Prediction of groundwater fluctuations in Chengalpattu district through machine learning

Ninu Praseetha N.S[1]*, *P.Kaythry*[2], *P.Sangeetha*[3], *Devavaram* Jebaraj[2], *Santhosh* Srinivas[2], *Karan K*[3] *and Devadharshini* [3]

[1 2] Department of ECE, Sri SivaSubramaniya Nadar College of Engineering, Chennai,603 110, TamilNadu, India

[3] Department of Civil Engineering, Sri SivaSubramaniya Nadar College of Engineering, Chennai,603 110, TamilNadu, India,

Abstract: Groundwater, found beneath the Earth's surface in saturated zones of soil, sediment, and rock, plays a crucial role in sustaining ecosystems and supporting human activities like agriculture and industry. Monitoring and managing groundwater resources are crucial for sustainable use. As of the latest update in January 2022, Chennai, a city in southern India, has been grappling with water scarcity issues. The city has faced recurrent water shortages due to various factors, including rapid urbanization, inadequate infrastructure, depleting groundwater levels, and irregular rainfall patterns. Chengalpattu district in Tamil Nadu, India, is known for its diverse geographical features, incorporating urban and rural landscapes, and is significant for agriculture and water resource management. This study focuses on predicting variations in groundwater levels in open wells at different locations in the Chengalpattu district, assessing the effectiveness of various machine-learning models. This paper utilizes the ARIMA model provided by the stats models library. This library is widely employed for statistical modelling and hypothesis testing in Python, and encompasses a range of tools for time series analysis, including the ARIMA model. In this context. ARIMA models are employed for predicting future depth which focus on depicting autocorrelations in the data. Additionally, a package consolidating various models, including Seasonal Naïve (a straightforward forecasting method for seasonal data that serves as a reliable benchmark by relying on the observation from the same period a season ago), was incorporated in this study.

Keywords – Groundwater level, forecasting, fluctuations, Predictions, Machine Learning.

*Corresponding Author : ninupraseetha2350520@ssn.edu.in

1.Introduction

The monitoring of groundwater levels is essential for acquiring data regarding the depth of the water column beneath the Earth's surface. Continuous groundwater monitoring yields detailed data, facilitating observations for ongoing development and future projections [1-4]. Piezometers, commonly known as level sensors, play a crucial role in gathering comprehensive information from beneath the Earth's surface, specifically tracking the water table's starting point. To assess the water content in sand pores and aquifers, piezometers can be strategically installed in wells, bore wells, and tube wells at desired depths [5][6]. The integration of electromagnetic flow meters and telemetry systems with the level sensor, connected via appropriate wiring to the surface level, enables real-time transmission of data to a server. This data is then accessible in real-time through data management software, regardless of location [6][7].

The primary objectives of this work is to identify the geographical positions of open wells within the Chengalpattu district, estimate and monitor groundwater levels in selected open wells over six months, and conduct a thorough analysis of the observed variations. Additionally, the study aims to employ diverse machine-learning techniques to forecast and model fluctuations in groundwater levels, thereby enhancing our understanding and ability to predict changes in this vital natural resource [7].

Some of the related works have been discussed below.

Dilip Kumar Roy., et al. [1], in their study, assess data-driven models utilizing various machine learning algorithms to predict fluctuations in groundwater levels for one, two, and three weeks ahead in Bangladesh's Godagari upazila. Used Minimum Redundancy Maximum Relevance (MRMR) algorithm. Despite a slight decline in performance with an increasing forecast horizon, evaluation indices affirm the superior performance of the Bayesian model averaging (BMA) ensemble. BMA-based heterogeneous ensemble emerges as a promising strategy to enhance multi-step ahead Ground water level (GWL) forecasts, both within the studied area and beyond.

