Water Table Depth Prediction Using Machine Learning

Shelley Gupta[1], **Sanjay Singh Bisht**[2], **Prakhar Jain**[3] and **Adarsh Mishra**[4]
Department of Information Technology, ABES Engineering College, Ghaziabad,
19th KM Stone, NH-09, Ghaziabad 201009 UP India
[1]shelley.g17@gmail.com, [2]sanjay.18bci1001@abes.ac.in, [3]prakhar.18bit1041@abes.ac.in, [4]adarsh.18bit1085@abes.ac.in

Abstract --**Water table depth prediction in agricultural regions is difficult due to complex and varied hydrogeological properties, boundary conditions, and human activities, as well as nonlinear interactions among these elements. As a result, as a substitute to costly models, this study established a model framed from layers of Long short-term memory network (LSTM) based innovative series of time framework along with fully connected layer. The first LSTM layer employed a dropout approach. The suggested model was tested and assessed using data from 14 years (2000–2013) in China's Hetao Irrigation District's five auxiliary field of the northern desert. The suggested model predicts water table depth based on diversion of evaporation, water diversion, temperature, time and precipitation. The experiment divides data of 14 year as training and validation dataset. The conventional feed-forward neural network (FFNN), that has acquired comparatively low (0.004–0.495) R2 scores, the suggested framework has acquired higher R2 scores of (0.789–0.952) in depth prediction of water, demonstrating that the suggested framework can conserve and gain past data well.**
The effectiveness of the dropout approach is further explored, as well as the design of the suggested model. The results of the experiments suggest that using the dropout strategy can greatly reduce overfitting. Furthermore, comparisons of the proposed model's R2 scores with the R2 scores of the Double-LSTM framework ranges from 0.170 - 0.864 which depicts the appropriateness of suggested architecture contributing to high capacity of learning over the data of series of time. As a result, the suggested model may be utilized to forecast depth of water table as an alternative to hydrogeological data, particularly in places where hydrogeological data is scarce.

Keywords: *Water table depth, Machine learning, Long short-term memory network, Recurrent neural network*

I. INTRODUCTION

THE primary source of fresh water is under-ground water that meets much of the commercial and residential water demands. People all throughout the world rely on ground water for drinking water and agricultural needs. Groundwater resource accessibility and availability are intricately linked to socio-economic development. For example, groundwater meets 22% of domestic freshwater demands, 69% of agricultural water needs, and 9% of industrial water needs (UNESCO Report, 2022). While worldwide water usage will rise considerably in the upcoming time, recent study shows that groundwater levels are falling in various regions of the globe. Nonetheless, seasonal variations in water table levels may occur owing to evapo-transpiration extraction, hydraulic qualities, and other natural processes. Depending on the season, the water table might alter (rise or decrease) in depth. During winter, when accumulated snow melts and rainfall is plentiful, surface water infiltrates into the earth, raising the water table. The water table drops due to evapo-transpiration as water-loving plants begin to develop again in the spring and precipitation gives place to scorching, dry summers. Due to this, predicting the depth of the groundwater table is a difficult undertaking.

Despite the fact that groundwater offers a great deal of potential for socioeconomic growth, it has received little attention. For many nations that are mostly reliant on groundwater may endure lengthy droughts in the future. To avoid this, water management decisions must be based on fast, trust-worthy, and responsive data. Improving tools for an accurate forecast of seasonal changing levels of groundwater is one option for better groundwater resource management.

Advances in modeling, computational skills, and data processing enabled us in improving understanding of very multiplex natural systems. The usage of machine learning approaches to the area of hydrology has received a lot of attention. Most of the strategies outlined in the literature are ineffective because of the involvement of noisy and sparse samples. As a consequence, this research developed a novel time-series model which is based on LSTM (Long short-term memory), computationally efficient alternative to artificial models available. The proposed model with layer of LSTM has utilized a dropout method. The major objectives of our research work are stated below:

- To extract dataset using web scraping in CSV format for training our model. The dataset is filtered which helps in removing anomalies like null values, etc.
- To derive a pattern from dataset by plotting a graph to further implement LSTM model.
- To define and implement a class-based LSTM model architecture.
- To train and validate the model in dataset along with plotting variation from actual values.

II. RELATED WORK

Machine learning models based on statistics are widely used in water depth prediction techniques and applied to analyze time-series data. Since 1991, hydrology has used the ANN paradigm of machine learning. According to the researchers [1], machine learning models beat artificial models available presently. The approach [2] contrasted particle-swarm optimized SVM [14, 15] which stands for Support Vector Machine with artificial models for groundwater depth forecasting. The models of ML have demonstrated that these models are straightforward which calibrate and interpret.

