CHENNAI WATER STORAGE MANAGER

FINAL REVIEW REPORT

for

Technical Answers for Real-world Problems (ITE3999)

submitted in partial fulfillment of

B.Tech (Information Technology)

by

Senthil Kumar R (18BIT0048)

Somya Sheti (18BIT0489)

Fall 2021

Under the Guidance of

Prof. Daphne Lopez

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Information Technology and Engineering

December 2021

Water Storage Manager: Using Artificial Intelligence to forecast water-level of Chennai's reservoirs

Senthil[1], Somya[2]

[1,2] **Department of Information Technology, VIT University, Vellore, Tamil Nadu, India**

Abstract

Deep Learning Networks provide cutting-edge solutions for real life problems that require human logic and reasoning. Water shortage in Chennai is a severe issue that needs to be addressed right away. In this paper we propose a hybrid LSTM and CNN1D model to effectively manage water supply to Chennai that optimises water storage level from Chennai's primary water-supply reservoirs

I. INTRODUCTION

Chennai is one of the most important metropolitan cities of India. It holds approximately 10.6 million people. The city does not have any perennial rivers flowing through it, and hence the city's municipal department relies on reservoirs and lakes located in and around the capital of Tamil Nadu.

Chennai's water potential relies on the annual monsoon rains that replenish the city's reservoirs and lakes. With the population increasing rapidly over the last few years, Chennai is prone to acute water shortages. The city has been unable to cope up with the growing demand for water, and got hit by the worst drought in the summer of 2019 (due to severe water shortage in the reservoirs). Surprisingly, the same city was also critically hit by the late monsoon rains of 2015.

Hence, an immediate need to properly manage Chennai's water sources has arisen. The following paper focuses on implementing Deep Learning Techniques to manage storage level of water from these four water-supply reservoirs. In the long run, the proposed techniques will be used as a driving force to address the Chennai water crisis.

II. BACKGROUND

By 2026, Chennai city's overall water demand will be 2248 MLD (Million Litres per Day), which includes Household water demand, industrial supply and other commercial/ non-commercial water requirements. With the civilian population crossing 1.20 crore, it is highly likely that the water-potential of the Chennai Metropolitan Area (CMA) will eventually get exhausted by 2030!

One of the major problems faced in Chennai Metropolitan Corporation (CMC) is the inefficiency of the administration to optimise water supply with respect to the increasing water demand. This problem is not confined to Chennai alone, several cities across the country face the same issue. Chennai is powered by 4 primary water reservoirs:

- ❖ Poondi reservoir
- ❖ Chembarambakkam reservoir
- ❖ Red Hills reservoir
- ❖ Chozhavaram reservoir

With the groundwater nearing to a complete exhaustion, almost 30% percent of Chennai's water demands are satisfied by water tankers. Chennai receives about 70% of the water supply from the lake's such as Poondi, Chozhavaram, Red Hills and Chembarambakkam.

Figure 1: Chennai City and its water supply reservoirs

III. Literature Survey

M. K. Tiwari and C. Chatterjee conducted research which utilized wavelet and bootstrapping techniques to develop a WBANN model[3] (hybrid Wavelet Bootstrap ANN). This model produces hourly forecasts of floods.Comparative analysis of three other ANN models was also done in this paper. Their results were such that BANN and WBANN models outperformed the other primitive models like ANN and WANN. It was also noted that the performance of the traditional ANN's were poor in most of the cases.

Changhyun Choi et al. published a paper on water level forecasting models by using various Machine Learning algorithms [6]such as ANN, SVM, Decision Tree and Random forest .The largest inland wetland in south korea known as the Upo was selected as the study area. The main objective was to develop prediction models using different machine learning models and survey them. Based on the experiments conducted in study,among all the other models tested, Random Forest was the best performing model.This model consisting of 492 different trees was used for simulating the water level in Upo wetland.