Junaid Khan.,et.al [2] comprehensively studied commonly employed traditional numerical methods, machine learning approaches, and deep learning models are commonly employed for the prediction of various outcomes showcasing notable advancements in prediction efficiency over the past two decades. A meticulous analysis of 109 research articles published from 2008 to 2022, investigating various modeling techniques, is undertaken. Concluded that machine learning and deep learning approaches are efficient for modeling GWL and recommended possible future research to enhance the accuracy of GWL prediction.

Ahmedbahaaaldin Ibrahem Ahmed Osman., et .al [3**]** studies have successfully modelled and predicted GWL in diverse regions using methods proposed by the authors. Reviewed in 40 studies in GWL forecasting using artificial intelligence (AI). The incorporation of a trial-and-error approach during certain phases of AI modelling has proven beneficial for testing in specific applications related to GWL modeling. Analyzing modeling methods used in all the reviewed studies it was estimated that the machine learning methods are efficient enough for modeling GWL.

Khabat Star Mohammed.,et.al.,[4] illustrate that the groundwater level fluctuations constitute a crucial aspect of the hydrogeological cycle and are essential variables in numerous water resources operation models. GA-ANN and ICA-ANN hybrid methods and the ELM and ORELM models are utilized. The output of the ORELM model has the best fit with observed data with a correlation coefficient equal to 0.96, and it also has the best and closest scatter points around the 45-degree line, and in this sense, it is considered the most accurate model.

Cristina Di Salvo.,[5] provides a comprehensive review that is focused specifically on the integration of numerical and machine learning methods for groundwater level modeling. Numerical and machine learning models can be successfully used as complementary to each other as a powerful groundwater management tool. Sudhakar Singha.,[6] suggested that for ensuring future safe drinking water sources necessitates understanding current groundwater quality and pollution levels. accurate prediction of water quality is crucial for controlling pollution and improving water management. The study proposes a deep learning (DL) model for predicting groundwater quality, comparing its performance with three machine learning (ML) models: random forest (RF), eXtreme gradient boosting (XGBoost), and artificial neural network (ANN). The DL model's reliability was confirmed by running the algorithm with randomized datasets ten times, showing minor deviations in performance metrics. Additionally, input variable importance analysis affirmed the DL model as the most realistic and accurate for predicting groundwater quality.

Arman Ahmadi.,et.al[7] conducted a thorough examination of the use of machine learning (ML) techniques in the groundwater level. The research topic and search technique are organized according to the Population, Intervention, Comparator, and Outcome (PICO) framework, which is used in a systematic review methodology. Through an extensive dataset analysis of peer-reviewed articles and conference proceedings, the report demonstrates the efficacy of machine learning algorithms in accurately predicting groundwater levels. The meta-analysis highlights the value of ML in groundwater resource management by demonstrating the ability of data-driven models to reliably estimate a variety of groundwater parameters. The results highlight the value of machine learning techniques in improving groundwater modeling and prediction, especially in dry and semi-arid areas where surface water resources are scarce.

Stephen Afrifa [8]., et.al studied the effects of climate change, such as rising temperatures, heavier precipitation, and an increase in the frequency of extreme weather events, calling for new methods of researching fluctuations in groundwater levels. These methods are necessary for stakeholders and researchers to make well-informed decisions. Although mathematical models and machine learning can forecast changes in groundwater levels, there are currently no thorough explanations of these techniques. By employing the PRISMA reporting methodology and conducting a thorough assessment of 117 papers from the Scopus database, this study fills this gap. According to the review, machine learning has become more popular and delivers higher accuracy rates than traditional mathematical models. In particular, random forest (RF), support vector machine (SVM), and artificial neural network (ANN) methods have gained appeal. Ensembles for machine learning also increase computing efficiency. The study recommends that in addition to mathematical methods, future research should use sophisticated machine-learning approaches.

2. Methodology

In any predictive project, pivotal stages involve gathering data from diverse locations within the district, reflecting varying climate conditions, soil types, rainfall rates, and more. This diversity contributes significantly to enhancing the predictive capabilities of the Deep Learning Model. Figure 1 illustrates the various steps involved in this work comprehensively.

