DAM MANAGEMENT THROUGH VARIOUS MACHINE LEARNING AND DEEP LEARNING MODELS: A REVIEW

*C.M. Nisha *, N. Thangarasu*

Abstract

Dams or reservoirs play a significant role in managing drought, flood, hydro power generation, irrigation for agriculture activities, drinking purposes etc. Data from meteorological department like daily prediction and weekly prediction of rainfalls in catchment areas, monsoon effects and water levels in dams, inflow level, outflow rate of water from dams, temperature etc. are major criteria for deciding dam openings. Correct and timely data helps to take correct decisions for managing dams. In this review, several machine learning and deep learning algorithms like ANFIS, CANFIS, decision tree, artificial neural network (ANN), random forest, Particle swarm optimization, gradient boosting, XG Boost, multilayer perceptron, Genetic algorithm, CNN, RNN, SMLA, Artificial Bee colony Optimization are presented for effective dam or reservoir management.

Keywords: Deep learning, dam management, Artificial neural network, Convolutional neural network.

I INTRODUCTION

About three fourth of earth's surface is covered by water in different forms like liquid, ice and glaciers but approximately two percentage of water is consumable since sea water mainly consists salt. Water is very essential for drinking, irrigation, cleaning and hydro power generation. Due to varying climatic conditions, water will not be available in all seasons. So dams play a pivotal role in making human lives easy. Dams are effectively utilized for controlling drought, flood, irrigation and hydro power generation. Earlier, people used to settle nearby rivers because of water availability but now a days it is not secure to settle nearby rivers due to climatic conditions and human related activities[1].Several parameters are to be taken for effective management of dams. Rainfall, water levels in dam, evaporation rates, inflow rate, outflow rates, temperature, wind speed etc. are some important parameters to be considered for opening or closing shutters. Evaporation is one of the main factors which determine the water level in dams. The variation in evaporation may be due to changes in temperature, humidity, pressure in atmosphere and solar radiation. An overview of various machine learning and deep learning models used for flood control as well as dam management and statistical methods are reviewed in this paper.

Department of Computer Science,
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
*Corresponding Author

II REVIEW OF LITERATURE

Dam and reservoir systems can be successfully operated with efficient simulation models with optimum use of water sources[2].Reservoir level changes due to atmospheric (climate, pressure, and air-mass movement), natural (slope and topography of the catchments), social (demand, distribution), and environmental (sedimentation, water quality) conditions [3].It will be possible to prepare plans for droughts and floods by establishing a water supply plan in the downstream area of the dam and unlike existing physical models, only weather data and flow data are required; thus, the applicability of the model will be superior in areas where data acquisition is limited[4].

Various machine learning and deep learning algorithms used for flood and dam management are reviewed in this paper. Important Machine learning algorithms are Linear Regression, Random forest, Naive Bayes algorithm, support vector machine(SVM),Decision tree, Gradient boosting

algorithm, Ada boosting algorithm, Logistic regression etc. Some of the deep learning algorithms are CNN (Convolutional Neural Network), LSTM (Long Short-Term Memory), Multilayer Perceptron, GAN (Generative Adversarial Networks),DBN (Deep Belief networks),Restricted Boltzmann Machines (RBMs), RNN (Recurrent Neural Networks), Auto encoders, RBFN (Radial Basis Function Networks) etc.

Two models, Artificial neural network and CANFIS, were used to predict inflow data into Haditha dam in Iraq by [5] and it showed that CANFIS model outperformed ANN model and ANFIS.

Six machine learning algorithm models were used to forecast the discharge amount of the Soyang River Dam, and their performance was studied, in which LSTM was evaluated to predict the discharge with the best accuracy among the six algorithms[4]. Random forest, gradient boosting, RNN-LSTM, MLP, Decision tree, CNN-LSTM were the six algorithms used for comparison in the discharge prediction of multipurpose dam.

Extreme Gradient Boosting (XGBoost) which is a novel method for hydrological prediction was chosen to build the daily and monthly reservoir inflow prediction models of Sirikit Dam in Thailand[6].The reservoir inflow prediction models, are built totally 54 models for one-ahead prediction and 54 models for one-month ahead prediction, was run in different combination of input data. The XGBoost model has learned the data series and demonstrated reliable performance in reservoir inflow prediction. The established model used monthly inflow and monthly precipitation to predict the inflow of this study area.