Authors [4] estimated future groundwater levels in Selangor, Malaysia, using time-series data. The Xgboost approach is comparable against ANN and SVR with regard to mean absolute error and root-mean-square deviation. Underground water level at Ljubljana polje Aquifer was examined by the research [3]. The research [5] is focused on predicting changes in India's changing coastal aquifer. According to [6], a technique for predicting depth and flowing levels of water in Shanon located in Ireland has been developed. Time series data was used as an input, spanning 30 years from 1983 to 2013. Depth of the water at lower part of Shanon water station are predicted using a convolutional neural network (CNN) from 2013 to 2080 by the system. To estimate groundwater levels in the Konan Kochi Prefecture, Japan, [8] utilized many artificial neural networks (ANN) and linear regression. 68% of the research area was made up of paddy fields. A machine learning model was developed utilizing multilayer feed-forward neural networks and the backpropagation method Levenbeerg–Marquardt. The authors compared MLR's prediction performance to that of ANN, finding that ANN was more precise. With this goal, a new 33-year record was set.

For monitoring and prediction, [9] investigated eight wells along a side of a river in South Korea. In terms of relevance and effect, three criteria were assessed for each of the eight wells. The ground-water height, the water channel, & the ground-water heating pump are the three requirements. There was also a link established between rainfall, dam discharge, and groundwater. Rainfall was found to play a little role. The input data was trained by using back-propagation and a ANN. Several options were tested instead of using a predetermined number of layer nodes which are hidden, rate of learning, or impulse. The performance of each of the eight wells was assessed using ME, NSE and RMSE - Root Mean Square Error, Correlation to estimate the groundwater level.

The research [10] explains two well-known methods for forecasting, controlling, and projecting water supplies, as well as employing insights to try and figure out why a drier period occurred. A research performed at the Santa Barbara's Ecological Station in Brazil's So Paulo State, utilizing auto regressive integrated moving average (ARIMA) and sequentially Gauss simulations (SGS). SGS is a geostatistical method, while ARIMA is a time-series method. The Akaiike Information Criteria (AIC) was utilized to optimize the ARIIMA boundaries (Akaike 1974). The ARIMA models are suggested for groundwater monitoring, regardless of the fact that per- haps the SGS has a marginally level of accuracy. SGS, unlike ARIMA, does not include an automated optimization feature.

In research [11] employed RNN and LSTM networks to analyze and predict table of groundwater reactivity to events of storm inside the Norfolk food-prone coastal city. This inquiry looked at period data from 2010 to 2018. Constant information and storming event data were the two types of data available. After statistical analysis of both models, the authors discovered that LSTM had superior prediction accuracy for the study region than RNN. The RMSE of LSTM was 0.09 m, whereas the RMSE of RNN was 0.14 m, according to the findings. Its LSTM that showed improvement, but it required three times the time to train.

Also, for multi-step forward forecasting utilizing time-series data format from hydrological processes, [12] observed the calibre of stochastic model such as ARIMA with ML approaches such as neural network models. The majority of research on groundwater level prediction was focused on short- term forecasts. A less research work has been done on anticipating seasonal changes in groundwater levels. We suggested a novel model based on LSTM as a reaction to the achievement of data science models in a variety of hydrological modelling applications. In our methodology, all that is required is a basic data pre-processing procedure. Using the dropout approach, over-fitting is minimized significantly.

III. METHODOLOGY

The purpose of the proposed research is to determine if deep learning can assist forecast two-dimensional water floods depth. With the explicit goal of providing quick flood level forecasts for recorded, in the training dataset rain events and geographical locations are not included this model exploits meteorological and typographical data. The stepping the approach proposed are discussed beneath and presented in Fig. 1.

i. Initialization (Web scrapping): The data is collected from Kaggle [6]. The dataset consists 170 water table levels. The dataset has columns like irrigation, rainfall, temperature, evaporation, actual water table depth, etc.

ii. Preprocessing of Dataset: In this step, we have removed null and incorrect values [16, 17, 18, 19]. Seaborn library of Python is utilized for converting our dataset into data frames. This dataset is further converted into splitted dataset.

iii. Data splitting: The procedure of splitting train-test dataset is utilized for to determining the performance of ML algorithms when they are used to make evaluations on data apart from train the model. In our work, training model consists of 70% of dataset and testing model consists of 30% of dataset.

iv. Deriving Mathematical Relation:
The mathematical relations derived from our data set for LSTM model are given below:

Blocks of inputs: This phase is dedicated to modifying the blocks input element, which coupled signal of input as x (t) coupled with result of that Long short - term memory unit y of the previous entry (t1). This may be done in the following way:

$$z(t) = g(Wzx(t) + Rzy(t-1) + bz) \quad (1)$$

wherein x(t) along with y(t1) are associated with the Wz and Rz weights, correspondingly, and bz is the biases scale parameter.