To understand and predict groundwater behavior in the Chengalpattu district by first identifying and mapping the geographical positions of all open wells and gathering their location data. Chengalpattu district is a district in the state of Tamil Nadu, India. Located in the southern part of the state, it covers an area of approximately 2,945 square kilometers. The district is named after its headquarters, Chengalpattu, which is a significant town with historical and cultural importance. The district is also recognized for its agricultural activities, with farming being a primary occupation for many residents. Over the past decade, the growing Chengalpattu town and its residents, located on the outskirts of Chennai, have been grappling with an uneven struggle against water scarcity. Water scarcity has reared its ugly head again in this rapidly urbanizing town located off GST Road.

2.1 Procedures Followed

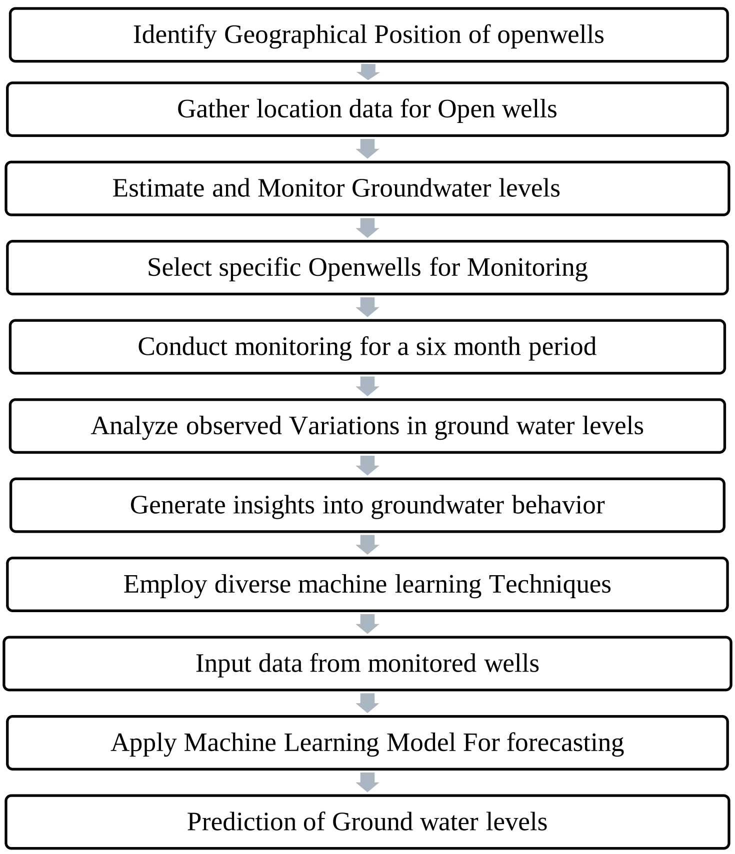

Fig.1. Procedures Followed

Specific wells are selected for detailed monitoring over six months to estimate and observe variations in groundwater levels. Authorization was granted to collect readings from 15 wells, including major and minor irrigation tanks, as well as private wells. Project members conducted readings every fortnight (15 days), commencing from September 2022 to April 2023. The collected data is analyzed to generate insights into groundwater behavior, considering factors such as seasonal changes and extraction rates. Figure 2 illustrates the various locations in the Chengalpattu district where the data are collected. The locations where readings were taken are enumerated below.

Thandalam, Rajiv Gandhi IT Expy, Karunguzipallam, Sirudavoor, Acharavakkam, Alapakkam, Thenur, Kattankolathur, Guduvancheri, Rathinamangalam, Kolapakkam, Kandigai, Kolathur-I, Kolathur -II, Senganmal.