Flash flood prediction of Golestan region in Iran was done [7] using CNN(convolution neural network) and RNN(Recurrent Neural network).Predictive models were developed and verified with a geo spatial database which contained records about past flood events and environmental aspects were constructed. The SWARA (step-wise weight assessment ratio analysis) method was used. The results showed that the CNN model performed slightly better than the RNN model in predicting future floods.

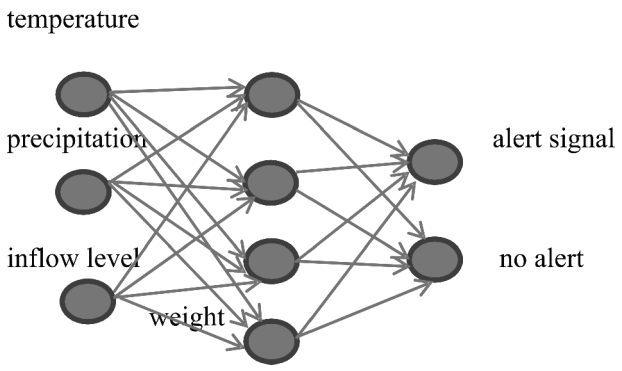
Soft computing is a combination of various intelligent computing methodologies which includes fuzzy logic, neuro computing and evolutionary computing. Conventional hard computing can perform on clear, precise and full truth data. But soft computing can work on data with imprecision, vagueness and partial truths[8].In a study conducted by [9] ,Shark Machine Learning Algorithm(SMLA algorithm) is used for reducing the deficit water volume in water releases from the dam and the water demand for drinking and irrigation. The results obtained after conducting a comparative study clearly indicated that SMLA algorithm is a better solution than Particle swarm optimization and Genetic algorithm for minimizing the shortage in demand and release of water in reservoir management system.

The different performance indicators for the models discussed are reliability, vulnerability, resiliency and shortage index in [2].Reliability is calculated by how many times demand is met while considering the simulation period. Vulnerability is the maximum deficit over simulation term and it is measured in percentage. Resiliency measures how the reservoir system recover from failures over total number of failures occurred. The shortage-index determines the model efficiency in meeting down steam flows.

Another set of statistical performance measures are Mean square error(MSE), Root mean square error (RMSE), coefficient of correlation (R) and coefficient of determination (R2) which are used for standalone models[10].

III METHODOLOGY

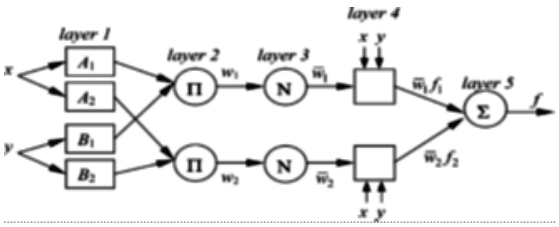
Artificial Neural Network(ANN): ANN is inspired and designed by the working of human brains which consists of billions of neurons. These neurons have cell body for processing information, dendrite to accept signals from other neurons and axons to pass the processed messages to other neurons. An ANN has nodes similar to neurons which accepts data and performs various operations on these data and passes the result to other nodes through links based on the activation function. Different parameters like inflow level, evaporation rate, temperature, precipitation etc. can be given as inputs and according to the priority different weights can be allocated and each node processes the information and decides whether to pass the information to the next node or not.



Fig(1):A Simple Design of ANN

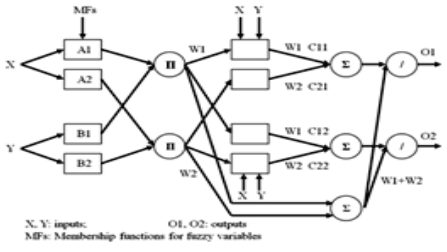
Multilayer Perceptron: A multilayer perceptron consists of 3 layers. The input layer which accepts values, hidden layers perform computation and output layer produces results. Several hidden layers process various parameters which help to solve complex problems. It was designed to overcome the drawback of perceptron which could not represent XOR gate. Bayesian regularization technique is used in the training of the MLPs. The weights and bias values are updated using Levenberg- Marquardt optimization technique. An amalgamation of squared errors and weights are reduced, and then produced the correct union so to construct a network[3].