Gate at the input: This step updated the gates of input, which integrates the signal of x(t) input, the LSTM result of y(t1) cell, and the value of previous cell c(t1) iteration's. The technique is depicted in the following equation:

$$i(t) = \sigma(Wix(t) + Riy(t-1) + pi \circ c(t-1) + bi) \quad (2)$$

Model Calibration: During model-calibration, the training procedure is improved by lowering the cost-function. The weights provided by model are learned using the applicable training dataset. A common fault, model overfitting, can occur at this phase, resulting in noise and a detrimental training impact. The hyper - parameters add to the model's complexity because they can't simply extract data from training dataset directly. As a consequence, all models' model hyperparameters were fine-tuned throughout the model calibration procedure.

Model-validation: RMSE as Root-Mean-square-error, NSE as efficiency of Nash Sutcliffe, RSR as RSME standard deviation ratio, MSE as Mean square error, (MAE) mean-absolute-error, determination coefficient (R2) & MAPE as mean-absolute-percentage-error are used to measure model correctness during training phase & validation phase. The model's performance may be classified as excellent (RSR 0.49 NSE>0.74), better (RSR as 0.59 along with 0.74 NSE greater than 0.64), Okayish (.59RSR 0.69 along with 0.64 NSE>0.490), or terrible (RSR greater than 0.70 along with NSE 0.50), according to Moriasiet (2007). When the MAPE number lowers, the model's accuracy improves.

Result evaluation: The uncertainty bounds were calculated using the quantile regression approach. The standard deviation varies based on the geological environment, with complicated moraine clay (formed from evaporation and temperature features of dataset taken) soils having a larger level of uncertainty. Quantile regression has the well-known flaw of requiring independent training for each quantile, which can lead to an erroneous distribution in time interval. The standard deviation reflects the deviation percentage of result of proposed model against the actual water table depth present in dataset.

Visualization: For result visualization, we have plotted our model actual results along y axis and predicted results along x-axis. The line plot denotes percentage of variation rate among actual and predicted values.

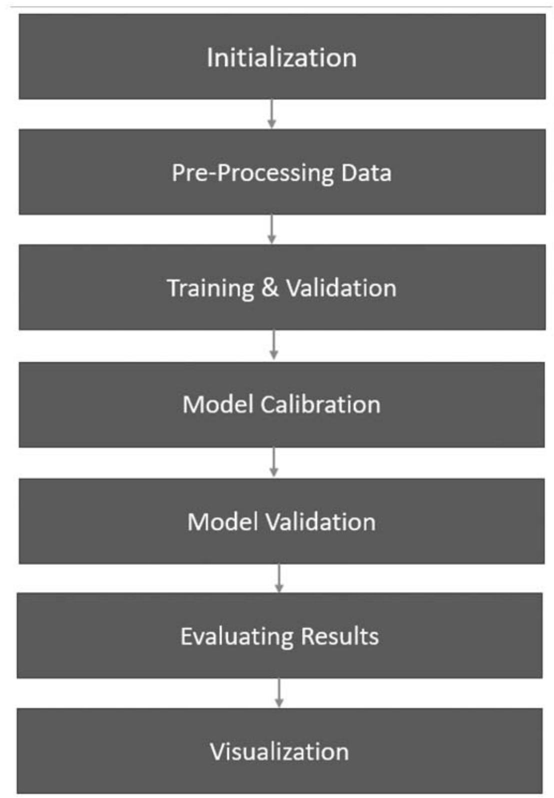

Figure 1. Proposed Methodology.

The proposed methodology is accompanied using the U-NET neural network architecture, which is commonly used for picture segmentation. Locations and rain events not included in the training are evaluated systematically. The results of the experiments suggest that the dropout strategy may greatly reduce overfitting.

IV. RESULTS AND DISCUSSION

The uncertainty bounds were calculated using the quantile regression approach. Uncertainty over intervals is 60 to 90% as depicted Fig. 2. The sigma value is 0.5q, suggesting that the depth of water table is less than half of standard variation. The standard deviation otherwise varies based on the geological environment, with complicated moraine clay soils having a larger level of uncertainty.

Quantile regression has the well-known flaw of requiring independent training for each quantile, which can lead to an erroneous distribution in time interval.