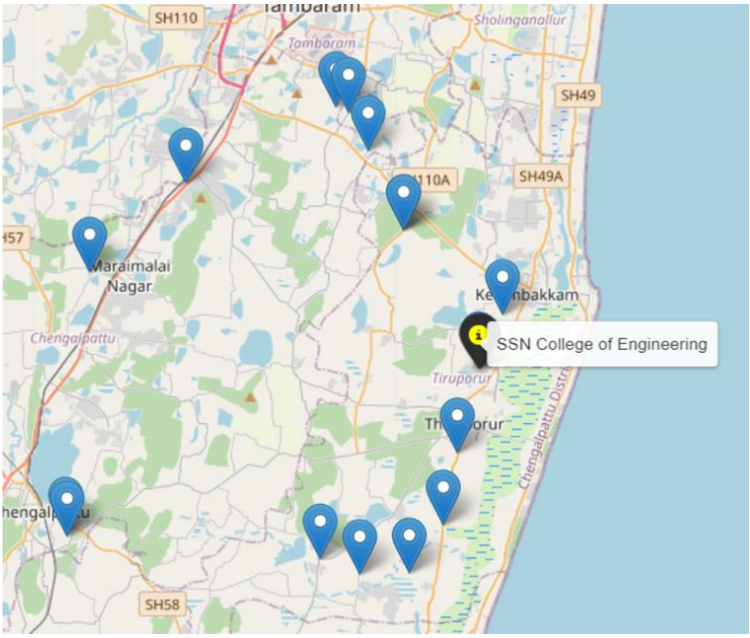

Fig.2. Various locations from which data is to be collected.

Utilizing a groundwater level meter, the water levels were measured from various pumps and wells. In addition to the aforementioned locations, readings were also obtained every fortnight from several wells situated within our college campus. Acquiring them early is essential to allow sufficient time for studying and designing our model, taking into account the variations in the collected data. After collecting the data, it is crucial to conduct timely verification to ensure its correct format and suitability for the specified date and time. Subsequently, this verified data can be utilized for the development of the time series forecasting model. In the context of monitoring rainfall and water levels, obtaining precise readings is imperative, as any inaccuracy may result in suboptimal outcomes. Figure 3 depict the openwells in Rajiv Gandhi IT Expressway and Karunguzhipallam and Figure 4 illustrates the working of ground water level sensor.

A groundwater level sensor is a device that detects and measures the depth of groundwater in wells or boreholes. It typically consists of a probe or sensor that is lowered into the well, and it detects the presence of water based on changes in pressure, conductivity, or other physical properties. Groundwater level sensors often provide continuous monitoring and may be connected to data logging systems for real-time data collection.

Diverse machine learning techniques are employed to handle complex patterns in the data, with the input data from monitored wells used to train and test these models. The models are then applied to forecast future groundwater levels, providing valuable predictions that can inform and enhance water resource management practices, ensuring sustainable usage and conservation of this vital natural resource.

Fig. 3. Certain wells were used for the timely recording of data.

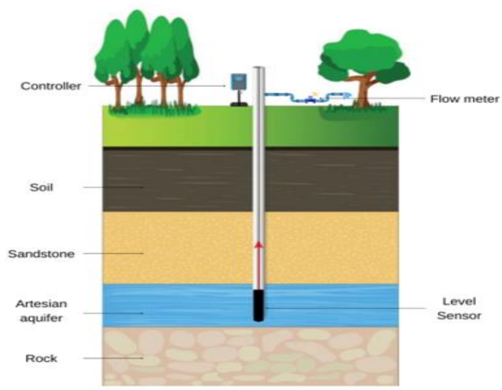

Fig.4. Working of a Ground Water Level Sensor

We utilized the ARIMA model, which encompasses a diverse range of Time Series models. By automating machine learning operations, the ARIMA model simplifies the development of applications with robust prediction performance. It enables the training and deployment of highly accurate machine learning and deep learning models for various data types, including images, text, time series, and tabular data, with just a few lines of code. For training, a 2-hour session was allocated. ARIMA (utilizing Autoregressive Integrated Moving Average) provides an alternative approach to time series forecasting, focusing on autocorrelations in the data.

