using RT for scenario 5 and observed outflow.

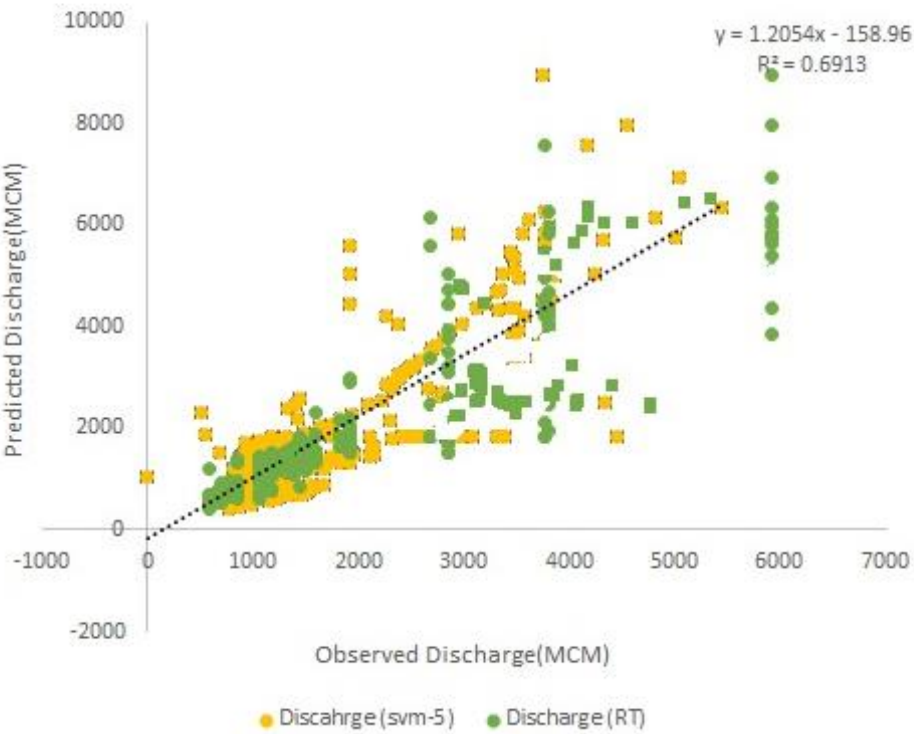

Figure 4.15: Scatter plots of forecasted outflow using SVM and RT models.

The observed and forecasted monthly streamflow’s by the SVM and RT models for the input combination data are shown in Figure. 4.14 in the form of scatterplot. It is clear from the scatterplots that the fit line coefficients a and b (assume that the fit line equation is (y=ax+b) of the SVM model are respectively closer to the 1 and 0 with a higher R2 value than that of RT model.

The estimates of the best models are compared in Figure. 4.15. It is obviously seen from the variation that graphs that SVM estimates closely follow the corresponding observed values while the Regression Tree considerably underestimate outflow values. Scatter plots clearly demonstrate that the SVM have less scattered estimates compared to Regression Tree models. therefore it is clear SVM is the best fit. The results indicated that the SVM performed better than the Regression tree models.

CHAPTER 5

CONCLUSION

5.1 CONCLUSIONS

Reservoir operation is a crucial part of reservoir management, and multi - reservoir theory and procedures have evolved through time as hydropower and reservoirs stations were created in the early twentieth century. Models based on the physical ideas and AI models developed from data mining technology are now categorised into two types based on the theory underlying the model: models based on the physical ideas and models based on physical concepts. A physical-based model, but in the other hand, is only useful if the simulation's operating rules truly represent how the system works in real life. In practise, several unknown factors, such as natural circumstances and artificial needs, influence the operation of a reservoir, and the operation often differs from the operating rules, restricting the applicability of such models. Artificial intelligence (AI) models, often defined as data-driven models, are capable of learning operating rules. Artificial intelligence (AI) models, also referred as data-driven models, may learn operating rules directly from a reservoir's historical operation data, increasing their robustness and able to handle with complicated situations.