Adaptive Neuro fuzzy system(ANFIS): It is a type of artificial intelligence which is a blend of fuzzy logic and neural networks. The ANFIS model consists of 5 layers[11].The first layer(fuzzy layer) nodes perform fuzzification by applying membership functions to the input data. The second layer is called product layer which is fixed and its output is generated from all incoming signals. The third layer(normalized layer) is also fixed and it calculate the sum of weights of all neurons in this layer. The fourth layer known as de-fuzzy layer calculates a parameter function on layer 3 output. The fifth layer is named as total output layer finds the aggregated sum of all the signals.



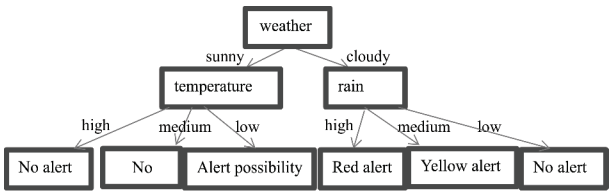
Fig(2):ANFIS structure[11]

Coactive Neuro-fuzzy inference system (CANFIS)-ANFIS has multiple inputs and single output. But CANFIS is a modification of ANFIS with multiple inputs and multiple outputs. A fuzzy based neuron is used in CANFIS. Fixed nodes are represented with circles and variable nodes with rectangles. The CANFIS model combines fuzzy inputs and modular neural network by which estimation of difficult functions are performed. The first layer consists of input variables, the middle layer performs fuzzy rules and the third layer denotes output variables[8].The learning capability of CANFIS model can be improvised by genetic algorithms which helps to select important features of the training data.



Fig(3):CANFIS Structure [12]

Decision Tree: Decision tree consists of a tree structure in inverted form where branches grow downwards. Leaf nodes are the final results or decisions. There are mainly two types of decision trees. One is classification trees which are used for predicting which category the outcome belongs. Second one is regression tree in which the outcome is of continuous type.



Fig(4):Example of Decision Tree

Random Forest: Random forest is made up of several decision trees which consists of summation of computational power of all decision trees. The random forests are trained using bagging method(bootstrap aggregation) and boosting method. Bagging method in which several training sets are used for creation of decision trees and average of these models are taken for prediction. In boosting method, net errors on the fitted models are calculated and at successive steps errors are reduced by changing weights. The RF is used to join several weak classifiers to build a strong classifier that can overcome the limitations of errors occurring due to single decision trees and boost the correctness of classification[13].

Gradient Boosting: Gradient boosting is a machine learning algorithm used for tabular datasets. Gradient descend is utilized in gradient boosting machines (GBM), for reducing loss functions by finding the weak learners in some certain areas where low performance is occurring[14]. Firstly the average of actual values is taken as prediction value. Then the difference of actual values and predicted value is calculated to find residuals as target. Fit a model on these residuals to find predicted residuals. The default prediction is updated using new predicted residuals.

New prediction=Default prediction +learning rate*New predicted residuals.

Then new predicted results are obtained and first iteration is completed. This process is repeated several times to create more trees for reducing the error rate and final prediction is obtained.

XGBoost: [6] points out that the prediction of reservoir inflow at time t (pit) by XGBoost is the additive sum of all previous predictions made by the model. It is implemented using Gradient boosted decision trees. All independent variables are assigned weights and fed into the decision tree for predicting results. This ensemble learning method has multiple learners and each learner's error is recognized and the errors are adjusted. A new metric called similarity score is used in XGBoost algorithm. For finding similarity score residuals are calculated by taking the difference of predicted values from observed values. Then similarity score is calculated using the formula,

$$\text{similarity score(sim)} = \frac{(\text{sum of residuals})^2}{\text{Number of residuals}+\lambda} \quad (1)$$

where λ is regularization parameter.