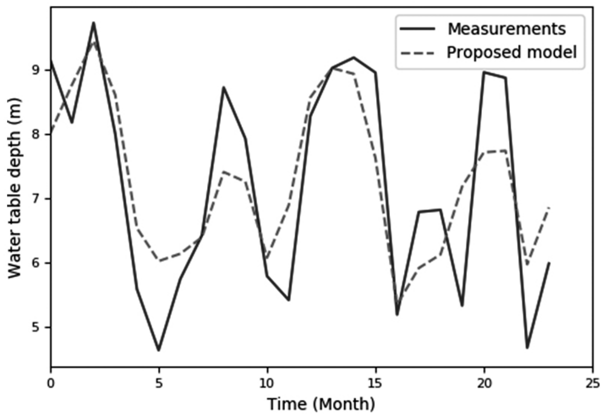

Figure 2. Result Visualization.

Figure 3 depicts the reinforcement learning of our model where our model learns from its previous iteration to improve its loss percentage.

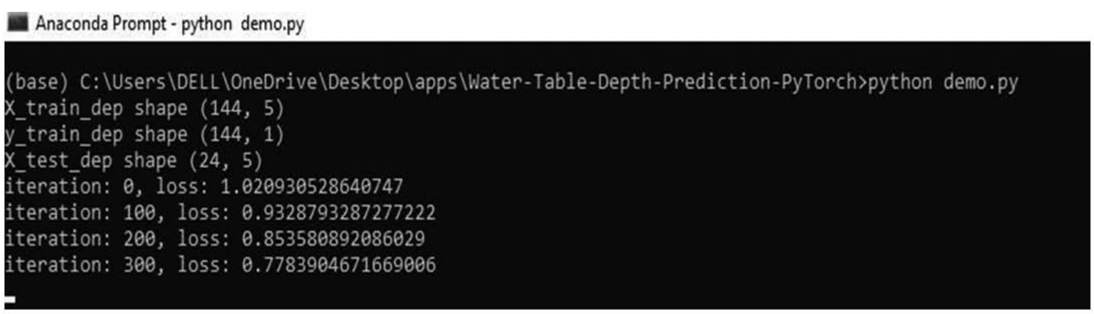

Figure 3. Reinforcement learning of model.

V. CONCLUSION

LSTM is a model of deep-learning, based on regression. LSTM excels in portraying variability in post-monsoon data. The highest NSE was found in MLP with a value of 0.980. In future observation can be done for groundwater wells for monsoon-prone Indian states which can be modelled using the proposed model. Putting the models to the test in piezometric wells and looking at groundwater level modelling in states with winter monsoons might improve the research.

REFERENCES

[1] T.M. Daniell, “Neural networks. Applications in hydrology and water resources engineering. In: *National Conference Publication,* Institute of Engineers, Australia, 1991.

[2] Z.L. Wei, D.F. Wang, H Y Sun and X Yan, “Comparison of a physical model and phenomenological model to forecast groundwater levels in a rainfall-induced deep-seated landslide”, *J. Hydrol.*586:124894, 2020. https://doi.org/10.1016/j.jhydrol.2020.124894

[3] Kenda K, Čerin M, Bogataj M, Senožetnik M, Klemen K, Pergar P, Laspidou C, Mladenić D (2018) Groundwater modeling with machine learning techniques: Ljubljana polje aquifer. *Multidiscip Digit Publ Inst Proc*, Vol. 2, no.11, p.697.

[4] Osman, A. I. A., Ahmed, A. N., Chow, M. F., Huang, Y. F., & El-Shafie, A. (2021). Extreme gradient boosting (xgboost) model to predict the groundwater levels in Selangor Malaysia. Ain Shams Engineering Journal, 12(2(2), 1545–1556. https://doi.org/10.1016/j.asej.2020.11.011

[5] S. Javadinejad, R. Dara & F. Jafary, “Modelling groundwater level fluctuation in an Indian coastal aquifer”, *Water SA*, Vol. 46, no.4, 2020, pp.665-671.

[6] H. Assem, S. Ghariba, G. Makrai, P. Johnston, L. Gill and F. Pilla, “Urban water flow and water level prediction based on deep learning. In *Joint European conference on machine learning and knowledge discovery in databases,* Sep 2017, pp. 317-329. Springer, Cham.

[7] S. Sahoo, and M.K. Jha, “Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment”, *Hydrogeology J.*, Vol. 21, no. 8, 2013, pp. 1865-1887.

[8] D. C. Ducker, “Photo-elastic way of Incidence separates stresses”, *Mechanics J.,* 1967.

[9] K.K. Lee and H. Yoon, “Using artificial neural network models for groundwater level forecasting and assessment of the relative impacts of influencing factors”, *Hydrogeology J.*, Vol. 27, no.2, 2019, pp.567-579.