3. Results

A classical time series analysis model was used to predict groundwater levels using data acquired through a groundwater level meter. The dataset showcases fluctuations across different regions, timeframes, and diverse climatic conditions. A segment of the collected data is presented below in Table 1.

Checking the efficacy of machine learning models refers to assessing how well a model performs in solving a particular task or making predictions on unseen data. It involves evaluating various aspects such as score value, score test and other relevant metrics depending on the nature of the problem. These metrics and parameters are essential for

Table 1. Small sample of the data collected.

Location/Date	11-09-2022	22-09-2022	15-10-2022	26-10-2022	05-11-2022	24-11-2022	10-12-2022	27-12-2022	13-01-2023	31-01-2023	09-02-2023	22-02-2023	13-03-2023	28-03-2023	05-04-2023
Thandalam	1.8	1.4	1.6	1.4	1.5	1.4	1.4	1.3	1.5	1.55	1.2	1.2	1.25	1.1	1.1
Rajiv Gandhi IT Expy	3.2	2.9	3.1	3.1	2.7	2.5	2.2	3.1	3.4	3.3	3.4	2.9	2.7	2.6	2.6
Karunguzhipallam	2.7	2.2	2.3	2.6	2.1	2	2.4	2.2	2.3	2.6	2.1	2.3	2.4	2.45	2.4
Sirudavoor	2.9	2.5	2.6	3.3	3	3.1	3.4	3.2	3.4	3.4	3.5	3	3.1	2.6	2.4
Acharavakkam	1.3	1.6	1.4	0.8	0.5	0.4	0.9	0.9	0.95	1.2	1.2	1.35	1.4	1	0.8
Alapakkam	1	0.7	0.5	0.6	0.2	0.3	0.4	0.5	0.4	0.5	0.4	0.6	0.4	0.3	0.3
Thenur	4.8	5.2	5	4.6	4.1	4.5	4.3	4	4	4.2	4.35	4.1	4.2	4.1	4
Kattankolathur	4.4	4.4	4.8	4.1	3.6	3.9	4	4.3	4.1	4.1	4	3.9	3.5	3.2	3.5
Guduvancheri	1.65	1.75	2.05	1.85	1.55	1.7	1.6	1.5	1.3	1.5	1.6	1.4	1.3	1.8	1.6
Rathinamangalam	1.7	1.9	2.2	1.7	1.2	1.6	1.4	1.5	1.3	1.6	1.5	1.4	1.3	1.2	1.7
Kolapakkam	1.2	0.9	0.7	1.2	0.9	0.9	0.6	0.9	0.5	0.7	1.2	1.2	1.4	1.3	1.1
Kandigai	0.4	0.1	0.2	0.1	0.6	1.2	1	1.3	1.4	1.2	1.4	1.65	1.25	1.2	1.4
Kolathur-I	0.3	0.6	0.4	0	0.6	0.3	0.4	0.55	0.85	0.4	0.3	0.4	0.55	0.4	0.6
Kolathur-II	1.5	1.4	1.8	1	0.7	0.4	0.6	0.4	0.5	0.4	0.6	0.7	1	1.3	1.2
Senganmal	1.35	1.45	1.55	0.95	0.45	0.95	0.55	0.6	0.85	1	1.35	1.2	1.4	1.5	1.2

evaluating both the performance and efficiency of machine learning models, providing insights into how well they solve the given problem and how computationally expensive they are to train and make predictions. In Table 2, the parameters of different models obtained from the collected data are presented.