Over the decades, traditional hydrological forecasting models have changed, and SVM has acquired prominence in this sector by offering accurate data forecasts for a variety of hydrological processes. The ability to accurately estimate changes in reservoir outflow can aid in the planning and management of reservoir water usage in the long run. In present study two distinct machine learning approaches; regression tree and SVM are used and compared to find the most accurate method for predicting reservoir outflow based on monthly hydrological records for the past 30 years. A variety of scenarios based on number of data inputs are evaluated to find the best parameters for reservoir outflow and for this evaluation, RMSE, MAE, MAPE, scatter index and R^2 indices are used to quantify the performance of the forecasted models. In summary, scenario 5 is the optimum combination of input data; it comprises of inflow, evaporation, water level, reservoir storage, previous inflow and previous outflow. The best SVM regression is with medium gaussian kernel function and is employed for the optimal scenario selection. The findings of present study revealed that all of the models performed well and were able to match the actual values.As a result of the comparative analysis of water outflow prediction through two algorithms, SVM is the best algorithms for outflow prediction. However, when performed individually the outflow of reservoir calculated from SVM algorithm is also compared with outflow data of Bhakra reservoir and outflow calculated from conventional method. It is observed that outflow of reservoir predicted by SVM algorithm give accurate results as compare to conventional method. Similarly, outflow of reservoir calculated from RT algorithm give more accurate result as compare to conventional method.This highlights its unique capabilities and benefits in detecting hydrological time series with nonlinear properties. The best input scenario is one that takes into account all input factors i.e. scenario-5. In comparison to Regression tree, the SVM model is able to estimate reservoir outflow precisely. This suggests that SVM has some generality and may be used as a model for reservoir outflow predictions.

5.2 Future scope

i) Future studies should include additional hydrological data, such as infiltration rates, transpiration rates, low inflow circumstances, and other pertinent characteristics, to produce more precise projections. Hopefully, data collectors will follow suit, enabling for the construction of a more robust SVM-based forecasting model.

ii) Furthermore, it is suggested that the methodology given in this study be used to future studies involving the modelling of other hydrological processes such as rainfall, rainfall, runoff, and so on.

iii) Use remote sensing data to test the ML-based method's generalizability in more ungauged basins in the future.

REFERENCES

[1] Hu Hussain, W.; Ishak, W.; Ku-mahamud, K.R.; Norwawi, N. Neural network application in reservoir water level forecasting and release decision. Int. J. New Comput. Archit. Appl. 2011, 1, 265–274.

[2] Valizadeh N, El-Shafe A, Mukhlisin M, El-Shafe AH (2011) Daily water level forecasting using adaptive neuro-fuzzy interface system with diferent scenarios: Klang Gate, Malaysia. In J Phys Sci 6(32):7379–7389

[3] Tokar AS, Markus M (2000) Precipitation-runoff modeling using artificial neural networks and conceptual models. J Hydrol Eng 5(2):156–161.

[4] Abdul Mokhtar S, Wan Ishak WH, Md Norwawi N (2014) Modelling of reservoir water release decision using neural network and temporal pattern of reservoir water level. In: Proceedings of the fifth international conference on intelligent systems, modelling and simulation, pp 127–130

[5] Seckin N, Cobaner M, Yurtal R, Haktanir T (2013) Comparison of artifcial neural network methods with l-moments for estimating food fow at ungauged sites. Water Resour Manage J 27(7):2103–2124

[6] Ahmadlou, M.; Karimi, M.; Alizadeh, S.; Shirzadi, A.; Parvinnejhad, D.; Shahabi, H.; Panahi, M. Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (anfis) and biogeography-based optimization (bbo) and bat algorithms (ba). Geocarto Int. 2019, 34, 1252–1272. [CrossRef]

[7] Naghibi, S.A.; Pourghasemi, H.R. A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping. Water Resour. Manag. 2015, 29, 5217–5236. [CrossRef].

[8] Bowden, G.J.; Maier, H.R.; Dandy, G.C. Optimal division of data for neural network models in water resources applications. Water Resour. Res. 2002, 38, 2-1-2-11. [CrossRef]

[9] Mullainathan, S.; Spiess, J. Machine learning: An applied econometric approach. J. Econ. Perspect. 2017, 31, 87–106. [CrossRef

[10] Najah, A.A.; El-Shafie, A.; Karim, O.A.; Jaafar, O. Integrated versus isolated scenario for prediction dissolved oxygen at progression of water quality monitoring stations. Hydrol. Earth Syst. Sci. Discuss. 2011, 8, 6069–6112. [CrossRef]

[11] Hipni, A.; El-shafie, A.; Najah, A.; Karim, O.A.; Hussain, A.; Mukhlisin, M. Daily forecasting of dam water levels: Comparing a Support Vector Machine (SVM) Model with Adaptive Neuro Fuzzy Inference System (ANFIS). Water Resour. Manag. 2013, 27, 3803–3823. [CrossRef]

[12] Hejazi, M.I., Cai, X.M., 2009. Input variable selection for water resources systems using a modified minimum redundancy maximum relevance (mMRMR) algorithm. Adv. Water Resour. 32 (4), 582–593.