Then gain values are calculated to know whether trees can be split or not using the formula

$$\text{Gain value}=\text{sim}_{\text{left}}+\text{sim}_{\text{right}}+\text{sim}_{\text{root}}. \quad (2)$$

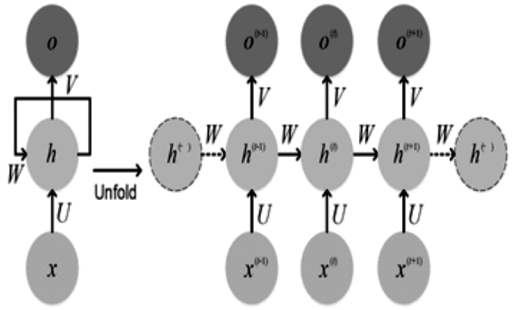
The trees are built until the required numbers of models have built. Then output values are calculated using the formula,

$$\text{Output value} = \frac{\text{sum of residuals}}{\text{Number of residuals}+\lambda} \quad (3)$$

IV RNN (RECURRENT NEURAL NETWORK)

Feed Forward Network(FFNN) has several disadvantages which was overcome by Recurrent neural network. FFNN can learn amid training phase and it depends on the current inputs. But RNN is able to remember the past

patterns from previous inputs and it depends upon past as well as present inputs[7]. RNN consists of mainly three types of layers. Input layer accepts data, middle layers where processing are done and output layer generates results. The middle layers consists of weights, bias and activation functions and it can utilize previous outputs as inputs. RNN has a variation namely Long short-term memory(LSTM) which is used for avoiding unnecessary parts using forget gate and remembering only essential inputs. GRU(Gated Recurrent Unit) is also a variation of RNN and works more fast compared to LSTM and needs less memory. Another variation of RNN is Bidirectional RNN which uses information from both ends to receive outputs. LSTM and GRU are used for overcoming the problem of vanishing gradient of RNN. Vanishing gradient problem occurs due to degeneration of learning by lowering of gradients which leads to lack of training of nodes in the layers.



Fig(5):An RNN architecture[15]

IV CNN (CONVOLUTIONAL NEURAL NETWORK)

Convolutional Neural Network(CNN) is mainly used for image identification and classification. CNN has three layers. The first layer is known as convolution layer has an input image which is in the form of matrix upon which a filter of 2x2 matrix is passed to get a feature map for extracting features. The feature map is then passed to the pooling layer to reduce the feature by max pooling or average pooling. The features extracted in previous layers are classified in the fully connected layer. CNNs can be utilized for extracting features from satellite images for flood prediction.[1]has shown that CNN is efficient in identifying areas of Iran which are easily susceptible to floods by taking various flood conditioning factors. The ten factors used for study were slope, height, aspect, bend, contour curving, precipitation, land characteristics, usage of land, remoteness from roads, and detachment from rivers.

SMLA (Shark Machine Learning Algorithm)

SMLA is a meta heuristic algorithm which works on the principle of Shark's ability to find prey by smelling and its movement. The shark has the ability to smell the presence of prey nearly one kilometre away and it can process the information in the brain to find the correct location[9].The movement of the shark, the smell sensing ability and its environment are mathematically modelled in the algorithm. The monthly water deficit is minimized by the expression

$$\text{Min } Z=\sum_{t=1}^{12} (D_t - R_t)^2. \quad (4)$$

Here t is used to represent months, Dt shows monthly water demands and Rt shows monthly water release. The reservoir storage capability values has to maintain a limit on storage which is expressed by

$$S_{(t+1)} = S_t + I_t - R_t - Lt \quad (5)$$

S(t+1) is the monthly end storage condition, St is monthly initial storage, It is monthly inflow, Rt is monthly water release, Lt is losses from surface water. Penalty function is used for handling constraints of optimization problem.

$$\text{penalty 1} = 0 \quad \text{if } St > S_{min}$$
$$= C1(S_{min} - S_t)^2 \quad \text{if } St < S_{min} \quad (6)$$

$$\text{penalty 2} = 0 \quad \text{if } St < S_{max}$$
$$= C1(Smin - St)^2 \quad \text{if } St > S_{max} \quad (7)$$

The final objective function is

Minimum value for Y = Z + penalty 1 + penalty 2 (8)

Here Y is used for representing accurate prediction of models working.