By evaluating various metrics by using various models ARIMA model was chosen because of its simple interpretation, flexibility and accuracy. While exponential smoothing models are based on a description of the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data. Seasonal Naïve in the case of seasonal data, there is a simple forecasting method that can be considered as a good benchmark in many situations. Similar to Naïve, Seasonal Naïve relies only on one observation, but instead of taking the most recent value, it uses the value from the same period a season ago. All these models are conveniently bundled within a single package for ease of use. During this timeframe and plot, predictions have been made for some samples from Thandalam and Rajiv Gandhi IT Expy. The identical approach can be utilized for the remaining regions as well.

Table 2. Parameters of various models obtained using the data collected.

	model	score_test	score_val	pred_time_test	pred_time_val	fit_time_marginal	fit_order
0	WeightedEnsemble	-0.102377	-0.058583	0.236483	102.471823	3.986929	13
1	ARIMA	-0.105073	-0.061705	0.681781	0.598753	0.001002	6
2	SimpleFeedForward\T1	-0.107704	-0.076466	0.171731	0.127995	52.954464	12
3	AutoETS	-0.109718	-0.062338	62.129611	60.631308	0.001002	7
4	TemporalFusionTransformer	-0.111110	-0.074443	0.181594	0.072997	58.173907	11
5	ETS	-0.125063	-0.068257	0.166349	0.163598	0.003000	3
6	SeasonalNaive	-0.131607	-0.095140	0.033231	0.053071	0.000982	2
7	Naive	-0.131607	-0.095140	8.470385	8.398296	0.001002	1
8	DynamicOptimizedTheta	-0.132089	-0.099735	64.797131	63.646260	0.001000	8
9	Theta\T2	-0.135328	-0.103295	0.764066	0.794916	0.000000	5
10	Theta\T1	-0.135328	-0.103295	0.798166	0.776831	0.000000	4
11	AutoARIMA	-0.141788	-0.105893	44.336594	40.950169	0.001999	10
12	DeepAR\T1	-0.174232	-0.144771	0.455149	0.384281	55.522120	9

3.1 Predicted Depth for Thandalam

The Predicted depth for Thandalam using the ARIMA model is visualized in Figure 4 (a) and (b). With this prediction, the future water levels of corresponding wells can be analysed. The data are given till April 2023 and the water level is predicted till November 2023. The predicted depth will be accurate if there are several data points.

	Date	Predicted_Depth
0	2023-04-27	1.097274
1	2023-05-12	1.108397
2	2023-05-27	1.058388
3	2023-06-11	1.079549
4	2023-06-26	1.070954
5	2023-07-11	1.075513
6	2023-07-26	1.059574
7	2023-08-10	1.072320
8	2023-08-25	1.064734
9	2023-09-09	1.068468
10	2023-09-24	1.062744
11	2023-10-09	1.068643
12	2023-10-24	1.064089
13	2023-11-08	1.066822
14	2023-11-23	1.064257