[13] Nwobi-Okoye CC, Igboanugo AC (2013) Predicting water levels at Kainji dam using artifcial neural networks. Niger J Technol 32(1):129–136

[14] Khan MS, Coulibaly P (2006) Application of support vector machine in lake water level prediction. J Hydrol Eng 11(3):199–205

[15] Shafaei M, Kisi O (2016) Lake level forecasting using Wavelet-SVR, Wavelet-ANFIS and Wavelet-ARMA conjunction models. Water Resour Manage 30(1):79–97

[16] Onur Hınçal A, Burcu Altan-Sakarya A, Ger M (2011) Optimization of multireservoir systems by genetic algorithm. J Water Resour Manage 25:1465–1487

[17] Yang, T., Gao, X., Sorooshian, S., Li, X., 2016. Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled crossvalidation scheme. Water Resour. Res. 52, 1626–1651

[18] Mjolsness E, DeCoste D (2001) Machine learning for science: state of the art and future prospects. Science 293:2051–2055

[19] Simon, H. A. H. A., & Langley, P. (1995). Applications of machine learning and rule induction. Communications of the ACM, 38(11), 54-64. ACM. doi:10.1145/219717.219768

[20] Brutsaert W (2005) Hydrology, an introduction. Cambridge University Press, NY.

[21] Khalil A, Almasari M, McKee M, Kemblowski MW, Kaluarachchi J (2005a) Applicability of statistical learning algorithms in groundwater quality modeling. Water Resour Res 41:W05010.

[22] Solomatine DP, Shrestha DL (2009) A novel method to estimate model uncertainty using machine learning techniques, Water Resour Res 45:W00B11.

[23] Guergachi A, Boskovic G (2008) System models or learning machines? Appl Math Comp 204:553–567.

[24] Milly, P C D, Julio Betancourt, Malin Falkenmark, Robert M Hirsch, Zbigniew W Kundzewicz, Dennis P Lettenmaier, and Ronald J Stouffer. 2008. Climate change - Stationarity is dead: Whither water management? Science 319, no. 5863: 573-574.

[25] Shumway RH and Stoffer DS (2011) Time series analysis and its applications. Third edition. Springer. USA.

[26] Pulido-Calvo I, Roldan J, Lopez-Luque R, and Gutierrez-Estrada J C. Demand forecasting for irrigation water distribution systems (2003). Irr Drain Eng 129(6):422-431.

[27] Nourani V, Mehdi K, Akira M (2009) A multivariate ANN-wavelet approach for rainfall–runoff modeling. Water Resour Manage 23:2877–2894

[28] Tomas, K. and Milos, S., 2021. Adaptive Management of the Storage Function for a Large Reservoir Using Learned Fuzzy Models. *Water Resources*, *48*(4), pp.532-543.

[29] Yaseen, Z.M., El-Shafie, A., Jaafar, O., Afan, H.A. and Sayl, K.N., 2015. Artificial intelligence based models for stream-flow forecasting: 2000–2015. *Journal of Hydrology*, *530*, pp.829-844.

[30] Gavahi, K., Mousavi, S.J. and Ponnambalam, K., 2018. Comparison of two data-driven streamflow forecast approaches in an adaptive optimal reservoir operation model. *EPiC Series in Engineering*, *3*(1), pp.755-763.

[31] Bai, Y., Chen, Z., Xie, J. and Li, C., 2016. Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *Journal of hydrology*, *532*, pp.193-206.

[32] Al-Sudani, Z.A., Salih, S.Q. and Yaseen, Z.M., 2019. Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation. *Journal of hydrology*, *573*, pp.1-12.

[33] Yilmaz, B., Aras, E., Nacar, S. and Kankal, M., 2018. Estimating suspended sediment load with multivariate adaptive regression spline, teaching-learning based optimization, and artificial bee colony models. *Science of the Total Environment*, *639*, pp.826-840.

[34] Su, J., Wang, X., Liang, Y. and Chen, B., 2014. GA-based support vector machine model for the prediction of monthly reservoir storage. *Journal of Hydrologic Engineering*, *19*(7), pp.1430-1437.