Genetic Algorithm (GA)

The Genetic algorithm is based on genetics formulated by Charles Darwin. He proposed survival of the fittest theory which is the back bone of Genetic algorithm. The main idea is that strong elements are accepted while weak elements are removed. The [2] review paper points out that the amalgamation of Genetic algorithm, artificial bee colony (ABC) and other models exposed the ability of rapport between climatic parameters and vaporization pattern. A population is derived From different generations until no improvement is generated[16].The paper discusses about selection which finds out parent for next generation. Next step is crossover which joins the parents to create children for consequent generation. Mutation is the next step which modifies parents to get variant children. In Genetic algorithm chromosomes are a series of ones and zeros and gene is a single bit inside a chromosome. Initialization of chromosomes is done inside the population and the fittest solution is achieved by successful modification and evolution through generations.

Particle Swarm Optimization (PSO)

Particle swarm optimization(PSO) depends upon the flying of swarms in which each swarm obtains the best position for flying[9].The birds fly in group with their position in such a way that the shape formed exist, sometimes change direction and fall on the prey in a comfortable manner. The PSO algorithm first initializes each element and operates on the position of each element by changing the position in each stage to find the fittest solution[10].

Artificial Bee Colony Optimization (ABC)

ABC algorithm is inspired by the working of bees where division of labour in searching food is copied. The idea used in ABC method is to allot multiple tasks for problems based on dam optimization[17].There are three types of bees associated with ABC method. The employed bee finds food and number of employed bees is the same as that of food sources. On lookers bee lookout for the performance of employed bees and determines whether to accept food sources. Scout bees search for new food sources. To implement ABC algorithm, firstly random positions are set for employee bees and for onlooker bees. Employee bee phase is done next by doing collection of food at their sources. Then onlooker bees perform the evaluation of food extraction. On exhaustion of food new exploration can be started.

V CONCLUSION

This review focuses on familiarization of major machine learning and deep learning algorithms for management of dams. Several other algorithms can be effectively utilized like Ant colony optimization(ACO),Differential evolution(DE) and Cat Swarm optimization(CSO) [18].Vaporization rates were collected on monthly basis from different stations in Iraq using artificial intelligence models by[19].

Nearly 75 percentage of the data of different parameters collected can be used for training models and remaining 25 percentage can be used for validating data. Different prediction models performance are compared using the statistical methods like RMSE,MSE,NSE,R2 and R [6].The main aim of models is to reduce the error rate and predict the possibility of dam opening and closing based on different parameters collected. Thus different models and algorithms can be used in evaluation and estimation of water volume, its possibility of change in water level and correctness can be evaluated by different statistical parameters.

REFERENCES

[1] K. Khosravi et al., “Convolutional neural network

approach for spatial prediction of flood hazard at national scale of Iran," J. Hydrol., vol. 591, p. 125552, Dec. 2020, doi: 10.1016/j.jhydrol.2020.125552.

[2] M. F. Allawi, O. Jaafar, F. Mohamad Hamzah, S. M. S. Abdullah, and A. El-shafie, "Review on applications of artificial intelligence methods for dam and reservoir-hydro-environment models," Environ. Sci. Pollut. Res., vol. 25, no. 14, pp. 13446–13469, May 2018, doi: 10.1007/s11356-018-1867-8.

[3] F. Unes, "DAM RESERVOIR LEVEL MODELING BY NEURAL NETWORK APPROACH: A CASE STUDY," Neural Netw. World, p. 15.

[4] J. Hong et al., "Comparison of Machine Learning Algorithms for Discharge Prediction of Multipurpose Dam," Water, vol. 13, no. 23, p. 3369, Nov. 2021, doi: 10.3390/w13233369.

[5] M. F. Allawi, I. R. Hussain, M. I. Salman, and A. El-Shafie, "Monthly inflow forecasting utilizing advanced artificial intelligence methods: a case study of Haditha Dam in Iraq," Stoch. Environ. Res. Risk Assess., vol. 35, no. 11, pp. 2391–2410, Nov. 2021, doi: 10.1007/s00477-021-02052-7.