Fig 4. a). Predicted data from Thandalam

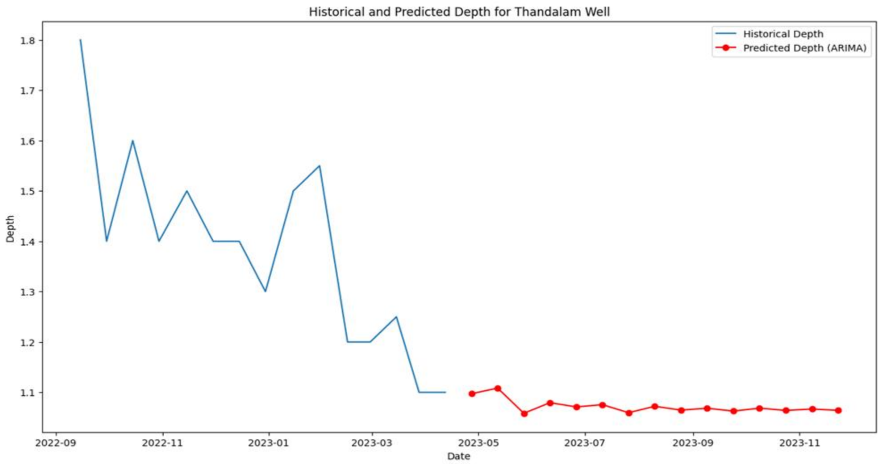

Fig 4. b)). Predicted plot from Thandalam

3.2 Predicted Depth for Rajiv Gandhi IT Expy

Figure 5 displays the predicted depth for Rajiv Gandhi IT Expy using the ARIMA model. This forecast allows for the analysis of future water levels in the corresponding wells. Here the water level is predicted till April 2024. Increasing the amount of data available would enhance the accuracy of the prediction.

	Date	Predicted_Depth
0	2023-04-12	2.696406
1	2023-04-27	2.974389
2	2023-05-12	3.030492
3	2023-05-27	3.001506
4	2023-06-11	2.891322
5	2023-06-26	2.771067
6	2023-07-11	2.788031
7	2023-07-26	2.842812
8	2023-08-10	2.909108
9	2023-08-25	2.942292
10	2023-09-09	2.906339
11	2023-09-24	2.866409
12	2023-10-09	2.840368
13	2023-10-24	2.843311
14	2023-11-08	2.872645
15	2023-11-23	2.891931
16	2023-12-08	2.895612
17	2023-12-23	2.884103
18	2024-01-07	2.867377
19	2024-01-22	2.862063
20	2024-02-06	2.866436
21	2024-02-21	2.875649
22	2024-03-07	2.882469
23	2024-03-22	2.881420
24	2024-04-06	2.876408

Fig. 5. a)). Predicted data from Rajiv Gandhi IT Expy

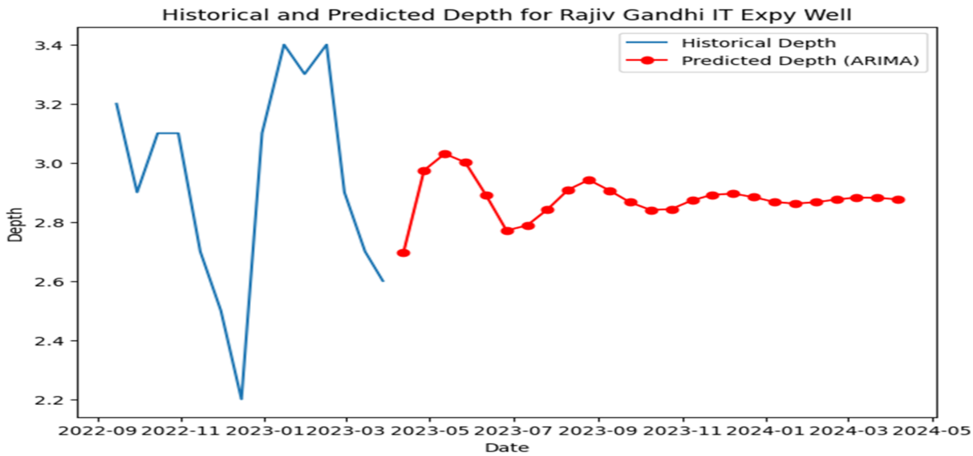

Fig 5. b)). Predicted plot from Rajiv Gandhi IT Expy

4. Conclusion and Future Work

With increasing global temperatures due to climate change, there is a growing concern about water shortages, which could be exacerbated by depletion in the ozone layer. Therefore, this project could be instrumental in estimating water shortages, allowing for better planning for future water needs. To improve the accuracy of the model, more data would be required, which could help in addressing the major water crisis that is expected to occur soon. This could have a significant impact on communities that are already struggling with water scarcity, particularly in rural areas where access to clean water is limited. By providing more accurate forecasts, communities can better prepare for water shortages, ensuring that they have access to clean water during times of drought or other water-related crises. As the project aims to develop a groundwater level forecasting model using machine learning, it has the potential to be applied to various locations in India where water scarcity is a significant issue. Overall, this project has significant potential to contribute to addressing one of the most pressing environmental challenges facing our world today.