F.-J. Chang modeled a real-time water level predicting system by using RNN as a solution to the problem in controlling urban floods [1] . BPNN, Elman NN, NARX are the networks used to forecast the water levels of which the latter two are dynamic and the other, Static. The resulting ANN are then evaluated by their performance.
BPNN considers only the observed data while producing the output whereas the dynamic ones also considers the time delay units. Hence it was observed that the dynamic networks are performing effectively in discovering the long-term dependencies and removing the fluctuations in the outputs when compared with the static network. The Elman network's backward connections could remember the past

values of the layers which makes it highly dependent on input values. This results in eliminating fluctuations in the output values when compared with the NARX Network's output layer . As a result, these three ANNs are coupled in a statistical manner to forecast the water level in order to control floods.

Guoyan Xu et al. published a paper on Water Prediction model based on ARIMA-RNN [4] .The dataset used in this paper contains the water level in lake Taihu along with the various environmental factors affecting the water level. This paper proposes a scheme which uses both ARIMA and RNN models,the linear component of data is predicted by the ARIMA whereas the RNN model is utilized to predict the non-linear components of data.

Y.-T. Chang et al. proposed a neuro-fuzzy based solution for water level prediction in Shihmen reservoir to mitigate the effects of floods [2]. The 5-layered ANFIS system formulates a 1 to 3 hour-ahead water level prediction mechanism by using fuzzy modelling with the help of the gauge stations dataset. The system, upon successful training, yields a correlation coefficient of 0.99 which is highly acceptable for a time-step-forward based prediction. Using this, the water level of the Shihmen reservoir can be accurately predicted for T+3 hours, thus solving the complex water-flooding problem with a neuro-fuzzy inference system.

Panjaporn Truatmoraka et. al. proposed a water level prediction system using BackPropagationNet[5]. It is a simple, yet highly effective neural network that uses backpropagation to predict the Chao Phraya river's water level, which flows across central Thailand. With outflow and capacity as its baseline, the 3 layered neural network uses backpropagation to predict the water level with a very high accuracy. This served as an inspiration to our ANN model due to its very low RMSE loss factor. The results infer that the Arima-RNN model has a better predictive effect on future water levels compared to other models and can more accurately understand the overall trend and amplitude fluctuations.
The following table suggests the diverse applications in watermarking schemes available offered by various techniques.

Paper no.	Name of algorithm/technique	Advantage	Disadvantage
1	Multiple samples of RNN	Predictions are both static and dynamic	Static models have time-varying trade-offs, dynamic models have disability in accuracy
2.	neuro-fuzzy ANFIS	Extremely good performance wrt. correlation coeffecient	Only calculates T+3 hours
3	Wavelet Bootstrap ANN	Performs better than BANN	Can only predict floods on an hourly basis
4.	ARIMA-RNN	Combines linearity of ARIMA with time-non-linearity of RNN	Slight amount of noise and distortion of results in prediction
5.	BackPropagationNet	Predicts river water level with high accuracy	Takes only inflow and capacity into account
6.	Various ML techniques	Random Forest performs well with Upo wetland	Cannot be remodelled to tropic conditions and metropolitan water resources

Table 1: Literary survey papers

IV. PROPOSED ALGORITHM

Aspect name	Specification
Data	Daily SIRO data for 2004-2018
Preprocessing	Min-Max normalisation, batch cleaning
Train-Test split	85%
Model	LSTM-CNN1D model
Layers	3 Layers
Trainable parameters	20284
No. of epochs	300
Average time taken (per sample)	187µs per sample

Table 2: Model specifications

Data is a key factor for any Deep Learning model. SIRO (Storage Inflow Rainfall Outflow) data of the four reservoirs is web-scraped from CMWSSB's public data portal. The daily SIRO data for 2004-2018 is normalised and batch-cleaning is also performed. A time series forecasting model which combines both LSTM and CNN1D is proposed. LSTM is chosen due to the fact that traditional RNN methods face the issue of exploding/vanishing gradients while the number of weights keeps on increasing (vanishing gradient problem). On the other hand, LSTM doesn't suffer from this problem as it uses a memory vector known as the forget gate which sorts out the relevant and irrelevant information and pushes forward only the relevant information towards the cell state.