[6] P. Dornpunya et al., "THE RESERVOIR INFLOW PREDICTION OF SIRIKIT DAM USING ARTIFICIAL INTELLIGENCE WITH MACHINE LEARNING: EXTREME GRADIENT BOOSTING TECHIQUE," p. 9, 2021.

[7] M. Panahi et al., "Deep learning neural networks for spatially explicit prediction of flash flood probability," Geosci. Front., vol. 12, no. 3, p. 101076, May 2021, doi: 10.1016/j.gsf.2020.09.007.

[8] M. Saemi and M. Ahmadi, "Integration of genetic algorithm and a coactive neuro-fuzzy inference system for permeability prediction from well logs data," Transp. Porous Media, vol. 71, no. 3, pp. 273–288, Feb. 2008, doi: 10.1007/s11242-007-9125-4.

[9] M. F. Allawi, O. Jaafar, M. Ehteram, F. Mohamad Hamzah, and A. El-Shafie, "Synchronizing Artificial Intelligence Models for Operating the Dam and Reservoir System," Water Resour. Manag., vol. 32, no. 10, pp. 3373–3389, Aug. 2018, doi: 10.1007/s11269-018-1996-3.

[10] K. S. M. H. Ibrahim, Y. F. Huang, A. N. Ahmed, C. H. Koo, and A. El-Shafie, "A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting," Alex. Eng. J., vol. 61, no. 1, pp. 279–303, Jan. 2022, doi: 10.1016/j.aej.2021.04.100.

[11] R. Shafaei, M. Rabiee, and M. Mirzaeyan, "An adaptive neuro fuzzy inference system for makespan estimation in multiprocessor no-wait two stage flow shop," Int. J. Comput. Integr. Manuf., vol. 24, no. 10, pp. 888–899, Sep. 2011, doi: 10.1080/0951192X.2011.597430.

[12] P. Tahmasebi and A. Hezarkhani, "A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation," Comput. Geosci., vol. 42, pp. 18–27, May 2012, doi: 10.1016/j.cageo.2012.02.004.

[13] X. Li, Z. Wen, and H. Su, "An approach using random forest intelligent algorithm to construct a monitoring model for dam safety," Eng. Comput., vol. 37, no. 1, pp. 39–56, Jan. 2021, doi: 10.1007/s00366-019-00806-0.

[14] S. C. Ibañez, C. V. G. Dajac, M. P. Liponhay, E. F. T.

Legara, J. M. H. Esteban, and C. P. Monterola, "Forecasting Reservoir Water Levels Using Deep Neural Networks: A Case Study of Angat Dam in the Philippines," Water, vol. 14, no. 1, p. 34, Dec. 2021, doi: 10.3390/w14010034.

[15] W. Feng, N. Guan, Y. Li, X. Zhang, and Z. Luo, "Audio visual speech recognition with multimodal recurrent neural networks," in 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, May 2017, pp. 681–688. doi: 10.1109/IJCNN.2017.7965918.

[16] W. El Harraki, D. Ouazar, A. Bouziane, and D. Hasnaoui, "Optimization of reservoir operating curves and hedging rules using genetic algorithm with a new objective function and smoothing constraint: application to a multipurpose dam in Morocco," Environ. Monit. Assess., vol. 193, no. 4, p. 196, Apr. 2021, doi: 10.1007/s10661-021-08972-9.

[17] S. M. Saab, F. B. Othman, C. G. Tan, M. F. Allawi, and A. El-Shafie, "Review on generating optimal operation for dam and reservoir water system: simulation models and optimization algorithms," Appl. Water Sci., vol. 12, no. 4, p. 73, Apr. 2022, doi: 10.1007/s13201-022-01593-8.

[18] O. M. Sorkhabi, B. Shadmanfar, and E. Kiani, "Monitoring of dam reservoir storage with multiple satellite sensors and artificial intelligence," Results Eng., p. 100542, Jul. 2022, doi: 10.1016/j.rineng.2022.100542.

[19] M. Al-Mukhtar, "Modeling the monthly pan evaporation rates using artificial intelligence methods: a case study in Iraq," Environ. Earth Sci., vol. 80, no. 1, p. 39, Jan. 2021, doi: 10.1007/s12665-020-09337-0.