. References

1. Dilip Kumar Roy, Tasnia Hossain Munmun, Chitra Rani Pau, Mohamed Panjarul Haque,Nadhir Al-Ansari and MohamedA.Mattar, "Improving Forecasting Accuracy of Multi-Scale Groundwater Level Fluctuations Using a Heterogeneous Ensemble of Machine Learning Algorithms" Water 2023, 15, 3624. https://doi.org/10.3390/w15203624
2. Junaid Khan ,EunkyuLee, Awatef Salem Balobaid and Kyungsup Kim, " AComprehensive Review of Conventional, Machine Leaning, and Deep Learning Models for Groundwater Level (GWL) Forecasting" Sci. 2023, 13, 2743. https://doi.org/10.3390/app13042743
3. Ahmedbahaaaldin Ibrahem Ahmed Osman,Ali Najah Ahmed,Yuk Feng Huang.Pavitra Kumar,Ahmed H. Birima ,Mohsen Sherif, Ahmed Sefelnasr, Abdel Azim Ebraheemand, Ahmed El-ShafiePast, "Present and Perspective Methodology for Groundwater Modeling-Based Machine Learning Approaches" International Center for Numerical Methods in Engineering (CIMNE) 2022
4. Khabat Star Mohammed,Saeid Shabanlou, Ahmad Rajabi,Fariborz Yosefvand, Mohammad Ali Izadbakhsh, "Prediction of groundwater level fluctuations using artificial

intelligence-based models and GMS" Applied Water Science (2023) 13:54 https://doi.org/10.1007/s13201-022-01861-7

5. Cristina Di Salvo "Improving Results of Existing Groundwater Numerical Models Using Machine Learning Techniques: A Review" Water 2022, 14, 2307. https:// doi.org/10.3390/w14152307
6. Sudhakar Singha, Srinivas Pasupuleti, Soumya S. Singha , Rambabu Singh , Suresh Kumar" Prediction of groundwater quality using efficient machine learningTechnique" Chemosphere 276 (2021) 130265
7. Arman Ahmadi ,Mohammadali Olyaei,Arash Ghomlaghi , Andre Daccache , Zahra Heydari Mohammad Emami ,Graham E. Fogg ,Amin Zeynolabedin , and MojtabaSadegh "Groundwater Level Modeling with Machine Learning: A Systematic Review and Meta-Analysis" Water2022, 14, 949. https://doi.org/ 10.3390/w14060949
8. Stephen Afrifa, Tao Zhang , Peter Appiahene ,Vijayakumar Varadarajan "Mathematical and Machine Learning Models for Groundwater Level Changes: A Systematic Review and Bibliographic Analysis, Future Internet 2022, 14, 259. https:// doi.org/10.3390/fi14090259
9. Fan Feng, Hamzeh Ghorbani, Ahmed E. Radwan,"Predicting groundwater levels using traditional and deep machine learning algorithms", Big Data, AI, and the Environment Volume 12 - 2024 | https://doi.org/10.3389/fenvs.2024.1291327
10. Ebenezer K, Siabia,b, Yihun Taddele Dilec, Amos T. Kabo-Bahb,d, Mark Amo- Boatenga, Geophery K. Anornue, Komlavi Akpotif, Christopher Vuue, Peter Donkorg, Samuel K. Mensahg, Awo B. M. Incoomh, Emmanuel K. Opokue, and Thomas Atta-Darkwa," Machine learning based groundwater prediction in a data-scarce basin of Ghana", APPLIED Artificial Intelligence 2022, VOL. 36, NO. 1, e2138130 (30 pages) https://doi.org/10.1080/08839514.2022.2138130