In order to fine-tune the model's prediction, the CNN1D is fused with the model, thereby automatically creating informative representations of the given data. The fused model, being noise-resistant, extracts deep and informative features that are independent of time.

The proposed model first takes in the input layer which is a 5x4 matrix (where rows represent the data of the previous 5 days and the columns represent SIRO parameters). The model processes this data parallelly using LSTM and CNN-1D and then concatenates the results from both.The feature maps are reduced to a single one-dimensional vector by using the flatten layer and hence the final level is predicted.

Storage level is predicted individually for each location and then combined at the end. This is done because, environmental factors vary a lot in each of these locations such as the temperature in Chozhavaram is normal whereas the temperature in Chembarambakkam is pretty high.So to improve the accuracy of our model we get the final level from the combination of the individual storage levels from each reservoir. The procedure to calculate them is depicted in Figure 2.

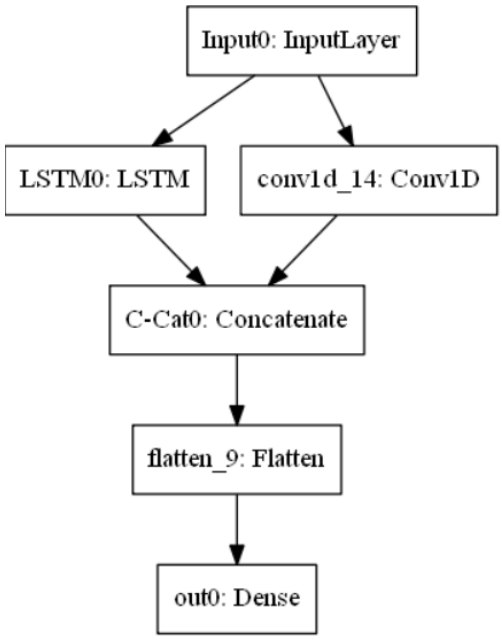

Figure 2: Layer-architecture of proposed Model

V. EXPERIMENTAL RESULTS

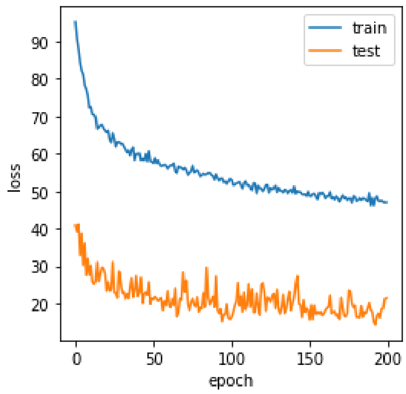

Figure 3: Loss-Epoch graph of the trained model

After 200 epochs, as shown in figure 3, the model is able to map a definite pattern between the input data and the storage level of each reservoir. Surprisingly, the model learns to adapt itself based on the season, although season-related information is not fed. As depicted in figure 4, the orange line (which shows level for all days in the year 2016 test data) almost overlaps the blue line (which shows real world level data). Hence, it can be conclusively said that the model is able to perform as perfectly as humans.

The method used for error estimation and analysis here is Root Mean Squared error (RMS). The total loss calculated from all reservoirs is 234.81 mcft (which is also due to the evaporation factor).The individual loss from each reservoir is presented in table 3.