[35] Hipni, A., El-shafie, A., Najah, A., Karim, O.A., Hussain, A. and Mukhlisin, M., 2013. Daily forecasting of dam water levels: comparing a support vector machine (SVM) model with adaptive neuro fuzzy inference system (ANFIS). *Water resources management*, *27*(10), pp.3803-3823.

[36] Fathian, F., Mehdizadeh, S., Sales, A.K. and Safari, M.J.S., 2019. Hybrid models to improve the monthly river flow prediction: Integrating artificial intelligence and non-linear time series models. *Journal of Hydrology*, *575*, pp.1200-1213.

[37] He, S., Gu, L., Tian, J., Deng, L., Yin, J., Liao, Z., Zeng, Z., Shen, Y. and Hui, Y., 2021. Machine Learning Improvement of Streamflow Simulation by Utilizing Remote Sensing Data and Potential Application in Guiding Reservoir Operation. *Sustainability*, *13*(7), p.3645.

[38] Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., Liu, X. and Zhuang, J., 2018. Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *Journal of Hydrology*, *565*, pp.720-736.

[39] Ahmadi, A., Karamouz, M. and Moridi, A., 2010. Robust methods for identifying optimal reservoir operation strategies using deterministic and stochastic formulations. *Water resources management*, *24*(11), pp.2527-2552.

[40] Hu, C.; Wan, F. Input Selection in Learning Systems: A Brief Review of Some Important Issues and Recent Developments. In Proceedings of the 2009 IEEE International Conference on Fuzzy Systems, Jeju Island, Korea, 20–24 August 2009; pp. 530–535

[41] Sharafati, A.; Khosravi, K.; Khosravinia, P.; Ahmed, K.; Salman, S.A.; Yaseen, Z.M.; Shahid, S. The potential of novel data mining models for global solar radiation prediction. Int. J. Environ. Sci. Technol. 2019, 16, 7147–7164. [CrossRef]

[42] Tools and Techniques with Java Implementations. San Francisco: Morgan Kaufmann Publishers

[43] Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer-Verlag New York, Inc., New York, NY, USA

[44] Collobert, R., Bengio, S., 2001. SVMTorch: support vector machines for large-scale regression problems. J. Mach. Learn. Res. 1, 143–160. http://dx.doi.org/10.1162/ 15324430152733142.

[45] Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapnik, V., 1996. Support vector regression machines. Adv. Neural Inf. Process. Syst. 155–161, doi:10.1. 1.10.4845.

[46] Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press.

[47] Schwefel, H.-P., 1981. Numerical Optimization of Computer Models. John Wiley & Sons, Inc.

[48] Kim, S.J.; Ryoo, E.C.; Jung, M.K.; Kim, J.K.; Ahn, H.C. Application of support vector regression for improving the performance of the emotion prediction model. J. Intell. Inf. Syst. 2012, 18, 185–202. [Google Scholar]

[49] Gunn, S.R. Support vector machines for classification and regression. ISIS Tech. Rep. 1998, 14, 5–16. [Google Scholar]

[50] Basak, D.; Pal, S.; Patranabis, D.C. Support vector regression. Neural Inf. Process.-Lett. Rev. 2007, 11, 203–224. [Google Scholar]

[51] 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. Int. J. Climatol. 30, 1256–1268. http://dx.doi.org/10.1002/joc.1954.

[52] Hong Y, Lee YJ (2009) Citing Websites. A loss function approach to model specification testing and its relative efficiency to the GLR test. Retrieved 3rd February 2013

[53] Smola AJ, Scholkopf B (2004) A tutorial on support vector regression. Stat Comput 14(3):1–24

[54] Zhang, D., et al., 2018a. Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. J. Hydrol.565, 720–736.

[55] Vens, C. et al. Inductive Databases and Constraint-Based Data Mining 365–387 (Springer, 2010)..

[56] Chavent, M. A monothetic clustering method. Pattern Recogn. Lett. 19, 989–996 (1998).

[57] Loh, W.Y.: Classification and regression trees. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. **1**(1), 14–23 (2011)

[58] Quinlan, J. Induction of decision trees. Mach. Learn. 1, 1–81 (1986).

[59] Breiman, L.: Classification and Regression Trees. Routledge, London (2017)

[60] Witten, I.H. and Frank, E. (2000). Data Mining: Practical Machine Learning

[61] Bahrami, S.; Wigand, E. Daily streamflow forecasting using nonlinear echo state network. Int. J. Adv. Res. Sci. Eng. Technol. 2018, 5, 3619–3625

[62] imdpune.gov.in

[63] www.bbmb.gov.in