[10] E.H.D.M. Takafuji, M. M. D. Rocha and R. L. Manzione, Groundwater level prediction/forecasting and assessment of uncertainty using SGS and ARIMA models: a case study in the Bauru Aquifer System (Brazil)”, *Natural Resources Research*, Vol. 28, no.2, 2019, pp. 487-503.

[11] B.D. Bowes, J.M. Sadler, M.M. Morsy, M. Behl and J.L. Goodall, “Forecasting groundwater table in a flood prone coastal city with long short-term memory and recurrent neural networks”, *Water*, Vol. 11, 5, 2019, p. 1098.

[12] G. Papacharalampous, H. Tyralis and D. Koutsoyiannis, “Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes”, *Stochastic Environmental Research and Risk Assessment*, Vol. 33, no.2, 2019, pp. 481-514.

[13] UNESCO World Water Assessment Programme, “The United Nations World Water Development Report 2022: groundwater: making the invisible visible; facts and figures”, 2022.

[14] S. Gupta, A. Singh and J. Ranjan, "Multimodal, multiview and multitasking depression detection framework endorsed with auxiliary sentiment polarity and emotion detection", *Int J. Syst Assur Eng Manag,* 2023. https://doi.org/10.1007/s13198-023-01861-z.

[15] S. Gupta, S. Bisht and S. Gupta, "Sentiment Analysis of an Online Sentiment with Text and Slang Using Lexicon Approach. In S.C. Satapathy, V. Bhateja, M.N. Favorskaya, T. Adilakshmi, (eds) *Smart Computing Techniques and Applications. Smart Innovation, Systems and Technologies*, Vol. 224, 2021. Springer, Singapore. https://doi.org/10.1007/978-981-16-1502-3_11.

[16] S. Gupta, A. Singh, J. Ranjan, "Sentiment Analysis: Usage of Text and Emoji for Expressing Sentiments. In: M. Kolhe, S. Tiwari, M. Trivedi and K. Mishra, (eds) *Advances in Data and Information Sciences. Lecture Notes in Networks and Systems*, Vol. 94, 2020, Springer, Singapore. https://doi.org/10.1007/978-981-15-0694-9_45.

[17] R. Mehrotra, O. Garg, S. Gupta and A. Singh, "Opinion Mining of Pandemic Using Machine Learning", In S. Tiwari, M.C. Trivedi, M.L. Kolhe, K. Mishra and B.K. Singh, (eds) Advances in Data and Information Sciences. Lecture Notes in Networks and Systems, Vol. 318, 2022. Springer, Singapore. https://doi.org/10.1007/978-981-16-5689-7_20.

[18] S. Gupta, O. Garg, R. Mehrotra and A. Singh, "Social Media Anatomy of Text and Emoji in Expressions", In S.C. Satapathy, V. Bhateja, M.N. Favorskaya, T. Adilakshmi (eds) *Smart Computing Techniques and Applications. Smart Innovation, Systems and Technologies*, Vol. 224, 2021. Springer, Singapore. https://doi.org/10.1007/978-981-16-1502-3_6.

[19] S. Gupta and A. Singh, "A Vocabulary-Based Framework for Sentiment Analysis", 2023, In *Computational Intelligence: Select Proceedings of InCITe 2022* (pp. 507-515). Singapore: Springer Nature Singapore.

Dr. Shelley Gupta, Ph.D. (CSE), is currently working as Assistant Professor in ABES Engineering College, Ghaziabad, India. Completed her B.Tech.. and M.Tech. degrees in Computer Science and Engineering from India in 2007 and 2012 respectively. Her research areas of interest include Natural Language Processing, Data Analytics, Data Mining, Algorithms, etc.

She has expertise in Artificial Intelligence, Pattern Recognition, Python Programming, etc. Also published numerous research papers and patents. Also serves as the reviewer of many reputed journals

Sanjay Singh Bisht is a software development engineer who loves to solve challenging problems. Obtained B.Tech. (IT) from ABES Engineering College, Ghaziabad. He is always on the lookout of new opportunities which can help him enrich his experience and learning.

Prakhar Jain he is a responsible and focused technophile, B.Tech (IT) from ABES Engineering College with knowledge of web development , Competitive Coding & Software development domain.

Passion to keep on learning various tools, emerging technologies and implementing them.

Adarsh Mishra obtained B.Tech. (IT) from ABES Engineering College, Ghaziabad. He is a Project Engineer who thrives on solving complex problems, always eager to discover new opportunities that can contribute to the enrichment of his experience and learning journey.