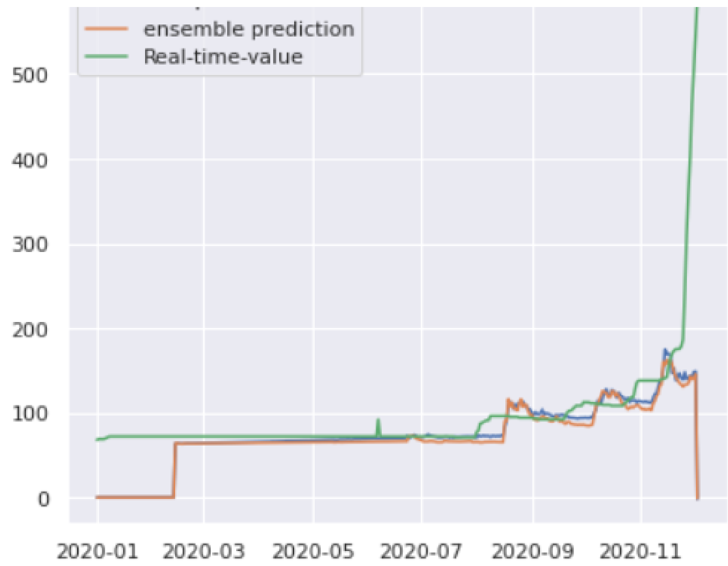

Figure 4: Real-world and predicted water level for Poondi reservoir in 2020

S.no	Reservoir	Absolute Error (mcft)
1.	Chembarambakkam	63.411
2.	Chozhavaram	8.498
3.	Poondi	115.057
4.	Red Hills	45.875

Table 3: Error statistics

VI. COMPARATIVE STUDY

The resulting system is portraying industry-grade accuracy with respect to predicting the future. The following are the features of the developed system which overcome the shortcomings of the previous works:

a. 15-day prediction

The developed system beats Wavelet Bootstrap ANN[3], neuro fuzzy ANFIS[2] and ARIMA-RNN[4] by giving a window of 15 days, thereby beating hourly forecasting with semi-month predictions.

b. Adaptable to weather dynamics

The model has been trained with data from 2004 to 2020, scaling under all scenarios. This allows the model to be highly useful as it predicts spontaneous showers, as well as seasonal climate trends, hence performing significantly better than various ML models previously pitched[1].

c. Precision

Models developed on reservoirs like Poondi and Cholavaram show incredibly high accuracy even in highly skewed moths like November and December, hence proving to be better than BPN[5] and RNN models[6]

d. **Flood-Drought awareness**
Outperforming all previous works, the 15-day forecast also gives us warnings of possible water-scarcity and water-flooding so that the government and its associated organisations can make preemptive steps beforehand.

VII. CONCLUSION AND FUTURE WORK

The proposed Deep-Learning model has allowed us to attain human-like performance. Thus, water supply for CMA can be intelligently managed by using the proposed Neural Network model. This model can be extended to imply water management systems in other metropolitan areas.

Our Future work will involve solving the problems mentioned below:

- Implementing the same proposed model to predict the level of water in rivers, lakes and dams for flood control and resource management.
- Extending the proposed model's utility to optimise electricity generation, irrigation and domestic water supply.

VIII. REFERENCES

[1] F.-J. Chang, P.-A. Chen, Y.-R. Lu, E. Huang, and K.-Y. Chang, "Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control," *Journal of Hydrology*, vol. 517, pp. 836–846, 2014.

[2] Y.-T. Chang and F.-J. Chang, "Adaptive neuro-fuzzy inference system for prediction of water level in reservoir," *Advances in Water Resources*, vol. 29, no. 1, pp. 1–10, 2006.

[3] M. K. Tiwari and C. Chatterjee, "Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach," *Journal of Hydrology*, vol. 394, no. 3-4, pp. 458–470, 2010.

[4] Guoyan Xu, Yi Cheng, Fan Liu,Ping Ping, Jie Sun, "A Water Level Prediction Model Based on ARIMA-RNN".

[5]Panjaporn Truatmoraka, Narongrit Waraporn, Dhanasite Suphachotiwatana,"Water Level Prediction Model Using Back Propagation Neural Network".

[6] Changhyun Choi, Jungwook Kim, Heechan Han, Daegun Han and Hung Soo Kim,"Development of Water Level Prediction Models Using Machine Learning in Wetlands: A Case Study of Upo Wetland in South Korea".